

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 7.056



IJCSMC, Vol. 9, Issue. 12, December 2020, pg.68 – 75

Multi-Classifer based Sentiment Analysis for Opinionated Data Posted in Social Networking

Mr. Partha Sarker¹; Dr. Siddhartha Roy²

¹Department of Computer Science, The University of Burdwan, West Bengal, India

²Department of Computer Science, The Heritage College, Kolkata, India

¹ partha_sarker_personal1989@yahoo.com; ² siddhartha.roy@thc.edu.in

DOI: 10.47760/ijcsmc.2020.v09i12.009

Abstract— *Huge quantities of information are available through web and social media platforms have made available to users worldwide. Social networking is the most valuable source of learning, getting ideas, reviews for a product or a service. Everyday millions of users post view about a product, person or a place, give their views, ideas on blog, twitter. Ideas use reviews, assessment surveys, and online networking as a device to obtain criticism on their items and services. The web has extremely huge number and size it is very difficult to handle and understand such reviews. Sentiment analysis is an emerging field in the research area is used to analyse and extracts the opinion from the given review and the analysis process includes natural language processing (NLP), computational linguistics, text analytics and classifying the polarity of the opinion. In the field of sentiment analysis there are many algorithms exist to tackle NLP problems. This immense volume of information may forces potential beneficial business related data, which when separated keenly and spoke to sensibly, can be a mine of gold for administrations Research &Development (R&D), attempting to add library an item in light of prominent popular opinion. In this paper, we use the multi-classifier model in the area of Sentiment Analysis using Support Vector Machine which worked on review /opinion posted on social network/ twitter. Our work aims at forfeiting these researches with greater accuracy and potential to excel in both retail and corporate forums*

Keywords— *Opinion mining, Sentiment analysis, sentiment target, Emotion analysis*

I. INTRODUCTION

Huge amount of information, reviews or opinions are posted in the websites of social media or e-services in the form of raw data every day. Proper method is required to work with those raw data. Sentiment analysis, analyses people's opinions, sentiments attitudes, appraisals, and emotions toward entities such as products, services, organizations, individuals, events, issues, or topics and their attributes expressed in written text[1]. It has Sentiment analysis are also named as sentiment analysis, opinion mining, opinion analysis, opinion extraction, sentiment mining, and review mining having slightly different tasks, all are under the umbrella of sentiment analysis. Natural language processing (NLP), text analysis, computational linguistics, and biometrics are some part of sentimental analysis which are used to systematically identify, extract, quantify and study sentimental states and subjective information[2]. Sentiment analysis is widely used analyze customer's reviews and survey responses, online and social media, and healthcare materials[3]._Nasukawa and Yi (2003) first introduced the term of sentiment analysis, and the term *opinion mining* first discussed in the paper written by in Dave et al. (2003).Sentiment analysis is a type of data mining that measures the preference of

people’s opinions through NLP, computational linguistics and text analysis, which are used to extract and analyze individual information from the social media and similar sources. The analyzed data quantifies the general community’s sentiments or reactions toward certain products, people or ideas and unveil the appropriate polarity of the information. Sentiment analysis systematically rates human emotional states according to positive or negative polarity or a neutral or mixed value, or according to mood, emotion, or feelings (angry, joyful, depressing, conceited, saddened, etc.) and to use sentiment data for business purposes. Prejudice and sentiment affluence to human communications, whether in conversations or posted online or to our social networks. Customer’s view. mood, opinion, and emotion carries enormous business opportunities when captured electronically. It has very broad application areas such as ,Social media monitoring, Brand monitoring, Voice of customer (VoC),Customer service, Market research and analysis, Workforce analytics and voice of employee and Market research and analysis.. With more rapid Internet, people typically search for information on the Internet.

