

Opinion Mining

Entwicklung eines Vorgehensmodells

Masterthesis

zur Erlangung des Grades Master of Science
im Studiengang Informationsmanagement

vorgelegt von

Ina Kimmling

Betreuer:

Dr. Michael Möhring, Institut für Wirtschafts- und Verwaltungsinformatik
(IWVI) im FB 4: Informatik der Universität Koblenz-Landau
Prof. Dr. Klaus G. Troitzsch, IWVI, FB 4, Universität Koblenz-Landau
Katja Deutsch, Consultant, Altran CIS
Dr. Jörg Reinnarth, Business Manager, Altran CIS

Erstgutachter:

Dr. Michael Möhring, IWVI, FB 4, Universität Koblenz-Landau

Zweitgutachter:

Prof. Dr. Klaus G. Troitzsch, IWVI, FB 4, Universität Koblenz-Landau

Koblenz, im Januar 2010

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden.

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu.

Koblenz, im Januar 2010

Ina Kimmling

Inhaltsverzeichnis

Abbildungsverzeichnis	v
Abkürzungsverzeichnis	vii
1 Einleitung	1
1.1 Problemstellung	2
1.2 Zielsetzung	3
1.3 Aufbau der Arbeit	4
2 Theoretische Grundlagen	6
2.1 Web 2.0 - Die nächste Generation Internet	6
2.1.1 Weblog	
Vormedialer Raum zwischen Individual- und Massenmedien	7
2.1.2 Virtuelle Community	
Meinungsführerschaften im sozialen Netzwerk	11
2.1.3 Online Forum	
Einfluss des Third Party Endorsement	13
2.1.4 Bewertungsportal	
Kundenrezensionen im Internet	13
2.2 Business Intelligence und ihre Methoden	15
2.2.1 Data Mining	16
2.2.2 Text Mining	18
2.2.3 Opinion Mining	20
3 Empirische Untersuchung	25
3.1 Entwicklung eines Vorgehensmodells	25
3.1.1 Business Understanding	25
3.1.2 Data Understanding	26
3.1.3 Data Preparation	27
3.1.4 Modeling	28
3.1.5 Evaluation & Deployment	30
3.2 Praktische Anwendung und Auswertung	33
3.2.1 Business Understanding	33
3.2.1.1 Auswahl eines Produktes	33
3.2.1.2 Auswahl eines Tools	34
3.2.2 Data Understanding	38
3.2.2.1 Erfassung der Kundenmeinungen per Webcrawling	38
3.2.2.2 Erstellen der Review-Datenbank	42

3.2.3 Data Preparation	47
3.2.3.1 Erstellen der Frequent Feature-Datenbank.....	48
3.2.3.2 Erstellen der Opinion Word-Datenbank	50
3.2.3.3 Data Preparation im RapidMiner.....	51
3.2.4 Analyse	59
3.2.4.1 Analyse im RapidMiner	59
3.2.4.2 Analyse der Ergebnisdatei	62
3.2.5 Evaluation & Deployment	73
3.2.5.1 Bewertung der Opinion Mining Ergebnisse.....	73
3.2.5.2 Prüfung der Validität	75
3.2.5.3 Bewertung des Opinion Mining Tools	76
3.3 Implikationen für Forschung und Praxis	78
3.3.1 Implikationen für die Forschung	78
3.3.2 Implikationen für die Praxis	81
4 Zusammenfassung und Ausblick	83
Literaturverzeichnis	88
Anhang.....	93

Abbildungsverzeichnis

Abbildung 1: Fakten vs. Meinungen	3
Abbildung 2: Aufbau der Arbeit	5
Abbildung 3: Die Web 2.0 Komponenten	7
Abbildung 4: Blogs zwischen Individual- und Massenkommunikation	8
Abbildung 5: Bekanntheit und Nutzung von Weblogs	9
Abbildung 6: Nutzungsmotive im Vergleich	9
Abbildung 7: Bedeutung von Blogs für die Meinungsbildung	10
Abbildung 8: Welche Blogs nicht glaubwürdig sind	10
Abbildung 9: Elemente virtueller Communities	11
Abbildung 10: Kundenrezensionen bei Amazon	14
Abbildung 11: Kundenrezensionen bei Ciao	15
Abbildung 12: Unterschiedliche Facetten der Business Intelligence	16
Abbildung 13: Der KDD-Prozess	17
Abbildung 14: Phasen des CRISP-DM Modells	18
Abbildung 15: Abgrenzung der Mining-Methoden	20
Abbildung 16: Primärforschung vs. Datensammlung im Opinion Mining	23
Abbildung 17: Business Understanding	26
Abbildung 18: Data Understanding	27
Abbildung 19: Data Preparation	28
Abbildung 20: Ablauf von Training und Test	29
Abbildung 21: Modeling	30
Abbildung 22: Evaluation & Deployment	31
Abbildung 23: Vorgehensmodell für das Opinion Mining	32
Abbildung 24: Text Mining Lösungen im Vergleich	37
Abbildung 25: Sequenzieller Ablauf des Webcrawlings	39
Abbildung 26: Methoden des Webcrawlings	40
Abbildung 27: Beantwortung der Forschungsfrage 1	42
Abbildung 28: Blog-Suche unter Google	43
Abbildung 29: Blog-Suche unter Technorati	44
Abbildung 30: Community-Suche unter Google	44
Abbildung 31: Forum-Suche unter Google	45
Abbildung 32: Bewertungssuche unter Amazon	46
Abbildung 33: Bewertungssuche unter Ciao	46
Abbildung 34: Ergebnisse der Social Website Suche	47

Abbildung 35: Vorgehensweise des Brill-Taggers.....	49
Abbildung 36: Adjektivstruktur: Synonyme und Antonyme.....	51
Abbildung 37: SplitSegmenter im RapidMiner.....	52
Abbildung 38: TextInput und Preprocessing Steps im RapidMiner	54
Abbildung 39: Ergebnis der Data Preparation.....	55
Abbildung 40: Auflistung der Frequent Features des iPhone	56
Abbildung 41: Überblick der positiven, neutralen und negativen Opinion Words ..	57
Abbildung 42: Beantwortung der Forschungsfrage 2	58
Abbildung 43: AttributeFilter_1 im RapidMiner	59
Abbildung 44: AttributeConstruction im RapidMiner.....	60
Abbildung 45: AttributeFilter_2 im RapidMiner	61
Abbildung 46: CSVExampleSetWriter im RapidMiner	61
Abbildung 47: Vergleich der FFs (alle Sätze mit unterschiedlichen FFs).....	67
Abbildung 48: Vergleich der FFs (ohne Sätze mit mehr als 5 FFs).....	67
Abbildung 49: Vergleich der FFs (ohne Sätze mit mehr als 3 FFs).....	68
Abbildung 50: Vergleich der FFs (ohne Sätze mit mehr als 1 FF).....	68
Abbildung 51: Verteilung der am stärksten positiv und negativ bewerteten FFs ...	69
Abbildung 52: Bewertung der Frequent Features des iPhone.....	70
Abbildung 53: Bewertung des iPhone 3G	71
Abbildung 54: Bewertung des iPhone 3GS	72
Abbildung 55: Vergleich der Bewertungen der Generationen 3G und 3GS	73
Abbildung 56: Kodierung einiger Sätze durch den RapidMiner.....	116

Abkürzungsverzeichnis

BI.....	Business Intelligence
CAR	Class Association Rule
CBA-CB	Classification Based on Associations-Classifier Builder
CBA-RG	Classification Based on Associations-Rule Generation
CRISP-DM	CRoss Industry Standard Process for Data Mining
EIS	Enterprise Information System
FF(s)	Frequent Feature(s)
IE.....	Information Extraction
KDD	Knowledge Discovery in Databases
KDT	Knowledge Discovery in Texts
KI.....	Künstliche Intelligenz
MIS.....	Management Information System
OLAP	OnLine Analytical Processing
OM	Opinion Mining
OW(s).....	Opinion Word(s)
PDA.....	Personal Digital Assistant
POS	Part-of-Speech Tagging
RM	RapidMiner
WWW.....	World Wide Web

1 Einleitung

„Alle traditionellen Medien werden derzeit heftig angefochten durch das Internet [...]. Die Medien werden demokratisiert.“ (Meckel 2006)

Die damit einhergehende Entwicklung des Konsumenten vom passiven Rezipienten zum aktiven Partizipanten führt zu einer Flut an Meinungsäußerungen, die im Internet frei zugänglich sind. Diese Meinungen sind insoweit unverfälscht, als dass sie keines Mediators bedürfen. Jedoch sind sie von einem stark subjektiven Charakter geprägt und auch nicht als repräsentativ zu betrachten, da sie ausschließlich der Internetcommunity entspringen. Trotzdem sind diese Meinungsäußerungen im Internet von immenser Bedeutung für die Unternehmen.

„People are going to talk about your product.“

(Bill Perry, Director of Public Relations bei Participate.com (vgl. Brewer 2000))

Wie von Perry auf den Punkt gebracht, ist es in der Netzwerkgesellschaft, in welcher wir heute leben, für Unternehmen unumgänglich das Internet auf Kundenmeinungen zu den eigenen Produkten und Dienstleistungen zu monitoren. Die Unternehmenskommunikation wandelt sich aufgrund des User Generated Content¹. Im Web 2.0 findet sich dieser Content in den Kommunikationsprozessen die durch Unternehmen angestoßen werden oder auch auf diese abzielen. Botschaften werden zu meinungsorientierten Kommunikationsangeboten, die in ein Netzwerk sozialer Verbindungen eingebettet sind, welche über die Bedeutung und Bewertung der Botschaften befinden. Somit wird das Ergebnis eines Kommunikationsprozesses durch das Zusammenspiel von Nutzern, der sogenannten „Wisdom of the Crowds“², bestimmt (vgl. Meckel 2008 [2]). Basierend auf der heutigen Peer-Production stellt es keine Seltenheit mehr dar, dass sich viele potenzielle Kunden vor dem Kauf über ein Produkt oder eine Dienstleistung im Internet informieren, welches ihnen etliche Bewertungen und Reputationen durch

¹ User Generated Content bezeichnet den Inhalt, den User im Netz selbst erstellen - in Blogs, Foren, Communities oder auf eigenen Webseiten. Im Gegensatz dazu steht der Business Generated Content, also redaktioneller Inhalt (Quelle: Duden Szenesprachenwiki, URL: <http://szenesprachenwiki.de/definition/ugc/>, Abruf am 03.07.2009).

² Unter Wisdom of the Crowds wird die Hypothese verstanden, dass die Antworten mehrerer Individuen zusammengenommen im Schnitt zutreffend und richtig sind, wenn die Voraussetzungen der Diversität, Unabhängigkeit, Dezentralisierung und Aggregation erfüllt sind. (Quelle: Surowiecki, J. (2005): The Wisdom of the Crowds: Why the Many are Smarter Than the Few, Little Brown Book Group, 2005.)

die Partizipanten zur Verfügung stellt, und sich auf deren sogenanntes „Third Party Endorsement“³ verlassen. Aufgrund dessen ist es für Unternehmen von beachtlichem Interesse, die verschiedenen Meinungen der Kunden zu ihren Produkten und Dienstleistungen zu kennen. Andernfalls ergibt sich die Gefahr, dass negative Meinungsäußerungen nicht erkannt werden, welche die Reputation des Unternehmens nachhaltig zerstören und sich zur Kommunikationsfalle entwickeln können. Ein bekanntes Beispiel dafür stellt der Fall „Dell-Hell“ (Case Study „Dell-Hell“ 2007) dar, in dem der Blogger Jarvis sich in seinem Blog „Buzzmachine“ über Servicemängel der Firma DELL beklagt und den Begriff „Dell-Hell“ kreiert. Seiner Meinung schlossen sich weitere zahlreiche frustrierte Dell-Kunden an. Mittlerweile findet sich dieser Begriff mehrere tausend Male bei Google.

„Wer die Kommunikationsfalle umgehen will, darf nicht stecken bleiben in Datenmengen und Informationsbergen.“ (Meckel 2007, S.46)

Um die von Meckel beschriebenen Datenmengen und Informationsberge des WWW zu bewältigen, bietet sich die Methode des Opinion Mining (OM) an, mit welcher sich diese Arbeit nachfolgend befasst.

1.1 Problemstellung

Wie zuvor beschrieben informieren sich potenzielle Kunden oftmals im Internet über die Fakten und Meinungen zu bestimmten Produkten und Dienstleistungen. Der Enzyklopädie Brockhaus nach handelt es sich bei Fakten um nachweisbare Tatsachen. Meinungen hingegen sind Einstellungen gegenüber Ereignissen, Umständen und Themen. Gegenüber den subjektiven Meinungen sind die objektiven Fakten relativ einfach zu erfassen und miteinander in Vergleich zu setzen. So können beispielsweise Größe und Gewicht bestimmter Produkte einander direkt gegenübergestellt werden. Unterschiedliche Meinungen hingegen sind nur schwer gegeneinander abzuwägen und gleichzusetzen. Zum besseren Verständnis verdeutlicht Abbildung 1 noch einmal den Unterschied.

³ Third Party Endorsement ist wie folgt definiert: „Solicited or unsolicited recommendation or testimonial from an entity (usually a customer or user) other than the manufacturer and seller of a product or service.“ (Quelle: Business Dictionary, URL: <http://www.businessdictionary.com/definition/peer-group.html>, Abruf am 03.07.2009).



Abbildung 1: Fakten vs. Meinungen

Quelle: in Anlehnung an Altran CIS (Faust 2009): *Opinion Mining. Bestimmung von Gefühlswelten und Meinungen im Web 2.0*

Das Opinion Mining stellt eine Methode dar, mit der Meinungen aus dem WWW gefiltert und intelligent aufbereitet werden können, um Stimmungsanalysen zu einzelnen Produkten und Dienstleistungen zu erstellen. Die Unternehmen sind somit nicht mehr auf – über den traditionellen Weg (bspw. Umfragen) – schwer zu erreichende Kundendaten angewiesen, sondern erhalten die Möglichkeit ohne größere Umwege die Meinungen ihrer Kunden aus dem WWW zu filtern. Dadurch entfallen zeit- und kostenintensive Forschungsaufträge. Opinion Mining bietet den Unternehmen somit eine innovative Möglichkeit ihre Marketingkommunikation planbarer und wirkungsvoller zu gestalten und ein deutlich erhöhtes Maß an Involvement und Relevanz innerhalb bestimmter Zielgruppen zu erreichen.

Beim OM handelt es sich um ein neues und aktuelles Verfahren im Bereich Business Intelligence. Aufgrund dessen findet sich in der Literatur eine Vielzahl unterschiedlichster Vorgehensweisen. Der Fokus dieser Arbeit liegt deswegen auf der Entwicklung eines – für die Zwecke dieser Arbeit optimalen – Vorgehensmodells für das Opinion Mining und der praktischen Anwendung dieses Modells auf ein Produkt.

1.2 Zielsetzung

In der Arbeit wird das Opinion Mining theoretisch aufgearbeitet und anschließend auf mindestens ein Produkt angewandt. Im Rahmen dieser Untersuchung sollen die nachfolgenden Forschungsfragen beantwortet werden. Ziel ist es den vormaligen Raum des Social Web (Diskussionsforen, Blogs, Communities) mit Hilfe eines Opinion Mining Tools zu scannen und diese Datenmengen intelligent zu analysieren, um unternehmensrelevante Ergebnisse, beispielsweise in Bezug auf Entscheidungshilfen, Prozessoptimierung und Vorhersagen zu erhalten.

- Forschungsfrage 1:

Wie erfolgt die Erfassung der benötigten Kundenmeinungen aus dem Internet?

Es werden verschiedene Methoden zur automatischen und manuellen Datenerfassung aus dem Internet vorgestellt. Anschließend erfolgt die Auswahl und Anwendung einer der Alternativen (vgl. Kapitel 3.2.2).

- Forschungsfrage 2:

Wie lassen sich die Kundenmeinungen per Opinion Mining bewerten?

Zur Bewertung der Kundenmeinungen ist es zunächst notwendig die Frequent Features eines Produkts zu kennen. Anschließend müssen die Opinion Words identifiziert werden, welche die Produkteigenschaften beschreiben (vgl. Kapitel 3.2.3).

- Forschungsfrage 3:

Wie gut unterstützt die ausgewählte Software das Opinion Mining?

Im Zusammenhang mit dieser Forschungsfrage wird das zuvor ausgewählte Opinion Mining Tool bewertet (vgl. Kapitel 3.2.5.3).

- Forschungsfrage 4:

Welche Implikationen ergeben sich aus dem Ergebnis?

Basierend auf den Ergebnissen werden Implikationen für die Forschung und Praxis ausgesprochen (vgl. Kapitel 3.3).

1.3 Aufbau der Arbeit

Zum besseren Verständnis werden zu Beginn der Arbeit, in Kapitel 2, zunächst die theoretischen Grundlagen des Opinion Mining erklärt. Dazu zählt zum einen die Erläuterung des Web 2.0 und seiner neuartigen Kommunikationsformen. Zum anderen wird die Business Intelligence mit den für diese Arbeit relevanten Methoden, einschließlich des Opinion Mining, erklärt. Anschließend folgt in Kapitel 3 die empirische Untersuchung, welche den Schwerpunkt der Arbeit darstellt. Diese beginnt mit der Entwicklung eines Vorgehensmodells. Anschließend erfolgt die praktische Anwendung dieses Modells auf ein Produkt. Daraufhin werden die Ergebnisse des Review-Scan in Hinblick auf Forschung und Praxis ausgewertet und diskutiert. In

diesem Schwerpunktkapitel werden auch die zuvor genannten Forschungsfragen behandelt. Die Arbeit schließt mit einem Resümee des methodologischen Vorgehens beim Opinion Mining und einem Ausblick, welcher Implikationen für Forschung und Praxis zusammenfassend darstellt. Abbildung 2 stellt das Vorgehen noch einmal grafisch dar.

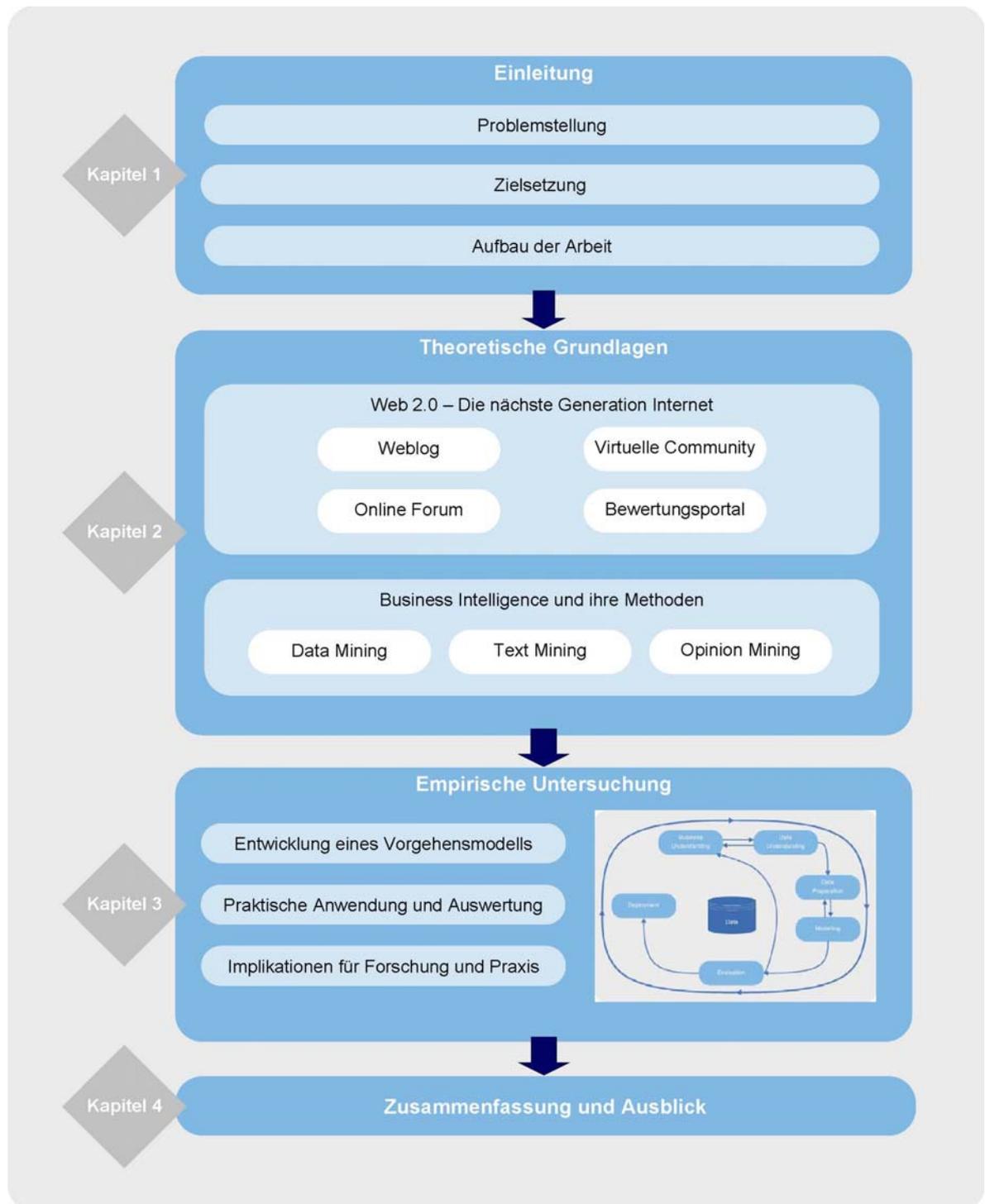


Abbildung 2: Aufbau der Arbeit

2 Theoretische Grundlagen

Dieses Kapitel beschreibt die notwendigen Grundlagen zum besseren Verständnis der Arbeit. Dazu werden verschiedene für die Arbeit notwendige Begrifflichkeiten näher erläutert, definiert und/oder gegeneinander abgegrenzt. Das erste Unterkapitel beschäftigt sich mit dem Web 2.0, der nächsten Generation Internet, und damit einhergehender Entwicklungen wie Weblogs, Communities und Foren. Das zweite Unterkapitel widmet sich der Business Intelligence (BI) sowie der Erläuterung einzelner BI-Methoden. Hier findet sich auch eine kurze Abhandlung der historischen Entwicklung des Opinion Mining.

2.1 Web 2.0 - Die nächste Generation Internet

Eine der heute bedeutsamsten Entwicklungen des WWW ist das Web 2.0, oftmals bezeichnet als die nächste Generation Internet, welches von Högg et al. (2006, S.13) wie folgt definiert wird:

„Web 2.0 is defined as the philosophy of mutually maximizing collective intelligence and added value for each participant by formalized and dynamic information sharing and creation.“

Stanoevska-Slabeva (2008, S.13-38) beschreibt das Web 2.0 folgendermaßen: Im Jahr 2004 wurde der Begriff Web 2.0 auf der gleichnamigen Konferenz von O'Reilly (2005) mit dem Ziel geprägt auf eine neue Entwicklung des Internets hinzuweisen. Zum einen handelt es sich um eine Reihe von Technologien und Anwendungen (bspw. neue Protokolle, die Sprachen vereinfachen und neue Austauschmöglichkeiten für Informationen (RSS) bieten), die mit Web 2.0 bezeichnet werden. Zum anderen steht dieser Begriff jedoch auch für eine gravierende Verhaltensänderung der Internetnutzer. Es entwickelte sich eine neue Art von Plattformen und Communities (bspw. YouTube, Wikipedia und MySpace). Im Gegensatz zu Web 1.0 Plattformen kreieren Web 2.0 Plattformen meist keine eigenen Inhalte, sie übernehmen keine redaktionellen Aufgaben. Die Plattform wird lediglich mit ihren Funktionalitäten und Benutzungsregeln zur Verfügung gestellt (vgl. Högg et al. 2006). Desweiteren stellt das Web 2.0 neue Instrumente und Werkzeuge wie beispielsweise Blogs oder Wikis zur Verfügung. Aus Anwendersicht bedeutet Web 2.0 eine neue Philosophie der Internetnutzung sowie eine neue Umgangsform mit Inhalten und der gegenseitigen Kommunikation (vgl. Högg et al. 2006). Die Benutzer beteiligen sich (inter-)aktiv an der „Mitmach-Plattform“ Internet, indem sie selbst Beiträge erstellen. Individuelle Meinungen werden frei geäu-

bert, vernetzen sich mit den Meinungen Gleichgesinnter und können sich schnell zu einflussreichen Meinungspools entwickeln. Die Komponenten des Web 2.0 sind in Abbildung 3 noch einmal im Überblick dargestellt.

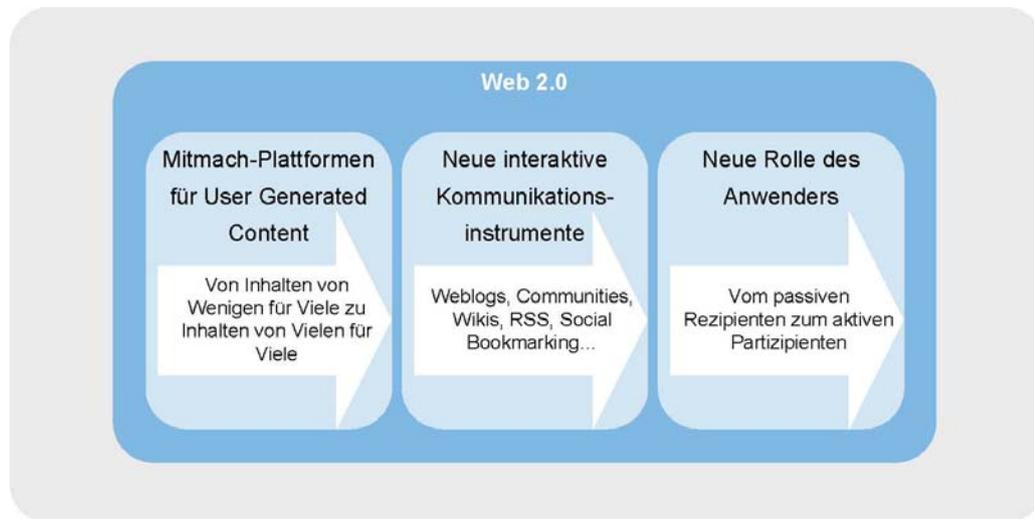


Abbildung 3: Die Web 2.0 Komponenten

Quelle: in Anlehnung an Stanoevska-Slabeva 2008, S.16

Das Web 2.0 birgt einige neue Kommunikationsformen wie Weblogs, Communities, Diskussionsforen und Bewertungsportale. Diese dienen in der Arbeit als Informationsquellen zum Auffinden von Kundenmeinungen über das zu untersuchende Produkt. Aufgrund dessen werden die genannten Kommunikationsformen nachfolgend näher erläutert. Dazu erfolgen jeweils eine Definition und eine Darstellung der verschiedenen Besonderheiten.

2.1.1 Weblog - Vormedialer Raum zwischen Individual- und Massenmedien

Der Weblog – kurz Blog – ist ein Kunstwort, das sich aus den Wörtern „Web“ und „Logbuch“ zusammen setzt. Kaiser (2008, S.22) gebraucht folgende Definition:

„A blog is a user-generated website on which the writers (bloggers) enter their remarks in journal style, which then appear in reverse chronological order. Blogs often provide commentary or news on a particular subject [...], although some function as more personal online diaries. [...] A typical blog combines text, images, and like to other blogs, Web pages, and other media related to its topic. The vast majority are primarily text, although some focus on photographs (photo-blog), video (vlog) or audio (podcasting, [...]), and are part of a wider network of social media.“

Über gegenseitige referenzielle Bezüge (Trackbacks⁴ und Pingbacks⁵) wird ein soziales Netzwerk, die sogenannte Blogosphäre, erschaffen. Diese unterstützt den Informationsaustausch und erweitert den Wissensstand aller Produzenten und Rezipienten der Blogs (Kaiser 2008, S.26 f.).

Weblogs sind von immenser Wichtigkeit für Unternehmen, da sie einen vormedialen Raum für Meinungen jeglicher Art schaffen. Sie bewegen sich fortwährend zwischen Individual- und Massenkommunikation, wie schon in der Einleitung durch den Case „Dell Hell“ veranschaulicht. Blogs können Grassroot-Bewegungen⁶ auslösen, indem Themen problematisiert und von Insidern diskutiert werden. Wenn sich das Thema anschließend in der Blogosphäre sehr stark verbreitet, wird es von den Massenmedien aufgegriffen und erscheint dort als Schlagzeile. Später verliert das Thema dann an Bedeutung und geht wieder in die Individualkommunikation über (vgl. Abbildung 4 und Zerfaß 2005, S.94 ff.).

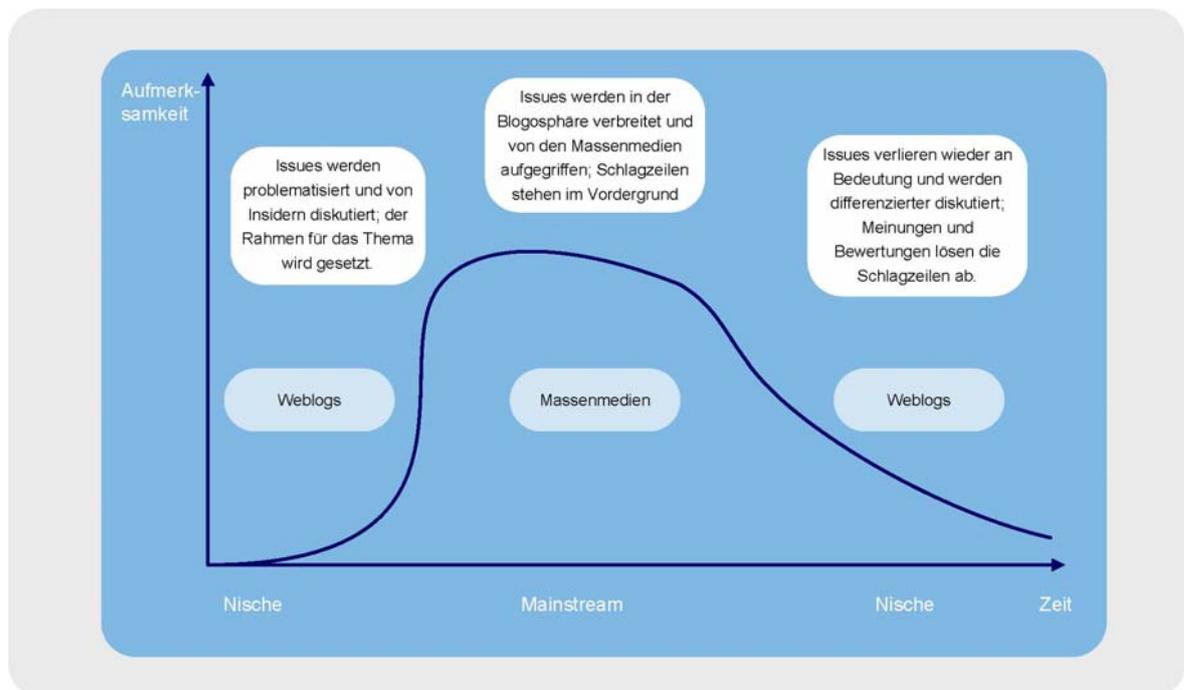


Abbildung 4: Blogs zwischen Individual- und Massenkommunikation

Quelle: Zerfaß 2005, S.96

⁴ Die Trackback-Funktionalität ermöglicht es den Bloggern zu ersehen, ob auf ihren Eintrag in andere Blogs eine Bezugnahme erfolgt. Trackbacks werden dann eingerichtet, wenn im betreffenden Artikel ein Verweis auf den Quellartikel besteht. Trackbacks werden manuell gesetzt. (Quelle: Kaiser 2008, S.26).

⁵ Die automatisierte Variante des Trackbacks durch die Blogsoftware wird als Pingback bezeichnet. (Quelle: Kaiser 2008, S.26).

⁶ Grassroot-Bewegungen sind Initiativen, die aus der Bevölkerung heraus entstehen. In Form von Bürgerjournalismus nehmen die Bürger durch eigene Medien am gesellschaftlichen Diskurs teil. (Quelle: Computerbase, URL: <http://www.computerbase.de/lexikon/Graswurzel-Journalismus>, Abruf am 07.07.2009).

Die enorme Bedeutung von Weblogs in der heutigen Zeit stellt die „Blogstudie 2007“ von Zerfaß und Bogosyan dar. Diese befragten über 600 Internetnutzer in Deutschland zur Informationssuche im Internet sowie zur Nutzung und Glaubwürdigkeit von Blogs. Laut der Studie kennen 90% der befragten Internetnutzer Weblogs und knapp 80% nutzen sie (in-)aktiv (vgl. Abbildung 5).

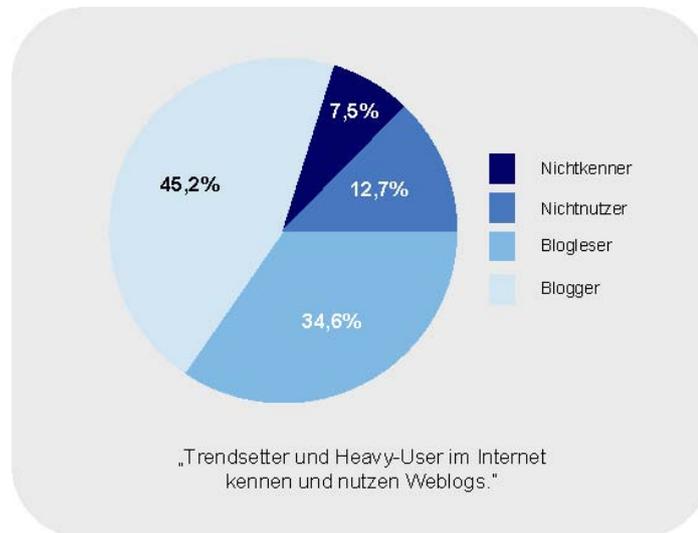


Abbildung 5: Bekanntheit und Nutzung von Weblogs

Quelle: Blogstudie 2007 (Zerfaß und Bogosyan 2007, S.3)

Abbildung 6 verdeutlicht, dass vor allem Blogleser das Angebot an Weblogs nutzen, um die dort enthaltenen Meinungen zu erfahren, welche von den Bloggern zur Kritik und Aufklärung veröffentlicht werden. Weblogs bieten sowohl für die Blogger als auch für die Leser ein Medium, welches sich von den klassischen Medien abgrenzt und neuartige Informationen und Möglichkeiten bietet.

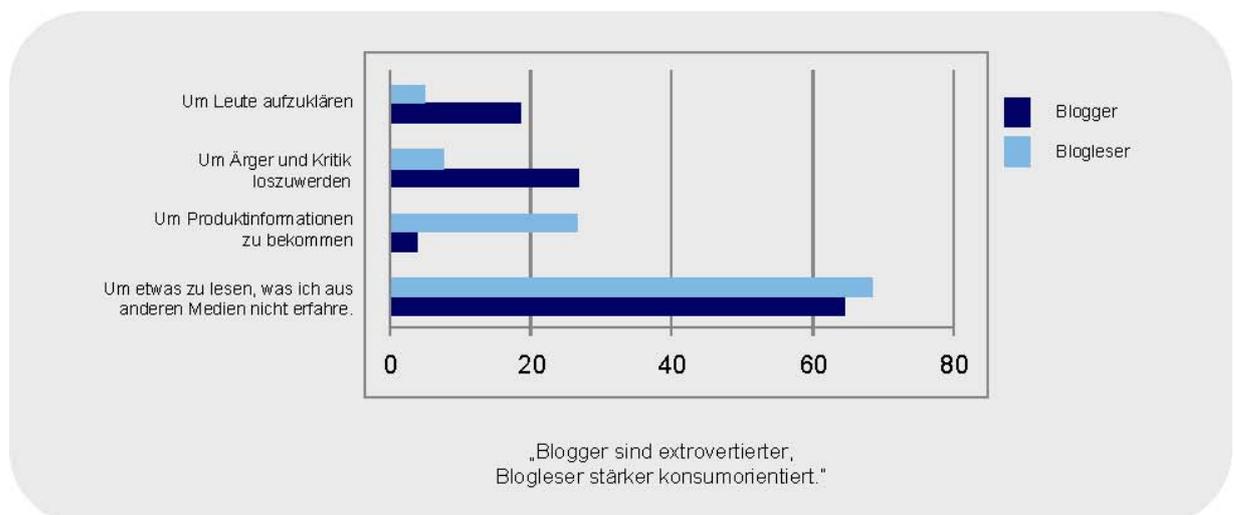


Abbildung 6: Nutzungsmotive im Vergleich

Quelle: Blogstudie 2007 (Zerfaß und Bogosyan 2007, S.6)

Zur Meinungsbildung erfreuen sich vor allem Fachblogs sowie journalistische- und Medienblogs großer Beliebtheit. Corporate Blogs werden hingegen weniger stark wahrgenommen (vgl. Abbildung 7).

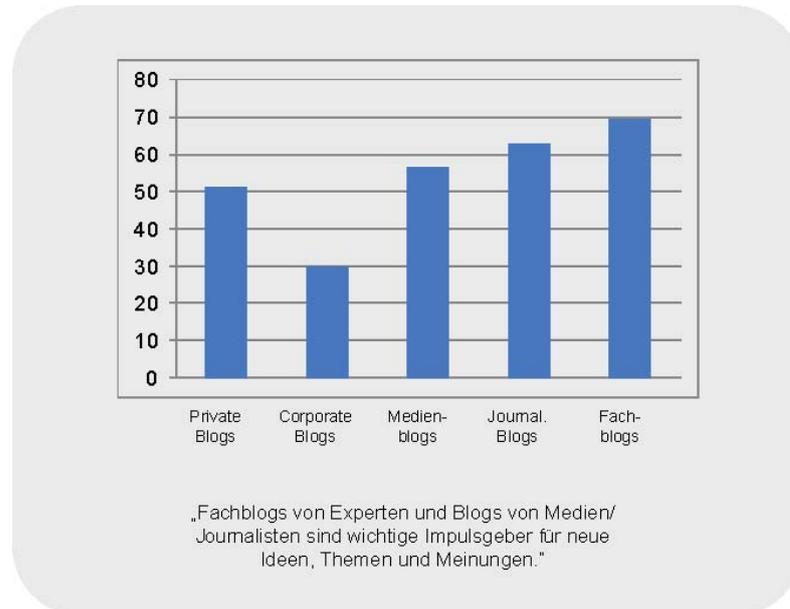


Abbildung 7: Bedeutung von Blogs für die Meinungsbildung

Quelle: Blogstudie 2007 (Zerfaß und Bogosyan 2007, S.9)

Abbildung 8 zeigt auf, dass Corporate Blogs von den Internetnutzern am wenigsten glaubwürdig eingestuft werden. Dies erklärt auch die zuvor bei dieser Blogart festgestellte geringe Beliebtheit bei der Meinungsbildung. Vor allem Fachblogs scheinen den Internetnutzern authentisch. Nur 2% der Befragten sind ihnen gegenüber skeptisch eingestellt.

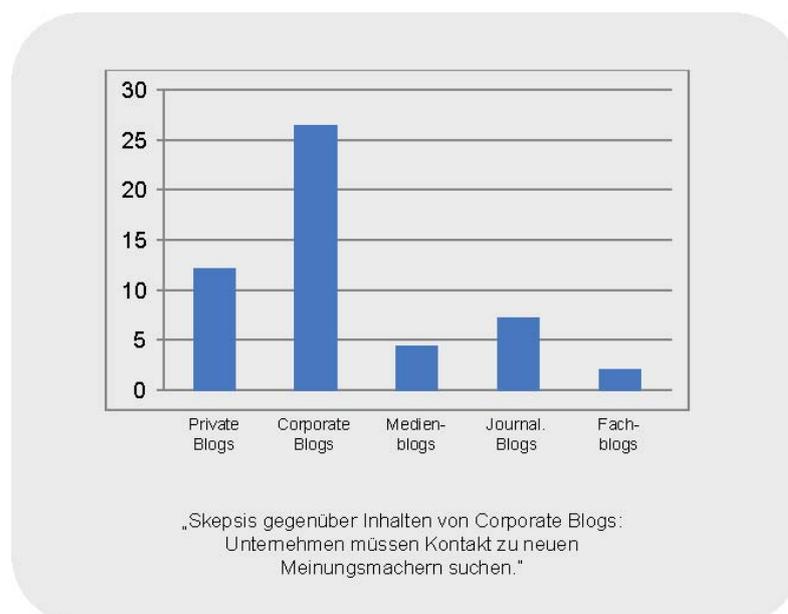


Abbildung 8: Welche Blogs nicht glaubwürdig sind

Quelle: Blogstudie 2007 (Zerfaß und Bogosyan 2007, S.10)

Zusammenfassend stellen Weblogs eine wichtige Informationsquelle zur Untersuchung von Produktmeinungen dar. Aufgrund möglicher Grassroot-Bewegungen der Blogosphäre sind Blogs von enormem Belang für Unternehmen, da sie somit Einfluss auf die Unternehmens- und Produktreputation nehmen können. Desweiteren nutzt ein Großteil der Internetnutzer (80%) das Angebot an Blogs, da sie gegenüber anderer Medien neuartige Meinungen und Informationen enthalten.

2.1.2 Virtuelle Community - Meinungsführerschaften im sozialen Netzwerk

Die wohl bekannteste Definition von virtuellen Communities, auch Online Communities genannt, stammt von Rheingold (1993, Introduction):

“Virtual communities are social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace.”

Virtuelle Communities sind demnach soziale Online Netzwerke, welche auf der Basis gemeinsamer Interessen entstehen. Communities bestehen nach Heitman et al. (2004, S.2) aus vier grundlegenden Elementen, die in Abbildung 9 dargestellt sind:

- einer gleichartigen Mitgliederbeschaffenheit
- einer Beziehung dieser Mitglieder untereinander, welche zur Wiederkehr und einem Mehrwert für den Einzelnen führt
- einem gemeinsamen Interesse an Marken und/oder Produkten
- und einem Interaktionsraum als unterstützendes Medium



Abbildung 9: Elemente virtueller Communities

Quelle: in Anlehnung an Heitman et al. 2004, S.2

Die Besonderheit von virtuellen Communities liegt in der (partiellen) Anonymität der Interaktionspartner, welche – neben der Freiwilligkeit der Mitgliedschaft – sowohl Beitritt als auch Austritt erleichtern. Online Communities sind durch Unverbindlichkeit, Offenheit und Vertrauen gegenüber den anderen Mitgliedern gekennzeichnet (vgl. Heitmann et al. 2004, S.3).

Neben Weblogs können Communities ebenfalls einen starken Einfluss auf die Unternehmens- und/oder Produktreputation nehmen. Dies geschieht über die sogenannten Community Communicators, die in ihren Extremformen sowohl als Reputationsevangelisten als auch als Reputationsterroristen auftreten können. Erstere setzen sich positiv für das Unternehmen/Produkt ein, Zweitere zielen darauf ab, dem Unternehmen/Produkt größtmöglichen Schaden zuzufügen und dessen Reputation zu zerstören. Weitere weniger extreme Typologien der Kommunikatoren sind Information-Provider und konstruktive Kritiker. Erst genannte sind daran interessiert, wichtige und relevante Informationen über ein Unternehmen/Produkt zur Verfügung zu stellen. Letztere möchten Fehler und Missstände aufdecken und geben erst dann nach, wenn diese aufgeklärt und behoben sind (vgl. Meckel [2] 2008, S.126 f.). Die Verbreitung der Informationen und Meinungen der Kommunikatoren in den sozialen Netzwerken geschieht über virales Marketing⁷. Dabei spielt die Qualität und Kreativität der Inhalte eine entscheidende Rolle, da User die Nachricht nur dann weiterleiten, wenn ihr Interesse durch diese geweckt wurde und sie sich persönlich angesprochen fühlen.

Ein Vorteil bei der Verbreitung von Nachrichten kommt vor allen den Meinungsführern⁸ im WWW zugute. Meist verfügen diese über ein Expertenwissen in Bezug auf bestimmte Themengebiete, aufgrund dessen sie eine gewisse Autorität in ihrem Fachgebiet darstellen und bei wichtigen Entscheidungen oftmals von ihren Mitmenschen um Rat gefragt werden. Ihre besondere Stellung im sozialen Netzwerk ermöglicht es den Meinungsführern, als Multiplikatoren aufzutreten und ihre Meinung reichweitenstark im WWW mitzuteilen (vgl. Langner 2007, S.81).

Virtuelle Communities stellen demnach neben den Weblogs ebenfalls eine wichtige Informationsquelle zur Datensammlung in dieser Arbeit dar. Community Communicators, speziell Meinungsführer, besitzen die Möglichkeit, ihre Informationen und Meinungen über Unternehmen/Produkte per viralem Marketing im

⁷ Virales Marketing ist "the internet version of mouth-to-mouth marketing – e-mail messages or other marketing events that are so infectious that customers will want to pass them along to others." (Quelle: Kotler und Armstrong (2006): Principles of Marketing, 11. Auflage, Academic Internet Publishers Incorporated, New Jersey Prentice Hill, S.571).

⁸ Meinungsführer sind „Personen, die mit ihren Aussagen Hunderte, Tausende oder gar Millionen von Konsumenten erreichen und beeinflussen können.“ (Langner 2007, S.81).

