

Sentiment Analysis of Customer Reviews on Amazon Electronics Product: Natural Language Processing Approach and Machine Learning

MSc Research Project MSc in Cloud Computing

Vipin Yadav Student ID: 19211791

School of Computing National College of Ireland

Supervisor: Prof. Divyaa Manimaran Elango

National College of Ireland Project Submission Sheet School of Computing



Student Name:	Vipin Yadav
Student ID:	19211791
Programme:	MSc in Cloud Computing
Year:	2021
Module:	MSc Research Project
Supervisor:	Divyaa Manimaran Elango www.ncirl.ie
Submission Due Date:	16-12-2021
Project Title:	Sentiment Analysis of Customer Reviews on Amazon Elec-
	tronics Product: Natural Language Processing Approach and
	Machine Learning
Word Count:	6622
Page Count:	24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

<u>ALL</u> internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Vipin Yadav
Date:	31-01-2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).			
Attach a Moodle submission receipt of the online project submission, to			
each project (including multiple copies).			
You must ensure that you retain a HARD COPY of the project, both for			
your own reference and in case a project is lost or mislaid. It is not sufficient to keep			
a copy on computer.			

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only		
Signature:		
Date:		
Penalty Applied (if applicable):		

Sentiment Analysis of Customer Reviews on Amazon Electronics Product: Natural Language Processing Approach and Machine Learning

Vipin Yadav 19211791

Abstract

There is a rapid rise in online shopping over the last few decades, people tend to buy things online rather than going to shop. E-commerce companies like Amazon, Flipkart, eBay are funding more innovative projects related to customers data. It's been estimated that Amazon's retail business tripled from last year. Most the eCommerce site seeks Customer' review and rating for their product. This practice of providing feedback is helpful for other customers and for the business too. Product feedback from the consumer will be helpful for another customer to get more insight into the product during their purchases, as well as for the eCommerce company that helps to know the quality and uses of their products and utilizes that feedback for improving their products. This paper aims to analyze the sentiments of customer's feedback against Amazon's product. Data will be imported from web sources for further analysis. As text data is always in an impure form, so data cleaning and pre-processing must be done. Natural language processing toolkit is one of the techniques that will be used for pre-processing data by removing stop words, nouns, pronouns, punctuation marks, and for the bag of words, the vectorization technique of NLP will be implemented. Customer's sentiments in their feedback will be categorized by labeling the data into three categories Positive, negative, and neutral. For further analysis on the cleaned and labeled data, machine learning models, the hybrid method will be used in which two or three algorithms of ML will be evaluated and compared with another algorithm of ML.

Keywords - NLP, Sentiment Analysis, Machine Learning, MNB, MLP, SGD Classifier, Count-vector, TD-IDF.

1 Introduction

Amazon is one of the most valued trillion-dollar companies. It has a catalog of over 12 million products. Amazon's sells its product all over the world. A variety of options are present in any products category which gives users a lot of options to choose from. If any user searches for mobile in the search engine of Amazon, more than 1000 search results will be displayed showing a different variety of products. With this many options of product, Amazon also provides an option to consumers to give Ratings and Reviews of their products. These rating and review text helps other consumers to get the user experience from past customers. This data is very important for the production company

also by getting useful insights from the consumer's review they can modify their product more effectively.

Data is considered as oil in today's world. Many multi-National companies demand user reviews and customer's feedback for benefiting their marketing strategies. Companies collect these data and analyze them by applying different machine learning models using NLP techniques on the text data. Sentiment analysis of the review data will help to categorize the customer's feedback into positive or negative or neutral sentiment.

1.1 Motivation and Background

Sentiment analysis has benefitted the social media industry by majoring in textual-based classification and helps to understand the sentiment of text by categorizing them into Positive, Negative, and Neutral categories. Sentiment analysis helps Marketing industries to understand the public feedback regarding their products. Sentiment analysis, machine learning models and natural language processing (NLP) techniques help to analyze the sentiments of the consumers.

It's been more than 2 decades since Amazon have been founded, and currently they have millions of products for the sale, out of the Thousands of products are sold daily. People tend to give their feedback against that product. Some research has been done before using different of techniques of NLP and evaluating the findings through machine learning models.

Very little research has been done with MLP classifier (multilayer perceptron) and SGD Classifier model. Implementing count vector and TF-IDF feature extractor techniques (Dalaorao et al.; 2019) with these models will help to analyze the sentiment of reviews more properly. The accuracy of these models will be drawn and a comparison of the performance of these models will be conducted.

