Contributions to Data Analytics Techniques with Applications in Forecasting, Visualization and Decision Support

Von der Wirtschaftswissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

- Doctor rerum politicarum -

genehmigte Dissertation

von

Dennis Eilers, M.Sc.

geboren am 20.06.1993 in Hannover

2018

Betreuer und Gutachter:	Prof. Dr. Michael H. Breitner
Weiterer Gutachter:	Prof. Dr. Hans-Jörg von Mettenheim
Vorsitzender der Prüfungskommission:	Prof. Dr. Stefan Wielenberg
Weiteres Mitglied (beratend):	Dr. Michelle Muraz
Tag der Promotion:	02. November 2018

Für meine Oma.
In Erinnerung, dass bei allem was wir tun, die Liebe das Wichtigste im Leben ist.

Danksagung

Begegnungen sammelt man auf seinem Weg mit vielen Menschen. Die Richtung allerdings prägen nur Wenige. Den Menschen die meinen Weg geprägt haben, möchte ich an dieser Stelle Danke sagen.

Einer dieser Menschen ist mein Doktorvater Prof. Dr. Michael H. Breitner. Sie haben mir nach dem ersten Semester die Möglichkeit gegeben bei Ihnen als Studentische Hilfskraft zu arbeiten und mir so frühzeitig einen Blick hinter die Kulissen der akademischen Welt gewährt. Für das Vertrauen bin ich Ihnen sehr dankbar. In den vergangenen Jahren habe ich immer versucht etwas davon zurückzugeben, auch wenn ich zuweilen sicher ein unbequemer Student und Doktorand war, der gerne die fachlich kritische Auseinandersetzung gesucht hat. Ihre Geduld, uneingeschränkte Förderung aber auch klar formulierte Kritik werden mir im Gedächtnis bleiben. Jeder dieser drei Punkte hat mich gleichermaßen vorangebracht.

Einen akademischen Mentor, Förderer und guten Freund habe ich während meiner Zeit am Institut in Prof. Dr. Hans-Jörg von Mettenheim gefunden. Bei dir hatte ich die Freiheit meine Ideen umsetzen zu können und gleichermaßen Leitplanken, die meine Arbeit auf den richtigen Weg gebracht haben. Unsere vielen Gespräche waren für mich eine sowohl fachliche als auch persönliche Bereicherung, die ich nicht mehr missen möchte. Vielen Dank!

Über den Satz "Na, bist du auch am IWI HIWI" von Daniel Olivotti freue ich heute noch besonders, denn als Freund, Leidensgenosse, Reisehelfer, Co-Autor, Küchenbauer, Gesprächspartner, Berater und nebenbei Kollege hast du mir nicht nur mit deinem Eis das Leben süßer gemacht. ;-) Für deine Unterstützung ein großes Dankeschön! Von der Tapete bis zur Wand und darüber hinausdenken, konnte ich am besten mit meinem Bürogenossen Rouven Wiegard. Danke für eine tolle Zeit und eine super Freundschaft! Mit Jean-Henrick Schünemann im Büro waren dann die großen Fragen der Menschheit kein Problem mehr. Als Freund, Bürogenosse und philosophischer Sparringspartner unschlagbar. Danke Joi! Bei meinen bisher noch nicht genannten Co-Autoren Christian L. Dunis, Dennis Gercke, Christoph Gleue und Cornelius Köpp möchte ich mich ebenfalls ganz herzlich bedanken! Bei allen Kollegen die ich im Laufe der Zeit kennenlernen durfte bedanke ich mich genauso herzlich! Dankbar bin ich auch für die vielen Studenten mit denen ich in Vorlesungen, Übungen, Seminaren und Abschlussarbeiten zusammenarbeiten durfte. Eure Ideen und Feedback haben mich angetrieben auch immer mein Bestes in der Lehre und Betreuung zu geben!

Und abschließend möchte ich mich bei den Menschen bedanken, die mir privat die Kraft und Unterstützung haben zukommen lassen, ohne die eine solche Arbeit nicht möglich ist. Allen voran meine Eltern, die mir immer die Sicherheit gegeben haben, die ich brauchte. Und bei meinen Freunden, die mich unterstützt und ausgehalten haben.

Abstract

This cumulative dissertation summarizes and critically discusses seven peer-reviewed publications where I was involved as a co-author. All publications contribute to data analytics techniques. The dissertation consists of four main sections.

(1) Machine Leaning in Finance: In this section a Decision Support Algorithm based in Reinforcement Learning is introduced which filters rule-based trading decisions. We contribute to the literature by describing the implementation of the algorithm. We also provide empirical evidence of financial market anomalies.

(2) Mining Customer Reviews: Opinions from customers about certain products are more and more expressed on social media platforms. Here we provide the first study which analyses YouTube comments as a data source for an aspect-based Sentiment Analysis. We also contribute to the literature by proposing a filtering method based on Google Trends which sorts product aspects according to their relevance for the customers.

(3) Forecasting Resale Prices of Used Cars: In this section we show how to efficiently forecast resale prices of used cars with Artificial Neural Networks. We provide lessons learned about long-term forecasts. We also provide insights in the importance of certain independent factors which determine the resale price.

(4) Visual Model Evaluation: The research in this section is mainly driven by the question of how to better incorporate human domain knowledge in data science. We develop a visualization technique based on heat maps which provides a more intuitive view on errors of a machine learning model. The visualization technique allows domain experts to discuss the results of machine learning models with data science experts on the same level of complexity.

Keywords: Reinforcement Learning, Artificial Neural Networks, Sentiment Analysis, Leasing, Used Cars, Feature Engineering, Domain Knowledge, Visualization.

Zusammenfassung

Diese kumulative Dissertation fast sieben von mir mitverfasste peer-reviewed Publikationen zusammen und diskutiert diese kritisch. Alle Publikationen leisten einen Beitrag zu Data Analytics Techniken. Die Dissertation ist in vier Hauptbereich eingeteilt.

(1) Maschinelles Lernen in Finance: In diesem Kapitel wird ein Entscheidungsunterstützungsalgorithmus basierend auf Bestärkendem Lernen vorgestellt, welcher regelbasierte Tradingentscheidungen filtert. Wir leisten einen Beitrag zur Literatur indem wir die Implementierung des Algorithmus beschreiben und zusätzlich empirische Hinweise auf Finanzmarktanomalien liefern.

(2) Mining von Kundenrezensionen: Meinungen von Kunden zu einem bestimmten Produkt, werden zunehmend auf sozialen Netzwerken gepostet. Wir liefern die erste Studie, die YouTube Kommentare als Datenquelle für eine aspektbasierte Sentiment Analyse untersucht. Wir leisten zudem einen Beitrag zur Literatur, durch die Entwicklung einer Filtermethode basierend auf Google Trends zur Sortierung von Produkteigenschaften in Abhängigkeit ihrer Relevanz für die Kunden.

(3) Vorhersage der Wiederverkaufspreise von Gebrauchtfahrzeugen: In diesem Kapitel zeigen wir, wie die Wiederverkaufspreise von Gebrauchtwagen mit Künstlichen Neuronalen Netzen effizient prognostiziert werden können. Wir liefern Erkenntnisse über langfristige Prognosen. Wir geben auch einen Einblick in die Bedeutung bestimmter unabhängiger Faktoren, die den Wiederverkaufspreis bestimmen.

(4) Visuelle Modellevaluation: Die Forschung in diesem Abschnitt wird hauptsächlich von der Frage geleitet, wie menschliches Domänenwissen besser in Data Science Projekte integriert werden kann. Wir entwickeln dazu eine Visualisierungstechnik basierend auf Heatmaps, die eine intuitivere Sicht auf Fehler eines maschinellen Lernmodells bietet. Die Visualisierungstechnik ermöglicht Domänenexperten, die Ergebnisse von Modellen des maschinellen Lernens mit Datenexperten auf der gleichen Komplexitätsebene zu diskutieren.

Schlüsselwörter: Bestärkendes Lernen, Künstliche Neuronale Netze, Sentiment Analyse, Leasing, Gebrauchtfahrzeuge, Feature Engineering, Domain Knowledge, Visualisierung.

Management Summary

Analyzing data for example to make better decisions, find hidden insights or improve customer satisfaction is one of the major challenges in modern, digitized industries. Almost every part of today's economy is affected by the possibilities of data analytics. Hence, there exist a tremendous interest in this topic, both in research and practice. This cumulative dissertation contributes to four research streams within the broad topic of data analytics.

1. Machine Learning in Finance
2. Mining Customer Reviews
3. Forecasting Resale Prices
4. Visual Model Evaluation

(1) The first part of this dissertation (Section 2) empirically investigates the existence of calendar anomalies and empirical regularities on financial markets and propose a new decision support algorithm based on Reinforcement Learning (RL) which allows to build an intelligent trading system which self-learns and adapts to new market situations. The findings from the empirical part of the paper show some weak evidence that certain anomalies like the new year and Easter effect exist in major stock indices. But the results also show that exploiting these anomalies by a simple trading strategy is hardly possible. To improve trading decisions, the main part of the publication proposes a new algorithm which analyzes the market based on certain characteristics like past returns and technical indicators and provides recommendations of how to position oneself. The basis for this algorithm is the idea of RL which is inspired by the learning process of humans. By rewarding good decisions and punishing bad decisions, the algorithm learns how to act such that the own actions lead to the highest expected reward. The process is visualized in Figure 1.

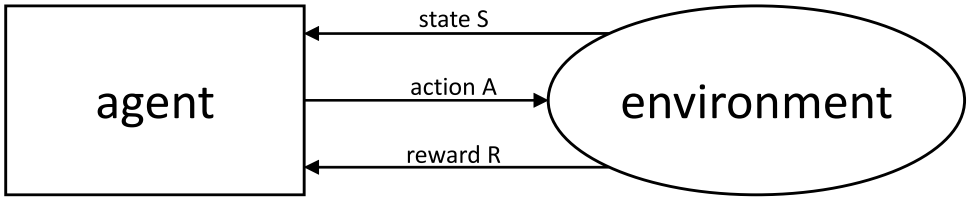


Figure 1: Idea of Reinforcement Learning

The results show that the trading strategy which is supported by RL, outperforms buy & hold as well as the naive anomaly strategies. Figure 2 provides an illustration of how the algorithm works in practice.

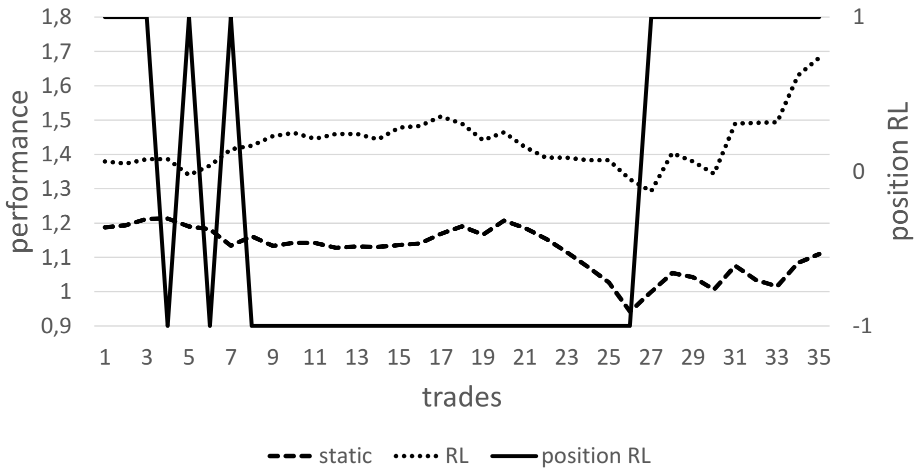


Figure 2: Trading Performance Improvements with Reinforcement Learning (RL)

The solid line represents the position of the trading strategy. A 1 indicate a long strategy, a -1 a short strategy and 0 a neutral position. The two dashed lines compare the return development of a static strategy which simply follow the anomaly exploitation rules from the literature and a RL strategy which is also based on the same rule but filters the results based on past experience. It can be seen that higher returns and also better reward to risk ratios (smaller maximum drawdowns) can be achieved by applying the RL filter.

(2) The next study (Section 3) contributes to the text mining literature. In e-commerce it becomes more and more important to understand what customers think about products or services to improve their experience. Opinions are often available in text form like conventional online reviews but also from an increasing number of social media data where users post their experience. Automatically analyzing the

unstructured text data in the context of product evaluations is known as Sentiment Analysis (SA) or Opinion Mining. SA is divided into several subtasks. Figure 3 illustrates the structure.

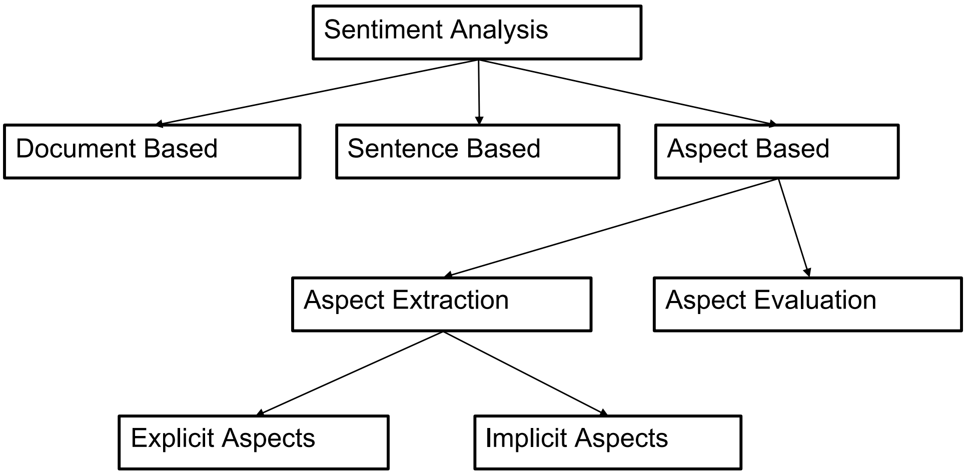


Figure 3: Structure of Sentiment Analysis

While document based and sentence based analysis try to extract the overall opinion about a product from a text/sentence, the aspect based analysis tries to extract opinions regarding certain important aspect of a product like the camera of a smartphone. This task is again divided into aspect extraction and aspects evaluation tasks. We are focusing on the aspect extraction task in our research which aims to find the aspect which are discussed in the text. Our first contribution to the literature is a comparison between conventional review data from Amazon reviews and noisier social media data from YouTube. We provide the first study which explicitly identifies YouTube comments to product related videos as a suitable source of information for an aspect based SA. Our comparisons show that a standard aspect based sentiment algorithm performs equally well on Amazon reviews and YouTube comments. Our second contribution is a filter mechanism which incorporates information from Google Trends about the search volumes of products in conjunction with their aspects. The assumption is that customers tend to search for products in conjunction with important aspects. Filtering potential aspects based on their search volume further increase the aspect extraction results. One particular problem remains namely the extraction of implicit aspects. There exist a difference between explicit aspect like a camera for a smartphone and implicit aspect like the weight of the smartphone which is only indirectly mentioned by adjectives like heavy and light.

(3) The third part of this dissertation (Section 4) discusses the findings from a research project with the goal to forecast resale prices of used cars. The initial situation is that a large car manufacturer faces the challenge of setting proper leasing rates for their cars. The main determinant for the leasing rate is the expected resale price of the car at the end of the contract. To improve the situation the car manufacturer started to collect all available contract data and merge them with the realized price on the used car market of the corresponding car. In the collaborative project we develop a forecasting model based on Artificial Neural Networks (ANN), which learns the dependencies. Figure 4 provides an overview of the implemented Decision Support System (DSS).

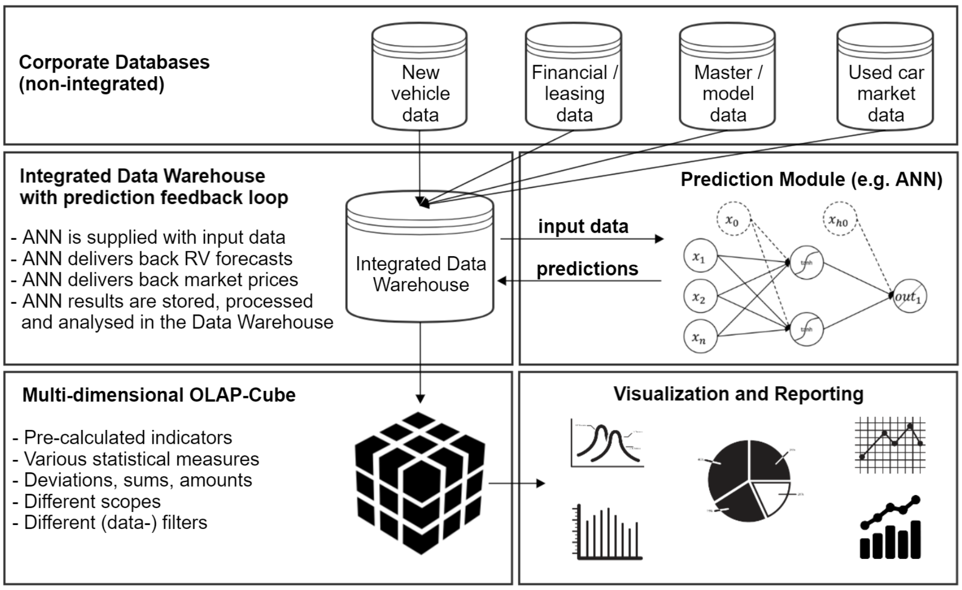


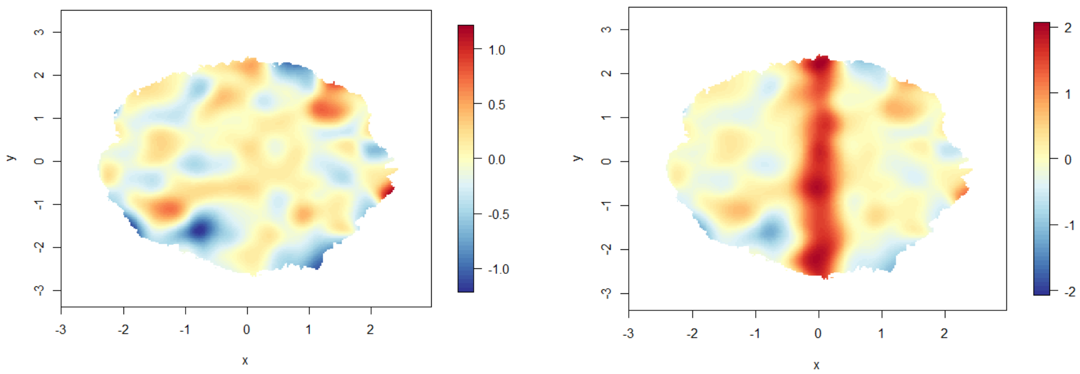
Figure 4: Decision Support System to Forecast Resale Prices

The contribution of this part is threefold. First, we show how the DSS is designed and implemented. Second, we investigate the influence/predictive power of the available independent variables. And third, we provide recommendations about how to design the forecasting application with ANNs while facing challenges like long-term forecast horizons, noisy data and time dependent variables.

The empirical findings show that the influence of external variables like oil price, stock index or consumer index are negligible. After controlling for all available information about the car and the contract specifications, none of the external factors show significant explanatory (linear mixed effects models and likelihood ratio test), nor predictive power (perturbation ranking algorithm). Another finding is the influence of the time factor on the forecasting performance. Our unique data set show that resale prices are subject to seasonal patterns and a trend. Again after controlling for all available information, the constructed time series of regression residuals are used for a time series decomposition which reveal higher resale prices during spring and lower resale prices during the end of the year. The implemented forecasting application then shows that ANNs with the capability of incorporating non-linear dependencies, outperform ridge linear regression models. The problem is to incorporate the time dependent variables. One possible solution is to combine ANNs to incorporate the non-linear dependencies in the data and incorporating the time dependent factors by a linear adjustment which provides a one and a half year unbiased out-of-sample performance in our tests.

(4) The forth part of this dissertation (Section 5) addresses the problem of the communication and understanding gap between data science experts and domain experts. In real world industry applications, the complex machine learning and data analytics task are carried out by highly skilled experts with a quantitative background. The problem is that these experts not necessarily have the required domain expertise to incorporate all relevant aspect and dependencies into the model. The result can be models which are biased or may lead to wrong conclusions in the worst case. On the other hand, domain experts who have worked many years in their field often lag understanding of the complex machine learning models which can result in mistrust and lack of acceptance. The consequence of both cases can be a non-optimal decision-making process. Therefore, our goal is to address this issue by proposing a new visualization technique for regression model performance based on heat maps. Heat maps are a familiar visualization tool for managers/decision

makers (domain experts) and data scientist alike. The idea is to use an intuitive visualization technique to enable a better communication between data scientist and domain experts about the current results of the model. With the help of this discussion, one can identify possible new independent variables (features) which might be previously overlooked. Figure 5 provides an impression of how the heat map visualization can provide an overview of the model performance in different regions of a data space.



(a) White Noise

(b) Bias

Figure 5: Heat Map Visualization of Model Performance

The heat map shows the error distribution (color scale) of a machine learning model depending on two different features. On the left-hand side (Figure 5a), we see a white noise with no clear patterns in the data. This is what can be expected if the models work reasonably well and no patterns remain in the data which are not explained by the model. On the right-hand side (Figure 5b), we see an example of a biased model which can be the result of a missing variable that is not properly reflected in the model specification.

The idea of model building, discussing and adjustment is an iterative process which should facilitate the so-called Feature Engineering (FE), a process of constructing a proper input space for a machine learning model. One example where the visualization technique is already successfully applied is the resale price forecast application from the previous part of the dissertation. Here we were able to identify three additional features which have improved the forecasting performance of the models in economically relevant magnitudes.