WWW zu verbreiten und einer Vielzahl an Personen mitzuteilen. Aufgrund des in sozialen Netzwerken bestehenden Vertrauens untereinander vermitteln die Nachrichten eine hohe Glaubwürdigkeit und beeinflussen den Leser bei seiner persönlichen Meinungsbildung.

2.1.3 Online Forum - Einfluss des Third Party Endorsement

Ein Online Forum⁹ ist nach Rutenbeck (2006, S.184)

„A virtual site for conducting asynchronous online discussions. A forum is a discussion site or space where Internet users can log in to read and send messages to members within the group. The posted discussion messages are available for all members of the group to read and respond to at their convenience.“

In Internetforen werden demnach Diskussionen über bestimmte Themen geführt. Sobald ein Teilnehmer einen Thread (Gesprächsfaden) eröffnet hat, folgen zu meist Beiträge anderer Teilnehmer in Form von Fragen, Anmerkungen oder auch Tipps. Im Forum herrscht Meinungsfreiheit und Anonymität. Falls keine Verletzung der Foren-Regeln vorliegt, erscheinen die einzelnen Beiträge unzensiert (vgl. Bö nisch 2008, S.24 ff.).

Online Foren werden oftmals aufgesucht, wenn Informationen und Meinungen zu bestimmten Themen benötigt werden. Potenzielle Kunden informieren sich in den Foren über Produkte sowie deren Bewertung und vertrauen häufig auf das Third Party Endorsement der dort auftretenden Teilnehmer. Somit stellen Foren neben Weblogs und Communities einen wichtigen Bestandteil des Social Web-Monitorings durch Unternehmen dar, da auch hier ein direkter Einfluss auf das Kaufverhalten potenzieller Kunden besteht.

2.1.4 Bewertungsportal – Kundenrezensionen im Internet

Eine weitere Informationsquelle der Arbeit zur Datensammlung stellen Kundenrezensionen im Internet dar. Diese finden sich beispielsweise auf Online-Marktplätzen (z. B. www.Amazon.de) oder Bewertungs- und Vergleichsportalen (bspw. www.Ciao.de). Im Rahmen dieser Arbeit sind diese beiden Formen in einem Unterkapitel zusammengefasst, da nur die dort vorhandenen Rezensionen, jedoch nicht das Produktangebot und Geschäftsmodell, von Interesse sind. Wie

⁹ auch Bulletin Board genannt

auch bei den Online-Foren werden die Kundenrezensionen auf Online-Marktplätzen und Bewertungsportalen von potenziellen Kunden zur Information vor dem Kauf eines Produktes genutzt.

Bei Amazon werden dazu Sterne vergeben. Je mehr Sterne ein Produkt besitzt, desto positiver sind seine Rezensionen. Neben der Vergabe von Sternen können die Rezensenten zusätzlich eine textuelle Produktbewertung schreiben. Die Beurteilungen selbst werden ebenfalls von Amazon-Kunden bewertet, so dass eine qualitative Ordnung der positiven und negativen Rezensionen entsteht (vgl. Abbildung 10).

Kundenrezensionen
Apple iPhone 3G 16GB - Schwarz

64 Rezensionen

Durchschnittliche Kundenbewertung
 ★★★★★ (64 Kundenrezensionen)

Sagen Sie Ihre Meinung zu diesem Artikel

Eigene Rezension erstellen

Kundenrezensionen suchen

Nur in den Rezensionen zu diesem Produkt suchen

Die hilfreichste positive Rezension

159 von 169 Kunden fanden die folgende Rezension hilfreich:

★★★★★ **Top Handy mit Entwicklungspotential**

Ich bin seit Jahren Kunde hier bei Amazon und nutze bei meinen Einkäufen die Rezensionen sehr intensiv. Vieles habe ich schon über Produkte erfahren, was ich so nie erfahren hätte. Habe bisher aber nie eine Rezension geschrieben, weil ich gedacht habe, es sind schon genügend andere da. Jetzt sehe ich mich aber denn noch genötigt mal eine eigene zu schreiben, weil ich noch...

[Vollständige Rezension lesen >](#)

Vor 1 Monat von kitzbiker veröffentlicht

> Weitere Rezensionen anzeigen: [5 Sterne](#), [4 Sterne](#)

Die hilfreichste negative Rezension

46 von 60 Kunden fanden die folgende Rezension hilfreich:

★★★★☆ **Geiles Design, einige Schwächen, zu teuer...**

Also ich habe mein iPhone jetzt 3 Monate und ich bin teilweise begeistert, teilweise sehr enttäuscht.

Keine Frage, dass iPhone hat die Handy-Welt revolutioniert - aber das eher durch das schicke Design und nicht wegen der Funktionen und der Ausstattung.

Was das Design angeht, kann kein Handy nur teilweise mit halten und auch die neueren Handys...

[Vollständige Rezension lesen >](#)

Vor 6 Monaten von schnellie4ever veröffentlicht

> Weitere Rezensionen anzeigen: [3 Sterne](#), [2 Sterne](#), [1 Sterne](#)

Abbildung 10: Kundenrezensionen bei Amazon

Quelle: http://www.amazon.de/product-reviews/B001AXA056/ref=sr_1_2_cm_cr_acr_img?ie=UTF8&showViewpoints=1, Abruf am 21.07.2009

Ciao bietet eine ähnliche Beurteilungsmöglichkeit wie Amazon. Zusätzlich können bestimmte Merkmale des Produkts mit Punkten bewertet werden. Je mehr Punkte das Produkt in einer Kategorie erhält, desto besser. Beim iPhone werden u.a. Ausstattung, Klang, Akkulaufzeit und Design bewertet (vgl. Abbildung 11).

1-15 von 72 Erfahrungsberichten über Apple iPhone

Sortieren nach: **Nützlichkeit** | Bewertung | Datum

★★★★★ Der einzige iPhone-Bericht, den ihr lesen müsst!

Bewertung für Apple iPhone von Magdalena

Pro: Design, Verarbeitung, Funktionsvielfalt, durchdachtes Multitouch-UI / Touchscreen, einfache Handhabung, Akku, WiFi, macht glücklich!

Kontra: Preis, Vertragskosten, kein UMTS, viele Funktionen fehlen, keine reale Tastatur, keine MMS

...kompatibel, was einfach unakzeptabel ist. Apple bietet einen Adapter an... :(!!! (Das Gewicht beträgt 135g und liegt damit deutlich über federleichten Mobiltelefonen wie dem Razr2, aber ungefähr im Rahmen von vergleichbaren Geräten wie dem Blackberry. Für meinen Geschmack ist es ein paar Gramm zu schwer, wirklich störend oder belastend wirkt sich das in der Praxis aber nicht aus, auch nicht bei längeren Gesprächen. Der Touchscreen Wow! ...

[Bericht lesen](#)

Ciao Mitglieder bewerteten diesen Erfahrungsbericht insgesamt als **besonders hilfreich**

★★★★★ Reden ist Silber - Hören ist Gold

Bewertung für Apple iPhone von Thingo

Pro: Innovativ, benutzerfreundlich, revolutionär, viele Möglichkeiten

Kontra: Software noch in den Kinderschuhen

...berichten. Wer die Produkte von Apple ein wenig kennt, weiß, dass diese immer innovativ, wie auch schlicht und äußerst bedienungsfreundlich sind. Auch bei dem ersten Mobiltelefon von Apple haben die Technischen Entwickler mit knapp 300 Patenten die angemeldet worden sind zukunftsweisende und trendangegebende Feinheiten konzipiert. Gekauft habe ich mir das iPhone aufgrund der vielseitigen Möglichkeiten die es in nur einem Gerät vereint. Der die ...

[Bericht lesen](#)

Ciao Mitglieder bewerteten diesen Erfahrungsbericht insgesamt als **sehr hilfreich**

★★★★ Apple iPhone erste Generation - Vergleich

Bewertung für Apple iPhone von mgrasek100

Pro: Gute Verarbeitung, großer Akku, WLAN, Visual Voice Box

Kontra: kein UMTS, kein MMS (auch nicht updatefähig)

...EU... Das iPhone von der Firma Apple ist das erste fähige volle mobile Internetbrowsergerät unter den Handys, es ist eigentlich nicht nur ein Telefon, es ist, dank seines integrierten MP3 Player und vielen Speichers und des Breitbandinternet ein Multimediergerät. Vorteile gegenüber des iPhone 3 G gibt es kaum, höchstens die ALU Verkleidung und vielleicht der Akku, der zwar in der Version nicht wechselbar ist aber immerhin mehr Energie speichern ...

[Bericht lesen](#)

Ausstattung: **sehr hilfreich**

Klang: **sehr hilfreich**

Empfangsqualität: **sehr hilfreich**

Verarbeitung: **sehr hilfreich**

Akkulaufzeit: **sehr hilfreich**

Design: **sehr hilfreich**

20.11.2007

sehr hilfreich

04.05.2008
(05.05.2008)

sehr hilfreich

01.05.2009

Abbildung 11: Kundenrezensionen bei Ciao

Quelle: http://www.ciao.de/Erfahrungsberichte/Apple_iPhone_3G_7864901, Abruf am 21.07.2009

2.2 Business Intelligence und ihre Methoden

Der Begriff Business Intelligence verfügt über eine Reihe unterschiedlicher Definitionen. Eine etwas allgemeinere und für die Zwecke dieser Arbeit ausreichende Definition stammt von Loshin (2003, S.6), der Business Intelligence folgendermaßen beschreibt:

“The process, technology, and tools needed, to turn data into information, information into knowledge, and knowledge into plans that drive profitable business action. Business Intelligence encompasses data warehousing, business analytical tools, and content/knowledge management.”

Unter Business Intelligence werden demnach sowohl Prozesse als auch Technologien und Tools verstanden, die zur Informationsanalyse im Hinblick auf unternehmerische Entscheidungen dienen. Im analyseorientierten Sinne umfasst Business Intelligence laut Kemper et al. (2006, S.2 ff.) sämtliche Anwendungen, bei denen der Entscheider direkt mit dem System arbeitet. Neben OLAP und MIS/EIS gehören hierzu unter anderem auch Systeme des Data Mining und Text Mining. Einen Überblick der Facetten von Business Intelligence gibt Abbildung 12.

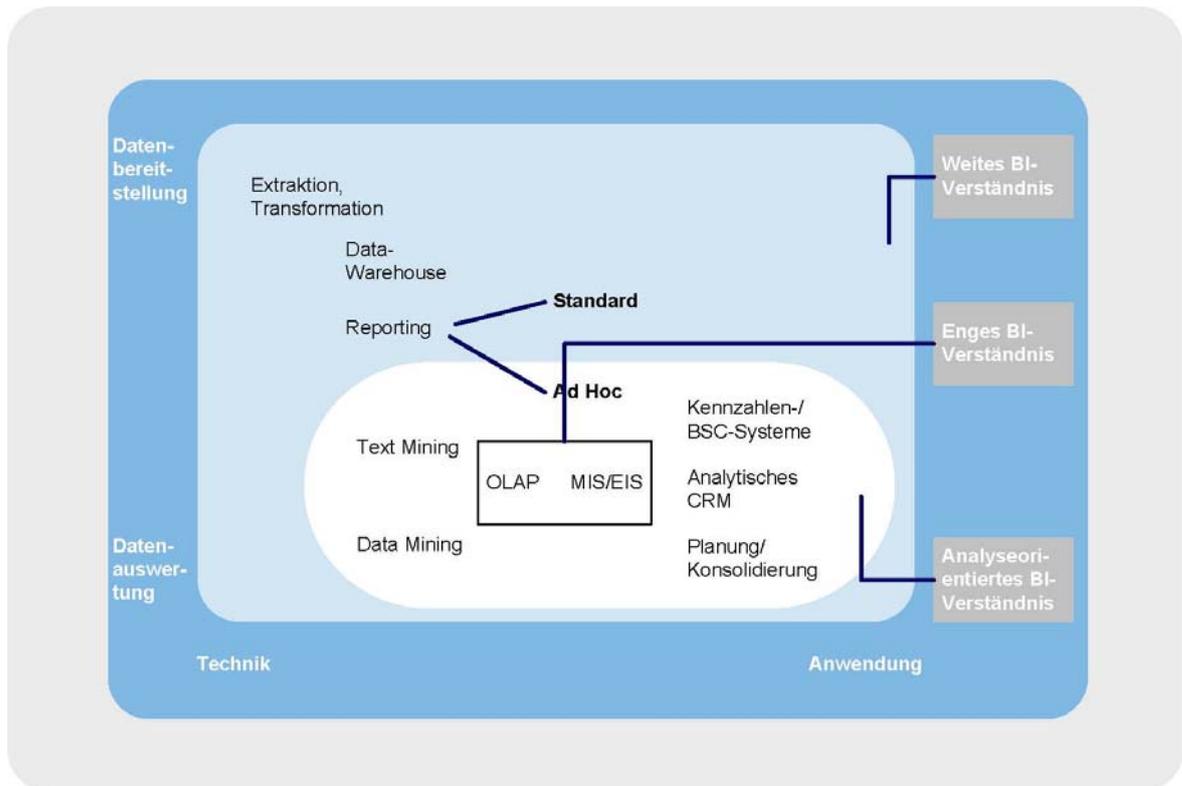


Abbildung 12: Unterschiedliche Facetten der Business Intelligence

Quelle: Gluchowski 2001, S.7

Anschließend werden verschiedene, für die Arbeit relevante Data Mining Methoden vorgestellt. Dazu zählen Data Mining, Text Mining und das daraus entstandene Opinion Mining, welches das Thema dieser Arbeit darstellt.

2.2.1 Data Mining

Als Wegbereiter des Data Mining gelten Frawley, Piatetsky-Shapiro und Matheus (Frawley et al. 1991). Sie beschreiben Data Mining als „die Extraktion und Entdeckung von implizitem, bisher nicht bekanntem und potenziell nützlichem Wissen aus Daten“ (Frawley et al. 1991, S.58). Fünf Jahre später verwenden Fayyad et al. (1996, S.40 f.) folgende erweiterte Definition für das Knowledge Discovery in Databases (KDD), welches das Data Mining umfasst:

„Knowledge Discovery in databases is the nontrivial process of identifying valid, novel, potential useful, and ultimately understandable pattern in data.“

Der Definition nach ist der Prozess ein neuer Analyseansatz (*nontrivial*), der statistisch sichere Aussagen beinhaltet (*valid*). Er fokussiert auf neue Erkenntnisse (*novel*), die nicht immer offensichtlich bewertbar, aber nützlich (*potential useful*) und für den Anwender verständlich (*ultimately understandable*) sind. Anstelle von In-

formation wird der Begriff „pattern“ verwendet. Demnach geht es primär um die Mustererkennung und erst sekundär um die Informationsgewinnung. Zudem sprechen Fayyad et al. vom „process“. Beim Data Mining handelt es sich also um einen Prozess von Datenanalysen (vgl. Petersohn 2005, S.8 ff.). Abbildung 13 stellt den Prozess des Knowledge Discovery in Databases (KDD) und den darin enthaltenen Unterprozess Data Mining im Überblick dar.

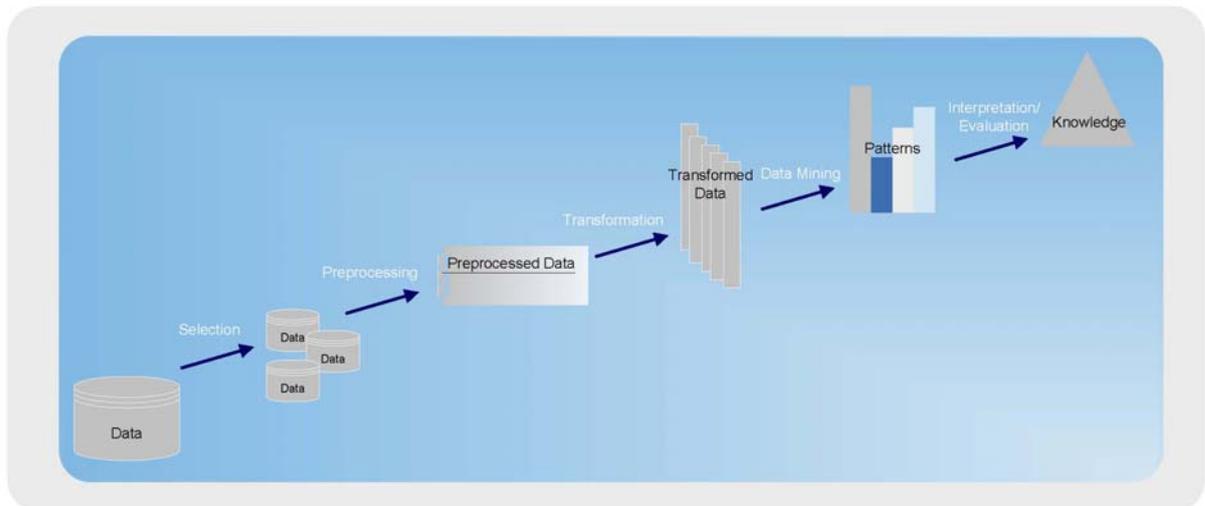


Abbildung 13: Der KDD-Prozess

Quelle: Fayyad et al. 1996, S.41

Die Unterscheidung zwischen den Begriffen Data Mining und Knowledge Discovery von Fayyad et al. ist jedoch umstritten. Die meisten Autoren verwenden die Begriffe bisher synonym (vgl. Adriaans und Zantinge 1998, S.5; Liu 2008, S.6).

Zur Durchführung des Data Mining gibt es eine Vielzahl unterschiedlicher Vorgehensmodelle. In dieser Arbeit wird der 1996 von DaimlerCrysler, SPSS, NCR und OHRA entwickelte *CRoss Industry Standard Process for Data Mining* (CRISP-DM, vgl. Abbildung 14) verwendet, weil das CRISP-DM Modell das bislang einzige standardisierte Prozessmodell ist, dessen Schwerpunkt weniger auf technischen, sondern vielmehr auf wirtschaftlichen Aspekten beruht. Ein weiterer Vorteil des CRISP-DM Modells ist dessen Verständnis des Data Mining als Prozess. Aufgrund dessen gewährleistet es einen stärkeren Anschluss der Data Mining-Ergebnisse an die unternehmerischen Vorhersagen und Entscheidungen (vgl. Kneip 2008, S.12).

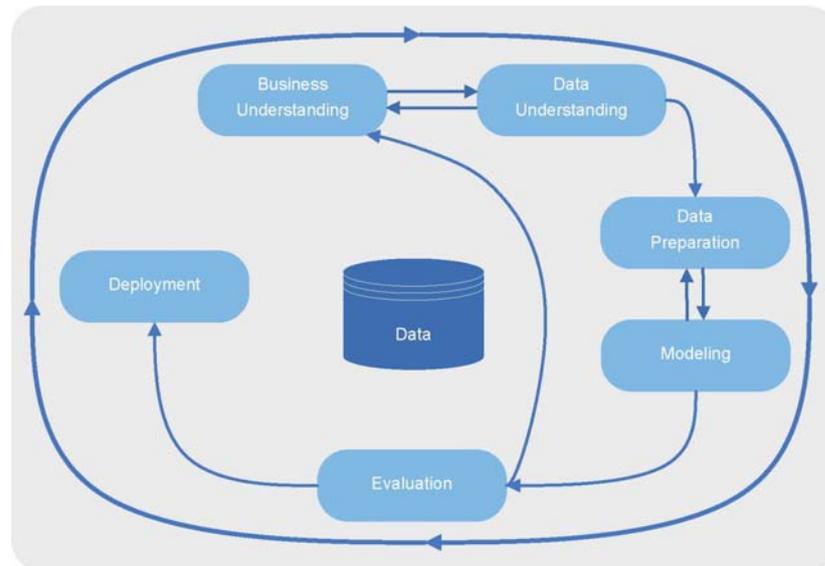


Abbildung 14: Phasen des CRISP-DM Modells

Quelle: <http://www.crisp-dm.org/Process/index.htm>

Das Modell gliedert sich in sechs iterative Phasen, welche zu einem späteren Zeitpunkt dieser Arbeit (in Kapitel 3) näher beschrieben werden.

2.2.2 Text Mining

Nachdem zuvor das Data Mining definiert wurde, folgt anschließend eine Erklärung des Text Mining, welches ebenfalls zu den analyseorientierten Business Intelligence-Methoden zählt. Das Text Mining ist aus der Inhaltsanalyse entstanden, welche durch Holsti (1969, S.16) folgendermaßen definiert wird:

„Content Analysis is any technique for making inferences by objectively and systematically identifying specified characteristics of messages.“

Inhaltsanalysen werden bereits seit mindestens 100 Jahren durchgeführt. Vorläufer waren textanalytische, textvergleichende oder auch hermeneutische Ansätze, wie beispielsweise Bibelanalysen und die Traumdeutungen Sigmund Freuds um 1900 (vgl. Mayring 2000). Laut Früh (2007) setzte die Entwicklung der Inhaltsanalyse zu einer bis heute wichtigen und eigenständigen Methode kurz nach der Jahrhundertwende ein. Ihren ersten Höhepunkt erreichte sie in der Zeit des Zweiten Weltkrieges mit der zunehmend massenhaften Verbreitung von Kriegspropaganda und der Expansion des Medienangebots. Zu dieser Zeit legten Lazarsfeld (im Office of Radio Research) und Lasswell (im Hoover Institute) in den USA die Grundlage einer quantitativ orientierten Analyse von Massenmedien. Nach dem Zweiten Weltkrieg fand in den USA dann auch eine intensive Auseinandersetzung mit den theoretischen Aspekten der Inhaltsanalyse statt. Im Jahr 1955 erfolgte an

der Universität in Illinois die „Allerton House Conference“, an der namhafte Inhaltsanalytiker teilnahmen. Ziel dieser Konferenz war es, die Methode einem größeren Kreis von Wissenschaftlern vorzustellen (vgl. Früh 2007, S.12). Es folgte eine interdisziplinäre Erweiterung und Differenzierung in den 60er Jahren als sich die Methode auch auf weitere Anwendungsgebiete ausweitete und eine Verfeinerung der Vorgehensweisen stattfand (vgl. Mayring 2000). Zur gleichen Zeit wurde unter Einfluss von Gerbner et al. (1966) auch die Entwicklung der maschinellen Inhaltsanalyse vorangetrieben. Seither erfolgte eine Verlagerung des methodischen Forschungsinteresses auf die elektronische, computerunterstützte Inhaltsanalyse. In den letzten drei Jahrzehnten gab es sowohl eine stetig wachsende Anwendungshäufigkeit der codebasierten Inhaltsanalyse als auch eine verstärkte Auseinandersetzung mit deren methodischen Vorgehen. Die Inhaltsanalyse wird heute in einer Vielzahl von Anwendungsgebieten (u.a. in der Soziologie, Psychologie, Informationstheorie, Ethnologie und den Kunst- und Geisteswissenschaften) eingesetzt (vgl. Früh 2007, S.15).¹⁰

Das aus der Inhaltsanalyse entstandene Text Mining zählt zur „Knowledge Discovery in Texts“ (KDT), welche eine Erweiterung des zuvor definierten KDD darstellt. Zur Begriffserklärung des KDT erweitert Feldman (1995) die Definition des KDD von Fayyad et al. (1996) folgendermaßen:

„[...] text analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns over the text.“

Demnach ist Text Mining zur Wissens- und Informationsentdeckung sowie Mustererkennung von textuellen Daten geeignet. Mit der zunehmenden Entwicklung des WWW und der Suchmaschinen wuchs die Bedeutung des Text Mining zu einer heute gewichtigen Business Intelligence-Methode (Mittra und Atcharya 2003, S.14 f.). Das Filtern und die Analyse von Meinungen aus Texten fallen unter die Methode des Opinion Mining. Sowohl beim Data- und Text- als auch beim Opinion Mining können die Analysedaten auch aus dem Internet bezogen werden. In diesem Fall handelt es sich um Web Mining (vgl. Abbildung 15).

¹⁰ Weiterführende Literatur zur Inhaltsanalyse u.a. Berelson und Lazarsfeld 1948 und 1952; Gerbner et al. 1969; Krippendorff 1980 und 2004; Lasswell 1948.

„Web mining refers to the use of data mining techniques to automatically retrieve, extract, and evaluate information for knowledge discovery from Web documents and services.” (Mitra und Atcharya 2003, S.16)

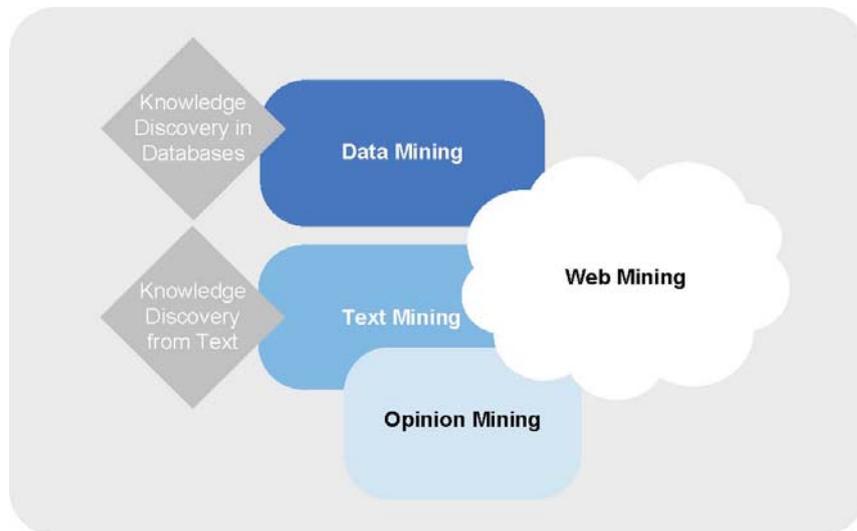


Abbildung 15: Abgrenzung der Mining-Methoden

2.2.3 Opinion Mining

Beim Opinion Mining¹¹ werden Informationen online aus dem WWW oder aus Offline-Dokumenten erfasst, extrahiert und ausgewertet. Jedoch liegt der Fokus auf der Analyse der dort enthaltenen Meinungen. Die Forschung im Bereich Opinion Mining begann mit der Identifikation von Meinungswörtern. Viele Forscher arbeiteten an der Auffindung solcher Wörter und deren semantischer Orientierung (positiv, negativ oder neutral). Es wurden linguistische Regeln erstellt, welche den Forschern dazu verhalfen, Meinungswörter und deren Orientierung aus großen Textkörpern zu filtern. Diese Methoden wurden angewandt, erweitert und optimiert, und gleichzeitig wurden Opinion Mining Probleme und deren Lösung aufgezeigt. Liu definiert das Opinion Mining folgendermaßen:

“Given a set of evaluative text documents D that contain opinions (or sentiments) about an object, opinion mining aims to extract attributes and components of the object that have been commented on in each document $d \in D$ and to determine whether the comments are positive, negative or neutral.”

¹¹ Synonyme für Opinion Mining: Web Content Mining (Liu, B. 2008, S.7), Sentiment Mining (Liu, B., S.1).

Eine weitere Definition stammt von Lee et al. (2008, S.1):

„The task of analyzing such data, collectively called customer feedback data, is known as opinion mining.“

Bewertende Texte können aus drei verschiedenen Perspektiven betrachtet werden (vgl. Liu 2008, S.411 f.). Folgende Herangehensweisen sind möglich:

- Sentiment Klassifikation
- Eigenschaften-basiertes Opinion Mining
- vergleichendes Sentence und Relation Mining

Die Sentiment Klassifikation behandelt Opinion Mining als Klassifikationsproblem auf dem Dokumenten-Level. Das heißt die Meinungstexte werden den Kategorien „positiv“ oder „negativ“ zugeordnet. Details über die Vorlieben oder Abneigungen der Rezensenten werden nicht aufgedeckt. Das Eigenschaften-basierte Opinion Mining befindet sich auf dem Satz-Level. Hier geht es um das Herausfinden von Details, also Aspekten der bewerteten Objekte, die von den Rezensenten genannt werden. Beispielsweise werden positiv oder negativ beurteilte Produkteigenschaften aufgedeckt. Im Satz *„Die Akkuleistung ist nicht ausreichend“* wird zum Beispiel die Akkuleistung negativ beurteilt. Das vergleichende Sentence und Relation Mining umfasst die direkte Gegenüberstellung von gleichartigen Objekten. Ein Beispiel hierfür ist der Satz *„Die Akkuleistung von Produkt A ist viel besser als die von Produkt B“*. In dieser Arbeit erfolgt die Ausarbeitung der zweiten Herangehensweise, dem Eigenschaft-basierten Opinion Mining (vgl. Liu 2008, S.417 ff.).

Kernpunkt des Opinion Mining ist die Identifikation von Stimmungen, welche in Texten ausgedrückt werden. Zusätzlich sind diese mit positiven, negativen oder neutralen Meinungen zu konnotieren. Opinion Mining umfasst die Identifikation von

- subjektiv geäußerten Stimmungen,
- Polarität und Stärke der Äußerungen
- und deren Beziehung zum untersuchten Thema

(vgl. u.a. Nasukawa und Yi 2003, S.71; Conrad und Schilder 2007, S.232; Ding et al. 2008, S.231). Wichtig ist dabei die linguistische Unterscheidung zwischen objektiven Wörtern, welche Fakten beschreiben, und subjektiven Wörtern, welche

Meinungen ausdrücken (vgl. Lee et al. 2008, S.1; vgl. auch Abbildung 1). Ziel des Opinion Mining ist die Extraktion von Kundenmeinungen (Feedback) über Produkte und die Nutzung dieses Wissens zur Optimierung der Produkte (vgl. Lee et al. 2008, S.1).

Das Opinion Mining als Methode mit prozessproduzierten Daten bietet gegenüber den klassischen Methoden mit forschungsproduzierten Daten einige Vorteile. Da sich die Wissensentdeckung beim Opinion Mining auf bestehende Texte stützt, entfällt die beschwerliche und kostenintensive Forschung und Analyse mit Fragebogendaten, welche vor allem in Zusammenhang mit großen Datenmengen und offenen Fragen als kritisch zu betrachten ist (vgl. u.a. Morinaga et al. 2002, S.341; Nasukawa und Yi 2003, S.70). Durch die Analyse und Auswertung frei geäußelter Rezensionen im Internet hilft Opinion Mining den Unternehmen, ihre Kunden besser zu verstehen, indem Trends und Stimmungen frühzeitig erkannt werden. Somit können bestehende und zukünftige Produkte nach den selbst genannten Anregungen und Wünschen der Kunden (um)gestaltet werden (vgl. Funk et al. 2008, S.1). Typische Probleme, die sich im Rahmen der herkömmlichen Primärforschungsmethoden ergeben, entfallen beim Opinion Mining. Als Nachteil muss jedoch die Repräsentativität und Generalisierbarkeit genannt werden. Da beim Opinion Mining im Web nur die Meinungen von Internetusern analysiert werden, findet hier eine gewisse Self Selection statt. Nicht-Internet-User werden von der Analyse ausgeschlossen. Die klassischen Methoden hingegen bieten eine Chance die Repräsentativität und Generalisierbarkeit zu erreichen, indem eine umfassende Stichprobe gezogen wird. Ein Vergleich der Methoden ist in Abbildung 16 dargestellt.

Methoden	Herkömmliche Primärforschung	Datensammlung im Web 2.0
<i>Persönliche Befragung</i>	Einfluss des Interviewers auf den Befragten	Freie Meinungsäußerung der Rezensenten im Internet → keine äußeren Einflüsse
	Überwindung räumlicher Distanzen	Meinungen sind im Internet für jeden einfach zugänglich
	Hoher Mitarbeiteraufwand → sehr kostenintensiv	Geringer Mitarbeiteraufwand → kostengünstig
<i>Schriftliches Interview</i>	Geringe Rücklaufquoten	Das Internet verfügt über eine Vielzahl an Meinungen
	Die benötigten Informationen müssen erst erhoben werden → langer Durchführungszeitraum	Die benötigten Informationen sind direkt verfügbar → relativ kurzer Durchführungszeitraum
<i>Geschlossener Fragebogen</i>	Vorgegebene Antworten, keine Alternativen möglich, kein Interpretationsspielraum → Einschränkung in der Antwort	Der Rezensent äußert seine Meinung ohne jegliche Vorgaben → keine Einschränkung
<i>Online Fragebogen</i>	Schnelligkeit, hohe Anonymität und geringe Kosten	Schnelligkeit, hohe Anonymität und geringe Kosten
	Bei geschlossenen Fragen: Einschränkung in der Antwort	Keine geschlossenen Fragen → Keine Einschränkung in der Antwort
<i>Offener Fragebogen</i>	Hoher und stark subjektiver Informationsgehalt	Hoher und stark subjektiver Informationsgehalt

Abbildung 16: Primärforschung vs. Datensammlung im Opinion Mining

Der Onlinefragebogen bietet mittlerweile ähnliche Möglichkeiten bei der Datensammlung wie das Opinion Mining, da diese Erhebungsmethode weniger beschwerlich und kostenintensiv gestaltet werden kann. Die Methode des Fragebogens mit offenen Fragen ist jedoch die einzige, mit der die Datensammlung des Opinion Mining vergleichbar ist. Beide besitzen einen hohen und stark subjektiven Informationsgehalt, sind aber auch schwer auszuwerten, da sehr viel Text zu analysieren ist.

Der Opinion Mining Prozess ist dem Data Mining Prozess sehr ähnlich, Unterschiede finden sich – wie zuvor aufgezeigt – lediglich in der Datensammlung.

Im Rahmen des Data Mining befinden sich die notwendigen Daten meist schon gesammelt und gespeichert in einem Data Warehouse. Beim Opinion Mining zählt das Sammeln von Daten zu einer wesentlichen Aufgabe, da das Webcrawling eine große Anzahl an Webseiten umfasst. Nach Abschluss der Datensammlung gleicht sich die Vorgehensweise beider Methoden wieder an (Liu 2008, S.7). Aus diesem Grund lehnt sich das Vorgehen in dieser Arbeit auch dem in Kapitel 2.2.1 vorgestellten CRISP-DM Modell an.

Auf Basis der dargestellten theoretischen Grundkenntnisse erfolgt im nächsten Kapitel die praktische Anwendung des Opinion Mining im Rahmen einer empirischen Untersuchung.

3 Empirische Untersuchung

Die empirische Untersuchung beinhaltet zunächst die Entwicklung eines Vorgehensmodells. Aufbauend auf diesem Modell erfolgt dessen praktische Umsetzung. Dazu wird das Opinion Mining auf das zu untersuchende Produkt angewandt. Anschließend folgt die Auswertung der im RapidMiner erhaltenen Ergebnisse. Das Kapitel schließt mit Implikationen für die Forschung und Praxis, welche sich aus den Opinion Mining Resultaten ergeben.

3.1 Entwicklung eines Vorgehensmodells

Beim Opinion Mining soll nach der Methode des aus sechs iterativen Schritten bestehenden CRISP-DM Referenzmodells (siehe Kapitel 2.2.1 in Abbildung 14) vorgegangen werden. Da es sich beim Opinion Mining jedoch um eine spezielle Form des Data Mining handelt, wird das CRISP-DM Modell leicht modifiziert angewandt. In diesem Kapitel erfolgt zunächst die konzeptionelle Beschreibung des Vorgehensmodells. Anschließend folgt in Kapitel 3.2 die praktische Anwendung des Konzepts auf das Opinion Mining.

3.1.1 Business Understanding

Die einführende Phase fokussiert auf das Verständnis der Anforderungen und Ziele des Projekts aus der Unternehmensperspektive, um anschließend eine Problemdefinition aus der Data Mining Perspektive zu ermöglichen und einen vorläufigen Projektplan zur Zielerreichung zu erstellen (vgl. u.a. Kneip 2008, S.12 f.; McCue 2006, S.50).

Das Business Understanding beginnt mit der Auswahl des zu untersuchenden Produktes und der anschließenden Produktbeschreibung. Es ist bedeutsam einen Aufschluss darüber zu geben, aus welchen betriebswirtschaftlichen Gründen die Analyse des ausgewählten Produktes von Interesse ist. Neben der Produktauswahl erfolgt zudem die Selektion und Bewertung verschiedener Opinion Mining Tools und die Entscheidung für eines der Tools, welches für die Analysezwecke am geeignetsten erscheint. Nachdem sowohl das Produkt als auch das passende Analysetool ausgewählt wurden, kann mit der eigentlichen Analyse begonnen werden (vgl. Abbildung 17).

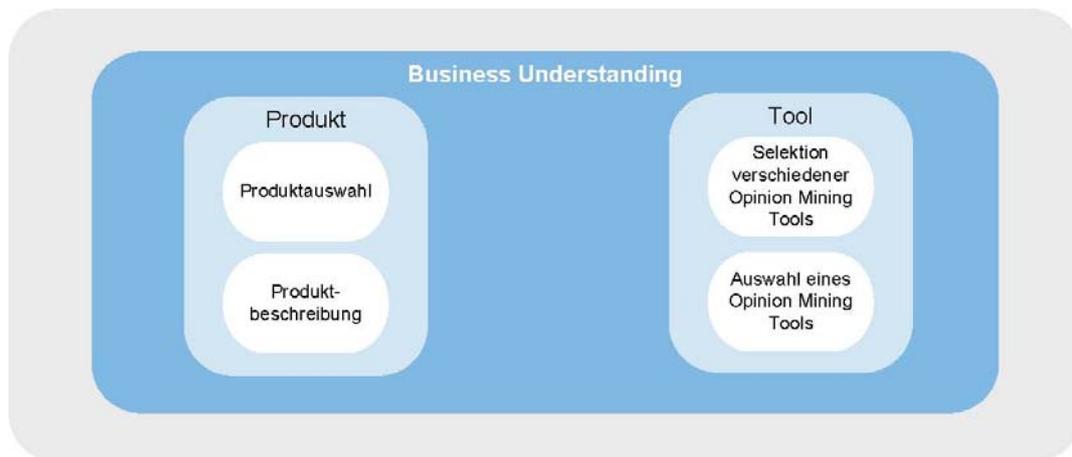


Abbildung 17: Business Understanding

Quelle: in Anlehnung an Hu und Liu 2004 [1] und [2], S.170

3.1.2 Data Understanding

Im Data Understanding stehen die Rohdaten im Vordergrund. Es beinhaltet die anfängliche Datensammlung und daran anschließend das Verstehen der Daten. Dazu werden diese beispielsweise auf zu überwindende Qualitätsprobleme oder deren Nützlichkeit und Verfügbarkeit für die jeweilige Problemstellung überprüft. Zudem können in dieser Phase erste Modelle zur Mustererkennung erstellt werden (vgl. u.a. Kneip 2008, S.13; Olsen und Delen 2008, S.9).

Im Rahmen des Opinion Mining beginnt das Data Understanding mit einer Daten-/Meinungssammlung über das zu analysierende Produkt im World Wide Web. Dazu werden passende Blogs, Communities, Diskussionsforen und Bewertungsportale selektiert. Die Social Websites bieten sich als ausgezeichnete Informationsquelle an, da sie (wie auch schon in Kapitel 2.1 beschrieben) einen starken Einfluss auf das Kaufverhalten potenzieller Kunden nehmen können. Grassroot-Bewegungen und virales Marketing führen zu einer rasanten Informations- und Meinungsverbreitung im Internet. Zudem führt die im virtuellen Raum vorherrschende Anonymität zu einem vermehrten Vertrauen der User untereinander, welches wiederum eine verstärkte Glaubwürdigkeit und ein gesteigertes 3rd Party Endorsement mit sich bringt. Nach Abschluss des Data Understanding finden sich die Daten/Meinungen gesammelt in einer Review-Datenbank (vgl. Abbildung 18).

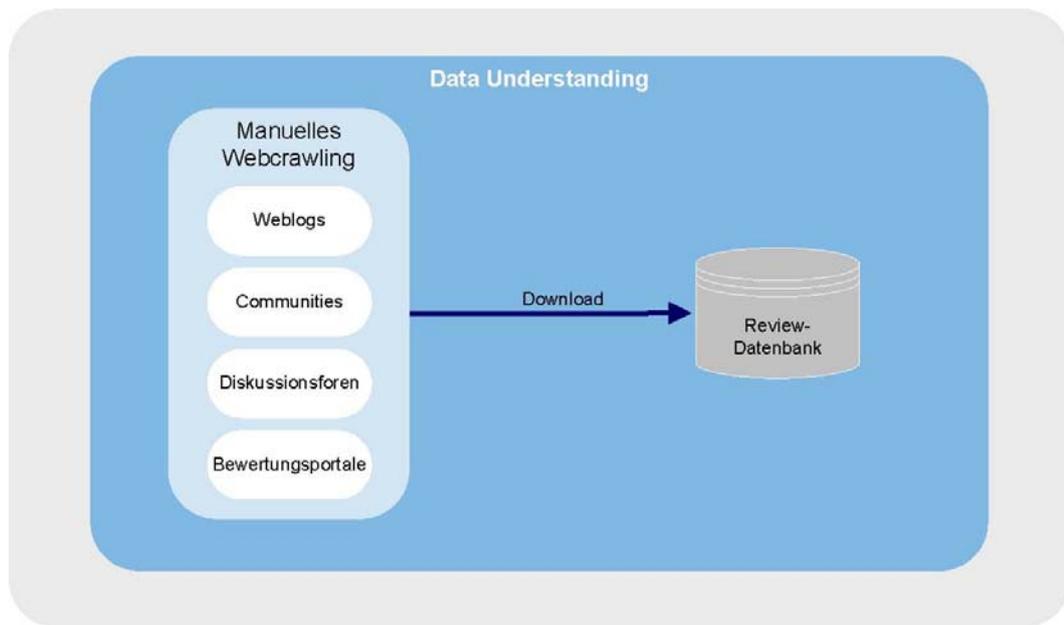


Abbildung 18: Data Understanding

Quelle: in Anlehnung an Hu und Liu 2004 [1] und [2], S.170

3.1.3 Data Preparation

Ziel dieser Phase ist die Konstruktion des finalen Datensets, welches eine möglichst hohe Datenqualität aufweisen sollte. Die Data Preparation umfasst Aktivitäten wie die Datenselektion, -bereinigung, -formatierung und -zusammenführung (vgl. u.a. Gabriel et al. 2009, S.129 f.; Olsen und Delen, S.9). Folgende Kriterien zählen zur Beurteilung der Datenqualität (vgl. Schaarschmidt und Hermann 2002):

- Vollständigkeit: die Verwendung sämtlicher für die Analyse notwendiger Daten und die Vermeidung von Leerwerten innerhalb des Datensets
- Konsistenz: die Widerspruchsfreiheit der Daten
- Genauigkeit: das Vorliegen der Daten in der für die Analyse erforderlichen Granularität
- Korrektheit: die Vermeidung fehlender und falscher Daten innerhalb des Datensets
- Aktualität: die Verwendung ausreichend aktueller Daten für die Analyse

Im Rahmen dieser Arbeit wird zunächst die Frequent Feature Datenbank erstellt. Dazu erfolgt das Part-of Speech (POS) Tagging¹², um die in den Meinungsätzen

¹² Part-of-Speech (POS) Tagging stellt eine Technik zur Aufbereitung der natürlichen Sprache dar. In Deutsch: Part of Speech = Wortart, Wortklasse, lexikalische Kategorie
Weiterführende Literatur zum POS-Tagging: Santorini 1990.

enthaltenen Wörter in ihre Wortklassen aufzuteilen. Anschließend werden die in den Rezensionen am häufigsten genannten Produkteigenschaften (Frequent Features) identifiziert. Darauf baut dann die Extraktion und Identifikation der Opinion Words (Datenbereinigung) auf. Dazu werden die Meinungswörter aus der Frequent Feature Datenbank gefiltert und anschließend positiv bzw. negativ konnotiert. Das abschließende Ergebnis der Data Preparation Phase ist eine Opinion Word Datenbank (Datenzusammenführung), die alle Meinungsworte beinhaltet, welche das zu analysierende Produkt beschreiben (vgl. Abbildung 19). Anschließend kann mit der eigentlichen Modellierung begonnen werden.

