How Technologies Will Change the Way Finance Departments Work – A Target Picture and Guidelines for Digital Finance

Von der Mercator School of Management, Fakultät für Betriebswirtschaftslehre, der

Universität Duisburg-Essen

zur Erlangung des akademischen Grades

eines Doktors der Wirtschaftswissenschaft (Dr. rer. oec.)

genehmigte Dissertation

von

Markus Andreas Eßwein

aus

Mannheim

Referent: Prof. Dr. Peter Chamoni

Korreferent: Prof. Dr. Jochen Gönsch

Tag der mündlichen Prüfung: 25.10.2019

Publications associated with this thesis

This thesis is based on and extends research that was published in different conference proceedings:

Esswein, M. and Chamoni, P. (2018). Business Analytics in the Finance Department – A Literature Review. Proceedings of the Multikonferenz Wirtschaftsinformatik, Lüneburg, Germany.

Mayer, J. H., Esswein, M., Razaqi, T., and Quick, R. (2018). Zero-Quartile Benchmarking – A Forward-Looking Prioritization of Digital Technologies for a Company's Transformation. Proceedings of the International Conference on Information Systems, San Francisco, USA.

Esswein, M., Mayer, J.H., Stoffel, S., and Quick, R. (2019). Predictive Analytics – A Modern Crystal Ball? Answers from a Cash Flow Case Study. Proceedings of the European Conference on Information Systems, Stockholm, Sweden

Submitted: Esswein, M., Mayer, J.H., Sedneva, D., and Albers, J.-P. (2019). Improving the Invoice Allocation in Accounting – An Account Recommender Case Study. Conference of the Italian Chapter of the Association for Information Systems

Submitted: Esswein, M., Reinersmann, M., and Chamoni, P. (2019). Machine Learning and Analytics Adoption – Why are the number-crunching accountants lagging behind? Hawaii International Conference on System Sciences

Submitted: Esswein, M., Wienand, M., and Mayer, J.H. (2019). The Value of Digital Enterprise Platforms for Performance Management – Assessing Efficiency, Effectiveness, and Experience. Hawaii International Conference on System Sciences

Additionally, research on other areas of digital finance was published or submitted in-parallel but is not incorporated in this thesis:

Krönke, B., Mayer, J. H., Quick, R., and Esswein, M. (2017). Manager App Portal at Continental: A Content, Collaboration, and Convenience App Selection that Works. *Controlling*, 29(5), 36-45.

Submitted: Mayer, J. H., Esswein, M., Hornung, K., and Steigner, M. (2019) The MENSCH in the Digital Age – Are next-generation managers triggering basic changes? *Controlling*, 31(tbd), tbd.

Submitted: Mayer, J.H., Stritzel, O., Quick, R., and Esswein, M. (2019). Towards Automation in Accounting: Results from an IFRS16 Leasing Case Study. Americas' Conference on Information Systems, Cancún, Mexico.

Table of contents

List of figures V

List of tables VII

Abbreviations VIII

PART I: QUO VADIS DIGITAL FINANCE 1

Introduction 1

1. The zero quartile is the new target state for finance organizations 5

1.1. Fundamentals: The four finance core processes are reshaped by digital technologies 5

1.1.1. Finance processes 5

1.1.2. Digital technologies 9

1.1.3. Benchmarking 14

1.2. Literature review: Finance technology benchmarking is mostly backward-looking 15

1.2.1. Search strategy 16

1.2.2. Overview of results 17

1.2.3. Framework for classification – Taxonomy of benchmarking 18

1.2.4. Results 20

1.3. Method: The zero quartile is constructed based on the best possible target state 25

1.3.1. Questionnaire 25

1.3.2. Rasch algorithm 27

1.4. Results: Four imperatives help to prioritize technologies and reach the zero quartile 29

1.4.1. Descriptive results 29

1.4.2. Synthesis of the results 33

2. Evaluating digitalization efforts goes beyond monetary benefits 40

2.1. Excursus: Many digitalization projects lack a traditional business case 40

2.2. Evaluation: Efficiency, effectiveness, and experience help to assess a value case for the zero quartile 41

2.3. Discussion: The zero quartile of digital finance is changing over time. 44

PART II: MACHINE LEARNING AS A CORNER STONE OF DIGITAL FINANCE 46

Introduction 46

3. Machine learning and analytics currently see a surge of interest from practice 47

3.1. Literature review: Machine learning and analytics adoption is still rather poor and research for finance is scarce 47

3.1.1. Search strategy 47

3.1.2. Framework for systemization.. 49

3.1.3. Gap analysis.. 50

3.2. Method: Exploring the drivers of machine learning adoption helps decision makers 54

3.2.1. Partial least squares structural equation modeling 54

3.2.2. Survey on machine learning and analytics adoption 56

3.2.3. Structural and measurement model 59

3.3. Results: Task characteristics are the most important driver of adoption 62

3.3.1. Descriptive statistics... 62

3.3.2. Model parameters 69

3.3.3. Synthesis... 76

4. Use case fundamentals: There is a range of algorithms for different problem types 78

4.1. Linear regression and LASSO 79

4.2. Exponential smoothing, Holt-Winters, and ARIMA 80

4.3. K-nearest neighbors 83

4.4. Gradient boosting 84

4.5. Neural networks 85

4.6. Auxiliary methods 88

4.6.1. Imputation 88

4.6.2. Cross-validation 89

4.6.3. Bootstrapping 89

5. Use case 1: Financial accounting can greatly benefit from an account recommender 90

5.1. Method: Machine learning is used to mitigate bottlenecks in daily business 91

5.1.1. Case study in a global chemical company 91

5.1.2. Requirements engineering 93

5.2. Results: Processing invoices with missing data becomes easier and faster 95

5.2.1. Accuracy vs. coverage trade-off 95

5.2.2. Guidelines 97

5.3. Evaluation: Amortization is less than two years and further applications are easy to identify 99

6. Use case 2: Cash flow forecasting is one of the big levers in management accounting 101

6.1. Method: Machine learning and analytics are applied to improve forecast accuracy 102

6.1.1. Case study in a global energy utility company 102

6.1.2. Cross-industry standard process for data-mining 103

6.2. Results: Machine-based forecasts outperform humans most of the time 108

6.3. Evaluation: Side-by-side use is currently most likely..................... 112

Discussion... 117

Conclusion 122

References.. 125

Appendix A: Mathematical description of the Rasch algorithm 145

Appendix B: Questionnaire for machine learning and analytics adoption....... 147

Appendix C: List of R packages used 154

Appendix D: Exemplary R code for PLS-SEM... 156

Appendix E: Exemplary R code for use case 2 157

Acknowledgements 162

List of figures

Figure 1. Overview of the topics addressed 4

Figure 2. Finance processes 7

Figure 3. Overview of automation categories 10

Figure 4. Varieties of analytics 13

Figure 5. Digital enterprise platform overview 14

Figure 6. Search results on a timeline with three phases 18

Figure 7. Taxonomy of benchmarking 20

Figure 8. Research approach and research output 22

Figure 9. Dyads based on the taxonomy for benchmarking 24

Figure 10. Triads combining function and object of the benchmarking taxonomy with the three most frequent elements of contribution type 24

Figure 11. Overview of the zero-quartile idea.. 25

Figure 12. Benefit circle for digital technologies 43

Figure 13. Citation pearl growing search terms 48

Figure 14. Classification for financial accounting... 50

Figure 15. Classification for management accounting.................................... 52

Figure 16. Technology acceptance model... 57

Figure 17. Full structural and measurement model, all paths are formative 60

Figure 18. Model constructs, latent variables, and measurement 61

Figure 19. Results for current use, sub-functions, use cases, and tasks......... 63

Figure 20. Results for data and algorithms used, decision-steps and rationale 65

Figure 21. Results for personal opinion and familiarity.................................... 67

Figure 22. Results for driver and tool support for analyses 68

Figure 23. Sub-model (1) coefficients and loadings 70

Figure 24: Sub-model (2) coefficients and loadings 71

Figure 25. Sub-model (3) path coefficients.. 72

Figure 26. Sub-model (4) path coefficients.. 75

Figure 27. Moderating effect of bottom-up and top-down driver....................... 75

Figure 28. Final model, significant paths, and total effects 77

Figure 29. Basic structure of an artificial neural network 86

Figure 30. P2P process of the reference company 93

Figure 31. Invoice processing for invoices with missing data 95

Figure 32. Accuracy vs number of neighbors 96

Figure 33. Accuracy and coverage of three prediction models........................ 97

Figure 34. Private customer installment, procurement, and cash flows......... 104

Figure 35. Operating cash flow over the fiscal years.................................... 104

Figure 36. Benchmark for cash flow forecast accuracy 108

Figure 37. Year-end operating cash flow prediction using ARIMA and ARMAX 109

Figure 38. Prediction of decomposed year-end operating cash flow using a method mix.. 111

Figure 39. Model fitting process 112

List of tables

Table 1. Search terms 17

Table 2. Zero-quartile questionnaire sample characteristics............................ 27

Table 3. Re-coding prior to Rasch algorithm.. 28

Table 4. Results of the Rasch algorithm and cluster analysis.......................... 31

Table 5. Benefit circle focus group characteristics.. 42

Table 6. Machine learning and analytics adoption sample characteristics....... 58

Table 7. Sub-model (3) loadings.. 73

Table 8. Sub-model (4) loadings.. 74

Table 9. Total effects on actual use in sub-model (4) 77

Table 10. Relevant invoice fields (independent variables)............................... 94

Table 11. Excerpt of the input factors 106

Table 12. Excerpt of the correlation matrix 106

Table 13. Overview of prediction accuracy 110

Abbreviations

General

AI	Artificial intelligence
ANN	Artificial neural network
AP	Accounts payable
AR	Accounts receivable
ARIMA	Autoregressive integrated moving average
ARMAX	Autoregressive integrated moving average with external indicators
BI	Business intelligence
CAPEX	Capital expenditures
CB-SEM	Covariance-based structural equation modeling
CV	Cross-validation
DEP	Digital enterprise platform
DL	Deep learning
DSR	Design science research
ELM	Extreme learning machine
EPM	Enterprise performance management
ERP	Enterprise resource planning
FTE	Full-time equivalent
GFT	Global finance transformation
GL	General ledger
HW	Holt-Winters
IS	Information systems
IT	Information technology
KNN	k-nearest neighbors (algorithm)
KPI	Key performance indicator
LASSO	Least absolute shrinkage and selection operator
MAR	Missing at random
MCAR	Missing completely at random
ML	Machine learning
ML&A	Machine learning and analytics
MLP	Multi-layer perceptron
MNAR	Missing not at random
NA	Not available
NLP	Natural language processing
O2C	Order-to-cash
OPEX	Operational expenditures

P2P	Purchase-to-pay
PLS-SEM	Partial least squares structural equation modeling
R2R	Record-to-report
RPA	Robotic process automation
RQ	Research question
TAM	Technology acceptance model
XGbar	Extreme gradient boosting time series forecasting

Specific to use case two

BCI	Business confidence index
CCI	Consumer confidence indicator
CF	Operating cash flow
CLI	Composite leading indicator
CNS	Average consumption
CUST	Total number of customers
ELE	Spot price electricity
FORD	Accounts receivables balance
GAS	Spot price gas
PPI	Producer price index
PRE	Precipitation
SUN	Sunshine duration
TMP	Temperature
WADTB.ELE	Weighted average days to bill for electricity
WADTM.ELE	Weighted average days to meter for electricity

PART I: QUO VADIS DIGITAL FINANCE

Introduction

Digitalization is on everyone's lips. It is no longer a question of "what for" or "why" but of "how". Unprecedented technology-driven changes are predicted to create opportunities and threats in all societies and professions over the course of the next decade (Roos, 2015). Moreover, there is an ongoing shift from business models built primarily around fixed (tangible) assets to intellectual capital (Guthrie et al., 2012) and companies increasingly focus on digital business models and distribution channels in order to remain competitive. This topic has been subject to research for more than a decade (e.g., Al-Debi et al. (2008)) and is often discussed in practice (Amit and Zott, 2012; Knickrehm et al., 2016). In line with it, 30% of IT managers in a recent study consider laying the foundations for digital business models one of their main tasks over the next couple of years. At the same time, the respondents are aware of the peril with 60% believing their organization is unprepared for what lays ahead (Forni and van der Meulen, 2016). Furthermore, the transition affects the entire company, individual departments need to transform as well. Yet, *transformation* efforts of individual departments show substantial differences. Marketing, for instance, has fully embraced digital channels like web pages and apps for mobile devices or terms like the "360 degree customer", who is entirely profiled and targeted with specific advertisement campaigns (Rouse, 2015; Chaffey and Ellis-Chadwick, 2019). In contrast, the finance department, once a frontrunner in driving change, is in danger of falling behind (Halper, 2014), although the Chief Financial Officer – the number-crunching conscience of the company – seems well-equipped to manage change at the intersection of business and technology.

Often lacking a clear future target state, *finance departments* are urged by analysts and consultants to pick up speed along four priorities: Firstly, a better support of corporate strategy is required; secondly, cost saving potentials need to be uncovered and leveraged; thirdly, digital transformation of accounting processes needs to be advanced; and fourthly, a shift in the mindset and employee profile needs to be realized to accommodate the new work demands

(Essaides et al., 2017). Different digital technologies have a strong impact on finance processes and the way employees in the finance department work. For instance, proficiency with new information technology (IT) and statistical analysis become more important for management accountants (Wadan et al., 2019).

Taking *analytics* as an example, qualitative research found out that most companies have not yet recognized the advantages and opportunities that analytics tools can provide (Sangster et al., 2009; IBM, 2013; Appelbaum et al., 2017). While there is a lot of ready-for-use software in order to solve specific, but restricted, forecasting problems, there is a lack of specific and customized forecasting models (Ord and Fildes, 2017). Since there is no one-size-fits-all approach, companies from different industries and with different processes will require individual and customized tools. On the other hand, due to the number of software solutions, which are increasingly available on the market for all kinds of problems, organizations often fail to choose those tools that best fit their specific requirements and to decide where to get the most out of it. There seems to be a lack of concrete case studies in literature, which may support companies with their decisions on which analytical tools to employ, how to set up an “analytics culture” (Davenport and Harris, 2007) and how to increase actual use, for which the “behavioral intention to use” is one of the key mediators (Ajzen, 1991). Moreover, there is no comprehensive study on the drivers of machine learning and analytics adoption, as it is common for more mature technologies like business intelligence (BI), e.g. by Scholz et al. (2010), Wieder et al. (2012), or Khan et al. (2010)

Similar chains of reasoning can easily be found for other digital technologies. Therefore, the aim of this thesis is to contribute to the *digitalization of the finance function* along three main research questions

(1) Which digital technologies are most relevant towards the medium-term future finance function (zero quartile)?
(2) What is the current state of machine learning and analytics adoption and which are the most relevant drivers (adoption)?
(3) How can companies leverage machine learning and analytics to advance their digital transformation and what are first guidelines (use cases)?

In order to address these questions, this work follows the tenets of *design science research* (DSR) in information systems (IS) and in particular the guidelines proposed by Hevner and Chatterjee (2010) and Gregor and Hevner (2013)[1]. The thesis will be split into two parts.

Part I covers research question (1) and proposes a maturity model, design guidelines, and a model for evaluating investments in digital technologies. Following the steps of the publication schema by Gregor and Hevner (2013), after an introduction into finance processes, digital technologies, and current benchmarking practices (introduction, sections 1 and 1.1), existing knowledge is reviewed and classified (literature review, section 1.2). Based on survey data, the Rasch algorithm is adapted (method, section 1.3) to derive eight design guidelines and four imperatives for leveraging digital technologies to improve finance process activities (results, section 1.4). In order to evaluate the results a benefit assessment model is proposed (evaluate, section 2.2). Finally, the results are compared with prior work, put into perspective, and limitations as well as avenues for future research are stated (discussion, section 2.3).

Part II covers research questions (2) and (3) and describes the current state and drivers of machine learning and analytics adoption as well as use cases for machine learning in financial accounting and predictive analytics in management accounting. It starts with an introduction into machine learning and analytics adoption (introduction, section 3). Afterwards, existing literature on machine learning and analytics in the finance department is reviewed (literature review, section 3.1). Applying partial least squares structural equation modeling (PLS-SEM) to survey data (method, section 3.2), descriptive results and most relevant drivers for adoption are presented (results, section 3.3). Subsequently, the

---

[1] Gregor and Hevner (2013) published a comprehensive overview of how researchers can develop and present their projects for maximum impact and proposed a seven-step publication schema for DSR in IS. They distinguish between two types of useful knowledge, descriptive (“what knowledge”) and predictive (“how knowledge”). Although the three research questions of this thesis are phrased “which, “what”, and “how”, they will all lead to descriptive and predictive knowledge to some degree. Furthermore, they consider a two-by-two grid of low or high solution maturity and low or high application maturity. This grid then leads to four contribution types: routine design (high solution maturity, high application maturity), improvement (low, high), exaptation (high, low), and invention (low, low). This work is mostly situated in the improvement quadrant, since the maturity of activities in the finance department is rather high. Finally, the publication schema consisting of introduction, literature review, method, artifact description, evaluation, discussion, and conclusion helps to answer the three main questions that all research projects should face: “is it new”, “is it interesting”, and “is it true”.

algorithms used for the two use cases are introduced (fundamentals, chapter 4) and the two use cases are presented as an extended form of demonstrating the validity and utility of machine learning and analytics in the finance department (evaluation, chapters 5 and 6). The thesis concludes with a broader reflection on whether the three main research questions could be answered, limitations, avenues for future research, and a short summary. Figure 1 provides a rough overview of the topics addressed in the two parts of this thesis.

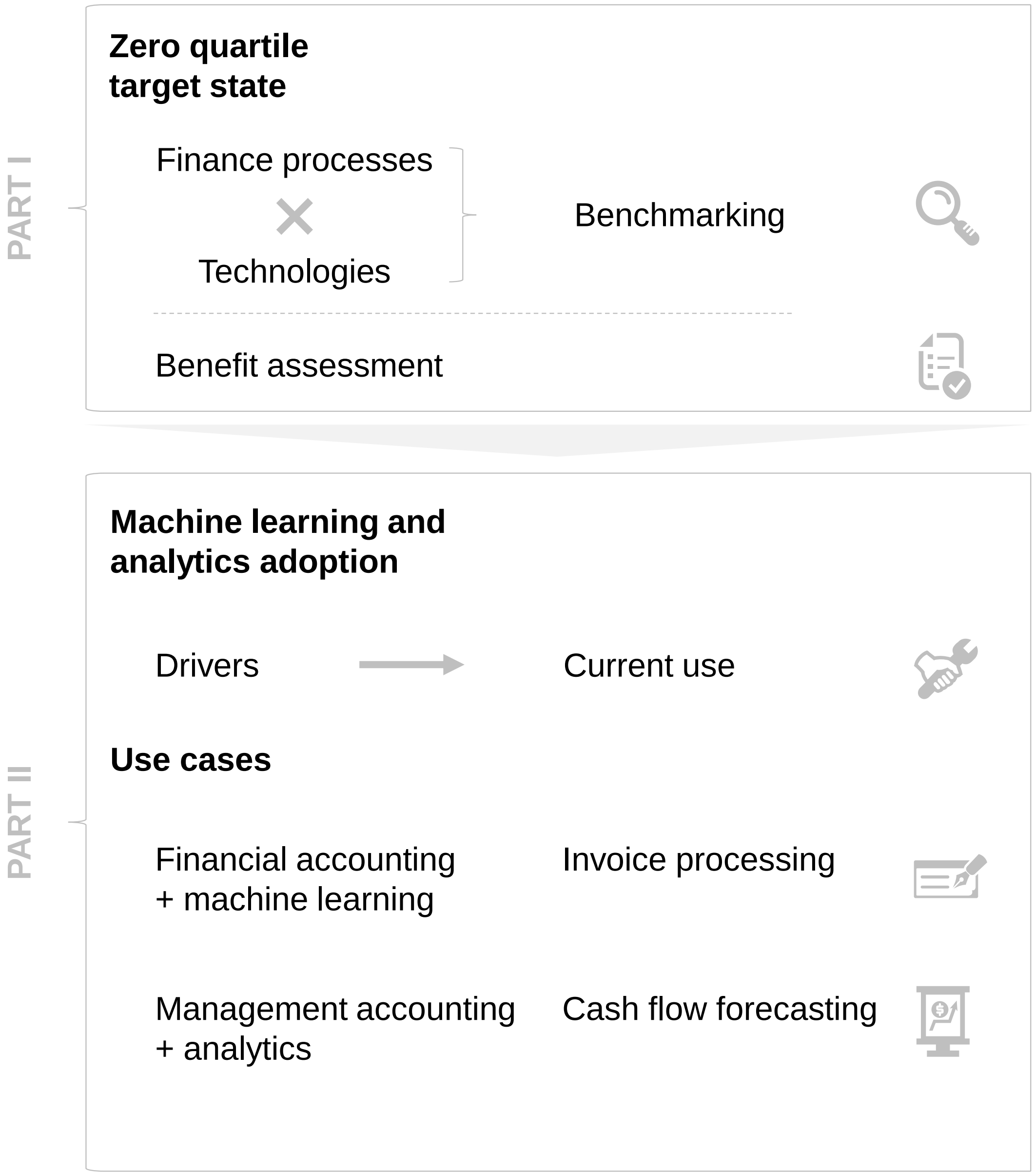


Figure 1. Overview of the topics addressed

1. The zero quartile is the new target state for finance organizations

Currently, there is no clear *guidance* for finance functions in which *technologies* to invest at what moment. "From bookkeeper to business partner" is a term slowly becoming overused in practice, but still addressed regularly by research, e.g. Wadan et al. (2019), as it is not fully ingrained in practice. Additionally, topics like outsourcing and shared service centers, which already became relevant in the early 2000s (Bangemann, 2005), are mostly considered a standard. Nonetheless, finance organizations are in a transition phase with a new wave of zero-based budgeting[2] to free up resources for investments in new technologies that change the way people work.

In the next two chapters this lack of guidance is addressed and a technology-driven target state for the future finance function is proposed.

1.1. Fundamentals: The four finance core processes are reshaped by digital technologies

The following chapter will lay the foundations for the remainder of part I and introduce finance processes, digital technologies, and the fundamentals of benchmarking.

1.1.1. Finance processes

In most companies, the finance department is responsible for activities like accounting and transaction processing, providing financial information, tax and cash management and financial controls (Smith and Payne, 2011). Four *core processes* help to cover these responsibilities.

---

[2] Zero-based budgeting is an approach to annual planning where managers have to justify their entire budget from a base of zero, hence there is no carry-over or lump sum based on last year's budget (Cheek, 1977), which is common in other budgeting approaches.

Covering eight activities, the *order-to-cash process* (O2C, Figure 2, top row) generates cash from sold products and services (Hall, 2018). (1) Deciding about the adequacy of sales on credit to a customer, the O2C process starts with credit authorization (Knechel and Salterio, 2017). (2) Once approved, master data for a new customer are created or existing data are updated. (3) Invoicing comprises the bill generation and distribution to the customer following the terms of trade like payment date (Romney and Steinbart, 2018). (4) Maintain accounts receivable (AR) ledger and apply cash comprises the processing of a payment receipt, deposition of the customer payment, and updating the customer's AR ledger as well as the general ledger (Hall, 2018). (5) Analyzing the customer account balances is covered by managing and processing collections. Uncollectable balances must be written-off. (6) Including a root-cause analysis, the dispute and deduction management prepares chargebacks and deductions. (7) This is followed by managing customer requests and inquiries and, finally, (8) the execution of revenue assurance activities. The latter covers revenue stream monitoring and leads to the implementation of preventive measures in case of leakage.

The *purchase-to-pay process*[3] (P2P, Figure 2, second row) converts the organization's cash into resources necessary to conduct business. It covers eight activities. (1) Requisition and fulfillment determines the order requirements (Bodnar and Hopwood, 2013) and selects the supplier typically by a standard procedure such as contract, purchase-order or purchase card. (2) After maintaining vendor master data, (3) manage inbound documents comprises the receipt, scan, and archiving of the inbound documents. (4) As goods are received the invoice processing follows. A receipt is created and must be approved by inventory controls. Inventory records are updated and the accounts payable (AP) department processes the invoice from the supplier. (5) Process payments covers the payment schedule (Deshmukh, 2006). The AP subsidiary ledger is updated, and summary totals are forwarded to the general ledger. (6) In case of discrepancies the management of vendor disputes succeeds. This is followed by the (7) reconciliation of the general ledger accounts and period-end closing. (8) The process ends with the management of the purchase card program, i.e. the

[3] Software vendors regularly refer to the process as procure-to-pay.

selection of card vendors, the issuance to selected employees and the maintenance of card-holder profiles[4].

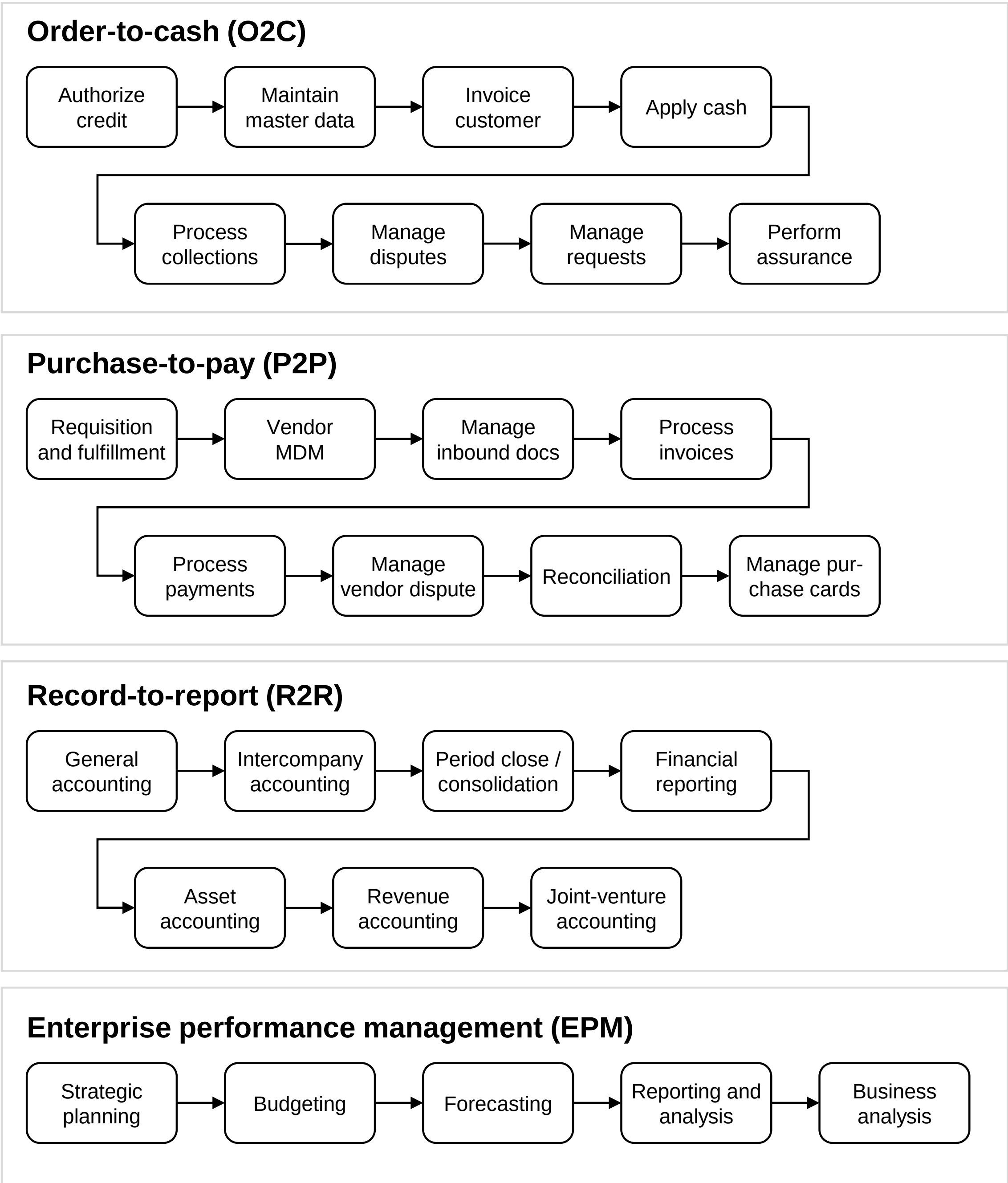


Figure 2. Finance processes

[4] Note, that purchase cards are common in the United Stated (US) but less so in Europe. Since the majority of accounting research is based in the US and the companies that were part of this research project all have branches in the US, purchase cards were included at this point.

After obtaining operating and accounting data, the *record-to-report process* (R2R, Figure 2, third row) prepares the financial reports. The process covers seven activities initiated by (1) general accounting that covers the recording of sales or purchase transactions in accordance with journal entry policies (Bodnar and Hopwood, 2013). (2) Intercompany accounting comprises the reconciliation of intercompany balances, followed by the (3) period closing activity, which is the basis for (4) financial reporting. As an example, the (5) intangible asset accounting measures, valuates, and records the intangible assets. This activity is succeeded by (6) revenue and cost accounting, which identifies performance obligations, transaction prices, and closing and reporting costs. (7) In case of a significant share in a joint venture and other third-party transactions, joint venture accounting accounts profits/losses of such an interaction.

Enterprise performance management (EPM, Figure 2, bottom row) represents activities designed for the execution of a business strategy (Eckerson, 2004). (1) Strategic planning covers finance activities like the development, review, and monitoring of strategic business plans as well as the support of the preparation of mid- and long-term plans (3-5 years) (Frolick and Ariyachandra, 2006). (2) Annual planning (Budgeting) contains activities that are associated with the organization´s annual financial planning efforts. This primarily includes the annual budgeting process and development of the capital plan (Mowen and Hansen, 2005). (3) Forecasting consists of periodic activities of monthly, quarterly and annual efforts as well as event-driven activities. This primarily includes year-end or rolling forecasts, development of profit and loss statements, balance sheets, cash-flow statements, and collaboration with operating units (Sorensen, 2013). The fourth process (4) reporting and analysis comprises processes and activities associated with reporting the company's short and long-term financial and non-financial performance. This includes the compilation and creation of standard and ad-hoc management reports (may be done by traditional paper-based reports, web-based reports, balanced scorecards delivered in an executive information system or a combination) (Frolick and Ariyachandra, 2006). Finally, the aim of (5) business analysis is to conduct complex analyses as the basis for decisions with more impact or the need for special data. This includes mid- to long-term decisions that cannot be delegated and ensure continuity of the company.

1.1.2. Digital technologies

There is always a broad spectrum of emerging, maturing, and mainstream technologies that organizations have to consider when planning their investments. For instance, looking at the Gartner hype cycles for *emerging technologies* of the past years (e.g., (Panetta, 2017)), there are technologies like machine learning, the internet of things, cloud computing, the blockchain, virtual and augmented reality, virtual assistants, and advanced analytics with self-service delivery[5]. However, while all of these technologies may have a strong impact on the future finance function, for the remainder of this work three topic areas will be in focus: Automation, machine learning and analytics (ML&A), and digital enterprise platforms. What they have in common is that they all have moved beyond the point of pure hype, but are still far from being adopted by the mainstream company.

The term *automation* is derived from the Greek language and signifies acting on its own, self-moving or self-directing. In the context of digitalization, it is defined as machines, tools, devices, and information systems performing a given set of activities without human intervention (Nof, 2009). Generally, there are two categories of automation that are relevant for the finance function (Figure 3, top). While *rule-based automation* (with its most prominent instantiation "robotic process automation", RPA[6]) targets routine tasks with structured data and deterministic outcomes (Lacity and Willcocks, 2016), *cognitive automation* intends to perform activities which were performed by human operators and require more cognitive capacities such as situation assessment, problem solving and pattern recognition (Fast-Berglund et al., 2013).

Closely related to cognitive automation is *artificial intelligence* (AI), a term that was coined by McCarthy et al. prior to the 1956 Dartmouth summer research project (McCarthy et al., 2006). It is generally described as a technique that

---

[5] There are also numerous other technologies that most likely do not have any direct impact on the finance function over the next years like 3D printing, autonomous vehicles or quantum computing and are therefore left out at this point.

[6] RPA enables the automation of business processes with the help of software robots. These software robots work on the presentation layer of software, hence on the same layer as humans do. They work similar to recorded macros and can be trained to read emails, open files, identify text fields, enter data and trigger workflows for further robots or human workers (Vasarhelyi and Rozario, 2018).

enables machines to mimic human behavior (Rich and Knight, 1991) and includes research fields like natural language processing (NLP), image and object recognition or text and speech recognition[7].

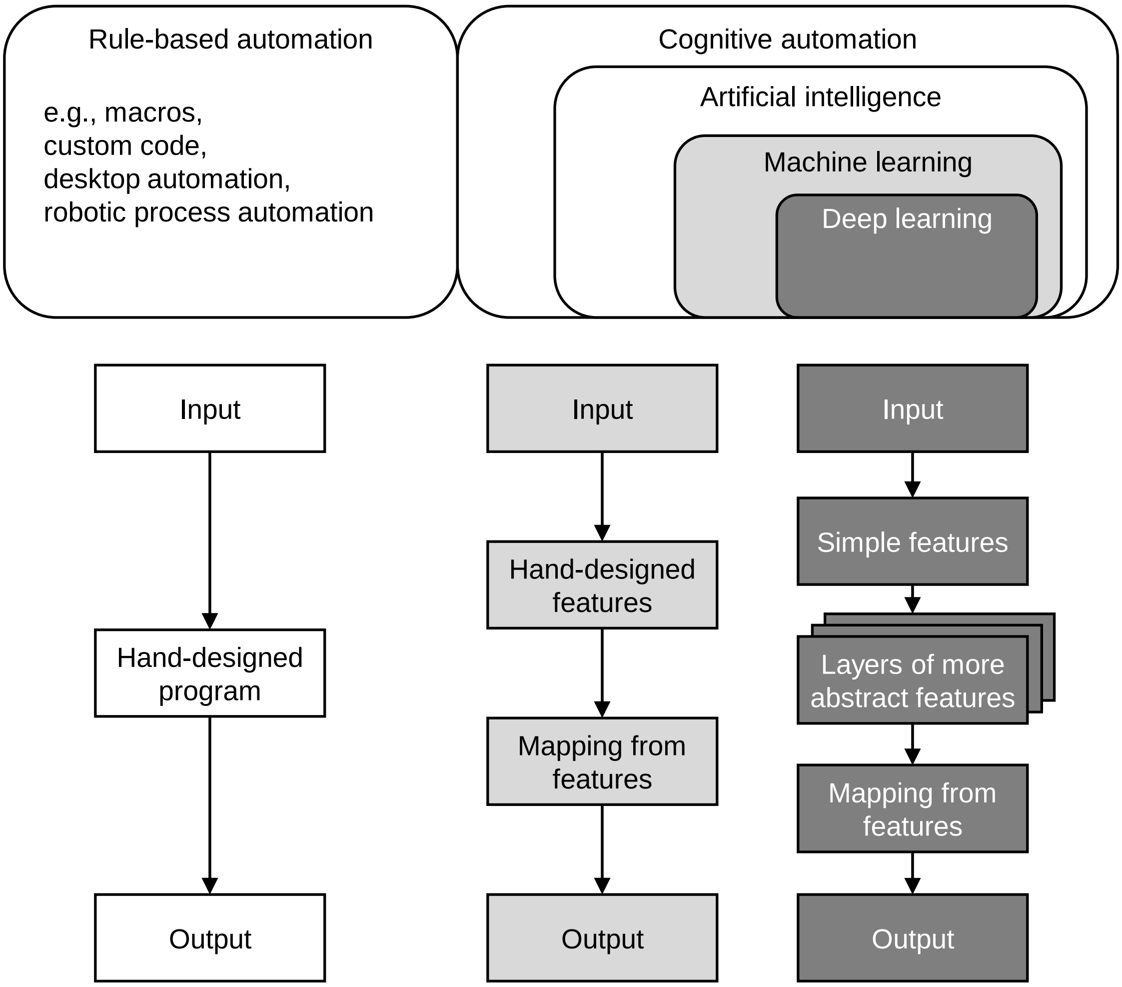


Figure 3. Overview of automation categories (modified after Goodfellow et al. (2016))

A cornerstone of AI is *machine learning* (ML), which enables machines to learn a task from a series of examples based on logical operations and perform it more effectively the next time (Simon, 1983). Machine learning algorithms can be grouped into three different categories: supervised, unsupervised, and reinforcement learning (Kacprzyk and Pedrycz, 2015). Regardless of the

[7] The potential applications of AI span basically all industries (Kurzweil, 2010). At this point, only some fundamentals will be laid out, while more specific details with respect to the work at hand will be introduced in part II. For more information on AI, NLP, and the ongoing discussion of when a machine or system can be considered intelligent, refer to Russell and Norvig (2016) or Nilsson (2014).

category, the aim is always to try and induce a model of or identify patterns in the entire dataset based on a fraction of it (Fürnkranz et al., 2012). The difference lies in how the algorithms classify data.

In *supervised learning* the correct output for training data is known. A teacher provides the output value of the target function for all labeled examples, so-called training data (Kotsiantis et al., 2007). The purpose of supervised algorithms is to find a general rule based on historical data that characterizes the input data in the best possible way. Examples for supervised learning approaches are the classification of customers based on their payment behavior or the classification of markets based on their characteristics (Fürnkranz et al., 2012). Most often, the classification is followed by a prediction, e.g., a prediction of the payment behavior of a customer based on certain character traits.

By contrast, in *unsupervised learning,* data is unlabeled. There is no information except the input values for training examples and the main task lies in the automatic creation of classification rules. Unsupervised learning algorithms search for similarity between sets of data to create groups – so-called clusters (Talwar and Kumar, 2013). In other words, algorithms in unsupervised learning try to identify homogeneous groups of examples that are similar to each other but differ from examples in other groups (Fürnkranz et al., 2012).

Lastly, *reinforcement learning* is a combination of unsupervised and supervised learning. After every (unsupervised) action, the learning process is “awarded” or “punished”. Therefore, it performs a series of actions in order to maximize the award. This makes it different to unsupervised learning as there is guidance from an external evaluation. On the other hand, unlike in supervised learning, the learner is only provided an evaluation of the action made. In addition to these fundamental learning paradigms, there are often hybrids. The best-known of them is semi-supervised learning, which uses both labeled data and unlabeled data for training (Kacprzyk and Pedrycz, 2015).

Artificial intelligence research quickly found solutions to very difficult problems (for humans) as long as they could be described by a set of mathematical rules. On the other hand, things that were solved intuitively by humans posed problems for these earlier solutions (Goodfellow et al., 2016). Therefore, more complex algorithms were developed which are referred to as *deep learning* (DL). As opposed to the simpler approach of rule-based automation (Figure 3, left) and the still relatively straight-forward ML, DL comprises several layers to break down the complexity of such problems into different levels of abstraction (LeCun et al., 2015). At this point, DL is only mentioned for completeness; it will not be required in the remainder of this work.

With its roots in the field of statistics, *analytics* derives insights from a vast amount of data (Agarwal and Dhar, 2014). As such, it makes use of statistical methods and machine learning. Since the late 2000s, the term analytics is used to describe business intelligence (BI) components that rely on advanced statistics or visualization (Chen et al., 2012). Analytics differs from classical BI in that it links a set of explanatory variables to a business response based on a given set of data either with the aim of predicting something or inferring causalities (Baesens et al., 2016). Only today, the fast pace at which transactions are moving online allows for the collection of vast amounts of data. Thus, analytics is becoming more relevant to practitioners (e.g., (Lavalle et al., 2011) and scholars (e.g., (Agarwal and Dhar, 2014) alike. Building on the omnipresence of data arising from all kinds of sources such as enterprise systems, social networks, mobile devices, public data, and the internet-of-things, analytics comprises three main areas: Descriptive, predictive, and prescriptive analytics (Delen and Demirkan, 2013).

Descriptive analytics answers the question of what exactly happened and/or why it happened[8]. It comprises standard reporting, periodic business reporting, and ad-hoc reporting to identify business opportunities and problems (Delen and Demirkan, 2013). In contrast, *predictive analytics* addresses the questions of what will happen next or what is likely to happen (Watson, 2014). With the aim of building and assessing models for making empirical predictions, predictive analytics employs a broad range of advanced methods from statistics and ML (Gandomi and Haider, 2015). Finally, *prescriptive analytics* answers the questions of what a company should do and why based on descriptive and predictive models (Souza, 2014). The goal is to determine the optimal decision based on

[8] In practitioner literature, this second aspect is often labeled diagnostic analytics, e.g., Gartner (2015).

several alternatives and thereby improve the business performance (Liebowitz, 2014). Figure 4 provides an overview of these three facets of analytics.

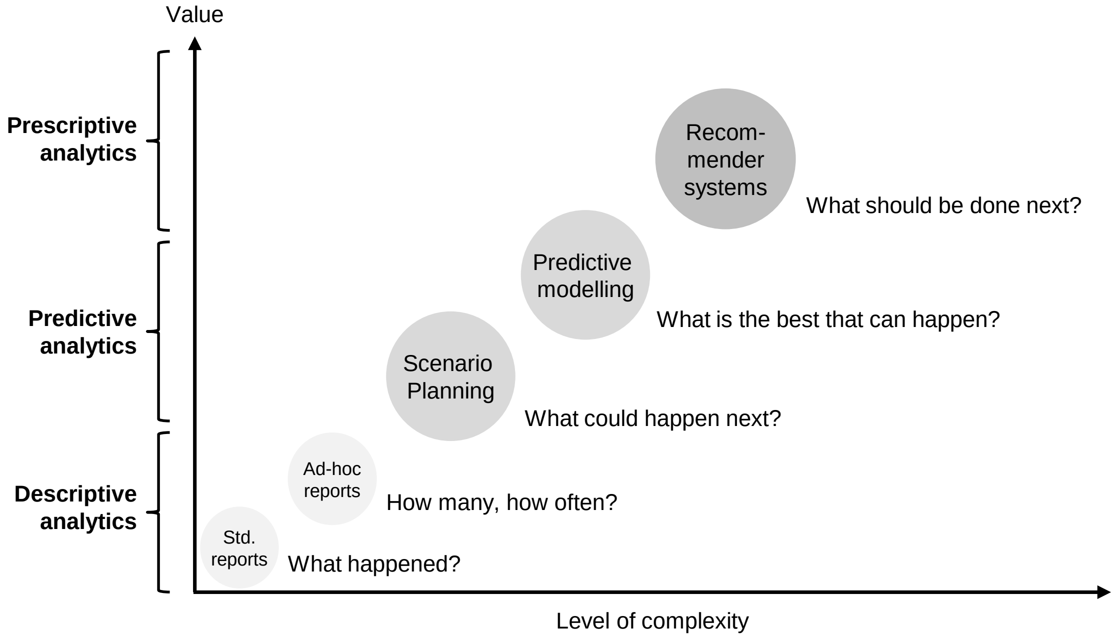


Figure 4. Varieties of analytics (modified after Gartner (2015))

Many authors discourage implementing analytics merely as a new technology, e.g., Franks (2014). Rather, they argue it is to be considered as a technology stack with the necessity to introduce an analytics organization and promote a so-called analytics culture (i.e. objective thinking and data-driven decision making). A *digital enterprise platform* can be seen as such a technology stack. It combines the concepts of digital transformation, the internet of things, and platforms-as-a-service (Lucas, 2016) and is sometimes referred to by terms like “next-generation architecture” (Hutchinson et al., 2009) in an IT context or “cloud enterprise resource planning” with a focus on core enterprise processes (Saeed et al., 2012). Combining in-memory engines, hot and cold storage concepts[9], cloud computing and master data management in one platform, DEPs provide the required simplification of IT and enhanced performance across ERP and BI systems. They aim at supporting the increasing process automation, more

[9] Hot data refers to frequently accessed records, which is often the minority in transactional databases, while cold data refers to less often accessed records (Levandoski et al., 2013).

intuitive information systems, and more fact-driven decisions on real-time data (Lucas, 2016). To this end, DEPs combine a database layer, an application layer with interfaces to internal and external data, and a device-agnostic frontend layer (Figure 5).

