Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Customer Services with the Help of Sentiment Analysis on Twitter Data

Manish Rana[1], Rahul Khokale[2], Mahendra S. Makesar[3], Darasing Ramrao Solanke[4], Monika Rajesh Walde[3]

[1]R Department of Computer Engineering, St. John College of Engineering & Management (SJCEM), Palghar, India
[2]G H Raisoni University (GHRU)Saikheda, Maharashtra –India
[3]Nagpur Institute of Technology (NIT), Nagpur, Maharashtra -India
[4]Sant Gadge baba Amravati University,Amravati(SGBAU)-India
Email: dr.manish_rana@yahoo.co.in

The recent advancements in computer technology have provided an immense boost to the marketing world across the globe. This paper aims to provide better customer services with the help of Sentiment Analysis on Twitter data. The goal is to use this data for developing a Sentiment Analysis model to help companies understand their product better and hence make necessary changes in their products based on customer requirements. Sentiment Analysis model helps in improving customer satisfaction and future plans of a company. This model uses various Machine Learning Techniques to predict the sentiment of a tweet.

Keywords: Machine Learning, Tweet, Sentiment, Naïve Bayes, Polarity, Tokenization.

1. Introduction

Over the years, the Customer – Provider services have been into limelight because of the inconsistency in the services. Customer requirements have also changed thanks to the enhancement in the fields such as Artificial intelligence and Machine Learning. Twitter can be used to make this service more efficient as it is widespread around the globe. Building a model which analyses user requirements based on their reviews can cause a tremendous advancement in the business field. As this task is manual and automated unlike the earlier system which was based on in person calling and hence was inconsistent. To make changes in a product as per customer review and provide the necessary services to a customer are one of the major goals of any firm. This model will perform Sentiment Analysis on the data extracted from Twitter and provide analysis based on customer needs. Social Media is an important tool. Twitter is used globally and due to its availability and features, users post their reviews, express their opinion regarding products. The sentiment can be can be categorized into

positive, negative and neutral comments. Sentiment analysis will help companies to understand their product better and with our model they can also analyze the product enhancement part.

We used Twitter dataset for implementing our Sentiment analysis model. The dataset basically comprises of textual data on which the Model would predict the sentiment of a Tweet.

Following are the steps in development of this Model:

1) Gathering Twitter Data

2) Preprocessing/Cleaning the Data

3) Transforming Data.

4) Model Building.

5) Training the model and validation.

Advantages of our Sentiment Analysis include "Product Enhancement" as the model can be used to work on Product Enhancement based on the sentiment which is achieved by checking the most occurred word in a positive or negative tweet and this word could be a good feature or a Bug based on the sentiment, Customer Service Enhancement is another advantage as the process is automated and accurate, Social Media Monitoring is another advantage as companies can monitor their product and make necessary changes to their product, marketing and sales strategy. Another advantage is that they can get analysis of almost any product due to Twitter's global reach and millions of users. Some common difficulties which were encountered while developing this model includes problems such as, "Domain dependence" wherein many features are not helpful due to the domain difference. "Handling Comparisons" as it is difficult to differentiate in the way the statement is conveyed. "Order Dependence" like how the form in which the statement is conveyed. "Sarcasm Detection" as sometimes the user's tone can be sarcastic but the model might evaluate it in a different way. Existing Sentiment analysis models can be further improved with more Semantic and Common Knowledge. The accuracy of model can also be worked upon with more efficient classification and analysis.

2. LITERATURE REVIEW

I. Sentiment Analysis: Uses in Business [1] Author : Martin Atwebembire Year : 2015		
Features	Future Scope	Gaps
• Qualitative content analysis was usedduring Data analysis. • The data was categorized to reflect the themes in the literature review as well as provide answers to theresearch questions. • Provides a better insight in corporateSouth Africa	• .Increasing the sample data. • To develop a sentiment analysis business value framework. • The no of responses can be increased from 13 to 50 corporates so that more valuable insights could be gained.	• Data gathering was limited due to time constraints. • Limited time meant fewer responses and therefore fewer data to analyze. • Choice of research method was changed from interviews to questionnaire due to inability to schedule interviews.
II. Business Intelligence analytics using Sentiment Analysis - a survey[2]		