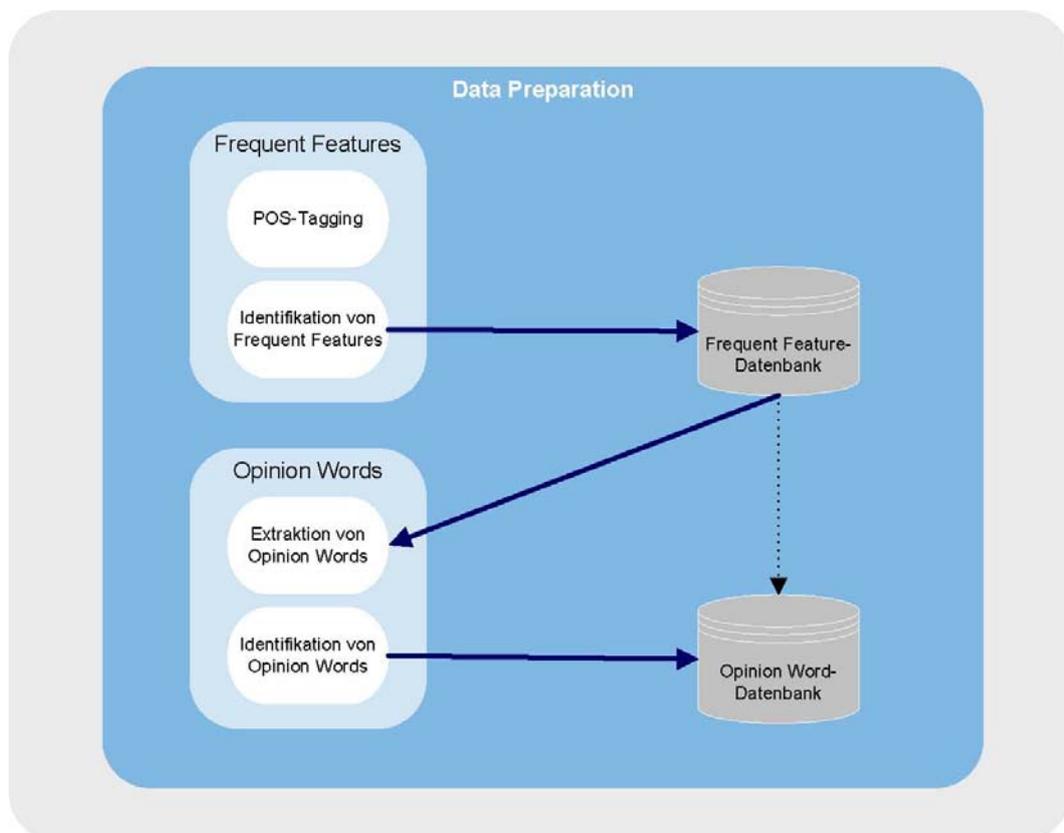


Abbildung 19: Data Preparation

Quelle: in Anlehnung an Hu und Liu 2004 [1] und [2], S.170

3.1.4 Modeling

Im Anschluss an die Datenvorbereitung steht das Modeling. In dieser Phase können verschiedene Data Mining Verfahren mit dem Zweck der Kategorisierung, Klassifikation, Prognose und Abhängigkeitsanalyse zum Einsatz kommen (vgl. Gabriel et al. 2009, S.134). Aufgrund der oftmals verschiedenen Voraussetzungen der Data Mining Techniken kann sich die Notwendigkeit ergeben, zur vorangegangenen Data Preparation Phase zurückzukehren (vgl. Kneip 2008, S.13).

Im Zuge des Opinion Mining umfasst das Modeling einen induktiven Lernprozess¹³. Analog zum Menschen lernen die Maschinen aus vergangenen Erfahrungen, um neues Wissen zur Steigerung der Leistungsfähigkeit zu erwerben. Da Maschinen jedoch keine „Erfahrungen“ besitzen, lernen sie von Daten, welche Ereignisse aus der Vergangenheit repräsentieren (vgl. Liu 2008, S.55). Die zum Lernen genutzte Datenmenge nennt sich Trainingsdaten¹⁴. Sie besteht aus einem Teil der ursprünglichen Daten. Der restliche Teil dieser Daten wird zum Test verwendet, diese Datenmenge beinhaltet dementsprechend die sogenannten Testdaten¹⁵. In der Trainingsphase bedient sich ein Lernalgorithmus der Trainingsdaten, um ein Klassifikationsmodell zu erstellen. In der nachstehenden Testphase erfolgt der Test dieses Modells anhand der Testdaten, um eine Aussage über die Genauigkeit (Accuracy) der Klassifikation zu ermöglichen (vgl. Liu 2008, S.58; McCue 2006, S.153 f.). Wenn die Genauigkeit zufriedenstellen ist, bildet das Modell die reale Wirklichkeit ab und kann zielgenaue Vorhersagen über die Einordnung neuer Klassen in bestehende Klassen treffen. Im Falle einer nicht zufriedenstellenden Genauigkeit sollte ein anderer Lernalgorithmus zur Modellerstellung gewählt werden. Eine Übersicht von Trainings- und Testablauf beinhaltet Abbildung 20.

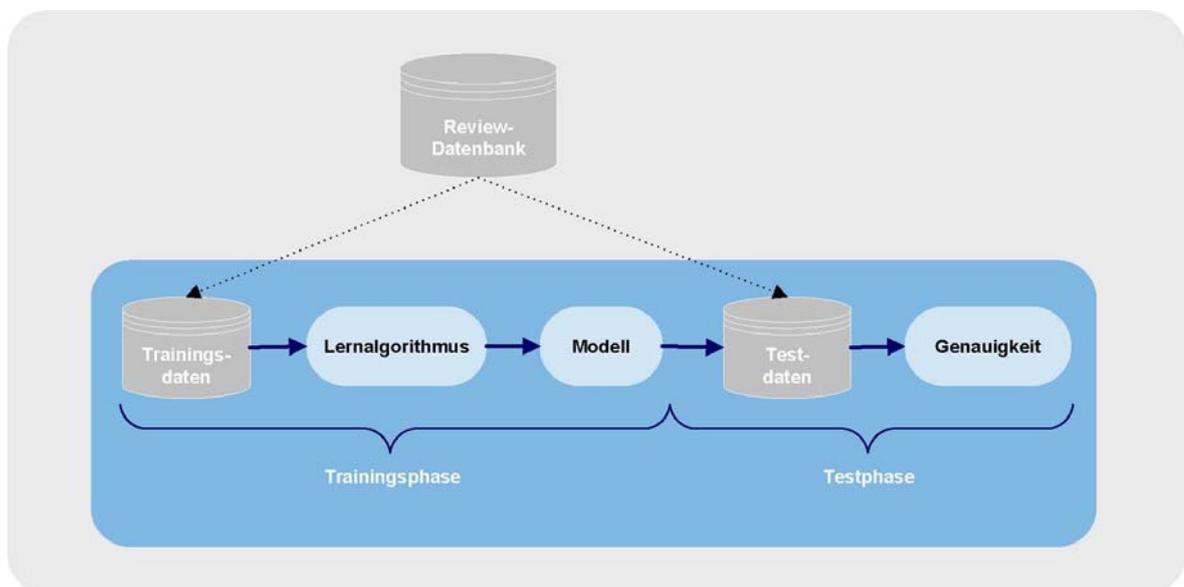


Abbildung 20: Ablauf von Training und Test

Quelle: in Anlehnung an Bramer 2007, S.80 und Liu 2008, S.58

¹³ induktives Lernen = supervised (beaufsichtigtes) Learning oder Klassifikation. Das Gegenteil zum supervised Learning stellt das unsupervised (unbeaufsichtigte) Learning dar. In diesem Fall sind die Klassen unbekannt und die Aufgabe des Lernalgorithmus besteht darin, die Klassen automatisch zu generieren.

¹⁴ weitere Bezeichnung: Trainingsmenge

¹⁵ weitere Bezeichnungen: Testmenge oder unseen (unsichtbare) data

Zur Umsetzung von Training und Test wird aus der Review Datenbank eine Trainingsmenge entnommen, die anschließend vorverarbeitet wird. Im Rahmen des Tests wird die Trainingsmenge in verschiedene Modelle eingespeist. Die jeweiligen Ergebnisse werden dann untereinander verglichen. Es folgt die Auswahl des optimalen Modells (vgl. Abbildung 21). Ziel des Modeling ist die Ermittlung eines Analyseergebnisses, welches die in der Phase des Business Understanding formulierten Anforderungen und Ziele abdeckt und sich durch eine hohe Aussage- und Interpretationsfähigkeit auszeichnet (vgl. Gabriel et al. 2009, S.13).

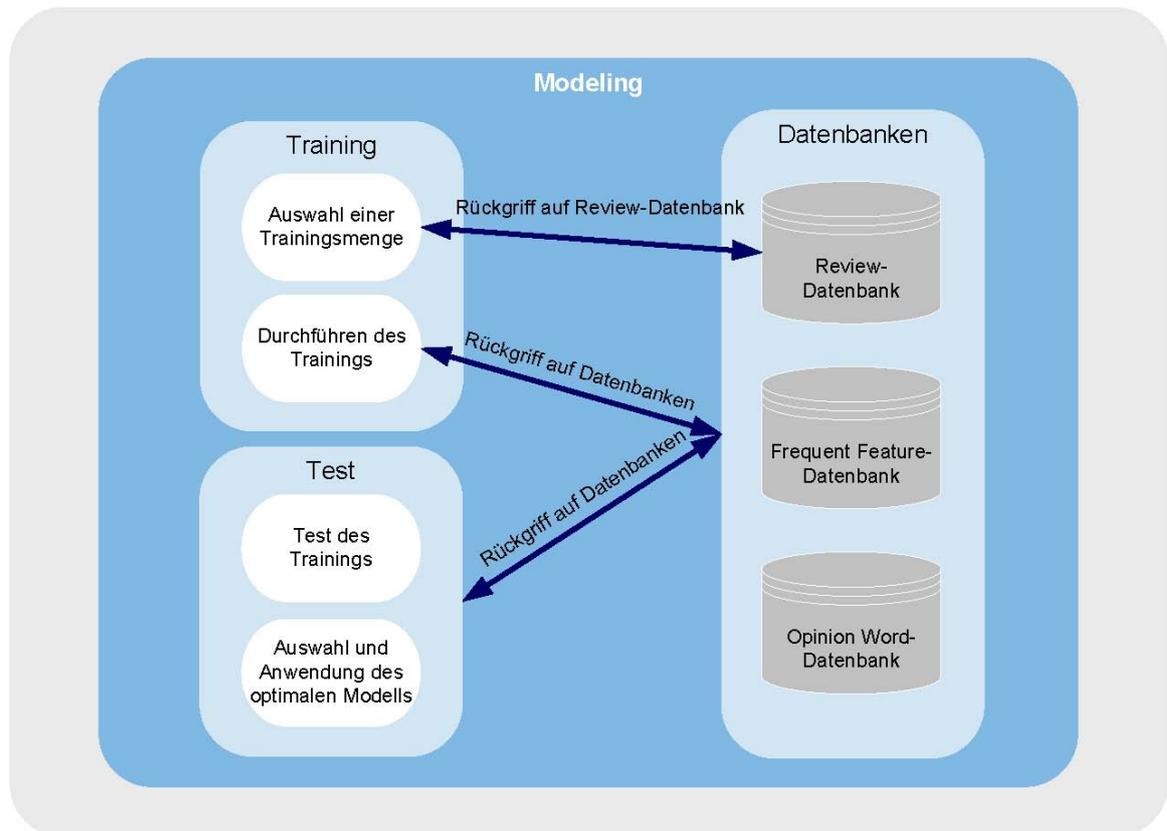


Abbildung 21: Modeling

Quelle: in Anlehnung an Hu und Liu 2004 [1] und [2], S.170

3.1.5 Evaluation & Deployment

Die Evaluation beinhaltet eine kritische Beurteilung des Vorgehens. Dazu zählen die Bewertung der Opinion Mining Ergebnisse ebenso wie die Bewertung des gesamten Vorgehensmodells (vgl. Gabriel et al. 2009, S.138). Die in der Modeling-Phase erhaltenen Ergebnisse sind hinsichtlich der im Business Understanding definierten Anforderungen und Ziele zu interpretieren und bewerten. Zudem erfolgt eine kritische Überprüfung der Qualität des Vorgehens. Dazu wird das ausgewählte Opinion Mining Tool beurteilt. Dieses wird gründlich auf Qualität- und Zielerreichung überprüft, um herauszufinden, ob es den im Vorgehensmodell entstandenen Anforderungen gerecht wurde. Dabei stehen vor allem die Aufdeckung von

Schwachstellen und die Erarbeitung von Verbesserungspotenzialen im Vordergrund.

Das Deployment umfasst die endgültige Auswertung der Ergebnisse für Fragestellungen der betrieblichen Praxis (vgl. Kneip 2008, S.13). Abschließend erfolgt die zielgruppengerechte Aufbereitung und Präsentation der durch das Opinion Mining erhaltenen neuartigen Informationen (vgl. Abbildung 22).

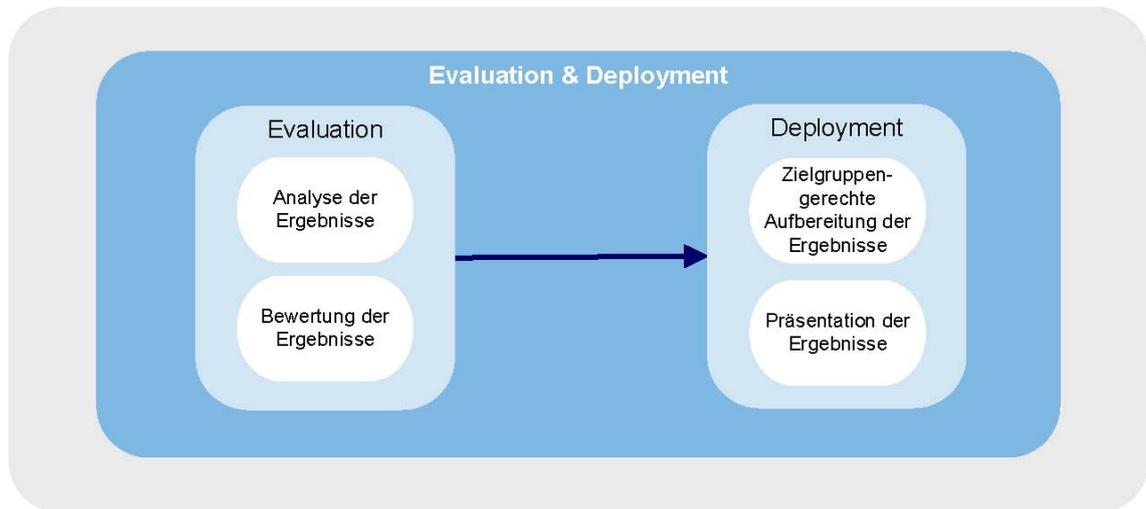


Abbildung 22: Evaluation & Deployment

Quelle: in Anlehnung an Hu und Liu 2004 [1] und [2], S.170

Zur Veranschaulichung stellt Abbildung 23 das Vorgehensmodell noch einmal im gesamten Überblick dar. Beginnend mit der Phase des Business Understanding erfolgt die Produkt- und Toolauswahl für das Opinion Mining. In der Data Understanding Phase werden über Social Websites Produktmeinungen gefiltert und in einer Review-Datenbank abgelegt. Mit Hilfe der Review-Datenbank kann in der anschließenden Data Preparation das POS-Tagging durchgeführt und die Frequent Features identifiziert werden, um die Frequent Feature-Datenbank zu erstellen. Es erfolgt die Extraktion und Identifikation von Opinion Words aus der Frequent Feature-Datenbank. Die Opinion Words werden nachfolgend in der Opinion Word-Datenbank gespeichert. In der Modeling Phase wird für das Training und Testing auf die erstellten Datenbanken zurückgegriffen. Beim Training wird eine gewisse Anzahl an Opinion Words manuell bestimmt. Anschließend werden die restlichen Daten anhand der Trainingsdaten positiv und negativ konnotiert. Nach erfolgreichem Test kann mit dem Scan aller Reviews begonnen werden. In der abschließenden Evaluation und dem Deployment werden die Ergebnisse der vorherigen Phase aus betriebswirtschaftlicher Sicht analysiert und bewertet. Das Vorgehensmodell schließt mit der zielgruppengerechten Aufbereitung und Präsentation der Ergebnisse.

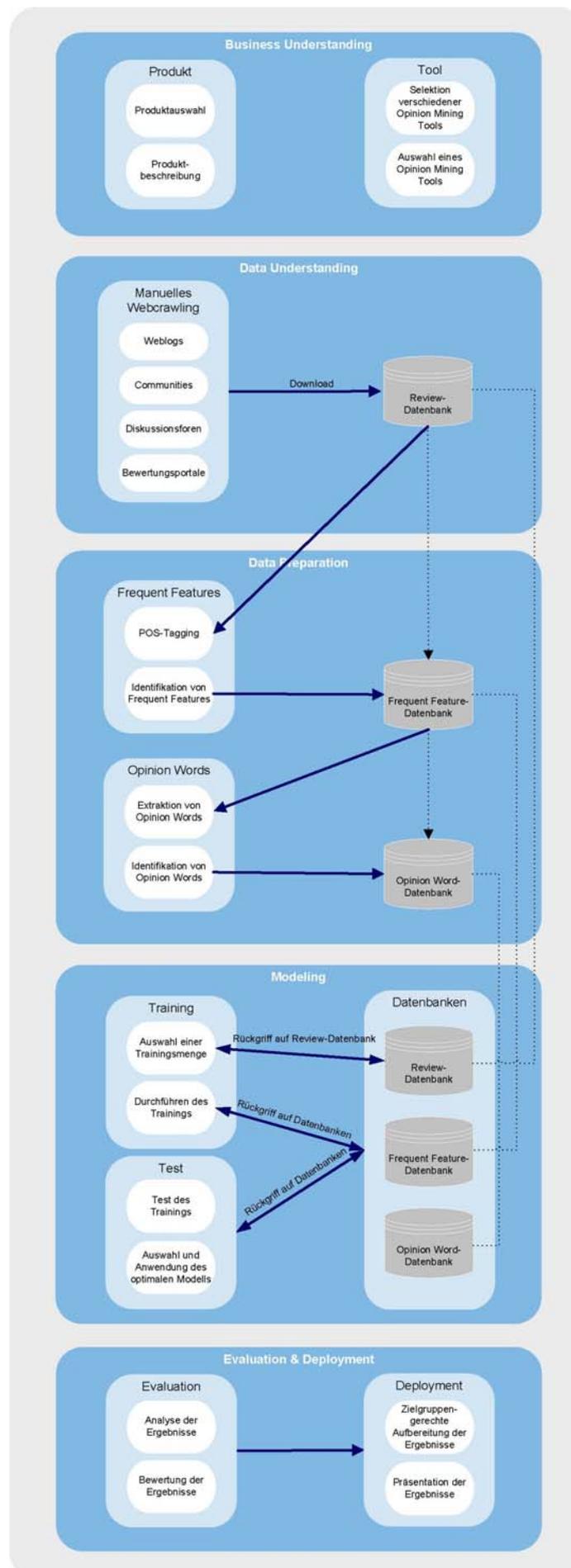


Abbildung 23: Vorgehensmodell für das Opinion Mining
 Quelle: in Anlehnung an Hu und Liu 2004 [1] und [2], S. 170

3.2 Praktische Anwendung und Auswertung

Dieses Kapitel erläutert die praktische Anwendung des zuvor dargestellten Vorgehensmodells. Wie auch schon im vorangegangenen Kapitel gliedert sich der Ablauf in die einzelnen Phasen des CRISP-DM Modells. Abschließend erfolgt in der letzten Phase (Evaluation & Deployment) die Auswertung und Aufbereitung der beim Opinion Mining erhaltenen Ergebnisse.

3.2.1 Business Understanding

Wie bereits im vorangehenden Kapitel erwähnt, fokussiert die einführende Phase auf das Verständnis der Anforderungen und Ziele des Projekts aus der Unternehmensperspektive, um anschließend eine Problemdefinition aus der Data Mining Perspektive zu ermöglichen.

Im Gegensatz zur Methode des Data Mining, bei der bereits in dieser ersten Phase ein Wissen darüber besteht, welche Informationen für die Analyse benötigt werden, ist es beim Opinion Mining zunächst notwendig alle Meinungen aufzunehmen und erst anschließend mit der Interpretation zu beginnen. Demnach erfolgt keine Selektion von schon bekannten Produktfeatures, stattdessen beginnt das Opinion Mining mit dem Herausfinden von noch unbekanntem Features, welche das zu untersuchende Produkt beschreiben.

3.2.1.1 Auswahl eines Produktes

Im Rahmen dieser Arbeit wird das Opinion Mining auf das Produkt iPhone angewandt. Es handelt sich dabei um ein Smartphone¹⁶ des Herstellers Apple, welches in seiner ersten Generation im Juni 2007 auf dem Mobilfunkmarkt in den USA eingeführt wurde. In Deutschland ist das iPhone seit November 2007 verfügbar, wobei T-Mobile, eine Tochtergesellschaft der Deutschen Telekom AG, das Exklusivrecht zur Vertragsbindung erwarb (vgl. Telekomprasse 2007). Somit ist das Smartphone in Kombination mit einem Vertrag bisher ausschließlich bei T-Mobile erhältlich.

Aufgrund einer großen Anzahl an Konkurrenzprodukten ist es sowohl für den Hersteller Apple als auch für den Händler T-Mobile unabdingbar, die Loyalität der Kunden zu „ihrem“ Produkt iPhone zu kennen. Diese kann beispielsweise anhand der Analyse von Kundenmeinungen durch das Opinion Mining beurteilt werden. Desweiteren bietet die Analyse die Möglichkeit, Kundeninformationen dar-

¹⁶ Ein Smartphone vereint den Leistungsumfang eines Mobilfunkgerätes und eines Personal Digital Assistant (PDA).

über zu erhalten, welche Verbesserungspotenziale das iPhone besitzt. Dazu zählen zum Beispiel wünschenswerte Features oder Korrekturen, deren Umsetzung dem Produkt auf dem Markt eine gesteigerte Konkurrenzfähigkeit gewährt.

Aus der Unternehmensperspektive heraus besteht das Ziel der Untersuchung aus der Analyse von Kundenmeinungen über das iPhone, um

- Kenntnisse über die Loyalität der Kunden zum Produkt zu erwerben,
- mögliche Verbesserungspotenziale aus Kundensicht zu erfahren und
- dem Produkt gegenüber seiner Konkurrenz eine Unique Selling Proposition zu verschaffen.



3.2.1.2 Auswahl eines Tools

Zur Erreichung der genannten Ziele ist die Auswahl eines passenden Opinion Mining Tools zur Analyse notwendig. Abschließend erfolgen eine kurze Darstellung und ein Vergleich gängiger Opinion Mining Tools auf dem Markt. Es besteht eine Vielzahl an textanalytischer Software¹⁷, sowohl kommerzieller als auch freier sowie Open-Source Produkte. Die Arbeit beschränkt sich auf die Vorstellung folgender Tools:

- Nstein Text Mining Engine (kommerziell)
- SAS Text Miner (kommerziell)
- SPSS PASW Text Analytics (kommerziell)
- RapidMiner Text Mining (open-source)

Nstein Text Mining Engine¹⁸ ist ein kommerziell angebotenes Produkt von Nstein Technologies. Das Unternehmen entwickelt vielsprachige Lösungen, welche die digitale Herausgabe von Zeitungen, Magazinen und inhaltsorientierte Organisationen ermöglichen. Die Text Mining Engine verfügt über das Sentiment Mining Modul Nsentiment¹⁹. Dieses identifiziert subjektive Strukturen innerhalb eines Doku-

¹⁷ Ein Überblick über die Text-Mining Software ist auf der Homepage von KDnuggets zu finden: <http://www.kdnuggets.com/software/text.html>.

¹⁸ Informationen zum Nstein Text Mining Engine finden sich unter: <http://www.nstein.com/products/tme/modules.html>.

¹⁹ Informationen zum Modul Nsentiment sind auf der folgenden Webseite zu finden: <http://www.nstein.com/products/tme/modules/sentiment-analysis.html>.

ments und filtert den Text anhand linguistischer Regeln. Jeder Satz wird auf Fakten und Meinungen geprüft. Meinungen werden positiv, negativ oder neutral konnotiert und erhalten einen Score, welcher den Level der Subjektivität (die Intensität der subjektiven Meinung) widerspiegelt.

Der SAS Text Miner²⁰ umfasst Tools zur Auffindung und Extraktion von Wissen und Meinungen aus Textdokumenten. SAS ist ein weltweit agierender Anbieter auf dem Business Intelligence Markt, dessen Softwarelösungen den Unternehmen helfen, konkrete Informationen für strategische Entscheidungen zu treffen. Der Text Miner stellt Werkzeuge und Methoden zur intelligenten und automatischen Verarbeitung textueller Informationen zur Verfügung, um Informationen und Meinungen aus Textdokumenten zu extrahieren. Die unstrukturierten Texte werden in ein Zwischenformat übertragen, um diese leichter zu klassifizieren, Beziehungen und Zusammenhänge untereinander herzustellen sowie Dokumente in Kategorien einzuordnen. Unstrukturierte Texte werden somit zu strukturierten Daten.

SPSS PASW Text Analytics (Text Mining for Clementine)²¹ ermöglicht die Extraktion von Schlüsselbegriffen, Stimmungen und Beziehungen aus textuellen, unstrukturierten Daten. Die Daten werden in ein strukturiertes Format konvertiert, welches die Kreation von Vorhersagemodellen unterstützt. SPSS ist eine Softwarefirma, die Statistik- und Analyse-Software entwickelt und vertreibt.

RapidMiner²² ist eine Software für maschinelles Lernen und Data Mining. Er beinhaltet mehr als 500 Operatoren für alle Aufgaben der Wissensentdeckung in Datenbanken, beispielsweise für Datenvorverarbeitung, maschinelles Lernen, Data-Mining, Text-Mining, Web-Mining und automatische Stimmungsanalyse aus Internet-Diskussionsforen. Das Unternehmen Rapid-I ist aus dem Lehrstuhl für künstliche Intelligenz der Technischen Universität Dortmund entstanden und bietet Softwarelösungen im Bereich Predictive Analytics, Data Mining und Text Mining an. Nachfolgende Abbildung zeigt die vier vorgestellten Softwareprodukte im Vergleich verschiedener Kriterien auf. Die Auswertung stützt sich auf eine bestehende Evaluation der Homepage www.wissensexploration.de. Für die Zwecke dieser Arbeit wurden die Ergebnisse der zuvor genannten Tools (SAS, SPSS und nStein) selektiert. Da Rapid-I nicht in der Evaluation berücksichtigt wurde, erfolgte die Be-

²⁰ Informationen zum SAS Text Miner finden sich unter: <http://www.sas.com/technologies/analytics/datamining/textminer/>.

²¹ Informationen zu SPSS PASW Text Analytics finden sich unter: <http://www.spss.com/software/modeling/text-analytics/>.

²² Weitere Informationen zum Rapid-I RapidMiner ist unter <http://rapid-i.com/content/view/73/148/> zu finden.

wertung der Kriterien für den RapidMiner anhand der Informationen auf der Unternehmenshomepage und der Erklärungen im RapidMiner Handbuch.

Folgende Kriterien²³ werden beim Vergleich der Opinion Mining Tools (s. Abbildung 24) herangezogen:

- Retrieval: die Informationsgewinnung durch Volltextsuche wird von den Tools entweder konzeptbasiert (+), basierend auf boolescher Algebra (o) oder gar nicht (-) unterstützt
- Crawler: die Tools besitzen einen eigenen Webcrawler (+), unterstützen die Anbindung von einzelnen Seiten (o) oder unterstützen kein Crawling von Internetquellen (-)
- Concept Linkage: die Entdeckung und Verknüpfung von gemeinsamen Konzepten in verschiedenen Dokumenten ist möglich (+) oder die Verknüpfung von Konzepten ist nicht durchführbar (-)
- Zusammenfassen: die Möglichkeit der Zusammenfassung mehrerer Texte (+), einzelner Texte (o) oder keine Zusammenfassung der Texte (-)
- Semantisches Wissen: die Erweiterung und Personalisierung von Taxonomien, Thesauern oder Ontologien (+), die Möglichkeit der Verwendung von vordefiniertem Wissen (o) oder keine Anwendungsmöglichkeit von Hintergrundwissen (-)
- Klassifikation: die automatische Zuordnung neuer Dokumente zu vordefinierten Kategorien (+) oder keine Unterstützung der Klassifizierung (-)
- Clustern: die Einordnung von Dokumenten in nicht vordefinierte, vom System automatisch gefundene, Kategorien bzw. Gruppen (+) oder das Fehlen dieses Clusterings (-)
- IE: die Unterstützung eigener Regeln oder annotierter Trainingsdaten und personalisierter Entitäten (+), die Extraktion vordefinierter Entitäten (o) oder das Fehlen einer IE Funktion (-)
- Topic Tracking: eine automatische Benachrichtigung und Verfolgung von Neuigkeiten und Veränderungen in den Daten (+) oder keine Möglichkeit der automatischen Themenverfolgung (-)
- Visualisierung: die Visualisierung textueller Daten durch mindestens zwei verschiedene Möglichkeiten (+), eine Visualisierungsmöglichkeit (o) oder keine (-)

²³ Eine zusätzliche Erklärung der Kriterien findet sich unter <http://wissensexploration.de/textmining-software-overview.php>.

- Freie Abfragen: die Unterstützung von Abfragen in natürlicher Sprache (+) oder das Fehlen einer derartige Funktion (-)
- Übersetzen: die Übersetzung von Dokumenten oder Abfragen (+), die Möglichkeit ausschließlich Dokumente zu übersetzen (o) oder keine Übersetzungsmöglichkeit (-)
- Demo: das Vorliegen einer installierbaren Demoversion zum Tool (+), die Online-Abrufbarkeit der Demo (o) oder das Fehlen einer Demoversion (-)

	Nstein	Rapid-I	SAS	SPSS
<i>Retrieval</i>	+	+	+	-
<i>Crawler</i>	-	+	-	-
<i>Concept Linkage</i>	+	+	+	+
<i>Zusammenfassen</i>	+	+	-	-
<i>Semant. Wissen</i>	+	?	+	+
<i>Kategorisierung</i>	+	+	+	+
<i>Clustern</i>	+	+	+	-
<i>Informationsextraktion</i>	o	+	+	-
<i>Topic Tracking</i>	+	-	-	-
<i>Visualisierung</i>	o	+	o	+
<i>Freie Abfragen</i>	+	-	-	-
<i>Übersetzen</i>	+	-	-	-
<i>Betriebssystem</i>	k.a.	W	W	WLS
<i>Demo</i>	-	+	o	-

+ vorhanden
 Unterstützung von: o teilweise vorhanden
 W Windows L Linux S Solaris

Abbildung 24: Text Mining Lösungen im Vergleich

Quelle: in Anlehnung an www.Wissensexploration.de – Text Mining Übersicht²⁴

Zum Einsatz in dieser Arbeit fiel die Entscheidung letztendlich aus nachfolgenden Gründen auf den RapidMiner:

- Die Software steht als Open-Source Produkt kostenfrei im Internet zum Download zur Verfügung.

²⁴ Die Evaluation der Text-Mining Tools durch wissensexploration.de sind unter folgendem Link einzusehen: <http://wissensexploration.de/Evaluation/Wissensexploration-TextMining-Marktuebersicht.pdf>. Die Informationen zum RapidMiner stammen von <http://rapid-i.com/>.

- Somit können bei Bedarf eigene (Java-)Knoten erstellt und eingefügt werden.
- Der RapidMiner ist Gewinner des Open-Source-Business Award 2008 (Pressebox 2009).
- Die kommerziellen Softwarelösungen von SAS, SPSS und nStein zeigen im Vergleich (s. Abbildung 24) keinen bedeutend größeren Funktionsumfang auf.



3.2.2 Data Understanding

Das Data Understanding beginnt mit einer Daten-/Meinungssammlung über das iPhone im World Wide Web. Dazu erfolgt ein manuelles Webcrawling der in Kapitel 2.1 näher erläuterten Kommunikationsformen des Web 2.0, um anschließend die Review Datenbank erstellen zu können. In diesem Kapitel erfolgt die Beantwortung der ersten Forschungsfrage aus Kapitel 1.2, indem verschiedene Möglichkeiten des Webcrawlings aufgezeigt werden. Abschließend wird eine Alternative zur Erfassung der Kundenmeinungen ausgewählt.

3.2.2.1 Erfassung der Kundenmeinungen per Webcrawling

Webcrawler²⁵ sind Programme, die Webseiten automatisch herunterladen. Dazu besuchen sie eine Vielzahl von Internetseiten, um dort Informationen zu sammeln, die anschließend entweder online oder offline (nach dem Speichern) analysiert werden. Am häufigsten werden Webcrawler durch Suchmaschinen wie beispielsweise Google oder Yahoo! genutzt, die Webseiten für ihre Indexierung sammeln. Aus diesem Grund stellen die Suchmaschinenbetreiber auch die hauptsächlichen Verbraucher von Internetbandbreiten dar (vgl. Menczer 2008, S.273). Das Vorgehen solcher vollautomatischen Webcrawler, wie sie von Suchmaschinenbetreibern genutzt werden, ist in Abbildung 25 dargestellt.

²⁵ Webcrawler werden auch Spider oder Robots genannt.

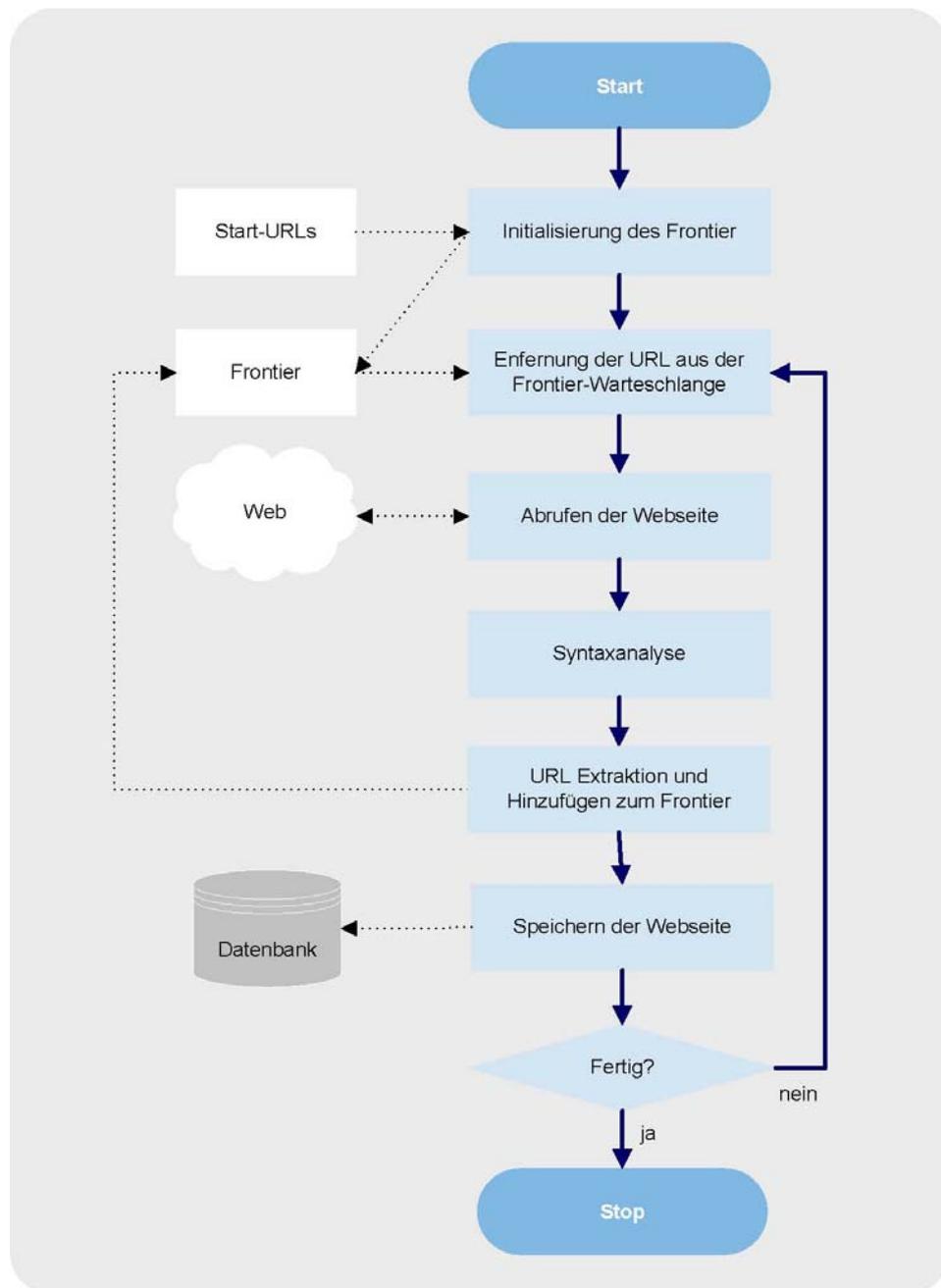


Abbildung 25: Sequenzieller Ablauf des Webcrawlings

Quelle: in Anlehnung an Menczer 2008, S.275

Am Anfang des Webcrawlings stehen Start-URLs, die der Crawler vom Nutzer oder von anderen Programmen erhält. Diese noch nicht besuchten URLs werden im Frontier, einer Warteschlange, abgelegt. Anschließend werden die URLs nach einer bestimmten Methode, beispielsweise FIFO²⁶ oder LIFO²⁷, aus der Frontier

²⁶ FIFO = first-in-first-out

Die URL, die als erstes in dem Frontier abgelegt wurde, wird auch wieder als erste aus dem Frontier entnommen.

²⁷ LIFO = last-in-first-out

Die zuletzt in dem Frontier abgelegte URL, wird als erste aus dem Frontier entnommen.

entnommen und im Internet abgerufen. Dazu agiert der Crawler wie ein Web Client, indem er eine http-Anfrage an den Hostserver der Webseite sendet und dessen Antwort liest. Er beginnt dann mit der Syntaxanalyse, indem er beispielsweise den jeweiligen HTML Code vom <head> bis zum </body> analysiert. Neu auftretende URLs (bspw. Links) auf der Webseite werden extrahiert und der Frontier-Warteschlange hinzugefügt. Es folgt das Speichern der Seite in einer Datenbank. Falls weitere URLs in dem Frontier enthalten sind, erfolgt eine Iteration des Ablaufes, welche mit der URL-Entnahme aus der Warteschlange beginnt. Der Zyklus endet, sobald das Frontier keine weiteren URLs mehr enthält (vgl. Menczer 2008, S.274 ff. und Pant et al. 2003, S.3 ff.). Spätestens nach dem Download einer Webseite beginnt der Crawler mit der Syntaxanalyse, in dem er beispielsweise den jeweiligen HTML Code vom <head> bis zum </body> analysiert.

Es gibt verschiedene Möglichkeiten des Webcrawlings. Einen Überblick bietet Abbildung 26.

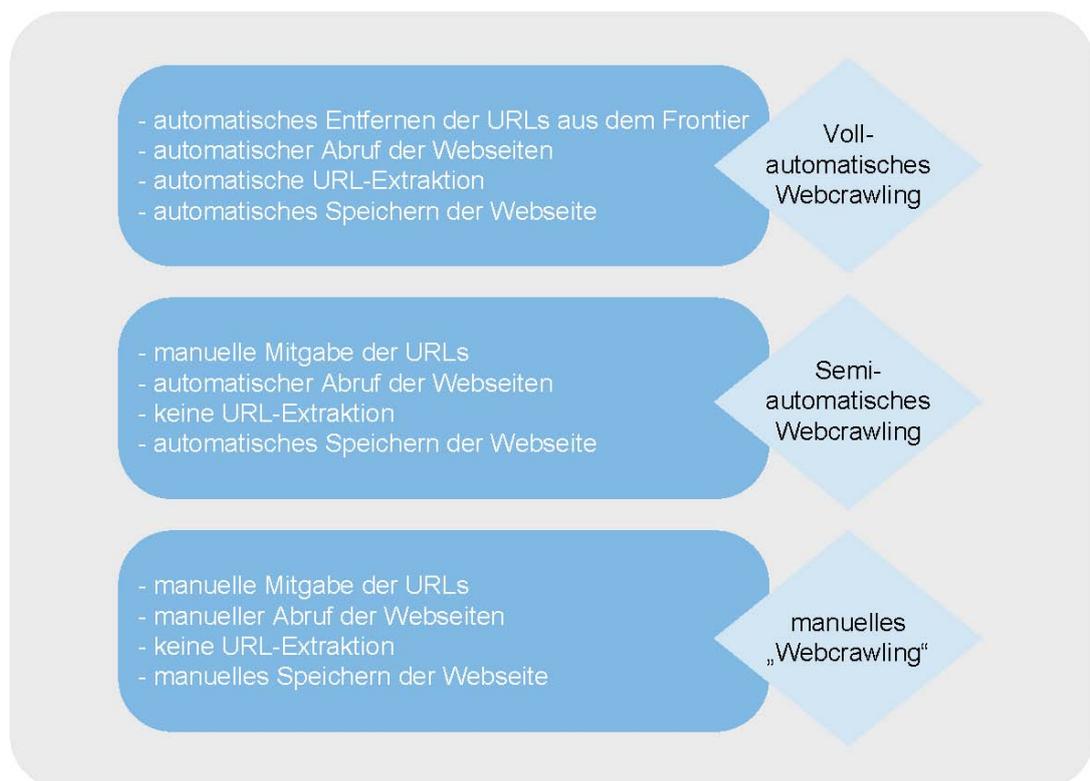


Abbildung 26: Methoden des Webcrawlings

Die Start-URLs können automatisch oder manuell vergeben werden. Beim vollautomatischen Webcrawling können über Suchmaschinen zum Beispiel die ersten einhundert Treffer automatisch zur Weiterverarbeitung in das Frontier gegeben werden. Der Crawler liest dann den HTML Code vom Head bis zum Body. Der in Abbildung 25 dargestellte Ablauf wird vom Crawler vollautomatisch übernommen. Zudem besteht die Möglichkeit einer manuellen Auswahl der Webseiten und eines

anschließenden automatischen Crawlens der Seiten. Dazu werden die URLs manuell in ein Crawlingtool eingefügt. Der Unterschied zur ersten Alternative besteht darin, dass ausschließlich die manuell mitgegebenen URLs gecrawlt werden. Es geschieht keine weitere URL-Extraktion innerhalb der Webseiten. Bei dieser Crawling-Variante entsteht jedoch das technische Problem, dass der Nutzer dem Crawler die Information mitgeben muss, wo er mit dem Lesen der Webseite beginnen und enden soll. Eine dritte Möglichkeit, Informationen aus dem Internet zu beziehen, besteht in der manuellen Auswahl der Webseiten und einer zusätzlich manuellen Bearbeitung dieser Seiten. Bei dieser Alternative handelt es sich also um ein manuelles „Webcrawling“ durch den Anwender.

Automatisches und semiautomatisches Webcrawling ist sowohl mit dem RapidMiner als auch mit externen Tools, wie beispielsweise HTTrack²⁸, möglich. Jedoch stellen dabei ethische Richtlinien ein Problem dar. Crawler können Webserver signifikant beanspruchen, indem sie deren Hauptbandbreite durch eine Vielzahl rapide hintereinander gestellter Seitenanfragen in Beschlag nehmen. Der Server ist dann so stark damit ausgelastet, dem Crawler zu antworten, dass sein Service gegenüber anderen Anfragen nachlässt. Im schlimmsten Fall führt dies zu einer Anfragenablehnung des Servers aufgrund der Serviceattacke durch einen einzigen Crawler (vgl. Menczer 2008, S.315).

Desweiteren erfordert die Crawler-Etikette die Befolgung des Robots Exclusion Protokolls²⁹, welches es den Webseitenbetreibern ermöglicht, den Crawlerzugriff auf Inhalte ihrer Seiten zu sperren. Das Protokoll stellt jedoch nur eine beratende Funktion dar. Aus diesem Grund unterscheidet man zwischen ethischen und nicht-ethischen Robots. Erstere lesen die sogenannte robots.txt Datei, welche vom Betreiber formulierte Zugriffsrichtlinien enthält, und befolgen diese während ihres Aufenthaltes auf der zugehörigen Webseite. Letztere umgehen jedoch das Robots Exclusion Standard Protokoll und greifen somit auch auf für das Crawling gesperrter Webseiteninhalte zu (Menczer 2008, S.315 f. und Sun et al. 2007, S.1123).

²⁸ HTTrack findet sich unter: <http://www.httrack.com/> (Abruf am 31.08.2009)

²⁹ Der Robots Exclusion Standard wurde vereinbart, um Robots (Crawler) von unerwünschten Bereichen einer Webseite fernzuhalten. Dem Standard entsprechend liest ein Robot zunächst die robots.txt Datei im Root-Verzeichnis des Webservers. Mit einem „Disallow“ wird dem Robot beispielsweise mitgeteilt, welche Bereiche der Seite beim Crawl nicht mit einbezogen werden dürfen (vgl. Hauser et al. 2007, S.878 f.). Weitere Informationen zum Robots Exclusion Standard sind unter www.robotstxt.org zu finden (Abruf am 02.09.2009).

Beispiel Amazon: Disallow: /gp/customer-reviews/write-a-review.html → Der Zugriff auf die Kundenrezensionen ist gesperrt (<http://www.amazon.de/robots.txt>, Abruf am 31.08.2009).

Um den zuvor genannten ethischen Bestimmungen gerecht zu werden und somit einen Ausschluss der IP-Adresse durch die Webseitenbetreiber zu vermeiden, wird in dieser Arbeit das manuelle „Webcrawling“ angewandt.

Zum Abschluss dieses Kapitels gilt die in Kapitel 1.2 erst genannte Forschungsfrage somit als beantwortet. Um die benötigten Kundenmeinungen aufzufinden, erfolgt zunächst eine Vorauswahl an Informationsquellen im Internet (z.B. Google, Twitter, Amazon). Nachfolgend werden die gewählten Weblogs, Communities, Diskussionsforen und Bewertungsportale auf Kundenrezensionen untersucht. Für die Untersuchung der Arbeit geeignete Social Websites werden abschließend in einer Review Datenbank gespeichert (vgl. Abbildung 27).



Abbildung 27: Beantwortung der Forschungsfrage 1

Wie erfolgt die Erfassung der Kundenmeinungen aus dem Internet?

3.2.2.2 Erstellen der Review-Datenbank

Dieses Kapitel beinhaltet die Erstellung der Review-Datenbank für das zuvor ausgewählte Produkt iPhone. Dazu findet das manuelle „Webcrawling“ Anwendung. Grund dafür ist die technische Problematik der Tag-Anbringung, um dem Crawler Informationen darüber zu geben, wo er mit dem Lesen der Webseite beginnen und enden soll. Desweiteren besteht das Bestreben die genannten ethischen Richtlinien zu befolgen, um nicht gegen die Crawler-Etikette zu verstoßen.

