International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.118

Opinion Mining Customer Reviews on Amazon With Machine Learning Techniques

Shruti Ramesh
Department of Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur, Chennai, India
sr9478@srmist.edu.in

A. Prithivi
Department of Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur, Chennai, India
ap5851@srmist.edu.in

Dr. Divya G
Department of Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur, Chennai, India
divyag3@srmist.edu.in

Abstract*—Online customer reviews providefeedback to businesses about their products and services, which can be very valuable. By analyzing these reviews, businesses can gain insights into customer satisfaction, identify areas for improvement, and make data-driven decisions to improve their products and services. Opinion mining, also known as sentiment analysis, can help businesses identify common themes, issues, and areas for improvement and can help organizations find the needs and requirements of its target audience. Therefore, in this paper, opinion mining is performed on online customer reviews from Amazon. A comparative study is also conducted to compare the effectiveness of various Machine Learning models such as Decision Tree, Random Forest, Naïve Bayes, KNN, SVM, and Logistic Regression.

Keywords—opinion mining, customer reviews, machine learning, decision tree, random forest.

I. Introduction

With customer reviews playing an integral role in a consumer's decision-making process while purchasing a product, it is imperative that businesses analyze these reviews to make better business decisions. Data driven decision making can significantly boost a company's business outcomes. Opinion mining is the process by which one can extract opinions and emotions from a target using NLP approach to obtain meaningful data. Thus, opinion mining and sentimental analysis play a crucial role in a business's ability to be profitable. These processes, along with an efficient feedback management system, can help businesses meet consumer demand and maintain their expectations so as to remain competitive in the market.

Opinion mining has a significant number of use cases and advantages from a business perspective. It can be used to effectively analyze customer reviews in order to gain meaningful insights. Stakeholders can learn and understand how consumers interact with the product, and can use the data that has been mined to improve the functionalities of these products. They can gain an insight into what the consumer likes or dislikes about the product, and can also use the feedback provided in a review to make decisions about their products. Moreover, stakeholders can use opinion mining to keep track of their competitors as well. By understanding how consumers feel about their competitors' products, businesses can ensure that they apply findings in such a manner that they outrank their competitors and stay one step ahead of the competition at all times. With significant leaps and bounds in AI and NLP technologies, opinion mining systems are getting better at processing relevant data and are proving to be crucial to the success of a business.

In this paper, opinion mining is utilized to process and analyse customer reviews from a range of different products on popular e-commerce website Amazon. The research objective of this project is mainly to showcase the effectiveness of opinion mining from a business perspective. Since opinion mining can be implemented in a number of different ways by using many different ML or NLP models, the objective is also to perform a comparative study to understand the efficacy of a few different machine learning models trained on the same dataset. In doing so, one can understand which approach is best for opinion mining and also ensure that an opinion mining model can be deployed that is most beneficial to a business. A good opinion mining system must be able to correctly analyse a review as positive or negative and extract the most meaningful data from the language. It must be able to classify language hurdles such as emoticons or sarcasm in the appropriate way as well.

Opinion mining can be an extremely valuable tool for any business. Thus, the aim is to utilize opinion mining to create a system that can effectively process and analyze online customer reviews on the popular e-commerce website Amazon. The comparative study examines the efficiency of various ML models such as Naïve Bayes, Random Forest, KNN, SVM, and Logistic Regression. The most efficient model can be used to deploy an opinion mining system or feedback management system that can have real time use cases in a business. The motivation behind this project is to enable businesses with a system wherein they can leverage opinion mining to make insightful data driven decisions. This will not only help them remain profitable, but will also help them beat competitors in the market. By understanding exactly what ML model provides the best outcome, it can be ensured thatthe most efficient and useful opinion mining system is created.

Existing opinion mining systems have a number of limitations that stop them from truly being useful in a business. The primary hurdle includes language barriers such as emoticons, slang words, and sarcastic comments. While NLP has advanced in recent times, these challenges still persist. In this paper, a model may be proposed that can overcome at least some of these challenges, in order to make opinion mining more holistic and advantageous.

