Final Report: Market Research Classification using Supervised Machine Learning

Franklin Liu Department of Computer Science
Virginia Tech
7054 Haycock Rd, Falls Church, VA 22043
lfranklin99@vt.edu

Jackson Prillaman Department of Computer Science
Virginia Tech
7054 Haycock Rd, Falls Church, VA 22043
jackp96@vt.edu

Abstract

Market research is gradually becoming a much greater part of our daily lives and machine learning plays an incredibly important role in this type of research by allowing researchers to process large amounts of customer data. However, it can be very difficult for machines to interpret certain kinds of customer data (i.e., whether a user review is positive or negative). Thus, we propose to utilize sentiment analysis along with various machine learning methods to classify Amazon user reviews. Our experiment results show that sentiment analysis is an effective way of classifying reviews as either positive or negative achieving over 80% accuracy and allowed us to determine that certain machine learning classification methods such as Support Vector Machines (SVM) are more effective than others like K-Nearest Neighbors.

1 Introduction

At its most basic, market research is the process of collecting data from customers (e.g. sales numbers, customer reviews, etc.) and analyzing that data to determine the viability of a service or product. Market research thus makes heavy use of various statistical techniques to draw inferences about consumers.

Machine learning aids in this process by helping to find useful patterns from data without the need for explicit programming. More and more researchers are relying on machine algorithms to collect data and discover new insights into consumer behavior. The use of machine learning in market research opens a plethora of new opportunities that researchers are only just beginning to discover.

One area where we hope to expand research is in using sentiment analysis in order to analyze customer reviews. While previous work has focused on attempting to classify customers based on data, to our knowledge this is the first attempt to use sentiment analysis to analyze customer perception of products.

2 Related Works

There is no small amount of research into sentiment analysis as the field is highly relevant to many different disciplines. The technology underlying sentiment analysis is constantly changing. As such the related works that were reviewed for this paper were chosen from more recent publications to better reflect the current state of sentiment analysis research.

Preprint. Under review.

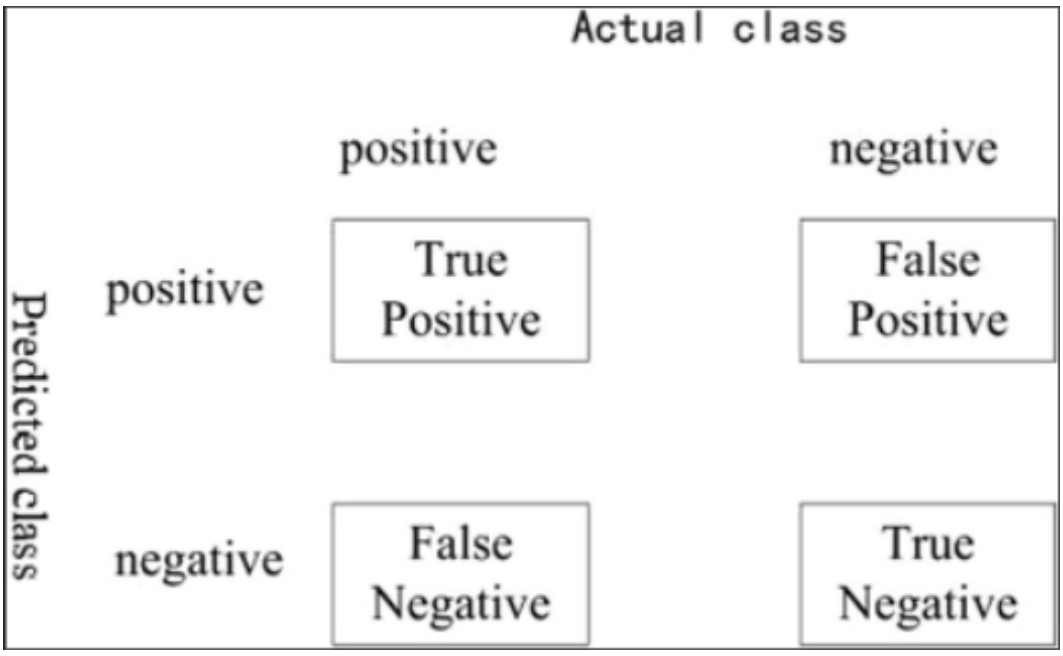


Figure 1: A confusion matrix demonstrating the distribution of negative and positive reviews