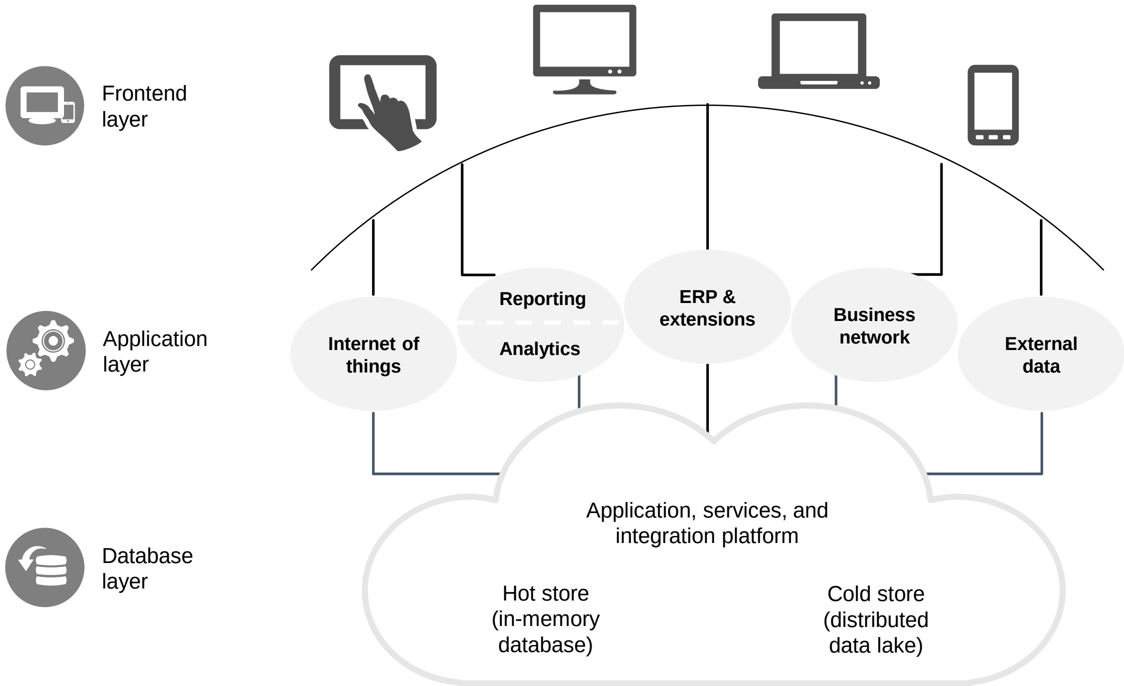


Figure 5. Digital enterprise platform overview

1.1.3. Benchmarking

In order to remain competitive, companies have to continuously improve strategy, organization, and information systems support. As an instrument to gauge the competitiveness, *benchmarking* is a continuous process of identifying highest standards of excellence for products, services, and processes (Bhutta and Huq, 1999). Challenged by the rising Japanese export rates and to overcome their own quality and cost problems, the first companies to use benchmarking were IBM in the late 1960s and Xerox in 1979 (Horvath and Herter, 1992; Jackson, 2001). Since then, a considerable body of literature has been created. In practice, most often a so-called *"first quartile"* is applied (The Hackett Group, n.d.). Best practices of high-performing companies define this first quartile (Joo et al., 2011), which serves as the top 25 percent target state (Dolan and Moré, 2002).

In general, benchmarking can be used to compare all functions of a company. Yet, the *finance function* was often in particular focus (Yasin, 2002). For example, in the 1990s Fisher (1994) stated that accountants have to stop looking only at data from the past. He highlighted benchmarking as a tool to get a picture of how up-to-date a finance department is (Guilding et al., 2000). The effects of IT on finance in general have been studied regularly (Bhimani, 2003; Deshmukh, 2006; do Céu F. Gaspar Alves, 2010). Practitioners are addressing the changes as well – either in a broader context (Axson, 2015; Lucas, 2016) or evaluating the impact of a digital technology, such as automation, on finance processes in detail (Plaschke et al., 2018). However, a rigorous literature review is missing.

The *process of benchmarking* covers five steps. Firstly, processes and benchmarking partners are selected. Secondly, key performance indicators (KPIs) and business processes are collected. Thirdly, the differences (or gaps) between the indicators and processes of benchmarking partners are analyzed. Fourthly, activities and processes, which should be changed or adapted, are determined. Whereas the comparison of KPIs can offer information on possible process improvements, information flow comparison and process logic comparison can shed light on how to close the gaps. Fifth and finally, the redesigned processes are implemented and monitored (Juan and Ou-Yang, 2005).

1.2. Literature review: Finance technology benchmarking is mostly backward-looking

Literature reviews are a widely accepted methodology for all fields of research. They are not only used as a first step for a research project, but also as a means to categorize existing research, present avenues for future research, and facilitate theoretical progress (Cooper, 1998; Webster and Watson, 2002; Vom Brocke et al., 2009).

1.2.1. Search strategy

For the literature review at hand, the recommendations given by vom Brocke et al. (2015) served as a starting point. Aiming for a comprehensive review, an *iterative search process* was employed. Starting with "benchmarking" and "digital technology", search terms were updated whenever new relevant aspects were identified in reviewed papers.

The focus laid on leading IS[10] research outlets and proceedings from major IS conferences[11]. Additionally, journals from business, management, and accounting[12] were included. ScienceDirect, EBSCOhost, JSTOR, and Google Scholar were used to access the journals and look at titles, abstracts, and key-words. Since the subject is of considerable practical interest as well, the work of practitioners was also included by looking at MIS Quarterly Executive, Harvard Business Review, and "grey literature" – led by EBSCOhost's Business Source Premier and the standard Google search.

Due to the large number of non-relevant hits, the *search string* was adjusted first by adding "finance (function)" as the case example with associated key words such as financial and management accounting, enterprise performance management, and the more general "business process" term. Table 1 gives an overview of the final search terms. Articles were considered relevant if their title, abstract, or keywords covered benchmarking (or one of the alternative search terms in line one of Table 1) and at least one of the finance- or technology-related search terms. This first step led to 37 relevant hits.

---

[10] Based on the senior scholars' basket of leading IS Journals: European Journal of Information Systems (EJIS); Information Systems Research (ISR); Information Systems Journal (ISJ); Journal of the Association for Information Systems (JAIS); Journal of Information Technology (JIT); Journal of Management Information Systems (JMIS); Journal of Strategic Information Systems (JSIS); MIS Quarterly.

[11] Americas Conference on Information Systems (AMCIS); European Conference on Information Systems (ECIS); International Conference on Information Systems (ICIS); Pacific and Asia Conference on Information Systems (PACIS).

[12] Top journals from Scimago JR. Subject area Business, Management and Accounting; subject categories Accounting, Business and International Management, Business, Management and Accounting, Management Information Systems, Management of Technology and Innovation, Organizational Behavior and Human Resource Management, Strategy and Management.

To look for articles cited in the best-fitting papers (backward) and newer articles citing the identified papers (forward), a *backward and forward search* was conducted. Synthesizing the results, this led to a total of *58 relevant publications*.

Table 1. Search terms

<table>
<tr><td></td><td colspan="6">OR</td></tr>
<tr><td rowspan="3">AND</td><td colspan="2">Benchmarking</td><td colspan="2">Maturity model</td><td colspan="2">Data envelopment analysis</td></tr>
<tr><td rowspan="2">OR</td><td>Digital technology</td><td>Automation</td><td>Analytics</td><td colspan="2">Digital Enterprise Platform: In-memory, data lake, cloud, master data management, collaboration, mobile (finance) apps, workflow management</td></tr>
<tr><td>Finance</td><td>Financial accounting</td><td>Management accounting</td><td>Business process</td><td>Enterprise Performance Management</td></tr>
</table>

1.2.2. Overview of results

Benchmarking has been of scientific interest for a couple of decades now. Earlier literature focused on introducing the idea and concepts behind benchmarking (Pryor, 1989; Henricks, 1993) or first applications in practice (Tucker et al., 1987; Shetty, 1993). In general, this first phase (Figure 6, left) is characterized by a limitation to single tasks or the implementation of isolated best practices, hence the name *emergence*. The second phase (Figure 6, middle) applied benchmarking to a wider range of industries (e.g., public sector) or more specific IT topics (e.g., IT costs, benefits achieved) – therefore named *exploration*. This is also the time when not only single tasks or financial KPIs, but also entire processes were benchmarked (Gleich et al., 2008). The third phase is characterized by the increasing pervasion of new digital technologies. Process benchmarking, on the other hand, is less prominent (only two publications as opposed to ten in the second phase), hence the name *evolution*. New methods for benchmarking were introduced in all three phases (Camp, 1989; Binder et al., 2006; Ketter et al., 2015) which is an indicator that there is still potential to improve benchmarking practices.

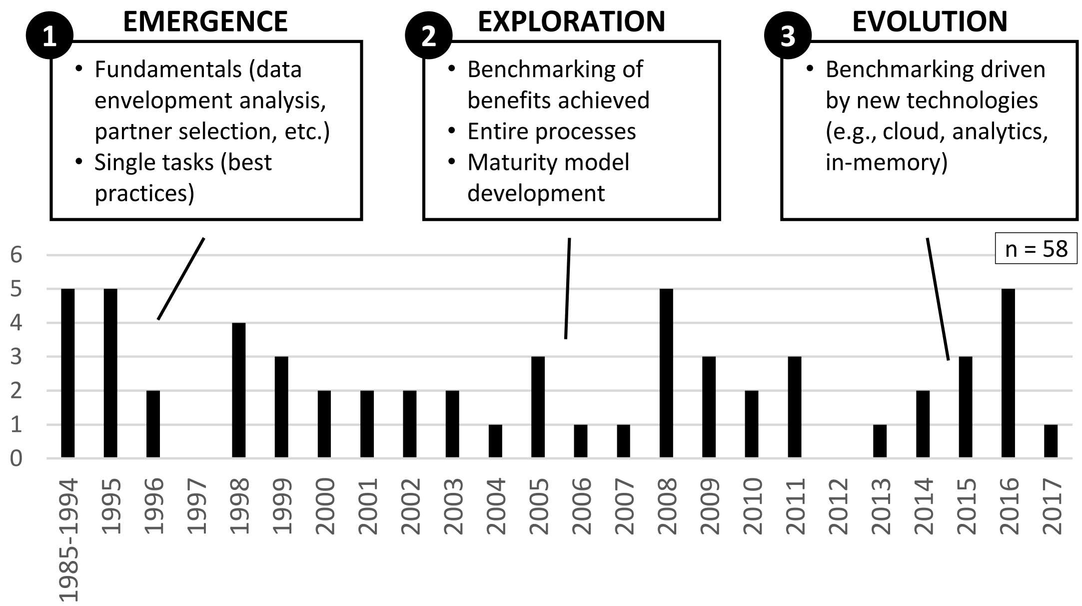


Figure 6. Search results on a timeline with three phases

1.2.3. Framework for classification – Taxonomy of benchmarking

Structuring the 58 reviewed articles on benchmarking a five-component *taxonomy* was developed. Some of the components and their component-levels were derived from the review of relevant articles, e.g. activity is based on Gleich et al.'s (2008) performance objects. However, several component-levels were grouped and technology-related drivers were added. In general, taxonomies represent the objects or concepts under consideration into an ordered classification for further studies, meaningful comparisons, and focused discussions (Milgram and Kishino, 1994). Following Ramaprasad and Syn (2015), the proposed taxonomy allows to concatenate the columns (components) into natural English sentences. (1) The first component looks at the type of *comparison* – either benchmarking in general or more from a mathematical standpoint like data envelopment analysis (Cooper et al., 2004) or with the goal of developing a maturity model (De Bruin et al., 2005). The component level "other" includes comparisons from outside without the aim to close the gap, as it is generally the case with benchmarking. (2) *Function* covers the back-office functions of a company (Sako, 2006). For the review at hand, the focus was laid on the financial accounting, management accounting, and information technology departments. (3) *Activity* describes the level of detail on which benchmarks are performed. Single tasks are most often so-called best-practices; hence they are outcome-based (Berghout et al., 2009). Process-based benchmarking considers entire processes and process chains making it also appropriate for overhead cost areas and, thus, for intra- and cross-industry approaches (Beretta et al., 1998; Anderson and McAdam, 2005). On the other hand, company performance is a more high-level form of benchmarking. It compares financial or process KPIs as it is often done in more practice-oriented research (Flaherty et al., 1995). (4) *Target* refers to the partner or database against which a comparison is made (Camp, 1989). Generally, this is either a peer or the best-in-class (across industries). Both categories are inherently retrograde, thus, benchmarking against the theoretical optimum was added as another option. Introducing this option, ideas from maturity models (assessment models for the maturity of social or technical systems) were drawn on where a company compares its process maturity, object maturity or people capability against a pre-defined set of achievement levels (Mettler, 2011). However, what is considered the theoretical optimum is not predefined by the researcher, but rather based on what experts consider the best possible future state leveraging digital technologies. Finally, the (5) *driver* that makes companies embark on a benchmarking journey was considered. The large range of potential drivers was clustered into internal (such as cost cutting or management decision), technology-driven (in particular, automation, analytics, and digital enterprise platform) or other external factors (like globalization efforts or external analyst reports).

Mapping the literature onto the taxonomy, the (a) *research approach* and (b) *methods applied* as well as the (c) *contribution type* of the examined articles were also looked at. Following a general distinction in the IS discipline, the research approach was declared to be either behavioral or design science research (Hevner and Chatterjee, 2010). Due to the generic nature of benchmarking, many articles also applied other research approaches (systems approach, theoretical analysis, mathematical deduction), which resemble DSR in some aspects. Nevertheless, they were kept separate as they did not reference any IS publications in their method section. Looking at the contribution type, models (abstractions and representations, (Hevner et al., 2004) and methods (algorithms and practices) were included based on the distinction proposed by March and Smith (March and Smith, 1995) – there were no constructs or instantiations, hence they were left out. In order to highlight the proximity of benchmarking and maturity models, the latter were explicitly looked at as well. Finally, guidelines (including frameworks) and survey results (where authors report on findings, but do not synthesize results via one of the other contribution types) were added.

1.2.4. Results

Examining the results mapped onto the taxonomy (Figure 7), a few things can be noted directly. Benchmarking is by far more often applied than maturity models or data envelopment analysis. The same holds true for "single tasks," "peer" as a benchmarking partner, and "internal" as driver. On the other hand, maturity models, financial accounting, benchmarking against the theoretical optimum and technology as driver are rarely mentioned.

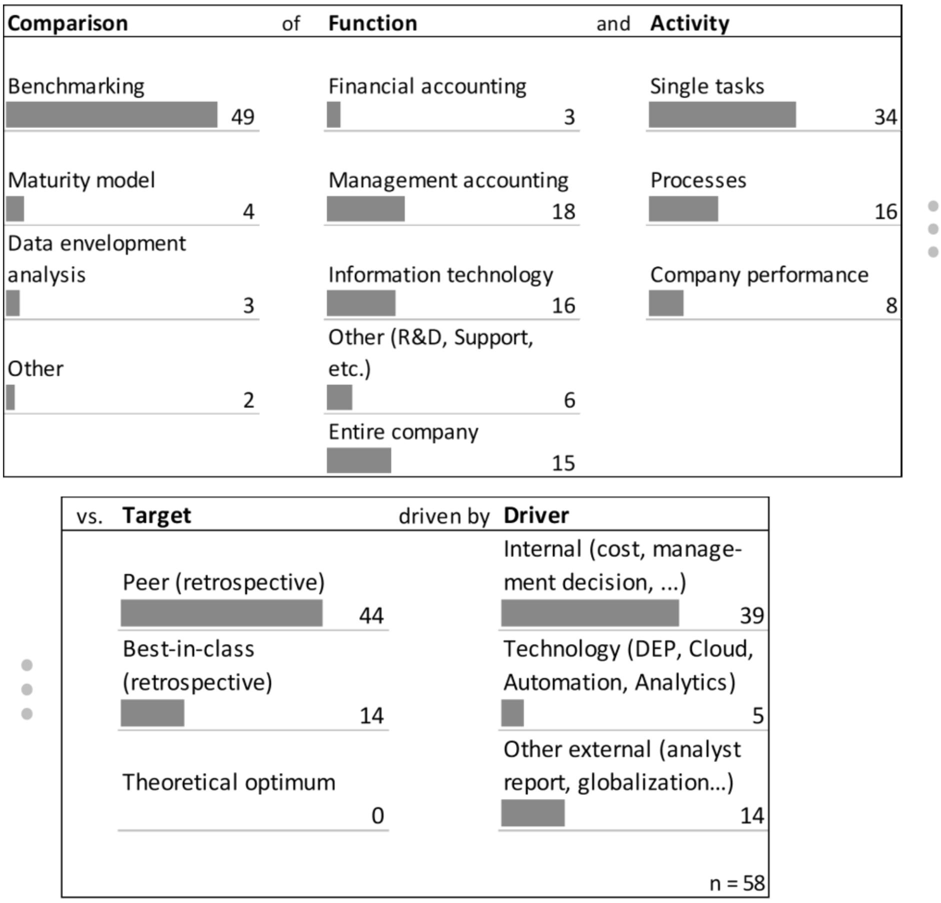


Figure 7. Taxonomy of benchmarking

Specifying *financial accounting*, Baliga (1995) addresses the purchase-to-pay process. Benchmarks were made regarding the number of transactions per each full-time employee. Blumenberg (2004) benchmarks the invoicing activity to emphasize efficiency in the order-to-cash process. There was no research covering the record-to-report process. Regarding *management accounting*, 18 articles were examined. For example, Jazayeri and Hopper (1999) discuss the influence of management accounting when adopting manufacturing best practices. Taking cost management as an example, Elnathan et al. (1996) propose a framework to measure the success of benchmarking projects by costs and non-financial measures such as speed to market. Eckerson (2004) describes activities within the EPM process and Marx et al. (2012) provide a maturity model to systematically align management accounting processes from an organizational, functional, and IT perspective.

Regarding the function to be benchmarked, a large fraction of sixteen articles focus on the *IT domain* and cover IT costs (Krotov and Ives, 2016), best practices of single IT applications (Cragg, 2002) or enterprise IT maturity levels (Seong Leem et al., 2008). Covering digital technologies, the majority of articles focus on cloud computing, e.g., Yuan et al. (2015) look at cloud storage. Automation is not new in accounting (Carlson, 1957), but gains increasing attention due to new rule- and cognitive solutions (Lacity et al., 2015). Predictive analytics on big data is mentioned by Ketter et al. (2015). Explaining the Digital Enterprise Platform, Jin et al. (2016) benchmark in-memory engines and Ferme et al. (2017) cover workflow management systems. Master data management is examined by Spruit and Pietzka (2015). However, all these publications only incorporate past data. More in detail, Chenhall and Langfield-Smith (1998) gathered data over a period of three years for the manufacturing industry. Flaherty et al. (1995) gathered backward-looking balance sheet data. Joo et al. (2011) relied on public data, which was filed with the Securities and Exchange Commission.

The research approach, methods applied, and contribution type show a rather even distribution at first glance (Figure 8). Only DSR and interviews are clearly underrepresented. DSR was used by two authors, whereas 30 references employed a behavioral approach. The latter are based on observations and apply empirical methods such as case studies, surveys or experiments (Urbach et al., 2009). The remaining twenty-six articles could not be assigned precisely to DSR or behavioral research (see explanation in “framework for classification”).

Regarding the type of contribution, Teuteberg et al. (2009) suggest an ontology-based model for automating the benchmarking process. Their model identifies differences in quantity (e.g., diverging performance indicators) and quality (e.g., the level of detail of processes). Ketter et al. (2015), in turn, present a novel method to address complex challenges of societal scale. They propose a step-wise research design of “compete, analyze, disseminate, and realign.” Gleich et al. (2008) developed a method to benchmark accounting activities and sub-process-related costs. They propose four steps “preparation, analysis, comparison, and stage” and also give insights into non-financial indicators such as process time or volume. *Maturity models* were developed for example to address the relationship between strategic priorities, management techniques, and management accounting (Chenhall and Langfield-Smith, 1998) or for IT maturity stages (Seong Leem et al., 2008). Marx et al. (2012) are the only authors to use the Rasch algorithm to derive a maturity model, although the algorithm seems well suited to the field of IS research. It derives the expected (to-be) state as well as the current (as-is) state, measures the gap between the two and shows development paths towards the to-be state.

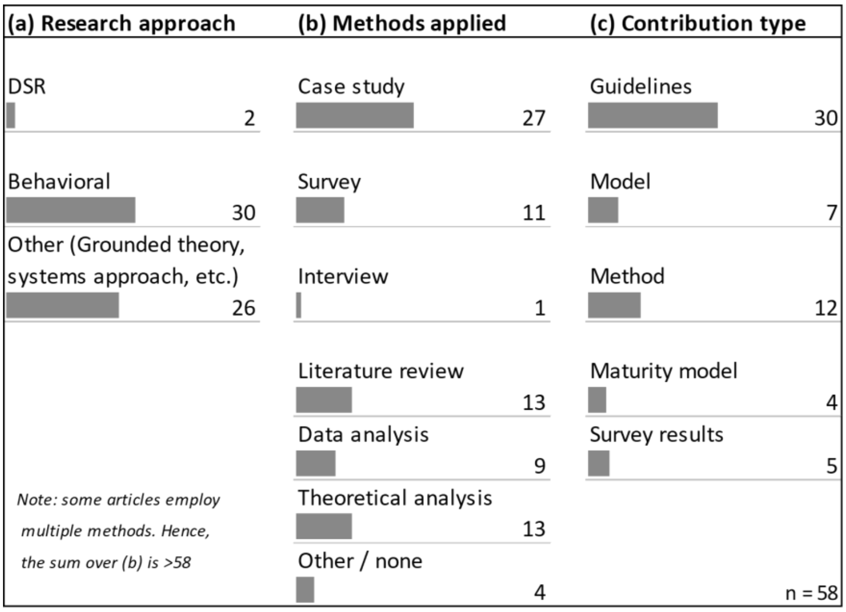


Figure 8. Research approach and research output

Many authors employ two or more methods. For example, Berghout et al. (2009) combine findings from a literature review with a case study. Spruit and Pietzka (2015) develop a maturity model based on a literature review. They benchmark the master data management of organizations.

To have a closer look at the coverage of specific areas, *dyads* between all components of the taxonomy were computed (Figure 9). These dyads represent the number of co-occurrences of each pair, e.g., financial accounting and benchmarking (3 co-occurrences, (Baliga, 1995; Blumenberg, 2004; Juan and Ou-Yang, 2005)). In addition to the insights gained from Figure 7, this visualization shows a couple of things. First, processes were mostly addressed by benchmarking (14 vs. 1 maturity model and 1 data envelopment analysis) and their spread over the different functions is relatively even. Moreover, a comparison against the best-in-class was mainly used for benchmarking (benchmarking 11, e.g., Cooper and Ezzamel (2013), maturity models 2, other 1), but never for financial accounting.

As a final step in the analysis, selected *triads* were computed combining function and object of the taxonomy with three contribution types, namely method, model, and guidelines as they were the most frequent ones (Figure 10).

This illustration shows that for example methods for a comparison against the best-in-class were only provided for the IT department (e.g., Doll et al. (2003), 3 in total). On the other hand, methods for comparison with peers (mostly benchmarking, as can be derived from Figure 9) are provided for all functions, except for financial accounting. Guidelines for peer comparison in financial accounting do exist, however.

		Comparison				Function					Activity			Target		
		Benchmarking	Maturity model	Data envelopment analysis	Other	Financial accounting	Management accounting	Information technology	Other (R&D, Support, etc.)	Entire company	Single tasks	Processes	Company performance	Peer (retrospective)	Best-in-class (retrospective)	Theoretical optimum
Function	Financial accounting	3	0	0	0											
	Management accounting	15	1	1	1											
	Information technology	12	3	0	1											
	Other (R&D, Support, etc.)	5	0	1	0											
	Entire company	14	0	1	0											
Activity	Single tasks	28	2	2	2	0	15	11	2	6						
	Processes	14	1	1	0	2	3	4	3	4						
	Company performance	7	1	0	0	1	0	1	1	5						
Target	Peer (retrospective)	38	2	3	1	3	14	11	6	10	28	12	4			
	Best-in-class (retrospective)	11	2	0	1	0	4	5	0	5	6	4	4			
	Theoretical optimum	0	0	0	0	0	0	0	0	0	0	0	0			
Driver	Internal (cost, ...)	33	2	3	1	1	12	10	4	12	24	11	4	31	8	0
	Technology (DEP, ...)	4	1	0	0	0	1	2	2	0	4	0	1	3	2	0
	External (Globalization,...)	12	1	0	1	2	5	4	0	3	6	5	3	10	4	0

Figure 9. Dyads based on the taxonomy for benchmarking

		Target (Method)			Target (Model)			Target (Guidelines)		
		Peer (retrospective)	Best-in-class (retrospective)	Theoretical optimum	Peer (retrospective)	Best-in-class (retrospective)	Theoretical optimum	Peer (retrospective)	Best-in-class (retrospective)	Theoretical optimum
Function	Financial accounting	0	0	0	0	0	0	3	0	0
	Management accounting	1	0	0	2	1	0	9	3	0
	Information technology	3	3	0	2	0	0	4	0	0
	Other (R&D, Support, etc.)	3	0	0	1	0	0	1	0	0
	Entire company	2	0	0	0	1	0	7	3	0

Figure 10. Triads combining function and object of the benchmarking taxonomy with the three most frequent elements of contribution type

Overall, while a benchmarking object can always be considered backward-looking (i.e., a company has established a certain process or technology) the benchmarking target against which a comparison is made, may be forward-looking. Thus, the objective of the next chapters is to motivate a new perspective on benchmarking the finance function and to propose a *new – forward-looking – approach accommodating the impact of digital technologies*. Instead of a backward-looking peer-group analysis, this "zero quartile" defines the expected (best possible) state of a benchmarking object – in this case Finance 2025 leveraging digital technologies (Figure 11).

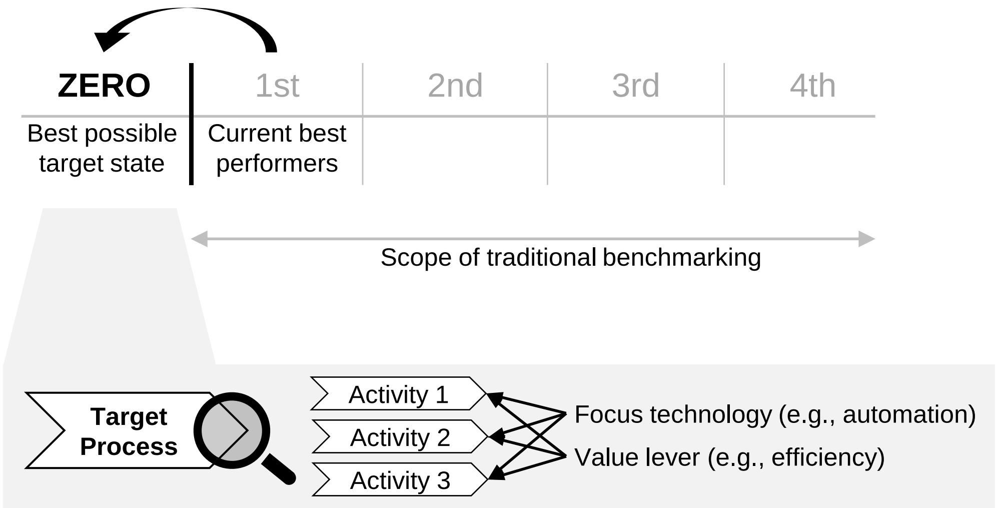


Figure 11. Overview of the zero-quartile idea

1.3. Method: The zero quartile is constructed based on the best possible target state

Surveys are a widely used method to obtain (generate) data about a larger group of people across all areas of interest in a standardized way (Fowler Jr, 2013).

1.3.1. Questionnaire

Setting up the questionnaire for the survey, the four finance processes were detailed on a process activity level. In doing so, the findings from the literature review were complemented with an auditor's process activity check list from their annual audit and the benchmarking schema of The Hackett Group (n.d.).

The value creation of the three technologies automation, analytics, and digital enterprise platform is covered by “profitable growth” and complementing targets (Couto et al., 2017): Assuming equal quality, (1) *efficiency* reflects cost savings (e.g., processing the same number of invoices with less people) or reduced time while processing the same amount of data. (2) *Growth* covers the capability to offer new services, reach out to new customers or understand customers’ behavior better. This can be driven by a broader range of information (e.g., complementing information from social media listening to identify cross-selling opportunities) or with more in-depth information (e.g., sales information not just on country, but on shop level). Complementing targets are as follows: (3) *Speed* covers the capability to improve the speed of information delivery and decision making (e.g., by reducing the number of interfaces along the information supply chain, (Davenport, 2010)). (4) *Flexibility* describes the ability to react to new, but anticipated requirements (Schelp and Winter, 2007), whereas agility additionally covers responses to unexpected requirements (Yusuf et al., 1999). (5) *Accuracy* covers the status of correct and up-to-date information (e.g., personnel data within an employee data base, (Olson, 2003)). (6) *Auditability* outlines the transparency of decision and process steps (Teppler, 2003). For example, manual activities are not recorded in practice, but will be captured automatically in case a software robot is in place. (7) *Security* comprises data security even in heterogeneous ERP/BI system architectures or the cloud (Takabi et al., 2010).

For the *main study*, the members of a manager working group hosted by the “Corporate Management” Competence Center at Darmstadt University of Technology[13] and other participants of a European event focusing on research in business economics, teaching and business practice organized by the Schmalenbach Gesellschaft were asked to answer the questionnaire. The data sample characteristics are given in Table 2[14]. Most members of the working group

---

13 This working group meets regularly to discuss trends in Finance topics. Further information can be found at: https://www.schmalenbach.org/index.php/arbeitskreise/finanz-und-rechnungswesen-steuern/digital-finance (German only).

14 The distinction between analyst and consumer managers helps to understand the intrinsic motivation of using new technologies and in the case of this survey helps to probe the full spectrum of quite diverging working styles. For more information, see, e.g., Mayer et al. (2012).

are heads of financial or management accounting departments or project managers for finance transformation.

Table 2. Zero-quartile questionnaire sample characteristics

Sector	No.	%	Working style	No.	%	Revenues [bn. $]	No.	%
Manufacturing	14	60	Analyst manager	17	74	< 2	6	26
Utilities	0	0	Consumer manager	6	26	2-9	9	39
Financial services	0	0				> 10	8	35
Telecommunications	1	5						
Others	8	35						
Total	23	100	Total	23	100	Total	23	100

1.3.2. Rasch algorithm

Developed by Rasch (1960), the algorithm was initially introduced to measure dichotomous data. The model assumes that answers to questionnaires depend on the respondents' individual ability, along with the item difficulty. It assumes that skilled individuals (companies in this case) implement more complex and more expected items than unskilled ones (Bond and Fox, 2015). For more details and mathematical descriptions of the Rasch algorithm, see Appendix A. In order to apply the Rasch algorithm to the construction of a Finance 2025 zero quartile, the original model was modified in three aspects based on Lahrmann et al. (2011). Firstly, a five-point Likert scale was employed to capture an *expressed opinion* instead of a single right or wrong in the dichotomous scale. Secondly, as the overall utility of an item at a company might not be monotonically increasing, but limited by an upper bound, the expected (*to-be) level* for an item's difficulty leveraging digital technologies was enquired in addition to the current *as-is situation*. Thus, the aim was a target state description from scratch, not only an update to existing technologies. Constituting a zero quartile Finance 2025 that is valid for the entirety of companies and stable for extreme values, *medians* were used to express the expected (to-be) level of an item for all companies.

The difference between the common expected value ($\widehat{D}_i$)[15] and the companies' individual actual value (B_{ci})[16] on the Likert scale represents the anticipated room

[15] Item index $i = 1{:}I$

[16] Company index $c = 1{:}C$

for improvement. A large positive gap, $X_{ci} = \widehat{D}_i - B_{ci}$, expresses a more difficult item regarding Finance 2025, as opposed to negative deltas which imply rather not so difficult items. Following common practice when applying the Rasch model, the data were recoded, collapsing the X-values into five distinct classes (Z) labeled descending from easy to difficult (Table 3). This re-coding is necessary to enable valid item estimates. Taking the re-coded values, the Rasch algorithm computes and iteratively refines *maximum-likelihood estimates* (*logits*).

Table 3. Re-coding prior to Rasch algorithm

X_{ci}	X_{ci}<0	$0 \leq X_{ci} < 1$	$1 \leq X_{ci} < 2$	$2 \leq X_{ci} < 3$	$X_{ci} \geq 3$
Z_{ci}	5	4	3	2	1
	Easy items				Difficult items

Thirdly, the Rasch algorithm does not yield maturity levels, but only a single ordinal scale representing the logit measure of each item and company. Thus, a complementing *cluster analysis* was performed which overcomes subjectivity in defining maturity levels. Using Ward's minimum variance method (Ward Jr, 1963)[17], items were differentiated based on their logit values separately for each of the three digital technologies. The best outcome was produced for $n = 5$ clusters assuring distinct levels based on significant differences between each level and between the individual domains. Each cluster of the final maturity model (column "Cluster", Table 4) represents a distinct stage.

For maturity level I – *digital beginners* – the logit values of items assigned to this level are between -5.42 and -2.37 across the entire sample set depending on the technology under consideration[18]. Such companies are newcomers leveraging digital technologies. Maturity level II comprises items with logit values between -2.4 and -0.15, again depending on the digital technology under consideration. In such *student companies*' technologies are used, however, no competitive advantages can be gained yet. Assessment items with a logit between -0.32 and 2.91 define maturity level III. *Digital practitioners* are confident

[17] As opposed to single linkage or group average methods, Ward's method (and later extensions) have the advantage of considering homogeneity within clusters and heterogeneity between clusters at the same time. For further information, see, e.g., Dubien and Warde (1979) or Szekely and Rizzo (2005)

[18] Note, that the lower and upper boundaries for maturity levels are not exactly the same for all digital technologies, i.e., they have differing "base difficulties." Hence, there are slight overlaps in the logit values constituting maturity levels across different digital technologies.

with digital technologies that already entered the stage of mass adoption and make good use of them. Maturity level IV covers more innovative features implemented by *digital drivers*. The logit values range from 2.36 to 4.49. Maturity level V covers *digital masters*. Such companies are strongly committed to the most desired but also most rarely used digital technologies. They have successfully implemented such technologies or plan to do so soon.

Independent estimates for the companies and item parameters are provided by the Rasch model as well. Hence, their validity can be tested with the help of the integrated fit statistics. Variations around the organization’s own maturity level are represented by the sensitivity of the *infit* statistic (see Appendix A), whereas *outfit* is more sensitive to items far from the company’s maturity level (outliers). Both values should have a value of around one. A value greater than two is regarded as critical and should be considered for removal (Linacre, 2002).

1.4. Results: Four imperatives help to prioritize technologies and reach the zero quartile

1.4.1. Descriptive results

The initial Rasch calculation showed poor infit and outfit statistics for items B.7 regarding automation, D.5 and D.6 regarding analytics, and D.3 regarding the DEP. Those items were removed from the data sample. The subsequent estimation performed well and none of the items exceeded the critical fit threshold of 2.0 (Linacre, 2002). Table 4 presents the final results of the logit measurement (column 4) and the infit/outfit statistics (columns 6, 7) on a process activity level (column 1) and for each digital technology (column 3). Based on the logit value (column 4), Table 4 also presents the cluster assigned to each technology and process activity (column 5).

Being the result of the cluster analysis, the distinct maturity levels I-V show an ascending ranking of the digital initiatives’ priority for all process activities. The clusters IV “digital driver” and V “digital master” indicate the greatest gap between the median for the expected to-be ratings and the as-is statuses. These process activities have the greatest potential for optimization and differentiation against peers. Finally, Table 4, shows the expected type of value creation per process activity level (column 2). The bold lines highlight main findings with logit values greater than 3.5 and a ranking of each process activity on maturity levels IV or V.

Starting with the **O2C process**, *automation* was the primary digital technology driving Finance 2025, since three of the eight O2C process activities have logit values greater than 3.5 and are in the clusters of digital masters or digital drivers. Automation is used for process activity A.2 “maintain master data” (logit: 4.58) to increase the accuracy of information in master data files. Aiming to facilitate growth, another suitable O2C process activity for automation is “manage customer requests and inquiries” (A.7, 4,58). Improving efficiency, the process activity “perform revenue assurance activities” (A.8, 3,67) should be supported by automation to align disputes between sales and O2C. A logit value of 6.61 for “credit authorization” (A.1) is the highest value of the entire survey. Applying *analytics,* credit authorization is a strong use case to drive growth.

Considering the **P2P process**, *analytics* was the primary digital technology towards Finance 2025. Two of the eight P2P process activities have logit values greater than 3.5 and are within the cluster of a digital driver or master. Analytics has its greatest value in “invoice processing” (B.4) with a logit of 5.27 driving growth. Addressing efficiency, analytics was also named for “manage requisition and fulfilment” (B.1, 5.27). In comparison to the O2C process, the best logit for automation just received a value of 3.03 for “vendor master data management” (B2) – not relevant for further examination.

The **R2R process** showed no logit values greater than 3.5. Thus, all R2R process activities indicate just a minor potential for improvements by leveraging digital technologies. Considering “revenue and cost accounting” (C.6), the best value for automation is 0.06. The other logits for automation are all negative and none of the activities was assigned to a cluster greater than III. Regarding analytics the best logit value in the R2R data set is negative (“intangible asset accounting”, C.5, -0.29). Only a DEP showed potential use cases, the best logits were driving the speed of “period close/ consolidation” (C.3) and “financial reporting” (C.4).

Focusing on speed, the Rasch output for the **EPM process** proposes a strong application of the *DEP* for both “action planning” (D.5) and “business decision support” (D.6, 5.28). *Analytics* is the driver for “forecasting” (D.3) aiming to enhance a company’s growth and “business performance reporting and analysis” (D.4). Automation has its highest logit value in “budgeting” (D.2, 0.45) corresponding to the maturity level of a digital practitioner. Hence, it is not perceived as a strong value driver in comparison to analytics and the DEP.

Table 4. Results of the Rasch algorithm and cluster analysis

Process activity		Value	Technology	Logit	Cluster	Infit	Outfit
A. Order-to-Cash							
A.1	Authorize Credit: Investigate customers to determine if sales on credit is appropriate.	Growth	Automation	-1.85	II	0.53	0.60
			Analytics	6.51	V	0.57	0.13
			DEP	-3.74	I	1.42	1.41
A.2	Maintain master data: The customer master data structure is designed and maintained according to master data policies and procedures.	Accuracy	Automation	4.58	V	0.54	0.15
			Analytics	-4.32	I	0.51	0.53
			DEP	3.33	IV	1.06	0.62
A.3	Invoice customer: The invoice is generated and distributed to the customer, followed by receivable entry posting, and recording revenue.	Speed	Automation	2.77	IV	1.03	0.68
			Analytics	-3.89	I	0.50	0.50
			DEP	-1.32	II	0.93	0.90
A.4	Maintain accounts receivable ledger and apply cash: Receive and deposit customer payments. Post and reconcile accounts receivables (AR) activity to GL (general ledger).	Efficiency	Automation	3.33	IV	0.64	0.41
			Analytics	-4.78	I	0.70	0.76
			DEP	2.64	IV	1.93	1.17
A.5	Manage and process collections: Analyze customer account balances and manage collections (e.g., proactive customer contacts, distribute dunning letters, etc.) according to treatment strategies. Write-off uncollectible balances.	Efficiency	Automation	2.77	IV	0.46	0.32
			Analytics	1.74	III	1.24	1.23
			DEP	-0.46	II	1.11	1.05
A.6	Manage and process disputes and deductions: Perform root cause analysis for dispute and deductions. If applicable, prepare chargeback invoices and process dispute/ deduction adjustments and write-offs.	Growth	Automation	-5.06	I	0.60	0.58
			Analytics	2.40	III	0.87	0.88
			DEP	-3.13	I	0.76	0.68
A.7	Manage customer requests and inquiries: Receive, record, and resolve/respond to cus-tomer requests and inquiries.	Growth	Automation	4.58	V	0.62	0.50
			Analytics	3.28	III	0.79	0.68
			DEP	0,34	III	0.64	0.57
A.8	Perform revenue assurance activities: Monitor revenue from booked order through cash collection and implement preventative measures in case of leakage points.	Efficiency	Automation	3.67	IV	1.13	0.45
			Analytics	-1.60	II	0.82	0.84
			DEP	-0.31	II	0.70	0.66

Process activity		Value	Technology	Logit	Cluster	Infit	Outfit
B. Purchase-to-Pay							
B.1	Manage requisition and fulfillment: Determine method: purchase card, purchase order or via contract and procure.	Efficiency	Automation	0,19	III	1.09	1.11
			Analytics	4,49	IV	1.63	0.93
			DEP	-3,74	I	0.69	0.65
B.2	Vendor master data management: Design/ maintain master data according to enterprise wide policies and procedures. Delete inactive vendors.	Accuracy	Automation	3.03	IV	0.82	0.44
			Analytics	-4.62	I	0.70	0.76
			DEP	2.11	IV	1.84	1.18
B.3	Manage inbound documents: Receive, scan, and archive documents. Reject inappropriate invoices and return them back to the vendor.	Efficiency	Automation	2.32	IV	1.39	1.15
			Analytics	-1.02	II	0.58	0.57
			DEP	2.36	IV	1.66	1.08
B.4	Invoice processing: Receive goods receipt, approve, and schedule payment.	Efficiency	Automation	1.44	III	1.36	1.03
			Analytics	5.27	IV	1.58	0.69
			DEP	-1.85	II	0.78	0.81
B.5	Process payments: Includes the generation of scheduled, automated payments, and the processing of rush payments.	Efficiency	Automation	2.54	IV	1.18	0.73
			Analytics	-2.03	II	0.81	0.82
			DEP	-2.37	I	1.49	1.54
B.6	Manage vendor dispute: Receive, record, process, and resolve vendor inquiries.	Efficiency	Automation	1.28	III	0.79	1.12
			Analytics	2.91	III	0.66	0.60
			DEP	-0.61	II	1.47	1.39
B.7	Reconciliation: Analyze and reconcile general ledger (GL) accounts and perform period end closing.	Efficiency	Automation	---	---	---	---
			Analytics	-4.47	I	0.70	0.62
			DEP	-0.31	II	1.38	1.25
B.8	Manage purchase card program: Select purchase card vendors and issue purchase card to employees.	Efficiency	Automation	-1.59	II	0.85	1.01
			Analytics	-0.88	II	1.18	1.15
			DEP	-5.42	I	0.77	1.96
C. Record-to-Report							
C.1	General Accounting: Record transactions according to journal entry policies.	Efficiency	Automation	-1.07	II	1.66	1.63
			Analytics	-3.89	I	1.54	1.54
			DEP	0.70	III	0.85	0.73
C.2	Intercompany accounting: Perform intercompany reconciliation, identify, and resolve issues and finalize intercompany balances.	Efficiency	Automation	-0.94	II	0.48	0.53
			Analytics	-0.88	II	1.82	1.80
			DEP	1.26	III	0.76	0.63
C.3	Period close/consolidation: Execute period-end accounting entries, close the general ledger and consolidate results. Approve financial results and perform post-closing activities.	Speed	Automation	-2.40	II	0.78	0.80
			Analytics	-0.59	II	1.07	1.04
			DEP	2.36	IV	0.72	0.48
C.4	Financial reporting: Prepare, distribute, review, and finalize financial reports.	Speed	Automation	-0.32	III	1.00	1.18
			Analytics	-1.17	II	1.15	1.13
			DEP	2.64	IV	0.89	0.51
C.5	Intangible asset accounting: Measure, record, and value intangible assets	Efficiency	Automation	-1.72	II	0.85	0,98
			Analytics	-0.29	II	0.96	0,95
			DEP	-2.88	I	0.36	0,34
C.6	Revenue and cost accounting: Identify performance obligetions, determine transaction price, recognize, close, and report revenue and costs.	Growth	Automation	0.06	III	0.91	0.89
			Analytics	-1.60	II	0.96	0.94
			DEP	1.26	III	0.36	0.50
C.7	Joint venture accounting and others: Accounting activities in relation to joint ventures and other business entities.	Accuracy	Automation	-4.89	I	0.97	0.99
			Analytics	-3.89	I	1.03	1.03
			DEP	-0.31	II	0.69	0.74

Process activity		Value	Technology	Logit	Cluster	Infit	Outfit
D. Enterprise Performance Management							
D.1	Strategic (business) planning support: Determine key business drivers and measures of success and create a strategic plan.	Growth	Automation	-4.22	I	1.69	1.71
			Analytics	2.74	III	1.43	1.39
			DEP	-2.50	I	1.71	1.72
D.2	Budgeting: Preparation of budgets (revenue, manpower, Opex, Capex etc.)	Growth	Automation	0.45	III	0.86	0.91
			Analytics	1.90	III	0.92	0.87
			DEP	-0.46	II	0.97	0.94
D.3	Forecasting: Develop, review, consolidate forecasts and track actions to close gaps between performance and budget/target.	Efficiency	Automation	-2.26	II	0.99	1.04
			Analytics	4.49	IV	1.84	1.26
			DEP	-0.15	II	1.21	1.14
D.4	Business performance reporting & analysis: Measure, report and analyze business performance.	Speed	Automation	0.06	III	1.35	1.34
			Analytics	4.19	IV	1.75	1.23
			DEP	---	---	---	---
D.5	Action planning: Perform root cause analysis, identify, and report required corrective actions.	Speed	Automation	-1.99	II	1.01	1.06
			Analytics	---	---	---	---
			DEP	5.28	V	0.74	0.13
D.6	Business decision support: Conduct business analysis to support decision making.	Growth	Automation	-4.72	I	1.06	1.04
			Analytics	---	---	---	---
			DEP	5.28	V	0.74	0.13

1.4.2. Synthesis of the results

Emphasizing iterative "build" and "evaluate" activities (Peffers et al., 2007), this section synthesizes the previous section in two sequential steps. (1) Based on the logit values (quantitative results obtained from the Rasch measurement) and complementing qualitative comments from the interviewees during the survey, in a first step *eight design guidelines* helping companies to develop themselves towards Finance 2025 by leveraging digital technologies are presented. (2) In a second step, the practical relevance of the proposed design guidelines is discussed with both a business expert of a leading automotive company (2017 revenue: 230 bn. EUR; 650,000 employees) and the head of Digital Finance of a chemical company (2017, revenue: 65 bn EUR, people: 115,490) as well as three experts from a benchmarking consultancy. With their input, finally *four imperatives* are synthesized (headlines to the next paragraphs) that may help companies accelerate their transformation journey towards Finance 2025.

