LOOKING TO THE STARS: A HYBRID VOTING APPROACH FOR STAR RATING PREDICTION OF AMAZON CUSTOMER REVIEWS

By

Storm Davis

July 2021

Director of Thesis: Dr. Nasseh Tabrizi

Major Department: Computer Science

ABSTRACT

Due to the continuous growth of E-Commerce platforms, more and more data is becoming widely available. Almost every online transaction, from buying products on Amazon to renting vacation houses on Airbnb, allows for users to leave feedback in the form of customer reviews.

Therefore, there is a need to provide an overview of the status of customer review analysis. The first part of this thesis systematically reviews research works conducted in the past ten years (2010-2020) to identify how customer reviews are being analyzed and how this analysis can serve a purpose to the consumer or the vendor. Common machine learning algorithms and datasets will be presented as well as the numerous applications of customer review analysis. This section aims to provide insight into what work has already been done in this field as well as discusses directions for future research.

The second part of this thesis proposes a novel approach to star rating classification using customer reviews. TF-IDF techniques were used to extract key features from the review text. After evaluating the performance of five baseline classifiers through 5-fold cross

validation, we incorporate the use of an ensemble hard voting classifier that employs multiple baseline classifiers to aid in final star rating prediction. Experimental results using an Amazon customer review dataset demonstrated that the best performing ensemble voting model was able to perform just as well as the best performing individual classifier while employing the use of both Logistic Regression and LinearSVC.

LOOKING TO THE STARS: A HYBRID VOTING APPROACH FOR STAR RATING PREDICITION OF AMAZON CUSTOMER REVIEWS

A Thesis

Presented to the Faculty of the Department of Computer Science

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Software Engineering

By

Storm Pierce Davis

July 2021

LOOKING TO THE STARS: A HYBRID VOTING APPROACH FOR STAR RATING PREDICITION OF AMAZON CUSTOMER REVIEWS

By

Storm Pierce Davis

APPROVED BY:

Director of Thesis ______________________________

Dr. Nasseh Tabrizi

Committee Member ______________________________

Dr. Mark Hills

Committee Member ______________________________

Dr. Rui Wu

Chair of the Department of Name ______________________________

Venkat Gudivada, PhD

Dean of the Graduate School ______________________________

Paul J. Gemperline, PhD

Table of Contents

LIST OF TABLES vi

LIST OF FIGURES vii

1. INTRODUCTION 1

 1.1. Research Contribution 3

 1.2. Thesis Structure 3

2. SYSTEMATIC LITERATURE REVIEW 5

 2.1. Introduction 5

 2.2. Systematic Review 6

 2.3. Customer Review Datasets 9

 2.4. Machine Learning Models 12

 2.5. Application Areas 15

  2.5.1. Sentiment Analysis 16

  2.5.2. Text Mining 16

  2.5.3. Attribute Extraction 17

  2.5.4. Prediction 17

  2.5.5. Other 18

 2.6. Research Directions 18

3. RELATED LITERATURE ON STAR RATING PREDICTION OF CUSTOMER REVIEWS 20

4. METHODOLOGY 23

4.1. Hardware/Software Setup 23

4.2. Amazon Customer Review Dataset 24

4.3. Machine Learning Classifiers 31

4.4. Evaluation of Baseline Models 32

4.5. Ensemble Voting Model approach and Validation 35

5. EXPERIMENTAL RESULTS 38

5.1. Model Comparison and Evaluation 38

6. CONCLUSIONS AND FUTURE WORK 45

BIBLIOGRAPHY 47

LIST OF TABLES

4.1 General description of data columns in the Amazon Review dataset 25

4.2 Sample Star_rating and Review_body data from Amazon Review dataset 26

4.3 Machine learning classifiers and parameters used 31

4.4 5-fold cross validation results for baseline classifiers 33

4.5 5-fold cross validation results for Voting classifiers 36

5.1 F1 scores for classification models 42

LIST OF FIGURES

2.1 Number/Type of papers published from 2010 to 2020 7

2.2 Number of papers by database with query ‘customer reviews analysis machine learning’ 8

2.3 Number of papers by database with query (("customer reviews") AND ("machine learning")) 8

2.4 Distribution of the different datasets used for customer review analysis 10

2.5 Count of papers using specific product areas for customer review analysis 12

2.6 Distribution of classifiers used for customer review analysis 14

2.7 Distribution of topic areas using customer review analysis 15

4.1 Count of Reviews by Star Rating 27

4.2 Word cloud generated for 1-star reviews 28

4.3 Word cloud generated for 2-star reviews 29

4.4 Word cloud generated for 3-star reviews 29

4.5 Word cloud generated for 4-star reviews 30

4.6 Word cloud generated for 5-star reviews 30

4.7 Strip plot generated for baseline classifiers for accuracy across cross validation folds . 34

4.8 Strip plot generated for voting classifiers for accuracy across cross validation folds 37

5.1 Accuracy of machine learning models on cross validation vs. test set 39

5.2 Accuracy of machine learning models on test set 40

5.3 Confusion matrices for baseline models 43

5.4 Confusion matrices for ensemble voting models 44

Chapter 1

Introduction

As data on the internet continues to grow at an exponential rate, customers are confronted with information overload [1]. Users are presented with huge amounts of information about products or services, making purchasing decisions even more difficult. The aim of customer reviews is to help possible future customers make informed decisions based on feedback left from past customers, however, due to the ever-growing number of customer reviews, this is certainly is not an easy task. Recently, the area of customer review analysis has surfaced, attempting to automate analysis processes of the reviews to aid consumers as well as merchants.

Although the analysis goal for each application may be unique, the primary subject areas used in analyzing customer reviews remain consistent. The areas of sentiment analysis [58] and text mining [59] are by far the most used tools for the process of analyzing customer reviews. Sentiment analysis, often referred to as opinion mining, is a process in which text is analyzed and given an objective sentiment score. In its most basic form, sentiment analysis should be able to determine if a text has a sentiment of either positive or negative. However, these sentiment classes do not have to be binary and can be expanded to any number of different sentiment labeling. Text mining is a subject area that can be used to complement sentiment analysis or as a tool on its own. Text mining is an artificial intelligence technology that uses natural language processing to transform unstructured text into a more structured and suitable form for analysis. Although there are certainly more areas used in customer review analysis than sentiment analysis and text mining, these are the two most common approaches used for analyzing textual data at a basic level.

Textual analysis, more specifically, customer review analysis has become an increasingly popular topic throughout the past 10 years due to an increase in available data from online services in addition to advances in machine learning tools [2-14]. In section one of this thesis, we systematically review the research in customer review analysis conducted from 2010-2020. The next few sections will detail the systematic steps taken during this review, define goals and research questions, discuss current trends in customer review analysis literature, as well as provide future research directions and draw conclusions.

As we will later discuss in section 2, there are very few studies that deal with predicting the star rating of customer reviews. In this thesis, we present a systematic literature review on the status of the customer analysis field as the basis of predicting star ratings in embedded in the understanding of customer review analysis. We then define the baseline models used as well as introduce a novel concept that has never been implemented in this area of research. Our best performing ensemble voting model was able to perform just as well as the best performing baseline classifier while employing the use of both Logistic Regression and LinearSVC.

1.1 Research Contributions

In this research work, we first present a systematic literature review on the field of customer review analysis. In this literature review, we systematically review trends in research from 2010 to 2020 in terms of common machine learning algorithms, customer review datasets, and the many ways customer review analysis techniques are being used. The goal of this literature survey on customer review analysis is to provide insight into this field of research as well as discuss directions for future research.

The second, and primary, contribution of this thesis is developing a novel, hybrid voting classifier that incorporates multiple different classifiers to generate a final classification value. The proposed classifier uses a “soft” voting technique where the output is based on the average probability of each class. The class with the maximum average probability across all the individual classifiers used is the final prediction. To determine which classifiers were best fit for use in the voting classifier, a baseline of five classifiers was performed using 5-fold cross validation to generalize individual performance. Then, a set of three voting models were created from employing multiple baseline classifiers. After evaluating these models with the same cross validation techniques, a final experiment was conducted to further evaluate each model using unseen Amazon customer review data.

1.2 Thesis Structure

The structure of this thesis implementation is defined as follows. Chapter 2 presents the systematic literature review conducted on the customer review analysis field. Chapter 3 provides some related works on the idea of using customer review analysis to aid in the prediction of star ratings. Chapter 4 details the proposed methodology in detail. Chapter 5 includes experimental results and analysis of the proposed voting model on an Amazon customer review dataset. Lastly, in Chapter 6, we draw conclusions and discuss future research opportunities

Chapter 2

Systematic Literature Review

2.1 Introduction

As data on the internet continues to grow at an exponential rate, customers are confronted with information overload [1]. Users are presented with huge amounts of information about products or services, making purchasing decisions even more difficult. The aim of customer reviews is to help possible future customers make informed decisions based on feedback left from past customers, however, due to the ever-growing number of customer reviews, this is certainly is not an easy task. Recently, the area of customer review analysis has surfaced, attempting to automate analysis processes of the reviews to aid consumers as well as merchants.

Although the analysis goal for each application may be unique, the primary subject areas used in analyzing customer reviews remain consistent. The areas of sentiment analysis [58] and text mining [59] are by far the most used tools for the process of analyzing customer reviews. Sentiment analysis, often referred to as opinion mining, is a process in which text is analyzed and given an objective sentiment score. In its most basic form, sentiment analysis should be able to determine if a text has a sentiment of either positive or negative. However, these sentiment classes do not have to be binary and can be expanded to any number of different sentiment labeling. Text mining is a subject area that can be used to complement sentiment analysis or as a tool on its own. Text mining is an artificial intelligence technology that uses natural language processing to transform unstructured text into a more structured and suitable form for analysis. Although there are certainly more areas used in customer review analysis than sentiment analysis and text mining, these are the two most common approaches used for analyzing textual data at a basic level.

Textual analysis, more specifically, customer review analysis has become an increasingly popular topic throughout the past 10 years due to an increase in available data from online services in addition to advances in machine learning tools [2-14]. In this thesis, we systematically review the research in customer review analysis conducted from 2010-2020. The following sections will detail the systematic steps taken during this review, define goals and research questions, discuss current trends in customer review analysis literature, as well as provide future research directions and draw conclusions.

2.2 Systematic Review

As mentioned above, this is a systematic review of recent studies in the customer review analysis and machine learning field. First and foremost, the goals of the review need to be defined in addition to the systematic process used in the collection of literature works. The goals of the review are defined as follows.

- Identifying machine learning algorithms/models employed to analyze customer reviews
- Identifying the datasets used for testing and validating machine learning models
- Identifying topic areas or application areas for customer reviews analysis

Based on the goals mentioned above, this systematic review focuses on answering the following guiding questions.

- What machine learning algorithms/models are being used to analyze customer reviews?
- What customer review datasets are being used to validate machine learning models?
- What topic areas or application areas are being studied in the field of customer review analysis?

In the following sections, we attempt to answer these questions to reach the goals of this systematic review.

