


**Datenanalysemethoden mit Anwendungen  
in der Akzeptanzforschung, bei Finanzmarktprognosen  
und der natürlichen Sprachverarbeitung**

Von der Wirtschaftswissenschaftlichen Fakultät der  
Gottfried Wilhelm Leibniz Universität Hannover  
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften  
– Doktor rerum politicarum –

genehmigte Dissertation  
von

Diplom-Ökonom Rouven-Benjamin Wiegard  


2019

Betreuer und Gutachter:	Prof. Dr. Michael H. Breitner
Weiterer Gutachter:	Prof. Dr. J.-Matthias Graf von der Schulenburg
Vorsitzender der Prüfungskommission:	Prof. Dr. Stefan Wielenberg
Weiteres Mitglied (beratend)	Dr. Ute Lohse
Tag der Promotion:	15.05.2019

# I. Abstract / Zusammenfassung

---

Abstract: In times of increasing digitalization and advanced analytics, the use of data will enable the development and delivery of cutting-edge services, innovative business models and management approaches. The key is to analyze the overwhelming amount of data available to organizations, companies, and science and interpret their implications correctly. The interdisciplinary scientific field of data science uses algorithms, processes and systems as well as scientific methods for the extraction of information and patterns. The present cumulative dissertation deals with different data analysis methods and shows application examples from acceptance research, financial market forecasting and natural language processing. Lack of user acceptance is an obstacle to the success of new information systems of progressive digitization. In order to analyze the complex relationships of the influencing, different acceptance determinants, multivariate analysis methods like structural equation modeling, are used. Furthermore, artificial neural networks are used to create financial forecasts with their function as a decision-supporting information and optimization function. Finally, this work deals with the automatic processing of text data and proposes a method to improve the text mining for Social Media data. In the respective research areas, the status quo in research and practice is shown, research gaps are identified, and exemplary studies are presented on the basis of concrete application examples. The derived recommendations for action and critical discussions in their respective discipline contribute to future research.

**Keywords:** Data Science, Technology Acceptance Model (TAM), Structural Equation Modeling (SEM), Artificial Neural Networks (ANN), Text Mining, Mobile Payment, Wearable Technologies, Option Price Forecasting, Social Media Data, Online-Reviews.

Zusammenfassung: In Zeiten zunehmender Digitalisierung und fortschrittlicher Analysemethoden, wird die Nutzung von Daten die Entwicklung und Bereitstellung modernster Dienstleistungen, innovativer Geschäftsmodelle und Managementansätze ermöglichen. Der Schlüssel besteht darin, die überwältigende Menge an Daten zu analysieren, die Organisationen, Unternehmen und der Wissenschaft zur Verfügung stehen und deren Auswirkungen richtig zu interpretieren. Das interdisziplinäre Wissenschaftsfeld der Datenwissenschaften (Data Science) nutzt dabei Algorithmen, Prozesse und Systeme sowie wissenschaftliche Methoden zur Extraktion von Informationen und Mustern. Die vorliegende kumulative Dissertation beschäftigt sich mit unterschiedlichen Datenanalysemethoden und zeigt Anwendungsbeispiele aus der Akzeptanzforschung, der Finanzmarktprognose und der natürlichen Sprachverarbeitung auf. Mangelnde Nutzerakzeptanz ist ein Hindernis für den Erfolg neuer Informationssysteme der fortschreitenden Digitalisierung. Um die komplexen Beziehungen der beeinflussenden, unterschiedlichen Akzeptanzdeterminanten zu analysieren, werden multivariate Analyseverfahren, wie bspw. Strukturgleichungsmodellierung verwendet. Weiterhin werden Künstliche Neuronale Netze eingesetzt, um mit ihrer Eigenschaft als entscheidungsunterstützende Informations- und Optimierungsfunktion Finanzprognosen zu erstellen. Abschließend beschäftigt sich diese Arbeit mit der maschinellen Verarbeitung von Textdaten und schlägt eine Methode vor, um das Text Mining für Social Media Daten zu verbessern. In den jeweiligen Forschungsbereichen wird der Status Quo in Forschung und Praxis aufgezeigt, Forschungslücken identifiziert und exemplarische Studien anhand konkreter Anwendungsbeispiele vorgestellt. Die abgeleiteten Handlungsempfehlungen und kritischen Diskussionen leisten in ihrer jeweiligen Disziplin einen Beitrag für zukünftige Forschung.

