Driving factors behind product ratings in the smartphone industry ‡

N.B. van den Berg§

This version: February 18, 2021

Academic supervisors:
Prof. Dr. Ph.H.B.F. Franses ¶
Prof. Dr. M.G. de Jong

Abstract

The smartphone and its functionalities can no longer be ignored in contemporary life. Since the introduction in 1992, it has completely changed our life and the way we do business. Given this, a highly competitive market has emerged of companies that produce smartphones. Given the approximately 2.5 billion smartphone users nowadays, not only a lot of money can be earned but also a lot of influence can be gained. At the same time, an increasing amount of user-generated content became available. The rise of machine learning has created possibilities to analyze this user-generated content (UGC) on a large scale and output insightful information. This thesis aims to analyze UGC about smartphones specifically. Future smartphone production is highly dependent on consumer preferences and therefore the information extracted from product reviews is very valuable for this industry. The main dataset is web scraped from a variety of online retailers and contains 121.080 reviews after preprocessing. The dataset has been expanded with information about smartphone characteristics. In this thesis, we focus primarily on sentiment analysis and topic modeling supported by regression analysis. This approach allows us to come up with specific business recommendations for players in the smartphone industry.

JEL classification: C25, C38, C53. C55, C88, M31
Keywords: User-Generated Content, LDA, Sentiment analysis, Marketing analytics

‡A thesis submitted in fulfillment of the requirements for the degree of *Master of Science* in Data Science and Marketing Analytics

§Corresponding author: Hoevestraat 33A, 3033 GB, Rotterdam (NLD) — E-mail: nielsvandenberg@msn.com — Faculty: Erasmus School of Economics — Erasmus University Student Number: 431278.

¶Special thanks to prof. dr. Ph.H.B.F. Franses for his flexibility, valuable comments, and guidance throughout the process.

Contents

1 Introduction

"You can't connect the dots looking forward; you can only connect them looking backwards. So you have to trust that the dots will somehow connect in your future. You have to trust in something - your gut, destiny, life, karma, whatever. This approach has never let me down, and it has made all the difference in my life." (Steve Jobs, 2005)

Nowadays we can not imagine life without our smartphones. Nevertheless, the history of this essential gadget in our daily lives only starts in 1992. This smartphone, Simon Personal Communicator, produced by IBM does not even come close to the device we are used to now. The first smartphone that did come close to that was presented by Steve Jobs in 2007. The iPhone had not only a revolutionary touch screen, it was also the first device that offered a full, un-watered down version of the internet. This device gave consumers the ability to browse the web just as they would on a desktop computer (Andrew, 2018).

Since the first iPhone was presented and the world became acquainted with the type of smartphone that we still know today, a lot has happened in the field of development and diversification in this market. Now, 25 iPhones have already been released and big competitors have entered the market making cell phones running on the Android system, which is Google's answer to the iPhone. With the rise of apps, being able to take professional-quality photos and messaging through wireless connections, we can safely say that smartphones changed our lives. Of the estimated 5 billion cellphone owners around the world, it's estimated that at least 2.5 billion of them own a smartphone. Of course, with the increasing economy and general wealth around the globe, this number is only predicted to rise (Park et al., 2014).

Besides that smartphones changed our lives it also changed how we do business. The total influence on business is too comprehensive to be summarized in this introduction, but the impact on modern-day business can be narrowed down to four different ways of impact. The creation of the 'right now' culture, farewell to office space, advertiser's delight, and social media. The first way of impact represents the opportunity the smartphone gives us to access the world's database of information within seconds. People expect quick replies and responses to all sorts of requests. With different communication tools that can be run by a smartphone, work can be done from everywhere and employees are far from dependent on office space. In the world of advertising, smartphones brought new space for advertisers to expose content. The revenue generated from mobile advertising has become a significant part of the income of many businesses. Nearly 80% of social media time consummation is on smartphones. This resulted in millions of new business that are built on social media and its possibilities (Tocci, 2019).

The smartphone rolled into our lives and changed the way we live and do business. Along with the rapid development of smartphones, there has also occurred an explosive growth of online word-of-mouth behavior (Liu et al., 2010). The internet has made it possible to express opinions on products all over the web. Accumulatively, this is called user-generated content (UGC). Consumers can now verify opinions from all over the world instead of only ask relatives and friends. Besides, smartphone manufactures can now connect the dots by looking at the past consumer experience to achieve success in the future. An approach, Steve Jobs already referred to at his Stanford University Commencement speech in 2005, as something that made all the difference in his life. Therefore, this relatively new source of data is very important to better understand products and consumer needs (Agarwal, 2020).

To be a market-leading company you need to recognize what the customer thinks of a product and how it qualifies the product. There are a lot of ways to classify product experience and to estimate the underlying reason for a certain classification. Pang and Lee (2005) explained that product rating is suitable as a class label. This is because the proxy represents the author's evaluation on a multi-point scale. These scales can be different in size (e.g. one to five stars ratings). Since these multi-point scale class labels represent various dimensions of a customer product experience, the underlying motivation of such ratings captures an accurate and broad evaluation of a product. Especially, when machine learning allows us to efficiently extract information from a big amount of review data (Liu et al., 2010).

Given this, the important question for companies that compete in the smartphone industry is what drives the certain product rating of a smartphone. Tirunillai and Tellis (2012) already showed in their research the relation between rating change and the impact on different stock performances like return, risk, and trading volume. In this thesis, the underlying reasons behind a product rating will be examined and arranged into specific business recommendations. This will provide valuable insides for companies in this industry when producing new smartphones and bringing them to market. Therefore, the following research question will be investigated:

Research question: *What drives the product rating of a smartphone?*

The examination of this research question contributes to current research in various ways. Conventional research mainly addresses parts of this research area. This is supported by the key relevant findings of table 1 that will be discussed in section 2. In this thesis, we will broaden the insights by researching smartphones in specific. Various proven effective methods will be used to deeply address the consumer's experience of smartphones. On top of that, a scarcely researched area is added to go beyond ephemeral sentiments.

This thesis will follow the subsequent structure. First, the relevant literature will be discussed. Based on this discussion, the research question can be divided into several topics and hypotheses can be made. The data and methodology is discussed in sections 3 and 4, after which the empirical results are discussed in section 5. The final section includes the conclusion and limitations of this study. Furthermore, recommendations based on the limitations are provided in the final section.

2 Theoretical framework

In this section, the relevant literature regarding the research question will be discussed. Section 2.1 will provide an overview of the effect of product rating on the sales of a product and the influences behind a product rating. These insights will provide the methodological framework used in this thesis. Thereafter, the current literature about the different aspects of this framework will be examined. Based on these findings, the research question can be divided into different hypotheses which will be tested in this thesis.

2.1 Factors impacting product rating and sales

Before we dive further into the smartphone industry specifically, a general understanding of product ratings and the methods used to approach the underlying factors is needed. Sun (2012) described that consumers seek opinions and experiences on all sort of channels before they are convinced into buying. According to Kee (2008), who evaluated an online survey of Forrester Research, 64% of the respondents want to see user ratings and reviews on the e-commerce websites they visit. Practitioners and scholars have shown that this need is only likely to increase over time (Chen, Fay, & Wang, 2011). This strongly stresses the importance for manufacturers to examine consumer reviews and their product rating since it influences future conversions. An overview of existing literature about the effect of product rating on sales and a variety of methods to excess the driving factors behind product rating is given in table 1.

This overview shows a significant influence of the product rating on the sales of the corresponding product. The sign can be positive or negative based on the value of the product rating. Not only did existing literature had common ground regarding the sign of the effect, but they also found the negativity bias. This implies that the negative product ratings have a more significant sales impact than the positively rated product (Chen, Wang, & Xie, 2011).

The diverse effects on sales within the scope of a product rating, motivates for a deeper understanding of what drives these product ratings. Hu et al. (2014) already introduced a deeper understanding by examining the sentiment of the reviews. The underlying emotions of a review do seem to have a significant effect on sales and product rating. Since they estimated a significant relationship between the sentiment score, the product rating, and sales. In this research, we will therefore further approach the relationship between the sentiment and the product rating in the smartphone industry to address our research question. Floyd et al. (2014) have also found that the nature of the product influences product rating and company performance. de Albornoz et al. (2011) supported this by showing quantitatively that the predicting accuracy of product ratings increases when the opinions on hotel features are taken into account. The product features and their influences will therefore be further discussed in the next section. Lastly, among others, McAuley & Leskovec (2013) emphasize the need for interpretable textual output when addressing the product rating. The key findings represented in table 1 show evidence for an increasing accuracy of rating classification by modeling the latent dimensions of review text. In addition, the output allows for specific business recommendations. The latter is important since we want to contribute to current research by providing useful recommendations for the smartphone industry.

Based on the above, three different research areas are used in this thesis. We will approach the research question by looking at the feature effect, review sentiments, and in-depth interpretation of the review text using Latent Direct Allocation (LDA). Therefore the following sections of the literature review will focus on these areas to formulate hypotheses for the smartphone industry specifically.

Table 1: Existing literature review table

Contributers	**(Main) Methodology**	**Dataset**	**Key (relevant) Findings**
Chen et al. (2011)	A natural experiment	Collection of digital camera info from amazon.com	The provided results showed that negative product ratings have a more significant sales impact than the products that were rated positive.
Cui et al. (2012)	A fixed effect multiple regression model	Panel data of 332 new products from amazon.com	This research confirms the negativity bias. In addition, it shows that the product rating has a higher influence on the new product sales of search goods then experience goods.
Chevalier et al. (2006)	Difference-in-Difference analysis	Sample of 3587 books collected from Amazon and B&N	An improvement of product rating results in higher sales. Consumers value reading reviews closely above summary statistics.
Hu et al. (2014)	Multiple equation model, sentiment mining	Sample of 4000 books collected from Amazon	Sentiments within a review have a significant direct effect on sales and product rating. Stresses the importance of sentiments since it provides a more context-specific explanation of the reviewers experiences.
Floyd et al. (2014)	A Meta-analysis	Approximately 400 articles	A Significant high effect of product rating on sales elasticity ($E_s = .78$) has been found. They also found that the nature of products influences the effect that online product reviews has on retailer performance.
de Albornoz et al. (2011)	Sentiment mining, Logistic, SVM and FT	1500 hotel reviews collected from booking.com	Showed that the predicting accuracy of product rating increases when the opinions on hotel features are taken into account.
McAuley & Leskovec (2013)	Latent factor recommender, LDA	>35m reviews extracted from various platforms	Evaluated based on a HFT, it was shown that topics within reviews give highly interpretable output. The nature of LDA allows therefore for new product recommendations.
Luo (2019)	LDA, GRU-CNN sentiment classification	two datasets, >4k user text data of 2 chinese platforms	The combined analysis of topic modelling and sentiment classification showed that the LDA effectively improved the accuracy of text sentiment classification.
Xu et al. (2020)	Aspect-Based sentiment analysis	Panel data of 295.628 product reviews of the Alibaba group	The Aspect-Based sentiment model improves the comprehensiveness and the prediction accuracy on sales. The improvement is over a baseline model that did not includes sentiment values.
Liang et al. (2015)	Multifacet sentiment analysis (MFSA), online word of mouth (eWOM)	Panel data on > 80 iOS apps collected from appanie.com	It was found that consumer textual reviews show a significant influence on app sales rankings. Based on the MFSA that was used, convincing support was found to understand eWOM by a sentiment analysis.