1.2 Research Question

"To what extent machine learning models (MNB, MLP, SGD classifier) can help to analyze the customer's review of Amazon products. "

The primary focus is to implement 3 different machine learning models (Multinomial Naive Bayes, MLP classifier (multilayer perception), and Stochastic Gradient Descent (SGD classifier) on customer's feedback of Amazon's product with the help of NLP techniques like Count-vector and TF-IDF.

1.3 Business Use-case

: It is important to focus on modernizing the existing systems to make products more user-friendly and time-saving. The study of user's sentiment analysis against the product can be useful for making better marketing and manufacturing strategies. Some research and approaches have proven that sentiment analysis may be extremely beneficial to both customers and organizations.

1.4 Novelty

This research work will compare three different machine learning classifier models (MNB, MLP, SGD) to understand the sentiments of customer feedback. NLP techniques, TF-

IDF and count vector were used for feature extraction for text.

1.5 Objectives

The main objective of the research will be divided into subtasks which will include Data collection; Pre-processing of data, Implementing NLP techniques and different machine learning models

Subtask 1: Identify the previous research done on customers' feedback sentiment and understand the NLP techniques and different machine learning models.

Subtask 2: Collect the data related to the consumer feedback on Amazon's product from a web source.

Subtask 3: Pre-process the collected data and apply NLP techniques for the text feature classification of customers review text.

Subtask 4: Implement 3 Machine learning models (Multinomial Naive Bayes (MNB), Multi-layer Perceptron(MLP classifier), and Stochastic Gradient Descent (SGD classifier) and visualize the accuracy of model.

2 Literature Review

In the Section, Different research papers were reviewed to understand the implementation of different machine learning algorithms, NLP techniques, and Data cleaning methods. Many studies have been done in order to understand the sentiment of customers reviews by implementing different machine learning techniques. Also, few research papers have included different feature extraction techniques to analyze the textual data.

2.1 Review of Sentiment Analysis of user feedback by implementing different Machine Learning Models

Sentiment analysis is one of the fastest-growing fields, making it difficult to keep up with all the new developments. Customer reviews on a product can be explained with the help of opinion mining and sentiment analysis. Amazon.com collected the customer's reviews over a large period of time. Sentiment analysis on Twitter, for example, can be used to assess the consensus on a specific hot issue. Companies and brands frequently use sentiment analysis to keep tabs on their brand's reputation on social media and the internet as a whole. (Chauhan and Sehgal; 2017) published his research on Customers' feedback on Amazon's product. They have implemented a POS Tagging technique in the data processing. The research doesn't include the implementation of a machine learning model. They just predict the data into positive, negative, and natural sentiment with the help of the Rating of the product.

Unsupervised and supervised learning approaches are the two most used methods for sentiment analysis. ML techniques such as Naive-Bayes, SVM, and Maximum-Entropy are used to classify tweets based on their sentiment. A classifier's efficiency is determined by which dataset is utilized for which categorization algorithms. The use of training data in the case of Supervised ML algorithms helps in the categorization of test data. There are a lot of research papers where different algorithms of machine learning are being used to predict the (Mashuri et al.; 2019) sentiments of public expressions. They acknowledged that Twitter API is the best method to crawl the tweets from Twitter by parsing the topic as input. Sentiment analysis is looking at how people feel about a given topic or issue. To assist the commencement of a text, NLP (Natural Language Processing) was employed. This was done by tokenizing emotion, removing stop words, and stemming from the data. Sentiment analysis is the center of this research, which employs lexicons and polarity multiplication. It was identified in this research that accuracy results were a bit different from using machine learning models. So, there is still a lot of room for improvement when it comes to the accuracy of outcomes. According to (Kumar et al.; 2021) there are a lot of python libraries that can be used for data pre-processing. Hydra tor tool was used to read the tweets and save them into the local machine for further analysis. Then, Python libraries like NumPy (different operations on the multidimensional array), pandas (data manipulation), Matplotlib (plot the graphs), seaborn (extension of Matplotlib for data visualization), Re (a regular expression is an NLP library that is used for text pattern match), nltk (natural language processing toolkit for sentiment analysis) were used to preprocess the data of tweets related Covid-19 pandemic.



