12-15-2021

Amazon Reviews using Sentiment Analysis

Ammar Rashed Hamdallah
arh9763@rit.edu

RIT

Amazon Reviews using Sentiment Analysis

by

Ammar Rashed Hamdallah

A Graduate Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology
RIT Dubai

(December 15, 2021)

RIT

Master of Science in Professional Studies:
Data Analytics

Graduate Capstone Approval

Student Name: Ammar Rashed Hamdallah

Capstone Title: Amazon Reviews using Sentiment Analysis

Graduate Capstone Committee:

Name: Dr. Sanjay Modak Date: December 15, 2021
Chair of committee

Name: Dr. Ioannis Karamitsos Date: December 13, 2021
Member of committee

Acknowledgments

My special heartfelt gratitude goes to the almighty God, whose grace enabled me to undertake and complete the study. I have the pleasure to thank all the people who contributed to this study in one way or the other, my supervisor Dr.Ioannis Karamitsos and Dr Sanjay Modak, for guiding me in this study, my classmates, and lastly, my family.

Abstract

The intense competition to attract and maintain customers online is compelling businesses to implement novel strategies to enhance the customer experiences. It is becoming necessary for companies to examine customer reviews on online platforms such as Amazon to understand better how customers rate their products and services. The purpose of this study is to investigate how companies can conduct sentiment analysis based on Amazon reviews to gain more insights into customer experiences. The dataset selected for this capstone consists of customer reviews and ratings from consumer reviews of Amazon products. Amazon product reviews enable a business to gain insights on customer experiences regarding specific products and services. The study will enable companies to pinpoint the reasons for positive and negative customer reviews and implement effective strategies to address them accordingly. The capstone project helps companies use sentiment analysis to understand customer experiences using Amazon reviews.

Table of Contents

List of Figures

List of Tables

Chapter 1

1.1 Background

This chapter entails the projects goals, aims and objectives , research methodology and limitations of the study. Amazon is among the largest online marketplace in the world for various products. Customers occasionally go through the products and their reviews just before they purchase a product. These reviews give them a level of information and opinion about the quality of the products they want to buy. Sometimes the reviews prove to be misleading because they are usually subjective and could not be giving the entire information about a product but just a first impression of a customer with the amazon product or service. Data from customer reviews is critical in today's information-driven business environment. Companies use customer reviews to gain meaningful insight into the consumption behaviors of customers. An in-depth analysis of customer sentiments enables a business to understand the market better and make rational decisions to address customer needs and concerns proactively. Sentiment analysis capitalizes on natural processing language statistics and text analysis to explore what the customers are saying, how they express it, and what they mean. Tweets, reviews, and comments are crucial sources of customer sentiments. The computations of sentiment analysis determine whether customer reviews are positive, negative, neutral, or mixed. According to Du et al. (2019), 91% of online shoppers read product reviews before purchasing products and services online. Product reviews play a crucial role in enhancing the customer purchasing experience. Besides, it is essential to enable a business to improve its products and services by comprehending customers' needs, preferences, and tastes. However, it is imperative to note that product reviews on Amazon platforms are vulnerable to quality control. According to Du et al. (2019), a recent study revealed that customers tend to restrict concentration on the first view reviews, irrespective of their helpfulness. The challenge is that access to extensive customer reviews makes it difficult for customers to identify useful information.

1.2 Statement of the problem

The retail industry is the backbone of the US economy. According to Statista projections, the sales revenue forecast for this industry in 2020 was $5.48 trillion (Statista.com, 2020). The retail sector creates millions of jobs annually, generates revenue, and contributes approximately 10% of the gross national product (Lim et al., 2019). The retail industry's health dramatically relies on the degree of customer satisfaction and confidence with the retailers of their preferred products and services. Advancements in technology enable customers to share their experiences by reviewing the products and services from specific retailers. Today, social media platforms and online networks allow businesses to mine genuine comments and reviews from customers worldwide. Customer reviews reveal customers' experiences regarding the prices, value, quality, customer service, ease of shopping, and more factors about what they shop online. The customer reviews are unstructured, and the sentiment analysis will help extract the sense of these unstructured texts efficiently and cost-effectively. This capstone project will study how companies can conduct sentiment analysis based on Amazon reviews to gain customer experience insights. Companies will gain more understanding of top-rated products and services, what customers value, and what they dislike using sentiment analysis. For a business to succeed in today's competitive and information-driven business environment, it is vital to understand what the customers feel about the products and services offered. A company must maintain positive reviews from customers and improve the neutral and negative reviews of the customers. For instance, most customers complain about the quality of products based on Amazon reviews, and the business must develop and implement effective strategies to enhance the quality of their offerings.

1.3 Project goals

The first goal is to get the sentiments expressed in the customer reviews and analyze the frequency of the sentiments. The second project goal is to build and train a machine learning model that can be used to classify customer reviews into two sentiments (positive or negative).

1.4 Aims and Objectives

The aim of the study is to classify customer reviews into positive or negative sentiment.

The objectives are as follows:

- To measure the intensity of the sentiments generated from the customer reviews
- To analyze the association between customer reviews concerning different amazon products.

1.5 Research Methodology

Sentiment analysis methodology will be used in this research. The method is also referred to as opinion mining. The process relies on machine learning (ML) algorithms and natural language processing (NLP) to determine the emotions behind online reviews. The research will focus on analyzing the sentiments on Amazon product reviews. To be specific, the platform has a feature where customers can review products on a five-scale rating. The ratings from 5 to 1 represent very positive, positive, neutral, negative, and very negative experiences (Chauhan et al., 2020). Five stars mean the customer is very positive about the product or service. One star indicates that the customer's experience is very negative.

Sentiment analysis follows the certain steps to collect, process, transform and convert raw data to simple corpus for sentiment classification. These steps are essential in achieving the best results in sentiment analysis.

STEP 1: A data set of Amazon product reviews within a specific period will be collected. Once the data set is collected, the next step is text preparation. This is the process in which the data extracted is filtered before analysis (Shankhdhar, 2019).

STEP 2: In this stage, non-textual content is identified and eliminated. Irrelevant content is also identified and removed from the data set. The objective is to ensure that only the required data set is analyzed.

STEP 3: Sentiment detection. The purpose of sentiment detection is to review each comment for subjectivity. Sentences or comments with objective expressions are eliminated, while those with subjective expressions are retained further.

STEP 4: Sentiment classification. This stage's data is categorized into two major groups, positive and negative (Shankhdhar, 2019). The data can be classified more to include like and dislike categories.

STEP 5: The final step is to convert the results of the analysis into meaningful information. The text results will be displayed on bar charts, pie charts, and line charts. The graphs will visualize the results for a better understanding of the trends in online product reviews.