A study conducted by Tarnowska and Ras [4] sought to use sentiment analysis to better understand customer reviews of heavy equipment repair services. They applied a strategy to transform the unstructured customer data into structured data using data mining. To do this, they made use of domain specific terminology to determine the elements relevant to the sentiment analysis, then broke down the data using an aspect-based process which extracted an opinion consisting of a sentiment and a target of the opinion.

Another study conducted by Jemai, Hayouni and Baccar [3] mined tweets to assess the accuracy of different methods available when one is pursuing sentiment analysis. They pulled 10000 tweets, pre-processed them, and used multiple types of Naïve Bayes (Regular, Multinomial and Bernoulli) as well as Logistic Regression and LinearSVC or Linear Support Vector Classifier. Their research found that regular Naïve Bayes was the most accurate in correctly classifying the tweets.

3 Problem Statements

We hypothesize that utilizing sentiment analysis with machine learning methods will allow us to achieve high accuracy in classifying user reviews. Sentiment analysis is the use of natural language processing and text analysis in machine learning. We also hypothesize that certain machine learning methods are more effective at classifying user reviews utilizing sentiment analysis than others and seek to determine which of the machine learning methods are most effective. To judge the effectiveness of utilizing sentiment analysis we define accuracy using the equation 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Accuracy is defined using the number of positive reviews the algorithm is able to correctly classify (true positives), how many negative reviews the algorithm is able to correctly classify (true negatives), and how well negative reviews are classified as positive (false positive) or positive reviews are classified as negative (false negative). A representation of the distribution of the negative and positive reviews can be seen in figure 1

4 Methodology

This section describes our methodology for performing sentiment analysis on Amazon reviews. We detail the experiments we will perform using our model and what metrics we will use to judge our model's effectiveness.

4.1 Data Set info

To evaluate our model, we utilize two data sets of labeled Amazon reviews.

Data Set 1 [1] is a collection of 34,686,770 Amazon reviews from 6,643,669 users on 2,441,053 products. The data spans a period of 18 years, including 35 million reviews up to March 2013. The data set is divided into several subsets with each containing 28,000 training samples and 12,000 testing samples. Each user review within the data set will contain the following attributes: a user rating of the product, a title, and the full text of the review. The amazon review will be regarded as positive if it has a star rating of 4 or 5 and will be regarded as negative if it has a star rating of 1 or 2. Samples of score 3 are ignored.

Data Set 2 [2] is a list of over 233.1 million consumer reviews for Amazon products. Each user review within the data set includes basic product information, rating, and review text. Like data set 1, we divide the data set into subsets with each containing 28,000 training samples and 12,000 testing samples. Each review will have a star rating of 4 or 5 indicates a positive review, a star rating of 1 or 2 indicates a negative review and samples of score 3 are ignored.

4.2 Pre-processing

Our program pre-processes the data, so it be analyzed by the different machine learning methods. There are several steps involved in our pre-processing.[3] The first step is to tokenize the data or break down every sentence and statement into component words. The second step is to remove punctuation from the dataset. The third step is to remove mentions of websites and email addresses. A sub-step to this is to remove any remaining free-standing non-alphanumeric characters. The fourth step is to remove empty spaces that may not have been removed from the dataset. The fifth step is to remove stop words, common words that do not lessen understanding if they are removed (i.e., "we", "and", "it", etc.). The final step is to convert the remaining words into lemmas which reduces words into a root word.