The purpose of this project is the create an opinion mining system that can correctly process, analyze, and classify online customer reviews on Amazon. The objective is to distinguish the most efficient ML model to carry out such a system by means of a comparative study. With this information, this paper proposes a model that can best

DOI: rp-9788770040723.118

perform opinion mining to allow businesses to make data driven business decisions.

II. LITERATURE REVIEW

In a paper written by T. Kim Phung et al [1], a study was conducted to try and use and apply the concept of supervised machine learning and its various methods. This study collected around 40000 reviews of travelers on hotels in Vietnam from 'agoda.com', a famous travel booking and review site. The next step consisted of training the machine learning models and figuring out which model works the best to forecast opinions. The results generated show that in the field of opinion mining, which is conducted here, three types of models have been effective in their performance, namely, Logistic Regression, Support Vector Machines and Neural Networks.

X. Xu et al [2] attempt to improve the aspect-level opinion mining that is generated for online customer reviews. Aspect-level opinion mining involves both the product and the sentiment that is generated for it. The new model that is proposed in this paper is to be used to jointly extract the aspects and the sentiment lexicons that are aspect-dependent, from the customer reviews available online. The results of this model are regarded to be more effective from the current standards and the practical values of the extracted lexicons are also increased.

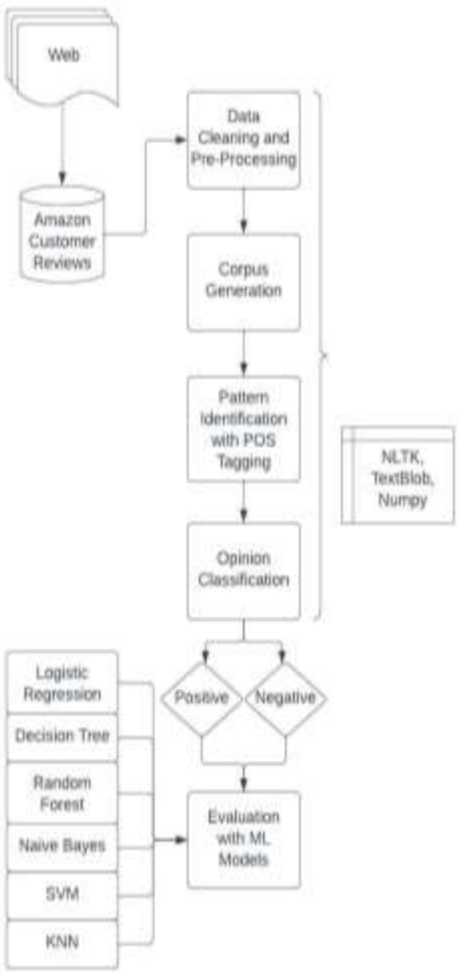

Fig. 1. Architecture Diagram

A process is explained by R. A. Laksono et al [3]where the reviews are classified using Naive Bayes technique to extract the sentiment of the customers, whether positive or negative. These reviews are often vital cogs of information for both the travel aggregator and the review site, as well as future consumers. The result from the research shows that the existing method of using TextBlob versus Naive Bayes, favors Naive Bayes by around 3%, but both methods are retrieving correct customer response.

Wu et al [4] conduct a study which aims to undertake and examine sentiment information from customer reviews and to predict and explore the potential of the same to enhance hotel demand forecasts. The customer reviews of four Macau luxury hotels are taken into consideration and their customer reviews are considered for the initial test. LSTM is used in this process and with the help of that, three indices are created that are checked out and are also evaluated for their effectiveness. The results of this paper are impressive and help in improving the forecasting accuracy.

