The Amazon Review Spectrum: A Sentiment Analysis Exploration

Sadiul Arefin Rafi, Irfanul Hoque, Pritam Barua, Nafisa Muhammad and Annajiat Alim Rasel

Brac University, Dhaka, Bangladesh
sadiul.arefin.rafi@g.bracu.ac.bd

Abstract. With time, most tasks have become digitized. The growth in the use of the internet along with smart devices brings all decision-making processes to the fingertips. Online shopping has become an easier alternative to in-person shopping now more than ever before. When customers are not able to rely on physical features to confirm a purchase, the next source of reliability comes from online product reviews. Such reviews not only help consumers but also help businesses make predictions about future trends. With the use of NLP, comments in bulk can be categorized with ease. This paper is based on an Amazon product review dataset focusing on gadgets. After tokenizing, in order to prepare the data for the machine learning model, reviews were transformed into numerical values. Ratings, content, and support votes are sectioned using a binary classifier to separate positive and negative reviews. The WordNet Lemmatizer from the NLTK library was used to make semantic associations between words. The NLTK POS tagger was used for nouns, verbs, and adjectives while treebank tags were mapped to WordNet parts of speech. The K-means clustering method was used to combine various reviews with similar sentiments. The Support Vector Machine (SVM) classifier demonstrated the highest pre-optimization accuracy (93.94%) among the various models tested. The SVM's performance increased after some fine-tuning, and it now has an accuracy of 94.08%. Comparatively, the Multinomial Naive Bayes model achieved 93.45% accuracy, while the logistic regression model showed a 93.70% accuracy.

Keywords: NLP, sentiment analysis, POS tagger, tokenization, lemmatization, K-means clustering

1 Introduction

In this day and age, online shopping is becoming more and more popular due to the amount of convenience that comes along with it. Smart devices allow using the internet with ease for all generations. Throughout the pandemic, with people being stuck within the walls of their homes, the advantages of online shopping have come to the surface more than ever before. The availability of online services became the only way to survive for numerous businesses. However, e-commerce websites that were popular long before 2020 were able to see the jump in sales even more clearly.

Amazon reached annual revenue of more than 457 billion dollars at the end of 2021. This was almost a 60% increase since 2019. As a trusted portal that has gained an immense amount of exposure, it became a valued option for many online shoppers.

When it comes to not being able to see or touch products, customer reviews hold the most weight for ensuring a purchase. It makes it easier for consumers to compare between brands and prices as well as an overall source for reliability. Not only do such reviews make the shopping experience smoother for customers, but it also helps

the seller understand the customer’s needs. With this information, predictions and strategies for future sales can be made in a systematic manner.
One application for natural language processing is sentiment analysis, also known as opinion mining. With the help of computational linguistics and text analysis, opinion mining is able to determine the emotion behind different types of writing. From a generic angle, the text is classified into two types. A text can be either an opinion or afact. “The bag is red”, would be a fact. “The bag is not useful”, would typically be more of an opinion. Customer reviews will mostly consist of opinions that can be categorized as good, bad, or neutral. This paper will be using a dataset of Amazon product reviews upon which sentiment analysis will determine the type of comments from various customers.

2. Related Works

Recent developments in opinion mining and sentiment analysis heavily emphasized product reviews. In their study [1], Elli, Maria, and Yi-Fan used survey responses and data analysis to create a business model. They made a point of saying that the tools they used produced highly accurate results. Their research focused on recognizing emotions in reviews, deciphering gender from names, and identifying bogus reviews. It was furthered by the use of business analytics. Their main programming languages were Python and R, and their classifiers were Support Vector Machine (SVM) and Multinomial Naive Bayesian (MNB).

Numerous classifiers were employed in a different study [2] to calculate the precision and recall levels. In paper [3], data from Amazon review datasets were used to improve sentiment analysis and NLP techniques already in use. They specifically targeted book reviews in Amazon's Kindle sector and used the Naive Bayesian classifier and a decision list to differentiate between good and negative evaluations.

In paper [4], classifiers such NV, SVM, and maximum entropy were used, and the results were displayed in statistical charts. No accuracy measurements, though, were offered. In their paper [5], the authors created a model that uses a bag of words to predict product ratings based on textual evaluations. They examined the effectiveness of bigrams and unigrams and discovered that bigrams underperformed unigrams by 15.89% due to their higher variation.

Last but not least, the methodology used in publication [6] was based on the naive Bayes classifier algorithm, which regrettably did not produce adequate outcomes. For easier comprehension, their methodology tended toward simpler algorithms. Despite having great accuracy, SVM has been found to be ineffective for large datasets. Additionally, decision trees and logistic regression were investigated. Paper [6] also used Tfidf for additional analyses. The researchers demonstrated that they could predict ratings from a collection of words, although the classifiers they used had several limitations. A linear regression model that highlighted the root mean square error was also included in their findings.