Authors : Prakash P. Rokade, Aruna D. Kumari Year : 2019		
Features	Future Scope	Gaps
• It discusses various approaches which could be used for sentiment analysis. • Like SVM, Naïve Bayes, Decision Trees, Rule based approach and various Lexicon based approaches. • Compares accuracy and efficiency of the approaches used in sentiment analysis.	• Complex interactions can be resolved to improve Decision tree performance. • Language Translation can be taken into consideration as Tweets are diverse in the languages used.	• Probabilistic Classifier (Bayesian Network) were used even when features were correlated which could have resulted in low accuracy. • Domain dependence.
III. An interpretation of Sentiment analysis for enrichment of business Intelligence[3]Authors : Nidhi Kushwaha, O.P.Vyas, Bharat Singh Year : 2016		
Features	Future Scope	Gaps
• It uses the NLP (Natural language Processing) approach. • Some outliers were considered whilefeature extraction. • Improvised lesk algorithms were alsoused.	• The Visualizations of the results can be further improved to a more attractive way. • To provide better framework for this model in form of a website or Android application.	• The Model uses Lesk algorithm which comprises of certain disadvantages like the dictionary definitions are often very short andjust do not have enough words for this algorithm to work well.
IV. Market Research: The role of Sentiment analysis [4] Authors : Meena Rambocas, Joao Gama Year : 2013		
Features	Future Scope	Gaps
• It takes into account the fact that howsentiment analysis as an alternative technique is capable of triangulating qualitative and quantitative methods.	• Better Opinion Mining software can be integrated in the near future. • The reviews data on products could be skewed hence may give different results under different classification models.	• One of the major gaps relates to the nature of classification. • As there is usually a limit of the number of groups and subgroups that can be extracted. • Also, Text based data are usually contex specific and Domain Dependent which means it is valid in specific placesat specific times.
• It triangulates qualitative and quantitative methods through innovative real time data collectionand analysis.	• So, the Dataset should be taken into context based on the type of classification we want to perform.	
V. Sentiment Research on Twitter Data [5] Authors : K A.Brahmananda Reddy, D.N.Vasundhara, P. Subhash Year : 2014		
Features	Future Scope	Gaps
• It discusses about Naïve Bayes used intheir classifier model. • Provides insights about various data Pre-processing on twitter data before training it. • Like removal of Elongated words	• Instead of going for Bag of Words wecan apply Word2Vec. • As it works for small as well as largedataset and is easy to scale.	• In Pre-processing part the stemming or lemmatization part is not done. • As this feature can be added in order to make the model more accurate.

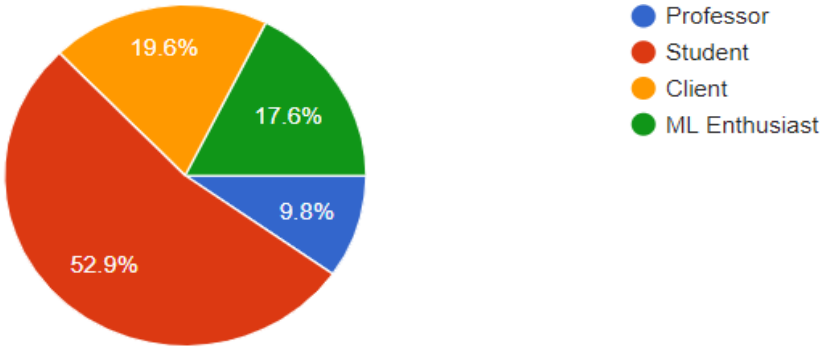
number, Emoticons, Last Token, Punctuation marks.		
VI. Sentiment Analysis on Twitter Data using Machine Learning algorithms[6] Authors : S.Siddharth , R.Darshini , Dr.M.Sujithra Year : 2019		
Features	Future Scope	Gaps
• Uses Pre-Processing, Feature Extraction, Feature selection and Classification Operations. • Combines them with Machine Learning algorithms like Naïve Bayes and Neural Networks. • The data was categorized to reflect the themes in the literature review as well as provide answers to the research questions.	• Existing Sentiment analysis models can be Further improved with more Semantic Knowledge • To use cost effective opinion mining software. • To Improve Accuracy of the model with more efficient classification and analysis.	• Irregularity in availability of opinion Mining Software as it is very expensive and only affordable to Big Organizations. • Domain dependence.
VII. Sentiment analysis of twitter data: A survey of techniques[7] Authors : Vishal A. Kharde , S.S.Sonawane Year : 2016		
Features	Future Scope	Gaps
• It uses ML based Supervised learning approach, Maximum Entropy. • Provides insights on Lexicon Based Approaches which comprises of Dictionary based and Corpus Based Approach.	• The Lexicon based approach gave an accuracy of 74.00 % which can be further improved by focusing on Tokenized Document Collection. • Dataset could be made to include a greater variety of uncommon sentences.	• Order Dependence. • Entity Recognition. • Domain Dependence.
VIII. Twitter Sentiment Analysis[8] Authors : Aliza Sarlan, Chayanit Nadam, Shuib Basri. Year : 2014		
Features	Future Scope	Gaps
• It explains the technique of extracting polarities through sentiment analysis. • The two main approaches were discussed for extracting sentiment automatically which are the lexicon based approach and Machine Learning Based approach.	• To develop a Web Application and to make user performance better. • To demonstrate the results in a more convenient way instead of the HTML page representation.	• The Program will be categorized into positive or negative which is represented in form of a chart. • Which means due to the limitation of Django it can work only on Linux or LAMP which also specifies that it cannot be realized.
IX. Sentiment analysis using product review data[9] Authors : Xing Fang*, Justin Zhan Year : 2015		
Features	Future Scope	Gaps
• Three classification models were selected for categorization. • Provides insights on Naïve Bayesian, Random Forest, and Support Vector Machine.	• After Performing various techniques like Slicing, Data Preprocessing this model can be applied on Twitter Data to gain more valuable insights. • Has onboard computation which is slow as compared to cloud and hence generalizing performance across all	• Data Gathering could be a problem as this paper used Review Data for Amazon products. • Work on scalability can be performed.

