IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ADVANCED MACHINE LEARNING BASED ASPECT LEVEL SENTIMENT ANALYSIS FOR FLIPKART PRODUCTS

[1]K.Siva Kumar,[2]B.Bharathi,[3]K.Susmitha, [4]Ch.Venkatesh, [5]G.Lavanya

[1]Assistant Professor,[2]Student,[3]Student, [4]Student, [5]Student

[1]Department of Computer Science & Engineering,

[1]Bapatla Engineering College, Bapatla, 522102, India

Abstract: Aspect-based sentiment analysis is a natural language processing technique that is used to analyse the sentiment expressed towards specific aspects or features of a product. The Existing System works on Amazon products using SVM. So, this paper presents a different algorithm on Flipkart products. The main purpose of this project is to develop a system a system to extract reviews from e-commerce websites and extract the aspects from the review and classify the reviews into very positive, positive, neutral , very negative , or negative. The work has been implemented on Flipkart Products. In this work some pre-processing steps are involved like tokenizing, stop word removal , stemming, and lower case. And finally comparing the accuracy of the algorithms which is working accurately

Index Terms **- Aspects, Sentiment Analysis, Multinomial Naive Bayes**

I. INTRODUCTION

Sentiment Analysis is a Natural Language Processing Technique(NLP). It is used to classify whether the given text is positive, neutral, or negative. Now a days many of the organizations are using sentiment analysis to know about the opinion of the customers on their products. Sentiment Analysis involves many steps like text pre-processing, feature extraction, data analysis, sentiment classification, calculating sentiment score. In Text pre-processing some more steps are there they are tokenization, stemming, stop word removal and lower case. Sentiment Analysis identify the strengths, weakness of the product or services which helps the organizations to increase the quantity of the products which people likes and overcome the weakness of the products.

Aspect-Based Sentiment Analysis is a technique of Natural Language Processing(NLP).Which is used to identify the specific features or aspects of a products and services and then it will analyse the opinion of each aspect. Whereas Sentiment Analysis focus on the overall sentiment of review but aspect-based sentiment analysis focus on the aspects of the product. For Example, I like Camera quality of Smart Phone. In this camera is an aspect.

In this paper a Supervised Machine Learning Technique is applied. The algorithms used in our system is Random Forest and Multinomial Naïve Bayes. These algorithms are used for both classification and regression but they are especially used for \classification purpose. Both algorithms are used for training and testing of dataset.

1.1 .LITERATURE SURVEY

Literature Survey is the summary of previously done work on topics. Which explains the what work has been done by other authors. The main purpose of this is to understand the existing work and it helps us for our present work.

Sentiment analysis on large scale Amazon product reviews was presented by Haque, T. U., Saber, N. N., & Shah, F. M. they combine the both manual and active learning approach to label the dataset. And also, they use two kinds of feature extraction approach.

Liu, Z., Yang, N., & Cao, S. (2016) presented a sentiment analysis on review text of micro-videos. They worked more than 20 domestic internet video platforms. Which is used to analyse preference of the viewers and find the users favourite video and suggest the film or videos. The system uses the MYSQL , PHP, APACHE to crawel the data of the videos from internet.

Elsevier B.V reports a paper on sentiment in form of a videos, photos, films etc. This paper shows a systematic approach of sentiment analysis using lexicon-based algorithm.

A typology of viral ad sharers was presented by Kulkarni, K. K., Kalro, A. D., Sharma, D., & Sharma, P to know the sentiment for the viral videos in social media by using advanced tools. The work has been done on 2 things .[1]for to know the viewers response and [2] is to know the four distinct segments about the viral advertisement.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. proposed an Aspect Based Sentiment Analysis . The system works on two domains i.e., restaurants , laptops. The work is mainly focus on the aspects of laptop or restaurants and they interested 163 submissions and 32 team works.

Alexander Pak, Patrick Paroubek have proposed a Twitter as a corpus for opinion mining and collect the data or corpus automatically from website twitter and analyse the data by using the TreeTagger for POS-tagging and find the differences among positive ,negative or neutral. Multinomial Naïve Base algorithm is used.

Recently Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi proposed a sentiment analysis for predicting the stock market using the twitter data. They find the correlation of stock price when it is rising and falling.