Bring repetitive O2C tasks such as maintain master data on a robot, so that there is more time for growth initiatives like authorizing credit by analytics on big data and managing customer requests with the help of chatbots

Employing analytics to enable growth via *“credit authorization”* (*A.1)* received the highest logit of all items examined with 6.51. An interviewee from a pharmaceutical company highlighted this item with the emergence of the Arab Spring until 2016. Their credit rating classified hospitals governed by the state department and countries in this region in general still as a safe business environment, whereas the analysis of online news and reports from social media listening gave a first picture of the actual risk. At the end, most of the hospitals did not pay their bills for a longer period as the state department froze their accounts. Thus, the interviewed manager continued that better information about potential customers’ creditworthiness is a large lever for improvements in the O2C process. Especially data (even unstructured) from *social media* should be more in focus as they could complement a company’s information and decision basis.

Another interviewee, in turn, doubted that analytics is a digital technology in focus. Relevant data might be available, however, only from expensive sources and not worthwhile to justify the investment in manpower and tools. The expert from the automotive company modified this statement as he outlined that gaining insights through analytics depends on the *quality of the data*. “Analytics is fancier than automation, however, the best algorithm will not produce any useful insights, if the (master) data are of poor quality.” As a first design guideline, it is argued that *better decisions provided by analytics* (on big data, including social media listening) have a strong potential:

Design guideline 1: Boosting a company’s growth, credit managers should go beyond traditional credit ratings and include analytics on big data, especially from social media listening.

To improve the accuracy of *“maintain master data”* (*A.2),* digital master companies leverage rule-based automation. One interviewee stated that *software robots* (RPA) helped them to set up and maintain customer master data and product hierarchies, especially to align them between finance, IT and other domains. These activities should be automated completely to eliminate human errors and help accountants focus on more value-adding tasks such as managing and processing disputes and deductions (A.6). A chemical group that answered the survey already has a finance software robot in place. Supporting contract agreements with customers, the robot receives an Excel sheet via Outlook that requests up-to-date information on customer contracts. The robot then searches for the required data in the SAP ERP system, copies and pastes them into a Microsoft Excel template, and sends it to the end user.

Design guideline 2: To improve accuracy, RPA can help to align master data such as customer information and product hierarchies across different domains.

Automation has a great potential to increase growth in *“managing customer request and inquiries” A.7. Chatbots* are a future automation technology that can answer customer inquiries regarding bills or product information. To handle customer requests of low complexity, in another company cognitive automation is already in place. In Finance 2025, such chatbots will become ubiquitous helpers, as the underlying technologies and methods like natural language processing will progress.

Design guideline 3: Receiving, resolving, and responding to customer requests and inquiries of lower complexity should be transferred from humans to chatbots.

Summarizing the findings regarding the O2C process, these first three initiatives of digitalization are a great chance for finance managers to improve their role as a *business partner* within their companies instead of just being number crunchers. In other words, they may become more influential by bringing value to the table. Thus, beyond years of cost cutting initiatives driving efficiency, a shift towards growth as an equal value creation target for finance besides efficiency may be appropriate.

Handle invoices without a purchase order by cognitive automation and drive working capital efficiency by leveraging analytics in managing requisitions and fulfilment

Using rule-based automation within the P2P process is most often just an initiative of a digital practitioner. The results show a focus on *cognitive automation* and *analytics* which is in line with the feedback from the interviewed benchmarking experts. They agreed that applying rule-based automation is already common practice for the P2P process, but an integration of cognitive automation and analytics will drive efficiency and growth.

“Management of requisition and fulfilment” (B.1) is a first digital use case in the P2P process towards Finance 2025. Analytics should help to continuously improve the working capital efficiency by optimizing delivery volumes and schedules, contract negotiations for framework agreements, and the handling of semi-finished or intermediate products on stock. One company analyzed the retention time of parts and other materials from vendors within their stocks. The results from their *pattern matching* indicated that parts and other materials ordered automatically are three days less in the company’s stock in comparison to the same items when ordered manually. Their explanation for this was that stock masters added a personal three-day buffer on top of the “typical” retention time in case something unexpected would appear that he or she might be made responsible for.

Design guideline 4: Pattern matching should be used to optimize working capital efficiency. This covers the purchase order procedure per se as well as the price and quantity of each order to optimize both stock volumes and retention time of parts and other materials.

“Invoice processing” (B.4) with a logit of 5.27 received an outstanding result in the survey. The rationale was to optimize a company’s payment behavior in a more fact-driven manner from goods received to approval, scheduling, and final payment. A case example of the chemical company are *invoices without a purchase order* („FI-Invoices – non-purchase order items“) where *machine learning* is applied to route documents to the right approvers and suggest cost centers at

the first go (for more details see chapter 5). With about five million invoices, whereof 10 percent come without a purchase order, and a proposed accuracy of the finance robot of around 75 percent, the work load of three full-time equivalents (FTE) can be taken over by a machine, thus, the amortization takes less than one year. Moreover, accountants can focus on more complex issues of invoice processing or value-adding activities of the P2P process such as managing vendor disputes (B.6).

Design guideline 5: Apply cognitive automation to optimize the handling of invoices without a purchase order.

Record-to-report will benefit from a digital enterprise platform, but start your journey with the upstream finance processes O2C and P2P

The survey did not yield any logit values above 4.0 for the entire set of R2R process activities. Thus, the R2R process is already quite advanced regarding Finance 2025, but it could benefit from improvements in other processes as well. This finding was addressed in the feedback discussions and three of the respondents stated that R2R activities currently have no crucial roadblocks and are therefore less important. Hence, related activities are lower on the priority list of the accounting department.

Rethink enterprise performance management with predictive analytics in forecasting and real-time business reporting and analysis even on ERP data. A Digital Enterprise Platform is the prerequisite for doing so

One of the most popular EPM activities *“business performance reporting and analysis” (D4)* such as sales, spend, performance or variance reports, but even Balanced Scorecards or Value-Driver Trees is driven by analytics. Conducting real-time analyses is key as one interviewee stated: “It will improve the acceptance of finance within the business when they deliver most important KPIs in real time.” Other comments named mobile solutions for approvals, monitoring, more user-centric dynamic dashboards, a data visualization for manager self service, and, finally, better collaboration with other functions to discuss the findings of the reports. In doing so, one automation use case is in a variance report which states differing actuals from the target. The current values are obtained from an ERP system by an accountant and transferred to an Excel file, which ultimately is sent via MS Outlook to the respective management accountant. The process is strictly rule-based, thus, the company leveraged RPA to automate the report creation.

Design guideline 6: As more and more standard reports and analyses are automated, in Finance 2025, user-centric, self-service business performance reporting and analysis is a standard.

A prerequisite for such a new kind of business performance reporting and analysis is a new IT backbone. As one interviewee stated, it is time to rethink the role of ERP and move away from a purely legal (historical) focus towards a forward-looking perspective in Finance 2025. This became obvious examining *"action planning" (D.5)* and *"business decision support" (D.6)*. Another interviewee confirmed, "a new digital backbone must become the go-to source for decision support." Especially ad-hoc reporting such as root-cause analyses to identify required corrective actions (D.5) need a drill-through directly to the ERP data and the possibility to analyze real-time in the ERP. Furthermore, an underlying in-memory engine and respective business application, for example SAP's S/4HANA Finance, provide the capability to conduct a better decision support (D.6) on line-item level. Thus, a *DEP* will help finance professionals perform supporting analyses on an event-driven basis (off the standard track). That covers contract analyses, customer evaluations or fluctuation reports for the business units.

Design guideline 7: Establish a DEP as the go-to source for decision support, ad-hoc and event-driven management support analyses. They are performed in real time with the help of a DEP.

Considering the results regarding *"forecasting" (D.3)* and taking the information received from the interviewees into account, an eighth and final design guideline was derived. *Predictive analytics* algorithms provide a value-add to create more accurate forecasts, but also to slim down forecasting processes. Where previously several iterations across all hierarchies were necessary, towards Finance 2025, analytics will provide a first top-down estimate upon which business units can build. Such a prototype recently went live in the finance department of one of the interview partners to improve quarterly cash flow forecasting (for more details see chapter 6). Not only internal (historical) data, but also leading external indicators like business confidence or price indices are fed into an ensemble of forecasting algorithms to provide the best possible estimate as a starting point. Overall, this approach adds a more fact-driven side-car to the already established experience-based forecast.

Design guideline 8: Develop, review, and consolidate forecasts by leveraging predictive analytics. Combine company-internal and external data sources as a starting point for faster and more accurate forecasting.

Summarizing these findings, EPM will evolve from an expense-controlling, spreadsheet-driven management accounting to a more powerful function covering both *real-time* business performance reporting and analysis (D.4), action planning (D.5), and business decision support (D.6) as well as *predictive analytics* in forecasting (D.3). A *digital enterprise platform* is the technology backbone and central source for such a new decision support.

2. Evaluating digitalization efforts goes beyond monetary benefits

2.1. Excursus: Many digitalization projects lack a traditional business case

Companies most often resort to business cases to gauge the financial impact of a project. Generally, it helps to identify the value potential and amortization period of an investment. Key elements of a business case are the cash flows that will be generated after successful completion of the project, the associated costs, further potential benefits and risks, as well as decision alternatives and their financial impact (Schmidt, 2009; Messner, 2013).

With the simultaneous emergence of the trends and technologies mentioned in the previous chapters, companies face the decision in which technology to invest first. On the one hand, budgets are often only granted if a valid business case is provided. On the other hand, digitalization initiatives do not always yield directly tangible and measurable benefits (Colas et al., 2014). As a result, weighing the benefits and costs of a new technology is not always possible in the traditional way. Hence, determining the (direct and indirect) value and costs of technology has been and will remain a relevant issue for the research and practitioner communities alike (Schryen, 2013). Although there have been several approaches over the past decades, they mostly neglect the intangible aspects. For instance, Post (1992) developed a model to estimate the value of group decision support systems based on the levers of efficiency and effectiveness. Wegen and Hoog (1996) combined activity-based costing, the information commodity approach and graph modelling to determine the object that should be evaluated, define what value is, and define in which way value can be measured. Another example are Nielsen and Persson (2017), who provided a method to improve the calculation of business cases for IS. Shang and Seddon (2002) developed a method that helps to assess the benefits of enterprise systems based on a five-dimensional framework that covers benefits like efficiency, effectiveness and the social aspects of an enterprise system. Addressing one of the three technologies in the focus of this work, Seddon et al. (2017) propose a “business analytics success model” – a combination of process model and variance model – that includes organizational aspects like leadership, analytics orientation, and enabling technologies.

None of the models mentioned above resembles a traditional business case, which is further evidence that such an approach is not suitable for digitalization projects. Nevertheless, some of the models neglect intangible benefits (Post, 1992), some are a bit too cumbersome for an application in practice (Seddon et al., 2017), and some do not consider the extend of organizational change that may be associated with digital transformation (Wegen and Hoog, 1996). Thus, as an extension of the zero-quartile results, in the following section a manageable model for the analysis of the value levers of a digital technology implementation will be proposed. While the proposed model leverages the scientific foundation laid by the aforementioned models, it was co-developed in a workshop with practitioners to cover their requirements more thoroughly.

2.2. Evaluation: Efficiency, effectiveness, and experience help to assess a value case for the zero quartile

Discussing the results of the zero-quartile study (section 1.4) with a couple of practitioners, they voiced the need for a reasonable set of value levers to justify a DEP implementation. Although the value levers of the zero-quartile study served as a good basis, they also lacked a category for intangible benefits (cf. the other models described in section 2.1).

Therefore, following the consortium research approach (Österle and Otto, 2010), a model was created in a workshop with a manager focus group. This close cooperation between researchers and practice – where all participants agree on a common definition of objectives, mutually exchange their knowledge, and iteratively design and test the artifact – has been demonstrated to particularly support DSR in IS. Moreover, this format provides direct suggestions and feedback between researchers and practice in a personal atmosphere and has the advantage that participants all share a high interest in the topic. The focus group for this model consisted of 25 mostly finance and finance-IT managers from ten different companies, who met three times in 2018 to discuss topics around the DEP and SAP's HANA platform in particular (Table 5)[19].

Table 5. Benefit circle focus group characteristics

Sector	No.	%	Department	No.	%	Revenues [bn. $]	No.	%
Manufacturing	14	56	Finance	18	72	< 2	4	16
Utilities	5	20	Finance IT	5	20	2-9	3	12
Financial services	0	0	Other	2	8	> 10	18	72
Telecommunications	2	8						
Other	4	16						
Total	25	100	Total	25	100	Total	25	100

In the course of the three workshops, the model was iteratively refined and in its final state shows a couple of differences to the zero-quartile value levers: Efficiency was detailed as described below. Growth was replaced by business insights as this term better covers the aspect of business partnering and providing helpful information that enable growth. Auditability was dropped due to its secondary relevance. Improvements in auditability were rather considered a positive side-effect of increased accuracy and speed. Finally, the new dimension experience was added. In the following, the model (Figure 12) will be described in more detail. Note, that some of the levers are very similar to those described in section 1.3.1.

Assuming equal quality, *efficiency* covers the monetary effects in terms of full-time equivalents (FTE) and costs (Couto et al., 2017). (1) FTE addresses the number of full-time employees that are working on the respective process. 2) Cost describes the discovery of currently hidden OPEX/CAPEX (short-term/long-term) balance sheet impacts or cost cutting opportunities[20].

Those effects of a technology implementation that are not directly tangible or measurable are covered by e*ffectiveness*. This level of IS value creation can be divided into five sub-levels: (1) *Business insights* covers the degree to which the

---

[19] The number is in line with Griffin and Hauser (1993), who argue that 25-30 users are sufficient to gather at least 90% of all user needs.

[20] Budget was discussed as a third category of efficiency but dropped in the second iteration. It comprises the funds of the finance department that are invested for digitalization (Secrett, 2012).

reports, analyses, and KPIs fit to the requirements of the information recipients and by which they are improved due to the use of digital technologies. (2) *Speed* covers the capability to improve the time it takes to collect and distribute information and finally for decision making, e.g., by implementing in-memory applications (Davenport, 2010). (3) *Flexibility* describes the ability to react to new, but mainly anticipated requirements (Bock et al., 2012). (4) *Accuracy* covers the provision of correct and up-to-date information (e.g., address data within the ERP) (Belhiah et al., 2015). (5) *Security* comprises data reliability even in heterogeneous ERP/BI architectures or the cloud (Takabi et al., 2010).

Experience comprises the (intangible) learning effect the company has (the employees have) by using the digital technologies. (1) The use of the digital technology paves the way for economies of scale, which in turn, lead to cost reduction, or a reduction of FTE and thus *enable efficiency*. (2) *Enabling effectiveness* addresses how the use of the digital technology shapes the organization for higher customer satisfaction by enabling improved business insights, speed, flexibility, accuracy or security.

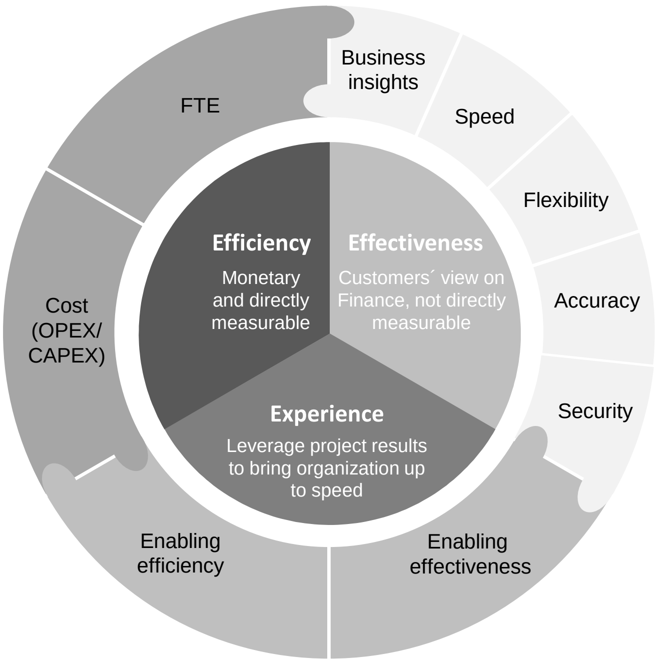


Figure 12. Benefit circle for digital technologies

2.3. Discussion: The zero quartile of digital finance is changing over time

Taking the impact of technologies on the finance function as a case example, the objective of this part was to examine a forward-looking benchmarking. From an interdisciplinary literature review, 29 assessment items for Finance 2025 were derived – the expected state of the benchmarking object. Data for the Rasch algorithm was obtained from a manager-working-group survey.

These results answer the question about which digital technologies are most relevant for the future finance function. Additionally, eight design guidelines were developed that help companies develop themselves towards Finance 2025 by leveraging digital technologies. In comparison to the latest practitioner literature at the intersection of finance, benchmarking, and digital technologies such as Axson (2015), Lucas (2016), Plaschke et al. (2018), a *rigorous literature review* provided the foundation for this work.

For practice, this benchmarking approach is more forward-looking and goes beyond the traditional ways (The Hackett Group, n.d.). Thus, it should better help companies to develop themselves towards Finance 2025 than a backward-looking peer comparison which is currently state of the art.

For research purposes, the approach at hand is *more comprehensive* than other publications like Blumenberg (2004) or Marx et al. (2012). (1) 29 activities of the four accounting core processes O2C, R2R, P2P, and EPM were covered. Furthermore, there seems to be no scholarly work on the R2R process, which is a gap addressed by this work as well. (2) The current body of knowledge focuses purely on efficiency and effectiveness by cost-cutting, which represents only one dimension of the *multidimensional target system* (benefit circle, Figure 12) proposed. (3) Finally, the relevance of outcomes was ensured by incorporating findings from a survey among a manager working group based on their *business perspective* on digital technologies. (4) Last, but not least, the Rasch algorithm serves as a rigorous starting point for all kinds of benchmarking beyond finance such as procurement, production, sales as well as human resources or logistics.

However, there are avenues for future research. Due to the relatively small data sample, four items showed fit statistics larger than two and had to be dropped in order not to degrade the Rasch measurement. Hence, a *broader sample of managers* with a more even balance of male and female participants should also ensure the further generalizability of findings. Furthermore, the same relevance for each of the four accounting processes and their process activities was assumed. However, the O2C and P2P processes might be more important than R2R and EPM to keep a company's operating model running. Moreover, *further finance sub-functions like tax and treasury* should be considered in the future.

Another avenue for research is to quantify the *impact of the zero-quartile findings*. The design guidelines give companies a good direction towards Finance 2025. However, a subsequent evaluation should indicate, what the required implementation time and achievable monetary benefit of these findings are. Case studies may be best suited for doing so. For example, recommending automation in the O2C process, accountants replaced by machines may help to quantify the impact. However, targets such as flexibility by using a DEP or better insights by applying predictive analytics are more complicated to measure. The benefit circle can be a first step towards such a quantification.

Overall, the results should be interpreted with caution and cannot be generalized per se due to differences in the way industries work and, thus, the way finance functions have to be set up. The rapid pace of digitization may lead to unforeseeable developments in the future. However, examining the digital finance roadmaps of managers in the Schmalenbach working group, the results should nevertheless be valid at least for the next three to five years. Hence, the proposed zero quartile should not be considered as cast in stone, but rather as the pillars of an evolving target state. Therefore, there is a need for a regular update of the findings. Finally, a continuous benchmarking for companies over time should also help to address today's far reaching changes.

Part II: Machine learning as a corner stone of digital finance

Introduction

The use of digital technologies, especially to handle large volumes of data, can lead to major improvements in the respective processes (Martin and Conte, 2012; Blanc and Setzer, 2015). Looking at a list of recent technological advances *machine learning and analytics*[21] (in the following ML&A) play an important role. The distinction between these terms and related areas like business intelligence, advanced analytics, and artificial intelligence is not always sharp. Especially in practice, the terms are often used interchangeably. Generally, however, they always aim to apply methods from statistics and computer science to enable better decision making. Thus, the finance department should benefit from an implementation in several areas.

Still, machine learning and analytics *adoption in the finance department* are rather poor and only gradually increasing despite clear statements to do so in the past (Halper, 2014). Some of the barriers preventing users from adoption are still the same as they were ten years ago, for example, high complexity, poor data quality, or lack of expertise (Eckerson, 2007). With respect to technology adoption, the Technology Acceptance Model (TAM, (Davis, 1985; Davis et al., 1989)) is one of the most influential models in the IS discipline. However, it was also subject to criticism based on the large body of extensions it entailed without clear guidance. Nevertheless, its main components perceived usefulness and perceived ease of use remain valid (Benbasat and Barki, 2007). In the field of ML&A, the motivation for adoption has already been studied ("rationale for business analytics" (Holsapple et al., 2014)) ranging from achieving a competitive advantage to knowledge production, but only on a qualitative basis. Further determinants of use may include the types of data available or the methods and algorithms employed (Bhimani, 2003). However, there is no comprehensive study on the drivers of adoption and the current state in the finance department.

[21] See chapter 1.1.2 for definitions.

3. Machine learning and analytics currently see a surge of interest from practice

Despite the potentials of ML&A briefly described in section 1.1.2, the actual use varies substantially among the different functions of a company. According to a 2014 study, 64% of respondents said they already use predictive analytics in marketing, with an additional 24% saying they will use it within the next three years (Halper, 2014). The finance department, on the other hand, was only mentioned by 39% and 26%, respectively. While this is still considerably more than, e.g., human resources (17% and 22%), it is noteworthy that the number-crunching finance department is apparently not the first stop for advanced statistical methods. The demand, however, for ML&A adoption is clear when looking at practitioner literature. For example, in its 2017 CFO Agenda, the Hackett Group states that the finance department needs to step in and support the company strategy facing more constraints on funding and headcount and, secondly, provide the organization with more and better information including insights from analytics (Essaides et al., 2017). In order to further investigate this discrepancy, a literature review was conducted at the intersection of ML&A and finance, followed by a survey on ML&A adoption to build a model of relevant drivers of adoption.

3.1. Literature review: Machine learning and analytics adoption is still rather poor and research for finance is scarce

3.1.1. Search strategy

As a first step of the literature review, a *journal search* in leading journals followed by a backward and forward search to look for articles cited in the identified papers (backward) and newer articles citing the identified papers (forward) was performed (Webster and Watson, 2002). Since the focus of this section is at the intersection of statistics and operations research on the one hand and accounting on the other hand, literature was searched “from both ends”.

Regarding accounting, the top ten accounting journals[22] were chosen in line with Nitzl (2016), complemented with the top IS journals from the Senior Scolar's Basket of Journals[23] and with AIS conferences[24]. The search terms used were "machine learning" and "analytics"[25]. With respect to statistics and operations research, five journals[26] from the list of top journals in scimago were selected based on their scope and the search terms "management (or managerial) accounting" and "financial accounting" were used.

As a second step of the literature search, the scope was broadened to a *comprehensive database search* in ScienceDirect, EBSCOhost, JSTOR, Wiley Online Library, and Google Scholar combining different search terms according to the *citation pearl growing* approach (Rowley and Slack, 2004). The initial search terms were "finance" and "machine learning" or "analytics" and then the search was widened to include different accounting, IS, and planning terms in the finance context and machine learning types and various forecast terms in the ML&A context (see Figure 13 for the search terms used).

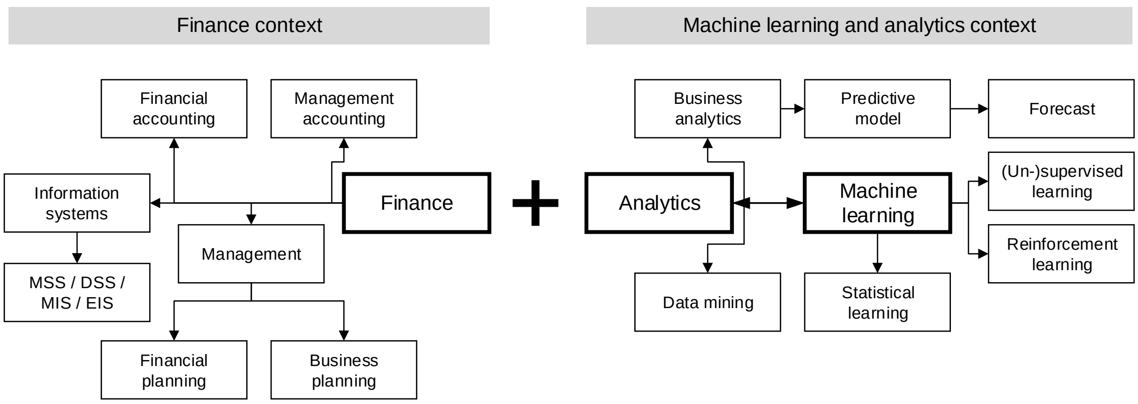


Figure 13. Citation pearl growing search terms

[22] Journal of Accounting and Economics; Journal of Accounting Research; The Accounting Review; Management Accounting Research; Journal of Management Accounting Research; Contemporary Accounting Research; Behavioral Research in Accounting; Accounting, Auditing & Accountability Journal; Accounting and Business Research; Accounting, Organizations and Society.

[23] European Journal of Information Systems; Information Systems Research; Information Systems Journal; Journal of the Association for Information Systems; Journal of Information Technology; Journal of Management Information Systems; Journal of Strategic Information Systems; MIS Quarterly.

[24] Americas' Conference on Information Systems; European Conference on Information Systems; International Conference on Information Systems; Pacific and Asia Conference on Information Systems.

[25] The term "adoption" was not included in the search terms as the search with the machine learning and analytics context already resulted in not too many hits. Additionally, this literature review is meant to serve as a basis for chapters 5 and 6 as well.

[26] International Journal of Forecasting; Journal of Forecasting; Operations Research; European Journal of Operational Research; International Journal of Production Economics

Due to the importance of the field to practitioners, a number of accounting organizations and consulting agencies have published surveys and point-of-view reports. In this work, however, the focus is on academic, peer reviewed, literature. While this may omit a number of recent developments, it should be justified for a literature review. For an overview of practitioner statements regarding management accounting see, for example, Nielsen (2017).

3.1.2. Framework for systemization

Based on the results of the literature search, a framework is proposed to classify the existing applications of ML&A in financial and management accounting. With the help of this framework, "hotspots" of current interest and potential hotspots of future interest are identified.

The framework has two dimensions: first, the accounting activities and second, the rationale for using ML&A with respect to a specific accounting activity. *Accounting activities* are the tasks that an accountant in financial or management accounting performs on a regular basis. Although, the scope of financial accounting is not the same for all companies, there is some common denominator in companies of a certain size. A list of three activities in financial accounting presented by Horngren et al. (2002) is followed: (1) Bookkeeping (incl. accounts payables, receivables, and credit management), (2) statutory reporting, and (3) consolidation. Equally, management accounting can be set up differently in an organization, but four core tasks are common as well, as described by Blocher et al. (2010) and Brands and Holtzblatt (2015): (1) Strategic (cost) management, (2) performance measurement, (3) planning and decision making, and (4) support in financial statement preparation.

Rationale for ML&A is the reason why they are applied in this specific situation. Generally, there are numerous possible nuances, however, a list of six endogenous elements summarized by Holsapple et al. (2014) will be used: (1) Achieving a competitive advantage, (2) support of strategic and tactical goals, (3) better organizational performance, (4) better decision outcomes, (5) knowledge production, and (6) obtaining value from data.

3.1.3. Gap analysis

Comparing the ML&A coverage in financial accounting and management accounting, it is clear that the latter has attracted more attention. While this is partly due to the type of work in each of the two domains, it should not lead to an exclusion of financial accounting from consideration. In the following, three categories of increasing interest are proposed – less relevant (white shading), relevant (light grey shading), and highly relevant (middle grey shading). Additionally, some of the applications of ML&A in each of these categories are highlighted. The categorization is based on the nature of the activity and the general potential for statistical methods as well as current literature coverage. Figure 14 shows the results for financial accounting with two highly relevant areas in bookkeeping, a couple of relevant and some less relevant areas.

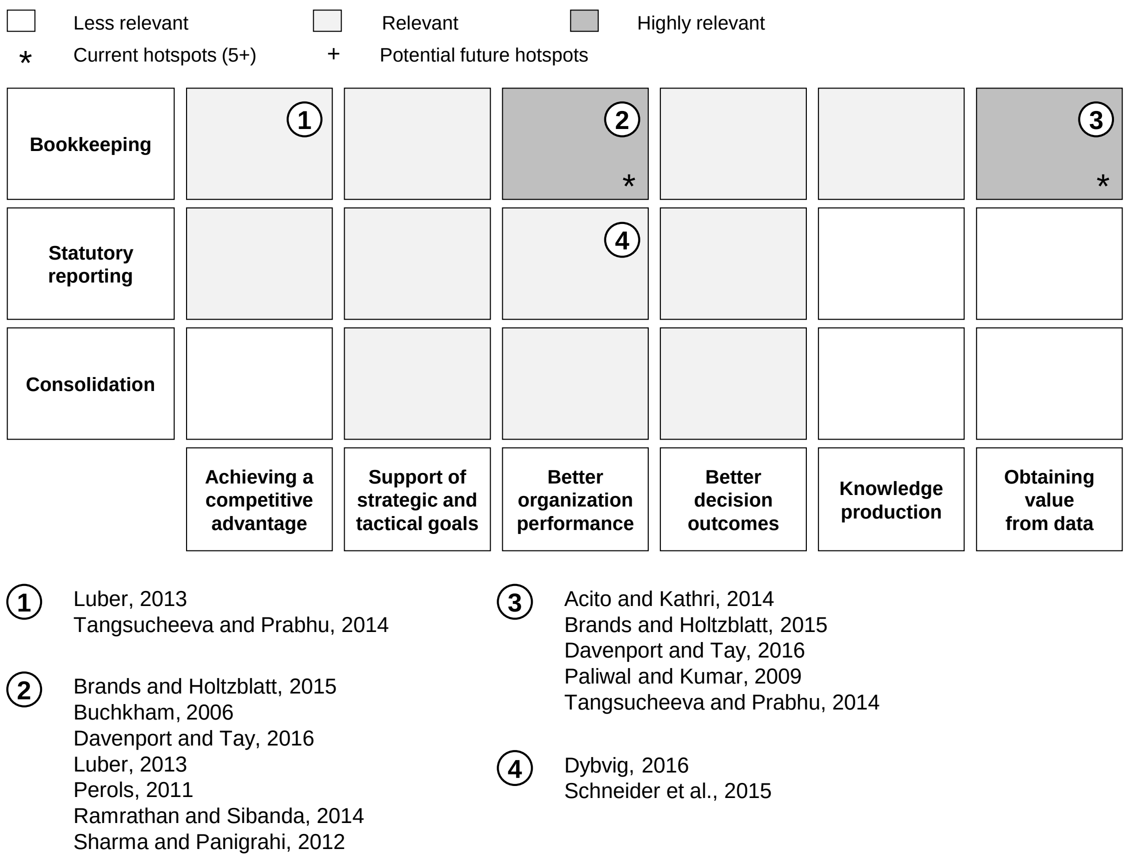
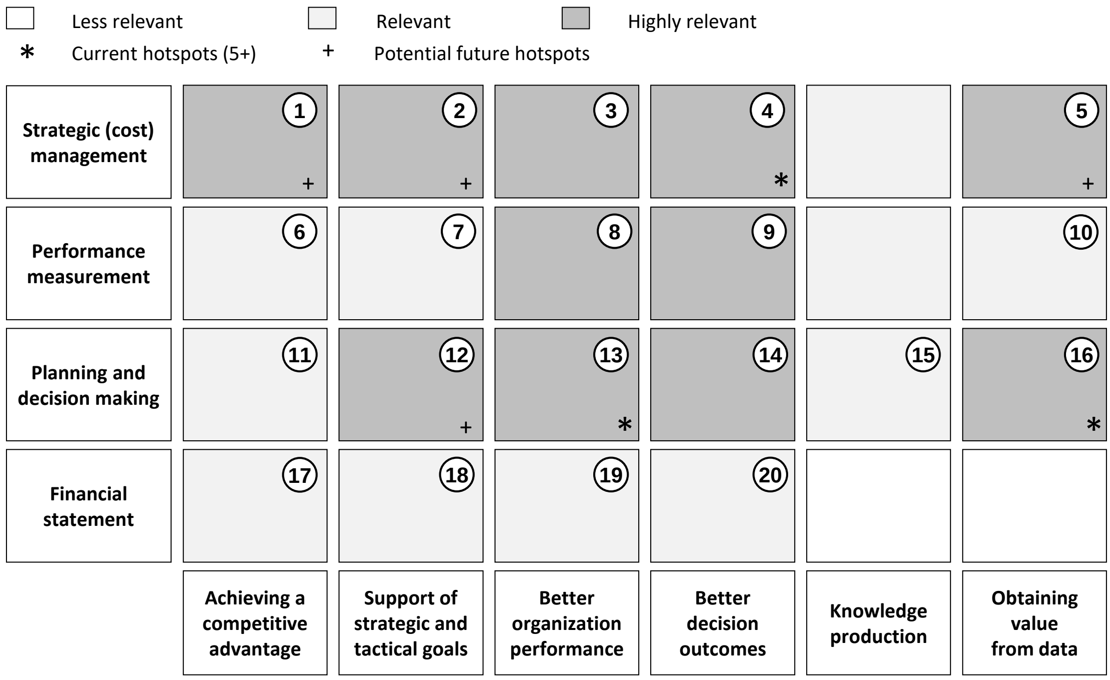


Figure 14. Classification for financial accounting

Brands and Holtzblatt (2015) address better organization performance in bookkeeping and state that accounts payable and payment monitoring can greatly benefit from an analytics integration. Analytics can also help in choosing and contacting the right customers in order to improve collections cash flows (Buckham, 2006). Achieving a competitive advantage and obtaining value from data are generally not directly associated with bookkeeping, but can become relevant goals when it comes to fraud detection (Perols, 2011; Sharma and Panigrahi, 2014), bankruptcy prediction, or credit default prediction (Paliwal and Kumar, 2009; Luber, 2013; Acito and Khatri, 2014; Ramrathan and Sibanda, 2014; Tangsucheeva and Prabhu, 2014). Dybvig (2016) proposes an optimized income statement improving organization performance in statutory reporting by including more accurate forecasts and Schneider et al. (2015) see potential for predictive analytics in an early identification of financial accounting discrepancies.

Currently, there are only two hotspots of research with the implementation of machine learning for fraud detection and the integration of external data to improve credit default and bankruptcy prediction (bookkeeping – better organization performance and obtaining value from data). There does not seem to be a real future hotspot. Amani and Fadlalla (2017) found 11% of data mining applications in financial accounting, 25% in managerial accounting, and 64% in assurance and compliance. The papers cited for financial accounting apply neural networks or other data mining techniques to predict, e.g., quarterly cash flows, risk factors in financial statements, or sentiments between different public statements. Yet, most of them take an external perspective, which is not the focus of this thesis.

With respect to management accounting, the overall picture is different (Figure 15). There is a number of highly relevant and only two less relevant areas in the proposed grid.



(1) Amani and Fadlalla, 2017
Marchant, 2013

(2) Bhimani and Willcocks, 2014
Cokins, 2013

(3) Schneider et al., 2015
Amani and Fadlalla, 2017
Quaadgras et al., 2014
Nielsen et al., 2014

(4) Amani and Fadlalla, 2017
Appelbaum et al., 2017
Cokins, 2013
Kriens et al., 1983
Nielsen, 2015
Nielsen, 2017
Schniederjans and Garvin, 1997

(5) Appelbaum et al., 2017
Nielsen, 2017

(6) Appelbaum et al., 2017

(7) Shanks et al., 2012

(8) Brands and Holtzblatt, 2015
Cao and Duan, 2012
Cokins, 2017
Schläfke et al., 2012

(9) Cao and Duan, 2012
Dino et al. 1982
Guo et al., 2016
Schläfke et al., 2012

(10) Sun et al., 2017

(11) Davenport and Ray, 2016
Payne, 2014

(12) Winklhofer and Diamantopoulos, 1996
Schneider et al., 2015

(13) Cao and Dunan, 2012
Duan and Xiong, 2015
Halper, 2014
Lu and Kao, 2016
Ramrathan and Sibanda, 2014
Zotteri et al., 2007

(14) Baesens et al., 2016
Bhimani and Willcocks, 2014
Quattrone, 2016
Dinan, 2015

(15) Bhimani and Willcocks, 2014

(16) Bhimani and Willcocks, 2014
Cao and Dunan, 2012
Halper, 2014
Shanks et al.., 2012
Sun et al. 2017
Tangsucheeva and Prabhu, 2014

(17) Bhimani and Willcocks, 2014

(18) Winklhofer and Diamantopoulos, 1996

(19) Bräuning et al., 2017
Dybvig, 2016

(20) Paliwal and Kumar, 2009

Figure 15. Classification for management accounting

More in detail, researchers mention almost all rationales for ML&A in connection with strategic (cost) management. Marchant (2013) states that management accountants are perfectly prepared to help management find ways to use data for a competitive advantage. Bhimani (2003) considers the impact of novel forms of information on corporate strategy and goals and even organizational structures. Better organization performance, for example through the creation and revision of business rules with the help of business analytics, is addressed by many authors (Marchant, 2013; Quaadgras et al., 2014; Schneider et al., 2015; Amani and Fadlalla, 2017). Likewise, better decision outcomes, for instance using the analytical hierarchy process for cost driver selection (Schniederjans and Garvin, 1997) or through a holistic view and integrated thinking (Nielsen, 2017), are covered sufficiently (Kriens et al., 1983; Cokins, 2013; Nielsen, 2015; Amani and Fadlalla, 2017; Appelbaum et al., 2017). Looking at performance measurement, there is less literature coverage of ML&A. Schläfke et al. (2012) provide a framework that consists of the four layers capture (performance drivers in inputs, processes, and outputs), couple (performance drivers), control (knowing cause-effect relationships and crucial levers), and communicate (internally and externally). Recent conference proceedings look at critical success factors for analytics in performance management to support strategic goals (Shanks et al., 2012) or the mechanisms through which business analytics supports strategic decision making (Cao and Duan, 2012). Further research emphasizes better decision outcomes or identifies ways to obtain value from data (Dino et al., 1982; Guo et al., 2016; Cokins, 2017; Sun et al., 2017). O'Leary (2018) sheds light on potential applications of an artificial intelligence like IBM's Watson, for example, reporting based on natural language questions by the user.

Planning and decision making is another area of high interest. However, at the current point, it is focused mainly on better organization performance (Zotteri et al., 2005; Cao and Duan, 2012; Halper, 2014; Ramrathan and Sibanda, 2014; Duan and Xiong, 2015; Lu and Kao, 2016) and better decision outcomes (Bhimani, 2003; Dinan, 2015; Baesens et al., 2016; Quattrone, 2016) due to more accurate and fact-based data, even for small and medium-sized companies. Analytics is also applied to planning and decision making to achieve a competitive advantage, for instance with the help of a generalized advanced analytics competency in the finance department (Payne, 2014; Davenport and Tay, 2016), or to support strategic goals with improved forecasting (Winklhofer and Diamantopoulos, 2002). Obtaining value from data was also covered from various angles like looking at what possible actions customers might take (Bhimani, 2003; Cao and Duan, 2012; Shanks et al., 2012; Halper, 2014; Tangsucheeva and Prabhu, 2014; Sun et al., 2017). On the other hand, financial statement preparation was covered only occasionally with articles focusing on the impact of machine learning support in selecting different accounting methods (Paliwal and Kumar, 2009) or better organization performance in preparing the statements (Dybvig, 2016; Bräuning et al., 2017).

Finally, there were a number of articles in top accounting or information systems journals that did not address accounting-specific benefits of ML&A. Some of them highlighted a better organization performance from a general business perspective (Wixom et al., 2013; Asadi Someh and Shanks, 2015) or obtaining value from data in the business functions (Lacity et al., 2011; Sharma et al., 2014).

3.2. Method: Exploring the drivers of machine learning adoption helps decision makers

3.2.1. Partial least squares structural equation modeling

As opposed to first-generation techniques like cluster analysis, exploratory factor analysis, analysis of variance, regression, and confirmatory factor analysis, structural equation modeling is a second-generation technique[27] (Hair Jr et al., 2016). In accounting research, PLS-SEM has not seen widespread adoption so far. Several reviews over the last years each identified less than 40 articles (Smith and Langfield-Smith, 2004; Lee et al., 2011; Nitzl, 2016). Despite that, Nitzl (2016) argues that *PLS-SEM is well suited to accounting research* as it supports the often exploratory, data-driven nature of having to build complex models for areas in which theory is weak and uncertainty often plays a role. Hence, in the

[27] First- and second-generation refers to the prevalence of these methods in psychology and social science research, e.g., as stated by Fornell (1987).

following sections, PLS-SEM will be applied to analyze the current state and drivers of machine learning and analytics adoption.

In case of a primarily confirmatory goal, covariance-based (CB) SEM is well-suited, whereas partial least squares (PLS) SEM is better suited to exploratory questions (Rigdon, 2012). Additionally, PLS-SEM does not make many assumptions about the distribution of the data (i.e., it is *non-parametric*), whereas a normal distribution is desirable for CB-SEM. Data should be on an ordinal scale, where coding can be used to assign numbers to categories in a manner that facilitates measurement, while observing the requirement for equidistance between the values on the scale.

Besides only few assumptions about the data, several characteristics make PLS-SEM very attractive for research projects like the one at hand. It performs well with small sample sizes, missing values (up to 5% missing per indicator do not affect the results much), and complex models (Cassel et al., 1999; Henseler et al., 2009). More in detail, the *sample size* can relatively easily be determined based on the guidelines provided in Cohen's (1992) "power primer"[28]. For instance, aiming for an acceptable quality for outer loadings (≥ 0.7), a certain level of complexity (maximum number of independent variables per single construct, say 4 in this example), the commonly used value for power of 80%, a significance level of 5%, and a minimum detectable coefficient of determination (R^2) of 0.25, at least 41 responses should be collected.

The path model for PLS-SEM consists of a *structural (inner) model and measurement (outer) models*. The former consists of all constructs (latent variables) and the latter of all indicators (manifest variables, directly measured proxy variables) (Hair Jr et al., 2016). There are two types of relations or ways to measure unobserved variables: Reflective (the construct causes the measurement or indicator), and formative (the indicator causes the construct). The latter are also called composite (Bollen and Bauldry, 2011) and can be assumed to be error-free (Diamantopoulos, 2011). Generally, theory and logic dictate the order of constructs in a structural model. In large complex models, several adjustments

[28] There are also easy rules of thumb like the "10-times rule" (Thompson et al., 1995), but they should only be used as a first estimate as they neglect the complexity of the model and desired significance levels.

may be necessary to best fit a model to the data. Thus, it may be advisable to test alternative models with different sequences of the constructs (Callaghan et al., 2007). Moreover, two concepts may need to be considered in a model: (1) Mediation, which is an indirect effect linking two constructs via a third (mediating) construct and (2) moderation which is a construct (variable) that affects the relationships between other constructs (exogenous and endogenous latent variables) in the model. For example, moderator effects can be used for multi-group analysis (Sarstedt et al., 2011).

Owing to the difficulty of capturing a construct as comprehensively as possible, over the past decades, several processes for developing indicators or measurements for a construct have been developed, e.g., by De Vellis (2011) or Hair et al. (2010). In contrast, the *process of deriving answers* from these measurements is relatively straightforward: (1) specify the structural model, (2) specify measurement models, (3) collect data and examine, (4) estimate the path model, (5) assess results of the measurement models, (6) assess results of the structural model, (7) perform advanced analyses, and finally, (8) interpret the results (Hair Jr et al., 2016; Henseler et al., 2017). The reliability of models can be assessed by the coefficient of determination of dependent constructs (i.e., the ones not at the beginning of the model), which is a measure for the proportion of variance explained (Nagelkerke, 1991) and by Dillon-Goldstein's composite reliability ρ_C, which should be above 0.7.

3.2.2. Survey on machine learning and analytics adoption

Addressing issues related to user acceptance of IS, models were developed to measure and predict system use. In the 1970s, technological progress resulted in a wave of increasing technology adoption across companies. However, this also lead to a growing number of system adoption failures. As a consequence, researchers sought to develop models of user acceptance and rejection and ultimately build systems catered for adoption (Chuttur, 2009).

Among these models is Davis' (1989) *TAM* which is based on the theory of reasoned action and has become one of the most widely used models in IS. To determine system adoption, two primary determinants are considered: *perceived* *usefulness and perceived ease of use* (King and He, 2006). These two fundamental factors influence the attitude towards using, which, in turn, can be seen as a prerequisite for actual system use (Davis, 1985). Over the years, several adjustments to the original TAM have been proposed. For instance, Venkatesh (2000), proposed subcategories to anchors and adjustments to better explain variance when measuring perceived ease of use variance. Figure 16 shows the basic TAM (Davis, 1985) in the middle with the additions of Venkatesh and Davis (2000) and Venkatesh (2000) to the left and the intention to use construct (Davis et al., 1989) to the right.

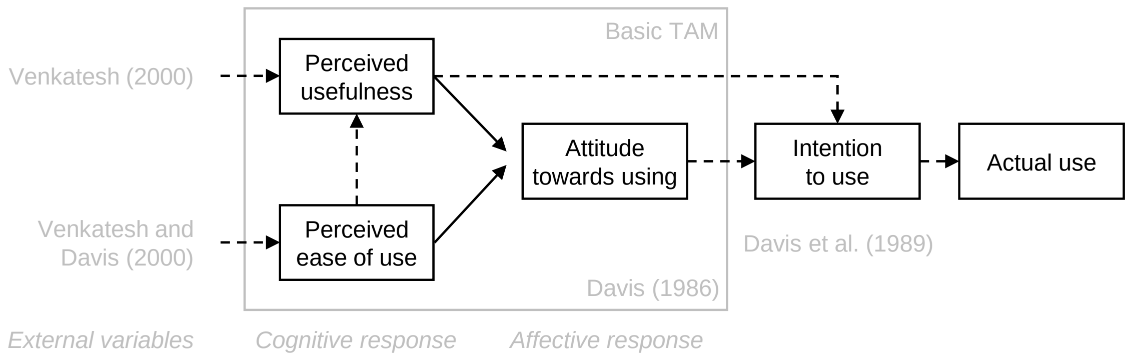


Figure 16. Technology acceptance model

Approaching the topic from a different angle, the *task technology fit* (TTF) model by Goodhue and Thompson (1995) looks at technology and performance. Four of the five model components, namely task characteristics (e.g., routine, or non-routine), technology characteristics (e.g., reliability), task-technology fit, and utilization are needed to predict the last component, performance impact. Additionally, individual characteristics and some precursors of utilization are considered in the model. While the TTF mostly provided theoretical guidance but lacked practical applicability, it was later-on integrated as determinants of the TAM constructs perceived usefulness and perceived ease of use (Dishaw and Strong, 1999; Gebauer and Shaw, 2004; Larsen et al., 2009).