	system.	
X. Sentiment analysis as a tool of Business Analytics in contemporary organizations [10] Authors : Leszek Ziora Year : 2018		
Features	Future Scope	Gaps
• It uses a very computationally effective model. • It reviews various approaches which can be used for sentiment analysis. • Provides insights on SVM, Naïve Bayes, Decision Trees, Rule based approach and various Lexicon based approaches.	• Language Translation can be taken into consideration as Tweets are diverse in the languages used. • So, the Dataset should be taken into context based on the type of classification we want to perform.	• Handling comparisons as this model describes a bag of words approach which means it doesn't take into account the relation towards the two parts of a sentence.

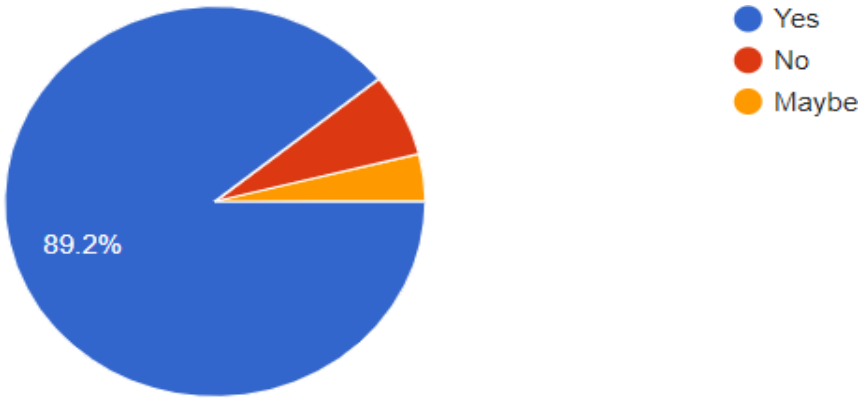
3. Results Of Survey Conducted

A Survey on Twitter Sentiment analysis was conducted. Total 102 responses were collected out of which more than 50% were students who have given their valuable responses. We reached out to other professionals like Professors, ML Enthusiasts, Clients. In order to understand their thoughts, demand and expectations.