3. Methodology

3.1 Data Preprocessing

The dataset was collected from Kaggle. It includes Amazon item reviews and information. Reviews (ratings, content, and support votes), item metadata (portraits, class data, value, brand, and image highlights), and linkages are all included in this collection. The following dataset contains mainly the reviews of gadgets such as Amazon kindle, Bluetooth speaker, TV, and Tablet. 'Reviewer ID', 'ASIN', 'Reviewer Name', 'Review title', 'Helpful', 'Summary', 'Rating', and 'Review time' are all properties in the files. It indicates that group ranks range from 1 to 5. We'll need to transform these ratings into two categories, 1 and 0 because we are going to use a binary classification model. Positive (1) ratings are those that are greater than or equal to three, while negative ratings (0) are those that are less than three. Because machine learning models work with numerical characteristics, we'll need to transform our review column into numerical values before we start creating our model. We will tokenize the text and conversions into tokens. Tokenization means the task of breaking down the text into smaller parts which can be words, characters, or subwords. Then we will remove the stop words. Things that are present in a phrase that is not required in any text mining area are stop words. We usually omit certain terms to improve the accuracy of the analysis. And Finally POS tagging. Parts of speech tag (POS) is a particular label applied to each token in a text corpus to denote the part of speech and, in certain cases, additional grammatical categories such as tense, number (plural/singular), case, and so on.

3.2 Model

In supervised algorithms, the sense of a word is represented by its assigned label, and input words are tagged in accordance with their senses. Techniques are used to infer these senses because unsupervised methods do not offer this labeling. The context-sense set is used as the training data in supervised approaches, which generally involve classifiers like naive Bayes. Lemmatization, which determines the root of words regardless of their appearance, was performed using the Wordnet Lemmatizer from the NLTK library. Additionally, part-of-speech tagging was done using the NLTK POS tagger, which had been pre-trained on the Treebank dataset. The textbase was divided into training and testing sets after lemmatization. For sentiment analysis, a number of classifiers were investigated, including decision trees, logistic regression, naive Bayes, multinomial naive bayes, support vector machines (SVM), and random forests. The goal was to determine positive or negative sentiments from product reviews.

3.3 Results

On the bag of words, the logistic model had an accuracy of 0.892 and a TF-IDF score of 0.88 on gadgets, while the Naive Bayes model had comparable values of 0.887 and 0.884. The ratio of tp / (tp + fp) is called precision, the meaning of tp and fp is true positive and false positive respectively. Precision gives the classifier the ability to avoid labeling a negative sample as a positive sample. The weighted harmonic mean of accuracy is the F-beta score where 0 is the lowest value and 1 is the highest value.

All models showed outstanding precision.

Logistic Regression: 93.70%
Multinomial Naive Bayes: 93.45%
SVM: 93.94%
Decision Tree: 90.18%

Random Forest: 93.46%

Due to its maximum accuracy, the SVM was chosen for additional optimization in light of the findings. The classifier can prevent false positives by using precision, which is defined as tp / (tp + fp). The harmonic mean of accuracy is represented by the F-beta score, which has a range of 0 to 1.

4 SVM Fine-Tuning

The SVM, which can differentiate between extreme examples of sentiment, was further refined. We used a grid search and experimented with various settings to enhance its performance. Additionally, the LinearSVC classifier pipeline was used to execute this search, making the most of CPU cores. After training, the optimized SVM was put to the test on random reviews to determine how well it would work in practice.

4.1 Results

The SVM showed an improved accuracy of 94.08% after optimization. Tests on arbitrary reviews confirmed the model's accuracy in identifying positive, neutral, and negative attitudes.

5 Conclusion

To conclude, reviews of products can vary over different factors and places. In the research conducted, we can see the utilization of models like Support Vector Machine both of which provided an almost precise result. Despite difficulties in data collection caused by the data protection policies of e-commerce sites, our study was effective in analyzing evaluations for well-known electronic devices. Active Learning, POS Tagging, and TF-IDF techniques have all been useful in this investigation. Future developments in models and algorithms should result in even more effective sentiment analysis.

References

1. Elli, Maria Soledad, and Yi-Fan Wang. "Amazon Reviews, business analytics with sentiment analysis." 2016
2. Xu, Yun, Xinhui Wu, and Qinxia Wang. "Sentiment Analysis of Yelp„s Ratings Based on Text Reviews." (2015).
3. Rain, Callen. "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning."Swarthmore College (2013)
4. Bhatt, Aashutosh, et al. "Amazon Review Classification and Sentiment Analysis." International Journal of Computer Science and Information Technologies 6.6 (2015): 5107-5110.
5. Nasr, Mona Mohamed, Essam Mohamed Shaaban, and Ahmed Mostafa Hafez. "Building Sentiment analysis Model using Graphlab." IJSER, 2017
6. Text mining for yelp dataset challenge; Mingshan Wang; University of California San Diego, (2017)
7. Amazon revenue 2006-2021: AMZN. Macrotrends. (n.d.). Retrieved January 15, 2022, from https://www.macrotrends.net/stocks/charts/AMZN/amazon/revenue

8. Klebnikov, S. (2021, June 28). 5 big numbers that show Amazon's explosive growth during the coronavirus pandemic. Forbes. Retrieved January 15, 2022, from https://www.forbes.com/sites/sergeiklebnikov/2020/07/23/5-big-numbers-that-show-amaz ons-explosive-growth-during-the-coronavirus-pandemic/?sh=47798fe04137
9. LIU, B. I. N. G. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions.* CAMBRIDGE UNIV PRESS.
10. Muhammad Ihsan Zul, Feoni Yulia, Dini Nurmalasari, “Social Media Sentiment Analysis Using K-means and Naïve Bayes Algorithm”, 2nd International Conference of Electrical Engineering and Informatics(Icon EEI 2018), October 2018