**Schlagworte:** Datenwissenschaften (Data Science), Technologieakzeptanzmodell (TAM), Strukturgleichungsmodellierung (SEM), Künstliche Neuronale Netze (KNN), Text Mining, Mobile Payment, Wearable Technologien, Optionspreisprognose, Social Media Daten, Online-Rezensionen.

## II. Management Summary

---

Das Bewusstsein, dass Analysetechniken verwendet werden können, um Einsichten aus Daten zu gewinnen und daraus Erkenntnisse abzuleiten, ist so alt wie das Gebiet der Statistik selbst und stammt aus dem 18. Jahrhundert. Ein offensichtlicher Unterschied zu heute ist jedoch das schnelle Tempo von wirtschaftlichen und sozialen Transaktionen in Online-Umgebungen und ermöglicht somit die digitale Erfassung von großen Datenmengen - auch Big Data genannt (Agarwal & Dhar 2014). Die Fähigkeit, die Struktur und den Inhalt des menschlichen Diskurses zu verstehen, hat die Dimensionalität der verfügbaren Datensätze erheblich erweitert. Zudem hat die Verbreitung von Smartphones, Netzwerkkonnektivität, intelligenten Geräten und IoT die Vielfalt und das Volumen sowohl strukturierter als auch unstrukturierter Daten enorm gesteigert und verursacht immer größer werdende Datenmengen. Einfache Tools sind nicht in der Lage, diese riesige Menge und Vielfalt an Daten zu verarbeiten. Aus diesem Grund werden komplexere und fortschrittlichere analytische Werkzeuge und Algorithmen für die Verarbeitung benötigt, die die Analyse und Gewinnung aussagekräftiger Erkenntnisse garantieren. Der Zugang zu Rechenkapazität und benutzerfreundlicher analytischer Software hat den Bereich der Datenwissenschaft (Data Science) demokratisiert, so dass viel mehr Wissenschaftler (und Praktiker) an den Möglichkeiten teilhaben können, die durch die großen Datenmengen ermöglicht werden. Angemessene Datenverarbeitung und -verwaltung kann neues Wissen freilegen und die zeitnahe Reaktion auf sich abzeichnende Chancen und Herausforderungen erleichtern (Chen et al. 2013). Einige Komponenten von Data Science und Business Analytics existieren bereits seit langer Zeit, aber es gibt bedeutende neue Fragen und Möglichkeiten, die durch die Verfügbarkeit von Big Data und große Fortschritte bei der maschinellen Intelligenz entstehen (Chen et al. 2018). Data Science verwendet dabei quantitative und qualitative Methoden, um relevante Probleme zu lösen und Ergebnisse vorherzusagen. Die quantitative Analyse umfasst dabei die Techniken, mit denen Forscher Daten in numerische Formen umwandeln und sie statistischen Analysen unterziehen.

Im Rahmen dieser kumulativen Dissertation werden anhand der entsprechenden Forschungsarbeiten bestimmte exemplarische Studien im Kontext der quantitativen Datenanalyse vorgestellt und diskutiert. Die Arbeit gliedert sich thematisch in drei Hauptteile:

Teil A beschäftigt sich mit der Analyse empirischer Studien in Form von Akzeptanzanalysen mobiler Dienste. Teil B befasst sich mit Finanzmarktdaten und der Prognose sowie der Entscheidungsunterstützung. Teil C fokussiert die Verarbeitung natürlicher Sprache und die Auswertung von Social Media Textdaten. Alle Forschungsbeiträge können unter der Rubrik Datenwissenschaften und -analyse (Data Science) zusammengefasst werden.

