



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 7, Issue 4 - V7I4-1275)

Available online at: <https://www.ijariit.com>

Sentiment and thematic analysis on E-commerce application for user reviews using Machine Learning

Rutuja Sanjay Mane

manerutuja8@gmail.com

JSPM Narhe Technical Campus,
Pune, Maharashtra

Aarti Sanjay Sahib

aartisahib0103@gmail.com

JSPM Narhe Technical Campus,
Pune, Maharashtra

Prachi Sunil Mulay

prachi202000@gmail.com

JSPM Narhe Technical Campus,
Pune, Maharashtra

Sumesh Santosh Mapara

sumeshmapara05@gmail.com

JSPM Narhe Technical Campus,
Pune, Maharashtra

Sulochana Sonkamble

sonkamblesulochana@gmail.com

JSPM Narhe Technical Campus,
Pune, Maharashtra

ABSTRACT

Over the year's we have experienced tremendous growth in the use of ECommerce Applications. Since the pandemic, there has been an escalation in the use of these applications. Hence, we must understand the factors that are affecting the effectiveness of the services. In this paper, we will be analyzing different ecommerce applications on Google Play and App Store by performing sentiment analysis on user reviews by machine learning and then perform thematic analysis to identify the themes of reviews. Sentiment analysis is the process of identifying and categorizing opinions expressed in the text, especially when we want to determine whether the attitude of the customer concerning the services provided is positive, negative, or neutral. Performing Sentiment analysis manually is a humongous task as there are millions of users. Hence we will be implementing different classifiers using supervised ML algorithms. These Classifiers will be trained and compared, then the classifier with the highest accuracy will be used to predict the sentiment polarity. Later on, we will be performing thematic analysis on positive and negative reviews to determine themes representing various factors affecting the effectiveness of e-commerce apps both positively and negatively. In the end, we will be proposing how to tackle the negative issues that are hampering the services

Keywords— E-Commerce, Machine Learning, Sentiment Analysis, Thematic Analysis, Applications

1. INTRODUCTION

Sentiment Analysis is a machine learning tool that is used to analyze the sentiment polarity of a text or a document into positive, negative, or neutral. Sentiment Analysis is mostly used in companies or businesses to detect their reputation on a social level or to understand the customer's emotion through user reviews. Nowadays customers have been very vocal about the products they use on social media or in the review section of the product or application. Now what this does is that it helps the companies understand what the customer needs or what are the additional inputs can make the customer happy. Sentiment Analysis is exceptionally important as it helps the

companies quickly understand the overall opinions of the customers and make accurate decisions.

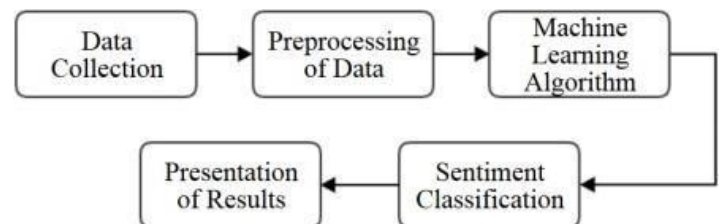


Fig 1.1

Fig 1.1 depicts the architecture of sentiment analysis. At first, the data is collected then Natural language processing is used for data pre-processing. After that the classifiers are trained using the supervised Machine learning algorithm then the best classifier is used for sentiment classification and later on the result is presented.

Thematic Analysis approach extracts themes from text. It is applied to set of texts or documents. In this paper, we will be using thematic analysis on both positive and negative reviews to understand the themes that are representing the factors that are affecting the effectiveness of the product or service. Later on, we will be proposing how to tackle the negative issues that are hampering the services.

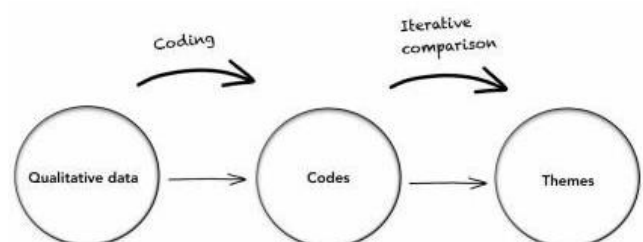


Fig 1.2

Fig 1.2 depicts the process of thematic analysis. The first step is to get to know the data. Second step is coding, now here coding means

specifying section from our text mostly it can be phrases or sentences and later on we have to generate codes to describe that section. Third Step is to generate themes ,In this we have to study the codes we have generated and identify patterns between them and then define themes. Themes are mostly combination of codes with similar pattern. Fourth step is to review the themes that we have created, check weather they are useful and are correct representation of the data. Fifth Step is to define the themes, Now that we have finalized thee themes we have to name and define like what they exactly do they mean. Step six is to write the analysis of whole data