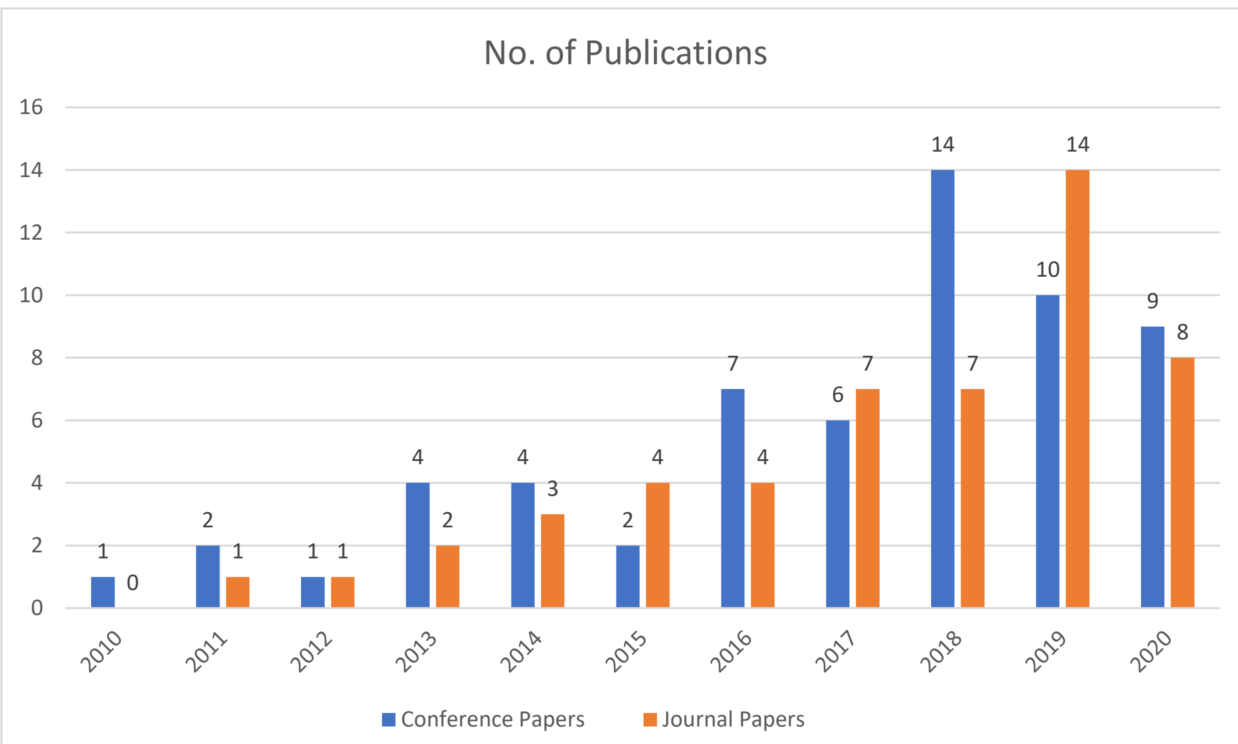

Figure 2.1: Number/Type of papers published from 2010 to 2020

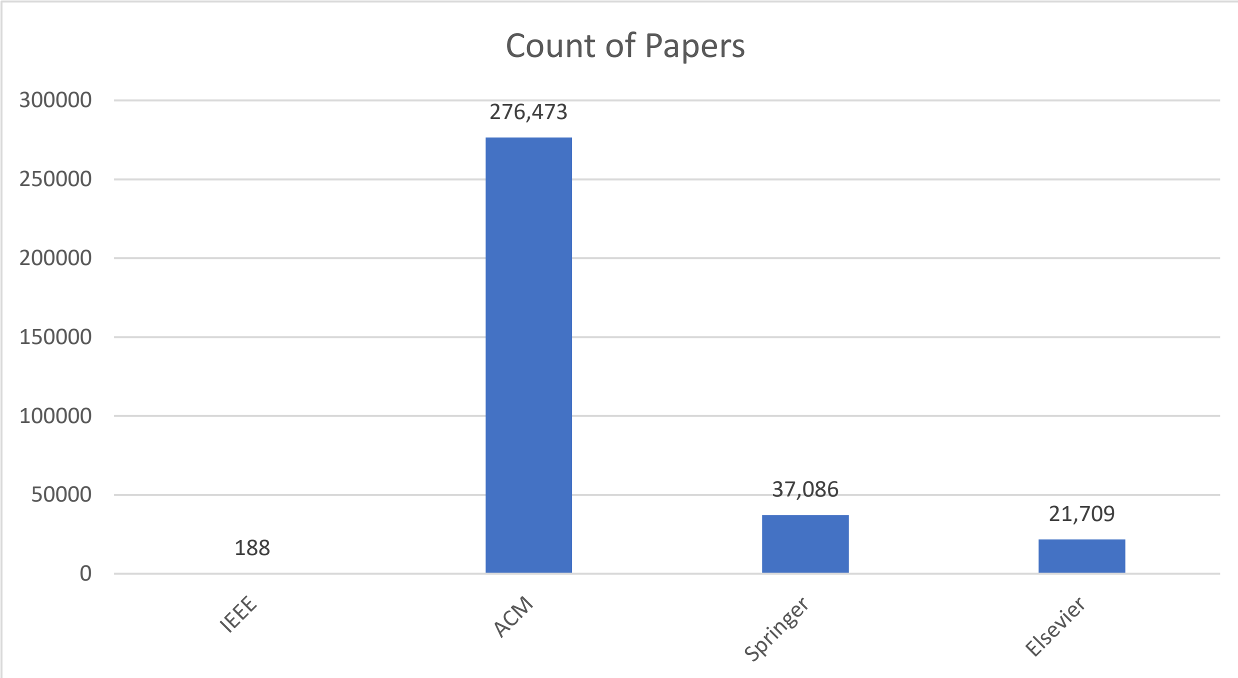

Figure 2.2: Number of papers by database with query ‘customer reviews analysis machine learning’.

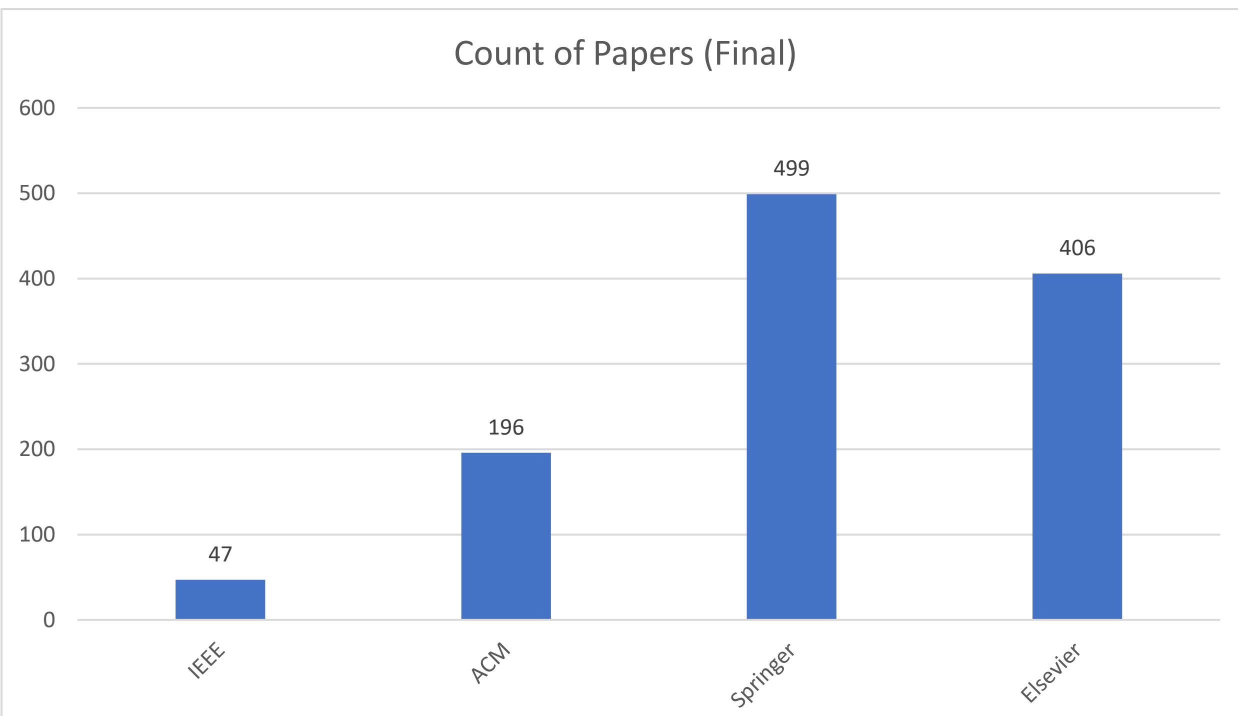

Figure 2.3: Number of papers by database with query (("customer reviews") AND ("machine learning")).

In terms of the systematic protocol, first, we searched for publications via four scientific databases IEEE, ACM, Springer, and Elsevier and noted the results. The initial search query was "customer reviews analysis machine learning", minus the quotes during search time. The results are shown in Fig. 2.2. However, the vast amount of papers returned by the search criteria hinted that the query was too broad. The final query was determined to be: (("customer reviews") AND ("machine learning")), quotes included, see Fig. 2.3.

As a second step of the protocol, the top 50 publications from each database, sorted by relevance, were collected. Next, exclusion criteria were performed on the remaining 200 papers. The exclusion criteria are detailed below.

- No book chapters or short/poster papers will be analyzed
- No other survey or review papers will be analyzed
- No papers using non-English datasets will be analyzed

As the final step of protocol, all remaining publications (after applying exclusion criteria) were thoroughly analyzed, and relevant data was extracted to answer the questions of this systematic review. A detailed table was constructed using relevant extracted information and was used as an aid to visualize the outcome of our analysis as shown in Fig. 2.1. The following sections discuss the findings of our analysis.

2.3 Customer Review Datasets

First and foremost, we wanted to understand what customer review datasets are being used and how these datasets are being collected. It is known [60] that several large E-commerce companies like Amazon provide publicly available datasets. However, to our surprise, many of the analyzed research (see Fig. 2.4) works opted for a manual data collection approach.

Approximately 61 papers opted for a manual approach, using web- scraping or web-crawling to collect customer reviews from popular E-commerce websites like Amazon. The other publications used publicly available datasets provided by companies, previous research works, or web-based communities like GitHub or Kaggle. As can be seen in Fig. 2.4, the distribution of datasets is greater than the number of papers analyzed. The reason for this is that several papers used multiple different datasets to study how a certain model performs on customer reviews from different product areas. Datasets under the category of “Other” include datasets from LG electronics, MPQA, Epinions, Enron, Ciao, as well as a few others.

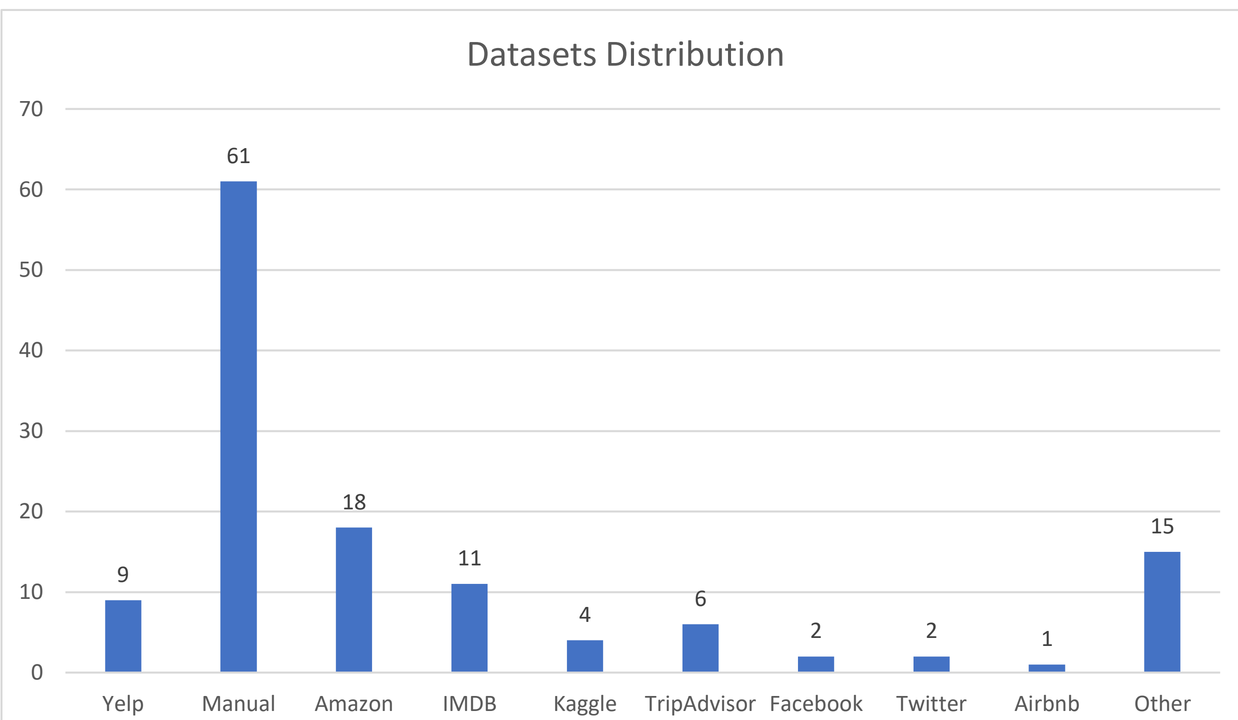

Figure 2.4: Distribution of the different datasets used for customer review analysis

Now that we know some information on what datasets are being used in the customer analysis field and how these datasets are being collected (manual collection vs. publicly available), it is important to delve deeper into these datasets. Some large E-commerce websites like Amazon offer product reviews for a whole slew of different product categories (electronics, kitchen, books, etc.). Stating that a study used the Amazon dataset does not detail what data was used. So, we also collected information about the product areas used in each work.

As shown in Fig. 2.5, datasets consisting of customer reviews for the electronics category are some of the most popular. The “General” category is used to define that multiple product categories were employed into the dataset. This is the most popular category with 34 papers using datasets that contained customer reviews from two or more product areas. Research works that used multiple different datasets for testing a model’s performance would fall under this “General” category in most scenarios. Moreover, works that used datasets from large E-commerce websites like Amazon would most likely also fall under this “General” category unless a sole product area was the focus. The reason for many different product areas used certainly has to do with validating a machine learning model. Although a model may perform well on customer reviews for electronic purchases, it may not perform well on customer reviews for restaurants. The use of multiple product areas in a work is an attempt of validating that a model is adaptive to any set of customer reviews, despite review area. The “Other” category is used to define less-common product areas such as: product safety, cosmetics, research articles, clothing, and a few others. Moreover, several papers did not define what product area was being used and these are also classified under the “Other” category.