Q1. Profession

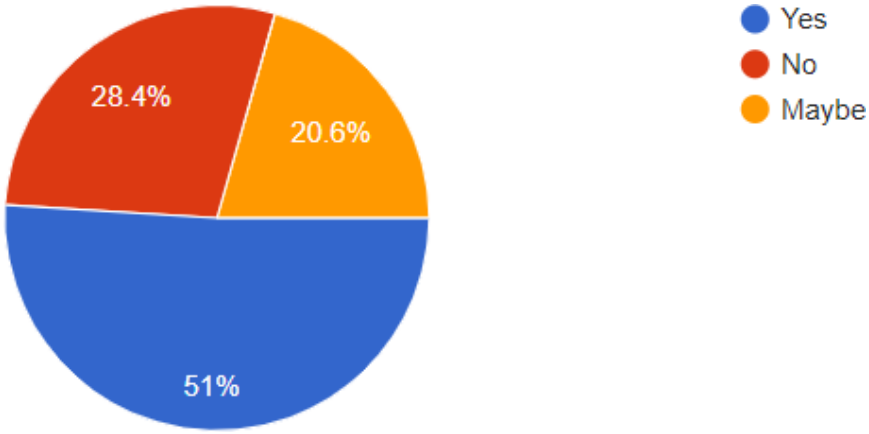


Q2. Do you think Sentiment Analysis is necessary in today's Social Media Marketing World?



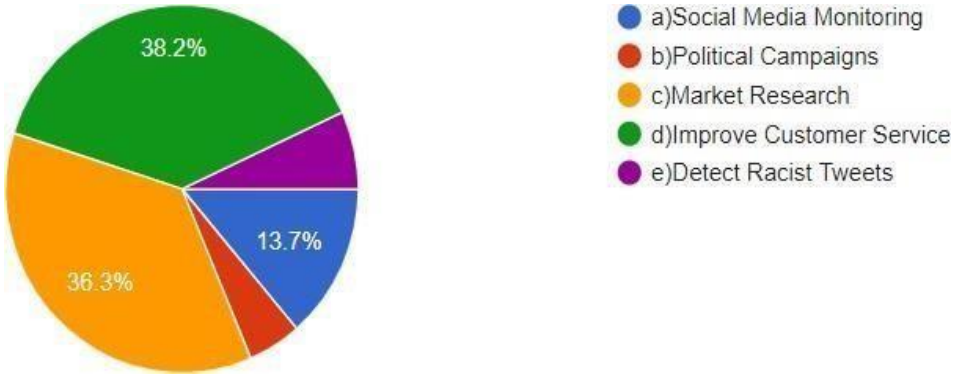
Majority of the entities think that Sentiment Analysis is necessary but there are very few who believe that such analysis is not necessary.

Q3. Do you think an Automated Agent is sophisticated enough to understand and decipher human emotions, in a sentence?



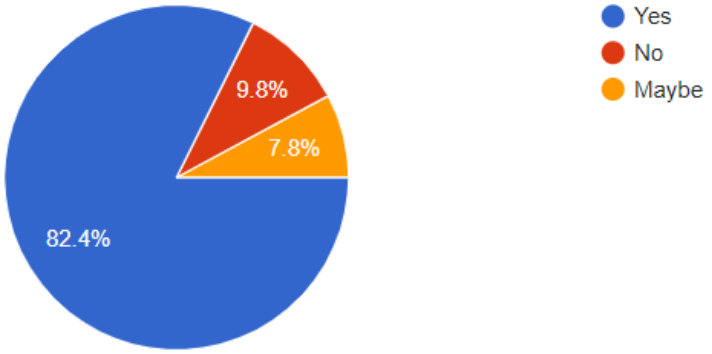
Around 51% respondents think that an Automated Agent can understand Human emotions or sentiments around 28.4% respondent conceive that such type of an Automated Agent cannot decipher the human emotions.

Q4. For what purposes can Sentiment Analysis be used?



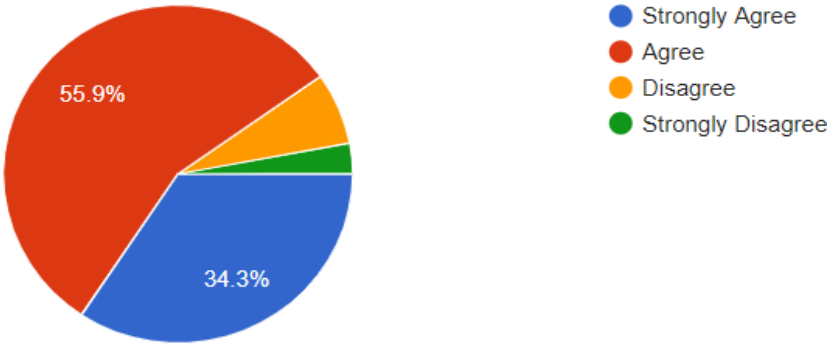
Around 38.2% of respondents approved that Sentiment Analysis can be used to Improve Customer Service that is understanding the customer reactions towards a product Whereas around 36.3% of respondents believe that Market Research can be conducted using Sentiment Analysis while a small majority of people convey that it can be used for Social Media Monitoring.