Contents

List of Figures

List of Tables

List of Abbreviations

AIS	Association of Information Systems
ANN	Artificial Neural Network
API	Application Programming Interface
BISE	Business & Information Systems Engineering
DAX	German Stock Index
DSS	Decision Support System
ECIS	European Conference on Information Systems
FE	Feature Engineering
FOMC	Federal Open Market Committee
ICIS	International Conference on Information Systems
IF	Impact Factor
MKWI	Multikonferenz Wirtschaftsinformatik
MLP	Multilayer Perceptron
Nm	Newton metre
OEM	Original Equipment Manufacturer
OVB	Omitted Variable Bias
RL	Reinforcement Learning
RMSE	Root Mean Square Error
S&P500	Standard & Poor's 500
SA	Sentiment Analysis

VHB	Verband der Hochschullehrer für Betriebswirtschaft
WI	Internationale Tagung Wirtschaftsinformatik
WKWI	Wissenschaftliche Kommission für Wirtschaftsinformatik

Overview of Publications and Task Allocation

This dissertation is based on seven peer-reviewed and accepted papers published in different conference proceedings or journals like the International Conference on Information Systems 2017, European Conference on Information Systems 2017, Internationale Tagung Wirtschaftsinformatik 2017, Multikonferenz Wirtschaftsinformatik 2018 and two international journals Business & Information Systems Engineering and Decision Support Systems. In total, eight co-authors are involved in the research contributions namely Michael H. Breitner, Christian L. Dunis, Dennis Gercke, Christoph Gleue, Cornelius Köpp, Hans-Jörg von Mettenheim, Daniel Olivotti and Rouven Wiegard (alphabetical order). All topics contribute to challenges which are related to data analytics with different applications like machine learning in finance, text mining, forecasting, decision support and visualization.

Table 1 provides an overview of the publications. Three different rankings or quality indicators are presented. First, the WI-Orientierungslisten 2008 from the Wissenschaftliche Kommission für Wirtschaftsinformatik (WKWI) which provides a perspective from the German-language Wirtschaftsinformatik community. Second, the JOURQUAL3 2015 ranking from the Verband der Hochschullehrer für Betriebswirtschaft (VHB). The third key figure is the Impact Factor (IF) from Thomson Reuters.

All publications were written in collaboration with different co-authors.

The paper "Intelligent Trading of Seasonal Effects: A Decision Support Algorithm Based on Reinforcement Learning" [Eilers et al., 2014] is based on the ideas of seasonal effects on financial markets and the decision support algorithm based on reinforcement learning. Prof. Dunis contributed the idea of investigating seasonal effects. He and Prof. von Mettenheim provided many papers for the literature analysis. Filtering the resulting trading decisions with a reinforcement learning approach was my idea. The empirical analysis, the development and implementation of the reinforcement learning algorithm, the Java implementation of the agent-based

Table 1: Overview of Publications

Year	Title	Co-Authors	Conference/Journal	WKWI[a]	JQ3[b]	IF[c]	Appendix
2014	Intelligent Trading of Seasonal Effects: A Decision Support Algorithm Based on Reinforcement Learning	Dunis C. L., von Mettenheim H.-J., Breitner M. H.	Decision Support Systems	A	B	3.222	A
2017	Decision Support for the Automotive Industry: Forecasting Residual Values Using Artificial Neural Networks	Gleue C., von Mettenheim H.-J., Breitner M. H.	Wirtschaftsinformatik 2017 Proceedings	A	C	-	B
2017	A Picture Is Worth a Thousand Words: Visual Model Evaluation in Data Science Applications	Breitner M. H.	Wirtschaftsinformatik 2017 Proceedings	A	C	-	C
2017	What Does YouTube Say about Your Product? An Aspect Based Approach	Wiegard R., Gercke D.	European Conference on Information Systems 2017 Proceedings	A	B	-	D
2017	It's Not a Bug, It's a Feature: How Visual Model Evaluation Can Help to Incorporate Human Domain Knowledge in Data Science	Köpp C., Gleue C., and Breitner M. H.	International Conference on Information System 2017 Proceedings	A	A	-	E
2018	Understanding Anomalies: Visualizing Sensor Data for Condition Monitoring of Manufacturing Machines	Olivotti D.	Multikonferenz Wirtschaftsinformatik 2018 Proceedings	C	D	-	F
2018	Decision Support for the Automotive Industry: Forecasting Residual Values Using Artificial Neural Networks	Gleue C., von Mettenheim H.-J., Breitner M. H.	(forthcoming) Business & Information Systems Engineering	A	B	3.392	G

[a] Wissenschaftliche Kommission für Wirtschaftsinformatik 2008 WI-Orientierungslisten
[b] JOURQUAL3 Verband der Hochschullehrer für Betriebswirtschaft
[c] Thomson Reuters Impact Factor 2016

trading system and the backtesting evaluation was mainly done by me. All three co-authors have provided technical support. Especially Prof. von Mettenheim for the empirical part and Prof. Breitner for the training of the Artificial Neural Networks and data preprocessing.

The paper "What Does YouTube Say about Your Product? An Aspect Based Approach" [Wiegard et al., 2017] is based in the ideas of an aspect-based Sentiment Analysis and a filtering method based on Google Trends. Mr. Wiegard was responsible for the literature analysis about Sentiment Analysis. He also developed the idea for the research design. Mr. Gercke was responsible for the empirical evaluation of the results from the algorithm and human annotators. I was responsible for the literature analysis about aspect extraction algorithms and implementation of the selected algorithm. I have also done the data gathering, preprocessing and the application of the algorithm on the data as well as result interpretation.

The paper "Decision Support for the Automotive Industry: Forecasting Residual Values Using Artificial Neural Networks" [Gleue et al., 2017] published in the proceedings of the Internationale Tagung Wirtschaftsinformatik 2017 is based on an implementation of a decision support system and a forecasting application for resale prices of used cars. Mr. Gleue had the idea of developing a forecasting application for the leasing business of the large German car manufacturer we were working with. He is responsible for the motivation, the description of the data gathering process and the structure of the decision support system. I was responsible for the forecasting application and the construction and optimization of the Artificial Neural Networks. The empirical study about the time factor in this forecasting task, the benchmark with the linear model and the evaluation and interpretation of results was also done by me. Prof. von Mettenheim provided technical support for the empirical part of the paper especially the data cleansing process, the design of the decision support system and the optimization of the Artificial Neural Networks. Prof. Breitner has worked on the design of the decision support system and provided technical support in the areas data preprocessing, data splitting and optimization of the Artificial

Neural Networks. The subsequent extension of the conference paper, which was submitted to the journal Business & Information Systems Engineering [Gleue et al., 2018] contains a complete redesign of the research methodology. We developed and described a case study with three different forecasting scenarios which were comprehensively tested and benchmarked. We also conducted a study about external explanatory factors and a feature importance analysis to show which factors have the greatest impact on the forecasting accuracy. The development of the three forecasting scenarios, the implementation and evaluation of the scenarios, as well as the empirical investigation of external factors based on linear mixed-effects models was done by me. I also implemented the program code to analyze the optimized Artificial Neural Networks with respect to important features. Prof. von Mettenheim provided support for the empirical study and monitored especially the methodological approach of linear mixed-effects models. Prof. Breitner and Mr. Gleue contributed to the design of the case study and Prof. Breitner also provided technical support for the analysis of the Artificial Neural Networks and feature importance analysis.

The paper "A Picture Is Worth a Thousand Words: Visual Model Evaluation in Data Science Applications" [Eilers and Breitner, 2017] is based on a discussion of how data science models can be visualized more intuitively for people with a non-technical background. The derivation of the discussed research gap was done by me based on a literature analysis. The discussion and implementation of a first heat map-based prototype which fulfills the requirements are also done by me. Prof. Breitner contributed to the implementation of the heat map approach and to the implications for the management. The second paper on this topic "It's Not a Bug, It's a Feature: How Visual Model Evaluation Can Help to Incorporate Human Domain Knowledge in Data Science" [Eilers et al., 2017] provide a more solid derivation of the research gap, an improved mathematical implementation and a comprehensive design science oriented research approach which was evaluated by artificial examples and within a real data science project. I was responsible for the complete evaluation part and implementation of the presented idea for all provided scenarios. I also conducted a comprehensive literature review about visualization, Feature Engineering and model evaluation. Based on that, I have derived the research gap. I was also responsible for the new implementation of the proposed technique. Dr. Köpp provided support for the mathematical implementation of the heat map-based approach and provided ideas for further improvements of the argumentation. Mr. Gleue was responsible for providing the data and evaluation scenarios and Prof. Breitner provided technical support for the optimization of the Artificial Neural Network optimization in the empirical evaluation part. Prof. Breitner was also responsible for the discussion of the paper and the general support in questions about data visualization and communication in teams. The third paper on this topic "Understanding Anomalies: Visualizing Sensor Data for Condition Monitoring of Manufacturing Machines" [Olivotti and Eilers, 2018] show the implementation of the proposed visualization technique for making sense of large amounts of sensor data in the context of condition monitoring. Mr. Olivotti was responsible for the data collection process and the design/construction of the demonstration machine. He also provided the literature review about condition monitoring techniques. I was responsible for the anomaly detection part which was realized based on Artificial Neural Networks. I also implemented the visualization for the given task. Together we interpreted and discussed the results.

1 Introduction and Motivation

The most perfect knowledge is that which is both adequate and intuitive.

Die vollkommenste Erkenntnis ist diejenige, die zugleich adäquat und intuitiv ist.

Gottfried Wilhelm Leibniz (1684)

Data Science, Big Data Analytics, Business Intelligence, Visual Analytics and many more buzzwords have emerged over the last few years. The leading journals in the field of information systems research have adopted this trend and publish an increasing number of papers and special issues in this area [Chen et al., 2012]. [Agarwal and Dhar, 2014] have published an editorial in the journal Information Systems Research with the title "Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research". In this article they try to define the role of information systems research as an independent research discipline within the broader context of the buzzwords mentioned in the title. One important question they try to answer is "What is our [as information systems researchers] competitive advantage?". One answer they provide is "[. . .] one could reasonably assert that the IS discipline has the longest history of conducting research at the nexus of computing technology and data in business and society". The authors also state that "[. . .] among all functional areas of business, IS researchers perhaps have the broadest perspective on the enterprise as a whole, and how different pieces fit together. This focus creates a tighter linkage between data and business models: we care deeply about business transformation and value creation through data, and less for algorithms or frameworks without a linkage to business value. Our research has been alternately praised and criticized for being too crossdisciplinary, but we believe this is strength and not a weakness in today's data rich environment. [. . .] This cross- and transdisciplinary nature of IS research to date positions us uniquely to exploit the big data opportunity". These ideas about the role of information systems research are the motivation for the main part of this dissertation.

The initial situation was a collaborative data analytics project we have worked on together with a large German car manufacturer. The challenge was to analyze and forecast the resale price of used cars based on certain independent variables like the age of the car and the mileage. While the analytics component of this project is in and of itself an interesting research topic, our observation show that the collaboration between experts in data analytics and domain experts like managers and decision makers was not optimal. But both groups are stakeholders in this project and depend on each others' knowledge and expertise. Experts in analyzing data require the experience and domain knowledge from the decision makers and in turn decision makers require properly designed DSS which they can trust. Based on the ideas from [Agarwal and Dhar, 2014] about the role of information systems research, we have defined our role in this project as bridge builders between both worlds. The goal was to use our unique advantage of crossdisciplinarity and broad perspective to better integrate human domain knowledge in the data science project. The design science-oriented research approach, lessons learned and a critical discussion of our work are presented in Section 5.

Overall, this cumulative dissertation is based on four parts which are all related to data analytics. First, I discuss our contributions to machine learning in finance (Section 2). The idea is to conduct an empirical study about the existence of calendar anomalies and empirical regularities on financial markets. To exploit potential anomalies, we develop a self-learning and adapting decision support algorithm based on RL which analysis the market based different input factors and then decides which anomalies at which point in time should be traded and which not. The next part (Section 3) is based on a study about analyzing written customer reviews. The idea is here to use text mining algorithms to extract important product aspects from customer reviews. In addition, this study is the first which applies aspect-based sentiment algorithms not only to conventional review data but also to social media data like YouTube. We document our aspect extraction results for three different smartphones and compare the precision on Amazon review data and YouTube comments. The third part (Section 4) discusses the resale price forecast project which is already introduced above. The focus of this part is the data analytics application and forecasting approach. This part lays the foundation for the fourth and final part (Section 5) of the dissertation which discusses the idea of how to better incorporate human domain knowledge in data science. Section 6 discusses the overall work and highlights several further research questions, while Section 7 provides a short conclusion. Figure 6 illustrates the overall structure of this dissertation and lists the journals and conference proceedings where the work is published.

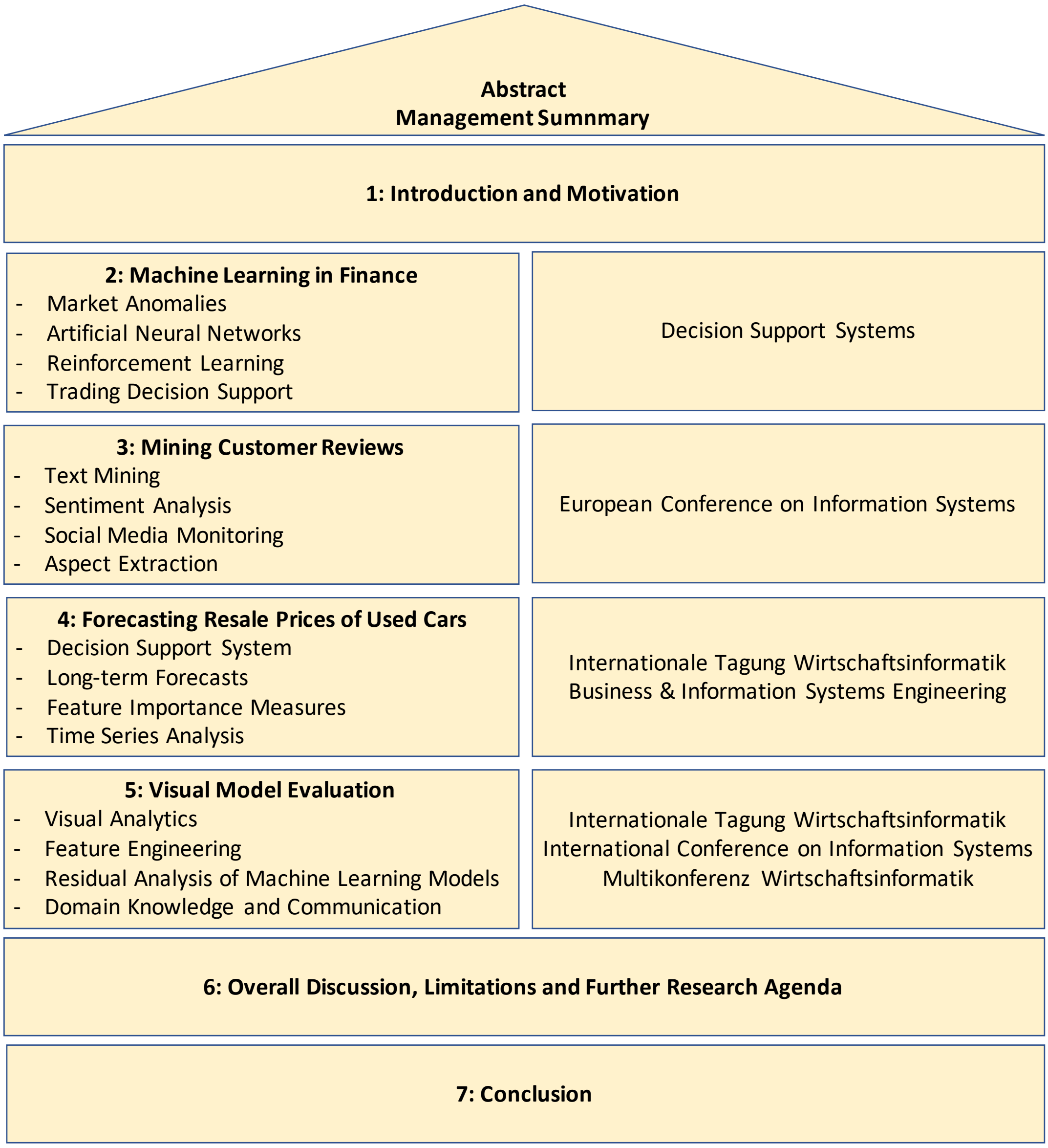


Figure 6: Structure of this Dissertation

This work may perhaps be seen as the summary of seven peer-reviewed and accepted publications. But my intention is rather to provide an independent, comprehensive and especially critical review of my own work regarding validity, methodology, relevance and rigorousness. I would like to achieve that this work has an independent utility and raison d'être. Therefore, my argumentation in each part is based on the current literature, reviewer opinions and common sense.

2 Machine Learning in Finance

In this section, the research paper with the title "Intelligent Trading of Seasonal Effects: A Decision Support Algorithm Based on Reinforcement Learning" is discussed [Eilers et al., 2014]. The paper is a collaboration of four authors, Christian L. Dunis, Hans-Jörg von Mettenheim, Michael H. Breitner and Dennis Eilers. A pilot study was first presented and discussed at the Forecasting Financial Markets Conference 2013 in Hannover. The resulting paper was submitted to the journal Decision Support Systems in 2014 and accepted after two revisions. Decision Support Systems is published by Elsevier. The Thomson Reuters Impact Factor in 2016 was 3.222. In the VHB-JOURQUAL3 ranking for Wirtschaftsinformatik (Information Systems Research) the journal is rated with a "B", while in the WI-Orientierungsliste of the WKWI it is rated with an "A" (highest possible score). In the VHB-JOURQUAL3 list for Wirtschaftsinformatik, Decision Support Systems receives the third largest number of ratings from the community (165) after the journals Business & Information Systems Engineering (168) and Management Information Systems Quarterly (201). The journal describes the requirements for submitted papers as follows:

"The common thread of articles published in Decision Support Systems is their relevance to theoretical and technical issues in the support of enhanced decision making. The areas addressed may include foundations, functionality, interfaces, implementation, impacts, and evaluation of decision support systems."[1]

2.1 Motivation and Theory

October. This is one of the peculiarly dangerous months to speculate in stocks in. The others are July, January, September, April, November, May, March, June, December, August, and February.

Mark Twain (1894)

[1] `https://www.journals.elsevier.com/decision-support-systems/` accessed: December 21. 2017

There exists a long and ongoing debate in the finance literature, how efficient financial markets are and to which extend an exploitation of certain anomalies by a profitable trading system is possible. One of the most influential theories is presented by [Fama, 1970]. He describes three types of information efficiency on financial markets, the weak, semi strong and strong form. If the weak form holds it would be impossible to generate excess returns based on an analysis of past prices since all information about past prices are already reflected in the current price of the market. Excess returns are here defined as the difference between the actual returns and the expected returns at the respective level of risk. If the semi strong form of information efficiency holds it would be impossible to generate excess returns by analyzing all publicly available information because they are also completely reflected in the current price and if the strong form holds even insider information are reflected in the current price. Trading strategies in general mostly assume that even the weak form of information efficiency does not always hold and chart analysis, technical indicators and support/resistance levels provide valuable information to generate excess returns in certain situations. There exist a bunch of literature which tries to exploit these anomalies by certain strategies [Baetje, 2018] or machine learning algorithms like ANNs [Dunis and Huang, 2002] [Sermpinis et al., 2012]. Empirical studies also show that there exist seasonal effects in different markets. Popular seasonal effect which are documented in the literature are the upward bias at the turn of the month [Ariel, 1987], during exchange holidays [French, 1980] and the pre-FOMC (Federal Open Market Committee) announcement drift [Lucca and Moench, 2015]. Several other potential anomalies are documented in the literature [Cadsby and Ratner, 1992], [Dickinson and Peterson, 1995], [Jaffe and Westerfield, 1989], [Atanasova and Hudson, 2010], [Lakonishok and Smidt, 1988], [Ogden, 1990], [Rozeff and Kinney, 1976]. [Hansen et al., 2005] provides a comprehensive overview. But again, the empirical evidence is ambiguous [Sullivan et al., 2001]. It also depends on the time horizon and the observational periods. We are interested in the question if it is possible to exploit these documented anomalies by implementing a profitable trading strategy. In a first step, we therefore analyze the German Stock Index (DAX) and the Standard & Poor's 500 (S&P500) to document possible seasonal effects. For our analysis, we focus on the three above mentioned anomalies turn-of-the-month, exchange holidays and the pre-FOMC announcement drift. Our results are in line with the previous literature and the overall effects of the investigated anomalies on the two major indices are rather small. But as documented in the literature we found strong patterns during certain times of the year, which also change over time [Eilers et al., 2014]. This might lead to the conclusion that some empirical regularities and seasonal effects are not persistent and change over time. A static trading strategy is therefore inappropriate to exploit these phenomena in the future. Therefore, the main research question is as follows.

RQ: If seasonal effect exits but are not persistent, is it possible to automatically identify situations in the market where the empirical regularities hold?

To answer the question our idea is to develop a self-learning DSS which is able to assess the current market situation and automatically filters the decisions of the static trading strategy. Such a self-learning filter application should be able to outperform the static seasonal strategies which are described in the literature. To realize this, we use the idea of RL to improve the reward to risk ratios of a baseline seasonality strategy. In the following, the ideas and methods are described. Afterwards the results are summarized and discussed.

2.2 Methods and Ideas

2.2.1 Reinforcement Learning

A man who has committed a mistake and doesn't correct it, is committing another mistake.

Confucius

The concept of RL is based on the idea of learning from reward and punishment.