Typically, a large volume of documents, webpages or learning objects are published electronically by search engine in no specific order. Online customer reviews are regarded as a important source of information that is helpful for both future consumers as well as the companies themselves In webpages, reviews are presented in natural language in unstructured format. The job of manual scanning through huge quantities of reviews is computationally expensive and is not practicable with regard to businesses. Hence, it is more effective to automate processing of those reviews and present the required information in an appropriate format. The high level issue of opinion summarization handles the determination of sentiments, attitudes or opinions which authors have conveyed in a natural language format with regard to a particular attribute Sentiment analysis can be applied at different levels of scope:

- *Document level*: sentiment analysis obtains the sentiment of a paragraph or entire document .*Sentence level* : sentiment analysis obtains the sentiment of a single sentence.
- *Sub-sentence level* : sentiment analysis obtains the sentiment of sub-expressions within a sentence.(Aspect based)

The following figure (Fig 1) illustrates the taxonomy of sentimental analysis.

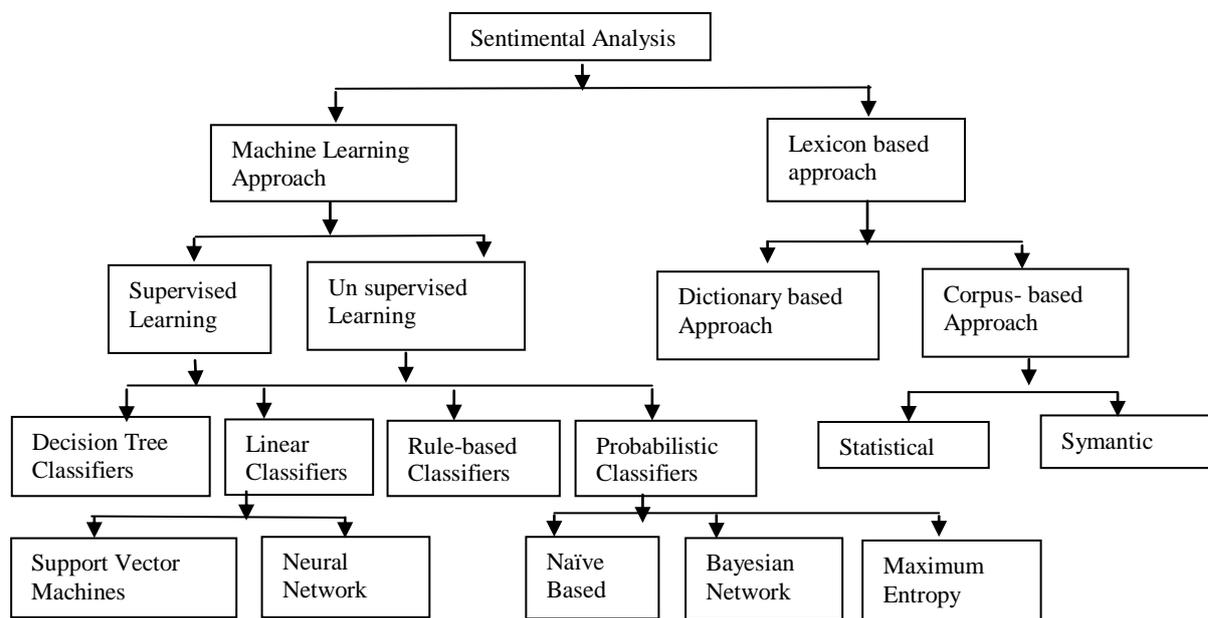


Fig 1 Taxonomy of Sentimental Analysis

This paper presents the sentimental analysis of some well-known methods or proposal of Sentiment Analysis. The new approach follows machine learning technique at document level . The Standard classifier like Naive Bayes (NB), is used to deduct result and for analysis. After introductory section, literature review is discussed in section II .Steps towards steps towards sentimental analysis is discussed in Section III. Present work flow is discussed in section IV. Simulation results discussed in section V where as section VI and Section VII we discuss the future scope and conclusion.