The data tools used in this capstone project are: R Programming Languages, Amazon APIs, and Visualization tools (Tableau) will be used in this research.

1.6 Limitations of the Study

- The study was limited when it comes to dealing with reviews from sarcastic customers who usually use ironic language, this would make it hard for the model to learn the correct sentiment elicited.

- Another limitation arose in the subjectivity of the customers, this would hinder correct sentiment extraction from the reviews because the subjectiveness changes from person to person, and some people would be very irrational when submitting their reviews.
- People’s opinions change over time and this could be due to mood change or even interaction with other products and customers, hence when collecting data the period is a factor in affecting the sentiments in a review.

Chapter 2 – Literature Review

This chapter highlights some of the previous studies in the field, citing existing gaps in knowledge on the how business benefit from the findings of sentiment analysis on customers.

2.1 Theoretical background of Online reviews

According to Sharma, Chakraborti, and Jha (2019), online shopping has gained global popularity over the past decade. This dramatic trend's primary reasons include the ease of internet access, availability of smartphones, increased awareness of e-commerce, and increased access to online shopping applications. Online shopping platforms such as Amazon enable customers to shop with convenience, save time, and get their products delivered at home within the shortest time possible. Competitive businesses prefer e-commerce over physical stores for the potential to reach more customers around the world (Sharma, Chakraborti, and Jha, 2019). E-commerce platforms also enable the company to save on enormous costs by setting up online stores. However, the customer demands on platforms such as Amazon on price and quality are drastically changing. The majority of customers are focusing more attention on high-quality and affordable products and services. In that regard, businesses must analyze the sentiments of customers to address these demands effectively.

Meire et al. (2019) argued that social media is a trending platform for markets to drive customer engagement today. Customer engagement initiatives enable businesses to boost their emotional bonds with customers located in different parts of the world. Product reviews also shape customer engagement (Nandal, Tanwar, and Pruthi, 2020). An analysis of the product reviews enables the business to gain insights into how customers feel about their products. According to Schoenmueller, Netzer, and Stahl (2020), online consumer reviews play a critical role in shaping customers' purchasing decisions online. A business must concentrate on analyzing and understanding these reviews to succeed in today's business environment.

Presently, businesses conduct sentiment analysis to enhance their competitiveness in the market. Karamitsos et al. (2019) argued that sentiment analysis enables companies to understand their products and services' views and experiences. As a result, companies can effectively design their marketing campaigns (Dong et al., 2017). Sentiment analysis also allows modern businesses to maximize the word-of-mouth marketing strategy. Competitive companies have gone the extra mile to utilize text mining methods to understand customer experiences better. Text mining enables the business to extract useful information from social media platforms, articles, and other sources. Zhao (2013) explained how to extract text from tweets using the code userTimeline (). Thus, a combination of sentiment analysis, text mining, and more methods plays a vital role in ensuring that businesses understand and capitalize on customer reviews online.

According to Jain, Kumar, and Mahanti (2018), sentiment extraction was an effective method of understanding customer suppositions online. The information gathered from online platforms and product review sites enable a business to enhance their marketing strategies. The product reviews also inform and shape customer purchasing decisions (Jain, Kumar, and Mahanti, 2018). Competitive companies such as Amazon capitalize on the information in decision-making. Lim et al. (2019) asserted that US top retailers bank on online product reviews to enhance their marketing campaigns and enhanced business processes (Jagdale, Shirsat, and Deshmukh, 2019). For instance, if a specific product receives many negative reviews, the company investigates the issue to address it immediately. If the negative reviews are linked to pricing or quality, the company ensures that the issue is solved immediately.

Schoenmueller, Netzer, and Stahl (2020) collected an extensive data set of more than 280 million reviews to study their distribution. The data was created by more than twenty-four million reviewers from twenty-five sites, including Amazon and Yelp. The reviews covered different products and services. The study found that most product reviews online are less polar and positively imbalanced (Schoenmueller, Netzer, and Stahl, 2020). The study also found that the distribution of product reviews for similar products varies from one platform to another (Vyas and Uma, 2019). The reviews' variation is linked determined by various factors, including the rating scale, the online platform's business model, and the frequency of reviews (Schoenmueller, Netzer, and Stahl, 2020). Hence, to succeed in capitalizing on online product reviews, companies such as Amazon must take advantage of these factors.

Moreover, it is profound to note that businesses maximize sentiment analysis to enhance business processes and improve customer retention. Govindaraj and Gopalakrishnan (2016) asserted that an analysis of product reviews enables a business to understand customer experiences. A customer can post a review to show whether they are satisfied or unsatisfied with a specific product or service. However, most of the product reviews fail to indicate the extent of customer satisfaction. As a result, Govindaraj and Gopalakrishnan (2016) conducted a study to categorize the extent of customer satisfaction based on online reviews. They created a method of categorizing customer satisfaction based on acoustic and linguistic features. They proposed a model of categorizing customer reviews as highly positive, positive, neutral, and highly negative (Govindaraj and Gopalakrishnan, 2016). This study's results are consistent with previous research conducted by Ghasemaghaei et al. (2018) on the impact of the length of reviews and online sentiments. Customers tend to concentrate on extensive and detailed product reviews before making the final purchase (Singla, Randhawa, and Jain, 2017). Therefore, for an accurate decision based on online product reviews, companies need to investigate the actual extent of customer satisfaction with their products and services.

Sharma, Chakraborti, and Jha (2019) conducted a study to investigate how online reviews drive book sales at Amazon. According to the study, customers consider online reviews to be a dependable source of information. Customers find reviews to be more accessible and detailed. The study found that online reviews significantly shape user experiences and product prices. The findings are consistent with a previous investigation conducted by Chong et al. (2016) on online reviews and sentiments. Another issue that the study explored is online review valance. Sharma, Chakraborti, and Jha (2019) note that online reviews' impact on sales is contradictory. Some studies found that online review valance significantly impacts sales, while others found minimal impact. The impact is also dependent on factors such as product categories and qualitative text features.

Du et al. (2019) conducted a study focused on 142.8 million customer reviews from Amazon. The study focused on identifying each review's helpfulness and unhelpfulness by analyzing the summary headline, comment on the product, and helpfulness information. To enhance the accuracy of the findings, the researchers filtered all blank and non-English product reviews. Only those with the highest votes were selected (Du et al., 2019). The study found that an analysis of online product reviews on Amazon plays a significant role in today's e-commerce. Helpful reviews provide detailed information regarding the specific product or service based on customer experience (Meenakshi, Intwala, and Sawant, 2020). The reviews with the highest votes revealed that customers depend on the information to make accurate purchasing decisions. The findings are consistent with Anh, Nagai, and Nguyen's (2019) investigation on how customer reviews influence online shopping. Positive product reviews encourage customers to gain more credibility for the products they plan to purchase online.