**Teil A:** Die zunehmende Verbreitung mobiler Informations- und Kommunikationstechnologien (IKT) gibt uns die Fähigkeit jederzeit und überall Informationen einzuholen und auszutauschen. Heutzutage sind Mobile Dienste kaum noch aus unserem Alltag wegzudenken. Die Beliebtheit von Smartphones in der gesamten Gesellschaft eröffnet der Wirtschaft neue Möglichkeiten mit Kunden in Kontakt zu treten und mit Hilfe neuer Technologien ein breiteres Publikum zu gewinnen. So haben Mobile Dienste wie M(obile)-Payment das Potenzial, die Zahlungsindustrie zu revolutionieren und die Ära der bargeldlosen Gesellschaften einzuläuten. Neue Technologietrends wie Fitness-Armbänder und intelligente Uhren, auch als Wearables bezeichnet, bieten Versicherten bspw. die Möglichkeit, Bewegungs- und Gesundheitsdaten aufzunehmen und für einen gesunden Lebensstil Gratifikationen ihrer Versicherung zu erhalten. Diese Mobilen Dienste werden als sogenannte Pay-as-you-live (PAYL) Dienste bezeichnet. Während sich PAYL noch in der Frühphase ihrer Entwicklung befindet, existieren für M-Payment bereits fortgeschrittene Lösungen. Jedoch gehört M-Payment in Deutschland überraschenderweise nicht zu den häufig genutzten Mobilen Diensten. Akzeptanzstudien können einen wichtigen Beitrag zum Verständnis der Faktoren zu Annahme oder Ablehnung leisten, indem Verhaltensmodelle aus interdisziplinären Bereichen zur Erklärung und Vorhersage des Verhaltens von Nutzern herangezogen werden.

Um zu verstehen, warum M-Payment Dienste in Deutschland immer noch hinter den Erwartungen zurück bleiben, werden zwei Akzeptanzstudien in verschiedenen Ländern durchgeführt und Faktoren analysiert, die die Akzeptanz von M-Payment beeinflussen und kulturelle Unterschiede diskutiert, die zwischen den unterschiedlichen Ländern bestehen. Hierfür werden erweiterte Technologieakzeptanzmodelle (TAM) entwickelt und evaluiert. Die Konstrukte der Forschungsmodelle ergeben sich aus qualitativen Experteninterviews und den Ergebnissen explorativer Literaturrecherchen. Zur Bewertungs- und Modellprüfung wurden quantitative Befragungen in den verschiedenen Ländern durchgeführt und mittels Strukturgleichungsmodellierung unter Verwendung von SmartPLS (Partial Least Squares) ausgewertet.

Ergebnisse der ersten Studie zeigen, dass Amerikaner, im Gegensatz zu Deutschen, nicht bereit sind, zusätzlich für einen M-Payment Dienst zu zahlen. Das Vertrauen in den Dienst sowie die wahrgenommene Nützlichkeit beeinflussen jedoch für beide Länder die Absicht M-Payment zu nutzen gleichermaßen. Auch die Einfachheit der Nutzung hat einen Einfluss auf die Nützlichkeit für beide Länder. Ein Einfluss der Einfachheit der Nutzung auf die Intention der Nutzung kann für Deutschland nicht beobachtet werden. Aus der zweiten Studie geht hervor, dass die Technologiebereitschaft einen Einfluss auf die Intention der Nutzung von M-Payment Diensten hat und es darüber hinaus signifikante kulturelle Unterschiede in Bezug auf die Akzeptanz gibt. Darüber hinaus können die Faktoren des Technologieakzeptanzmodells für alle untersuchten Länder bestätigt werden und bieten eine Grundlage für differenzierte Diskussionen und Verbesserungsvorschläge in der Konzeption mobiler Bezahldienste.