Zur Auswahl der Social Websites wird die Suchmaschine von Google (www.google.de), das Ermittlungssystem Technorati (www.technorati.de), das Social Bookmarking Tool Mister Wong (www.mister-wong.de) und das soziale Netzwerk Twitter (www.twitter.com) verwendet. Weitere Kundenmeinungen werden aus den „Bewertungsportalen“ Ciao (www.ciao.de) und Amazon (www.amazon.de) gefiltert.

Zur Suche nach Social Websites erfolgt im Google Suchfeld jeweils die Eingabe der Wortkombinationen „iPhone Blog“, „iPhone Community“ und „iPhone Forum“. Um ausschließlich Webseiten in deutscher Sprache zu erhalten, beschränkt sich die Suchanfragen auf „Seiten auf Deutsch“.

Die Blog-Suche gibt ca. 8.440.000 Treffer aus (vgl. Abbildung 28). Davon fällt die Auswahl für die Untersuchung auf 29 Blogs. Der letzte für die Untersuchung verwendbare Blog befindet sich unter Ergebnis 156. Der Suchabbruch erfolgt zirka 100 Treffer später, bei Ergebnis 260. Nicht ausgewählte Google-Treffer beinhalten hauptsächlich Blogs anderer Themengebiete, in denen der Begriff „iPhone“ nur als Tag hinterlegt ist und die somit für die Untersuchung nicht relevant sind.



Abbildung 28: Blog-Suche unter Google

Quelle: www.google.de (Abruf am 12.07.2009)

Eine weitere Hilfe bei der Blog-Suche stellt die Blogsuchmaschine Technorati (www.technorati.com) dar. Unter den Reitern „Channels“ und „Search Blog“ erfolgt die Eingabe des gesuchten Produkts „iPhone“. Die Suche ergibt 49.939 Ergebnisse (vgl. Abbildung 29). Eine Vielzahl der ausgegebenen Blogs ist jedoch in Fremdsprachen verfasst, da eine Beschränkung auf ausschließlich deutsche Ergebnisse bei Technorati nicht möglich ist. Zudem werden einige Ergebnisse mehrfach als Treffer ausgegeben. Schließlich fällt die Auswahl für die Untersuchung in dieser Arbeit auf nur drei hilfreiche Blogs.

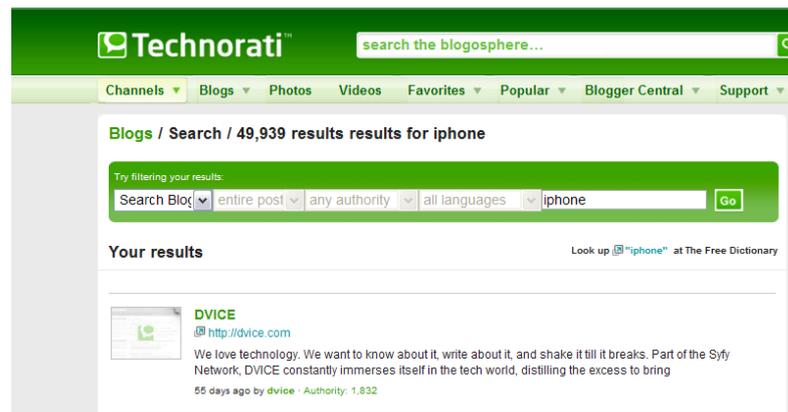


Abbildung 29: Blog-Suche unter Technorati

Quelle: www.technorati.com (Abruf am 12.07.2009)

Die Community-Suche unter Google erzeugt ca. 6.740.000 Treffer (vgl. Abbildung 30). Unter Eintrag 143 befindet sich die letzte von zehn für die Analyse nutzbaren Communities. Der Suchabbruch erfolgt zirka 100 Treffer später bei Ergebnis 250. Trotz der Suche nach „Seiten auf Deutsch“ sind etliche der ausgegebenen Communities in englischer Sprache verfasst. Desweiteren handelte es sich oftmals um Communities anderer Themengebiete, in denen das iPhone nur am Rande erwähnt oder getaggt ist.



Abbildung 30: Community-Suche unter Google

Quelle: www.google.de (Abruf am 18.08.2009)

Insgesamt zirka 8.120.000 Treffer ergibt die Forum-Suche unter Google (vgl. Abbildung 31). Davon werden 13 Foren für die Analyse ausgewählt. Das letzte Forum befindet sich bei Ergebnis 143. Der Suchabbruch erfolgt etwa 100 Treffer später bei Ergebnis 260.



Abbildung 31: Forum-Suche unter Google

Quelle: www.google.de (Abruf am 18.08.2009)

Im Zuge der Community- und Forum-Suche stellt sich eine fehlende Abgrenzung der beiden Social Website-Arten im Internet dar. Bei der Community-Suche werden neben den gesuchten Communities ebenfalls Foren ausgegeben und umgekehrt.

Weitere Suchen nach Social Websites finden im Social Bookmarking Tool Mister Wong (www.mister-wong.de) und im sozialen Netzwerk Twitter (www.twitter.com) statt. Dort ergeben sich jedoch keine neuen Webseiten, die nicht schon zuvor bei Google oder Technorati ausgewählt wurden.

Zusätzlich erfolgt eine Suche nach Kundenmeinungen unter Amazon (www.amazon.de) und Ciao (www.ciao.de). Bei beiden Portalen wird jeweils der Suchbegriff „iPhone“ eingegeben und anschließend das Ergebnis gefiltert. Unter Amazon ist eine Aufteilung der Ergebnisse in die einzelnen Produktgenerationen möglich. Für die verschiedenen Farbausprägungen und Speicherkapazitäten des Apple-Smartphones sind die Kundenrezensionen die gleichen. Für das iPhone 3G sind zum gegebenen Zeitpunkt 67 und für die Generation 3GS zwölf Kundenrezensionen hinterlegt (vgl. Abbildung 32).

Kundenrezensionen
Apple iPhone 3G 8GB



Kundenrezensionen
Apple iPhone 3Gs 32GB weiß



Die hilfreichste positive Rezension

185 von 200 Kunden fanden die folgende Rezension hilfreich:

★★★★★ **Top Handy mit Entwicklungspotential**
Ich bin seit Jahren Kunde hier bei Amazon und nutze bei meinen Einkäufen die Rezensionen sehr intensiv. Vieles habe ich schon über Produkte erfahren, was ich so nie erfahren hätte. Habe bisher aber nie eine Rezension geschrieben, weil ich gedacht habe, es sind schon genügend andere da. Jetzt sehe ich mich aber denn noch genötigt mal eine eigene zu schreiben, weil ich noch...

[Vollständige Rezension lesen >](#)
Vor 3 Monaten von kitzbiker veröffentlicht

> Weitere Rezensionen anzeigen: [5 Sterne](#), [4 Sterne](#)

Die hilfreichste positive Rezension

10 von 12 Kunden fanden die folgende Rezension hilfreich:

★★★★★ **Bestes Handy das ich je hatte**
Meiner Meinung nach ist das Iphone 3 GS das beste Telefon das ich je hatte. Die Einstellmöglichkeiten haben fast keine Grenzen! Fast alles erdenklich kann eingestellt und auf die jeweiligen Bedürfnisse angepasst werden.

+ Punkte:
< Sehr gelungene Optik
< fast unendlich viele Einstellmöglichkeiten
< viele Apps können auch...

[Vollständige Rezension lesen >](#)
Vor 1 Monat von G. Benjamin veröffentlicht

> Weitere Rezensionen anzeigen: [5 Sterne](#), [4 Sterne](#)

Abbildung 32: Bewertungssuche unter Amazon
Quelle: www.amazon.de (Abruf am 26.08.2009)

Unter Ciao ist ebenfalls eine Einteilung in die einzelnen Generationen des iPhone möglich. Zusätzlich gibt es allgemeine Kundenmeinungen, die sich nicht auf eine bestimmte Generation beziehen. Für die Generation 3G gibt es zum gegebenen Zeitpunkt 92 Erfahrungsberichte und 17 für das Modell 3GS. Zum iPhone allgemein finden sich zusätzlich 72 Berichte (vgl. Abbildung 33).

1-15 von 92 Erfahrungsberichten über Apple iPhone 3G (8GB - 16GB) Sortieren nach: [Nützlichkeit](#) | [Bewertung](#) | [Datum](#)

★★★★★ **Warum man das iPhone kaufen sollte und warum nicht**
Bewertung für Apple iPhone 3G (8GB - 16GB) von explore.ur.world

Pro: Design, Bedienung, Spitzen-Audioplayer, GPS, Internet überall, Zusatzprogramme
Kontra: schlechte Kamera, keine Videofunktion, Fingertapsen, Akku nicht austauschbar

...aller Munde und wird von Apple als revolutionäres Handy verkauft. In der Tat hebt es sich von anderen Handys ab, aber was steckt wirklich dahinter? Ich habe mir zur Einführung des neuen iPhones auch ein solches Gerät gekauft und möchte nachfolgend meine Erfahrungen mit dem iPhone zu schildern, um anschließend allen zu helfen, die sich noch nicht sicher sind, ob sie sich ein iPhone 3G kaufen sollen. UPDATE-Hinweis: Der Bericht wurde am 23.06.2009 ... [Bericht lesen](#)

Ciao Mitglieder bewerteten diesen Erfahrungsbericht insgesamt als **besonders hilfreich**

Ausstattung:	★★★★★	besonders
Klang:	★★★★★	hilfreich
Empfangsqualität:	★★★★★	hilfreich
Verarbeitung:	★★★★★	
Akkulaufzeit:	★★★★★	04.08.2008
Design:	★★★★★	(23.06.2009)

1-15 von 17 Erfahrungsberichten über Apple iPhone 3GS (16GB - 32GB) Sortieren nach: [Nützlichkeit](#) | [Bewertung](#) | [Datum](#)

★★★★★ **iPhone 3GS - der beste Kompromiss den es gibt**
Bewertung für Apple iPhone 3GS (16GB - 32GB) von langre

Pro: Bedienung, Geschwindigkeit und vieles mehr....
Kontra: Batterie nicht auswechselbar

...der Rückseite befindet sich das Apple Logo und die Beschriftung – insbesondere die Speichergroße. Auf der Vorderseite ist natürlich das Display. Oberhalb des Displays befindet sich der Lautsprecher für das Telefonieren, unter dem Display die Home-Taste. Standardmäßig dabei sind ein Netzladadapter, das USB-Kabel, ein Headset mit Gesprächsannahme Tasten - auch zum lauter und leiser stellen - und Kurzinfo sowie Sim-Karten-Slot Öffner. Anschaffung ... [Bericht lesen](#)

Ciao Mitglieder bewerteten diesen Erfahrungsbericht insgesamt als **sehr hilfreich**

Ausstattung:	★★★★★	sehr
Klang:	★★★★★	hilfreich
Empfangsqualität:	★★★★★	hilfreich
Verarbeitung:	★★★★★	
Akkulaufzeit:	★★★★★	20.07.2009
Design:	★★★★★	

1-15 von 72 Erfahrungsberichten über Apple iPhone Sortieren nach: [Nützlichkeit](#) | [Bewertung](#) | [Datum](#)

★★★★★ **Der einzige iPhone-Bericht, den ihr lesen müsst!**
Bewertung für Apple iPhone von Magdal.ena

Pro: Design, Verarbeitung, Funktionsvielfalt, durchdachtes Multitouch-UI / Touchscreen, einfache Handhabung, Akku, WiFi, macht glücklich!
Kontra: Preis, Vertragskosten, kein UMTS, viele Funktionen fehlen, keine reale Tastatur, keine MMS

...kompatibel, was einfach unakzeptabel ist. Apple bietet einen Adapter an.... :(((((Das Gewicht beträgt 135g und liegt damit deutlich über federleichteren Mobiltelefonen wie dem Razr2, aber ungefähr im Rahmen von vergleichbareren Geräten wie dem Blackberry. Für meinen Geschmack ist es ein paar Gramm zu schwer, wirklich störend oder belastend wirkt sich das in der Praxis aber nicht aus, auch nicht bei längeren Gesprächen. Der Touchscreen Wow! ... [Bericht lesen](#)

Ciao Mitglieder bewerteten diesen Erfahrungsbericht insgesamt als **besonders hilfreich**

Ausstattung:	★★★★★	besonders
Klang:	★★★★★	hilfreich
Empfangsqualität:	★★★★★	hilfreich
Verarbeitung:	★★★★★	
Akkulaufzeit:	★★★★★	20.11.2007
Design:	★★★★★	

Abbildung 33: Bewertungssuche unter Ciao
Quelle: www.ciao.de (Abruf am 26.08.2009)

Die nachfolgende Abbildung 34 stellt die Suchergebnisse noch einmal im Überblick dar. Es wurden insgesamt 33 Blogs, zehn Communities, 13 Foren und 260 Kundenrezensionen für die Untersuchung in dieser Arbeit ausgewählt.

	Blogs	Communities	Foren	Kunden- rezensionen
<i>Google</i>	29	10	13	-
<i>Technorati</i>	4	-	-	-
<i>MisterWong</i>	-	-	-	-
<i>Twitter</i>	-	-	-	-
<i>Amazon</i>	-	-	-	79
<i>Ciao</i>	-	-	-	181
Gesamt:	33	10	13	260

Abbildung 34: Ergebnisse der Social Website Suche³⁰

Nachdem die manuelle Suche der URLs abgeschlossen ist, erfolgt das manuelle Crawlen der selektierten Internetseiten. Dazu werden die einzelnen Texte aus den Quellen zunächst in eine Textdatei eingefügt, die anschließend in einem Ordner Review-Datenbank abgespeichert wird.

Die Blogs, Communities und Foren werden im Folgenden vernachlässigt, da sie den kleinsten Anteil der gesammelten Kundenmeinungen ausmachen. Ein weiterer Grund für diese Entscheidung ist technischer Natur. Die einzelnen Einträge besitzen meist zugehörige Kommentare, welche oftmals nur in Zusammenhang mit der vorherigen Diskussion verständlich sind. Diese Zusammenhänge sind jedoch nur begrenzt oder nicht maschinell erfassbar. Demnach stützt sich die Arbeit von nun an ausschließlich auf die Kundenrezensionen von Amazon und Ciao.

3.2.3 Data Preparation

Ziel der Data Preparation ist die Konstruktion eines finalen Datensets. Da in dieser Arbeit das eigenschaftsbasierte Opinion Mining angewandt wird, welches in Kapitel 2.2.3 näher erläutert wurde, ist zu diesem Zweck zunächst einmal das Auffinden der häufigsten Produkteigenschaften (Frequent Features) des iPhones notwendig (vgl. Liu 2008, S.418). Um die expliziten Features in Form von Nomen und Satzphrasen aufzufinden, wird eine nicht überwachte (unsupervised) Methode

³⁰ Eine Auflistung der Links zu allen ausgewählten Social Websites befindet sich im Anhang (A).

angewandt (vgl. Liu 2008, S.429). Dazu erfolgen das POS-Tagging, welches die Worte innerhalb der Sätze ihren Wortklassen zuordnet und das Association Rule Mining, um die Frequent Features von den Infrequent Features abzugrenzen. Anschließend folgt die Extraktion der Opinion Words aus der Frequent Feature Datenbank anhand der Adjektive, die sich in unmittelbarer Nähe zu den Frequent Features befinden. Die semantische Zuordnung der Opinion Words wird durch die Verwendung von schon fixen Synonymen und Antonymen in Lexika vollzogen. Das Ergebnis dieser Phase ist die Opinion Word Datenbank. An dieser Stelle sei schon einmal angemerkt, dass es sich bei den Theorien um Ideallösungen handelt, welche in der Praxis nicht unbedingt umsetzbar sind. So kann das POS-Tagging beispielsweise nicht vom RapidMiner erfüllt werden, wie in Kapitel 3.2.3.3 deutlich wird.

3.2.3.1 Erstellen der Frequent Feature-Datenbank

Die Frequent Feature-Datenbank baut auf der in Kapitel 3.2.2 erstellten Review-Datenbank auf, welche eine ausgewählte Sammlung an Kundenmeinungen zum iPhone beinhaltet. Zunächst besteht die Aufgabe darin, die sogenannten Frequent Features in der Datenbank aufzufinden. Dabei handelt es sich um genau die Produkteigenschaften, welche von einer Vielzahl an Rezensenten beurteilt werden. Diese stellen meist Nomina oder Nominalphrasen innerhalb von Meinungssätzen dar. Entscheidend ist hier das POS Tagging. Die Wortart ist eine linguistische Kategorie, welche sich über die syntaktische und morphologische Wortform definiert (vgl. Liu 2008, S.413). Gängige deutsche Wortarten sind zum Beispiel Nomina, Verben, Adjektive, Pronomen, Adverbien, Präpositionen und Konjunktionen. Zusätzlich bestehen Gattungen, welche aus den verschiedenen Formen der linguistischen Kategorien entspringen. Beispielsweise werden Verbformen im Deutschen konjugiert, so dass sich der Wortstamm verändert oder Affixe anhängen. Unter Zuhilfenahme des POS-Taggings kann jedes einzelne Wort innerhalb eines Satzes adäquat seiner Wortart zugeteilt werden. Die getaggten Sätze werden anschließend in einem Transaction-Set gespeichert.

Ein POS-Tagger der deutschen Sprache ist der Brill-Tagger³¹ (nach Brill 1993), welcher durch die Computervisualistik Gruppe an der Universität Zürich für das Deutsche trainiert wurde. Der Brill-Tagger funktioniert durch das automatische

³¹ URL für den Brill Tagger in Deutsch: <http://www.ifi.uzh.ch/arvo/cl/tagger/> (Abruf am 22.09.2009).

URL für den NLProcessor (POS-Tagger) in Englisch: <http://www.infogistics.com/textanalysis.html> (Abruf am 23.09.2009)

Erkennen und Beheben seiner Schwächen. Zunächst ordnet er jedem Wort eine Wortform zu, von welcher er ausgeht, dass diese am ehesten dem Wort entspricht. Zu dieser Bewertung beinhaltet der Tagger einen umfangreichen kontextunabhängigen Korpus, der eine Vielzahl an Tags enthält. Zur Leistungssteigerung besitzt der anfängliche Tagger zwei Verfahren. Eines der Verfahren stützt sich auf die Begebenheit, dass großgeschriebene Worte, die der Trainingskorpus nicht beinhaltet, in ihrer Zugehörigkeit zur Kategorie der Nomina tendieren und somit dieser Kategorie zugeordnet werden. Das zweite Verfahren bezieht sich auf die Wortendungen, die oftmals Aufschluss über die Wortart geben, welcher das Wort anschließend zugeteilt wird (vgl. Brill 1993, S.112 f.). Anhand dieses Algorithmus lernt der Brill-Tagger und ordnet die einzelnen Worte den Wortklassen zu, denen sie mit der größten Wahrscheinlichkeit angehören. Ein Beispiel des Brill-Taggers stellt Abbildung 35 dar.

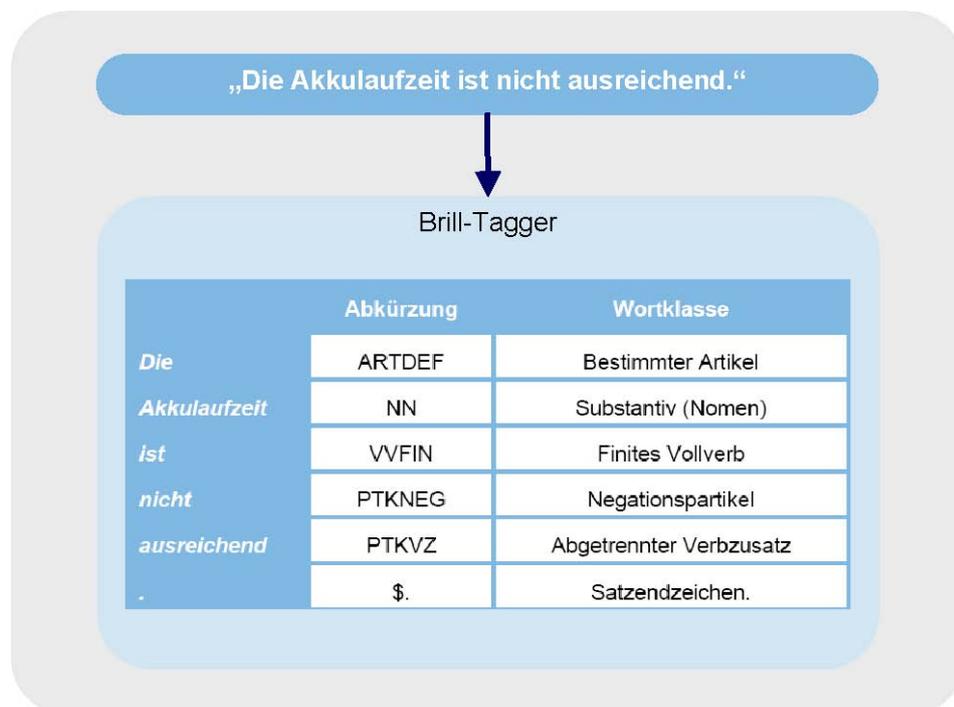


Abbildung 35: Vorgehensweise des Brill-Taggers

Der Tagger bearbeitet den eingegebenen Satz „Die Akkulaufzeit ist nicht ausreichend“, indem er ihn in seine einzelnen Tags unterteilt und jedes Wort einer bestimmten Wortklasse zuordnet. So erkennt er beispielsweise die „Akkulaufzeit“ als Nomen oder das „nicht“ als Negation.

Ein weiterer Schritt zur Identifikation der Frequent Features stellt das Association Mining dar. Der Grund dafür besteht darin, dass Kundenrezensionen häufig Mitteilungen enthalten, die nicht direkt mit den relevanten Produkteigenschaften in Beziehung stehen. Das Association Mining hilft nach dem POS-Tagging, die Fre-

quent Features von den Infrequent Features, also den Nicht-Produkteigenschaften, zu unterscheiden.

In der Literatur findet sich eine Vielzahl von Ansätzen und Vorgehensweisen für das Association Mining (vgl. u.a. Agrawal et al. 1993; Hipp et al. 2000; Liu 2008, S.13–54; Rauch und Šimůnek 2002). In dieser Arbeit soll kurz der Association Miner CBA (Classification Based on Associations) von Liu et al. 1998 vorgestellt werden. Er besteht aus einem Regelgenerator (CBA-RG) und einem Klassifikationsentwickler (CBA-CB). Aufgabe des Regelgenerators ist die Erstellung von häufig vorkommenden Regelobjekten (ruleitems). Dazu wird das zuvor erstellte Transaction-Set in mehreren Durchläufen auf diese Objekte durchsucht. Die dazu angewandte Vorgehensweise stützt sich auf den von Agrawal und Srikant (1994, S.5 ff.) entwickelten Apriori Algorithmus. Aus der Menge der ausgewählten Regelobjekte werden mit Hilfe von bestimmten Pruning-Methoden (vgl. Hu und Liu 2004 [2], S.171) abschließend die Regeln (Class Association Rules: CARs) erstellt. Der Klassifikationsentwickler nutzt anschließend die CARs, um den besten Klassifizierer innerhalb der Menge an Regeln zu finden. Eine weitere Möglichkeit zur Klassifikation stellt etwa der C4.5 Algorithmus dar. Jedoch bewiesen Liu et al. (1998), dass CBA im Vergleich zu C4.5 akkuratere Klassifizierer produziert.

Der Association Rule Miner CBA findet alle Wörter oder Phrasen aus den zuvor getaggtten Sätzen, die zusammengehören und möglicherweise eine Produkteigenschaft darstellen (Hu und Liu 2004 [1]). Die ausgewählten Wörter und Phrasen zählen somit zu den Frequent Features (Hu und Liu 2004 [2], S.171).

3.2.3.2 Erstellen der Opinion Word-Datenbank

Die Opinion Words werden aus der gesamten Review-Datenbank extrahiert. Zu den Opinion Words zählen die Wörter, derer sich die Rezensenten bedienen, um ihre positive oder negative Meinung auszudrücken. Meist befinden sich diese Wörter innerhalb eines Satzes nahe an der zu bewertenden Produkteigenschaft. In verschiedenen Studien (u.a. Bruce und Wiebe 2000 und Wiebe et al. 1999) konnte nachgewiesen werden, dass eine statistisch positive Korrelation zwischen der Subjektivität und dem Beisein von Adjektiven besteht. Zur Vorhersage, ob es sich bei einem Satz um eine Meinung handelt, ist demnach die Präsenz von Adjektiven nützlich, welche die Opinion Words darstellen. Darauf basierend werden aus der Frequent Feature-Datenbank die Sätze auf Adjektive untersucht, welche die zuvor erschlossenen Frequent Features beinhalten. Die Adjektive, die sich in unmittelbarer Nähe zu den Frequent Features befinden, sind die gesuchten Opinion Words.

Für jedes Opinion Word wird dann seine semantische Orientierung bestimmt, worauf sich die Vorhersage der semantischen Orientierung der Meinungsätze und Rezensionen aufbaut. Worte mit einem wünschenswerten Status (z. B. schön, großartig) besitzen eine positive Orientierung, wohingegen Worte mit nicht erstrebenswertem Status (z. B. hässlich, schlecht) eine negative Orientierung aufweisen. Neben positiven und negativen Adjektiven gibt es auch neutrale Adjektive, die keiner semantischen Orientierung angehören (vgl. Hu und Liu 2004 [2], S.172). Die Bestimmung der semantischen Orientierung erfolgt anhand der Verwendung von schon fixen Synonymen und Antonymen in Lexika. Dazu kann beispielsweise das Online Lexikon³² Woxikon (www.synonyme.woxikon.de, Abruf am 23.09.2009) genutzt werden. Dort besteht die Möglichkeit nach einem bestimmten Adjektiv zu suchen. Anschließend werden die Synonyme und Antonyme dieses Adjektivs ausgegeben (vgl. Abbildung 36). Anhand von Stammadjektiven, deren semantische Bestimmung bekannt ist, kann die semantische Zuordnung unbekannter Adjektive vollzogen werden (vgl. Hu und Liu 2004 [2], S.172).

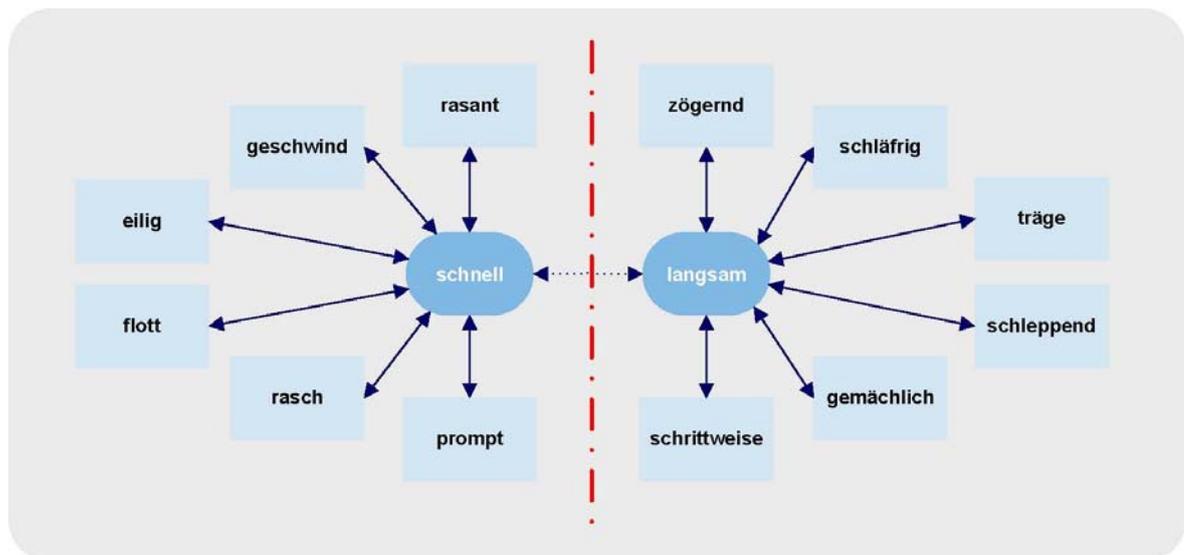


Abbildung 36: Adjektivstruktur: Synonyme und Antonyme

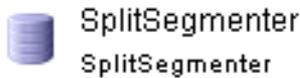
Quelle: in Anlehnung an Hu und Liu 2004 [2], S.172 und
<http://synonyme.woxikon.de/synonyme/schnell.php> und
<http://synonyme.woxikon.de/synonyme/langsam.php>, Abruf am 23.09.2009

3.2.3.3 Data Preparation im RapidMiner

Um mit dem RapidMiner arbeiten zu können, ist es zu Beginn notwendig, die Meinungen aus der Review Datenbank in das Tool einzulesen und den Text in seine

³² Ein Online Lexikon für Synonyme und Antonyme im Englischen ist bspw. WordNet (<http://wordnet.princeton.edu/>), Abruf am 23.09.2009.

einzelnen Sätze zu untergliedern. Dazu werden die Operatoren SplitSegmenter und TextInput angewandt.



- Unter „texts“ im *SplitSegmenter* erfolgt die Pfadeingabe zur Review-Datenbank. Der Operator entnimmt dieser den Meinungstext. Anschließend wird dieser Text in seine einzelnen Sätze aufgeteilt.³³ Dazu werden unter „split_expression“ Ausdrücke festgelegt, nach welchen der Split vorgenommen werden soll (vgl. Abbildung 37). Die Sätze werden abschließend unter dem im „Output“ angegebenen Pfad jeweils in einer einzelnen Textdatei abgelegt.

Als „split_expression“ werden in dieser Arbeit ausschließlich Satzzeichen gewählt. Weitere Satzzeichen werden nicht als Trennzeichen verwendet. Einen Grund dafür stellen beispielsweise Aufzählungen dar. Hier würde der Satz bei jedem Komma, also bei jeder einzelnen Aufzählung, gesplittet und es bestünde die Gefahr, dass zusammengehörende Frequent Features und Opinion Words getrennt würden.

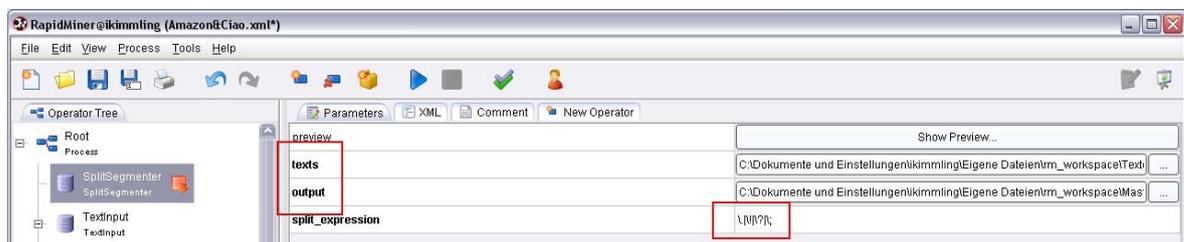
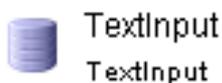


Abbildung 37: SplitSegmenter im RapidMiner³⁴

(Quelle: RapidMiner Version 4.3)

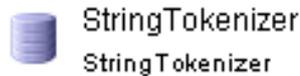


- Mit dem *TextInput* werden die zuvor generierten Sätze in den RapidMiner eingelesen. Dazu gibt man unter „texts – Edit List“ (vgl. Abbildung 38) den Pfad zur erstellten Review Datenbank mit.

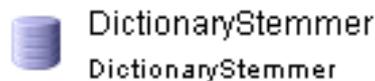
³³ Die Meinungssätze zum iPhone allgemein sind auf der beigefügten CD unter „RapidMiner Modelle → Rezensionen insgesamt“ einzusehen.

³⁴ Das RapidMiner Modell befindet sich auf der beigefügten CD unter „RapidMiner Modelle → Amazon&Ciao.xml“

Zusätzlich ist eine Vorverarbeitung der Daten erforderlich, um unterschiedliche Wortvarianten zu berücksichtigen, Rechtschreibfehler zu reduzieren und weiteren Datenverzerrungen (Noisy Data) vorzubeugen. Der RapidMiner stellt dazu eine Vielzahl an Operatoren zur Verfügung. Folgende werden im Modell dieser Arbeit angewandt (vgl. Abbildung 38):



- Der *StringTokenizer* zerlegt den String in mehrere Tokens. Die einzelnen Wörter sind so jeweils einer Variablen zugeordnet.



- Der *DictionaryStemmer* ermöglicht das Ersetzen von Wörtern anhand zuvor manuell festgelegter Regeln. Diese befinden sich in einer Textdatei, die mit Hilfe des Operators in das Modell eingelesen wird.³⁵



- Der *GermanStopwordFilter* entnimmt dem String sogenannte deutsche Stopwords³⁶. Dazu zählen Artikel, Präpositionen, Konjunktionen und Pronomen, welche häufig in unserer Sprache vorkommen, für die Untersuchung jedoch nicht signifikant sind.

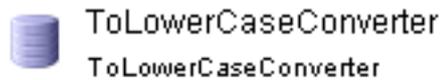


- Die *StopwordFilterFile* enthält zusätzlich zum *GermanStopwordFilter* deutsche Stopwords. Diese werden manuell in einer Textdatei gespeichert und anhand des Operators eingelesen.³⁷

³⁵ Das im Modell angewandte Dictionary ist auf der beigelegten CD unter „RapidMiner Modelle → Dictionary“ einzusehen.

³⁶ Beispiele für deutsche Stopwords: als, an, bei, bis, da, das, dein, doch, ein, er, für, hat, hier, ich, ihr, in, jeder, kann, mein, muss, nach, oder, seine, sind, soll, und, von, was, wir, zu (vgl. <http://www.ranks.nl/stopwords/german.html>, Abruf am 15.09.2009).
Die im Modell angewandte StopwordFilterFile kann auf der beigelegten CD unter „RapidMiner Modelle → StopwordFilterFile“ eingesehen werden.

³⁷ Die deutschen Stopwords in dieser Datei stammen aus den folgenden Quellen:
<http://www.promomasters.at/blog/2009/01/13/stop-words/>,
http://www.phpbar.de/w/Stopwortliste_deutsch,
<http://www.ranks.nl/stopwords/german.html>, Abruf am 28.09.2009.



- Der *ToLowerCaseConverter* sorgt für ein einheitliches Bild, indem er alle Buchstaben in Kleinbuchstaben konvertiert.

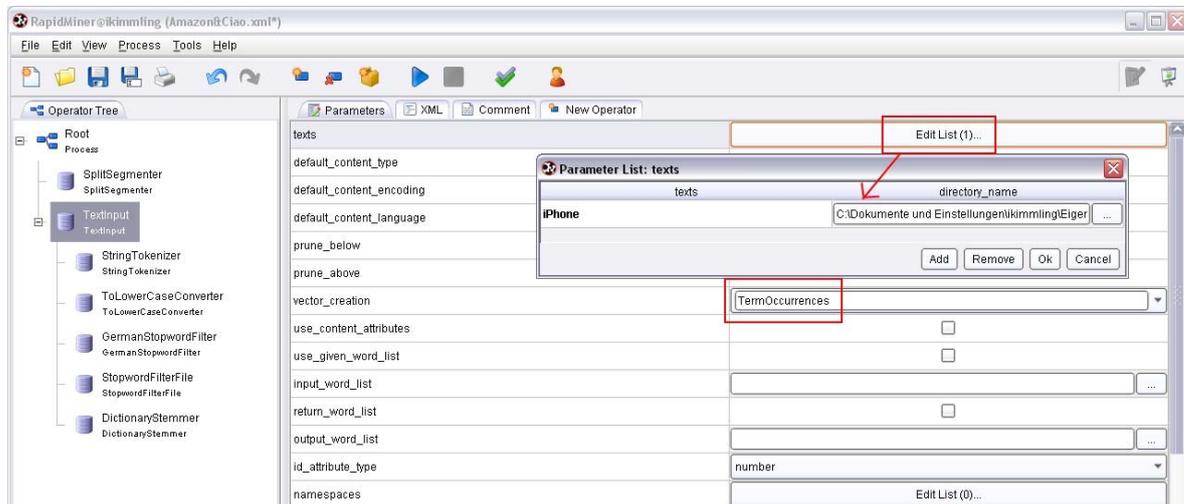


Abbildung 38: TextInput und Preprocessing Steps im RapidMiner

Quelle: RapidMiner Version 4.3 (Operatoren: SplitSegmenter, TextInput, StringTokenizer, ToLowerCaseConverter, GermanStopwordFilter, StopwordFilterFile, DictionaryStemmer)

Zusätzlich zu diesen Operatoren könnte der *GermanStemmer* angewandt werden. Dieser reduziert die im String enthaltenen Worte auf ihren Wortursprung. In vielen Sprachen besitzen die Worte, je nach Kontext, in dem sie verwendet werden, verschiedenartige syntaktische Formen (bspw. Pluralformen bei Nomen oder infinite und finite Verben), die jedoch auf den gleichen Wortursprung zurückgehen. Nach Durchlauf des *GermanStemmer* bleibt ausschließlich der Wortstamm, ohne Präfixe und Suffixe, zurück. Die Gefahr des Stemming im RapidMiner besteht jedoch darin, dass der Stemmer auch Worte mit einbezieht, bei deren Reduktion der Affixe ein Wortrest zurückbleibt, den es in der deutschen Sprache nicht gibt (z.B. Katze – Katz, Leben – Leb). Um solche Fehler zu vermeiden, wird das Stemming vernachlässigt.

Da der RapidMiner nicht das notwendige POS-Tagging zur Erstellung des Transaction-Sets unterstützt, müssen die Frequent Features auf eine andere Weise herausgefiltert werden. Dazu wird die „Vector_Creation“ im TextInput Operator auf „TermOccurences“ eingestellt, um im Ergebnis eine Auflistung der Auftrittshäufigkeit (unter „Sum“) der einzelnen Worte innerhalb der Texte zu erhalten (vgl. Abbildung 39).

ExampleSet (12227 examples, 2 special attributes, 12780 regular attributes)

Type	Name	Value Type	Statistics	Range	Sum
regular	mensur	real	avg = 0.000 +/- 0.022	[0.000 ; 1.000]	0
regular	produkte	real	avg = 0.002 +/- 0.043	[0.000 ; 2.000]	21
regular	erfahren	real	avg = 0.001 +/- 0.029	[0.000 ; 2.000]	8
regular	nie	real	avg = 0.007 +/- 0.088	[0.000 ; 2.000]	89
regular	verschiedene	real	avg = 0.002 +/- 0.047	[0.000 ; 1.000]	27
regular	kategorien	real	avg = 0.000 +/- 0.022	[0.000 ; 1.000]	6
regular	nutzer	real	avg = 0.003 +/- 0.055	[0.000 ; 2.000]	35
regular	unterscheiden	real	avg = 0.000 +/- 0.013	[0.000 ; 1.000]	2
regular	iphone	real	avg = 0.179 +/- 0.427	[0.000 ; 10.000]	2183
regular	systemssoftware	real	avg = 0.000 +/- 0.020	[0.000 ; 1.000]	5
regular	vibrationsalarm	real	avg = 0.000 +/- 0.018	[0.000 ; 1.000]	4
regular	einstellen	real	avg = 0.003 +/- 0.059	[0.000 ; 1.000]	42
regular	ausgeschaltetem	real	avg = 0.000 +/- 0.009	[0.000 ; 1.000]	1
regular	ton	real	avg = 0.001 +/- 0.035	[0.000 ; 2.000]	13
regular	losgeht	real	avg = 0.000 +/- 0.009	[0.000 ; 1.000]	1
regular	wahre	real	avg = 0.001 +/- 0.026	[0.000 ; 1.000]	8
regular	apple	real	avg = 0.046 +/- 0.236	[0.000 ; 10.000]	566

Abbildung 39: Ergebnis der Data Preparation

In Excel erfolgt eine Sortierung dieser Summenwerte, so dass die am häufigsten genannten Worte in der Liste oben stehen. Manuell werden dann die am häufigsten genannten Produktmerkmale aus der Liste herausgefiltert. Abbildung 40 beinhaltet eine Zusammenfassung der 15 am häufigsten genannten Merkmale. Da die Produktmerkmale von den Kunden zum Großteil durch mehrere verschiedene Wörter ausgedrückt werden, erfolgt ein Zusammenschluss dieser Worte unter einem jeweils zutreffenden Oberbegriff³⁸. Weitere Produktmerkmale werden nicht betrachtet, da sie nur noch ≤ 150 Mal genannt werden und somit in dieser Arbeit nicht mehr zu den Frequent Features zählen.

³⁸ Das Display wird bspw. auch als Bildschirm, Touchscreen oder Handydisplay bezeichnet. Ein Überblick der Zusammenfassung einzelner Wörter zu einem Frequent Feature ist im Anhang (B) einzusehen.

Lfd.	Frequent Features	Häufigkeit im Text
1	<i>Display</i>	626
2	<i>Musik</i>	584
3	<i>Funktionalität</i>	525
4	<i>Fotofunktion</i>	419
5	<i>Emailfunktion</i>	416
6	<i>Applicationen</i>	389
7	<i>iTunes</i>	384
8	<i>Internetfunktion</i>	374
9	<i>Videofunktion</i>	332
10	<i>SMS</i>	303
11	<i>Kamera</i>	293
12	<i>Akku</i>	238
13	<i>Bedienbarkeit</i>	207
14	<i>Vertrag</i>	197
15	<i>Betriebssystem</i>	177

Abbildung 40: Auflistung der Frequent Features des iPhone

Der nächste Schritt umfasst das Auffinden der Opinion Words. Dazu dient ebenfalls die im Ergebnis der Data Preparation summierte Auftrittshäufigkeit der einzelnen Worte. Statt sie wie zuvor auf häufig auftretende Nomen zu untersuchen, werden nun alle vorliegenden Meinungsworte (Adjektive) gefiltert und ihrer Zugehörigkeit nach in positiv, negativ und neutral sortiert. Abbildung 41 beinhaltet einen Ausschnitt der Opinion Words als Tag-Cloud.



Abbildung 41: Überblick der positiven, neutralen und negativen Opinion Words³⁹

³⁹ Grün = positives Opinion Word; gelb = neutrales Opinion Word; rot = negatives Opinion Word.
Eine komplette Auflistung der Opinion Words befindet sich im Anhang (C).

Somit gilt auch die zweite Forschungsfrage aus Kapitel 1.2 als beantwortet. Theoretisch lassen sich die wichtigsten Eigenschaften eines Produkts und die zugehörigen Opinion Words durch das POS-Tagging bestimmen. In dem der Satz in einzelne Wörter gesplittet wird und diese ihren Wortklassen zugeteilt werden, können Nomen und Adjektive identifiziert werden. Die am häufigsten genannten Nomen zählen meist zu den Frequent Features. Die Adjektive stellen die Opinion Words dar (vgl. Abbildung 42).

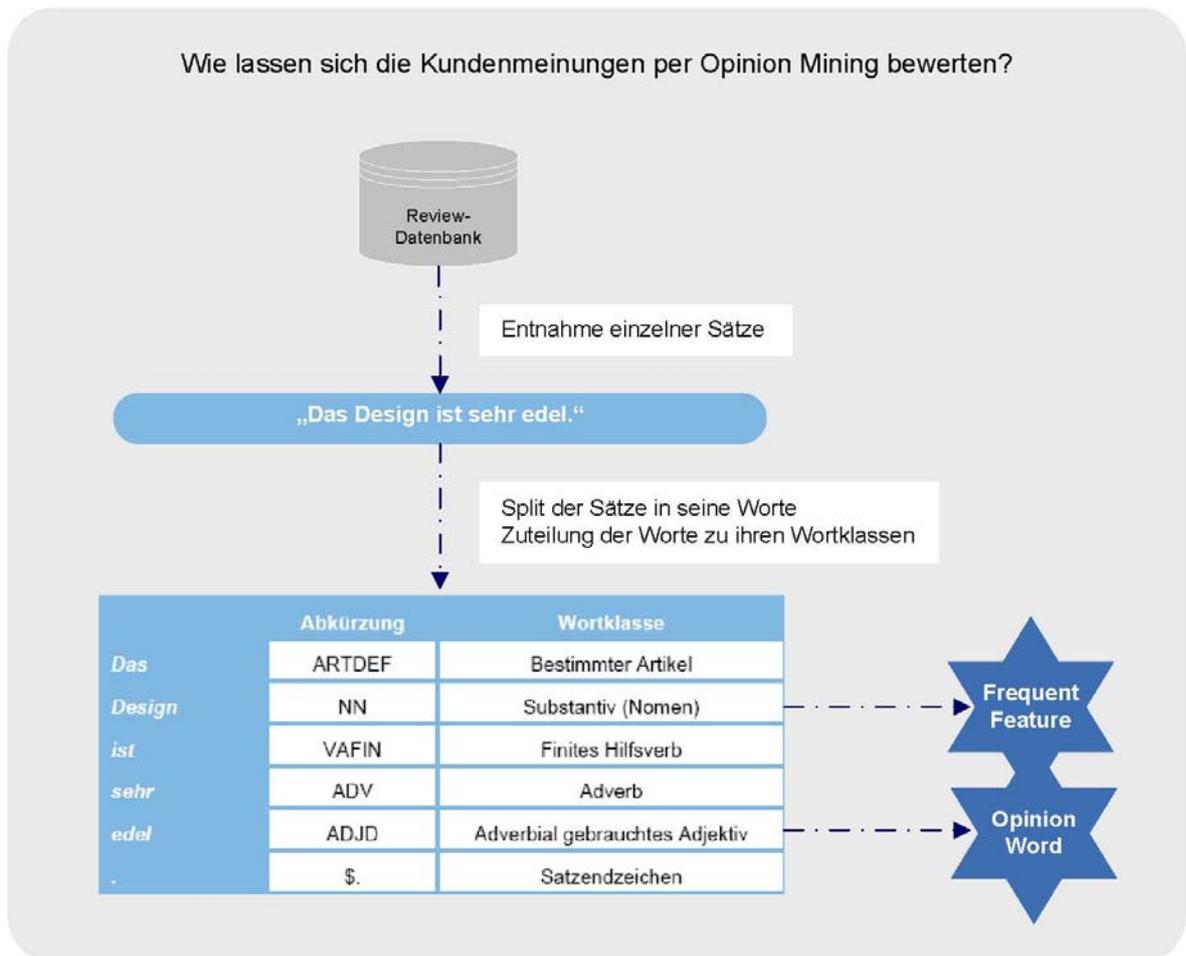


Abbildung 42: Beantwortung der Forschungsfrage 2

Wie lassen sich die Kundenmeinungen per Opinion Mining bewerten?⁴⁰

Da der RapidMiner das POS-Tagging nicht unterstützt, wird in der Praxis auf eine Alternative zurückgegriffen. Anhand der Auftrittshäufigkeit der einzelnen Worte, welche im Ergebnis einzusehen ist, können die Frequent Features (Nomen) und Opinion Words (Adjektive) manuell bestimmt werden.