Chapter 3- Research Methodology

3.1 Sources of Data

The primary source of data for this project is Amazon product reviews available online. Amazon has an online feature where customers are presented with a rating scale to rate the products and services purchased from the platform. Customers can also comment to specifically present what they considered when rating the product. A data set containing many of these product reviews will be used for sentiment analysis in this study. The data for use in this research will be collected from the website's Amazon product review section. Users can rate different products and services sold on Amazon's online store. Apart from the rating, customers also can comment about their experiences (Dey et al., 2020). The comment reveals a range of issues, including quality, durability, price, and customer service.

3.2 Stylometric Variables

Stylometric variables are used to breakdown the text data into corpus-based features. The methods used to attribute origin of text from online sources are style markers including tidings, document connections, HTML labels and spelling. The Stylometric variables fetch specific aspects such as punctuation, and n-grams text arrays from a review. In the review dataset the text is converted into a corpus and these aspects are identified and removed.

3.3 Lexical

This analysis aims at labelling words with the sentiments using a vocabulary based technique which utilizes a semantic score assigned to the reviews for measuring the extreme nature of the review in order to assign the right grammatical feature labeling.

3.4 Structural

This research is focused on breaking down the amazons reviews into simple words, cleaning, and normalizing the data. The data addresses the opinion of customers about Amazon's products thus the hence it requires algorithm that accurately assigns sentiments to the words used.

3.5 Syntactic

The language used in product reviews has certain particulars that will hinder accurate sentiment analysis. The syntax of the reviews contains hash tags, @ symbols and other no-word symbols that slow down text mining algorithm. These items have to be removed by syntactic algorithms like lemmatization and syntactic parsing. The character cases also has be converted to lower case and ensure that words are stemmed to reduce them to a solitary event and get rid of several forms of the same word. The objective is to finally have a lemmatized standardized textual data.

3.6 N-grams

The text data has to undergo feature extraction and selection. The methods used to perform this is the TF-IDF rating to show the n-gram feature vectors. Other approaches use existing sentiment dictionaries and leverage the unigram sentiment word as the feature. For this study we build n-gram sentiment features by extracting sentiments and multiplying their intensifiers by TF-IDF rating to find the feature score.

3.7 Text representation

The text is represented as the frequency of characters per review or corpus, the numbers of times words have been used in the reviews, the number of common words used in the review that can be found in the dictionary. These analysis are carried out using word clouds, and scatter plots to show the frequency distribution of words.

3.8 Corpus design

Our dataset comprises of two attributes; the polarity which has 2 levels of measurement 1 stands for negative reviews and 2 stands for positive reviews. The next attribute was the review text which holds the product reviews from customers. The review text is forms the data text data that goes through the processing and cleaning then a text mining package is used to generate a corpus from the text. This corpus is then used to create document term matrix. This is a matrix of frequencies where the number of occurrence of words in each review are stored. The document term matrix is the quantitative data that will be used to train a classification model on whether the review is positive or negative.

Chapter 4- Data Analysis

4.1 Text preparation

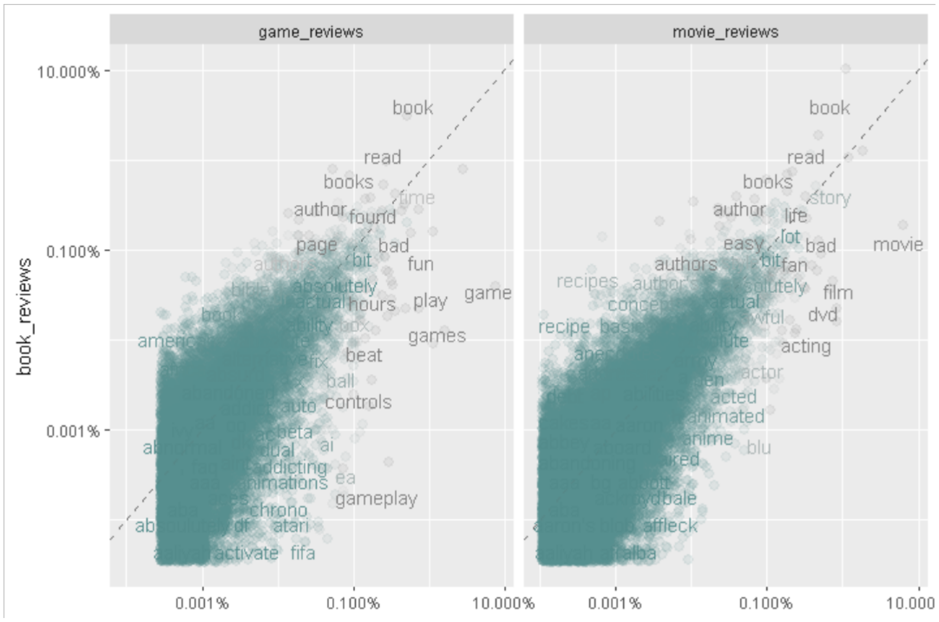
The first step was to load the amazon dataset into R-studio environment. The size of the data was 1.5GB with dimensions of 3 variables and 3.6 Million observations. Due to computational constraints the data was sampled using random sampling and a sample of 5% of the entire dataset was used for the rest of the analysis, this was a data.frame of dimensions 3 variables and 180,000 observations. The dataset was cleaned using functions in the “tidyverse” package and general substring functions to remove punctuation, numbers, whitespaces and stop words.

The next step was to extract three data frames from the dataset representing book reviews, movie reviews, and game reviews. The purpose of these datasets was to analyze and compare reviews of different amazon products for example book reviews against movie reviews, and game reviews. The next stage shows the relationship between book reviews and movie reviews or game reviews.

