I.J. Information Technology and Computer Science, **2021, 2, 38-48**
Published Online April 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2021.02.04

Investigation of Different Machine Learning Algorithms to Determine Human Sentiment Using Twitter Data

Golam Mostafa[1], Ikhtiar Ahmed[2], Masum Shah Junayed[2,3,*]
[1]East West University, Dhaka, Bangladesh.
[2]Daffodil International University, Dhaka, Bangladesh.
[3]Bahcesehir University, Istanbul, Turkey
E-mail: gmostafa1210@gmail.com, ikhtiar15-5257@diu.edu.bd, masumshahjunayed@gmail.com

Received: 17 July 2020; Accepted: 26 February 2021; Published: 08 April 2021

Abstract: In recent years, with the advancement of the internet, social media is a promising platform to explore what going on around the world, sharing opinions and personal development. Now, Sentiment analysis, also known as text mining is widely used in the data science sector. It is an analysis of textual data that describes subjective information available in the source and allows an organization to identify the thoughts and feelings of their brand or goods or services while monitoring conversations and reviews online. Sentiment analysis of Twitter data is a very popular research work nowadays. Twitter is that kind of social media where many users express their opinion and feelings through small tweets and different machine learning classifier algorithms can be used to analyze those tweets. In this paper, some selected machine learning classifier algorithms were applied on crawled Twitter data after applying different types of preprocessors and encoding techniques, which ended up with satisfying accuracy. Later a comparison between the achieved accuracies was showed. Experimental evaluations show that the Neural Network Classifier' algorithm provides a remarkable accuracy of 81.33% compared with other classifiers.

Index Terms: Sentiment Analysis, Tweet, Twitter, Sentiment, Social Media, Machine learning, Natural Language Processing.

1. Introduction

Sentiment Analysis is a way to identify people's opinions, attitudes, and emotions from text data, generally which is available on different platforms on the internet. It is the most well-known and leading area in Natural Language processing [1]. Sentiment analysis has become progressively significant in recent years and is widely known for the processing of data that can be collected from social networking sites, blogs, Wikis, microblogging websites, and other shared online media [2]. With the rapid growth of internet users, people frequently express their feelings over the internet on these platforms and through reviews. Because of this emergence in textual data, the principle of communicating sentiments needs to be investigated. Big companies and marketing agencies often use sentiment analysis to find out new business strategies and marketing campaigns.

Sentimental Analysis is mainly a method of evaluating the individuals' opinions or mood, calculated in mathematical terms as positive, negative, or neutral. Data mining is also another term for the study of sentiments analysis [3]. Identify the emotional viewpoint is very useful in many sectors such as the business industry, politics, and public behavior. So, it is really important to consider the thoughts and feelings of the consumers or users to grow a business or the relevant areas. Sometimes sentiment analysis can be so influential and powerful that, when it comes to politics, it can be used to predict the election outcome. Moreover, it can be used to change people's point of view and way of thinking.

Of all the sources to collect textual data, Twitter is the most popular one. It is a platform where people can share their own opinions and thoughts publicly with a limited number of words known as 'Tweet'. People from almost every class including celebrities, politicians, and popular persons all over the world use Twitter as an opinion sharing platform. The number of Twitter users is increasing rapidly every day. According to Internet Live Stats [4], currently, over 6000 tweets are posted every second on average which leads to almost 500 million tweets every day and without any doubt, the number is huge. This huge number of tweets carry a lot of important information and messages. It is very easy to crawl runtime tweets from Twitter with their tags and can be used for analyzing as well as other purposes. For the above reasons, Twitter data is a good choice for sentiment analysis. Sentiment analysis uses different methods and algorithms of Natural Language Processing [5]. The automatic systems of sentiment analysis deal with different machine learning

techniques to get sentiment from data. Our approach helps to easily find the Positive Review, Negative Review, and Neutral Review. We analyze our output from different performance metrics like timing, precision, and memory.

In this paper, live tweets were collected using Twitter API. We used TextBlob for pre-processing data to perform the data several NLP tasks. In our experiment, to process the data we use some features like Tokenization, StopWord Filter, and WordNet Lemmatizer. Then those encoded numeric forms were fed to different Machine Learning Classifier algorithms to determine the accuracy. The main objective of this paper is to study the People's reaction to Twitter data and also provide a theoretical comparison of the approaches. This Paper organized several classification techniques (e.g. Naïve Bayes, Neural Network, K-Nearest Neighbors, K-Nearest Centroid, Logistic Regression, and Logistic RegressionCV, SVM) which is used in sentiment analysis.