Human action is the result of an analysis of the current environment and an assessment about the consequences of a certain behavior in this environment. Humans generally try to improve their behavior by analyzing the feedback of the environment on a certain action. The goal of a homo oeconomicus is to perform the actions which maximize the own reward over time. An easy example is a human trader. Based on experience from the past about gains and losses, the trader tries to find the right investment strategy for the current situation which most likely maximizes his/her reword in the future. In this setting, the feedback from the environment are gains and losses of a certain strategy. Trading rules which continuously produce losses will be adjusted or removed by the trader. The larger the profit, the better the decision in the current market situation. Therefore, the feedback on financial markets is easy to interpret and can directly be used as a basis for a reward and punishment system of a self-learning system. In the long run the goal is that the system learns which action in a current market situation would lead to the largest reward (profit). Actions with bad experiences should be avoided. Figure 7 illustrates the idea of RL graphically.

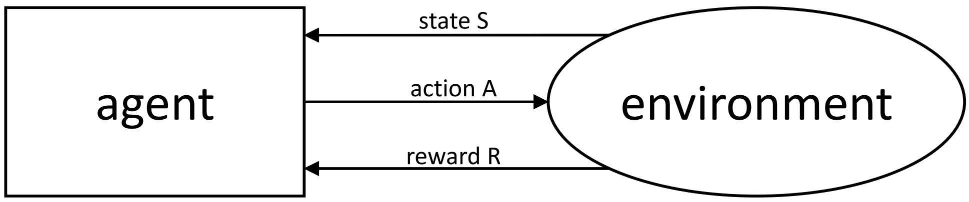


Figure 7: Concept of Reinforcement Learning

RL in the general case can be formally described as follows. An artificial agent executes an action (a_t) in the current environment which is in state (s_t) which leads the agent to the next state (s_{t+1}).

$$State\ s_t \in S \qquad s_t \xrightarrow{a_t} s_{t+1} \tag{1}$$

to achieve this the δ function is used as transition function.

$$s_{t+1} = \delta(s_t, a_t) \tag{2}$$

Since RL is based on the principle of reward and punishment, this can be modeled by a function which depends on the current state and the performed action.

$$r_t = r(s_t, a_t) \tag{3}$$

In a trading application the following applies for the reward:

$$r_t > 0 : positive\ reward \tag{4}$$

$$r_t = 0 : no\ reward \tag{5}$$

$$r_t < 0 : negative\ reward \tag{6}$$

Since in general the long term reward should be maximized the agent has to find a suitable policy which projects the appropriate actions to each state. To find this policy is the learning task of the agent.

$$policy\ \pi : S \rightarrow A \tag{7}$$

This results in a maximization problem. To find the best policy, the weighted sum of the single rewards of a policy must be maximal. A discount factor γ is used to weight future rewards. It is also important to ensure that the sum converges. The sum converges under the condition that r is limited ($|r| \leq$ a constant B where $B < \infty$) and $0 < \gamma < 1$. A policy π^* is optimal if the following applies to all states:

$$V^{\pi^*}(s) \geq V^{\pi}(s) \tag{8}$$

The value function:

$$V^{\pi}(s_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + ... = \sum_{i=0}^{\infty} \gamma^i r_{t+i} \tag{9}$$

Convergence:

$$|\sum_{t=0}^{\infty}\gamma^i r_t| = \sum_{t=0}^{\infty}\gamma^i |r_t| < \sum_{t=0}^{\infty}\gamma^i B = B\sum_{t=0}^{\infty}\gamma^i = B\frac{1}{1-\gamma} \tag{10}$$

RL in general tries to maximize the reward over time. This is because in applications for example in the field of robotics, the actions are not independent from each other. An action leads to a new state transition and the new state needs to be evaluated again. In a trading application, only the immediate reward should be maximized. This eliminates the typical credit assignment problem which frequently occur if certain actions have an impact on future states like playing a chess game. If we assume, that the actions (trades) have no or only negligible influences on the resulting state of the market the situation is much easier to interpret. Only a method for selecting the best action in a certain situation is needed. The following simple relationship remains:

$$V^{\pi}(s_t, a_t) = r_t = r(s_t, a_t) \tag{11}$$

In this case, only the immediate reward needs to be maximized. An optimal decision policy is therefore given by

$$\pi^*(s_t) = \operatorname{argmax}_{a_t} r(s_t, a_t) \tag{12}$$

while the state s_t is defined by the market situation and a_t represents a combination of different order parameters which needs to be optimized for a one-step decision process.

To realize this, it is necessary to define a certain scope of possible actions for the agent. These actions could be the tradable product, the position (long/short) or a certain lever. The agent then starts with no prior experience and randomly selects a possible action to get the feedback from the market. The corresponding profits and losses are now associated with a so-called State-Action-Pair. In this case the State-Action-Pair is the current state of the market and the selected action which was performed. The association between profits/losses and a State-Action-

Pair represents an input-output relationship. The condition of the market can be described by historical price data or technical indicators. This leads to an infinite number of possible states. To learn certain patterns in the data even with an infinite number of possible states, a machine learning technique like ANNs can be applied.

2.2.2 Artificial Neural Networks

ANNs are applied for forecasting in finance for many years [Leigh et al., 2002], [Lam, 2004]. ANNs can be viewed as a method for non-linear function approximation. For the presented application feed forward networks are used. This type of networks consists of several layers of neurons and is therefore assigned to the class of Multilayer Perceptrons (MLP). A first layer, the input layer defines the independent variables which describes a certain phenomenon under investigation. The neurons are fully connected by weights with the next layer of neurons, the hidden layer. Fully connected means that each input neuron is connected to each neuron of the hidden layer. The hidden layer is again fully connected to the next layer which is either the output layer or another hidden layer by weights. The number of hidden layers and neurons inside of each hidden layer is unlimited in theory and the only restriction is the computing power which is necessary to optimize the network. A simple three-layer feed forward ANN is defined by

$$h_\theta(X) = \theta_2 \tanh(\theta_1 X) \tag{13}$$

while $h\theta(X)$ represents the estimated dependent variable or the hypotheses about the real world. The hidden neurons transform the weighted sum of their input by a nonlinear function which enables the ANN to learn complex nonlinear dependencies in the data. In this example, we use the hyperbolic tangent:

$$tanh(X) = 1 - \frac{2}{e^{2X} + 1} \tag{14}$$

A graphical illustration of an ANN is presented in Figure 8.

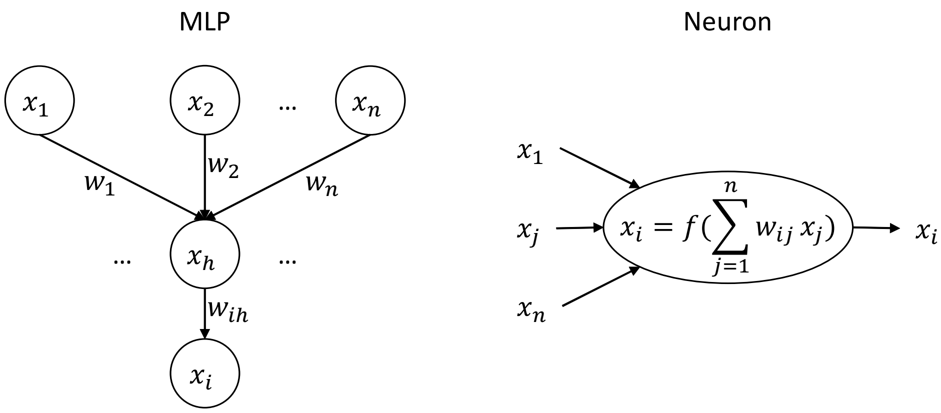


Figure 8: Artificial Neural Network Structure

It is also possible to directly connect the input layer with the output layer by so called shortcuts. This can help the ANN to focus on learning the non-linear dependencies within the deep structure of hidden neurons, since the linear dependencies are learned by the shortcut connections.

To perform a function approximation based on given Input-Output patterns, the ANN is trained with training patterns (input x_k, output y_k, represent training pattern k). The training is an iterative process of updating the randomly initialized weights inside the network to minimize the difference between the actual realized values and the output values of the network approximation. A general problem which can occur with this procedure is overfitting. If the data contains noise, the function approximation has the goal to generalize from the given data. This means that measurement errors, or simply random noise should have no influence on the network structure. However, by minimizing the difference between the real value and the estimation, the network probably learns these noisy patterns by hard which on the one hand leads to a better performance on the training data set but deteriorate the generalization capabilities of the network which is essential in real world applications. There exist several methods to mitigate the problem. Dropout [Srivastava et al., 2014] for example excludes a certain percentage of weights in each training iteration which reduces the risk that certain weights become experts for only one specific type of data (for example one particular measurement error). This strengthens the generalization capabilities of the whole network. Another method is the early stopping approach. To perform early stopping the training data is first split up into an actual training data (I_t) set to perform the iterative weight update process and a validation data set (I_v) which is used to continuously monitor the estimated out of sample error. The validation data is not part of the weight update process and therefore serves as an approximation for the error the network will produce on new and previously unseen data. If the error began to increase on the validation data set after a certain amount of iterations, the training is stopped, even if the error on the training data could be further decreased. This method stops the training before the network starts to overfit which is indicated by an increase of the error on an out of sample data set. The approximation capabilities of a trained network is evaluated by calculating the training and validation error functions:

$$\varepsilon_t = \frac{1}{2} \sum_{k \in I_t} (h_\theta(x_k) - y_k)^2 \tag{15}$$

$$\varepsilon_v = \frac{1}{2} \sum_{k \in I_v} (h_\theta(x_k) - y_k)^2 \tag{16}$$

The next section describes the implemented trading system which combines RL with ANNs.

2.3 Agent Based Trading System

The basic idea is to use a seasonal effects trading strategy as a baseline and then filter the decisions by a self-learning system. In this case we use the three previously mentioned effect, turn-of-the-month, exchange holiday and pre-FOMC announcement drift. In the literature, an upward bias is documented for all the three event categories. Therefore, a simple baseline strategy is to open a long position at the close price of the day before the event and close the position to the close price of the next trading day after the event. For the decision support algorithm, we implement an artificial agent which represents a trader on the markets. The agent starts with no prior knowledge or experience. Its task is to decide whether a trade of the static seasonal effects strategy should be executed, not executed, leveraged, held for a further day or even reversed such that the exact opposite of the original strategy is executed. The agent is allowed to automatically choose between these different options based on the state of the current market. Figure 9 illustrates the possible decisions.

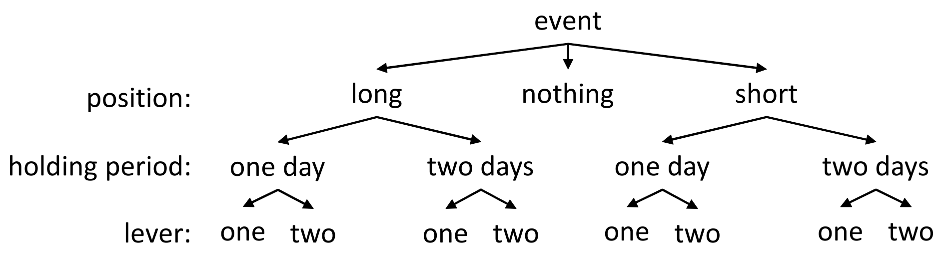


Figure 9: Options for the Agent

The state of the market is described by certain indicators and historic price data. Table 2 provides an overview of the variables which describe the current market situation, the actions and the reward. These variables serve as input and output for the agent.

Table 2: ANN Layout

	description	**neurons**
state	Event: turn-of-the month, FOMC or exchange holidays	1
	OHLC of the present day	4
	Close-values from the past three days	3
	SMA of the last five close-values	1
	RSI of the last six close-values	1
	the number for the current month	1
action	position	1
	holding period	1
	leverage	1
output	immediate reward (the profit of each order)	1

The agent begins with a random guess. The action is performed on the market and the respective reward (profit/loss) of the trade is calculated after the position is closed. The resulting profit/loss serves as a label for the given state of the market and the performed action. Together, these information form the first training data set for an ANN which learns the dependencies between the State-Action-Pair as the inputs and the reward as the output (in this case the return of the trade). The network is then trained for the first time with this new data set. For the next decision, the agent observes the current state of the market, and then calculates the estimated output of the network for each possible action. The actions are represented as additional dummy variables with respective input neurons during the training process. The agent then chooses the action which promises the highest return. After the trade is closed the next training data set is generated. A graphical illustration of this process is presented in Figure 10.

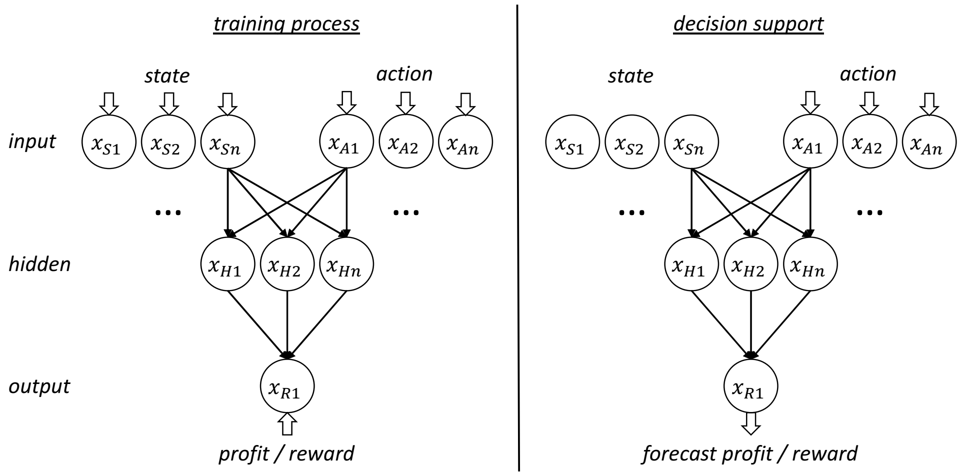


Figure 10: Training and Decision Process

RL applications always face the trade off between exploration and exploitation. To find better solutions and to generate more diverse training samples, the agent should act against its own estimations to explore possibly better regions of the solution space. This is called exploration. It is necessary to adapt new developments over time but it decreases the immediate performance since most of the time these actions lead to bad decisions in the short run. Following the calculated recommendations is called exploitation. To balance the trade off between exploration and exploitation is one of the most challenging parts in RL. In our application, we develop a dynamic approach. The probability that a certain trade is executed based on exploration decreases with an increasing number of available training patterns [Eilers et al., 2014]. The results of this implementation are discussed in the next section.

2.4 Conclusion, Discussion and Outlook

To illustrate the described decision support algorithm we use a backtesting approach to simulate the performance of the system during a trading period of 2000 to 2012 for the DAX and the S&P500. The results outperform the benchmarks (buy and hold, and a static seasonality strategy). Figure 11 provides an intuition of how the agent operates.

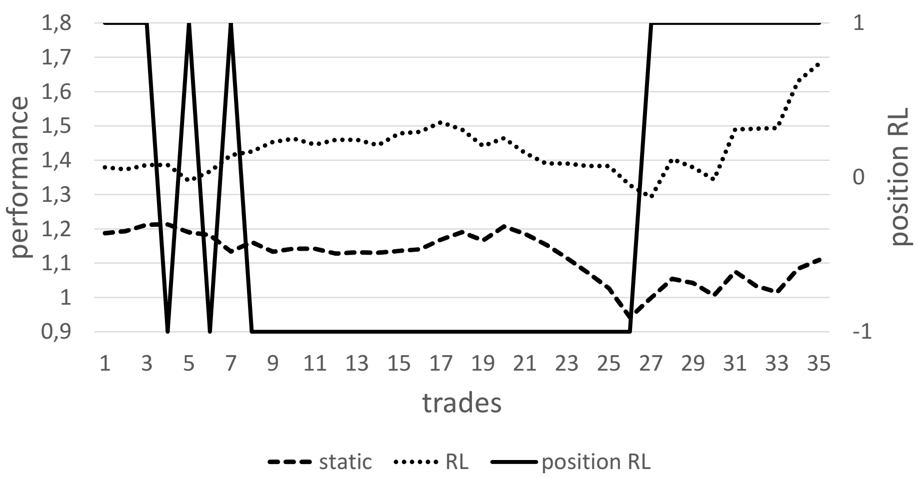


Figure 11: Comparison of Reinforcement Learning with Static Strategy

The solid line shows if the agent approves the static seasonal effect strategy (indicated by a 1 for a long position) or if the agent chooses the exact opposite in certain phases of the market. One can see that after a certain time of exploiting the new market situation the agent chooses to switch the rules of the seasonality strategy automatically. The moment the strategy produces positive returns again, the agent switches back to the original rules.

The results show that it is indeed possible to improve a basic seasonal effects trading strategy by a self-learning agent which automatically identifies market situations where the strategy is appropriate. Despite the promising results there is much room for improvements. Currently we have only implemented a very limited action space. Increasing the possible options for the agent might improve the results but this would need a higher amount of trades for exploration. Therefore, a study on high frequency data would be interesting as well. Several other possibilities for improving the results need to be discussed, for example the trade off between exploration and exploitation. All in all it would be interesting to test the presented approach in other domains which are not primarily related to finance, since the technique is generally applicable to any kind of self-learning decision support problem.

2.5 Reviewer Opinions, Contribution and Critical Comment

This section serves as a critical reflection of the work. One important criterion is the possible impact on the research field and practice. The following quote is based on the comment of the first anonymous reviewer if the paper after the initial submission:

"The paper is interesting and well written. The introduction of a filtering technique in calendar trading strategies is something that would be of interest not only to researchers but also to practitioners. Calendar trading strategies are still heavily applied in the fund management industry. Thus, I believe that this study will have a significant impact on the field. In addition to the above, it is worth noting that the paper is easily replicable."

The research aims to focus on an important practical and theoretical question. The reviewer therefore points out that the paper can have a significant impact on the field. For example, fund managers can use the ideas to support their investment decisions. The developed solution provides empirical evidence as well as practical implementation guidelines to reproduce the results. Therefore, the study can be considered as relevant and rigorously conducted.

However, the question arises if trading systems in general are appropriate research topics. In a simulated back testing environment, there are a lot of assumptions like immediate order execution which are not always fulfilled in real world applications. Another question is if the research really addresses a socially important problem. Even if some fund managers can benefit from the generated insights, the overall impact could be considered as rather small. Nevertheless, in hindsight, I would like to summarize the overall contribution of the article as follows:

Contribution: We provide further empirical evidence for the existence of certain market anomalies like calendar effects and propose a self-learning decision support algorithm which allows to better exploit these phenomena.

3 Mining Customer Reviews

In this section the research paper with the title "What does YouTube say about your Product? An Aspect based Approach" is discussed [Wiegard et al., 2017]. Three authors have contributed to this work: Rouven Wiegard, Dennis Gercke and Dennis Eilers. The paper was submitted and accepted after one revision for publication at the European Conference on Information Systems (ECIS) 2017 in Guimaraes, Portugal. In the VHB-JOURQUAL3 ranking, the proceedings of the peer-reviewed conference are rated with a "B". In the WI-Orientierungsliste of the WKWI it is rated with an "A" (highest possible score). The ECIS is an official Conference of the Association of Information Systems (AIS) which describes the ECIS as one of their leading Conferences. The theme of the ECIS 2017 was "Information Systems for a smart, sustainable and inclusive world".

3.1 Motivation and Research Background

The trend towards electronic commerce provide a bunch of new opportunities for customers as well as companies. Especially the huge amount of data generated by customers can be used to better understand purchasing decisions and success or failure of a product. In this study, we focus on analyzing customer reviews to automatically extract opinions about certain aspects of a product. The research field of automatically analyzing customer opinions (also often mentioned as SA) is divided into three different subcategories.

- Document based SA: An analysis on document level provides the highest abstraction. The result is an assessment of the whole review if the opinion expressed by the customer is rather positive or negative. [Pang et al., 2002] [Turney, 2002]
- Sentence based SA: The sentence level provides more individual information about the opinion of the customer since the analysis is carried out on each

sentence in the review. The functionality is basically the same except from the fact that each sentence is treated as a whole document. [Wilson et al., 2005]

- Aspect based SA: Analyzing reviews on aspect level extracts the most valuable information from a text since the opinion of the customer is analyzed with respect to each mentioned product aspect. Therefore, it is possible to distinguish the overall opinion from the assessment of certain aspects. For example, if a customer is satisfied with a product but criticizes the camera of a smartphone as the only weak aspect of the product, an analysis on aspect level is able to detect these patterns [Hu and Liu, 2004]. Aspect based SA again requires three subtasks. (1) Aspect extraction which identifies the relevant terms in a text which describes the investigated product. (2) Aspect evaluation which afterwards analyze the opinion with respect to each of the mentioned aspects. (3) Summarization which collects all the analyzed opinions to form groups of aspects or to assess the overall opinion about the product.

[Pang and Lee, 2008] provide a comprehensive overview of the field of SA. Literature reviews are provided by [Liu, 2012] and [Liu and Zhang, 2012]. Especially [Tsytsarau and Palpanas, 2012] derive further research opportunities in this field.

But users not only use online shop websites to express their opinions, feelings and thoughts about a product. Especially social media platforms more and more become a platform for this purpose. Research fields like social media monitoring arise to address these developments. It can be shown that companies which use social media monitoring can outperform their competitors [Zikopoulos et al., 2013]. Social media comments offer the advantage that they provide an immediate opinion expression of the user in real time, while conventional reviews often have a certain time delay. The time factor can be critical for companies in fast paced times which is the reason why more and more companies invest in better solutions for analyzing the user opinions in social media. But while the automated analysis of review data is in and of itself a very difficult task, which requires in depth knowledge about text mining and natural language processing, the content posted on social media websites is even more challenging to analyze due to the more informal and often product unrelated comments. Therefore, the main research question of this study is:

RQ: Is it possible to perform a meaningful aspect based SA on unstructured social media data and what are the performance differences compared to conventional data sources like Amazon review data?