Notes: The key findings of these papers provide the theoretical background that results in the analysis framework for this thesis. The analysis framework is used to solve the research question.

2.2 Feature effect of smartphones

As introduced in section 2.1, we will approach the research question in three different ways. In this section, we will look at the current literature regarding smartphone features. Based on this, we can examine which feature variables to include in our models and hypothesize on the effect they have on product rating.

Cecere, Corrocher, and Battaglia (2015) studied the evolution in hardware components of all smartphones released between 2004 and 2013. Their goal was to investigate whether there has been developed a dominant design over the years. This research can therefore provide insides in which features are considered important to consumers by the manufacturing companies. Cecere et al. (2015) looked at the smartphone market both in terms of technological innovations and in terms of industrial dynamics. Their empirical analysis found that the number of new products has increased substantially over time and the technical features have improved as well as new features have been introduced to smartphones. This continuous introduction of new features both on hardware and on software sides indicates a relation between consumer preference and smartphone features which suggests looking at both these categories of features in this paper. Although there was no dominant design found due to different innovation strategies by the leading companies the research does outline certain features that have been improved a lot and therefore considered important by the manufacturing companies. The features that are important for innovation are operating software, size, weight, touchscreen quality, camera, connections, and internet connectivity.

Han and Cho (2016) aimed to extract relevant features of a modern-day smartphone out of the technological evolution of smartphones. Smartphone manufacturers launched new devices with a great variety of technical features since the arrival of the first iPhone. Understanding the evolution of these features is essential for all the stakeholders in the smartphone industry. They addressed the problem by looking at the specifications of the smartphones and how it developed over time with the corresponding success factor. In that way, it is possible to identify which features of a smartphone are important. Han and Cho (2016) provided as an example that the enhanced web browser and touchscreen technology are the main attributes of Apple's early success. In addition to the features already examined by Cecere et al. (2015), the number of cores and random excess memory are found to be important. Through this, several important smartphone features are added to the data frame which values can be found in table 2.

The current discussed literature on smartphone features also states that the manufacturers adjust new products towards consumer preferences. This is supported by Jean (2017), who showed evidence that smartphone manufacturers that were competitive in terms of customers also had strong innovation activities. Therefore, we hypothesize that the development of the features that happened over time have a positive effect on product rating.

Hypothesis 1: *The improvement of smartphone features have a positive effect on product rating.*

2.3 Sentiment analysis on product rating

Previous research has shown the interaction between review sentiments and product rating (Hu et al., 2014; de Albornoz et al., 2011). Given this interaction, we have chosen to use sentiment analysis to research the driving factors behind smartphone ratings. This section will further elaborate on how sentiment analyses are used in current literature. This way, hypotheses can be made on how sentiment mining will provide answers to our research question.

As already mentioned in the introduction, the availability of user-generated content increases the need for methods that can extract opinions from UGC. The consumer opinions about products and their competitors are valuable information when evaluating products and services. It is unmanageable for companies to manually analyze context data and organize them into usable forms. Therefore, automated techniques are needed to summarize these embedded opinions into practical information. Sentiment analysis approaches this need by computational study the underlying emotion, sentiments, and opinions expressed in text (Liu et al., 2010). These analyses can be applied to many different kinds of texts including customer reviews (Pang Lee, 2009; Liu Zhang, 2012). There are a variety of applications to study the sentiment of a review text and examine the relation with product rating.

Alessia et al. (2015) addressed that the classification of sentiments can be done with different criteria. Deriving the polarity score and machine learning techniques are methods to classify review text data based on sentiment. The polarity score represents a text to be positive, negative, or neutral on different classification levels. These different levels represent the sentiment of a total document, sentence or a specific entity of the review (Fang & Zhan, 2015). Barbosa et al. (2015) studied the accuracy of predicting ratings based on sentiment analysis. They used over one million reviews and their corresponding ratings from TripAdvisor. The classification was based on the polarity score. Supportive evidence was found for a positive correlation between the sentiment classifier and the actual ratings. This indicates that the sentiment, based on polarity score, is a good method to estimate the driving factors behind product rating. As Barbosa et al. (2015) propose, the polarity sentiment classification can also be used to further examine what features of product results in a certain product rating.

Lak and Turetken (2014) confirmed the relation between polarity score and ratings by comparing the calculated sentiments scores with the actual ratings. The authors used chi-square analyses and two tail bivariate correlation analyses to compare the sentiment scores with the corresponding rating. A combination of panel data from different platforms and different products were used. The results show that the sentiment score can be used as a substitute for the ratings. So according to Lak and Turetken (2014) the polarity score has explanatory value to product rating. In our research, the polarity score will be used as a classification metric. This will form the basis of deeper text mining. It is therefore important to establish the relationship between polarity score and ratings within our data. Based on research done, as described above, we can formulate the first part of hypothesis 2 as follows:

Hypothesis 2a: *The polarity score of smartphone reviews is positively correlated with the product rating.*

The sentiment of a review can be captured by emotions classifications (Mohammad, 2016) . Ortony et al. (1990) already argued that all emotions are valenced, either positive or negative. This implies there could be a relation between the emotions that are represented by a review and the rating. Garcia and Schweitzer (2011) researched this subject by extracting emotions from review data with a lexicon-based classifier. On a panel dataset with many different products from *amazon.com*, they found a relationship between emotions and how a consumer rates a product. Positive emotions have a strong bias towards positive classified reviews. Negative emotions, on the other hand, are more equally distributed as regards rating. This indicates that reviewers tend to also be positive when they express negative emotion.

Singla et al. (2017) used the NRC dictionary to derive the underlying emotions within the review text. Compared to this thesis, they also investigated consumer reviews of smartphones. Based on the polarity score, the emotions can be classified as positive or negative (Singla et al., 2017). Joy, trust, anticipation, and surprise are classified as positive. On the contrary are disgust, fear, sadness, and anger classified as negative. The authors found supportive evidence for the correlation between positive emotions and product ratings. Insights about the effect of negative classified emotions on product ratings were not found. The following two hypotheses can be prepared about review emotions and the interaction with smartphone ratings:

Hypothesis 2b: *The positive emotions (i.e. joy, trust, anticipation, and surprise) have a positive effect on the product rating of smartphones.*

Hypothesis 2c: *The negative emotions (i.e. disgust, fear, sadness and anger) have no direct effect on the product rating of smartphones.*

As already recommended earlier by Barbosa et al. (2015), the sentiment classification by polarity score allows to further research the effect of product features on rating classification. Extracting features using a sentiment analysis is becoming an active and important area of research (Asghar et al., 2014). Feature-based sentiment analysis provides understanding about the costumer's opinion on a certain object in a review (Liu et al., 2010).

A certain approach was used by Wang et al. (2014). They suggested a feature-based vector model and a novel weighting algorithm for sentiment analysis of Chinese product reviews. The results outperformed conventional research on classification accuracy. This addresses the importance of smartphone features when investigating the driving factors behind the product rating. Models where features are incorporated are better at predicting classes. Therefore, the features that are related to a certain class of sentiment and rating provide insightful information for manufactures.

Kangale et al. (2016) addresses the importance of sentiments about product features, for both consumers and manufacturers. Consumers can use it as guidance on which product to buy. Where the manufactures can use it to learn from their costumers and improve their products. The dataset, used by Kangale et al. (2016), contains review and feature data for two smartphone types. They found through sentiment mining that some feature words appear more in positive sentences and others more in negative sentences. Based on a classification model, they looked at the effect of smartphone features on the product rating and sentiment. Since previous literature convinces on the effect of sentiment on product rating, their analysis can be used to generalize a possible effect of features on product rating.

We looked at the effect Kangale et al. (2016) found for smartphone features on the product rating. They found that customers tend to talk more negative about *camera*, *battery capacity* and *audio*. Therefore, the last part of the second hypothesis will concentrate on the classification of smartphone features within our sentiment analysis.

Hypothesis 2d: *Words related to camera, batteries and audio/sound will occur more often within negative classified sentences.*

On the other hand, they found that reviewers tend to talk more positive about the *design*, *software*, *display* and *weight*. Therefore, the following hypothesis can be formulated.

Hypothesis 2e: *Words related to design, software, display and weight will occur more often within positive classified sentences.*

Based on section 2.3, we will address the effect of review sentiments on smartphone ratings with three subtopics. First, the effect of the polarity score on ratings will be quantified. Thereafter, the interaction between the underlying emotions within the review text and smartphone rating will be examined. Lastly, the sentiment scores will be used to extract which smartphone features are classified as positive or negative.

2.4 Topics within consumer reviews

Topic modeling is an advanced technique that focuses on the latent dimension of review texts. Limited research is done on the interaction between topic modeling and product rating of smartphones. Therefore, this section will identify how topic modeling is currently used concerning consumer reviews. Given this and augmented with substantiated assumptions, one can hypothesize on the interaction with smartphone ratings.

Luo (2019) examined that LDA improves the accuracy of text sentiment classification. This indicates a relation between found topics and consumer sentiments. Adding to that, Section 2.1 and 2.3 already elaborated on the existing interaction between sentiments and product rating. Therefore, we can assume that findings on the effect of topic modeling on sentiments also interact with the product rating. This thesis aims to find the deeper meaning behind product ratings in the smartphone industry. This provides valuable information for the innovation of new smartphones. The tools used, therefore, need to extract specific representation of this deeper meaning, classified by ratings.

Onan et al. (2016) argued that making classification only by sentiments indeed falls short concerning implementable information. Hereby is meant, that companies can use the subjects found by analysis for innovative purposes. These sentiment text classifications suffer from the high dimensional feature space and feature sparsity problems. As a solution, Onan et al., (2016) propose the LDA to represent collections of discrete data. This paper examines the performance of LDA in-text classifications. A dataset of over 4000 reviews and several classification methods is used to estimate the performance. Onan et al. (2016) found that LDA-based representation utilized in the predicting models is a viable method to represent text collections in a compact yet efficient way. This can be explained by looking at the function of topic modeling.

Topic modeling is explained by Blei et al. (2009) as the method of defining the underlying semantic structure of a text using a Bayesian hierarchical analysis. The accelerating availability of digital content is unstructured (Blei, 2012). The algorithms of topic modeling can structure these content into clear output. The techniques have been successfully incorporated in many applications by discovering topics through entity relationships and automatic classification (Daud et al., 2010). Through this, topic modeling covers a wider dimension of the data which logically results in better classification accuracy. There are many methods of topic modeling but the latent Dirichlet allocation is one of the most widely used methods (Alghamdi & Alfalqi, 2015).