Based on these two fundamental models, a model for machine learning and analytics adoption was developed in this work. In order to do so, a questionnaire (see Appendix B) was issued to the members of the Schmalenbach working group “Digital Finance” and other practitioners. Answers were given on different types of five-point Likert scale[29]. For questions relating to *frequency*, the choices were [1] “never”, [2] “rarely”, [3] “sometimes”, [4] “often”, and [5] “always”. For questions relating to *agreement*, the choices were [1] “strongly disagree”, [2] “disagree”, [3] “neutral”, [4] “agree”, and [5] “strongly agree”. Finally, for questions relating to a *degree* of something, the choices were [1] “not at all”, [2] “slightly”, [3] “somewhat”, [4] “very much”, and [5] “extremely”. In total, 39 people responded to the survey[30] (sample characteristics can be found in Table 6). This number is roughly in line with the sample size recommendation stated in the previous section for a minimum detectable R^2 of 0.25 with a maximum of four independent variables per construct[31]. The survey questions were developed along the guidelines provided by Fowler Jr (2013) and each serves as a proxy (measurement variable) for one of the constructs (latent variables) of the model. In the following section, the structural and measurement models will be explained in detail.

Table 6. Machine learning and analytics adoption sample characteristics

Position	**No.**	**%**	**Department**	**No.**	**%**	**Analytics-related role**	**No.**	**%**
Upper management	5	13	Finance	26	67	End user	20	51
Middle management	12	31	Business	4	10	Power user	12	31
Lower management	11	28	IT	5	13	Developer	4	10
Staff	11	28	Other	4	10	Other	3	8
Total	39	100	Total	39	100	Total	39	100

Employees in the company	**No.**	**%**	**Employees in the department**	**No.**	**%**
<250	0	0	<25	9	23
250-10,000	3	8	25-1,000	26	67
10,001-100,000	23	59	>1,000	4	10
>100,000	13	33			
Total	39	100	Total	39	100

[29] Lissitz and Green (1975) showed that increasing the number of scale points beyond five does not increase reliability to the same degree. An additional “don’t know” answer option is sometimes discouraged as it may be seen as a “lazy option”, but it was included in this work to avoid random answers.

[30] Note that 34 single answers (from a total of 2106) were not provided by the respondents, which equals a missing (not available, NA) ratio of 1.6%. In section 3.3.1, these missing values will lead to some row sums being less than 100%. In section 3.3.2, the missing values will also be imputed for comparison.

[31] The recommendations with 80% statistical power, 5% significance level, and a minimum detectable R^2 of 0.50 are: 16 (max 3 independent variables), 18 (max 4 independent variables), 20 (5). For a minimum detectable R^2 of 0.25 they are: 37 (3), 41 (4), 45 (5). For a minimum detectable R^2 of 0.10 they are above 100 in all cases (Cohen, 1992; Hair Jr et al., 2016).

3.2.3. Structural and measurement model

The structural model for this work follows the *joint TAM-TTF model* proposed by Dishaw and Strong (1999). The TTF constructs individual characteristics, technology characteristics, task characteristics, and task-technology fit serve as input to the two TAM constructs perceived ease of use and perceived usefulness, which, in turn, have an influence on the attitude towards using and the intention to use. The "dependent variable" of the model is the actual use (Figure 17). While this structural model is somewhat universal and can be applied to many technology adoption objectives, the measurement model is what makes this combination tailor-made to machine learning and analytics adoption. For instance, the measurement model for technology characteristics comprises the data types used for ML&A – either structured internal, structured external, semi- or unstructured internal, or semi- or unstructured external (Halper, 2014) – and the methods applied – prediction, classification, time-series forecasting, association, and segmentation (cf. section 4). Tools as the third latent variable added to the measurement model as a potential moderator effect. Similarly, the measurement model for perceived usefulness consists of three latent variables: First, the rationale for use (Holsapple et al., 2014) – achieving a competitive advantage, supporting strategic and tactical goals, improving performance, making better decisions, having better decision processes, producing knowledge, or obtaining value from data. Second, the level of support (in the style of the query items in Davis et al. (1989)) – e.g., beating human experience or making human workers complete tasks more quickly or more accurately. Third, the decision steps according to Simon (1997) – identification of problems (intelligence) and opportunities, development of solutions (decision alternatives, design), and evaluation of decision alternatives (choice). A list of all variables in the measurement model and their respective form of measurement[32] can be found in Figure 18. All measurement models are set up with formative relations between the constructs and their proxies. The abbreviations indicated in the third column (e.g., AppA1) are used in the R model (see Appendix D). In section 3.3.2, the model parameters

[32] See section 3.2.2 for a description of Likert scale items.

will be described both for the entire model (as in Figure 17) and for various sub-models.

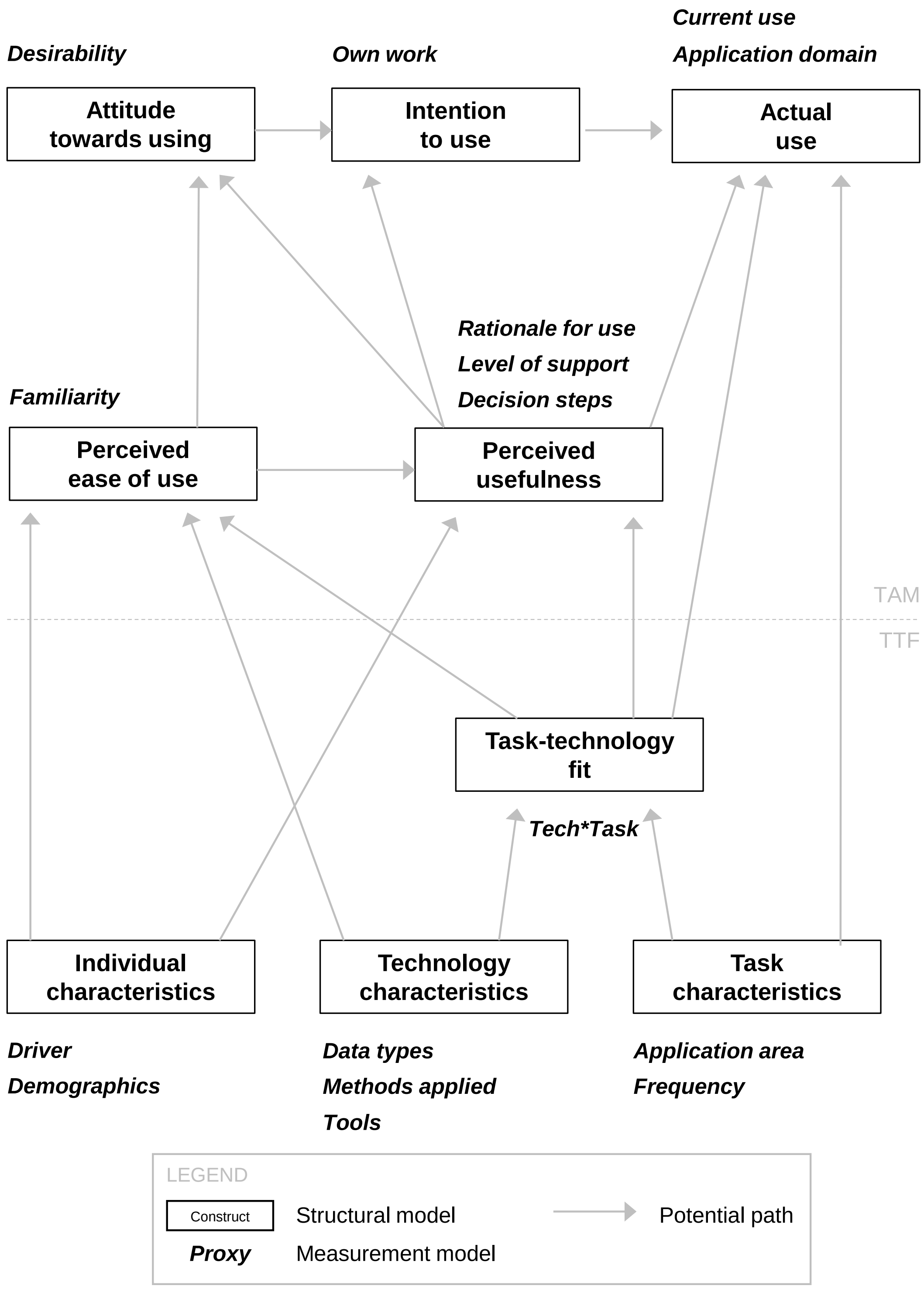


Figure 17. Full structural and measurement model, all paths are formative

Construct	Latent variables		Measurement
Task characteristics	Application area	Revenues (AppA1) Cost (AppA2) Working capital (AppA3) Risk (AppA4)	Likert frequency
	Frequency	Planning (FREQ1) Budgeting Forecasting	Likert frequency
Technology characteristics	Data types	Structured internal (DAT1) Unstructured internal Structured external Unstructured external	Likert frequency
	Methods applied	Prediction (MTD1) Classification Time-series forecasting Association Segmentation	Likert frequency
Individual characteristics	Demographics	Employees company (DEMO1) Position Department Employees finance Analytics role	Multiple choice
Task-technology fit	Task*technology	(TTF)	Orthogonal interaction
Perceived ease of use	Familiarity	Descriptive statistics (FAM1) Linear regression Time-series models Machine learning	Likert degree
Perceived usefulness	Rationale for use	Achieve a competitive advantage (RfU1) Support our strategic and tactical goals Improve our performance Make better decisions Have better decision processes Produce knowledge Obtain value from data	Likert agreement
	Level of support	Not beat human experience (SUP1) Support human workers in their tasks Replace most of the efforts of human workers Make human workers complete tasks more quickly Make human workers complete tasks more accurately	Likert agreement
	Decision steps	Identification of problems (DEC1) Development of solutions Evaluation of alternatives	Likert degree
Attitude towards using	Desirability	Planning (DES1) Budgeting Forecasting	Likert agreement
Intention to use	Own work	Stick to the tools and methods (OWO1) Use the machine as a side-car Transfer all I can to the machine	Likert agreement
Actual use	Current use	Single prototypes (CU1) Side-by-side use Full-scale	Likert frequency
	Application domain	Financial accounting (AppD1) Management accounting Treasury	Likert frequency
Moderators	Tools Driver	(TLS) (DRV)	Multiple choice

Figure 18. Model constructs, latent variables, and measurement

3.3. Results: Task characteristics are the most important driver of adoption

3.3.1. Descriptive statistics

Four questions addressed the intensity of ML&A use. Looking at the responses, it is evident that single prototypes are still most common with 70% of the participants using them sometimes, often, or always as opposed to 51% for side-by-side use, and only 21% full-scale adoption (Figure 19a). At the other end of the spectrum, only 10% never use single prototypes, while 33% never use machine learning or analytics full-scale for certain use cases. With respect to *sub-functions* (Figure 19b), management accounting is the most advanced with 8% always using ML&A, 28% often, and 26% sometimes. Financial accounting is trailing behind a bit with 3%, 13%, and 26%, respectively. The large fraction of 28% “don’t know” answers makes an assessment in treasury less reliable. Due to this insecurity and the general focus of this thesis on financial and management accounting, treasury will not be the covered in much detail in the following. The spread over the *use cases* revenues, costs, working capital, and risk is relatively even (Figure 19c), with all of them used often by around 20% and sometimes by around 30%. The highest divergence is between revenues and risk as there are 10% respectively 21% who never use ML&A and 15% respectively 3% who always use it. Finally, the *tasks* planning, budgeting, and forecasting (Figure 19d) differ the most in the categories never (23%, 33%, 13%) and often (15%, 10%, 26%). Hence, budgeting is supported the least by ML&A with 50% using it never or rarely, while forecasting sees the highest use with almost two thirds (64%) sometimes, often, or always.

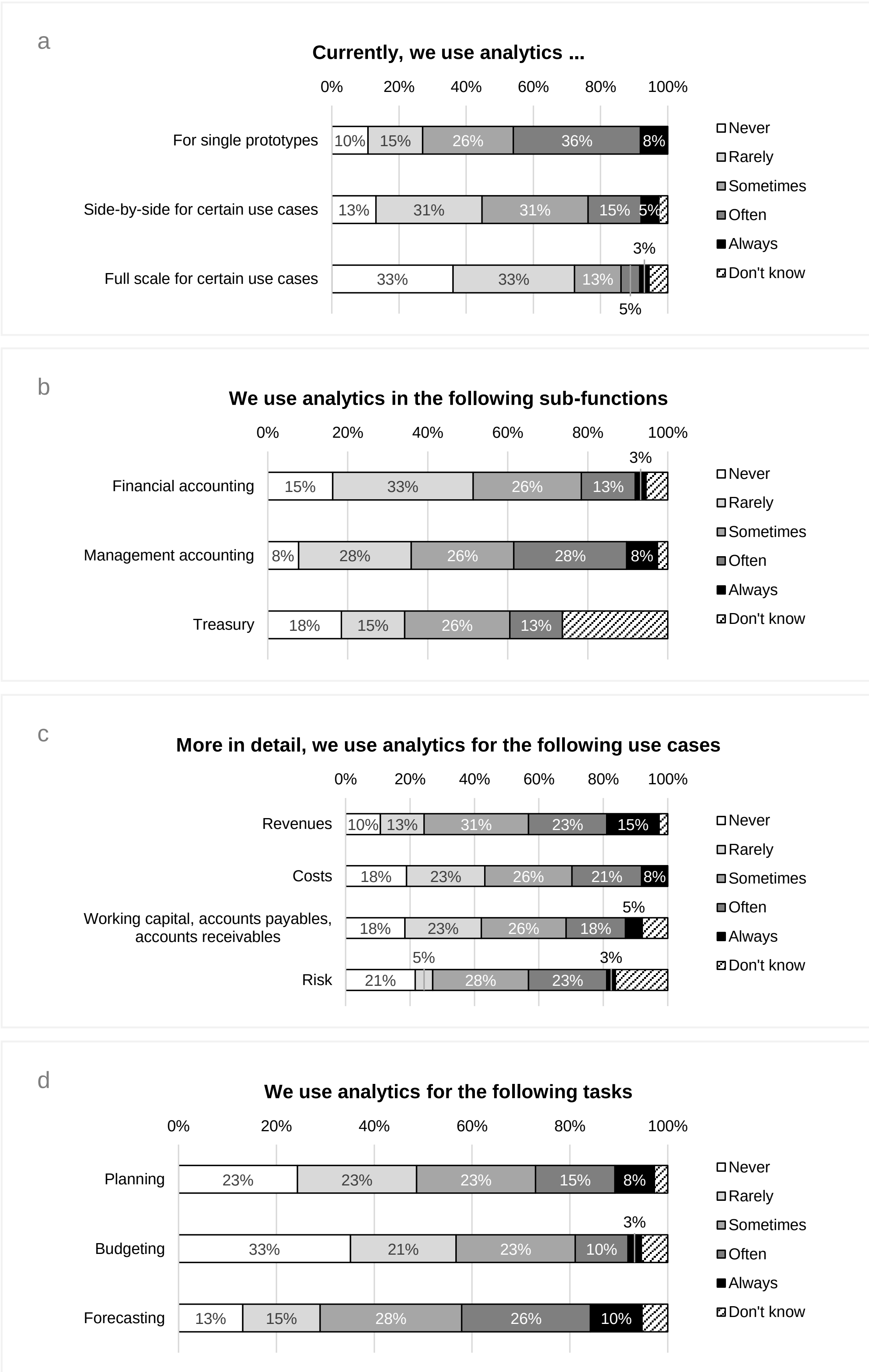


Figure 19. Results for current use, sub-functions, use cases, and tasks

The next four sets of questions covered the data and algorithms used, the decision steps supported, and the motivation to use ML&A. More in detail, the *data used as a basis for ML&A* (Figure 20a) are most often structured internal data like transactions recorded in ERP systems or aggregates thereof in BI systems (67% often or always). In contrast, 0% of respondents always use structured external data like indicators from third-party databases, 23% often, 18% sometimes, and more than 50% rarely or never. Chapter 6 will argue for a more frequent inclusion of the right external indicators to improve forecast accuracy with a mix of internal and external data. Semi- or unstructured external data is used the least frequently with almost two thirds using it never or only rarely. The *categories of algorithms used* (Figure 20b) all show a significant amount (>20%) of don't know answers. Only methods from time-series forecasting (subject of chapter 6) are used always or often by at least one third of the respondents. Roughly one fifth use prediction or classification methods (subject of chapter 5) often or always. The verdict on the usefulness of ML&A is similar for different *steps of decision making* with around 30% responding very much or extremely and around 10% not at all (Figure 20c). However, the number of undecided answers increases from 10% in the case of identifying problems and opportunities, over 13% when developing decision alternatives to 18% when evaluating these alternatives. The responses regarding the *rationale for ML&A use* demonstrate that there is indeed a wide variety of reasons with relatively equal importance (Figure 20d). For instance, 46% of respondents agreed and 36% strongly agreed that achieving a competitive advantage and improving performance are among the reasons for ML&A use. Similarly, 51% and 62% strongly agree that ML&A is used to make better decisions and obtain value from data, respectively. Adding 36% respectively 33% of respondents who agree leads to 87% and 95% of agreement and makes them the two most important rationales for using ML&A. While the feedback was mostly in favor of the response options, only having better decision processes and producing knowledge were subject to some disagreement (both 10%).

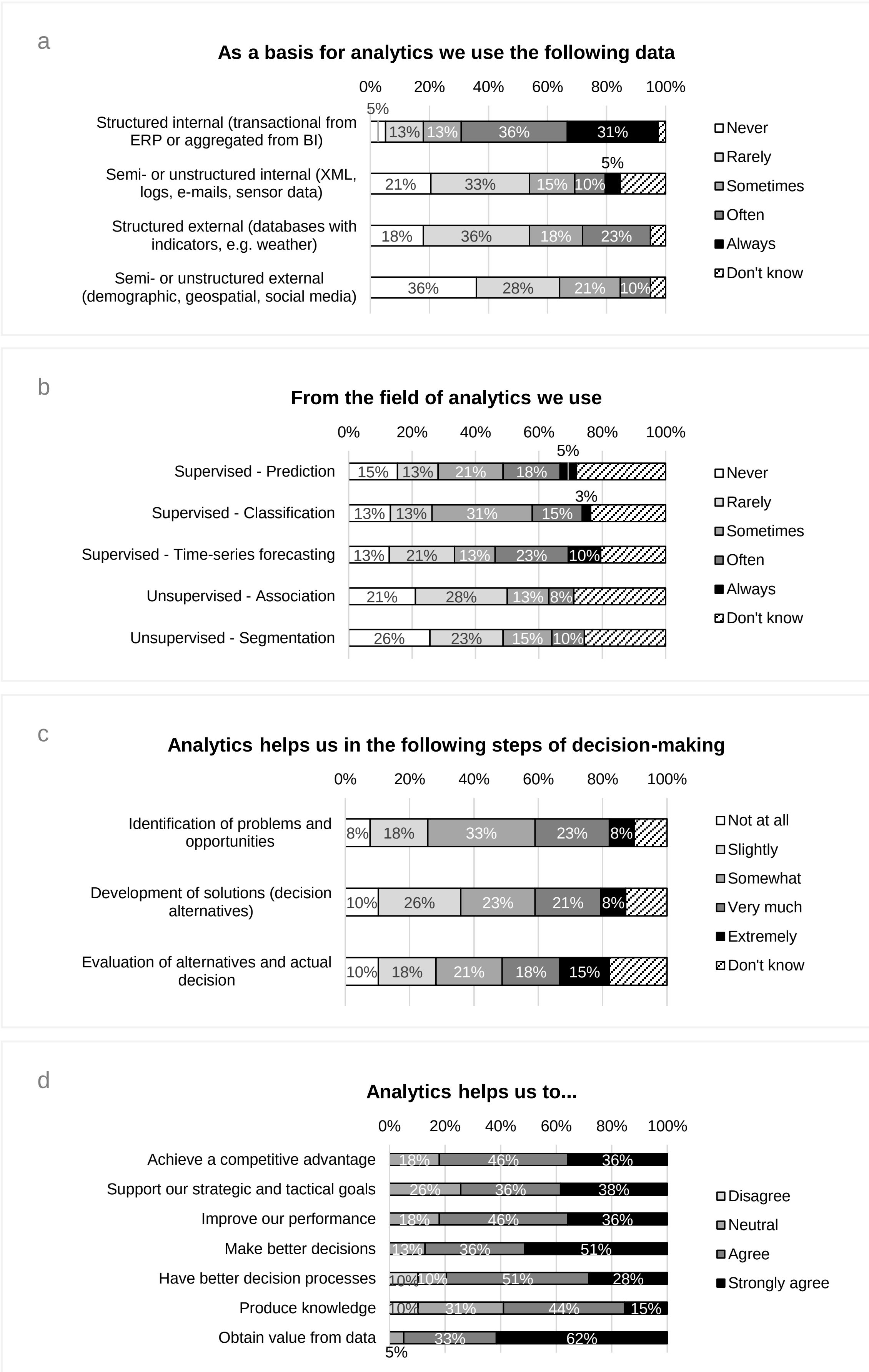


Figure 20. Results for data and algorithms used, decision-steps and rationale

Afterwards, four sets of questions were directed at the personal opinion of the respondents and their familiarity with different algorithms. The first question asked if respondents believed machines could not beat human experience and received an almost normal distribution of responses with 8% strongly disagreeing, 23% disagreeing, 28% neutral, and the rest agreeing or strongly agreeing (Figure 21a). Supporting human workers in their tasks, on the contrary, was unanimously agreed upon by all participants. A similar response pattern was given for making human workers complete tasks more quickly and more accurately (>80% agree or strongly agree). With respect to replacing most of the efforts of human workers, respondents were roughly split into one third disagreeing, one third neutral, and one third agreeing. Asked what they would do for their own work (Figure 21b), 62% of respondents said they would use the machine as a side-car (agree or strongly agree), whereas only 8% said they would stick to the tools and methods they are used to. Transferring all they can to the machine is an attractive option for 62% of respondents (agree or strongly agree), while 13% are reluctant to do so (disagree or strongly disagree). This might also have to do with *familiarity* (Figure 21c). As far as descriptive statistics and correlations are concerned 85% responded that they were either extremely, very, or moderately familiar with it. This figure dropped to 75% for linear regression models, and, more drastically, to 44% for time-series models like ARIMA and 36% for machine learning algorithms like neural networks. Consequently, the share of respondents not at all familiar with these algorithms increased from 10% (descriptive statistics and linear regression models) to 31% (time-series models) and 36% (machine learning models). In light of the responses to the current use for the tasks planning, budgeting, and forecasting (Figure 19d), the *desirability* shows a different picture (Figure 21d). 95% consider ML&A desirable for forecasting, 90% for planning, and 77% for budgeting. With respect to the latter, 15% are neutral and 8% disagree. Thus, the large majority is very much in favor of introducing or continuing to use ML&A for planning, budgeting, and forecasting alike.

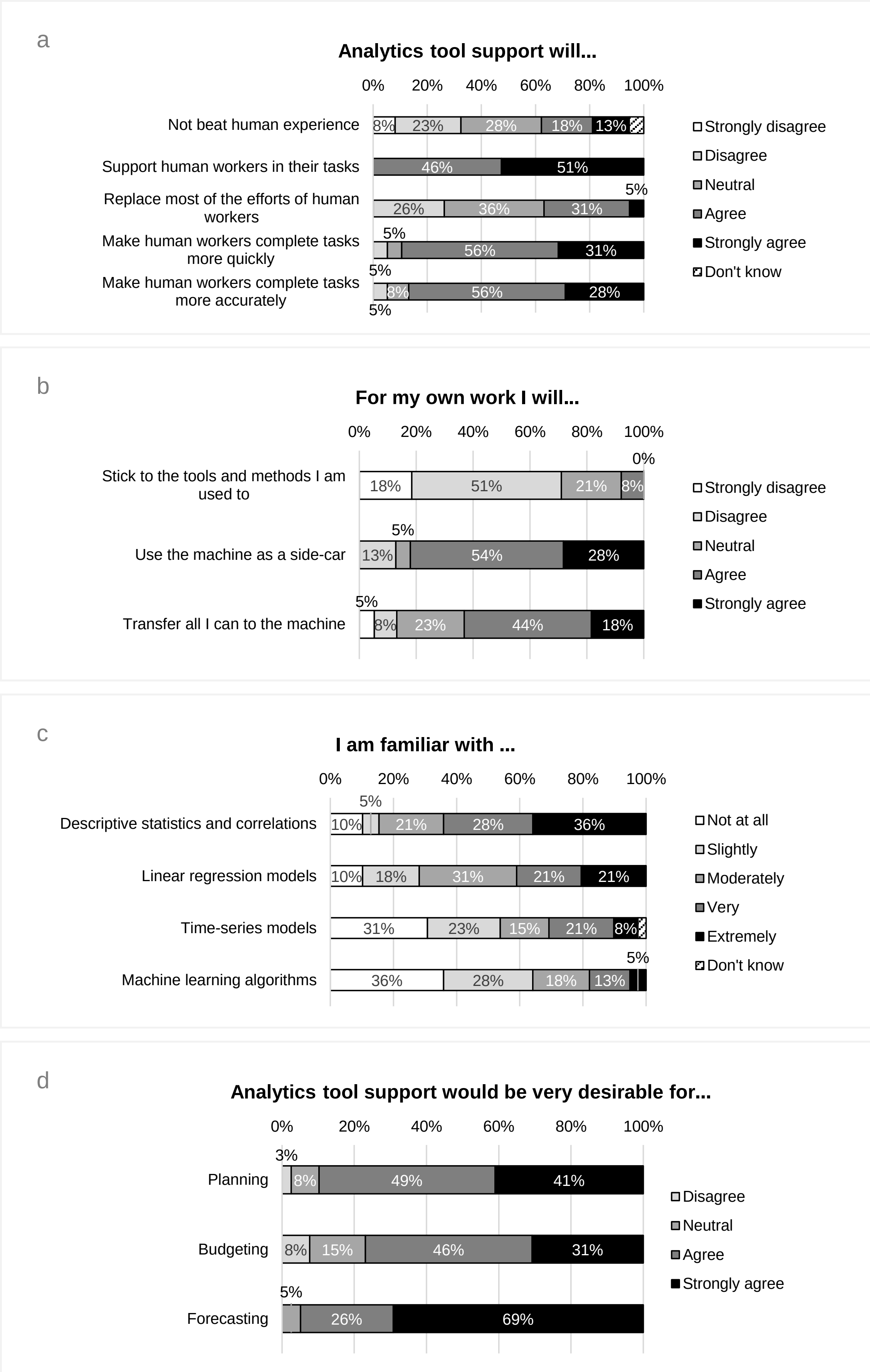


Figure 21. Results for personal opinion and familiarity

Finally, the last two questions asked for the way analytics is driven inside the organization and what kind of tools are used to support analyses. 18% answered that in their organization ML&A is driven top-down with an analytics strategy, 33% that it is driven bottom-up via pain points and use cases in the department, and 44% said it is a combination of both (Figure 22a). Regarding the tools in use (Figure 22b), not surprisingly, spreadsheets like Microsoft Excel were mentioned by 95%. This was followed by 74% dashboarding and visual analytics tools (e.g., Tableau, PowerBI), 51% enterprise BI solutions (e.g., SAP BusinessObjects, MicroStrategy) as well as open source platforms (e.g., Hadoop, R), and 28% analytics platforms (e.g., Teradata, SAS).

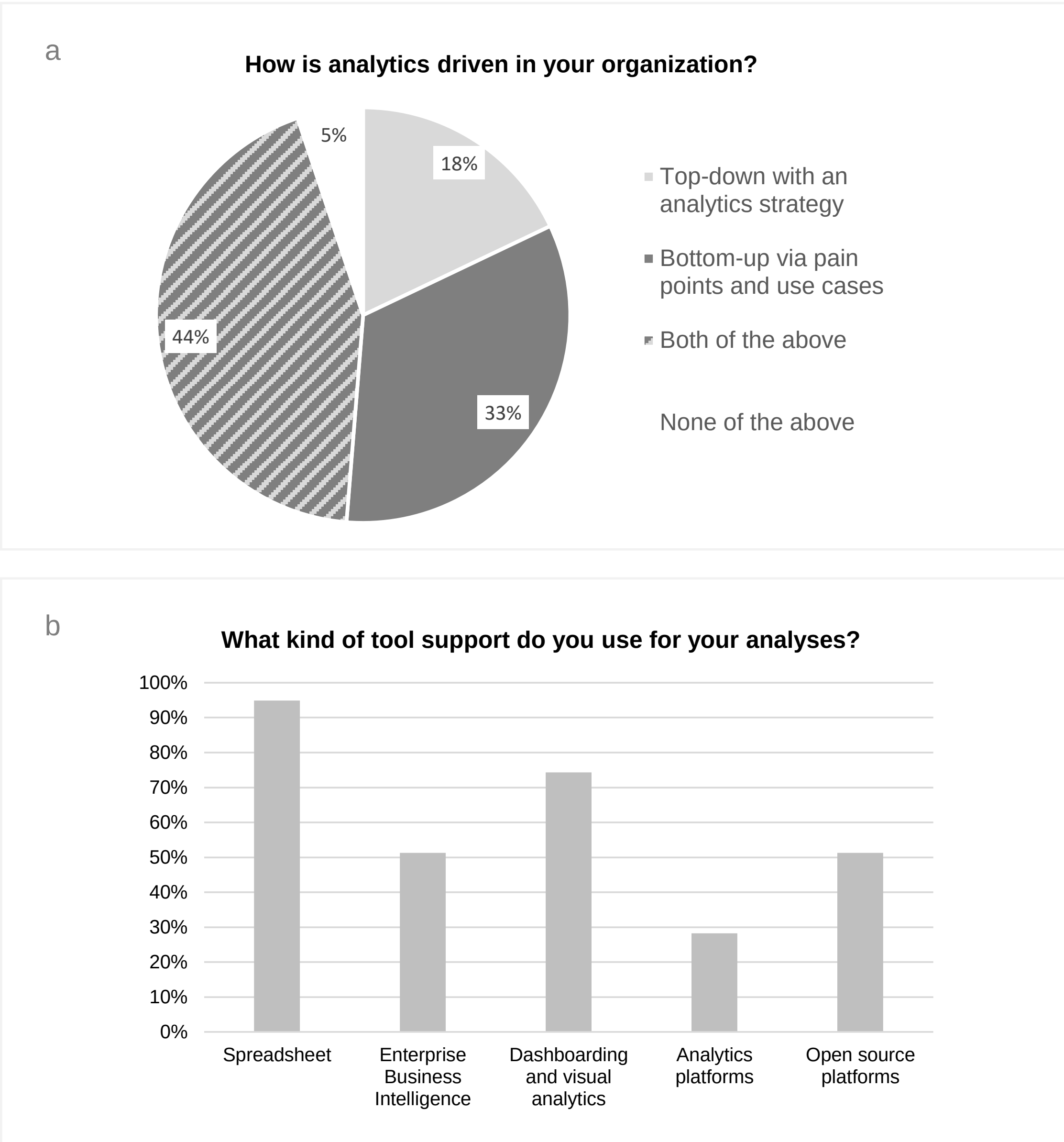


Figure 22. Results for driver and tool support for analyses

3.3.2. Model parameters

Building on these descriptive statistics, in the following, the results of PLS-SEM will be described. In a first step, the full model as depicted in Figure 17 was estimated[33]. The R package `SEMinR` by Ray and Danks (2019) was used for all modeling and the guidance by Hair Jr et al. (2016), Monecke and Leisch (2012), and Sanchez (2013) was followed (see Appendix D for exemplary R code). However, due to the large number of variables and the limited number of observations – i.e. an overdetermined model – the resulting model was an essentially perfect fit and the coefficients unreliable. Without imputation (32 observations), bootstrapping the model (1000 resamples) led to an error due to singularity. With random forest imputation[34] (39 observations), the model still produced an “essentially perfect fit” error, but parameters could be estimated. However, none of the coefficients was statistically significant and none of the loadings of the TTF measurement model were above 0.7. As a consequence, various sub-models were derived, each with a number of adjustments, such as:

- Removing the task-technology interaction term, which accounts for 90 variables if all task and all technology proxies are included
- Removing proxies that produce errors if simultaneously included in the model, such as “stick to the tools and methods I am used to” (OWO1) and “transfer all I can to the machine” (OWO3)
- Removing proxies that are related to less-advanced states of ML&A adoption, such as “For single prototypes” (CU1)
- Removing proxies with poor loadings, particularly for constructs with many proxies, such as perceived usefulness

In the following, four such sub-models will be described in detail: (1) The TAM with its constructs perceived ease of use, perceived usefulness, attitude towards using, and intention to use. (2) A greatly simplified version of the TAM+TTF model with the constructs task characteristics, technology characteristics, and perceived

[33] All computations for this section and the remainder of this work were performed using R x64 3.5.1 and R Studio 1.0.143 on a Lenovo laptop with Intel Core i5 6440HQ, 8 Gigabyte RAM, and Windows 10 64bit.

[34] See section 4.6.1 for more details. The data suggest a case of “missing completely at random”.

usefulness. (3) An extension of sub-model (2) with the constructs task characteristics, technology characteristics, perceived ease of use, perceived usefulness, and intention to use. Finally, (4) an almost full model with the added attitude towards using construct of the original and TAM model and the individual characteristics construct.

Looking at sub-model (1), four path coefficients of the seven tested were significant (again bootstrapping with 1000 resampling iterations): perceived ease of use has an influence on perceived usefulness on a 0.1% significance level, perceived usefulness has an influence on actual use on a 1% and on actual use on a 0.1% significance level, and attitude towards using has an influence on intention to use on a 1% level (Figure 23). Some of the loadings were below the 0.7 lower bound and were dropped in a second iteration. However, the reliability of the constructs was relatively poor in both iterations with adjusted R^2 values between 0.06 (attitude towards using) and 0.45 (actual use). Note that due to the incompatibility of a simultaneous inclusion of more than one intention to use proxy, only use the machine as a side-car (OWO2) and transfer all I can to the machine (OWO3) were tested with the former leading to better results.

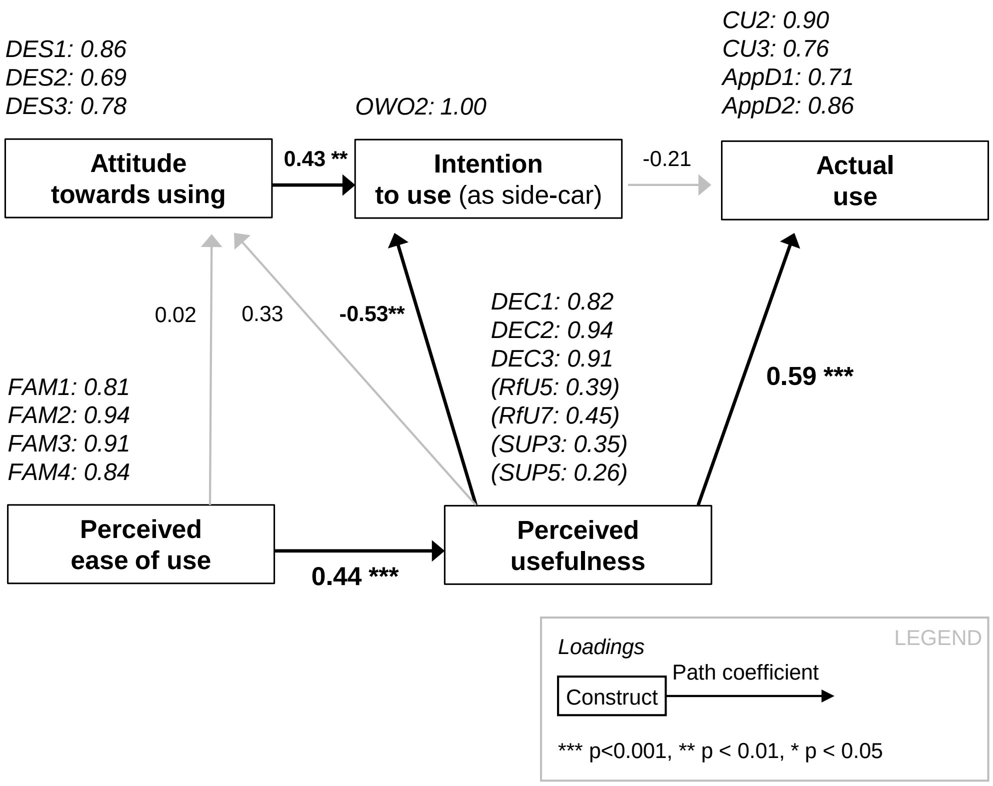


Figure 23. Sub-model (1) coefficients and loadings

The paths in sub-model (2) with only four constructs were slightly modified in comparison to the full model in that it also includes a direct path from technology characteristics to actual use (Figure 24).

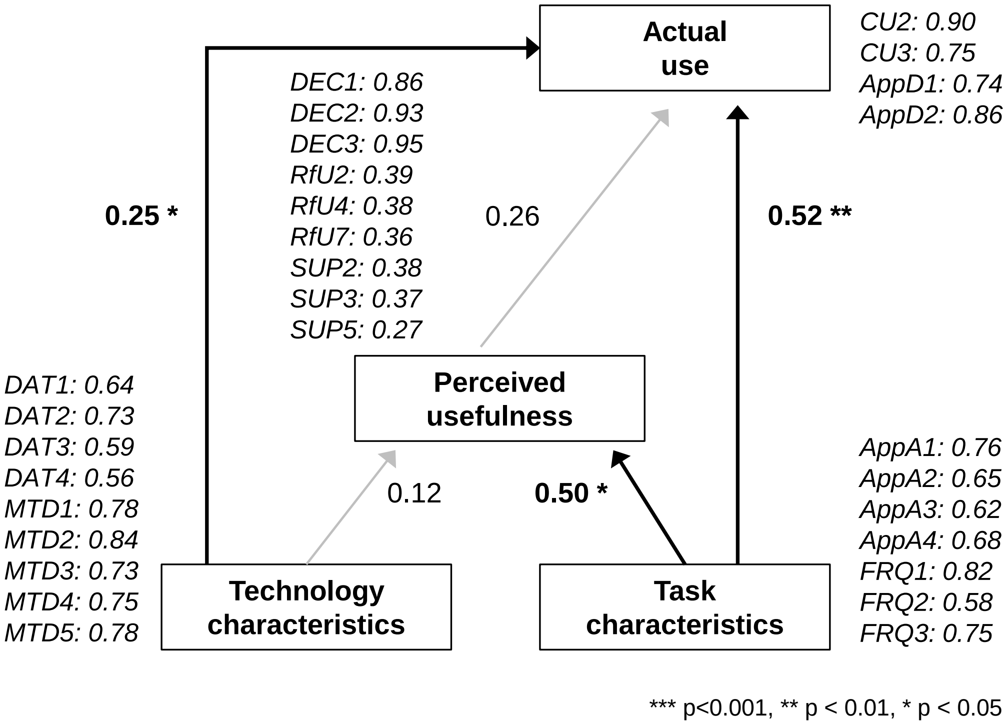


Figure 24: Sub-model (2) coefficients and loadings

The results based on 1000 bootstrap samples with three of the five path coefficients being statistically significant showed that technology characteristics can have a significant influence on actual use and is not only mediated by perceived ease of use or perceived usefulness like suggested in the TAM-TTF model by Dishaw and Strong (1999) and the full model at the beginning of this section. Additionally, the loadings of the proxies were mostly on a higher level than in sub-model (1) and Dillon-Goldstein's ρ_C was higher than 0.7 for all constructs (0.81 for perceived usefulness, 0.87 for task characteristics, 0.89 for actual use, and 0.90 for technology characteristics). More importantly, adjusted R^2 for actual use increased to 0.78. The p-value for the path from perceived usefulness to actual use was 0.053, hence on the verge of being significant, which is why sub-models (3) and (4) were developed.

Testing sub-model (3), there were six statistically significant path coefficients (Figure 25), with technology characteristics ($p<0.01$), task characteristics ($p<0.01$), and intention to use ($p<0.05$) all having an influence on actual use. Additionally, perceived usefulness is influenced by task characteristics ($p<0.05$) and perceived ease of use ($p<0.05$), and intention to use is influenced by perceived usefulness ($p<0.05$). The coefficient of determination for actual use was slightly increased (adjusted R^2 of 0.81). Composite reliability ρ_C scores were again all above 0.80 and the measurement model loadings mostly above 0.7 (see Table 7)[35]. Note that, as in sub-model (1), the use as side-car proxy for intention to use leads to negative path coefficients, whereas replacing it with the transfer all I can proxy would lead to positive coefficients, but no statistical significance.

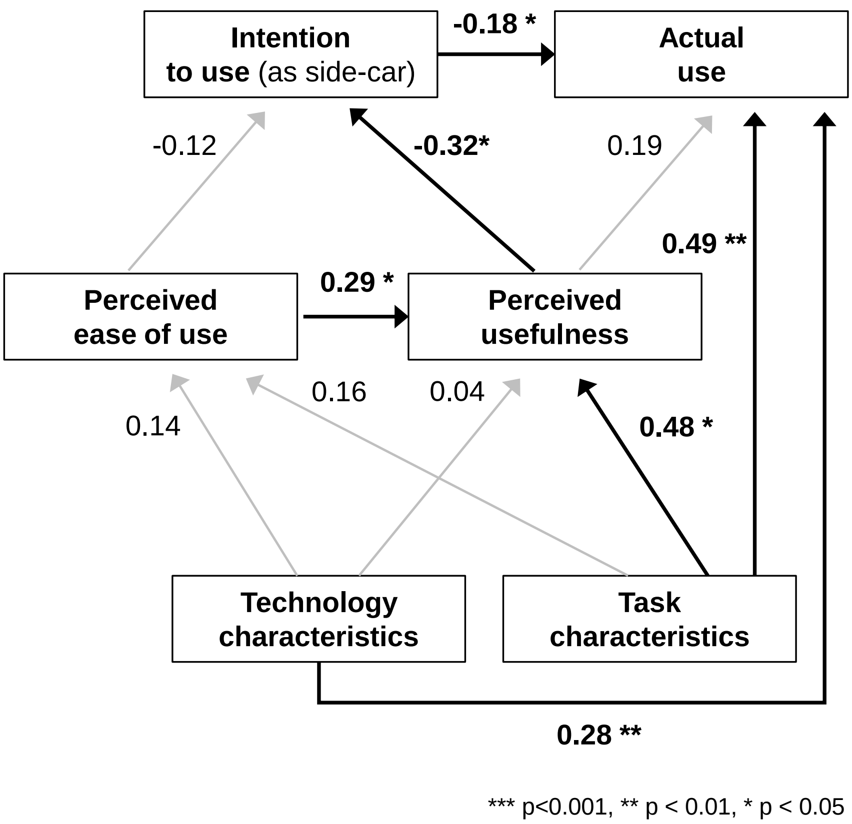


Figure 25. Sub-model (3) path coefficients

[35] Rationale for use and level of support were kept despite their low loadings since they improved model reliability and statistical significance of the paths from perceived usefulness and since composite reliability was not affected much.

Table 7. Sub-model (3) loadings

	PU	TskChar	TecChar	PEoU	ItU	Use
CU2	0	0	0	0	0	0.90
CU3	0	0	0	0	0	0.75
AppD1	0	0	0	0	0	0.74
AppD2	0	0	0	0	0	0.86
DEC1	0.85	0	0	0	0	0
DEC2	0.95	0	0	0	0	0
DEC3	0.93	0	0	0	0	0
RfU5	0.33	0	0	0	0	0
RfU7	0.39	0	0	0	0	0
SUP3	0.35	0	0	0	0	0
SUP5	0.26	0	0	0	0	0
FAM1	0	0	0	0.83	0	0
FAM2	0	0	0	0.94	0	0
FAM3	0	0	0	0.90	0	0
FAM4	0	0	0	0.83	0	0
AppA1	0	0.78	0	0	0	0
AppA2	0	0.66	0	0	0	0
AppA3	0	0.62	0	0	0	0
AppA4	0	0.69	0	0	0	0
FRQ1	0	0.82	0	0	0	0
FRQ2	0	0.58	0	0	0	0
FRQ3	0	0.74	0	0	0	0
DAT1	0	0	0.64	0	0	0
DAT2	0	0	0.73	0	0	0
DAT3	0	0	0.59	0	0	0
DAT4	0	0	0.55	0	0	0
MTD1	0	0	0.78	0	0	0
MTD2	0	0	0.84	0	0	0
MTD3	0	0	0.74	0	0	0
MTD4	0	0	0.75	0	0	0
MTD5	0	0	0.79	0	0	0
OWO2	0	0	0	0	1.00	0

Sub-model (4) with all but the TTF interaction term shows that none of the constructs in the model is irrelevant, as they all have a statistically significant influence on one or more constructs (Figure 26). Adjusted R^2 for actual use was again 0.81, but adjusted R^2 for perceived ease of use increased from 0.02 to 0.27 due to the included individual characteristics construct. Composite reliability scores ρ_C were again all above 0.80. Loadings for the measurement model proxies are given in Table 8.