Figure 1: Implementation of SDG classifier Flow Diagram

Machine Learning has made document/text classification a major focus. There are many ways to anticipate accurately the category of any new text/document under examination. The content of the news can now be used to judge the reader's reaction to the item, as the newspaper's news pieces now feature a wide range of moods and inclinations. One of the methods used to analyze these textual data by hybrid methodology. A paper presented by (Singh et al.; 2019)The Nave Bayes model is employed for their research to produce predictions. Bayes' theorem is a statistical classification procedure that uses Bayes' theorem. For datasets with fewer records (312 in this example), this study found that Multinomial Nave Bayes performed marginally better than Bernoulli Nave Bayes, although it only achieved an accuracy of 73 percent, which isn't great. The algorithms mentioned in this article both require large amounts of data in order to reach high accuracy, and this is true for both algorithms.

2.2 Review of the ML models for sentiment analysis

To increase the accuracy for better results, (Mandloi and Patel; 2020) researched and used different methods of machine learning. Opinion mining and Material polarity are the decision-makers that tell the type of sentiments like positive, negative, and neutral. To analyze sentiments, different machine learning models are implemented and compared. Naive Bayes is a classification algorithm that leverages Bayes' theorem. Text classification is a common application and one that takes advantage of a large dataset for high-dimensional training. Classification is the primary goal of the Support Vector Machine classifier. By constructing a hyperplane between the classes of data, we can classify the data in n-dimensional space into separate groups. It is a probabilistic classifier known as the Maximum Entropy Classifier (Roy and Ojha; 2020). Entropy is considered while selecting data to be included in the model. A wide range of applications is possible, from sentiment analysis and language identification to picture classification and text classification. Assumptions regarding the previous distribution can be quite dangerous, thus we employ this method when we have very little information to work with. Compared to the Naive Bayes technique, it took longer to train the data set. The accuracy of Naive Bayes, Support Vector Machine, Maximum Entropy Classifier is 86 percent, 75 percent, and 82 percent respectively.

In (Imran et al.; 2020), the author has conducted the sentiment classification on the Covid-19 tweets from different countries using deep learning approaches. A baseline model of deep learning was applied to the sentiments extracted data. All deep learning models DNN, LSTM Fast Text, LSTM Glove, LSTM Glove Twitter, and LSTM without pretrained embedding got the validation accuracy, precision, F1 score. Fast Text model outperforms well. Then, the author classified the emoticons of tweets into positive (joy and surprise) and negative (sad, anger and fear) and implemented the above models and found that the Glove Twitter model has good accuracy i.e., 70 percent

Perfoamce of any model will depend on the accuracy and F1 score value. Different types of machine learning models are being used in order to get the accuracy of the models. A research paper was submitted by Singla et al. (2017) on sentiment analysis of customer reviews on mobile phones. They have classified the customer's review in a positive and negative review. After that they applied 3 different machine learning models : Naïve Bayes, SVM and Decision Tree. They have achieved the highest accuracy of 81.75 percent with SVM Model.

With the help of the SGD classifier model, research was conducted on Bangla text documents-(Kabir et al.; 2015). They have implemented the machine learning model (SVM and NB model) and compared the accuracy of the model with the accuracy of the SGD classifier model. The document they used is a text data of Indian originated language - Bangla, and they also added feature extraction technique like TF-IDF. Documents are categorized into one or more specified categories by their content, and this task is called document classification. The conclusion of their research show a higher F1-Score and the execution time is less plus the precision o model is depicted as high. Below is the high-level architecture of their implementation.

2.3 Sentiment Analysis using Natural Language Processing

For the sentiment analysis of any data, we use different data cleaning processes to clean the data. NLP techniques are a widely concept for any textual image-based classification. NLP-based preprocessed data framework includes different techniques Tokenization, stemming, and lemmatization; POS tagging; Bag of Words (BOW) model, and the term (TF-IDF). (Hasan et al.; 2019a) has implemented NLP technique on Twitter Data. For this research they defined positive and negative polarity in the tweets acquired using the Twitter streaming API, the main goal is to detect the tweet's emotion. These tweets are used as raw data. In order to make an informed decision on whether to purchase a product or service, a customer can use sentiment analysis. Data extraction, tokenization, lemmatization, stemming, stop word removal, parts of speech tagging and named entity recognition, generating a data frame, text modeling (Hasan et al.; 2019b), and a classifier model are some methods used to analyze sentiment in Twitter data. The use of TF-IDF to identify relevant terms in tweets to predict sentiment has been widely employed. Pickle's object serialization module is used to build a classifier model.