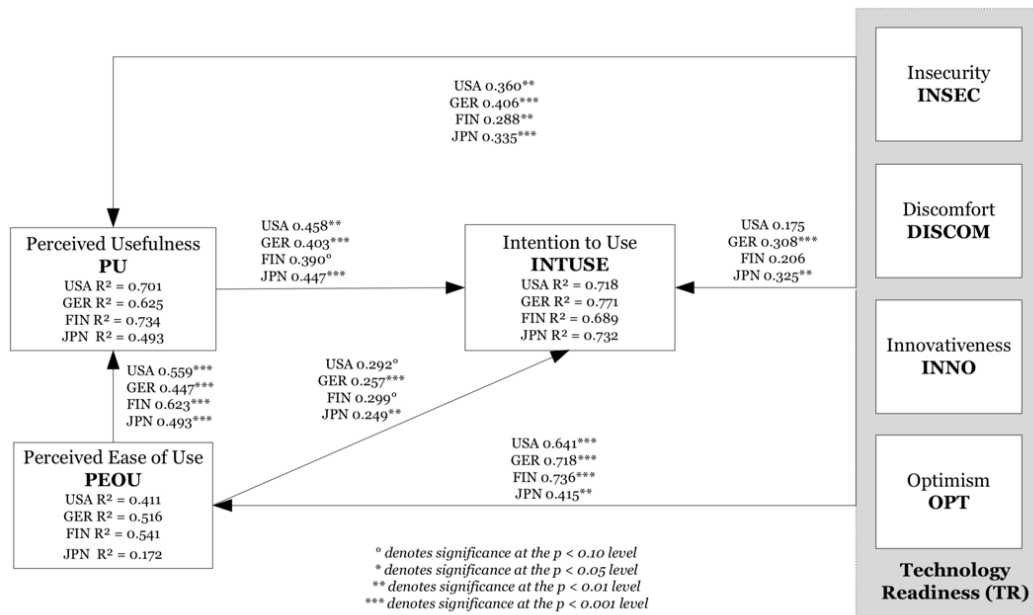
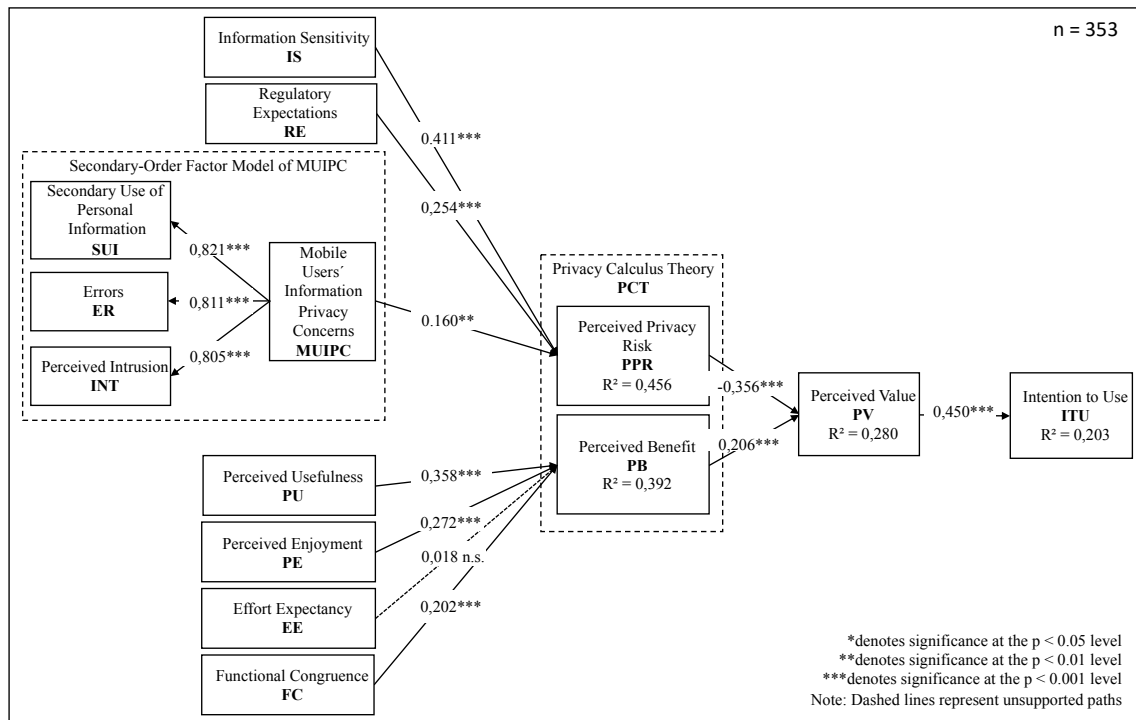


Abbildung 1: Forschungsmodell zu M-Payment

Das Ziel der Studie zu PAYL-Diensten ist es, die Bereitschaft von Kunden zu untersuchen, PAYL-Dienste mit Hilfe von Wearable-Technologie zu nutzen, indem wahrgenommene Datenschutzrisiken und wahrgenommene Vorteile durch die Nutzung und Gratifizierungen verglichen werden. Es werden Kundenbedürfnisse untersucht, um daraus Empfehlungen für Versicherungsunternehmen abzuleiten. Auch hier basiert das Forschungsmodell auf qualitativen Experteninterviews und einer strukturierten Literaturrecherche. Daten aus einer Onlineumfrage werden genutzt, um mittels Strukturgleichungsmodellierung das Forschungsmodell zu evaluieren und Handlungsempfehlungen abzuleiten. Die Ergebnisse zeigen u.a., dass aktuelle Risikofaktoren einen größeren Einfluss auf den

wahrgenommenen Wert einer Nutzung von PAYL-Diensten haben als Vorteile, die bei der Verwendung entstehen.