2. Literature Review

In this section, we provide a summary of previous work on the social media context based on the machine learning algorithm for sentiment analysis. After the exploration of some previous work, we think that Neural Network, Naïve-Bayes, and support vector machine (SVM) is the most standard machine learning techniques for solving sentiment analysis [6].

Regarding the paper [7] researchers used a machine-learning algorithm to analyze Twitter comments on political views. Naïve Bayes Classifier and SVM classifier utilize the training set to build an analysis model. Based on the model, used three different types of sentiment analyzer such as TextBlob, SentiWordNet, and WSD. They analyze the Positive, Negative, Neutral results whereas TextBlob provides the highest positive accuracy result compared to others approximately 54.08%. Also, the Naïve Bayes classifier works better than Neural Network which is to analyze more positive tweets. Mainly the researchers focused on the comparison of several sentiment lexicons.

As explained in Abdullah and Mohammad [8] implement three different types of techniques to identify Twitter content based on the expressions. In the supervised machine learning approaches, a classifier trained datasets in three different ways like positive, negative, and unbiased tweets where SVM, Bayesian, and Entropy classifier used to identify the sentiment polarity of tweets. But on the other hand, ensemble approaches are used to multiple classifiers to precise and accurate predictions. The main drawback of the approaches is it takes more time when required separate predictions. Lexicon based approaches depend on the polarity of the text. To overcome the problem, they used external resources such as SentiWordNet, MPQA, and WordNet-Affect to analyze tweets. After that, they used an ensemble and hybrid-based approach to classify text compared to the supervised machine learning techniques and achieve the highest classification accuracy around 85%.

In a different work [19], the researcher developed a tweet classification system using deep learning-based sentiment analysis techniques to identify extremist or non-extremist tweets. In the experiment, they used multiple machine learning classifiers and also implement CNN, LSTM, and CNN+LSTM. to investigate the classical features. Process the Twitter reactions they used some techniques like tokenization, stop word removal, and special removal. In the training process, implement CNN and also used some layers to finalize the data. They compare the extremist and non-extremist tweets using the LSTM+CNN model and ML and DL classifiers.

In our work, we used seven types of classification technique which is different from other paper such as Naïve Bayes, Support Vector Machine, K Neighbors Classifier, Logistic Regression CV, Logistic Regression, Neural Network Classifier, and K Nearest Centroid to analyze social context. To compare them, the Support Vector Machine, Naïve Bayes, and Neural Network always provide better accuracy than others. In our experiment, we found that the Neural Network Classifier algorithm achieve the highest accuracy of 81.33%.

3. Classification Techniques

In the wide-area of machine learning, lots of classification techniques have been developed. Classification techniques are used to classify the unleveled data item. Nowadays, classification is the most useful technique to identify unstructured data [10]. The main purpose of the classification is to identify classes or groups [11]. Classification techniques we used depend on the polarity because every classifier works differently from one to another.

$$\text{Polarity} = \frac{\text{P(Positive\_Word)/P(Total\_Words)}}{\text{P(Positive\_Negetive)/P(Total\_Words)}} \quad (1)$$

In our paper, we have implemented several machine learning classifiers that are given below.

- Naïve Bayes (NB)
- Support Vector Machines (SVM)
- Neural Network (NN)
- K-Nearest Neighbors (KNN)

- K-Nearest Centroid (KNCen)
- Logistic Regression (LR)
- Logistic Regression CV (LRCV)

3.1. Naïve Bayes

The Naïve Bayes algorithm is based on the Bayes theorem and consists of several popular features. Mostly, Naïve Bayes predicts the data based on various attributes. It is the most popular classification technique to solve real-world problems.

The algorithm requires a small amount of training data to identify the necessary parameters and it is also a fast classifier compared to other classifiers [11]. Recently, Naïve Bayes is the most common technique for text classification into multiple classes but recently it is utilized for sentiment analysis [8].

$$P(a/b) = \frac{P\left(\frac{b}{a}\right) * P(a)}{P(b)} \tag{2}$$

Here, P(a/b) is the posterior probability of class given predictor and P(b/a) is the likelihood probability of predictor given class. On the other hand, P(a) is the prior probability class and P(b) is the prior probability of predictor.

3.2. Support Vector Machine (SVM)

Support Vector Machine is a supervised machine learning algorithm used for both classification and regression problems [12]. But it is more popular to solve the classification problems. SVM use kernel functions for categorizing data, text, images as well as vectors [13]. In the SVM algorithm, they utilize hyper-plane to separate the two classes [12]. Hyper-plane easily differentiates data points and also helps to find the maximum distance between the nearest data point.