To answer the question, in this study we focus on the first task of an aspect based SA which is aspect extraction. Aspect extraction forms the foundation of the subsequent tasks like aspect evaluation. The goal is to find out if this first task of the process provides meaningful results by using social media postings as a basis for the analysis. Therefore, we carry out an exploratory analysis on different social media platforms like Facebook, Twitter and YouTube to analyze their suitability for an aspect based SA task. As an example, we choose three smartphones for our investigations. The sampling approach, data and algorithms are summarized in the next section.

3.2 Sampling Approach, Data and Algorithm

The first step of the analysis is the identification of a suitable data source. In this case we focus on three leading social media platforms, Facebook, Twitter and YouTube. All three provide Application Programming Interfaces (APIs) to access their content. The question arises, how to sample posts and comments which are directly related to a certain product. For Facebook and Twitter, we choose to filter all post/tweets which mention the concrete name of the investigated product. A manual screening of the results shows that especially tweets have a very high number of advertising content. In addition, tweets but also Facebook posts are very short which results in a more general opinion expression about a product. An example: "I love my new iPhone".

Specific aspects are rarely mentioned which is the basic requirement for a detailed aspect based SA. YouTube, however, is different to Facebook and Twitter. First, it is not possible to extract comments directly from the whole platform by searching for specific key words. Therefore, in a first place, relevant videos need to be identified. For our purposes, we choose videos which explicitly name the product in the title. These videos are most often reviews of product testers who include the whole name of the product and the respective version directly in the title to generate a high number of views. The advantage in this case is that the comments of these videos are most often also related to the presented product even if the comments themselves do not name the product again. Typical patterns include for example: "The camera is great!"

The comments of review or unboxing videos also often provide a platform for discussions between the users about certain aspect. For example, question and answer patterns can be observed where users ask the community about their opinion. Another interesting observation is that the video content provides the opportunity to experience the product even without owning the product. This fact results in comments which are solely based on the product presentation. This leads to the problem that the comments can be biased by the opinion of the content creator, but it also provides the opportunity to discuss aspects of a new product which is not even available on the market if the video is a preview. This can be an important indicator for a company to receive an immediate feedback even before the release. Therefore, YouTube offers very unique properties as a data source for social media monitoring and especially for an aspect based SA on social media data.

We therefore choose YouTube comments for our analyzes and implement the following sampling approach:

1. We select three exemplary smartphones: LG G5, Samsung Galaxy S7 and the Apple iPhone 6S to illustrate the approach. The products are well known and therefore many comments and reviews are available.

2. For each of the products, we use the YouTube API to extract the top 50 English-language videos (ranked by YouTube) which contain the exact name

of the product.

3. For the LG G5, the 1215 latest comments and responses are extracted from the selected videos. For the Samsung Galaxy S7 1281 comments are extracted and for the Apple iPhone 6S 1135 comments.

4. From the comments, the text corpus is constructed for each product which is used for the subsequent analyses.

As a benchmark, the 250 most useful Amazon reviews (sorted by Amazon) are selected for all three smartphones. To perform the aspect extraction task, we use a state of the art algorithm proposed by [Eirinaki et al., 2012]. We adjust the algorithm for our purposes and implement the solution in the programming language R (for more details see page 5 of [Wiegard et al., 2017]). The following section summarizes the results of the aspect extraction performance for the two data sources.

3.3 Summary of results

As a quality criterion, we define the precision of the algorithm as the number of relevant aspects in the top-N retrieved aspects. The assessment is based on two independent human annotators. Table 3 presents the results for Amazon reviews and Table 4 for YouTube comments.

Table 3: Precision Amazon Reviews

top-N retrieved	iPhone 6S	Samsung S7	LG G5	row average
10	0.60	0.50	0.50	0.53
20	0.55	0.40	0.50	0.48
30	0.50	0.40	0.40	0.43
40	0.50	0.35	0.38	0.41
50	0.46	0.36	0.38	0.40
column average	0.52	0.40	0.43	0.45

The overall results are rather poor and rarely exceed a precision of 50% which means that often more than half of the extracted aspects are unrelated to the product. In fact the best results can be achieved for the LG G5 with YouTube comments. An exploratory analysis reveal that YouTube comments for the LG G5 contain on

Table 4: Precision YouTube Comments

top-N retrieved	iPhone 6S	Samsung S7	LG G5	row average
10	0.50	0.50	0.60	0.53
20	0.40	0.40	0.50	0.43
30	0.43	0.40	0.50	0.44
40	0.35	0.33	0.48	0.38
50	0.28	0.34	0.42	0.35
column average	0.39	0.39	0.50	0.43

average 2.02 sentences with 11.12 words per sentence. For the Samsung Galaxy S7 (1.75 sentences, 9.12 words) and especially for the Apple iPhone 6S (1.59 sentences, 8.89 words) these values are lower. The comments of the LG G5 also contain the lowest number of swearwords (0.99%) compared to the Apple iPhone 6S (1.34%) and the Samsung Galaxy S7 (1.21%). These indicators can be a hint that the discussion in the comments about the LG G5 contain more serious discussions which improves the quality of the results. To filter the relevant comments and aspect are therefore a major challenge for the research community.

In the paper, we therefore also propose a new approach to filter the extracted aspects. We use the information provided by Google Trends to sort the potential aspects of a product based in their search volume in conjunction with the product name. The assumption is that the search volume for a product in conjunction with a relevant aspect will be higher than the search volume for irelavent aspect. Figure 12 illustrates the approach with the graphical user interface from Google Trends. For this research purpose we have used an R script to automate the download and comparison of the values (for more details about the implementation of the filtering algorithm see [Wiegard et al., 2017] page 8).

Tables 5 and 6 present the new and improved precision values after the filtering procedure for Amazon reviews and YouTube comments respectively.

3.4 Conculsion, Discussion and Outlook

The paper aims to answer the question if it possible to perform a meaningful aspect based SA on unstructured social media data and what are the performance differ-

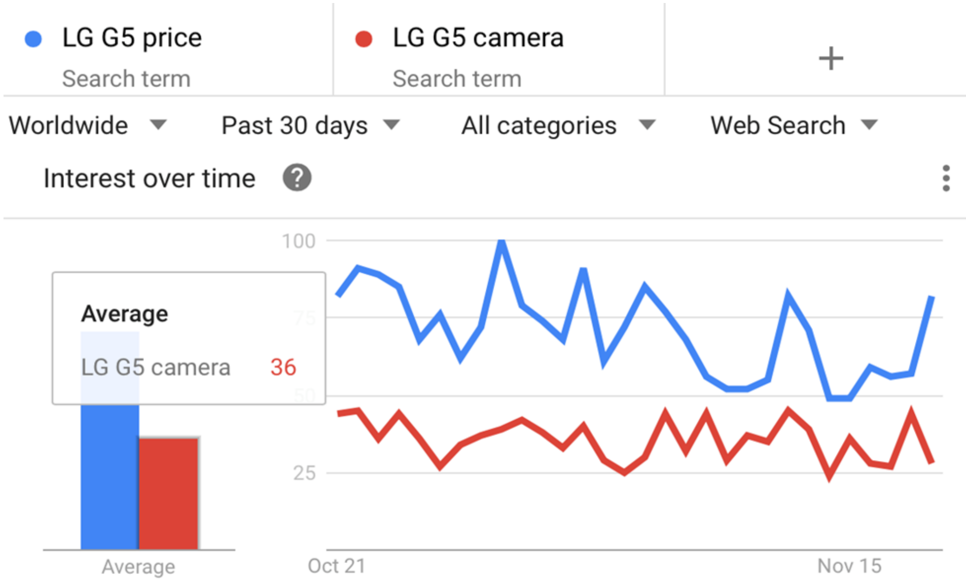


Figure 12: Aspect Filtering with Google Trends

Table 5: Precision Amazon Reviews after Filtering

top-N retrieved	iPhone 6S	Samsung S7	LG G5	row average
10	0.60	0.70	0.80	0.70
20	0.60	0.60	0.65	0.62
30	0.57	0.57	0.57	0.57
40	0.55	0.48	0.53	0.52
50	0.52	0.46	0.54	0.51
column average	0.57	0.56	0.62	0.58

Table 6: Precision YouTube Comments after Filtering

top-N retrieved	iPhone 6S	Samsung S7	LG G5	row average
10	0.80	0.60	0.80	0.73
20	0.60	0.65	0.70	0.65
30	0.53	0.53	0.63	0.57
40	0.48	0.43	0.50	0.47
50	0.40	0.40	0.46	0.42
column average	0.56	0.52	0.62	0.57

ences compared to classical data sources like Amazon review data. We show that Facebook and Twitter are hardly suitable for such analyses due to their high advertisement content. YouTube however have different characteristics, which makes it more appropriate for an efficient filtering and subsequent analysis of useful comments. Therefore, we use YouTube comments and compare the aspect extraction performance of a state-of-the-art algorithm with the results based on conventional Amazon reviews. We can show that the precision values only slightly differ between the two data sources. Hence, YouTube comments can indeed be seen as a suitable data source for aspect based SA or at least for the first task, the extraction of relevant aspects.

However, the study is limited due to the fact, that we only use one algorithm for aspect extraction. Further research must show if the results hold for different approaches like more advances machine learning algorithms. Since the submission of the paper this section is based on, many interesting papers have gained greater interest in the research community. One example is [Poria et al., 2016]. The paper presents the first approach for aspect extraction which is based on convolutional neural networks. The athors further extend their approaches and present in their highly topical paper ideas for combining visual, audio and textual modalities [Poria et al., 2017]. So far, the research field of SA has also lacked a systematic literature review. Previous works provide only simple reviews without a clear methodological approach. This gap has been closed in the meantime by [Qazi et al., 2017].

A further limitation is that the overall results of the proposed approach are rather poor. Developing aspect based SA approaches is therefore still a major challenge and current solutions are by far not suitable to be applied in practice. We propose a new approach to filter the extracted aspects, which is able to improve the results but there are many open questions and technological boundaries to be addressed in further research. Despite the improvement of learning algorithms and data analysis technologies, further research questions include:

- How can sarcasm, fake reviews or product unrelated comments be efficiently filtered before the analysis?

- How can the aspect extraction and evaluation be individualized for each customer? For example, small/large smartphones are not necessarily positive/negative.
- How can implicit aspects which are not directly mentioned like the weight of a product be addressed more efficiently?
- Are there reginal differences which need to be taken into account?
- Which impact does the new phenomenon of influencer marketing on social media channels have?
- How to deal with products for which there exist only little information?

3.5 Reviewer Opinions, Contribution and Critical Comment

In this section I critically reflect the contribution and real impact of the presented work together with statements from reviewers of the underlying paper. The first of the anonymous reviewers' states that

"Overall, the whole idea is applied on a promising area with potential high impact. The paper is well-written and easy to read. It is well-organized and each section is well developed step by step. Also, the sections regarding the core analysis are very strongly written".

The second anonymous reviewer summarizes his opinion as follows:

"I would argue for accepting this submission. Because of the influential topic and the idea to filter derived product aspects through a different source of knowledge (Google trends)."

The reviews show, that the topic is indeed relevant and the results might have a real impact on the research filed and in practice. It is also mentioned that the analysis is rigorously carried out and documented. Nevertheless, from my point of view, the paper lacks comprehensive benchmark results. For example, different algorithms and different product categories. As a conference paper, the presented ideas are the focus of the work but for a subsequent journal publication the analysis part of the paper must be extended and strengthened. The evaluation of the performance is also a major issue. For this study we use two independent human annotators to manually extract the relevant aspects from the text. This procedure is very subjective and it also prevents the reader from reproducing the results. In addition, the YouTube comments which are used in this study are just a snapshot of a very short period in time and might not be representative. Overall the results of the paper should be treated with extreme caution and only further studies can underpin their validity. Therefore, I see great potential for improvement in the methodological approach even if the main purpose of the paper is to present new ideas and possible new directions of research. In retrospect, I would like to summarize the contribution as follows:

Contribution: We provide the first study which identifies YouTube comments as a suitable data source for social media monitoring and aspect based SA. We also show how to efficiently sample the data. We provide an in depth exploratory analysis of the text characteristics. Our results show, that YouTube comments provide several advantages compared to conventional online reviews with a comparable aspect extraction accuracy. We also propose a new method to filter out irrelevant aspect based on Google Trends.

4 Forecasting Resale Prices of Used Cars

In this section two papers are discussed. Both publications have the title "Decision Support for the Automotive Industry: Forecasting Residual Values Using Artificial Neural Networks". Four authors have contributed to this work: Christoph Gleue, Hans-Jörg von Mettenheim, Michael H. Breitner and Dennis Eilers. The first study about this topic was submitted and accepted after one revision at the Internationale Tagung Wirtschaftsinformatik (WI) 2017 in St.Gallen [Gleue et al., 2017]. In the VHB-JOURQUAL3 ranking, the proceedings of the peer-reviewed conference are rated with a "C". In the WI-Orientierungsliste of the WKWI it is rated with an "A" (highest possible score). It is the main conference of the German language information systems community. The theme of the WI 2017 was "Towards Thought Leadership in Digital Transformation". This conference paper was further developed and subsequently submitted and accepted after two revisions in the journal Business & Information Systems Engineering (BISE) [Gleue et al., 2018]. The Thomson Reuters Impact Factor in 2016 for BISE was 3.392. In the VHB-JOURQUAL3 ranking for Wirtschaftsinformatik (Information Systems Research) the journal is rated with a "B", while in the WI-Orientierungsliste of the WKWI it is rated with an "A" (highest possible score). In the VHB-JOURQUAL3 list for Wirtschaftsinformatik, BISE receives the second largest number of ratings from the community (168) after the journal Management Information Systems Quarterly (201). BISE describes the requirements for publishing in the journal as follows: "Research published in the journal examines relevant problems in the analysis, design, implementation, and management of information systems. BISE has been the flagship journal of the German-language Information Systems community for almost 60 years. It is now one of the leading European journals in the field".[2]

[2] http://www.bise-journal.com/, accessed February 2., 2018

4.1 Motivation and the Question of Generalizability

As a researcher in the field of data analytics, one depend on unique data sources which offer the opportunity to invent and evaluate new analytics techniques or to reveal new findings by empirical research methods. Data from practice often enables unique insights into real world processes. Cooperation with partners from practice can therefore ensure relevance of the research but it is always a walk along a razor's edge. Research aims to produce generalizable findings and ideas which needs to be ensured even if a cooperation with a company can lead to ungeneralizable case studies with no relevance and impact on the research field as a whole.

The basis for the research of this section is a cooperation with a large German car manufacturer which also operates a leasing business. The idea is to use data from past leasing contracts to forecast resale prices of used cars at the end of the leasing period. To have a solid idea about the resulting resale price of a car already at the beginning of the leasing period is important because the loss in value is the main factor for setting a proper leasing rate. While the leasing rate is fixed over the whole period of use, the resale price is uncertain which leads to the resale price risk. It is especially important to forecast the resale price as exactly as possible because either an overestimation or underestimation would have negative consequences. An overestimation of the resale price would lead to lower leasing rates because the loss in value which needs to be compensated by the leasing rate is underestimated. Even if these lower leasing rates lead to a competitive advantage, the car manufacturer will realize a loss when the car is resold on the used car market in the end of the leasing period. An underestimation of the resale price leads in turn to higher leasing rates and therefore to a competitive disadvantage.

The idea is therefore to develop a DSS to forecast the expected resale price based on past data and calculate the required leasing rate to compensate the loss in value. The question is, if there are any generalizable ideas from which readers of the research filed can benefit even without an application in the leasing market? Or are there any generalizable findings which can be produced based on the provided data?

Just a forecast of resale prices would not have any impact on the data analytics or information systems community. Therefore, the following sections explicitly focus on the generalizable contributions which were developed during the analysis process besides the implementation of the DSS.

4.2 Related Work

Especially in the automotive market, leasing is an important business model. 65% (178.2 billion Euros) of all leasing contracts in the European leasing market 2014 are related to automotive assets.[3] For car manufacturers and financial institutions, the question of setting a sufficient leasing rate is therefore a critical success factor. The problem of fixed leasing rates and uncertain resale prices is known in the literature as the residual value risk [Prado and Ananth, 2012]. In the context of this work we adopt the term resale price risk to avoid term confusion because of the focus of our research in the area of regression residuals. Because of the exclusive data, research in resale price risk or resale price forecasting is still rare. One of the main contributions in this field is the work of [Lessmann et al., 2010]. 124,386 past leasing contracts of the same vehicle type from a large car manufacturer are used to forecast resale prices. Support vector regression is adopted to build their DSS. 176 attributes are available which describe the objects. Their focus is on the investigation of transaction specific variables. For example, the authors investigate the influence of typical characteristics of the customer. They demonstrate the benefits of using this only internally available information in their forecasting model. Based on their findings we are particularly interested in the question which factors are actually important to perform a proper forecast on the leasing market. Since the submission of the papers for this dissertation, the authors of the previously presented study further extend their ideas to improve the forecasting accuracy [Dress et al., 2017]. The approach is to use asymmetric cost functions since over-/underestimations not necessarily have the same consequences for the respective business context. Another

[3] `http://www.leaseurope.org/uploads/documents/FF%20Leaseurope%202014.pdf` accessed January 13. 2018

paper from [Wu et al., 2009] implements a forecasting system of resale prices for the Taiwanese market. They propose a new model type named adaptive neuro-fuzzy inference system in combination with ANNs. [Lian et al., 2003] focus on the time series perspective. They found evidence for high resale price at the beginning of the year and low resale prices during the end of the year. Most studies use macroeconomic factors to explain price developments. [Prado, 2009] uses the diesel price and the industrial production index as explanatory variables. But comprehensive benchmark studies about important variables or appropriate model types are still missing. [Fan Hongqin et al., 2008] provide a comparison between AutoRegressive Trees, ANNs and linear regression with data from heavy construction machines. But only little information is given about the actual implementation of the models. Hence, the validity of the results is questionable. The presented papers focus on empirical methods with real-world data. Another stream of research focuses more on theoretical models in the leasing business and the resulting implications for risk management. This is beyond the scope of this work but I refer to [Rode et al., 2002], [Storchmann, 2004] and [Smith and Jin, 2007] for further reading.

The following section covers the main contribution of our work in this field and highlights the unique value we provide compared to other studies. The focus is on the question of generalizable findings and methods which have a real impact not only for one company but also for the forecasting and data analytics and especially for the leasing research community.

4.3 Research Gap: Time Factor and Feature Importance

An earlier version of the paper was submitted to the Journal of Forecasting in 2016. In the submitted draft we have focused on the forecasting application and methodology using ANNs. The paper was on a technical level and described how to implement and correctly optimize ANNs for the purpose of resale price forecasts. After the first review cycle the paper was rejected. The reason of the associate editor can be summarized as follows: "With this paper, whilst the topic is interesting, and the methodology is leading edge, there is not a sufficient research contribution to forecasting methodology in the paper. It is a good empirical exercise to undertake this study, but in the end, the reviewer felt that the research contribution was not sufficient". This shows the conflict of practical oriented research which is often relevant for a small part of the community but without generalizable contributions. Therefore, we have completely redesigned our research approach to focus more on generalizable findings and ideas while using the data just as an example application to validate the approach in form of a case study. Hence, the focus now is not anymore on the application itself but on the general applicable ideas which were tested in a real-world example.

Our first generalizable approach is to develop an approach to incorporate time dependent variables in in forecasting applications of the leasing market. In the conference paper for WI 2017 we present a two-step approach. First, a linear regression is applied to all data point to explain resale prices by obvious independent variables like the mileage and the age of the car. Second, we focus on the residuals of the resulting model. By sorting the residuals according to time, filtering outliers and grouping them together on a monthly basis, we received a time series of unexplained information over four years of observations. This time series is further analyzed by the Seasonal and Trend decomposition using Loess (STL) method [Cleveland et al., 1990] which allows us to clearly distinguish between seasonal and trend effects. Figure 13 show a graphical representation of the decomposition.

The results reveal interesting findings. The leasing market obviously follows a clear trend and a seasonal pattern. Even if the seasonal effects are rather small, it seems that resale prices tend to be higher during spring and reach a minimum during the end of a year. Domain experts explain the results by contract fulfillment requirements of the used car dealers. Therefore, in the end of the year often promotion measures are undertaken to reach the sales targets. The effect in spring is a market reaction to supply and demand. Incorporating a time factor and monthly dummy variables lead to the following regression model:

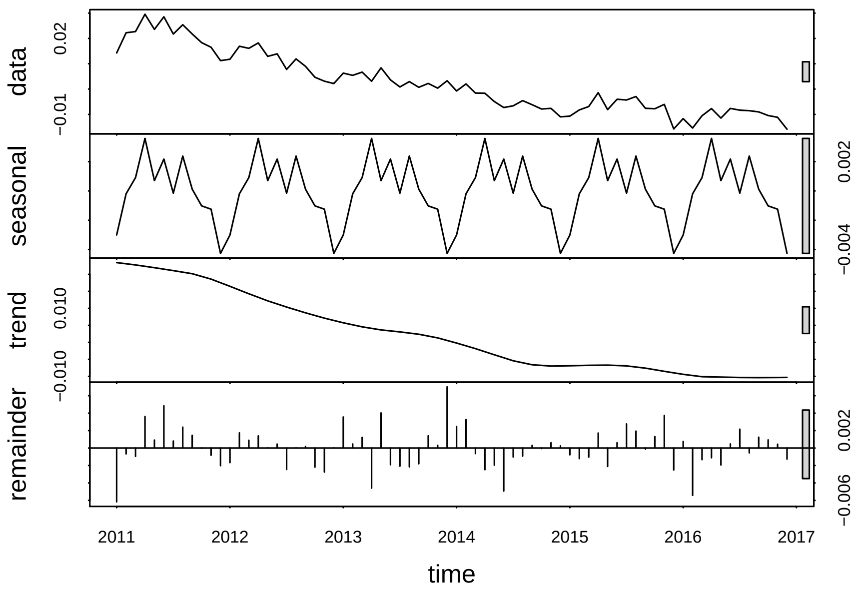


Figure 13: Residual Time Series Decomposition

$$rv_k = c + \sum_{j=1}^{13} (\beta_j \cdot feature_{j,k}) + \gamma \cdot Time_k + \sum_{m=1}^{11} (\delta_m \cdot Month_{m,k}) + \varepsilon_k \qquad (17)$$

with rv as resale price, which is explained by 13 features (independent variables), an increasing time factor and monthly dummy variables. Using the BDS (Broock, Dechert and Scheinkman) test [Broock et al., 1996] we show that even with the new model, a misspecification in the residuals can be observed due to non-linearities in the data. Therefore, a simple linear model was not sufficient to be used in a forecasting application.