(Lobur et al.; 2011) The natural language processing (NLP) area of machine learning is employed in text analytics, according to (Lobur et al.; 2011). NLP, or natural language processing, is addressed by the Natural Language Toolkit (NLTK), a component of the Python standard library. In addition to text analytics, natural language processing plays a key role in research based on human language analysis. Computational linguistics deals with the creation of models based on human languages for the sake of study. To begin learning about natural language processing, utilizing NLTK is a time-saver because it simplifies the learning process for even the most inexperienced programmers. Numerous advantages of utilizing NLTK include 60 real-world data corpora, a library of grammatical rules, trained models, and a set of functions that may be used for basic natural language processing. Natural language processing is depicted in the following table1. To help its users, NLTK uses a variety of distinct corpora. It's possible to perform natural language processing jobs with the help of various programming languages.

POS Tagging Technique

(Finch and Sumita; 2007) have presented a new approach, they provided a novel paradigm for sequence tagging based on tagging in units of the word sequence. With the use of two different kinds of sequence data, they were able to demonstrate the value of this strategy. For various data types, the phrase-only model outperforms our baseline ME model in terms of performance, according to our testing. For terms where there is no multi-word phrase in its phrasing table, it is weak. With the use of both the phrase-based and classic ME tagging models, this problem can be solved. The performance of the tagger including both models is significantly superior to the performance of either model on its own. Though the two methods use a lexical/tag context that is comparable, they both provide unique information that is useful to the other.

Sentilyzer:Aspect-Oriented Methodology It's been researched by (Wladislav et al.; 2018) Sentilyser library will more be fruitful for textual-based sentiment analysis. This methodology is divided into 2 sub-goals. Textual features that are discussed with the help of the tuples is the second sub-goal. Since Amazon has lot of users and a huge number of reviews are being given by customers against Amazon's product. The customer also gets an option to rate and post the text of the review against that product. This review text is being used for the text-based analysis. Lexicon generator along with sentiment analysis tool has been used for the analysis.

Conclusion of Related Work

Several research work has been examined on sentiment analysis of the customer's review. Most of this research was implemented using the following machine learning models -Naïve bayes, SVM, Random Forest, decision tree, and Regression models. Also, these models provide good performance based on accuracy and precision value. Very little research has been done using the SGD classifier, Multinomial Naive Bayes (MNB) and Multi-layer Perceptron (MLP Classifier) for analysing the sentiment of customer's feedback. Also, literature research helped to understand different NLP techniques that can be used for data cleaning. No major or direct study has been done for determining the sentiment of customers reviews using MNB, MLP and SGD classifier ML models using TF-IDF and Count vector technique. So, the proposed work collects amazon's product data and applies different NLP techniques to analyse the sentiments.

The main reason for implementing machine learning model has two primary goals: first, to classify data using models that have been developed, and second, to predict future outcomes based on these models. It may be used in a hypothetical algorithm for classifying data to train and test it and to achieve the better predictions.



3 Methodology

KDD Process flow for Sentiment Analysis

Knowledge Discovery in Database (KDD)

KDD is the most popular technique for the data mining task. It will include a collection of Raw data, Data cleaning, transforming, and utilizing this data by finding the meaningful pattern using algorithms. The below diagram depicts the flow of the KDD process for sentiment analysis.

3.1 Data Selection

Dataset used for this research is created by famous researcher Julian McAuley from University of California, San Diego. In total, more than 233.1 million reviews are present in the Dataset. Different categories of products are present like Beauty, Art, Electronic, Home and Kitchen.

For this research project, Data set of the product category - Electronics is chosen. The data set is present in JSON format. In total, 20,994,353 reviews are present in the dataset. Also, it includes information like review text, rating, review time, reviewerID.

Google-Colab is being used for implementing the code since Google-colab provides GPU to implement machine learning and deep learning algorithms faster. A new Notebook

is created, and then Google Drive needs to be mounted to google-colab then the dataset can be downloaded directly to google drive with the help of Google-colab.

3.2 Data Preprocessing and Data Transformation

Raw data needs to be cleaned before implementing any algorithms. Data Cleaning will include removal of null values, removing unnecessary columns, changing the date-time format. Different NLP techniques will be used like stop word and Stemming and tokenization. All the important libraries for NLP need to be imported.

Tokenization- With this technique, the whole sentence will be split into words, and later basic cleaning processes will be done like removing punctuation marks and converting text into lower case.

Stop words- In this technique, all the unuseful words get removed from the sentences like: "for", "where", "when". This helps to shorten the data and ultimately helps to reduce the execution time.