The substantiated assumption was made that topic modeling interacts with product rating. This will be called the interaction assumption. In section 2.3 was found by Kangale et al. (2016) that consumers tend to talk more or less about certain aspects based on classification. In our research, the dataset will be split into product rating categories so that the LDA can be applied to these different categories. Based on the findings of Kangale et al. (2016) and the interaction assumption we can set up the following hypotheses:

Hypothesis 3a: *Words related to camera, batteries and audio/sound will occur more often within the topics of negative rating categories.*

Hypothesis 3b: *Words related to design, software, display and weight will occur more often within within the topics of positive rating categories.*

The hypotheses will examined by finding latent topic words within different rating categories. This aims to discover which subjects consumers talk about per rating category. Since the LDA proved to be a viable method to classify text collections, the output topics are highly relevant to understand why consumers classified the product as they did. Section 4.4 will further elaborate upon the execution and details of the methodology.

3 Data

In this section, the data used for the analyses will be discussed. This is done in the following sequence. First, information is provided on how the data is collected and its corresponding sources. Thereafter, an outlay is given on how the fundamental dataset is supplemented with product feature information. Lastly, an explanation is given on how the raw data is processed and modified so that it can be used for text analytics. The fundamental data used is acquired from kaggle[‖].

3.1 Base

At the basis of the final dataset used for the analyses are six phone user review datasets web scraped from online platforms all over the world. These sets combined are narrowed towards English written reviews only since we mainly use text analytics to approach the research question. Based on a quantity threshold regarding reliability of the results, the set is brought down to smartphone models of Apple, Samsung, and Lenovo. The sample of Samsung products contains the S3, S4, S5, S6, and S7 of the Galaxy series. The Apple models consist of the iPhone 4, 4s, 5, 5s, 6, 6s, and 7. The sample is completed with the Vibe K4 and Vibe K5 of Lenovo with a total of 121.080 reviews. The quantity threshold is set on 2500 reviews per device. This is an absolute minimum and for most devices, the data carries sufficiently more reviews. The dataset consists of the following variables:

phone_url *The domain code for the specific website page*

date *Specific date the review was written by the consumer. The window is between 01 January 2013 and 9 September 2016 with a few exceptions considered negligible.*

lang *the language in which the review was written. For this paper, we only consider reviews written in English.*

country *The reviewers country of origin. The data contains reviewers originating from Australia, Canada, Great Britain, United States, New Zealand, India, and Zambia.*

source *The channel where the review was posted. More than 90 % of the information was web scrabbed from the official domains of Samsung and Amazon. The rest comes from a wide variety of internet sources.*

score *The product rating given by the reviewer which has a range from 1 to 10.*

extract *The textual content of the review.*

author *the self-chosen name of the reviewer that is used with activities on the domain of the source.*

product *The brand and product type that is the subject of the review.*

When looking at the summary statistics it is seen that the average product rating of the total sample lays around approximately 8. This indicates that customers tend to leave a review when they had a positive experience with the smartphone (Singla et al., 2017). This tendency needs to be taken into account when the deeper insights behind the texts are investigated. 78.360 of the reviewers are from the USA, 27.848 from India, and 12.706 from Great Britain. The rest of the remaining reviews are relatively small amounts dividend among the other countries mentioned in the variable explanation.

[‖]Kaggle is a subsidiary of Google LLC. Kaggle is an online community of data scientists and machine learning practitioners that allows for data publishing for non-commercial purposes. *https://www.kaggle.com/datasets*

3.2 Features

Apart from the analysis of the extract, the research question requires an investigation of the interaction between specific characteristics of each model and product rating. Therefore given each product type, the corresponding features are added to the data.

The Samsung product features are retrieved from Android Planet (2020), Apple's from iPhoned (2020) and Lenovo's from GSM Arena (2020). The different values for the features added are represented in table 1. The back camera, battery capacity, and weight are programmed as integer values. Screen size, front camera, and RAM size as numbers. The resolution and chipset as categorical variables. For further analyses, the resolution is calculated based on the multiplication belonging to the category so that the exact effect of the pixels on the product rating can be estimated.

Table 2: Feature values

Back camera	**Screensize**	**Battery capacity**	**Front camera**
8	4.7 inch	2100 mAH	1.9 Megapixels
12	5 inch	2550 mAH	2.1 megapixels
13	5.1 inch	2600 mAH	5 megapixels
16		2800 mAH	
		3000 mAH	
Weight	**Chipset**	**RAM size**	**Resolution**
130 grams	single-core	0.5 GB	640x960 pixels
133 grams	dual-core	1 GB	720x1280 pixels
138 grams	quad-core	2 GB	1136x640 pixels
145 grams	octa-core	3 GB	1080x1920 pixels
152 grams		4 GB	1440x2560 pixels

Notes: The variety of values added to the datafram based on their corresponding smartphone product type.

3.3 Preprocessing

In this paper, text analytics plays an important role. Therefore, the original webs crabbed extract needs to be modified so that the data allows for the use of these methods. First, all letters are changed to lower case letters. Thereafter, are among other things excess spacing, expressive characters, time indications, and other signs that have no explanatory value removed. For the sentiment analyses and LDA, the stop words, frequent and infrequent words are removed. This is done by the join and anti join function of the *dplyr package* in R. The LDA method requires words to be stemmed. This means that only the meaning full part of the word will be used. A stem is the root or roots of a word, together with any derivational affixes, to which inflectional affixes are added (Crystal, 1985). After these modifications, some extract entries had no content anymore and were removed.

As explained by Konrad (2016), topic modeling with the LDA requires to build a Document Term Matrix (DTM) that contains the number of term (i.e. word) occurrences per review. A row of the DTM represents a review and the columns represent the whole vocabulary, i.e. the set of words that are mentioned at least once in all reviews. The last mentioned modifications are done by the DocumentTermMatrix function of the *tm package*.

The preprocessing as described allows us to test the hypotheses. The methodology used will be discussed in the next section. The final dataset consists of smartphone ratings, reviews, and other review information. Also, smartphone feature data is added and the text corpus is modified into structured data, which is also stored into a DTM for topic modeling.

4 Methodology

This part describes the methodology of the performed analyses in this thesis. The outcome aims to give further insights into the driving factors behind product ratings in the smartphone industry. The implemented methods are discussed in more technical detail after which the results are presented in the next section.

4.1 Regression analysis

The first hypotheses make statements about the interaction between product rating, polarity score, feature characteristics, and emotions. To test these hypotheses, a model is required that can capture these relationships. As the purpose of this research is to identify what drives the interaction, and not necessarily building excellent performing predictive models, it makes sense to start with more interpretable and relatively restrictive models. A simple and commonly used approach is by using linear regression. The results of multiple linear regression models with a different subset of predictors will help to explain the product rating of smartphones.

4.1.1 The model

A multiple linear regression model is an extension of the simple model by containing more than one predictor. This model is a linear approach to model the relationship between a scalar response and explanatory variables. The scalar response is also known as the dependent variable and the predictors as an independent. The relationships between the dependent and independent variables are modeled using linear predictor functions. The unknown parameters in these models are estimated using the data. These estimated parameters represent the relationship between the response and explanatory variables of the used data. Given a dataset $\{y_i, x_{i1}, ..., x_{ip}\}_{i=1}^{n}$ of n rows, a linear regression model assumes a linear relationship between the dependent variable and its predictors. The parameters are calculated by the following formula for the p-vector of regressors $\mathbf{x}$:

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \epsilon_i = \mathbf{x}_i^T \beta + \epsilon_i, \qquad i = 1, ..., n \tag{1}$$

Within this formula, $\mathbf{x}_i^T \beta$ represents the inner product between vectors $\mathbf{x}_i$ and β. The parameters of the multiple regression model are estimated by minimizing the sum of the squared residual. This method is called ordinary least squares which produces estimates under certain assumptions. The first assumption is that the Data Generating Process is described by a linear model. However, the regressors of vector $\mathbf{X}$ must all be linearly independent of each other. The next assumption states that the error term is completely independent from the independent variables $E[\epsilon|X] = 0$. The last assumptions are that the error term has the same variance in each observation $Var(\epsilon_i|\mathbf{X}) = \sigma_\epsilon^2$ and that the errors are uncorrelated between observations $Cov(\epsilon_i, \epsilon_j) = 0$.

4.1.2 Predictors

To test hypothesis 1, hypothesis 2a, hypothesis 2b, and hypothesis 2c different multiple regression models are derived. The models are distinguished by the predictors that are included in the model. The feature characteristics represented in table 2 are used to test hypothesis 1. The next section explains how the polarity scores and the emotions within the reviews are extracted. These values are incorporated into different regression models to answer the first part of hypothesis 2.

4.2 Sentiment analysis

Sentiment analysis is referred to as the process of determining the opinions of authors about specific entities (Feldman, 2013). As described in among others the introduction, sentiment analysis becomes an important topic of research. This is because of the rise of UGC, natural language processing (NLP), and machine learning techniques that allow analyzing sentiments on a large scale. By analyzing and assembling the information into resumptive content, companies can learn from a wide audience of customers for new products. Sentiment analysis by the use of text analytics represents the process of assigning weighted sentiment scores to the entities, topics, themes, and categories within different document levels. The different levels of classification are document-level, sentence-level, or aspect-level. The document-level classification aims to find a sentiment score for the entire review, whereas the sentence-and aspect level classification represents the sentiment score for both levels respectively.

To assign a sentiment score to a word various dictionaries can be used (Kwartler, 2017). In this thesis, we use the tidytext package that provides access to several sentiment lexicons. These lexicons contain many English words and are assigned a proxy for the sentiment. The lexicons used are based on unigrams. To calculate the polarity score, the bing lexicon (Liu, 2012) is used to categorizes words into positive and negative categories. The technical details of this process are provided in section 4.2.1. Besides the bing lexicon, the NRC lexicon (Mohammad & Turney, 2010) is used to categorize words in a binary fashion of emotions. This way, the emotions can be incorporated in the multiple regression models. The NRC dictionary categorizes words into ten emotions; positive, negative, joy, anticipation, surprise, trust, fear, disgust, sadness, and anger. The lexicon provides every word with the label positive or negative and one label from the other eight specific emotions.

To test hypothesis 2d and hypothesis 2e, we will use a combination of the polarity score and word counts. Based on the polarity score, we can obtain word frequencies within positive and negative classified documents. To perform an even more in-depth analysis of the review sentiments, we will examine word frequencies on a sentence classification level and examine the differences in word frequencies.

4.2.1 Polarity function

The polarity function by the *qdap package* of R programming is used to derive the polarity scores. The bing dictionary allows words to be classified as positive or negative (Hu and Liu, 2004). The equation used by the algorithm to assign value to the polarity of each sentence, first uses the bing dictionary to tag polarized words. The preprocessing, as described in section 3.3, is necessary so that the function excludes meaningless parts of a text regarding sentiments. For example, punctuation has no explanatory value over the sentiment of the reviewer.