Q5. Can Twitter Sentiment Analysis Improve the Understanding of Product Reviews?



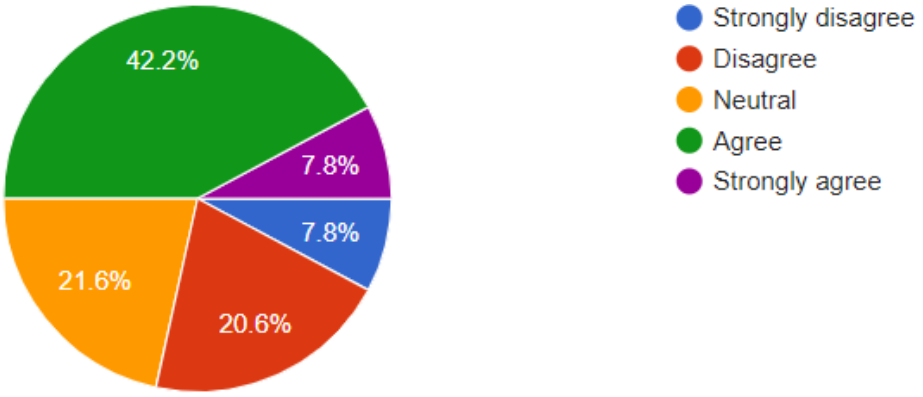
Twitter Sentiment Analysis will help in understanding the Product reviews believe majority of the respondents according to the Survey. As basically getting a proper feedback and proper understanding of the product will definitelyhelp the clients to make wise business decisions.

Q6. Can Sentiment Analysis replace the In - Person Customer Service provided by Brands for Product Improvement?



The results for this question showed that the respondents are divided on the automation of the Customer Care Services as there are still good companies like apple who provide better customer services in - person.

Q7. Sentiment Analysis Models are the future of Business Planning.



The majority of the respondents think that Sentiment Analysis is the future for business planning because in order to get customer reviews social media is the best place as many of the users share their thought at such platforms which can be advantageous for a company to understand the customer's sentiment.

4. SWOC Model

5. IMPLEMENTATION

It involves 5 steps –

1) Gathering Data
2) Pre-processing/Cleaning the information
3) Transforming Data
4) Model Building
5) Training the model and validation

1. Gathering Data:

Sentiment Analysis models can predict the long run adjustments required in any business model which suggests that it has to be optimal moreover as accurate. We are using Twitter dataset to implement our Sentiment analysis model. The dataset basically comprises of textual data on which the Model would predict the sentiment of a Tweet. For data gathering we used Twitter APIs using the Tweepy Library. The keys included Consumer key, Consumer Secret,

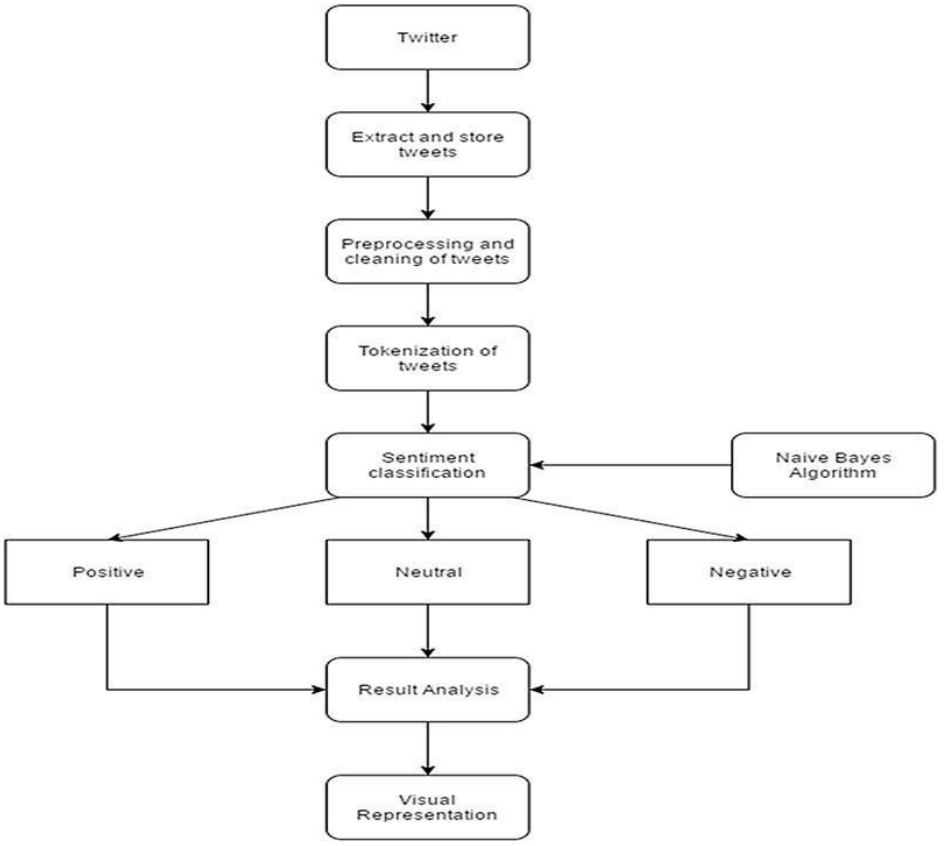
Access Token, Access Token Secret. Initially, we created an Authentication Object, Set the Access Token and Access Token Secret and created an API Object while passing in the auth information.