An article by A. Adak et al [5]dwells on the fact that in the COVID-19 pandemic, the consumer had a change in how their food was delivered to them. Food Delivery Services took over the game and their growth has been rapid since doorstep delivery was preferred. These organizations have customers review them, which also can determine their company performance. This review is undertaken to find and explain machine learning, deep learning and explainable AI methods that are being used to predict customer sentiments. Key findings show a gap between DL and XAI methods and recommend to integrate them to perform in a better manner.

Like X. Xu et al [4]., S. Vanaja and M. Belwal [6] also use an aspect-based approach in their paper. This paper is an attempt to explain how to use sentiment analysis to analyze textual data and derive the sentiment from the same. This is due to the fact that in today's world it is seen that almost every minute there is some or the other form of text data being generated. This paper deals with the e-commerce reviews that can help retailers and the companies to understand the expectations of the customer, to provide a holistic interaction and to add to their sales. Aspect terms are used here, along with parts-of-speech and also applying classification algorithms to get the score of the review.

III. METHODOLOGY

A. Overview

In this paper, the objective is to test various machine learning models to check which one of the models provides the most amount of accuracy in the end. The performances of models such as random forest, decision trees, logistic regression and so on are compared. The algorithm with the best performance can then go on to classify customer reviews as good or bad with reasonable accuracy and confidence. The methodology proposed in this paper is fairly simple: It begins with some basic data preprocessing and cleaning methods, following which a corpus of biwords and triwordsis created for pattern generation. Next, POS tagging is performed and the dataset is trained. Finally, the

DOI: rp-9788770040723.118

model is tested and validated against the proposed machine learning models to compare their performances.

B. Data Cleaning and Pre-Processing

The dataset is obtained of Amazon reviews for a range of products by rudimentary web scraping methods. However, this data needs to be processed and cleaned before any actions can be performed on it to train the model. Data pre-processing is an important step in any NLP task. Tasks such as lemmatizing and tokenizing the words in the dataset are performed as initial data pre-processing steps. In order to obtain a clean dataset that can be trained, stopwords and duplicates are removed. Moreover, the spelling of incorrectly spelt words in the dataset is corrected. To accomplish these tasks the pandas and nltk libraries in Python are used. For spelling correction, the TextBlob library is utilized. The TextBlob library is particularly useful as it helps simplify text-based tasks for natural language processing in Python. Once some basic data cleaning has been performed, a corpus of words is created. Further actions on this corpus are performed in the subsequent steps.

C. Generating Biwords and Triwords

The next step is to add biwords and triwords to the corpus that was created, in order to generate patterns. A corpus is generated with biwords and triwords in order to simplify language tasks and train the model effectively. The goal is to avoid any erroneously tagged opinions while the user searches for the opinion of a particular product. This step is the precursor to POS tagging.

D. Parts-of-Speech Tagging

POS Tagging, or Parts of Speech Tagging, is a process that labels or categorizes words in a sentence to their corresponding parts of speech. Thus, it is a language task that helps the algorithm understand natural language in a better way, with its correct syntax and grammatical structure [8]. By performing parts of speech tagging on the reviewed text, one can extract essential linguistic features of the text to generate patterns. In this paper, for each reviewed text, eight patterns are generated using a rule-based approach. Using these eight patterns, it becomes possible to generate opinion words and their corresponding opinion targets. Since a rule-based approach is being followed in this paper, it is essential to define a set of language rules for different parts of speech in a sentence. Then, these rules are applied to the text corpus to identify patterns. The eight patterns generated are adjective + noun, adjective + noun + noun, adverb + adjective, adverb + adverb + noun, adverb + verb, adverb + adverb + adjective, verb + noun, and verb + adverb. These patterns can help identify opinions in reviews more easily.