Stemming- Stemming is a method of removing affixes from words in order to get to their root form. It's like chopping down a tree to its roots. In the case of sleeping, for example, the root word is sleep. The words are indexed via stemming, which is done by search engines. Because of this, search engines can only record the stems of words, rather not all of a word's variations. A stemmed index is smaller and more accurate because of this.

Bag of Words -

BoW is one of the commonly used feature extraction techniques. It can be used to extract text features for use in machine learning techniques. It mainly focuses on the number of occurrences of a word in a document. It is referred to as a "bag" of words since the document's order and structure are important in this technique. In the paradigm, the only thing that matters is whether known words appear in the document.

TF-IDF(Term frequency-inverse document frequency)- Tf-IDF is mainly used for textual data. It is used for showing the importance of a particular word in the document. TF-IDF can be calculated by multiplying 2 metrics. Its been highlighted by Dalaorao et al. (2019) in their research paper with the use of the Tf-IDF technique for increasing the accuracy of their Machine learning models (RandomForest, MultinomialNB and SVM.) The result showed a 10 percent rise in the performance when collocation is integrated into the enhancement of this process.

 $\ensuremath{1-}$ $\ensuremath{\text{Term}}$ frequency - With this, calculate the frequency of a particular word that is commonly used.

2- Inverse document frequency - IDF is a measure of the term's usefulness. In this, the most common terms, such as stop words, will have a low IDF because they appear in nearly every document. So if the word have IDF value near to zero, that means it is the most common word used.

TF-IDF = Term Frequency (TF) * Inverse Document Frequency (IDF)

3.3 Data Mining

After pre-processing of data, 3 different machine learning models will be implemented Multinomial Naive Bayes, MLP (multilayer perceptron), Stochastic Gradient Descent (SGD) Classifier.

3.4 Interpretation and Evaluation

– In this, the results and the accuracy of different ML models will be evaluated. In every model, a confusion matrix and Classification report will be used to show the result of the algorithm. The implementation of the Algorithm will provide F1 score, precision, recall value.

Evaluation metrics

Accuracy Accuracy will be the ratio of number of correctly classified data instances over the total number of data instances.

Precision - Precision is the ratio of correctly predicted positive classes with the total number of positive predictions present in the model. Ideally, precision value should be 1 (high) for a good classifier.

Recall - Predicted positive instances divided by all predicted positive instances is called Recall. Recall can also be called as Sensitivity or true positive rate follows:

F1 Score: F1 score can be defined as the ratio of Precision and Recall.

4 Design Specification

Design specification of implantation of machine learning models to review text data. The below diagram shows the Research flow diagram.

- Collection of data from Amazon's web source.
- Data cleaning needs to be done by removing null values, removing unnecessary columns.

- Feature extraction technique like Count vector, TF-IDF is applied to the textual data.

- Data is divided in positive, Negative, and neutral ratings.

- machine learning models were implemented (MNB, MLP, and SGD) and with the help of confusion matrix and classification report different results like accuracy, precision, F1 Score, Recall were drawn from each model.

- Result is drawn by comparing and visualizing the accuracy of all 3 models.



Figure 2: High Level flow Chart of our analysis

5 Implementation and Evaluation of Machine Learning Models

This section contains the implementation of machine learning models on the pre-processed data (structured review text) with the help of NLP techniques like feature extraction by TF-IDF, count vector is applied. Also, the data is categorized in 3 sub-parts Positive, negative, and neutral with the help of "Overall" ratings. The following ML models are Multinomial Naive Bayes, MLP (multilayer perceptron) and Stochastic Gradient Descent (SGD classifier). The performance of the models will be presented by using the Confusion matrix and Classification reports graph. Also, the Accuracy, Precision, F1-score and Recall values will be calculated which will help in comparing the performance of all 3 Machine Learning models.

The following programming language and various tools and approaches are used in this research work:

- Tools Google Collab (for code implementation).
- Programming Language Python 3.6.9.
- Libraries sklearn, nltk, stopwords, confusion matrix, RegexpTokenizer, Tokenizer, WordCloud, seaborn and Matplotlib, countVectorizer, TfidfTransformer, TfidfVectorizer.
 Tashnigues Natural language processing with NLTK library Count mater. TE IDE
- Techniques Natural language processing with NLTK library, Count vector, TF-IDF.