4.2 Association and Correlation analysis

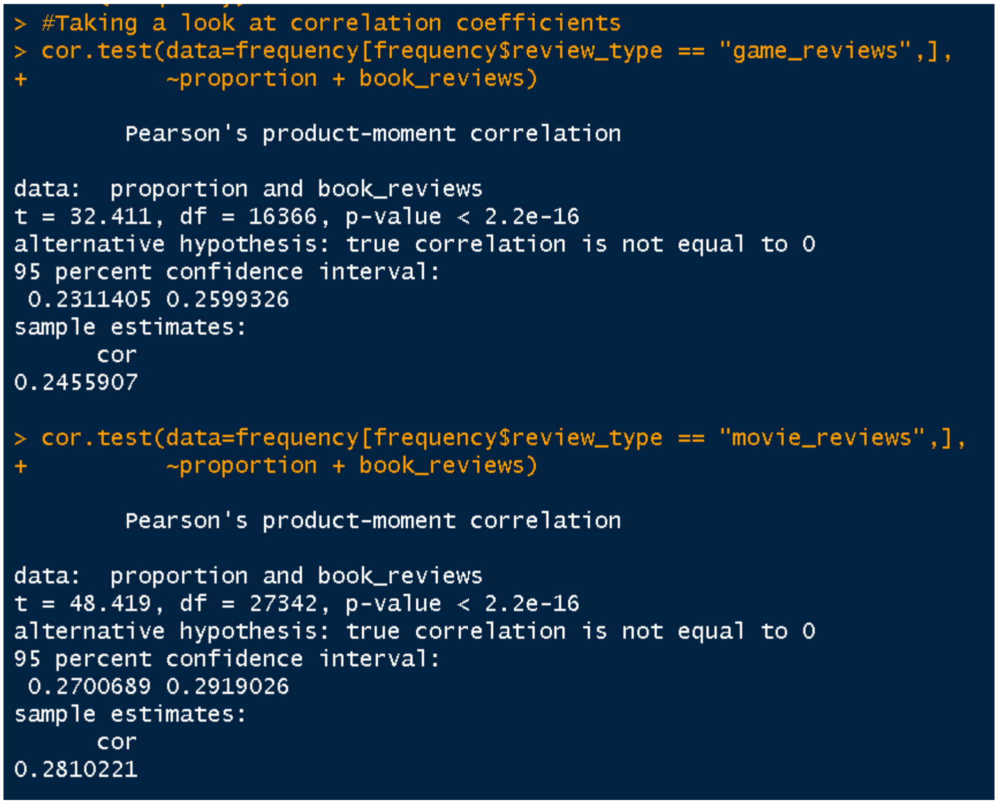
The reviews were filtered to obtain reviews about amazon books, movies and games. The data was tokenized and frequency data.frame generated from these tokens. The frequency and proportions data were used to plot correllograms as shown below;

Figure 1: Correlation plots of book reviews against game and movie reviews



The plots show that both game and movie reviews had similar correlation with book reviews. The frequency of words used in the book reviews was similar and also the words that are shared in both reviews appear on the best fit dotted line, for instance words like “life”, and “story” are used by both movie and book reviewers. On the other hand words like “absolutely”, and “time” are used by both book and game reviewers in similar fashion.

Figure 2: Correlation test of book reviews against movie and game reviews



```
> #Taking a look at correlation coefficients
> cor.test(data=frequency[frequency$review_type == "game_reviews",],
+          ~proportion + book_reviews)

        Pearson's product-moment correlation

data:  proportion and book_reviews
t = 32.411, df = 16366, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2311405 0.2599326
sample estimates:
      cor
0.2455907

> cor.test(data=frequency[frequency$review_type == "movie_reviews",],
+          ~proportion + book_reviews)

        Pearson's product-moment correlation

data:  proportion and book_reviews
t = 48.419, df = 27342, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2700689 0.2919026
sample estimates:
      cor
0.2810221
```

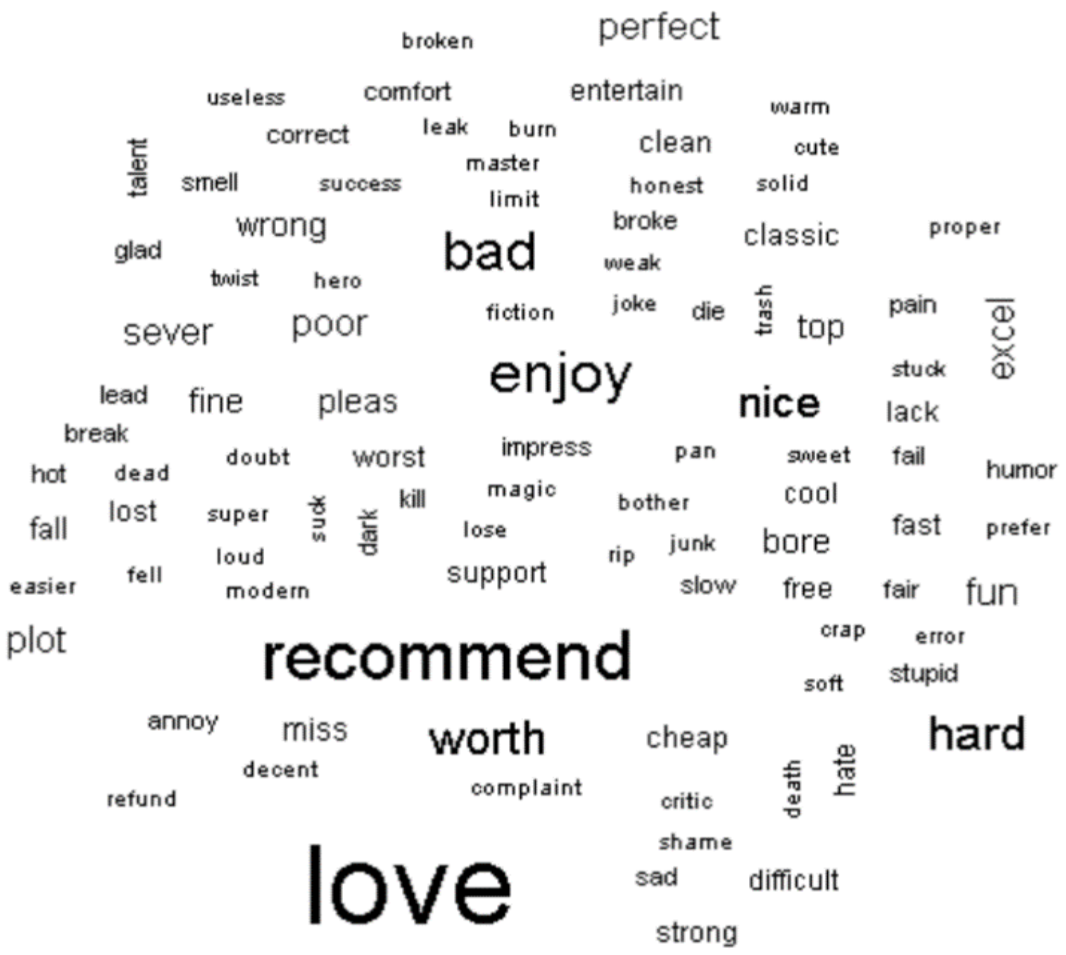
Figure 3 shows the Pearson’s correlation test between book reviews and movie reviews, and another correlation test between game reviews and book reviews . The p-value of both tests was below 0.05, hence the tests were statistically significant in determining a correlation between the these sets of reviews. From the correlation test in Figure 3, it’s evident that the book reviews and movies had statistically significant correlation of 0.2 in terms of frequencies of the word used in the reviews. This was the case also in correlation between book reviews and game reviews.

4.3 Sentiment detection

Bad reviews that have nothing to do with rating an Amazon product were eliminated. The remaining reviews words were tokenized and frequency analysis performed to analyze the number of words and sentiments in the reviews. The sentiment analysis was carried out using R packages like dplyr, tidyverse, tm, wordcloud, tidytext etc. These packages rendered insightful functions that cleaned, tokenized and summarized the reviews to transform them into meaningful information. The first step of analysis was to plot a word cloud to visualize the distribution of the most frequent words.

From the word cloud its visible that words like “recommend”, “love”, ”enjoy”, “nice”, “bad” and “hard” dominated the reviews of most amazon customers. Word clouds are important especially in quick visualization of textual data and they give brief result of textual analysis, portraying the tone, sentiment or emotion in written media.

Figure 3: Word cloud showing the most frequent words in all the reviews



The next sentiment analysis shown in figure 4 involved utilizing “nrc” lexicons(Explain what is NRC lexicon and why you select this lexicon and not AFINN) to assign sentiments to words that have been tokenized. Now, the previous tokenization involved three groups but this time the whole data reviews are tokenized and the sentiments of the words are generated and used to form a new data frame. This data frame is used to plot several bar chart showing the most frequent words in the reviews grouped into different sentiments.

Figure 4:Contribution of these most occurring words to the various sentiment

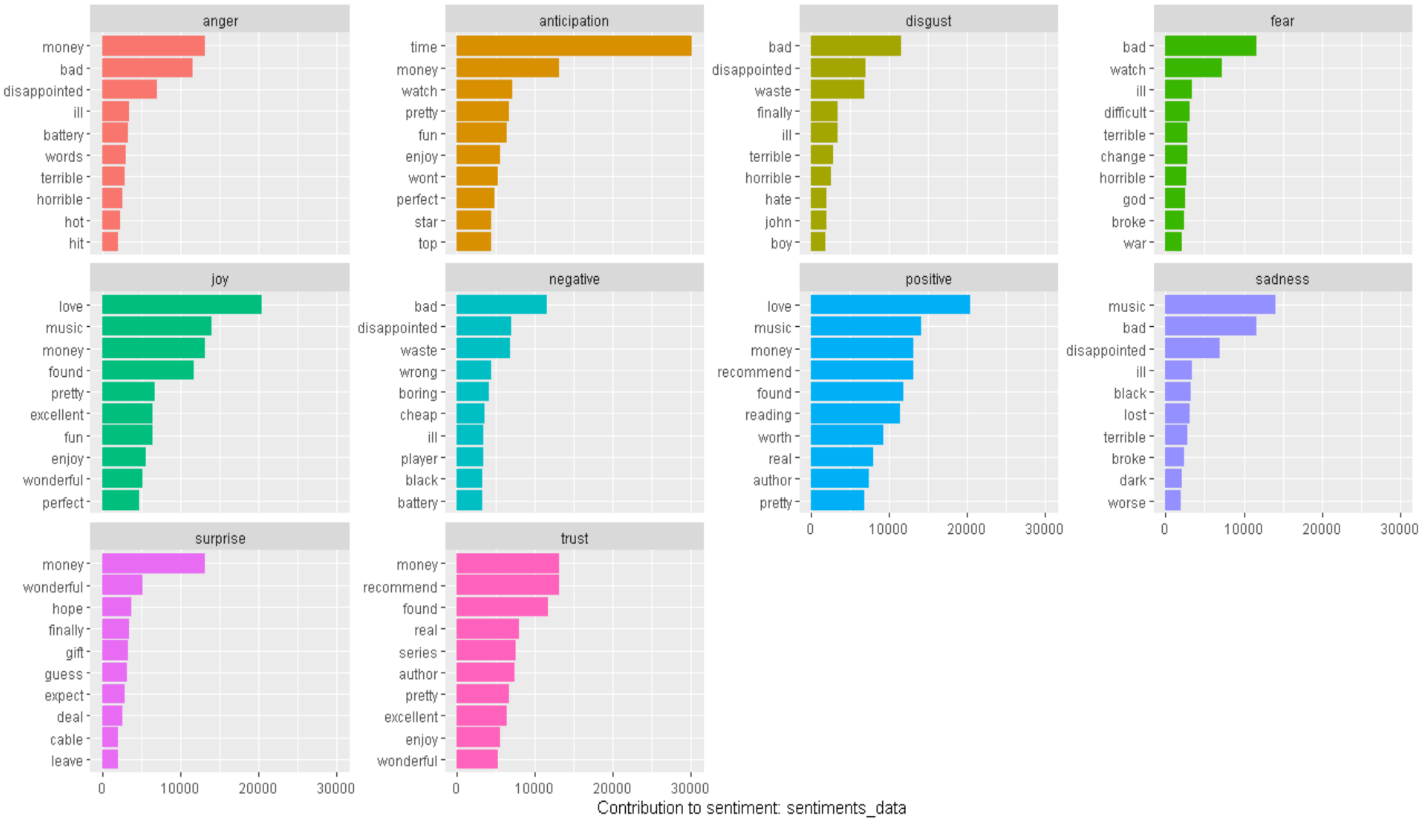


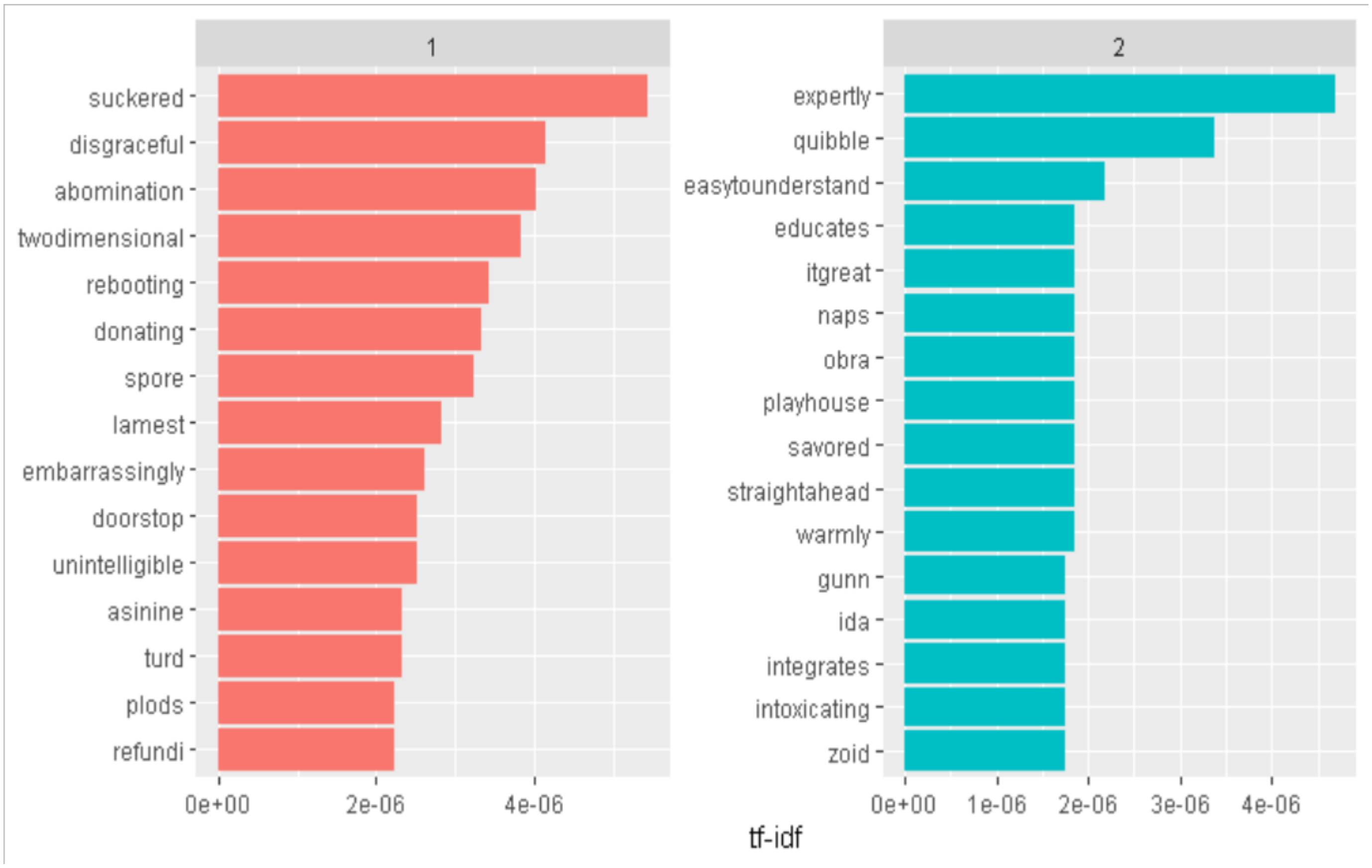
Figure 5 shows the ten sentiments assigned to the tokenized words. Looking at the graphs; anger is mostly portrayed when a customer is talking about money. Secondly customers who expressed anticipation were mostly mentioning (time) in their reviews. Disgusted customers used words like (“bad” and “disappointed”) in their reviews. Customers who expressed fear in their reviews used words like (“watch” and “bad”) most likely they were addressing movies. Customers with joy sentiments used words like (love, music and money) this most likely was addressed to music products from amazon. Therefore, it’s statistically sound to assume that music triggers joy in most amazon customers. The same music also brings sad sentiments to customers who express disappointment in most of their reviews.

Another sentiment expressed by the customers was ‘surprise’, now most customers showed this sentiment when using words like (“money”, “wonderful”, “gift”, ”guess”, “expect”) this can be attributed to prices perhaps. Whereby customers visit the site expecting different prices on certain products and are surprised to find the contrary. Another sentiment that was assigned to the words used in the reviews was trust. Customers portraying trust mostly used words like (“recommend”, ”found”, ”real”, ”money”). Now the clear picture here is that for customers to recommend products it takes trust and also where money and realness is involved customers would express trust in these products. Finally positive sentiments and negative sentiments also are assigned to the reviews. The positive sentiment was liked to use of words like (“love”, ”recommend”, ”music”) while negative sentiments were expressed when words like (“bad”, ”waste”, ”disappointed”) were used.

Term frequency analysis [ref??] was used to calculate the frequency of certain terms in the reviews in relation to the total terms extracted from the whole data. These frequencies show the extend of usage of terms in the either negative reviews or positive reviews. -This is the part of your research methodology.