The function separates each review ($r_i = s_1, s_2, ...s_n$) composed of sentences, into element sentences ($s_i, j = w_1, w_2, ...w_m$) where w are the words within sentences. Each sentence (s_j) is on its turn divided into a an ordered bag of words. The exact location of the word is represented by $w_{i,j,k}$, where i, j, k stands for the place of the word, place of the sentence and number of the review respectively. The words $w_{i,j,k}$ in each sentence are searched and compared to the bing dictionary of polarized words. Positive (w_i, j, k^+) words are tagged +1 (x_i^+) and negative (w_i, j, k^-) words are tagged -1 (x_i^-). Words can also be tagged neutral (x_i^0), negator ($x_i\neg$) or amplifier ($x_i^\uparrow$). Neutral words do not have a value in the equation but do affect word count ($\mathbf{n}$) in equation 2 and 3. These two equations represent how these elements interact to result in a polarity score (δ) between -1 and 1.

$$\delta_i = \frac{\sum(x_i^0, \quad x_i^{\uparrow} + x_i^{+} \cdot (-1)^{\sum(x_i \neg)}, \quad x_i^{\uparrow} + x_i^{-} \cdot (-1)^{\sum(x_i \neg)})}{n} \tag{2}$$

$$x_i^{\uparrow} = \frac{1}{n-1} \tag{3}$$

As described in section 4.2, the polarity scores are used to classify reviews on different levels. Given these classifications, we can derive the word frequencies within these levels. Since the conventional literature acknowledges the interaction between sentiment score and product rating, the word frequencies provide information about frequently used words given the sign of product rating classification. In this thesis, the acknowledgment of the interaction is also tested with hypothesis 2a.

4.3 Latent Dirichlet Allocation

In this thesis, LDA is used to detect topics within different product rating categories. With this, results can be obtained about which topics belong to a certain product rating classification. This information provides a multidimensional answer to what drives the product rating of smartphones. The theoretical framework, in section 2, already showed that classification models based on LDA have satisfactory predictive performance. These topics derived by LDA contain a set of words. Therefore, the topics show what consumers talk about given a certain classification.

The LDA model is a generative probabilistic topic model where each review is represented as a random mixture of latent topics and each topic is represented as a distribution over a secure set of words. Based on the observed data, the model aims to identify the underlying topic structure. For each review in the corpus, words are generated through two steps. First, a randomly chosen distribution over the topics is assigned. Given this distribution, a topic is randomly selected from the distribution of topics for each word of the document (Blei, 2012). With LDA, words are discrete data from an indexed vocabulary by $\{1, ..., V\}$ and a chain of N words by $w = (w_1, w_2, ..., w_n)$. A corpus is a group of M documents indicated by $D = \{w_1, w_2, ..., w_M\}$. Figure 1 shows a summary of the generative process of LDA.

Figure 1: The generative process of LDA (Blei et al., 2003)

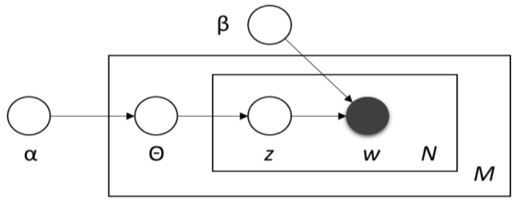
For each document w in a corpus D:

1. Choose $N \sim$ Poisson (ξ)
2. Choose $\Theta \sim$ Dir (α)
3. For each of the N words w_n:
   (a) Choose a topic $z_n \sim$ Multinomial (Θ)

Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

The LDA process can be modeled by a three-level Bayesian graphical model. As can be seen in figure 2, random variables are represented by nodes, and possible dependencies between variables are represented by edges. In this figure, α refers to the Dirichlet parameter, Θ refers to document-level topic variables, z refers to the topic assignment for each word, where w refers to the observed word and β refers to the topics. As can be seen from the three layers of figure 2, α and β parameters are sampled once while generating the corpus. The sampling of the document-level topic variables occurs for each document and word-level variables for each word of the document (Blei & Jordan, 2003).

Figure 2: The graphical representation of LDA



The LDA generation process shows the joint distribution of random variables. Equation 4 shows how the probability density function of a k-dimensional Dirichlet random variable is computed. Equation 2 estimated the joint distribution of a topic mixture and equation 3 derives the probability of a corpus (Blei & Jordan, 2003).

$$p(\Theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \Theta_1^{\alpha_1 - 1} ... \quad \Theta_k^{\alpha_k - 1} \tag{4}$$

$$p(\Theta, z, w|\alpha, \beta) = p(\Theta|\alpha) \prod_{n=1}^{N} p(z_n|\Theta) p(w_n|z_n, \beta) \tag{5}$$

$$p(D|\alpha, \beta) = \prod_{d=1}^{M} \int p(\Theta_d|\alpha) (\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\Theta_d)(p(w_{dn}|z_{dn}, \beta)) d\Theta_d \tag{6}$$

For a specific document, calculating the posterior distribution of hidden variables is an essential task in LDA. The calculation of the exact inference of the posterior distribution of the hidden variables is a challenging problem. In our model, this problem is approximated with the Gibbs sampling algorithm.

4.3.1 Hyperparameter tuning

Hyperparameter tuning, also known as automatic model tuning, approaches the best version of the model by running a series of hyperparameters on the dataset. Based on an objective metric, the best model is chosen by testing a range of values for each chosen tunable hyperparameter. In this thesis, a combination of cross-validation and the perplexity score is used to obtain the best model. The hyperparameter tuning approaches the best values of α and the number of topic k-dimensions. The methodology is applied to all four smartphone rating categories subsets. The objective metric used is the perplexity measure. Perplexity is a measure of how well a probability model predicts a sample. The perplexity of a set of word $(W_d|\alpha_d)$ where d is a member of corpus D, is represented by the following formula (Yarnguy & Kanarkard, 2018; Pleplé, 2013).

$$perplexity(W_d|\ \Phi, \alpha_d) = exp[-\frac{\sum_d log\ \ p(W_d|\ \Phi, \alpha_d)}{N_d}] \tag{7}$$

The condition p represents the likelihood given topic model Φ and α the hyperparameter value which is also entered by topic k-dimensions. To test the predictive performance using perplexity, the resampling method of cross-validation is used. The perplexity function of the package *Topicmodels* returns a single number. The lowest number indicates the best model given the decreasing function of the log-likelihood. The same package of R provides the LDA function to execute the Latent Dirichlet Allocation.

5 Results

In this section, the results following the methodology will be discussed. The results are reviewed based on the hypotheses formulated in section 2. First, the regression results of the three estimated models are discussed. Thereafter, there will be further elaborated on the outcome of the sentiment analysis on different levels. Lastly, the most occurring topic words by LDA per product rating category will be presented to approach the multidimensional reasoning of product ratings. In the next section, we will conclude with the findings and provide specific recommendations for the smartphone industry.

5.1 Regression results

To test hypotheses 1, 2a, 2b, and 2c, we employ three different linear models. Model 1 consists of smartphone features that can be found in table 2. Model 2 is an extended version of model 1 by also including the estimate of the polarity score effect. Model 3 is the most extensive model where the sentiment polarity score is replaced by emotion scores for all reviews. When addressing the regression diagnostics, no significant correlation between the predictors and the errors were found. Besides, the error terms are uncorrelated with each other. To assess the hazard of multicollinearity, the Variance Inflation Factor (VIF) is used. For all three models, the square of the generalized VIF of battery capacity was excessively high, with a value above 5. This indicates a risk of multicollinearity by this predictor. Therefore, the variable of battery capacity is manually removed from all three models. The regression results are summarized in table 3.

Table 3: Regression models

	Model 1		**Model 2**		**Model 3**	
Variable	*Estimate*	*S.E.*	*Estimate*	*S.E.*	*Estimate*	*S.E.*
(Intercept)	8.847***	(0.149)	7.114***	(0.127)	8.197***	(0.142)
Backcamera	-0.025***	(0.005)	-0.017***	(0.004)	-0.021***	(0.005)
Screensize	-0.393***	(0.038)	-0.214***	(0.032)	-0.277***	(0.036)
Frontcamera	0.231***	(0.010)	0.157***	(0.009)	0.193***	(0.010)
Weight	-0.002*	(0.001)	-2.25e^{-4}	(0.001)	-0.002	(0.000)
Chipsetdual-core	-	-	-	-	-	
Chipsetocta-core	-1.268***	(0.047)	-0.961***	(0.040)	-0.981***	(0.046)
Chipsetquad-core	-0.034	(0.035)	-0.060*	(0.030)	-0.029	(0.034)
Chipsetsingle-core	-0.424***	(0.062)	-0.287***	(0.053)	-0.408***	(0.059)
RAMsize	-0.013	(0.027)	0.002	(0.023)	0.005	(0.027)
pixels	7.44e^{-7} ∗ ∗∗	(2.26e^{-8})	4.93e^{-7} ∗ ∗∗	(1.94e^{-8})	5.62e^{-7} ∗ ∗∗	(2.19e^{-8})
polarity			2.693***	(0.012)		
anger					-0.755***	(0.023)
anticipation					-0.138***	(0.018)
disgust					-0.586***	(0.070)
fear					-0.977***	(0.039)
joy					1.268***	(0.019)
surprise					-0.774***	(0.029)
trust					0.256***	(0.026)
N	120.939		120.672		120.932	
Adjusted R^2	0.049		0.314		0.137	
p-value F-statistic	2.2e-16		2.2e-16		2.2e-16	

Notes: Significance: *p<0.05 **p<0.01 *** p<0.001

Hypothesis 1 requires an understanding of the improvement in smartphone features. The different feature values can be found in table 2. All variables except *weight* are considered to be improved when the value increases. *Weight* on the other hand, is assumed to be improved when the value declines. For the variables that are included in all three models, we find almost the same estimate values and significance levels. The exceptions are the variables *weight* and *chipsetquad-core* where the latter represents the smartphones processor. *Weight* only has a small negative effect at the lowest significance level in model 1 where this is only the case in model 2 for *chipsetquad-core*. Therefore, we assume that the effect of these predictors on product rating is negligible. The estimates of *backcamera*, *screensize* and *chipsetocta-core* contradict with our hypothesis, since our models predict a significant negative effect on product rating. The chipset variables are categorical and their effects are relative to reference variable *chipsetdual-core*. Given the sequence of performance, we hypothesized that *chipsetocta-core* should have a positive relationship with product rating. As said, our models indicate that this cannot be accepted. the effect of *chipsetsingle-core* does follow the hypothesis since a single-core is a less performing chipset then a dual-core. The other software variable *RAMsize* has no significant effect in any model. The predictors *pixels* and *frontcamera* follow the hypothesis since a significant positive relation with product rating is estimated.

To test hypothesis 2a, we look at the effect of the polarity score. Based on a great variety of literature we could already assume that a polarity score has a positive effect on product rating. The expected effect is confirmed in model 2. In this thesis, this is an important assumption which we can now confirm. Namely, the sentiment analysis in the next section uses the polarity score to classify the consumer experience. Therefore we can relate these findings to the influence on product rating which we aim to achieve. Model 3 also contains the effect of different emotions on product rating. The emotion scores are assigned to reviews by the NRC dictionary that classifies words to emotions. The coefficients and significance levels reveal that *anger*, *anticipation*, *disgust*, *fear*, and *surprise* have a negative effect on product rating. Therefore we can reject hypothesis 2c, since we hypothesized that there will be no direct effect. The effect we found for the emotions is endorsed by the nature of the emotions. Except for the estimate of *surprise* and *anticipation*, that represent positive emotions, is against expectations. Our model indicates that these particular feelings of a customer has a negative effect on product rating. This could imply that customers do not like to be surprised and want to know exactly what they buy in advance. The estimates of *joy* and *trust* have a positive effect based on their estimates and significance level. The effect of the emotion *trust* follows with the effect of *surprise*. Consumers seem to value the fact that they can rely on a manufacturer of its smartphones. These interactions are in line with hypothesis 2b, which is partly rejected due to the negative effect of *anticipation* and *surprise*.