2. RELATED WORK

In[2] implementation of machine learning algorithms is done which process textual ,statistical data provided by yelp data set. Their main aim was to classify reviews into graded categories in a way that they can be sorted in an order from positive to negative. They have used nltk (Natural Language Tool Kit) which is a natural language processing library in python used to fetch the algorithms. To ensure the quality of training data ,over 16000 reviews are classified as positive or negative and the machine is trained to higher levels of precision

In [5], Comparison between Naïve Bayes and Support Vector Machine base on sentiment analysis is done. They have stated that Naïve Bayes Classifier give us an excellent result when it is used for text data analysis

.Naive Bayes is probabilistic classifier. Whereas they have also mentioned that Naïve Bayes and SVM are the best machine learning algorithms for sentiment analysis

In [8],they have used “Beautiful soup” python library for removing punctuations. Stop Words were removed using NLTK(Natural Language Toolkit).In this they have used word2vec model so there was no need for lemmatization because it deals with each word . Their result showed that deep learning approach complicated rather than simple methods as bag of words and even then it shows slightly better results.

3. IMPLEMENTATION DETAILS ALGORITHM

- 1.By using cmd open the desktop Application.
- 2.Input the dataset using UI element.
- 3.The dataset will be displayed on the screen.
- 4.Then the user needs to click on Analyse button to calculate the accuracy and count of tweets that are positive , negative e or neutral.
- 5.A new window opens up where the user can see the number of positive reviews that are present in dataset ,negative reviews as well as neutral review
- 6.Also there is another method to check our sentiment analysis model and that is by entering the review manually in the text box provided.
- 7.Along with this 2 graphs are generated that depicts the effectiveness of a word in the dataset
- 8.Stop

All this process will happen with the help of Logistic Regression Model.

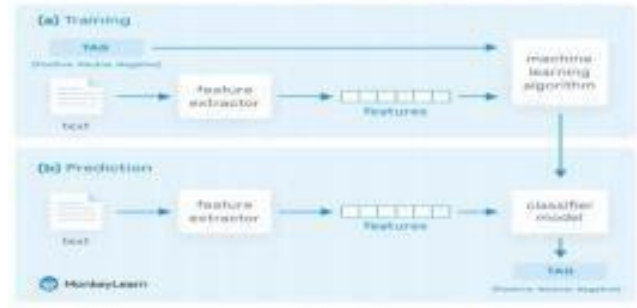
Dataset: The Dataset was provided by kaggle

- It is a training dataset which we are going to use to train the data..
- We split the dataset into two datasets for training and for testing

Proposed Method

The main objective is to understand the factors that are affecting the quality of services.In this paper, we will be performing sentiment and thematic analysis on e- commerce applications using machine learning algorithms.

How Does Sentiment Analysis Work?



Text Input: We have collected the user reviews of various E-Commerce applications.

Tokenization: In this, we have removed punctuations,special characters and extra spaces

Stop Word Filtering: In this step the words that don't make any difference in defining the polarity are removed (e. g – the, an, will, shall, let, may). Stop Word Filtering will be done using NLTK (Natural Language Toolkit).

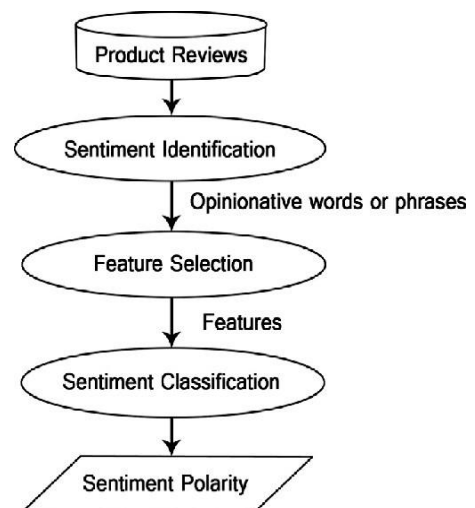
Stemming: In this step we have used WordNet Lemmatizer which is also a part of nltk module, so that words can be converted to their root forms for e.g better becomes good.

Data Vectorization: Here we have applied bag of words approach which helps in extracting unique words from the corpus and vectorize each document using TF- IDF(Term Frequency – Inverse Document Frequency)

Sentiment Classification: In this, we developed Machine Learning models to classify user reviews into positive or negative . we have implemented classifiers using supervised ml algorithms which are used for text classification and then we have trained and evaluate each classifier and the classifier with best accuracy has been used for sentiment classification.