2. Pre-processing/Cleaning the information:

The steps involved within the Pre- processing part include:

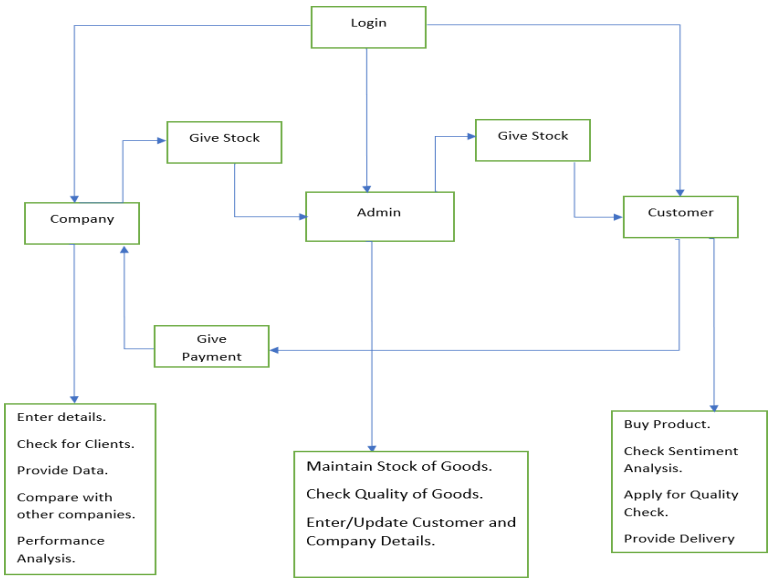
- Removal of Stop words
- Removal of Punctuations
- Removal of Hyperlinks
- Removal of Repeating Characters
- Removal of Numeric Numbers
- Removal of Retweets

3. Transforming Data:



During the Preprocessing of data, we accessed the tokenization of a tweet, Applied Stemming & Lemmatization to give it a meaningful nature. After this process, we spilt the data into training and testing. Finally, the data was transformed using TF-IDF Vectorizer.

4. Model Building:



1. Plot of Polarity:

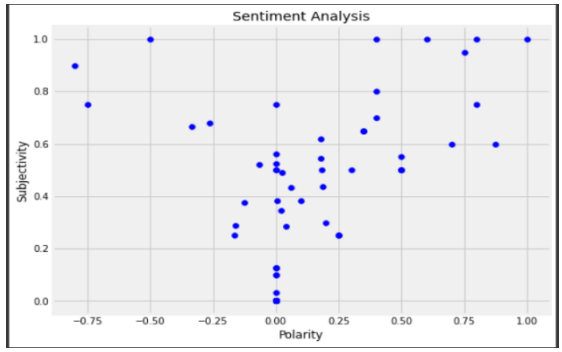
After transforming the data, the next step is Model Building. Here, we used the Bernoulli Naïve Bayes by using the training and testing data. The reason behind using the Bernoulli Naïve Bayes is due to its flexibility, reliability, accuracy and ability to work with real time data.

5. Training the model and Validation:

Once the features are generated the foremost relevant ones are selected, each of the tweets from the training dataset is going to be expressed in terms of the attributes. During the training process the presence of every attribute is checked for every of the classes (positive,negative and neutral).