Table 8. Sub-model (4) loadings

	PEoU	PU	AtU	ItU	TskChar	TecChar	IndChar	Use
CU2	0	0	0	0	0	0	0	0.90
CU3	0	0	0	0	0	0	0	0.75
AppD1	0	0	0	0	0	0	0	0.74
AppD2	0	0	0	0	0	0	0	0.86
DEC1	0	0.84	0	0	0	0	0	0
DEC2	0	0.94	0	0	0	0	0	0
DEC3	0	0.92	0	0	0	0	0	0
RfU5	0	0.36	0	0	0	0	0	0
RfU7	0	0.41	0	0	0	0	0	0
SUP3	0	0.35	0	0	0	0	0	0
SUP5	0	0.26	0	0	0	0	0	0
OWO2	0	0	0	1.00	0	0	0	0
FAM1	0.84	0	0	0	0	0	0	0
FAM2	0.94	0	0	0	0	0	0	0
FAM3	0.90	0	0	0	0	0	0	0
FAM4	0.82	0	0	0	0	0	0	0
DES1	0	0	0.86	0	0	0	0	0
DES2	0	0	0.69	0	0	0	0	0
DES3	0	0	0.78	0	0	0	0	0
AppA1	0	0	0	0	0.78	0	0	0
AppA2	0	0	0	0	0.66	0	0	0
AppA3	0	0	0	0	0.62	0	0	0
AppA4	0	0	0	0	0.68	0	0	0
FRQ1	0	0	0	0	0.82	0	0	0
FRQ2	0	0	0	0	0.58	0	0	0
FRQ3	0	0	0	0	0.75	0	0	0
DAT1	0	0	0	0	0	0.64	0	0
DAT2	0	0	0	0	0	0.73	0	0
DAT3	0	0	0	0	0	0.58	0	0
DAT4	0	0	0	0	0	0.55	0	0
MTD1	0	0	0	0	0	0.78	0	0
MTD2	0	0	0	0	0	0.84	0	0
MTD3	0	0	0	0	0	0.74	0	0
MTD4	0	0	0	0	0	0.75	0	0
MTD5	0	0	0	0	0	0.79	0	0
DEMO_code2	0	0	0	0	0	0	1.00	0

Due to the relatively small number of observations, subsets for moderating effects of tools and driver could not be tested with one of the four sub-models presented in this section. In order to see if there was any effect of driver at all, a much simplified model with only the three constructs task characteristics, perceived usefulness, and actual use was tested with the two subsets for ML&A driver: top-down and bottom-up. There was a noticeable change in the path coefficient from perceived usefulness to actual use, which shows that such factors may have an influence on the model parameters. Therefore, this observation will be addressed in the discussion at the end of this work.

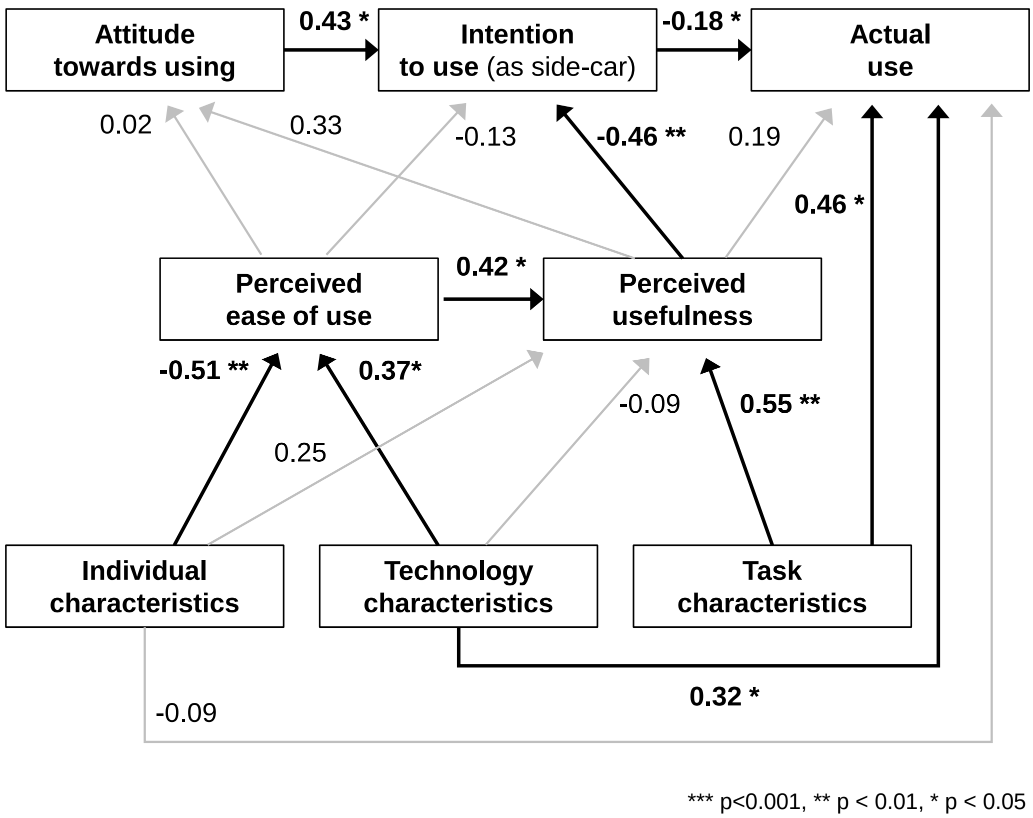


Figure 26. Sub-model (4) path coefficients

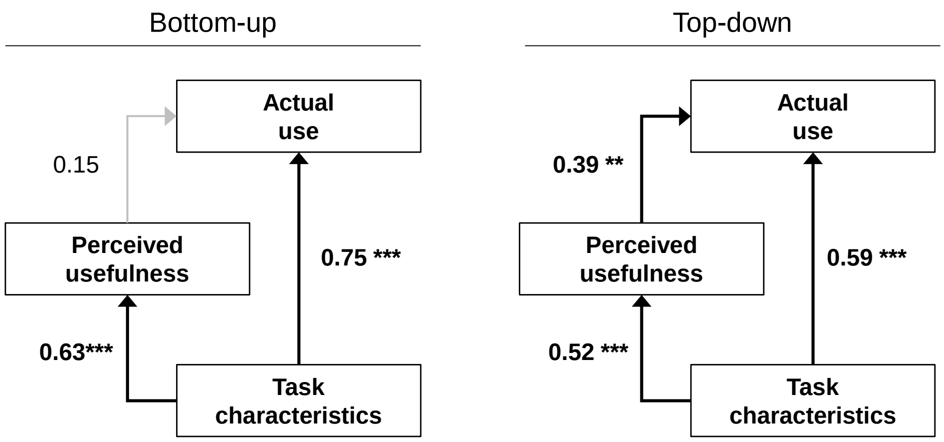


Figure 27. Moderating effect of bottom-up and top-down driver

In the following section, a couple of conclusions will be drawn based on the constructs and their relations to each other and to actual use of ML&A.

3.3.3. Synthesis

From the path coefficients in a model, the total effect of the constructs on actual use can be calculated (Bollen, 1987) by summing over the direct and indirect effects. For instance, the total effect for intention to use is equal to its path coefficient to actual use while the total effect for perceived usefulness is equal to the sum of its direct effect and the indirect effects via attitude towards using and intention to use. More generally, in recursive models without loops or bidirectional relations, the total effect can be calculated as follows. Let $\mathcal{M}$ be the set of all mediators for one construct i, $w_{i,0}$ the path coefficient between construct i and the dependent variable (in this case actual use) of a model, and $w_{i,j}$ the path coefficient between construct i and its mediators $j \in \mathcal{M}$, then the total effect is

$$TE_i = w_{i,0} + \sum_{j \in \mathcal{M}} w_{i,j} \cdot w_{j,0}$$

Table 9 shows the total effects for all constructs in sub-model (4). Task characteristics have the strongest effect, followed by technology characteristics and perceived usefulness. Hence, the higher the perceived usefulness of analytics, the more likely is a company using ML&A side-by-side or full-scale for certain use cases. The negative total effect of intention to use (as a side-car) also allows for a reasonable interpretation: intending to use ML&A only as a side-car limits the actual use to situations in which the machine can easily be used alongside an existing approach without too much effort. On the other hand, a full-scale implementation requires more commitment than just the intention to use as a side-car.

Similarly, a straightforward interpretation for the negative path coefficient from individual characteristics to perceived ease of use is: The higher a manager is in the hierarchy (that is a high value for individual characteristics), the lower the perceived ease of use. On the other hand, perceived ease of use, in particular familiarity with ML&A, has a positive effect on actual use.

Generally, the model explains more than 80% of the variance in the actual use data and can be considered a good fit without the danger of overfitting. Figure 28 summarizes the results with all statistically significant path coefficients and the total effect of each construct on actual use. Note that the total effects were calculated including the mediation via non-significant path coefficients as indicated in column three of Table 9.

Table 9. Total effects on actual use in sub-model (4)

	Actual use	Components
Attitude towards using	-0.07	Indirect (ItU)
Individual characteristics	-0.10	Indirect (PEoU, PU)
Intention to use	-0.18	Direct
Perceived ease of use	0.13	Indirect (PU, Atu, ItU)
Perceived usefulness	0.27	Direct, indirect (AtU, ItU)
Technology characteristics	0.35	Direct, indirect (PEoU, PU)
Task characteristics	0.59	Direct, indirect (PU)

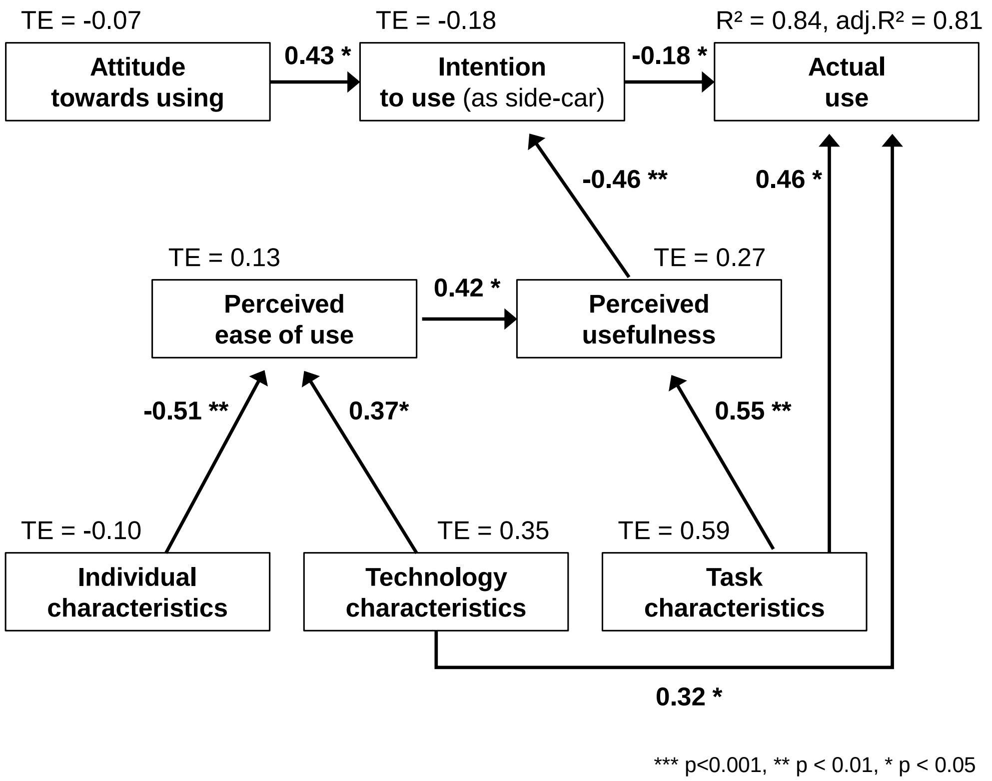


Figure 28. Final model, significant paths, and total effects

4. Use case fundamentals: There is a range of algorithms for different problem types

In the following, the algorithms used in the remainder of this work, especially the use cases in chapters 5 and 6, will be introduced in detail. Certainly, there are countless algorithms that could have been applied to the use cases as well, but were left out mostly due to practical considerations, such as required input format or missing experience in the project team, and the prototype nature of the use cases. Therefore, for a more comprehensive overview, the reader is referred to books like Witten et al. (2016), James et al. (2013), or Bishop (2006).

In general, three categories of problems can be distinguished: regression and classification (both supervised learning[36]), and clustering (unsupervised learning). Sometimes association or time-series forecasting are added as further categories. *Regression* aims at identifying underlying structures in data with continuous outcome variables (Kuhn and Johnson, 2013). Models for regression include linear regression and penalized versions of it like the LASSO (section 4.1), less simple models like support vector machines, *k*-Nearest-Neighbors (section 4.3[37]), gradient boosting (section 4.4) and decision trees, neural networks (section 4.5), and, in the case of time series data, exponential smoothing and ARIMA models (section 4.2). In *classification* problems, the outcome variable is of a categorical nature (Kuhn and Johnson, 2013). Some of the models for regression problems can also be applied to classification, like *k*-Nearest-Neighbors, neural networks, support vector machines, and decision trees. Additionally, there are specific classification algorithms like discriminant analysis. *Clustering or segmentation* aims to partition the data into subgroups or clusters of similar properties (James et al., 2013). Algorithms for this task include, among others, *K*-means and hierarchical clustering. Finally, *association* (rule mining) seeks to identify relations (rules) between variables in data sets (Agrawal et al.,

[36] See section 1.1.2 for descriptions of supervised and unsupervised learning.

[37] In this work, *k*-Nearest-Neighbors is used for classification purposes, but it can also be applied to regression problems.

1993). For the use cases in the remainder of this work, only regression and classification algorithms will be applied.

4.1. Linear regression and LASSO

Linear regression is one of the most basic statistical techniques for identifying relationships between a dependent variable and one or more independent variables. The aim of regression analysis is to fit a plane (in as many dimensions as there are independent variables) to the data such that the sum of squared residuals is minimized (Montgomery et al., 2012). This can be written as[38]

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{N} ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||_2^2 \right\} \tag{1}$$

where $\boldsymbol{\beta}$ denotes the $p \times 1$ vector of coefficients for the independent variables, N the number of observations, $\mathbf{y}$ the $N \times 1$ vector of dependent variable observations, $\mathbf{X}$ the $N \times p$ matrix of independent variable observations, and $||\cdot||_2^2$ the squared L_2 norm.

While linear regression is easy to use and understand, one major disadvantage is the assumption of linearity. If the true relationship is indeed not linear, the accuracy of the model suffers considerably and conclusions may be biased (James et al., 2013). Additionally, linear regression cannot cope with overdetermined models. If the number of observations N is smaller than the number of independent variables p, which is often the case in practical applications, the solution to equation (1), which can be written as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{2}$$

is not unique as the matrix $\mathbf{X}^T\mathbf{X}$ does not have full rank and cannot be inverted. However, even if N is larger than p, interpretability of the results is limited if there are plenty of non-zero coefficients.

[38] In this work, normal letters (Greek and Latin) denote scalars, bold lower case letters denote vectors, and bold upper case letters denote matrices.

Addressing these weaknesses of poor interpretability or singularity, Tibshirani (1996) proposed the least absolute shrinkage and selection operator (LASSO) which imposes a constraint on the coefficients in addition to the normal linear regression. This L_1-norm penalty term leads to a selection of those explanatory variables that fit the model best, coefficients for other variables are set to zero (instead of almost zero as is the case in linear regression) and therefore helps avoiding an overdetermined model. The linear regression part of LASSO is the same as stated above (Hans, 2009), which leads to the following equation for regression using LASSO:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{\frac{1}{N}||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_1\right\} \tag{3}$$

where λ is a tuning parameter for a more severe (higher values) or more relaxed (lower values) L_1 penalty. In practice, the value for λ is usually determined via cross-validation (see section 4.6). The resulting coefficient vector is most often sparse[39].

Overall, there is a wide variety of methods for subset selection (e.g., stepwise forward or backward), shrinkage (e.g., ridge regression) or dimension reduction (e.g., principal component analysis). However, at this point, only those methods used in the remainder of this work will be described in more detail. For all others, the reader is referred to James et al. (2013) or Friedman et al. (2001)

4.2. Exponential smoothing, Holt-Winters, and ARIMA

Exponential smoothing models are amongst the simplest and most widespread time series models in economics, finance and business analysis (Chatfield and Yar, 1988). First proposed in the 1950s and 60s by Brown (1959), Holt (1957), and Winters (1960) they motivated a number of very successful forecasting methods. Overall, in exponential smoothing weighted averages of past

[39] Besides unique solutions to underdetermined systems, sparse vectors and matrices can also be saved and processed much more efficiently in computer programs (Gilbert et al., 1992). Instead of keeping every entry it is possible to save only the non-zero ones and their respective index.

observations are summed, while weights for older observations decrease exponentially (Hyndman and Athanasopoulos, 2018).

For example, with additive trend and seasonal component, exponential smoothing can be written in its recursive form as

$$\hat{y}_t = \alpha y_{t-1} + \alpha(1-\alpha)\hat{y}_{t-1} \tag{4}$$

which expands to (Hyndman and Athanasopoulos, 2018)

$$\hat{y}_t = \alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + (1-\alpha)^2 y_{t-3} + \dots \tag{5}$$

where α denotes the smoothing parameter, $\hat{y}_t$ the smoothed estimate at time t, and y_{t-i} the observed value $i = 1..t$ periods ago.

Based on a classification by McCormick (1969), there are 15 combinations of additive or multiplicative trend and seasonal components. *Holt-Winters* (HW), as one of those extended forms of simple exponential smoothing, approximates data with the help of a local mean or level (l_t), a local trend (b_t), and a local seasonal factor (s_t) (Winters, 1960; Chatfield and Yar, 1988)[40]. The additive form of the HW model comprises three terms for component updates (6b-6d) and one for the new estimate of the time series (6a):

$$\hat{y}_{t+1} = l_t + b_t + s_{t-m+1} \tag{6a}$$

$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} - b_{t-1}) \tag{6b}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{6c}$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \tag{6d}$$

where m denotes the periodicity of the seasonality, e.g., 1 for yearly and 4 for quarterly effects (Hyndman and Athanasopoulos, 2018). Equation (6a) shows the additive nature of this form of HW model. The component updates all make use of the decomposition into the three components, e.g., the seasonal component (s_t) is composed of the weighted sum of what would be the seasonal component

[40] It is also called double exponential smoothing, as seasonal and trend component are each modeled with an exponential term (Goodwin, 2010). There are additional models like triple exponential smoothing to deal with multiple seasonal patterns, e.g., a weekly and a yearly seasonality, which will not be addressed in this work.

with last period's mean and trend $(y_t - l_{t-1} - b_{t-1})$ plus the applicable seasonal component m periods ago, s_{t-m}.

A different approach to modeling time series data are autoregressive-integrated-moving-average *(ARIMA)* models, which describe the autocorrelation in the given data set instead of the mean, trend, and seasonal component (Hyndman and Athanasopoulos, 2018). In order to do so, they use the autocorrelation and the partial autocorrelation function (Zhang, 2003). The three parts of ARIMA all address different possible characteristics of a time series. (1) A time series is said to be autoregressive if its value at time t depends on one or more (in the following p with $p \geq 1$) of its past values (Walker, 1931; Hyndman and Athanasopoulos, 2018)

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t \quad (7)$$

where *c* denotes an arbitrary constant, φ_i the model parameters, ε_t a white noise error term, and p the order of the autoregressive model.

(2) In contrast, in a moving average model, the time series at time t depends on one or more (in the following q with $q \geq 1$) of its past (white noise) error terms (Theodoridis and Koutroumbas, 2009)

$$y_t = c + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (8)$$

where θ_i denotes the model parameters and q the order of the moving-average model.

(3) Integrated refers to differencing[41] a time series if it is not stationary[42], i.e., its properties change over time (Enders, 2008). Depending on the characteristics of the data, differencing may be applied to the entire time series or only to seasonal or trend components

[41] Differencing means computing the difference between consecutive observations throughout the entire time series (Hyndman and Athanasopoulos, 2018).

[42] Stationarity of a time series is most often tested with unit root tests like the Dickey-Fuller test or adaptations of it (Dickey and Fuller, 1981; Leybourne, 1995), which will not be part of this thesis.

$$\Delta y_t = y_t - y_{t-1} \tag{9}$$

Note that the resulting differenced time series has one value less than the original time series. For certain applications, second order differencing (differences of the differences in equation (9)) may be necessary. Hence, the order of differencing, d, is indicated in ARIMA models as well.

With the three parameters, p, d, and q, an ARIMA model can be specified as follows and applied to the data

$$\Delta y_t = c + \sum_{i=1}^{p} \varphi_i \Delta y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{10}$$

4.3. K-nearest neighbors

The *k*-nearest neighbors (*k*-NN) algorithm is among the simpler classification algorithms. In this method, a new object is assigned the class that occurs most frequently among the *k* observations in a given radius or neighborhood of the new observation (Cover and Hart, 1967). To determine this neighborhood, different distance measures can be used, among them the Euclidian distance between the two vectors $\mathbf{x_1}$ and $\mathbf{x_2}$

$$d(\mathbf{x_1}, \mathbf{x_2}) = \sqrt{(\mathbf{x_1} - \mathbf{x_2})^T (\mathbf{x_1} - \mathbf{x_2})} \tag{11}$$

Let $T = (\mathbf{x}_i, y_i), i = 1..N$ be the training set with N observations where $\mathbf{x}_i \in \mathbb{R}^{\boldsymbol{p}}$ are the training data (vectors) and $y_i \in \mathbb{R}$ the training labels (scalars). Also, let $\tilde{\mathbf{x}} \in \mathbb{R}^{\boldsymbol{p}}$ be the vector of observed data to be classified and $\tilde{y} \in \mathbb{R}$ the class label to be determined. With this, the objective function for *k*-NN classification is

$$\tilde{y} = \arg\max_{y} \sum_{(\mathbf{x}_i, y_i) \in \boldsymbol{N}_{\tilde{\mathbf{x}}}} \delta(y = y_i) \tag{12}$$

where $\delta(\cdot)$ is the Dirac delta function that returns 1 if $y = y_i$ and 0 otherwise, and $N_{\tilde{\mathbf{x}}} \subseteq T$ is the neighborhood of the observed data to be classified.

As opposed to the simple *k*-NN algorithm, the *distance-weighted k*-nearest neighbor rule (*dwk*-NN) weighs the contribution of each of the *k* neighbors according to their distance to the new observation, giving greater weight to closer neighbors (Dudani, 1976). The weights are taken from the interval [0 to 1]. The closest neighbor gets a weight of 1, the furthest of 0 and the others are scaled linearly on the interval in-between. Equation (12) then becomes (Gou et al., 2012)

$$\tilde{y} = \arg\max_{y} \sum_{(\mathbf{x}_i, y_i) \in \boldsymbol{N}_{\tilde{\mathbf{x}}}} \widetilde{w}_i \delta(y = y_i) \tag{13}$$

On a side note, a general problem with *k*-NN and *dwk*-NN algorithms is the quickly increasing complexity when searching for nearest neighbors in higher-dimensional space. Addressing this, more efficient searching algorithms were proposed, e.g., based on principal component analysis[43] (McNames, 2001) and search trees.

4.4. Gradient boosting

Gradient boosting is a family of highly customizable algorithms that constructively form an ensemble of models (Natekin and Knoll, 2013)[44]. Boosting refers to the process of iteratively deriving weak hypotheses (so-called base learners) about the distribution of a selected subset of the data to finally combine them into an accurate prediction (Freund and Schapire, 1997). In each iteration, the new base learner is (numerically) chosen such that it has the highest possible correlation with the negative gradient of the loss function, therefore its name gradient boosting (Friedman, 2001). Over the last decades, gradient boosting algorithms have been derived for a number of different loss functions with a continuous response (regression) variable or with a categorical response (classification) showing accurate predictions and/or good performance at variable selection

---

[43] Principal component analysis is an approach for dimensionality reduction or feature extraction with the help of an orthogonal projection into lower dimensional space (Bishop, 2006).

[44] As opposed to other ensemble methods like random forests or manual ensembles (see section 4.6) where ensembles are averaging "without a strategy".

(Schmid and Hothorn, 2008). *Extreme gradient boosting* (XGbar[45]) is an algorithm that fits an ensemble of decision trees in order to perform predictions and is well suited to time series forecasting (Chen and Guestrin, 2016).

The prediction algorithm can be described as follows. Let $T = (\mathbf{x}_i, y_i), i = 1..N$ again be the training set where $\mathbf{x}_i \in \mathbb{R}^p$ are the training data and $y_i \in \mathbb{R}$ the training labels. The algorithm uses *K* additive functions f_k out of the space of regression trees to predict the labels

$$y_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i) \tag{14}$$

The objective $\mathcal{L}$ of the algorithm then can be written as (Chen and Guestrin, 2016)

$$\mathcal{L} = \sum_{k=1}^{K} l\left(y_i, \hat{y}_i^{(k-1)} + f_k(\mathbf{x}_i)\right) + \Omega(f_k) \tag{15}$$

where l is a differentiable, convex loss function, $\hat{y}_i^{(k)}$ is the prediction of the *i*-th observation at the *k*-th iteration and $\Omega(f_k)$ is a penalty term to avoid overfitting. From equation (15), the greedy nature of the algorithm can be seen in that it iteratively adds the regression tree f_k that improves the prediction the most.

4.5. Neural networks

Artificial neural networks (ANNs) are a simple mathematical representation of the human brain[46] based on neurons arranged in an input layer, one or more hidden layers, and an output layer (Figure 29). These neurons (or nodes) calculate the weighted sum of their input values and return a thresholded version of that result (Dinov, 2018).

[45] XGbar is an open source project and is applied in many forecasting competitions. For example, 17 out of the 29 winning solutions published on Kaggle's blog in the 2015 competitions used XGbar. Eight thereof exclusively used XGbar to train the model while the remaining nine combined it with neural networks (Chen and Guestrin, 2016).

[46] First proposed by McCulloch and Pitts (1943), the idea quickly gained momentum due to its flexibility for different applications. However, until the 1960s only one hidden layer was considered, which resulted in a decline in interest after Minsky and Papert (1969) proved serious flaws in the predictive capabilities of ANNs. Only in the late 1980s research interest returned with the proposition of backpropagation to train multi-layer ANNs and increasingly powerful computers (Abbott, 2014).

Thresholding is realized with a specific activation function, e.g., a sigmoidal function like the logistic $(1/(1-e^{-x}))$ or the hyperbolic tangent $((e^{x}-e^{-x})/(e^{x}+e^{-x}))$ function (Zhang and Zhang, 1999)

$$y = S(\mathbf{w}^T\mathbf{x} + b) \tag{16}$$

where *y* is the output, *S* an arbitrary sigmoid function, **x** the input vector, **w** the vector of weights given to each input and *b* the bias of the neuron. This way, the information is fed forward from input to output through the network.

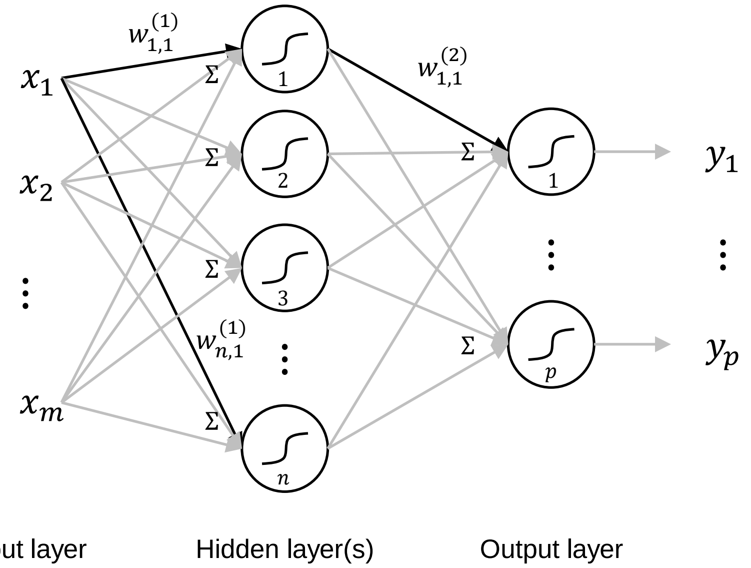


Figure 29. Basic structure of an artificial neural network (modified after Abbott (2014); Dinov (2018))

ANNs can handle both linear and complex nonlinear relationships within the data depending on the network structure and the number of hidden layers (Zhang, 2003; Hyndman and Athanasopoulos, 2018). Adaptation of an ANN to a certain task is realized by iteratively adjusting the weights and biases at each neuron until a minimum in the objective function is reached (Abbott, 2014). Weights are updated via error backpropagation (Dreyfus, 1990), which is a form of supervised learning and a special case of gradient-descent optimization.

New weights are calculated as

$$w_{j,i,new}^{(l)} = w_{j,i,old}^{(l)} - \eta \frac{\partial E_l}{\partial w_{j,i,old}^{(l)}} \tag{17}$$

where $w_{j,i,old}^{(l)}$ is the current weight between node i in layer $l-1$ and node j in layer l, η is an arbitrary tuning parameter (so-called learning rate), and E_l is the error function at node l.

In the remainder of this work, a *multi-layer perceptron* (MLP) and an *extreme learning machine* (ELM) will be used. Thus, descriptions will be provided only for those two types of ANN. For other types like recurrent or convolutional neural networks, see, e.g., Dinov (2018), Goodfellow et al. (2016), or Bishop (2006).

MLPs are among the simplest forms of feedforward ANNs – there are no cycles as in recurrent neural networks and all nodes of one layer are connected to all nodes of the layer before and after them (Friedman et al., 2001). Figure 29 shows an MLP with *n* input variables, one hidden layer with *m* hidden nodes, and an output layer with *p* output variables. The flexibility of MLPs lies in the ability to approximate any continuous function (Cybenko, 1989). Although ANNs can be applied to most problems, the accuracy of basic ANNs like MLP was sometimes limited, which is why research has also focused on developing hybrids of MLPs and other models like ARIMA. For instance, the hybrid proposed by Khashei and Bijari (2010) generates the residuals based on an ARIMA model in the first stage and then fits an MLP to the linear and non-linear relationships in the residuals and original data.

ELMs are a form of single-hidden-layer ANN that address one of the major bottlenecks in the application of ANNs, the time it takes to train the network through many repetitions of a gradient-descent algorithm (Huang et al., 2006). Instead of iterating through all weights, the ELM randomly chooses one of the hidden nodes and assigns input weights and a bias. It then calculates the output matrix and corresponding output weights (analytically). Overall, an ELM can easily be a hundred times faster than backpropagation (Huang et al., 2012).

4.6. Auxiliary methods

While the algorithms described before all share the purpose of making predictions or identifying patterns in data, they require data to be in a certain format and often do not provide built-in statistics. Addressing this shortcoming, some auxiliary algorithms for dealing with missing data (imputation) or preventing overfitting and computing significance levels (cross-validation and bootstrap) exist.

4.6.1. Imputation

When collecting samples of data, most often not all observations (or individual records) are complete, that is, some values for some of the variables either have default values that are not helpful in the analysis or missing values. Reasons for this can be manifold, such as incomplete sampling, deliberate omission to protect privacy, or a subject’s inability to provide an answer (Horton and Kleinman, 2007). In general, there are three different patterns of missing data, missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), which need to be treated differently in an analysis (Enders, 2010; Little and Rubin, 2019). In the case of MCAR, there is no relationship between the missing value and any of the other data points in the data set, missing or observed. For instance, if in a survey a person chooses not to provide details about her age, this is not related to the age, gender, or any other variable in the model. In contrast, in the case of MAR, the missing value is conditional on one or more variables in the model. Therefore, the missing value can (and should) be imputed accordingly to avoid bias. Thirdly, in the case of MNAR, the missingness is related to the missing data itself. For example, respondents below 18 years of age may not provide their age if the survey is only meant for adults.

The most straightforward way of handling missing data is to delete observations with missing values. However, this may introduce a bias if data is not MCAR (Schafer, 1999). Hence, in order to address the problem, different methods of imputation have been developed, like substituting with the mean, maximum likelihood estimation (Little, 1992), *k*-NN estimation (Crookston and Finley, 2008), or random forest imputation (Stekhoven and Bühlmann, 2011). The latter iteratively fits random forests to the missing data and works well with both continuous and categorical data.

4.6.2. Cross-validation

Due to limited data samples, in practice, estimating the parameters of a predictive model is most often a trade-off between accurately reconstructing the training data with as many training observations as possible and getting a good feeling for predictive performance with as many validation observations as possible (Bishop, 2006). Cross-validation is a technique of re-using data (*resampling without replacement*) to get the most out of a limited sample and gather all patterns in the data while simultaneously avoiding overfitting (Abbott, 2014). *K*-fold cross validation (CV) divides the sample into *k* disjoint groups (so-called folds) and consecutively treats each one of these groups as a validation set while the remaining *k*-1 groups are used to estimate the parameters of the model (James et al., 2013). At the end, the parameter estimates and error statistics of the *k* folds are averaged. In practice, most often 5 or 10 folds are used[47].

4.6.3. Bootstrapping

In contrast to CV, bootstrapping is a *resampling* method *with replacement* and a form of nonparametric Monte Carlo method (Efron, 1992)[48]. The idea is to estimate the actual probability distribution $(\mathcal{F})$ and related statistics like the mean, median, or confidence intervals based on the empirical probability distribution $(\hat{\mathcal{F}})$ (Efron and Tibshirani, 1993). In brief, the bootstrap estimate $\hat{\theta}$ for a parameter θ is calculated as follows. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ denote a random sample and $\mathbf{x} = (x_1, x_2, \dots, x_n)$ its realization. Draw repeated bootstrap samples with replacement $\mathbf{X}^{(b=1..B)}$ and repeated realizations $\mathbf{x}^{(b)} = \left(x_1^b, x_2^b, \dots, x_n^b\right)$ and calculate $\hat{\theta}^{(b)}$ from them. Then, use these *B* estimates to calculate the estimate of the true parameter's probability distribution (Efron, 1992; Bishop, 2006).

---

[47] The most robust case of *k* equal to the sample size (*n*-fold), where each observation is used once for validation (leave-one-out CV), is computationally too expensive for most applications (James et al., 2013).

[48] In addition to the aforementioned *k*-fold CV, there is also Monte Carlo CV, which allows for varying sizes of the training and validation sets (see, e.g., Girard (1989)). A potential advantage of *k*-fold CV is that all observations are used exactly once as validation, whereas the random subsetting in Monte Carlo CV may leave out single observations.

5. Use case 1: Financial accounting can greatly benefit from an account recommender

Invoices are an integral part of business transactions as they specify the details between a buyer and a seller (Taylor, 1985). Especially when companies are an integrated part of a comprehensive value chain, the yearly amount of invoices can easily reach millions.

Most often, not all incoming invoices can be directly matched to a purchase order indicating an order number, prices, and quantities. Thus, finding the account to be charged in the ERP system for such *(supplier) invoices without a purchase order* is a very repetitive task. For accountants, finding patterns to allocate such invoices to the “right” account according to a predefined rule is a very tedious task. In terms of cost per booking, it is not efficient for companies to use qualified human workers for such a repetitive type of work.

Finding *patterns* in data and, more forward-looking, encapsulating them in theories that can be used for predicting what will happen in new situations is a good use case for automation (Witten et al., 2016). On the one hand, pattern matching is not new: Rules for information extraction and the completion of predefined fields in a template were already addressed in the late 1990s (Califf and Mooney, 1999). On the other hand, automation received new momentum over the last years by leveraging *machine learning* (Onken and Schulte, 2010).

Over the last couple of years, finance departments across organizations most prominently applied automation in general R2R process activities, not yet addressing the potential in accounts payable and accounts receivable (Plaschke et al., 2018). Based on findings from a survey about finance process activities to be digitalized, the focus of this section lies on the invoice processing activity in the P2P process. The objective is to propose first design guidelines[49] for cognitive automation to handle supplier invoices without a purchase order. For that,

[49] In addition to the four types of DSR artifacts identified by March and Smith (1995) and Hevner et al. (2004) – constructs, models, methods, and instantiations – design guidelines contribute to theories that specify how IS artifacts should be designed based on kernel theories (Vaishnavi and Kuechler, 2015). Design guidelines contribute to both models and methods.

an accounts recommender prototype in a leading chemical company serves as case example. The following two research questions are answered:

(1) Compared to manual work, what is the potential of machine learning to improve accounting accuracy and process efficiency?
(2) What are the most important design guidelines to successfully implement and, finally, run machine learning in the accounting domain?

5.1. Method: Machine learning is used to mitigate bottlenecks in daily business

5.1.1. Case study in a global chemical company

Following Dul and Hak (2008) case studies allow researchers to study artifacts in natural settings and observe the situation in which activities take place. Thus, case studies enable researchers to learn from practice (build on people's experiences and practices), understand the nature and complexity of the process, and leverage the possibility of iteratively testing results in a real environment.

In comparison to broader surveys, case studies provide in-depth information, and recognize the complexity and embeddedness of activities (Yin, 2017). Multiple sources for data collection such as documentations and archival records to gather information from one or a few entities are employed. The results are analyzed in a *qualitative manner* (Dul and Hak, 2008). The literature review indicated clear gaps (section 3.1), and a case approach is a proven way to conduct research where only few studies exist to finally *build theory* (Benbasat et al., 1987). In the IS and accounting disciplines, case studies are a generally accepted research approach and have been employed for decades. Especially with respect to value creation through the use of data, a number of data analytics case studies just recently showed evidence of how companies can leverage their own data to successfully generate value (Hopf et al., 2017).

For this part a *single case study* in a *leading chemical company* with revenues of more than 50bn USD and over 50,000 employees served as reference. The financial processes are standardized worldwide and continuously optimized. Global Finance Transformation (GFT) as the central governance unit harmonizes and optimizes financial processes running on a single SAP system. With the help of a virtual network of process experts around the world, GFT is constantly seeking new (digital) technologies and promoting the development of new solutions for global process improvements in the finance department.

Following the tenets of *requirements engineering* (Kotonya and Sommerville, 1998), first, internal documents were examined, followed by four semi-structured expert interviews in the reference company. The statements from these interviews were analyzed along the guidelines of qualitative content analysis (Kohlbacher, 2006; Mayring, 2014) and requirements for a prototype were derived. Interviewees were the head of the O2C process, the senior expert of the R2R[50] process, the head of the P2P process and the senior expert for accounts payable. To find evidence that the prototype outperforms “classic” manual human work (RQ 1), a commonly used machine learning algorithm was implemented (Dietrich et al., 2015).

As described in section 1.1.1, the P2P process generally starts with a requisition, comprises several steps from purchase order to invoice settling and most of the time ends with administrative tasks as illustrated in Figure 2 (Monczka et al., 2015; The Hackett Group, n.d.). Figure 30 is a simplified representation of this process in the reference company highlighting the focus of this section on *invoice processing*. While many rule-based tasks in the P2P process of the reference company are already automated, the actual invoice processing still requires a lot of human input. The company receives around 4.8 million invoices from suppliers per year, out of which over 600,000 (13%) are not purchase-order-related. This means materials and services are requested without a purchase requisition and no purchase order is available. After approval regarding amount, goods received, and supplier of the invoice without purchase order, the invoice is posted to a manually determined general ledger (GL) account.

[50] In the reference company, this process is called book-to-report

GFT set up a project team to develop a machine learning prototype that suggests a GL-account matching the content of the invoice (account recommender). This should lead to both an accuracy increase in accounting and a more efficient P2P process by reducing the manual workload.

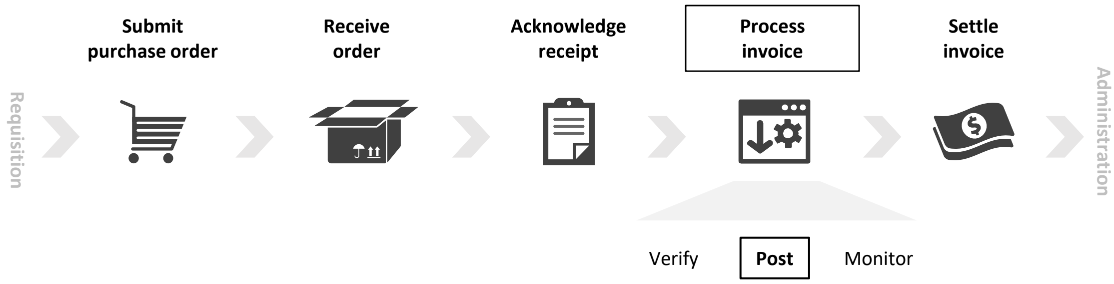


Figure 30. P2P process of the reference company

5.1.2. Requirements engineering

In addition to non-functional requirements for usability and response times (Sommerville, 2007), the following *six functional requirements* were derived:

(1) New invoices without purchase order should be transmitted to the machine learning cloud right after arrival. They should be taken out of the standard workload basket for accounts payable and the usual robotic process automation should be stopped. There should be a possibility to still do the process steps manually in case of urgent issues or a faulty interface.
(2) The account recommender should be able to handle all invoices, regardless of type and issuer.
(3) Determine the matching GL-account by comparing the billing attributes of the new invoice with the historical data and their related postings to identify the best-fitting account.
(4) If the estimate for the account has a likelihood of 80% or higher, directly write the account to the ERP system. If the likelihood is below 80%, return the three best options as a choice.
(5) The approver should have the ability to override all propositions, even those above 80% likelihood.
(6) The algorithm should remember account choices for retraining, which takes place every month

Based on these requirements, a data set was put together using historical invoice data of more than 2 million records in the period from 2010 to 2018. Each invoice includes over 200 fields. One of the most important steps in the preparation of data is the selection of *relevant features*. The reduction in the number of variables (the rejection of attributes that are weakly correlated with the target variable) not only increases the accuracy of the prediction, but also lowers the requirements for the computing resources. The potentially most relevant variables for the use case were chosen after getting acquainted with the process of invoice processing. 11 fields were gathered that are necessary for the AP employees to choose a matching GL account. Here "GL_ACCOUNT" is the dependent variable and the remaining features are independent variables and listed in Table 10.

Table 10. Relevant invoice fields (independent variables)

Field name	Description	Examples
WC_USER	Invoice approver name	SMITHJ
VENDOR_NO	Account number of vendor or creditor	123456
REF_DOC_NO	Invoice number from vendor system	XX 12345
COMP_CODE	Company code	US01
CURRENCY	Currency key	USD
NET_AMOUNT	Net amount in document currency	500
GROSS_AMOUNT	Gross amount in document currency	600
INVOICE_IND	X if it is a normal invoice, blank if it is something else (e.g., a credit note)	X
CP_DOC_TYPE	Defines a kind of invoicing process	ID
SUPCOUNTRY_ISO	Country of the invoice (from vendor side)	US
DOC_DATE	Date of invoices	20180101

Based on these features, the supervised learning algorithm predicts the matching GL-account for a new invoice without purchase order. Following the reasoning that some fields are more relevant for prediction than others, different weights were iteratively assigned to them. Afterwards, the full dataset was randomly split into K subsets of approximately equal size for *K-fold cross-validation*. K-1 blocks were used to estimate the parameters of the model (train) and the one remaining block was used for testing the model's accuracy (test). The process was repeated K times, and each of the blocks was used once as a test set. Finally, the parameters of the resulting K models were averaged to get one estimate.

5.2. Results: Processing invoices with missing data becomes easier and faster

5.2.1. Accuracy vs. coverage trade-off

The processing of invoices (step 4 in Figure 30) always starts with data extraction and validation (Figure 31). If all data are available, the standard approval process is triggered. Otherwise, the type of data that is missing needs to be identified. In case of a missing GL account, the *account recommender* performs the necessary actions of (1) preparing the data, (2) computing a *k*-NN estimate, and (3) returning the GL account. Note that if the confidence levels for a single GL account are too low, the account recommender instead returns three candidate accounts from which the user has to select one. For missing data other than the GL account, manual handling is still necessary. However, missing approvers have already been identified as an extension (see evaluation in section 5.3).

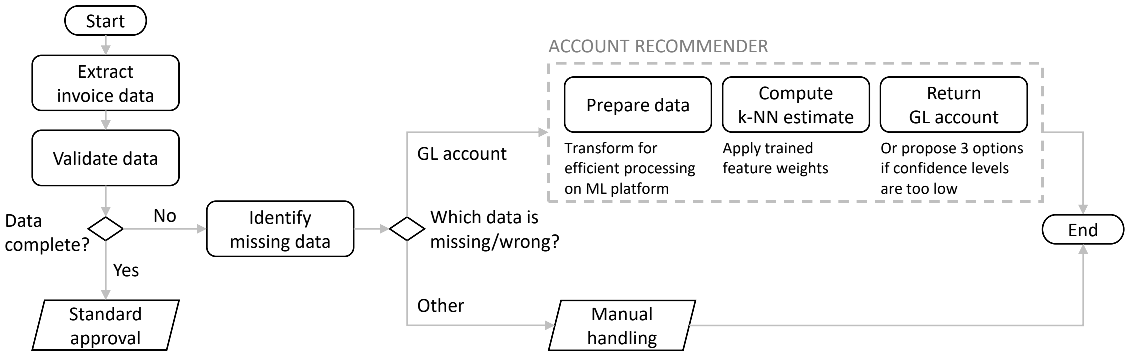


Figure 31. Invoice processing for invoices with missing data

To determine the *optimal number of neighbors*, the percentage of correct predictions at *k* from 1 to 20 was computed. Figure 32 shows the results, e.g., if *k* = 1 is chosen, the accuracy of the proposed forecast will be approximately 60%, whereas in the case of *k* = 20, the accuracy is 88%. A decision for k should be a trade-off between the following two aspects: On the one hand, *k* should be large enough to avoid noisy decision boundaries that occur at very small *k*. On the other hand, it should be small enough so that only nearby samples are included. Choosing *k* too large will lead to over-smoothed boundaries and longer computation times without a proportional accuracy increase.