Data is divided in 3 sub-parts based on the rating of the product. Whichever products have an" Overall" rating 5 and 4 will be considered as positive, 3 will stand for neutral and 2 and 1 will be for negative. A distribution graph is being drawn to compare the ratio of Neutral, Negative, and Positive ratings. The percentage of Positive words is 80.279 Percent, on the other hand, percentage of neutral and negative words rating are 8.421 and 11.299 Percent as shown in the diagram.

Overall Rating	Category		
5 and 4	Positive		
3	Neutral		
2 and 1	Positive		

Table 1: Categorization of Data



Figure 3:



Most Repeated words in negative reviews

Figure 4: List of Positive, negative and neutral common used word

The list of most repeated words of neutral, positive, and negative reviews words is generated from the data and it's been displayed together. Below is the list of most repeated words in positive, negative and natural reviews.

In Below Bigram plot (Figure 16) of Top 5 most common words of each positive, negative, and neutral reviews vs count of that word used is being visualized in the below

bigrams



Figure 5:

The Evolution of applied machine learning models can do with the help of different matrices like Accuracy, Precision, recall, and F1 score. A total of 142257 reviews were presents after cleaning the data. This data is divided in Train- Test split with the ratio of 80: 20. For implementing a machine learning algorithm we will be using this 20 Percent of test data and with the help of confusion matrix and classification report the result of these algorithm will be predicted.

5.1 Model 1- Multinomial Naive Bayes (MNB)

For Natural Language Processing (NLP), the multinomial Naive Bayes algorithm (MNB) is a probabilistic learning method (NLP). Predicting a text's tag is possible thanks to an algorithm built on top of the Bayes theorem. The tag with the highest probability is output after the probability of each tag for a given sample is calculated. In the Naive Bayes Classifier, each feature being classified is unrelated to any other feature. It is a collection of numerous algorithms based on this one premise.

5.1.1 Experiment 1: Classification of the model with Count vector

Confusion matrix and Classifier classification report will help us to visualize the data. Figure 6 shows the classification report of MNB models. The accuracy of model is 63.87 Percent using Count Vectorization, this accuracy shows the performance of our model. Also, Precision, Recall and F1-score values are 0.62, 0.67, and 0.64 respectively As shown in figure 6.



Figure 6: Classification report of MNB model



Figure 7: Classification report of MNB model

A plot of the Actual label and Predicted label is depicted in below Figure:7. The MNB model correctly predicted 18976 Negative reviews, 15204 Neutral reviews, and 20341 Positive reviews. Also, there are 4122 reviews that wrongly predicted as positive reviews but in reality, they are negative reviews

5.1.2 Experiment 2: Classification of the model with TF-IDF

The accuracy of the MNP model is 65.528 with the TF-IDF feature. Which is more compared to the Count Vector model of MNB. The precision, Recall and F1 value are

0.64, 0.66, 0.65. The performance of this MNB with TF-IDF shows better performance than with Count Vector. Below Figure: 8, show the classification report of the model.



Figure 8: Classification report of MNB model with TF-IDF



Figure 9: Confusion Matrix with TF-IDF

The confusion matrix of the Actual label vs the Predicted label is shown in Figure 9. In these 18823 reviews are negative reviews, 16239 are neutral and 20870 are positive reviews. Also, there are 3254 reviews that are wrongly predicted as positive reviews but in reality, they are negative reviews which is more accurate as compared with the Count vector model. In Last, it can be concluded that the accuracy of the Multinomial Naive Bayes model shows better accuracy with the TF-IDF vector.

5.2 Multi-layer Perceptron (MLP Classifier)

MLP Classifier stands for Multi-layer Perceptron which is a Neural Network in the term itself. Classification is handled by MLP Classifier via a Neural Network, which is different from previous algorithms. Using a feedforward artificial neural network model, MLP predicts the outputs that will be produced based on the data inputs.

5.2.1 Experiment 1: Evaluation of result with Count vector

Figure 10 shows the classification report of the MLP model with a count vector. The accuracy of the MLP model is 70.379. MLP model shows the highest accuracy recorded among all models. The Precision, Recall and F1 score values are 0.73, 0.63 and 0.70 respectively.



Figure 10: Classification report of MLP model with count vector



Figure 11: Confusion Matrix with count vector

As show in Figure 11 the confusion matrix of actual label against Predicted label. It can be seen 19047 reviews are actually negative and 23457 reviews are positive. Also, there are 1347 reviews that were wrongly predicted as positive reviews but, in reality, they are negative reviews. In this result, we can conclude that MLP model performed well than MNB model.