Here, two types of empirical work must be distinguished: inference and prediction. Empirical papers which try to explain certain phenomena, use scientific methods which need to be interpretable. But interpretability and forecasting ability of non-linear dependencies often face a trade-off. In this study we therefore use ANNs for the forecasting application and linear regression for the inference about certain factors. In the empirical part, however, we show that ANNs have problems to capture the time factor in out-of-sample data. Forecasting horizons of more than half a year lead to statistically significant biases in case of ANNs. To deal with this problem, we have developed a forecasting ensemble which combines both model types. The idea is to first filter out the time factors by a linear regression approach. The data points cleaned from the time factors are used to produce time independent, non-linear forecast by using ANNs. Afterwards, the time factor is added bag to the results depending on the forecast horizon. The empirical results show, that our approach receives statistically significant better results than each model type separately. The forecasts are statistically unbiased for more than one and a half year which is three times longer than using just ANNs. The rule-based ensemble therefore significantly improves the forecasting results.

As already mentioned, forecast ability of non-linear dependencies often comes with less interpretability of the resulting models. For the BISE journal paper, we have conducted a comprehensive benchmark of different methods for the forecasting application. We were also able to validate our results on two more years of data from the car manufacturer. Ridge regression represents the linear model type. Ridge regression prevents the linear model from overfitting by penalizing many large coefficients in the model already during the optimization. Compared to a standard linear regression, not only the residual sum of squares is minimized but also a second term which includes the coefficients of the model. The optimization can be formalized as follows:

$$\min_{\theta} \sum_{i=1}^{n} (Y_i - h_\theta(X_i))^2 + \lambda \sum_{j=1}^{p} \theta_j^2 \tag{18}$$

With Y_i as real observation and $h_\theta(X_i)$ as hypothesis of the model which depends in a parameter vector θ. λ is now an additional parameter which controls the influence of the penalization. We select λ by a naive grid search. A detailed description about the results can be found in [Gleue et al., 2018]. The main finding is that ANNs receive the best results when it comes to forecasting applications. To better understand the reasons for the results and to make the results more comprehensible by human who have to apply the models subsequently, we implement a feature importance analysis based on the idea of [Breiman, 2001]. The procedure can be described as follows:

1. Train an ANN with training data
2. Calculate an overall evaluation criterion like the Root Mean Square Error (RMSE)
3. Perform a perturbation on one input variable by randomly shuffle the values in an out-of-sample data set
4. Calculate the evaluation criterion for each randomized variable
5. Calculate feature importance as $\frac{error\ for\ variable\ i}{\max(error)}$
6. Sort the list of variables in descending order by their feature importance

With this algorithm implementation we are able to analyze which variables have the highest influence on the quality of the model. Table 7 show the results.

Table 7: Feature Importance

Factor	RMSE	Importance
Age	0.187	1.000
Mileage	0.100	0.536
Number of Gears	0.081	0.431
Initial List Price	0.075	0.399
Financing Type	0.071	0.380
Distribution Center	0.068	0.364
Vehicle Type Age	0.067	0.358
Engine Capacity	0.067	0.356
Additional Equipment Price	0.067	0.356
Horsepower	0.066	0.354

The so generated findings reveal a highly topical fact which is controversially discussed in the literature about leasing markets, especially car leasing market. During the analysis process we have investigated many possible variables which might have an influence on the resale price of cars. This also includes, besides contract and car specific information, the information about the macroeconomic situation. Therefore, we have included variables like the oil, diesel and gasoline price, a consumer price index and interest rates. We used linear mixed effects models to estimate the explanatory power of such external factors. We find that none of these factors have a significant influence on the resale price after controlling for all internal factors [Gleue et al., 2018]. Also the feature importance analysis of forecasting application reveal that after applying a method which is able to capture non-linear dependencies in the data, the information about external factors, have no relevant impact on the results of the forecasting model. Non of the external factors are ranked in the top ten variables presented in Table 7. The change in the RMSE of a correct specified ANN (0.061) was not different from the RMSE of a model with any randomized external factor.

4.4 Conculsion, Discussion and Outlook

The findings lead to the conclusion that external factors during our time period of investigation have no influence if the model exploits all internally available sources of information. This is in contrast to many other studies in this field which heavily rely on external data sources. Our recommendation we draw from the study is therefore as follows:

1. Incorporate seasonal and trend variables by combining linear and non-linear models.

2. Exploit as many internally available variables which describes the contract and the leasing object since this could lead to a competitive advantage compared to other forecasting providers which heavily rely on externally available information.

However our findings are only valid for the sample we have considered (2011-2017). Long-term forecast face the risk of external shocks like new regulatory measures or technology disruptions. The time window of the analysis is also embossed by a healthy market condition and only moderate economic fluctuations. More data over a whole business cycle would be necessary to investigate how the dependencies change in more volatile market situations.

A direction for further research is the idea of using the designed data base and forecasting system to implement a stress testing system. This would allow the user to adjust certain parameters to observe the potential impact on the resale price of the cars. Based on simulations of different scenarios, one could design special financial products or derivatives which rate the resale price risk.

One increasingly important business model is the idea of pay-as-you-drive [Desyllas and Sako, 2013]. The idea is to adjust insurance premiums for car owners who drive responsible. This concept can also be applied to leasing. In our empirical analysis, we currently have only included contract and car specific information. If it becomes possible to include further variables like the driving behavior or the general use of the car measured by certain sensors like temperature, vibration etc. one could also adjust the leasing rate of a car for certain customer groups. Or new models of dynamically adjusted leasing rates can be established which incentivise the user of a car to behave as gentle to materials as possible.

4.5 Reviewer Opinions, Contribution and Critical Comment

In this section I discuss the overall contribution of this practical oriented research project and the generalizability of the ideas and findings. I will also provide a critical view on the research process and methodology.

To assess the contribution, I provide again comments from the anonymous reviewers. This ensures an independent and external view on the topic. One reviewer of the initial conference paper states the following:

"They properly highlight how and why this work and the results obtained are of economic importance to an [Original Equipment Manufacturer] (OEM) in order to derive unbiased and accurate expectations."

The comment show that the paper has real value for the practical oriented reader. As the reviewer points out, we have tried to focus on the how and why our methodological approach and the obtained results can be used as generalizable findings not only for the car manufacturer we have worked with.

After the first submission to the BISE journal, we received the following comment from another reviewer.

"The contribution of the paper is more on the practical side (high relevance) than on the theoretical side. [...] I think that the authors should definitely clarify the research goal and the scientific contribution of the submission."

This comment shows that we have not fully succeeded in providing general scientific results. In the revised version of the paper we have focused even more on the time factor in our discussion of the results to show the scientific contribution in forecasting resale prices of leased objects in general. Our unique data set enables us to provide insights in how to deal with such a time dependent forecasting problem over a long horizon. The second anonymous reviewer state the following about the contribution of the work.

"The specific way in which this data is employed (cleaning, preparation, etc.) adheres established scientific standards and guarantees external validity to the largest degree possible. The discussion of different variables and analysis or their relative predictive value, which draws inspiration from the random forest variable importance, represents another strength of the paper. Prediction methods are also used in a skillful way and illustrate how to produce accurate and reliable forecasts from empirical data. One particular strength is given by the provision of different use cases for residual value estimation and design of corresponding experiments. All in all, the motivation and setup of the research task, the design of the empirical analysis, and the interpretation of observed results display excellence and make this paper a highly remarkable piece of research."

The reviewer points out all the contributions we have tried to address during the project, especially the design of different use cases and the experimental design. But this comment is very much in favor of our paper and provides only little value for further improvements. The following comment provide a more critical point of view on the contribution.

"I have no doubt that the paper will draw an audience. Especially practitioners involved with pricing durable goods, forecasting, and fleet management, to name a few, will find the paper highly valuable. However, the degree to which it contributes to the academic literature is moderate. The paper is not meant to build novel theory. Moreover, the prediction methods employed are well-established. Consequently, no methodological contribute is made to the forecasting literature. The contribution of the paper lies in the case study and the empirical results provided therein."

Comparing all reviewers' comments highlights the contrast between practical and scientific contribution. Overall, the reviewers agree on the practical contribution and the high relevance of the findings and developed forecast procedures. The main weakness is the scientific contribution. We have not invented a new method or improved any established standard in the forecasting literature. Our goal with this paper is to provide scientifically validated recommendations for a practical audience with similar long horizon forecasting problems. In further research, we will focus more on methodological improvements for the scientific forecasting literature.

From my point of view the work must be seen very critically. Maybe more critical than the reviews suggest. Even if we have reached practical impact and relevance the overall goal of research remains to provide scientific contributions. Practical contributions are important as well, maybe even more important to make real use of scientific discoveries. But the question is if this is the claim for scientist to provide this kind of value? Is it a task of a researcher to show how to practically implement already developed methods in a certain type of application? After finishing the research project, I am still not sure about it. At least it is important to ask the question as a researcher, what is the real justification of my work and is that the task of a researcher?

The overall contribution can be summarized as follows:

Contribution: We provide one of the few studies which use real leasing data sets. The unique data is an asset of the paper which allows us to

investigate which factors have an influence on the resale price of used cars. We also provide lessons learned when it comes to long-term resale price forecasts and provide recommendations for practical forecast applications beyond the leasing domain.

5 Visual Model Evaluation

This section covers my main research topic. Three papers have been published to this topic. The first paper has the title "A Picture is Worth a Thousand Words: Visual Model Evaluation in Data Science Applications" [Eilers and Breitner, 2017] which was submitted and accepted after one revision at the at the Internationale Tagung Wirtschaftsinformatik (WI) 2017 in St.Gallen. The second paper with the title "It's not a Bug, it's a Feature: How Visual Model Evaluation can help to incorporate Human Domain Knowledge in Data Science" [Eilers et al., 2017] was submitted and accepted after one revision at the International Conference on Information Systems (ICIS) 2017 in Seoul. In the VHB-JOURQUAL3 ranking, the proceedings of the peer-reviewed conference are rated with an "A". In the WI-Orientierungsliste of the WKWI it is rated with an "A" (highest possible score). The theme of the conference was "Transforming Society with Digital Innovation". The ICIS is the leading conference of the AIS and can be considered as "the most prestigious gathering of information systems academics and research-oriented practitioners in the world".[4] The third paper with the title "Understanding Anomalies: Visualizing Sensor Data for Condition Monitoring of Manufacturing Machines" [Olivotti and Eilers, 2018] was submitted and accepted after one revision at the Multikonferenz Wirtschaftsinformatik (MKWI) 2018 in Lüneburg. In the VHB-JOURQUAL3 ranking, the proceedings of the peer-reviewed conference are rated with a "D". In the WI-Orientierungsliste of the WKWI it is rated with a "C". The theme of the conference was "Data driven X - Turning Data into Value".

5.1 Motivation and Purpose

The research in this area can be considered as a consequence of the general findings from the three previously described fields. Developing decision support algorithms in a finance context, applying text mining algorithms to analyze review data and

[4] `http://aisnet.org/page/ICISPage` accessed November 20. 2017

designing a forecasting system for resale prices of used cars based on ANNs are in and of itself interesting research topics. But the question arises, why it is necessary for an information systems researcher to work on these questions. Are finance experts, computer scientist and business administration scholars potentially more suited to work in these fields? Where is the justification for a cross-sectional discipline like information systems research to provide contributions which are relevant for other scientific communities? What can we provide that has an impact on the broad filed of data analytics which cannot be provided by any other field of research?

As discussed in the introduction and motivation section, information systems research is an interdisciplinary research field which not only combines methods and tools from management and computer science but provides its own view and solution approaches. From my point of view, one of the main purposes of information systems research is to build bridges between the involved disciplines like management and computer science which depend on each other but often speak different languages. The same is true in the emerging field of data analytics. Highly skilled data scientists develop machine learning models and statistical analysis approaches for their companies to provide for example a better customer satisfaction or reveal competitive advantages. On the other hand, there are decision makers and managers who have worked for many years in a certain area and therefore have tremendous experience and so-called domain knowledge. A respectful and serious interaction between these groups is important because data scientist need the domain knowledge about the topic that they are investigating to understand the dependencies in the data and consequently to develop unbiased models which cover all relevant influences. The decision makers, however, need to trust the models and solutions, the data scientist came up with to actually make the right decision. The prerequisite for this is a proper explanation of the model behavior and therefore communication between data scientist and domain experts on the same level of complexity. Research in the area of how to integrate domain experts and data scientist is surprisingly rare in the literature. Many studies argue in favor for comprehensible models and the importance of understandable results. [Ionescu et al., 2016] for example develop a data driven detection of decision rules for daytime airplane delay trends. They argue that an interpretable model can help to better understand the delay mechanisms. But explicitly addressing the question of how to bridge the gap between data science and domain experts is still not represented in the scientific literature as an own research topic. One reason for this could be the focus of computer scientist to develop improved algorithms and the focus of management-oriented scholars on behavioral research. This situation is especially surprising since in the data science practice, the idea that domain knowledge could be a critical success factor, is already widely discussed.[5] Therefore, I think a better understanding of the role of domain knowledge in data science is a research gap especially for information systems research as a field which builds bridges between different areas. I would like to point out that it is important not to see information systems research as a bad copy of computer science/statistics or management science. Information systems research can address challenges which are not or cannot be addressed by any other research filed on its own. And I think one of the challenges is to investigate the question of how to better integrate human domain knowledge and data science. The results should be more than the sum of its parts. Any data related research should benefit from the findings. These thoughts were the starting point for me to think about ideas to fill this research gap from the view of an information systems researcher. The following background shows the process of how to build the ideas by analyzing the existing literature in related fields.

5.2 A Literature based Justification of the Research

The starting point for the research was the project about forecasting resale prices (Section 4). During the work, we observed that our forecasting models performed reasonably well on out-of-sample data but only after a very restrictive outlier elimination beforehand. The experts explained the unusual number of outliers with a

[5] See for example `https://www.datacamp.com/community/tutorials/feature-engineering-kaggle` accessed January 19. 2018

manual data input on the part of the contract deals. But manual errors usually lead to errors which are completely unrealistic like a negative resale price or a reported car age of more than 100 years. The outliers we observed however often lay in a reasonable data range, but the results didn't fit into the patterns of previous observations. Declaring these data set as outliers is the easiest way to handle the situation because the task was to forecast a typical used car. Special cases explicitly shouldn't be included, since they would bias the models. The goal was to develop models for usual cases. An alternative explanation for these patterns would be unobserved predictor variables. Unobserved variables which have an influence on the dependent variable but are not included in the model can lead to unexplained errors. A way to investigate if important predictor variables are missing in the model specification is a residual analysis. Unobserved variables lead to remaining information in the residuals of a forecasting model. This can result in certain patterns which let the residuals deviate from a normal distribution. Residual analyses are known from linear regression applications. The typical approach is to use QQ-Plots or residual plots (residuals plotted against fitted values) to investigate if there still remain information in the residuals of the model [Kuhn and Johnson, 2013]. But these standard methods do not provide a simple way to reveal what the reason for a certain type of pattern can be or which variable is actually missing. Identifying the right variables for a model is always a complex task and relies on human intuition and experience. Summarized, this situation leads to the following problem derivation:

1. Unexplainable patterns are represented more frequently than would have been expected if they were simple outliers (measurement errors etc.). Standard residual analysis techniques support this first impression.

2. Is it possible to identify missing variables which might explain these patterns?

3. Identifying important variables requires domain knowledge. But domain experts are not necessarily experts in machine learning techniques. So how can we incorporate the domain knowledge in our models?

4. To realize this, a better communication about the models, application domain and data needs to be established. How can we facilitate communication between different expert groups?

The problems and questions we are facing require a comprehensive literature search process to identify the areas where we can find possible solution approaches. First, we start by search for research in the area of omitted variable biases. The knowledge from previous studies lead to the term Feature Engineering (FE) which is a frequently used term in the machine learning literature. While the literature about omitted variable biases is mainly a statistical issue, FE is a more practical approach to construct a proper input space for a machine learning algorithm. Domain knowledge is often mentioned as an important factor for a successful FE. The topic of FE leads to the last question of how to better incorporate and communicate with domain experts about complex mathematical models. In the literature, visualizations are a common approach to build bridges between different expert groups. It is also often used as a communication vehicle. Therefore, we focus on visualization as a set of tools to realize our ideas. The following subsections represent the search and idea development process and how we combine ideas and different approaches to solve the described problem.

5.2.1 Literature about the Omitted Variables Bias Problem

The starting point for our literature search is the typical Omitted Variable Bias (OVB) problem since our initial goal is to find the right variables which explain the dependencies in the data. The problem of omitted variables is a highly discussed topic in the econometric literature [Clarke, 2005] [Lee, 1982]. The bias is defined for a linear regression analysis as the problem of an over-/underestimation of certain coefficients in the model due to a missing factor. This factor must be correlated with the depended variable and one or more independent variables. There exists test for identifying omitted variables like the Regression Equation Specification Error Test [Ramsey, 1969]. But there exists no clear approach to identify which variables are actually missing.

Other methods for analyzing the performance of regression models are the typical residual and QQ-plots. Residual plots provide hints on misspecifications of the model if there are clear patterns visible in the distribution (residuals plotted against predicted values). QQ-plots show the residuals plotted against a theoretical quantile from the normal distribution. Obvious deviations from the normal distribution (deviations from the straight line) / fat tails may indicate (depending on the amount of available data) that the interpretation of the p-values can be wrong due to outliers and errors which cannot be explained by the model in its current specification. But there is again no easy way to identify the reason for a certain observation.

Hence there remain two problems after a first screening of the econometric literature about omitted variables. The most important issue is the fact that there exist no simple method to really understand the reasons for a certain bias in the model and second, the research focuses on inference and not on prediction. The reason why it is important to avoid an OVB in a linear regression context is the fact that the coefficient can be biased and therefore the inference from the models can lead to wrong conclusions. But this literature stream provides no further ideas for more prediction-oriented approaches. Therefore, we redirect our search activities towards omitted variables in the context of machine learning. A keyword which appears in this context more and more often is FE which is analyzed in the next subsection.

5.2.2 Literature about Feature Engineering and the Role of Domain Knowledge

In machine learning the term feature is defined as an independent variable of a prediction model. FE is then defined as the construction of a proper input space for a machine learning model which includes variable preprocessing like scaling, constructing dummy variables for categorical variables or find and combine new features which can improve the performance of the model. This process and not the algorithms themselves can be seen as critical success factor for machine learning since the features are domain specific and change in each application while algorithms can be generally replaced [Domingos, 2012]. Several studies in highly ranked information systems research journals justify the success of their models by an intelligent FE [Lash and Zhao, 2016] [Ghiassi et al., 2016]. FE is often described as a creative process which involves human intuition and experience. It is usually an iterative procedure compromising model construction, evaluation and adjustment. In the machine learning oriented literature it is shown that engineering the right feature can outperform sophisticated algorithm optimizations with less suitable features [Yu et al., 2010] [Bengio et al., 2013] [Heaton, 2016]. Hence it can be seen as a key success factor for machine learning applications. FE is a relatively new research filed. Many studies therefore base on ideas from FE but research which explicitly focuses on methods for improving FE is rare. Analyzing the few paper about this topic show two streams which emerge within this area. The automated FE literature and the human/domain knowledge-oriented literature. I will provide an overview of both streams in the following.

Automated Feature Engineering: Automated FE, or Feature Learning referrers to the ideas of how to construct a proper input space for a model automatically based on already existing features [Heaton, 2017]. [Coates et al., 2011] show how a simple single layer neural network learns features automatically for further higher-level tasks like classification. Other common techniques, especially used for dimensionality reduction but also for FE include Principal Component Analysis [Timmerman, 2003] and t-distributed stochastic neighbor embedding [Maaten and Hinton, 2008]. [Coates and Ng, 2012] use K-means clustering for learning large-scale representations/features of images. [Kanter and Veeramachaneni, 2015] develop the Deep Feature Synthesis algorithm to derive the right features for relational data sets. [Heaton, 2017] introduce the first genetic programming algorithm for FE. As shown, the current literature focuses on constructing new features based on already existing data. It is shown that these methods can improve the performance of the subsequent machine learning algorithms. But completely new features, which incorporate further information beyond the available data are not constructed. The second research stream therefore focuses more on human intuition and domain knowledge to improve the FE process.

Human/Domain Knowledge-oriented Feature Engineering: Contributions to this field are even more sparse and often only indirectly mentioned. [Lash and Zhao, 2016] use intelligent FE for predicting movie profitability. They show that their new developed features make weighty contributions to the prediction. This study is one of the first which explicitly focus on FE as a key success factor for their model. The contribution also shows how to develop new features which is in contrast to the typical approach of combining already existing features. [Ghiassi et al., 2016] present a study about SA of brand-related Twitter posts. They comprehensively describe the process of FE. Their approach is based on a manual process of developing the reasonable features which results in a set of seven dimensions. They show that their results outperform state of the art algorithms. Even if the focus of the work is the development of the SA system, the study provides interesting insights in the process of FE and why it was important to creatively think about and construct new and reasonable features which are suitable for the proposed application.