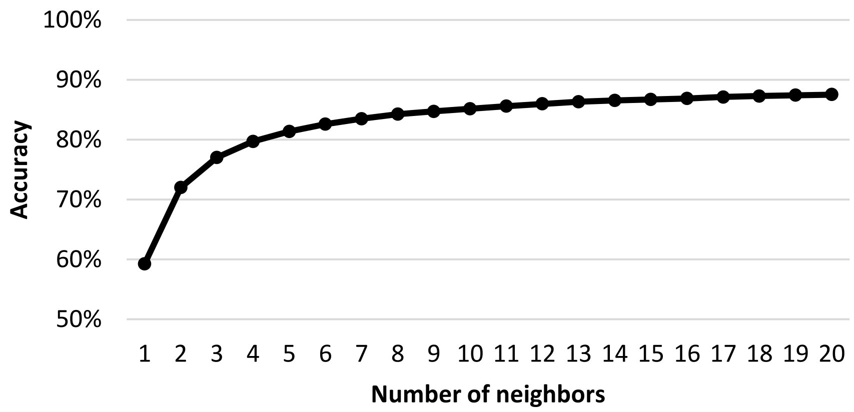


Figure 32. Accuracy vs number of neighbors

One key tuning parameter in the prototype is *coverage*, which means the ratio of invoices that are handled automatically by the prototype versus those for which three candidates are proposed. The performance was tested for three scenarios, 50%, 75%, and 100% coverage. Figure 33 clearly shows that it is impossible to have high accuracy and high coverage at the same time. While the accuracy for 50% coverage is around 87%, it is only 59% for 100% coverage. In the trade-off scenario, the accuracy is 72%.

Currently, the manual process requires at least five full time employees and can take up to one day of time. In contrast, for the three steps indicated in Figure 31, the account recommender only needs a few seconds. However, *training* the model and retraining it with delta loads every week consumes more time and is quite resource-intensive. On a standard laptop, the process of initial model fitting and estimating the test set takes around 24 hours, which is a strong argument for cloud computing.

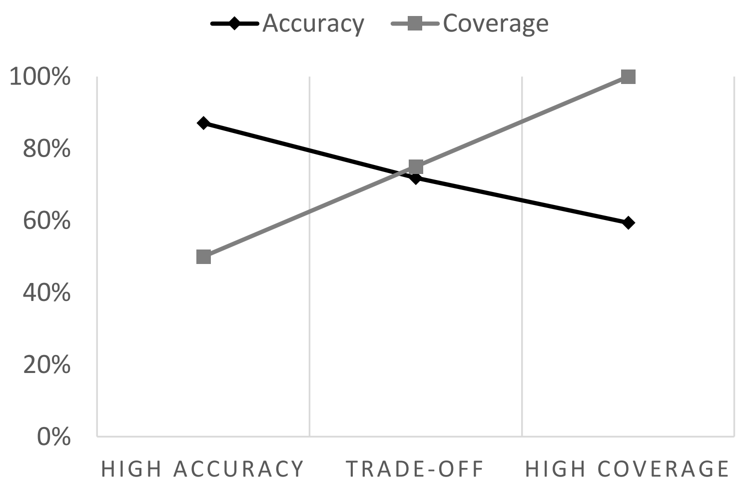


Figure 33. Accuracy and coverage of three prediction models

5.2.2. Guidelines

Based on the results described in the previous section and insights gained during the development of the prototype, the following five design guidelines were synthesized.

Design guideline 1: Truly understand the business need. Get yourself acquainted with the process and automate with the best possible user experience in mind.

This first guideline addresses the fact that with a sensitive topic like replacing human labor with machines, the solution has to be all the better. More specifically, turning affected and most often reluctant accounting clerks into engaged innovators may be the best facilitator for this transformation.[51] However, this can only be achieved if the developer has a good business understanding.

Design guideline 2: More data can only get you so far. While a broader training set generally leads to more accurate predictions, do not forget to provide good-quality data.

[51] The SSC employees of the reference company were the ones to identify use cases and develop an agenda for the broad introduction of (rule-based and cognitive) automation in their daily business.

Especially when training machines based on historical data, the quality of the training examples is a crucial factor[52]. Although some algorithms are relatively robust against missing or wrong data, the risk of replicating past mistakes hundreds of times always exists. Hence, cleansing the initial training set from flawed data is still preferable.

Design guideline 3: Give the machine a good starting position. Using a combination of expert judgment and machine learning for feature selection and assigning weights speeds up the training process.

Nowadays, most machine learning algorithms come paired with powerful feature selection and iteratively adjust parameters with methods like gradient descent (cf. section 4). Nevertheless, most of the approaches to feature and weight selection are heuristics that start with a random or an all-equal-to-one initial configuration and often only reach a local optimum. In the presence of experts who have performed the task hundreds of times, initial configurations can be deduced that improve speed towards and accuracy of the final configuration.

Design guideline 4: Computing power is crucial. Processing large amounts of data with machine learning requires resources that a cloud-based architecture is better suited for than local hardware.

Among the reasons why machine learning was not as prominently applied as it is today are the increasing hardware requirements with increasing data volumes and complexity of tasks. Although business users were used to data loads over the night from traditional data warehouses, they are no longer as patient. While there has been significant progress in terms of single processing units, parallelization has had an even bigger impact on performance increases. Thus, distributing the workload of a machine learning use case over several machines is the most reasonable choice. A cloud architecture provides just that, which is why the reference company chose to implement the prototype on one of the leading cloud platforms.

---

[52] Training machine learning algorithms with biased and incomplete data has been subject to research for more than twenty years (Cortes et al., 1995). In the context of big data analytics, patching and cleansing data in real time has become an important tool (Saha and Srivastava, 2014).

Design guideline 5: Do not burn your bridges yet. Keep a fallback solution where human workers can override propositions and help the machine learn from experts' decisions.

Despite the intelligence attributed to the machine and its far superior speed, there are always situations that require human judgement in a process as complex as invoice processing. As a result, in many cases the machine can only provide likely alternatives or make a choice based on probabilities of past data. Therefore, the accountant should still have the option of overriding entries. Additionally, this may help build confidence since the accountant can experience first-hand what the machine does well and where it lags behind. In turn, this knowledge can then be used to adjust parameters and improve the accuracy and coverage of the algorithm (in a reinforcement learning approach).

5.3. Evaluation: Amortization is less than two years and further applications are easy to identify

Gregor and Hevner (2013) suggest to evaluate artifacts along a number of dimensions such as validity, utility, quality, and efficacy. Besides these goal- and activity-related dimensions, Prat et al. (2014) add environment, structure, activity, and evolution as further evaluation criteria. The following section will elaborate on the evaluation of the proposed account recommender prototype based on a selection of these criteria – namely *validity* (i.e. does the artifact do what it is meant to do), *utility* (i.e., is the artifact useful beyond its original development environment), and *evolution* (i.e., is the artifact robust and can adapt to new requirements).

Validity of the prototype has been demonstrated in the results section (5.2.1). The prediction accuracy of the prototype was very promising in its original case environment. An *implementation* for daily use in the reference company is ongoing. The head of the GFT department said that the business case for the account recommender is very straight forward and amortization should be within a maximum of two years. Hence, the focus will be on utility and evolution in the following.

A possible *evolution* of the account recommender prototype was evaluated in the reference company with the use case "*approver recommender*." Based on a very similar model in which only the dependent variable "GL_ACCOUNT" is exchanged with "WC_USER" (the approver), this second use case is already on the list of future projects for GFT. Thus, the evolution criterion can be considered as satisfied as well.

In order to assess the *utility* of the prototype, three *interviews* were conducted during a workshop of a manager focus group. Interview partners were the head of accounting of a multinational energy utility company, the head of Digital Finance and a partner of an audit firm who had recently started implementing a number of machine learning prototypes for finance as well. Overall, the feedback of all interviewees was very positive. The head of accounting said that, especially in conjunction with an approver recommender, the automation should mitigate one of the bottlenecks in the finance back office. According to him, the process of determining approver and GL account (with cost center or project reference) can sometimes take up to three weeks. The audit partner pointed out that his company began training a neural network that is fed all information on an invoice, including visuals, text positions, and the text information used in the prototype proposed in this work. He did not have comparable figures yet, but could acknowledge the eagerness of two of his clients to try it out.

6. Use case 2: Cash flow forecasting is one of the big levers in management accounting

Companies use *forecasts* to gauge developments and anticipate uncertainties to predict deviations from their plans (Valentin, 2014). Applying methods from fields such as statistics and machine learning (Mishra and Silakari, 2012), ML&A aim to be a "modern crystal ball" for future events and their impact on financial outcomes such as net sales, earnings before interest and taxes, and cash flow.

As described in section 3.1, there are several articles covering machine learning and analytics (on big data). For example, using ARIMA models, Subramaniyan et al. (2018) develop a period-based algorithm to predict bottlenecks in the production. Fang et al. (2016) propose a new model to forecast customer profitability by applying random forest regression. However, leveraging predictive analytics in the context of cash flow forecasting gained little attention. In practice, most of the forecasting processes still rely on "classic" experience-based human judgement (Bacon-Gerasymenko et al., 2016). Such and other approaches suffer from a number of things, among them (1) limited analytical capacity of human minds, (2) inconsistent use of non-time-series information (e.g., promotional campaigns), (3) the complexity of seasonal growth or decline and basic trends and, finally, the (4) distortions from group processes (Goodwin, 2002).

Combining the best of both worlds, experience-based human judgement and the advantages of a fully data-driven organization leveraging predictive analytics, the objective of this section is to lay out first *design guidelines for a cash flow prediction model*. Taking a company from the utility sector as case example, a more accurate and efficient approach by applying well-proven forecasting methods (Dietrich et al., 2015) and state-of-the-art prediction algorithms is proposed. Due to drastic changes in the utility industry and society in general and the required investments in renewable infrastructure, cash flow improvements are essential to drive change in the utility industry (Jacobson and Delucchi, 2011). Two research questions are answered:

- Compared to experience-based human judgement, do ML&A improve forecasting accuracy and process efficiency?
- Which are first design guidelines to successfully implement ML&A?

6.1. Method: Machine learning and analytics are applied to improve forecast accuracy

6.1.1. Case study in a global energy utility company

Like in the first use case (chapter 5) a *single case* was studied, this time with a utility company as reference. Following Eisenhardt (1989), (1) internal documents were examined, semi-structured expert interviews conducted within the reference company, the statements from these interviews analyzed along the guidelines of qualitative content analysis (Kohlbacher, 2006; Mayring, 2014) and relevant external and internal input factors for the prediction model derived. (2) To find evidence that the model outperforms "classical" experience-based human judgement, commonly used forecasting methods (Dietrich et al., 2015) were complemented with state-of-the-art machine learning algorithms. Modeling and coding was done in the software R[53] (see Appendix E for exemplary R code). (3) Closure was reached once adding further internal and external factors or minor model modifications only led to incremental or no improvements.

The reference company supplies electricity and natural gas to private and business customers, who differ in terms of the billing processes. For business customers, the amount of energy consumed is determined every 15 minutes. Hence, an exact invoice can be issued every month and the calculated payment is received from the customer thereafter. In turn, *private customers* pay a fixed sum each month. This so-called *installment* is calculated based on three things: first, the individual previous year (electricity) or 10-year-average (natural gas)

---

[53] R is a software environment well suited to statistical data analysis and visualization (Ihaka and Gentleman, 1996; R Core Team, 2013). It offers a broad range of libraries (so-called packages) for most statistical applications including forecasting (Gentleman et al., 2004). For a list of packages used in this work, see Appendix C.

consumption, second, the current electricity and natural gas prices, and third, other factors such as a renewable energy surcharge.

6.1.2. Cross-industry standard process for data-mining

To ensure best results, the artifact design was structured along the *Cross-Industry Standard Process for Data Mining* (CRISP-DM; Chapman et al., 2000). It consists of six steps which were applied as follows. In a first step, information for a sound and comprehensive business understanding was gathered by working full-time within the management accounting department of the reference company. In order to ensure in-depth data understanding, the available data were analyzed in a second step. To obtain a usable data set (third step), data preparation covered cleaning actions such as transforming the data into a homogenous format. In a fourth step, a forecasting prototype was implemented (modeling) and revised in several discussions whenever aspects for improvement were identified. Afterwards, the results were evaluated in project group meetings (evaluation). In a sixth and final step, the prototype was transferred to the reference company (deployment) and will be integrated in the day-to-day operational business in late 2019.

Business understanding

Depending on the month of the year, there are different effects on the operating cash flow – defined as the cash flow generated through everyday activities (Walker, 2009): (1) C*ustomer perspective:* While much energy is delivered to customers in the cold seasons, there is a significant reduction in energy consum-ption during the summer. Due to the constant installments over the year, there is a negative cash flow effect in the cold seasons, because customers require more energy than they pay for and a positive cash flow effect in the warmer months, because customers need less energy than they pay for. (2) *Company perspective:* Payments for energy procurement vary depending on how much has already been purchased as a forward transaction and how much needs to be spent additionally on the spot market to account for consumption peaks.

Hence, there is a constant, customer-specific, cash flow until the final, individual, annual settlement has been carried out. Procurement, on the other hand, is consumption-based, which is why the resulting cash flow follows a “zig-zag” curve (i.e., it shows strong seasonality) as can be seen in the schematic view in Figure 34[54].

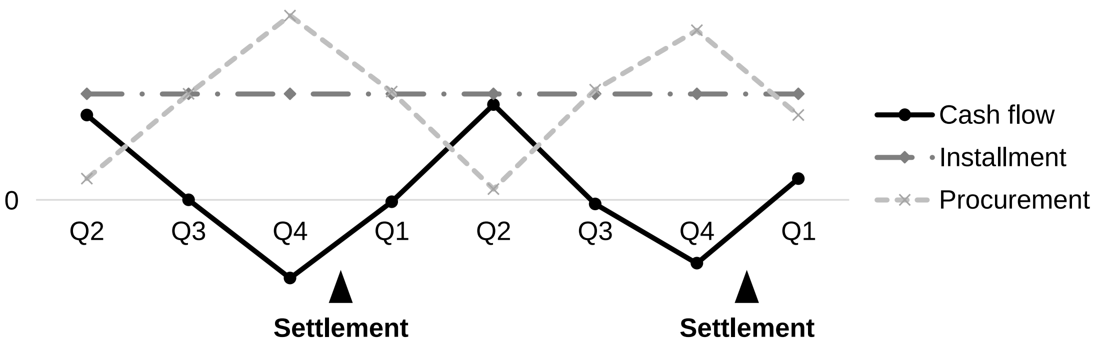


Figure 34. Private customer installment, procurement, and cash flows

In addition to the seasonal pattern, the cash flow also shows a slight decline over the years, which can be attributed, among other things, to higher standards in energy efficiency (e.g., with the help of light emitting diodes, LEDs) and more conscious energy consumption. After removing a number of singular special effects (e.g., acquisitions and restructurings), the operating cash flow (Figure 35) was usable for the modeling phase.

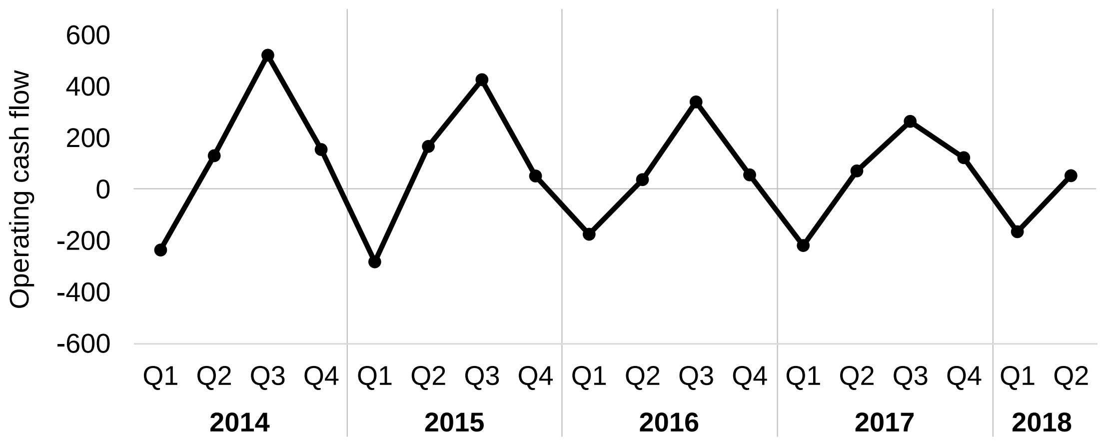


Figure 35. Operating cash flow over the fiscal years

[54] Note that for confidentiality reasons no exact numbers are provided in figures or tables throughout this chapter.

The current *experience-based planning and forecasting process*, which served as a benchmark and starting point for this research project, requires the input of different departments and employees. For example, the energy economics department provides monthly information about future sales. Every quarter, all information is gathered in an MS Excel file, which is the basis for an operating cash flow forecast report for the company's management. Since forecasts deviated strongly from actual year-end cash flow in 2017, the company set up a project team of two internal people, a student, and the author at hand to improve the current forecasting process by integrating predictive analytics.

Data understanding and preparation

Complemented with findings from the literature review (section 3.1), expert interviews were conducted to obtain factors that could potentially influence the operating cash flow. In doing so, three people from the reference company were consulted: the Finance Director, an employee from claim management, and an expert from the energy economics department. Analyzing their statements, the findings can be clustered into *internal data* (financials and non-financials that can be influenced) and *external data* (which cannot be influenced). External data can belong to one of three categories: (1) Spot prices traded at the European Energy Exchange, (2) macroeconomic indicators (e.g., as provided by OECD), such as the composite leading indicator (CLI), which provides early signals of turning points in business cycles (OECD, 2018), and (3) environmental input factors. In total, *21 input factors* were considered relevant. Table 11 summarizes the results.

Next, correlations in the data set were examined. Taking into account that many algorithms cannot handle multicollinearity (Hyndman and Athanasopoulos, 2018), a *correlation matrix* helps to prevent incorporating highly correlated input factors into the prediction model (Mukaka, 2012). Table 12 shows an extract of the correlation matrix. Each cluster from Table 11 is represented with one variable.

Internal data	Financials	Operating cash flow (CF) Accounts receivables balance (FORD) Sales electricity (SALES.ELE)
	Non-financials	Total number of customers (CUST) Average consumption (CNS) Weighted average days to bill for electricity segment (WADTB.ELE) Weighted average days to meter for electricity segment (WADTM.ELE)
External data	Spot Prices	Spot price electricity (ELE) Spot price gas (GAS)
	Macroeconomic indicators	Business confidence index (BCI) Consumer confidence indicator (CCI) Composite leading indicator (CLI) Producer price index (PPI)
	Environment	Temperature (TMP) Sunshine duration (SUN) Precipitation (PRE)

Table 11. Excerpt of the input factors

	Accounts receivables balance	Average consumption	Spot price electricity	Business confidence index	Temperature	Temperature shifted by two quarters
Operating cash flow	-0.80	**-0.73**	-0.19	0.01	**0.86**	**-0.86**
Accounts receivables balance		0.69	-0.31	-0.49	-0.61	0.65
Average consumption			0.22	-0.23	**-0.88**	0.88
Spot price electricity				0.50	-0.34	0.27
Business confidence index					-0.01	-0.05
Temperature						-0.98

Table 12. Excerpt of the correlation matrix

For example, the matrix reveals a strong negative correlation between operating cash flow and average consumption (CNS: -0.73). Temperature, on the other hand, has a strong positive correlation with operating cash flow (TMP: +0.86). Hence, both indicators have a linear relationship with operating cash flow, however in a different direction. Average consumption and temperature themselves depict strong negative correlation (CNS and TMP: -0.88). Due to the problem of multicollinearity, these variables should not be included in the model at the same time. Shifting temperature (TMP) by two quarters leads to an inversion of the correlation coefficient (TMP.lag2: -0.86). This is due to fact that temperature in its original course is very similar to the operating cash flow and the shift leads to an anticyclical course. Summarizing the findings from business understanding, data understanding and preparation, a first design guideline is proposed preventing multicollinearity in a cash flow prediction model:

Design guideline 1: To avoid the pitfall of incorporating highly correlated data at the same time, combine your findings from a thorough business understanding with a correlation matrix.

Modeling

To find evidence that the new model outperforms the experience-based approach currently in place, commonly used forecasting methods and state of the art algorithms were implemented. As a starting point, four methods were applied *using only historical operating cash flow data*: (1) Holt-Winters (HW), (2) ARIMA, (3) extreme gradient boosting for time series forecasting (XGbar), and two types of neural networks, (4a) ELM and (4b) MLP, both of them with one hidden layer and a logistic activation function (cf. section 4 for details).

As a next step, *internal and external data* were incorporated into the prediction model. The following four forecasting methods were used: (1) Linear regression (LIN), (2) LASSO, (3a) ARMAX with regular exogenous factors and (3b) ARMAX with a Fourier transformation[55] of the input factors, and (4) neural networks, this time with external factors. Given the multitude of prediction methods, *ensemble forecasts*, that is combinations of multiple forecasting methods, were computed to enhance prediction accuracy and reliability (Montgomery et al., 2015). To this end, medians and means were tested for ensemble calculation with the latter prevailing due to more consistent results. Summarizing the findings from the modelling phase, a second guideline can be stated as:

---

[55] The Fourier transform is a technique to transform a signal from the time domain into the frequency domain. In this particular instance, it feeds the seasonality pattern as external variables into the ARIMA model. For more information on the Fourier transform in general, see, e.g., Bracewell (1986). For more information on using the Fourier transform on regression variables, see Hyndman (2010).

Design guideline 2: Applying historical cash flow data typically leads to a good first prediction. Based on this, understand predictive analytics as a kind of art that requires the "right" combination of internal and external input factors. The more the better is not always true.

6.2. Results: Machine-based forecasts outperform humans most of the time

To find evidence that applying predictive analytics to cash flow forecasting improves accuracy, the performance of the new prediction model was analyzed against a benchmark set by the reference company as follows: With the year-end CF marking 100%, the weighted sum of deviations from actual CF of all four forecasts over the year should not be higher than 40%. In order to account for the importance of the last forecast (FC4), it is weighted twice as high as the other three forecasts (FC1-FC3). **Figure 36** illustrates this benchmark.

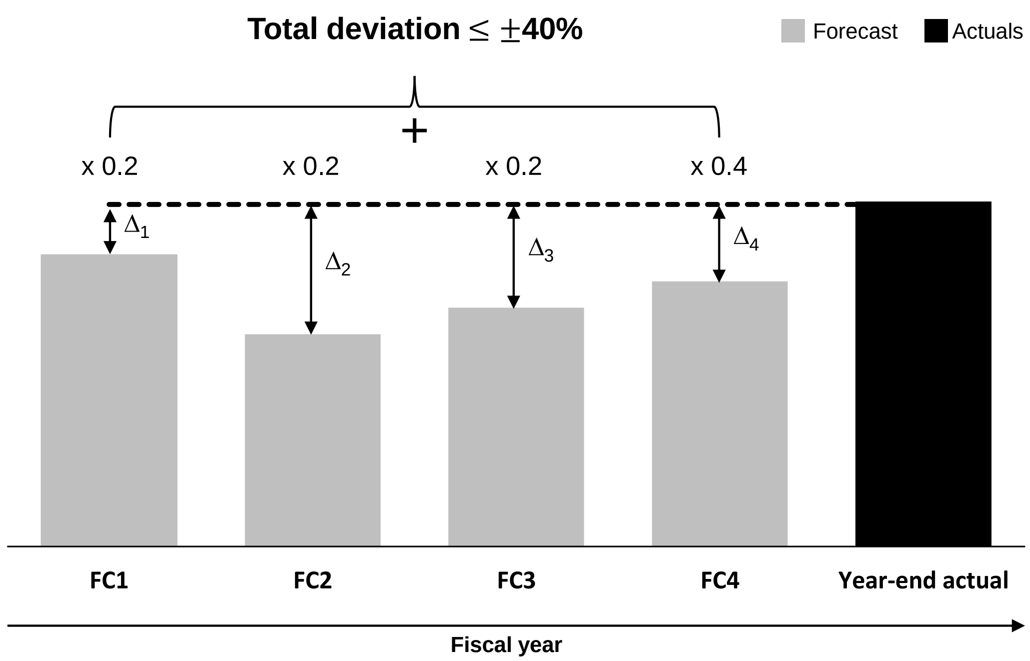


Figure 36. Benchmark for cash flow forecast accuracy

Computing predictions, an *initial training set* of twelve quarters (Q1/2014 – Q4/2016) was used, leaving four quarters for an out-of-sample prediction (test set), Q1-Q4/2017. In a first step, operating cash flow values for 2017 were predicted on a quarterly basis, leading to a four-quarter-ahead forecast (FC1 in **Figure 37**). Summing over the four quarters led to a final year-end operating cash flow prediction. To imitate a company's available information throughout the fiscal year, the initial training set was extended by incorporating Q1/2017 in a second step (FC2). Hence, the test set shortened by one quarter, now only including Q2-Q4/2017. The following prediction therefore reflected a three-quarter ahead forecast. To calculate the final prediction for 2017, the actual value of Q1/2017 and the three-quarter prediction were summed up. The same principle of progressively more available data was used for FC3 and FC4. Figure 37 shows the results for ARIMA and ARMAX (using TMP as external indicator).

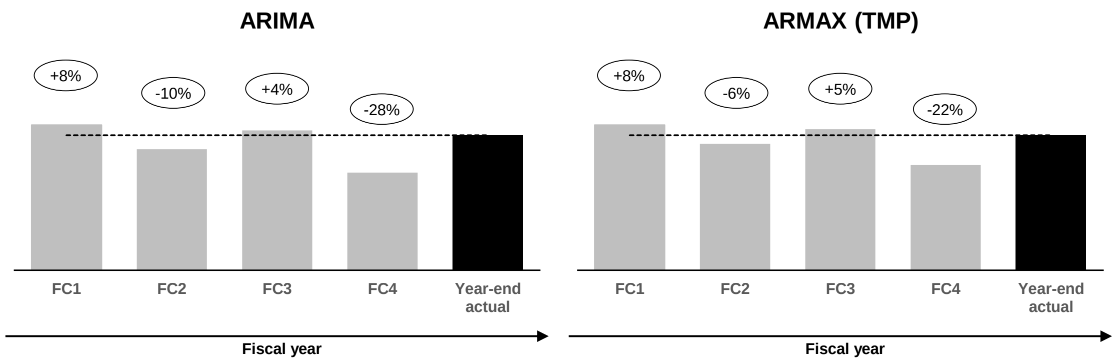


Figure 37. Year-end operating cash flow prediction using ARIMA and ARMAX

In view of this, a third design guideline is:

Design guideline 3: To get even business users involved, evaluate the accuracy of cash flow prediction models against a commonly accepted benchmark from their business department.

With ARIMA, the FC1 operating cash flow prediction only deviated by 8% from year-end actuals. FC2 and FC3 both delivered relatively accurate results (-10% / +4%) as well. Although the final year-end (FC4) forecast underestimated the actual CF by 28%, the forecasting approach still outperformed the benchmark's threshold. Incorporating temperature (TMP) into the prediction model enhanced the overall prediction quality slightly. While FC1, FC2, and FC3 stayed almost constant (+8%/-6%/+5%), FC4 prediction improved by six percentage points, deviating from actuals by 22%. Hence, ARMAX enhanced prediction quality and outperformed the benchmark. Thus, both forecasting algorithms can be considered superior to the model currently in use.

Ensemble forecasts by averaging over a set of algorithms led to more constant deviations. Variance within the fiscal year clearly declined (except Q4, which was underestimated by all algorithms). For example, using the mean of ARIMA, XGbar, and ARMAX led to good results (+7%/+4%/+7%/-26%). Table 13 summarizes the results. The algorithms marked in grey all outperformed the benchmark.

		FC1 vs. act. year-end operating cash flow [%]	FC2 vs. act. year-end operating cash flow [%]	FC3 vs. act. year-end operating cash flow [%]	FC4 vs. act. year-end operating cash flow [%]
Historical cash flow data	HW	-63	-63	-31	-47
	ARIMA	8	-10	4	-28
	XGbar	5	29	13	-29
	MLP	-158	-40	-1	-48
	ELM	39	-5	33	-71
	ELM.50	8	-24	25	-71
External and internal data	LIN	39	24	-54	-82
	LASSO	178	168	102	-53
	ARMAX.fou	65	-60	-212	-202
	ARMAX.ext	8	-6	5	-22
	ELM	12	-22	29	-37
Ensemble	ARIMA+XGbar+ARMAX	7	4	7	-26

Table 13. Overview of prediction accuracy

Summarizing these findings, a fourth design guideline is proposed:

Design guideline 4: ARIMA as a commonly used prediction method is an easy start. Combining several algorithms in ensemble forecasts can mitigate the outlier-proneness of individual algorithms.

In a final step, operating cash flow was predicted by *decomposing the data* and leveraging a method mix. In doing so, given data was broken down into its *seasonal, trend, and remainder components*. After several pairwise comparisons, ARIMA was chosen to predict the seasonal component and LASSO to predict the trend and remainder components. Quite intuitively, over the course of the (simulated) year, the accuracy improved considerably. FC1 and FC2 deviated by 21% and 15% respectively from year-end actuals. With a deviation of 4% and -1% respectively, both FC3 and FC4 predictions were highly accurate. Overall, this approach delivered the best prediction results and can be considered superior to current forecasting approaches of the reference company (Figure 38). Summarizing the findings from the result section, a fifth guideline is proposed:

Design guideline 5: If the data hints at seasonal and trend components decompose the data and predict the individual cash flow components using a method mix with different input factors.

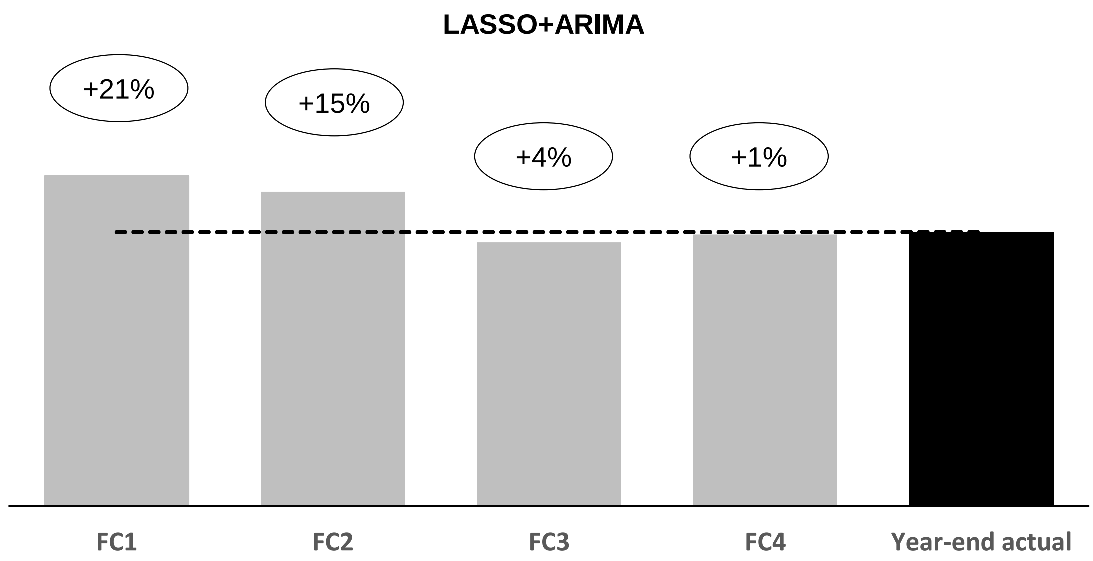


Figure 38. Prediction of decomposed year-end operating cash flow using a method mix

Finally, the reference company considered the process of developing a predictive analytics prototype along the steps of CRISP-DM helpful for understanding the challenges in forecasting their cash flow. It helped to free themselves from preconceptions that business users had. As a result, the forecasting process will be redesigned to require less people and time in the future. This leads to a sixth and final guideline:

Design guideline 6: Adopt a clean-sheet approach to not be overwhelmed by the large amount of potential input factors. A slim model entails a more efficient forecasting process.

Figure 39 shows the four steps of the model fitting process that will be applied by the reference company in the future. After (1) decomposing the data into its trend, seasonal, and random component, they will (2) start with historical operating cash flow data. In a second step, (3) internal and external input factors will be included (Table 11), the model parameters estimated, and (4) ensemble forecasts calculated. After each iteration, the results will be analyzed for accuracy and potential biases, and model modifications made if necessary. This process will be performed several times until saturation, which means model modifications only lead to minor improvements.

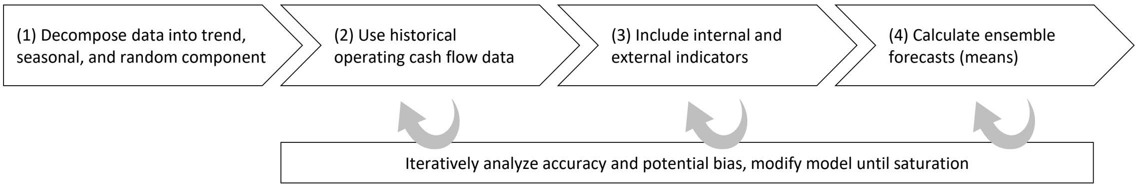


Figure 39. Model fitting process

6.3. Evaluation: Side-by-side use is currently most likely

As described in section 5.3, evaluating the *relevance* of new artifacts is a major activity in DSR (Venable et al., 2016). Regarding RQ1, whether the approach presented here outperforms the experience-based approach of the reference company in terms of prediction accuracy and forecasting process efficiency, the results in the previous section are conducive. The prototype's accuracy is superior and, at the same time, maintaining the new approach requires less people and time. Addressing RQ 2, the design guidelines were evaluated in *manager interviews* with (a) three employees from the accounting department of another utility company, (b) two consulting project managers from an audit firm who had just finished setting up a predictive analytics demo platform, and (c) an expert for analytics from academia. With their input, *three imperatives* summarizing the guidelines can help companies to accelerate the design of cash flow prediction models as follows:

The art of modelling is about the right input factors and their meaningful combination. A deep business understanding and insights from a correlation matrix will help.

The first design guideline received full support by the analytics expert who pointed out that deriving information from a *correlation matrix* provides a good first orientation. However, he continued, the interpretation of the correlation matrix may be difficult for two reasons: (1) Without a deep business understanding an analyst may not be able to select the relevant input factors in the first place and (2) without a basic understanding of mathematics and statistics, a correlation matrix is difficult to understand, sometimes even for experts drawing the right conclusions is complex.

The employees of the utility company emphasized that an overall *business understanding* is crucial to find the right use case and cluster potential input factors. Just relying on statistical relationships without a business understanding can result in misleading conclusions, since relationships might exist within subsidiaries. However, on a group level they may not be relevant. As an example, they outlined that the cash flow characteristics of the use case were valid for sales entities, but opposing effects, such as payments to grid-operators and payments received from customers, balanced each other out on group level. The managers of the audit firm gave the advice to divide the input factors in three clusters such as macroeconomics (e.g., consumer price index), industry-specific factors (e.g., market growth), and internal factors (e.g., order intake). This is roughly in line with the approach proposed in this work.

Furthermore, one manager of the audit firm reported from her last project that she typically starts with a lot *more input variables* (up to 200 macroeconomic, up to 200 industry, and more than 100 internal input factors, as opposed to the smaller number as given in Table 11). In day-to-day business, this does not make a difference, since most of the input factors are pulled automatically from online data sources such as Reuters and, thus, are fed automatically into the model.

Going forward, it can be kept in mind that the number of input factors could be considerably larger, once the dataflow is automated. However, overall, the experts acknowledged the first design guideline as relevant.

The second design guideline describes predictive analytics as a kind of art – refraining from a "the more external and internal input factors, the better" approach. This guideline was endorsed as an alternative way to go by the audit managers. They outlined that someone may either use a large set of input factors as a starting point and shrink them towards a reasonable amount using algorithms and interviews. Alternatively, the approach of starting with fewer, but wisely chosen input factors is appropriate as well. In the end, in either case, only a handful of input factors should be used for the model. This guarantees understandability and interpretability even by common users in their day-to-day business. In line with design guideline 2, the analytics expert pointed out that, as of today, transparency and interpretability are two of the major reasons why statistically-based forecasting is only applied rarely in practice. He added that even academia should focus more on system handling and a nice frontend. Otherwise their prototypes would not be relevant for practitioners.

The business department should set a clear target for the model in advance. Transparency of the algorithms and delivering tangible results are more important than the last inch of accuracy.

Driven by the aim of a better prediction accuracy and even more tangible results, the third design guideline stated that a clear benchmark provided by the business department is key. One interviewee from the audit firm outlined that the most *fundamental benchmark* are a random walk and simple linear regression. If either of these delivers more accurate results than a predictive analytics model, it is futile to invest resources in a forecast using predictive analytics. To use another example, paying a surplus for an actively managed fund is not necessary when exchange-traded funds always yield higher returns. One interviewee from the utility company outlined a general problem with incentives: most business managers get a monetary reward for target (over-) achievement and thus tend to underestimate their operating cash flow. Therefore, to assess the prediction accuracy and, more importantly, to validate the forecast reliability, a well-defined benchmark should be set in advance. A manager from the audit firm added that continuously communicating the performance of experience-based versus machine-based forecasts increases managers' ambition to provide more accurate forecasts and "beat the machine." It is also useful in the first periods to help people identify with the new forecasts. Thus, the third design guideline was approved as well.

The fourth design guideline proposes to go for an entire set of algorithms combined in an *ensemble*. The expert from academia outlined that to the best of his knowledge, ARIMA is the algorithm most often applied in practice. However, it also has a limited pervasion. Furthermore, one manager from the audit firm remarked that the actual day-to-day user in operating business later faces the trade-off between accuracy and transparency of the applied algorithms while delivering tangible results. Thus, the fourth design guideline was adjusted and now states:

Adjusted design guideline 4: Transparency of the algorithms and delivering tangible results are more important than the last inch of accuracy. Always keep this in mind when modeling.

Leveraging data decomposition and a meaningful method mix, a clean-sheet approach often takes less time and people than existing approaches.

The fifth design guideline proposed to *decompose data* and predict trend and seasonal cash flow components using a different method for the individual components. The analytics expert emphasized the correctness of this design guideline. He pointed out, that the first thing he thought about was to decompose the data.

Finally, the sixth guideline received strong support. Just one person is currently needed to update the model in the reference company with new data, which is a strong indicator for a future efficiency gain. However, the process redesign will only be started once the prototype has been used for a couple of periods alongside the traditional forecasts. The managers of the audit firm also reported on such a side-car approach, where business users only gradually changed their working style. In the end, however, they relied on the machine as a basis and turned their attention to explaining exceptions that could not be foreseen by the machine.

In conclusion, all interviewees acknowledged the design guidelines – even from their different perspectives. However, one issue was brought up during the interviews, which is currently not handled by the research at hand: Prediction models are typically developed in a coding environment, like R Studio. However, accountants from the business department in practice generally have no coding skills and are reluctant to work with coding environments. Thus, future research should include a *smart visualization of results* and a more *user-driven IS handling* as well.

Discussion

Following the seven step publication schema (Gregor and Hevner, 2013), this section will reflect to what degree the research questions have been answered and how the results can be put into the context of existing literature. Addressing the *first research question* of this work, “which digital technologies are most relevant towards the medium-term future finance function”, part I proposed a new zero quartile for the technology-driven future finance function and a way of benchmarking the current state of an organization against it. The implications of this first part were discussed in section 2.3.

The *second research questions*, “what is the current state of machine learning and analytics adoption and which are the most relevant drivers”, was covered in chapter 3 with a comprehensive literature review in accounting and information systems publications and a survey among 39 managers of a focus group. The survey results were used to build and test a combined TAM+TTF model for drivers of ML&A adoption and several sub-models in a PLS-SEM approach. It could be confirmed that a combined model of TAM+TTF provides a better explanation of actual use variance than the TAM alone, as has been suggested by Dishaw and Strong (1999). Likewise, it could also be reaffirmed that perceived ease of use only affects perceived usefulness in the TAM part of the model, not so much the intention to use or attitude towards using. However, this thesis showed that, in an ML&A context, perceived usefulness has a significant influence on intention to use as opposed to the attitude towards using link stated by Davis et al. (1989) and Dishaw and Strong (1999). In terms of actual use variance that could be explained, the model proposed in this thesis outperforms other models by at least ten percentage points.

For research, this work may have several benefits: it provides a more comprehensive and methodically founded model of ML&A adoption than the currently existing practitioner literature. Additionally, it shows a real application of PLS-SEM in accounting, not only an appeal to do so, as for instance in Nitzl (2016). *For practice*, this comprehensive list of potential drivers and their positive or negative effects on actual use can help in designing better artifacts in the future.

It can also help to reduce the number of adoption failures, in particular with a set of technologies deemed as important as ML&A.

However, there are also a number of avenues for future research. While this work could provide a detailed model for drivers of adoption on a general basis, the moderating effects of top-down or bottom-up analytics strategies, of an individual's background, or of different departments could not really be evaluated. Especially the latter lacked with respect to the equidistance requirement for PLS-SEM variables. Hence, a comparison of the impact of these potential moderator variables could support a more situated IS design. Additionally, some of the proxies had poor loadings and some of the constructs poor reliability scores which could be a sign for hidden factors that were not covered in the questionnaire. An exploratory analysis, for instance in an interview-based approach such as grounded theory, could shed light on these missing factors.

The *third research question*, “how can companies leverage machine learning and analytics to advance their digital transformation and what are first guidelines”, was answered with two concrete use cases from case studies in a chemical company and an energy utility company.

Based on a single case study in a chemical company, the objective of the *first use case* in chapter 5 was (1) to gauge the potential of machine learning in improving accounting accuracy and process efficiency in comparison to the manual approaches typically in place in practice and (2) to lay out design guidelines to successfully implement and run such a solution. Six functional requirements as well as relevant fields in invoices were identified and distance-weighted *k*-nearest neighbors was applied to determine the most probable general ledger accounts. Three scenarios with different levels of coverage and accuracy were presented. Finally, the results were evaluated along the criteria utility, validity, and evolution and five design guidelines were derived.

For practice, the set of requirements and invoice fields should help companies to advance the implementation of cognitive automation and, thus, especially help managers improve the accuracy and efficiency of their finance back office. In addition, the design guidelines should help companies get started with machine learning for accounts payable. *For research* purposes, the work contributes to theories that specify how artifacts should be designed based on kernel theories. The approach was more comprehensive than mostly literature-based references like Fung (2014). As opposed to articles like Lacity and Willcocks (2016), DSR's iterative "build and evaluate" activities were followed in this case study and a prototype was developed for more in-depth research. Last, but not least, the relevance of the prototype and design guidelines was ensured by evaluating the findings with experts from different domains as opposed to articles without an evaluation like Bräuning et al. (2017) or Caruana and Niculescu-Mizil (2006).

Studying a single case and taking a company from the utility sector as reference, the objective of the *second use case* in chapter 6 was (1) to find evidence whether predictive analytics can outperform experience-based cash flow forecasts and (2) – to start the discussion – to lay out first design guidelines to successfully implement predictive analytics for cash flow forecasting. The artifact design was structured along CRISP-DM. Relevant input factors were identified and commonly used forecasting methods as well as state-of-the-art algorithms were implemented in the software R. Finally, the results were evaluated by discussing the design guidelines in manager interviews with another utility company, an audit firm, and an analytics expert from academia. Three imperatives summarized the guidelines for a more accurate and efficient cash flow prediction model.

For practice, the proposed set of internal and external data should help utility companies to enhance their cash flow forecasting and, thus, help managers to make better decisions. Furthermore, the design guidelines give concrete advice and, thus, should help both digital beginner companies to get started with a cash flow prediction model and experienced companies to improve their forecasting process efficiency across all industries. *For research* purposes, this work also contributes to theories that specify how artifacts should be designed based on kernel theories. As opposed to literature-based references like Lorek (2014) the approach was more comprehensive in that commonly used forecasting techniques such as ARIMA were complemented with state-of-the-art algorithms (e.g., extreme gradient boosting). Moreover, in comparison to authors like Seebach et al. (2011), publicly available data was extended with internal company data by using a case study as research approach. Last, but not least, the relevance of the proposed design guidelines was evaluated with experts from different domains contrary to publications without an evaluation like Brown and Rozeff (1979) or Orpurt and Zang (2009).

However, both use cases also reveal avenues for future research. Although single case study analysis offers a broad range of advantages, Willis (2014) points out one limitation in that external validity and generalizability are always difficult to prove. Thus, future research should approach machine learning and analytics in financial and management accounting with the help of a quantitative approach or a multiple case study. Additionally, the results described for use case 1 cover only the *k*-nearest-neighbors algorithm. However, as Merkert et al. (2015) pointed out, 30% of machine learning applications use artificial neural networks, which would be an interesting choice for a future prototype[56].

