IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT
Impact Factor: 6.252

Volume 2, Issue 8, May 2022

Opinion Mining on Twitter Data using Machine Learning: A Case Study

S. Farjana Farvin[1] and Darshini Dhanapal[2]
Assistant Professor, Department of Computer Science and Engineering[1]
Student, Department of Computer Science and Engineering[2]
Anjalai Ammal Mahalingam Engineering College, Thiruvarur, India
farzana@aamec.edu.in[1] and darshinidhan2000@gmail.com[2]

Abstract: *Sentimental analysis, also referred to as opinion mining or emotion extraction is the classification of emotions within a textual data. This technique has been widely used over the years in order to determine the sentiments, emotions within a particular textual data. Twitter is a social media platform that has been mostly used by the people to express emotions for particular events. In this paper we have collected the tweets automatically in a csv file for a number of events, analyzed them using a number of machine learning algorithms like Multinomial Naïve Bayes, Logistic regression, TF-ID vectorizer, Decision tree classifier and Support vector machine(SVM) and then compared the results. And also the exploratory Data analysis for positive, negative and neutral results are displayed by Bar graph, heat map, scatter plot, word cloud and a line graph.*

Keywords: Twitter sentiment analysis; supervised approach; Multinomial naïve bayes; Logistic regression; Decision tree; SVM; TF-ID Vectorizer

I. INTRODUCTION

Sentimental analysis or emotion AI refers to the use of natural language processing, computational linguistics to systematically extract emotions, sentiments, opinions i.e. the subjective information in a piece of textual data. Opinion mining has found its use mainly within the market research allowing a business to understand the sentiment regarding their products, services [1]. It not only allows the monitoring of opinions but also the likes and dislikes of people in general.

Over the years Twitter has developed as one of the leading social media platforms and sentiments with respect to an event, incident in the form of tweets. By monitoring these tweets, a company or a political party can easily understand that how they are being perceived and what improvements could be made. In order to extract sentiment from tweets, sentiment analysis is used. The results from this can be used in many areas like analyzing and monitoring changes of sentiment with an event, sentiments regarding a particular brand or release of a particular product, analyzing public view of government policies etc.

It has significantly altered the appearences, not only by supplying all necessary information but also by allowing people to express themselves freely. As a result when it comes to market research, social media could be considered one of the greatest tools to monitor peoples's thoughts. The ability to predict sentiment within textual data has improved because to sentiments with a high degree of accuracy thanks to these algorithms .Machine learning enables the acquisition of new skills without the need for explicit training or programming. Sentiment analysis models can be used to predict not only the sentiment of a piece of text, but also other subjective information.

For forcasting emotions and attitudes, several machine learning techniques could be utilized. I employed methods like Multinomial naïve bayes, Decision tree, Logistic regression, SVM, TF-ID Vectorizer in this paper. This is a case study that includes a variety of recent occurences such as Indian muslims attack, EkMokoKejriwalNe, and RCBvsPBKS. Tweets about such events were gathered, fed into algorithms and results are compared.

II. LITERATURE REVIEW

Hearst [3] and Kessler et al [4] were the first to classify text based on sentiments or emotions in 1992 and 1997, respectively. The lexical method and the machine learning technique are the two basic approaches for classification of a piece of text . The capacity to extract characteristics and capture context has sparked interest in machine learning approaches [5]. These methods are typically used to determine the sentiment of a whole document, rather than just a few sentences.

Because it not always possible to detect sentiment based on a single word in sentiment analysis, Pederson presented and proved the effectiveness of n-gram extraction in 2001 [6]. Other methods required merely selecting a subset of the words recognized by a part-of-speech(POS) recognizer. Later on, a step by step classification strategy was created, which began with the removal of objective statements [1].

Many studies have been based on tweets gathered from twitter throughout the years. Barbosa and feng evaluated twitter data in 2010 by evaluating the polarity of tweets based on symbols, retweets, emoticons, and even the grammar of the tweets. The sentiments of movie reviews were initially analysed using the naïve bayes method [7]. Tong and koller were the first to apply the support vector machine to further increase prediction accuracy.

III. SENTIMENT ANALYSIS

With the rise of social media platforms, the number of people expressing their thoughts and ideas has increased. Labelling these sentiments can be very valuable for people who want to use these opinions to enhance their products, services, and so on. [1] Sentiment Analysis is the process of analyzing and finding sentiment and opinion within it. Unsupervised and supervised methodologies, as well as machine learning approaches, can be used to determine emotions.

Unsupervised Approach

It uses an unlabeled dataset and built –in libraries such as TextBlob to predict if a piece of text is positive, negative, or neutral.