5.2 Sentiment analysis

The sentiment analysis performed in this section is used to examine the underlying sentiments of a consumer review. Based on the estimated polarity score on different classification levels, we have derived word frequency plots that reflect the sentiment. The different levels by the sentiment analysis are review, sentence, and differences. The differences level means that based on word frequencies between negative and positive classified sentences, the differences between words are estimated. This level is used to test hypotheses 2d and 2e since this level most closely represents the actual sentiment for opposed classification.

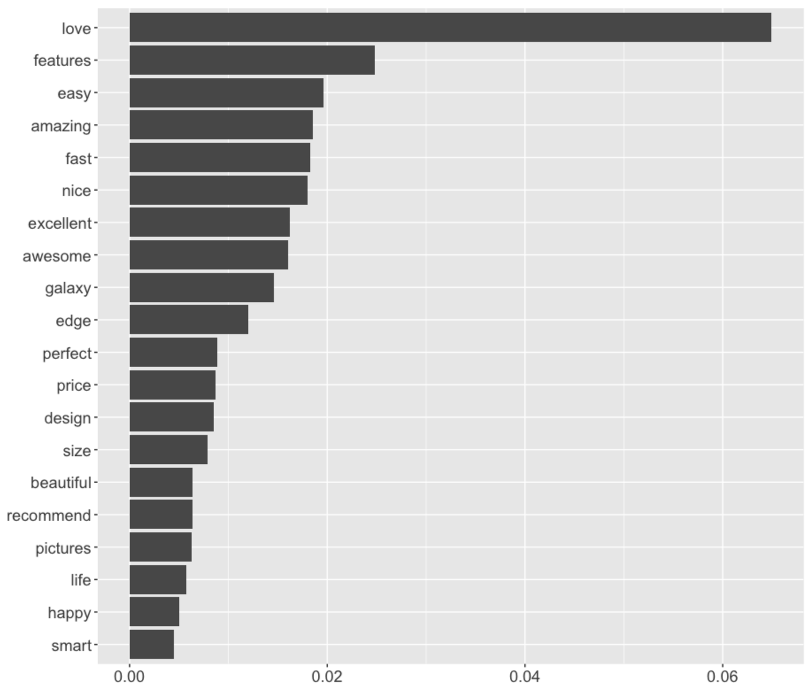
Appendix A shows the polarity score distribution categorized by happy and unhappy reviewers. As mentioned in section 3, the data used contains outstanding more positive rated reviews than negative. This is supported by Singla et al. (2017), who found evidence that customers tend to leave a review when they had a positive experience with the smartphone. The polarity score distribution of our dataset confirms this phenomenon. First, we start by analyzing the word frequencies on a document level which can be found in figure 6 of Appendix B. The reviewers discourse the *battery*, *camera* and *screen* in both positive and negative classified sentences. Besides the similarities, there are also interesting differences. Consumers talk mainly positive about the *price*, *design*, *features*, and *speed*.

So these are factors that consumers value when classifying a smartphone as positive on a document level. The chatter about *charging*, *chargers*, *money*, *apps*, *card*, and smartphone heat mainly occurs in negative classified reviews. particularly, charging, apps, and the sim card are considered fundamentals of the smartphone product. This indicates that smartphone users tend to negatively review a smartphone when the fundamental functions of a smartphone do not perform properly.

The following word frequencies discussed are based on a sentence level classification. The word frequency plots can be found in figure 7 of Appendix B. This approach provides more insights about what consumers chatter about when classifying a smartphone since it looks at a review on a deeper level. The word similarities between positive and negative correspond to the document level classification of figure 6 (Appendix B). This also largely applies to the differences between positive and negative classified sentences. The chatter does differ on this level with words as *easy* occurring more in a positive sentence where words as *touch* and time indications occur more in a negative sentence. Noticeable is also that by positive classified levels, the chatter is more about words that represent the positive feeling. Whereby negative classified levels, words that point out specific aspects of the smartphone experience occur more.

To test hypotheses 2d and 2e, both figure 3 and 4 are used. The interpretation of these figures requires a clear explanation of the definition of specific positive and negative words. We have defined a word as specifically positive or negative when a word appears more at one classification side based on the differences. Figure 3 represents the differences in word frequencies for specific positive words and figure 4 for specific negative words. The classification of the document and sentence level does have some limitations. One could argue some standard words show up and that negative words appear more often by definition. To solve this issue and increase the informative value, we look at the differences in word frequencies.

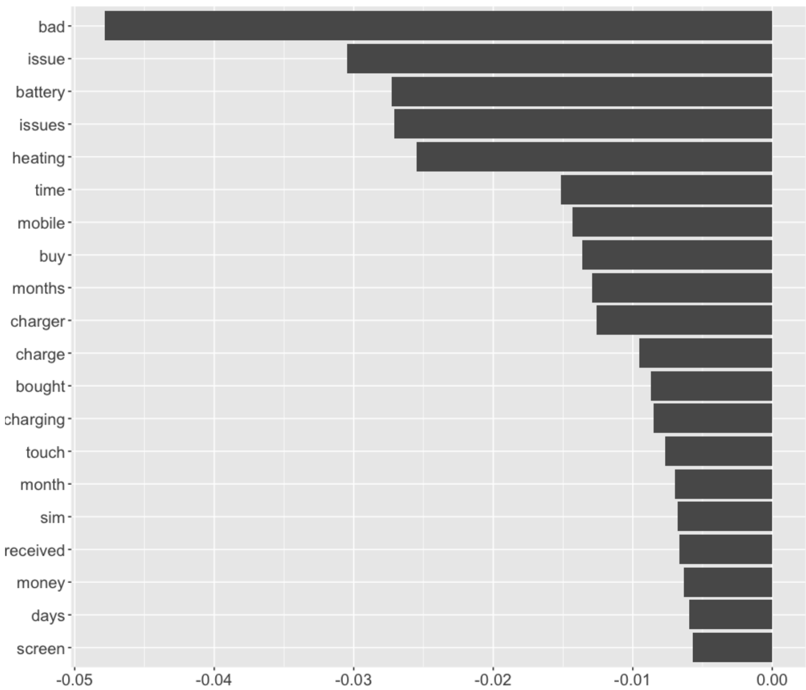
Figure 3: Specific positive words: difference in word frequency (pos-neg)



Regarding specific positive words, we have hypothesized that words related to design, software, screen, and weight will occur more often. According to figure 3, we can not fully reject the hypothesis. We do find that reviewers use the words *design* and *size* more often, which is in line with the hypothesis. By using the words *easy*, *fast* and *pictures*, the reviewers show the preference for the functionality of the smartphone. In addition, specific positive words are *price* and *features.* In contradiction with hypothesis 2d, it seems that consumers are more likely to use specific positive words relating to hardware than software.

Next, we will zoom in on the specific negative words in figure 4. This figure provides information on the actual motives to classify a smartphone as negative. Based on the literature review in section 2, we expected words related to camera, batteries, and sound to occur more in negative classified sentences based on differences in word frequencies. The findings of figure 4 show that we can not fully reject the hypothesis 2e. We do find that words related to batteries, namely *battery*, *heating* and *charging*, are specific negative words. The combination of the specific negative words *touch* and *screen* reveals that the quality of the display is more used in negative classifications. This is remarkable because we expected it to be used more as specific positive words. Lastly, *sim* is classified as a specific negative word. The disappointing performance of the functionality of a sim card appears to be an important indicator to rate the smartphone experience as negative.

Figure 4: Specific negative words: difference in word frequency (pos-neg)



The execution of the sentiment analysis in section 5.2 provides the driving factors behind smartphone product rating based on word frequencies. Companies in the industry need to follow up on these sorts of findings. As is confirmed by the literature, UGC becomes more important and available to both companies and consumers. This means that it will get even easier to leave product experiences behind on online platforms. Therefore the analysis of online chatters approaches consumer experiences very closely. In the next section, we will further look into these driving factors by looking at a multidimensional space through topic modeling.

5.3 LDA

The last part of the results will present the topics within the product rating categories and overall striking aspects of the topics. The product rating is categorized based on four categories. We divided the ratings into high, above average, below average, and low. The numerical meaning behind these categories can be found in table 4. First, we will discuss the hyperparameter tuning executed to obtain a better model. Thereafter, we will further elaborate on the topics by LDA.

5.3.1 Hyperparameter tuning

As discussed earlier in this thesis, we have chosen the α and number of topic k-dimensions as hyperparameters. The tuning is done by a combination of cross-validation and the subsequent calculation of the perplexity score. As described, the lowest perplexity score represents the probability model with the best predicting performance. Using cross-validation, the lowest perplexity score based on different topic k-dimensions is calculated. After the optimal amount of topics is found, the same method approaches the best α. We executed this for all four categories which results can be found in table 4.

Table 4: Hyperparameter values of the LDA models

Product rating category	α	**K-topics**
High ($\geq$8)	0.1	30
Above average ($\geq$6 & $<$8)	0.1	20
Below average ($\geq$4 & $<$6)	0.1	30
Low ($<$4)	0.1	20

Notes: The parameters values are chosen based on cross-validation and perplexity score.

The represented hyperparameters values will be used for the models that extract the topic words within a product rating category in the following section.

5.3.2 Latent topic words

To test hypothesis 3a and 3b, a clarification is needed on the definition of topic classification. We define the rating category *low* and *below average* as negative rating categories. The categories *above average* and *high* are defined as positive rating categories. First, we will evaluate the negative rating categories. As explained in section 3.3, the LDA method requires words to be stemmed and this means that only the meaningfull part of the word will be used. Table 5 shows us that the topics with the highest average topic probability are *return*, *product* and *startup* in rating category low.

Table 5: Topic words - rating category: low

Return	**Product**	**Startup**
charg	month	sim
month	galaxi	card
week	bui	unlock
start	purcha	mobil
return	replac	activ
issu	price	servic
box	product	bought
take	dai	bit
note	love	fine
receive	app	start

Notes: The topic words with the highest average topic probability.

The reviewers within this category tend to talk mostly about returning the product, the product itself, or the problems when starting up and activating the device. Beyond the figure, we find that the word *disappointed* (Appendix C) is commonly used in a lot of topics. This indicates that the device falls short of expectations which is in line with the estimates of the emotions *trust* and *surprise* in table 3. Table 6 shows us the topics found in the rating category: below average. These topics contain words about *battery*, *product* and *primary features*. Specifically, the disappointing performance of the charger, call options, the functionality of the sim-card, and sound appears to be a reason for a negative classification.