E. Training The Model

After performing parts of speech tagging to generate patterns in the text, the rest of the model training procedures can be commenced. A semi-supervised approach is used to create opinion targets from the list of Amazon products. Next, using wordnet from nltk corpus, sets of words that are similar in meaning to the titles of the products are obtained. This step is necessary in order to make it possible for users to look for product opinions without having knowledge of the precise keywords associated with it. The existing list of product names is thus appended with the similar words generated from wordnet. Next, for each opinion target, i.e., the product name or its corresponding similar word, opinion words from the dataset of reviews are found. Then, for each of these opinion words, words similar in meaning to these opinion words are further searched for. The same procedure as earlier is followed to do this, using wordnet from nltk corpus. For the final extracted opinion words, polarities are generated using the TextBlob library. The polarity of a word is an indication of its emotional sentiment, and is usually on a scale of -1 to +1, with -1 being the lowest value or overall negative sentiment and +1 being the highest or overall positive sentiment. Generating the polarity of each opinion word allows the model to categorize the resulting opinion as negative or positive, which is the crux of this paper. The polarities of similar opinion words are averaged out and assigned an overall score. This score is then used to categorize the target as overall positive or overall negative. Since ratings on Amazon are in a range of 1 to 5 stars, ratings of 1, 2, and 3 stars are considered to be overall negative and ratings of 4 and 5 stars to be overall positive. Thus, the model is trained to generate either a positive sentiment or a negative sentiment for a given product. A confusion matrix is also generated to evaluate the performance of the model thus far and obtain the r and f scores, as well as precision and accuracy. With this, the opinion mining model is trained. The dataset consists of 34000 reviews for analysis, out of which around 10000 can be used for testing.

F. Testing The Model

In order to test the model, reviews are introduced and assigned a positive or negative sentiment class to each of the reviews. Then, the same data preprocessing techniques as earlier are performed. Next, using the corresponding opinion words, the score for the review is generated. This is done using the TextBlob and numpy libraries. The resulting score depicts whether the product has an overall positive or negative sentiment associated with it. For 10000 reviews, the model achieved an overall accuracy of 78.67%.

G. Validating with Machine Learning Models

To validate the model with different machine learning models that were proposed, thesklearn library in Python is used, along with the bag of words model.LogisticRegression, DecisionTreeClassifier, Naive Bayes, RandomForestClassifier, KNeighborsClassifier, and SCV are fit to the training dataset. With each of these models, the test dataset results are predicted and k-fold cross-validation is performed, where in this paper k = 10. This is done to measure the performance of the classifiers.

Since the random forest algorithm usually has an upper hand against other traditional machine learning algorithms [12], it may have the best performance in this model as well. The working of the algorithm is hence detailed in the next subsection.

H. Random Forest Algorithm

In a machine learning context, a random forest is a collection of decision trees. Decision tree algorithm is a

DOI: rp-9788770040723.118

popular machine learning technique used for sentiment analysis, which involves determining the sentiment or emotion behind a text document. In the context of sentiment analysis, a decision tree algorithm works by using a tree-like model to represent possible decisions and their possible consequences. The algorithm builds a decision tree based on a set of training data, which consists of input text data and their corresponding sentiment labels (e.g., positive, negative, neutral).

The decision tree algorithm uses the training data to identify the most important features or attributes that are most predictive of sentiment. These features can include word frequency, the presence of specific words, or other linguistic features. The algorithm then creates a series of decision nodes based on these features, which determine the path that the algorithm will take through the decision tree. As the algorithm traverses the decision tree, it assigns sentiment labels to the input text data based on the decisions made at each node. For example, if the decision tree contains a node that tests for the presence of a specific word (such as "great"), the algorithm will assign a positive sentiment label if the word is present in the input text, and a negative sentiment label if it is not. Once the decision tree has been trained, it can be used to classify new text data based on their sentiment. The algorithm uses the decision tree to traverse the tree and assign a sentiment label to the input text data. Overall, decision tree algorithms are a useful technique for sentiment analysis because they are relatively easy to understand and interpret, and they can be trained on large datasets with high accuracy. However, they may not perform as well as more complex machine learning techniques for highly nuanced or complex sentiment analysis tasks. This is because decision trees are highly prone to over fitting. So to make this model more robust, an ensemble of decision trees can be used. Decision trees are bagged together to increase their performance, leading to Random Forest.

