A Novel Approach to Predict the Real Time Sentimental Analysis by Naive Bayes & RNN Algorithm during the COVID Pandemic in UAE

Abdulrahman Radaideh
Collage of Engineering
Engineering Management
University of Sharjah
Sharjah, UAE
u18105563@sharjah.ac.ae

Fikri Dweiri
Industrial Engineering &
Engineering Management
University of Sharjah
Sharjah, UAE
fdweiri@sharjah.ac.ae

Mohammad Obaidat
Computing and Informatics
Fellow of IEEE & Fellow of SCS
University of Sharjah
Sharjah, UAE
mobaidat@sharjah.ac.ae

Abstract— **At present, each and every part of the globe are facing a COVID crisis, which is affecting an individual physically, mentally and on the other hand, it is affecting the nation economically. Also, the unemployment scenario will be at its peak in the upcoming years as reported by UNGA. To combat this scenario, all the country are working on fostering their fiber network and so the sectors apart from the manufacturing, will tend to work from their home and contribute to the economy. But there are many problems arising to implement this culture practically, since it affects the mindsets of the people who have to endure this transformation within a very short span. Hence, in this research work, it has been decided to focus on this current issue for which the usage of certain apps in UAE such as zoom, totok, botim for internet calling have been identified since this is the only way of connectivity with the outside world. To perform this analysis, the tweets from December to July have been collected by converting the image to text and analyzed using two algorithms such as Naive Bayes Classifier (NBC) and Recurrent Neural Networks (RNN). The sentimental analysis found that 630 tweets were positive and people in UAE feels secured, satisfied and internet calling is very useful for them in the prospect of work, education, etc. Only 48 tweets has negative impact because people feel little bit harder in sudden change of culture with in short period of time and 155 tweets has impact that both positive and negative were view and said to be natural. The study found that NB (84%) is more accurate, user friendly and takes less time than RNN (79%) to perform the analysis. Finally, the sentimental analysis reveals that people in UAE were accepting the new culture of internet calling and it is useful for them in the prospect of work and education.**

*Index Terms—***COVID19, Internet calling, Naive Bayes, Recurrent neural networks, Sentimental analysis**

I. Introduction

The current scenario of the world is uncertain not only in terms of health but also in terms of future approach. The development activities in each and every sector have taken a leap backward and all the industries are working to find a solution to cope up with the crisis without any loss (Shen et al, 2020). It has been found that internet is the best way to communicate with the colleagues to perform any sort of work where the information can be shared and received in return (Chamola et al, 2020). Also, it has to be noted that the Industrial 5.0 is in papers frequently which will create a major revolution in the industrial sector for which the internet will be a key (Okorie, 2020). will be the mindset of the individuals who will be performing the day-to-day activities irrespective of any change happening. Before, launching any new technology, the individual concern is very much important to attain the specified result. Universally, it has been accepted that the sentimental analysis will be mandatory to bring any change in the society (Gang et al, 2010). The feedbacks of the people have to be gathered and the development has to be made accordingly to be successful. To perform the analysis, social media is the better choice where the individuals are posting their opinions without any hesitation which encourages this research to use the data source of twitter application for the sentimental analysis (Yuan, 2013). A novel method to understand the sentimental analysis was given by **Ye Wu** (Ren, 2011) which highlights the correlation between the influencing and influenced probabilities. **Batool (2013)** highlighted that the tweet classification initially is needed before performing the analysis which will help in focusing the exact tweets needed for evaluation instead of all the tweets. The tweets can be classified efficiently by extracting the root words, different figures of speech and synonyms which will enhance the efficiency and information needed. In order to further narrow down the tweet classification.

Mohammad (2019) evaluated the wrapper and filter approach for improving the sentimental analysis. It has been concluded that the approach yielded better results in most cases while reducing the features. **Shahana (2015)** evaluated the various features on sentimental analysis by considering the function words, unigram, POS tag of

words, bigram as the feature set. Based on the comparison, it was confirmed that the unigram is found to be better to extract the review with better accuracy. In order to improve the efficiency, **Rathi (2018)** investigated the sentiment analysis of tweets using machine learning approach and the results prove that the developed model has shown improvement in the overall classification accuracy and f-measure of sentiment prediction as compared to traditional existing techniques for classification.

Pandey (2017) proposed a novel metaheuristic hybrid cuckoo search method to perform the sentimental analysis both theoretically and statistically. Based on the validation, it was found that this model has a tendency to generate conclusive reviews on any social issues. To analyze the predictability of the model, **Goel (2018)** performed a real time analysis of the social media dynamics to create a unique desktop application that showcases the tweets priorly to users based on the most used keyword for a particular user from the same timeline. Since it involves, real time analysis, the preference of users varies for each and every hour and hence, an optimized model is still an area to be considered.

Fırat et al (2018) used a three way classification to perform the sentimental analysis such as positive, neutral and negative. From the analysis, it has been concluded that the three way classification is found to be providing more insights about the individual opinion and will act as a tool for the betterment of any product. Also, this method is found to be widely accepted and many algorithms have been used and validated.