**Abbildung 2:** Forschungsmodell zu PAYL

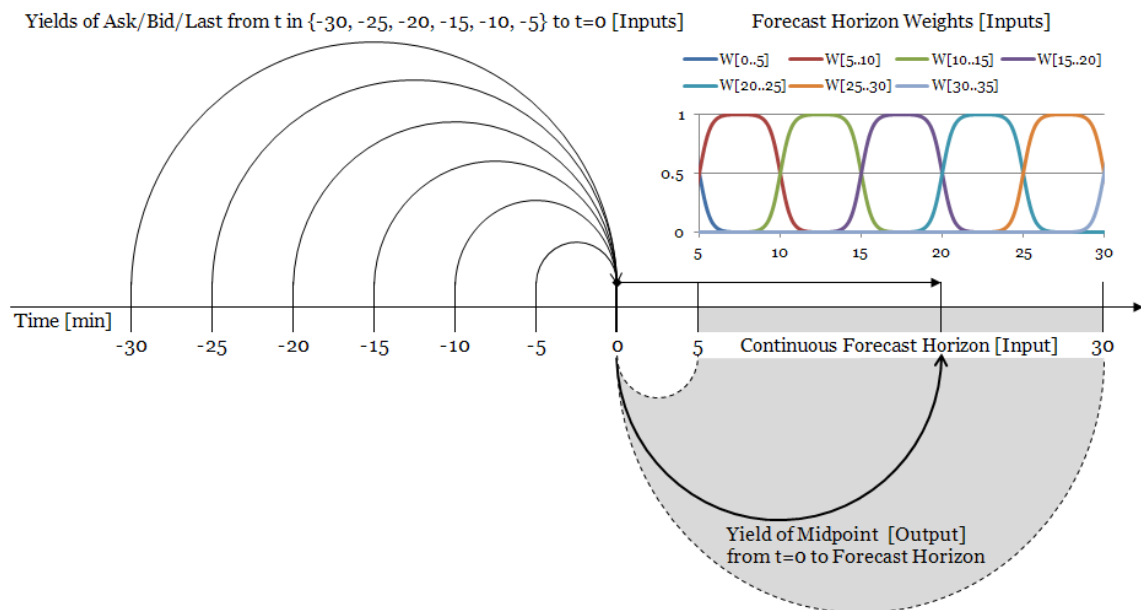
Zudem ist es die erste Studie, die die direkte Beziehung von Mobile Users' Information Privacy Concerns (MUIPC) auf die wahrgenommenen Datenschutzrisiken untersucht und bestätigt.

**Teil B:** Im modernen Finanzmarkt nutzen Marktteilnehmer verschiedene Analysen, um wertvolle Informationen aus historischen Marktdaten für eine bessere Entscheidungsfindung zu gewinnen. Der Finanzmarkt wird dabei als ein komplexes System angesehen, das aus vielen interagierenden Hochfrequenzhändlern besteht, die auf der Grundlage der relativen Stärken dieser Interaktionen Entscheidungen treffen. Dabei gewinnen Finanzderivate als Sicherungsinstrument gegen viele Risiken zunehmend an Bedeutung. Ihr Wert und die zukünftige Entwicklung sind jedoch schwer zu prognostizieren. Aufgrund langfristiger Trends, zyklischer Schwankungen, saisonaler Schwankungen und unregelmäßiger Bewegungen sind Finanzzeitreihen komplizierter als andere statistische Daten (Box et al. 2015). Es werden insbesondere modellgestützte Entscheidungsunterstützungssysteme benötigt, um das zugrunde liegende Risiko zu quantifizieren und es dem Endnutzer ermöglichen, angemessene Risikozuschläge zu finden (de Fortuny et al. 2014, Geva & Zahavi 2014, von Mettenheim & Breitner 2010).