After reviewing the sparse literature about the topic one question remains if the definition of FE as an iterative improvement process holds. How to engineer the right features without knowing reasons for certain unexplainable observations? How can we identify features which are important to explain currently unexplainable patterns? The problem of how to find possible omitted features which might explain certain patterns in the data still remains. This literature stream, however, provides hints that creativity, intuition, experience in combination with domain knowledge is essential to solve this problem. Therefore, the human who is responsible for constructing the models plays a much more important role. But there is still no clear direction to clearly answer the question of how to approach this problem properly.

5.2.3 Literature about Visualization and Communication

The resulting question of how to do it properly leads to a literature search about other areas where similar problems have been already discussed. If domain knowledge is really as important as stated in the literature about human-oriented FE, methods are needed to incorporate human domain knowledge in the models. One research field which deals with similar problems is visual analytics. Visual analytics is the use of visualization in an analytics context to generate insights through human perception. [Thomas and Cook, 2006] first introduce the term as the science of analytical reasoning facilitated by interactive visual interfaces. Most approaches use visualizations of large data sets to make sense for the human user [Keim et al., 2008]. [Tam et al., 2017] use visualizations to enable the user to generate decision trees for detecting facial expressions and emotions. The models generated by humans outperform the machine learning approaches in their study. For reviews about the field of visual analytics I refer to [Sun et al., 2013] and [Endert et al., 2014]. The overall idea is the integration of human knowledge in the analysis process for joint reasoning of humans and machines in order to achieve better results. In the literature, other positive effects of applying visualizations have been documented. [Bresciani and Eppler, 2009] for example show that using visualizations have a positive impact on the communication in a team which is essential for the overall team performance.

In summary, we have therefore derived the following solution idea:

1. Unexplainable patterns could be caused by missing information/missing variables in the provided data.

2. The research field which deals with the problem of finding proper input variables is called FE which can be divided into two streams namely the automated FE which is based in the already available input space and the human/domain knowledge-oriented FE which explicitly focuses on the intelligent and creative construction of new features which are suitable for the task at hand. To find

the right features for our models, we need to focus on the second research stream.

3. Then, to solve the problem that data analysts not necessarily have the required domain knowledge and domain experts not necessarily have the required analytical skills we have identified visual analytics as a research filed which deals exactly with these problems.

4. Therefore, we propose the idea of using visualizations to facilitate domain knowledge-oriented FE.

5.3 Methodology: Design Science Oriented Research Approach

After the identification of the problem and the derivation of an idea, we choose a design science oriented research approach to develop an artifact that can be used in real data science projects. The process follows the guidelines proposed by [Hevner et al., 2004].

Design as an Artifact: Our goal is to develop a visualization method which can be applied and tested in a real data science project.

Problem Relevance: We have identified a communication and understanding gap between data science experts and domain experts. However, both groups depend on each other since data science experts need the domain knowledge to build proper models and domain experts may trust and accept the models if they are actively involved in the development process. From our point of view the identified research gap is suitable for information systems research and design science because it addresses a relevant business problem.

Design Evaluation: We will evaluate the developed visualization method by means of artificial examples with controlled data and in a real industry case study.

Research Contributions: The designed artifact should be suitable for any kind of data science project. It should be easy to implement, use and interpret. To strengthen the contribution, we provide the code for the artifact online in a freely accessible repository.[6]

Research Rigor: The literature research shows that visualizations are suitable methods to fulfill our goals since they are already widely accepted and applied to similar problems. We also want to evaluate the visualization method as rigorous as possible by means of multiple application examples with controlled data sets and a real-life case study which are accepted methods to fulfill the requirement.

Design as a Search Process: We see our research as an iterative process of finding and developing a proper solution for the problem at hand. The goal is to find a "good" solution which works for our investigated cases, but which is also applicable for any other case. We want to iteratively improve the developed solution and by uploading the code on GitHub we also encourage the community to adapt the method and propose new ideas for improvements or extensions.

Communication of Research: We present the research for a technical oriented audience by providing the documented source code and the mathematical implementation. We address the management-oriented audience by providing examples of how to use the artifact, how to work with it and how to interpret the results in a comprehensible way.

5.4 Critical Discussion of the Mathematical Implementation

In the paper "It's not a Bug, it's a Feature: How Visual Model Evaluation can help to incorporate Human Domain Knowledge in Data Science" we show why heat maps are a suitable visualization technique for our purpose. The resulting artifact visualizes the residuals of the forecasting model in a two-dimensional data space. The visualization in the same data space as scatter plots allows us to highlight regions in the data space where the models perform good or bad. The mathematical implementation is comprehensively described in the paper. Compared to the paper, this section critically discusses the mathematical implementation and its shortcomings.

The idea is that each data point represents the local model error in a certain

[6] `https://github.com/eilersde/residualheatmap`

region around its position. As shown in [Eilers and Breitner, 2017] each data point is assigned a two-dimensional Gaussian kernel which represents the local influence of a residual. With increasing distance, the influence of a residual decreases. Figure 14 illustrates the weighting procedure. Darker red regions represent a higher influence of the residual on its surrounding grid points. Figure 15 shows an example scatter plot from a two dimensional normal distribution (15a) and the corresponding residual heat map which visualizes a white noise on a diverging color palette with 64 divisions (15b). The color scale is chosen to be interpretable even for color blind users [Wong, 2011] (improvement from [Eilers and Breitner, 2017] to [Eilers et al., 2017]). [Borland and Taylor, 2007] show why the use of a standard rainbow color scale is not recommended and can lead to wrong conclusions.

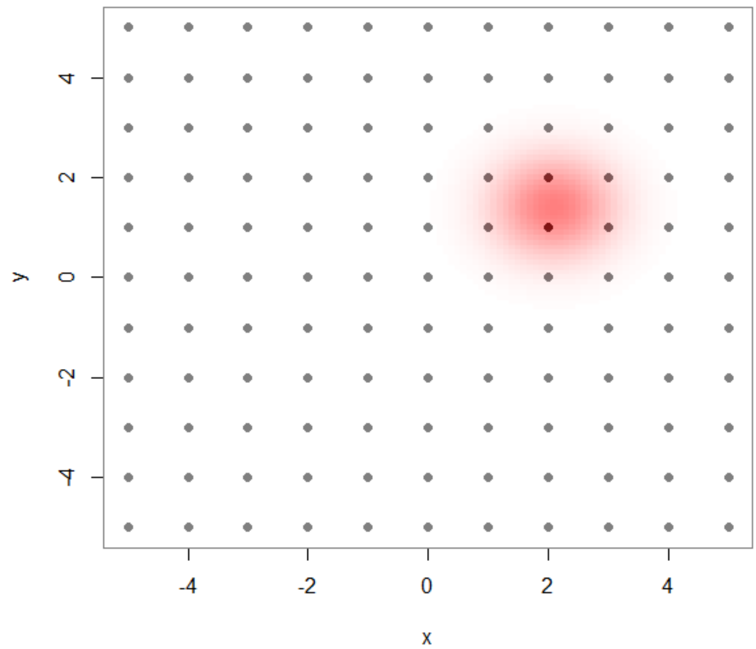
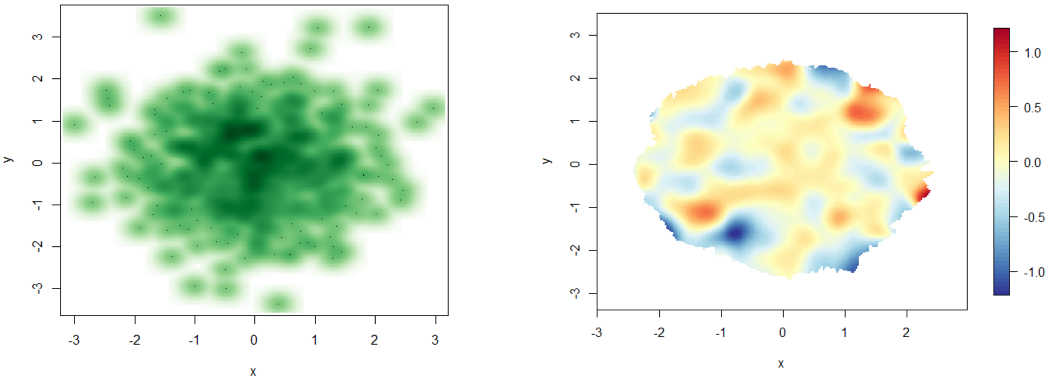


Figure 14: Residual Weighting based on a Gaussian Kernel

The first problem with this implementation can be analyzed based on Equation 19 (Equation 5 in the original paper) where $w^i_{m,n}$ is the weight for residual i at grid point m, n which is divided by the sum of all weights at grid point m, n and $c \in N$ is the number of residuals which defines how many residuals must have a relevant influence on a certain grid point to be colored.



(a) Scatterplot

(b) Heat Map of White Noise Residuals

Figure 15: Scatter Plot and corresponding Heat Map Visualization

$$h_{m,n} = \begin{cases} \sum_i \left(r_i \frac{w^i_{m,n}}{\sum_j w^j_{m,n}} \right), & \text{if } \sum_i [w^i_{m,n} > 0] \geq c \\ NA, & \text{otherwise} \end{cases} \tag{19}$$

In case that only two residuals $a1$ and $a2$ have an influence on the grid point m, n and $c > 1$, the current implementation would lead to colored grid points only between the actual data points where the Gaussian kernels intersect and represent values greater than the threshold parameter. But the grid points which are closest to the data points may not be considered due to the restrictions of the threshold parameter and c. This contradicts the idea of data points being representative for their local region.

The second problem with this implementation is the possible misinterpretation in areas of transitions between two main influencing residuals. For example, if two residuals $a1$ and $a2$ are close together with a high negative value for $a1$ and a high positive value for $a2$, the region of the transition would indicate an area of zero derivation from the real value, since the values neutralize each other. The interpretability of the results in this region is at least questionable because the absolute values of the residuals are very high.

Another problem is the Gaussian kernel itself. In the current implementation each grid point is inserted into the generated kernel for each data point. The density function of a Gaussian kernel however reaches zero only asymptotically which leads to the problem that residuals would have an influence over the whole data space. To mitigate the problem, the threshold parameter is introduced which sets all weights below this value to zero. An alternative would be to use a different kernel. The Epanechnikov kernel might be a reasonable alternative [Epanechnikov, 1969].

In addition, the current implementation is computationally very expensive. Generating Gaussian kernels for each data point and inserting each grid point in each kernel to receive the final weights of the heat map leads to tremendous computing times. A heat map with of the dimension 200x200 grid points and 2000 data points in the distributions already needs 196 seconds for calculating all the weights on an Intel Core I5 2,4 GHz. 7. Generation (R Version 3.4.2). Possible alternatives would be to use lookup tables for the Gaussian kernel, implement the procedure in a faster programming language than R or develop a more efficient way to calculate the weights.

5.5 Evaluation: An Observational and Descriptive Approach

The evaluation of this method is one of the most critical points regarding the design science-oriented research approach. As [March and Storey, 2008] state a real design science research contribution requires a demonstration that no adequate solution exist in the knowledge base and afterwards a rigorous evaluation of the developed artifact which enables an assessment of its utility. [Hevner et al., 2004] summarize possible evaluation methods. These methods need to be appropriate for the presented artifact, however, they allow different strong evidence. We choose descriptive and observational methods. First, we construct typical scenarios with artificial data to demonstrate the utility. Therefore, we reproduce the experiment from [Heaton, 2016] and train an ANN with 2000 random samples from the following ratio-difference function:

$$f(x_1, x_2, x_3, x_4) = \frac{x_1 - x_2}{x_3 - x_4} \tag{20}$$

To evaluate the performance, we use the heat map approach on 1000 out-of-sample data points and choose a sequential color scale with 64 divisions because we use the absolute values of the residuals. Figure 16 show the results. With the heat map approach we are able to visualize the exact region where the model performs rather poor (when the denominator tends towards zero and the equation is undefined). The heat map for the variables x_1 and x_2 show no clear pattern in Figure 17.

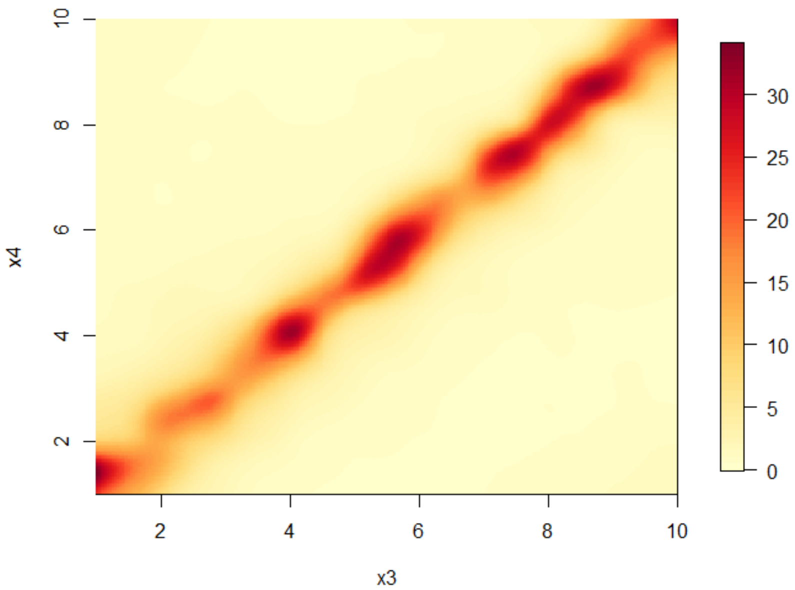


Figure 16: Bias of the ANN if the Denominator is Undefined

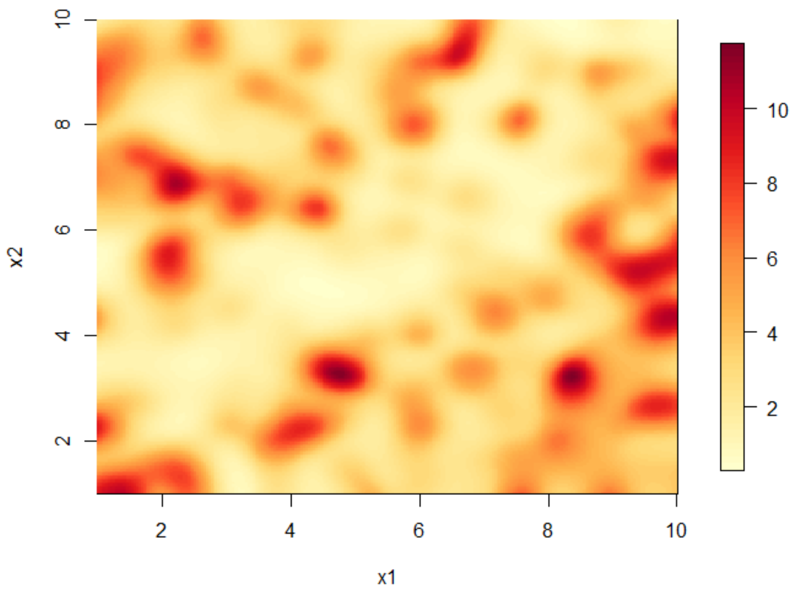


Figure 17: Evaluation Heat Map with Random Noise

This is one of the weakest types of evaluation methods science the constructed scenarios might be explicitly designed for the artifact and cannot be generalized for other applications. But as [Hevner et al., 2004] state, these descriptive methods can be used for especially innovative artifacts. We decide to incorporate this methods, not only do evaluate the artifact, but also to make the application of the artifact more comprehensible. The descriptive approach allows us to explain the functionality of the artifact in a controlled environment. We are therefore able to point out explicit ideas, strength and weaknesses of how and when to apply this technique to a certain problem. Therefore, it serves as an additional explanation approach and at the same time as first evidence for the utility which is essential for a successful evaluation process. But since this is a weak form of evaluation we decided to add another method to evaluate the artifact from an observational standpoint. As suggested by [Hevner et al., 2004] we study the artifact in depth in a real business environment. The method can be seen as a single case study which is conducted in cooperation with the car manufacturer from which the data of Section 4 come from.

Here I present one example of how we have used the heat map approach during the development of the resale price forecasting model. Figure 18 shows a residual heat map which plots the dependent variable (resale price) against the mileage of the cars (scaled values for anonymization).

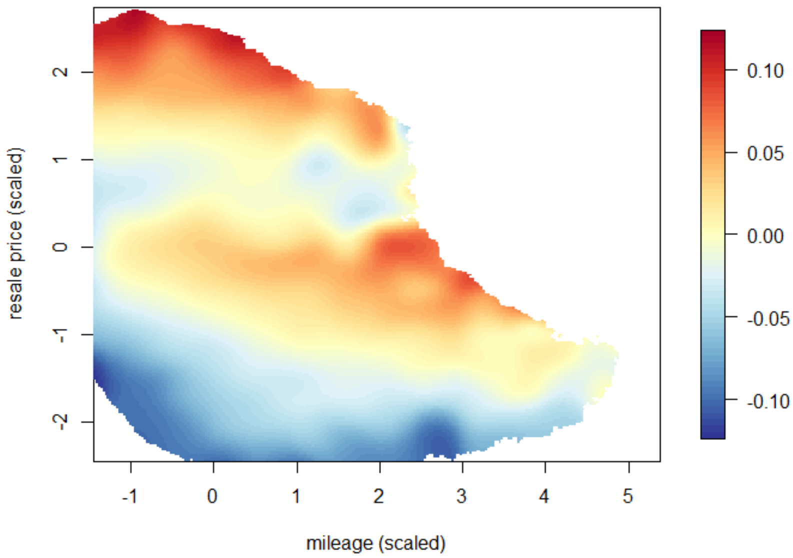


Figure 18: Residual Heat Map for the Resale Price Forecast

True outliers and/or measurement errors are represented in the red region at the top of the distribution and in the blue region at the bottom of the distribution. But there is another pattern which cannot be explained by the outlier assumption. The red region around zero of the resale price which increase in magnitude with increasing mileage. This regions shows an underestimation of the resale price. The pattern is still puzzling but the identification of such an anomaly has triggered a diverse discussion around domain experts about possible reasons. One hypothesis is that the competition in rural areas for used cars is much lower than in urban areas. Therefore, higher resale prices of cars with a higher mileage could be achieved in rural areas which is currently not reflected in the data. This example show how our approach of visual model evaluation can help to incorporate human domain knowledge in data science.

Such observational evaluation approaches are more rigorous since the artifact is applied in a real business environment with real challenges which are not artificially constructed. The limitation is still the question of generalizability since this evaluation technique is based on a single case study. Hence the solution might also "overfitted" to the problem at hand. To further increase the validity with observational evaluation approaches, more cases and field studies, as suggested by [Hevner et al., 2004], are necessary. We therefore develop new ideas about how to apply the presented method in other use cases. One idea is discussed in the following subsection.

5.6 Future Applications: Understanding Anomalies

The thoughts presented in this subsection are based on the paper "Understanding Anomalies: Visualizing Sensor Data for Condition Monitoring of Manufacturing Machines" [Olivotti and Eilers, 2018]. After the application of the developed artifact in artificial scenarios and in a real-world industry context for improving the intuitive model evaluation, the idea is now to further validate the utility of the idea in different contexts. Since the idea tries to visualize model errors to improve the performance with human domain knowledge, we are not limited to apply it to model errors which are the reasons of possible misspecifications but also to anomalies which arise over time in the data source. One example is the analysis of sensor data from manufacturing machines. Sensor data can be used for example in predictive maintenance applications where the generated information is used to predict future failures of machines. These models can be used to prevent these predicted failures by an early maintenance of the affected machines. There exists a complete literature stream about condition monitoring and predictive maintenance. Here I refer to an inspiring overview of condition based maintenance and time based maintenance techniques [Ahmad and Kamaruddin, 2012].

One challenge is to use sensor data to identify the reason for a certain change in the operation of the machine. Predicting future failures is in and of itself an important task but to identify and solve possible problems which are the reason for a future failure is even more challenging and requires domain knowledge and experience. The initial situation is therefore similar to the previously described problem. Therefore, we apply the following procedure.

1. We collect sensor time series data from a demonstration machine. In addition, we measure values which indicate the operation of the machine, like torque.

2. We develop a one step ahead time series forecast application based on ANNs which learns the patterns for a normal operation of a machine.

3. These forecasting models are then applied during real life operation and the difference between the predication and the measured values are used as the basis for the already introduced residual analysis.

4. After calculating the residuals, the heat map approach is used to visualize the errors depending on the measured sensor data. The idea is therefore to link operation deviations with deviations in the sensor time series.

An example of the torque time series can be seen in Figure 19.