Table 6: Topic words - rating category: below average

Battery	**Product**	**Primary features**
batteri	batteri	charg
charg	life	charger
charger	purcha	replac
time	run	monei
condit	note	fine
sound	disappoint	call
bad	excel	condit
life	switch	sim
purcha	app	sound
month	edg	absolut

Notes: The topic words with the highest average topic probability.

Other noticeable subjects found in topics in this rating category are about the lack of meeting up to expectations and heating (Appendix C). In line with hypothesis 3a, we did found that batteries and sound are mentioned in the negative rating topics. Related words to the quality of the camera were not found.

As second, we will evaluate the positive rating categories. Table 7 shows us that the topic words with the highest average topic probability are *battery*, *reception* and *primary features* in the above-average rating category. It has common ground with the below-average rating category which could indicate that reviewers are more pronounced in extreme categories.

Table 7: Topic words - rating category: above average

Battery	**Reception**	**Primary features**
batteri	charger	sim
dai	charg	card
charg	month	expect
life	replac	mobil
time	return	time
issu	bad	updat
week	batteri	sound
condit	dai	charger
user	receiv	edg
love	purcha	return

Notes: The topic words with the highest average topic probability.

Again, we see that consumers do value to receive what they expected and chatter about this subject to justify their product rating. Within other topics (Appendix C), it strikes that words as *screen*, *touch*, *video* and *picture* arises. This stresses the importance of good functioning displays and cameras for a positive rating. The topics with the highest topic probability in the highest rating category can be found in table 8. The word combinations provide the topics with the headers *hardware*, *first experience* and *reception*.

Table 8: Topic words - rating category: high

Hardware	**First experience**	**Reception**
price	excel	product
mobil	perfect	start
bui	price	box
product	product	bui
month	fast	purcha
hand	week	servic
app	absolut	smart
camera	activ	unlock
design	amaz	absolut
light	android	activ

Notes: The topic words with the highest average topic probability.

The way a device is brought to the consumers appears to be an important aspect of a positive experience. The hardware and appearance also stimulate a positive experience. This is supported by words that occur multiple times in the topics (Appendix C) such as *light* and *design*. Besides these words, the word *fast* also appears frequently. Hypothesis 3b expected words related to the design, software, and display to be present in the positive rating category. Based on the finding by LDA, we can argue that words related to design and display indeed occur in the positive product rating. However, we do not find evidence that software is a subject of chatter in the positive smartphone rating category.

6 Conclusion

This final section will provide a summary of the results from the analysis performed in this thesis. After summarizing the main results, we will further elaborate on the contribution to academics and the implication for managers in the smartphone industry. At last, the limitations of this research will be given which results in recommendations for further research.

6.1 Discussion

In this thesis, we investigated what drives the consumer experience of smartphones by answering the following research question: "What drives the product rating of a smartphone?". We have chosen to take the product rating of a smartphone as the objective to measure consumer experience. The product rating directly represents the consumer experience since the consumers individually rate the smartphone on the internet themselves.

To answer the research question, we first established the positive interaction between sales and ratings by examining previous literature. This interaction addresses the importance of this research for smartphone manufacturers. We used an extensive literature review to construct the methodology framework used for answering the research question. This framework is based on the proven interaction by previous literature between different types of methods and product rating. The different methods are used to answer the hypotheses which contribute to the conclusion on the main research question.

To answer hypothesis 1, about the relationship between smartphone characteristics improvements and product rating, a significant negative effect was found between the variables *backcamera*, *screensize*, *chipsetsingle-core*, *chipsetocta-core* and product rating. Only the predictors *pixels* and *frontcamera* have a significant positive relation with product rating. Therefore, hypothesis 1 can be partly rejected. Model 2 found a significant positive association between the sentiment polarity score and rating. This interaction was evaluated to test hypothesis 2a, which we can accept. The positive relation between the sentiment score and rating is fundamental in this thesis. Namely, the polarity score is used to classify sentiments in the sentiment analysis. This way, the results of the sentiment analysis can be connected to the product ratings of the devices. Continuing with hypothesis 2b and 2c, which aims to establish the interaction between emotions and product rating, a significant negative interaction between *anger*, *anticipation*, *disgust*, *fear*, *surprise* and rating was found. The variables *joy* and *trust* positively interact with product rating. Thus, we reject hypothesis 2c and partly reject hypothesis 2b.

Proceeding with hypothesis 2d and 2e, stating that certain words occur more often in positive classified sentences and some words more often in negative classified sentences, evidence was found based on a sentiment analysis. Words as *design*, *size*, *easy*, *fast*, *pictures*, *price* and *features* appear more often in positive classified sentences. As specific negative words, we found among others *battery*, *heating*, *charging* and *sim*. The combination of the specific negative words *touch* and *screen* reveals that the quality of the display is more used in negative classifications. Thus, both hypotheses can be partly rejected.

At last, testing hypothesis 3a and 3b, resulted in specific topics per rating category. In the negative rating categories, chatters are about returning the device, heating, battery, the smartphone itself, or the problems when starting up and activating the device. A remarkable subject in the negative rating categories is also that consumers are dissatisfied because the product did not meet up to expectations. In the positive rating categories, chatters are mainly about the reception, the first experience, hardware, the touchscreen, video, and pictures. Thus, both the hypotheses cannot be fully accepted. In the next section, the findings mentioned above will be used to construct business recommendations for managers in the smartphone industry.

6.2 Contribution and implication

This study uniquely combines a variety of data analysis methods to research the driving factors behind product ratings for the smartphone industry specifically. The research contributes to current literature since it uses the proven validity of methods by other researchers and shows the practical implication these methods can have. Section 2, showed that methods as sentiment mining and topic modeling are highly suitable for classification problems. Nevertheless, research that goes beyond analytical valence and actually uses these methods to deeply interpret consumer experience, is scarce. The framework proposed in this research can be used for other industries and future research can expand this framework, resulting in more extensive recommendations.

The business recommendation for managers in the smartphone industry can be divided into four types of implications. The technical development of new devices, the communication to consumers, the trends that need to be actively monitored, and the reception experience. First, future development of smartphones should be focused on the quality of the touchscreen, the front camera, and the battery. Future devices need to be fast, easy to use and there need to be no complications when activating a new device. As second, correct communication with consumers appears to be very important. Managers need to make sure that consumers will not be surprised according to the expectations set by the smartphone manufacturers. This way, the trust in the company by consumers increases which is valued highly. The trends that need to be actively monitored and implemented in future devices are design, size, price, and features. These aspects are trend-sensitive and need to move with consumer preferences. Price is not necessarily trend-sensitive but the willingness-to-pay does change over time. Lastly, consumers have shown the importance of the first experience. This includes the first experience of the product and the reception. Therefore, manufactures should improve among others packaging and all other sorts of first experience elements so that the consumers feel satisfied with the product from the start.

6.3 Limitations and further research

Based on the limitations of this thesis, recommendations for further research can be given. First, this research relies on the assumption that the product rating interacts with the sales performance of a smartphone manufacturer. Although this was indicated by prior research, a reliable test specifically for this industry is missing. To increase the confidence in this assumption, we suggest to excess the interaction between sales and product rating for the smartphone industry specifically.

Secondly, our data is globally gathered. Since consumer preferences differ around the globe, the findings from such a sample are not necessarily useful for specific countries. Therefore, we recommend further researchers to execute this research for more specific datasets. This will provide a broader understanding of consumer preferences characterized by geography or other demographics.

Lastly, the text analytics executed in this thesis is mainly based on unsupervised learning. The next step in establishing the interaction between consumer chatters and product rating is supervised learning. The analysis performed in this study has shown the relevance of this research area. By using supervised learning, future research will provide further insights into the actual predicting value of certain areas of text analytics on product rating in the smartphone industry.

References

Agarwal, M. (2020). Importance of user generated content as a part of social media marketing that drives customer s brand awareness and purchase intentions. *The International journal of analytical and experimental modal analysis*, *12*(2).

Alessia, D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, *125*(3).

Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, *6*(1).

Andrew, O. (2018). *The history and evolution of the smartphone: 1992-2018.* Retrieved from `https://www.textrequest.com/blog/history-evolution-smartphone/`

Android Planet. (2020). *Samsung telefoons en tablets vergelijken.* Retrieved from `https://www.androidplanet.nl/samsung/`

Asghar, M. Z., Khan, A., Ahmad, S., & Kundi, F. M. (2014). A review of feature extraction in sentiment analysis. *Journal of Basic and Applied Scientific Research*, *4*(3), 181–186.

Barbosa, R. R. L., Sánchez-Alonso, S., & Sicilia-Urban, M. A. (2015). Evaluating hotels rating prediction based on sentiment analysis services. *Aslib Journal of Information Management*.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. *Text mining: classification, clustering, and applications*, *10*(71), 34.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Cecere, G., Corrocher, N., & Battaglia, R. D. (2015). Innovation and competition in the smartphone industry: Is there a dominant design? *Telecommunications Policy*, *39*(3-4), 162–175.

Chen, Y., Fay, S., & Wang, Q. (2011). The role of marketing in social media: How online consumer reviews evolve. *Journal of interactive marketing*, *25*(2), 85–94.

Chen, Y., Wang, Q., & Xie, J. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of marketing research*, *48*(2), 238–254.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, *43*(3), 345–354.

Crystal, D. (1985). *A dictionary of linguistics and phonetics, 2nd edn updated and enlarged.* Oxford, Basil Blackwell in association with André Deutsch.

Cui, G., Lui, H.-K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, *17*(1), 39–58.

Daud, A., Li, J., Zhou, L., & Muhammad, F. (2010). Knowledge discovery through directed probabilistic topic models: a survey. *Frontiers of computer science in China*, *4*(2), 280–301.

De Albornoz, J. C., Plaza, L., Gervás, P., & Díaz, A. (2011). A joint model of feature mining and sentiment analysis for product review rating. In *European conference on information retrieval* (pp. 55–66).

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, *2*(1), 5.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82–89.

Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How online product reviews affect retail sales: A meta-analysis. *Journal of Retailing*, *90*(2), 217–232.

Garcia, D., & Schweitzer, F. (2011). Emotions in product reviews–empirics and models. In *2011 ieee third international conference on privacy, security, risk and trust and 2011 ieee third international conference on social computing* (pp. 483–488).

GSM Arena. (2020). *Lenovo phones.* Retrieved from `https://www.gsmarena.com/lenovo-phones-73.php`

Han, Q., & Cho, D. (2016). Characterizing the technological evolution of smartphones: insights from performance benchmarks. In *Proceedings of the 18th annual international conference on electronic commerce: e-commerce in smart connected world* (pp. 1–8).

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Aaai* (Vol. 4, pp. 755–760).

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision support systems*, *57*, 42–53.

iPhoned. (2020). *iphone koopgids.* Retrieved from `https://www.iphoned.nl/kopen/`

Jean, A. T. (2017). An analysis of the impact of innovation on the competitiveness of smartphone manufacturers. *International Journal of Management Research and Reviews*, *7*(9), 872–892.