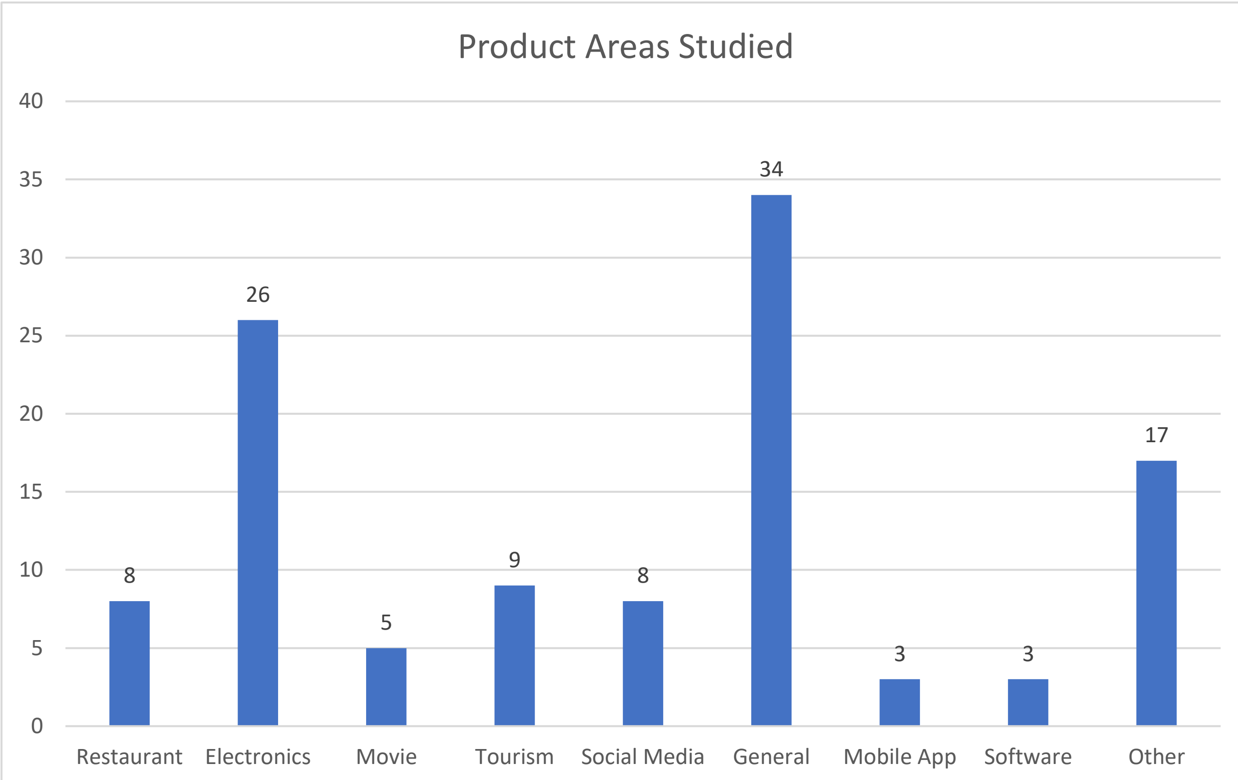

Figure 2.5: Count of papers using specific product areas for customer review analysis

2.4 Machine Learning Models

Now that the various datasets and product areas for customer review analysis have been defined, it is time to explore what machine learning models or methodologies have been used to analyze customer reviews.

Before delving too deeply into the specific models and methodology applied on the datasets detailed in the previous section, we wanted to know if this data was manipulated or pre-processed for training purposes. Out of the 111 papers analyzed in detail, every single one utilized some form of pre-processing techniques. Although the focus of this is not to detail pre-processing techniques used, we will mention a few popular approaches used for pre-processing purposes. They are as follows:

- Stemming - provides mapping of connected words into a standard root form [15].
- Data extraction - method of removing the unwanted information fields from information files before the cleanup method [16].
- Information cleanup - removing the noisy data from files [16].
- Stop word removal – removing common terms that convey little information

These, of course, are not all the pre-processing techniques that were present in the 111 papers analyzed but they were some of the most popular approaches in our systematic analysis of these works.

Now, let us consider the most popular machine learning approaches that were employed for customer review analysis. The focus of this analysis would be to understand what machine learning classifiers are being used in this field, despite the purpose for classification (sentiment analysis, product ranking, etc.). Fig. 6 depicts the most common classifiers used and how many papers utilized each classifier. The purpose of many of these research works was to determine what machine learning model performs best on customer reviews so the ratio of papers to classifiers is not 1:1.

As shown in Fig. 2.6, Support Vector Machines [61], Neural Networks [62], and Logistic Regression [63] models are some of the most widely used classifiers for customer reviews. Classic machine learning classifiers like Naïve Bayes [64], and Decision Tree [65] are still being used, but it seems they are being outperformed or out used by the newer classification models. When it comes to textual analysis, Naïve Bayes is known to be the standard. Most papers that incorporated Naïve Bayes into their work were using it as a baseline to determine the performance when compared with other classifiers/models.

However, there is no such thing as “one classifier fits all” when it comes to classification. Some papers use multiple classifiers [3] to determine which classifier performs best on a certain dataset. Moreover, the purpose of the classification affects which classifiers are the best. For example, research works that were using customer reviews for the purpose of classifying them into positive or negative categories mostly used standard classifiers like Naïve Bayes, decision tree, etc. for binary classification purposes. However, other works that attempt multi-class classification use more robust classifiers like Neural Network or Ensemble classifiers.

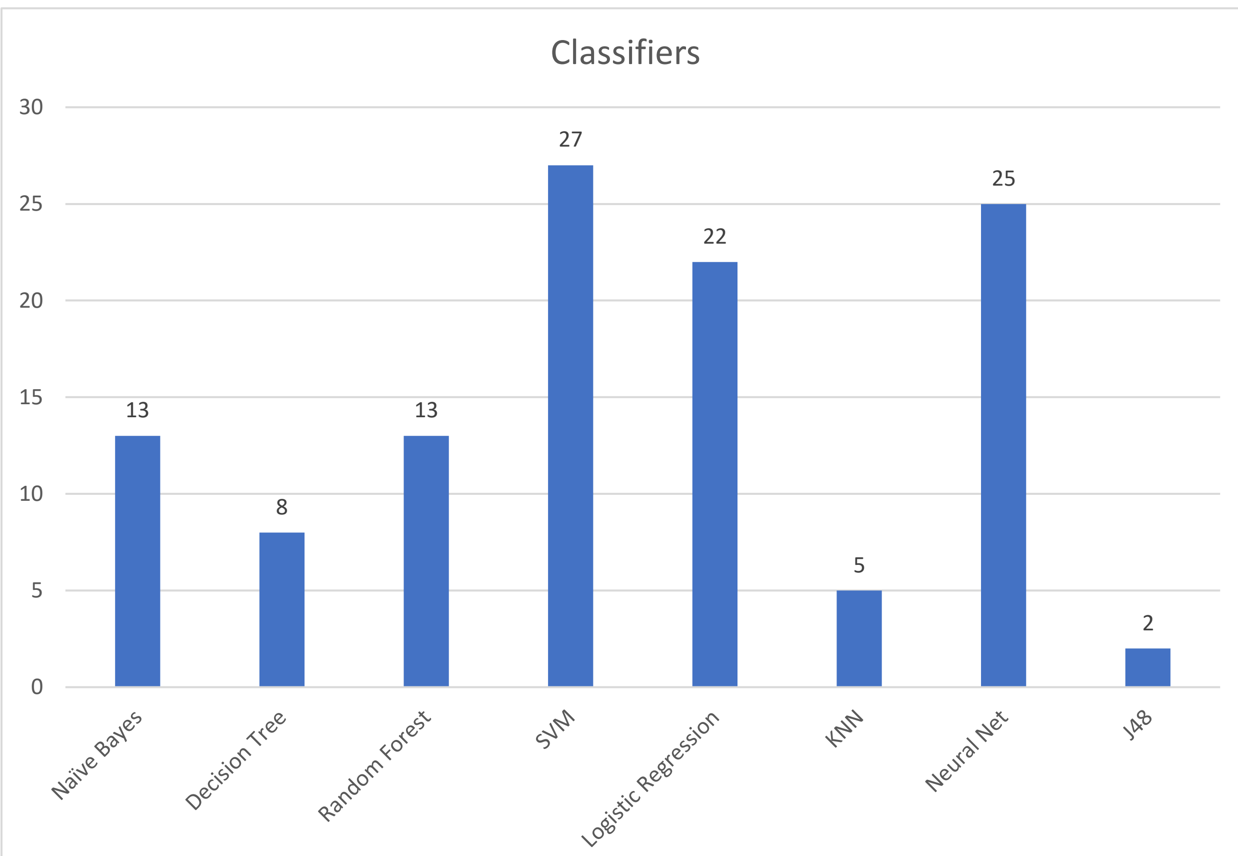

Figure 2.6: Distribution of classifiers used for customer review analysis

In the next section, we consider the overarching purpose of these works in the customer review analysis field. We will define the goals of the research and categorize these in a way that allows readers to understand what the various research pathways in customer review analysis currently exist.

2.5 Application Areas

As defined by this systematic analysis, customer review analysis applications are current areas of research that are being investigated using machine learning tools. While analyzing the set of papers, the purpose of each paper was extracted. Although each paper strives to serve a unique purpose, we group the papers that use a similar underlying methodology into categories to generalize the many topic areas of customer review analysis. The Fig. 2.7 details some popular topic areas currently being studied. In the next section we will introduce specific application areas that fall under each topic area's domain.

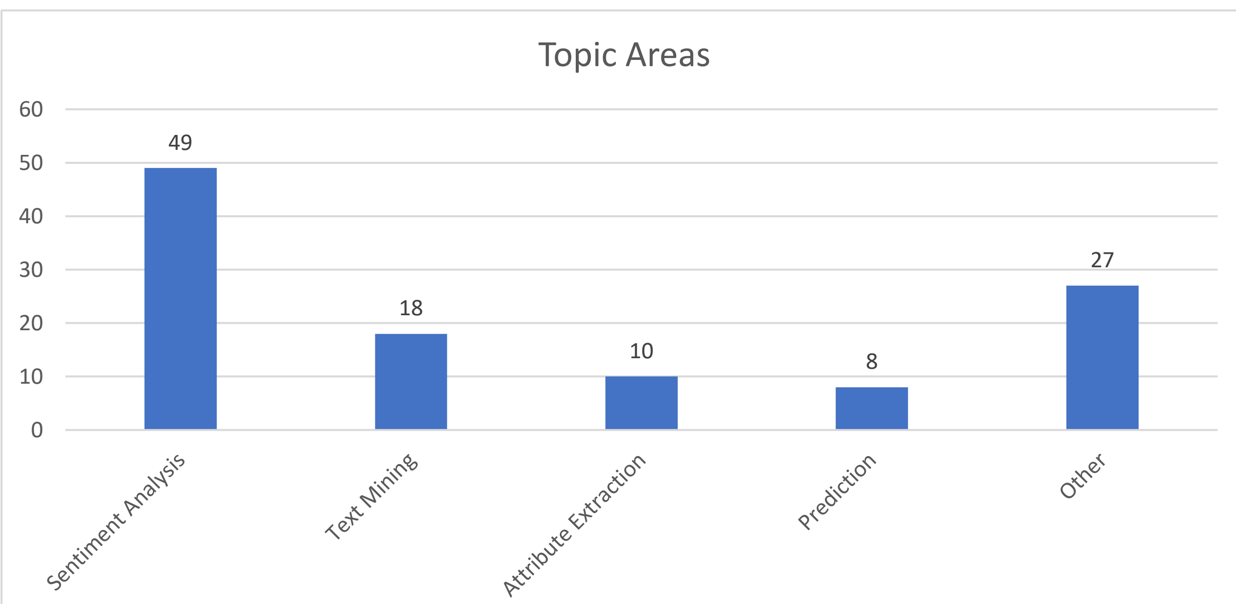

Figure 2.7: Distribution of topic areas using customer review analysis

2.5.1 Sentiment Analysis

Sentiment analysis is an ever-so popular topic area in the customer review analysis field. Rather than detail the specific purpose of each paper that incorporated sentiment analysis, we will discuss a few application areas that employed sentiment analysis techniques for customer review analysis. Some papers used sentiment analysis techniques to demonstrate new or effective classification of customer reviews using already existing/ensemble classifiers or new approaches that incorporate incremental learning, semi-supervised learning, hybridization, or other techniques [17-23]. Other works used sentiment analysis to simply compare classifiers on customer review data as in the case of [3] where SVM [56] and Naïve Bayes [57] classification models are comparatively studied. Sentiment analysis is certainly a topic area of customer review analysis that is of current interest.

2.5.2 Text Mining

Text mining is another prominent topic area in the realm of customer review analysis. Some of these works classified under text mining could also fall under the sentiment analysis category. For example, the authors in [24-27] use text mining as an approach for determining customer opinions on products or services in a way that is like sentiment analysis. On the other hand, researchers have also used text mining for determining review helpfulness or usefulness [28-30] , product ranking [31-34], building customer review process models [35], as well as constructing unique tools that aid in customer review analysis [36].

2.5.3 Attribute Extraction

Attribute extraction is another category that could fall under the realm of text mining or opinion analysis in some cases. In the case of [37, 38] attribute extraction is used to determine customer opinions from reviews. In [39] attribute extraction techniques are used to elicit usability and user-experience information about a product from reviews. In [40-43] attribute extraction has been used to determine popular product features, categorize college and universities based on feedback, analyze research works on information system success and failure, as well as attempt to disambiguate customer reviews, respectively. Lastly attribute extraction has also been used in customer review summarization. [44-46] used attribute extraction methods to extricate product features or opinions detailed in customer reviews to aid in generating concise and informative summaries.

2.5.4 Prediction

The prediction category encompasses any work with a goal of predicting a feature other than sentiment. This is a topic area that seems to be under-explored and has plenty of room for future research. [4, 47] both deal with predicting the success of a product based on customer reviews. Rather than dealing with the success of a product, [48] predicts the success of near-weekend ticket sales. Others [49] attempt to predict market segmentation and travel choices for Spa hotels rather than product success. As has been noted previously, determining helpfulness of customer reviews is a current area of interest and [28] studies how review helpfulness can be predicted using a scripts-enriched regression model. Interestingly, [50] demonstrates a computational intelligence approach to predict customer review ratings using the textual reviews left by customers. Product rating suggestion is a real-world application of this research and has great potential for E-commerce sites.