1.2. EXISTING SYSTEM

In existing system, they perform sentiment analysis on Amazon product reviews using Support Vector Machine(SVM) which is a supervised Machine Learning Algorithm. SVM is used for both classification and regression but in existing system they used for classification. To know the opinion of the customers on product. Which helps the owners to understand about their products.

The given below is the workflow of the existing system. They work on the aspects of products that means they take the aspects of product from each review and find the sentiment of the review and evaluate the score.

First , Data have to collect from the website i.e., Amazon website. Then extract the aspects which are the product properties and next map the sentiments for ratings. Now discussing about the work flow of existing first collect the customer reviews from website and the second step is pre-processing has to done on that data pre-processing have some steps like stemming, tokenization and stop word removal. And next extracted the aspects from the review and applied the SVM algorithm which is a classifier and map them and finally evaluated the score for the reviews. They use some metrics for evaluation like learning rate, mean squared error (MSE), accuracy, precision, recall, confusion matrix androc curves.

3.1 Workflow of the Existing system

I. RESEARCH METHODOLOGY

In this paper we propose a sentiment analysis on Flipkart products using supervised machine learning techniques i.e., Random Forest and Multinomial Naïve Bayes, to determine whether the given text or review is very positive ,positive, neutral, very negative, very negative.

A. Multinomial Naive Bayes Algorithm

Multinomial Naive Bayes classification algorithm tends to be a baseline solution for sentiment analysis task. The basic idea of Naive Bayes technique is to find the probabilities of classes assigned to texts by using the joint probabilities of words and class. It provides an ability to perform the classification, using small training sets, not requiring to be continuously re-trainedes.The Multinomial Naive Bayes is one of the variants of the Naive Bayes algorithm in machine learning. It is very useful to use on a dataset that is distributed multinomially. This algorithm is especially preferred in classification tasks based on natural language processing.

$$Posterior\ Probability = \frac{Conditional\ Probability * Prior\ Probability}{Predictor\ Prior\ Probability}$$

$$P\left(\frac{A}{B}\right) = \left(\frac{P(A \cap B)}{P(B)}\right) = \frac{P(A) * P\left(\frac{B}{A}\right)}{P(B)}$$

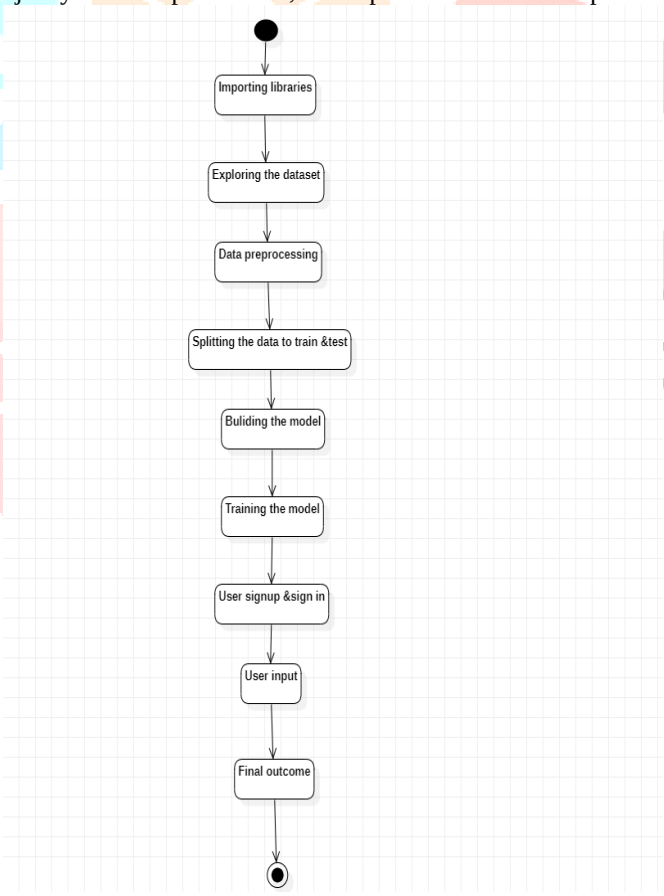
where,
PA= the prior probability of occurring A
PBA= the condition probability of B given that A occurs
PAB= the condition probability of A given that B occurs
PB= the probability of occurring B

The posterior probability, can be interpreted as: What is the revised probability of an event occurring after taking new information into consideration?It is a better reflection of the underlying truth of a data generating process because it includes more information.

