10.48047/jocaaa.2023.31.01.37

From Tweets to Trends: Utilizing Supervised Learning Algorithms for Social Media-Driven Business Analytics

Lohit Banakar

Lecturer, E&CE Department, Government Polytechnic, Sorab, Karnataka, India
lohitbanakar@gmail.com, https://orcid.org/0000-0002-1710-2156

Abstract

The opinions of people's feedback, emotions, attitudes, evaluations, which are provided in natural language on various social media websites. The domain that analyzes these opinions is called sentiment analysis. It is one of the latest research domains in analyzing natural language and is extensively used in text mining, web mining, and data mining. The significance of this study not only constrained to computer science, but also to social studies and business studies. The development of social media with reviews, posts, feedbacks, discussions in forums, blogs has correspondingly enhanced the relevance of sentiment analysis. A large amount of data is collected and analyzed in a digital nature for the first time in history. Nowadays most of the business and social fields are affected by the opinions of customers. Hence sentiment analysis for these fields is required. Most of the things in the world are evaluated based on the how public see those things and how the public accepted those things. Due to this, by observing the opinion of the public we are making decisions. The same thing can even be applied to the companies also. Companies can also alter their business strategies by using analytical data provided by the opinions of the public in social media. Furthermore, this analytical data is also helpful for companies to make decisions. In order to determine trends and patterns, the application uses machine learning classifiers. In this paper, a popular mobile brand is used to analyze the customer tweets from the Twitter website. For analyzing the tweets, different classifier methods are used. Initially, customer tweets are collected with the help of Twitter API(Application Peripheral Interface). These tweets are given as an input for the binary tree classifier to identify the positive or negative polarity. The paper outcomes are very advantageous for companies to enhance their business and technical aspects using Twitter data. The paper also suggests futuristic opportunities in this developing domain.

Keywords- Sentiment analysis, Supervised learning, Twitter data, Business analytics
Machine learning classifiers

10.48047/jocaaa.2023.31.01.37

1. INTRODUCTION

Companies can strategize their business processes through the redefined nature with the help of social media. Business comprehension such as customer sentiments and content analysis is based on a huge amount of unstructured data such as reviews, user posts, comments, tweets, blogs, and forum discussions. Most companies use social media networking online service for promotion and marketing tool. In this paper static twitter raw data is used for the analysis. Twitter data contain subjective information in the form of texts that are used to analyze and make decisions based on sentiments for a particular issue.

Without investing in costly public studies and interviews the subjective information of twitter posts are sufficient for industries to collect feedback about their products and services from a business perspective. Data miners face difficulties during analysis and extracting unstructured data. When a huge amount of data is involved it is very difficult for humans to analyze the data, as compared to limited data where customers can freely identify trends and patterns in data. Such a challenge of analyzing the huge amount of data is achieved with the aides of computational approaches.

The focus of social media research is gaining more importance currently due to the business strategies of the industry that is based on sentiment analysis. The market has a growth rate of 17.2% with a rise to 8.79 billion dollars for sentiment analysis [8]. There is much need for industries to have analytics of social media data for futuristic analytics, and to modify their business strategies. The companies can financially gain profits using the opinions of the public in developing their services, company reputation, or observing their brand.

The value of business strategies cannot be neglected, particularly by customer support and marketing and other units of a company. The paper mainly analyses the use of customer posts in various business operations. The paper also demonstrates the analysis of text in inspecting the public opinions towards a particular mobile brand that can be used for making decisions using hidden knowledge (example business and customer insights). Text analytics of Twitter data has limited academic literature surrounding text data, hence this paper tries to share with this advancing domain by equipping a practical knowledge on customers' posts in twitter and also mining the data.

10.48047/jocaaa.2023.31.01.37

2. LITERATURE REVIEW

To understand customers' opinions concerning about the services or products "Text analysis of data in social media for business analytics" application is used by companies. Alrence Santiago Halibas, Abubucker Samsudeen Shaffi and Mohamed Abdul Kader Varusai Mohamed [1] concluded that "Having a text analysis of customer feedback and reviews allows effective quality management. With sentiment analysis, companies can now strategically reposition their businesses according to customers' sentiments".

According to V. A. Kharde and S. S. Sonawane[2], Sentiment Analysis is "a process that automates the mining of attitude, opinions, views, and emotions from text, speech, tweets, and database sources through Natural Language Processing (NLP)".