The forecasting procedure is comprehensively discussed in [Olivotti and Eilers, 2018]. In this dissertation I present the resulting heat map in Figure 20. Two

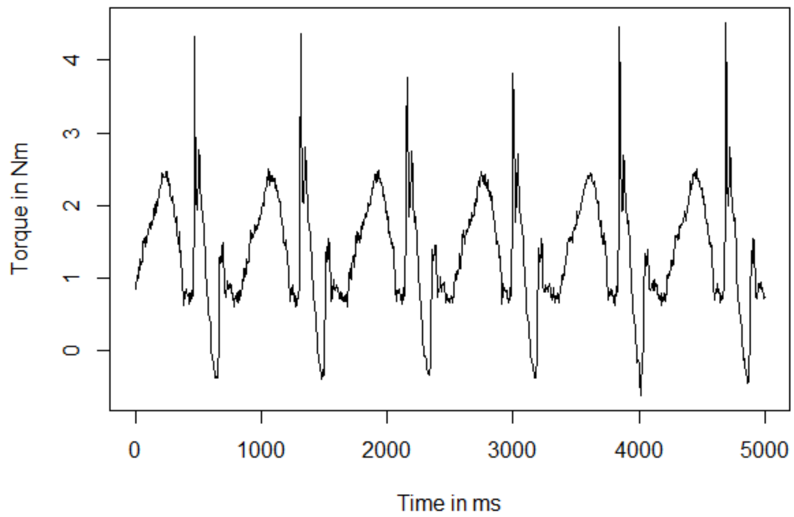
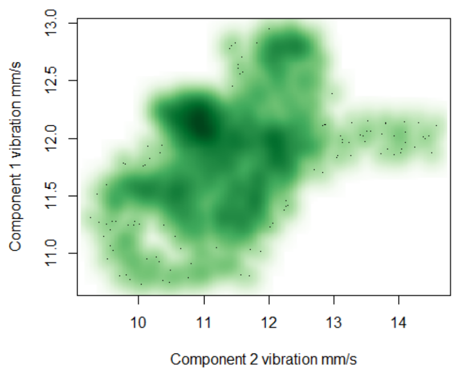
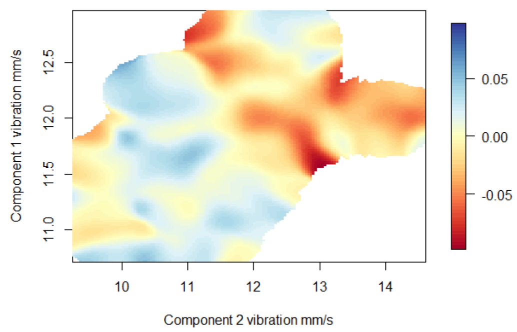


Figure 19: Time Series of Torque in Newton Metre (Nm)

simulated vibration sensors of two different components are plotted against each other. The color scale represents the forecasting error of the torque time series in Nm. Blue regions represent high positive deviations (actual value in this state is higher than expected by the model) while red regions represent high negative deviations (actual value in this state is lower than expected by the model).



(a) Scatter Plot of Sensor Data



(b) Heat Map of Forecasting Errors

Figure 20: Heat Map Visualization of Forecast Errors Depending on Sensor Data

The resulting visualization is similar to the presented approaches. The only difference in this application is the way of how to calculate the residuals and how to interpret them. Here, we assume that our models represent the normal operation of a machine and any deviation which shows systematic patterns are a hint for a change in the productivity of the machine and can be analyzed in an early stage by means of the heat map approach. Figure 20b shows for example that the residuals are high for higher vibrations. In this case this might be an obvious finding that increased vibration causes higher deviations from the usual machine operation, but since we are now able to visualized two different sensors for example from two different components and link them to an intuitive measure for behavior of the machine, domain experts can discuss the reasons for certain deviations on a much simpler level without understanding the underlying machine learning models. This also enables the user to investigate developments over time by comparing the appearance of the same heat map at different points in time.

The idea is to use this application as a further proof of the utility of the heat map approach. But again, the success of this case must be evaluated properly. In the current state of our research we develop ideas of how to measure the utility for this specific application. If we are able to prove the utility, this in turn would strengthen the evaluation part of the heat map method itself. In the following overall discussion, I will present possible ideas, strengths and weaknesses of certain approaches.

5.7 Conclusion, Discussion and Outlook

The papers presented in this section are the consequence of the findings from previous research. The idea is to think about new ways for incorporating domain knowledge in data science. Based on a design science oriented research approach, we have developed a heat map based visualization technique which has shown to provide value for pure machine learning applications, real industry applications where it is important to incorporate human domain knowledge for a successful FE and anomaly detection or understanding of reasons for certain anomalies.

The problem of the evaluation is already discussed but a general question remains. How important is domain knowledge really? The complete section in this work is based on the assumption that domain knowledge actually plays a major role in data science. Therefore, one can argue that the ideas I have presented are the second step before the first. Based on our experience and the overall general agreement that domain knowledge is important, the assumptions might be valid to justify the research effort. But one important step for further research must be to conduct a study on critical success factors for data science projects. Currently no comprehensive study exists which provides quantitative or qualitative evidence for the impact of domain knowledge on the outcome of data science projects. Maybe it heavily depends on the application, hence generalizable evidence is impossible. But from my point of view on the current state of the data science literature, it is worth to develop a research design which addresses this question.

Another important aspect is the derivation of the solution. Based on an iterative literature review, starting from the problem statement, over OVB, FE, domain knowledge to visualization, we have developed the presented idea for our problems at hand. But this is definitely not the global optimum of how to solve the described problems. But following a design science-oriented approach explicitly allows us to develop "satisfactory solutions" [Hevner et al., 2004]. The question remains who defines satisfactory. The approach we follow is from an epistemological point of view the idea of continuous falsification. The developed approach must be tested and evaluated over and over again to refine it or maybe to disprove its suitability. The presented applications are by far not enough. This is a time-consuming process of further research.

A further outlook is to concentrate more on the topic of FE which is currently underrepresented in the scientific literature while it plays a major role in the practice-oriented data science community. An idea can be to use again a design science research approach to develop an artifact which in this case are design principles of how to engineer proper features. The result could be a process model for example.

5.8 Reviewer Opinions, Contribution and Critical Comment

In this section again I will critically reflect the overall contribution of my work, discuss the impact and comment on different reviewer opinions from the submission process of the presented papers.

As already mentioned above, the methodological approach of the presented research can be considered controversial. The first reviewer of the ICIS paper states the following:

"The methodology is quite sound and I particularly appreciate the candid nature that the paper revealed preprocessing choices such as outlier removal, and more difficult decisions".

The second reviewer is more critical about the methodological execution of the research:

"While apparently the author uses design science method in the paper, the design evaluation methods apparently were not well-executed. As Hevner et al. (2008) suggested that the utility, quality, and efficacy of a design artifact in the design science research must be rigorously demonstrated via well executed evaluation methods starting from observational, analytical, experimental, testing, and descriptive. The authors did quite well on the observational, analytical, however less clear on the experimental, testing, and descriptive. The author has clearly stated the problems and IT artifact (solutions), however, the evaluation method is insufficient (March and Storey, 2008). For example, the author did not demonstrate the feasibility of heat maps compared to the available alternatives of visualizations techniques. Recall, The author proposes the use of heat maps as the proper visualization technique to facilitate the communication between data scientist and domain experts. However, the author did not demonstrate the extent to which heat maps can facilitate such communication."

We see this comment as a valuable hint for the future development of the ideas. The critique about the comparisons is definitely justified since the requirement for a valid design-oriented research process is to demonstrate that no other solution exists in the knowledge base, as mentioned before. This point must be improved in future research. Even if we have argued based on the current literature that there exists no solution approach, maybe other already existing techniques can be applied to this kind of problem as well which makes our developed technique unnecessary. But comparing different techniques is a challenging task since it is already difficult to measure success or quantify the utility of these methods in a specific project.

But the reviewer also mentioned a possible idea of how to handle this problem.

"The paper, however, should be more interesting when the author can experiment and test the proposed visualization via controlled experiment or simulation. For example, apart from conducting experiments for feature engineering using artificial data, the author could investigate the proposed visualizations in a controlled environment to ensure whether the visualizations can provide intuitive views and familiar views of the machine learning results to the management. Additionally, the author could experiment the proposed visualization to data scientist. The author, therefore, can show the effectiveness or the usability of the proposed visualization in mitigating the gap between data scientist and domain experts. In such a way, this research will make a much clearer contribution to practice."

The idea of controlled experiments would allow to explicitly compare different visualization techniques and exactly quantify the utility. But the question is how to design such an experimental study. One could think of different data science experts which are randomly assigned to groups with different toolsets they are allowed to use for the analysis of a data set. But also controlled experiments have weaknesses compared to the methods presented in this work. Usually experiments suffer from a lower external validity. Is the experimental setup comparable to real world situations? Especially data science projects are hard to be investigated by controlled experiments since the analysis is typically a longer process. Breaking down the task for data scientist, such that a laboratory experiment is possible, can be conducted but would decrease the external validity of the results tremendously.

It becomes clear that there is no perfect solution which leads to the question if it is maybe rigorous enough to propose different ideas which have worked in certain applications. The reviewers of our paper strongly disagree on that point. While the above mentioned reviewer was rather critical regarding the methodology, another reviewer of the ICIS paper represents the opposite opinion.

"Thank you for your contribution. Overall this was an extremely well-written

submission, with the added benefit of having both an interesting topic area and excellent application. The paper is relevant to all three overarching themes of the track in its importance to data science, decision support and information visualization. [...] The introduction builds a solid foundation for the paper. We quickly appreciate the generalized problem of a gap between professionals, and using visualization as a means for allowing bilateral understanding and two-way dialogue between data scientists and managers taking business decisions. Since visualization's two main roles by definition are understanding and explaining, the purpose is well founded: that both parties can explore and communicate with each other given the right visualization method. [...] Overall I appreciated the conscious visual encoding effort to remain intuitive, without sacrificing complexity. The quote from Cairo reminds me a principle in The Truthful Art on maintaining complexity, by prioritizing clarity (as shown here) over simplicity (which implies constant stripping away of data). I particularly enjoyed the effort to use visualization as a vehicle for cross-discipline coordination, rather than a means for high-level communication within (management) teams. The positive effects for organizations is clear. For the data scientist it offers affordances related to calibration of the model, and a powerful feedback loop potential within their role. And for the management there is an opportunity to un-black-box models and alleviate mistrust."

The comment focuses on the developed visualization itself rather than on the methodology and evaluation. Again, the relevance and idea are pointed out as the major strength of the paper.

Compared to the other research areas presented in this work, the idea presented in this section can be the most relevant one from my point of view. The relevance and impact on research and practice was also not questioned by any reviewer of all three papers. I would argue that the contribution is not only the visualization technique itself but also the identified research stream and the combination of FE and Visualization. It also tries to justify why we need information systems research in the data science community since we have the tradition of analyzing sociotechnical systems. Therefore, the question of how to integrate machine learning techniques, experts in quantitative methods and domain experts is a suitable question for information system researchers which is not addressed by any other research field. Nevertheless, the greatest weakness is the question of scientific rigor. One can argue that it is just a nice idea applied to some very specific problems. But I think research also has the duty to provide ideas which afterwards can be applied and tested in practice. If we always wait to communicate our research results and ideas until the evidence about the utility and validity is certain, it would be impossible to generate a real impact and think ahead.

The overall contribution can be summarized as follows: **Contribution: We identify the research gap of human/domain knowledge-oriented FE in the information systems research literature. Based on a comprehensive literature analysis in various fields like visualization and communication, we design an artifact (visualization technique) which should be able to better incorporate human domain knowledge in data science. The artifact is applied in different scenarios like resale price forecasting and anomaly detection with sensor data to evaluate the utility.**

6 Overall Discussion, Limitations and Further Research Agenda

It is not about getting the answer right but about making progress in answering the right questions that will accelerate the progress of our field and the contributions we collectively make in areas of broader impact."

[Rai, 2017] in Management Information Systems Quarterly

6.1 Future Research Questions and How to Find Them

Each part in this dissertation already provides an outlook to different further research opportunities. In this section, however, I will present more general ideas of possible research directions based on the findings of the presented articles.

Information systems research is by definition an interdisciplinary field. Thinking about what distinguishes us from other disciplines leads to the conclusion that exactly this interdisciplinary is our competitive advantage. Therefore, generating ideas about how to integrate knowledge of different expert groups like data science experts and domain experts is a suitable field for information systems research. This also leads to the idea of how to integrate machine knowledge and human expertise. In recent years a whole stream of literature has emerged which deals with such problems. A recent paper by [Nushi et al., 2017] provides an interesting idea of how human intellect can be used for troubleshooting of machine learning systems. They demonstrate their human-in-the-loop methodology in a case study with the machine learning task of capturing images. The ideas can be summarized as Hybrid Intelligence. [Kamar, 2016] provides an overview of state-of-the-art approaches. The paper also illustrates possible further research directions. One interesting idea is the approach of developing new machine learning strategies which incorporate the findings from teamwork literature. She explicitly recommends "to develop new representations and decision-making approaches that can reason about the hybrid nature of human-computer teamwork". This again shows the huge potential of information systems research in this field since we have a research tradition in Human-Computer-Interaction, technology acceptance and the design of technological artifacts which are human-oriented.

Hybrid intelligence also imply that humans working together with such systems accept and trust them. These questions are especially suitable for information systems researchers. How systems must be designed to be accepted by their human co-workers? How much trust and mistrust in such systems is optimal for an optimal decision making? In general, which factors lead to trust and acceptance in machine learning models? How can we prevent that people see these systems or machines as competitors rather than actual colleges/co-workers? What are the consequences of mistrust but also, what are the consequences of too much trust (a blind reliance on machines without questioning)? Another idea for further research is to find out which factors actually determine the success of such hybrid systems or data analytics projects in general. Again, information systems research has a long tradition and well-developed methods to analyze critical success factors for certain systems. One idea is to measure the actual influence of domain knowledge of people on a data science project. Is domain knowledge really that important? And are there other hidden factors which can improve the outcome of such projects? For example, which role does creativity plays?

The thought about creativity closes this subsection but leads to another general remark about how research is understood and conducted in this dissertation. This work does not contain any kind of structured literature analysis. The reason for this is that my understanding of research is embossed by the idea of creativity. There already exist research paper which demonstrate how to support literature analyzes with machine learning [Koukal et al., 2014]. One can argue that these systems provide support for the subsequent creative work of defining research gaps for example. But these systems are limited (1) by the scope of the database and (2) by their ability to construct new connections between papers where previously no connection was evident. Again, this is not a criticism of structured literature reviews themselves. But the question for which purpose we do a literature analysis. Structured literature reviews are important to better understand the state-of-the-art and to reveal potential research gaps within a research stream. It is important to build new ideas "on the shoulders of giants" [vom Brocke et al., 2015]. But finding research gaps within a stream of literature can already be done by automated systems as well as categorizing papers into different topics. From my point of view, the real intellectual challenge is rather to connect different ideas from many different research streams to create something new based on human creativity. Therefore, studying the literature is of tremendous importance but structured literature analyses are (1) not a task which necessarily must be done by humans or which requires human action and exploratory spirit and (2) do not necessarily lead to interdisciplinary thinking beyond the borders of a certain subfield. In contrast my research results are all based on combinations of ideas from different subfields. For example, combining seasonal effect on financial markets with a decision support algorithm which is based on RL or using visualizations for FE. Both examples show that the combined ideas were not included in any structured literature review about the other subfield. I want to emphasize that my research is just another approach of doing research which has a raison d'être and should not be interpreted as a criticism of other research approaches.

6.2 Critical Comment and General Limitations

When I started my studies for this dissertation, the idea of making real impact with my research was the main driver of my motivation. My goal was to find topics which are worth working on. Interesting and relevant for a broad readership. But soon I found out that even within the information systems community there is an ongoing debate if our discipline is relevant at all. Screening the track descriptions of the International Conference on Information Systems 2018 provides a first impression of topics which are currently discussed in our community. One track is entitled "Practice-Oriented Demand-driven IS"[7]. My first thought was, should not every single track have exactly the purpose to seek for practice-oriented and demand driven research? The first sentences of the track description are as follows: "The field of IS as most other research fields have become obsessed with writing publications, because we are measured on number of publications. Irrespectively of whether they have value to any of our stakeholders (students, industry or society). As New York Times wrote recently, we are not doing research; we have become 'paper writing factories' turning out papers that fewer and fewer reads. The mindless strive for theory instead of real knowledge, rigor instead of relevance and citations rather than impact/use is not sustainable." Of course, this is just one opinion about the research practice. As mentioned already in the introduction [Agarwal and Dhar, 2014] for example have an almost enthusiastic view on the field of information systems research when it comes the question of how to define our role in the era of big data, even though there is also no critical reflection in their work. They just state that information systems research has a competitive advantage compared to other disciplines. But it demonstrates that there exist different opinions on such topics about our research field itself.

I mention these thoughts in this this discussion, because even after finishing my studies for this dissertation, I am not sure how to think about this problem. My impression after writing seven accepted publications is that indeed the critique is justified and necessary. As [Ioannidis, 2005] already states in his widely recognized (5544 citations in February 2018 based on Google Scholar) proof of why most published research findings are false, conflicts of interest and prejudges may increase bias. He states that "Many otherwise seemingly independent, university-based studies may be conducted for no other reason than to give physicians and researchers qualifications for promotion or tenure." So, are we really driven by the motivation of doing research which has an impact, or are we biased by financial and nonfinancial incentives which let us become "paper writing factories"? In this discussion I

[7] `http://icis2018.aisnet.org/?page_id=111` accessed February 2. 2018

can only speak for myself and against this background I have to critically reflect my work. An honest answer to the question of the real motivation for each of my presented papers must indeed include the tremendous pressure of getting published in high ranked journals and conference proceedings to fulfill the requirements for this dissertation. Does that mean each paper is ultimately a product of the pressure to publish and have no intrinsic value? At least I hope this is not the case. Each paper is the result of a long process of thinking about real problems and how to solve them. But the nonfinancial incentive of getting published affects the way of conducting and writing down the research. By reading more and more already published research articles, one develops an idea of how to write down own research in such a way that reviewers and editors argue in favor for the paper. In this dissertation for example I have discussed the design science-oriented research approach for the heat map visualization method in Section 5. One gets the impression that any research has to follow a clear methodology, a rigorously executed step-by-step process which results at the end in a well-designed artifact. But observing a certain problem in a real-world data science application, thinking about a possible solution and implement it seems to be not enough to be called research. To get ideas published, a proper methodology has to be used and discussed. This should not be misunderstood as a criticism of rigorous research. The question is rather what we define as rigorous research. Depending on the topic, research must be carried out rigorous enough to be relevant. But rigorous applied methods cannot excuse irrelevant or bad ideas. Answering wrong questions using right methods is called Type III error [Rai, 2017]. At least I have tried to avoid this kind of error in all my published articles. But do we always want relevant and impactful articles? Is not research without a purpose the most perfect research? So, are there any wrong questions at all? In this critical reflection I am still not able to come to a conclusion. Regardless of this discussion, it cannot hurt if the research is made available to a wider audience. Therefore, to close this section I would like to point out an interesting approach of how to incorporate practitioners into the research process and result communication. One example is the European Journal of Information Systems. The idea is here to facilitate a better dialog between research and practice. "In order for researchers to influence practice, they need to engage in dialog with practitioners. It is as simple as that." [Te'eni et al., 2017]. Articles which are accepted for publication can be selected for a Science-to-practice Editorial Process. Together with the platform `http://www.science2practice.org/` a broader audience should be reached without lowering scientific standard.

7 Conclusion

It is, in fact, nothing short of a miracle that the modern methods of instruction have not yet entirely strangled the holy curiosity of inquiry: for this delicate little plant, aside from stimulation, stands mainly in need of freedom: without this it goes to wrack and ruin without fail.

Albert Einstein

In this cumulative dissertation four research topics within the broad field of data analytics were discussed. The first part shows the implementation of a decision support algorithm based on RL which was applied in a trading system. In the second part a text mining approach was presented to analyze customer reviews from online shops and social media platforms like YouTube. The third part introduces a forecasting system for resale prices of used cars and the fourth part show how a developed visualization technique can help to better incorporate human domain knowledge in this data science process.

The main purpose of this work is to provide a critical reflection about the research process and results during my studies. Overall, the main contribution is the idea of using visualization as a tool for a domain knowledge-oriented FE. The application within the resale price forecast project provides the opportunity to evaluate and refine the idea within a real-world data science application. We show that in such complex projects the human knowledge and expertise still plays a major role which can be fostered by the right techniques and tools. How important human domain knowledge and other human factors like creativity actually are, in general or in different applications, is a question for future research.

References

[Agarwal and Dhar, 2014] Agarwal, R. and Dhar, V. (2014). Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, 25(3):443–448.

[Ahmad and Kamaruddin, 2012] Ahmad, R. and Kamaruddin, S. (2012). An overview of time-based and condition-based maintenance in industrial application. *Computers & Industrial Engineering*, 63(1):135–149.

[Ariel, 1987] Ariel, R. A. (1987). A monthly effect in stock returns. *Journal of Financial Economics*, 18(1):161–174.

[Atanasova and Hudson, 2010] Atanasova, C. V. and Hudson, R. S. (2010). Technical trading rules and calendar anomalies – Are they the same phenomena? *Economics Letters*, 106(2):128–130.

[Baetje, 2018] Baetje, F. (2018). Does a lot help a lot? Forecasting stock returns with pooling strategies in a data-rich environment. *Journal of Forecasting*, 37(1):37–63.

[Bengio et al., 2013] Bengio, Y., Courville, A., and Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.

[Borland and Taylor, 2007] Borland, D. and Taylor, R. M. (2007). Rainbow Color Map (Still) Considered Harmful. *IEEE Computer Graphics and Applications*, 27(2):14–17.

[Breiman, 2001] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

[Bresciani and Eppler, 2009] Bresciani, S. and Eppler, M. J. (2009). The Benefits of Synchronous Collaborative Information Visualization: Evidence from an Experimental Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1073–1080.

[Broock et al., 1996] Broock, W. A., Scheinkman, J. A., Dechert, W. D., and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3):197–235.

[Cadsby and Ratner, 1992] Cadsby, C. B. and Ratner, M. (1992). Turn-of-month and pre-holiday effects on stock returns: Some international evidence. *Journal of Banking & Finance*, 16(3):497–509.

[Chen et al., 2012] Chen, H., Chiang, R. H., and Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly*, 36(4):1165–1188.