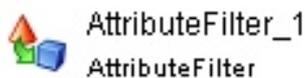
⁴⁰ Das Wort "edel" ist im Zusammenhang dieses Beispielsatzes grammatikalisch kein adverbial gebrauchtes Adjektiv sondern ein prädikativ gebrauchtes Adjektiv.

3.2.4 Analyse

Das Ziel dieser Phase besteht darin, aussagekräftige Analyseergebnisse zu ermitteln, welche sich durch eine hohe Aussage- und Interpretationsfähigkeit auszeichnen. Statt wie in Kapitel 3.1.4 eine Phase des Modeling anzuschließen folgt in der praktischen Umsetzung vielmehr eine Textanalyse. Demnach erfolgen keine Vorhersagen auf Basis von Trainings- und Testdatensätzen (wie eigentlich im Vorgehensmodell vorgesehen), sondern lediglich Berechnungen, die auf den im vorherigen Kapitel präparierten Daten basieren. Grund für diese Entscheidung ist zum einen, dass die Trainings- und Testmodelle im RapidMiner nur eingeschränkt anwendbar sind. Zum anderen fehlt den Ergebnissen die notwendige Repräsentativität, da die Klassifikation von noch nicht eingeordneten Sätzen durch die gewählte Vorgehensweise im Opinion Mining Tool nur selten fehlerfrei erfolgt. Vor allem neutrale Sätze werden oftmals fälschlicherweise in positive oder negative Meinungen eingeordnet, wodurch das Ergebnis erheblich manipuliert wird.

3.2.4.1 Analyse im RapidMiner

Zur Vereinfachung der Berechnungen werden die Attribute im RapidMiner gefiltert und es werden zusätzlich neue Attribute geschaffen.



AttributeFilter_1
AttributeFilter

- Dem *AttributeFilter* werden unter „parameter_string“ die Attribute mitgegeben, die ausschließlich im weiteren Verlauf des Modells betrachtet werden sollen (vgl. Abbildung 43). Dazu zählen die Opinion Words, welche durch den DictionaryStemmer positiv, negativ und neutral zusammengefasst wurden sowie die Frequent Features.

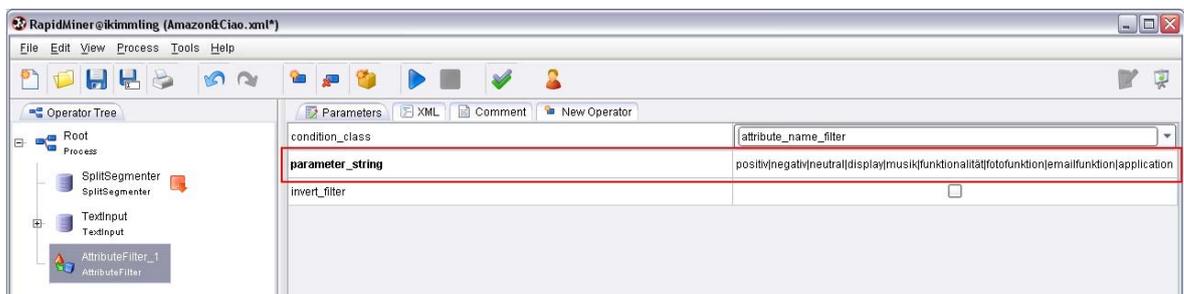


Abbildung 43: AttributeFilter_1 im RapidMiner

Quelle: RapidMiner Version 4.3 (Operatoren: SplitSegmenter, TextInput (inkl. Operatoren zur Vorverarbeitung, AttributeFilter)



- Der Operator *AttributeConstruction* ermöglicht das Zusammenfügen von mehreren Attributen. Unter „function_descriptions“ werden die neu zu generierenden Attribute benannt. Die Formel zur Berechnung dieser Attribute wird unter „function_expressions“ eingegeben (vgl. Abbildung 44). Ziel ist es, neue Attribute aus den Opinion Words und den Frequent Features zu generieren (bspw. Display_positiv).⁴¹ Der Wert dieser Attribute gibt Auskunft darüber, wie oft das jeweilige Frequent Feature in einem Satz positiv, negativ oder neutral bewertet wurde.

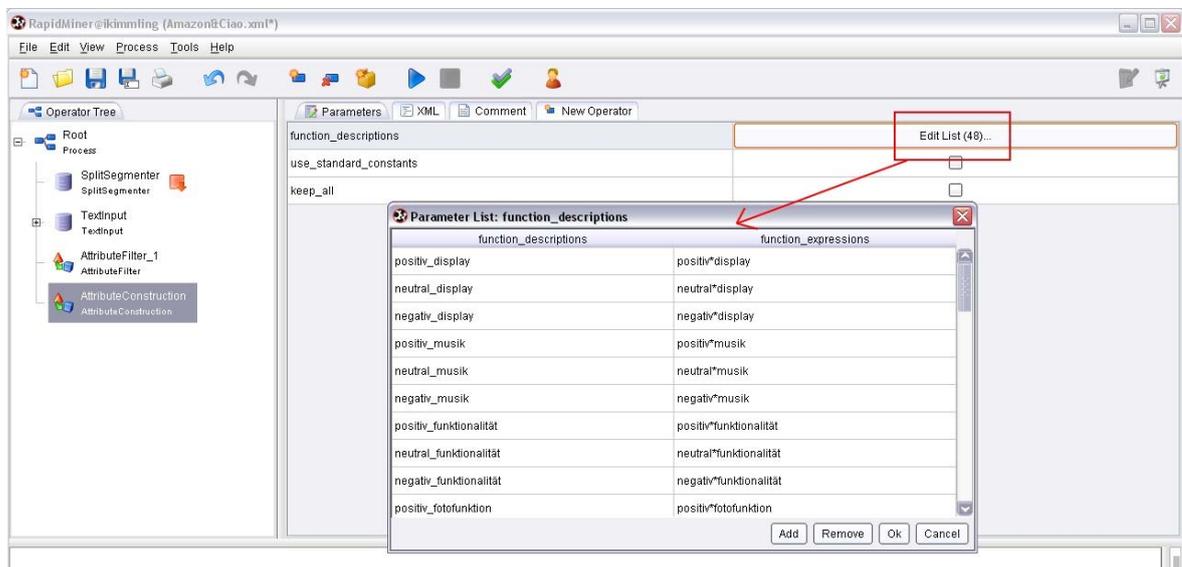
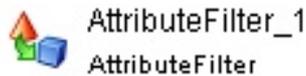


Abbildung 44: AttributeConstruction im RapidMiner

Quelle: RapidMiner Version 4.3 (Operatoren: SplitSegmenter, TextInput (inkl. Operatoren zur Vorverarbeitung, AttributeFilter, AttributeConstruction))

⁴¹ Neu generierte Attribute sind:

Display_positiv, Display_negativ, Display_neutral, Musik_positiv, Musik_negativ, Musik_neutral, Funktionalität_positiv, Funktionalität_negativ, Funktionalität_neutral, Fotofunktion_positiv, Fotofunktion_negativ, Fotofunktion_neutral, Emailfunktion_positiv, Emailfunktion_negativ, Emailfunktion_neutral, Applications_positiv, Applications_negativ, Applications_neutral, iTunes_positiv, iTunes_negativ, iTunes_neutral, Internetfunktion_positiv, Internetfunktion_negativ, Internetfunktion_neutral, Videofunktion_positiv, Videofunktion_negativ, Videofunktion_neutral, SMS_positiv, SMS_negativ, SMS_neutral, Kamera_positiv, Kamera_negativ, Kamera_neutral, Akku_positiv, Akku_negativ, Akku_neutral, Bedienbarkeit_positiv, Bedienbarkeit_negativ, Bedienbarkeit_neutral, Vertrag_positiv, Vertrag_negativ, Vertrag_neutral, Betriebssystem_positiv, Betriebssystem_negativ, Betriebssystem_neutral.



- Ein weiterer *AttributeFilter* wird abschließend benötigt, um neben den Opinion Words ausschließlich die neu generierten Attribute im Ergebnis ausgegeben zu bekommen (vgl. Abbildung 45).

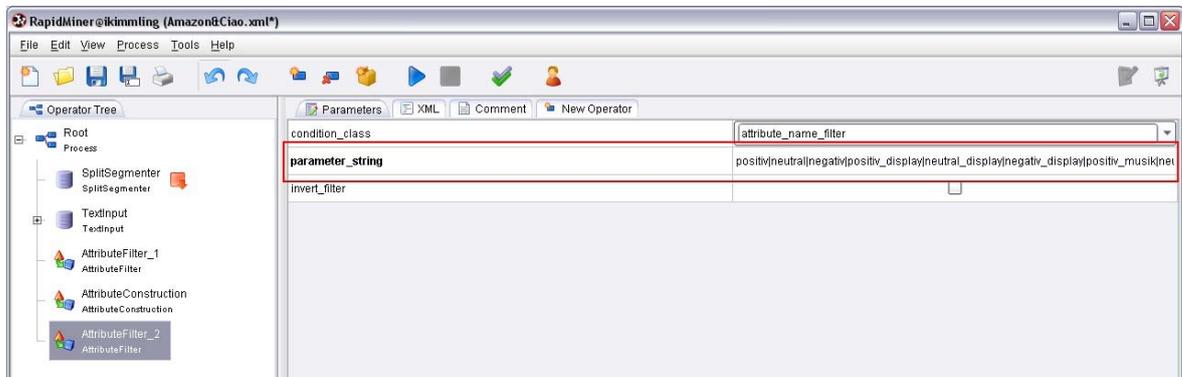
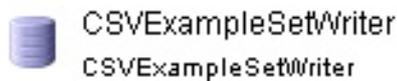


Abbildung 45: AttributeFilter_2 im RapidMiner

Quelle: RapidMiner Version 4.3 (Operatoren: SplitSegmenter, TextInput (inkl. Operatoren zur Vorverarbeitung, AttributeFilter, AttributeConstruction, Attribute Filter)



- Dem CSVExampleSetWriter wird unter „parameter_string“ ein Pfad mitgegeben, wo eine CSV-Datei abgelegt wird, welche die Ergebnisse des Modells beinhaltet (vgl. Abbildung 46).

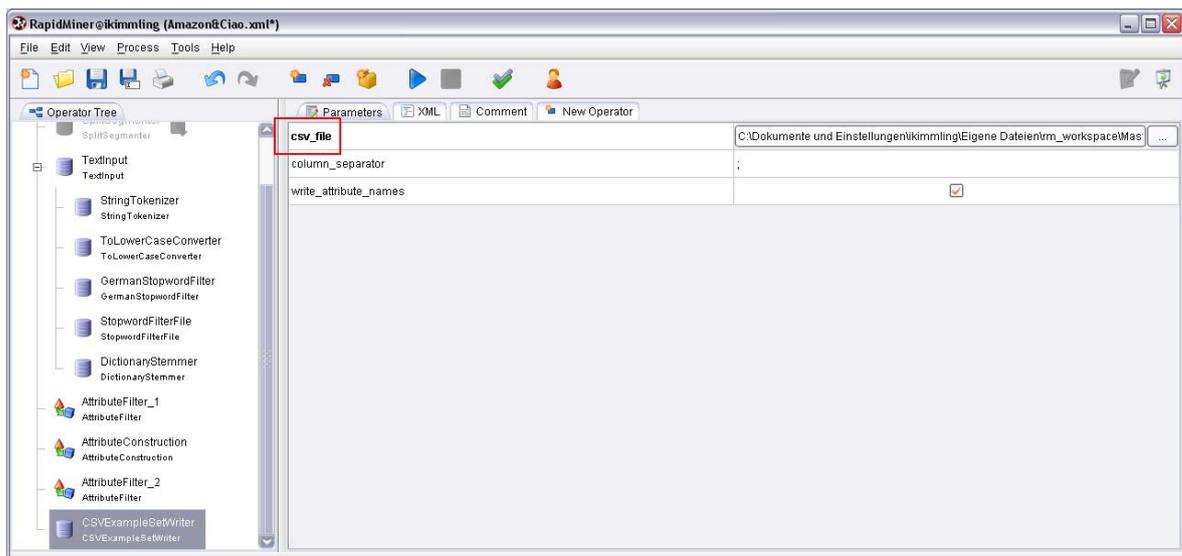


Abbildung 46: CSVExampleSetWriter im RapidMiner

Quelle: RapidMiner Version 4.3 (Operatoren: SplitSegmenter, TextInput (inkl. Operatoren zur Vorverarbeitung, AttributeFilter, AttributeConstruction, Attribute Filter, CSVExampleSetWriter)

3.2.4.2 Analyse der Ergebnisdatei

Basierend auf der generierten CSV-Datei⁴² erfolgen die Berechnungen (bspw. in Excel) zur Analyse der Ergebnisse. Jede Ergebniszeile beinhaltet jeweils einen Satz des anfänglichen Meinungstextes. Insgesamt beinhaltet das Ergebnis 12.227 Sätze. Die Datei wird dahingehend bearbeitet, dass letztendlich ausschließlich die Zeilen bestehen bleiben, welche mindestens ein Opinion Word und die zuvor per AttributeConstruction generierten Attribute (Frequent Features) enthält. Somit erhält man ausschließlich Sätze mit Meinungen bezüglich der Frequent Features. Als problematisch erweisen sich jedoch die Sätze, in denen zum einen verschiedenen konnotierte Opinion Words oder zum anderen mehrere Frequent Features auftreten.

Im Folgenden wird anhand einer zufälligen Auswahl einiger gelöschter Sätze⁴³, welche Opinion Words mit verschiedenartiger semantischer Bestimmung enthalten, aufgezeigt, dass diese zu einer Verfälschung des Ergebnisses führen.

Satz Nr. 2436:

*„Nur bei der ersten Generation gab es eine Dockingstation dazu(Kostenpunkt 30 Euro bei den neuern Modellen) - das erste Modell hat eine Stahlrückseite ist somit einerseits **robuster** und weniger anfällig für Kratzer als das 3g oder das 3gs, andererseits sieht es **hochwertiger** aus und ist es auch - der **Akku** ist besser als beim 3g, da das **fehlende** Umts keinen so hohen Stromverbrauch hat.“*

- FF: Akku
- OW positiv: robuster, hochwertiger
- OW negativ: fehlende

Satz Nr. 6441:

*„Man ist also immer **bestens** verbunden und kommt fast an jedem Ort mit dem iPhone ins **Internet**, was sehr **praktisch** ist, aber in kleineren Tarifen **teuer** werden kann (siehe Punkt 8) T-Mobile hat als Netzbetreiber in Deutschland ein sehr **gut** ausgebautes, wenn nicht sogar das **beste** Mobilfunknetz.“*

- FF: Internetfunktion
- OW positiv: bestens, praktisch, gut, beste
- OW negativ: teuer

⁴² Auf der beigefügten CD ist unter „Ergebnismodelle → Ergebnisse iPhone insgesamt → Ergebnis_unbearbeitet.csv“ die ursprüngliche Ergebnisdatei abgelegt.

⁴³ Die gelöschten Sätze stammen aus der Ergebnisdatei des iPhone allgemein.

Anhand dieser beiden Sätze ist zu erkennen, dass Vergleiche im Opinion Mining nur schwer zu bewerten sind. Zum einen treten sowohl positive als auch negative Opinion Words auf, die dem richtigen Frequent Feature zugeordnet werden müssen, um keine falschen Aussagen zu treffen. Zum anderen ist in Satz Nr. 2436 von verschiedenen Produktgenerationen die Rede, so dass die Beurteilung dem richtigen Produkt zuzuordnen ist, um Verzerrungen zu vermeiden.

Satz Nr. 3707:

„**Innovative** Multi-Touch-Steuerung

- **Stylische** Menüs
- **Simple** Einrichtung - überhaupt geht alles super-**easy**
- sehr durchdachte **Bedienung**
- Schnelle **Internet**-Verbindung
- Kaum Verzögerungen
- Sehr **gute** Empfangseigenschaften
- Sehr **gute** Sprachverständlichkeit
- **relativ** handlich
- sehr **gute** Haptik
- **AppStore** mit vielen, vielen - teils sehr **coolen** - **Programmen**

Wie bekannt fehlen dem iPhone jedoch eine ganze Reihe von **Features**, die uns Apple teilweise absichtlich vorenthält:

- Kein Flashplayer (bei meinem auf 200MB beschränkten Tarif vielleicht auch besser)
- Bluetooth nicht zum Datenaustausch geeignet
- MMS nicht möglich (**Software** aber im **Appstore** erhältlich oder unlocken)
- Keine **Videoaufzeichnung** (oder unlocken)
- Keine Sprachaufzeichnung (Voice-Recorder) -oder schaut mal in den **Appstore**“
 - FF: Bedienbarkeit, Internetfunktion, Applications, Funktionalität, Betriebssystem, Videofunktion
 - OW positiv: innovative, stylische, easy, durchdachte, gute, coolen
 - OW neutral: simple, relativ

Satz Nr. 4740:

„**Wunderschönes** Gerät mit vielen Nachteilen: nur eine 2-Megapixel-**Kamera** und kann keine MMS versenden sowie keine Diktate oder **Videos** aufzeichnen, langsame Aktivierung über **Itunes** (**Ipod** ist leider völlig abhängig von **Itunes**), lang-

same Aufbau der Internetseiten, **schwache Akkuleistung**, zu kleine Tastatur für **SMS**, noch keine Navigation per GPS, nur Anzeige des Standorts.“

- FF: Kamera, iTunes, Musik, Akku, SMS
- OW positiv: wunderschönes
- OW negativ: schwache

Satz Nr. 6607:

“Es sieht **elegant** aus, ist etwas größer als **normal**, besticht durch eine **revolutionäre Bedienung** und durch seine Einfachheit und seinem Dasein als bester **Musikplayer** der Welt. “

- FF: Bedienbarkeit, Musik
- OW positiv: elegant, revolutionäre
- OW neutral: normal

Aufzählungen führen ebenfalls zu dem Problem der Zuordnung der richtigen Opinion Words zu den Frequent Features. Eine falsche Zuordnung führt zu verfälschten Ergebnissen.

Satz Nr. 608:

“Das Adressbuch, hier "Kontakte" kann **sauber** und **einwandfrei** mit Microsofts Outlook synchronisiert werden und **funktioniert** ohne Probleme, und nette Gimmicks sind implementiert: Wird man beispielweise von einem "Kontakt" angerufen, bei dem ein **Foto** hinterlegt ist, so wird dieses bei einem angehenden Anrufe Bildschirmfüllend angezeigt. “

- FF: Fotofunktion
- OW positiv: sauber, einwandfrei, funktioniert

Satz Nr. 7039:

“Schon der Start mit diesem Apple Handy ist nicht leicht: Man muss gleichzeitig (so war es bisher) einen **teuren T-Mobile-Vertrag** erfüllen, der für den Mobilfunkbetreiber vielleicht sinnvoll ist, für die meisten Nutzer aber sicher nicht: Denn mit 50 Inklusivminuten kommt man nicht sehr weit, es gibt keine **Frei-SMS** und der **billigste Vertrag** ist mit 29 Euro im Monat auch nicht wirklich **billig**. “

- FF: Vertrag, SMS
- OW neutral: billigste, billig
- OW negativ: teuren

Satz Nr. 7427:

“Die Tatsache, dass die **Software** dem iPhone nicht direkt beiliegt ist meiner Meinung nach nicht weiter **schlimm**, da sie ja **kostenlos** zum Download bereitsteht, außerdem würde die Beigabe einer CD die Verpackung des iPhones **unnötig** vergrößern, was nicht nur auf die **Optik** geht, sondern auch zu Lasten der Umwelt, von daher meiner Meinung nach eine **gute** Lösung von Seiten Apples“

- FF: Betriebssystem, Design
- OW positiv: kostenlos, gute
- OW negativ: schlimm, unnötig

Diese sehr langen und stark verschachtelten Sätze sind schon für einen menschlichen Leser nicht einfach zu verstehen. Ihren korrekten Inhalt durch einen maschinellen „Leser“ (bspw. ein Opinion Mining Tool) zu erfassen, ist problematisch, da auch hier wieder die den Frequent Features zugehörigen Opinion Words fehlerfrei eingeordnet werden müssen.

Satz Nr. 4549:

“Auf den zweiten Blick ist das iPhone aber nur die **billigste Technik** mit **minimalistischen Funktionen** in eine **schöne** Hülle gepackt. “

- FF: Betriebssystem, Funktionalität
- OW positiv: schöne
- OW neutral: billigste
- OW negativ: minimalistischen

Satz Nr. 7562:

“Meiner Meinung nach ist die **"Fotos" App** sehr **beeindruckend** in der **Handhabung**, aber die **Kamera** einfach zu **schwach** um zu glänzen, deshalb ein wenig Kritik an dieser Stelle. “

- FF: Fotofunktion, Bedienbarkeit, Kamera
- OW positiv: beeindruckend
- OW negativ: schwach

Satz Nr. 10.403:

“ (Ja, man kann sich all das (Bildverbesserer für die **mickerige 2MP-Cam**, Rekorder für milchige & **unterbelichtete** Filmchen) laden, und installieren, aber ich rede erst einmal vom Auslieferungszustand, ansonsten wird es zu umfangreich)

*Also würde ich, auf dem Teppich wieder angekommen, mal locker behaupten, dass das iPhone ein featurelastiges Durchschnittshandy in der wohl **geilsten** Verpackung der Welt ist, und das ist für mich persönlich als Kaufanreiz/ Kaufargument eigentlich auch schon **vollkommen** ausreichend. “*

- FF: Kamera
- OW positiv: geilsten, vollkommen
- OW negativ: mickerige, unterbelichtete

Diese Beispiele zeigen, dass in einem Satz sowohl positive als auch negative Aussagen bezüglich der Frequent Features getroffen werden können. Hier stellt sich wieder einmal das Problem der einwandfreien Zuordnung von Frequent Features und Opinion Words.

Nach der Elimination von Sätzen mit verschiedenen konnotierten Opinion Words aus dem Ergebnis, stellt sich die Frage, wie mit Sätzen umzugehen ist, welche mehrere Frequent Features beinhalten. Problematisch ist hier, dass jedes Frequent Features mit im Satz ebenfalls auftretenden Opinion Words gewichtet wird, obwohl das Opinion Word eventuell gar nicht in Zusammenhang zu dem Frequent Feature steht. Die folgenden Anteilsberechnungen sollen einen Aufschluss darüber geben, ob Sätze mit mehreren Frequent Features aus der weiteren Betrachtung ausgeschlossen werden sollten und ab wie vielen Frequent Features in diesem Falle der Ausschluss sinnvollerweise stattfinden sollte. Es wird zwischen vier Varianten unterschieden⁴⁴:

- Variante 1 (vgl. Abbildung 47) beinhaltet die Verteilung der Frequent Features, wenn alle Sätze mit mehreren unterschiedlichen Frequent Features betrachtet werden.
- Variante 2 (vgl. Abbildung 48) beinhaltet die Verteilung unter Einbezug der Sätze mit weniger als sechs (≤ 5) unterschiedlichen Frequent Features.
- Variante 3 (vgl. Abbildung 49) beinhaltet die Verteilung unter Einbezug der Sätze mit weniger als vier (≤ 3) unterschiedlichen Frequent Features.
- Variante 4 (vgl. Abbildung 50) beinhaltet die Verteilung unter Einbezug der Sätze mit genau einem (≤ 1) Frequent Feature.

⁴⁴ Auf der beigefügten CD ist unter „Ergebnismodelle → Ergebnisse iPhone insgesamt → Ergebnis_alleFFs.xls“ ist die bearbeitete Ergebnisdatei ohne mit allen FFs abgelegt. Unter „Ergebnismodelle → Ergebnisse iPhone insgesamt → Ergebnis_FF kleiner= 5.xls“ ist die bearbeitete Ergebnisdatei mit FFs ≤ 5 abgelegt. Unter „Ergebnismodelle → Ergebnisse iPhone insgesamt → Ergebnis_FF kleiner= 3.xls“ ist die bearbeitete Ergebnisdatei mit FFs ≤ 3 abgelegt. Unter „Ergebnismodelle → Ergebnisse iPhone insgesamt → Ergebnis_FF kleiner= 1.xls“ ist die bearbeitete Ergebnisdatei mit FFs ≤ 1 abgelegt.

Die Produkteigenschaften werden in absteigend positiver Reihenfolge aufgezählt.

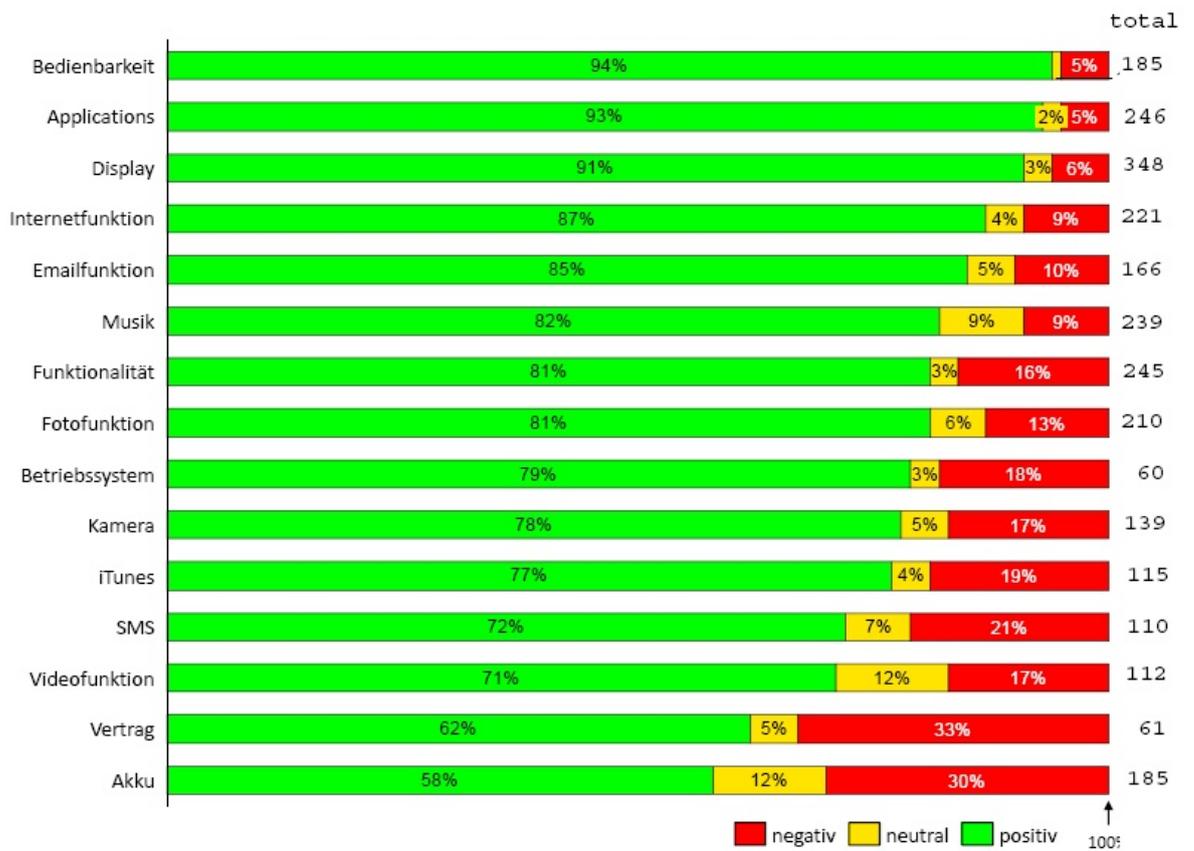


Abbildung 47: Vergleich der FFs (alle Sätze mit unterschiedlichen FFs)

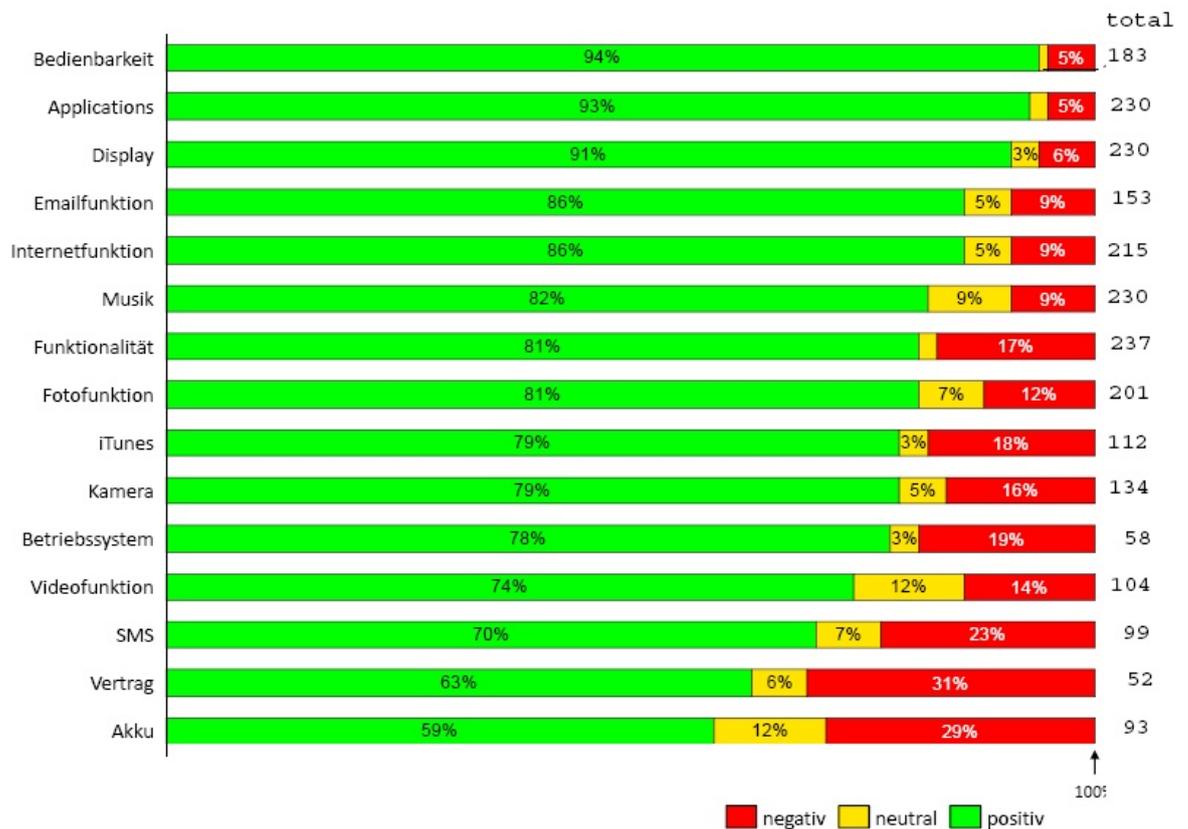


Abbildung 48: Vergleich der FFs (ohne Sätze mit mehr als 5 FFs)

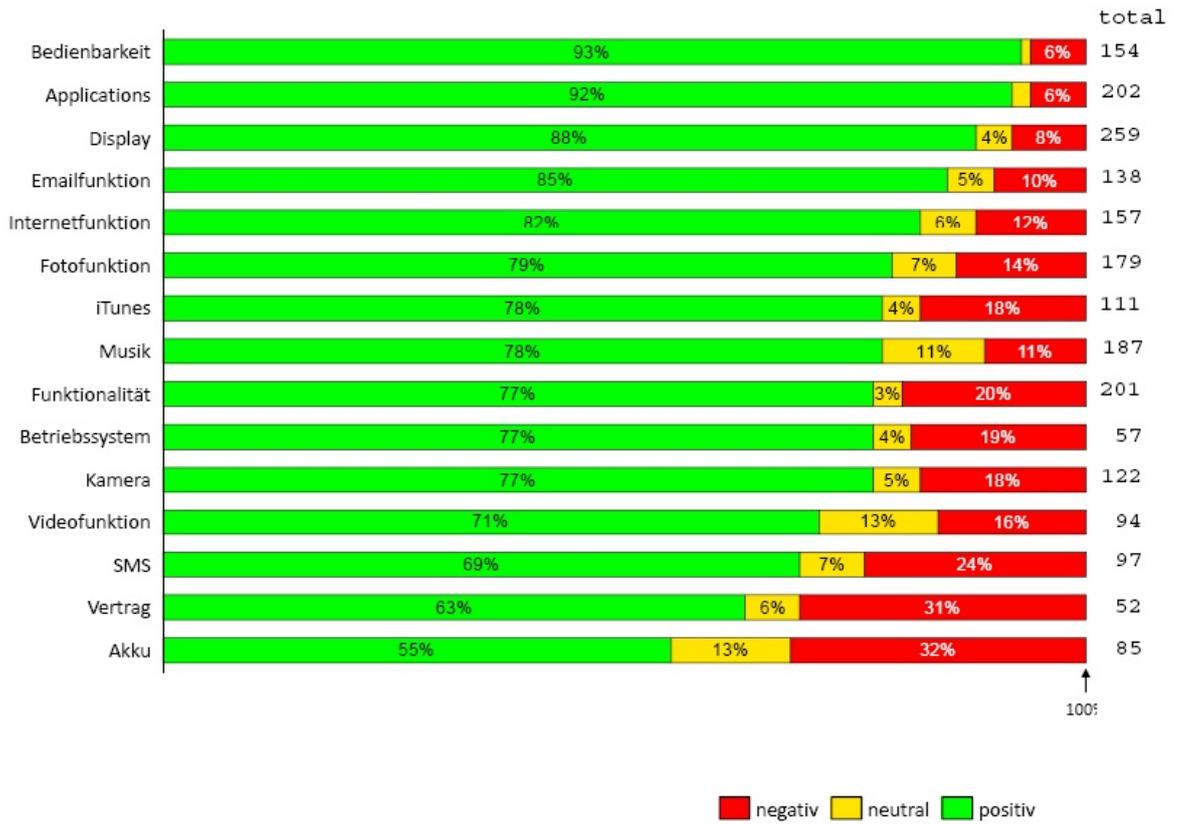


Abbildung 49: Vergleich der FFs (ohne Sätze mit mehr als 3 FFs)

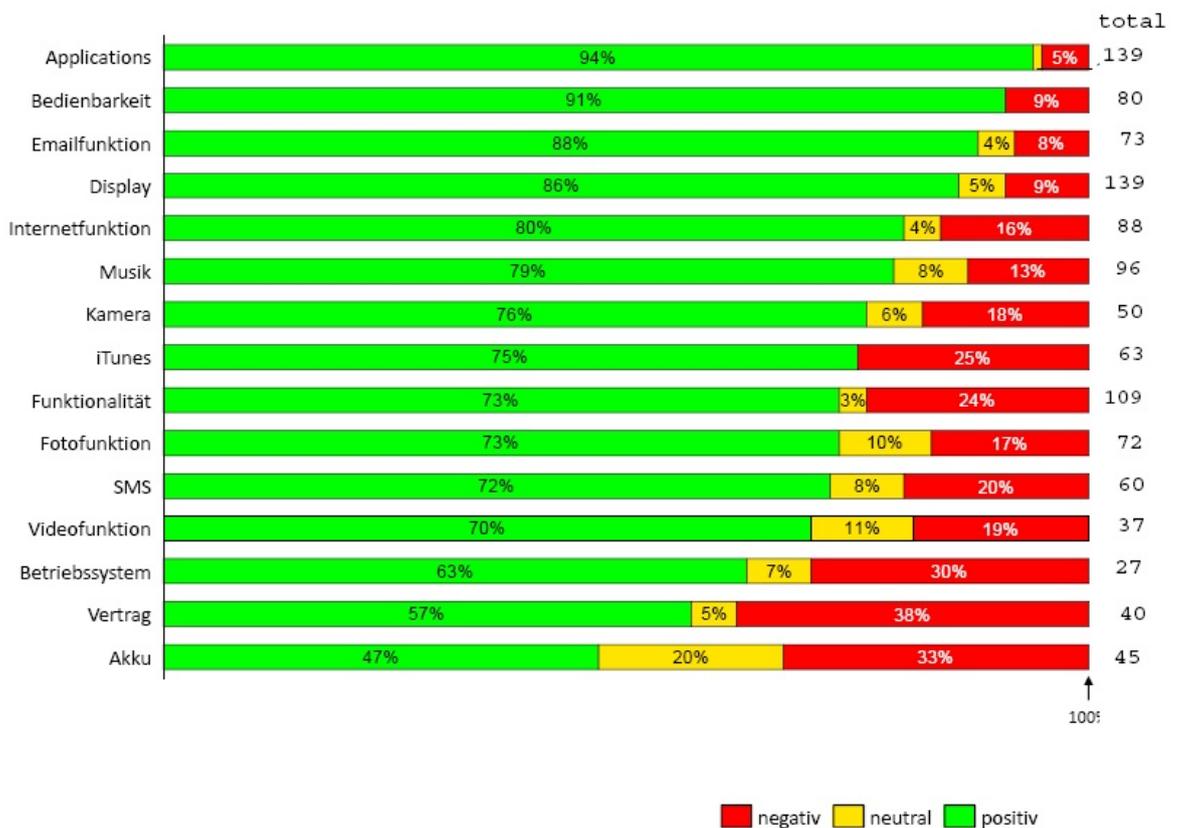


Abbildung 50: Vergleich der FFs (ohne Sätze mit mehr als 1 FF)

Am stärksten positiv bewerten die Rezensenten die Produkteigenschaften Applications, Bedienbarkeit und Display. Am wenigsten zufrieden sind sie mit dem Akku und der Vertragsbindung.

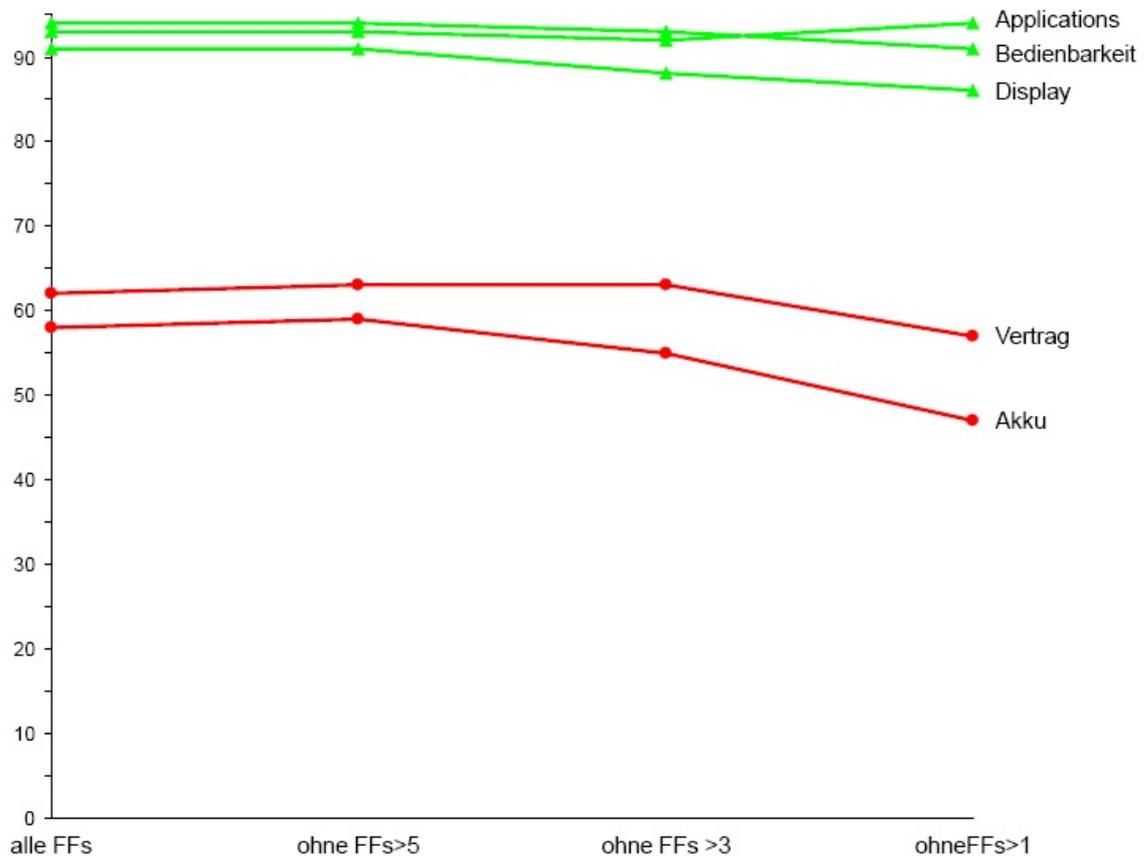


Abbildung 51: Verteilung der am stärksten positiv und negativ bewerteten FFs

Im Vergleich der zuvor vorgestellten vier Varianten der Frequent Feature-Betrachtung (s. Abbildung 51) ist zu erkennen, dass sich die Verteilung der wichtigsten Produkteigenschaften nur minimal verändert. Im Folgenden werden demnach auch Sätze mit mehreren Frequent Features (Variante 1) betrachtet. Zwar wird in Variante 4 die „Bedienbarkeit“ als beliebteste Produkteigenschaft von den „Applications“ verdrängt, jedoch fällt die Entscheidung trotzdem darauf, auch Sätze mit mehreren verschiedenen Frequent Features in Betracht zu ziehen, da sonst etwa 40% weniger Sätze in das Ergebnis mit einbezogen würden. Zudem hat das Löschen der Sätze keinen nennenswerten Einfluss auf die Interpretation des Ergebnisses. Demnach kann gefolgert werden, dass in der hier gewählten Vorgehensweise keine bedeutsamen Unterschiede in der Kodierung längerer und kürzerer Sätze auftreten.

Aus der ursprünglichen Ergebnisdatei werden demnach folgende Sätze entfernt:

- Sätze, die kein FFs oder OWs beinhalten und
→ 10.960 gelöschte Sätze
- Sätze; die OWs verschiedenartiger semantischer Bestimmung enthalten.
→ 143 gelöschte Sätze

Insgesamt werden also 11.103 Sätze (90%)⁴⁵ aus dem ursprünglichen Ergebnis entfernt. Die überarbeitete Ergebnisdatei⁴⁶ besteht somit schlussendlich noch aus 1.124 Sätzen. Abbildung 52 gibt noch einmal einen Überblick über die Bewertung der wichtigsten Produkteigenschaften des iPhone.