Supervised Approach

It uses a labelled dataset and machine learning techniques such as Multinomial naïve bayes, SVM, TF-ID Vectorizer, and others.

Multinomial Naive Bayes – This is a supervised method that uses a probabilistic technique to categorize text into one of the two categories: positive or negative [9] . This algorithm assesses the probability of each word in the dataset before classifying tweets or text into certain groups. The bayesrule is used in this algorithm.

$$P(x|y) = \frac{[P(y|x)\ P(x)]}{P(y)}$$

Where, x,y = events

P(x|y) = Probability of x given y is true

P(y|x) = Probability of y given x is true

P(x),P(y) = Independent probabilities of x and y

Steps involved:

a. Data separation into training and test sets.
b. Creating a vocabulary from the words in the training sets.
c. Matching the content of the tweet with the vocabulary
d. Creating a feature vector.
e. Training the classifier using the feature vector i.e.,training the model.
f. Testing the model with the test set.

SVM - Support Vector Machine is a supervised learning technique that has been used to solve problems in classification and regression. SVM classifies the data dispersed in the n-dimensional space by determining a hyperplane. The categorization is based on mathematical functions known as kernels, which are then utilized to find a hyperplane. On opposite sides of the hyperplane, two different classes exist, and so this plane might be considered a decision boundary that could aid in the straightforward classification of the data points given.

The equation of the hyperplane is as follows:

$$w.x - b = 0$$

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 8, May 2022

IJARSCT
Impact Factor: 6.252

Where w = weight vector

x = Input vector

b = bias

The steps involved are :

a. Collecting the practice and the test sets
b. Data vectorization.
c. Developing a training model SVM Model.
d. Putting the model to the test on the set.

TF-ID Vectorizer - Term Frequency and Inverse Document Frequency is a text vectorizer that converts text into a vector that can be used .Term Frequency (TF) and Document Frequency(DF) are combined in this formula.

The term frequency refers to the number of times a term in a document .The frequency of a term in a document reflects its importance . Term frequency depicts each text in the data as a matrix with the number of documents in the rows and the number of distinct terms in the columns.

The amount of papers that include a specific term is known as document frequency. The frequency of documents reveals how common a term is. The weight of a term is determined by the inverse document frequency(IDF), which seeks to minimize the weight of a term if the term's occurences are dispersed throughout all the documents. The following can be used to calculate IDF:

$$idf_i = \log(n/df_i)$$

$$w_{i,j} = tf_{i,j} * idf_i$$

IV. METHODOLOGY

Sentiment analysis using machine learning algorithms entails a number of phases, the majority of which involve the processing of twitter data. The steps are as follows:

- Data collection
- Data preprocessing
- Parts of speech tagging
- Sentimental Analysis using Machine Learning Algorithms

4.1 Data Collection

This stage entails pulling data from twitter in the form of tweets, which necessitates the creation of a twitter account. Apart from that, twitter permission is required to gather tweets. We extracted tweets and saved them in .csv files after acquiring permission.

4.2 Data preprocessing

Once the data has been collected in .csv files, it is critical to preprocesses the data and remove any extraneous information. There are several pre-processing procedures involved, including the following:

1. Tokenization: URLs, Hashtags, and at mentions are removed, and string is broken down into a list of tokens. It is mostly used to tally the number of times words appear.
2. N-grams Extraction: This is done by grouping related words into phrases known as n-grams. This is done in order to increase sentiment analysis quality.
3. Stemming: Reading is replaced by read, and words are replaced by their stems or roots.
4. Stop words removal: It entails removing prepositions and articles that occur frequently but have no bearing on the text's overall sentiment..

4.3 Part-of-speech Tagging

This is the process of tagging each word in terms of its parts of speech.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 8, May 2022

4.4 Sentiment analysis using Machine Learning Algorithm

Machine learning algorithms are used to a structured dataset in this method. We compared the results of Multinomial naïve bayes, SVM, Logistic regression, Decision tree classifier, and TF-IF Vectorizer in this study.

V. CASES

5.1 Indian Muslims Attack

Indian Muslims attack were the important events that occurred, last year, a number of tweets were recorded with this hashtags. we collected around 15000 tweets in the raw form, fed them to an in-built library and then fed them to Machine learning Algorithms like Multinomial Naive Bayes, SVM, Logistic regression, Decision tree, TF-ID Vectorizer to obtain the results more accurately.

```
#LogisticRegression(C=1, penalty='l1', solver='liblinear')

model.fit(X_test, y_test)
predictions_train = model.predict(X_train)
predictions_test = model.predict(X_test)
print('Train Accuracy:', accuracy_score(y_train, predictions_train)*100)
print('Test Accuracy:', accuracy_score(y_test, predictions_test)*100)

Train Accuracy: 79.8890429958391
Test Accuracy: 81.7174515235457
```

Fig1. Indian Muslims attack accuracy using Logistic Regression.

```
[ ] from sklearn.naive_bayes import MultinomialNB
    model = MultinomialNB()
    model.fit(X_train, y_train)
    predictions_train1 = model.predict(X_train)
    predictions_test1 = model.predict(X_test)
    print('Train Accuracy:', accuracy_score(y_train, predictions_train1)*1
    print('Test Accuracy:', accuracy_score(y_test, predictions_test1)*100)

    Train Accuracy: 79.16088765603328
    Test Accuracy: 77.00831024930747
```

Fig2. Indian Muslims attack accuracy using MultinomialNB.

```
[ ] vectorizer = TfidfVectorizer(min_df=2)
    train_term = vectorizer.fit_transform(x_train)
    test_term = vectorizer.transform(x_test)
    #vectorizer.get_feature_names()[:20]

    #X_train
    model1 = MultinomialNB()
    model1.fit(train_term, y_train)
    predictions_train2 = model1.predict(train_term)
    predictions_test2 = model1.predict(test_term)
    print('Train Accuracy:', accuracy_score(y_train, predictions_train2)*100)
    print('Test Accuracy:', accuracy_score(y_test, predictions_test2)*100)

    Train Accuracy: 88.10679611650485
    Test Accuracy: 81.57894736842105
```

Fig3. Indian Muslims attack accuracy using TF-ID Vectorizer.

```
[ ] from sklearn.tree import DecisionTreeClassifier
    classifier = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
    model3 = classifier.fit(X_train, y_train)
    predictions_train3 = model3.predict(X_train)
    predictions_test3 = model3.predict(X_test)
    print('Train Accuracy:', accuracy_score(y_train, predictions_train3)*100)
    print('Test Accuracy:', accuracy_score(y_test, predictions_test3)*100)

    Train Accuracy: 91.8862690707351
    Test Accuracy: 77.42382271468145
```

Fig 4. Indian Muslims attack accuracy using Decision tree classifier.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 8, May 2022

```
[ ] from sklearn.svm import SVC
    svc_ml = SVC(kernel = 'linear', random_state = 0)
    model4 = svc_ml.fit(X_train, y_train)
    predictions_train4 = model4.predict(X_train)
    predictions_test4 = model4.predict(X_test)
    print('Train Accuracy:', accuracy_score(y_train, predictions_train4)*100)
    print('Test Accuracy:', accuracy_score(y_test, predictions_test4)*100)

    Train Accuracy: 81.62274618585298
    Test Accuracy: 80.47091412742382
```

Fig5. Indian Muslims attack accuracy using Support Vector Machine.

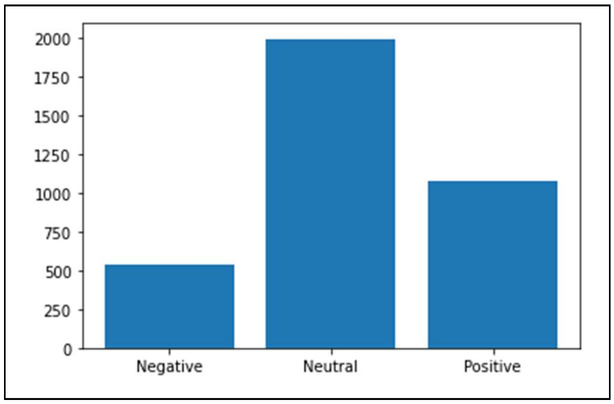

Fig6. Indian Muslims attack EDA by Bar Graph.

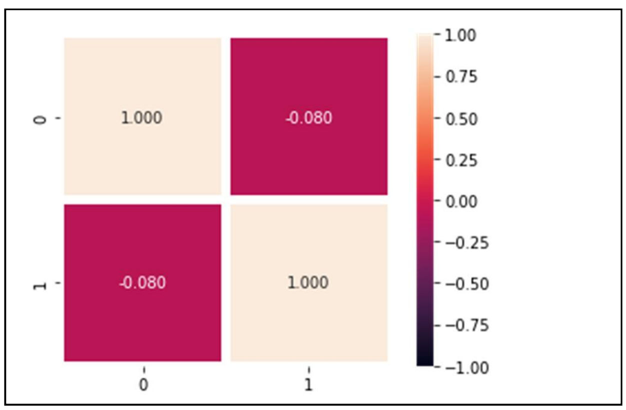

Fig7. Indian Muslims attack EDA by Heat Map.

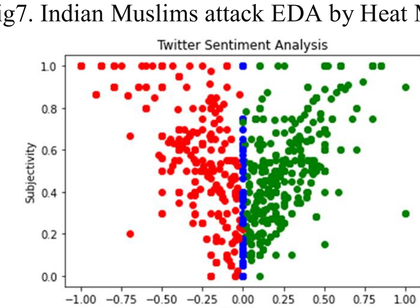

Fig8. Indian Muslims attack EDA by Scatter Plot.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
Impact Factor: 6.252

Fig9. Indian Muslims attack EDA by WordCloud.

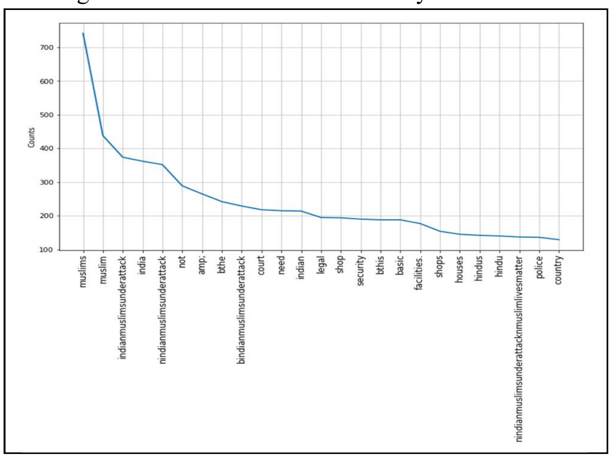

Fig9. Indian Muslims attack EDA by Line Graph.

5.2 EkMokoKejriwalNe

EkmokoKejriwalNe were the important events that occured last year, a number of tweets were recorded with this hashtags. We collected around 8000 tweets in the raw form, fed them to an in-built library and then fed them to Machine learning Algorithms like Multinomial Naive Bayes, SVM, Logistic regression, Decision tree, TF-ID Vectorizer to obtain the results more accurately.

```
#LogisticRegression(C=1, penalty='l1', solver='liblinear')

model.fit(X_test, y_test)
predictions_train = model.predict(X_train)
predictions_test = model.predict(X_test)
print('Train Accuracy:', accuracy_score(y_train, predictions_train)*100)
print('Test Accuracy:', accuracy_score(y_test, predictions_test)*100)

Train Accuracy: 53.77358490566038
Test Accuracy: 61.25000000000001
```

Fig11. EkmokoKejriwalNe accuracy using Logistic Regression.

```
[31] from sklearn.naive_bayes import MultinomialNB
     model = MultinomialNB()
     model.fit(X_train, y_train)
     predictions_train1 = model.predict(X_train)
     predictions_test1 = model.predict(X_test)
     print('Train Accuracy:', accuracy_score(y_train, predictions_train1)*100)
     print('Test Accuracy:', accuracy_score(y_test, predictions_test1)*100)

     Train Accuracy: 64.15094339622641
     Test Accuracy: 61.25000000000001
```

Fig12. EkmokoKejriwalNe accuracy using Multinomial NB.

```
[33] vectorizer = TfidfVectorizer(min_df=2)
     train_term = vectorizer.fit_transform(x_train)
     test_term = vectorizer.transform(x_test)
     #vectorizer.get_feature_names()[:20]

     #X_train
     model1 = MultinomialNB()
     model1.fit(train_term, y_train)
     predictions_train2 = model1.predict(train_term)
     predictions_test2 = model1.predict(test_term)
     print('Train Accuracy:', accuracy_score(y_train, predictions_train2)*100)
     print('Test Accuracy:', accuracy_score(y_test, predictions_test2)*100)

     Train Accuracy: 93.08176100628931
     Test Accuracy: 82.5
```

Fig13. EkmokoKejriwalNe accuracy using TF-ID Vectorizer..

```
[35] from sklearn.tree import DecisionTreeClassifier
     classifier = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
     model3 = classifier.fit(X_train, y_train)
     predictions_train3 = model3.predict(X_train)
     predictions_test3 = model3.predict(X_test)
     print('Train Accuracy:', accuracy_score(y_train, predictions_train3)*100)
     print('Test Accuracy:', accuracy_score(y_test, predictions_test3)*100)

     Train Accuracy: 67.29559748427673
     Test Accuracy: 56.25
```

Fig14. EkmokoKejriwalNe accuracy using Decision tree classifier.

```
[37] from sklearn.svm import SVC
     svc_ml = SVC(kernel = 'linear', random_state = 0)
     model4 = svc_ml.fit(X_train, y_train)
     predictions_train4 = model4.predict(X_train)
     predictions_test4 = model4.predict(X_test)
     print('Train Accuracy:', accuracy_score(y_train, predictions_train4)*100)
     print('Test Accuracy:', accuracy_score(y_test, predictions_test4)*100)

     Train Accuracy: 62.893081761006286
     Test Accuracy: 61.25000000000001
```

Fig15. EkmokoKejriwalNe accuracy using Support Vector machine.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
IJARSCT
Impact Factor: 6.252

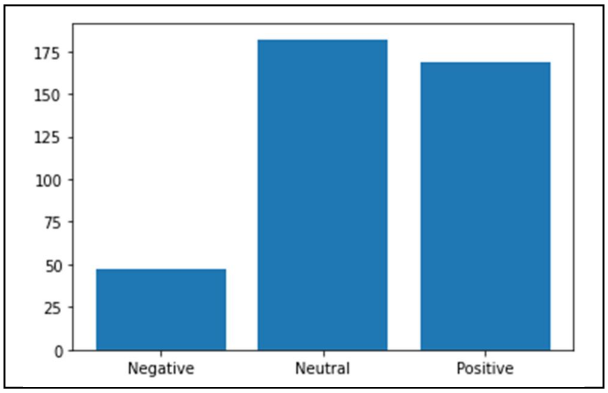

Fig16. EkmokoKejriwalNe EDA by Bar Graph.

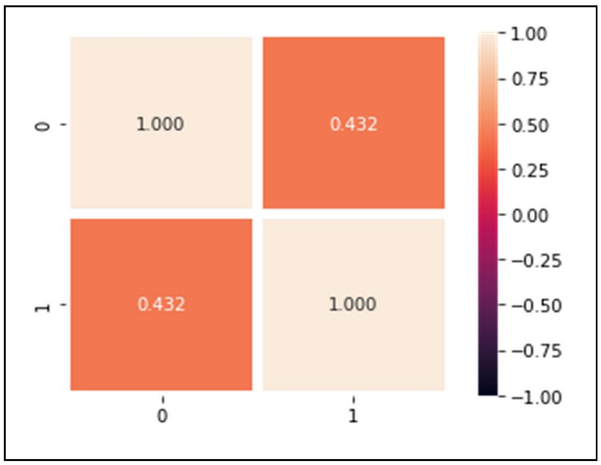

Fig17. EkmokoKejriwalNe EDA by Heat Map.

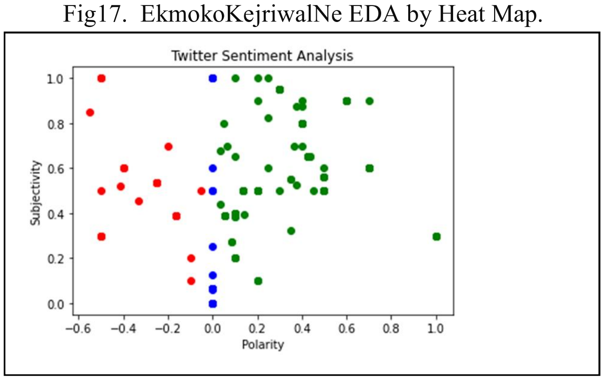

Fig18. EkmokoKejriwalNe EDA by Scatter plot.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 8, May 2022

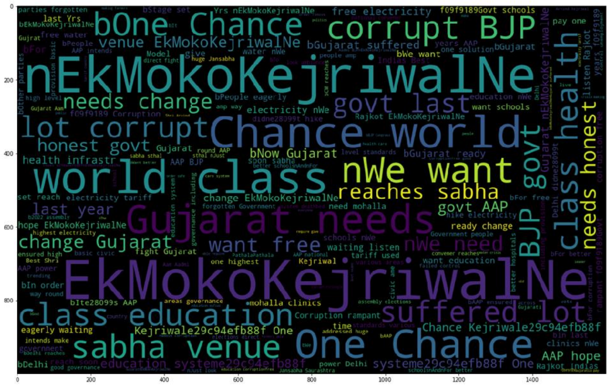

Fig19. EkmokoKejriwalNe EDA by WordCloud.

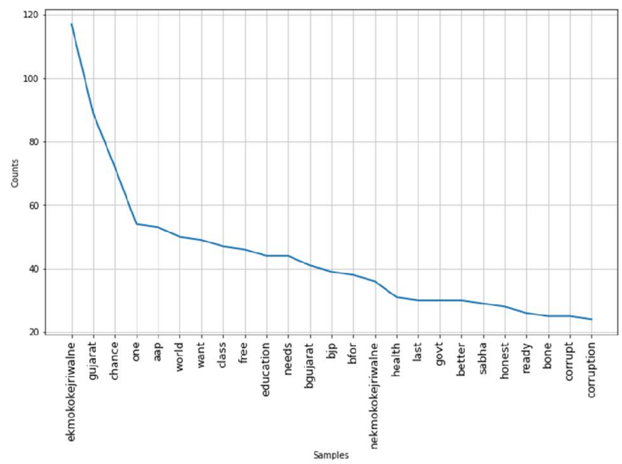

Fig20. EkmokoKejriwalNe EDA by LineGraph

5.3 RCBvsPBKS

RCBvsPBKS were the important events that occured last year, a number of tweets were recorded with this hashtag. we collected around 15000 tweets in the raw form, fed them to an in-built library and then fed them to Machine learning Algorithms like Multinomial Naive Bayes, SVM, Logistic regression, Decision tree, TF-ID Vectorizer to obtain the results more accurately.

```
#LogisticRegression(C=1, penalty='l1', solver='liblinear')

model.fit(X_test, y_test)
predictions_train = model.predict(X_train)
predictions_test = model.predict(X_test)
print('Train Accuracy:', accuracy_score(y_train, predictions_train)*100)
print('Test Accuracy:', accuracy_score(y_test, predictions_test)*100)

Train Accuracy: 63.841524573721166
Test Accuracy: 74.72417251755266
```

Fig21. RCBvsPBKS accuracy using Logistic regression.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
Impact Factor: 6.252

```
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB()
model.fit(X_train, y_train)
predictions_train1 = model.predict(X_train)
predictions_test1 = model.predict(X_test)
print('Train Accuracy:', accuracy_score(y_train, predictions_train1)*100)
print('Test Accuracy:', accuracy_score(y_test, predictions_test1)*100)

Train Accuracy: 67.42728184553661
Test Accuracy: 67.90371113340021
```

Fig 22. RCBvsPBKS accuracy using MultinomialNB.

```
vectorizer = TfidfVectorizer(min_df=2)
train_term = vectorizer.fit_transform(x_train)
test_term = vectorizer.transform(x_test)
#vectorizer.get_feature_names()[:20]

#X_train
model1 = MultinomialNB()
model1.fit(train_term, y_train)
predictions_train2 = model1.predict(train_term)
predictions_test2 = model1.predict(test_term)
print('Train Accuracy:', accuracy_score(y_train, predictions_train2)*100)
print('Test Accuracy:', accuracy_score(y_test, predictions_test2)*100)

Train Accuracy: 83.32497492477432
Test Accuracy: 74.62387161484455
```

Fig23. RCBvsPBKS accuracy using TF-ID Vectorizer.

```
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
model3 = classifier.fit(X_train, y_train)
predictions_train3 = model3.predict(X_train)
predictions_test3 = model3.predict(X_test)
print('Train Accuracy:', accuracy_score(y_train, predictions_train3)*100)
print('Test Accuracy:', accuracy_score(y_test, predictions_test3)*100)

Train Accuracy: 91.17352056168505
Test Accuracy: 62.28686058174524
```

Fig24. RCBvsPBKS accuracy using Decision tree classifier.

```
from sklearn.svm import SVC
svc_ml = SVC(kernel = 'linear', random_state = 0)
model4 = svc_ml.fit(X_train, y_train)
predictions_train4 = model4.predict(X_train)
predictions_test4 = model4.predict(X_test)
print('Train Accuracy:', accuracy_score(y_train, predictions_train4)*100)
print('Test Accuracy:', accuracy_score(y_test, predictions_test4)*100)

Train Accuracy: 67.67803410230691
Test Accuracy: 68.7061183550652
```

Fig 25. RCBvsPBKS accuracy using Support Vector Machine.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 8, May 2022

IJARSCT
Impact Factor: 6.252

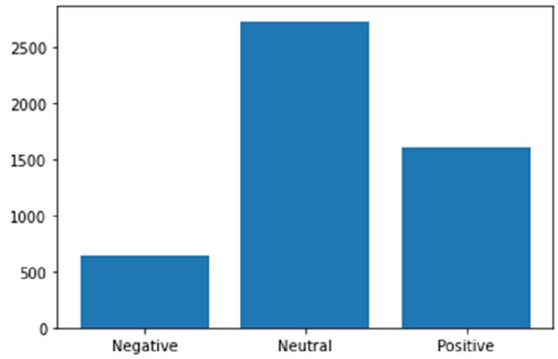

Fig26. RCBvsPBKS EDA by Bar Graph

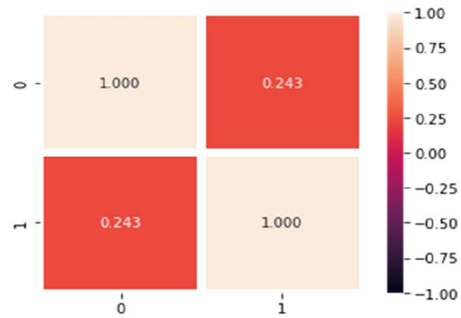

Fig 27. RCBvsPBKS EDA by Heat map.

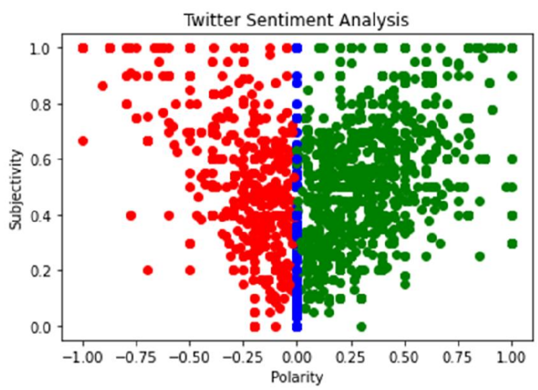

Fig 28. RCBvsPBKS EDA by Scatter plot.

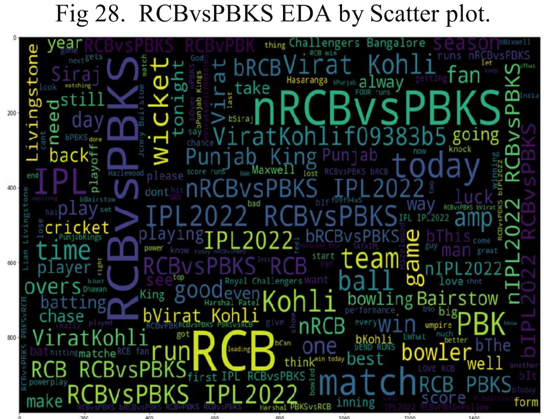

Fig 29. RCBvsPBKS EDA by Wordcloud.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 8, May 2022

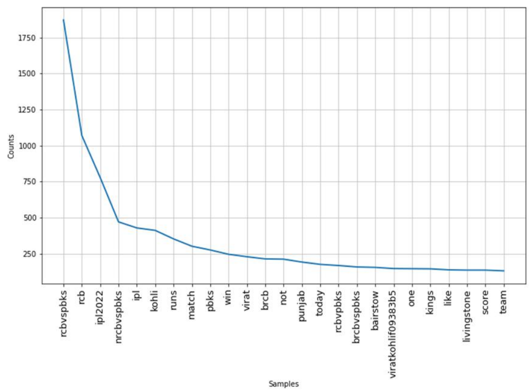

Fig30. RCBvsPBKS EDA by Line Graph.

V. RESULT

The final result obtained after feeding the different datasets to all the mentioned Machine Learning algorithms is mentioned in table I

TABLE I: SENTIMENTAL ANALYSIS

MACHINE LEARNING ALGORITHM	INDIAN MUSLIM ATTACK	EKMOKOKEJRIWALNE	RCBvsPBKS
MultinomialNB	77%	61.25%	67.9%
Logistic Regression	81.7%	61.25%	74.7%
Decision tree	77.4%	56.25%	62.2%
SVM	80.4%	61.25%	68.7%
TF-ID Vectorizer	81.5%	82.5%	65.7%

Here I have predicted the sentiments for some Hashtags like INDIAN MUSLIM ATTACK, EKMOKOKEJRIWALNE, RCBvsPBKS using various algorithms .For INDIAN MUSLIM ATTACK I got the highest accuracy using TF-ID Vectorizer as 81.5%, For EKMOKOKEJRIWALNE again I got the highest accuracy using TF-ID Vectorizer as 82.5% and For RCBvsPBKS again I got the highest accuracy using TF-ID Vectorizer as 65.7%. among all five algorithms I am getting highest accuracy using TF-ID Vectorizer only.

VI .CONCLUSION

This study found that employing Machine learning and Deep learning methods, such as Multinomial naïve bayes, SVM, Logistic regression, Decion tree classifier, and TF-ID Vectorizer, attitudes may be predicted with greater accuracy. However, the accuracy with which each of them is able to predict the sentiment differs. These have found their usage in the field of Natural language processing, particularly in sentiment analysis or to discern subjectiveinformation like opinion, emotion inside a piece of text. It can be used for social media monitoring, customer assistance, customer feedback analysis, brand monitoring and reputation, product analysis, market research, competitive research, and so on.

REFERENCES

[1]. B. Pang, L. Lee, and S.Vaithiyanathan, Thumbs up?:sentiment classification using machine learning techniques, Proceedings of the ACL-02 conference in Empirical methods in natural language processing, vol.10,2002,pp.79-86.

[2]. Edosomwan, Simeon,et al. "The history of social media and its impact on business." Journal of Applied Management and entrepreneurship 16.3(2011):79-91.

[3]. Hearst M A . Direction-based text interpretation as an information access refinement. In :Jacobs P,ed., Text-based intelligent systems: current research and practice in information extraction and retrieval. LawrenceErlbaum A ssociates, Inc , Mahwah, NJ,pp.1992,257-274.

[4]. Kessler B,Nunberg G and Sch"itze H.Automatic detection of text genre. In :Proceedings of the 35th Annual Meeting of the European chapter of the association for Computational Linguistics. Association for Computational Linguistics, Somerset, New Jersey,pp.1997,32-38.

[5]. Polanyi L and Zaenen A . Contextual valence shifters.In :Shanahan J, Qu Y and Wiebe J,eds., Computing attitude and affect in text: Theory and applications.Springer,pp.2006, 1-10.

[6]. Pedersen T () .A decision tree of bigrams is an accurate predictor of word sense. In :Proceedings of the second Annual Meeting of the North American Chapter of the Association for Computational Linguistics. 2001,pp.79-86.

[7]. Boiy,E., Moens, M.A machine learning approach to sentiment analysis in multinomial web texts.Inf Retrieval 122,526-558, 2009.

[8]. Tong S and Koller D . Support vector machine active learning with applications to text classification.Journel of machine learning Research,2:45-66.,2002.

[9]. Sentiment Analysis Using Naïve Bayes classifier Kavya Suppala,Narasinga Rao IJITEE 2019

[10]. A.Pak and P. Paroubek. „Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326

[11]. R. Parikh and M. Movassate, "Sentiment Analysis of User- GeneratedTwitter Updates using Various Classi_cation Techniques",CS224N Final Report, 2009

[12]. Go, R. Bhayani, L.Huang. "Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper,2009

[13]. L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitterfrom Biased and Noisy Data". COLING 2010: Poster Volume,pp. 36-44.

[14]. Bifet and E. Frank, "Sentiment Knowledge Discovery inTwitter Streaming Data", In Proceedings of the 13th InternationalConference on Discovery Science, Berlin, Germany: Springer,2010, pp. 1-15

[15]. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011Workshop on Languages in Social Media,2011, pp. 30-38

[16]. Dmitry Davidov, Ari Rappoport." Enhanced Sentiment Learning Using Twitter Hashtags and Smileys". Coling 2010: Poster Volumepages 241{249, Beijing, August 2010

[17]. Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social MediaData", IEEE 14th International Conference on Mobile Data ManagementMilan, Italy, June 3 - 6, 2013, pp 91-96, ISBN: 978-1-494673-6068-5, http://doi.ieeecomputersociety.org/10.1109/MDM.2013.

[18]. Pablo Gamallo, Marcos Garcia, "Citius: A Naive-BayesStrategyfor Sentiment Analysis on English Tweets", 8thInternationalWorkshop on Semantic Evaluation (SemEval2014), Dublin, Ireland,Aug 23-24 2014, pp 171-175.

[19]. Neethu M,S and Rajashree R," Sentiment Analysis in Twitter using Machine Learning Techniques" 4th ICCCNT 2013,at Tiruchengode, India. IEEE – 31661

[20]. P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417–424, Association for Computational Linguistics, 2002.

[21]. Li, S., Xue, Y., Wang, Z., & Zhou, G.."Active learning for cross-domainsentiment classification". In Proceedings of the Twenty-Thirdinternational joint conference on Artificial Intelligence (pp. 2127-2133).AAAI Press,2013

[22]. Bollegala, D., Weir, D., & Carroll, J.. Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus. Knowledge and Data Engineering, IEEE Transactions on, 25(8), 1719-1731,2013

[23]. Pang, B.and Lee, L. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts". 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04). 2004, 271-278.

[24]. V. M. K. Peddinti and P. Chintalapoodi, "Domain adaptation in sentiment analysis of twitter," in Analyzing Microtext Workshop, AAAI, 2011.

[25]. Halima Banu S and S Chitrakala, "Trending Topic Analysis Using Novel Sub Topic Detection Model", (IEEE) ISBN- 978-1-4673-9745-2, 2016.

[26]. Shi Yuan, Junjie Wu, Lihong Wang and Qing Wang, "A Hybrid Method for Multi-class Sentiment Analysis of Micro-blogs", ISBN- 978-1-5090-2842-9, 2016.

[27]. Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau, "Sentiment Analysis of Twitter Data" Proceedings of the Workshop on Language in Social Media (LSM 2011), 2011.

[28]. Neethu M S and Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", IEEE – 31661, 4th ICCCNT 2013