From figure 1, we can see that the H1 hyper-plane doesn't separate the two groups. H2 separate with a small margin. But, the H3 hyper-plane separates the classes with the maximal distance.

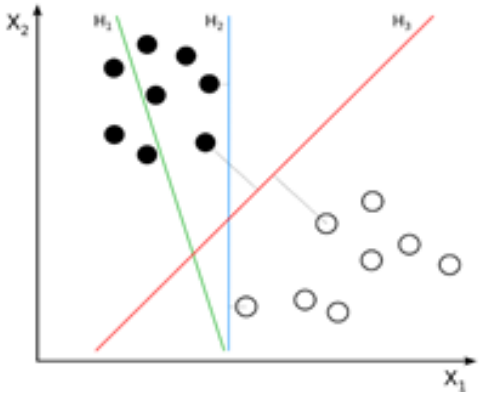

Fig.1. Support Vector Machine.

3.3. Neural Network

The neural network is a learning process of the human brain and consists of a wide range of layers of neurons. In every layer, accepting inputs from previous layers and calculate the output from it again passing the outputs to the next layers [13]. It is a continuous process and in the last stage gets the final outputs.

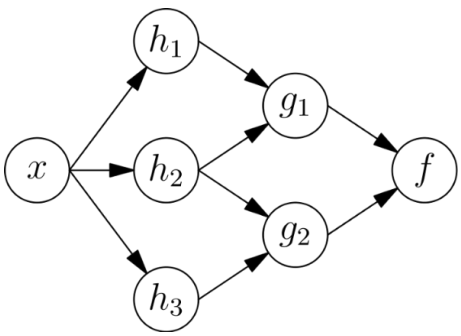

Fig.2. Working Process of Neural Network.

The above figure shows the working process of a neural network. Suppose, x is the initial input of the process then it passed the neurons to the first layer (h1). After that, the first layer (h1) receives the input from the previous and generate an output. Then, this output is again passed to the second layer (g1), it is calculated the output based on the first layer. At last, the combined output of the second layer is called the final output of the model.

3.4. K-Nearest Neighbor

K-nearest neighbors' algorithm is the simplest and effective rule for pattern classification. It is assigned to a class level to each query pattern which is compared with the nearest centroid value of the classifier. In figure 3, the classifier always uses feature similarity to identify the nearest data points to predict the values.

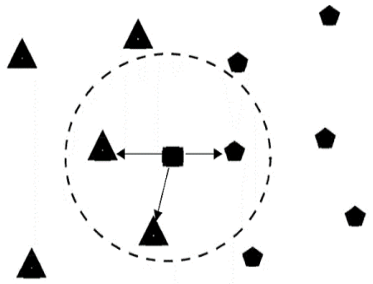

Fig.3. K-Nearest Neighbor.

3.5. K- Nearest Centroid

K-Nearest Centroid is a classification that assigns objects whose mean or centroid is nearest to the observation. The centroid presents every class, with test samples assigned to the nearest centroid level. It helps to distribute information to its neighbors. Nearest-centroid classifiers were widely used in various high-dimensional applications, in recent genomics. In addition to considering the proximity and geographic distribution of k neighborhoods, the proposed scheme also uses the local mean vector of k neighbors from each group to make classification decisions.

3.6. Logistic Regression

Logistic regression works when the dependent factor is binary and it is the correct regression analysis to perform. It is a statistical method, like all regression analyses. It is used to characterize data and illustrate the relationship with one dependent binary variable and one or more independent nominal, ordinal, interval, or ratio-level variables [14].

3.7. Logistic Regression CV

Logistic Regression CV is a scikit learning model where CV functions allow the k-fold cross-validation. It is normally used to separate the data into two sections such as training and validation which is helping to improve the result.

4. Proposed Architecture

We have already described which classifiers have been used for this research work in the 'Classification Techniques' section of this paper. Figure 4 gives us a very basic idea of our proposed architecture or methodology of this work.

Here classification techniques were applied based on the following structure.

First, we cleaned the runtime tweets and converted them into CSV format. Then Tokenization was used to tokenize the tweet's word by word. After that Stopword filter was used to remove the unnecessary words from the tokenized tweets. Then WordNet Lemmatizer was used to turn a word into its base word (e.g. went → go, doing → do).

Now comes the encoding part. Here level Encoder was used to encode the sentiment and OneHot Encoder was used to encode the tweets. Lastly, the encoded numeric form was fed in the selected classification algorithms to achieve our result. Later we will briefly describe how they were used in the 'Methodology' section.