Furthermore, the artifacts themselves face limitations. Combining both findings from literature and expert interviews, six functional requirements respectively 21 relevant internal and external input factors were derived. Despite a deep business understanding, relevant requirements or input factors could have been overlooked in the prototypes. Thus, future research could look into extending these sets of requirements and input factors. For the cash flow forecasting prototype, data availability became evident as one of the major limitations in practice. As the utility reference company underwent many structural changes, data availability was limited to the years between 2014 and 2017 (adjusted 2018 data becoming available only throughout the project). Thus, since the research is based on quarterly data, the initial test set only covered twelve observations. Therefore, one could argue that the applied prediction algorithms (e.g., neural networks) require many more observations in order to produce robust results. Future research should therefore also investigate a longer time period or incorporate data on a monthly basis instead and compare the robustness of the algorithms against shorter time samples.

[56] A manuscript for journal publication based on such a comparison of different algorithms for an account, approver, and cost center recommender is currently in preparation by the author of this work and two co-authors. However, due to time constraints, in chapter 5 only the first case study is reported.

Overall, the research results of all the chapters in this work should be interpreted carefully. Generalizability across companies and especially across industries may not be possible due to differences in the way industries work. For example, the financial industry and public sector were not covered in this work. Instead, the focus was laid on companies in the manufacturing industry. Furthermore, digitalization and its transformational effects may lead to unforeseeable developments in the future with new business models replacing long-standing traditions. Nevertheless, digital technologies are a rising topic and their application will become a real game changer for the finance department and beyond.

Conclusion

The aim of this work was to contribute to the digitalization of the finance function along three main research questions. In doing so, it followed a DSR in IS approach and proposed contributions to the knowledge base in several respects.

Chapter 1 was centered on the development of a *future target state for the finance function* and a way of benchmarking an organization against it. After an introduction to finance processes, digital technologies, and benchmarking, a literature review revealed that there is currently no benchmarking that considers the transformational nature of digital technologies. Instead, most benchmarking is retrospective and fails to acknowledge that even the first quartile of organizations is not a good measure for what an organization should be aiming for. Based on this literature review, a questionnaire was administered to a group of finance managers and their responses were analyzed with the help of the Rasch algorithm. Doing so helped to highlight the technology potential for all process activities in each of the four core finance processes and derive four imperatives that help companies with their transformation journey.

Building on the discussions with practitioners that were part of the research for the first chapter, *chapter 2* proposed a more holistic way of looking at digital technology investments. The three dimensions efficiency (monetary and directly measurable), effectiveness (not directly measurable), and experience (intangible) were merged into the *benefit circle* that should help companies evaluate projects more thoroughly where traditional business cases fail.

Chapter 3, as the first chapter in part two of this work, took a closer look at *machine learning and analytics adoption*, two of the currently most promising digital technologies for finance departments. Starting with a literature review of current applications and motivations to use ML&A in financial and management accounting, this chapter showed that there is currently no model for drivers of adoption. Therefore, a questionnaire was developed and answered by 39 finance managers. Based on the results and using structural equation modeling, several models were tested showing that task and technology characteristics play the most important role. It also revealed, among other things, that the higher in the hierarchy of an organization a manager is, the lower is his perceived ease of use.

In order to provide better guidance for how to implement ML&A in the finance department, the following chapters focused on use cases. *Chapter 4* laid the *foundations* with detailed descriptions of algorithms like the LASSO and *k*-nearest neighbors. It also provided explanations of auxiliary methods like imputation, cross-validation, and bootstrapping.

Chapter 5 introduced a *use case in financial accounting* with the implementation of machine learning for better process efficiency and accuracy in invoice processing. Taking a chemical company as reference, a prototype for automatic account recommendation was developed to handle over 500.000 invoices without purchase order per year. Three scenarios of accuracy versus coverage were evaluated and five guidelines were derived that help companies implement cognitive automation. The guidelines addressed the interplay of man and machine for such a prototype, the importance of data preparation and for setting up the machine with the best possible starting position, and recommended a cloud architecture and failsafe for manual override.

Chapter 6 presented a second *use case*, this time in *management accounting* for improved cash flow forecasting. Taking a utility company as reference and following CRISP-DM, a prototype for an automated year-end forecast was developed. Based on expert interviews and literature research, a selection of internal and external indicators made it possible to outperform the manual forecasts currently in place. Six guidelines, subsumed under three umbrella imperatives were derived from the case study and should help companies similarly to the guidelines from the first use case. Likewise, the guidelines for forecasting also covered the interplay of man and machine and suggested a combination of expert opinion and correlation analysis for selecting relevant indicators. Further guidelines also highlighted the importance of involving the employees in the business department when selecting a benchmark for the prototype, and of decomposing the data into trend, seasonal, and random component.

In conclusion, it can be said that interesting times for finance departments lay ahead. With wave after wave of digitalization hitting all organizations, efficiency, effectiveness, and experience will play a key role. Besides, there are a couple of aspects organizations, and in particular finance departments, should consider. *Investments*: Making the right selection for what to spend the limited budget on is crucial since doing all at once is not an option. The benefit circle may help in doing that and finance departments can be among the first to dispose of the fixation on traditional business cases. *Implementation*: There is a renewed trend to insourcing. Building the knowledge for digital technologies inside the company is a strong differentiator and reduces the dependence on external parties. The number-crunching finance department should not hesitate to train machine learning and analytics experts. *Interfaces*: Cumbersome front-ends should be a thing of the past. A combination of rule-based automation, cognitive automation, and chatbots is likely to make data analysis much more intuitive over the next years. Finance professionals can greatly benefit from this development. *Incentives*: For a company to succeed, its managers need to act in concert. A motivating bonus system is one of the means to this end. However, managers are often not in a position to shape an entire market or economy. Hence, seasonality and trend components are dictated by external forces. Ultimately, only the remainder component can be affected by the manager and incentives should be set accordingly. *Intelligence*: Human workers will not be replaced so easily by machines, at least in some areas. Currently, often only prototypes are tested in practice. Moving forward, a side-car approach is a good starting point to prepare the next big step, to build confidence, and to engage employees. The use cases presented show how this can be done. Yet, in the longer run, a full-scale implementation will be necessary to remain competitive. Finally, until the predictions of Stephen Hawking, Elon Musk, and other experts for artificial intelligence (FLI, 2015) come true and machines transcend human intelligence, companies should not play too safe but rather shape the change themselves.

References

Abbott, D. (2014). Applied predictive analytics: Principles and techniques for the professional data analyst. Indianapolis, IN, USA: John Wiley & Sons.

Acito, F. and V. Khatri (2014). "Business analytics - Why now and what next?" Business Horizons 57(2), 565–570.

Agarwal, R. and V. Dhar (2014). "Big data, data science, and analytics: The opportunity and challenge for IS research." Information Systems Research 25(3), 443-448.

Agrawal, R., T. Imieliński and A. Swami (1993). "Mining association rules between sets of items in large databases." ACM Sigmod Record 22(2), 207-216.

Ajzen, I. (1991). "The theory of planned behavior." Organizational Behavior and Human Decision Processes 50(2), 179–211.

Al-Debi, M. M., R. El-Haddadeh and D. Avison (2008). "Defining the business model in the new world of digital business." In: Americas Conference on Information Systems Proceedings

Amani, F. A. and A. M. Fadlalla (2017). "Data mining applications in accounting - A review of the literature and organizing framework." International Journal of Accounting Information Systems 24, 32–58.

Amit, R. and C. Zott (2012). "Creating value through business model innovation." MIT Sloan Management Review 53(3), 41-49.

Anderson, K. and R. McAdam (2005). "An empirical analysis of lead benchmarking and performance measurement: Guidance for qualitative research." International Journal of Quality & Reliability Management 22(4), 354-375.

Appelbaum, D., A. Kogan, M. Vasarhelyi and Z. Yan (2017). "Impact of business analytics and enterprise systems on managerial accounting." International Journal of Accounting Information Systems 25, 29-44.

Asadi Someh, I. and G. Shanks (2015). "How Business Analytics Systems Provide Benefits and Contribute to Firm Performance?" In: European Conference on Information Systems Proceedings

Axson, D. A. (2015). Finance 2020: Death by digital. The best thing that ever happened to your finance organization. https://www.accenture.com/t20150902T015110__w__/us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Dualpub_21/Accenture-Finance-2020-PoV.pdf (visited on 12 Feb 2019).

Bacon-Gerasymenko, V., R. Coff and R. Durand (2016). "Taking a second look in a warped crystal ball: explaining the accuracy of revised forecasts." Journal of Management studies 53(8), 1292-1319.

Baesens, B., R. Bapna, J. R. Marsden, J. Vanthienen and J. L. Zhao (2016). "Transformational Issues of Big Data And Analytics in Networked Business." MIS quarterly 40(4).

Baliga, W. (1995). "Accounting practices benchmarking study spots mistakes companies make." Journal of Accountancy 179(3), 24.

Bangemann, T. O. (2005). Shared services in finance and accounting. Aldershot, UK: Gower Publishing, Ltd.

Belhiah, M., B. Bounabat and S. Achchab (2015). "The impact of data accuracy on user-perceived business service's quality." In: Iberian Conference on Information Systems and Technologies. IEEE. Aveiro, Portugal. 1-4.

Benbasat, I. and H. Barki (2007). "Quo vadis TAM?" Journal of the association for information systems 8(4), 7.

Benbasat, I., D. K. Goldstein and M. Mead (1987). "The case research strategy in studies of information systems." MIS quarterly 11(3), 369-386.

Beretta, S., A. Dossi and H. Grove (1998). "Methodological strategies for benchmarking accounting processes." Benchmarking for Quality Management & Technology 5(3), 165-183.

Berghout, E., P. Schuurman and D. v. Wingerden (2009). "Benchmarking IT benefits: Exploring outcome-and process-based approaches." Americas Conference on Information Systems, 778-785.

Bhimani, A. (2003). Management accounting in the digital economy. Oxford, UK: Oxford University Press.

Bhutta, K. S. and F. Huq (1999). "Benchmarking–best practices: an integrated approach." Benchmarking: An International Journal 6(3), 254-268.

Binder, M., B. Clegg and W. Egel-Hess (2006). "Achieving internal process benchmarking: guidance from BASF." Benchmarking: An International Journal 13(6), 662-687.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. New York, NY, USA: Springer.

Blanc, S. M. and T. Setzer (2015). "Analytical debiasing of corporate cash flow forecasts." European Journal of Operational Research 243(3), 1004-1015.

Blocher, E. J., D. E. Stout and G. Cokins (2010). Cost management - A strategic emphasis. 5th edition. New York, NY, USA: McGraw-Hill/Irwin.

Blumenberg, S. (2004). "Benchmarking Financial Chain Efficiency-the Role of Economies of Scale for Financial Processes." In: Pacific Asia Conference on Information Systems Proceedings

Bock, A. J., T. Opsahl, G. George and D. M. Gann (2012). "The effects of culture and structure on strategic flexibility during business model innovation." Journal of Management studies 49(2), 279-305.

Bodnar, G. H. and W. S. Hopwood (2013). Accounting information systems. 11th edition. Upper Saddle River, CA, USA: Pearson.

Bollen, K. A. (1987). "Total, direct, and indirect effects in structural equation models." Sociological methodology 17, 37-69.

Bollen, K. A. and S. Bauldry (2011). "Three Cs in measurement models: Causal indicators, composite indicators, and covariates." Psychological methods 16(3), 265-284.

Bond, T. G. and C. M. Fox (2015). Applying the Rasch model: Fundamental measurement in the human sciences. New York, NY, USA: Routledge.

Bracewell, R. N. (1986). The Fourier transform and its applications. New York, NY, USA: McGraw-Hill.

Brands, K. and M. Holtzblatt (2015). "Business Analytics: Transforming the Role of Management Accountants." Management Accounting Quarterly 16(3).

Bräuning, M., E. Hüllermeier, T. Keller and M. Glaum (2017). "Lexicographic preferences for predictive modeling of human decision making - A new machine learning method with an application in accounting." European Journal of Operational Research 258(1), 295–306.

Bräuning, M., E. Hüllermeier, T. Keller and M. Glaum (2017). "Lexicographic preferences for predictive modeling of human decision making: A new machine learning method with an application in accounting." European Journal of Operational Research 258(1), 295-306.

Brown, L. D. and M. S. Rozeff (1979). "Univariate time-series models of quarterly accounting earnings per share: A proposed model." Journal of Accounting Research 17(1), 179-189.

Brown, R. G. (1959). Statistical forecasting for inventory control. New York, NY, USA: McGraw-Hill.

Buckham, C. (2006). "Using Predictive Modelling to Improve Collections." Credit Control 27(1), 54–59.

Califf, M. E. and R. J. Mooney (1999). "Relational learning of pattern-match rules for information extraction." In: National Conference on Artificial Intelligence.

Callaghan, W., B. Wilson, C. M. Ringle and J. Henseler (2007). Exploring causal path directionality for a marketing model using Cohen's path method. In: H. Martens and T. Naes (Eds.), Proceedings of PLS'07 – The 5th International Symposium on PLS and Related Methods, p. 132-135.

Camp, R. C. (1989). Benchmarking: the search for industry best practices that lead to superior performance. Milwaukee, WI, USA: ASQC/Quality Resources.

Cao, G. and Y. Duan (2012). "The Affordances of Business Analytics for Strategic Decision-Making and Their Impact on Organisational Performance." In: Pacific Asia Conference on Information Systems Proceedings

Carlson, A. E. (1957). "Automation in Accounting Systems." The Accounting Review 32(2), 224.

Caruana, R. and A. Niculescu-Mizil (2006). "An empirical comparison of supervised learning algorithms." In: Proceedings of the 23rd international conference on Machine learning. ACM. 161-168.

Cassel, C., P. Hackl and A. H. Westlund (1999). "Robustness of partial least-squares method for estimating latent variable quality structures." Journal of applied statistics 26(4), 435-446.

Chaffey, D. and F. Ellis-Chadwick (2019). Digital marketing. Strategy, implementation and practice. 7th edition. London, UK: Pearson.

Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth (2000). CRISP-DM 1.0: Step-by-step data mining guide.

Chatfield, C. and M. Yar (1988). "Holt-Winters forecasting: some practical issues." The Statistician 37(2), 129-140.

Cheek, L. M. (1977). Zero-base budgeting comes of age. New York, NY, USA: Amacom American Management Association.

Chen, H., R. H. Chiang and V. C. Storey (2012). "Business intelligence and analytics: From big data to big impact." MIS quarterly 36(4).

Chen, T. and C. Guestrin (2016). "XGboost: A scalable tree boosting system." In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Proceedings

Chenhall, R. H. and K. Langfield-Smith (1998). "The relationship between strategic priorities, management techniques and management accounting: an empirical investigation using a systems approach." Accounting, Organizations and Society 23(3), 243-264.

Chuttur, M. (2009). "Overview of the Technology Acceptance Model: Origins, Developments and Future Directions." Sprouts: Working Papers on Information Systems 9(37).

Cohen, J. (1992). "A power primer." Psychological bulletin 112(1), 155-159.

Cokins, G. (2013). "Top 7 trends in management accounting." Strategic Finance 95(6), 21–30.

Cokins, G. (2017). "Enterprise Performance Management (EPM) and the Digital Revolution." Performance Improvement 56(4), 14-19.

Colas, M., J. Buvat, K. V. J. Subrahmanyam and S. Nigam (2014). Measure for Measure: The Difficult Art of Quantifying Return on Digital Investments. Capgemini Consulting.

Cooper, D. J. and M. Ezzamel (2013). "Globalization discourses and performance measurement systems in a multinational firm." Accounting, Organizations and Society 38(4), 288-313.

Cooper, H. M. (1998). Synthesizing research: A guide for literature reviews. Thousand Oaks, CA, USA: Sage.

Cooper, W. W., L. M. Seiford and J. Zhu (2004). Data envelopment analysis. In: W. W. Cooper, L. M. Seiford and J. Zhu (Eds.), Handbook on data envelopment analysis, p. 1-39. Boston, MA, USA: Springer.

Cortes, C., L. D. Jackel and W.-P. Chiang (1995). "Limits on learning machine accuracy imposed by data quality." In: International Conference on Knowledge Discovery and Data-Mining Proceedings

Couto, V., E. Wiley, J. Plansky and D. Caglar (2017). Fit for Growth: A Guide to Strategic Cost Cutting, Restructuring, and Renewal. Hoboken, NJ, USA: John Wiley & Sons.

Cover, T. M. and P. E. Hart (1967). "Nearest neighbor pattern classification." IEEE transactions on information theory 13(1), 21-27.

Cragg, P. B. (2002). "Benchmarking information technology practices in small firms." European Journal of Information Systems 11(4), 267-282.

Crookston, N. L. and A. O. Finley (2008). "yaImpute: an R package for kNN imputation." Journal of Statistical Software. 23 (10). 16 p.

Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function." Mathematics of control, signals and systems 2(4), 303-314.

Davenport, T. and A. Tay (2016). "Ramping Up the Role Of Analytics Leader." CFO 32(9), 14–16.

Davenport, T. H. (2010). How Do You Speed Up Information Delivery? https://hbr.org/2010/05/how-do-you-speed-up-informatio (visited on 20 Feb 2019).

Davenport, T. H. and J. G. Harris (2007). Competing on analytics. The new science of winning. Boston, MA, USA: Harvard Business School Press.

Davis, F. D. (1985). "A technology acceptance model for empirically testing new end-user information systems: Theory and results." Massachusetts Institute of Technology.

Davis, F. D., R. P. Bagozzi and P. R. Warshaw (1989). "User acceptance of computer technology: a comparison of two theoretical models." Management science 35(8), 982-1003.

De Bruin, T., R. Freeze, U. Kaulkarni and M. Rosemann (2005). "Understanding the main phases of developing a maturity assessment model." In: Australian Conference on Information Systems Proceedings

De Vellis, R. F. (2011). Scale development. Thousand Oaks, CA, USA: Sage Publications.

Delen, D. and H. Demirkan (2013). "Data, information and analytics as services." Decision support systems 55(1).

Deshmukh, A. (2006). Digital accounting: The effects of the internet and ERP on accounting. Hershey, PA, USA: IGI Global.

Diamantopoulos, A. (2011). "Incorporating formative measures into covariance-based structural equation models." MIS quarterly 35(2), 335-358.

Dickey, D. A. and W. A. Fuller (1981). "Likelihood ratio statistics for autoregressive time series with a unit root." Econometrica 49(4), 1057-1072.

Dietrich, D., B. Heller and B. Yang (2015). Data science & big data analytics: discovering, analyzing, visualizing and presenting data. Indianapolis, IN, USA: John Wiley & Sons.

Dinan, T. P. (2015). "Predictive Analytics Can Move You from Scorekeeper to Proactive Manager." Pennsylvania CPA Journal 86(3), 9–10.

Dino, R. N., D. E. Riley and P. G. Yatrakis (1982). "The role of forecasting in corporate strategy: The Xerox experience." Journal of Forecasting 1(4), 335-348.

Dinov, I. D. (2018). Data science and predictive analytics: Biomedical and health applications using R. Cham, Switzerland: Springer.

Dishaw, M. T. and D. M. Strong (1999). "Extending the technology acceptance model with task–technology fit constructs." Information & Management 36(1), 9-21.

do Céu F. Gaspar Alves, M. (2010). Management accounting and information technology–some empirical evidence. In: M. J. Epstein, J.-F. Manzoni and A. Davila (Eds.), Performance Measurement and Management Control: Innovative Concepts and Practices, p. 429-455. Bingley, UK: Emerald Group Publishing Ltd.

Dolan, E. D. and J. J. Moré (2002). "Benchmarking optimization software with performance profiles." Mathematical programming 91(2), 201-213.

Doll, W. J., X. Deng and J. A. Scazzero (2003). "A process for post-implementation IT benchmarking." Information & Management 41(2), 199-212.

Dreyfus, S. E. (1990). "Artificial neural networks, back propagation, and the Kelley-Bryson gradient procedure." Journal of Guidance, Control, and Dynamics 13(5), 926-928.

Duan, L. and Y. Xiong (2015). "Big data analytics and business analytics." Journal of Management Analytics 2(1), 1-21.

Dubien, J. L. and W. D. Warde (1979). "A mathematical comparison of the members of an infinite family of agglomerative clustering algorithms." Canadian Journal of Statistics 7(1), 29-38.

Dudani, S. A. (1976). "The distance-weighted k-nearest-neighbor rule." IEEE Transactions on Systems, Man, and Cybernetics(4), 325-327.

Dul, J. and T. Hak (2008). Case Study Methodology in Business Research. Oxford, UK: Butterworth-Heinemann.

Dybvig, A. (2016). "A New Application of TRUE Math Programming Optimization for Finance - The Optimized Income Statement (OIS)." Journal of Corporate Accounting & Finance 27(3), 11–16.

Eckerson, W. W. (2004). "Best practices in business performance management: Business and technical strategies." TDWI report series, 8-23.

Eckerson, W. W. (2007). Predictive analytics: Extending the Value of Your Data Warehousing Investment.

Efron, B. (1992). Bootstrap Methods: Another Look at the Jackknife. In: S. Kotz and N. L. Johnson (Eds.), Breakthroughs in Statistics: Methodology and Distribution, p. 569-593. New York, NY, USA: Springer.

Efron, B. and R. J. Tibshirani (1993). An introduction to the bootstrap. Boca Raton, FL, USA: CRC Press.

Eisenhardt, K. M. (1989). "Building theories from case study research." Academy of Management Review 14(4), 532-550.

Elnathan, D., T. W. Lin and S. M. Young (1996). "Benchmarking and management accounting: A framework for research." Journal of Management Accounting Research 1, 37-54.

Enders, C. K. (2010). Applied missing data analysis. New York, NY, USA: Guilford Press.

Enders, W. (2008). Applied econometric time series. Hoboken, NJ, USA: John Wiley & Sons.

Essaides, N., T. Willman and J. O'Connor (2017). The CFO Agenda: Finance's Top Four Strategic Priorities in 2017. The Hackett Group.

Fang, K., Y. Jiang and M. Song (2016). "Customer profitability forecasting using big data analytics: A case study of the insurance industry." Computers & Industrial Engineering 101, 554-564.

Fast-Berglund, Å., T. Fässberg, F. Hellman, A. Davidsson and J. Stahre (2013). "Relations between complexity, quality and cognitive automation in mixed-model assembly." Journal of manufacturing systems 32(3), 449-455.

Ferme, V., J. Lenhard, S. Harrer, M. Geiger and C. Pautasso (2017). "Workflow management systems benchmarking: unfulfilled expectations and lessons learned." In: 39th International Conference on Software Engineering Companion Proceedings

Fisher, J. S. (1994). "The new finance." Journal of Accountancy 178(2), 73-76.

Flaherty, D. J., R. A. Zimmerman and M. A. Murray (1995). "Benchmarking against the best." Journal of Accountancy 180(1), 85.

FLI (2015). Research Priorities for Robust and Benficial Artificial Intelligence. https://futureoflife.org/ai-open-letter/?cn-reloaded=1 (visited on 27 May 2019).

Fornell, C. G. (1987). "A second generation of multivariate analysis: Classification of methods and implications for marketing." Review of Marketing 51, 407-450.

Forni, A. A. and R. van der Meulen (2016). Gartner Survey Finds That Two-Fifths of IT Professionals Consider Their IT Organisation Ready for Digital Business. https://www.gartner.com/en/newsroom/press-releases/2016-07-13-gartner-survey-finds-that-two-fifths-of-it-professionals-consider-their-it-organization-ready-for-digital-business (visited on 16 Apr 2019).

Fowler Jr, F. J. (2013). Survey research methods. 5th edition. Thousand Oaks, CA, USA: Sage Publications.

Franks, B. (2014). The analytics revolution: How to improve your business by making analytics operational in the big data era. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Freund, Y. and R. E. Schapire (1997). "A decision-theoretic generalization of on-line learning and an application to boosting." Journal of computer and system sciences 55(1), 119-139.

Friedman, J., T. Hastie and R. Tibshirani (2001). The elements of statistical learning. New York, NY, USA: Springer Series in Statistics.

Friedman, J. H. (2001). "Greedy function approximation: a gradient boosting machine." Annals of Statistics 29(5), 1189-1232.

Frolick, M. N. and T. R. Ariyachandra (2006). "Business performance management: One truth." IS Management 23(1), 41-48.

Fung, H. P. (2014). "Criteria, use cases and effects of information technology process automation (ITPA)." Advances in Robotics & Automation 3(3), 1-10.

Fürnkranz, J., D. Gamberger and N. Lavrač (2012). Foundations of rule learning. Berlin, Germany: Springer Science & Business Media.

Gandomi, A. and M. Haider (2015). "Beyond the hype: Big data concepts, methods, and analytics." International Journal of Information Management 35(2), 137-144.

Gartner (2015). The Analytics Continuum. https://insideanalytics.com.au/2017/05/28/confusion-consulting/ (visited on 25 Feb 2019).

Gebauer, J. and M. J. Shaw (2004). "Success factors and impacts of mobile business applications: results from a mobile e-procurement study." International Journal of Electronic Commerce 8(3), 19-41.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge and J. Gentry (2004). "Bioconductor: Open software development for computational biology and bioinformatics." Genome Biology 5(10), R80.

Gilbert, J. R., C. Moler and R. Schreiber (1992). "Sparse matrices in MATLAB: Design and implementation." SIAM Journal on Matrix Analysis and Applications 13(1), 333-356.

Girard, A. (1989). "A fast ‘Monte-Carlo cross-validation’ procedure for large least squares problems with noisy data." Numerische Mathematik 56(1), 1-23.

Gleich, R., J. Motwani and A. Wald (2008). "Process benchmarking: a new tool to improve the performance of overhead areas." Benchmarking: An International Journal 15(3), 242-256.

Goodfellow, I., Y. Bengio and A. Courville (2016). Deep learning. Boston, MA, USA: MIT Press.

Goodhue, D. L. and R. L. Thompson (1995). "Task-technology fit and individual performance." MIS quarterly 19(2), 213-236.

Goodwin, P. (2002). "Integrating management judgment and statistical methods to improve short-term forecasts." Omega 30(2), 127-135.

Goodwin, P. (2010). "The holt-winters approach to exponential smoothing: 50 years old and going strong." Foresight 19, 30-33.

Gou, J., L. Du, Y. Zhang and T. Xiong (2012). "A new distance-weighted k-nearest neighbor classifier." Journal of Information & Computational Science 9(6), 1429-1436.

Gregor, S. and A. R. Hevner (2013). "Positioning and presenting design science research for maximum impact." MIS quarterly 37(2), 337-355.

Griffin, A. and J. R. Hauser (1993). "The voice of the customer." Marketing science 12(1), 1-27.

Guilding, C., K. S. Cravens and M. Tayles (2000). "An international comparison of strategic management accounting practices." Management Accounting Research 11(1), 113-135.

Guo, L., F. Shi and J. Tu (2016). "Textual analysis and machine leaning - Crack unstructured data in finance and accounting." The Journal of Finance and Data Science 2(3), 153–170.

Guthrie, J., F. Ricceri and J. Dumay (2012). "Reflections and projections: a decade of intellectual capital accounting research." British Accounting Review 44(2), 68-92.

Hair, J. F., W. C. Black, B. J. Babin and R. E. Anderson (2010). Multivariate data analysis. Englewood Cliffs, NJ, USA: Prentice Hall.

Hair Jr, J. F., G. T. M. Hult, C. Ringle and M. Sarstedt (2016). A primer on partial least squares structural equation modeling (PLS-SEM). Thousand Oaks, CA, USA: Sage Publications.

Hall, J. A. (2018). Accounting information systems. 10th edition. Mason, OH, USA: Cengage Learning.

Halper, F. (2014). Predictive analytics for business advantage. TDWI Research.

Hans, C. (2009). "Bayesian lasso regression." Biometrika 96(4), 835-845.

Henricks, M. (1993). "How do you measure up?" Small Business Reports 18(6), 29-39.

Henseler, J., G. Hubona and P. A. Ray (2017). Partial Least Squares Path Modeling: Updated Guidelines. In: H. Latan and R. Noonan (Eds.), Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications, p. Cham, Switzerland: Springer.

Henseler, J., C. M. Ringle and R. R. Sinkovics (2009). The use of partial least squares path modeling in international marketing. In: R. R. Sinkovics and P. N. Ghauri (Eds.), New Challenges to International Marketing, p. 277-319. Bingley, UK: Emerald Group Publishing Ltd.

Hevner, A. and S. Chatterjee (2010). Design research in information systems: theory and practice. Boston, MA, USA: Springer Science & Business Media.

Hevner, A. R., S. T. March, J. Park and S. Ram (2004). "Design science in information systems research." Management Information Systems Quarterly 28(1), 75-105.

Hevner, A. R., S. T. March, J. Park and S. Ram (2004). "Design science in information systems research." MIS Quarterly 28(1), 75-105.

Holsapple, C., A. Lee-Post and R. Pakath (2014). "A unified foundation for business analytics." Decision support systems 64, 130-141.

Holt, C. C. (1957). Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. Carnegie Institute of Technology.

Hopf, K., S. J. Riechel, M. Sodenkamp and T. Staake (2017). "Predictive Customer Data Analytics – The Value of Public Statistical Data and the Geographic Model Transferability." In: Thirty Eigth International Conference on Information Systems Proceedings

Horngren, C. T., G. L. Sundem, J. A. Elliott and D. R. Philbrick (2002). Introduction to financial accounting. 8th edition. Upper Saddle River, NJ, USA: Prentice Hall.

Horton, N. J. and K. P. Kleinman (2007). "Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models." The American Statistician 61(1), 79-90.

Horvath, P. and N. Herter (1992). "Benchmarking: Comparison with the best of the best." Controlling 4(1), 4-11.

Huang, G.-B., H. Zhou, X. Ding and R. Zhang (2012). "Extreme learning machine for regression and multiclass classification." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42(2), 513-529.

Huang, G.-B., Q.-Y. Zhu and C.-K. Siew (2006). "Extreme learning machine: theory and applications." Neurocomputing 70(1-3), 489-501.

Hutchinson, C., J. Ward and K. Castilon (2009). "Navigating the next-generation application architecture." IT professional 11(2), 18-22.

Hyndman, R. J. (2010). Forecasting with long seasonal periods. https://robjhyndman.com/hyndsight/longseasonality/ (visited on 11 Apr 2019).

Hyndman, R. J. and G. Athanasopoulos (2018). Forecasting: Principles and practice. https://otexts.org/fpp2/ (visited on 27 Sep 2018).

IBM (2013). Descriptive, predictive, prescriptive: Transforming asset and facilities management with analytics. Choose the right data analytics solutions to boost service quality, reduce operating costs and build ROI. https://static.ibmserviceengage.com /TIW14162USEN.PDF (visited on 17 Apr 2019).

Ihaka, R. and R. Gentleman (1996). "R: A language for data analysis and graphics." Journal of Computational and Graphical Statistics 5(3), 299-314.

Jackson, N. (2001). "Benchmarking in UK HE: an overview." Quality Assurance in Education 9(4), 218-235.

Jacobson, M. Z. and M. A. Delucchi (2011). "Providing all global energy with wind, water, and solar power, Part I: Technologies, energy resources, quantities and areas of infrastructure, and materials." Energy policy 39(3), 1154-1169.

James, G., D. Witten, T. Hastie and R. Tibshirani (2013). An introduction to statistical learning. New York, NY, USA: Springer.

Jazayeri, M. and T. Hopper (1999). "Management accounting within world class manufacturing: a case study." Management Accounting Research 10(3), 263-301.

Jin, C., Y. Kong, Q. Kang, W. Qian and A. Zhou (2016). "Benchmarking in-memory database." Frontiers of Computer Science 10(6), 1067-1081.

Joo, S.-J., D. Nixon and P. A. Stoeberl (2011). "Benchmarking with data envelopment analysis: a return on asset perspective." Benchmarking: An International Journal 18(4), 529-542.

Juan, Y. and C. Ou-Yang (2005). "A process logic comparison approach to support business process benchmarking." The International Journal of Advanced Manufacturing Technology 26(1-2), 191-210.

Kacprzyk, J. and W. Pedrycz (2015). Springer handbook of computational intelligence. Dordrecht, Netherlands: Springer.

Kelderman, H. (1984). "Loglinear Rasch model tests." Psychometrika 49(2), 223-245.

Ketter, W., M. Peters, J. Collins and A. Gupta (2015). "Competitive benchmarking: an IS research approach to address wicked problems with big data and analytics." MIS quarterly 40, 1057-1080.

Khan, A. M. A., N. Amin and N. Lambrou (2010). "Drivers and barriers to business intelligence adoption: A case of Pakistan." In: European and Mediterranean Conference on Information Systems Proceedings

Khashei, M. and M. Bijari (2010). "An artificial neural network (p, d, q) model for timeseries forecasting." Expert Systems with applications 37(1), 479-489.

King, W. R. and J. He (2006). "A meta-analysis of the technology acceptance model." Information & Management 43(6), 740-755.

Knechel, W. R. and S. E. Salterio (2017). Auditing: Assurance and risk. 4th edition. New York, NY, USA: Routledge.

Knickrehm, M., B. Berthon and P. Daugherty (2016). Digital Disruption: The Growth Multiplier. Optimizing digital investments to realize higher productivity and growth. https://www.accenture.com/_acnmedia/PDF-14/Accenture-Strategy-Digital-Disruption-Growth-Multiplier-Brazil.pdf (visited on 16 Apr 2019).

Kohlbacher, F. (2006). "The use of qualitative content analysis in case study research." Forum: Qualitative Social Research 7(1), 1-30.

Kotonya, G. and I. Sommerville (1998). Requirements engineering: processes and techniques. Hoboken, NJ, USA: Wiley Publishing.

Kotsiantis, S. B., I. Zaharakis and P. Pintelas (2007). Supervised Machine Learning: A Review of Classification Techniques. In: I. Maglogiannis, K. Karpouzis, M. Wallace and J. Soldatos (Eds.), Emerging artificial intelligence applications in computer engineering, p. 3-24. Amsterdam, Netherlands: IOS Press.

Kriens, J., J. T. van Lieshout, J. Roemen and P. Verheyen (1983). "Management accounting and operational research." European Journal of Operational Research 13(4), 339-352.

Krotov, V. and B. Ives (2016). "IT cost benchmarking: Drawing the right conclusions." Business Horizons 59(6), 645-653.

Kuhn, M. and K. Johnson (2013). Applied predictive modeling. New York, NY, USA: Springer.

Kurzweil, R. (2010). The singularity is near. When humans transcend biology. 2nd edition. London, UK: Gerald Duckworth & Co.

Lacity, M., L. P. Willcocks and A. Craig (2015). Robotic process automation at Telefonica O2.

Lacity, M. C., S. Solomon, A. Yan and L. P. Willcocks (2011). "Business process outsourcing studies - A critical review and research directions." Journal of Information Technology 26(4), 221-258.

Lacity, M. C. and L. P. Willcocks (2016). "A New Approach to Automating Services." 58, 40-49.

Lahrmann, G., F. Marx, T. Mettler, R. Winter and F. Wortmann (2011). "Inductive design of maturity models: applying the Rasch algorithm for design science research." In: International Conference on Design Science Research in Information Systems Proceedings

Larsen, T. J., A. M. Sørebø and Ø. Sørebø (2009). "The role of task-technology fit as users' motivation to continue information system use." Computers in Human Behavior 25(3), 778-784.

Lavalle, S., E. Lesser, R. Shockley, M. Hopkins and N. Kruschwitz (2011). "Special report: Analytics and the new path to value." MIT Sloan Management Review 52(2), 22-32.

LeCun, Y., Y. Bengio and G. Hinton (2015). "Deep learning." Nature 521, 436.

Lee, L., S. Petter, D. Fayard and S. Robinson (2011). "On the use of partial least squares path modeling in accounting research." International Journal of Accounting Information Systems 12, 305-328.

Levandoski, J. J., P.-Å. Larson and R. Stoica (2013). "Identifying hot and cold data in main-memory databases." In: 29th IEEE International Conference on Data Engineering Proceedings

Leybourne, S. J. (1995). "Testing for unit roots using forward and reverse Dickey-Fuller regressions." Oxford Bulletin of Economics and Statistics 57(4), 559-571.

Liebowitz, J. (2014). Business analytics: An introduction. Boca Raton, FL, USA: CRC Press.

Linacre, J. M. (1999). "Understanding Rasch measurement: estimation methods for Rasch measures." Journal of outcome measurement 3(4), 382-405.

Linacre, J. M. (2002). "What do infit and outfit, mean-square and standardized mean." Rasch Measurement Transactions 16(2), 878.

Linacre, J. M. (2004). "Rasch model estimation: Further topics." Journal of applied measurement 5(1), 95-110.

Lissitz, R. W. and S. B. Green (1975). "Effect of the number of scale points on reliability: A Monte Carlo approach." Journal of Applied Psychology 60(1), 10-13.

Little, R. J. (1992). "Regression with missing X's: a review." Journal of the American statistical association 87(420), 1227-1237.

Little, R. J. and D. B. Rubin (2019). Statistical analysis with missing data. 3rd edition. Hoboken, NJ, USA: John Wiley & Sons.

Lord, F. and M. R. Novick (2008). Statistical theories of mental test scores. Charlotte, NC, USA: Information Age Publishing.

Lorek, K. S. (2014). "Trends in statistically based quarterly cash-flow prediction models." Accounting Forum 38(2), 145-151.

Lu, C.-J. and L.-J. Kao (2016). "A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server." Engineering Applications of Artificial Intelligence 55, 231–238.

Luber, M. (2013). "Employ Statistics and Analytics to Up Your Collections Game." Controller's Report 8, 8–9.

Lucas, S. (2016). The Benefits of the SAP Digital Enterprise Platform. https://blogs.saphana.com/2016/02/03/the-benefits-of-the-sap-digital-enterprise-platform/ (visited on 12 Feb 2019).

Lucas, S. (2016). Introducing the SAP Digital Enterprise Platform. https://blogs.saphana.com/2016/01/11/the-sap-digital-enterprise-platform-part-1-delivering-connected-infrastructure-insight-to-drive-the-digital-enterprise/ (visited on 25 Feb 2019).

March, S. T. and G. F. Smith (1995). "Design and natural science research on information technology." Decision support systems 15(4), 251-266.

Marchant, G. (2013). "Management Accounting in the 21st Century - A Profession for which the Time Has Come." Journal of Applied Management Accounting Research 11(2), 1–4.

Martin, J. and T. Conte (2012). "Data Quality in Financial Planning-an Empirical Assessment Based on Benford's Law." In: European Conference on Information Systems Proceedings

Marx, F., F. Wortmann and J. H. Mayer (2012). "A maturity model for management control systems." Business & information systems engineering 4(4), 193-207.

Mayer, J. H., S. Bischoff, R. Winter and T. Weitzel (2012). "Extending Traditional EIS Use to Support Mobile Executives Online and Offline." MIS Quarterly Executive 11(2).

Mayring, P. (2014). "Qualitative content analysis: theoretical foundation, basic procedures and software solution."

McCarthy, J., M. L. Minsky, N. Rochester and C. E. Shannon (2006). "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955." AI magazine 27(4), 12.

McCormick, G. P. (1969). "Communications to the editor—exponential forecasting: some new variations." Management science 15(5), 311-320.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity." The bulletin of mathematical biophysics 5(4), 115-133.

McNames, J. (2001). "A fast nearest-neighbor algorithm based on a principal axis search tree." IEEE Transactions on Pattern Analysis and Machine Intelligence 23(9), 964-976.

Merkert, J., M. Mueller and M. Hubl (2015). "A Survey of the Application of Machine Learning in Decision Support Systems." In: ECIS.

Messner, W. (2013). Making the compelling business case: Decision-making techniques for successful business growth. Houndmills, Hampshire, UK: Palgrave MacMillan.

Mettler, T. (2011). "Maturity assessment models: a design science research approach." International Journal of Society Systems Science 3(1/2), 81-98.

Milgram, P. and F. Kishino (1994). "A taxonomy of mixed reality visual displays." IEICE TRANSACTIONS on Information and Systems 77(12), 1321-1329.

Minsky, M. and S. Papert (1969). Perceptron: an introduction to computational geometry. Cambridge, MA, USA: MIT Press.

Mishra, N. and S. Silakari (2012). "Predictive analytics: a survey, trends, applications, opportunities & challenges." International Journal of Computer Science and Information Technologies 3(3), 4434-4438.

Monczka, R. M., R. B. Handfield, L. C. Giunipero and J. L. Patterson (2015). Purchasing and supply chain management. 6th edition. Mason, OH, USA: Cengage Learning.

Monecke, A. and F. Leisch (2012). "semPLS: structural equation modeling using partial least squares." Journal of Statistical Software 48(3), 1-32.

Montgomery, D. C., E. A. Peck and G. G. Vining (2012). Introduction to linear regression analysis. 5th edition. Hoboken, NJ, USA: John Wiley & Sons.

Montgomery, J. M., F. M. Hollenbach and M. D. Ward (2015). "Calibrating ensemble forecasting models with sparse data in the social sciences." International Journal of Forecasting 31(3), 930-942.

Mowen, M. M. and D. R. Hansen (2005). Management Accounting: The Cornerstone for Business Decisions. Mason, OH, USA: Thomson/South-Western.

Mukaka, M. M. (2012). "A guide to appropriate use of correlation coefficient in medical research." Malawi Medical Journal 24(3), 69-71.

Nagelkerke, N. J. (1991). "A note on a general definition of the coefficient of determination." Biometrika 78(3), 691-692.

Natekin, A. and A. Knoll (2013). "Gradient boosting machines, a tutorial." Frontiers in Neurorobotics 7(21).

Nielsen, P. A. and J. S. Persson (2017). "Useful business cases: value creation in IS projects." European Journal of Information Systems 26(1), 66-83.

Nielsen, S. (2015). The Impact of Business Analytics on Management Accounting. https://ssrn.com/abstract=2616363

Nielsen, S. (2017). New and Interesting Perspectives for the Management Accountant in a World of Data. https://pure.au.dk/portal/files/114248939/Working_paper_Man_Accountant_June_2017.pdf

Nilsson, N. J. (2014). Principles of artificial intelligence. Reprint. San Francisco, CA, USA: Morgan Kaufmann.

Nitzl, C. (2016). "The use of partial least squares structural equation modelling (PLS-SEM) in management accounting research: Directions for future theory development." Journal of Accounting Literature 37, 19-35.

Nof, S. Y. (2009). Automation: What it means to us around the world. In: S. Y. Nof (Eds.), Springer Handbook of Automation, p. 13-52. Berlin, Heidelberg, Germany: Springer.

O'Leary, D. E. (2018). Big Data and Knowledge Management with Applications in Accounting and Auditing: The Case of Watson. https://ssrn.com/abstract=3203842

OECD (2018). Composite leading indicator (CLI). https://data.oecd.org/leadind /composite-leading-indicator-cli.htm#indicator-chart (visited on 3 Sep 2018).

Olson, J. E. (2003). Data quality: the accuracy dimension. San Francisco, CA, USA: Morgan Kaufmann Publishers.

Onken, R. and A. Schulte (2010). "System-Ergonomic Design of Cognitive Automation : Dual-Mode Cognitive Design of Vehicle Guidance and Control Work Systems."

Ord, K. and R. Fildes (2017). Principles of business forecasting. 2nd edition. New York, NY, USA: Wessex Press.

Orpurt, S. F. and Y. Zang (2009). "Do direct cash flow disclosures help predict future operating cash flows and earnings?" The Accounting Review 84(3), 893-935.

Österle, H. and B. Otto (2010). "Consortium research." Business & information systems engineering 2(5), 283-293.

Paliwal, M. and U. A. Kumar (2009). "Neural networks and statistical techniques - A review of applications." Expert Systems with applications 36(1), 2-17.

Panetta, K. (2017). Top Trends in the Gartner Hype Cycle for Emerging Technologies, 2017. https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/ (visited on 20 Feb 2019).

Payne, R. (2014). "Discussion of ‘Digitisation, ‘Big Data’ and the transformation of accounting information’ by Alnoor Bhimani and Leslie Willcocks (2014)." Accounting and Business Research 44(4), 491–495.

Peffers, K., T. Tuunanen, M. A. Rothenberger and S. Chatterjee (2007). "A design science research methodology for information systems research." Journal of management information systems 24(3), 45-77.

Perols, J. (2011). "Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms." AUDITING: A Journal of Practice & Theory 30(2), 19-50.

Plaschke, F., I. Seth and R. Whiteman (2018). Bots, algorithms, and the future of the finance function. https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/bots-algorithms-and-the-future-of-the-finance-function (visited on 12 Feb 2019).

Post, B. Q. (1992). "A business case framework for group support technology." Journal of management information systems 9(3), 7-26.

Prat, N., I. Comyn-Wattiau and J. Akoka (2014). "Artifact Evaluation in Information Systems Design-Science Research-a Holistic View." In: Pacific Asia Conference on Information Systems Proceedings

Pryor, L. S. (1989). "Benchmarking: A self-improvement strategy." Journal of Business Strategy 10(6), 28-32.

Quaadgras, A., P. Weill and J. W. Ross (2014). "Management commitments that maximize business impact from IT." Journal of Information Technology 29(2), 114-127.

Quattrone, P. (2016). "Management accounting goes digital - Will the move make it wiser?" Management Accounting Research 31, 118–122.

R Core Team (2013). A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org/

Ramaprasad, A. and T. Syn (2015). "Ontological meta-analysis and synthesis." Communications of the Association for Information Systems 37, 138-153.

Ramrathan, D. and M. Sibanda (2014). "Impact of Analytics in Financial Decision Making - Evidence from a Case Study Approach." Acta Universitatis Danubius: Oeconomica 10(5), 124–135.

Rasch, G. (1960). "Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests."

Ray, S. and N. Danks (2019). SEMinR. https://cran.r-project.org/web/packages/seminr/vignettes/SEMinR.html (visited on 18 May 2019).

Rich, E. and K. Knight (1991). Artificial Intelligence. 2nd edition. New York, NY, USA: McGraw-Hill.