From figure 6 the red bar chart shows the distribution of terms in the negative reviews and the three most used terms in are “suckered” , “disgraceful” and “abomination”. These terms most likely address products with content and plot for example movies, books etc.. on the other hand the positive reviews used terms like “expertly”, “quibble”,” easytounderstand”, “educates”. Similarly, these terms target content based products like movies games books and works of art.

Figure 5: Term Frequency – Inverse Document Frequency graph


1
suckered
disgraceful
abomination
twodimensional
rebooting
donating
spore
lamest
embarrassingly
doorstop
unintelligible
asinine
turd
plods
refundi
0e+00
2e-06
4e-06
2
expertly
quibble
easytounderstand
educates
itgreat
naps
obra
playhouse
savored
straightahead
warmly
gunn
ida
integrates
intoxicating
zoid
0e+00
1e-06
2e-06
3e-06
4e-06
tf-idf

4.4 Sentiment classification

The classification of sentiments is executed following procedures that ensure the raw data is cleaned, converted into a corpus (a collection of words) and a term document matrix (a matrix showing the frequency of words in the corpus). These objects make it easy for the classification models to learn the relationship between the words and the sentiments. The following steps explain how the entire text mining is done using machine learning:

Step 1: Data cleaning, Corpus and Document term matrix

The data was fitted into a Naïve Bayes model and the model was trained to learn to classify new reviews into either positive or negative polarity. The dataset was first transformed into a corpus using the tm package. The corpus was then cleaned by removing numbers, punctuation, whitespace etc. The corpus is then used to create a DTM sparse matrix.

Step 2: Splitting the DTM into training and test DTM

Then to confirm that the subsets are representative of the complete set of data, we compare the proportion of positive and negative reviews using the polarity variable in the training and test data frames

Step 3: Visualizing text data using word clouds,

The corpus is visualized using word clouds where the most frequent words are plotted.

Step 4: Creating indicator features for frequent words

The indicator features will help in model fitting as the model will focus on those with the best predictive features. Then we create train and test data sets for frequent words. A function that converts the polarity attached to frequent words into categories of (positive and negative). Then apply the function to the train and test frequent features data.

Step 5: Train a Naive Bayes model on the frequent features.

A Naive Bayes model is a collection of classification algorithms that use the Bayes theorem in their methodology. These algorithms assume that the features in a model are independent and equal in their contribution to the target variable. There different naive Bayes classifiers which differ mainly by the assumptions they make regarding the distribution of P(xi | y). There are those classifiers that use discrete feature and others that are applicable to continuous classifiers. These a referred to as Gaussian Naive Bayes classifiers. The assumption of these classifiers is that the features are normally distributed. Some of the advantages of Naive Bayes are:

- They require little data to train.
- They are fast, especially because they alleviate the problem of dimensionality by making each feature distribution to be estimated as one-dimensional distribution.

A Naïve Bayes model uses Naïve Bayes theorem which is based on an event occurring given that another event has already occurred. Naive Bayes classifier uses the Bayes Theorem. It predicts membership probabilities for each category such as the probability that given record or data point belongs to a particular group. The class that is selected as the most likely class is the one with maximum a posteriori probability. In this model we applied the Bernoulli Naïve Bayes algorithm which analyzes features in binary form.

The model was evaluated using a confusion matrix. This is a contingency table of predicted and actual labels which shows how the model performed in predicting the polarity of the test reviews. The metrics for that include accuracy, Kappa, sensitivity, and specificity.

Accuracy is a metric that measures the fraction of predictions our model got right. It can be calculated using the following formula:

$$\boldsymbol{Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}}$$

Accuracy can also be expressed as true positives and negatives, and false positives and negatives.

$$\boldsymbol{Accuracy = \frac{TP + TN}{TP + TN + FP + FN}}$$

Kappa statistic is a measure of how close the classified data is to the ground truth labels. It is calculated using the following formula:

$$\boldsymbol{Kappa = \frac{Total\ accuracy - Random\ accuracy}{1 - Random\ accuracy}}$$

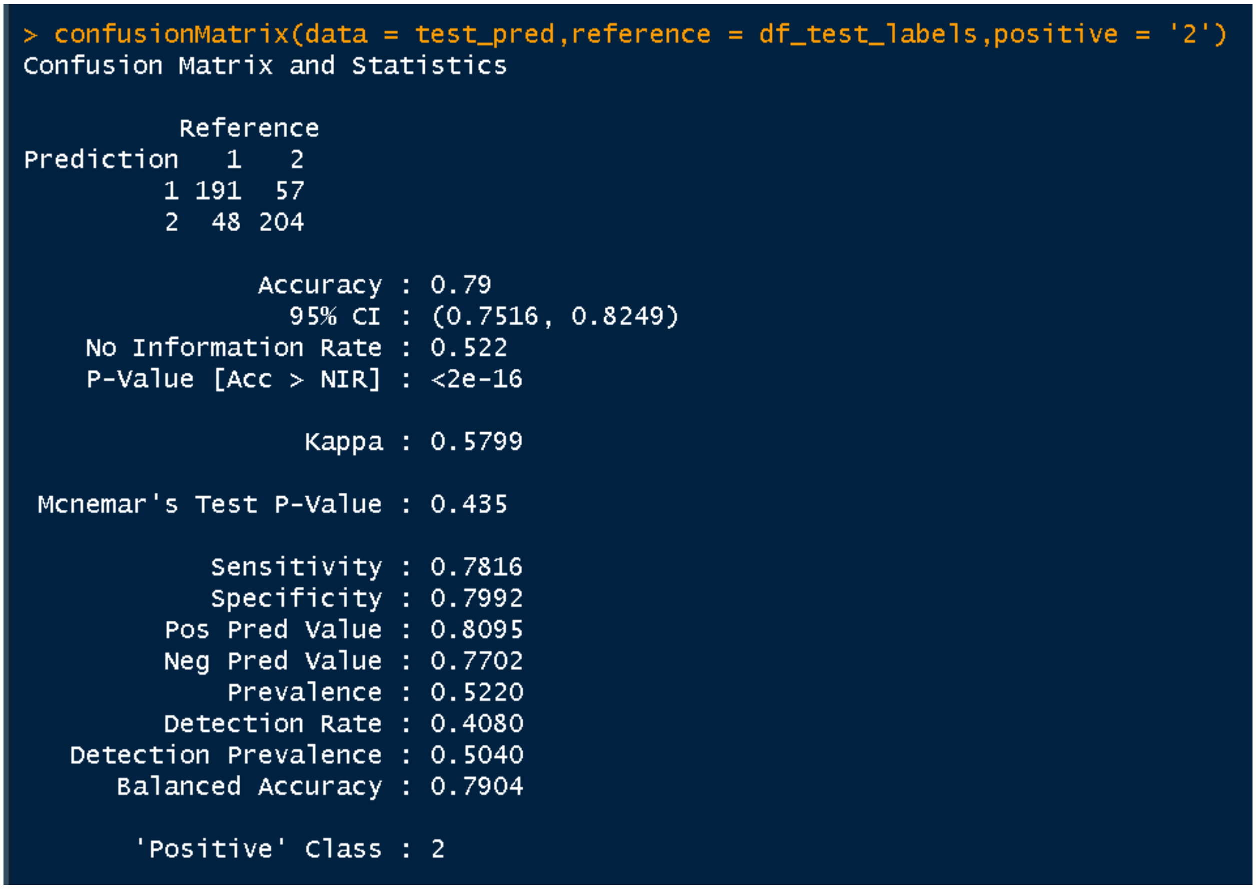
Sensitivity is the rate of the model in correctly classifying the data of class 1 (the positive class). The formula used to calculate sensitivity as follows:

$$\boldsymbol{Sensitivity = \frac{TP}{TP + FN}}$$

Specificity is the rate of the model correctly classifying the data of class 0 (the negative class). The formula used to calculate specificity as follows:

$$\boldsymbol{Specificity = \frac{TN}{TN + FP}}$$

Figure 6: Model Evaluation – Confusion matrix



```
> confusionMatrix(data = test_pred,reference = df_test_labels,positive = '2')
Confusion Matrix and Statistics

          Reference
Prediction   1   2
         1 191  57
         2  48 204

               Accuracy : 0.79
                 95% CI : (0.7516, 0.8249)
    No Information Rate : 0.522
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.5799

 Mcnemar's Test P-Value : 0.435

            Sensitivity : 0.7816
            Specificity : 0.7992
         Pos Pred Value : 0.8095
         Neg Pred Value : 0.7702
             Prevalence : 0.5220
         Detection Rate : 0.4080
   Detection Prevalence : 0.5040
      Balanced Accuracy : 0.7904

       'Positive' Class : 2
```

Explanation of the data in figure 7.

The first part of the output in figure 7 shows the contingency table with predicted values and reference values as the actual true values. These tables have the true positives (204), negatives (191) and false positives (48) and negatives (57). A true positive is when the model predicted positive and the actual result is positive. A true negative is when the model predicted negative and the actual result is negative. A false positive is when the model predicted positive and the actual result is negative. A false negative is when the model predicted negative and the actual result is positive.

The second part shows the Accuracy, this means the model was able to predict correctly 79% of the new reviews data. The third part was the Kappa statistic, which means that the model was

57.99% of the times in agreement with the baseline random classifier, this shows that our model agreement with ground truth was only 57.99%. This Kappa value shows that model still has long way to truthfully classify reviews correctly. Finally, there was sensitivity score, which means that the model had a 78.16% rate at correctly classifying positive reviews. Specificity score, indicating that the model had a 79.92% rate at correctly classifying negative reviews.