The benefits and importance of customer sentiment analysis for business analytics were pointed out by L. Ziora [3][4]. The majority of the purposes of sentiment analysis are for decision support and business emendation.

According to N. Yussupova, M. Boyko, and D. Bogdanova[5], the use of qualitative and quantitative methods in customer satisfaction research was proven to be inefficient as compared to a decision support system using sentiment analysis with data mining. The main factor that determines decision making is that customer satisfaction the study pointed out. Thus, it is significant here to manage the quality of the product by knowing customers' sentiments.

According to S. K. Markham, M. Kowolenko, and T. L. Michaelis[6] revealed that product decisions can be made by using unstructured text analytics. To create trust in the information for evaluative results the study indicates the concern of data from various unstructured data, including newspaper articles.

According to O. Muller, I. Junglas, S. Debortoli, and J. Von Brocke[7] Florida State University to obtain the incoming service requests stream across business units using text analytics by the Information Technology Services (ITS). The university offers an effective IT service administration by its success in gathering insight information from the texts.

According to P. Khobragade and V. Jethani[8], the customers are satisfied when tweets are positive. Therefore, customers' data provides huge benefits for industries/companies by knowing the tweets they provided.

10.48047/jocaaa.2023.31.01.37

3. METHODOLOGY

A ststic raw data collected from twitter website related to mobile brand in notepad is used for the analysis,screenshot is attached bellow figure 1.


Sample Pos and Neg - Copy (2) - Notepad

File Edit Format View Help

Nice Product, good performance. Amazing display, recommended for Samsung galaxy on 8. Battery
best phone Battery
Nice phone Battery
Good product..... thanks Amazon Battery
Nice Product, good performance. Amazing display, recommended for Samsung galaxy on 8. Battery
Received phone in good condition. Very nice phone in the price range. Brand name really matters.
very good product in this price with brand Samsung. all feature are good but screen mirroring not available.
VERY EXCELLENT MOBILE IN AMAZON TEAM NICE MOBILE WORKING GOOD
I used this for 1 wieek.really i love this phone.the design is awesome.ringing vol is little low.
Good stylish appearance. The golden tinge is understated.
Excellent battery life. With the kind of usage listed above, it lasts for two days quite easily.
Call volume is good and alarm and reminder functions work smoothly.
I bought 2 samsung on8 because under 15K this is feature rich phone .
Good device no problems it has very high speed it has 3gb ram i am a satisfy with this product
Waste of cost
Horrible heating problem while playing games and using any app.
Slow functioning, hang while playing games
Heats too much at left side of the phone while using 3g/4g Internet.
Wrost camera and worst charger and
totally disappointed with the response
Its is not nice allways down i want to replace
Very bad experience about this product
Very bad product supplied.Donot purchase any samsung phone online.
Guys dont waste money on samsung their models are now idle go
ery disappointed with Amazon
PHONE RINGING IS VERY VERY LOW
There was some problem with the phone speaker and mic since the first day but as the problem was minor it was neglected.
Please avoid this model because its sound output quality is very bad.

Figure 1. Sample static raw data collected from twitter

The following procedure is used to accomplish business analytics.

A) Using Twitter API (Application Program Interface) required data of a particular mobile brand is collected.

B) The collected data from Twitter API is preprocessed such that it can be suited for the mining process. In this paper, the Twitter data already collected and the same has been processed for running the paper.

C) Using supervised machine learning approaches a classifier is built. In this paper, two different supervised machine learning approaches that are the random forest with multipolarity and naïve Bayes with multipolarity is used.

D) Using a huge amount of data collected from Twitter, training, and testing of the built classifier is done.

10.48047/jocaaa.2023.31.01.37

E) Using the data collected from the Twitter output of the different classifiers is computed.

Later business analytics of a mobile brand is calculated and the results of the classifiers are compared and a graph is plotted that shows the trend of positive, more positive, negative, more negative. Figure 2. Shows the detailed flow diagram of system. To preferred understand the present market and upgrade their brand companies can understand more about public opinions about their products and services. Based on the public opinions the companies can review their strategies in marketing and do the campaigns.