2.5.5 Other

The “Other” category encompasses works that could not be categorized by the four chosen categories. Rather than having many categories, we wanted to focus on the most prominent topic areas and what applications these topic areas contain when it comes to customer review analysis. Some works categorized as “Other” include:

- Rule indication [51]
- Semantic analysis [52]
- Product recommendation [53]
- Hierarchy construction [54]
- NLP vs. injection attacks in customer reviews [55]

Although there are many others, these are a few examples of works contained in the “Other” column.

2.6 Research Directions

There are several research directions for future research in the field of customer review analysis. One of the most prevalent issue with customer review analysis is the amount of dimension and noise associated with free-text, unstructured reviews. In every application area, there are vast amounts of customer reviews that must be analyzed, and the challenges discussed prior make detailed analyses of large amounts of data less efficient. Some of these issues are addressed in [50], but there is certainly room for more research in this area.

A more interesting future research direction could be the application of certain customer review analysis tools in real-world systems. For example, some research works have used customer reviews to enhance product ranking systems [31, 33, 34], determine helpfulness of reviews [1, 28, 30], as well as a few other application areas. However, most papers reviewed in this analysis solely dealt with basic topics like sentiment analysis or text mining without discussing how to implement these studies into real-world applications. The field of using customer reviews to predict/classify other review attributes is a field that has not been studied too deeply. Moreover, with the help of customer review analysis and predictive modeling, other review attributes can be suggested to the user as well as using attribute extraction of textual reviews to highlight keywords or phrases for other users to see without needing to read an entire review.

Lastly, future work could include evaluating existing frameworks using big data analytics or applying existing customer review analysis frameworks to solve classification or predictive modeling tasks in areas beyond customer reviews. Moreover, improving performance of existing models or frameworks by utilizing evolutionary computation algorithms should also be considered.

Chapter 3

Related Literature on Star Rating Prediction of Customer Reviews

Customer review analysis is a very broad field with many related sub-areas. Now that we have a better understand of what falls into the customer review analysis area of research, let's delve into a more related sub-area consisting of review rating prediction for the English language. This section aims to analyze four highly related works with a goal of predicting customer review star ratings.

The authors of [66] present a neural network approach. They apply a transfer learning technique from a large volume of Amazon customer reviews onto three mobile application review datasets from the Google Play Store. This technique adopts the use of Long Short-Term Memory networks (LSTMs) where the model takes in a training sentences as an input sequence and outputs a sentiment analysis or polarity score from 1 to 5. This method was able to achieve a predictive accuracy of 87.61% compared to the original ratings of the users which is quite promising. The authors also noted that the correlation of SI with UR was significant for both predicted and original ratings but there was a stronger correlation with predicted UR as opposed to the original one

The use of various single and ensemble classifiers to identify polarity of reviews from a 2017 Yelp customer review dataset is investigated in [67]. The single models utilized in this study consisted of Naïve Bayes, Nearest Neighbor, Generalized Linear Model, Support Vector Machine, and Multilayer Perceptron. On the ensemble side, the authors used Bagging

Predictors, Random Forests, ADA Boost, and Gradient Boosting. Experimental results demonstrated that SVM, Logistic Regression, and Multilayer Perceptron were the top performing single models in terms of accuracy, precision, recall, and F-measure but their performance could be increased by implementing these classifiers as a baseline into an ensemble model. In addition to identifying top performing classifiers, the authors noted that users who give a 3-star rating tend to give a more positive review in terms of polarity.

The authors of [68] also investigate the performance of various classifiers. In this work, they evaluate the performance and computational time of classifiers like Multinomial Naïve Bayes, Bigram Multinomial Naïve Bayes, Trigram Multinomial Naïve Bayes, Bigram-Trigram Multinomial Naïve Bayes, and Random Forest. This work utilized customer reviews from Yelp to train and test the models. When performance of each of these classifiers is compared, Random Forest obtains the highest accuracy but at the expense of computational time. Bigram-Trigram Multinomial Naïve Bayes obtained an accuracy score of just 2% less than Random Forest with 12% faster computation. The accuracy scores for Random Forest and Bigram-Trigram Multinomial Naïve Bayes were 0.675 and 0.698, respectively.

A computational intelligence framework for efficiently learning ratings is proposed in [69] by addressing two main problem areas: dimension and imprecision of customer ratings data. Specifically, the authors integrate techniques like Singular Value Decomposition (SVD), Fuzzy C-Means (FCM) and ANFIS to build a synergetic approach. For experimentation, a movie dataset with textual reviews and numerical ratings was used. The results demonstrated that their combined approach yields better learning performance than other rating predictors based on the artificial neural network and FCM algorithm.

At the time of this writing, there is no study that incorporates an ensemble averaging or majority voting approach to customer rating prediction for reviews written in the English language. This study provides a novel contribution to literature by implementing this. We present the proposed approach in the following Chapter.

Chapter 4

Hybrid Voting Classifier Methodology

This chapter describes our hybrid voting classifier methodology and details the Amazon customer review dataset, the baseline classification approach, as well as the eventual star rating prediction with the help of a novel voting classifier. The details for the novel voting approach to star rating prediction can be seen in section 4.5. The following sections will elaborate on what data was used, how textual customer reviews were pre-processed, what tools were used for feature extraction, as well as what machine learning models were used and how these models were evaluated. In the following chapter, we will detail the experimental results by comparing the voting classifier to the baseline classifiers in terms of accuracy.

4.1 Hardware/Software Setup

Many of the procedures detailed in the following sections were extremely computationally intensive, and in return, required a significant number of hours to be performed. To alleviate some of this pressure from a personal machine, a dedicated virtual machine was used. Hardware wise, the virtual machine was hosted by a Gigabyte R282-Z93-00 server using two AMD EPYC 7282 16-Core Processors with 32 threads and a base clock speed of 3.2 Ghz. Moreover, this machine was equipped with 32 GB of DDR4 RAM running at 2667 MT/s.

On the software side, the virtual machine utilized VMware vSphere 7.0 as the hypervisor and the was running the Windows 10 Enterprise 64-bit operating system.

4.2 Amazon Customer Review Dataset

The customer review data used in this experimental study were obtained from [70]. This resource contains several US Amazon customer review datasets based on product category. This dataset contains the customer review text with accompanying metadata from Amazon.com marketplace from the year 1995 until 2015 for products listed under the video game category. Although not all data columns were used in this experiment, they can be seen in Table 4.1. Overall, the Amazon Customer review dataset for the video game category contained over 1.7 million customer reviews. With hopes of eliminating reviewer bias, only rows marked as “verified purchase” were considered which brought down the review count to slightly over 1.1 million. In this study, only the Star_rating and Review_body columns were used as input for machine learning models. We present some sample data for these two columns in Table 4.2.

Now that we understand a little more about what is contained in this dataset, let’s take a deeper look into star rating distribution. The distribution of reviews based on star rating are detailed in Figure 4.1. As can be seen, there are a far greater number of 5-star reviews than there are lower star-rated reviews. Well over 50% of the data is comprised of 5-star reviews where 2-star reviews make up the minority class.

Table 4.1: General description of data columns in the Amazon Review dataset

Marketplace	2 letter country code of the marketplace where the review was written.
Customer_id	Random identifier that can be used to aggregate reviews written by a single author.
Review_id	The unique ID of the review.
Product_id	The unique Product ID the review pertains to.
Product_parent	Random identifier that can be used to aggregate reviews for the same product.
Product_title	Title of the product.
Product_category	Broad product category that can be used to group reviews (also used to group the dataset into coherent parts).
Star_rating	The 1-5 star rating of the review.
Helpful_votes	Number of helpful votes.
Total_votes	Number of total votes the review received.
vine	Review was written as part of the Vine program.
Verified_purchase	The review is on a verified purchase.
Review_headline	The title of the review.
Review_body	The review text.
Review_date	The date the review was written.

Table 4.2: Sample Star_rating and Review_body data from Amazon Review dataset

Star_rating	**Review_body**
5	Used this for Elite Dangerous on my mac, an amazing joystick. I especially love that you can twist the stick for different movement bindings as well as move it in the normal way.
5	Loved it, I didn't even realise it was a gaming mouse, I typed in "silent mouse" and selected this one. It is perfect and looks pretty cool as well. Now my boyfriend's gaming is wonderfully comfortably silent :) . Think I might just get one for myself.
1	poor quality work and not as it is advertised.
3	nice, but tend to slip away from stick in intense (hard pressed) gaming sessions.
4	Great amiibo, great for collecting. Quality material to be desired, since its not perfect.

Figure 4.1: Count of Reviews by Star Rating

Next, to better analyze commonalities across reviews of the same star rating, we generate word cloud images for each star rating. Word clouds are used to give an intuitive visual overview of a text by depicting the most often occurring words. This summarization is helpful to learn about the number and kind of topics present in a body of text. This sort of statistical overview is achieved by correlating the font size of depicted words with word frequency [71]. The generated word clouds for one, two, three, four, and five star rated reviews can be seen in figures 4.2, 4.3, 4.4, 4.5, and 4.6 respectively. Since the dataset is comprised of reviews from the video games category, it is natural that we would see words like "game" or "play" in large text for every star rating class. Looking deeper into each word cloud image, we see that words like "good" or "bad", which carry a neutral or negative sentiment, being prominent in the lower star rating classes. Likewise, higher star rating classes contain words like "love" or "fun", which carry a strong positive sentiment.