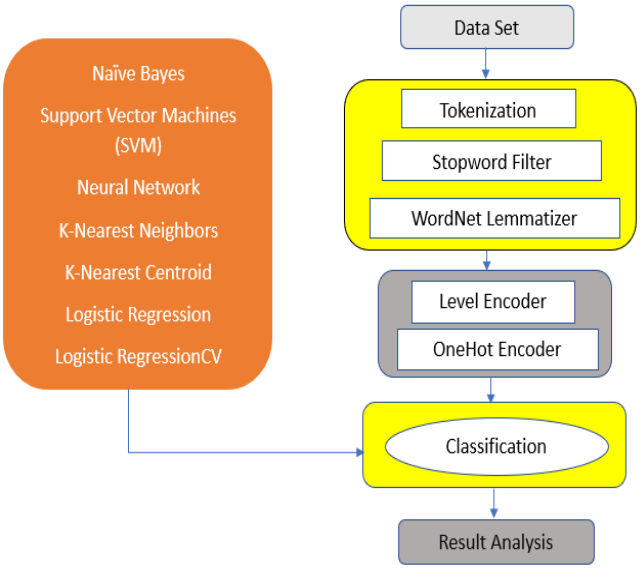

Fig.4. Structure of the Proposed Architecture.

5. Methodology

Now a day's Twitter is a widely used social network where people can express their opinion easily. Hundreds of million tweets in a single day. So, it is a good platform to collect real-time data to analyze people's sentiment. In this part, we describe the process of our work. It takes five steps to complete the full process.

5.1. Data Collection

To collect the data, at first, we need a Twitter account where the Twitter provider gives access to data from a Twitter account, and this data, we can use our purpose. After that, we need to create an application for streaming tweets according to the necessary details. Using API, we can easily get customer key, customer secret key, access token key, and access secret key which is used to authenticate the user to access Twitter data [6].

To extract data from Twitter, we have used Tweepy because Tweepy is the most powerful python library which helps to easily gather data from Twitter. Also, send to a connection request to the API we stored streaming data to the database. Then, we used TextBlob (Python library) to perform the data several NLP (Natural Language Processing) tasks.

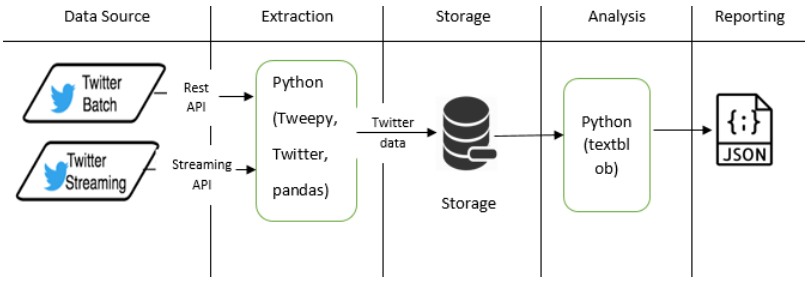

Fig.5. The Architecture of Data Collection.

5.2. Pre-process Data

In our experiment, we need a large amount of data to complete the process. From Twitter, using Twitter Batch and Twitter streaming we collect approximately 500000 data. The main goal of pre-processing is to remove the unwanted sign. A code which is written in C++ programming was used to get a CSV file from the JSON format. After pre-processing the data, we get the final dataset in CSV format with 448013 data.

From the architecture in figure 6, we can see the flow of pre-processing data.

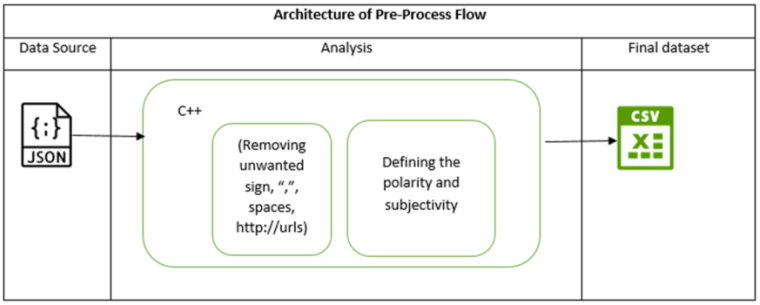

Fig.6. The Architecture of Pre-process data.

- Tokenization

Tokenization is the process of breaking a sequence of string that provides data security. The main goal of tokenization is to secure the method of storing information. It always separates the logical information from the components. Tokenization is language-specific and every language has its tokenization [15]. It breaks the language in every single word using the python spilled function. Tokenization converts sensitive data to one kind of new token ID's and stores sensitive information remotely and redundantly [16]. In our work, we use tokenization to break the text individually and remove the data redundancy.