The first step in using the random forest algorithm is to prepare the training data. This involves selecting the features that will be used to make predictions and preprocessing the data to ensure that it is in a suitable format for the algorithm. Next, the random forest algorithm builds a set of decision trees. To build each decision tree, the algorithm selects a random subset of the training data and a random subset of the features. It then uses these subsets to train a decision tree. Once the decision trees are built, the random forest algorithm can make predictions for new data. To do this, the algorithm passes the new data through each of the decision trees and averages the results to produce a final prediction. For classification tasks, the final prediction is the mode of the predictions made by the individual decision trees. For regression tasks, the final prediction is the mean of the predictions made by the individual decision trees. Finally, the performance of the random forest algorithm is evaluated using a test dataset. This involves comparing the predictions made by the algorithm to the true values in the test dataset. Metrics such as accuracy, precision, recall, and F1 score can be used to evaluate the performance of the algorithm. The key advantage of random forest algorithm is that it can handle a large number of features and complex datasets, and can provide an accurate prediction even in the presence of missing data or noisy data. Additionally, by using multiple decision trees, the algorithm is less prone to overfitting and provides a more robust and stable prediction. However, random forest algorithm can be computationally expensive and may require significant computational resources to build and evaluate the model. This is due to the fact that multiple decision trees have to be built and run when the model is trained and used.

IV. RESULTS

The accuracy of the model is found to be 78.67%. If the project were to be scaled and the training data increased, may result in better accuracy. Sarcastic comments are also able to be classified as actually positive or negative with considerable success. Upon validating the model against various machine learning models, it is found that decision trees had the best performance, with an accuracy of 83.26%. The random forest algorithm was a close second, with an accuracy of 83.17%. This indicates that Random Forest and Decision Tree algorithms are almost comparable in performance. For a simple opinion mining system that is run entirely on machine learning algorithms, random forest or decision tree would provide the best results to the user. Logistic regression, gaussian naive bayes, multinomial naive bayes, bernoulli naive bayes, k nearest neighbor and SVM each yielded an accuracy of 59.23%, 74.65%, 82.70%, 77.24%, 79.77%, and 80.05% respectively. Logistic regression was hence found to be the weakest machine learning model for opinion-mining customer reviews. After cross-validation, the precision was found to be 0.81, recall 0.86 and the f score was found to be 0.83. These results are summarized in Table I.

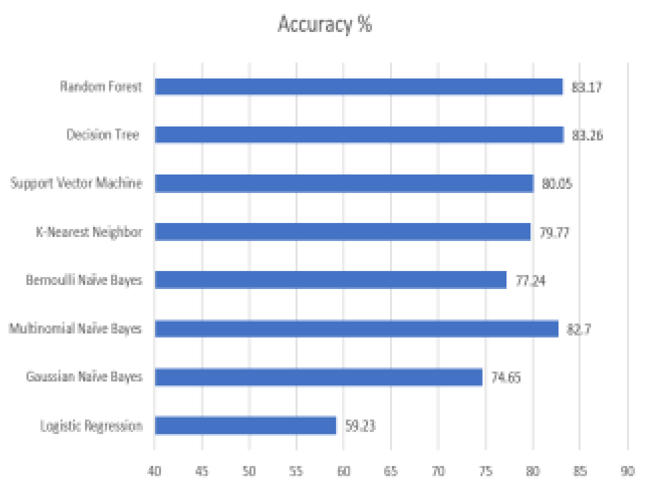

Fig. 2. Horizontal Bar Chart For Algorithm Comparison

TABLE I. ACCURACY OF ML ALGORITHMS

Algorithms	Accuracy %
Logistic Regression	59.23
Gaussian Naïve Bayes	74.65
Multinomial Naïve Bayes	82.70
Bernoulli Naïve Bayes	77.24
K-Nearest Neighbor	79.77
Support Vector Machine	80.05
Decision Tree	83.26
Random Forest	83.17

DOI: rp-9788770040723.118

V. Future Scope

This paper, for the most part, allows forthe comparison of various machine learning models and their performance on a model trained for opinion mining online customer reviews. Although Amazon reviews are used, this model can be applied to any sort of e-commerce or business platform, as long as customer reviews are available online. The approach in this paper is quite simple, however, with the advent of deep learning and advanced technologies such as GPT-4, opinion mining as a whole can be greatly improved. Opinion mining still has a long way to go. In the future, such a model should be quicker in processing data and should also be able to do so with high accuracy. It should also be able to overcome present-day limitations such as the use of emoticons, sarcasm, and other language barriers. Fake reviews are another hinderance that opinion mining algorithms must be able to deal with [7]. One major disadvantage is that the same kind of model can yield varying levels of accuracy for comments in different languages. Future opinion mining systems must be able to detect comments of different languages within the same dataset and classify them as such. Additionally, they must also be equally accurate for each of the different languages. Since slang or internet vernacular keeps evolving, corpuses in opinion mining models must be able to update themselves to keep up with changes in user language. There is still much research to be conducted on these aspects of opinion mining, but a more robust opinion mining or feedback management system will be easily achievable with advancements in NLP technology.

References

[1] T. Kim Phung, N. An Te and T. Thi Thu Ha, "A machine learning approach for opinion mining online customer reviews," 2021 21stACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), Ho Chi Minh City, Vietnam, 2021, pp. 243-246, doi: 10.1109/SNPDWinter52325.2021.00059.

[2] Sitharthan, R., Vimal, S., Verma, A., Karthikeyan, M., Dhanabalan, S. S., Prabaharan, N., ...&Eswaran, T. (2023). Smart microgrid with the internet of things for adequate energy management and analysis.Computers and Electrical Engineering, 106, 108556.

[3] R. A. Laksono, K. R. Sungkono, R. Sarno and C. S. Wahyuni, "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes," 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 2019, pp. 49-54, doi: 10.1109/ICTS.2019.8850982.

[4] D. C.Wu,S. Zhong, R. T. R. Qiu, and J. Wu, "Are customer reviews just reviews? Hotel forecasting using sentiment analysis," Tourism Economics, vol. l28, no. 3, pp. 795–816, 2022, https://doi.org/10.1177/13548166211049865

[5] A. Adak, B. Pradhan, and N. Shukla, "Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review," Foods, vol. 11, no. 10, p. 1500, May 2022, doi: 10.3390/foods11101500.

[6] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 1275-1279, 2018,doi: 10.1109/ICIRCA.2018.8597286.

[7] T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: A review," 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, pp. 298-301, 2017,doi: 10.1109/ICICCT.2017.7975207.

[8] Rajesh, M., &Sitharthan, R. (2022). Introduction to the special section on cyber-physical system for autonomous process control in industry 5.0.Computers and Electrical Engineering, 104, 108481.

[9] T. Hastie, J. Friedman, and R. Tibshirani, "Boosting and Additive Trees. In: The Elements of Statistical Learning," Springer Series in Statistics. Springer, New York, NY, 2001,https://doi.org/10.1007/978-0-387-21606-5_10

[10] Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim and Sung Hyon Myaeng, "Some Effective Techniques for Naive Bayes Text Classification," in IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 11, pp. 1457-1466, Nov. 2006, doi: 10.1109/TKDE.2006.180.

[11] Pádraig Cunningham and Sarah Jane Delany. 2021. K-Nearest Neighbour Classifiers - A Tutorial. ACM Comput. Surv. 54, 6, Article 128,p. 25, July 2022,https://doi.org/10.1145/3459665.

[12] A. Gupte, S. Joshi, P. Gadgul, A. Kadam, and A. Gupte, "Comparative study of classification algorithms used in sentiment analysis," International Journal of Computer Science and Information Technologies, vol. 5, no. 5, pp. 6261-6264, 2014.