The important decision making in business analytics can be achieved using analytical data which is collected with a huge amount of unstructured data from the Twitter website. In making strategic decisions of the product, the application uses different machine learning approaches to find out outcomes and trends to gain further necessary actions. Random forest with multipolarity which is a decision tree based supervised learning approach is of prime interest in this paper. According to Alpaydin, who is scientist in the area of machine learning concluded that decision tree is a nonparametric way for classification and forecasting of data in an efficient manner[14]. He also stated that decision tree is a machine learning approach that uses a data structure in the form of hierarchical manner. The decision trees are mainly used in this paper, since they learn and behave instantaneously and very easily explainable. Apart from this decision trees proved to be an efficient analytical classification method with very high accurate results.

The experimental evidences proved that Naive Bayes' machine learning classifier method has 86% accuracy in comparison with decision tree machine learning classifier method that has 100% accuracy result. Due to decision tree's inheritable characteristics such as accuracy result, easy to perceive and easy to classify the data this machine learning approach is used in this paper. The system proposed in this paper consists of file system with big data technique in an efficient way.

The data from the twitter social website is processed with faster idea and can be used easily in order to achieve customers' sentiment application. Twitter website contains customers' posts which is unstructured data in the form of feelings on several issues namely products, films, political issues, books, and games and so on.