Thematic Analysis: It is applied to set of texts or documents. In this paper, we will be using thematic analysis on both positive and negative reviews to understand the themes that are representing the factors that are affecting the effectiveness of the service

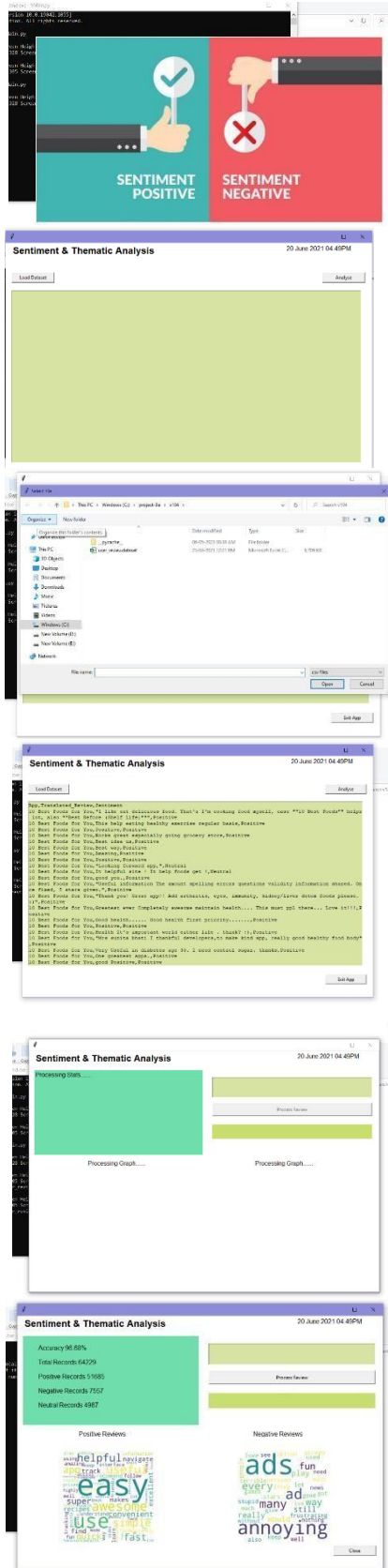
DataFlow:



4. RESULTS

The trained model is able to classify text into positive, negative and neutral .Trained model has an accuracy of 96.68 % .

- (a) Accuracy: It is calculated using the ratio of total no. true positive and true negative to the ratio of the total observations.
- (b) F1-Score: It is the weighted average of recall and precision.



company understand it. Later on we will be proposing how to tackle the negative issues that are hampering the services

REFERENCES

[1] Reddy,Dr& Jaidev&uma.”A review on the concept of sentiment analysisand its role in marketing strategies for e-commerce”. 7. 216-224.”

[2] H. S. and R. Ramathmika, "Sentiment Analysis of Yelp Reviews by Machine Learning,"2019 International Conference on Intelligent Computing and Control Systems (ICCS),Madurai, India,2019, pp. 700-704, doi: 10.1109/ICCS45141.2019.9065812.

[3] A, P., & Priya, K. S. (2020). “A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques”. 2020 6th International Conference on Advanced Computingand Communication Systems (ICACCS). doi:10.1109/icaccs48705.2020.9074312 .

[4] X. Zhang and X. Zheng, "Comparison of Text Sentiment Analysis Based on Machine Learning," 2016 15th International Symposium on Parallel and Distributed Computing(ISPDC), Fuzhou, 2016, pp. 230-233, doi: 10.1109/ISPDC.2016.39. 32

[5] Rathi, M., Malik, A., Varshney, D., Sharma, R., & Mendiratta, S. (2018). Sentiment Analysisof Tweets Using Machine Learning Approach. 2018 Eleventh International Conference on Contemporary Computing (IC3). doi:10.1109/ic3.2018.8530517.

[6] A. M. Rahat, A. Kahir and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2019, pp. 266-270, doi: 10.1109/SMART46866.2019.9117512

[7] U. Hassan, J. Hussain, M. Hussain, M. Sadiq and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2017, pp. 138- 140, doi: 10.1109/ICTC.2017.8190959.

[8] X. Fan, X. Li, F. Du, X. Li and M. Wei, "Apply word vectors for sentiment analysis of APP reviews," 2016 3rd International Conference on Systems and Informatics (ICSAI), Shanghai,2016, pp. 1062-1066, doi: 10.1109/ICSAI.2016.7811108.

[9] Zharmagambetov, A. S., & Pak, A. A. (2015). Sentiment analysis of a document using deep learning approach and decision trees. 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO). doi:10.1109/icecco.2015.7416902

[10] Y. Chandra and A. Jana, "Sentiment Analysis using Machine Learning and Deep Learning," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 1-4, doi:10.23919/INDIACom49435.2020.9083703.

[11] A. Rahman and M. S. Hossen, "Sentiment Analysis on Movie Review Data Using Machine Learning Approach," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2019, pp. 1-4, doi:10.1109/ICBSLP47725.2019.201470.

5. CONCLUSION

In this project we will be investing the task of sentiment analysis as a classification problem .This paper depicts methods for carrying out sentiment analysis of user reviews .We are going to perform sentiment analysis using machine learning (ML) approach to understand user’s emotions regarding E-Commerce apps with the aim of predicting positive and negative sentiments. Later on we will be performing Thematic analysis on those positive and negative reviews to determine the themes that are affecting the effectiveness of applications and help