Rigdon, E. E. (2012). "Rethinking partial least squares path modeling: In praise of simple methods." Long Range Planning 45(5-6), 341-358.

Romney, M. B. and P. J. Steinbart (2018). Accounting information systems. 14th edition. Harlow, UK: Pearson, .

Roos, G. (2015). "Technology-driven productivity improvements in the professional services industry." Academic Leadership Series 6, 41-50.

Rouse, M. (2015). 360-degree customer view. http://searchsalesforce.techtarget.com/definition/360-degree-customer-view (visited on 16 Apr 2019).

Rowley, J. and F. Slack (2004). "Conducting a literature review." Management research news 27(6), 31-39.

Russell, S. J. and P. Norvig (2016). Artificial intelligence: a modern approach. Malaysia: Pearson Education Limited.

Saeed, I., G. Juell-Skielse and E. Uppström (2012). Cloud enterprise resource planning adoption: Motives & barriers. In: C. Møller and S. Chaudhry (Eds.), Advances in Enterprise Information Systems II, p.

Saha, B. and D. Srivastava (2014). "Data quality: The other face of Big Data." In: IEEE International Conference on Data Engineering Proceedings

Sako, M. (2006). "Outsourcing and offshoring: Implications for productivity of business services." Oxford Review of Economic Policy 22(4), 499-512.

Sanchez, G. (2013). PLS path modeling with R. Berkeley, CA, USA: Trowchez Editions.

Sangster, A., S. A. Leech and S. Grabski (2009). "ERP implementations and their impact upon management accountants." Journal of Information Systems and Technology Management 6(2), 125-142.

Sarstedt, M., J. Henseler and C. M. Ringle (2011). Multigroup analysis in partial least squares (PLS) path modeling: Alternative methods and empirical results. In: M.

Sarstedt, M. Schwaiger and C. R. Taylor (Eds.), Measurement and research methods in international marketing, p. 195-218. Bingley, UK: Emerald Group Publishing Ltd.

Schafer, J. L. (1999). "Multiple imputation: a primer." Statistical Methods in Medical Research 8(1), 3-15.

Schelp, J. and R. Winter (2007). "Towards a methodology for service construction." In: Hawaii International Conference on System Sciences Proceedings

Schläfke, M., R. Silvi and K. Möller (2012). "A framework for business analytics in performance management." International Journal of Productivity and Performance Management 62(1), 110–122.

Schmid, M. and T. Hothorn (2008). "Flexible boosting of accelerated failure time models." BMC bioinformatics 9(1), 269.

Schmidt, M. J. (2009). Business case essentials: A guide to structure and content. Boston, MA, USA: Solution Matrix Limited.

Schneider, G. P., J. Dai, D. J. Janvrin, K. Ajayi and R. L. Raschke (2015). "Infer, Predict, and Assure - Accounting Opportunities in Data Analytics." Accounting Horizons 29(3), 719–742.

Schniederjans, M. J. and T. Garvin (1997). "Using the analytic hierarchy process and multi-objective programming for the selection of cost drivers in activity-based costing." European Journal of Operational Research 100(1), 72–80.

Scholz, P., C. Schieder, C. Kurze, P. Gluchowski and M. Böhringer (2010). "Benefits and Challenges of Business Intelligence Adoption in Small and Medium-Sized Enterprises." In: European Conference on Information Systems Proceedings

Schryen, G. (2013). "Revisiting IS business value research: what we already know, what we still need to know, and how we can get there." European Journal of Information Systems 22(2), 139-169.

Secrett, M. (2012). Brilliant Budgets and Forecasts: Your Practical Guide to Preparing and Presenting Financial Information. Harlow, UK: Pearson.

Seddon, P. B., D. Constantinidis, T. Tamm and H. Dod (2017). "How does business analytics contribute to business value?" Information Systems Journal 27(3), 237-269.

Seebach, C., I. Pahlke and R. Beck (2011). "Tracking the digital footprints of customers: How firms can improve their sensing abilities to achieve business agility." In: ECIS 2011 Proceedings.

Seong Leem, C., B. Wan Kim, E. Jung Yu and M. Ho Paek (2008). "Information technology maturity stages and enterprise benchmarking: an empirical study." Industrial Management & Data Systems 108(9), 1200-1218.

Shang, S. and P. B. Seddon (2002). "Assessing and managing the benefits of enterprise systems: the business man-ager's perspective." Information Systems Journal 12(4), 271-299.

Shanks, G., N. Bekmamedova and L. Willcocks (2012). "Business Analytics - Enabling Strategic Alignment and organisational Transformation." In: European Conference on Information Systems Proceedings

Sharma, A. and P. K. Panigrahi (2014). "A Review of Financial Accounting Fraud Detection based on Data Mining Techniques." International Journal of Computer Applications 39(1), 37-47.

Sharma, R., S. Mithas and A. Kankanhalli (2014). "Transforming decision-making processes - A research agenda for understanding the impact of business analytics on organisations." European Journal of Information Systems 23(4), 433–441.

Shetty, Y. (1993). "Aiming high: competitive benchmarking for superior performance." Long Range Planning 26(1), 39-44.

Simon, H. A. (1983). Why should Machines Learn? In: R. S. Michalski, J. G. Carbonell and T. M. Mitchell (Eds.), Machine Learning, p. 25-37. San Francisco, CA, USA: Morgan Kaufmann.

Simon, H. A. (1997). Administrative Behavior. A Study of Decision Making Processes in Administrative Organizations. 4th edition. New York, NY, USA: Free Press.

Smith, D. and K. Langfield-Smith (2004). "Structural equation modeling in management accounting research: Critical analysis and opportunities." Journal of Accounting Literature 23(1), 49-86.

Smith, P. and R. Payne (2011). The finance function: A framework for analysis. ICAEW.

Sommerville, I. (2007). Software engineering. Boston, MA, USA: Addison-Wesley.

Sorensen, D. (2013). "EPM in manufacturing: finally coming of age." Strategic Finance 95(3), 39-46.

Souza, G. C. (2014). "Supply chain analytics." Business Horizons 57(5), 595-605.

Spruit, M. and K. Pietzka (2015). "MD3M: The master data management maturity model." Computers in Human Behavior 51, 1068-1076.

Stekhoven, D. J. and P. Bühlmann (2011). "MissForest—non-parametric missing value imputation for mixed-type data." Bioinformatics 28(1), 112-118.

Subramaniyan, M., A. Skoogh, H. Salomonsson, P. Bangalore and J. Bokrantz (2018). "A data-driven algorithm to predict throughput bottlenecks in a production system based on active periods of the machines." Computers & Industrial Engineering in press.

Sun, Z., K. Strang and S. Firmin (2017). "Business Analytics-Based Enterprise Information Systems." Journal of Computer Information Systems 57(2), 169–178.

Szekely, G. J. and M. L. Rizzo (2005). "Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method." Journal of classification 22(2), 151-183.

Takabi, H., J. B. Joshi and G.-J. Ahn (2010). "Security and privacy challenges in cloud computing environments." IEEE Security & Privacy 8(6), 24-31.

Talwar, A. and Y. Kumar (2013). "Machine Learning: An artificial intelligence methodology." In: International Journal Of Engineering And Computer Science.

Tangsucheeva, R. and V. Prabhu (2014). "Stochastic financial analytics for cash flow forecasting." International Journal of production economics 158, 65–76.

Taylor, G. J. (1985). Accounting for Business Organisations: A Practical Approach. London, UK: Palgrave.

Teppler, S. W. (2003). "Digital data and the meaning of'audit'." The CPA Journal 73(2), 70.

Teuteberg, F., M. Kluth, S. Smolnik and F. Ahlemann (2009). "Semantic Benchmarking of Process Models-An Ontology-Based Approach." In: International Conference on Information Systems Proceedings

The Hackett Group (n.d.). Best Practices and Metrics for Next-Generation P2P. https://www.slideshare.net/carrfraser/hackett-tradeshift-webinar-final (visited on 08 Mar 2019).

The Hackett Group (n.d.). Finding the Right Best Practices Benchmarking Provider. https://www.thehackettgroup.com/best-practices-benchmarking/ (visited on 28 Apr 2019).

Theodoridis, S. and K. Koutroumbas (2009). Pattern Recognition. 4th edition. Burlington, MA, USA: Academic Press.

Thompson, R., D. Barclay and C. A. Higgins (1995). "The partial least squares approach to causal modeling: personal computer adoption and use as an illustration." Technology studies: special issue on research methodology 2(2), 284-324.

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society: Series B (Methodological) 58(1), 267-288.

Tucker, F. G., S. M. Zivan and R. C. Camp (1987). "How to measure yourself against the best." Harvard Business Review 65(1), 8-10.

Urbach, N., S. Smolnik and G. Riempp (2009). "The state of research on information systems success." Business & information systems engineering 1(4), 315-325.

Vaishnavi, V. K. and W. Kuechler (2015). Design science research methods and patterns: innovating information and communication technology. Boca Raton, FL, USA: CRC Press.

Valentin, E. K. (2014). Business planning and market strategy. Thousand Oaks, CA, USA: SAGE Publications.

Vasarhelyi, M. A. and A. M. Rozario (2018). "How Robotic Process Automation Is Transforming Accounting and Auditing." CPA Journal 88(6), 46-49.

Venable, J., J. Pries-Heje and R. Baskerville (2016). "FEDS: A framework for evaluation in design science research." European Journal of Information Systems 25(1), 77-89.

Venkatesh, V. (2000). "Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model." Information Systems Research 11(4), 342-365.

Venkatesh, V. and F. D. Davis (2000). "A theoretical extension of the technology acceptance model: Four longitudinal field studies." Management science 46(2), 186-204.

Vom Brocke, J., A. Simons, B. Niehaves, K. Riemer, R. Plattfaut and A. Cleven (2009). "Reconstructing the giant: On the importance of rigour in documenting the literature search process." In: European Conference on Information Systems Proceedings

Vom Brocke, J., A. Simons, K. Riemer, B. Niehaves, R. Plattfaut and A. Cleven (2015). "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research." Communications of the Association for Information Systems 37, 205-224.

Wadan, R., F. Teuteberg, F. Bensberg and G. Buscher (2019). "Understanding the Changing Role of the Management Accountant in the Age of Industry 4.0 in Germany." In: Hawaii International Conference on System Sciences Proceedings

Walker, G. T. (1931). "On periodicity in series of related terms." Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 131(818), 518-532.

Walker, J. (2009). Accounting in a Nutshell: Accounting for the Non-specialist. 3rd edition. Amsterdam, Netherlands: CIMA Publishing.

Ward Jr, J. H. (1963). "Hierarchical grouping to optimize an objective function." Journal of the American statistical association 58(301), 236-244.

Watson, H. J. (2014). "Tutorial: Big data analytics: Concepts, technologies, and applications." Communications of the Association for Information Systems 34, 65.

Webster, J. and R. T. Watson (2002). "Analyzing the past to prepare for the future: Writing a literature review." MIS quarterly 26, xiii-xxiii.

Wegen, B. v. and R. D. Hoog (1996). "Measuring the economic value of information systems." Journal of Information Technology 11(3), 247-260.

Wieder, B., M. Ossimitz and P. Chamoni (2012). "The impact of business intelligence tools on performance: a user satisfaction paradox?" International Journal of Economic Sciences and Applied Research 5(3), 7-32.

Willis, B. (2014). The advantages and limitations of single case study analysis. https://www.e-ir.info/2014/07/05/the-advantages-and-limitations-of-single-case-study-analysis/ (visited on 28 Apr 2019).

Winklhofer, H. M. and A. Diamantopoulos (2002). "Managerial evaluation of sales forecasting effectiveness." International Journal of Research in Marketing 19(2), 151–166.

Winters, P. R. (1960). "Forecasting sales by exponentially weighted moving averages." Management science 6(3), 324-342.

Witten, I. H., E. Frank, M. A. Hall and C. J. Pal (2016). Data Mining: Practical machine learning tools and techniques. Burlington, MA, USA: Morgan Kaufmann.

Wixom, B. H., B. Yen and M. Relich (2013). "Maximizing Value from Business Analytics." MIS Quarterly Executive 12(2), 111-123.

Wright, B. D. and G. N. Masters (1982). Rating scale analysis. Chicago, IL, USA: MESA Press.

Yasin, M. M. (2002). "The theory and practice of benchmarking: then and now." Benchmarking: An International Journal 9(3), 217-243.

Yin, R. K. (2017). Case study research and applications: Design and methods. 6th edition. Thousand Oaks, CA, USA: Sage Publications.

Yuan, D., X. Liu and Y. Yang (2015). "Dynamic on-the-fly minimum cost benchmarking for storing generated scientific datasets in the cloud." IEEE Transactions on Computers 64(10), 2781-2795.

Yusuf, Y. Y., M. Sarhadi and A. Gunasekaran (1999). "Agile manufacturing:: The drivers, concepts and attributes." International Journal of production economics 62(1-2), 33-43.

Zhang, G. P. (2003). "Time series forecasting using a hybrid ARIMA and neural network model." Neurocomputing 50, 159-175.

Zhang, L. and B. Zhang (1999). "A geometrical representation of McCulloch-Pitts neural model and its applications." IEEE Transactions on Neural Networks 10(4), 925-929.

Zotteri, G., M. Kalchschmidt and F. Caniato (2005). "The impact of aggregation level on forecasting performance." International Journal of production economics 93, 479–491.

Appendix A: Mathematical description of the Rasch algorithm

The Rasch model is a logit-linear model and can be estimated with a number of non-iterative (e.g. the normal approximation algorithm) and iterative (e.g., joint or conditional maximum likelihood estimation) methods (Wright and Masters, 1982). For a detailed comparison of statistical properties like accuracy, performance in the presence of missing data, bias, and consistency, see Linacre (2004). In the following, the model for a dichotomous response variable that assumes values 0 for failure and 1 for success will be described.

Given an individual's response $X_{n,i} \in \{0,1\}$ to I items where $n = 1..N$ denotes the individual with ability B_n and $i = 1..I$ denotes the item with difficulty D_i, the Rasch model states that

$$\log\left(\frac{\mathrm{P}_{n,i}^{1}}{\mathrm{P}_{n,i}^{0}}\right) = B_n - D_i \tag{A1}$$

where $P_{n,i}^{1}$ is the probability that individual n succeeds at item i (Linacre, 1999).

Based on equation (A1), the probability of response $X_{n,i} = x_{n,i}$ can be written as (Kelderman, 1984)

$$\mathrm{P}(X_{n,i} = x_{n,i}) = \frac{e^{x_{n,i}(B_n - D_i)}}{1 + e^{B_n - D_i}} \tag{A2}$$

One of the assumptions of this model is unidimensionality, which means that if an individual's performance depends on a single underlying trait, no further information on his performance can be obtained once his value of that single trait is known (Lord and Novick, 2008). Derived from this assumption, the infit mean square statistic (χ^2 statistic divided by the degrees of freedom) for Rasch models is the ratio of observed variance and expected variance (Wright and Masters, 1982; Linacre, 1999)

$$Infit_i = \frac{\sum_{n=1}^{N}\left(x_{n,i} - \tilde{x}_{n,i}\right)^2}{\sum_{n=1}^{N}\tilde{x}_{n,i}\left(1 - \tilde{x}_{n,i}\right)} \tag{A3}$$

where $x_{n,i}$ denotes the observed and $\tilde{x}_{n,i}$ the expected response of individual n for item i.

Similarly, the outfit mean square statistic (also a χ^2 statistic divided by the degrees of freedom) can be written as

$$Outfit_i = \frac{1}{N}\sum_{n=1}^{N}\frac{\left(x_{n,i} - \tilde{x}_{n,i}\right)^2}{\tilde{x}_{n,i}\left(1 - \tilde{x}_{n,i}\right)} \tag{A4}$$

Appendix B: Questionnaire for machine learning and analytics adoption

Machine learning and analytics in the finance function

The use of digital technologies, especially to handle large volumes of data, can lead to major improvements in the respective processes. In particular, planning, budgeting and forecasting should benefit from an implementation of analytics while accounting operations should benefit from machine learning. Still, machine learning and analytics (in the following we will refer to it as analytics only) adoption in the finance department is rather poor and only gradually increasing despite clear statements to do so in the past.

Against this background, we aim to understand two things:

- Do functional aspects like application area, methods applied, and type of data
- Do behavioral aspects like perceived usefulness, perceived ease of use, and rationale for use

have a significant influence on machine learning and analytics adoption?

We would like to get your expert input on this with the help of a survey. The questionnaire should take no longer than 15 minutes. We plan to show the results during our next public working group meetings. We will also gladly share results via e-mail if you send us your contact details. Of course, we will handle your answers as confidential.

Thank you for your support!

Prof. Dr. Peter Chamoni, Prof. Dr. Jörg Mayer,
Markus Eßwein,
Chair of Information Systems and Business Intelligence, University of Duisburg-Essen

1. [Current use] Currently, we use analytics ...

Note: By analytics we understand the use of mathematical methods, in particular from the field of statistics (e.g., correlations, predictive modeling) and computer science methods, in particular from the field of machine learning (e.g., supervised and unsupervised learning), to find patterns in data, better understand cause-and-effect chains, and make predictions.

We consider full-scale adoption to be the case when the results of the machine are taken as default values and human workers only build exceptions around this basis.

By use we mean regular actual use (not just the availability).

	Never	Rarely	Sometimes	Often	Always	Don't know
For single prototypes	○	○	○	○	○	○
Side-by side for certain use cases	○	○	○	○	○	○
Full-scale for certain use cases	○	○	○	○	○	○

2. [Application domain] We use analytics in the following sub-functions

	Never	Rarely	Sometimes	Often	Always	Don't know
Financial accounting (Rechnungswesen)	○	○	○	○	○	○
Management accounting (Controlling)	○	○	○	○	○	○
Treasury	○	○	○	○	○	○

3. [Application area] More in detail, we use analytics for the following use cases

	Never	Rarely	Sometimes	Often	Always	Don't know
Revenues	○	○	○	○	○	○
Costs	○	○	○	○	○	○
Working capital, accounts payables, accounts receivables	○	○	○	○	○	○
Risk	○	○	○	○	○	○

4. [Frequency] How often do you use analytics methods for the following tasks?

	Never	Rarely	Sometimes	Often	Always	Don't know
Planning	○	○	○	○	○	○
Budgeting	○	○	○	○	○	○
Forecasting	○	○	○	○	○	○

5. [Data types] As a basis for analytics we use the following data

	Never	Rarely	Sometimes	Often	Always	Don't know
Structured internal (transactional from ERP or aggregated from BI)	○	○	○	○	○	○
Semi- or unstructured internal (XML, logs, e-mails, sensor data)	○	○	○	○	○	○
Structured external (databases with indicators, e.g. weather)	○	○	○	○	○	○
Semi- or unstructured external (demographic, geospatial, social media)	○	○	○	○	○	○

6. [Methods applied] From the field of analytics we use

Note: In supervised learning the correct output for training data is known. A teacher provides the output value of the target function for all training examples ("labeled examples"). By contrast, in unsupervised learning data is unlabeled. There is no information except the input values for training examples and the main task is the automatic creation of classification rules.
By use we mean regular actual use (not just the availability).

	Never	Rarely	Sometimes	Often	Always	Don't know
Supervised - Prediction (e.g., credit default risk)	○	○	○	○	○	○
Supervised - Classification (e.g., customer credit-worthiness)	○	○	○	○	○	○
Supervised - Time-series forecasting (e.g., revenue forecasting)	○	○	○	○	○	○
Unsupervised - Association (e.g., market basket analysis)	○	○	○	○	○	○
Unsupervised - Segmentation (e.g., customer groups based on payment behavior)	○	○	○	○	○	○

7. [Decision steps] How much does analytics help you in the following steps of decision-making

Note: (1) Identification means the identification and definition of problems and opportunities. (2) Development means finding ready-made solutions or developing new solutions for the problem or opportunity. This step most often results in several decision alternatives. (3) Selection means screening and evaluating the alternatives and, finally, making a decision

	Not at all	Slightly	Somewhat	Very much	Extremely	Don't know
Identification of problems and opportunities	○	○	○	○	○	○
Development of solutions (decision alternatives)	○	○	○	○	○	○
Evaluation of alternatives and actual decision	○	○	○	○	○	○

8. [Rationale for use] Analytics helps us to...

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Don't know
Achieve a competitive advantage	○	○	○	○	○	○
Support our strategic and tactical goals	○	○	○	○	○	○
Improve our performance	○	○	○	○	○	○
Make better decisions	○	○	○	○	○	○
Have better decision processes	○	○	○	○	○	○
Produce knowledge	○	○	○	○	○	○
Obtain value from data	○	○	○	○	○	○

9. [Level of support] Analytics tool support will...

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Don't know
Not beat human experience	○	○	○	○	○	○
Support human workers in their tasks	○	○	○	○	○	○
Replace most of the efforts of human workers	○	○	○	○	○	○
Make human workers complete tasks more quickly	○	○	○	○	○	○
Make human workers complete tasks more accurately	○	○	○	○	○	○

10. [Intention to use] For my own work I will...

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Don't know
Stick to the tools and methods I am used to	○	○	○	○	○	○
Use the machine as a side-car	○	○	○	○	○	○
Transfer all I can to the machine	○	○	○	○	○	○

11. [Perceived ease of use] I am familiar with ...

	Not at all	Slightly	Moderately	Very	Extremely	Don't know
Descriptive statistics (mean, maximum, etc.) and correlations	○	○	○	○	○	○
Linear regression models and related statistics (t-test, etc.)	○	○	○	○	○	○
Time-series models (Exponential smoothing, moving averages, Holt-Winters, ARIMA, etc.)	○	○	○	○	○	○
Machine learning algorithms (neural networks, gradient boosting, etc.)	○	○	○	○	○	○

12. [Attitude towards using] In my opinion, analytics tool support would be very desirable for...

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Don't know
Planning	○	○	○	○	○	○
Budgeting	○	○	○	○	○	○
Forecasting	○	○	○	○	○	○

13. [Tools] What kind of tool support do you use for your analyses?

Note: Multiple answers are possible

- ☐ Spreadsheet (Excel)
- ☐ Enterprise Business Intelligence (SAP BO, MicroStrategy)
- ☐ Dashboarding and visual analytics (Tableau, PowerBI, etc.)
- ☐ Analytics platforms (Teradata, SAS)
- ☐ Open source platforms (Hadoop, R, Python)

14. [Driver] How is analytics driven in your organization?

Note: Multiple answers are possible

- ☐ Top-down with an analytics strategy
- ☐ Bottom-up via pain points and use cases
- ☐ None of the above

15. [Demographics 1/5] How many employees are in your company?

- <250
- 251-10,000
- 10,001-100,000
- >100,000

16. [Demographics 3/5] What is your position in the company?

- Upper management
- Middle management
- Lower-level management
- Workhorse

17. [Demographics 4/5] In which department are you?

- Finance department
- Business department
- IT department
- Sonstiges

18. [Demographics 2/5] How many employees are in your department of the company?

- <25
- 25-1,000
- >1,000

19. [Demographics 5/5] What is your role with respect to analytics

- End user
- Power user (incl. data scientist)
- Developer

Appendix C: List of R packages used

`>library(abind)` is a combination of `cbind` (column) and `rbind` (row) to combine multidimensional arrays (matrices).

`>library(caret)` comprises different functions for training and plotting classification and regression models.

`>library(corrplot)` allows to graphically display (correlation) matrices and do basic reordering of rows and columns.

`>library(devtools)` is a set of tools for the development of packages.

`>library(DiagrammeR)` is a package for graph or network visualization with a number of functions for adding and deleting nodes and edges.

`>library(forecast)` is a set of tools for the display and analysis of univariate time series and time-series forecasting.

`>library(forecastxgb)` is a package for efficient time-series forecasts using extreme gradient boosting.

`>library(ggplot2)` is a graphics framework that provides all kinds of visualizations for different types of data.

`>library(glmnet)` comprises efficient forms of the Lasso or elastic-net regularization path for linear regression, logistic and multinomial regression models, Poisson regression, and the Cox model. It also contains functions for cross-validation and boostrapping.

`>library(Hmisc)` collection of miscellaneous functions for data analysis, manipulation, visualization, and conversion.

`>library(leaps)` provides different methods for subset selection like forward and backward stepwise as well as exhaustive search.

`>library(missForest)` includes functions for the imputation of mixed (categorical and numerical) missing values using random forests.

`>library(mlr)` is a large collection of classification and regression techniques from the field of machine learning. It also includes resampling methods like cross-validation and bootstrap.

`>library(nnfor)` offers time-series forecasting with neural networks that can be specified automatically or manually.

`>library(pls)` is a package for partial least squares and principal component regression.

`>library(RColorBrewer)` provides color palettes that are well suited to graphical visualization.

`>library(seminr)` provides a syntax for partial least squares structural equation modeling that is close to natural language.

`>library(tidyverse)` is a collection of R packages for data science that all follow the same design philosophy and provide methods for data wrangling. The core packages include `ggplot2`, `tidyr`, `dplyr`, and `tibble`.

`>library(VSURF)` is a package for variable (subset) selection using random forests. Selection is performed in three steps: eliminate irrelevant variables, select variables related to response, refine selection by eliminating redundancy.

`>library(xgboost)` is an efficient implementation of a gradient boosting framework with linear model solvers and tree learning algorithms and highly parallelizable.

`>library(xlsx)` is one of the R connectors to Microsoft Excel's (all versions) .xls and .xlsx file formats. It is Java-based and allows to read, write, and format files.

Appendix D: Exemplary R code for PLS-SEM

```
library(xlsx) # Excel import
library(mlr) # machine learning
library(tidyverse) # data wrangling
library(missForest) # random forest imputation
library(seminr) # sem-pls estimation

### --------- Data load -------- ###

setwd("~/R")
surveyData <- read.xlsx("raw data/20190204_Model and answers.xlsx",
sheetName = "data (coded)")
surveyData[surveyData == 12345] <- NA

### --------- Imputation (not always used) -------- ###

# convert numerics to factors for categorial imputation
modelData <- surveyData[,-c(48:55)]

modelData.factors <- surveyData %>%
  select(-c(48:55)) %>%
  mutate_if(is.numeric, as.factor)

modelData.imp.lst <- missForest(modelData.factors)
modelData.imp.lst$OOBerror # check predicted error, here around 25%

# convert back to numeric, as required by sem-pls (conversion to
# character needed to get values instead of factor level labels)
modelData.imp <- modelData.imp.lst$ximp %>%
  mutate_if(is.factor, as.character) %>%
  mutate_if(is.character, as.numeric)

### --------- Estimate TAM only -------- ###

mm.TAM <- constructs(
  composite("Use", multi_items("CU", 1:3), weights = mode_B),
  composite("PU", multi_items("DEC", 1:3), weights = mode_B),
  composite("PU", multi_items("RfU", c(1,3,5,6)), weights = mode_B),
  composite("PU", multi_items("SUP", c(4,5)), weights = mode_B),
  composite("ItU", single_item("OWO2")),
  composite("PEoU", multi_items("FAM", 2:3), weights = mode_B),
  composite("AtU", multi_items("DES", 1:3), weights = mode_B))

sm.TAM <- relationships(
  paths(from = "PEoU", to = c("PU", "AtU")),
  paths(from = "PU", to = c("AtU", "ItU", "Use")),
  paths(from = "AtU", to = "ItU"),
  paths(from = "ItU", to = "Use"))

pls.TAM <- estimate_pls(data = modelData, #or modelData.imp
                           measurement_model = mm.TAM,
                           structural_model = sm.TAM,
                           inner_weights = path_weighting)

boot.TAM <- bootstrap_model(pls.TAM, nboot = 1000, cores = 4)
boot.TAM$path_coef
boot.TAM$loadings
summary(boot.TAM)
```

Appendix E: Exemplary R code for use case 2

```
# General information
Investigate different training and forecasting periods

Training period   | Test period
----------------- | -------------
Q1/2014 - Q4/2016 | Q1 - Q4/2017 (four quarters)
Q1/2014 - Q1/2017 | Q2 - Q4/2017 (three quarters)
Q1/2014 - Q2/2017 | Q3 - Q4/2017 (two quarters)
Q1/2014 - Q3/2017 | Q4/2017 (year-end forecast)

### --------- Set up environment and load data -------- ###
# Load libraries
library(ggplot2) # Plotting graphs
library(glmnet) # LASSO
library(forecast) # Predictions
library(plyr) # Data transformation
library(tseries) # Time series analysis
library(leaps) # Subset selection
library(caret) # Preprocessing
library(Hmisc) # P values for correlations
library(zoo) # YYYY-MM to date conversion
library(caret) # Linear combination detection
library(tidyr) # Reshaping of tables
library(reshape2) # Reshaping of tables
library(xgboost) # Extreme gradient boosting
library(DiagrammeR) # Visualizing tree from XGB

# next two required for computation of optimal lambda in glmnet cross-
validation
library(parallel) # Parallel execution of commands
library(data.table) # Data tables

library(devtools)
#devtools::install_github("ellisp/forecastxgb-r-package/pkg")
library(forecastxgb)

library(pls) # principal component analysis
library(VSURF) # random forest subset selection
library(corrplot) # Correlation plots (needs to be loaded after pls)
library(nnfor) # Neural network time series forecast
library(RColorBrewer) # For specific color

# Set working directory and seed
setwd('[...]CF Forecasting/R')
set.seed(100)

# Load external data
# Load indices and other
BCI.raw <- read.csv("External Data/BCI.csv", header = TRUE, sep = ";")
# Business confidence indicator
CCI.raw <- read.csv("External Data/CCI.csv", header = TRUE, sep = ";")
# Consumer confidence index
CLI.raw <- read.csv("External Data/CLI.csv", header = TRUE, sep = ";")
# Composite leading index
CPI.raw <- read.csv("External Data/CPI.csv", header = TRUE, sep = ";")
# Consumer price index (Inflation)
```

```
HPI.raw <- read.csv("External Data/HVPI.csv", header = TRUE, sep =
";")   # Harmonisierter Verbraucherpreisindex
PPI.raw <- read.csv("External Data/PPI.csv", header = TRUE, sep = ";")
# Producer price index

# Load weather data
TMP_MUC.raw <- read.csv("External Data/TMP_MUC.csv", header = TRUE,
sep =";")   # Temperature Munich
PRE_MUC.raw <- read.csv("External Data/PRE_MUC.csv", header = TRUE,
sep =";")   # Precipitation Munich
SUN_MUC.raw <- read.csv("External Data/SUN_MUC.csv", header = TRUE,
sep =";")   # Sunshine duration Munich
SNO_MUC.raw <- read.csv("External Data/SNO_MUC.csv", header = TRUE,
sep =";")   # Snowfall Munich

#Load spot price for electricity and gas
ELE.raw <- read.csv("External Data/ELE.csv", header = TRUE, sep = ";")
# Spot electriciy (day ahead)
GAS.raw <- read.csv("External Data/GAS.csv", header = TRUE, sep = ";")
# Spot NCG Gas (day ahead)

# Load internal Data
CF.raw <- read.csv("Internal Data/CF.csv", header = TRUE, sep = ";")
# Cash flow
CNS.raw <- read.csv("Internal Data/AVG_CONS.csv", header = TRUE, sep =
";")                 # Average consumption
SALES_ELE.raw <- read.csv("Internal Data/SALES_ELE.csv", header =
TRUE, sep = ";")         # Gesamtabsatz Elektrizität Energiebilanz
SALES_GAS.raw <- read.csv("Internal Data/SALES_GAS.csv", header =
TRUE, sep = ";")         # Gesamtabsatz Gas Energiebilanz
CUST_ELE_P.raw <- read.csv("Internal Data/CUST_ELE_P.csv", header =
TRUE, sep = ";")       # Total number of P electricity customer
CUST_GAS_P.raw <- read.csv("Internal Data/CUST_GAS_P.csv", header =
TRUE, sep = ";")       # Total number P gas customer
CUST_ELE_SME.raw <- read.csv("Internal Data/CUST_ELE_SME.csv", header
= TRUE, sep = ";")   # Total number SME electricity customer
CUST_GAS_SME.raw <- read.csv("Internal Data/CUST_GAS_SME.csv", header
= TRUE, sep = ";")   # Total number SME gas customer
WADTB_ELE.raw <- read.csv("Internal Data/WADTB_ELE_SLP.csv", header =
TRUE, sep = ";")     # Weighted average days to bill electricity slp
WADTB_GAS.raw <- read.csv("Internal Data/WADTB_GAS_SLP.csv", header =
TRUE, sep = ";")     # Weighted average days to bill gas slp
WADTM_ELE.raw <- read.csv("Internal Data/WADTM_ELE_SLP.csv", header =
TRUE, sep = ";")     # Weighted average days to meter electricity slp
WADTM_GAS.raw <- read.csv("Internal Data/WADTM_GAS_SLP.csv", header =
TRUE, sep = ";")     # Weighted average days to meter gas slp
FORD.raw <- read.csv("Internal Data/FORD_third_party.csv", header =
TRUE, sep = ";")       # Forderungen a LuL dritte

### --------- Preprocess data -------- ###
# Lag functions
# Lag pad
lagpad <- function(x, k) {
  if (k>0) {
    return (c(rep(NA, k), x)[1 : length(x)] );
  }
  else {
    return (c(x[(-k+1) : length(x)], rep(NA, -k)));
  }
}
```

```
# Lag single
lagExtSingle <- function(extInd){
  extInd.lags <- list()
  extInd.lags$lag0 <- lagpad(extInd, 0)
  extInd.lags$lag1 <- lagpad(extInd, 1)
  extInd.lags$lag2 <- lagpad(extInd, 2)
  extInd.lags$lag3 <- lagpad(extInd, 3)
  extInd.lags$lag4 <- lagpad(extInd, 4)
  return(extInd.lags)
}

# Lag external and internal data
# Lag all external indicators by 0-4 periods
BCI <- lagExtSingle(BCI.raw[,-c(1:3)])
CCI <- lagExtSingle(CCI.raw[,-c(1:3)])
CLI <- lagExtSingle(CLI.raw[,-c(1:3)])
CPI <- lagExtSingle(CPI.raw[,-c(1:3)])
HPI <- lagExtSingle(HPI.raw[,-c(1:3)])
PPI <- lagExtSingle(PPI.raw[,-c(1:3)])

# Lag weather data by 0-4 periods
TMP_MUC <- lagExtSingle(TMP_MUC.raw[,-c(1:3)])
PRE_MUC <- lagExtSingle(PRE_MUC.raw[,-c(1:3)])
SUN_MUC <- lagExtSingle(SUN_MUC.raw[,-c(1:3)])
SNO_MUC <- lagExtSingle(SNO_MUC.raw[,-c(1:3)])

# Lag spot price for electricity and gas by 0-4 periods
ELE <- lagExtSingle(ELE.raw[,-c(1:3)])
GAS <- lagExtSingle(GAS.raw[,-c(1:3)])

# Drop N/A values
# Drop values for 2013 (only used for lags)
BCI <- lapply(BCI, function(x){x[-c(1:4)]})
CCI <- lapply(CCI, function(x){x[-c(1:4)]})
CLI <- lapply(CLI, function(x){x[-c(1:4)]})
CPI <- lapply(CPI, function(x){x[-c(1:4)]})
HPI <- lapply(HPI, function(x){x[-c(1:4)]})
PPI <- lapply(PPI, function(x){x[-c(1:4)]})
TMP_MUC <- lapply(TMP_MUC, function(x){x[-c(1:4)]})
PRE_MUC <- lapply(PRE_MUC, function(x){x[-c(1:4)]})
SUN_MUC <- lapply(SUN_MUC, function(x){x[-c(1:4)]})
SNO_MUC <- lapply(SNO_MUC, function(x){x[-c(1:4)]})
ELE <- lapply(ELE, function(x){x[-c(1:4)]})
GAS <- lapply(GAS, function(x){x[-c(1:4)]})

### --------- Set up model and compute statistics -------- ###
# Build model
# Save factors as data frame. Hereby able to put into CF.model in the
next step
BCI = as.data.frame(BCI)
CLI = as.data.frame(CLI)
CCI = as.data.frame(CCI)
CPI = as.data.frame(CPI)
HPI = as.data.frame(HPI)
PPI = as.data.frame(PPI)
ELE = as.data.frame(ELE)
GAS = as.data.frame(GAS)
TMP_MUC = as.data.frame(TMP_MUC)
SUN_MUC = as.data.frame(SUN_MUC)
PRE_MUC = as.data.frame(PRE_MUC)
SNO_MUC = as.data.frame(SNO_MUC)
```

```
#Build final CF model for further investigation
CF.model <- data.frame(CF = CF.raw$VALUE[1:16], CNS.lag0 =
CNS.raw$VALUE[1:16],
                       SALES_ELE = SALES_ELE.raw$VALUE[1:16],
SALES_GAS = SALES_GAS.raw$VALUE[1:16],
                       CUST_ELE_P = CUST_ELE_P.raw$VALUE[1:16],
CUST_GAS_P = CUST_GAS_P.raw$VALUE[1:16],
                       CUST_ELE_SME = CUST_ELE_SME.raw$VALUE[1:16],
CUST_GAS_SME = CUST_GAS_SME.raw$VALUE[1:16],
                       FORD = FORD.raw$VALUE[1:16],
                       BCI = BCI[1:16,], CLI = CLI[1:16,],
                       CCI = CCI[1:16,], CPI = CPI[1:16,],
                       HPI = HPI[1:16,], PPI = PPI[1:16,],
                       ELE = ELE[1:16,], GAS = GAS[1:16,],
                       TMP_MUC = TMP_MUC[1:16,], SUN_MUC =
SUN_MUC[1:16,],
                       PRE_MUC = PRE_MUC[1:16,], SNO_MUC =
SNO_MUC[1:16,])


# Compute fits for historical cash flow data
getFits.hist <- function(CF.model, train){
    x = model.matrix(CF ~ ., data = CF.model[train,])
    x.ts <- ts(data = x, frequency = 4, start = c(2014,1))
    x.fourier <- fourier(x.ts, K = 2)
    rownames(x.fourier) <- rownames(x)

    y = CF.model[train, 1]
    y.ts <- ts(data = y, frequency = 4, start = c(2014,1))

    fit.HW <- HoltWinters(y.ts) # Holt-Winters
    fit.arima <- auto.arima(y.ts) # ARIMA
    fit.xgbar <- xgbar(y.ts, maxlag = 4, nrounds = 10, seas_method =
"decompose") # Extreme gradient boosting
    fit.mlp <- mlp(y.ts) # Multi-layer perceptron neural network
    fit.elm <- elm(y.ts, reps = 100) # Extreme learning machine neural
network with Lasso estimation of output layer weights
    fit.elm.step <- elm(y.ts, type = "step", reps = 100) # ELM NN with
stepwise estimation of output layer weights
    fit.elm.50 <- elm(y.ts, hd = 50, reps = 100) # ELM NN with 50
instead of 5 hidden nodes (all in one layer)

    fits.hist <- list()
    fits.hist <- list(HW = fit.HW, ARIMA = fit.arima, XGbar =
fit.xgbar, MLP = fit.mlp,
                      ELM = fit.elm, ELM.step = fit.elm.step, ELM.50 =
fit.elm.50)

  return(fits.hist)
}

# Compute predictions with historical cash flow data
getPreds.hist <- function(fits.hist, CF.model, train){
  pred.HW <- forecast(fits.hist[[1]], h = 16-length(train))
  pred.arima <- forecast(fits.hist[[2]], h = 16-length(train))
  pred.xgbar <- forecast(fits.hist[[3]], h = 16-length(train))
  pred.mlp <- forecast(fits.hist[[4]], h = 16-length(train))
  pred.elm <- forecast(fits.hist[[5]], h = 16-length(train))
  pred.elm.step <- forecast(fits.hist[[6]], h = 16-length(train))
  pred.elm.50 <- forecast(fits.hist[[7]], h = 16-length(train))
```

```
  preds.hist <- list()
  preds.hist <- list(HW = pred.HW, ARIMA = pred.arima, XGbar =
pred.xgbar, MLP = pred.mlp, ELM = pred.elm,
                     ELM.step = pred.elm.step, ELM.50 = pred.elm.50)

  return(preds.hist)
}

# Compute deviations
getDevs.hist <- function(preds.hist, CF.model, train){
  dev.HW <- preds.hist[[1]]$mean - CF.model[-train, 1]
  dev.arima <- preds.hist[[2]]$mean - CF.model[-train, 1]
  dev.xgbar <- preds.hist[[3]]$mean - CF.model[-train, 1]
  dev.mlp <- preds.hist[[4]]$mean - CF.model[-train, 1]
  dev.elm <- preds.hist[[5]]$mean - CF.model[-train, 1]
  dev.elm.step <- preds.hist[[6]]$mean - CF.model[-train, 1]
  dev.elm.50 <- preds.hist[[7]]$mean - CF.model[-train, 1]

  devs.hist <- data.frame(HW = dev.HW, ARIMA = dev.arima, XGbar =
  dev.xgbar, MLP = dev.mlp, ELM = dev.elm, ELM.step = dev.elm.step,
  ELM.50 = dev.elm.50)

  return(devs.hist)
}

train_Q4_2016 <- seq(1:12)
train_Q1_2017 <- seq(1:13)
train_Q2_2017 <- seq(1:14)
train_Q3_2017 <- seq(1:15)

datevec = seq.Date(as.Date("2014/1/1"), as.Date("2017/12/1"),
"quarter") # Create date vector with Year and quater ("YYYY-Qx")
quarters = as.character(as.yearqtr(2014 + seq(0,15)/4))
# Create date vector with Year and quater ("YYYY Qx")

fits.hist_Q4_2016 <- getFits.hist(CF.model, train_Q4_2016)
preds.hist_Q4_2016 <- getPreds.hist(fits.hist_Q4_2016, CF.model,
train_Q4_2016)
devs.hist_Q4_2016 <- getDevs.hist(preds.hist_Q4_2016, CF.model,
train_Q4_2016)
round(devs.hist_Q4_2016,2)

### --------- Visualize -------- ###
ggplot(aes(x = datevec[-train_Q4_2016], y = CF, color = "Actual"),
data = CF.model[-train_Q4_2016,]) + geom_point() +
  geom_line(aes(y = preds.hist_Q4_2016[[1]]$mean, color = "Holt-
Winters")) +
  geom_line(aes(y = preds.hist_Q4_2016[[2]]$mean, color = "ARIMA")) +
  geom_line(aes(y = preds.hist_Q4_2016[[3]]$mean, color = "XGbar")) +
  geom_line(aes(y = preds.hist_Q4_2016[[4]]$mean, color = "MLP")) +
  geom_line(aes(y = preds.hist_Q4_2016[[5]]$mean, color = "ELM")) +
  geom_line(aes(y = preds.hist_Q4_2016[[6]]$mean, color = "ELM.step"))
+
  geom_line(aes(y = preds.hist_Q4_2016[[7]]$mean, color = "ELM.50")) +
  ggtitle("CF Prediction vs. Actuals") +
  xlab("Time") +
  ylab("Cash flow [mEUR]") +
  scale_x_date(breaks = datevec[-train_Q4_2016], labels = quarters[-
train_Q4_2016]) +
  theme(axis.text.x = element_text(face="bold", angle=90))
```

Acknowledgements

First of all, I would like to thank Peter Chamoni for having me as his last PhD student, for giving me all the freedom to shape my project and to travel a lot for it, for his excellent supervision, and for introducing me to numerous people along the way. I also want to thank Jochen Gönsch for providing guidance and the second assessment of my thesis.

Moreover, my gratitude goes to my colleagues, Silke Bandyszak and Daniel Erkal for the nice company and discussions, Lena Grünhagen and Stefan de Dios Panal for the nice company during lunch breaks and beyond. Additionally, I would like to thank our “seniors” for all the nice chats, for diversion, and for making the work so much more enjoyable. In particular, I would like to thank Illa Frigge for all the administrative support, for caring for my plants, for meticulously repairing my flower pot, and for bringing me down to earth when I was new; Martina Reinersmann for entrusting me with her zeb and FOM students, for her excellent mentoring and thoughtful feedback, and for seeing so many things the way I see them; Stefan Krebs for being the kind soul of the MSM, for his passion when wrapping gift cards, for his help with manual labor and the finalization of this manuscript, and for the relaxing hours on our balcony.

I am also grateful to Jörg Mayer for re-igniting the PhD spark in me, for mentoring me throughout, for providing me with all the opportunities to gather data and work with practitioners, and for entrusting me with the administration of our Schmalenbach working group “Digital Finance” and helping me become an integral part of it. Many thanks go to the members of this working group, without whom I could not have achieved anything like this thesis.

Furthermore, I am thankful to my master’s students who helped me with many of the building blocks of my project and who made me learn so much about supervision, in particular Jana de Leeuw, Diana Sedneva, Sebastian Stoffel, Oliver Stritzel, Jessica Wewer, and Mareen Wienand.

Making my time in Duisburg such a pleasure, I would like to thank the Duisburg Lacrosse team for the super team spirit on and off the field. Especially, Maggie

Quested for the great time and for trying to teach me British English with rather limited success, and Nathalie Junghannß for our Ritterkampf group and Spätzle.

Special thanks to my parents for their love, a lifetime of encouragement, support, and faith in me, for setting me up with the brains for such an endeavor, and for letting me grab an almost entire household for Duisburg without having to spend a penny, and to my grandfather for being so enduring, amiable, and a role model.

Finally, I would like to express my most profound thanks to Domenica Martorana, for all her support and love, for telling me about matching colors, for accepting my sometimes spontaneous ideas like doing a PhD, for providing me with an endless stream of delicious cake, for making me a quarter Italian, and for always being on my side.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbständig und ohne unerlaubte Hilfe angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht.

Duisburg, den 08.07.2019

DOI: 10.17185/duepublico/70800
URN: urn:nbn:de:hbz:464-20191127-142800-9