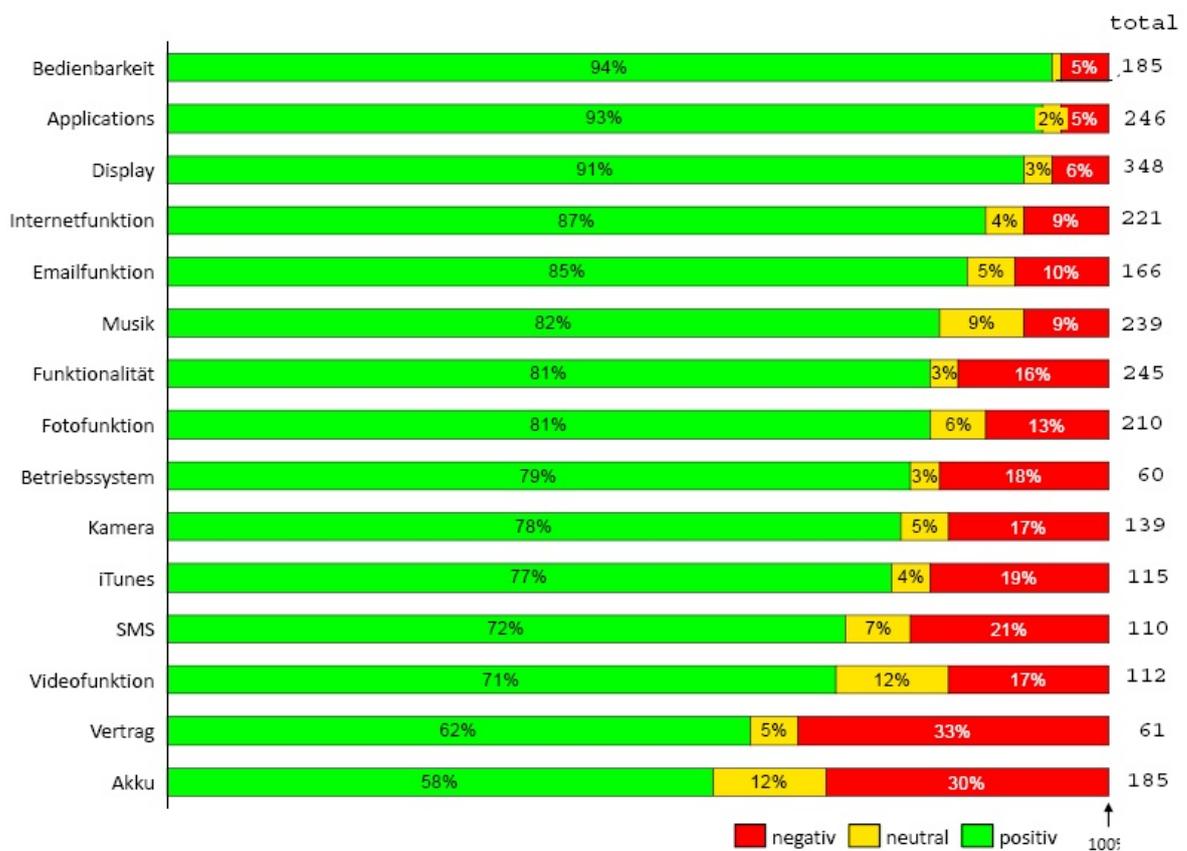


Abbildung 52: Bewertung der Frequent Features des iPhone

Die Rezensenten sind mit der Produkteigenschaft „Bedienbarkeit“ (94%) am zufriedensten. Ebenfalls stark positiv bewertet werden die „Applications“ des iPhone (93%) und das „Display“ (91%). Mit 33% sowie 38% werden „Vertrag“ und „Akku“ am schlechtesten beurteilt.

⁴⁵ Auf der beigefügten CD unter „Ergebnistabellen → Ergebnisse iPhone insgesamt → gelöschte Sätze aus dem ursprünglichen Ergebnis → Sätze die keine FFs enthalten.xls“ und unter „Ergebnistabellen → Ergebnisse iPhone insgesamt → gelöschte Sätze aus dem ursprünglichen Ergebnis → Sätze mit verschiedenen konnotierten OWs.xls“ sind die gelöschten Sätze einzusehen.

⁴⁶ Die überarbeitete Ergebnisdatei, auf der die Analyse basiert, ist auf der beigefügten CD unter „Ergebnistabellen → Ergebnisse iPhone insgesamt → bearbeitet_alleFFs.xls“ einzusehen.

Neben einer allgemeinen Auswertung des iPhone, welche alle Rezensionen des iPhone erfasst, kann auf die gleiche Weise eine Auswertung einzelner Modellgenerationen erfolgen. Nachfolgend werden die Generationen 3G und 3GS des iPhone analysiert. Dazu erfolgt ausschließlich die Betrachtung von Kundenrezensionen bezüglich dieser Modelle. Die Anteile verteilen sich dabei wie folgt (vgl. Abbildung 53 und Abbildung 54):

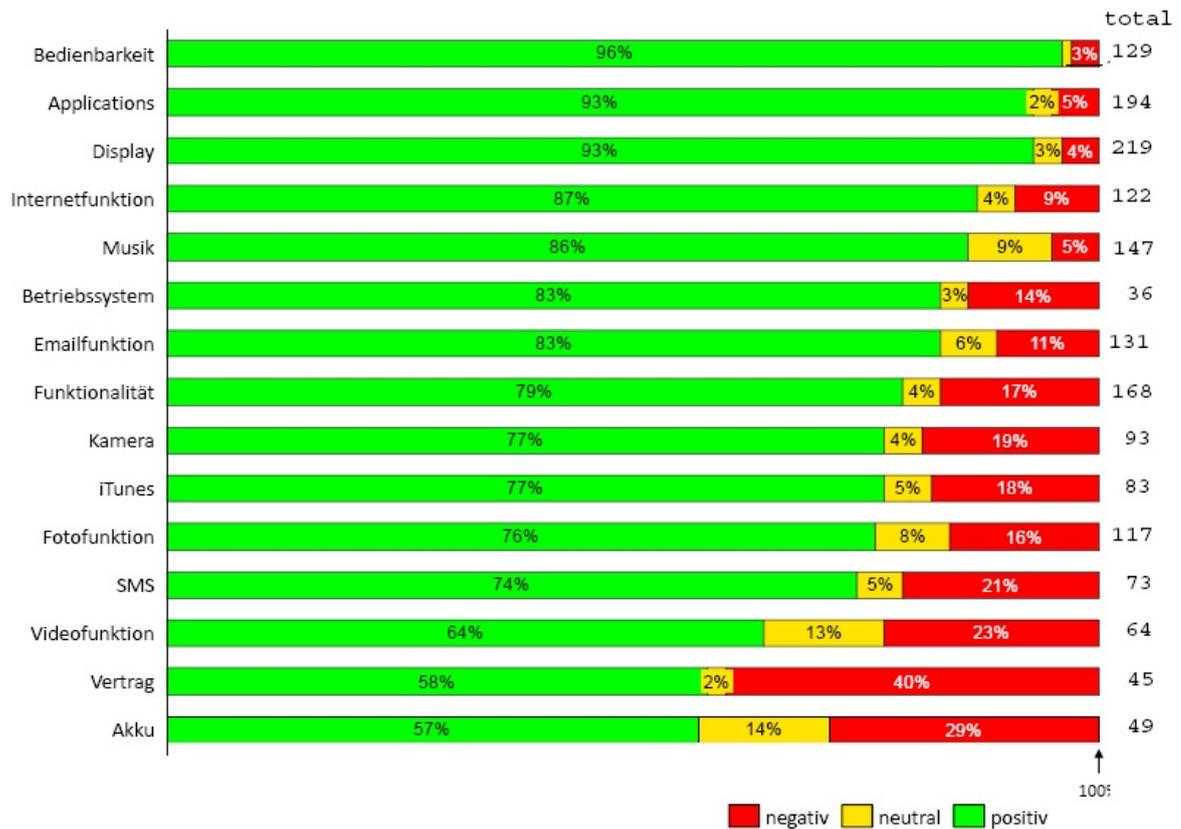


Abbildung 53: Bewertung des iPhone 3G

Die Verteilung ergibt sich aus 696 Sätzen. Die ursprüngliche Ergebnisdatei beinhaltet 7.729 Sätze, von denen 6.949 keine Frequent Features und 84 Opinion Words mit verschiedenartiger semantischer Bestimmung enthalten. Demnach werden 7.033 Sätze entfernt.⁴⁷ Die beliebtesten Produkteigenschaften der Generation 3G sind die „Bedienbarkeit“ (96%), die „Applications“ (93%) und das „Display“ (93%). Am schlechtesten schneiden – wie auch schon bei der allgemeinen Betrachtung des iPhone – der „Vertrag“ (40%) und der „Akku“ (29%) ab.

⁴⁷ Das RapidMiner Modell für die Auswertung des iPhone 3G findet sich auf der beigefügten CD unter „RapidMiner Modelle → Amazon&Ciao 3G.xml“. Die zugehörigen Ergebnisdateien sind unter „Ergebnistabellen → Ergebnisse iPhone 3G“ abgelegt.

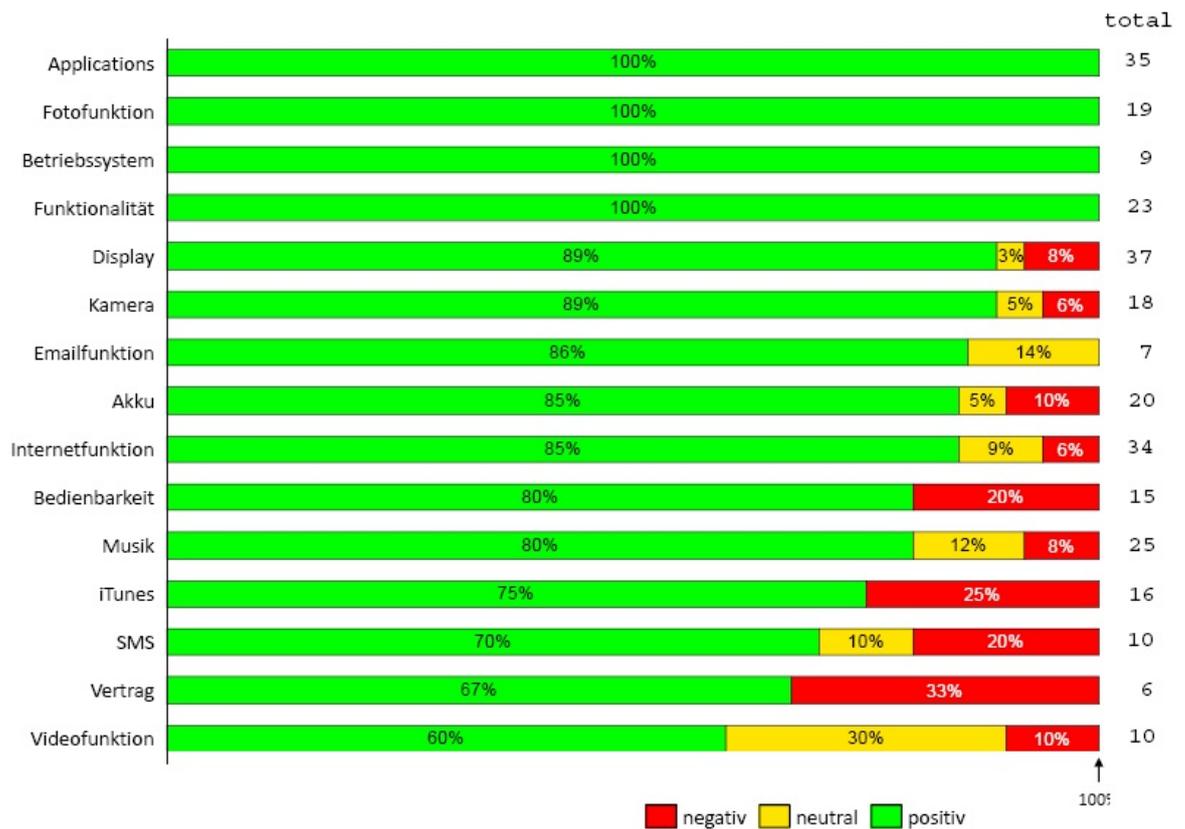


Abbildung 54: Bewertung des iPhone 3GS

Die Verteilung ergibt sich aus 125 Sätzen. Die ursprüngliche Ergebnisdatei beinhaltet 1.274 Sätze, von denen 1.138 keine Frequent Features und 11 Opinion Words mit verschiedenartiger semantischer Bestimmung enthalten. Demnach werden 1.149 Sätze entfernt.⁴⁸ Die Besitzer des iPhone 3GS bewerten die Produkteigenschaften „Applications“, „Fotofunktion“, „Betriebssystem“ und „Funktionalität“ zu 100% positiv. Die schlechteste Beurteilung erhält der „Vertrag“ mit 33%. Jedoch wird er von fast 70% der Nutzer immer noch positiv bewertet, so dass in diesem Fall nicht von einer negativen Beurteilung gesprochen werden kann. Somit werden die Frequent Features des iPhone 3GS durchweg positiv bewertet. Wirklich schlechte Beurteilungen lassen sich nicht ausmachen.

Im Vergleich der beiden Generationen (vgl. Abbildung 55) ist zu erkennen, dass die Produkteigenschaften „Applications“, „Betriebssystem“, „Fotofunktion“, „Akku“ und „Vertrag“ in der nachfolgenden Generation besser bewertet werden. Die stärkste Meinungsänderung (+28%) findet in Bezug auf den Akku statt. An

⁴⁸ Das RapidMiner Modell für die Auswertung des iPhone 3GS findet sich auf der beigefügten CD unter „RapidMiner Modelle → Amazon&Ciao 3GS.xml“. Die zugehörigen Ergebnisdateien sind unter „Ergebnistabellen → Ergebnisse iPhone 3GS“ abgelegt.

Ansehen verloren haben die Frequent Features „Display“, „Bedienbarkeit“ und „Videofunktion“, wobei die Bedienbarkeit mit -16% am stärksten betroffen ist.

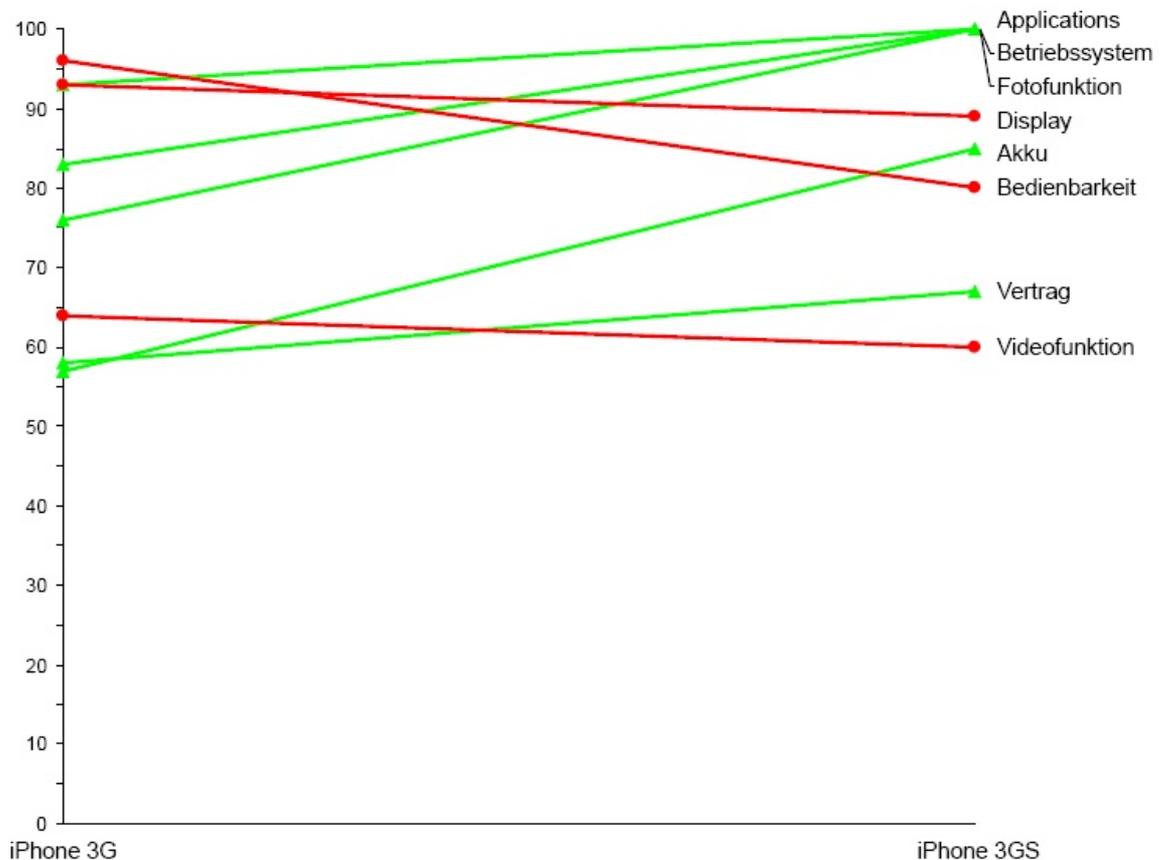


Abbildung 55: Vergleich der Bewertungen der Generationen 3G und 3GS

3.2.5 Evaluation & Deployment

Dieses Kapitel umfasst zum einen die Bewertung der Analyseergebnisse aus Kapitel 3.2.4. Zum anderen erfolgt eine Kritik am verwendeten Analysetools Rapid-Miner.

3.2.5.1 Bewertung der Opinion Mining Ergebnisse

Die Ergebnisse des Opinion Mining zeigen auf, dass das iPhone seinen Nutzern größtenteils sehr zusagt. Die besten Beurteilungen erhalten die Bedienbarkeit, die Applications und das Display. Das iPhone verfügt über eine intuitive Bedienung, die auch weniger technikaffine Anwender meist problemlos beherrschen. Ein großes Angebot an möglichen Anwendungen, welche im sogenannten App Store gesammelt zur Verfügung stehen, stellt eine attraktive Besonderheit des iPhone dar. Somit sind Apps, teilweise auch kostenfreie, schnell und einfach für jedermann zugänglich und steigern die Anwendungsmöglichkeiten des iPhone enorm. Das für

ein Smartphone recht große Display (3,5 Zoll) stellt einen weiteren Anreiz dar. Es besitzt eine hohe Pixelauflösung und verfügt zudem über einen Touch-Screen, welcher einfach per Fingerdruck zu bedienen ist.

Einzige Mängel sind der Exklusivvertrag des iPhone mit T-Mobile und der Akku. Wer das doch sehr teure Smartphone lieber günstiger in einem Vertrag erhalten möchte, muss den Vertragsbedingungen von T-Mobile zustimmen. Bei anderen Vertragsanbietern, welche unter Umständen persönlich besser eingeschätzte Konditionen anbieten, ist das iPhone nicht erhältlich. Desweiteren wird oftmals die zu geringe Akkulaufzeit beklagt und dass der Akku nicht selbstständig ausgetauscht werden kann.

Im Vergleich der Generationen 3G und 3GS fällt auf, dass es erhebliche Unterschiede in der Bewertung gibt. Diese können beispielsweise daraus resultieren, dass Apple Verbesserungspotenziale bei der Entwicklung der Nachfolgeneration ausgeschöpft hat. So scheint sich zum Beispiel die Akkulaufzeit des iPhone verbessert (+28%) zu haben. Zudem wurden Bugs behoben und fehlende Funktionen (z. B. Copy and Paste) ergänzt, so dass die Bewertung des Betriebssystems um 25% auf 100% positiver Rezensionen steigt. Ebenso wird die Fotofunktion des iPhone 3GS zu 100% positiv beurteilt. Ein Grund dafür kann beispielsweise die Erweiterung um eine Zoomfunktion sein. Die ständig steigende Anzahl an nützlichen Applications im App Store führt ebenfalls zu einer zu 100% positiven Kritik. Die Meinungsunterschiede können jedoch auch durch eine Anforderungsänderung der Nutzer entstanden sein. Das kann sich sowohl positiv als auch negativ auf die Beurteilung auswirken. Beispielsweise steigen die positiven Meinungen zur Vertragsbindung um knapp 10% an. Die positivere Einstellung könnte damit einhergehen, dass die iPhone Nutzer beim Kauf der Generation 3GS schon durch ihre Kenntnisse der Generation 3G auf die Vertragsbindung vorbereitet sind. Sie nehmen diese somit mit weniger Beschwerde hin als noch zur Zeit des Vorgängermodells, sie rekapitulieren gewissermaßen. Ebenso kann davon ausgegangen werden, dass die Ansprüche der Anwender an das Display, die Bedienbarkeit und die Videofunktion durch die Generation 3GS nicht erfüllt werden. Die genannten Produkteigenschaften sinken in ihrem Ansehen, da die Nutzer sehr wahrscheinlich Besseres erwartet haben. Hier kann Apple seine Kunden nicht genügend zufrieden stellen.

3.2.5.2 Prüfung der Validität

Nach der Vorstellung der Ergebnisse des Opinion Mining in Bezug auf das iPhone ist es interessant herauszufinden, wie stark diese Resultate der Realität entsprechen. Um eine Aussage darüber treffen zu können, ist eine Validierung notwendig. Ein möglicher Anhaltspunkt für die Validität ist die untersuchte Satzanzahl, also die Stichprobengröße. Je weniger Sätze analysiert werden, desto geringer ist die Validität, da eine geringere Anzahl an Meinungen in das Ergebnis einfließt. Dies ist auch daran zu erkennen, dass die Verteilung der Meinungen über das iPhone allgemein der über die Generation 3G sehr ähnlich ist. Eine Begründung dafür ist, dass ein Großteil (62%) der analysierten Sätze in Bezug auf das iPhone allgemein aus Rezensionen zur Generation 3G stammt. Den Rest stellen Rezensionen zum iPhone allgemein und zum iPhone 3GS dar.

Eine weitere Beurteilung der Validität ist anhand eines manuellen Tests möglich. Anhand von 50 zufällig ausgewählten Sätzen der bearbeiteten Ergebnisdatei⁴⁹ wird aufgezeigt, wie valide die Resultate in Wirklichkeit sind.⁵⁰ Bei 25 der 50 vorgestellten Sätze werden Probleme oder grenzwertige Fälle aufgezeigt. Demnach findet bei der Hälfte der überprüften Sätze eine korrekte Kodierung statt. Die restlichen 50% werden von der gewählten Vorgehensweise im RapidMiner falsch eingeordnet. Als typische Fehlerquellen treten vor allem die falsche Zuordnung von Opinion Words und Frequent Features auf. Zum einen, da sich die Opinion Words auf weitere im Satz genannte Eigenschaften des untersuchten Produktes beziehen. Zum anderen stehen die Opinion Words aber manchmal auch im Zusammenhang zu Eigenschaften von Konkurrenzprodukten, die vergleichsweise im Satz erwähnt werden. Ein weiteres Problem stellen Opinion Words dar, denen eine Negation vorangeht. Diese werden semantisch gegenteilig und folglich fehlerhaft bewertet. Zusätzlich kann auch die Auswahl der Splitargumente zur Aufteilungen eines Textes zu Beginn des Opinion Mining Probleme hervorrufen. So wird beispielsweise Satz Nr. 972 innerhalb der Phrase und nicht am Satzende vom restlichen Text getrennt. Die exakte Zuordnung von Frequent Features und Opinion Words stellt in solchen Fällen ein Problem dar. Es kann nicht immer sicher davon ausgegangen werden, dass das Opinion Word wirklich das genannte Frequent Feature statt einer anderen Produkteigenschaft, die im weiteren Verlauf des Satzes genannt würde, beschreibt. Problematisch sind auch fehlende Satzendzeichen, da dann eine Trennung der Sätze mit dem Splitargument nicht möglich ist.

⁴⁹ Die Sätze stammen aus der bearbeiteten Ergebnisdatei vom iPhone allgemein.

⁵⁰ Eine detaillierte Auflistung der untersuchten Meinungssätze findet sich im Anhang (D).

Solche Phrasen führen oftmals zu Problemen, da zu viele Frequent Features und Opinion Words in ihnen enthalten sind, um eine genaue Zuordnung durchführen zu können.

Neben der zuvor dargestellten Stichprobe erfolgt zusätzlich eine Überprüfung der Kodierung des Frequent Features „Akku“.⁵¹ Diese soll ebenfalls Aufschluss über die Repräsentativität der Ergebnisse geben.⁵² Von 61 Meinungssätzen bezüglich des Akkus kodierte die gewählte Vorgehensweise im RapidMiner 18 Sätze (29,5%) richtig, 26 Sätze (42,6%) falsch und 17 Sätze (27,9%) grenzwertig. Dieses Ergebnis ist wesentlich schlechter als das der vorherigen Validierung, da nur noch knapp 30% der Sätze vom RapidMiner semantisch richtig eingeordnet werden.

Die vorangegangene Prüfung der Validität würde eigentlich dafür sprechen, dass die Kodierung durch die gewählte Vorgehensweise im RapidMiner nicht der Realität entspricht. Jedoch zeigt Abbildung 51 sehr deutlich auf, dass eine falsche Zuordnung von Frequent Features und Opinion Words nicht unbedingt zu Problemen führt. Stattdessen verändert sich die Verteilung der wichtigsten Produkteigenschaften nur minimal zwischen der ersten (alle Sätze mit unterschiedlichen FFs) und der vierten Variante (ohne Sätze mit mehr als 1 FF). Dafür spricht auch die Auffälligkeit, dass bei näherer Betrachtung der grenzwertig oder falsch kodierten Sätze diese (zumindest teilweise) spätestens in Variante 4 (≤ 1 FF) aus der Analyse ausgeschlossen werden. Trotzdem sollte die Validität der Resultate in jedem Fall kritisch betrachtet werden. Zum einen ist die jeweilige Anzahl an untersuchten Sätzen zu beachten. Je mehr Meinungen vorliegen, desto zuverlässiger wird das Ergebnis. Zum anderen sollte die semantische Zuordnung der Sätze durch das Opinion Mining Tool mit einer gewissen Vorsicht betrachtet werden. Es wird empfohlen stichprobenartige Überprüfungen durchzuführen, da diese einen groben Überblick über die Validität der Ergebnisse liefern.

3.2.5.3 Bewertung des Opinion Mining Tools

Dieses Kapitel beinhaltet die Beantwortung der Forschungsfrage 3 (vgl. Kapitel 1.2), indem eine Einschätzung dahin gehend erfolgt, inwieweit die ausgewählte Software das Opinion Mining unterstützt.

⁵¹ Auf der beigefügten CD unter „Ergebnistabellen → Stichprobe Akku → Akku.xls“ können die Ergebnisse des Stichprobentests eingesehen werden.

⁵² Eine detaillierte Auflistung der Meinungssätze zum Frequent Feature „Akku“ und die Beurteilung der Kodierung finden sich im Anhang (E).

Der RapidMiner, welcher in dieser Arbeit als Opinion Mining Tool ausgewählt wurde, befindet sich in einem steten Wandel. Seit der Einführung der ersten Version im Jahr 2004, bekannt unter dem Namen YALE, ist bis heute eine Vielzahl an weiteren Versionen verfügbar. Ende September diesen Jahres kam die bisher neueste RapidMiner Version Beta 5.0 auf den Markt.⁵³ Die schnelle Weiterentwicklung des Tools spricht dafür, dass die Entwickler von Rapid-I ständig Verbesserungspotenziale erkennen und diese in neuen Anwendungen innerhalb des Tools umsetzen. Beim Auffinden von Bugs und fehlenden Anwendungsmöglichkeiten ist ihnen sicherlich das Forum von Rapid-I⁵⁴ eine sehr große Unterstützung. Dort findet ein reger Austausch von Fragen, Antworten, Tipps und Verbesserungsvorschlägen statt. Vor allem die Beantwortung der Fragen durch die Entwickler des RapidMiner erweist sich für den Anwender als besonders hilfreich. Innerhalb von zwei Werktagen kann in der Regel mit einer professionellen Antwort gerechnet werden. Neben dem Forum von Rapid-I steht jedoch nur noch das Tutorial als zusätzliche Hilfe zur Verfügung. Dieses ist aber unvollständig (einige der Operatoren sind dort nicht aufzufinden) und die Erklärungen sind oftmals nicht ausreichend. Immer wieder steht der Anwender vor dem Problem, dass er Details (wie bspw. Einstellungen) der Operatoren nicht ausreichend kennt oder versteht. Abhilfe schafft hier meist nur das selbstständige Ausprobieren verschiedener Alternativen, bis die richtige Einstellung für das gewünschte Ergebnis gefunden ist. Da der RapidMiner zudem nur unzureichend intuitiv zu bedienen ist, stellt sich häufig die Frage, welche Operatoren überhaupt anzuwenden sind, um ein bestimmtes Ergebnis zu erhalten. Problematisch ist dabei vor allem, dass einige wichtige Opinion Mining Methoden mit den vorhandenen Operatoren überhaupt nicht unterstützt werden. Dazu zählt beispielsweise das POS-Tagging, welches zum Auffinden der Frequent Features und Opinion Words benötigt wird.

Auch der Umgang mit Negationen innerhalb der Rezensionen stellt sich als schwierig dar, da dieser nicht durch den RapidMiner unterstützt wird. Das Tool müsste die Fähigkeit besitzen, innerhalb eines Satzes zu erkennen, ob sich in der Nähe eines Opinion Words eine Negation befindet. In diesem Fall wäre eine Umkehr der Konnotation dieses Adjektivs notwendig (von positiv in negativ und umgekehrt). Ohne die Berücksichtigung der Negationen findet eine Verzerrung der Ergebnisse statt, da positiv ausgedrückte Meinungen (zum Beispiel „nicht übel“)

⁵³ Die bisherigen Versionen des RapidMiner finden sich unter: <http://sourceforge.net/projects/yale/files/> (Abruf am 03.11.2009).

⁵⁴ Das Rapi-I Forum kann unter <http://forum.rapid-i.com/> abgerufen werden (Abruf am 03.11.2009).

negativ und umgekehrt negative Meinungen (beispielsweise „nicht gut“) positiv bewertet werden. Zur Beseitigung dieser Verzerrung ohne einen speziellen Operator kann alternativ im gesamten Text nach Negationen gesucht werden. Diese werden dann mit dem darauf folgenden Opinion Word zusammengefasst. So würde zum Beispiel „nicht übel“ in „nichtübel“ oder „nicht gut“ in „nichtgut“ umgeformt. Anschließend kann die semantische Klassifikation dieser neuen Wörter erfolgen, die Verzerrung im Ergebnis wäre somit beseitigt.

Neben den Negationen führt Opinion Spam ebenfalls zu einem verzerrten Ergebnis. Darunter werden übertrieben positive oder negative Meinungen verstanden, die für persönliche Zwecke veröffentlicht werden. Negative Rezensionen führen häufig dazu, dass sich potenzielle Kunden in Richtung eines Konkurrenzproduktes orientieren. Positive Bewertungen hingegen führen oftmals zum Kauf eines Produkts. Demnach gibt es zwei Hauptgründe dafür Spam-Bewertungen zu schreiben. Der Reputation Evangelist (z. B. Produzent) möchte ein Produkt stark anpreisen. Der Reputation Terrorist⁵⁵ (z. B. Konkurrent) verbreitet Spam, um der Reputation eines Produktes Schaden zuzufügen (vgl. Liu 2008, S.441 f.). Für dieses Thema stellt der RapidMiner ebenfalls keine Lösung zur Verfügung. Jedoch befindet sich die Forschung zu diesem Aspekt auch noch am Anfang. Liu (vgl. ebd. 2008, S.444 ff.) stellt mögliche Alternativen zur Aufdeckung von Opinion Spam dar. Dazu zählt unter anderem die Beobachtung der Zeitfolge. Beispielsweise weist eine größere Anzahl an ähnlichen Bewertungen zu einer annähernd gleichen Zeit auf möglichen Spam hin. Desweiteren sind Spammer oftmals die ersten Rezensenten eines Produkts, da die frühen Bewertungen dazu tendieren, einen stärkeren Einfluss auf die Meinungsbildung zu besitzen.

3.3 Implikationen für Forschung und Praxis

Basierend auf den Ergebnissen des vorherigen Kapitels (3.2) folgt in diesem Abschnitt die Beantwortung der vierten Forschungsfrage (vgl. Kapitel 1.2), indem Implikationen für die Forschung und Praxis ausgesprochen werden.

3.3.1 Implikationen für die Forschung

Weiterer Forschungs- und Entwicklungsbedarf ist vor allem im Bereich der Opinion Mining Tools notwendig. Die Bewertung des RapidMiner ergibt, dass das Tool noch stark ausbaufähig ist. Aufgrund der noch am Anfang stehenden Forschung zum Thema Opinion Mining ist jedoch zu bezweifeln, dass Konkurrenzprodukte

⁵⁵ Der Reputation Terrorist verfolgt das Ziel der Reputation eines Unternehmens Schaden zuzufügen (vgl. Rand).

eine bessere Alternative darstellen. Die linguistischen Besonderheiten einer jeden Sprache erschweren die Entwicklung einer fehlerfreien und intuitiv zu bedienenden technischen Umsetzung enorm. Die rasche Weiterentwicklung des RapidMiner lässt jedoch hoffen, dass dieser in naher Zukunft ein Tool darstellt, welches den praktischen Gebrauch des Opinion Mining für den Anwender stufenweise vereinfacht. Bis dahin könnte beispielsweise eine einfache Unterscheidung in Meinung und Nicht-Meinung eine denkbare Alternative darstellen. Die Meinung selbst kann positiver oder negativer Natur sein. Die Nicht-Meinung hingegen beinhaltet die neutralen Äußerungen. Damit entsteht eine simple Auflistung an Meinungen, welche anhand persönlicher Einschätzungen positiv oder negativ bewertet wird. Die Entwickler eines Produkts sollten dieses hinreichend kennen, um die Meinungen richtig einordnen zu können. Somit wäre eine realere Klassifikation der Meinungen denkbar.

Ein ausgereiftes Opinion Mining Tool müsste die Fähigkeit besitzen, die natürliche Sprache mit ihren charakteristischen Merkmalen zu verstehen. Dazu gehören:

- ihr umfangreicher Wortschatz und ihre vielfältigen generellen Ausdrucksmittel,
- die Unschärfe sowohl von Wörtern als auch von ganzen Äußerungen,
- Mehrdeutigkeiten auf verschiedenen sprachlichen Ebenen,
- die Einbettung von Äußerungen in bestimmte sprachliche und situative Kontexte,
- ihre Ausrichtung auf die Herstellung von Sachbezügen, sowie die Übernahme entsprechender Ausdrucksmittel zur Darstellung abstrakter Beziehungen und generalisierter Aussagen und
- ihre hohe Fehlertoleranz (vgl. Lehmann 1981).

Desweiteren ist die Eindämmung sprachlicher Vagheit notwendig. Diese entsteht beispielsweise in Folge von individuellem lexikalisch-semanticem Wissen einzelner Personen. Die persönliche Auffassung und Bedeutung bestimmter Wörter ist sehr stark von den unterschiedlichen Erfahrungen der Personen abhängig. Ein einzelnes, spezielles Wort kann also verschiedene Bedeutungen für unterschiedliche Personen besitzen. Ein weiteres Problem stellt die zeitliche Sprachveränderung dar. Die Bedeutung der Worte bleibt nicht immerzu konstant, sondern kann sich im Zeitverlauf verändern und neue Ausprägungen annehmen (vgl. Burkert 1996, S.198).

Rechtschreibfehler und grammatikalisch falsch formulierte Sätze führen ebenfalls zu Problemen, die es zu beseitigen gilt. Eine Abhilfe schaffen KI-basierte Korrekturprogramme. Dabei findet eine Unterscheidung zwischen der Korrektur von „Nicht-Wörtern“, kontextabhängiger Korrektur sowie grammatikalischer Berichtigung statt. Bei ersteren handelt es sich um Wörter, die dem System unbekannt sind. Hierfür können dem Nutzer Korrekturen angeboten werden, welche er manuell auswählt und die im Programm hinterlegt werden. Somit erlernt dieses schrittweise, die Korrekturen automatisch durchzuführen. Eine kontextabhängige Korrektur ist dann notwendig, wenn durch Eingabefehler statt einem „Nicht-Wort“ ein gültiges Wort entsteht. Dieses kann beispielsweise durch n-Gramm-Wahrscheinlichkeiten aufgedeckt werden, also der Wahrscheinlichkeit, dass jeweils n benachbarte Wörter gemeinsam auftreten (vgl. Fliedner 2001, S.411 ff. und Lewandowski 2005, S.113 ff.). Grammatikalische Fehler können unter anderem durch zuvor festgelegte Bedingungen aufgefunden werden. Wenn auf einen Satz keine der Bedingungen angewandt werden kann, liegt ein grammatikalischer Fehler vor. Ebenso können auch Fehlermuster erstellt werden. Wenn ein Satz beim Abgleich einem Muster entspricht, liegt ein Fehler vor (vgl. Fliedner 2001, S.411 ff.).

Die Computerlinguistik⁵⁶ beinhaltet verschiedene Ansätze zur computerunterstützten Kodierung der natürlichen Sprache. Diese umfassen die Morphologie, die Syntax und die Semantik der Sprache. Zudem verfügt die Computerlinguistik über eine Reihe von Ressourcen, welche die Textkodierung ermöglichen und infolgedessen durch Opinion Mining Tools unterstützt werden sollten. Dazu zählen unter anderem Textkorpora, Baumbanken und lexikalisch-semantische Wortnetze. Textkorpora sind linguistisch aufbereitete Texte in geschriebener oder gesprochener Sprache, die elektronisch gespeichert vorliegen. Ihre Aufbereitung ist bisher nicht vollautomatisch ausführbar, da es bislang keine allgemein gültige Konvention für das Kodieren von Dokumenten gibt. Sie besteht aus der Tokenisierung (Zerlegung des Satzes in definierte Einheiten) und dem POS-Tagging (Bestimmung der Wortarten) (vgl. Evert und Fitschen 2001, S.369 ff.). Baumbanken stellen eine spezielle Form von Textkorpora dar. Sie beinhalten eine Sammlung von Einheiten (meist Sätzen), deren syntaktische Satzstruktur annotiert ist. Wie der Begriff schon verrät ist die Satzstruktur meist in Form einer Baumstruktur kodiert (vgl. Lezius 2001, S.377 ff.). Lexikalisch-semantische Wortnetze bilden die am häufigsten auf-

⁵⁶ Weiterführende Literatur zur Computerlinguistik und automatischen Spracherkennung: Ruske 1994, Schukart-Talamazzini 1995, Jurafsky und Martin 2000

tretenden und somit wichtigsten Wörter einer Sprache und ihre Beziehungen zu anderen Wörtern der Sprache ab (vgl. Kunze 2001, S.386 ff.). Sie zeigen beispielsweise Synonyme und Antonyme sowie Hyperonyme und Hyponyme⁵⁷ auf.

Im Bereich textueller Datenbanken im XML-Format (bspw. aus dem WWW) kann die Auswertung und Aufbereitung der linguistischen Datenbestände auf vier Ebenen erfolgen (Ide und Véronis 1994):

- bibliografische Erfassung, verwendete Zeichensätze aufführen, usw.
- Auszeichnung textueller Einheiten wie Band, Kapitel, Abschnitt, Fußnote, usw.
- Strukturen innerhalb von Abschnitten werden markiert, zum Beispiel Sätze, Wörter, Abkürzungen, Eigennamen, usw.
- Markierung syntaktischer Kategorien oder morphologischer Einheiten

Ziel dieser Auswertung und Aufbereitung ist das Semantic Web⁵⁸, welches eine umfangreiche Strukturierung sowie effiziente und robuste maschinelle Auswertung von Dokumenten zulässt.

3.3.2 Implikationen für die Praxis

Nachdem auf Basis der Ergebnisse Empfehlungen für die weitere Forschung ausgesprochen wurden, folgen abschließend einige Implikationen für die Praxis. Diese können einerseits für das Unternehmen Apple als Entwickler des iPhone ausgesprochen werden. Andererseits ergeben sich aber auch Empfehlungen für die Verkäufer des iPhone, also in Deutschland die Telekom mit ihrem Tochterunternehmen T-Mobile. Für die weitere Entwicklung des iPhone sind aus Sicht der Kunden vor allem Verbesserungspotenziale bezüglich des Akkus, der Kamera, iTunes und der SMS vorhanden. Der Akku sollte beispielsweise eine längere Laufzeit aufweisen und auch vom Käufer auswechselbar sein. Desweiteren scheinen die Bilder und Videos der Kamera nicht zufriedenstellend zu sein. An iTunes wird bemängelt, dass hierüber die Freischaltung und Synchronisation mit dem Computer stattfinden muss und iPhone-Nutzer somit gezwungen sind, iTunes ebenfalls zu verwenden. Bezüglich der SMS wird oft kritisiert, dass sie nicht einzeln zu löschen sind, sondern direkt alle SMS eines Kontakts gelöscht werden

⁵⁷ Ober- und Unterbegriffe

⁵⁸ Definition Semantic Web: „a web of data that can be processed directly or indirectly by machines [...] We will solve large analytical problems by turning computer power loose on the hard data of Semantic Web.“ (s. Berners-Lee 1999, S.177 ff.).

müssen. Ein weiterer Mangel scheint zu sein, dass es keine Zähleranzeige bezüglich der schon geschriebenen Zeichen beim Verfassen einer SMS gibt. Somit weiß der Nutzer nicht, wann die 160 Zeichen für eine SMS überschritten sind. Es kann jedoch davon ausgegangen werden, dass bestimmte von den Kunden bemängelte Funktionen zu Apples Strategie zählen. Dazu gehört zum Beispiel der Akku, welcher ausschließlich von Apple selbst ausgetauscht werden kann. Desweiteren wird Apple seinen iTunes-Kundenstamm durch die iPhone-Nutzer erweitern und diese an ihr Multimediaprodukt binden wollen, indem die Freischaltung und Synchronisation nur hierüber möglich ist. Ein zufriedenstellendes Ergebnis für alle iPhone-Kunden scheint hier also auch in Zukunft nicht erreichbar zu sein.

Neben den genannten Produkteigenschaften stößt zudem der T-Mobile Exklusivvertrag bei vielen Kunden auf Ablehnung. Vor allem die Wechselbereitschaft von Kunden, die zuvor mit anderen Mobilfunkanbietern zufrieden waren, ist sehr gering. Hier kann Apple vor allem dadurch profitieren, dass das iPhone als Kult-Handy ein sehr hohes Ansehen in der Bevölkerung besitzt und nur dadurch eine Wechselbereitschaft zu T-Mobile entsteht. Jedoch ist dieser Grund nicht für alle potenziellen Kunden ausreichend. Bei einigen ist die Loyalität zu ihrem bisherigen Anbieter so beständig, dass sie auf das iPhone verzichten und sich für ein anderes Smartphone entscheiden. Hier ist es sowohl an Apple als auch an T-Mobile gelegen, diese Kunden von sich zu überzeugen. Apple könnte dies tun, indem es den Exklusivvertrag mit T-Mobile nicht verlängert und somit allen anderen Mobilfunkanbietern den Weg zum Verkauf des iPhone ebnet (vgl. Maier 2009). Nachteilig für Apple wären dann allerdings die sinkenden Preise des iPhone aufgrund des steigenden Wettbewerbs der Händler. Falls der T-Mobile Exklusivvertrag verlängert werden sollte, müsste vor allem T-Mobile daran interessiert sein, die potenziellen Kunden, welche sich noch bei anderen Mobilfunkanbietern befinden, für sich zu gewinnen. Dies wäre sehr wahrscheinlich nur über vertragliche Anpassungen und Veränderungen möglich, um die Verträge denen der Konkurrenz anzugleichen. Es ist jedoch fraglich, ob eine solche Lösung im Sinne von T-Mobile ist und ob diese Veränderungen wirklich lohnenswert sind. Das ist vor allem davon abhängig, wie viele potenzielle Kunden zu T-Mobile übergehen und ob sie – im Verhältnis zum Aufwand für die Umstrukturierung – genügend Mehrwert für das Unternehmen schaffen würden.

4 Zusammenfassung und Ausblick

Hinsichtlich der immensen Bedeutung von Kundenmeinungen für die Unternehmens- und Produktreputation ist es in der heutigen Zeit des Web 2.0 unabdingbar, das Internet auf Rezensionen zu monitoren. Ziel dieser Arbeit war die Entwicklung einer Vorgehensweise, um mit Hilfe von Opinion Mining die im Internet befindlichen Textinhalte auf Meinungen zu überprüfen und deren semantische Orientierung zu bestimmen. Dazu wurden zu Beginn in Kapitel 2.1 die wichtigsten Grundlagen des Web 2.0 und seiner neuartigen, interaktiven Kommunikationsformen (Weblog, Virtuelle Community, Online Forum, Bewertungsportal) vorgestellt. Desweiteren erfolgte in Kapitel 2.2 eine Erläuterung des theoretischen Basiswissens über die Business Intelligence und eine Abgrenzung ihrer Methoden Data Mining, Text Mining und Opinion Mining. Nach Abschluss des Grundlagenkapitels konnte die Entwicklung des Vorgehensmodells für das Opinion Mining im anschließenden Kapitel 3.1 erfolgen, welches sich an den fünf Phasen des CRISP-DM Modells orientiert. Kapitel 3.2 beinhaltet die praktische Anwendung und Auswertung des Modells. Innerhalb dieses Kapitels konnten drei der zu Beginn der Arbeit genannten Forschungsfragen beantwortet werden. Eine Antwort auf die vierte Forschungsfrage folgte in Kapitel 3.3, welches Implikationen für die Forschung und Praxis beinhaltet.

- Forschungsfrage 1:

Wie erfolgt die Erfassung der benötigten Kundenmeinungen aus dem Internet?

Die erste Forschungsfrage wurde in der Phase des Data Understanding in Kapitel 3.2.2 beantwortet. Hier erfolgte die Erfassung der Kundenmeinungen per Webcrawling und die anschließende Erstellung der Review Datenbank. Es wurden verschiedene Methoden zur automatischen und manuellen Datenerfassung aus dem Internet vorgestellt. Nach einer Abwägung der Methoden fiel die Entscheidung letztendlich auf die Durchführung eines manuellen „Webcrawlings“ durch den Anwender, der die Webseiten manuell auswählt und bearbeitet.

Zum Ende des Kapitels 3.2.2.2 wurde eine Einschränkung der Analyse auf die Kundenrezensionen von Amazon und Ciao vorgenommen, da sie den größten Teil der gesammelten Kundenmeinungen darstellten. Desweiteren ergab sich bezüglich der Blogs, Communities und Foren das Problem, dass die einzelnen Einträge meist zugehörige Kommentare besitzen, welche oftmals nur in Zusammenhang mit der vorherigen Diskussion verständlich sind. Diese Zusammenhänge

sind jedoch nur begrenzt oder nicht maschinell erfassbar und stellten somit ein technisches Problem dar.

- Forschungsfrage 2:

Wie lassen sich die Kundenmeinungen per Opinion Mining bewerten?