II. LITERATURE REVIEW

Various works have been carried out on sentiment classification and opinion mining are reviewed. In the first section various techniques used for Opinion mining and sentiment classification is investigated. Subsequently feature selection techniques which is critical for unstructured data classification is reviewed. The final sections deal with classification and optimization techniques. Balahur et al (2009) presented a comparison on the techniques as well as resources which may be utilized for mining opinions[4]. Somprasertsri and Lalitrojwong (2010) proposed a method for searching product attributes as well as opinions on the basis of considering syntactic as well as semantic information[5]. Through application of dependency relationships as well as ontological knowledge with probabilistic based models, outcomes of experiment showed that the proposed method was more flexible as well as efficient. Schneider et al., (2009) proposed a novel matrix learning strategy for extending relevance learning vector quantization (RLVQ), an effective prototype-based classification protocol, toward a general adaptive measure. Through introduction of a full matrix of relevance factors in the distance metric, correlations between various attributes as well as their significance for classification occurs at the time of training. Page Layout When contrasted with weighted Euclidean measure utilized in RLVQ as well as its variants. Li et al., (2010) proposed an opinion mining system which put emphasis on polarity calculating techniques and helps opinion information from reviews. Features lexicon and sentiment lexicon were constructed for mining attributes as well as affective entities. The experimental results showed the system is feasible and effective. Sentiment mining and retrieval system which mines functional knowledge from product reviews found in the paper written by Ji et al., (2010), where a comparison between positive and negative evaluation towards sentiment orientation were presented in the system and experimental results on a real-world dataset have shown quite satisfactory and effective. Veeraselvi and Saranya (2014) presented opinion detection and organization subsystem, based on Genetic-Based Machine Learning (GBML)[6] technique which integrates into proposed larger question-answering system. The classification of a review was estimated through the average semantic orientation of phrases and experimental results of the proposed techniques were satisfactory. Jia et al., (2008) proposed a new method of semantic similarity calculation. The concepts were classified into three classes: simple concept; complex concept and combined concept. To different concept, different method was designed and then transformed the similarity calculation of concept into the similarity calculation of the sememe. The similarity of the sememe was computed by the hyponymy of the sememe in the sememe tree. Experiments showed the new approach was effective to the similarity calculation and outperformed the conventional computed approaches. A novel type of tree known as opinion tree was proposed and described by Ding et al., (2009). An opinion tree based flexible opinion mining system was formulated and OM were realized in a unified flexible model. Finally, an experiment on how to construct the opinion tree was finished, and overall opinions of one internet topic was generated in the opinion tree. A review encompassing approaches to ensure opinion oriented text data was suggested by Khan et al., (2009) which dealt with sentiment analysis and utilized classification methods were identified for opinionated documents to assist further research. Binali et al., (2009), which proposed an opinion mining model and revealed novel research areas. A sentiment phrase classification vector based web Opinion mining algorithm was presented by Han et al., (2010). Through sentiment phrase classification techniques the algorithm compared similarity between document vectors, mined the document's theme and judged document theme attributes. Experiments revealed the algorithm's better effectiveness and practicality. Related studies regarding market prediction were surveyed by Nassirtoussi et al., (2014) on the basis of online-text-mining for producing a picture of generic elements which all have. This comparative systems analyses extends into theoretical as well as technical foundations and assists research in structuring the field as well as identifying particular features requiring more research. The creation of words in English as well as Chinese from alphabets and characters to words as well as phrases correspondingly was delineated by Chen and Su(2009). The keyless research model acquires deferential quantity of characters in a phrase, from several Chinese texts. Synonyms are discovered from synonymous procedure. Around fifty top keywords as well as fifty bottom keywords delineated quality of class teaching by students in an open-ended writing component in a questionnaire was utilized for evaluations. It is proven that keywords are from texts when keywords are not described prior, frequency of keyword occurrence in text documents is acquired and the model is applied to other domains. An opinion mining model which extracted opinions as well as views of customers and examined them for providing concrete market flow with confirmed statistical data was proven by Shandilya and Jain (2009). The software utilized classification, clustering as well as lingual knowledge-based opinion mining for providing the attributes. The rapid growth of computer based high-throughput method has offered unparalleled opportunities for but, huge amounts of high dimensional data are possess a challenge to data mining methods. Features selection is an important stage in data mining applications that can efficiently decrease data dimensionality through removal of non-relevant attributes For filling this gap, Zhao et al., (2010) presented a features selection archive that was formulated for collecting the most famous protocols which have been formulated in the features selection research for serving as a platform to facilitate their application, comparison as well as joint study. Yu and Liu (2004) showed that feature relevance alone is not sufficient for effective features selection of high-dimensional data. Features redundancy was defined