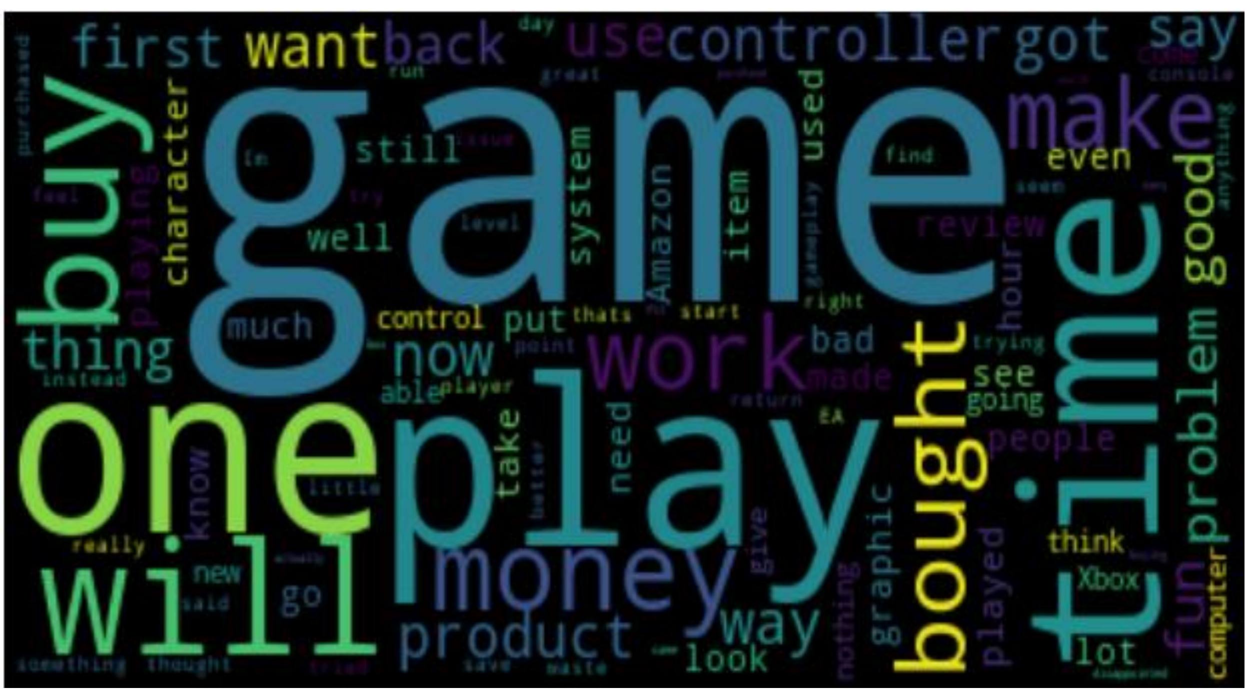

Figure 4.2: Word cloud generated for 1-star reviews

Figure 4.3: Word cloud generated for 2-star reviews

Figure 4.4: Word cloud generated for 3-star reviews

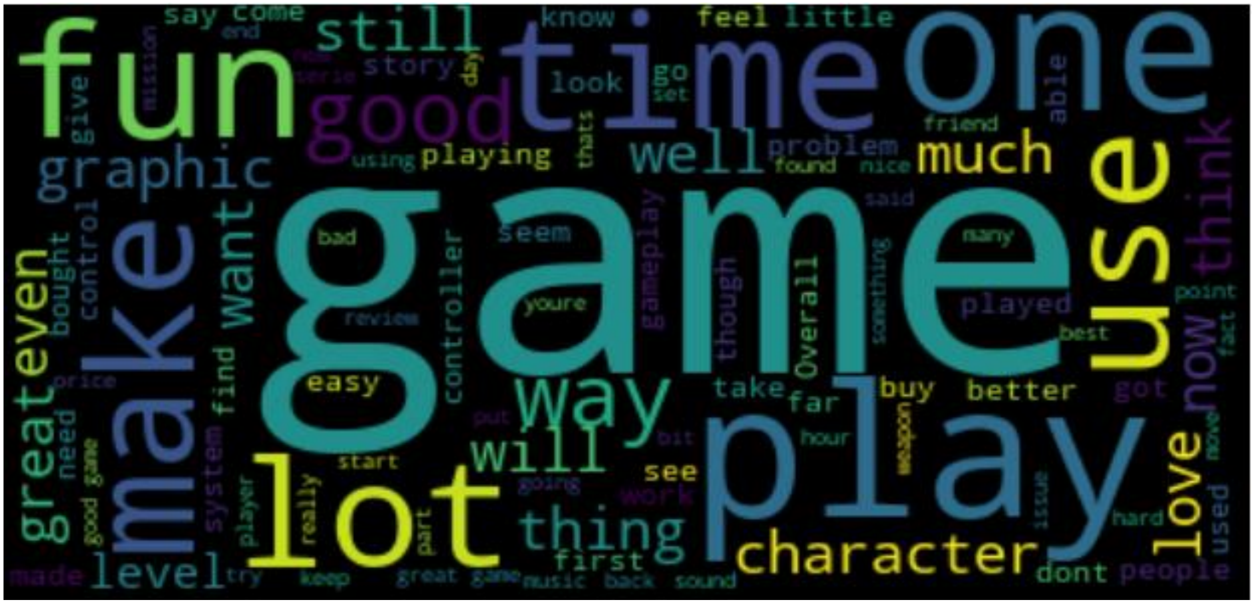

Figure 4.5: Word cloud generated for 4-star reviews

Figure 4.6: Word cloud generated for 5-star reviews

4.3 Machine Learning Classifiers

A set of five machine learning classifiers were chosen to define a baseline for the experiment. These classifiers were selected based on results from previous works discussed in Chapter 3. The classifiers consisted of the following: Random Forest [72], Linear Support Vector (SVC) [73], Multinomial Naïve Bayes [74], as well as Logistic Regression [75] and XGBoostClassifier [76]. The collection of these 5 classifiers, as well as their respective parameters, can be seen in table 4.3. Every classifier used was obtained from the Scikit-learn repository [77].

Table 4.3: Machine learning classifiers and parameters used

Classifier	**Parameters**
RandomForestClassifier	n_estimators=200, max_depth=3
LinearSVC	NONE
MultinomialNB	NONE
LogisticRegression	max_iter=100000
XGBClassifier	learning_rate=0.1

Before feature extraction, it was necessary to first clean or pre-process the reviews. The set of customer reviews that were used for experimental purposes exceeded 1.1 million. First, the text of all reviews was converted to lowercase. Then, all punctuation and stop words were removed from the review text. Data cleaning was done to ensure the text was in a more appropriate form for feature extraction.

For extracting features, a TF-IDF vectorizer was used. TfidfVectorizer is also a component of the Scikit-learn repository. Using the product of term frequency and inverse document frequency, the vectorizer converts a collection of raw documents into a matrix of features. These features are then used as input for the machine learning models.

4.4 Evaluation of Baseline Models

Firstly, it was necessary to evaluate the performance of the five baseline classifiers on the selected customer reviews. To do this, k-fold cross validation was utilized [78]. Five was chosen for the value of K in this evaluation. The purpose of performing 5-fold cross validation on the classifiers was to estimate the skill of the model on unseen data. This approach involves randomly dividing the training set into k groups, or folds, of approximately equal size. The first fold is treated as a validation set, and the method is fit on the remaining $k - 1$ folds. From this process, we can obtain metrics for each classifier across all folds. The mean accuracy and standard deviation of each classifier can be seen in table 4.4. Likewise, a strip plot for the accuracy of the 5 baseline classifiers across the folds of cross validation was generated and can be seen in figure 4.7. Logistic Regression and LinearSVC were the top performers in this cross validation with high accuracy scores, when compared to the others, as well as modest values for standard deviation. Surprisingly, Random Forest obtained the lowest accuracy score out of the set. As we will discuss in the next section, the results of this cross validation determine the selected classifiers for use in the ensemble voting model.

Table 4.4: 5-fold cross validation results for baseline classifiers

Classifier	Accuracy	Standard Deviation
LinearSVC	0.697	0.0009
LogisticRegression	0.713	0.0007
MultinomialNB	0.652	0.0003
RandomForestClassifier	0.622	0.0000
XGBClassifier	0.667	0.0006

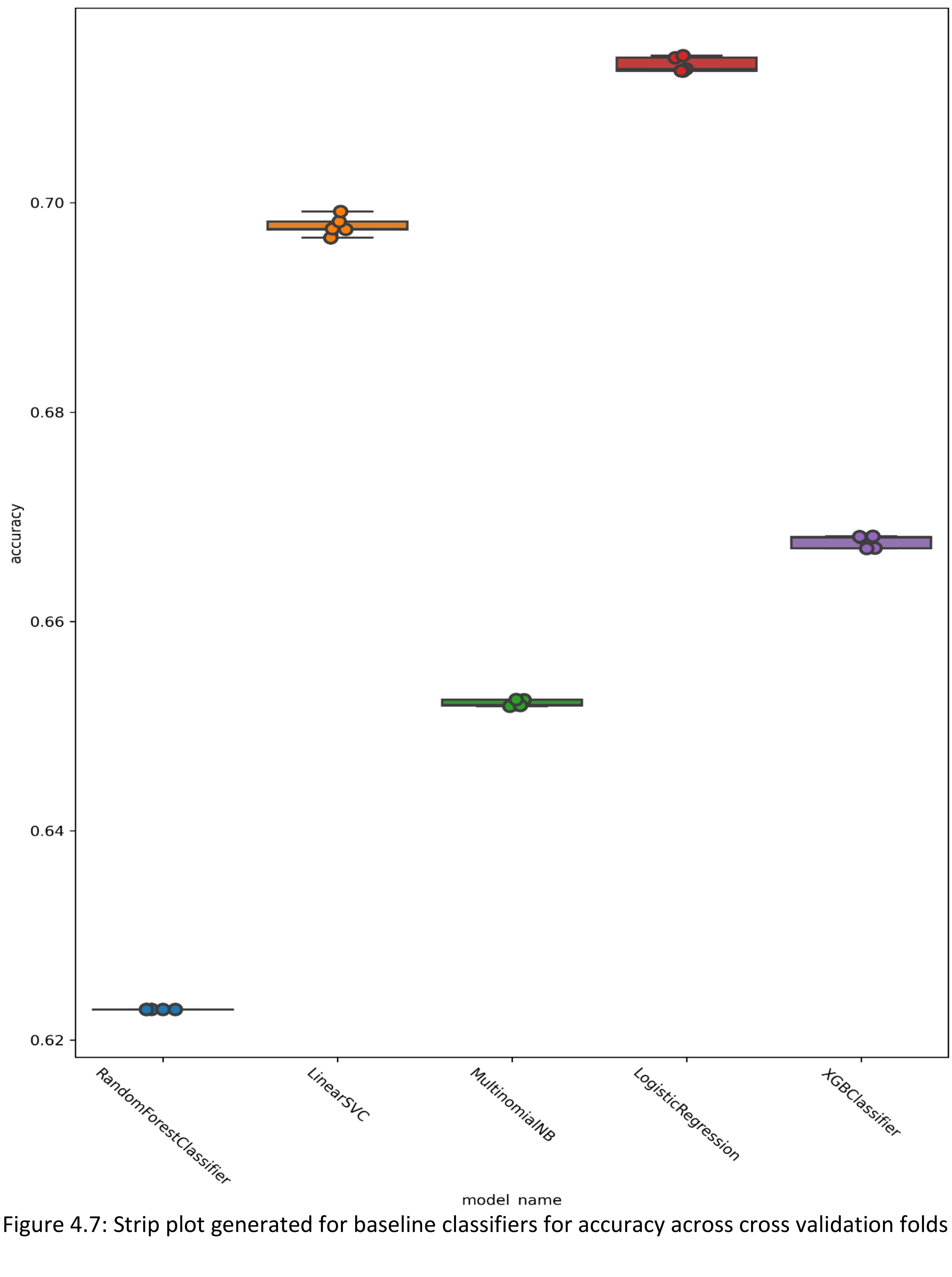

Figure 4.7: Strip plot generated for baseline classifiers for accuracy across cross validation folds

4.5 Ensemble Voting Model Approach and Validation

The ensemble voting model is the novel contribution of this work. Before detailing the baseline classifiers chosen to be incorporated into this ensemble, we need to investigate how an ensemble voting model works. Voting classifiers support two main types of voting techniques: hard voting and soft voting. With hard voting, the predicted output class is the class with the highest probability of being predicted by each of the individual classifiers. In soft voting, the predicted output class is based on the average of probability assigned to each class by each classifier. Although these voting techniques differ in the details, the majority class wins and is the final prediction of the voting classifier. The voting classifier utilized was obtained from Scikit-learn. The big idea behind a voting classifier is that if we include a variety of models into it, any error made by one classifier can possibly be resolved by the other(s). For this study, only the hard-voting technique was incorporated into the voting classifier.

Rather than just taking the best x classifiers and using them in a voting classifier, it was decided that three voting classifiers would be generated. First, a voting classifier with the two best performing classifiers in the cross validation step, Logistic Regression and LinearSVC, was created. Then, a voting classifier consisting of the two worst classifiers, Naïve Bayes and Random Forest, was made. Lastly, a balanced model was made. This voting classifier consisted of Logistic Regression, XGBClassifier, and Random Forest. For simplicity, these voting classifiers are labeled 1, 2, and 3, respectively. As can be seen in table 4.3, these classifiers comprised the top, middle, and bottom of the baseline cross validation test in terms of accuracy.

To get an estimate of the performance of each of these three, very different, voting classifier configurations, the same 5-fold cross validation test was performed. For each fold in the cross validation process, an accuracy score is generated, and these values can be seen in figure 4.8. The results, in the form of mean accuracy and standard deviation, for each of these classifiers can be seen in table 4.5.

It is important to keep in mind that these accuracy values were generated for the purpose of estimating model skill for baseline model selection and voting model validation. However, as we will detail in section 5, these results are comparable to those that will be generated from testing the model on the test set of the data. Out of the 1,164,785 reviews, approximately 80% (931,828) reviews were used for training and validation purposes and 20% (232,957) reviews were used to obtain final results via model testing.

Table 4.5: 5-fold cross validation results for Voting classifiers

Classifier	Accuracy	Standard Deviation
Voting Classifier 1	0.712	0.0004
Voting Classifier 2	0.652	0.0003
Voting Classifier 3	0.671	0.0004

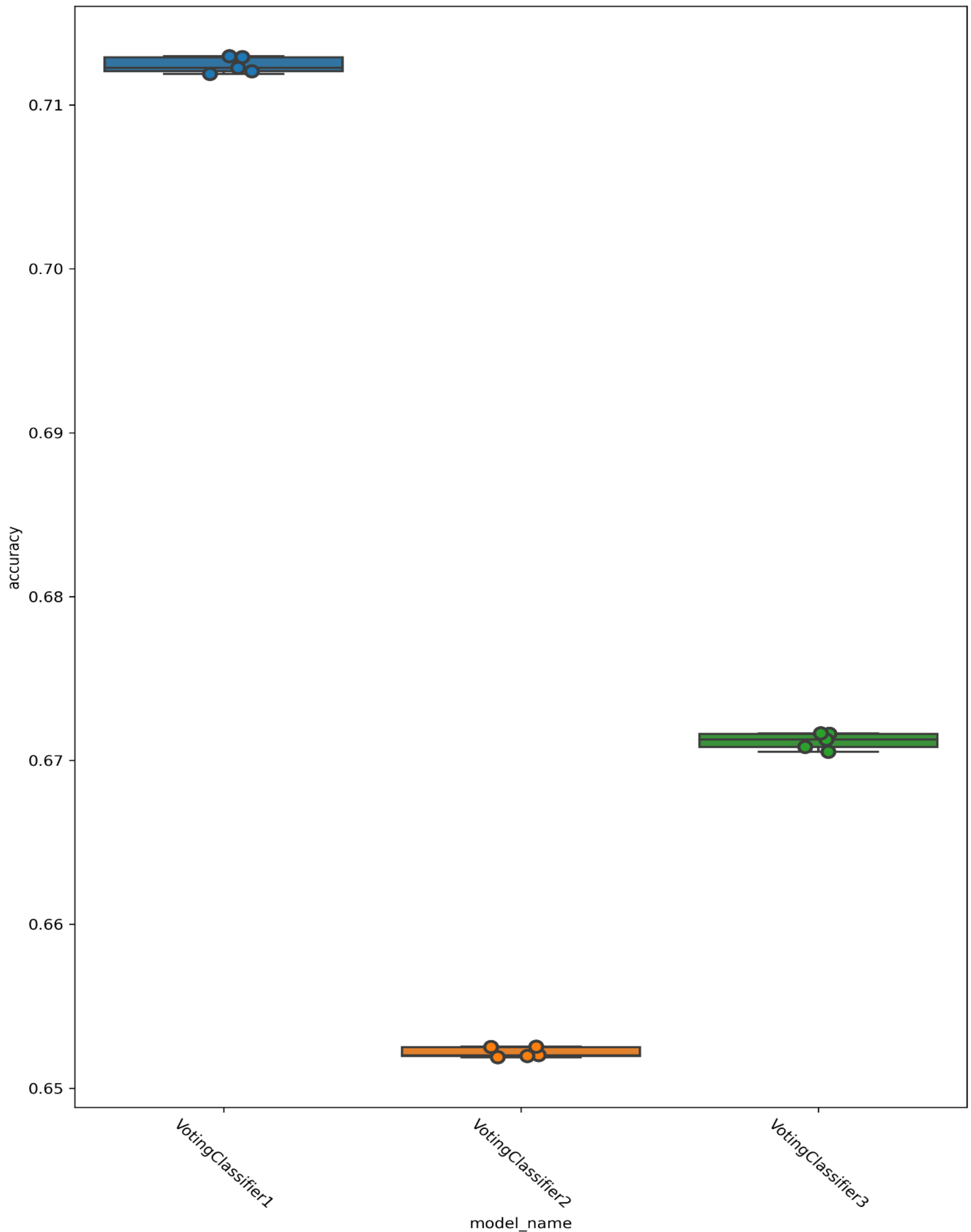

Figure 4.8: Strip plot generated for voting classifiers for accuracy across cross validation folds

Chapter 5

Experimental Results

In this section, the results obtained from the machine learning models are analyzed and discussed. For final evaluation, the various baseline models, as well as the ensemble voting models, are analyzed via their predictive performance. This experiment involved training the models on the training set (the same set used for 5-fold cross validation) and testing them on the held-out 20% of unseen Amazon customer review data. To quantify each model's performance, accuracy and F1 scores were generated.