Chapter 5. Conclusion

5.1 Conclusion

In conclusion the analysis findings show from the customer reviews that language and words used to review books, were likely to be used also by customers who review movies and game products from amazon. The correlation analysis using the word frequency between the three products was statistically significant indicating that the frequency of words used in the three products was similar. Anticipation was one of the sentiments that was elicited from the customer reviews and most customers talked about time as most frequently, hence anticipation can be illustrated in customer reviews when they talk about timing of products. To be able to predict disgust, joy, fear, sadness, surprise, trust, positive and negative sentiments from customer reviews the most frequent words associated with these sentiments are taken into account. Another analysis that is essential in portraying the most positive and negative of words used is the Term frequency – inverse Document frequency analysis, this method plots the most frequent terms in both positive and negative sentiments hence providing a clear picture of the reviews data.

In the classification of reviews dataset into either positive or negative sentiment group, a Naïve Bayes classification model was used. The model performed quite well. The metrics used to measure the model performance were accuracy, Kappa, sensitivity and specificity. These metrics show that the model had an accuracy of 79% in classifying reviews and the accuracy similar to the sensitivity and specificity scores, but the Kappa score was quite low. This is worrisome because our model was quite ambitious, and the Kappa score shows there is still some improvement to do to ensure the model is in agreement with ground truth instead of just focusing on true positive and true negatives.

5.2 Recommendations

After careful consideration and looking at the analysis findings, it is recommended that more powerful GPU to be used in training the machine learning model in order to get a more accurate model. Also, the dataset needs to be taken through extensive cleaning to remove subjective reviews, so that the sentiments generated are the objective opinions of customers.

Bibliography

Anh, K.Q., Nagai, Y. and Nguyen, L.M., 2019. Extracting customer reviews from online shopping and its perspective on product design. *Vietnam Journal of Computer Science*, *6*(01), pp.43-56.

Chong, A.Y.L., Li, B., Ngai, E.W., Chang, E. and Lee, F., 2016. Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations & Production Management*.

Chauhan, U.A., Afzal, M.T., Shahid, A., Abdar, M., Basiri, M.E., and Zhou, X., 2020. A comprehensive analysis of adverb types for mining user sentiments on amazon product reviews. *World Wide Web*, pp.1-19.

Du, J, Rong, J, Michalska, S, Wang, H & Zhang, Y. 2019. Feature selection for helpfulness prediction of online product reviews: An empirical study', PloS one, 14(12), p. e0226902.

Dey, S., Wasif, S., Tonmoy, D.S., Sultana, S., Sarkar, J., and Dey, M., 2020. A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews. In *2020 International Conference on Contemporary Computing and Applications (IC3A)* (pp. 217-220). IEEE.

Dong, L., Huang, S., Wei, F., Lapata, M., Zhou, M. and Xu, K., 2017, April. Learning to generate product reviews from attributes. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 623-632).

Govindaraj, S. and Gopalakrishnan, K. 2016. Intensified Sentiment Analysis of Customer Product Reviews Using Acoustic and Textual Features. *ETRI Journal*, 38(3), pp. 494–501.

Ghasemaghaei, M., Eslami, SP, Deal, K. and Hassanein, K., 2018. Reviews' length and sentiment as correlates of online reviews' ratings. *Internet Research*.

Han, H, Zhang, Y, Zhang, J, Yang, J & Zou, X. 2018. Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias', *PloS one*, 13(8), p. e0202523.

Jain, V.K., Kumar, S. and Mahanti, P., 2018. Sentiment recognition in customer reviews using deep learning. *International Journal of Enterprise Information Systems (IJEIS)*, *14*(2), pp.77-86.

Jagdale, R.S., Shirsat, V.S. and Deshmukh, S.N., 2019. Sentiment analysis on product reviews using machine learning techniques. In *Cognitive Informatics and Soft Computing* (pp. 639-647). Springer, Singapore.

Kabir, A.I., Ahmed, K. and Karim, R., 2020. Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. *Informatica Economica*, *24*(4), pp.55-71.

Karamitsos, I., Albarhami, S. and Apostolopoulos, C., 2019. Tweet Sentiment Analysis (TSA) for cloud providers using classification algorithms and latent semantic analysis. *Journal of Data Analysis and Information Processing*, *7*(4), pp.276-294.

Lim, J., Park, M., Anitsal, S., Anitsal, M.M. and Anitsal, I. 2019. 'Retail Customer Sentiment Analysis: Customers' Reviews of Top Ten US Retailers' Performance,' *Global Journal of Management and Marketing*, 3(1), 124+.

Manchaiah, V, Amlani, AM, Bricker, CM, Whitfield, CT & Ratinaud, P. 2019. Benefits and Shortcomings of Direct-to-Consumer Hearing Devices: Analysis of Large Secondary Data Generated from Amazon Customer Reviews. *Journal of Speech, Language & Hearing Research*, 62(5), pp. 1506–1516.

Meire, M, Hewett, K, Ballings, M, Kumar, V & Van den Poel, D. 2019. The Role of Marketer-Generated Content in Customer Engagement Marketing. *Journal of Marketing*, 83(6), pp. 21–42.

Meenakshi, A.B., Intwala, N., and Sawant, V., 2020. Sentiment analysis of amazon mobile reviews. *ICT Systems and Sustainability: Proceedings of ICT4SD 2019, Volume 1*, *1077*, p.43.

Nandal, N., Tanwar, R. and Pruthi, J., 2020. Machine learning-based aspect level sentiment analysis for Amazon products. *Spatial Information Research*, *28*(5), pp.601-607.

Schoenmueller, V., Netzer, O. and Stahl, F. 2020. The Polarity of Online Reviews: Prevalence, Drivers and Implications. *Journal of Marketing Research (JMR)*, 57(5), pp. 853–877.

Sharma, S. K., Chakraborti, S. and Jha, T. 2019. Analysis of Book Sales Prediction at Amazon Marketplace in India: A Machine Learning Approach. *Information Systems and e-Business Management*, 17(2–4), pp. 261–284.

Shankhdhar, G. 2019. Sentiment Analysis Methodology. Retrieved from https://www.edureka.co/blog/sentiment-analysis-methodology/

Singla, Z., Randhawa, S. and 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). IEEE.

Shelke, N., Deshpande, S. and Thakare, V., 2017. Domain independent approach for aspect-oriented sentiment analysis for product reviews. In *Proceedings of the 5th international conference on frontiers in intelligent computing: Theory and applications* (pp. 651-659). Springer, Singapore.

Vyas, V. and Uma, V., 2019. Approaches to sentiment analysis on product reviews. In *Sentiment Analysis and Knowledge Discovery in Contemporary Business* (pp. 15-30). IGI Global.

Zhao, Y., 2013. *R and data mining: Examples and case studies*. Academic Press.