as well as proposed for performing explicit redundancy analysis in features selection. A novel framework was suggested which decouples relevance as well as redundancy analyses. A correlation based technique was developed for relevance as well as redundancy analysis, and conducted an empirical study of its efficacy contrasting it with other representative techniques. Wang et al., (2010) proposed an effective method of semantic role labeling on the basis of hybrid comparative patterns for Chinese comparative sentences. In the proposed method, the original hybrid comparative patterns were constructed as per the syntactic structures of comparative sentences. The outcomes of experiments indicated the efficacy of the suggested method. Shein and Nyunt (2010) suggested an ontology based combination method for classifying sentiments. The suggested technique merged NLP methods, ontology based on FCA design, as well as SVM for classification of software reviews as positive, neutral or negative. Cho et al., (2014) proposed a Systemic analyses frameworks for g Korean Twitter data for mining temporal as well as spatial trends of brand images. Sentiment classification is carried out by a SVM as well as multi-nominal Naive Bayes classifier. Liang et al., (2010) proposed mining for user opinions on products on the basis of item taxonomy by experts It looked into personalized item recommendations on the basis of user opinions. Ma et al., (2011) implemented an opinion mining tool which hybridized three different methods: for semantic patterns, which simplified the structure of the natural language syntax; weighted sentiment lexicon, which used as semantic feature words; and finally based on traditional KNN or SVM classification method. Three supervised machine learning protocols which are Naïve Bayes, SVM as well as character based N-gram model were contrasted for sentiment classification of reviews on travel blogs for seven popular travel destinations in the US as well as Europe by Ye et al., (2009). Experiments proved that the SVM as well as N-gram methods performed better than Naïve Bayes and when training data sets had several reviews, all three methods displayed a minimum of 80% accuracy. Sentiment classification methods are included in mining reviews domain from travel blogs in this research.

III.STEPS IN SENTIMENTAL ANALYSIS

Sentiment analysis is the progress of a classification system to classify a piece of text as “positive” or “negative”. Sometime this classification system may includes “neutral”, or even positivity or negativity on a scale of 1 to 5, or 1 to 10. A very clear methodology to developing sentiment classifiers exist in terms of NLP. Various steps towards sentimental analysis is discussed elaborately In the following section.

Step 1: Text preprocessing:

This is the first step . In this step unnecessary aspects of the text information which are unlikely provide any additional information to the model is filtered out. It also allows to streamline the text in the model with more information. Some examples **are as follows :-**

1. Replacing all positive emoticons with “HAPPY”. Removing unnecessary punctuation. Punctuation is unlikely to provide additional information.
2. Removing stop-words, such as “to”, “and”, “do”. These occur very frequently and can create a onfusion of the model, causing it to perform worse.

Step 2: Feature generation

Once pre-processing of text is done it is ready to be put into a model, the next step is to generate features that can be used in the model. This means that we need to convert the text of each document into some numerical value that can be interpreted by the model. There are a number of popular features extracting tasks exist for NLP tasks:-

1. Word embeddings. Create a feature vectors for the document using pre-trained word embeddings such as GoogleNews word2vec or GloVe or InferSent
2. Lexicons. Most frequently word is chosen as Lexicons . NLTK is used to extract the most commonly occurring words in positive text. Count of the number of positive/negative lexicons in each document which can be treated as a features.
3. TF-IDF. This is a weighting metric used provide information on word importance across documents.

Step 3: Initial classification

Once the feature array set up is done, it can feed it into models to make predictions. There are various options are there we need to select any one of them . All are discussed below

1. Train/test split: Split entire data up into two parts say 90% data for training the model, and rest 10% for testing. The results obtained from testing are final results.
2. Cross-validation: This involves iteratively creating different train/test splits (called fold) across the data which gives different scores from testing. Then the average score is taking for prediction. Overfitting can be avoided by this method.

3. **Stratified cross-validation:** In this method it ensures that each fold has the same distribution of classes. For example, if 70% of data is classed as positive, and 30% negative, then each fold will have 70% positive documents and 30% negative.

To obtain the scoring matrix, various parameters such as Accuracy, F1-Score which is mainly useful for binary classification and Macro-averaged / Micro-averaged F1-score which is useful for multilabel classification with an unequal number of each class are used. There are various algorithms exist such as Naive Bayes, Logistic Regression, Support Vector Machine, Random Forests and Neural Networks. Different algorithm is used for various tasks, for example, Logistic Regression with word embeddings and lexicons as features.

Step 4 : Hyperparameter tuning

With hyperparameter tuning re-implementing the best model from step 3 can be obtained, but tuning the hyperparameters of the model to try to improve performance. Once it performs really well, test it again on the test set to make sure that tuned the model to the point of over-fitting the data.

Step 5 : Final Visualisation

In this stage visualization of graphical representation over the various graphs is obtained (including with the percentage of three sentiments which will be in the given text).

IV. PRESENT WORK

In our work we have used Naïve base classifier as it requires a small amount of training data to estimate the test data. So, the training period is less. It is easy to implement and this algorithm works quickly and can save a lot of time. It is also suitable for solving multi-class prediction problems. Since in our model features are independence, it can perform better than other models and requires much less training data. Moreover Naive Bayes is better suited for categorical input variables than numerical variables. To implement the project we have used Java script in the front end and python 3.7.3 with python library such as numpy, pyppass, pystocks[7],[8],[9] with backend as SQLite (V3) DB Browser. We run the code in windows 7. Algorithm for and workflow (Fig 2(a) and 2(b)) four proposed work is shown as follows

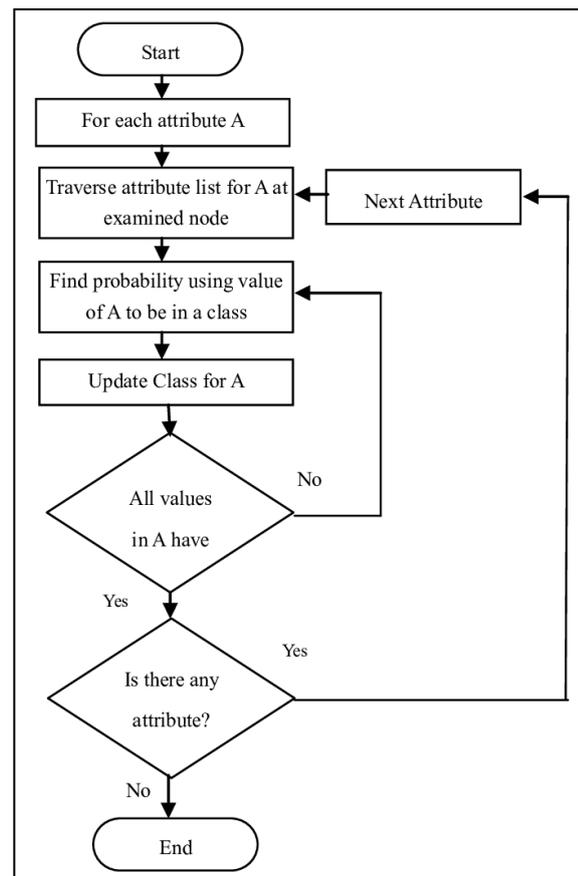
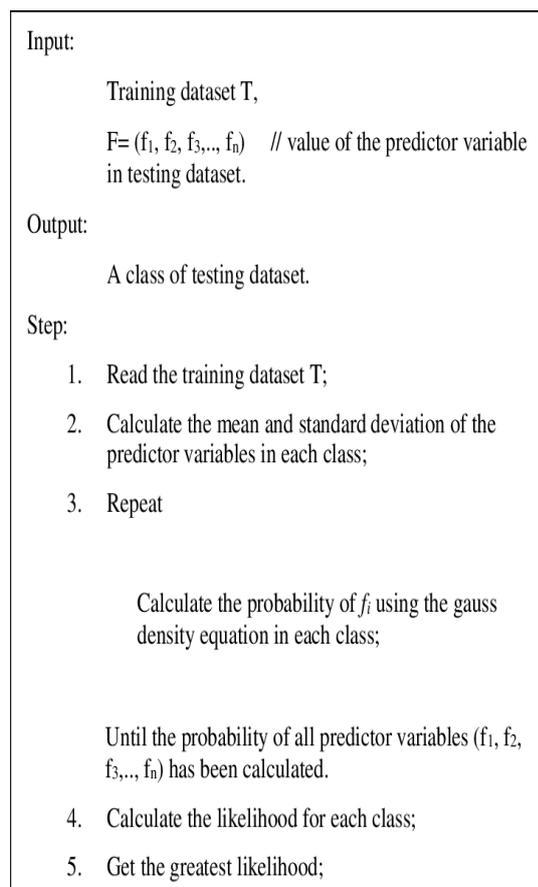