- StopWord Filter

StopWord filter is commonly used in natural language processing. The way of converting data sometimes needs to pre-process the data. Because, in natural language processing, cannot convert the useless words and at that time, it referred to stop the unusual word [16]. The main purpose is used to StopWord filter remove useless data such as ("the", "a", "an", "in") because the search engine ignores this kind of word.

- WordNet Lematizer

Lemmatization is the way to change the word into its base form. Sometimes Lemmatization works with stemming because lemmatization converts the text into meaningful structure whereas stemming just deletes some characters to solve the spelling errors [17]. For example:

'Sharing' → Lemmatization → 'Share'
'Nearing' → Lemmatization → 'Near'
'Caring' → Stemming → 'Care'

The main goal of both stemming and lemmatization is to remove the inflectional forms [18]. For Instance:

Am, are, is → be
Book, books, book's, books' → book

If we can see the one sentence using lemmatization and stemming –
The girl's bags are different colors →The girl bag be differing color
In our experiment, we use lemmatization because of getting accurate results and upgrade our accuracy.

5.3. LabelEncoder

LabelEncoder is a popular encoding technique which is used to convert categorical data into numeric form. Mostly it is used to encode the target labels into a range between 0 to and n_classes-1 [19]. In this paper, LabelEncoder was used to encode the sentiment (e.g. positive, negative, neutral).

5.4. OneHotEncoder

OneHotEncoder is another popular method that is used for encoding something known as a one-hot numeric array from different categorical features. The input to which is to be encoded should be an array of strings or integers, indicating the attributes assumed by categorical features [20]. In this experiment, OneHotEncoder was used to encode the tweets from the dataset.

5.5. Classification Approach

Classification techniques were implemented based on the training dataset. Table 1 applies to all variables with specific values. These parameters can be defined to restore the results we have obtained in this research. From the table, we can see that we used 85% data to train our model and 15% for testing.

Table 1. Required Parameters and Values

Parameters	Values	Parameters	Values	Parameters	Values
Iteration	3000	Number of Layers	5	Number of Weights	10
Train_Size	85%	Nodes Per Layer	10	No of Biases	10
Rate of Learning	0.01	Activation Function	Tanh	Initial Value (Biases)	1
Loss Function	Binary Cross Entropy	Output Activation Function	Sigmoid		
Range of Weights	(-1, 1)	Initial Value (Weights)	1		

In our experiment, we have implemented only one approach. Many searchers have done their experiment by Naïve Bayes Classifier, Support Vector Machine, and Maximum Entropy [8]. But we have used several machine learning techniques that have been developed already. For a comparative study of the performance, we have implemented seven different types of classifiers such as Naïve Bayes, Support Vector Machines (SVM), Neural Network, K-Nearest Neighbors, K-Nearest Centroid, Logistic Regression, and Logistic RegressionCV.

6. Dataset Overview

"Sentimental Analysis" has been obtained from the Tweeter developer account. We collect live data using Twitter API. In the dataset, there are two columns and 448013 rows, which means there are 448013 data of the Twitter users. In the very beginning, the sentiment was given as a numeric form such as (-1 to +1). It means if the polarity becomes negative range it means that the sentence is negative or negative sentiment. On the other hand, if the polarity becomes positive then it means its positive sentiment. If it becomes zero, then the sentence stands for the neutral. In figure 7 we can see the attributes of our dataset.

	A	B
1	Tweets	Sentimen
2	2500 retweets and no test !! help our ap class out\ud83d\ude2d\ud83	Neutral
3	We gave @Saweetie a lie detector test and we were for real Was the	Positive
4	Read the latest fishing news reports & offers - 19th November\n	Positive
5	Do you think degrees are getting easier? \ud83e\udd14\n\nHead back	Neutral
6	Gbam!!!! \nTachaXJack\nTachaXJack 6	Neutral
7	Take a look inside the Popcornopolis test kitchen O	Neutral
8	like fuck a rap carrier let\u2019s test these streets an make a mill	Negative
9	Failure is a teacher that gives you the test first and lesson after Remer	Negative
10	This is incredibly bizarre & doesn\u2019t pass the smell test \u2b	Positive
11	@billygil hope your stomach is feeling better One thing you'll find exc	Positive
12	$5 @ggvertigo SITE CREDITS\n\nROLLING ON STREAM RNNNNNN\n\n'	Neutral
13	@RepRatcliffe's line of questioning is the first moderately effective o	Positive
14	SO UH there's an unused test program in Twinkle Star Sprites a super s	Negative
15	$roku Symmetrical triangle break-out & re-test g	Neutral
16	Pass the small test \ud83d\udc40 h	Negative
17	PRESS PRESS PRESS PRESS PRESS my BBY DADDY give me loads of stres	Negative
18	Recreational facilities including ice arenas should use good ventilatio	Positive
19	StayStrongSidharth\n\n\"Life will test you but remember this \nWher	Neutral

Fig.7. Attributes in Dataset.