5.2.2 Experiment 2: MLP model with TF-IDF

Below Figure 12 shows the accuracy of MLP model is 66.8631 with TF-IDF feature. Which is very less compared to the accuracy of the Count vector method. Also, the values like Precision(0.66), Recall (0.66) and F1(0.66) score is also less.



Figure 12: Classification report of MLP model with TF-IDF



Figure 13: Confusion Matrix with TF-IDF

The confusion matrix shown in above figure 13, shows the plot of actual label with the predicted label. It also depicts actual negative reviews (18821), positive reviews (21685) and neutral reviews (16565). Overall it's been observed that MLP model is more successful than MNB model in terms of performance and accuracy.

5.3 sklearn.linear model or SGD Classifier

To implement linear classification, we will use the SGD Classifier from scikit-learn. Stochastic Gradient Descent (SGD classifier) is a popular numerical method for finding a function's local minimum (Specially its about the loss function, which calculates how far away from our boundary each instance is. By minimizing the loss function, the algorithm will learn the coefficients of the hyperplane.

Regularized linear models with stochastic gradient descent (SGD Classifier) learning are implemented by sklearn.linear model estimator, which updates the model with a decreasing strength schedule while estimating the loss gradient for each sample (aka learning rate). Dense or sparse arrays of floating-point values can be used in this method. One of several possible models it will fit is a linear support vector machine; this can be changed by altering the loss parameter (SVM).

5.3.1 Experiment 1: SGD classifier model with vector count.

Below figure shows the classification report of SGD classifier model with vector count technique. The accuracy of the SGD classifier model is 68.067 with the vector count technique. The value of Recall (0.73), Precision (0.67), F1 score (0.70).



Figure 14: Classification report of SGD model with count vector



Figure 15: Confusion Matrix with count vector

Actual label vs Predicted label plot is drawn above figure 15. The graph accurately predicts 20851 negative reviews, 24347 Positive reviews and 12901 natural reviews. The performance of this model is little less than MLP model, but it has better accuracy than MNB model.

5.3.2 Experiment 2: SGD Classifier with TF-IDF:

The accuracy of SGD classifier model is 70.4364 which is very high as compared to Countvector method. Also, it has the highest accuracy as compared with other 2 models. The value of Precision (0.70), Recall (0.73) and F1 score (0.71). Below figure 16 shows the above records.



Figure 16: Classification report of SGD model with TF-IDF



Figure 17: Confusion Matrix with TF-IDF

Above figure shows the plot between actual label vs predicted label. The model shows in total 20869 reviews are actual negative, 23765 are actual positive reviews and 15487 actual neutral reviews are present. SGD classifier model with TF-IDF is more successful than MNB and MLP model.

The purpose of the evaluation of machine learning models is to predict the sentiments from the customer reviews on Amazon products. By implementing the different models, we got to know which model gives more accuracy to analyse the sentiments so that customers could save their time and money while searching for their product. Additionally, the findings of this study of user sentiment analysis against the product may be used to improve marketing and production plans.

6 Result Discussion

The implementation of Machine learning (Multinomial Naive Bayes, MLP (multilayer perceptron) and Stochastic Gradient Descent (SGD classifier) models for the sentiment analysis of customer feedback for Amazon's product has shown good accuracy value. SGD classifier model with TF-IDF technique has shown maximum accuracy of 70.43 percent. Since much research is not being done using these machine models on customers review, this research paper might help to analyze the sentiment of customers reviews using one of these models in more depth.

Below figure shows the comparison of the accuracy of different machine learning models. MLP model has shown the maximum accuracy of 70.379 with the Count vector feature extractor and SGD classifier model has an accuracy of 70.436 with ID-TDF feature extractor.



Figure 18: Accuracy vs Algorithm with count vector extractor technique



Figure 19: Accuracy vs Algorithm with TF-IDF extractor technique

7 Comparison between machine learning models implemented for Sentiment Analysis

Below table shows the Performance comparison of different machine learning models which were implanted in this research.