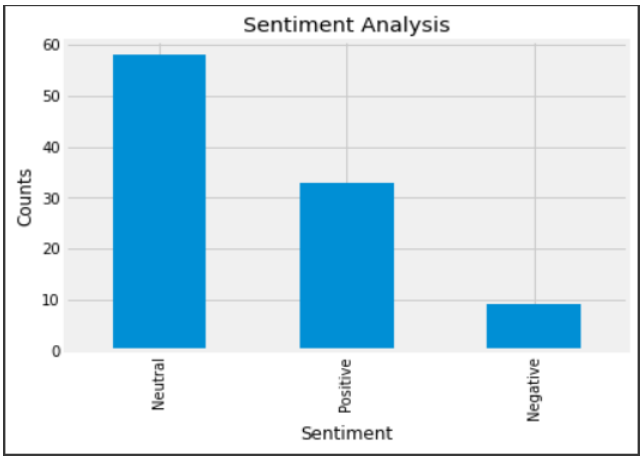
6. Result & Discussion

1. Plot of Polarity:



The figure 6.1 shows the polarity plot in which most of the instances lie around 0.

2. Plot of Sentiment



The figure 6.2 shows the Sentiment plot in which the tweets are classified into Positive, Negative and Neutral. Here, 58% of the tweets are Neutral, 33% are Positive and 9% are Negative.

7. Conclusion

The Sentiment analysis model categorized the gathered Twitter dataset into positive, negative or neutral sentiments so that a proper study of customer reviews can be done. This will help companies to study about their product with the help of public tweets and can provide them information so that necessary changes or updates in a product can be done and gain insights in a very short time. The irony and sarcasm type of tweets were categorized into neutral sentiment so that the outliers are removed. At times it is difficult to categorize a statement in which two products are compared so it becomes quintessential to categorize such statement which helps to understand a customer more accurately. Techniques of Machine Learning approach were utilized to obtain the desirable output. This will help the companies to study about their product with the help of customer opinion and can provide them the information to ensure that necessary changes or updates in a product can be done and insights can be gained in a short span of time.

References

1. R. Suchdev, P. Kotkar, and R. Ravindran, "Sentiment Analysis on Twitter Data Using Machine Learning Algorithms in Python," in Proc. International Conference on Advances in Computing Applications (ICACA-18), NIT Uttarakhand, Feb. 2018.
2. A. Rasool, R. Tao, and T. Nadeeb, "Twitter Sentiment Analysis: A Case Study for Brand Apparels," IOP: Journal of Physics, 2018.
3. S. Swamy, "Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach," Computer Engineering, VES Institute of Technology, University of Mumbai,

Mumbai, India. J. Y. Wang, M. Zhang, Z. X. Gao, and B. Adhikari, "Smart storage technologies applied to fresh foods: A review," *Crit. Rev. Food Sci. Nutr.*, vol. 58, pp. 2689–2699, 2018, doi: 10.1080/10408398.2017.1323722.

4. N. Kushwaha, B. Singh, and O. P. Singh, "An interpretation of sentiment analysis for enrichment of Business Intelligence," in Proc. 36th IEEE Tenon 2016, Singapore, 2016.
5. K. S. and E. Hovy, "Determining The Sentiment Of Opinions," in Proc. International Conference on Computational Linguistics (COLING'04), 2004.
6. B. Liu, "The science of detecting fake reviews," 2014. [Online]. Available: [URL]. [Accessed: Dec. 7, 2014.
7. N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. 2008 International Conference on Web Search and Data Mining (WSDM '08), New York, NY, USA: ACM, 2008, pp. 219–230.
8. A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in Proc. 21st International Conference on World Wide Web (WWW '12), New York, NY, USA: ACM, 2012, pp. 191–200.
9. R. Kosala and H. Blockeel, "Web mining research: A survey," SIGKDD Explor. Newsl., vol. 2, no. 1, pp. 1–15, Jun. 2000.
10. J. Han, Data Mining: Concepts and Techniques. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.
11. R. Cooley, B. Mobasher, and J. Srivastava, "Web mining: information and pattern discovery on the world wide web," in Proc. Ninth IEEE International Conference on Tools with Artificial Intelligence, Nov. 1997, pp. 558–567.
12. W. Jicheng, H. Yuan, W. Gangshan, and Z. Fuyan, "Web mining: knowledge discovery on the web," in Proc. IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC '99), vol. 2, 1999, pp. 137–141.