Figure 8 describes the occurrence of each significance. Here among 448013 significances in the dataset, 33% of the whole dataset is classified as a positive sentiment. After that, 49% is classified as Neutral. Moreover, 18% are classified as Negative. Therefore, 33% of people are bringing their sentences as positive however the number is less of the negative statements is 18%.

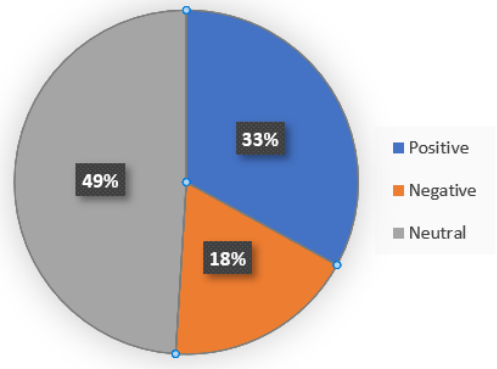

Fig.8. Occurrence of Each Significance.

7. Result Analysis

In this section, analyze the results that we have achieved after implement several classifiers. The results of the current segment and the output of all the methods considered were assessed and analyzed using different performance metrics like timing, precision, and memory requirements. Moreover, the iteration cost of the graph has also been described graphically and theoretically for comparative study. Additionally. On the other hand, training time and memory use of time has also been conducted. All these performance evaluation metrics have been concentrated to find the superior approach among all the approaches applied here in this work. All the approaches of Comparative studies are as follows.

Here, we have decided on the accuracy of each algorithm. As a result, it can compare to each other to determine which algorithm is most efficient to determine the most accuracies. The table of algorithm accuracies is shown below. From table 2, it can be observed that we have achieved higher accuracy from the Neural Network.

Table 2. Accuracies of the Algorithms

Algorithm Name	Accuracy (%)
Naive Bayes (NB)	77.16
Support vector machine (SVM)	79.00
K Neighbors Classifier (KNCl)	71.83
Logistic Regression CV (LRCV)	73.33
Logistic Regression (LR)	76.66
Neural Network (NN)	81.33
K nearest Centroid (KNCen)	67.53

Based on the information provided in table 2, we can generate a graph from the table. That performance of almost all the approaches is quite the same rather than a little bit different. It can be easily notified that Neural Network has the Compared to other category alternatives, 2-3 percent higher accuracy compared to this research. The correlation of accuracy was shown graphically in Figure 9.

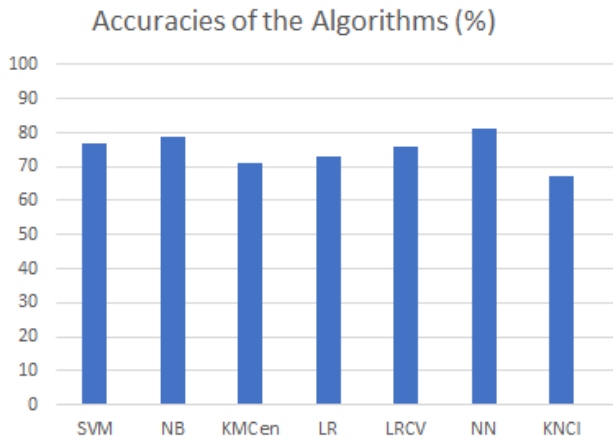

Fig.9. Comparing the Algorithms Accuracy.

In the meantime, we found the Binary Cross Entropy framework for cost estimation to implement the value function. We extracted the costs per iterations graph referred to in Figure 10 for better understanding.