Feature Extractor	Model	Accuracy	Precision	Recall	F1 score
[Multinomial Naive Bayes(MNB)	0.64	0.62	0.67	0.64
	Multi Layer Percepton (MLP)	0.7	0.73	0.67	0.7
Count Vector	Stochastic Gradient Descent (SGD)	0.68	0.67	0.73	0.7
	Multinomial Naive Bayes(MNB)	0.66	0.64	0.66	0.65
	Multi Layer Percepton (MLP)	0.67	0.66	0.66	0.66
TD-IDF	Stochastic Gradient Descent (SGD)	0.7	0.7	0.73	0.71

Figure 20:



Figure 21: Accuracy vs Algorithm with TF-IDF extractor technique

8 Conclusion and Future Work

The main objective of this research paper is to apply different Machine learning models like Multinomial Naive Bayes, MLP (multilayer perceptron) and Stochastic Gradient Descent (SGD CLASSIFIER) for the sentiment analysis of user's review for amazon's product. Since the main data is in textual form, feature extraction techniques- Count vector and TF-IDF are being used. Its accuracy of each model has shown successful performance. MLP model has shown the accuracy of 70.379 Percent with the Count vector feature extractor and SGD Classifier model have maximum accuracy of 70.436 percent with ID-TDF feature extractor. The output of these models has been visualized through confusion matrix and Classification report.

In the Future, a dual sentiment analysis (DSA) technique can also be implemented, which will help to analyze the data more accurately and can provide better performance of machine learning models. Also, we can perform the same machine learning model on larger dataset which will help to calculate the accuracy of models more accurately and there is scope of applying Deep learning model in the future.

8.1 Acknowledgement

I would like to give special thankyou to my supervisor Prof. Divyaa Manimaran Elango for her guidance throughout my research project. She helped me to complete my project in time also has been a great source of advice and encouragement. I would also like to show my gratitude to National college of Ireland and Department of MSc of cloud computing for their guidance.

References

Chauhan, C. and Sehgal, S. (2017). Sentiment analysis on product reviews, 2017 International Conference on Computing, Communication and Automation (ICCCA), pp. 26– 31.

- Dalaorao, G. A., Sison, A. M. and Medina, R. P. (2019). Integrating collocation as tf-idf enhancement to improve classification accuracy, 2019 IEEE 13th International Conference on Telecommunication Systems, Services, and Applications (TSSA), pp. 282–285.
- Finch, A. and Sumita, E. (2007). Phrase-based part-of-speech tagging, 2007 International Conference on Natural Language Processing and Knowledge Engineering, pp. 215–220.
- Hasan, M. R., Maliha, M. and Arifuzzaman, M. (2019a). Sentiment analysis with nlp on twitter data, 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), pp. 1–4.
- Hasan, M. R., Maliha, M. and Arifuzzaman, M. (2019b). Sentiment analysis with nlp on twitter data, 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), IEEE, pp. 1–4.
- Imran, A. S., Doudpota, S. M., Kastrati, Z. and Bhatra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning–a case study on covid-19, arXiv preprint arXiv:2008.10031.
- Kabir, F., Siddique, S., Kotwal, M. R. A. and Huda, M. N. (2015). Bangla text document categorization using stochastic gradient descent (sgd) classifier, 2015 International Conference on Cognitive Computing and Information Processing(CCIP), pp. 1–4.
- Kumar, A., Yun, K., Gebregzabiher, T., Tesfay, B. Y. and Adane, S. G. (2021). Covid19 tweeter dataset sentiment analysis, 2021 Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT), IEEE, pp. 110–115.
- Lobur, M., Romanyuk, A. and Romanyshyn, M. (2011). Using nltk for educational and scientific purposes, 2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), pp. 426–428.
- Mandloi, L. and Patel, R. (2020). Twitter sentiments analysis using machine learning methods, 2020 International Conference for Emerging Technology (INCET), IEEE, pp. 1–5.
- Mashuri, M. et al. (2019). Sentiment analysis in twitter using lexicon based and polarity multiplication, 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), IEEE, pp. 365–368.
- Roy, A. and Ojha, M. (2020). Twitter sentiment analysis using deep learning models, 2020 IEEE 17th India Council International Conference (INDICON), IEEE, pp. 1–6.
- Singh, G., Kumar, B., Gaur, L. and Tyagi, A. (2019). Comparison between multinomial and bernoulli naïve bayes for text classification, 2019 International Conference on Automation, Computational and Technology Management (ICACTM), pp. 593–596.
- Singla, Z., Randhawa, S. and Jain, S. (2017). Sentiment analysis of customer product reviews using machine learning, 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1–5.

Wladislav, S., Johannes, Z., Christian, W., André, K. and Madjid, F. (2018). Sentilyzer: Aspect-oriented sentiment analysis of product reviews, 2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 270–273.