Das Vorgehen zur Bewertung der Kundenmeinungen wurde in der Data Preparation in Kapitel 3.2.3 vorgestellt. Voraussetzung dafür war das Auffinden der Frequent Features und Opinion Words des untersuchten Produkts. Dazu wurden die Sätze aus der Review-Datenbank in ihre Worte gesplittet. Anschließend hätte das POS-Tagging erfolgen sollen. Da dies im RapidMiner jedoch nicht möglich war, wurden die Frequent Features und Opinion Words anhand ihrer Auftrittshäufigkeit manuell gewählt. Häufig genannte Substantive wurden den Frequent Features und die Adjektive den Opinion Words zugeordnet. Desweiteren erfolgte die semantische Bestimmung der Opinion Words. Mithilfe von Satzanalysen konnte abschließend mit einer Genauigkeit von etwa 50% eine Bewertung der Frequent Features durch ihre zugehörigen Opinion Words durchgeführt werden.

- Forschungsfrage 3:

Wie gut unterstützt die ausgewählte Software das Opinion Mining?

Im Rahmen von Evaluation und Deployment in Kapitel 3.2.5 erfolgte die Bewertung der Opinion Mining Vorgehensweise und des angewandten Tools. Grundsätzlich besteht die Möglichkeit bis zu einem gewissen Grad Opinion Mining mit dem RapidMiner durchzuführen. Jedoch bedarf er einer umfangreichen weiterführenden Entwicklung, damit wichtige fehlende Funktionen und Vorgehensweisen ermöglicht werden. Am wesentlichsten fiel in dieser Arbeit das Fehlen eines POS-Taggers auf, der zum Auffinden der Frequent Features und Opinion Words notwendig gewesen wäre. Stattdessen mussten diese manuell durch einfaches Zählen der vorkommenden Wörter erfasst werden.

- Forschungsfrage 4:

Welche Implikationen ergeben sich aus dem Ergebnis?

Kapitel 3.3 zeigt verschiedene Implikationen für die Forschung und Praxis im Bereich Opinion Mining auf. Weiterer Forschungs- und Entwicklungsbedarf ist vor allem im Bereich der Opinion Mining Tools und somit auch in der künstlichen Intel-

ligenz notwendig. Fehlende Funktionen sind beispielweise das Verständnis der natürlichen Sprache mit ihren charakteristischen Merkmalen, die Eindämmung sprachlicher Vagheit oder auch das automatische Erkennen und Beheben von Rechtschreibfehlern und grammatikalisch falsch formulierten Sätzen.

Praktisch ergeben sich Implikationen sowohl für den Entwickler Apple als auch für den Händler T-Mobile. Durch die Rezensenten angesprochene Produktmängel bieten Verbesserungspotenziale für die weitere Entwicklung und den Verkauf des iPhone. Durch die Behebung der angesprochenen Mängel und Schwächen können die Kundenloyalität gesteigert und potenziellen Kunden akquiriert werden.

Zum Abschluss der Arbeit erfolgt eine Auflistung der Chancen und Grenzen des Opinion Mining im Web 2.0 (vgl. Früh 2007, S.41 f. und Zerfaß et al. 2008, S.251 ff.), welche sich aus dieser Arbeit ergeben. Das Opinion Mining umfasst folgende Chancen für die Forschung und Praxis:

- Die Forscher sind nicht auf die Kooperation von Versuchspersonen angewiesen.
- Web 2.0 Inhalte sind unbeeinflusst von Forschern, die beispielsweise als Interviewer oder Beobachter auftreten.
- Darüber hinaus stehen die Inhalte jedem jederzeit in hohen Fallzahlen zur Verfügung.
- Meinungen und die Entstehung von Trends werden mit Hilfe von Web 2.0 Anwendungen erkennbar.
- Es wird möglich, interessante Informationen über Produkte und das Unternehmen zu gewinnen.
- Zudem können Erkenntnisse über die Beliebtheit oder das Image eines Produkts, einer Marke oder eines Unternehmens erworben werden.
- Ebenso ist eine Wettbewerbsanalyse möglich.
- Anhand der Durchführung eines kontinuierlichen Markenmonitoring ist die Beobachtung der Meinungsentwicklung über einen bestimmten Zeitraum hinweg denkbar.
- Netnografie⁵⁹ im Social Web ermöglicht herauszufinden, welche Einstellung die Webuser gegenüber Produkten, Marken und Unternehmen haben.

⁵⁹ Netnografie = Zusammenspiel von Ethnografie und Internet

Die Übertragung des ethnografischen Forschungsansatzes auf Online-Communities oder die Blogosphäre.

- Darüber hinaus können zukünftige Trends und Bedürfnisse frühzeitig erkannt werden, die wertvolle Informationen für Produktinnovationen und -modifikationen bieten.
- Durch die Identifikation und gezielte Ansprache von Meinungsführern können Unternehmen oftmals von deren starker Integration und meinungsverstärkender Funktion profitieren (virales Marketing).

Neben den Chancen zeigen sich auch einige Grenzen des Opinion Mining, die nachfolgend aufgeführt sind:

- Die auf dem Markt vorhandenen Opinion Mining Tools befinden sich oftmals noch in der Entwicklung und verfügen somit nicht über alle notwendigen Funktionen und Anwendungsmöglichkeiten.
- Bestimmte erforderliche technische Lösungen – vor allem im Bereich der Computerlinguistik und automatischen Sprachverarbeitung – sind bisher praktisch nicht umgesetzt.
- Die Wahrung der rechtlichen und ethischen Grundsätze (zum Beispiel des Robots Exclusion Protokolls) erfordert einen sensiblen Umgang bei den Forschungsaktivitäten.
- Die Validität ist oftmals aufgrund der fehlenden technischen Reife nicht gewährleistet.
- Auch die Repräsentativität ist aufgrund der Self-Selection nicht immer gesichert, da die Nutzer des Web 2.0 nur einen begrenzten Teil der Gesamtbevölkerung darstellen. Längst nicht alle Nutzer des jeweiligen untersuchten Gegenstandes veröffentlichen ihre Meinung in schriftlicher Form, so dass sie per Opinion Mining analysiert werden könnte.
- Desweiteren stellen die Reputation Evangelists und Terrorists ein Problem dar, da sie bewusst gefälschte Aussagen im Internet verbreiten, um ihre persönlichen Ziele zu verfolgen.

Zusammenfassend ist anzumerken, dass das Opinion Mining eine interessante und für die Zukunft sicherlich sehr wichtige Forschungsmethode darstellt. Vor allem im Bereich der Online Meinungsforschung zeigt diese Methode eine Vielzahl an Möglichkeiten auf. Aufgrund des angesprochenen Problems der Repräsentativität von Internetinhalten wird jedoch empfohlen, weitere Daten aus der Markt- und Sozialforschung heranzuziehen, die aus anderweitigen Informationsquellen stammen. Abschließend kann dem Opinion Mining ein großes Forschungspotenzial

zugesprochen werden. Für die Zukunft sind sicherlich noch viele weitere interessante und aufschlussreiche Ergebnisse aus dem Gebiet der Meinungsforschung zu erwarten.

Literaturverzeichnis

- Adriaans, P.; Zantinge, D. (1998): Data Mining, Addison-Wesley Professional.
- Agrawal, R.; Imielinski, T.; Swami, A. (1993): Mining Association Rules between Sets of Items in Large Databases, in: ACM SIGMOD Conference, Washington DC, USA, Mai 1993, URL: <http://eprints.kfupm.edu.sa/50864/1/50864.pdf>, Abruf am 28.09.2009.
- Agrawal, R.; Srikant, R. (1994): Fast Algorithm for Mining and Association Rules, in: VLDB'94, URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.7506>, Abruf am 23.09.2009.
- Berelson, B.; Lazarsfeld, P. (1948): The Analysis of Communication Content, University of Illinois, Chicago.
- Berelson, B.; Lazarsfeld, P. (1952): Content Analysis, in Communication Research, New York.
- Berners-Lee, T. (1999): Weaving the Web – The original Design and Ultimate Destiny of the World Wide Web by its Inventor, Harper San Francisco.
- Bönisch, B. (2008): Code-Switching in Internetforen am Beispiel von „Polen im Pott“, Grin Verlag.
- Bramer, M. (2007): Principles of Data Mining, 1. Auflage, Springer Verlag, Berlin.
- Brewer, C. (2000): Deepening Brand Loyalty, in: ComputerUser, Oktober 2000, URL: <http://www.computeruser.com/articles/1910,2,3,1,1001,00.html>, Abruf am 01.07.2009.
- Brill, E. (1993): A simple rule-based Part of Speech Tagger, Department of Computer Science, University of Pennsylvania, URL: <http://ucrel.lancs.ac.uk/acl/H/H92/H92-1022.pdf>, Abruf am 22.09.2009.
- Bruce, R.; Wiebe, J. (2000): Recognizing Subjectivity: A Case Study of Manual Tagging, in: Natural Language Engineering 1 (1), S. 1–16, URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.8785>, Abruf am 23.09.2009.
- Burkert, G. (1996): Repräsentation von lexikalisch-semantischem Wissen in einem System zur Verarbeitung natürlicher Sprache, Infix Verlag.
- Carstensen, K.; Ebert, C.; Endriss, C.; Jekat, S.; Klabunde, R.; Langer, H. (2001): Computerlinguistik und Sprachtechnologie – Eine Einführung, Spektrum Akademischer Verlag, Heidelberg.
- Case Study “Dell-Hell” (2007): Responding to Crisis Using Social Media, in: Market Sentinel, London, Mai 2007, URL: <http://www.marketsentinel.com/files/Crisisresponseusingsocialmedia.pdf>, Abruf am 01.07.2009.
- Conrad, J. G.; Schilder, F. (2007): Opinion Mining in Legal Blogs, in: ICAIL '07: Proceedings of the 11th international conference on Artificial intelligence and law, ACM, New York, S. 231–236, URL: <http://delivery.acm.org/10.1145/1280000/1276363/p231-conrad.pdf>, Abruf am 28.05.2009.
- CRISP-DM Model: Cross Industry Standard Process for Data Mining, Process Model, URL: <http://www.crisp-dm.org/Process/index.htm>, Abruf am 02.07.2009.
- Ding, X.; Liu, B.; Yu, P. S. (2008): A Holistic Lexicon-Based Approach to Opinion Mining, in: WSDM '08: Proceedings of the international conference on Web search and web data mining, ACM, New York, USA, S. 231–239, URL: <http://delivery.acm.org/10.1145/1350000/1341561/p231-ding.pdf>, Abruf am 28.05.2009.

- Evert, S.; Fitschen, A. (2001): Textkorpora, in: Carstensen et al. (Hrsg.): Computerlinguistik und Sprachtechnologie – Eine Einführung, S. 369–376, Spektrum Akademischer Verlag, Heidelberg.
- Fayyad, U. M.; Piatetsky-Shapiro G.; Smyth, P. (1996): From Data Mining to Knowledge Discovery: An Overview, in: Association for the Advancement of Artificial Intelligence, 1997, Juli 27–31, S. 37–54, URL: <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>, Abruf am 06.07.2009.
- Feldman, R.; Dagan, I. (1995): Knowledge Discovery in Textual Databases (KDT), in: Proceedings of the 1st international Conference on Knowledge Discovery (KDD-95), URL: <http://www.aaai.org/Papers/KDD/1995/KDD95-012.pdf>, Abruf am 06.07.2009.
- Fliedner, G. (2001): Korrekturprogramme, in: Carstensen et al. (Hrsg.): Computerlinguistik und Sprachtechnologie – Eine Einführung, S. 411–417, Spektrum Akademischer Verlag, Heidelberg.
- Frawley, W. J.; Piatetsky-Shapiro G.; Matheus C. J. (1991): Knowledge Discovery in Databases: An Overview, in: AI Magazine (AAAI), Vol.13, No.2, 1992, URL: <http://www.aaai.org/ojs/index.php/aimagazine/article/view/1011/929>, Abruf am 06.07.2009.
- Früh, W. (2007): Inhaltsanalyse. Theorie und Praxis, 6. Auflage, UVK Verlag.
- Funk, A.; Li, Y.; Saggion, H.; Bontcheva, K.; Leibold, C. (2008): Opinion Analysis for Business Intelligence Applications, in: OBI'08, October, Karlsruhe, Germany, URL: <http://delivery.acm.org/10.1145/1460000/1452570/a3-funk.pdf>, Abruf am 28.05.2009.
- Gabriel, R.; Guchowski, P.; Pastwa, A. (2009): Data Warehouse und Data Mining, 1. Auflage, W3I Verlag.
- Gerbner, G; Holsti, O.; Krippendorff, K.; Pasiley, W.; Stone, P. (Hrsg.) (1969): The Analysis of Communication Content, Wiley Verlag.
- Gluchowski, P. (2001): Business Intelligence. Konzepte, Technologien und Einsatzbereiche, in: HMD – Praxis der Wirtschaftsinformatik, 38 Jg., Heft 222, Dezember 2001, S. 5–15.
- Hauser, T.; Wenz, C.; Maurice, F. (2007): Das Website Handbuch, 2. Auflage, Markt und Technik Verlag.
- Heitman, M.; Prycop, C.; Aschmoneit, P. (2004): Using Means-end Chains to Build Mobile Brand Communities, in: Proceedings of the 37th Hawaii International Conference on System Sciences, 2004, URL: <http://www2.computer.org/plugins/dl/pdf/proceedings/hicss/2004/2056/07/205670196c.pdf>, Abruf am 10.07.2009.
- Hipp, J.; Güntzer, U.; Nakhaeizadeh, G. (2000): Algorithms for Association Rule Mining – A General Survey and Comparison, in: ACM SIGKDD, Juli 2000, Vol. 2, No. 1, S. 58–64, URL: <http://arbor.ee.ntu.edu.tw/~chyun/dmpaper/jochaf00.pdf>, Abruf am 28.09.2009.
- Högg, R.; Martignoni, R.; Meckel, M.; Stanoevska-Slabeva, K. (2006): Overview of Business Models for Web 2.0 Communities, Alexandria, Hochschule St. Gallen, URL: <http://www.alexandria.unisg.ch/EXPORT/DL/31412.pdf>, Abruf am 07.07.2009.
- Holsti, O. (1969): Content Analysis for the Social Sciences and Humanities, Addison-Wesley Verlag.
- Hu, M.; Liu, B. [1] (2004): Mining Opinion Features in Customer Reviews, in: AAAI (American Association for Artificial Intelligence) 2004, URL: <http://www.cs.pitt.edu/~wiebe/courses/CS3730/Fall08/hu2004a.pdf>, Abruf am 16.09.2009.

- Hu, M.; Liu, B. [2] (2004): Mining and Summarizing Customer Reviews, in: KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, NY, USA, S. 168–177, URL: <http://delivery.acm.org/10.1145/1020000/1014073/p168-hu.pdf>, Abruf am 28.05.2009.
- Ide, N.; Véronis, J. (1994): MULTEXT: Multilingual Text Tools and Corpora, in: COLING'94 – The 15th International Conference on Computational Linguistics, Band 1, S. 588–592, URL: <http://www.cs.vassar.edu/~ide/papers/MULTEXT.pdf>, Abruf am 30.11.2009.
- Jurafsky, D.; Martin, J. (2000): Speech and Language Processing – An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall, New Jersey.
- Kaiser, R. (2008): Bibliotheken im Web 2.0 Zeitalter, Dinges & Frick Verlag.
- Kemper, H. G.; Mehanna, W.; Unger, C. (2006): Business Intelligence - Grundlagen und praktische Anwendungen: Eine Einführung in die IT-basierte Managementunterstützung, 2. Auflage, Vieweg+Teubner Verlag.
- Kneip, M. (2008): Data Mining, Grin Verlag.
- Kunze, C. (2001): Lexikalisch-semantische Wortnetze, in: Carstensen et al. (Hrsg.): Computerlinguistik und Sprachtechnologie – Eine Einführung, S. 386–393, Spektrum Akademischer Verlag, Heidelberg.
- Krippendorff, K. (1980): Content Analysis. An Introduction to Its Methodology, Sage Verlag.
- Krippendorff, K. (2004): Content Analysis. An Introduction to Its Methodology, 2. Auflage, Sage Verlag.
- Langner, S. (2007): Viral Marketing, 2. Auflage, Gabler Verlag.
- Lasswell, H. (1948): The Structure and Function of Communication in Society, in: Lyman Brison (Hrsg.): The communication of ideas, New York, S. 37–52.
- Lee, D.; Jeong, O.-R.; Lee, S.-g. (2008): Opinion Mining of Customer Feedback Data on the Web, in: ICUIMC '08: Proceedings of the 2nd international conference on Ubiquitous information management and communication, ACM, New York, USA, S. 230–235, URL: <http://delivery.acm.org/10.1145/1360000/1352842/p230-lee.pdf>, Abruf am 28.05.2009.
- Lehmann, E. (1981): Computersimulation des Verstehens natürlicher Sprache, in: Nova acta Leopoldina, 54 (254), S. 125–174.
- Lewandowski, D. (2005): Web Information Retrieval: Technologien zur Informationssuche im Internet, Dinges&Frick Verlag, Wiesbaden.
- Lezius, W. (2001): Baubanken, in: Carstensen et al. (Hrsg.): Computerlinguistik und Sprachtechnologie – Eine Einführung, S. 377–385, Spektrum Akademischer Verlag, Heidelberg.
- Liu, B.: Opinion Mining, URL: <http://www.cs.uic.edu/~liub/FBS/opinion-mining.pdf>, Abruf am 03.07.2009.
- Liu, B. (2008): Web Data Mining. Exploring Hyperlinks, Contents, and Usage Data, 2. Auflage, Springer Verlag.
- Liu, B.; Hsu, W.; Ma, Y. (1998): Integrating Classification and Association Rule Mining, KDD'98, URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.7506>, Abruf am 23.09.2009.
- Loshin, D. (2003): Business Intelligence. The Savvy Manager's Guide, Morgan Kaufman Publ. Inc. Verlag.
- Maier, A. (2009): i-Phone-Monopol vor dem Aus, in: Manager Magazin, 24.06.2009, URL: <http://www.manager-magazin.de/it/artikel/0,2828,632181,00.html>, Abruf am 02.12.2009.

- Mayring, P. (2000): Qualitative Inhaltsanalyse, in: FORUM Qualitative Sozialforschung, Vol. 1, No. 2, Art. 20.
- McCue, C. (2006): Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis, Butterworth Heinemann Verlag.
- Meckel, M. (2006): Was vom Tage übrig bleibt, Miriam Meckel Blog, 29. Nov. 2006, URL: <http://www.miriammeckel.de/2006/11/29/was-vom-tage-uebrig-bleibt/>, Abruf am 21.07.2009.
- Meckel, M. (2007): Das Glück der Unerreichbarkeit. Wege aus der Kommunikationsfalle, Murmann Verlag.
- Meckel, M. [2] (2008): Reputationsevangelisten und Reputationsterroristen. Unternehmenskommunikation 2.0, in: Meckel, M.; Stanoevska-Slabeva, K. (Hrsg.): Web 2.0. Die nächste Generation Internet, Nomos Verlag, S. 109–128.
- Menczer, F. (2008): Webcrawling, in: Liu, B. (2008): Web Data Mining. Exploring Hyperlinks, Contents, and Usage Data, 2. Auflage, Springer Verlag, S. 273–321.
- Mitra, S.; Acharya, T. (2003): Data Mining: Multimedia, Soft Computing, and Bioinformatics, Wiley & Sons Verlag.
- Morinaga, S.; Yamanishi, K.; Tateishi, K.; Fukushima, T. (2002): Mining Product Reputations on the Web, in: KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, USA, S. 341–349, URL: <http://delivery.acm.org/10.1145/780000/775098/p341-morinaga.pdf>, Abruf am 28.05.2009.
- Nasukawa, T.; Yi, J. (2003): Sentiment Analysis: Capturing Favorability Using Natural Language Processing, in: K-CAP '03: Proceedings of the 2nd international conference on Knowledge capture, ACM, New York, USA, S. 70–77, URL: <http://delivery.acm.org/10.1145/950000/945658/p70-nasukawa.pdf>, Abruf am 28.05.2009.
- Olsen, D. L.; Delen, D. (2008): Advanced Data Mining Techniques, Springer Verlag.
- O'Reilly (2005): What is Web 2.0?, September 2005, URL: <http://www.oreilly.de/artikel/web20.html>, Abruf am 03.07.2009.
- Pant, G.; Srinivasan, P.; Menczer, F. (2003): Crawling the Web, URL: <http://dollar.biz.uiowa.edu/~pant/Papers/crawling.pdf>, Abruf am 08.09.2009.
- Petersohn, H. (2005): Data Mining: Verfahren, Prozesse, Anwendungsarchitektur, Oldenbourg Verlag.
- Pressebox (2009): Gewinner des Open Source Business Award 2008 – Rapid-I – auch im Jahr 2008 wieder auf dem Open Source Meets Business Kongress vertreten, Pressebox, 22.01.2009, URL: <http://www.pressebox.de/pressemeldungen/rapid-i-gmbh/boxid-232325.html>, Abruf am 03.06.2009.
- Rand, P.: Understanding and Managing Negative Word of Mouth, Ketchum, URL: http://www.ketchum.com/paul_rand_managing_negative_word_of_mouth_article, Abruf am 18.11.2009.
- Rauch, J.; Šimůnek, M. (2002): Alternative Approach to Mining Association Rules, in: FDM 2002, The Foundation of Data Mining and Knowledge Discovery, The Proceedings of the Workshop of ICDM02, S.157–162, URL: http://users.info.unicaen.fr/~bruno/asdisco/journees/doc/Rauch290304Presentation/ICDM02_TFDM_publ.pdf, Abruf am 28.09.2009.
- Rheingold, H. (1993): The Virtual Community, URL: <http://www.rheingold.com/vc/book/intro.html>, Abruf am 10.07.2009.
- Ruske, G. (1994): Automatische Spracherkennung: Methoden der Klassifikation und Merkmalsextraktion, 2. Auflage, Oldenbourg Verlag, München.

- Rutenbeck (2006): Tech Terms: What Every Telecommunications and Digital Media Professional Should Know, Butterworth Heinemann Verlag.
- Santorini, B. (1990): Part-of-Speech Tagging Guidelines for the Penn Treebank Retrieval, Technical Report MS-CIS-90-47, Department of Computer and Information Science, University of Pennsylvania, URL: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1603&context=cis_reports, Abruf am 22.09.2009.
- Schaarschmidt, R.; Herrmann, U. (2002): Daten lügen nicht – oder doch? Vorgehensweise zur Verbesserung der Datenqualität für Business Intelligence, in: Praxis der Wirtschaftsinformatik (HMD), Heft 226, S. 110–116.
- Schenk, M.; Taddicken, M.; Welker, M. (2008): Web 2.0 als Chance für die Markt- und Sozialforschung?, in: Zerfaß, A.; Welker, M.; Schmidt, J. (2008): Kommunikation, Partizipation und Wirkungen im Social Web, Band 1: Grundlagen und Methoden: Von der Gesellschaft zum Individuum, Herbert von Halem Verlag, Köln.
- Schukat-Talamazzini, E. (1995): Automatische Spracherkennung: Grundlagen, statistische Modelle und effiziente Algorithmen, Vieweg Verlag, Braunschweig/Wiesbaden.
- Stanoevska-Slabeva, K. (2008): Web 2.0 – Grundlagen, Auswirkungen und zukünftige Trends, in: Meckel, M.; Stanoevska-Slabeva, K. (Hrsg.): Web 2.0. Die nächste Generation Internet, Nomos Verlag, S. 13–38.
- Sun, Y.; Zhuang, Z.; Giles, C. L.; (2007): A Large Scale-Study of Robots.txt, in: International World Wide Web Conference 2007, Mai 8–12, S.1123–1124, URL: <http://delivery.acm.org/10.1145/1250000/1242726/p1123-sun.pdf>, Abruf am 02.09.2009.
- Telekompresse (2007): iPhone: Exklusivrecht bleibt bei T-Mobile, 04.11.2007, URL: http://www.telekom-presse.at/channel_mobile/news_31205.html, Abruf am 20.08.2009.
- Wiebe, J.; Bruce, R.; O'Hara, T. (1999): Development and Use of a Gold-Standard Data Set for Subjectivity Classifications, in: Proc. of ACL'99, URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.7584>, Abruf am 23.09.2009.
- Zerfaß, A.; Bölter, D. (2005): Die neuen Meinungsmacher. Weblogs als Herausforderung für Kampagnen, Marketing, PR und Medien, Nausner & Nausner Verlag.
- Zerfaß, A.; Bogosyan J. (2007): Blogstudie 2007. Informationssuche im Internet – Blogs als neues Recherchetool, Institut für Kommunikations- und Medienwissenschaft, Universität Leipzig, URL: http://www.blogstudie2007.de/inc/blogstudie2007_ergebnisbericht.pdf, Abruf am 07.07.2009.
- Zerfaß, A.; Welker, M.; Schmidt, J. (2008): Kommunikation, Partizipation und Wirkungen im Social Web, Band 1: Grundlagen und Methoden: Von der Gesellschaft zum Individuum, Herbert von Halem Verlag, Köln.

Anhang

A. Folgende Links wurden im Rahmen des Data Understanding für die Review Datenbank ausgewählt (vgl. Kapitel 3.2.2.2):

Blogsuche:

bei Google: Abruf am 12.07.2009

<http://www.iphone-blog.ch/>

<http://www.iphone-fan.de/>

<http://www.phonereloaded.de/>

<http://www.iphone-blog.eu/>

<http://www.iphone-news.org/>

<http://www.iphone-benutzer.de/>

<http://myblog-iphone.de/>

<http://iphone-news-tipps.blogspot.com/>

<http://iphone.haus.dj/>

<http://www.iphone-tipper.de/>

<http://iphone4ever.eu/>

<http://blogpingr.de/blog/10024-myblog-iphone/>

<http://www.stefanhaab.de/iphoneblog>

<http://blog.thorsten-stark.de/>

<http://www.apfelnet.de/>

<http://dasiphone.wordpress.com/>

<http://www.stepsblog.at/>

<http://ei-phone.blogspot.com/>

<http://forum.iphone-blog.ch/>

<http://paul.uhlenbrock.over-blog.com/>

<http://www.theiphoneblog.com/>

<http://iphone.blogvasion.com/>

<http://www.iphone-online.ch/blog/>

<http://www.iphone-essentials.ch/>

<http://spegeliphone.wordpress.com/>

<http://www.mfx.ch/>

<http://iphone.volblog.at/>

<http://apfelblog.ch/>

<http://www.blog-iphone.de/>

bei Technorati: Abruf am 12.07.2009

<http://freshzweinull.de/tag/iphone/>

<http://www.iphone-notes.de/>

<http://www.iphoneblog.de/>

Communities:

bei Google: Abruf am 18.08.2009

<http://www.iphonecity.de/>

<http://iszene.com/>

<http://www.maccommunity.de/beitraege>

<http://www.sevenmac.de/community/forum.php?c=category&op=index&cid=20>

<http://board.macnotes.de/6-iphone-und-ipod-touch/>

<http://forum.modopo.com/apple-iphone/>

<http://www.kwick.de/forum/10/164890>

<http://www.macosxhints.ch/forums/forumdisplay.php?f=88>

<http://www.bsmparty.de/gruppe/1293/forum>

<http://www.usp-forum.de/apple-iphone-forum/>

Foren:

bei Google: Abruf am 18.08.2009

<http://www.das-iphone-forum.ch/>

<http://www.mobilfunk-talk.de/iphone-forum/>

<http://www.iphone-handy.de/forum/index.html>

<http://apfelportal.de/forum/>

<http://apfelblog.ch/forum/forum/iphone>

<http://www.iweb-forum.de/board/index.php?page=Board&boardID=17>

<http://www.hardwareluxx.de/community/forumdisplay.php?f=213>

<http://www.iphorum.de/forum/>

<http://www.apfelphone.net/forum/2g-iphone/>

<http://www.iphonehome.ch/index.php/forum.html>

<http://www.iphone-forum.eu/>

<http://www.pocketpc.ch/iphone/>

<http://www.connect.de/connect-Forum/apple-iphone/>

Amazon:

iPhone 3G: 64 Kundenrezensionen, Abruf am 26.08.2009

http://www.amazon.de/product-reviews/B001ANGMSA/ref=dp_top_cm_cr_acr_txt?ie=UTF8&showViewpoints=1

iPhone 3GS: 12 Kundenrezensionen, Abruf am 26.08.2009

http://www.amazon.de/product-reviews/B002IPGOLE/ref=dp_top_cm_cr_acr_txt?ie=UTF8&showViewpoints=1

Ciao:

iPhone 3G: 92 Erfahrungsberichte, Abruf am 26.08.2009

http://www.ciao.de/Erfahrungsberichte/Apple_iPhone_3G_S_8182265

iPhone 3GS: 17 Erfahrungsberichte, Abruf am 26.08.2009

http://www.ciao.de/Erfahrungsberichte/Apple_iPhone_2502000

iPhone: 72 Erfahrungsberichte, Abruf am 26.08.2009

http://www.ciao.de/Erfahrungsberichte/Apple_iPhone_3G_7864901

B. Auflistung der Frequent Features aus der Data Preparation (vgl. Kapitel 3.2.3.3):

1 DISPLAY						
regular	bildschirm	real	avg = 0.006	[0.000 ; 2.000]		79
regular	bildschirmauflösung	real	avg = 0.000	[0.000 ; 1.000]		3
regular	bildschirme	real	avg = 0.000	[0.000 ; 1.000]		2
regular	bildschirmes	real	avg = 0.000	[0.000 ; 1.000]		1
regular	bildschirms	real	avg = 0.000	[0.000 ; 1.000]		4
regular	display	real	avg = 0.021	[0.000 ; 3.000]		258
regular	displays	real	avg = 0.002	[0.000 ; 1.000]		25
regular	handydisplay	real	avg = 0.000	[0.000 ; 1.000]		2
regular	multitouch	real	avg = 0.002	[0.000 ; 2.000]		20
regular	multitouchdisplay	real	avg = 0.000	[0.000 ; 1.000]		1
regular	multitouchscreen	real	avg = 0.000	[0.000 ; 1.000]		1
regular	touch	real	avg = 0.010	[0.000 ; 2.000]		121
regular	touchdisplay	real	avg = 0.001	[0.000 ; 1.000]		7
regular	touchdisplays	real	avg = 0.000	[0.000 ; 1.000]		1
regular	touchpad	real	avg = 0.000	[0.000 ; 1.000]		2
regular	touchscreen	real	avg = 0.007	[0.000 ; 2.000]		85
regular	touchscreendisplay	real	avg = 0.000	[0.000 ; 1.000]		1
regular	touchscreens	real	avg = 0.001	[0.000 ; 1.000]		13
SUMME						626
2 MUSIK						
regular	audiowiedergabe	real	avg = 0.001	[0.000 ; 1.000]		13
regular	audiowiedergabe	real	avg = 0.001	[0.000 ; 1.000]		13
regular	ipod	real	avg = 0.024	[0.000 ; 3.000]		290
regular	ipods	real	avg = 0.002	[0.000 ; 1.000]		29
regular	musik	real	avg = 0.017	[0.000 ; 2.000]		211
regular	musikabspielgerät	real	avg = 0.000	[0.000 ; 1.000]		2
regular	musikhören	real	avg = 0.001	[0.000 ; 1.000]		10
regular	musikplayer	real	avg = 0.000	[0.000 ; 1.000]		4
regular	musikplayers	real	avg = 0.000	[0.000 ; 1.000]		1
regular	musikwiedergabe	real	avg = 0.001	[0.000 ; 1.000]		11
SUMME						584
3 FUNKTIONALITAT						
regular	feature	real	avg = 0.001	[0.000 ; 1.000]		13
regular	features	real	avg = 0.003	[0.000 ; 1.000]		31
regular	funktion	real	avg = 0.016	[0.000 ; 2.000]		199
regular	funktionalitaet	real	avg = 0.000	[0.000 ; 1.000]		1
regular	funktionalität	real	avg = 0.002	[0.000 ; 1.000]		24
regular	funktionalitäten	real	avg = 0.000	[0.000 ; 1.000]		3
regular	funktionen	real	avg = 0.018	[0.000 ; 2.000]		218
regular	funktionsumfang	real	avg = 0.001	[0.000 ; 2.000]		18
regular	funktionsumfang	real	+/- 0.040	[0.000 ; 2.000]		18
SUMME						525
4 FOTO						
regular	bild	real	avg = 0.004	[0.000 ; 1.000]		48
regular	bilder	real	avg = 0.013	[0.000 ; 2.000]		164
regular	foto	real	avg = 0.003	[0.000 ; 1.000]		39
regular	fotofunktion	real	avg = 0.000	[0.000 ; 1.000]		2
regular	fotos	real	avg = 0.012	[0.000 ; 3.000]		142
regular	iphoto	real	avg = 0.001	[0.000 ; 1.000]		7
regular	photo	real	avg = 0.000	[0.000 ; 1.000]		2
regular	photofunktion	real	avg = 0.000	[0.000 ; 1.000]		1
regular	photos	real	avg = 0.001	[0.000 ; 2.000]		14
SUMME						419

	5 EMAIL					
regular	email	real	avg = 0.005	[0.000 ; 3.000]		62
regular	emailclient	real	avg = 0.000	[0.000 ; 1.000]		2
regular	emailfunktion	real	avg = 0.000	[0.000 ; 1.000]		2
regular	emailfunktionen	real	avg = 0.000	[0.000 ; 1.000]		1
regular	emails	real	avg = 0.004	[0.000 ; 1.000]		43
regular	emailservice	real	avg = 0.000	[0.000 ; 1.000]		1
regular	mail	real	avg = 0.016	[0.000 ; 4.000]		198
regular	mails	real	avg = 0.009	[0.000 ; 2.000]		107
				SUMME		416
	6 APPLICATION					
regular	anwendung	real	avg = 0.002	[0.000 ; 1.000]		24
regular	applicationen	real	avg = 0.000	[0.000 ; 1.000]		1
regular	appstore	real	avg = 0.006	[0.000 ; 3.000]		73
regular	appstores	real	avg = 0.000	[0.000 ; 1.000]		1
regular	gratisprogramme	real	avg = 0.000	[0.000 ; 1.000]		1
regular	programmen	real	avg = 0.003	[0.000 ; 2.000]		33
regular	programme	real	avg = 0.013	[0.000 ; 2.000]		160
regular	programmstore	real	avg = 0.000	[0.000 ; 1.000]		1
regular	spiel	real	avg = 0.001	[0.000 ; 2.000]		13
regular	spiele	real	avg = 0.006	[0.000 ; 1.000]		69
regular	zusatzapplikationen	real	avg = 0.000	[0.000 ; 1.000]		1
regular	zusatzprogramm	real	avg = 0.000	[0.000 ; 1.000]		1
regular	zusatzprogramme	real	avg = 0.000	[0.000 ; 1.000]		5
regular	zusatzprogrammen	real	avg = 0.000	[0.000 ; 1.000]		6
				SUMME		389
	7 ITUNES					
regular	itunes	real	avg = 0.031	[0.000 ; 4.000]		384
				SUMME		384
	8 INTERNET					
regular	internet	real	avg = 0.025	[0.000 ; 3.000]		311
regular	internetanschluss	real	avg = 0.000	[0.000 ; 1.000]		2
regular	internetbrowser	real	avg = 0.001	[0.000 ; 1.000]		7
regular	internetfunktion	real	avg = 0.000	[0.000 ; 1.000]		2
regular	internetfunktionen	real	avg = 0.000	[0.000 ; 1.000]		3
regular	internetnutzung	real	avg = 0.001	[0.000 ; 1.000]		18
regular	internets	real	avg = 0.000	[0.000 ; 1.000]		3
regular	internetverbindung	real	avg = 0.001	[0.000 ; 1.000]		17
regular	internetverbindungen	real	avg = 0.000	[0.000 ; 1.000]		1
regular	internetzugang	real	avg = 0.001	[0.000 ; 1.000]		10
				SUMME		374
	9 VIDEO					
regular	video	real	avg = 0.007	[0.000 ; 3.000]		84
regular	videoaufnahme	real	avg = 0.000	[0.000 ; 1.000]		5
regular	videoaufnahmefunktion	real	avg = 0.000	[0.000 ; 1.000]		2
regular	videoaufnahmen	real	avg = 0.001	[0.000 ; 1.000]		7
regular	videoaufzeichnung	real	avg = 0.000	[0.000 ; 1.000]		3
regular	videofunktion	real	avg = 0.002	[0.000 ; 2.000]		26
regular	videofunktionalität	real	avg = 0.000	[0.000 ; 1.000]		1
regular	videofunktionen	real	avg = 0.000	[0.000 ; 1.000]		1
regular	videokamerafunktion	real	avg = 0.000	[0.000 ; 1.000]		1
regular	videoplayerfunktion	real	avg = 0.000	[0.000 ; 1.000]		2
regular	videoqualität	real	avg = 0.000	[0.000 ; 1.000]		3
regular	videos	real	avg = 0.015	[0.000 ; 2.000]		180
regular	videowiedergabe	real	avg = 0.001	[0.000 ; 1.000]		17
				SUMME		332

	10	SMS					
regular		kurznachrichten	real	avg = 0.000	[0.000 ; 1.000]		6
regular		sms	real	avg = 0.024	[0.000 ; 4.000]		297
					SUMME		303
	11	KAMERA					
regular		cam	real	avg = 0.001	[0.000 ; 2.000]		9
regular		camera	real	avg = 0.000	[0.000 ; 1.000]		5
regular		digicam	real	avg = 0.000	[0.000 ; 1.000]		5
regular		digitalkamera	real	avg = 0.001	[0.000 ; 2.000]		13
regular		fotokamera	real	avg = 0.000	[0.000 ; 1.000]		4
regular		kamera	real	avg = 0.021	[0.000 ; 3.000]		257
					SUMME		293
	12	AKKU					
regular		akku	real	avg = 0.014	[0.000 ; 2.000]		166
regular		akkukapazität	real	avg = 0.000	[0.000 ; 1.000]		5
regular		akkuladung	real	avg = 0.000	[0.000 ; 1.000]		3
regular		akkulaufzeit	real	avg = 0.004	[0.000 ; 2.000]		48
regular		akkulaufzeiten	real	avg = 0.000	[0.000 ; 1.000]		1
regular		akkuleistung	real	avg = 0.001	[0.000 ; 1.000]		15
regular		akkus	real	avg = 0.001	[0.000 ; 1.000]		7
regular		akkutausch	real	avg = 0.000	[0.000 ; 1.000]		3
regular		akkuwechsel	real	avg = 0.000	[0.000 ; 1.000]		2
regular		handyakku	real	avg = 0.000	[0.000 ; 1.000]		1
					SUMME		251
	13	BEDIENUNG					
regular		bedienung	real	avg = 0.011	[0.000 ; 2.000]		134
regular		bedienungsanleitung	real	avg = 0.002	[0.000 ; 2.000]		22
regular		bedienungsfreundlichkeit	real	avg = 0.000	[0.000 ; 1.000]		2
regular		bedienungskomfort	real	avg = 0.000	[0.000 ; 1.000]		1
regular		benutzerfreundlichkeit	real	avg = 0.000	[0.000 ; 1.000]		5
regular		benutzung	real	avg = 0.001	[0.000 ; 1.000]		15
regular		gerätebedienung	real	avg = 0.000	[0.000 ; 1.000]		1
regular		handhabung	real	avg = 0.002	[0.000 ; 2.000]		24
regular		menübedienung	real	avg = 0.000	[0.000 ; 1.000]		1
regular		usability	real	avg = 0.000	[0.000 ; 1.000]		1
regular		userfriendlyness	real	avg = 0.000	[0.000 ; 1.000]		1
					SUMME		207
	14	VERTRAG					
regular		telekom	real	avg = 0.003	[0.000 ; 1.000]		31
regular		telekomverträgen	real	avg = 0.000	[0.000 ; 1.000]		1
regular		tmobile	real	avg = 0.001	[0.000 ; 1.000]		7
regular		vertrag	real	avg = 0.009	[0.000 ; 2.000]		114
regular		verträge	real	avg = 0.001	[0.000 ; 1.000]		13
regular		verträgen	real	avg = 0.001	[0.000 ; 1.000]		10
regular		vertrages	real	avg = 0.001	[0.000 ; 1.000]		8
regular		vertrags	real	avg = 0.000	[0.000 ; 1.000]		6
regular		vertragsbindung	real	avg = 0.000	[0.000 ; 1.000]		4
regular		mobilmfunkvertrag	real	avg = 0.000	[0.000 ; 1.000]		2
regular		mobilmfunkverträgen	real	avg = 0.000	[0.000 ; 1.000]		1
					SUMME		197

15		BETRIEBSSYSTEM				
regular	betriebssystem	real	avg = 0.002	[0.000 ; 2.000]	19	
regular	betriebssystemen	real	avg = 0.000	[0.000 ; 1.000]	1	
regular	betriebssysteme	real	avg = 0.000	[0.000 ; 2.000]	6	
regular	software	real	avg = 0.009	[0.000 ; 1.000]	106	
regular	technik	real	avg = 0.003	[0.000 ; 1.000]	32	
regular	technologie	real	avg = 0.001	[0.000 ; 1.000]	10	
regular	technologien	real	avg = 0.000	[0.000 ; 1.000]	3	
SUMME					177	

C. Auflistung der Opinion Words aus der Data Preparation (vgl. Kapitel 3.2.3.2):

Positive Opinion Words:

allerfeinsten allerliebste alleskönnendes alltagstaugliche angenehm angenehme angenehmen angenehmer angenehmes angesagt anpassbar ansprechend ansprechendem ansprechende anstandslos ästhetisches attraktiv attraktive attraktiver attraktivsten aufheiterndem ausgefeiltes ausgeklügelte ausgereift ausgereifte ausgereiften ausgereifter ausgereifteste ausgezeichnet ausgezeichnete ausgezeichneten ausgezeichneteter

bedienerfreundlich bedienerfreundlicher bedienerfreundlichste bedienungsfreundlich beeindruckend beeindruckende beeindruckendsten begehrt begehrten begeistert beliebt beliebte beliebtesten benutzerfreundlich benutzerfreundlicher bequem bequeme bequemer bequemes besondere besonderem besonderen besonderes besondern beste bestem besten bestens bester bestes bevorzugte bewegungsempfindlich blitzschnell brauchbar brauchbare brillant brillanter brilliante brillante clever cool coole coolen cooler cooles coolste

detailliert detaillierte detaillierte detailreiche dezent dezenter durchdacht durchdachte durchdachtes durchgestyltes

easy edel edeles edle edlen edler edles effektiv effektiven effektiver eigenständig einfache einfachem einfachen einfacher einfaches einfachsten einwandfrei einwandfreie einwandfreies einzigartig einzigartige elegant elegante eleganten eleganter eleganteren elegantes empfehlenswert empfehlenswertes erfolgreich erfolgreichen erfolgreiches erfolgreichsten erfreuliche erstklassige erstklassigen erstklassiger etabliert exklusiv exklusives extraklasse exzellent exzellente exzellentes

fantastisch fantastischen faszinierend faszinierende faszinierendste fasziniert fehlerfrei fettabweisend fettabweisende fettabweisenden flach flexibel fließend flott flotte freudigen freudigerweise frisch frisches funktional funktionale funktionales funktionell funktionelle funktionellen funktioniert funktioniert funktionierende funktionsfähig funktionsfähigen funktionsfähiger funktionstüchtig funktionstüchtiges geeignet geeignete geeigneter geeignetes gefallen gefällt geil geile geilen geiles geilsten gelungen gelungene gelungenen genaue genauem genauen genauer genauere genauestens genaustens genial geniale genialen genialer geniales genialste gepflegt gepflegte geschickt geschicktes gigantischen glänzend glänzendem glänzenden glücklich glücklichen glücklicher goldrichtig grandios grandiose grandioses graziles grossartig großartige großartiger grossartiges günstig günstige

günstigen günstiger günstigere günstigeren günstigeres günstigste günstigste
günstigsten gute gutem guten guter gutes

hammerschnell handlich handlicher handliches harmonischen herausragend her-
vorragend hervorragende hervorragenden herzliches hilfreich hilfreiche hochauflö-
send hochauflösendem hochauflösenden hochglänzend hochglänzendem hoch-
glänzenden hochsensible hochwertig hochwertige hochwertigem hochwertigen
hochwertiger hochwertige hochzuverlässigen hosentaschentauglich hosenta-
schentaugliches hübsch hübsche hübschen hübsches

ideal ideale idealen idiotensicher individuelle individuellen individueller innovativ
innovativer innovativste intelligente interessant interessante interessanten interes-
santes intuitiv intuitive intuitiver intuitivere

kinderleicht kinderleichte klarverständliche klasse komfortabel komfortable komfor-
tablen komfortabler kompakt kompakte kompakten kompetent kompetenten konk-
urrenzlos konstruktive kontrastreich korrekt kostenfrei kostenfreie kostenlos kos-
tenlose kostenlosen kostenloser kostenloses kratzfest kratzfesten kratzfester
kratzfestes kratzresistent

langlebig legendäre leistungsfähig leistungsfähiger leistungsfähigere leistungsfä-
higerer leistungsfähigste lesbar lesbaren locker logisch logischerweise lohnens-
wertes lückenlos lustige

magischen makellos modern modernen moderner modernes modernste multifunk-
tional multifunktionelle

nahtlose nobel nützlich nützliche nützlichem nützlichen nützlicher nützliches nütz-
lichste

optimal optimale optimalen optimaler optimales optimierten ordnungsgemäß origi-
nell originelle originellen

passgenau perfekt perfekte perfekten perfekter perfektes perfektioniert pflege-
leichteste pflegeleicht pflegeleichtem phantastisch positiv positive positiven positiver
positives praktisch praktische praktischen praktischer praktischerweise prakti-
sches präzise preisgünstigen preiswerte preiswerten problemlos problemlose
problemloser professionelle professioneller punktgenau punktgenaue

reibungslos rentabel revolutionär revolutionäre revolutionären revolutionäres rich-
tig richtige richtigem richtigen richtiger richtiges richtungsweisend riesigen robust
robuster robustes

sagenhaft sagenhaften sanft sauber saubere sauschnell scharf schick schicke
schicken schicker schickes schlank schlanken schlankes schlauer schlaues
schlicht schlichtem schlichten schlichtes schnell schnelle schnelltem schnellen
schneller schnellere schnelleren schnelles schnellste schön schöne schönen

schöner schönere schönes schönste selbsterklärend selbsterklärende sensibel
 sensibles sexy simpel simpelsten simple simplen simpler simples sorglos spacig
 spitzenklasse spitzenmäßig stabil stabilen stabiles stylisch stylische stylischen
 stylischer stylischeres stylish superflott superschnell superschön sympathisch
 sympathischer

tadellos tauglich taugliche toll tolle tolleren toller tolles tollsten top traumhaft trend-
 angehende trendy

überdurchschnittlich überdurchschnittliche überglücklich überragend überragende
 überragenden übersichtlich übersichtlichen übersichtlicher übertroffen überzeu-
 gend überzeugende überzeugenden ultimative umfangreich umfangreiche um-
 fangreicher umwerfende unbegrenzt unbegrenzt unbegrenzten unbegrenzter un-
 eingeschränkt uneingeschränkte unempfindlich unempfindlicher unerreicht
 unglaublich unglaubliche unglaublichen unglaublicher unkompliziert unkomplizierte
 unproblematisch unschlagbar unschlagbare unschlagbaren unstrittig unübertroffen
 unverzerrt unverzichtbar unverzichtbaren

verbessert verbesserte verbesserten verbesserter veredelt vereinfacht vereinfach-
 te vereinfachtes verlässlich verlässlicher verlockend verlockendes vernünftiges ver-
 nünftig vernünftige vernünftiger vertraut vertrautes vielfältig vielfältige vielgeprie-
 senen vielschichtig vielseitig vielseitigen vielseitiger vielversprechend vollkommen
 vollkommene vollständig vollständige vollständigen vollständige
 vollwertig vollwertige vollwertigen vollwertiger vollwertiges vorteilhaft

wegweisend wegweisendes weltklasse witzige wunderbar wunderbare wunderba-
 res wunderschön wunderschöne wunderschönes wundervoll wundervolle

zufrieden zufriedener zufriedenstellend zukunftsweisend zukunftsweisende zuver-
 lässig zuverlässiger zuvorkommend

negative Opinion Words:

abgewetzt albern anfällig anfälliger ärgerlich ausgespart

bedauerlich bedauerlicherweise befremdliche bemängelt bemängelte bestenfalls
 blöd blöde blödsinnigen böse böser

defekt defekte defekten doof dumm dummen dumpf dumpfer dürrer

eingeschränkt eingeschränktes empfindlich empfindliche empfindlichen empfindli-
 cher enttäuschend enttäuschenden erschrocken exzessiv exzessiver

falsch falsche falschen fehlende fehlenden fehlender filigran fragwürdiger frustie-
 rend

gewöhnungsbedürftig gewöhnungsbedürftig gezwungen gezwungene gezwungener
 gezwungenermaßen grässlich graus

hell helle hellem heller

kaputt katastrophal klobig klobigen knapp knappe knifflig komische komische
komplex komplexen kompliziert komplizierte komplizierten komplizierter kompli-
ziertere kostenpflichtem kostenpflichtig kostenpflichtige kostenpflichtigen kosten-
pflichtiges kostenpflichtlich kratzanfällig kratzempfindlich kratzempfindliche krat-
zeranfällig

lächerlich lächerliche lahm langsam langsame langsamen langsamer langsamere
langsameren lästig lästige leidlich

mager mageren mangelhaft mangelhafte mangelhafter mangelnden mickerige
mickrige mickriges mies miese miserabel miserable miserablen mühsam mühselig
nachteilig negativ negative negativen negativer nervenraubend nervig nervige ner-
vigen nüchtern nutzlos

oberflächlich

pfui primitiv problematisch

restriktiven riskant

schade schlecht schlechte schlechtem schlechten schlechter schlechtere schlech-
teren schlechterer schlechtes schlechteste schlimm schlimmer schmerzlicher
schmutzig schockierend schwach schwache schwacher schwächer schwergängig
schwerwiegend schwerwiegende schwerwiegender schwierig schwieriger sinnlos
sinnloses skeptisch störend störende störender superhell

teuer teuerm teuern teuerste teuerster teure teuren teurer teurere teureren teuers-
ten träge träges traurig

übel überarbeitungsbedürftig überbelichtete überhitzt überteuert überteuert über-
teuerte überteuertem überteuerten umständlich umständliche umständlichen um-
ständlicher umständliches umstrittenen unakzeptabel unansehnlich unbrauchbar
unerreicht unerträglich unflexiblen unfreiwillig ungeeignet ungenau ungepflegt un-
glücklich unglückliche unglücklicherweise unhandlicher unkomfortable unleserlich
unlogisch unlösbare unmöglich unnötig unnötige unnötigen unnützen unpassend
unpraktisch unqualifiziert unqualifizierten unscharf unscheinbar unschön unschöne
unschönen unsinnigen unterbelichtete unterdurchschnittlich unübersichtlich un-
übersichtlichsten unverschämt unverständlich unverständlicher unverständliche
unvollständig unzufrieden unzureichend unzureichendes unzuverlässig unzuver-
lässige umständlich umständliche umständliches

veralteten verbesserungswürdig verbesserungswürdige verdammt verfehlt ver-
kratzt verramscht verschleißt verschwommen verwackelt verwackeln verzeifelt
verzögert verzweifelt

widerrechtlicher witz wucher wünschenswert

zerbrechlich zerbrochen zerkratzt zweifelhaft zweitklassig

neutrale Opinion Words:

akzeptabel akzeptable akzeptablen akzeptabler akzeptiert angemessen angemessene annehmbare annehmbaren

bedienbar billig billige billigen billiger billigere billigeren billiges billigste billigsten durchschnittlich

gewöhnlich gewöhnlichen gewöhnliches

herkömmliche herkömmlichen herkömmlicher herkömmliches

mäßig mäßigen mäßiges mittelmäßig mittelmäßige mittelprächtigt

nett nette netten netter nettes neutral neutraler normal normale normalem normalen normaler normalerweise normales

ok

passable passablen

relativ relative

selbstverständlich solide

unwesentlich

verbilligt verschmerzbar

zweckmässig zweckmäßig

D. Stichprobentest zur Bewertung der Validität (Kapitel 3.2.5.1):

Satz Nr. 102:

„Das Update für der Systemsoftware wurde für vorhandene iPhone-Handys wie gewohnt von **iTunes** automatisch durchgeführt und **funktionierte** auch wie gehabt völlig **problemlos**.“

- FF: iTunes
- OW positiv: funktioniert, problemlos

Beschreibung: iTunes funktioniert problemlos.

Satz Nr. 169:

„Auf der linken Seite des Mobiltelefons findet man die manuelle Lautstärkekontrolle sowie einen Stummschalttaster (kein Klingeln mehr bei Anrufen) - eine **Funktion**, die **unglaublich hilfreich** ist, weil so peinliches Klingeln während einer Vorlesung nun leicht zu vermeiden ist.“

- FF: Funktionalität
- OW positiv: unglaublich, hilfreich

Beschreibung: Die Funktion ist unglaublich hilfreich.

Satz Nr. 225:

„Das Konzept finde ich **genial**, denn so **schnell** und leicht habe ich mich noch auf keiner Benutzeroberfläche eines Mobiltelefons zurechtgefunden - und so **effektiv** konnte ich mich auch noch nie in einem Menü bewegen und die gewünschte **Funktion** aufrufen.“

- FF: Funktionalität
- OW positiv: genial, schnell, effektiv

Problem: Die OW „genial“ und „schnell“ beziehen sich auf die „Benutzeroberfläche“ und nicht auf das OW „Funktionalität“.

Satz Nr. 377:

„Im dunklen sind die **Bilder** körnig und zeigen **überbelichtete helle** Stellen.“

- FF: Fotofunktion
- OW negativ: überbelichtet, hell

Beschreibung: Die Bilder sind überbelichtet und hell.

Satz Nr. 460:

„Batterieleistung/**Akku**

*Es ist etwas schwer für mich, hier eine **genaue** Aussage zur Batterieleistung zu treffen, denn meine Nutzung des iPhones variiert täglich.“*

- FF: Akku
- OW positiv: genau

Problem: Das OW „genau“ bezieht sich nicht auf das FF „Akku“.

Satz Nr. 814:

*„Zwar ist es nicht immer **Optimal** dicke Männer Pranken zu haben, aber ein so **sensibles** und **funktionales** Menü/**Display** hab ich noch nie gehabt.“*

- FF: Display
- OW positiv: optimal, sensibel, funktional

Problem: Das OW „optimal“ bezieht sich nicht auf das FF „Display“.

Satz Nr. 882:

*„Später **funktionierte** der Online Support wieder und **Itunes** setzte mein Gerät wieder in den Werks Zustand zurück und es lief wieder **einwandfrei**.“*

- FF: iTunes
- OW positiv: funktioniert, einwandfrei

Problem: Die beiden OWs beziehen sich nicht auf das FF „iTunes“. Das OW „funktioniert“ steht in Zusammenhang zum „Online Support“ und als „einwandfrei“ wird die Funktionsweise des iPhone nach dem Reset beschrieben.

Satz Nr. 972:

*„Je nach Verbindung (WLAN oder EDGE) werden die **Videos** in **guter** bzw.“*

- FF: Videofunktion
- OW positiv: gut

Problem: Das Split erfolgt nicht am Satzende sondern irgendwo inmitten des Satzes. Demnach ist nicht sicher, ob das OW „gut“ auf das FF „Videofunktion“ bezieht und ob das FF im fehlenden Satzteil eventuell noch anderweitig beurteilt wird.

Satz Nr. 1448:

*„:)) Nein also im ernst, ich fand das Handy wirklich **gut**, hatte ein **klasse** Tastatur und **Kamera**, aber dann habe ich das Iphone von meinem Bruder in die Hände bekommen und erstmal damit rumgespielt.“*

- FF: Kamera

- OW positiv: gut, klasse

Problem: Die OWs „gut“ und „klasse“ beziehen sich nicht auf das iPhone, sondern auf ein anderes Handy im Vergleich.

Satz Nr. 1628:

„Dem **Internet** sei Dank hab ich folgendes herausgefunden: Man kann doch auch ohne WLAN ins **Internet**, es ist nur nicht ganz so **schnell**, das heisst, es ist kein UMTS Netz.“

- FF: Internetfunktion
- OW positiv: schnell

Grenzwertiger Fall: Das OW wird hier in Zusammenhang mit einer eingeschränkten Negation verwendet. Zwar wird es noch immer positiv bewertet, jedoch wäre eine noch schnellere Variante unter Umständen wünschenswert.

Satz Nr. 1817:

„Die anfängliche Begeisterung allerdings wich **schnell** zumindest etwas zurück, denn das iPhone sieht zwar extrem **cool** aus und dürfte wohl auch zukünftig den Maßstab für **einfache Bedienung** setzen - aber das sagt leider nicht unbedingt etwas über die Praxistauglichkeit aus.“

- FF: Bedienbarkeit
- OW positiv: schnell, cool, einfach

Problem: Nur das OW „einfach“ bezieht sich auf die „Bedienbarkeit“ des iPhone.

Satz Nr. 2213:

„Das **Display** ist sehr **kratzfest**.“

- FF: Display
- OW positiv: kratzfest

Beschreibung: Das Display ist kratzfest.

Satz Nr. 2678:

„Das iPhone von Apple ist ideal für den geschäftlichen Umgang, da die **Funktionen einfach** und **übersichtlich** sind.“

- FF: Funktionalität
- OW positiv: ideal, einfach

Beschreibung: Die Funktionalität ist einfach und übersichtlich.

Satz Nr. 3008:

„**Empfehlenswert** ist dann allerdings der Abschluss eines Mobilfunk-**Vertrags** mit großem Datenvolumen oder - besser noch - mit einer Daten-Flatrate.“

- FF: Vertrag
- OW: empfehlenswert

Beschreibung: Der Vertrag ist empfehlenswert.

Satz Nr. 3248:

„Dafür hat man in jedem (Complete M, L, XL) der **preisgünstigen** **Verträge eMail Flat**.“

- FF: Vertrag, Email
- OW positiv: preisgünstig

Beschreibung: Der Vertrag und die Email(flat) sind preisgünstig.

Satz Nr. 3661:

„Das wirklich **Geniale** am iPhone 3G ist aber die auf das Wesentliche reduzierte **Bedienung** über die **schnelle** graphische Benutzeroberfläche.“

- FF: Bedienbarkeit
- OW positiv: genial, schnell

Grenzwertiger Fall: Es ist nicht ganz sicher, ob sich das OW „schnell“ auf das FF bezieht. Dazu müsste die „Bedienbarkeit“ die „graphischen Benutzeroberfläche“ mit umfassen.

Satz Nr. 4163:

„Multi **Touch** macht die **Bedienung** einfach nur **intuitiv**.“

- FF: Display, Bedienbarkeit
- OW positiv: intuitiv

Grenzwertiger Fall: Die Bedienung ist intuitiv aufgrund des Multitouch. Es ist nicht ganz klar, ob der Multitouch ebenfalls intuitiv ist.

Satz Nr. 4481:

„Mein Smartphone hat eine **vollwertige** QWERTZ-Tastatur und dafür ein kleines **Display** und muss über Menues oder Hotkeys bedient werden.“

- FF: Display
- OW positiv: vollwertig

Problem: Das OW „vollwertig“ bezieht sich nicht auf das FF „Display“ sondern auf die Tastatur.

Satz Nr. 4363:

„Ich finde allerdings **iTunes** garnicht so **schlecht**, denn man hat dadurch immer Ordnung auf dem iPhone.“

- FF: iTunes
- OW negativ: schlecht

Problem: Das OW „schlecht“ wird in Zusammenhang mit der Negation „nicht“ verwendet. Demnach findet der Rezensent iTunes eigentlich gut, der RapidMiner interpretiert die Aussage jedoch gegenteilig.

Satz Nr. 4696:

„Alleine das **Internet funktioniert** kongurenzlos **fantastisch**.“

- FF: Internetfunktion
- OW positiv: funktioniert, fantastisch

Beschreibung: Das Internet funktioniert fantastisch.

Satz Nr. 4862:

„Sowie die Erstellung unterschiedlicher Wiedergabelisten, der Genius-**Funktion** und natürlich dem **toll** anzusehenden Cover-Flow.“

- FF: Funktionalität
- OW positiv: toll

Problem: Das OW „toll“ bezieht sich nicht auf das FF „Funktionalität“, sondern auf den „Cover-Flow“.

Satz Nr. 5037:

„Die **Bedienung** ist nicht immer so einfach und **intuitiv** wie Apple behauptet, ein Handbuch wäre manchmal **hilfreich**.“

- FF: Bedienbarkeit
- OW positiv: intuitiv, hilfreich

Problem: Das OW „intuitiv“ wird von einer Negation „nicht immer“ begleitet. Demnach kodiert der RapidMiner die gegenteilige Meinung des Rezensenten. Zudem bezieht sich das OW „hilfreich“ auf das „Handbuch“.

Satz Nr. 5229:

„**Sms**-menü ist **unpraktisch** weil alle **sms** angezeigt werden, ganz seltsam.“

- FF: SMS
- OW negativ: unpraktisch

Beschreibung: SMS ist unpraktisch.

Satz Nr. 5582:

„Dazu kommt noch die **einfache** und **problemlose Bedienung** des Gerätes mit den Fingern von welcher ich wirklich absolut **begeistert** bin.“

- FF: Bedienbarkeit
- OW positiv: einfach, problemlos, begeistert

Beschreibung: Die Bedienbarkeit ist einfach und problemlos. Der Rezensent ist von der Bedienbarkeit begeistert.

Satz Nr. 5671:

„Aber wenn es klappt, kann man sich direkt über den AppStore **Programme, Spiele** und andere **Apps** kaufen, einiges ist sogar **kostenlos**.“

- FF: Applications
- OW positiv: kostenlos

Beschreibung: Applications sind kostenlos. (Die Aussage trifft zumindest auf einiger der Apps zu.)

Satz Nr. 5788:

„3) Lieferumfang + Preis

Da mein **Mobilfunkvertrag** glücklicherweise zum Zeitpunkt ausgelaufen ist, als das neue iPhone auf den Markt kam, habe ich die Chance sofort genutzt und mir das **schicke** Gerät für einen Euro geholt.“

- FF: Vertrag
- OW positiv: schick

Problem: Das OW „schick“ beschreibt das iPhone und nicht das FF „Mobilfunkvertrag“.

Satz Nr. 5981:

„Display----- Das **Display** kann sehr, sehr **hell** sein.“

- FF: Display
- OW negativ: hell

Beschreibung: Das Display ist hell.

Satz Nr. 6094:

„Ja Sie haben **richtig** gehört per **e-mail**.“

- FF: Emailfunktion
- OW positiv: richtig

Problem: Das OW „richtig“ steht nicht in Bezug zum FF „Emailfunktion“.

Satz Nr. 6232:

„**Spiele** gibt es natürlich auch , die Steuerung bei manchen **Spiele**n ist echt **cool** man muss das Handy bewegen um die Figur im Spiel zu bewegen.“

- FF: Applications
- OW positiv: cool

Beschreibung: Die (Spiele-)Applications sind cool.

Satz Nr. 6492:

„***** **Fotos** anschauen ist **fantastisch**, **Fotos** machen eher nicht ***** **Fotos** auf dem iPhone sind wirklich sensationell.“

- FF: Fotofunktion
- OW: fantastisch

Grenzwertiger Fall: Eigentlich sagt der Rezensent nur aus, dass das „Anschauen“ von Fotos „fantastisch“ ist. Der Satzteil „Fotos machen eher nicht“ könnte jedoch darauf hin deuten, dass die Fotofunktion des iPhone nicht zufriedenstellen ist. Da der darauf folgende Satzteil „Fotos auf dem iPhone sind wirklich sensationell“ jedoch das Foto an sich positiv bewertet, kann dieser Satz trotzdem positiv kodiert werden.

Satz Nr. 6531:

„Auch wenn das iPhone 3G ein Gerät ist wie kein anderes und eine **revolutionäre Bedienung**, sowie einige andere Dinge bietet, die kein anderes Handy hat, so gibt es dennoch ein paar Minuspunkte.“

- FF: Bedienbarkeit
- OW positiv: revolutionär

Beschreibung: Die Bedienbarkeit ist revolutionär.

Satz Nr. 6879:

„Die Ladezeiten der Seiten sind dabei einen Hauch länger als bei der DSL-**Verbindung**, aber doch **schneller** als bei einem Modem-Internet-Zugang.“

- FF: Internetfunktion
- OW positiv: schnell

Beschreibung: Die Internetverbindung ist schnell.

Satz Nr. 7354:

„Und so stellte ich letzte Woche **iPod**, wie auch Handy bei Ebay rein, verabschiedete mich von zwei Geräten und sagte Hallo zum iPhone 3G 16GB in Schwarz, dass ich mir im Zuge einer **besonderen** Aktion von T-Mobile (<http://www>“

- FF: Musik
- OW: besonders

Problem: Das OW „besonderen“ bezieht sich auf eine „Aktion von T-Mobile“ und nicht auf das FF „iPod“.

Satz Nr. 7553:

„Der **iPod** ist in Zusammenhang mit den 16GB Speicherplatz ein wahres Multimedia-Center und auf dem iPhone ebenso **bequem** und **praktisch** zu bedienen wie immer.“

- FF: Musik
- OW positiv: bequem, praktisch

Beschreibung: Der iPod ist bequem und praktisch.

Satz Nr. 7787:

„Wie aber bereits beim Vorgängermodell möchte ich dem IPHONE 3GS bescheinigen, dass es über die **beste** und **ausgereifteste Technik** und Qualität verfügt, welche im Bereich der Medienwiedergabe und -verwaltung mit einem Mobiltelefon erhältlich ist.“

- FF: Betriebssystem
- OW positiv: beste, ausgereift

Beschreibung: Es handelt sich um die beste und ausgereifteste Technik.

Satz Nr. 8240:

„Mittlerweile habe ich schon ein recht **gutes** Gespür dafür bekommen wie lang eine **SMS** ist.“

- FF: SMS
- OW positiv: gut

Problem: Das OW „gutes“ steht nicht in Zusammenhang zum FF „SMS“, sondern bezieht sich auf das „Gespür“.

Satz Nr. 8538:

„Alles lässt sich **problemlos** mit den Fingern bedienen und das **Display** reagiert **schnell** und **Präzise**.“

- FF: Display
- OW positiv: problemlos, schnell, präzise

Problem: Das OW „problemlos“ steht nicht im Bezug zum FF „Display“, stattdessen jedoch zum FF „Bedienbarkeit“, die in diesem Falle jedoch nicht als FF erkannt wird, da sie nicht als Substantiv im Satz auftritt.

Satz Nr. 9223:

„Ich versende Dateien einfach per **E-Mail**, was deutlich **schneller**, ebenfalls **kostenlos** und zukunftsorientierter ist.“

- FF: Emailfunktion
- OW positiv: schnell, kostenlos

Beschreibung: Die Emailfunktion ist schnell und kostenlos.

Satz Nr. 9676:

„Die Synchronisation aller Daten **funktioniert** in der Regel über **iTunes** auf dem PC oder Mac, das **funktioniert problemlos**.“

- FF: iTunes
- OW positiv: funktioniert (2x), problemlos

Beschreibung: iTunes funktioniert problemlos.

Satz Nr. 10.181:

„Spulen innerhalb jeglicher Datei **funktioniert** ebenso **schön** wie man es von allen **iPods** kennt – und auch genauso **schnell**, egal ob 3MB Musikstück, 600MB großer Film oder ein 4GB großes Audiobook.“

- FF: Musik
- OW positiv: funktioniert, schön, schnell

Beschreibung: Der iPod funktioniert schön und ist schnell.

Satz Nr. 10.293:

„**Selbstverständlich** findet man am unteren Rand des **Displays** weitere Icons für zusätzliche Optionen wie Favouriten, Anrufliste, die erwähnten Kontakte, den traditionellen Ziffernblock sowie Voicemail.“

- FF: Display
- OW positiv: selbstverständlich

Grenzwertiger Fall: Das OW „selbstverständlich“ bezieht sich auf die Icons, die sich auf dem FF „Display“ befinden. Da die Icons jedoch zum Display gehören, könnte man das OW auch in Zusammenhang zum Display sehen.

Satz Nr. 10.297:

„**Störend** finde ich dagegen das **Umständliche** Löschen der Nachrichten, da man jede **E-Mails** einzeln lösen muss.“

- FF: Emailfunktion
- OW negativ: störend, umständlich

Beschreibung: Das Löschen von Emails ist störend und umständlich.

Satz Nr. 10.756:

„- **Mail** kann, wie jeder **vernünftige** Mailclient am Computer/Mac auch, auf alle POP/IMAP (zwei Transportprotokolle) Mail-Konten zugreifen und darüber **eMails** senden und empfangen.“

- FF: Emailfunktion
- OW positiv: vernünftig

Beschreibung: Die Emailfunktion ist vernünftig.

Satz Nr. 10.843:

„Im Downloadcenter kann man sich **einfach** und **schnell** neue **Programme** und Zusatzsoftware dazu installieren was an sich richtig **genial** ist.“

- FF: Applications
- OW positiv: einfach, schnell, genial

Beschreibung: Die Programme sind einfach, schnell und genial.

Satz Nr. 10.923:

„Dazu kann man sich am PC **kostenlos** die AlbumCover laden, so dass man mit dem **iPod** immer **schön** die Alben in der Bibliothek hat.“

- FF: Musik
- OW positiv: kostenlos, schön

Problem: das OW „kostenlos“ beschreibt das AlbumCover und nicht die Musik.

Satz Nr. 11.394:

„Die Zoom **funktion** reagiert sehr **langsam**.“

- FF: Funktionalität
- OW negativ: langsam

Beschreibung: Die Funktion ist langsam.

Satz NR. 11.634:

„:D Ich bin seit 3 Wochen im Besitz des neuen iPhone 3G Ein **grandioses** Gerät mit unglablich vielen **Funktionen**.“

- FF: Funktionalität
- OW positiv: grandios

Problem: Das OW „grandios“ steht im Zusammenhang zum iPhone und nicht speziell zum FF „Funktionen“.

Satz Nr. 11.679:

„Der **Akku** ist viel zu **schwach**, daher ist es wirklich nichts für Personen die viel mit dem iPhone arbeiten / spielen wollen und dabei auch lange unterwegs sind.“

- FF: Akku
- OW negativ: schwach

Beschreibung: Der Akku ist schwach.

Satz Nr. 12.136:

„Das Iphone ist **handlich** und die **Software** hat mich echt überzeugt.“

- FF: Betriebssystem
- OW positiv: handlich

Problem: Das OW „handlich“ bezieht sich nicht auf das FF „Software“, sondern auf das iPhone.

Satz Nr. 12.201:

„Super Handy aber leider ohne **Vertrag** zu **teuer**.“

- FF: Vertrag
- OW negativ: teuer

Problem: Der RapidMiner kodiert den Vertrag als teuer. Der Rezensent sagt jedoch gegenteiliges aus: Er ist der Meinung, dass das Handy „ohne Vertrag“ zu teuer ist.

Satz Nr.	Kodierung	Satz Nr.	Kodierung
102	Korrekt	5788	Falsch
169	Korrekt	5981	Korrekt
225	Falsch	6094	Falsch
377	Korrekt	6232	Korrekt
460	Falsch	9492	Falsch
814	Falsch	6531	Korrekt
882	Falsch	6879	Korrekt
972	Falsch	7354	Falsch
1448	Falsch	7553	Korrekt
1628	Grenzwertig	7787	Korrekt
1817	Falsch	8240	Falsch
2213	Korrekt	8538	Falsch
2678	Korrekt	9223	Korrekt
3008	Korrekt	9676	Korrekt
3248	Korrekt	10.181	Korrekt
3661	Grenzwertig	10.293	Grenzwertig
4163	Grenzwertig	10.297	Korrekt
4481	Falsch	10.756	Korrekt
4363	Falsch	10.843	Korrekt
4696	Korrekt	10.923	Falsch
4862	Falsch	11.394	Korrekt
5037	Falsch	11.634	Falsch
5229	Korrekt	11.679	Korrekt
5582	Korrekt	12.136	Falsch
5671	Korrekt	12.201	Falsch

Abbildung 56: Kodierung einiger Sätze durch den RapidMiner

E. Stichprobe Akku (Kapitel 3.2.5.2)

Satz Nr. 461:

„Batterieleistung / Akku Es ist etwas schwer für mich, hier eine **genaue** Aussage zur Batterieleistung zu treffen, denn meine Nutzung des iPhones variiert täglich“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 659:

„Akku Leistung Der in das iPhone integrierte Akku kann auch dem ersten Blick recht **schwach** erscheinen, besonders dann wenn man gerne das Display **hell** erstrahlen lassen möchte, die W-Lan und Bluetooth aktiviert ist“

→ Der Akku ist schwach → **richtig**

Satz Nr. 662:

„Für alle die das iPhone geschäftlich nutzen kann die Akku-Leistung manchmal **ärgerlich** sein, aber es bleibt ja jeden selbst überlassen, welche Funktionen man ständig aktiviert hat“

→ Die Akkuleistung ist ärgerlich → **richtig**

Satz Nr. 663:

„Das **schnelle** Wechseln des Akkus ist nicht möglich“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 1292:

„80 EUR für einen neuen Akku außerdem auch noch recht **teuer**“

→ Der Akku ist teuer → **richtig**

Satz Nr. 1468:

„dann ist der Akku vielleicht **schneller** leer, aber wenn einige Sachen beachtet werden (nicht benötigte Funktionen, wie Wifi, BT ausschalten, Display Sperre auf 1 Minute usw), dann hält der Akku schon ca“

→ Es handelt sich um eine neutrale Aussage („schnell“ ist als OW schlecht gewählt) → **falsch**

Satz Nr. 1498:

„Man sollte beachten, dass man selber nur **relativ** aufwändig der Akku tauschen kann (Kostenpunkt ca: 35-40 EUR), bei Apple kostet ein Akkutausch ca“

→ Es handelt sich um eine negative Aussage → **falsch**

Satz Nr. 1575:

„Ich hatte es neulich auf der Autofahrt als Ipod an mein Autoradio angeschlossen, durchgehend damit Musik gehört, dazwischen auch noch telefoniert, als ich acht Stunden später an meinem Ziel angekommen war, ging das Akku **langsam** der Neige zu“

→ Das OW bezieht sich zwar nicht auf das FF, trotzdem ist die Aussage insgesamt negativ → **grenzwertig**

Satz Nr. 1749:

„Meiner Meinung nach, ist das Akku für so viel Power ein wenig zu **schwach**“

→ Der Akku ist schwach → **richtig**

Satz Nr. 1826:

„Viel **schlimmer** finde ich, dass die Akkulaufzeit sehr zu wünschen übrig lässt“

→ Schlimm ist die Akkulaufzeit, da sie zu wünschen übrig lässt → **richtig**

Satz Nr. 1851:

„der kurzen Akkulaufzeit mit aktiviertem WLAN und **fehlender** VoIP-Fähigkeit“

→ Das OW bezieht sich zwar nicht auf das FF, trotzdem ist die Aussage insgesamt negativ → **grenzwertig**

Satz Nr. 1873:

„Und seit keine SIM mehr eingelegt und damit das Mobilfunkmodul nicht mehr aktiv ist, hat sich die Akkulaufzeit auf ein für mich "**normales**" Maß eingependelt“

→ Die Akkulaufzeit ist normal → **richtig**

Satz Nr. 2266:

„die Geschwindigkeit mit Wlan habe ich noch gar nicht getestet +Akkulaufzeit: Mein iPhone hält trotz gelegentlichen Musikhören und Internet surfen 4-5 Tage, außerdem wird täglich telefoniert und Emails abgerufen (Automatisch jede Stunde) der Akkustand zeigt zwar **schnell** halb leer an hält hinten raus aber noch ziemlich lang“

→ Das OW bezieht sich zwar nicht auf das FF, trotzdem ist die Aussage insgesamt eher positiv → **grenzwertig**

Satz Nr. 2352:

„Nach einem **Defekt** muss also das gesamte iPhone eingeschickt und ein neuer Akku eingesetzt werden“

→ Bei dem Wort „Defekt“ handelt es sich nicht um ein Adjektiv und somit ist es kein OW. Trotzdem ist die Aussage insgesamt negativ → **grenzwertig**

Satz Nr. 2475:

„Natürlich ziehen Spiele eine Menge Kraft vom Akku, deswegen kann ich nicht vieles dazu sagen außer: Wenn ihr auf einem 3Tage Trip oder ähnlichem seid, nehmt euch euer Aufladegerät mit :-). Meine Meinung zum iPhone Einfach **Top**“

→ Das OW bezieht sich zwar nicht auf das FF, trotzdem ist die Aussage insgesamt eher positiv → **grenzwertig**

Satz Nr. 2814:

„Nach ersten Tests beträgt die Sprechzeit über sieben Stunden, Musik kann man **knapp** über 22 Stunden lang hören, und bei Internetnutzung entlädt sich der Akku in etwas mehr als neun Stunden“

→ Es handelt sich eher um eine neutrale Aussage, da ausschließlich Fakten genannt werden („knapp“ ist als OW ist schlecht gewählt) → **falsch**

Satz Nr. 3019:

„Vorzüge des iPhones gegenüber anderen Handys&PDA's: - Sehr scharfes Display- Das Design ist **einzigartig** und sehr gelungen-Akkulaufzeit: 8Std“

→ Das OW bezieht sich nicht auf das FF, trotzdem ist die Aussage insgesamt positiv → **grenzwertig**

Satz Nr. 3181:

„Der Akku hält nicht sehr lange, das Touchscreen ist auch sehr **mühsam**“

→ Das OW bezieht sich zwar nicht auf das FF, trotzdem ist die Aussage insgesamt negativ → **grenzwertig**

Satz Nr. 3221:

„Darüber war ich dann auch sehr **froh**, als mein iPhone sich nach dem Kauf als nicht **vollständig** okay erwies (der Akku hatte nicht die vollständige Kapazität)“

→ In Bezug auf den Akku handelt es sich um eine negative Aussage → **falsch**

Satz Nr. 3753:

„Schliesslich ist das iPhone auch iPod, und trotzdem ist meiner Meinung nach der Akku zu **schnell** leer“

→ Es handelt sich um eine negative Aussage („schnell“ ist als OW ist schlecht gewählt) → **falsch**

Satz Nr. 3849:

„+ es **funktioniert** halt alles und ist auch alles aufeinander abgestimmt so, jetzt aber kontra und so leid mir das tut, muss ich die pro's wieder schlechreden: - der akku, was ist das denn bitte“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 3942:

„Das Telefon hat eine super **funktionierende** und **coole** Navi-Funktion - wenn es am Strom hängt (bzw Auto Ladekabel) oder aber solange der Akku hält“

→ Die OWs beziehen sich nicht auf das FF → **falsch**

Satz Nr. 4333:

„- **schwache** Antenne - Akku reicht ca“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 4342:

„Aber mal ein Tipp an alle, die denken der Akku wäre zu **schwach**: geht doch mal in die settings --> general --> usage und vergleicht die Usage mit Standby“

→ Das OW bezieht sich zwar auf das FF, jedoch findet der Rezensent der Aussage nach den Akku gar nicht zu schwach, sondern nennt nur die Meinung anderer → **grenzwertig**

Satz Nr. 4343:

„Aber mal ein Tipp an alle, die denken der Akku wäre zu **schwach**: geht doch mal in die settings --> general --> usage und vergleicht die Usage mit Standby“

→ Das OW bezieht sich zwar auf das FF, jedoch findet der Rezensent der Aussage nach den Akku gar nicht zu schwach, sondern nennt nur die Meinung anderer → **grenzwertig**

Satz Nr. 4376:

„Einziges Manko: der Akku ist einfach zu **schnell** leer“

→ Es handelt sich um eine negative Aussage („schnell“ ist als OW schlecht gewählt) → **falsch**

Satz Nr. 4377:

„Hab mir jetzt noch ein zweites **"normales"** Handy zugelegt dass ich bei leerem Akku oder wenn ich keine Möglichkeit hab das Iphone zu laden trotzdem noch erreichbar bin“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 4395:

„Faktum ist, dass ich beinahe täglich **gezwungen** bin, den Akku wieder aufzuladen, obwohl ich mich zu einem "Wenig-Telefonierer" zähle, genauso wie ich außerhalb von zu Hause (Wi-Fi) das mobile Internet überhaupt nicht nutze“

→ Der Akku zwingt einen zum Aufladen → **richtig**

Satz Nr. 4508:

„Klar das der Akku öfter ran muss zum laden, da das Display auch größer, **heller** und besser in der Auflösung ist als bei anderen Handys“

→ Das OW bezieht sich nicht auf das FF. Der Rezensent scheint Verständnis zu haben → **grenzwertig**

Satz Nr. 4581:

„Bei aller berechtigter Kritik an den Mängeln des iPhones (hoher Preis, kein austauschbarer Akku, kein richtiges Bluetooth, keine Videotelefonie, keine **vollwertige** Navigationslösung, iTunes-Bindung etc“

→ Das OW bezieht sich nicht auf das FF. Zudem wird es in Verbindung mit einer Negation verwendet und ist demnach negativ zu bewerten. → **falsch**

Satz Nr. 4643:

„Akkukapazität für ein Handy kläglich, für ein Smartphone **OK**“

→ Der Akku ist ok → **richtig**

Satz Nr. 4748:

„) Zu guter letzt wäre da noch der Akku, dessen Leistung auch **dürftig** ist“

→ Die Akkuleistung ist dürftig → **richtig**

Satz Nr. 5004:

„kein großes design, erleichtern das leben jedoch ungemein und sind **kostenlos** wundervolle technik, aber leider hält der akku nicht mit“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 5026:

„Dennoch: - Die Standzeit des Akku ist **unerträglich** kurz“

→ Die Standzeit des Akkus ist unerträglich → **richtig**

Satz Nr. 5112:

„Ich zähle nur einiges auf: - sehr **schnelle** Internetverbindung - alle Anhänge sind zu öffnen - unzählige zusätzliche Apps aus dem App-Store - **tolle** Spiele in **richtigem** 3D - **genialer** Touch-Screen - **gute** Sprchqualität - **gute** Akkulaufzeit - **toller** Internetbrowser - Copy & Paste Funktion zwischen eigentlich allen Apps - Navigation mit Navigon und TomTom möglich - Google Maps usw“

→ sechs der sieben OWs beziehen sich nicht auf das FF. Die Aussage ist jedoch durchweg positiv → **grenzwertig**

Satz Nr. 5145:

„Ich finde es **cool** dass ich 3G deaktivieren kann, falls ich nicht Surfen mag reicht mir GSM, da wird der Akku auch nicht zu **schnell** verbaucht“

→ Die OWs beziehen sich nicht auf das FF → **falsch**

Satz Nr. 5237:

„akku-zeit, naja, einer findets **ok**, einer nicht“

→ Die Akkulaufzeit ist ok → **richtig**

Satz Nr. 5547:

„**Schade** dass der Akku fest verbaut und nicht einfach austauschbar ist“

→ Die feste Verankerung des Akkus ist schade → **richtig**

Satz Nr. 5798:

„- Es ist **schneller** als seine Vorgänger (Arbeitsgeschwindigkeit und Internetverbindung)- Längere Akkulaufzeit- Verbesserte 3D Grafik 6) Fazit Das iPhone ist ein echter Allrounder“

→ Das OW bezieht sich zwar nicht direkt auf das FF, trotzdem ist die Aussage insgesamt positiv → **grenzwertig**

Satz Nr. 6029:

„Fazit:-----Das neue iPhone bietet einige neue und **tolle** Funktionen, ist **schneller** und der Akku hält wenigstens etwas länger“

→ Die OWs beziehen sich nicht auf das FF → **falsch**

Satz Nr. 6074:

„Akku **Positiv** überrascht hat mich der **verbesserte** Akku“

→ Der Akku wurde verbessert und wird positiv bewertet → **richtig**

Satz Nr. 6606:

„)•Verwaltet E-Mail, Kalender und Kontakte•Zeigt Fotos und Videos in Super-Qualität an Kontra:•Fingerabdrücke auf dem Gehäuse•Vertrag nur mit T-Mobile•kein erweiterbarer Speicher•kein austauschbarer Akku•**unzureichendes** Bluetooth•Modemnutzung nur als Extra-Option (nicht im Vertrag inklusive)•nur mit iTunes verwendbar•keine Videofunktion, eher **schlechte** Fotokamera•Kein Flash-Player 11“

→ Die OWs beziehen sich zwar nicht direkt auf das FF, trotzdem ist die Aussage zum Großteil negativ → **grenzwertig**

Satz Nr. 7342:

„Die Akkulaufzeit ist halbwegs **ok**“

→ Die Akkulaufzeit ist ok → **richtig**

Satz Nr. 8505:

„2009 Sooo, und wieder ein Update zur Akkulaufzeit, weswegen ich auch eben diese als **negative** Eigenschaft zurück ziehe :-) Mittlerweile hat sich mein Nutzverhalten eingependelt“

→ Aufgrund des Updates hat sich die Meinung des Rezensenten geändert und er sieht den Akku nicht mehr als „negative Eigenschaft“ an → **falsch**

Satz Nr. 8533:

„Also lange Rede kurzer Sinn, wie so häufig bei mir :-D Ich habe letztens Zeiten erreicht von: •5 Tage 16 Stunden Standby•8 Stunden Nutzung Momentan bin ich bei 4 Tagen Standby und 4 Stunden Nutzung und der Akku nähert sich **langsam** aber sicher dem Ende“

→ Das OW bezieht sich nicht auf das FF („langsam“ ist als OW schlecht gewählt) → **grenzwertig**

Satz Nr. 8929:

„Diese Funktion ist bei mir **unnötig** und kostet nur Akkuleistung“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 9010:

„Die einzigen Punkte, die man als **Negativ** aufführen könnte neben dem Preis, sind der Akku und die Kamera“

→ Der Akku wird negativ betrachtet → **richtig**

Satz Nr. 9344:

„Unter Verwendung von UMTS, also mit aktivierten 3G Netz (**schnelle** Datenverbindung), beläuft sich die Akkulaufzeit bei einer Nutzung von 4h pro Tag, auf ca“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 9353:

„Nicht so **gut** finde ich dass der Akku sich nicht **einfach** so auswechseln lässt und die Akkulaufzeit bei hoher Nutzung sehr schnell leer ist“

→ Die OWs stehen beiden in Verbindung zu einer Negation. Die Aussage über den Akku ist negativ. → **falsch**

Satz Nr. 9430:

„Die Akkulaufzeit des IPHONE 3G würde ich als mehr als **angemessen** bezeichnen“

→ Aufgrund der Formulierung „mehr als angemessen“ findet der Rezensent die Akkulaufzeit nicht nur durchschnittlich, sondern gut. → **falsch**

Satz Nr. 9490:

„Bedauerlicherweise finde ich es nicht gut dass man den Akku nicht selber austauschen kann, das wird wohl heißen sich in zwei Jahren wieder ein neues iPhone zu kaufen, weil wahrscheinlich der Austausch **teurer** ist als sich ein neues anzuschaffen“

→ Der Akkutauch ist teuer → **richtig**

Satz Nr. 9690:

„Die Akku-Laufzeit ist bei „**normalem** Telefonbedarf“ bei maximal 2 Tagen zu sehen“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 10205:

„Die Akkuleistung ist mit knappen 24h im 3G Betrieb **locker** ausreichend, wenn nun die Push Notification und Navi Software mit dem OS 3“

→ Das OW bezieht sich nicht auf das FF, trotzdem ist die Aussage insgesamt positiv („locker“ ist als OW schlecht gewählt) → **grenzwertig**

Satz Nr. 10852:

„Akkulaufzeit Die meines Achtens einzigen wirklich Mängel, die das Telefon aufweist sind zu einem die Akkulaufzeit, die bei häufiger Telefonbenutzung doch mal **schnell** in die Knie geht“

→ Es handelt sich um eine negative Aussage („schnell“ ist als OW schlecht gewählt) → **falsch**

Satz Nr. 11113:

„kurze Zusammenfassung der Vor- und Nachteile: Vorteile:- GPS- **einfache** Menüführung- **schnelles** 3G- sehr **gute** Sprachqualität- man kann sich tausende Programme zusätzlich auf das iPhone laden (zum Teil **kostenlos**)- sehr großes Display- **richtige** Tastatur- lange Akkulaufzeit Nachteile:- keine Sprachführung - nicht mit jedem Bluetooth-Handy so leicht zu verbinden
- vll die Größe (mir gefällt) Insgesamt bereue ich den Kauf nicht, es ist ein wirklich **tolles** und **durchdachtes** Gerät“

→ Die OWs beziehen sich nicht auf das FF. Trotzdem wird der Akku positiv beschrieben → **grenzwertig**

Satz Nr. 11149:

„Akku:Akku-Typ Lithium-Ionen Standby-Zeit 12 Tage, 12 Stunden Sprechzeit 10:00 Stunden Mindest-Sprechzeit 3:59 Stunden Mein Freund hat dieses iPhone ich find es **schick**, aber mir fehlen da noch ein paar Sachen“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 11177:

„Akku:Der Akku ist mit Abstand das **schlechteste** am Iphone, er muss fast jeden Abend aufgeladen werden“

→ Der Akku ist schlecht → **richtig**

Satz Nr. 11236:

„Aber wenn man nur ab und zu **schnell** telefoniert, eben **schnell** News nachguckt und mal ne SMS schreibt, hält der Akku 5-7 Tage“

→ Das OW bezieht sich nicht auf das FF („schnell“ ist als OW schlecht gewählt).
Trotzdem ist die Aussage eher positiv formuliert → **grenzwertig**

Satz Nr. 11344:

„ - der Akku hält bei „**normaler** benutzung“ nur einen Tag, höchstens 1,5 Tage“

→ Das OW bezieht sich nicht auf das FF → **falsch**

Satz Nr. 11680:

„Der Akku ist viel zu **schwach**, daher ist es wirklich nichts für Personen die viel mit dem iPhone arbeiten / spielen wollen und dabei auch lange unterwegs sind“

→ Der Akku ist schwach → **richtig**

Satz Nr. 12201:

„Dies begünstigt die Akkulaufzeiten auch nicht besonders, was sich **schnell** bemerkbar macht, wenn man sich mit dem Gerät bewegt, zum Beispiel auf einer längeren Fahrt mit dem Auto oder aber der Bahn, ganz besonders wenn 3G (UMTS) aktiviert ist oder man mit dem iPhone surft“

→ Es handelt sich um eine negative Aussage („schnell“ ist als OW schlecht gewählt) → **falsch**

Ergebnis:

Von 61 Meinungssätzen zum Akku hat der RapidMiner

18 Sätze (29,5%) richtig,

26 Sätze (42,6%) falsch und

17 Sätze (27,9%) grenzwertig kodiert.