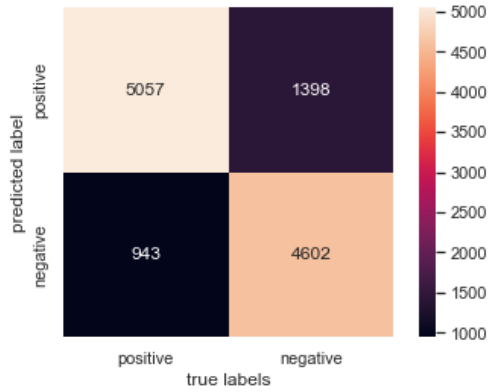
4.3 Experiment design

To analyze data sets we will apply the model to each of the subsets. We will extract word clouds (or frequency tables) based on whether the rating of the products is positive or negative. For instance, if words such as love, wonderful, useful, etc. are associated with positive reviews we will group them together into one word cloud while we group words associated with negative reviews into a separate word cloud. The word cloud will form the basis of a feature set for different reviews in the data set.

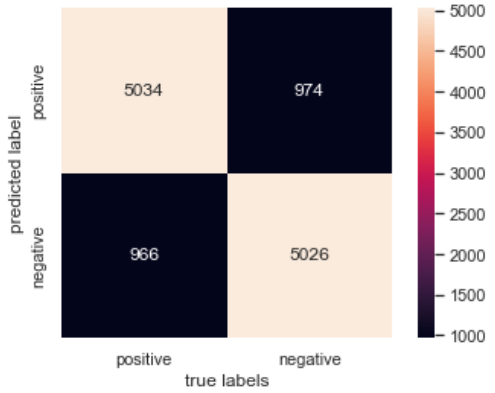
Our model will then utilize one of several machine learning methods (either a Naive Bayes Classification, K-Nearest Neighbors, or Support Vector Machine with a linear kernel) to classify each review in the test samples as either a positive or negative review based on the words used in the review.

4.4 Experiment Results

Unfortunately, K-Nearest Neighbors proved unsuitable for analyzing the large date set of user reviews. Our program was unable to finish running after several days. The reason for this is likely due to curse of dimensionality. The curse of dimensionality refers to a problem whereby the size of the data space grows exponentially with the number of dimensions. K-Nearest Neighbors is especially vulnerable to this problem because it requires a point to be close in every single dimension or feature. Our data represents a large swath of user generated reviews with a very large vocabulary. Even after pre-processing, the feature set for the data is still massive. The difference in feature sets across reviews with the same label can be very large.

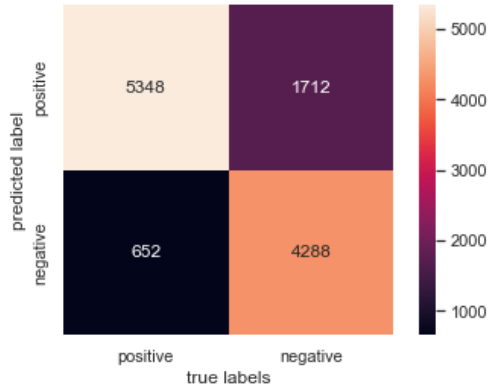


(a) A confusion matrix demonstrating the distribution of negative and positive reviews for Naive Bayes

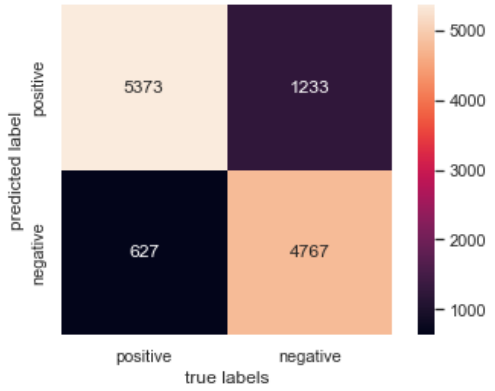


(b) A confusion matrix demonstrating the distribution of negative and positive reviews for SVM

Figure 2: Data Set 1 Results



(a) A confusion matrix demonstrating the distribution of negative and positive reviews for Naive Bayes



(b) A confusion matrix demonstrating the distribution of negative and positive reviews for SVM

Figure 3: Data Set 2 Results

On the other hand, both Naive Bayes and SVM both proved effective at classifying the user reviews using sentiment analysis. The confusion matrices representing the results of Naive Bayes and SVM on data from Data Set 1 and Data Set 2 can be seen in figure 2 and figure 3 respectively.