10.48047/jocaaa.2023.31.01.37

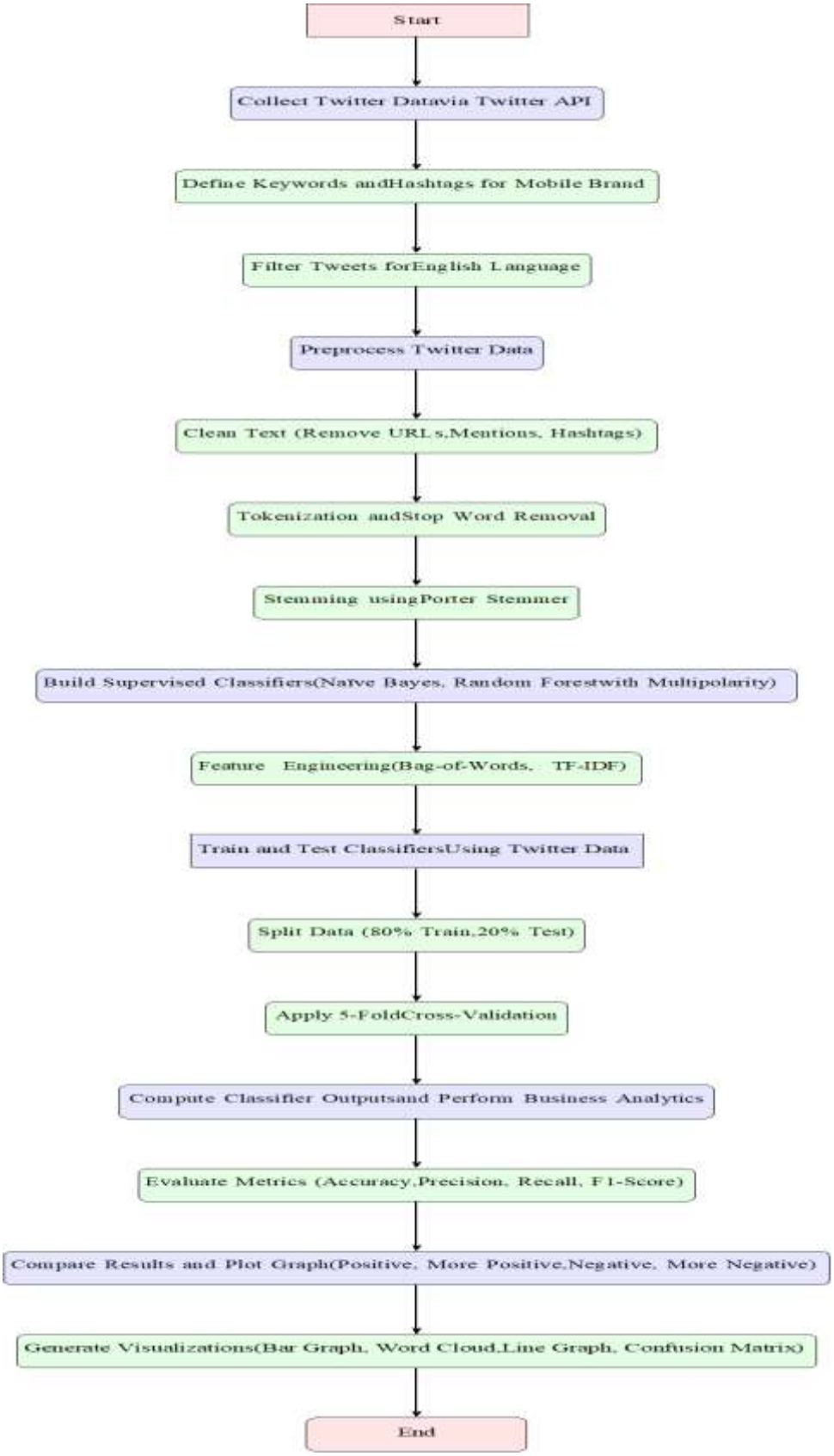

Figure 2. Flow diagram of existing system

10.48047/jocaaa.2023.31.01.37

These twitter posts from customers' are used to achieve sentiment analysis by analyzing the insight aspects. By processing the natural language data, symbols using various computational machine learning classifiers either who will win or lost is achieved. For classification approaches supervised learning method is used in order to find out the different ways the twitter posts are posted and to analyze the data.

The decision trees are based on binary classification method which is either of the two cases. Those cases are positive and negative with customers' sentiments in it. The primary advantage of binary classification shows positive outcomes with excellent view of issues related to product or negative outcomes with bad poor view of issues related to the same.

The proposed system is described as given bellow.

A) Browse the content – The text data in the natural language stored in the notepad and will be browsed through application.

B) Remove the stop words - Certain words in the natural language must be removed before it is processed.

Example of stop words are "the", "a", "an", "in".

Using "stopwords" algorithm presents in "NLTK(Natural Language Tool Kit) corpus" library, stopwords in the text data will be removed.

C) Remove the stemming words – Stemming words are words that are reduced to their base words or root words that are written in the natural language form.

Example of stemming words for the base word "like" is "likes", "liked", "likely".

Using "Porterstemmer " algorithm presents in NLTK(Natural Language Tool Kit) stem library, stemwords in the text data will be removed.

D) Calculate positive and negative words – With the assistance of the wordnet dictionary, the sentimental dictionary is created by uploading the positive and negative words.

E) Display word cloud - Word clouds are used to highlight critical data text points.

For displaying text data in which each word size indicates number of times the word has occurred and it's importance.

F) Classify the results - Based on the naive baye's and random forest classification algorithms results are classified into positve, more positive, negative and more negative.

10.48047/jocaaa.2023.31.01.37

G) Display the results in graph - The result of the text analysis is plotted in bar graph showing in different polarities in different colors.

4. APPROACH

Following two approaches are used.

A) Naive Baye's classifier

B) Random forest classifier

A) Naive Baye's classifier is a supervised learning algorithm used for classification. It is based on the Naive Baye's theorem. Naive Baye's based machine learning classifier algorithm is as follows –

i) Find the posterior probability for available labels of class.

ii) Calculate the likelihood probability with each attribute for each class.

iii) Mention these values in Baye's formula and find out subsequent probability.

iv) Based on the available input belongs to the maximum probability class, decide the class which has a maximum probability.

Naive Baye's theorem is based on the following probability equation,

$$P(\text{label}|\text{features}) = \frac{P(\text{label})*P(\text{features}|\text{label})}{P(\text{features})}$$

Where,

- P (label) is the posterior probability of a label or likelihood probability.
- P (features | label) is the posterior probability that a available feature set is being categorized as a label.
- P (features) is the posterior probability of a features.
- P(label | features) is the subsequent probability.

Some of the text classes chosen from the given dataset in notepad, represented in table 1.

DOC	TEXT CLASS	CLASS
1	best phone Battery	+
2	Waste of cost	-
3	Nice phone Battery	+
4	Very bad experience about this product	-
5	Good product thanks Amazon	+

Table .1 Text classes from dataset in notepad

10.48047/jocaaa.2023.31.01.37

From the above table 1. Total words belong to "+" class = 10.

< best, phone, Battery, Nice, phone, Battery, Good, product, thanks, Amazon, Battery >.

Total Words belongs to "-"class = 09.

< Waste, of, cost, Very, bad, experience, about, this, product >.

Total unique words are = 16.

<best, phone, Battery, Waste, of, cost, Nice, Very, bad, experience, about, this, product, Good, thanks, Amazon>

Convert the given document into feature sets, where the attributes are available words, and the number of times a word present indicated by the numbers as shown in the Table 2.