1) Conditional Probability:
The probability of one event A occurring when another event B with some relationship to A has already occurred is called conditional probability.This expression is valid only when P(A) is greater than zero.

2) Prior Probability
This probability can be defined as the prior knowledge or belief i.e., the probability of an event computed before the collection of new data. This probability is revised as new information becomes available to produce more accurate results.

B. Random Forest Algorithm
"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average
to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.



4.1 Activity of a Proposed system

At first , we import the libraries which are required to project and then explore the dataset . We are using the filpkart product dataset i.e., Earphones. The next step is Data pre-processing that includes data cleaning i.e., stemming, tokenization , stop word removal and lowercase. The next step is splitting the dataset into two sets they are training set and testing set. Then we build a model and train the models that means training the dataset with the algorithms Random Forest and Multinomial Naïve Bayes. Later user will sign or sign up with their details and gives input means comment and finally it will predict the output.

IV. RESULTS AND DISCUSSION

In this paper , we are introducing Advanced machine learning based aspect level sentiment analysis for flipkart products. The algorithms Random Forest, Multinomial Naïve Bayes are used to determine the sentiment of review whether it is very positive, positive, very negative, negative or neutral. At last, comparing the accuracy of the existing system algorithm i.e., SVM and proposed system algorithms i.e., Random Forest and Multinomial Naïve Bayes . Finally, we get accuracy for SVM is 84.9% and for Random Forest we get 85.1% and For Multinomial Naïve Bayes we get 87% accuracy . We compare the accuracy of three algorithms and find that Multinomial Naïve Bayes get more accuracy.

Sentiment Analysis is the up-to-date challenge area to work. The presented work in this paper is sentiment analysis on Flipkart products i.e., Earphones. The future work has to been done on some other challenging places like ecommerce , movie areas, spammers, fake reviews etc. with the up-to-date tools. Use advance algorithms to get efficiency result .

II. ACKNOWLEDGMENT

We would like to thank the college management and staff for providing the resources to complete the project in time.

REFERENCES

[1] Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: "Analyzingand comparing opinions on the web. In Proceedings of international conference on world wide web".

[2] Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining In: Proceedings of the Seventh conference on International Language Resources and Evaluation.. European Languages Resources Association, Valletta, Malta.

[3] M.Pontiki,D. Galanis,J. Pavlopoulos, H. Papageorgiou,I.Androutsopoulos & S. Manandhar . SemEval-2014task 4: Aspect based sentiment analysis. In Proceedings of the 8thinternationalworkshop on semantic evaluation (pp. 27–35)(2014).Dublin

[4] Xing Fang & Justin Zhan (2015) Sentiment using product review data

[5] E. M. Soledad and Y.-F. W. , "Amazon Reviews, business analytics withsentiment analysis," ResearchGate, 2016.

[6] Fouad M, Gharib T, Mashat A (2018)." EfficientTwitter Sentiment Analysis System with FeatureSelection and Classifier Ensemble

[7] Mitchell, R. (2018)."Web scraping with python :collecting more data from the modern web. Newton".

[8] Alsaeedi, A., & Khan, M. Z. (2019). A study on sentimentanalysis techniques ofTwitter Data . International Journal of Advanced Computer Science and Applications,10(2), 361-374

[9] F. Abid, M. Alam, M .Yasir, & C. Li, "Sentiment analysis through recurrent Variants latterly on convolutionalNeural networkofTwitter," FutureGener.Comput. Syst.,vol.95,pp.292-308,Jun.2019

[10] D. Shailendra Narayan Singh &Twinkle Sarraf(2020),"Sentiment Analysisof a Productbased On User Reviews using Random Forests Algorithm,"IEEE, pp.112-116

[11] Andreea-Maria Copaceanu (2021)Sentiment Analysis Using MachineLearning Approach.

[12] Dr. D. Sivaganesan , Dhruv Aggarwal , Sridhar K , Arunkumar M (June 2022). Feature based Sentiment Analysis for Product ReviewsVolume 11,Issue 06

[13] AnkitGoyal,AmeyParulekar "Sentiment AnalysisForMovie Reviews"

[14] LI YANG , YING LI , JIN WANG , SIMON SHERRATT .R(2020) Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning