International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Aspect Based Opinion Mining on Mobile Product

Shiji Abraham, Minu P. Abraham, Uday Kumar Reddy K. R., Anisha P. Rodrigues

Abstract: ***In this era, the web technology is growing quickly. Many people express their feedback related to products, social issues and services. As e-commerce site is becoming more popular, the customer review related to the product grows quickly. Due to this growth it is very difficult for the customer to read huge amount of data and make a decision whether to buy a product or not. It is also very difficult for the manufacturer of the product in-order to manage and focus on customer opinions. In this research we focus on mobile product review which is extracted from Kaggle site. In this experiment we have focused on one particular mobile product review that is Samsung. After data collection we do pre-processing, and further we extract aspect and corresponding opinion using Natural language processing and then categorize whether the extracted opinion is positive or negative by finding polarity for each extracted opinion of words. Finally performance evaluation is done by using two machine learning algorithm i.e. Multinomial Naive Bayesian (MNB) and K-Nearest Neighbour (K-NN) algorithm. This performance evaluation is calculated based on bag of words. Out of two algorithms K-NN gave best accuracy compared to Multinomial Naive Bayesian.***

Keywords: ***aspect, opinion, Natural language processing (NLP), Multinomial Naive Bayesian (MNB), K-Nearest Neighbour (K-NN)***

I. INTRODUCTION

In past decades, the community of researchers had witnessed the improvement in technology and increase in the internet actives such as E-commerce sites, social Medias and websites that has provided positive impact on research activity. Sentiment Analysis or Opinion mining is one of the field which is benefited from this advanced technology and the internet which has been defined as the computational process for recognition, identifying the people opinion for different entities. The people post their feedback based on the products, services, hotels and political issues of their interest. The applications in sentiment analysis are used in monitoring and analysis of the public opinions related to political issues. Sentiment analysis is also used for measuring market intelligent for whether the user is satisfied on the services or product and improving the product according to the customer feedback.

Sentiment analyses are classified into three types; they are Document Level, Aspect Level and Sentence Level. Document level is used to identify whether the review is positive or negative. In sentence level is used to identify whether every sentence is positive or negative and in aspect level entities and their aspects/features sentiment positive or negative.

1.1 Document Level: Here the documents are considered entirely; the resultants of each opinion are presented via the opinion holder and categorized into positive, negative and neutral [1]. Primarily based at the view of single opinion holder, a hypothesis is made for each document that expresses a particular subject sentiment from their view. Document level analysis isn't always employed as comparative sentences may be found in forms and blogs. Opinion holders may additionally contrast an entity which has features that is comparative.

1.2 Sentence Level: Here the document is splited into sentences, each sentences are considered as a single entity and make sense that if every communicated sentence is positive, negative or neutral sentiment. Here neutral means there is no opinion about the sentence. This level of investigation is related to the subjective arrangement which recognizes the sentence called as target sentence that expresses genuine information from the sentence called subjective sentences. It is not possible to compare subjective of supposition with same number of target sentence for example: "We purchased new car a month ago and the windshield wiper has tumbled off".

1.3 Aspect Level: Aspects are nothing but features of product for example "Camera is awesome". Here Camera is considered as aspect. The aspect in the sentence can be explicit or implicit. Consider an example "The voice quality of this phone is great". In this example the customers are speaking about the voice quality of the phone. In this example the feature is voice quality and this feature is specified directly in this sentence these are known as explicit aspects. Consider another example "This camera is too large" in this statement the customer is speaking about size of the camera. But it is not specified in this sentence this is called as implicit aspects.

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
Shiji Abraham*, Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, India.
E-mail: shijiabraham96@gmail.com.
Minu P. Abraham, Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, India.
E-mail: minupjuly12@nitte.edu.in.
Uday Kumar Reddy K. R., Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, India.
E-mail: krudaykumar@nitte.edu.in.
Anisha P. Rodrigues, Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, India.
E-mail: anishapr@nitte.edu.in.

II. RELATED WORK

P. Kalaivani et.al [2] applied three supervised machine learning algorithm such as K-Nearest Neighbour algorithm, SVM and Naive Bayes.

These algorithms are used for the compression of classification of sentiment based on movie reviews which has thousand of positive and negative reviews. Here the SVM approach is performed well compared to K-NN and Naive Bayes. The accuracy of SVM approach is reached up to 80%.

Anurag et al [3] has presented a new method called combined approach. This approach is used to categorise the reviews based on the sentiment that are present in reviews. Further they were able to improve the anticipated classification results with the help of two classifier combination rules. They also explained and introduced a method of handling the slag words as well as smiley, which overall generate a best classification of sentiment with higher accuracy. Jayashri Khairnar et al. have concentrated on the SVM that performs task of sentiment classification and also calculates the sentiment classification accuracy. With the support of small datasets in two dimensional feature space the concept of SVM is explained. It categorizes the datasets in high dimensional space by using kernel methods.

Anuja P et.al [4] introduced a method in order to perform sentiment analysis on datasets such as BNL, Apple and Bank. The size of size datasets ranges from 200 to 4000. In this experiment they have splited the datasets as 3/4th are used for training and 1/4th for testing. They used two machine learning algorithms they are Decision tree and Multinomial Naïve Bayes classifier. The pre-processing is done by extracting the features of tweets. They used a framework called Apache Spark that gave quick accurate result and also scalable. The decision tree gave precision, recall, f1 score and accuracy of 100%.

Dey et.al [5] they used movie and hotel reviews datasets. They worked on two classifiers that are Naive Bayesian and K-NN. The main aim of this classifier to find which classifier gives a best result for both datasets. Their finial result proves that Naive Bayesian gave best accuracy in case of movie review datasets. In hotel review datasets both Naïve Bayesian and K-NN gave approximate results. Finally they concluded that Naive Bayesian is best classifier for classification of movie review.

Bac Le et.al [6] they construct a version with the intention to examine twitter sentiment the usage of machine learning strategies like powerful set of features and enhances unigram, bigram accuracy and object orientated features. The twitter categories are achieved primarily based on algorithms including Naive Bayes (NB) and Support Vector Machine (SVM), these algorithm accuracies are examined based on the calculations of precision, recall and f1-score and also these algorithms gave similar accuracy.

Sayali P et.al [7] they created datasets based on twitter API and gathered all twitter datasets based totally on game called blue whale. The primary intention of this venture to carry out evaluation on sentiment tweets. Here they have used machine learning techniques that including Naïve Bayesian (NB), Support Vector Machines (SVM), Maximum Entropy and Ensemble classifier. SVM and Naive Bayes are applied by means of using MATLAB which is constructed with the help of functions and maximum Entropy classifier are applied with the help of software program called MaxEntsoftware. via evaluating the result Naive Bayesian has higher precision and low recall and accuracy this is 89% and other classifiers has equal accuracy of 90%.

Neethu M S et.al [8] accomplished sentiment analysis by means of considering about tweeter datasets primarily based on specific area the use of exclusive machine learning method. They focused on the issues which prompted at some stage like detection of emotional keywords from keywords that are multiple and also difficulties in misspelling and slag of phrases handling. So characteristic vector is created and accuracies are calculated based on SVM, Naive Bayesian, and Maximum Entropy and Ensemble classifiers.

Su et.al [9] had explained about implicit extraction of features as a product feature that doesn't occur explicitly but from the surrounding opinion word can be inferred. They also introduced a method for mutual reinforcement method for simultaneous clustering of product features and opinion of words. In subsequent work different methodologies are introduced in order to identify the association of opinion of words and aspect terms, thus implicit aspects also inferred from aspect word opinion of term for mapping.

Muhammad Abbas et.al [10] introduced a method in order to solve some problems with Multinomial Naive Bayes (MNB) which addresses both systematic problems and this problem arises when text is not actually case generated according to multinomial model. A MNB classifier is a type of Naive Bayes classifier that is used as a baseline for classification of text. They have used movie reviews as a dataset. This review contains a notice in the form of text numerical score. The datasets are further pre-processed with the help of natural language processing. After pre-processing the next step is bag of words. Here the bag of words are used to calculate the number of tokens present in each document and returns a matrix with sequential property. After this step they used Multinomial Naive Bayes classifier that gave 91% of accuracy.

Mohammad et al [11] described 2 SVM edge cutting classifier that is Tweets and SMS(message-level-task) which recognizes the sentiments in message and another recognizes a feelings of term inside data(term-level-task). In both the performance of this task is good. A variety of surface-form, semantic and opinion highlights with opinion word hash tags, and twitter data with their emoticons have been implemented. In initial stage the gain of 5 F-score points that provides by dictionary based features over all others.

III. PROPOSED WORK

The proposed method for opinion mining on product reviews has several levels of processing of the opinionated reviews by means of various techniques; these techniques are shown in fig 1

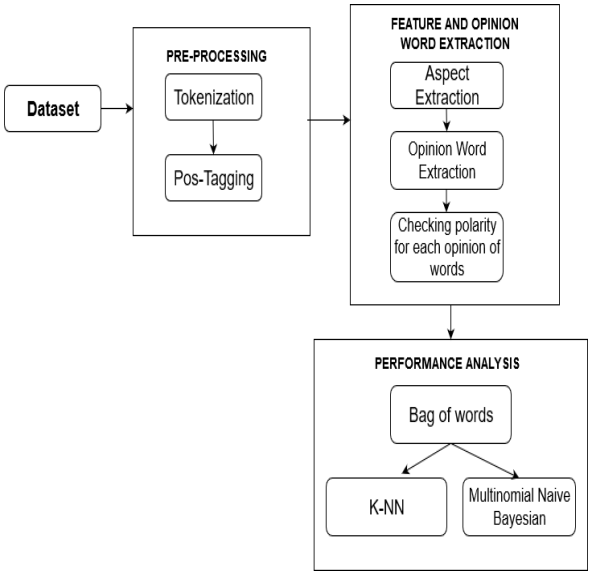

Fig 1 Conceptual Model for feature based opinion mining

Conceptual Model for feature based opinion mining Extracting and mining the product reviews are done using the following steps:

3.1 Collection of Datasets: The product review datasets are extracted from the Kaggle sites. Kaggle is a website which consists of Scripts, assortment of public datasets, as well as a dedicated forum for conversation and collaboration between data scientists working on a given dataset, which is extensively used for research purpose. The data set will be in the form of CSV file format which contains the various mobile product reviews. This project focus on Samsung mobile product reviews with different features.

Product N	Brand Nar	Price	Rating	Reviews	Review Votes
"CLEAR CL	Samsung	199.99	5	I feel so LUCKY to have found this used (phone	1
"CLEAR CL	Samsung	199.99	4	nice phone, nice up grade from my pantach re	0
"CLEAR CL	Samsung	199.99	5	Very pleased	0
"CLEAR CL	Samsung	199.99	4	It works good but it goes slow sometimes but i	0
"CLEAR CL	Samsung	199.99	4	Great phone to replace my lost phone. The onl	0
"CLEAR CL	Samsung	199.99	1	I already had a phone with problems... I know	1
"CLEAR CL	Samsung	199.99	2	The charging port was loose. I got that soldere	0
"CLEAR CL	Samsung	199.99	2	Phone looks good but wouldn't stay charged, I	0
"CLEAR CL	Samsung	199.99	5	I originally was using the Samsung S2 Galaxy fc	0
"CLEAR CL	Samsung	199.99	3	It's battery life is great. It's very responsive to t	0
"CLEAR CL	Samsung	199.99	3	My fiance had this phone previously, but caus	0
"CLEAR CL	Samsung	199.99	5	This is a great product it came after two days o	0

Fig 2 Snapshot of collected reviews from Kaggle site

3.2 Pre-processing: After collection of data we are pre-processing the data in order to change a raw data into justifiable format. Some of the data pre-processing steps include tokenization, POS Tagging

3.2.1 Tokenization: Tokenization is to split a sentence into words, expression, letters etc. these units are called as tokens. Here the reviews that are saved in a CSV file are given as an input for tokenization. And these tokens must be saved in a CSV file for further process.

sl.no	review
1	Innovative', 'product', '.', 'I', 'am', 'able', 'to', 'charge', 'my', 'spare', 'battery', 'of', 'mo
2	'The', 'unit', 'is', 'working', 'perfectly', 'and', 'due', 'to', 'the', 'slow', 'charge', 'the', 'bat
3	'very', 'bad', '.', 'battery', 'heats', 'up', 'very', 'fast'
4	'Works', 'well', '.', 'Slight', 'slow', 'charging', 'hence', 'good', 'for', 'battery', 'life', '.'
5	'Of', 'all', 'the', 'three', 'batteries', 'I', 'have', 'used', 'this', 'is', 'the', 'best', '.'
6	'I', 'tried', 'many', 'battery', 'but', 'this', 'battery', 'is', 'just', 'awsome', '...', '.please', 'g
7	Excellent', 'genuine', 'product', '.', 'Packaging', 'was', 'very', 'good', 'and', 'the', 'produ
8	Product', 'is', 'genuine', 'and', 'effective', '.', 'Original', 'battery', 'of', 'Jio', 'Fi'
9	'Working', 'fine', 'and', 'perfectly', 'fit', 'size', 'battery'
10	'battery', 'backup', 'is', 'to', 'good', '.', 'is', 'working', 'original', 'battery', 'and', 'it', 'tak

Fig 3 Snapshot of tokenized reviews

3.2.2 Pos-Tagging: Pos tagging is assigning the parts of speech for each review in CSV file. This Pos-Tagging is used for extracting the aspects and corresponding opinion of words from a review which is stored in a CSV file.

sl.no	review
1	[('Innovative', 'JJ'), ('product', 'NN'), ('.', '.'), ('I', 'PRP'), ('am', 'VBP'), ('able', 'JJ'), ('to',
2	[('The', 'DT'), ('unit', 'NN'), ('is', 'VBZ'), ('working', 'VBG'), ('perfectly', 'RB'), ('and', 'CC
3	[('very', 'RB'), ('bad', 'JJ'), ('.', '.'), ('battery', 'NN'), ('heats', 'VBZ'), ('up', 'RP'), ('very', '
4	[('Works', 'NNP'), ('well', 'RB'), ('.', '.'), ('Slight', 'NNP'), ('slow', 'JJ'), ('charging', 'VBG'),
5	[('Of', 'IN'), ('all', 'PDT'), ('the', 'DT'), ('three', 'CD'), ('batteries', 'NNS'), ('I', 'PRP'), ('hav
6	[('I', 'PRP'), ('tried', 'VBD'), ('many', 'JJ'), ('battery', 'NN'), ('but', 'CC'), ('this', 'DT'), ('ba
7	[('Excellent', 'JJ'), ('genuine', 'JJ'), ('product', 'NN'), ('.', '.'), ('Packaging', 'NN'), ('was', '
8	[('Product', 'NN'), ('is', 'VBZ'), ('genuine', 'JJ'), ('and', 'CC'), ('effective', 'JJ'), ('.', '.'), ('O
9	[('Working', 'VBG'), ('fine', 'JJ'), ('and', 'CC'), ('perfectly', 'RB'), ('fit', 'JJ'), ('size', 'NN'), (
10	[('battery', 'NN'), ('backup', 'NN'), ('is', 'VBZ'), ('to', 'TO'), ('good', 'JJ'), ('.', '.'), ('is', 'VB

Fig 4 Snapshot of Pos tagging reviews

3.3 Extracting the features:

3.3.1 Extracting all aspect terms: Aspect extraction is also known as feature extraction. In this project all the nouns are considered as aspects. Later these aspects are stored in a CSV file. To do this procedure this SpaCy package is used in python. SpaCy is an open source library used for natural language processing.

3.3.2 Extracting all opinion of words: In opinion mining all the adjectives is considered as opinion. Consider an example "battery is good". In this example good is an adjective so good is taken as opinion of words.

3.3.3 Finding polarity for extracted opinion of words: Polarity in sentiment analysis are used to identify whether the extracted opinion of words are positive or negative. If the opinion of word is greater than or equal to zero then the opinion of word is considered as positive. If the opinion of word is less than zero then it the opinion of word is considered as negative. This process can be done with the support of text blob. TextBlob is a python library that offers a simple API to access its methods to perform various NLP tasks.

	text	sentiment
0	fit	0.400000
1	Slow	-0.300000
2	easy	0.433333
3	handy	0.600000
4	good	0.700000
5	bad	-0.700000
6	best	1.000000
7	awesome	1.000000
8	Excellent	1.000000
9	genuine	0.400000
10	effective	0.600000
11	original	0.375000

Fig 5 Calculating the polarity of each word to identify whether the given opinion of words are positive or negative

sl.no	opinion	aspect	sentiment	review
1	Spare	Charge ba	positive	Innovative product. I am able to charge my spare bat
2	Slow	Unit charg	negative	The unit is working perfectly and due to the slow cha
3	easy hand	battery	positive	very bad . battery heats up very fast
4	bad	battery	negative	Works well. Slight slow charging hence good for batt
5	good	Battery	positive	Of all the three batteries I have used this is the best.
6	best	batteries	positive	I tried many battery but this battery is just awsome ..
7	awesome	battery	positive	Excellent genuine product. Packaging was very good
8	Excellent	Packing p	positive	Product is genuine and effective. Original battery of J
9	genuine e	Battery pr	positive	Working fine and perfectly fit size battery
10	fine fit	Battery siz	positive	battery backup is to good. is working original battery

Fig 6 Extracted aspects, opinion and sentiment

3.4 Performance Analysis:

3.4.1 Bag of words: A BoW is used to extract a feature from text that is used for machine learning algorithm modelling. It uses a unigram model of features based on the number of occurrence of each features. This further can be used for the classification.

label	good_bov	great_bow	excellent_	nice_bow	awsom_b	best_bow	fast_bow	easy_bow
0 positive	0	0	0	0	0	0	0	0
1 negative	0	0	0	0	0	0	0	0
2 positive	1	0	0	0	0	0	0	1
3 negative	0	0	0	0	0	0	0	0
4 positive	1	0	0	0	0	0	0	0
5 positive	0	0	0	0	0	0	0	0
6 positive	0	0	0	0	0	1	0	0
7 positive	1	0	1	0	0	0	0	0
8 positive	0	0	0	0	0	0	0	0
9 positive	0	0	0	0	0	0	0	0

Fig 7 Bag of words for opinion of words

3.4.2 Multinomial Naive Bayesian: MNB makes an assumption that the document consists of a bag of words and takes the frequency of each word and information into account. MNB uses supervised machine learning algorithm that makes use of probability and focuses on text classification. For conditional probability this approach makes use of multinomial distribution [10]. By the use of multinomial distributions, this algorithm is used for text cases by way of changing to the normal form which are computed as integer values. The probability calculation are given below

$P(c|d) \propto (c) \Pi P(tk|c)$ .....(1) $1 \leq k \leq nd$

The word tk conditional probability is $P(tk|c)$ which appears within the document that is having class c. The similar probability of tk at class c is $P(tk|c)$. (c) is the prior probability of a document that appears in class c. The determination of probability used for comparison the posterior class end result that is received. Similarly the class which has the most important posterior probability is taken for prediction of end result. The system for prior probability is proven in formula 2:

$P(c) = Nc/N$..... (2)

The category sum of c is Nc and sum of all category is N . The formula for likelihood probability are given as

$(tk|c) = Ttc/\Sigma Tct'' t' \in V$..... (3)

The occurrence of word t is Ttc in the document that has class c. The total number of occurrences of all words in class c is $\Sigma t' \in V\ Tct''$

3.4.3 K-Nearest Neighbour: K-NN is also called as lazy learning algorithm. It is a non-parametric method for regression or classification. The class membership is an ouput for classification. The object is classified based on majority of votes from its neighbor, with the help of class representation object are assigned for each query from K-NN the majority labels are from training set.

The Nearest Neighbour rule (NN) is a form of K-NN rule when k=1. Given training and sample set that is unlabelled. The distance between unlabelled samples and training set are computed. The distance with smallest value corresponds to the sample in training set closest to unlabelled samples. Based on this the unlabelled samples are categorized based on its NN classification.

K-NN is simple algorithm for analyses, implementation, and it's a powerful tool for disposal of sentiment analysis. K-NN is powerful because it does not make any assumption about data, other than distance measure can be calculated consistently between two instances, this process is called as non-linear or non-parametric as it does not make any assumption of functional form.

IV. RESULTS AND DISCUSSION

4.1 Confusion Matrix: A confusion matrix provides the performance description of classification model on test datasets for which true values are already known [13].

4.1.1 Precision: Precision is used to measure the exactness of a classifier. The precision can be calculated by using formula.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Table-I: Precision for positive and negative sentiment

Model	Positive	Negative
Multinomial Naive Bayesian	0.92	0.5
K-Nearest Neighbour	0.98	0.71

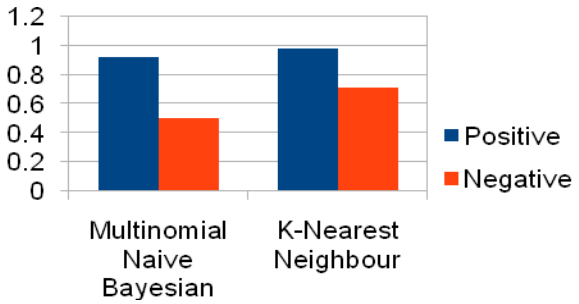

Fig 8 Precision for MNB and K-NN

Fig 8 shows the precision for predicted values and actual value. The graph shows precision for positive and sentiment.

4.1.2 Recall: Recall is used to measure the completeness or sensitivity of the classifier. The recall can be calculated by using a formula.

$$\text{Recall} = \frac{TP}{TP+FN}$$

Table-II: Recall for positive and negative sentiment

Model	Positive	Negative
Multinomial Naive Bayesian	0.92	1.0
K-Nearest Neighbour	0.96	0.83

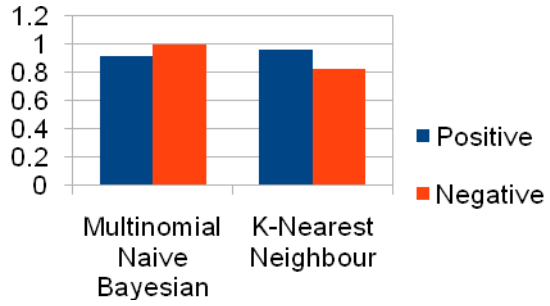

Fig 9 Recall for MNB and K-NN

4.1.3 F1 Score: F1 scores are an average, weighted precision and recall. It can be calculated by using a formula

$$\text{F1 Score} = \frac{2*(Precision*Recall)}{Precision+Recall}$$

Table-III: F1 score for positive and negative sentiment

Model	Positive	Negative
Multinomial Naive Bayesian	0.96	0.5
K-Nearest Neighbour	0.97	0.76

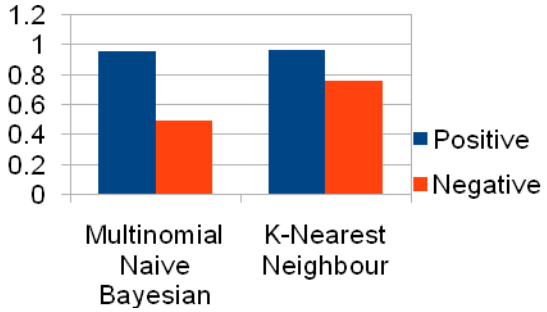

Fig 10 F1 Score for MNB and K-NN

4.1.4 Accuracy: Accuracy is a ratio of number of correct predictions to the total number of prediction. Fig 11 shows accuracy of MNB and K-NN. By comparing this 2 algorithm, K-NN gave best accuracy of 94.9%.An accuracy can be calculated by using the formula:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Table-III: Accuracy

Model	Accuracy
Multinomial Naive Bayesian	93.2
K-Nearest Neighbour	94.9

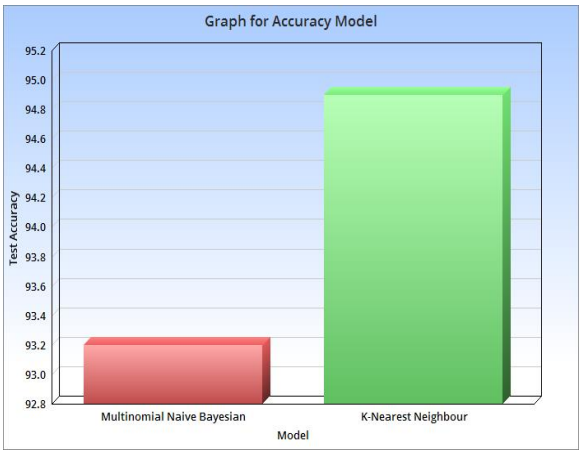

Fig 11 Accuracy for MNB and K-NN

V. CONCLUSION AND FUTURE WORK

In this era, the technologies are growing rapidly. Due to these technologies, many people write their feedback of the product which they purchased in online. The reviews are not a small amount; it's a huge bulk of data. By reading one or two reviews people cannot conclude that the product is good or bad. In this research we focus on Samsung mobile product review that is collected from Kaggle site. Here we have used 200 mobile reviews for experiment. After collecting data next step is to perform tokenization and POS tagging for each review.

Later we extract aspects, opinions and perform polarity on the extracted opinion to find whether the opinion is positive or negative. Further we used machine learning algorithms for performance analysis. In this experiment we have considered two machine learning algorithm i.e. Multinomial Naive Bayesian and K-Nearest Neighbour algorithm. The KNN gave best accuracy compared to MNB.

Here we have considered only one mobile product review future work can be done by using many product review and can also apply some more machine learning and deep learning algorithm for performance analysis.

REFERENCES

1. B. Liu, Web data mining: Exploring hyperlinks, contents, and usage data, 2nd ed. Berlin: Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2006.
2. P.Kalaivani, Dr.K.L.Shunmuganathan, "Sentiment classification of movie review by supervise machine learning approach", Indian Journal of Computer Science and Engineering (IJCSE) Vol. 4 No.4 Aug-Sep 2013.

3. Anurag Mulkalwar, Kavita Kelkar Sentiment "Analysis on Movie Reviews Based on Combined Approach", International Journal of Science and Research, Volume 3 Issue 7, July 2014.
4. Anuja Prakash Jain and Padma Dandannavar "Application of machine learning techniques to sentiment analysis" 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) pages 628-632 article{Jain2016ApplicationOM}.
5. Dey, Lopamudra&Chakraborty, Sanjay & Biswas, Anuraag& Bose, Beepa& Tiwari, Sweta. (2016). "Sentiment Analysis of Review Datasets Using Naïve Bayes and K-NN Classifier". International Journal of Information Engineering and Electronic Business. 8. 54-62. 10.5815/ijieeb.2016.04.07.
6. Le B., Nguyen H. (2015) "Twitter Sentiment Analysis Using Machine Learning Techniques". In: Le Thi H., Nguyen N., Do T. (eds) Advanced Computational Methods for Knowledge Engineering. Advances in Intelligent Systems and Computing, vol 358. Springer, Cham.
7. Sayali P. Nazare, Prasad S. Nar, Akshay S. Phate, Prof.Dr. D. R. Ingle"Sentiment Analysis in Twitter"International Research Journal of Engineering and Technology (IRJET)Volume: 05, Jan-2018.
8. M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, 2013, pp. 1-5.
9. Qi su , Xinying Xu,Honglei Guo, Xian Wu, 2008 Hidden Sentiment Association in Chinese Web opinion Mining, Processing of international conference on world wide web, pages 959-968.
10. Muhammad Abbas, Kamran Ali Memon,Abdul Aleem Jamali, Saleemullah Memon, Anees Ahmed, "Multinomial Naive Bayes Classification Model for Sentiment Analysis", IJCSNS International Journal of Computer Science and Network Security, VOL.19 No.3, March 2019 .
11. S. M. Mohammad, S. Kiritchenko, and X. Zhu, ―NRC-Canada: Building the state-of-the-art in Sentiment Analysis of Tweets, Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013), 2013.
12. Dey, Lopamudra&Chakraborty, Sanjay & Biswas, Anuraag& Bose, Beepa& Tiwari, Sweta. (2016). "Sentiment Analysis of Review Datasets Using Naïve Bayes and K-NN Classifier". International Journal of Information Engineering and Electronic Business. 8. 54-62. 10.5815/ijieeb.2016.04.07.
13. https://www.geeksforgeeks.org/confusion matrix-machine-learning/

AUTHORS PROFILE

Shiji Abraham, BE in Computer Science and Engineering. Pursuing MTech in Computer Science and Engineering at NMAM Institute of Technology, Nitte, Karkala, India. Her area of interests includes Natural Language Processing, Machine Learning and Image Processing. She has published her journal paper on Image forgery detection using DCT and Quantization Matrix in IJEAT journal. Her research work carried out in Aspect Based Opinion Mining on Product Reviews.

Minu P Abraham, working as Assistant Professor in the Department of CSE, NMAM Institute of Technology, Nitte. She has completed her B.Tech in Computer Science and Engineering from Kerala University and M.Tech in Computer Science and Engineering from Visvesvaraya Technological University. She has 9 years of teaching experience and 2 years of research experience. She has published 4 research papers in reputed journals. She has also presented 4 conference papers in national and international conferences. She is a lifetime member of Indian Society for Technical Education. Her fields of interests are Data mining, Text mining, Natural language processing and Artificial Intelligence.

Dr. Udaya Kumar Reddy K.R, In 1993 he completed his Diploma in Computer Science and Engineering from Siddaganga Polytechnic, Tumkur, Bangalore University. In 1998, he completed his Bachelor of Engineering in Computer Science and Engineering from Golden Valley Institute of Technology, K.G.F, Bangalore University, India. In 2002, he completed his Master of Engineering in Computer Science and Engineering from University Visvesvaraya College of Engineering, Bangalore, India. In 2011, he completed his Ph.D. in Computer Science and Engineering from National Institute of Technology, Trichy. He has over 20 years of teaching and research experience and currently he is Professor and Head of Computer Science and Engineering department at NMAMIT, Nitte. He has published many research papers in reputed journals.His fields of interests are Algorithmic graph theory, Design of Compilers, Theory of computation, and Data Mining.

Anisha P Rodrigues, working as Assistant Professor in the Department of Computer Science and Engineering at NMAM Institute of Technology, Nitte, Karkala, India. Her research interests include Natural Language Processing, Machine Learning, and Big Data Analytics. She is a member of ISTE.