DOC	Best	Phone	Battery	Waste	Of	cost	Nice	Very	bad	Experience	About	this	Product	Good	Thanks	Amazon	CLASS
1	1	1	1														+
2				1	1	1											-
3		1	1				1										+
4								1	1	1	1	1	1				-
5													1	1	1	1	+

Table 2. Text classes represented in both +ve or -ve

Following Table 3. Shows the sentiments with only "+" class.

DOC	Best	Phone	Battery	waste	of	Cost	Nice	Very	bad	experience	about	this	Product	Good	Thanks	Amazon	CLASS
1	1	1	1														+
3		1	1				1										+
5													1	1	1	1	+

Table 3. Text classes with +ve outcomes

10.48047/jocaaa.2023.31.01.37

Computing the probabilities,

- p (+) = 3 / 5 = 0.6

Compute probability of each word belongs to "+" Class.

p (Best|+); p (Phone|+); p (Battery|+); p (waste|+); p (of|+); p (cost|+); p (Nice|+); p (Very|+); p (bad|+); p (experience|+); p (about|+); p (this|+); p (product|+); p (Good|+); p (Thanks|+); p (Amazon |+).

Using Laplace Smoothing equation we are calculating probability of all the words belongs to +ve class.

$$p(w_k/+)=\frac{nk+1}{2n+|Vocabulary|}$$

Where,

k the number of times word k occurs in these cases (+).

2n be the number of words in the (+) class: 10.

|Vocabulary| = number of unique words.

The probability of all the words belongs to +ve class are calculated as follows.

p (Best|+) = (1 + 1) / (10 + 16) = 0.0769

p (bad|+) = (0 + 1) / (10 + 16) = 0.038

p (Phone|+) = (2 + 1) / (10 + 16) = 0.115

p (experience|+)=(1 + 1)/(10 + 16) = 0.0769

p (Battery|+) = (2 + 1) / (10 + 16) = 0.115

p (about|+) = (1 + 1) / (10 + 16) = 0.0769

p (waste|+) = (0 + 1) / (10 + 16) = 0.038

p (this|+) = (1 + 1) / (10 + 16) = 0.0769

p (of|+) = (0 + 1) / (10 + 16) = 0.038

p (product|+) = (1 + 1) / (10 + 16) = 0.0769

p (cost|+) = (0 + 1) / (10 + 16) = 0.038

10.48047/jocaaa.2023.31.01.37

p (Good|+) = (1 + 1) / (10 + 16) = 0.0769

p (Nice|+) = (1 + 1) / (10 + 16) = 0.0769

p (Thanks|+) = (1+ 1) / (10 + 16) = 0.0769

p (Very|+) = (0 + 1) / (10 + 16) = 0.038

p (Amazon |+) = (1 + 1) / (10 + 16) = 0.0769

Following Table 4. Shows the sentiments with only "-" class.

DOC	Best	Phone	Battery	waste	Of	cost	Nice	Very	bad	experience	about	this	Product	Good	Thanks	Amazon	CLASS
2				1	1	1											-
4								1	1	1	1	1	1				-

Table 4. Text classes with -ve outcomes

Compute probability of each word belongs to "-" Class.

p (-) = 2 / 5 = 0.4

Compute probability of each word belongs to "-" Class.

p (Best|-); p (Phone|-); p (Battery|-); p (waste|-); p (of|-); p (cost|-); p (Nice|-); p (Very|-); p (bad|-); p (experience|-); p (about|-); p (this|-); p (product|-); p (Good|-); p (Thanks|-); p(Amazon |-).

Using Laplace Smoothing equation,

$$p(w_k/-)=\frac{nk+1}{2n+|Vocabulary|}$$

Where,

k the number of times word k occurs in these cases (-).

2n be the number of words in the (-) class: 9.

10.48047/jocaaa.2023.31.01.37

|Vocabulary| = number of unique words.