Using the equation, we find that Naive Bayes has an accuracy of 80.49% and SVM has an accuracy of 83.83% for Data Set 1. For Data Set 2, Naive Bayes has accuracy of 80.3% and SVM has an accuracy of 84.5%. Thus, SVM performs the best of all the machine learning classification methods but Naive Bayes is very close in terms of accuracy

5 Conclusion

Through our experiment results we determined that using sentiment analysis with machine learning methods is an effective method of classifying Amazon reviews providing accuracy of greater than 80% in several cases. We also determined that of the three machine learning methods (K-Nearest Neighbors, Naive Bayes, and SVM) that SVM is the most effective followed by Naive Bayes while K-Nearest Neighbors is unsuitable for classification using large data sets.

6 Broader Impacts

Our work has potential to improve upon previous market research techniques and greatly enhance the ability of researchers to analyze how customers perceive their products. Researchers will be able to utilize our model to perform full sentiment analysis on a wide range of customer reviews and identify whether the consumers view the product in a positive light. Researchers will also know what kinds of machine learning methods are most effective in performing classification with sentiment analysis on user reviews. This will allow researchers to provide useful analysis more efficiently on a wider range of customer reviews.

Market research plays a very important role in supporting the development of new services and products aimed at improving our daily lives. It is a very important part of how decision makers at major companies meet consumer needs. By enhancing this ability, our work will allow companies to better meet the needs of their consumers.

7 Future Work

In regards to future work, we hope examine how different methods of pre-processing can affect our model. There are several different types of sentiment analysis such as as objective based, lexicon based, aspect based and fine grained. The type of sentiment analysis could have impacted our results and by exploring this further we hope to better understand the effectiveness of our model.

Another possible direction we could take is to explore how altering the different parameters of the machine learning methods affects accuracy. For instance we utilized a linear kernel for our implementation but utilizing a non-linear kernel could affect the effectiveness of various machine learning methods in classifying reviews and change which method is most effective.

8 Contributions

During this work Franklin Liu worked to find the data sets from Amazon reviews and determine how they could be used for sentiment analysis. Franklin also worked to find pre-existing python libraries to aid in sentiment analysis and helped to write the code for conducting review classification. Franklin also helped to write large portions of the Spotlight presentation.

Jackson Prillaman conducted a review of various papers related to sentiment analysis and market research to understand what work had been done previously in this area. Additionally, Prillaman wrote large portions of the code and conducted many of the experiments on his personal computer.

Both worked to write this report and explain what work had been accomplished since the milestone report.

References

[1] Xiang Zhang (xiang.zhang@nyu.edu), Character-level Convolutional Networks for Text Classification. Advances in Neural Information Processing Systems 28 (NIPS 2015),
https://www.kaggle.com/kritanjalijain/amazon-reviews

[2] Jianmo Ni, Jiacheng Li, Julian McAuley, Empirical Methods in Natural Language Processing (EMNLP), 2019,
https://nijianmo.github.io/amazon/index.html

[3] Jemai, Hayouni & Baccar, Sentiment Analysis Using Machine Learning Algorithms, 2021 International Wireless Communications and Mobile Computing (IWCMC), 2021, pp. 775-779, doi: 10.1109/IWCMC51323.2021.9498965.

[4] Tarnowska K.A, & Ras Z.W., Sentiment Analysis of Customer Data, Web Intelligence, vol. 17, no. 4, 2019, pp. 343–363., doi:10.3233/WEB-190423.

[5] H. Zhang (2004), The optimality of Naive Bayes. Proc. FLAIRS.
https://scikit-learn.org/stable/modules/naive_bayes.html