Traditionelle statistische Modelle für Finanzprognosen waren in der Vergangenheit zwar einfach, wiesen jedoch aufgrund der Nichtlinearität der Daten einige Mängel auf. Die Verwendung von Künstlichen Neuronalen Netzen (KNN) zur Vorhersage des Verhaltens und der Tendenzen von Finanzmärkten hat sich als eine praktikable Alternative zu bestehenden konventionellen Techniken erwiesen (Andrade de Oliveira & Nobre 2011, Naeini et al. 2010, Song et al. 2007, Lee und Chen 2007, Ma et al. 2010). Das Teilgebiet der Wechselkursprognosen ist dabei für jede auf dem Finanzmarkt tätige Einheit von beträchtlichem Wert.

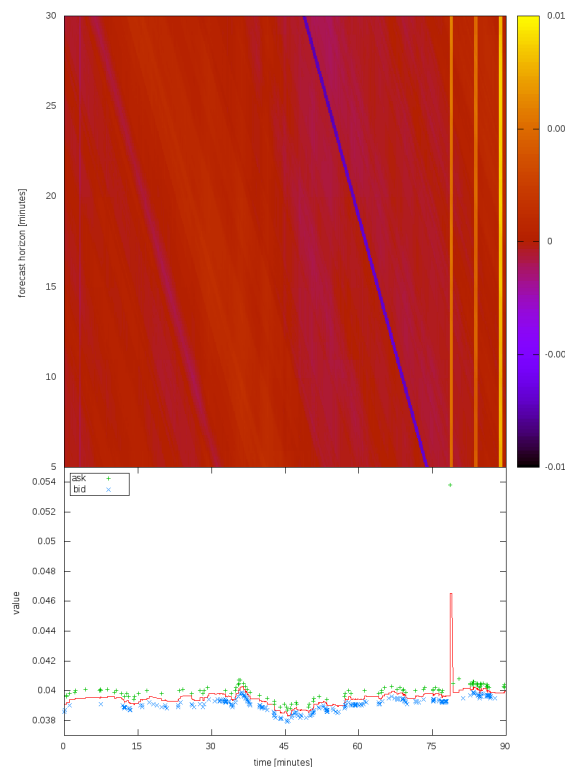
Daher wird in der Studie, dieser vorliegenden Dissertation, ein Entscheidungsunterstützungssystem entwickelt, welches die Anforderungen an ein adäquates Modell zur Prognose von Währungsoptionen anhand von Intraday-Tick-Daten abbildet. Unter Verwendung eines KNN wird der optimale Zeitpunkt für den Kauf und Verkauf sowie der wahrscheinliche Preis einer Währungsoption für einen Zeitraum innerhalb der nächsten dreißig Minuten bestimmt.



**Abbildung 3:** Darstellung des Prognosealgorithmus

Die hier diskutierte Studie konzentriert sich auf unbedingte Termingeschäfte, insbesondere Währungsoptionen. Es werden wahrgenommenen Schwächen von Modellen adressiert, die heute bei Finanzinstituten und (Finanz-) Entscheidungsträgern eingesetzt werden (Bookstaber 2007, Khandani & Lo 2007, Laïdi 2008, Turban et al. 2010).

Es wird ein Ensemble aus neuronalen Netzen trainiert, die 5 besten Netze anhand des Gesamtfehlers des Trainingsdatensatzes ausgewählt und der Mittelwert dieser 5 Netze als Prognoseergebnis für die Out-of-Sample-Daten genutzt.