Calculating the probability of all the words belongs to -ve class

p (Best|-) = (0 + 1) / (9 + 16) = 0.04

p (bad|-) = (1 + 1) / (9 + 16) = 0.08

p (Phone|-) = (0 + 1) / (9 + 16) = 0.04

p (experience|-) = (1 + 1) / (9 + 16) = 0.08

p (Battery|-) = (0 + 1) / (9 + 16) = 0.04

p (about|-) = (1 + 1) / (9 + 16) = 0.08

p (waste|-) = (1 + 1) / (9 + 16) = 0.08

p (this|-) = (1 + 1) / (9 + 16) = 0.08

p (of|-) = (1 + 1) / (9 + 16) = 0.08

p (product|-) = (1 + 1) / (9 + 16) = 0.08

p (cost|-) = (1 + 1) / (9 + 16) = 0.08

p (Good|-) = (0 + 1) / (9 + 16) = 0.04

p (Nice|-) = (0 + 1) / (9 + 16) = 0.04

p (Thanks|-) = (0 + 1) / (9 + 16) = 0.04

p (Very|-) = (1 + 1) / (9 + 16) = 0.08

p (Amazon |-) = (0 + 1) / (9 + 16) = 0.04

To classify new sentence “Good battery, very nice product” whether it belongs to +ve or –ve. Following probabilities are used.

- p (Good battery, very nice product) = p (Good) * p (battery) * p (very) * p (nice) * p (product).
- p (Good battery, very nice product |+) = p (Good|+) * p (battery |+) * p (very |+) * p (nice |+) * p (product |+)
- p (Good battery, very nice product |-) = p (Good|-) * p (battery |-) * p (very |-) * p (nice |-) * p (product |-) .

10.48047/jocaaa.2023.31.01.37

The above process is called feature engineering.

For the Sentence "Good battery, very nice product". Probabilities of each words are calculated as follows.

p (Good) = 1/10 = 0.1, p (battery) = 1/10 = 0.1, p (very) = 0/10 = ?, p (nice) = 1/10 =0.1, p (product|+) = 1/10 = 0.1.

Probabilities of each words for the sentence "Good battery, very nice product" belongs to "+" class are calculated as follows.

p(Good battery, very nice product|+)=p(Good|+)*p (battery|+)*p (very|+)*p (nice|+)*p (product |+) = 0.0769 * 0.0769 * 0.0384 * 0.0769 * 0.0769 = **$1.348 * 10^{-6} = 0.000001348$**

Probabilities of each words for the sentence "Good battery, very nice product" belongs to "-" class are calculated as follows.

p(Good battery, very nice product |-)=p (Good|-)*p(battery|-)*p(very|-)*p(nice|-)*p (product|-) = 0.04 * 0.04 * 0.08 * 0.04 * 0.08 = **$4.06 \text{ X } 10^{-7} = 0.000000406$**

From the above solutions, since p(Good battery, very nice product|+) is **high** it belongs to therefore sentence belongs to "**+" class**.

B) Random forest classifier is a decision tree based algorithm where each node represents a feature (or attribute), each link represents a decision (rule) and each leaf represents an outcome.

Random forest based machine learning classifier algorithm is given bellow.

i) From the available dataset select some samples of data randomly.

ii) For each sample of data construct a decision tree.

iii) For each constructed decision tree predict the outcome.

iv) Accomplish a choice, based on each predicted outcome.

v) Get the predicted outcome with the majority choices as the last prediction.

5. RESULTS AND DISCUSSION

The use of the Naïve Bayes and Random Forest classifiers on the preprocessed Twitter data is highlighted in the results section. According to the experimental data referenced in the methodology, the Random Forest classifier obtained 100% accuracy, whereas the Naïve Bayes classifier obtained 86% accuracy. These accuracy numbers imply that tweet sentiment can be more accurately classified into the specified polarities using the Random

10.48047/jocaaa.2023.31.01.37

Forest classifier. By leveraging the Twitter API, a sizable dataset of tweets was gathered, preprocessed (stop words and stemmed terms were removed using NLTK tools), and then used to train the classifiers. After that, the classifiers calculated outputs that were utilized to determine business analytics for the mobile brand. Based on the classifier outputs, Figure 3, which is cited in the findings section, most likely displays a bar graph that illustrates the distribution of sentiment polarities (positive, more positive, negative, and more negative). In accordance with the methodology's explanation of presenting findings in a bar graph with various polarities, the graph would separate these categories using different colours. The quantity of tweets categorized into each sentiment category would serve as the basis for the graph's data points, which would show trends in public opinion on the mobile brand. If there were 1,000 tweets in the dataset, for instance, the graph might display the number or percentage of tweets in each polarity (40 percent positive, 20 percent more positive, 25 percent negative, and 15 percent more negative).