5.1 Model Comparison and Evaluation

Figure 5.1 compares the accuracy scores generated from the cross validation and final experiment phases. As can be seen, the cross validation did an excellent job at estimating model performance. At the most, there was a discrepancy of 1 percentage point. This discrepancy can be seen in the values for Random Forest and VotingClassifier2.

The final test accuracy values for each individual classifier, as well as the voting classifiers, can be seen in figure 5.2. To better understand the performance of each of the ensemble voting classifiers, we will compare their performance against the baseline classifiers that were used in the voting approach.

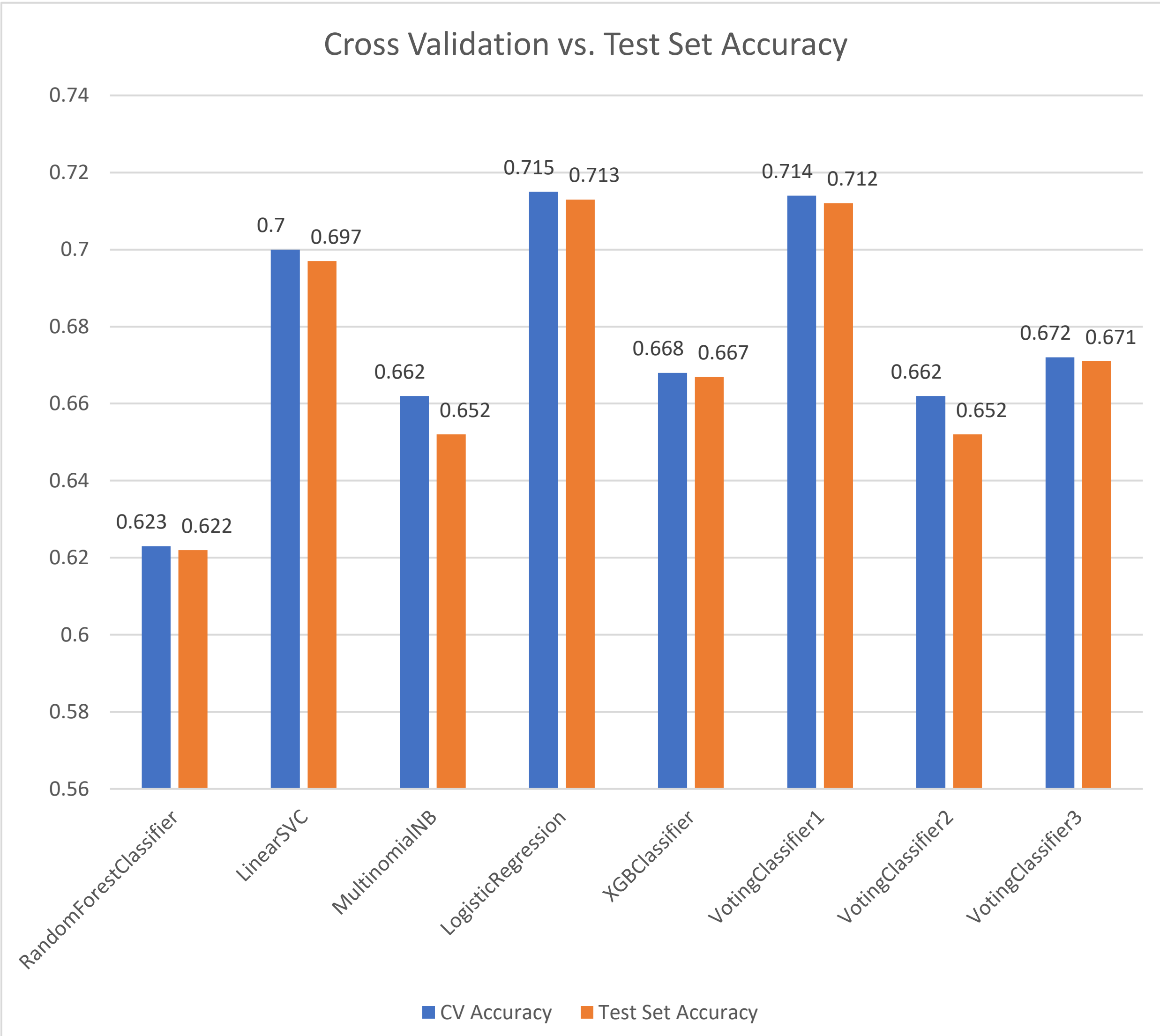

Figure 5.1: Accuracy of machine learning models on cross validation vs. test set

VotingClassifier1, as previously noted, was comprised of both a logistic regression and linear support vector classifier. Looking at the accuracy values for VotingClassifier1, LogisticRegression, and LinearSVC, we see that this voting ensemble performed comparably to the individual performance of LogisticRegression while slightly outperforming LinearSVC. The accuracy values for VotingClassifier2 and VotingClassifier3 seem to follow a similar trend.

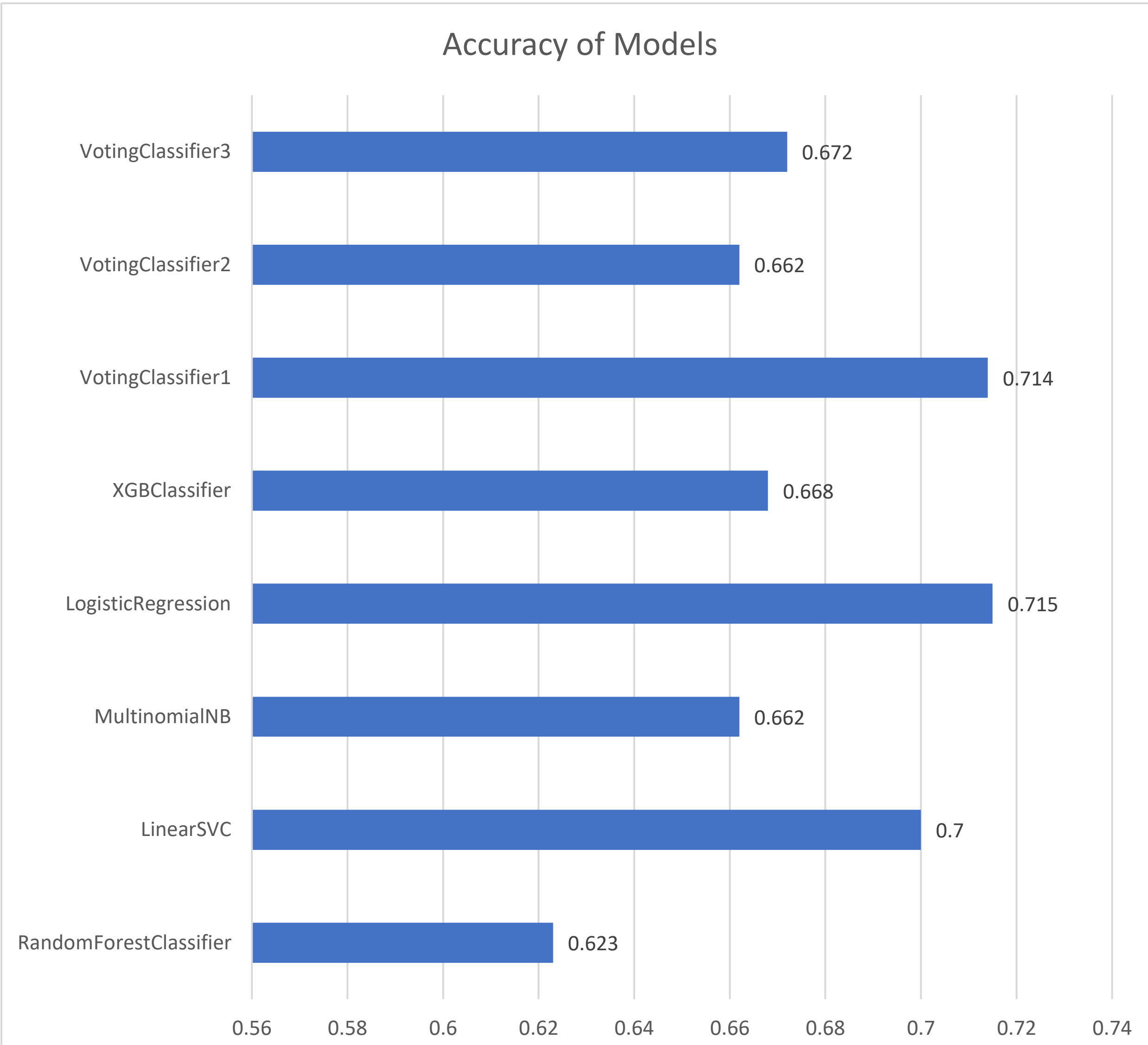

Figure 5.2: Accuracy of machine learning models on test set

VotingClassifier2 was comprised of RandomForestClassifier and MultinomialNB. As can be seen, VotingClassifier2 was able to match the performance of MultinomialNB exactly while outperforming RandomForestClassifier. The final voting model, VotingClassifier3, which was designed to consider the best, worst, and middle performing classifiers, fell short of the best performing classifier but was able to outperform the middle and worst performing baseline models.

Although the hopes of this experiment were to generate a classifier that outperformed the results of the baseline in terms of accuracy, this was not entirely the case. The average accuracy for the five baseline models came out to be approximately 67%. Taking that into account, VotingClassifier1 and VotingClassifier2 were able to match and exceed that average accuracy, respectively. Averaging the accuracy for all three voting models, they were able to outperform the average baseline accuracy value by 0.8%. However, despite the voting models being able to outperform baseline averages, no voting model was able to exceed the accuracy score of the Logistic Regression model on its own.

However, accuracy alone is not sufficient to determine the sheer performance of a model, especially given the unbalanced dataset that was used. To better evaluate model performance, an F1 score was computed for each model. This metric is a weighted average of precision and recall, taking both false positives and false negatives into account. The F1 scores for each model can be seen in table 5.1. Like accuracy, LinearSVC and Logistic Regression classifier have the highest F1 score. Likewise, the two worst performing classifiers in terms of accuracy, Random Forest and Naïve Bayes, generated the lowest F1 score.

Accuracy and F1 score are common metrics for evaluating the performance of machine learning models but they do not always tell the full story. Especially in the case where the majority class comprises over 70% of the dataset, it is very likely that some models are overfitting. By looking at the confusion matrices presented in figure 5.3, there are some clear indications of overfitting in one the baseline models. RandomForest predicted that every customer review was rated 5 stars. While it was the case that most reviews were rated 5 stars, it is obvious the RandomForest is overfitting. Although the rest of the models show no signs of obvious overfitting, it should be noted that Naïve Bayes and VotingClassifier2 have the exact same scores for accuracy and F1 as well as matching confusion matrices. From this, we can conclude that all prediction votes from Random Forest were rejected when building VotingClassifier2. The confusion matrices for the voting classifiers can be seen in figure 5.4. Although VotingClassifier1 yields similar results when compared with its baseline classifier, Logistic Regression, the confusion matrices are not exact and show that the ensemble voting model is taking both baseline model predictions into account. Similarly, VotingClassifier3 seems to be taking all three baseline classifiers into consideration.