**Abbildung 4:** Evaluation der Prognose

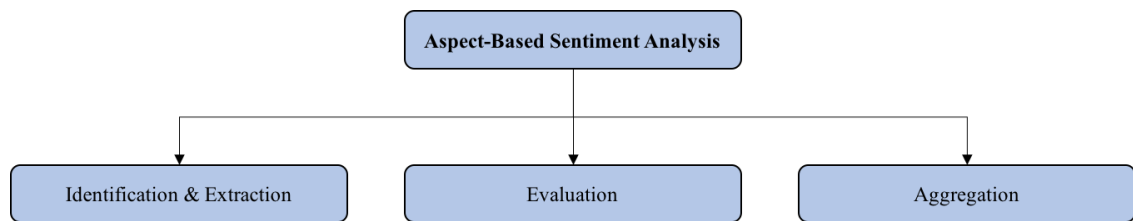
Als Ergebnis zeigt sich, dass der Prognosefehler oft um den Nullpunkt schwankt. Dieser Effekt impliziert, dass die Prognose in den meisten Fällen den Zielmittelpunkt genau erreicht. Der in der Abbildung angezeigte Spike zeigt einen signifikanten Prognosefehler, welcher durch einen kurzfristigen Anstieg des Ask-Kurses zustande kommt. Dennoch wird gezeigt, dass der vorgestellte Ansatz in der Lage ist, ganze Märkte zu modellieren, ohne die Modellkomplexität unnötig zu erhöhen. Die Studie zeigt, dass KNNs für eine kurzfristige Prognose von Optionspreisen geeignet sind. Mit Hilfe des entwickelten Entscheidungsunterstützungssystems wird nicht nur der faire Optionspreis berechnet, sondern auch der optimale Zeitpunkt ermittelt, zu dem diese Option gekauft oder verkauft werden sollte.

**Teil C:** In Social-Media-Daten gibt es eine enorme Menge an Informationen. In den vergangenen Jahrzehnten bezahlten Unternehmen Marktforschungsinstitute, um Konsumenten zu befragen und Fokusgruppen zu führen, um die Art von Informationen zu erhalten, die die Verbraucher jetzt bereitwillig an öffentliche Social-Media-Plattformen senden. Das Problem besteht darin, dass diese Informationen in Form von Freitext und natürlicher

Sprache vorliegen, also die Art von unstrukturierten Daten, die von Analysealgorithmen traditionell verwendet werden. Natural Language Processing (NLP) ist eine Schnittstelle zwischen Informatik, Künstlicher Intelligenz und Computerlinguistik und befasst sich mit Verarbeitung menschlicher (natürlicher) Sprache. Die Geschichte begann mit dem 1950 von Alan Turing herausgegebenen Aufsatz "Computing Machinery and Intelligence". Vier Jahre danach, im Jahre 1954 startete ein Experiment zur vollautomatischen Übersetzung von mehr als sechzig Sätzen aus dem Russischen ins Englische. Damals dachte man, die automatische Übersetzung würde in naher Zukunft gelöst werden (Koehn 2009).

Mit dem Fortschritt von maschinellem Lernen und künstlicher Intelligenz ist es für Unternehmen leichter geworden, die Informationen in Social-Media-Posts skalierbar zu quantifizieren. Bei Social-Media-Daten geht es aber nicht nur um quantifizierbare Statistiken. Während Meinungen zu Produkten häufig in Textform als herkömmliche Online-Rezensionen zu Verfügung stehen, äußern Kunden immer mehr auch in sozialen Netzwerken ihre Meinungen, Gefühle und Bedenken zu Produkten und Dienstleistungen. Dieses wachsende Volumen an Meinungsäußerungen anderer Nutzer spielt eine wichtige Rolle bei der Gestaltung von Produktwahrnehmungen und hilft Kunden bei Kaufentscheidungen (Jensen et al. 2013). Unternehmen können diese Art von Rezensionen als wichtige Ressource in der Produktentwicklung, im Marketing und im Kundenbeziehungsmanagement nutzen (Yu et al. 2011).