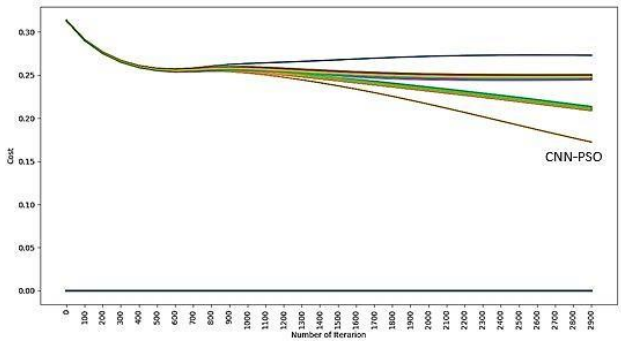

Fig.10. Cost Per Iteration Graph.

Figure 10 Shows that Logistic Regression CV reduces costs more than any method in this research which assures our approaches ' superior actions. Also, we have a variety of memory usage and timing requirements in the training phase for a much more vivid view of the efficacy and efficiency of our suggested approach.

Table 3. Comparison of Performance of the Proposed Approach

Name of the Approach	Memory Usage (MB)	Time Required for Training (sec)	Accuracy (%)
Neural Network	413.31	1916.66	70.33
Support vector machine	814.34	20833.33	70.00
Logistic RegressionCV	1001.21	13333.33	70.33
Na ïve Bayes	812.34	411.11	68.16
K-nearest Centroid	856.32	399.11	62.53
K Neighbor Classification	756.32	2833.33	61.83
Logistic Regression	701.00	535.833	68.66

Besides, Table 3 refers to the total time and memory needed while on our dataset training systems. Our classification methods seem more optimistic and accurate in the event of time and accuracy. Moreover, our classification approach's memory requirement is quite a higher support vector machine (SVM), but the total presentation notes that the best performance is revealed by our classification method.

The experiment reveals that the Neural Network has given the most promising result of all. Meanwhile, it took less time and higher. Figure 11 gives us an overall idea of a comparison between accuracy, memory, and time.

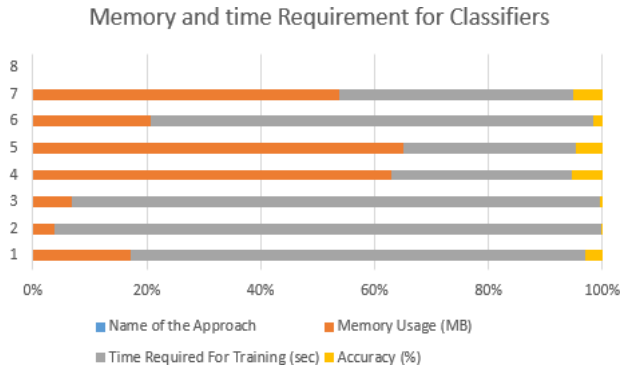

Fig.11. Comparison Between Accuracy, Memory and Time.

8. Conclusion and Future Work

Data is a very powerful thing in this generation. A vast amount of data with valuable information and messages are generated on Twitter as well as other platforms on the internet every day. For this reason, text mining or sentiment analysis has become very popular research nowadays. In this paper, our main goal was to find a comparison between different algorithms and so different machine learning classifier algorithms were used to analyze a large amount of

runtime Twitter data after using different preprocessors and encoding techniques as mentioned earlier. Now accuracies can be varied depending on the dataset and way of using algorithms. Among all the classifiers, Neural Network Classifier was the optimal one for greater accuracy (81.33%) and precision here.

The single most significant and amusing thing about research is that there is no best solution. There is always a better solution for the existing solved problems. Twitter data was used in this paper which can be used as a helping tool for different critical tasks easily like finding patterns or activities such as detecting hate speech about homophobia and racism, spam filtering, data security, and more. There are some limitations and lots of further working scope on this paper. This work can be done on an even bigger data set to obtain better accuracy. The bigger dataset will be used the stronger model can be achieved. This work can be extended by determining subjectivity as the machine can determine a person's sentiment whether it is objective or subjective. By the dataset of this paper, a module can be made using an embedded system, where a person will input his voice and after analyzing the sentence the module will reply automatically about the person's sentiment. This work can be used to build a model that can detect sentences about multiple senses. Also, a web-based application of sentiment analysis can be made in the future.

References

[1] A. Tripathy, A. Agrawal, S K. Rath "Classification of Sentimental Reviews Using Machine Learning Techniques", Procedia Computer Science: 57(2015)821-829, Issue: 2015

[2] I. Chatruvedi, E. Cambria, Roy E. Welsch, Francisco Herrera "Distinguishing between facts and opinions for sentimental analysis: Survey and Challenges", Information Fusion: 44(2018)65-77, Issue:2018

[3] S. Siddharth, R. Darsini, Dr. Sujithra "Sentiment Analysis on Twitter Data Using Machine Learning Algorithms in Python", Vol.5, Issue: 2 February, 2018. Twitter Uses Statistics https://www.internetlivestats.com/twitter-statistics/, Last Accessed: 08 July, 2020.