Table 5.1: F1 scores for classification models

Model Name	F1 Score
LinearSVC	0.648
LogisticRegression	0.670
MultinomialNB	0.558
RandomForestClassifier	0.479
XGBClassifier	0.580
Voting Classifier 1	0.670
Voting Classifier 2	0.558
Voting Classifier 3	0.579

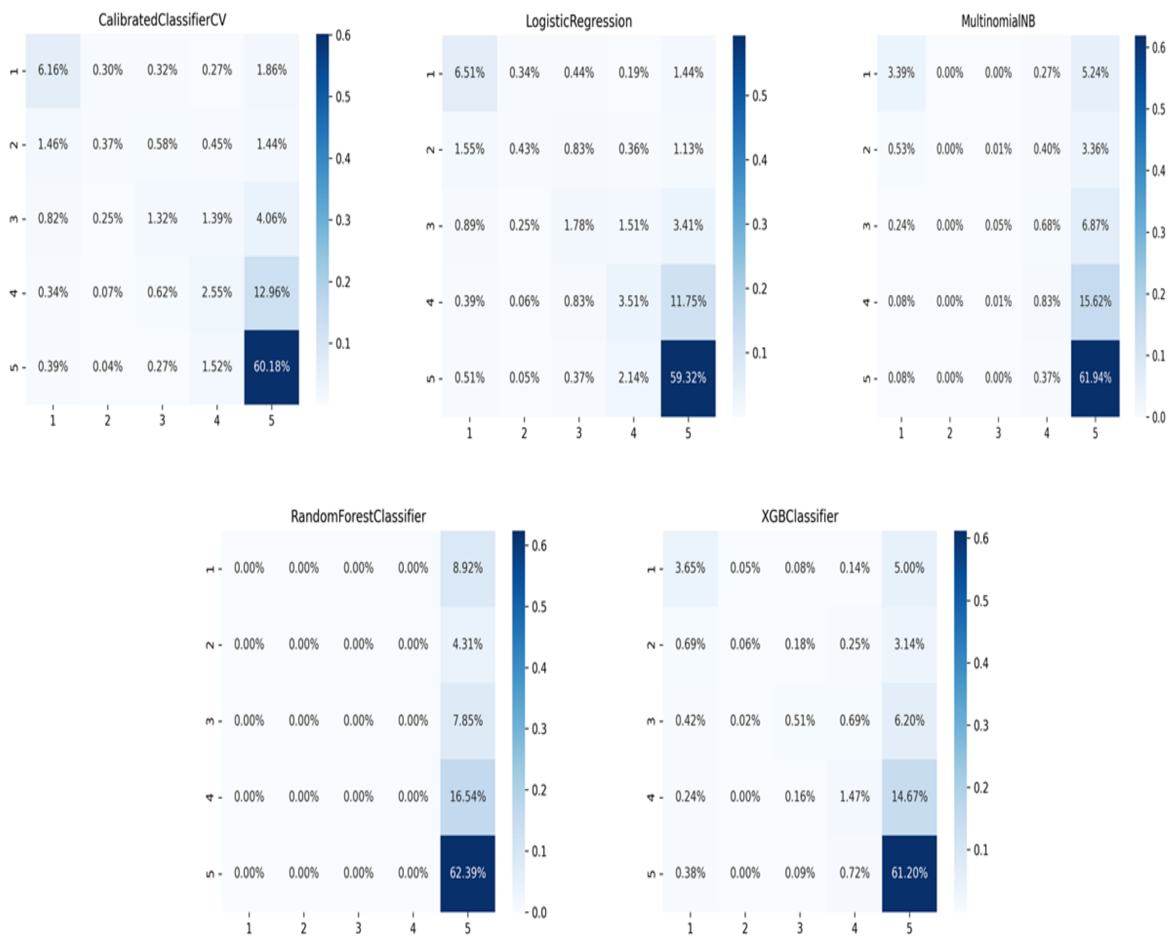

Figure 5.3: Confusion matrices for baseline models

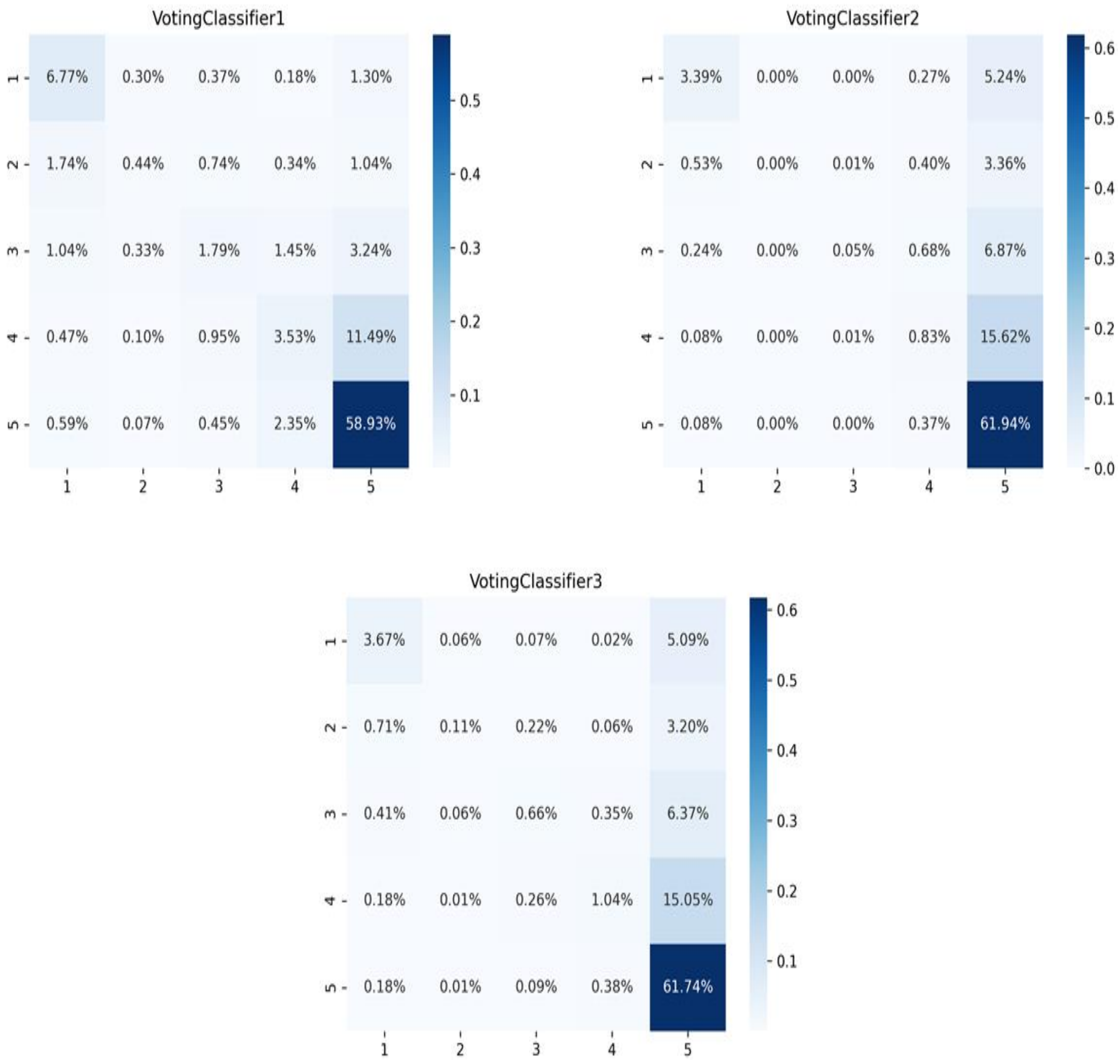

Figure 5.4: Confusion matrices for ensemble voting models

Chapter 6

Conclusions and Future Work

Due to the continuous increase in readily available online data as well as the growing importance of customer reviews, customer review analysis and star rating prediction are attracting increasingly more attention. However, there are very few works that have reviewed the current state of the customer review analysis area of research. As it is such, we systematically defined goals, questions, and protocols to guide the review process towards information extraction and analysis. We first identified common machine learning models that are employed on customer reviews. Next, we detailed common customer review datasets that these models incorporate. And lastly, we discussed the many topic/application areas that are incorporating customer review analysis.

In addition to the contribution of a literature review, we proposed an ensemble voting approach for star rating prediction of customer reviews. In our proposed model, we incorporate various baseline machine learning classifiers into an ensemble model that uses the highest probability class of each individual classifier in the ensemble to generate a final prediction. The hope of this approach is that by incorporating the predictions from more than one classifier, any error made by a single classifier can possibly be resolved by the other(s). We evaluated the performance of our novel approach using an Amazon customer review dataset, containing over a million reviews in total. We also compared the accuracy, F1 score, and confusion matrices of our ensemble models with the scores generated from the baseline classifiers. Our best performing ensemble voting model was able to perform just as well as the best performing individual classifier while employing the use of both Logistic Regression and LinearSVC.

In terms of future research directions, we are first going to increase the number of classifiers contained in the baseline. Likewise, the number of classifiers contained in the hard-voting classifier would increase. However, instead of applying the same weight to every classifier, we would assign voting weight for each classifier by analyzing results from the cross-validation stage to improve the predictive power of our model. Also, in this cross-validation stage, parameter tuning would be performed to fully optimize the capability of each individual classifier. Lastly, the use of a more balanced dataset or data scaling techniques will be utilized.

Bibliography

[1] B. M. Shoja and N. Tabrizi, “Tags-Aware Recommender Systems: A Systematic Review,” presented at the 2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD), May 2019, doi: 10.1109/bcd.2019.8884850.

[2] S. Tammina and S. Annareddy, “Sentiment Analysis on Customer Reviews using Convolutional Neural Network,” presented at the 2020 International Conference on Computer Communication and Informatics (ICCCI), Jan. 2020, doi: 10.1109/iccci48352.2020.9104086.

[3] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, “A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews,” presented at the 2020 International Conference on Contemporary Computing and Applications (IC3A), Feb. 2020, doi: 10.1109/ic3a48958.2020.233300.

[4] S. Narayanan, P. Samuel, and M. Chacko, “Improving prediction with enhanced Distributed Memory-based Resilient Dataset Filter,” J Big Data, vol. 7, no. 1, Feb. 2020, doi: 10.1186/s40537-020-00292-y.

[5] S.-W. Lee, G. Jiang, H.-Y. Kong, and C. Liu, “A difference of multimedia consumer’s rating and review through sentiment analysis,” Multimed Tools Appl, Mar. 2020, doi: 10.1007/s11042-020-08820-x.

[6] Y. Lee, J. Park, and S. Cho, “Extraction and prioritization of product attributes using an explainable neural network,” Pattern Anal Applic, vol. 23, no. 4, pp. 1767–1777, May 2020, doi: 10.1007/s10044-020-00878-5.

[7] G. Quattrone, A. Nocera, L. Capra, and D. Quercia, “Social Interactions or Business Transactions?What customer reviews disclose about Airbnb marketplace,” presented at the WWW ’20: The Web Conference 2020, Apr. 2020, doi: 10.1145/3366423.3380225.

[8] A. Sinha and V. Subrahmaniam, “Cerebro,” presented at the ICMLSC 2020: The 4th International Conference on Machine Learning and Soft Computing, Jan. 2020, doi: 10.1145/3380688.3380701.

[9] H. Liu and B. Wang, “Mitigating File-Injection Attacks with Natural Language Processing,” presented at the CODASPY ’20: Tenth ACM Conference on Data and Application Security and Privacy, Mar. 2020, doi: 10.1145/3375708.3380310.

[10] J. Kazmaier and J. H. van Vuuren, “A generic framework for sentiment analysis: Leveraging opinion-bearing data to inform decision making,” Decision Support Systems, vol. 135, p. 113304, Aug. 2020, doi: 10.1016/j.dss.2020.113304.

[11] S. Chatterjee, “Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach,” International Journal of Hospitality Management, vol. 85, p. 102356, Feb. 2020, doi: 10.1016/j.ijhm.2019.102356.

[12] S. Khedkar and S. Shinde, "Deep Learning and Ensemble Approach for Praise or Complaint Classification," Procedia Computer Science, vol. 167, pp. 449–458, 2020, doi: 10.1016/j.procs.2020.03.254.

[13] T. Van Nguyen, L. Zhou, A. Y. L. Chong, B. Li, and X. Pu, "Predicting customer demand for remanufactured products: A data-mining approach," European Journal of Operational Research, vol. 281, no. 3, pp. 543–558, Mar. 2020, doi: 10.1016/j.ejor.2019.08.015.

[14] F. R. Lucini, L. M. Tonetto, F. S. Fogliatto, and M. J. Anzanello, "Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews," Journal of Air Transport Management, vol. 83, p. 101760, Mar. 2020, doi: 10.1016/j.jairtraman.2019.101760.

[15] Rakesh Kumar and Vibhakar Mansotra, "Application of Stemming Algorithms in Data Retrieval- A Review", International Journal of Advanced Analysis in Technology and Code Engineering, pp. 418-423, 2016.

[16] Surbhi Anand and Rinkle Rane Aggarwal, "An Economical Algorithamic Rule for Information Cleanup of Net Files Mistreatment File Extensions", Internation Journal Of PC Application, vol. 48, no. 8, 2012.

[17] A. Sajeevan and L. K. S, "An enhanced approach for movie review analysis using deep learning techniques," presented at the 2019 International Conference on Communication and Electronics Systems (ICCES), Jul. 2019, doi: 10.1109/icces45898.2019.9002043.

[18] N. Aleebrahim, M. Fathian and M. R. Gholamian, "Sentiment classification of online product reviews using product features," The 3rd International Conference on Data Mining and Intelligent Information Technology Applications, Macao, 2011, pp. 242-245.

[19] A. Alrehili and K. Albalawi, "Sentiment Analysis of Customer Reviews Using Ensemble Method," presented at the 2019 International Conference on Computer and Information Sciences (ICCIS), Apr. 2019, doi: 10.1109/iccisci.2019.8716454.

[20] T. Doan and J. Kalita, "Sentiment Analysis of Restaurant Reviews on Yelp with Incremental Learning," presented at the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Dec. 2016, doi: 10.1109/icmla.2016.0123.

[21] M. Ghosh and G. Sanyal, "An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning," J Big Data, vol. 5, no. 1, Nov. 2018, doi: 10.1186/s40537-018-0152-5.