[Clarke, 2005] Clarke, K. A. (2005). The Phantom Menace: Omitted Variable Bias in Econometric Research. *Conflict Management and Peace Science*, 22(4):341–352.

[Cleveland et al., 1990] Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73.

[Coates et al., 2011] Coates, A., Ng, A., and Lee, H. (2011). An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 215–223.

[Coates and Ng, 2012] Coates, A. and Ng, A. Y. (2012). Learning Feature Representations with K-Means. In *Neural Networks: Tricks of the Trade*, Lecture Notes in Computer Science, pages 561–580. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-35289-8_30.

[Desyllas and Sako, 2013] Desyllas, P. and Sako, M. (2013). Profiting from business model innovation: Evidence from Pay-As-You-Drive auto insurance. *Research Policy*, 42(1):101–116.

[Dickinson and Peterson, 1995] Dickinson, A. and Peterson, D. R. (1995). Expectations of weekend and turn-of-the-month mean return shifts implicit in index call option prices. *Journal of Financial and Strategic Decisions*, 8(3):69–76.

[Domingos, 2012] Domingos, P. (2012). A Few Useful Things to Know About Machine Learning. *Commun. ACM*, 55(10):78–87.

[Dress et al., 2017] Dress, K., Lessmann, S., and von Mettenheim, H.-J. (2017). Residual Value Forecasting Using Asymmetric Cost Functions. *arXiv:1707.02736 [q-fin, stat]*. arXiv: 1707.02736.

[Dunis and Huang, 2002] Dunis, C. L. and Huang, X. (2002). Forecasting and trading currency volatility: an application of recurrent neural regression and model combination. *Journal of Forecasting*, 21(5):317–354.

[Eilers and Breitner, 2017] Eilers, D. and Breitner, M. (2017). A Picture is Worth a Thousand Words: Visual Model Evaluation in Data Science Applications. *Wirtschaftsinformatik 2017 Proceedings*.

[Eilers et al., 2014] Eilers, D., Dunis, C. L., von Mettenheim, H.-J., and Breitner, M. H. (2014). Intelligent trading of seasonal effects: A decision support algorithm based on reinforcement learning. *Decision Support Systems*, 64:100–108.

[Eilers et al., 2017] Eilers, D., Köpp, C., Gleue, C., and Breitner, M. (2017). It's not a Bug, it's a Feature: How Visual Model Evaluation can help to incorporate Human Domain Knowledge in Data Science. *ICIS 2017 Proceedings*.

[Eirinaki et al., 2012] Eirinaki, M., Pisal, S., and Singh, J. (2012). Feature-based opinion mining and ranking. *Journal of Computer and System Sciences*, 78(4):1175–1184.

[Endert et al., 2014] Endert, A., Hossain, M. S., Ramakrishnan, N., North, C., Fiaux, P., and Andrews, C. (2014). The human is the loop: new directions for visual analytics. *Journal of Intelligent Information Systems*, 43(3):411–435.

[Epanechnikov, 1969] Epanechnikov, V. (1969). Non-Parametric Estimation of a Multivariate Probability Density. *Theory of Probability & Its Applications*, 14(1):153–158.

[Fama, 1970] Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383–417.

[Fan Hongqin et al., 2008] Fan Hongqin, AbouRizk Simaan, Kim Hyoungkwan, and Zaïane Osmar (2008). Assessing Residual Value of Heavy Construction Equipment Using Predictive Data Mining Model. *Journal of Computing in Civil Engineering*, 22(3):181–191.

[French, 1980] French, K. R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics*, 8(1):55–69.

[Ghiassi et al., 2016] Ghiassi, M., Zimbra, D., and Lee, S. (2016). Targeted Twitter Sentiment Analysis for Brands Using Supervised Feature Engineering and the Dynamic Architecture for Artificial Neural Networks. *Journal of Management Information Systems*, 33(4):1034–1058.

[Gleue et al., 2017] Gleue, C., Eilers, D., Mettenheim, H.-J. v., and Breitner, M. (2017). Decision Support for the Automotive Industry: Forecasting Residual Values using Artificial Neural Networks. *Wirtschaftsinformatik 2017 Proceedings*.

[Gleue et al., 2018] Gleue, C., Eilers, D., Mettenheim, H.-J. v., and Breitner, M. H. (2018). Decision Support for the Automotive Industry. *Business & Information Systems Engineering*, pages 1–13.

[Hansen et al., 2005] Hansen, P. R., Lunde, A., and Nason, J. M. (2005). Testing the Significance of Calendar Effects. SSRN Scholarly Paper ID 388601, Social Science Research Network, Rochester, NY.

[Heaton, 2016] Heaton, J. (2016). An empirical analysis of feature engineering for predictive modeling. In *SoutheastCon 2016*, pages 1–6.

[Heaton, 2017] Heaton, J. (2017). *Automated Feature Engineering for Deep Neural Networks with Genetic Programming.* PhD thesis, Nova Southeastern University.

[Hevner et al., 2004] Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28(1):75–105.

[Hu and Liu, 2004] Hu, M. and Liu, B. (2004). Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

[Ioannidis, 2005] Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Med*, 2(8):696–701.

[Ionescu et al., 2016] Ionescu, L., Gwiggner, C., and Kliewer, N. (2016). Data Analysis of Delays in Airline Networks. *Business & Information Systems Engineering*, 58(2):119–133.

[Jaffe and Westerfield, 1989] Jaffe, J. and Westerfield, R. (1989). Is there a monthly effect in stock market returns?: Evidence from foreign countries. *Journal of Banking & Finance*, 13(2):237–244.

[Kamar, 2016] Kamar, E. (2016). Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence. In *IJCAI*, pages 4070–4073.

[Kanter and Veeramachaneni, 2015] Kanter, J. M. and Veeramachaneni, K. (2015). Deep feature synthesis: Towards automating data science endeavors. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–10.

[Keim et al., 2008] Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., and Melançon, G. (2008). Visual Analytics: Definition, Process, and Challenges. In Kerren, A., Stasko, J. T., Fekete, J.-D., and North, C., editors, *Information Visualization*, number 4950 in Lecture Notes in Computer Science, pages 154–175. Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-70956-5_7.

[Koukal et al., 2014] Koukal, A., Gleue, C., and Breitner, M. (2014). Enhancing Literature Review Methods - Evaluation of a Literature Search Approach based on Latent Semantic Indexing. *ICIS 2014 Proceedings*.

[Kuhn and Johnson, 2013] Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*. Springer Science & Business Media.

[Lakonishok and Smidt, 1988] Lakonishok, J. and Smidt, S. (1988). Are Seasonal Anomalies Real? A Ninety-Year Perspective. *The Review of Financial Studies*, 1(4):403–425.

[Lam, 2004] Lam, M. (2004). Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. *Decision Support Systems*, 37(4):567–581.

[Lash and Zhao, 2016] Lash, M. T. and Zhao, K. (2016). Early Predictions of Movie Success: The Who, What, and When of Profitability. *Journal of Management Information Systems*, 33(3):874–903.

[Lee, 1982] Lee, L.-F. (1982). Specification error in multinomial logit models: Analysis of the omitted variable bias. *Journal of Econometrics*, 20(2):197–209.

[Leigh et al., 2002] Leigh, W., Purvis, R., and Ragusa, J. M. (2002). Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support. *Decision Support Systems*, 32(4):361–377.

[Lessmann et al., 2010] Lessmann, S., Listiani, M., and Voß, S. (2010). Decision support in car leasing: A forecasting model for residual value estimation. *ICIS 2010 Proceedings*.

[Lian et al., 2003] Lian, C., Zhao, D., and Cheng, J. (2003). A fuzzy logic based evolutionary neural network for automotive residual value forecast. In *International Conference on Information Technology: Research and Education, 2003. Proceedings. ITRE2003*, pages 545–548.

[Liu, 2012] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

[Liu and Zhang, 2012] Liu, B. and Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer.

[Lucca and Moench, 2015] Lucca, D. O. and Moench, E. (2015). The Pre-FOMC Announcement Drift. *The Journal of Finance*, 70(1):329–371.

[Maaten and Hinton, 2008] Maaten, L. v. d. and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605.

[March and Storey, 2008] March, S. and Storey, V. (2008). Design Science in the Information Systems Discipline: An Introduction to the Special Issue on Design Science Research. *Management Information Systems Quarterly*, 32(4):725–730.

[Nushi et al., 2017] Nushi, B., Kamar, E., Horvitz, E., and Kossmann, D. (2017). On Human Intellect and Machine Failures: Troubleshooting Integrative Machine Learning Systems. In *AAAI*, pages 1017–1025.

[Ogden, 1990] Ogden, J. P. (1990). Turn-of-Month Evaluations of Liquid Profits and Stock Returns: A Common Explanation for the Monthly and January Effects. *The Journal of Finance*, 45(4):1259–1272.

[Olivotti and Eilers, 2018] Olivotti, D. and Eilers, D. (2018). Understanding Anomalies: Visualizing Sensor Data for Condition Monitoring of Manufacturing Machines. In *Multikonferenz Wirtschaftsinformatik*.

[Pang and Lee, 2008] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

[Pang et al., 2002] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Poria et al., 2016] Poria, S., Cambria, E., and Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49.

[Poria et al., 2017] Poria, S., Peng, H., Hussain, A., Howard, N., and Cambria, E. (2017). Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. *Neurocomputing*, 261:217–230.

[Prado, 2009] Prado, S. (2009). The European used-car market at a glance: Hedonic resale price valuation in automotive leasing industry. *Economics Bulletin*, 29(3):2086–2099.

[Prado and Ananth, 2012] Prado, S. and Ananth, R. (2012). Breaking Through Risk Management, a Derivative for the Leasing Industry. *Journal of Financial Transformation*, 34:211–218.

[Qazi et al., 2017] Qazi, A., Raj, R. G., Hardaker, G., and Standing, C. (2017). A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges. *Internet Research*, 27(3):608–630.

[Rai, 2017] Rai, A. (2017). Editor's Comments: Avoiding Type III Errors: Formulating IS Research Problems that Matter. *Management Information Systems Quarterly*, 41(2):iii–vii.

[Ramsey, 1969] Ramsey, J. B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2):350–371.

[Rode et al., 2002] Rode, D. C., Fischbeck, P. S., and Dean, S. R. (2002). Residual Risk and the Valuation of Leases Under Uncertainty and Limited Information. *The Journal of Structured Finance*, 7(4):37–49.

[Rozeff and Kinney, 1976] Rozeff, M. S. and Kinney, W. R. (1976). Capital market seasonality: The case of stock returns. *Journal of Financial Economics*, 3(4):379–402.

[Sermpinis et al., 2012] Sermpinis, G., Dunis, C., Laws, J., and Stasinakis, C. (2012). Forecasting and trading the EUR/USD exchange rate with stochastic Neural Network combination and time-varying leverage. *Decision Support Systems*, 54(1):316–329.

[Smith and Jin, 2007] Smith, L. D. and Jin, B. (2007). Modeling exposure to losses on automobile leases. *Review of Quantitative Finance and Accounting*, 29(3):241–266.

[Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

[Storchmann, 2004] Storchmann, K. (2004). On the Depreciation of Automobiles: An International Comparison. *Transportation*, 31(4):371–408.

[Sullivan et al., 2001] Sullivan, R., Timmermann, A., and White, H. (2001). Dangers of data mining: The case of calendar effects in stock returns. *Journal of Econometrics*, 105(1):249–286.

[Sun et al., 2013] Sun, G.-D., Wu, Y.-C., Liang, R.-H., and Liu, S.-X. (2013). A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges. *Journal of Computer Science and Technology*, 28(5):852–867.

[Tam et al., 2017] Tam, G. K. L., Kothari, V., and Chen, M. (2017). An Analysis of Machine- and Human-Analytics in Classification. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):71–80.

[Te'eni et al., 2017] Te'eni, D., Seidel, S., and Brocke, J. v. (2017). Stimulating dialog between information systems research and practice. *European Journal of Information Systems*, 26(6):541–545.

[Thomas and Cook, 2006] Thomas, J. J. and Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13.

[Timmerman, 2003] Timmerman, M. E. (2003). Principal Component Analysis. *Journal of the American Statistical Association*, 98(464):1082–1083.

[Tsytsarau and Palpanas, 2012] Tsytsarau, M. and Palpanas, T. (2012). Survey on Mining Subjective Data on the Web. *Data Min. Knowl. Discov.*, 24(3):478–514.

[Turney, 2002] Turney, P. D. (2002). Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.

[vom Brocke et al., 2015] vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. (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(1).

[Wiegard et al., 2017] Wiegard, R., Eilers, D., and Gercke, D. (2017). What does YouTube say about your Product? An Aspect based Approach. *ECIS 2017*.

[Wilson et al., 2005] Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural*

Language Processing, HLT '05, pages 347–354, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Wong, 2011] Wong, B. (2011). Points of view: Color blindness. *Nature Methods*, 8(6):441–441.

[Wu et al., 2009] Wu, J.-D., Hsu, C.-C., and Chen, H.-C. (2009). An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. *Expert Systems with Applications*, 36(4):7809–7817.

[Yu et al., 2010] Yu, H.-F., Lo, H.-Y., Hsieh, H.-P., Lou, J.-K., McKenzie, T. G., Chou, J.-W., Chung, P.-H., Ho, C.-H., Chang, C.-F., Wei, Y.-H., Weng, J.-Y., Yan, E.-S., Chang, C.-W., Kuo, T.-T., Lo, Y.-C., Chang, P. T., Po, C., Wang, C.-Y., Huang, Y.-H., Hung, C.-W., Ruan, Y.-X., Lin, Y.-S., Lin, S.-d., Lin, H.-T., and Lin, C.-J. (2010). Feature Engineering and Classifier Ensemble for KDD Cup 2010. *KDD Cup and Workshop 2010 Proceedings.*

[Zikopoulos et al., 2013] Zikopoulos, P., deRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D., and Giles, J. (2013). *Harness the power of big data: The IBM big data platform.* McGraw Hill Professional, New York.

Appendices

A Intelligent Trading of Seasonal Effects: A Decision Support Algorithm Based on Reinforcement Learning

Citation: Eilers, D., Dunis, C.L., von Mettenheim, H.-J., Breitner, M.H., 2014. Intelligent trading of seasonal effects: A decision support algorithm based on reinforcement learning. Decision Support Systems 64, 100–108.

Link: http://www.sciencedirect.com/science/article/pii/S0167923614001523

DOI: https://doi.org/10.1016/j.dss.2014.04.011

Abstract: Seasonalities and empirical regularities on financial markets have been well documented in the literature for three decades. While one should suppose that documenting an arbitrage opportunity makes it vanish there are several regularities that have persisted over the years. These include, for example, upward biases at the turn-of-the-month, during exchange holidays and the pre-FOMC announcement drift. Trading regularities is already in and of itself an interesting strategy. However, unfiltered trading leads to potential large drawdowns. In the paper we present a decision support algorithm which uses the powerful ideas of reinforcement learning in order to improve the economic benefits of the basic seasonality strategy. We document the performance on two major stock indices.

B Decision Support for the Automotive Industry: Forecasting Residual Values Using Artificial Neural Networks (WI)

Citation: Gleue, C., Eilers, D., Mettenheim, H.-J. von, Breitner, M., 2017. Decision Support for the Automotive Industry: Forecasting Residual Values using Artificial Neural Networks. Wirtschaftsinformatik 2017 Proceedings.

Link: http://aisel.aisnet.org/wi2017/track12/paper/1/

Abstract: The leasing business is one of the most important distribution channels for the automotive industry. This implies that forecasting accurate residual values for the vehicles is a major factor for determining monthly leasing rates: Either a systematic overestimation or underestimation of future residual values can incur large potential losses in resale value or, respectively, competitive disadvantages. In this paper, an operative DSS with the purpose of facilitating residual value related management decisions is introduced, with a focus on its forecasting capabilities. Practical implications are discussed, a multi-variate linear model and an artificial neural network approach are benchmarked and further, the effects of price trends and seasonal influences are investigated. The analysis is based on more than 150,000 data sets from a major German car manufacturer. We show that artificial neural network ensembles with only a few input variables are capable of achieving a significant improvement in forecasting accuracy.

C A Picture Is Worth a Thousand Words: Visual Model Evaluation in Data Science Applications

Citation: Eilers, D., Breitner, M., 2017. A Picture is Worth a Thousand Words: Visual Model Evaluation in Data Science Applications. Wirtschaftsinformatik 2017 Proceedings.

Link: https://www.wi2017.ch/images/wi2017-st_paper_2.pdf

Abstract: Besides programming and mathematical or statistical skills, domain knowledge about the investigated problem is an important factor in data science. In practical applications, however, there often exist a gap between data scientists who have the technical skills in advanced analytics methods and the domain experts like managers and decision makers with substantial knowledge in their field. Less participation of domain experts during the complex model building and evaluation process of the data science pipeline can lead to acceptance problems and prejudices against the results. Moreover, data scientists depend on the domain knowledge of experts when it comes to evaluation and problem identification of their models. In this paper we address this issue by introducing an easytounderstand heat map visualization technique for model evaluation. It enables all parties to discuss the analysis on the same level of complexity. The benefits are demonstrated based on a real world business example.

D What Does YouTube Say about Your Product? An Aspect Based Approach

Citation: Wiegard, R., Eilers, D., Gercke, D., 2017. What does YouTube say about your Product? An Aspect based Approach. European Conference on Information Systems 2017 Proceedings.

Link: http://aisel.aisnet.org/ecis2017_rp/28/

Abstract: Nowadays, customers have a variety of options to gather information about products, which can sup-port their purchasing decisions. More and more customers use YouTube reviews or unboxing videos to get a first impression of different products and interact or discuss with other users in the comment section. Automatically analyzing these comments to gain a better insight about the important product aspects remains a major challenge in the field of social media monitoring because the text data is unstructured and noisier compared to conventional review data for example from Amazon. In this study, we focus on the automated aspect extraction task to answer the question, which characteristics of products are important from the (potential) customer view. We show that YouTube comments are a valuable data source for this purpose with an aspect extraction precision comparable to conventional Amazon reviews. To improve aspect extraction in general, we propose a new aspect sorting method based on Google Trends. Incorporating the search volume of products combined with aspects into the extraction procedure improves the precision results especially for noisier text data. To illustrate the analysis results, we choose Amazon reviews and YouTube comments about three exemplary smartphones.

E It's Not a Bug, It's a Feature: How Visual Model Evaluation Can Help to Incorporate Human Domain Knowledge in Data Science

Citation: Eilers, D., Köpp, C., Gleue, C., Breitner, M., 2017. It's not a Bug, it's a Feature: How Visual Model Evaluation can help to incorporate Human Domain Knowledge in Data Science. International Conference on Information Systems 2017 Proceedings.

Link: http://aisel.aisnet.org/icis2017/DataScience/Presentations/15/

Abstract: The question of how to incorporate human domain knowledge in practical data science projects is still a major challenge. While machine learning tasks are usually carried out by technically skilled data scientists, these analysts do not necessarily have the required domain knowledge concerning a particular business problem to explain certain phenomena. In real-world data science applications, this may result in models that do not adequately reflect relationships in the data. We address this issue by introducing a heat map based technique for model error visualization to facilitate discussions of the results between data scientists and domain experts. By discussing model errors with domain experts during the iterative analysis process, the generated insights can be used for engineering new features (explanatory variables) which better represent the problem and therefore improve the results. We demonstrate the visualization approach based on artificial data and in the context of a real-world industry example.

F Understanding Anomalies: Visualizing Sensor Data for Condition Monitoring of Manufacturing Machines

Citation: Olivotti, D., Eilers, D., 2018. Understanding Anomalies: Visualizing Sensor Data for Condition Monitoring of Manufacturing Machines. Multikonferenz Wirtschaftsinformatik 2018 Proceedings.

Link: http://mkwi2018.leuphana.de/wp-content/uploads/MKWI_309.pdf

Abstract: In a more and more service oriented manufacturing industry new data driven challenges like predictive maintenance arise. For example, machine learning models can use sensor data to predict anomalies during machine operation. Such models usually base on experiences from past data to train these algorithms. However, since components are often used in different machines with different and partly unique or new domains, experiences about mutual interferences are missing. In this study we try to address this issue by introducing a visualization technique for an intuitive anomaly detection which allows domain experts and engineers to monitor the condition of a machine over time. The heat map based visualization highlights unusual operation measurements dependent on different sensor data combinations. With domain and engineering knowledge, the insights can be used to identify case based reasons for a changing behavior. The application is tested with a demonstration machine.

G Decision Support for the Automotive Industry: Forecasting Residual Values Using Artificial Neural Networks (BISE)

Citation: Gleue, C., Eilers, D., Mettenheim, H.-J. von, Breitner, M., (forthcoming). Decision Support for the Automotive Industry: Forecasting Residual Values using Artificial Neural Networks. Business & Information Systems Engineering.

Link: https://link.springer.com/article/10.1007/s12599-018-0527-3

DOI: https://doi.org/10.1007/s12599-018-0527-3

Abstract: In the automotive industry, it is very common for new vehicles to be leased rather than sold. This implies forecasting an accurate residual value for the vehicles, which is a major factor for determining monthly leasing rates. Either a systematic overestimation or underestimation of future residual values can incur large potential losses in resale value or, respectively, competitive disadvantages. In this paper an operative Decision Support System is introduced for the purpose of facilitating residual value related management decisions with emphasis on its forecasting capabilities. We demonstrate the use of Artificial Neural Networks for this application in a case study based on more than 250,000 data sets of completed leasing contracts between 2011 and 2017 from a major German car manufacturer. The importance of determining price factors and the effect of different time horizons on forecasting accuracy are investigated and practical implications are discussed. In addition, we neither found a significant explanatory nor predictive power of external economic factors, which underlines the importance of collecting and taking advantage of vehicle-specific data or, in more general terms, the exclusive data of corporations, which is often only available internally.