Kangale, A., Kumar, S. K., Naeem, M. A., Williams, M., & Tiwari, M. K. (2016). Mining consumer reviews to generate ratings of different product attributes while producing feature-based review-summary. *International Journal of Systems Science*, *47*(13), 3272–3286.

Kee, T. (2008). Majority of online shoppers check at least four reviews before buying. *Online Media Daily*(February 19).

Konrad, M. (2016). *Creating a sparse document term matrix for topic modeling via lda.* Retrieved from `https://datascience.blog.wzb.eu/2016/06/17/creating-a-sparse-document-term-matrix-for-topic-modeling-via-lda/`

Kwartler, T. (2017). *Text mining in practice with r.* John Wiley & Sons.

Lak, P., & Turetken, O. (2014). Star ratings versus sentiment analysis–a comparison of explicit and implicit measures of opinions. In *2014 47th hawaii international conference on system sciences* (pp. 796–805).

Liang, T.-P., Li, X., Yang, C.-T., & Wang, M. (2015). What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach. *International Journal of Electronic Commerce*, *20*(2), 236–260.

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

Liu, B., et al. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, *2*(2010), 627–666.

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

Luo, L.-x. (2019). Network text sentiment analysis method combining lda text representation and gru-cnn. *Personal and Ubiquitous Computing*, *23*(3-4), 405–412.

McAuley, J., & Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th acm conference on recommender systems* (pp. 165–172).

Mohammad, S., & Turney, P. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the naacl hlt 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26–34).

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement* (pp. 201–237). Elsevier.

Onan, A., Korukoglu, S., & Bulut, H. (2016). Lda-based topic modelling in text sentiment classification: An empirical analysis. *Int. J. Comput. Linguistics Appl.*, *7*(1), 101–119.

Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions.* Cambridge university press.

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *arXiv preprint cs/0506075*.

Pang, B., & Lee, L. (2009). Opinion mining and sentiment analysis. *Comput. Linguist*, *35*(2), 311–312.

Park, S., Cho, K., & Lee, B. G. (2014). What makes smartphone users satisfied with the mobile instant messenger?: Social presence, flow, and self-disclosure. *International Journal of Multimedia and Ubiquitous Engineering*, *9*(11), 315–324.

Pleplé, Q. (2013). *Perplexity to evaluate topic models.* Retrieved from `http://qpleple.com/perplexity-to-evaluate-topic-models/`

Singla, Z., Randhawa, S., & Jain, S. (2017). Statistical and sentiment analysis of consumer product reviews. In *2017 8th international conference on computing, communication and networking technologies (icccnt)* (pp. 1–6).

Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, *58*(4), 696–707.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? dynamics of user-generated content and stock performance. *Marketing Science*, *31*(2), 198–215.

Tocci, M. (2019). *History and evolution of smartphones.* Retrieved from `https://simpletexting.com/where-have-we-come-since-the-first-smartphone/`

Wang, H., Liu, L., Song, W., & Lu, J. (2014). Feature-based sentiment analysis approach for product reviews. *Journal of software*, *9*(2), 274–279.

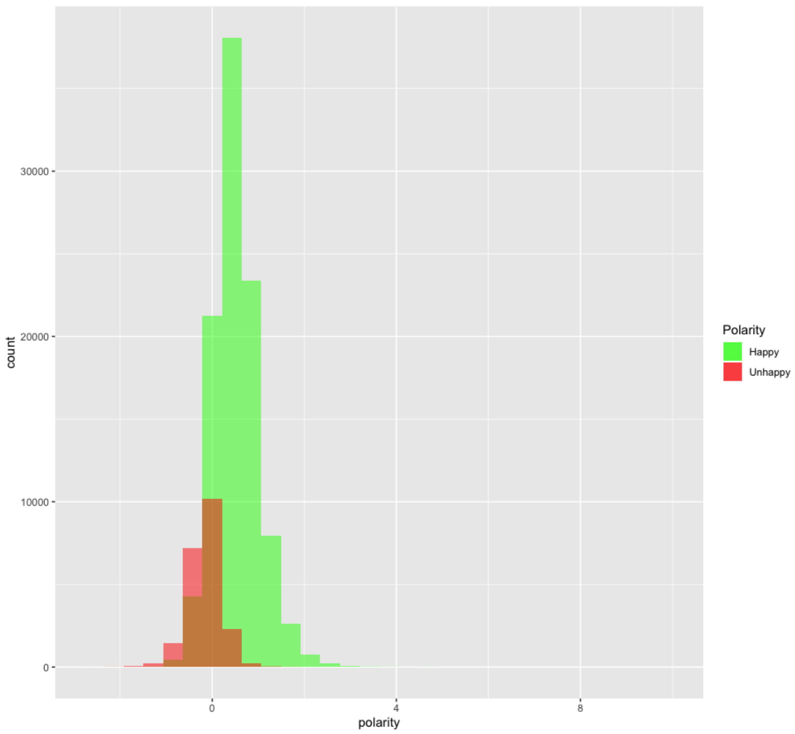
Xu, J., Tang, B., wen Liu, X., & Xiong, F. (2020). Automatic aspect-based sentiment analysis (aabsa) from customer reviews. In *Affcon@aaai.*

Yarnguy, T., & Kanarkard, W. (2018). Tuning latent dirichlet allocation parameters using ant colony optimization. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-9), 21–24.

Appendix

Appendix A

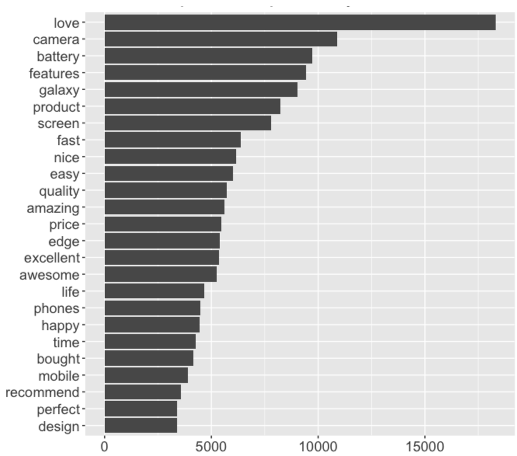
Figure 5: Histogram of sentiment split of Happy/Unhappy on polarity



Appendix B

Figure 6: Word frequencies

(a) Positively classified reviews



(b) Negatively classified reviews

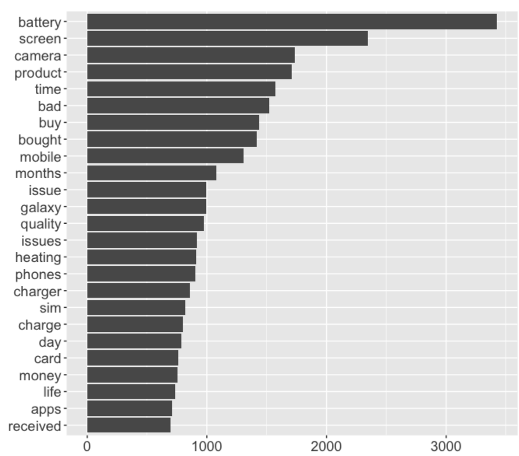
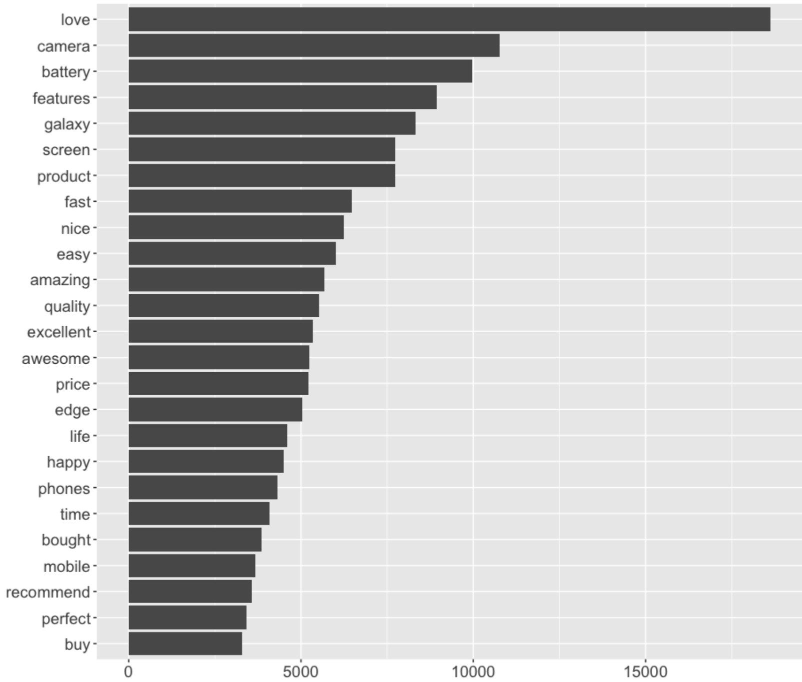
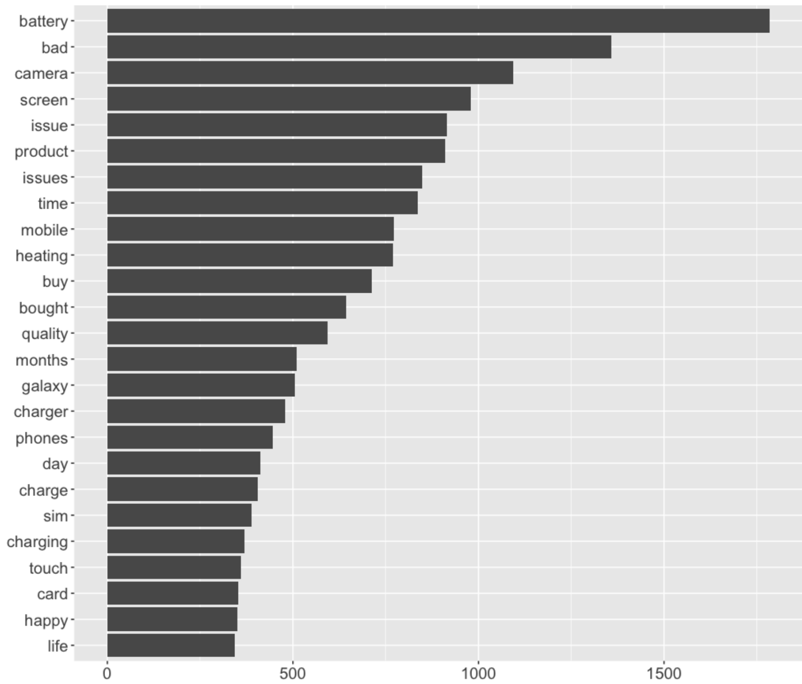


Figure 7: Word frequencies

(a) Positively classified sentences



(b) Negatively classified sentences



Appendix C

Table 9: LDA topic words - rating category: low

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
batteri	charg	mobil	bui	batteri	galaxi	charger	app	call	month
charg	month	picture	brand	charg	bought	charg	time	time	galaxi
month	week	receiv	recommend	purchase	model	bui	start	receiv	bui
disappoint	start	bought	unlock	time	receiv	bought	iphon	dai	purcha
bought	return	monei	monei	service	smartphon	week	screen	sound	replac
switch	issu	camera	heat	happi	box	replace	month	take	price
love	box	disappoint	receiv	bad	charg	happi	purcha	app	product
return	take	box	bad	screen	disappoint	review	android	featur	dai
user	note	condit	hand	active	unlock	condit	galaxi	smart	love
fast	receive	take	disappoint	iphon	return	brand	fine	easi	app
Topic 11	**Topic 12**	**Topic 13**	**Topic 14**	**Topic 15**	**Topic 16**	**Topic 17**	**Topic 18**	**Topic 19**	**Topic 20**
return	sim	dai	purcha	unlock	product	monei	set	batteri	screen
week	card	camera	screen	bad	model	time	devic	life	bought
activ	unlock	expect	galaxi	receiv	issu	bui	updat	dai	review
happi	mobil	nice	activ	fine	recommend	activ	water	bui	touch
bad	activ	feel	batteri	color	charg	call	disappoint	bought	time
fast	servic	devic	dai	bought	qualiti	disappoint	worth	purcha	box
receiv	bought	start	expect	feature	receive	run	batteri	monei	function
call	bit	displai	brand	love	monei	bad	recommend	activ	light
video	fine	bad	perform	return	sound	lot	switch	condit	user
absolut	start	picture	servic	servic	excel	smart	user	disappoint	price

Notes: This table contains the topic words derived by the LDA analysis for smartphones that were rated 1,2 or 3.