[4] Sentiment Analysis https://monkeylearn.com/sentiment-analysis/, Last Accessed: 08 July, 2020.

[5] S. Bhargava, S. Choudhary "BEHAVIORAL ANALYSIS OF DEPRESSED SENTIMENTAL OVER TWITTER: BASED ON SUPERVISED MACHINE LERANING APROACH", Issue: 2018.

[6] A. Hasan, S. Moin, A. Karim, S. Shamshirband "Machine Learning-Based Sentiment Analysis for Twitter Accounts", Issue: 27 February, 2018.

[7] A. Alsaeedi, M Z. Khan "A Study on Sentiment Analysis Techniques of Twitter Data", Vol. 10, Issue: February 2019.

[8] Shuo Xu "Bayesian Naïve Bayes Classifiers to text Classification", Vol.44, Issue:2018

[9] Machine Learning Techniques "https://www.edureka.co/blog/classification-in-machine-learning/" Last Accessed: 08 July, 2020

[10] Types of Classification "https://analyticsindiamag.com/7-types-classification-algorithms/" Last Accessed: 08 July, 2020

[11] Understanding Support Vector Machine (SVM) algorithm from examples "https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/", Last Accessed: 09 July, 2020

[12] Classification using neural Networks "https://towardsdatascience.com/classification-using-neural-networks-b8e98f3a904f", Last Accessed: 09 July, 2020

[13] Logistic Regression "https://www.statisticssolutions.com/what-is-logistic-regression/", Last Accessed: 09 July, 2020

[14] Text Analytics Functions "https://www.lexalytics.com/lexablog/text-analytics-functions-explained", Last Accessed: 10 July, 2020

[15] Tokenization "https://auricsystems.com/info-pages/tokenize-what-matters", Last Accessed: 10 July, 2020

[16] Lemmatization "https://www.machinelearningplus.com/nlp/lemmatization-examples-python/", Last Accessed: 10 July, 2020

[17] Stemming and Lemmatization "https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html", Last Accessed: 10 July, 2020

[18] Shakeel Ahmad, Muhammad Zubair Asghar, Fahad M. Alotaibi, Irfanullah Awan "Detection and Classification of Social Media-based Extremist Affiliations Using Sentiment Analysis Techniques", Human-centric Computing and Information Sciences, Issue: 2019

[19] J. Samuel, G.C.Md. Nawaz Ali, Md. Mokhlesur Rahman, Ek Esawi and Yana Samuel "COVID-19 Public Sentiment Insight and Machine Learning for Tweets Classification", Issue: 11 July, 2020

Authors' Profiles

Golam Mostafa completed his bachelor's degree in Computer Science and Engineering (CSE) at East West University in December 2019. Currently, he is working as a Research Assistant at the Vision Research Lab of AI (VRLAI). Also, working as a Jr. Software Developer in a software company. He aims to work and research in Data Science, Natural Language Processing, and Machine Learning in the future.

Ikhtiar Ahmed received his B.Sc. degree in Computer Science and Engineering (CSE) from Daffodil International University in 2019. Currently, he is working as a Research Assistant at the Vision Research Lab of AI (VRLAI). Also, working as a Jr. Software Developer in a software company. He aims to work and research in Data Science, Artificial Intelligence, and Machine Learning in the future.

Masum Shah Junayed is continuing his M.Sc. degree at the Department of Computer Engineering at Bahcesehir University (BAU). He obtained his B.Sc. degree in Computer Science and Engineering from Daffodil International University in 2019. Recently, he is working as a Research Assistant at BAU Computer Vision Lab and Gradient Lab for AI Research (GLAIR). He is also a Graduate Research Fellow at BAU, Member of Machine Intelligence Research Labs (MIR Labs). He is the founder and director of the Vision Research Lab of AI (VRLAI). He is much fond of research. He has had several publications in international journals and conference proceedings. He has worked as a reviewer of an international journal. His main research interest is in Artificial Intelligence, especially Machine Learning, Computer Vision, Natural Language Processing, Medical Image Analysis.

How to cite this paper: Golam Mostafa, Ikhtiar Ahmed, Masum Shah Junayed, "Investigation of Different Machine Learning Algorithms to Determine Human Sentiment Using Twitter Data", International Journal of Information Technology and Computer Science(IJITCS), Vol.13, No.2, pp.38-48, 2021. DOI: 10.5815/ijitcs.2021.02.04