Figure .3. Screenshot of result

10.48047/jocaaa.2023.31.01.37

The results' implications for business analytics are further discussed. The Random Forest classifier is a useful tool for businesses to gauge how the public views their products because of its high accuracy (100%) which shows that it is dependable in detecting customer attitudes. With an accuracy of 86%, the Naïve Bayes classifier likewise does well but is less accurate, indicating that it might be applied as a backup technique or for datasets where computing efficiency is more important than perfect accuracy. Businesses can determine the opinions of their customers by looking at the sentiment distribution shown in Figure 3. For example, a high percentage of negative sentiments may highlight areas for improvement, while a preponderance of positive thoughts may indicate strong brand loyalty.

According to the article, these insights help businesses improve their business strategy, like modifying marketing campaigns or improving product features in response to unfavorable reviews (e.g., resolving concerns like "bad experience" or "waste of cost" noted in the dataset). In contrast to traditional surveys, using unstructured Twitter data provides a scalable and affordable way to collect customer feedback, which is consistent with the literature review's focus on the effectiveness of sentiment analysis for decision support. With social media being dynamic, the conversation also emphasizes the possibility of further research, such investigating other classifiers or incorporating real-time Twitter data to capture changing patterns.

A crucial tool for decision-making, the graph in Figure .3 gives stakeholders a clear visual cue. A rise in "more negative" sentiments, for example, can call for quick action, like a public relations campaign, but a balanced distribution might indicate that specific changes are needed. With a predicted market increase of $8.79 billion, the paper's results highlight the growing significance of social media analytics and its applicability to businesses looking to use consumer feedback to gain a competitive edge.

6.CONCLUSION

The findings and discussion show that supervised learning algorithms—Random Forest in particular—provide a reliable way to analyze Twitter data in order to facilitate business analytics. Figure 3's trends give businesses concrete data that help them match their strategy to the feelings of their customers. In addition to improving product development and marketing, this strategy establishes sentiment analysis as a fundamental component of contemporary business intelligence, with substantial room for further improvement.

10.48047/jocaaa.2023.31.01.37

7.REFERENCES

1. Halibas, A. S., Shaffi, A. S., & Mohamed, M. A. K. V. (n.d.). Application of text classification and clustering of Twitter data for business analytics. [Unpublished manuscript or source details unavailable].

2. Kharde, V. A., & Sonawane, S. S. (2016). Sentiment analysis of Twitter data: A survey of techniques. *International Journal of Computer Applications*, 139(11), 975–8887.

3. Ziora, L. (2016). The sentiment analysis as a tool of business analytics in contemporary organizations. *Studia Ekonomiczne*, 234–241.

4. Yaram, S. (2017). Machine learning algorithms for document clustering and fraud detection. In *Proceedings of the 2016 International Conference on Data Science and Engineering (ICDSE 2016)*.

5. Yussupova, N., Boyko, M., & Bogdanova, D. (2015). A decision support approach based on sentiment analysis combined with data mining for customer satisfaction research. *International Journal of Advanced Intelligent Systems, 1*(2).

6. Markham, S. K., Kowolenko, M., & Michaelis, T. L. (2015). Unstructured text analytics to support new product development decisions. *Research-Technology Management*, 58(2), 30–39.

7. Muller, O., Junglas, I., Debortoli, S., & Von Brocke, J. (2016). Using text analytics to derive customer service management benefits from unstructured data. *MIS Quarterly Executive*, 15(4), 64–73.

8. Khobragade, P., & Jethani, V. (2017). Sentiment analysis of movie review. *International Journal of Advanced Research in Computer Science*, 8(5), 1941–1948.

9. Marketsandmarkets.com. (2017). *Text analytics market by component (software, services), application (customer experience management, marketing management, governance, risk and compliance management), deployment model, organization size, industry vertical, region - Global forecast to 20*. https://www.marketsandmarkets.com/Market-Reports/text-analytics-market-121114266.html

10. Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

10.48047/jocaaa.2023.31.01.37

11. Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). SentiBench - A benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1), Article 1.

12. Moreno, A., & Redondo, T. (2016). Text analytics: The convergence of big data and artificial intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*, 3(6), 57.

13. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.

14. Alpaydin, E. (2004). *Introduction to machine learning*. MIT Press.

15. Khobragade, P., & Jethani, V. (2017). Sentiment analysis of movie review. *International Journal of Advanced Research in Computer Science*, 8(5), 1941–1948.

16. Kamruzzaman, S. M., Haider, F., & Hasan, A. R. (2010). Text classification using data mining. *Science*, 19.

17. Pang-Ning, T., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Pearson Education.