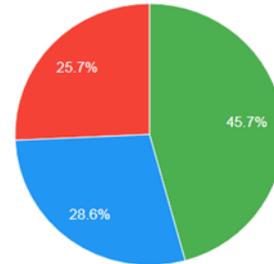
Fig 2(a) Algorithm for our proposed work

Fig 2(b) Workflow for our proposed work

V. RESULT AND ANALYSIS

In the following section we have extracting the text message from twitter as well as product review from social media. Apply sentimental analysis and shown the graphical images using pie chart as well as bar chart and classify into three categories namely, positive, neutral and negative opinion.

Result & Analysis for Twitter :-



● Positive ● Neutral ● Negative

Fig 3(a)- Pie Chart on Sentiment Analysis of Tweet

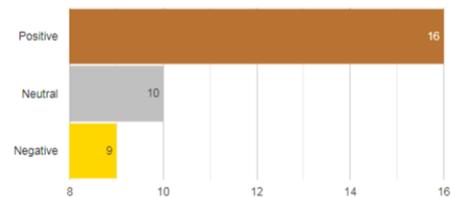
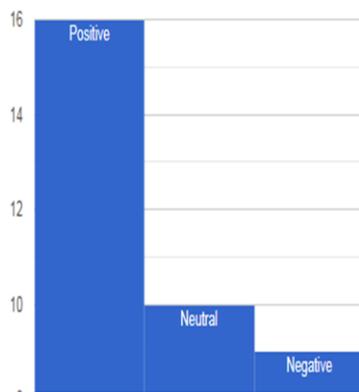


Fig 3(b) Column Chart on Sentiment Analysis of Tweet Fig 3(c):- Bar Chart on Sentiment Analysis of Tweet

Product Review :- (Sony Extra Bass SRS-XB10 Portable Splash-Proof Wireless Speakers with Bluetooth and NFC (Blue))

It had a great impression on me at very first, but later after 5-7 days I came to know the qualities of product which is very bad. It is not at all working properly though it has a good battery life but I am feeling totally bad to buy this product. Poor sound quality and also not working properly now.

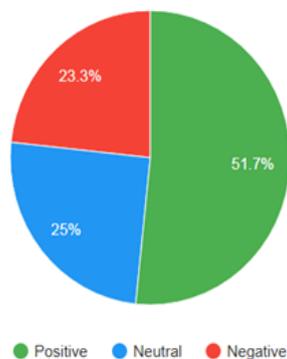


Fig 4(a):- Pie Chart on Sentiment Analysis of Amazon Product Review

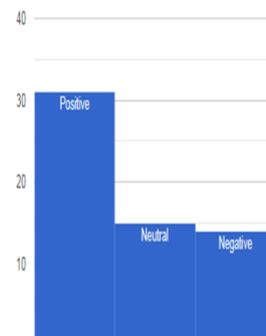


Fig 4(b) Column Chart on Sentiment Amazon Product Review

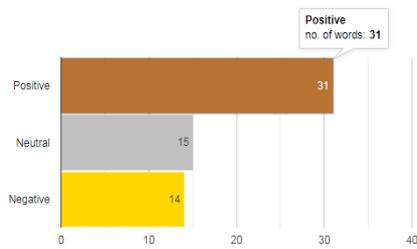


Fig 4(c):- Bar Chart on Sentiment Analysis of Amazon Product Review (Bluetooth headphone)

Everything is fine but the main problem is it's battery. It doesn't last long. Max 3-4 hours and that even I was continuous just calling.