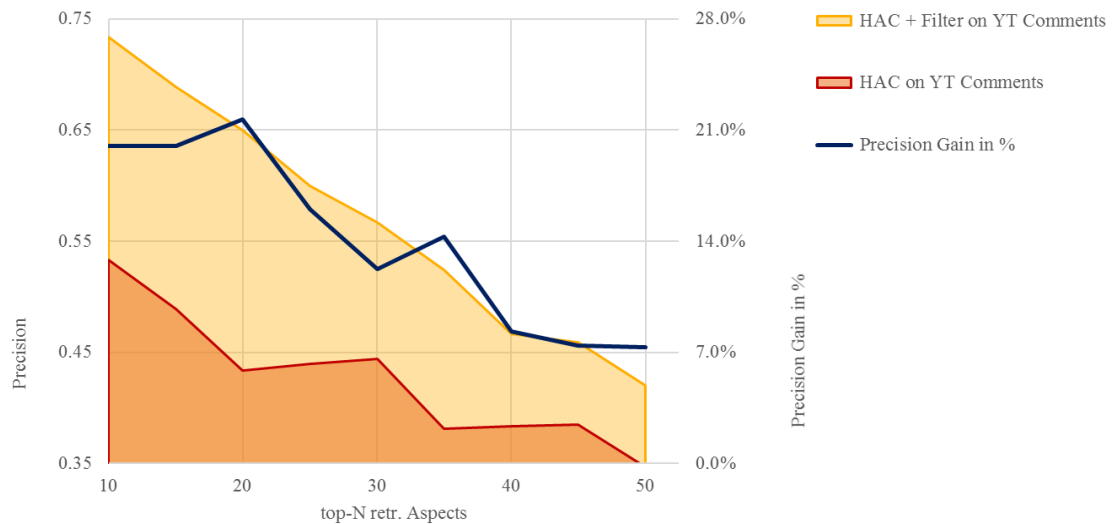
Bei Sentiment Analysen (Sentiment Analysis) oder Meinungsanalysen (Opinion Mining), bei denen maschinelles Lernen und Verarbeitung natürlicher Sprache zum Einsatz kommen, kann festgestellt werden, wie die Öffentlichkeit eine Marke wahrnimmt, welche Art von Produkten die Verbraucher mögen und nicht mögen und allgemein, wo die Märkte hinführen. Laut Liu (2015) ist die Sentiment Analyse das Untersuchungsfeld, in dem Meinungen, Gefühle, Einschätzungen, Einstellungen und Emotionen von Menschen zu Entitäten und ihren enthaltenen Attributen aus schriftlichem Text analysiert werden (Liu 2015). Dabei gibt es verschiedene Klassifizierungsstufen der Sentiment Analyse: Dokumentebene, Satzebene und Aspektebene (Liu 2012). Nach Tsyrarou et al. (2012) besteht die Analyse auf Aspektebene aus drei Schritten. Identifizierung & Extraktion, Bewertung und Aggregation.



**Abbildung 5:** Drei Schritte der Aspekt-basierten Stimmungsanalyse

Zunächst müssen die Aspekte bzw. Stimmungsziele ermittelt und extrahiert werden. Im nächsten Schritt müssen die identifizierten Aspekte ihrem jeweiligen Sentiment entsprechend klassifiziert werden. Schließlich sollte ein klarer Überblick über die aggregierten Stimmungswerte der einzelnen Aspekte gegeben sein. Diese unstrukturierten Daten, wie sie in Social Media Plattformen fast ausschließlich vorkommen, bleiben jedoch bei der Datenanalyse immer noch eine große Herausforderung (Singh et al. 2015). Die hier diskutierte Studie konzentriert sich daher auf die Aspektextraktionsaufgabe und zielt darauf ab, Aspekte für Produkte zu identifizieren, die in Social-Media-Textdaten diskutiert werden.

Die Studie leistet einen Beitrag beim Vergleich zwischen herkömmlichen Rezensionsdaten aus Amazon-Bewertungen und Social-Media-Daten von YouTube. Dies ist die erste Studie, in der YouTube-Kommentare zu produktbezogenen Videos explizit als geeignete Informationsquelle für eine aspektbasierte Sentimentanalyse identifiziert werden. Die Vergleiche zeigen, dass ein auf Seitenansichten basierender Sentiment-Algorithmus bei Amazon-Rezensionen und YouTube-Kommentaren gleichermaßen gut funktioniert. Der zweite Forschungsbeitrag ist die Entwicklung eines Filtermechanismus, der Informationen von Google Trends über das Suchvolumen von Produkten in Verbindung mit ihren Aspekten enthält.



**Abbildung 6:** Ergebnisse des Filteralgorithmus

Es wird davon ausgegangen, dass Kunden dazu neigen, Produkte in Verbindung mit wichtigen Aspekten zu suchen. Durch das Filtern potenzieller Aspekte basierend auf ihrem Suchvolumen werden die Ergebnisse der Aspektextraktion weiter erhöht. Ein besonderes Problem bleibt die Extraktion impliziter Aspekte. Es gibt einen Unterschied zwischen expliziten Aspekten wie einer Kamera für ein Smartphone und impliziten Aspekten wie dem Gewicht des Smartphones, das nur indirekt von Adjektiven wie schwer und leicht erwähnt wird.