Table 10: LDA topic words - rating category: below average

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
camera	batteri	month	batteri	start	batteri	qualiti	function	iphon	product
nice	charg	product	llife	month	time	service	smart	recommend	pictur
pictur	charger	easi	prcha	purcha	bad	galaxi	featur	displai	receiv
memori	time	bought	run	fine	happi	issu	light	switch	bad
take	condit	bui	note	love	charg	dai	lot	bad	color
batteri	sound	mobil	disappoint	charge	run	amaz	purcha	call	time
bought	bad	purcha	excel	pictur	mobil	note	run	beauti	bui
design	life	absolut	switch	review	return	perform	bit	featur	card
size	purcha	app	app	upgrad	take	sound	dai	perfect	charg
devic	month	expect	edg	android	box	take	excel	return	condit

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
screen	dai	charg	week	bad	galaxi	sim	unlock	call	dai
touch	expect	charger	return	app	issu	card	call	time	call
bit	set	replac	time	brand	bui	call	receiv	activ	week
perform	issu	monei	fast	disappoint	android	happi	box	user	issu
qualiti	return	fine	disappoint	happi	bad	bit	fine	nice	absolut
easi	bought	call	displai	start	product	recommend	monei	easi	brand
receiv	box	condit	love	life	mobil	absolut	issu	featur	android
android	love	sim	set	activ	review	activ	light	perfect	bad
app	monei	sound	unlock	displai	updat	amaz	qualiti	feel	displai
batteri	nice	absolut	camera	memori	appl	android	sound	price	easi

Topic 21	Topic 22	Topic 23	Topic 24	Topic 25	Topic 26	Topic 27	Topic 28	Topic 29	Topic 30
updat	bui	bought	bought	month	price	featur	screen	galaxi	time
app	week	mobil	upgrad	batteri	bui	camera	devic	time	camera
bought	love	review	galaxi	replac	mobil	charg	box	activ	lot
featur	model	sim	month	hand	batteri	condit	issu	lot	dai
heat	smartphon	card	disappoint	switch	call	bad	fast	love	nice
color	condit	bui	start	disappoint	happi	set	fine	month	disappoint
receiv	featur	perfect	charg	life	disappoint	week	receiv	disappoint	qualiti
charg	screen	product	heat	app	awesom	android	return	updat	iphon
price	beauti	upgrad	model	note	heat	feel	size	dai	love
replac	function	app	switch	purcha	lot	happi	color	issu	monei

Notes: This table contains the topic words derived by the LDA analysis for smartphones that were rated 4 or 5.

Table 11: LDA topic words - rating category: above average

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
camera	app	galaxi	product	love	screen	model	call	batteri	unlock
pictur	memori	disappoint	fine	featur	touch	bui	featur	dai	product
disappoint	iphon	call	price	size	set	time	size	charg	disappoint
bit	lot	function	bui	price	receiv	design	fine	life	easi
issu	run	week	switch	camera	satisfi	recommend	screen	time	android
monei	smartphon	sound	android	screen	light	servic	easi	issu	box
worth	charg	upgrad	bit	brand	purcha	bought	lot	week	excel
function	disappoint	activ	purcha	smart	review	mobil	perform	condit	happi
qualiti	expect	box	bad	awesom	smart	color	excel	user	hand
update	month	recommend	month	lot	upgrad	happi	perfect	love	purcha
Topic 11	**Topic 12**	**Topic 13**	**Topic 14**	**Topic 15**	**Topic 16**	**Topic 17**	**Topic 18**	**Topic 19**	**Topic 20**
batteri	batteri	nice	devic	charger	galaxi	sim	expect	bought	fast
life	charg	qualiti	purcha	charg	issu	card	displai	month	batteri
bad	love	galaxi	time	month	review	expect	lot	heat	mobil
perform	screen	easi	bought	replac	replac	mobil	mobil	video	app
app	run	screen	disappoint	return	week	time	price	take	featur
activ	bui	touch	feel	bad	love	updat	dai	hand	month
bui	replac	size	amaz	batteri	time	sound	nice	nice	easi
featur	take	receiv	bit	dai	product	charger	time	pictur	amaz
start	happi	sound	start	receiv	fast	edg	app	app	lot
bit	receiv	upgrad	bui	purcha	perform	return	fine	bad	updat

Notes: This table contains the topic words derived by the LDA analysis for smartphones that were rated 6 or 7

Table 12: LDA topic words - rating category: high

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
featur	price	excel	displai	easi	iphon	bought	perfect	featur	galaxi
screen	mobil	perfect	batteri	recommend	price	bui	galaxi	purcha	batteri
excel	bui	price	charg	android	product	month	sim	recommend	model
fine	product	product	light	nice	beauti	batteri	app	screen	design
issu	month	fast	awesom	screen	bit	devic	bui	camera	sim
time	hand	week	dai	set	call	screen	call	charg	absolut
touch	app	absolut	design	excel	displai	updat	issu	dai	activ
absolut	camera	activ	easi	purcha	life	upgrad	absolut	excel	amaz
activ	design	amaz	monei	receiv	return	worth	activ	function	android
amaz	light	android	recommend	smart	upgrad	absolut	amaz	love	app
Topic 11	**Topic 12**	**Topic 13**	**Topic 14**	**Topic 15**	**Topic 16**	**Topic 17**	**Topic 18**	**Topic 19**	**Topic 20**
charg	happi	love	bought	excel	sim	product	product	amaz	expect
start	fast	bought	upgrad	fast	batteri	start	time	galaxi	happi
awesom	fine	featur	featur	qualiti	amaz	box	heat	android	amaz
batteri	card	feel	heat	servic	card	bui	bit	perfect	function
brand	iphon	iphon	lot	camera	model	purcha	brand	time	hand
dai	price	life	love	edg	note	servic	absolut	updat	perform
displai	purcha	monei	recommend	price	perfect	smart	activ	absolut	screen
return	absolut	smart	absolut	receiv	return	unlock	amaz	activ	absolut
servic	activ	absolut	activ	sim	video	absolut	android	app	activ
time	amaz	activ	amaz	updat	absolut	activ	app	appl	android
Topic 21	**Topic 22**	**Topic 23**	**Topic 24**	**Topic 25**	**Topic 26**	**Topic 27**	**Topic 28**	**Topic 29**	**Topic 30**
displai	mobil	size	qualiti	camera	galaxi	love	android	expect	mobil
life	fast	sim	color	displai	replac	happi	awesom	fast	nice
time	perform	unlock	week	nice	servic	excel	week	call	box
excel	smartphon	card	absolut	app	amaz	disappoint	bought	android	call
featur	activ	dai	activ	pictur	expect	hand	featur	app	issu
month	bought	featur	amaz	color	satisfi	perform	issu	featur	love
nice	expect	fine	android	mobil	nice	review	replac	feel	run
receiv	lot	galaxi	app	note	receiv	absolut	absolut	light	absolut
absolut	recommend	lot	appl	take	size	activ	activ	month	activ
activ	screen	nice	awesom	absolut	absolut	amaz	amaz	perfect	amaz

Notes: This table contains the topic words derived by the LDA analysis for smartphones that were rated 8 or higher.

Appendix D

Table 13: Average topic probability - rating category: low

Topic	*Average topic probability*	**Topic**	*Average topic probability*
1	0.05289238	11	0.04852906
2	0.05526576	12	0.05963990
3	0.04510058	13	0.04680156
4	0.04882222	14	0.04715848
5	0.05117282	15	0.05194113
6	0.05278984	16	0.04806825
7	0.05223834	17	0.04478042
8	0.04915199	18	0.03947784
9	0.04939103	19	0.05253628
10	0.05557073	20	0.04867138

Table 14: Average topic probability - rating category: below average

Topic	*Average topic probability*	**Topic**	*Average topic probability*
1	0.03753149	16	0.03630395
2	0.04866343	17	0.03248477
3	0.02655349	18	0.03512069
4	0.04065559	19	0.03395384
5	0.04065820	20	0.03099415
6	0.03226299	21	0.03489987
7	0.02548092	22	0.03378494
8	0.02871683	23	0.03382397
9	0.02578584	24	0.03095483
10	0.03311034	25	0.03412637
11	0.03836665	26	0.03715757
12	0.02944500	27	0.02651661
13	0.04105395	28	0.03359516
14	0.03081215	29	0.03230196
15	0.02694088	30	0.02794356

Table 15: Average topic probability - rating category: above average

Topic	*Average topic probability*	**Topic**	*Average topic probability*
1	0.05007646	11	0.04982207
2	0.04061351	12	0.05218287
3	0.04438832	13	0.04793012
4	0.04892541	14	0.04965841
5	0.04879641	15	0.06255092
6	0.04826292	16	0.05184949

7	0.05035354	17	0.05477710
8	0.04821903	18	0.04932823
9	0.06156692	19	0.04577554
10	0.04855960	20	0.04636313

Table 16: Average topic probability - rating category: high

Topic	*Average topic probability*	**Topic**	*Average topic probability*
1	0.02587207	16	0.02976554
2	0.05001928	17	0.04104381
3	0.04329309	18	0.02606762
4	0.03719457	19	0.02493171
5	0.03564183	20	0.03012096
6	0.02929642	21	0.03070628
7	0.03080053	22	0.03425806
8	0.03029006	23	0.03680747
9	0.03597166	24	0.02417850
10	0.03359070	25	0.03595887
11	0.02964725	26	0.03761600
12	0.03086501	27	0.03347166
13	0.04048810	28	0.02852088
14	0.02996313	29	0.03795826
15	0.03499695	30	0.03066374