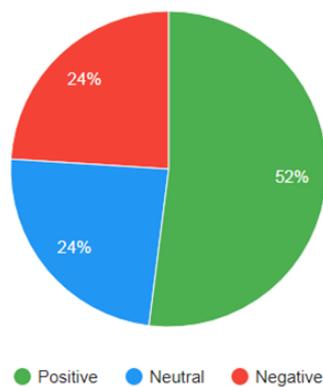


Fig 5(a):- Pie Chart on Sentiment Analysis of Analysis of Flipkart Product Review

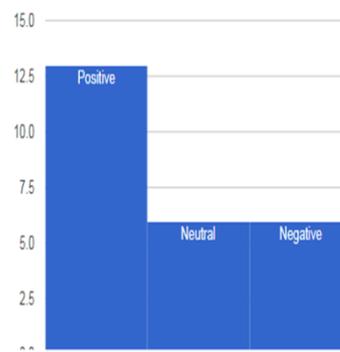


Fig 5(b)- Column Chart on Sentiment Flipkart Product Review

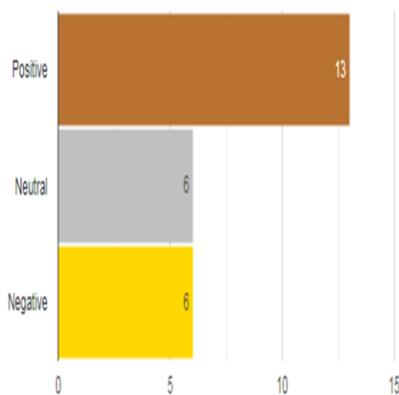


Fig 5(c):- Bar Chart on Sentiment Analysis of Flipkart Product Review

VI.LIMITATION AND FUTURE SCOPE

We have implemented using Naive Bays classifier method for some advantages .But this algorithm has some limitation also. Practical limitation of Naive Bayes is that all the attributes i.e, all features are independent, which rarely happens in real life. This is the reason that Naïve based should not use in real world problems. Another limitation of this algorithm is that This assigns zero probability to a categorical variable whose category in the test data set wasn't available in the training dataset is called 'zero-frequency problem'. So implementing with other classifier algorithm and comparison study is the future scope of our work. Moreover feature extraction is yet to be done. It can be more user friendly as well as more customized in future. There is no

refreshment functionality in the history of analysis; it can be refreshed (when certain limit has been reached, it will be refreshed time to time) in future. In this paper we have used three sentimental segments (Positive, Negative, Neutral) in result and analysis section but in future it can be added more sentimental segments like Extreme positive and Extreme Negative along with the used sentiment segment (Positive, Negative, Neutral).

VII. CONCLUSION

Field of sentiment analysis is an exciting new research direction due to large number of real-world applications. Document-level sentiment analysis is one of the significant components of this area where discovering people's opinion is important in better decision-making. Recently, people have started expressing their opinions on the Web that has increased the need of analyzing the opinionated online content for various real-world applications. In this paper three sentimental segments (Positive, Negative, Neutral) are implemented using SVM for detecting sentiment from the text. Text is mainly extracted from twitter and other social media.

REFERENCES

- [1]. B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 1st Edition, Cambridge University Press, June 2015.
- [2]. Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS)*, 2008. 26(3).
- [3]. Cambria, E; Schuller, B; Xia, Y; Havasi, C (2013). "New avenues in opinion mining and sentiment analysis". *IEEE Intelligent Systems*. 28 (2): 15–21.
- [4]. Alexandra Balahur, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen, Mijail Kabadjov, "Opinion Mining on Newspaper Quotations", Conference: Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence and International Conference on Intelligent Agent Technology - Workshops, Milan, Italy, 15-18 September 2009.
- [5]. Gamgarn Somprasertsri, Pattarachai Lalitrojwong, "Automatic product feature extraction from online product reviews using maximum entropy with lexical and syntactic features", Conference: Information Reuse and Integration, 2008.
- [6]. S. J. Veeraselvi, C. Saranya, "Semantic orientation approach for sentiment classification", 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)
- [7]. R. Nageswara Rao *Core Python Programming*, 2018.
- [8]. J. P. Mueller, L. Massaron *Machine Learning (in Python and R) For Dummies 2016 Introduction to Machine Learning with Python: A Guide for Data Scientists – 2016*
- [9]. M. Lutz *Learning Python*, 5th Edition