[22] S. M. Nagarajan and U. D. Gandhi, "Classifying streaming of Twitter data based on sentiment analysis using hybridization," Neural Comput & Applic, vol. 31, no. 5, pp. 1425–1433, Apr. 2018, doi: 10.1007/s00521-018-3476-3.

[23] S. Soni and A. Sharaff, "Sentiment Analysis of Customer Reviews based on Hidden Markov Model," presented at the 2015 International Conference, 2015, doi: 10.1145/2743065.2743077.

[24] A. Kamal and M. Abulaish, “Statistical Features Identification for Sentiment Analysis Using Machine Learning Techniques,” presented at the 2013 International Symposium on Computational and Business Intelligence (ISCBI), Aug. 2013, doi: 10.1109/iscbi.2013.43.

[25] P. Manivannan and C. S. K. Selvi, “Pairwise relative ranking technique for efficient opinion mining using sentiment analysis,” Cluster Comput, vol. 22, no. S6, pp. 13487–13497, Feb. 2018, doi: 10.1007/s10586-018-1986-5.

[26] B. Selvam, S. Ravimaran, and S. Selvam, “Parallelized root cause analysis using cause-related aspect formulation technique (CRAFT),” J Supercomput, vol. 75, no. 9, pp. 5914–5929, Jun. 2019, doi: 10.1007/s11227-019-02893-8.

[27] X. Lv and N. El-Gohary, “Text Analytics for Supporting Stakeholder Opinion Mining for Large-scale Highway Projects,” Procedia Engineering, vol. 145, pp. 518–524, 2016, doi: 10.1016/j.proeng.2016.04.039.

[28] T. L. Ngo-Ye and A. P. Sinha, “Analyzing Online Review Helpfulness Using a Regressional ReliefF-Enhanced Text Mining Method,” ACM Trans. Manage. Inf. Syst., vol. 3, no. 2, pp. 1–20, Jul. 2012, doi: 10.1145/2229156.2229158.

[29] A. Felbermayr and A. Nanopoulos, “The Role of Emotions for the Perceived Usefulness in Online Customer Reviews,” Journal of Interactive Marketing, vol. 36, pp. 60–76, Nov. 2016, doi: 10.1016/j.intmar.2016.05.004.

[30] M. Malik and A. Hussain, “An analysis of review content and reviewer variables that contribute to review helpfulness,” Information Processing & Management, vol. 54, no. 1, pp. 88–104, Jan. 2018, doi: 10.1016/j.ipm.2017.09.004.

[31] E. Najmi, K. Hashmi, Z. Malik, A. Rezgui, and H. U. Khan, “CAPRA: a comprehensive approach to product ranking using customer reviews,” Computing, vol. 97, no. 8, pp. 843–867, Feb. 2015, doi: 10.1007/s00607-015-0439-8.

[32] N. Gobi and A. Rathinavelu, “Analyzing cloud based reviews for product ranking using feature based clustering algorithm,” Cluster Comput, vol. 22, no. S3, pp. 6977–6984, Feb. 2018, doi: 10.1007/s10586-018-1996-3.

[33] Zhang, K., Cheng, Y., Liao, W., & Choudhary, A. (2012). Mining millions of reviews. Proceedings of the 13th International Conference on Electronic Commerce - ICEC ’11. the 13th International Conference. https://doi.org/10.1145/2378104.2378116

[34] Krestel, R., & Dokoohaki, N. (2015). Diversifying customer review rankings. Neural Networks, 66, 36–45. https://doi.org/10.1016/j.neunet.2015.02.008

[35] Varudharajulu, A. K., & Ma, Y. (2018). Feature-based Restaurant Customer Reviews Process Model using Data Mining. Proceedings of the 2018 International Conference on Computing and Big Data - ICCBD ’18. the 2018 International Conference. https://doi.org/10.1145/3277104.3277113

[36] Flory, L., Osei-Bryson, K.-M., & Thomas, M. (2017). A new web personalization decision-support artifact for utility-sensitive customer review analysis. Decision Support Systems, 94, 85–96. https://doi.org/10.1016/j.dss.2016.11.003

[37] Alfrjani, R., Osman, T., & Cosma, G. (2017, October). Exploiting domain knowledge and public linked data to extract opinions from reviews. 2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA). 2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA). https://doi.org/10.1109/ickea.2017.8169909

[38] Nandedkar, S. C., & Patil, J. B. (2018, April). Co-Extracting Feature and Opinion Pairs From Customer Reviews Using Hybrid Approach. 2018 3rd International Conference for Convergence in Technology (I2CT). 2018 3rd International Conference for Convergence in Technology (I2CT). https://doi.org/10.1109/i2ct.2018.8529462

[39] Hedegaard, S., & Simonsen, J. G. (2013, April 27). Extracting usability and user experience information from online user reviews. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '13: CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/2470654.2481286

[40] Bing, L., Wong, T.-L., & Lam, W. (2016). Unsupervised Extraction of Popular Product Attributes from E-Commerce Web Sites by Considering Customer Reviews. ACM Transactions on Internet Technology, 16(2), 1–17. https://doi.org/10.1145/2857054

[41] Soriano, L. T., & Palaoag, T. D. (2018). A Machine Learning-Based Topic Extraction and Categorization of State Universities and Colleges (SUC) Customer Feedbacks. Proceedings of the 8th International Conference on Information Communication and Management - ICICM '18. the 8th International Conference. https://doi.org/10.1145/3268891.3268897

[42] Akin-Adetoro, A., & Seymour, L. (2019). Analysing research on Information Systems success and failure. Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019 on ZZZ - SAICSIT '19. the South African Institute of Computer Scientists and Information Technologists 2019. https://doi.org/10.1145/3351108.3351121

[43] Singh, A., & Tucker, C. S. (2017). A machine learning approach to product review disambiguation based on function, form and behavior classification. Decision Support Systems, 97, 81–91. https://doi.org/10.1016/j.dss.2017.03.007

[44] Jin, K., Zhang, X., & Zhang, J. (2019, December). Learning to Generate Diverse and Authentic Reviews via an Encoder-Decoder Model with Transformer and GRU. 2019 IEEE International Conference on Big Data (Big Data). 2019 IEEE International Conference on Big Data (Big Data). https://doi.org/10.1109/bigdata47090.2019.9006577

[45] Hanni, A. R., Patil, M. M., & Patil, P. M. (2016, September). Summarization of customer reviews for a product on a website using natural language processing. 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI). https://doi.org/10.1109/icacci.2016.7732392

[46] Jmal, J., & Faiz, R. (2013). Customer review summarization approach using Twitter and SentiWordNet. Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics - WIMS '13. the 3rd International Conference. https://doi.org/10.1145/2479787.2479813

[47] Lu, X., Qu, J., Jiang, Y., & Zhao, Y. (2018, July 22). Should I Invest it? Proceedings of the Practice and Experience on Advanced Research Computing. PEARC '18: Practice and Experience in Advanced Research Computing. https://doi.org/10.1145/3219104.3229287

[48] Moon, S., Bae, S., & Kim, S. (2014, August). Predicting the Near-Weekend Ticket Sales Using Web-Based External Factors and Box-Office Data. 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). https://doi.org/10.1109/wi-iat.2014.113

[49] Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Weaven, S. (2019). Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. International Journal of Hospitality Management, 80, 52–77. https://doi.org/10.1016/j.ijhm.2019.01.003

[50] Cosma, G., & Acampora, G. (2016). A computational intelligence approach to efficiently predicting review ratings in e-commerce. Applied Soft Computing, 44, 153–162. https://doi.org/10.1016/j.asoc.2016.02.024

[51] Das, T. K., Acharjya, D. P., & Patra, M. R. (2014, November). Business intelligence from online product review-a rough set based rule induction approach. 2014 International Conference on Contemporary Computing and Informatics (IC3I). 2014 International Conference on Contemporary Computing and Informatics (IC3I). https://doi.org/10.1109/ic3i.2014.7019662

[52] Veremyev, A., Semenov, A., Pasiliao, E. L., & Boginski, V. (2019). Graph-based exploration and clustering analysis of semantic spaces. Applied Network Science, 4(1). https://doi.org/10.1007/s41109-019-0228-y

[53] Pai, H.-T., Lai, H.-W., Wang, S.-L., Wu, M.-F., & Chuang, Y.-T. (2017, October 17). Recommendations for mobile applications. Proceedings of the 1st International Conference on Internet of Things and Machine Learning. IML 2017: International Conference on Internet of Things and Machine Learning. https://doi.org/10.1145/3109761.3109771

[54] Tu, D., Chen, L., & Chen, G. (2015). Automatic multi-way domain concept hierarchy construction from customer reviews. Neurocomputing, 147, 472–484. https://doi.org/10.1016/j.neucom.2014.06.038

[55] Liu, H., & Wang, B. (2020, March 12). Mitigating File-Injection Attacks with Natural Language Processing. Proceedings of the Sixth International Workshop on Security and Privacy Analytics. CODASPY '20: Tenth ACM Conference on Data and Application Security and Privacy. https://doi.org/10.1145/3375708.3380310

[56] C. Cortes and V. Vapnik, “Support-vector networks,” Mach Learn, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[57] Hand, D. J. and Yu, K. (2001), Idiot's Bayes—Not So Stupid After All?. International Statistical Review, 69: 385-398.

[58] F. Sánchez-Rada and C. A. Iglesias, “Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison,” Information Fusion, vol. 52, pp. 344–356, Dec. 2019, doi: 10.1016/j.inffus.2019.05.003..

[59] S. M. Weiss, N. Indurkhya, and T. Zhang, “Overview of Text Mining,” in Texts in Computer Science, Springer London, 2010, pp. 1–12.

[60] Amazon Customer Reviews Dataset. [Online]. Available: https://s3.amazonaws.com/amazon-reviews-pds/readme.html. [Accessed: 16-Feb-2021].

[61] B. Scholkopf et al., "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," in IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2758-2765, Nov. 1997.

[62] D. F. Specht, "A general regression neural network," in IEEE Transactions on Neural Networks, vol. 2, no. 6, pp. 568-576, Nov. 1991.

[63] A. I. Schein and L. H. Ungar, “Active learning for logistic regression: an evaluation,” Mach Learn, vol. 68, no. 3, pp. 235–265, Aug. 2007, doi: 10.1007/s10994-007-5019-5.

[64] D.D. Lewis, Naïve (Bayes) at forty: the independence assumption in information retrieval, Machine Learning: ECML-98, Springer, Berlin Heidelberg, 1998, pp. 4–15.

[65] V. P. B. felean, "Analysis and Predictions on Students' Behavior Using Decision Trees in Weka Environment," 2007 29th International Conference on Information Technology Interfaces, Cavtat, 2007, pp. 51-56.

[66] B. Gezici, N. Bolucu, A. Tarhan, and B. Can, “Neural Sentiment Analysis of User Reviews to Predict User Ratings,” presented at the 2019 4th International Conference on Computer Science and Engineering (UBMK), Sep. 2019. doi: 10.1109/ubmk.2019.8907234.

[67] G. S. Budhi, R. Chiong, I. Pranata, and Z. Hu, “Predicting rating polarity through automatic classification of review texts,” presented at the 2017 IEEE Conference on Big Data and Analytics (ICBDA), Nov. 2017. doi: 10.1109/icbdaa.2017.8284101.

[68] Ch. S. C. Reddy, K. U. Kumar, J. D. Keshav, B. R. Prasad, and S. Agarwal, “Prediction of star ratings from online reviews,” presented at the TENCON 2017 - 2017 IEEE Region 10 Conference, Nov. 2017. doi: 10.1109/tencon.2017.8228161.

[69] G. Acampora, G. Cosma, and T. Osman, “An extended neuro-fuzzy approach for efficiently predicting review ratings in E-markets,” presented at the 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Jul. 2014. doi: 10.1109/fuzz-ieee.2014.6891829.

[70] Amazon Wireless Services. 2020. Amazon Customer Reviews Dataset; https://s3.amazonaws.com/amazon-reviews-pds/readme.html.

[71] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, “Word Cloud Explorer: Text Analytics Based on Word Clouds,” presented at the 2014 47th Hawaii International Conference on System Sciences (HICSS), Jan. 2014. doi: 10.1109/hicss.2014.231.

[72] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[73] C. Cortes and V. Vapnik, “Support-vector networks,” Mach Learn, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[74] Hand, D. J. and Yu, K. (2001), Idiot's Bayes—Not So Stupid After All?. International Statistical Review, 69: 385-398.

[75] G. I. Webb et al., “Logistic Regression,” in Encyclopedia of Machine Learning, Springer US, 2011, pp. 631–631. doi: 10.1007/978-0-387-30164-8_493.

[76] T. Chen and C. Guestrin, “XGBoost,” presented at the KDD ’16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2016. doi: 10.1145/2939672.2939785.

[77] Pedregosa, Fabian & Varoquaux, Gael & Gramfort, Alexandre & Michel, Vincent & Thirion, Bertrand & Grisel, Olivier & Blondel, Mathieu & Prettenhofer, Peter & Weiss, Ron & Dubourg, Vincent & Vanderplas, Jake & Passos, Alexandre & Cournapeau, David & Brucher, Matthieu & Perrot, Matthieu & Duchesnay, Edouard & Louppe, Gilles. (2012). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 12.

[78] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer New York, 2009. doi: 10.1007/978-0-387-84858-7.