Based on the conclusion of various researchers, it has been found that the evolution of various machine learning approaches has provided a major breakthrough for generating better accuracy during sentimental analysis. But there is still an area of concern when exposed to real time application since the mindset of people changes dynamically. Though the three way classification was found to be more reliable, an appropriate algorithm is still a matter of concern. Therefore, in this research work, to provide a solution to all of the above issues, a real time twitter sentimental analysis has been performed with reference to the internet calling in the UAE amidst this COVID pandemic using the Naive Bayes Classifier (NBC) and Recurrent Neural Networks (RNN). Finally, a novel algorithm suitable for the real time analysis will be predicted which will serve as a guideline to know how the mood swings of the people for the Industrial revolution 5.0 in the upcoming years.

II. METHODOLOGY

Under this section, we used dataset from twitter and analyzed it. The model used in the study was NBC and RNN along with clustering of sentimental analysis. The accuracy of the both the method NBC and RNN will be collected and the method which as high accuracy will cluster to find the factors for the sentimental analysis. Using the datasets collected, the training will be done and the corresponding output will be generated.
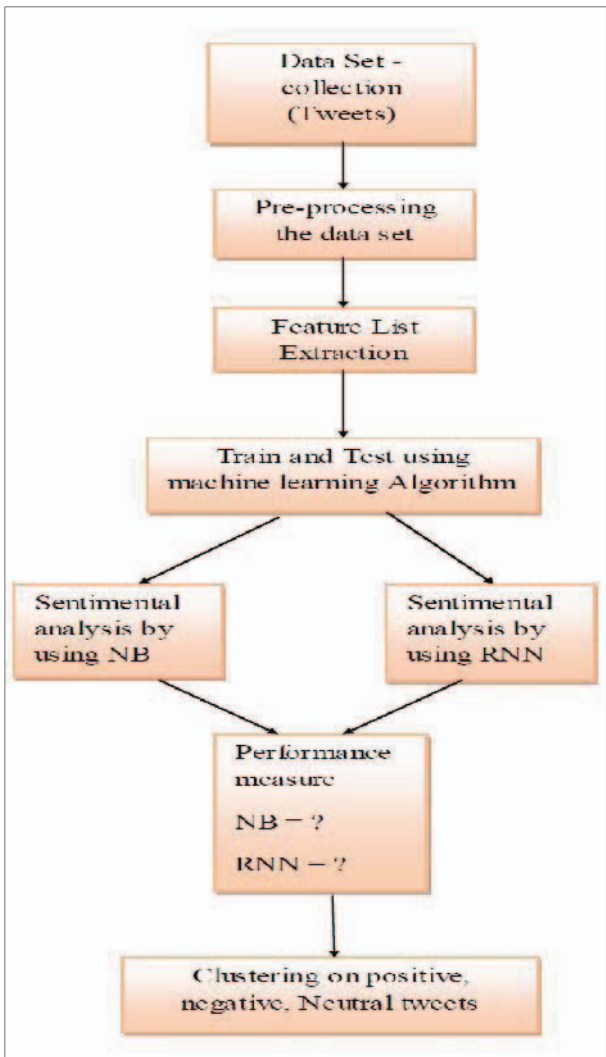

Fig. 1 – Research Procedure

A. Data Pre-processing

It is the first step for data analysis for which the various data has been collected from the twitter from December to July after the onset of COVID pandemic and it has been classified with reference to internet calling. This is an important step because the quality of the data will lead to more reliable results. Preprocessing a Twitter dataset involves a series of tasks like removing all types of irrelevant information like emojis, special characters, and extra blank spaces. It can also involve making format improvements, delete duplicate tweets, or tweets that are shorter than three characters. There

are eight steps for Data Pre – Processing Tweets. They are:

1. Letter casing: Converting all letters to either upper case or lower case.
2. Tokenizing: Turning the tweets into tokens. Tokens are words separated by spaces in a text.
3. Noise removal: Eliminating unwanted characters, such as HTML tags, punctuation marks, special characters, white spaces etc.
4. Stop word removal: Some words do not contribute much to the machine learning model, so it's good to remove them. A list of stop words can be defined by the nltk library, or it can be business-specific.
5. Normalization: Normalization generally refers to a series of related tasks meant to put all text on the same level. Converting text to lower case, removing special characters, and removing stopwords will remove basic inconsistencies. Normalization improves text matching.
6. Stemming: Eliminating affixes (circumfixes, suffixes, prefixes, infixes) from a word in order to obtain a word stem. Porter Stemmer is the most widely used technique because it is very fast. Generally, stemming chops off end of the word, and mostly it works fine. Example: Working -> Work
7. Lemmatization: The goal is same as with stemming, but stemming a word sometimes loses the actual meaning of the word. Lemmatization usually refers to doing things properly using vocabulary and morphological analysis of words. It returns the base or dictionary form of a word, also known as the lemma .Example: Better -> Good.
8. Vectorizing Data: Vectorizing is the process to convert tokens to numbers. It is an important step because the machine learning algorithm works with numbers and not text.

B. Training and Classification

Supervised learning is an important technique for classification problems. In the study we used two supervised tool for sentimental analysis. The two supervised tools said to be Naive Bayes and recurrent neural networks.

III. Implementation and Result

We used Python to train and classify Naïve Bayes and RNN. Out of 833 tweets 250 data were used for training the model. The Fig 2 shows the overall flow of the process
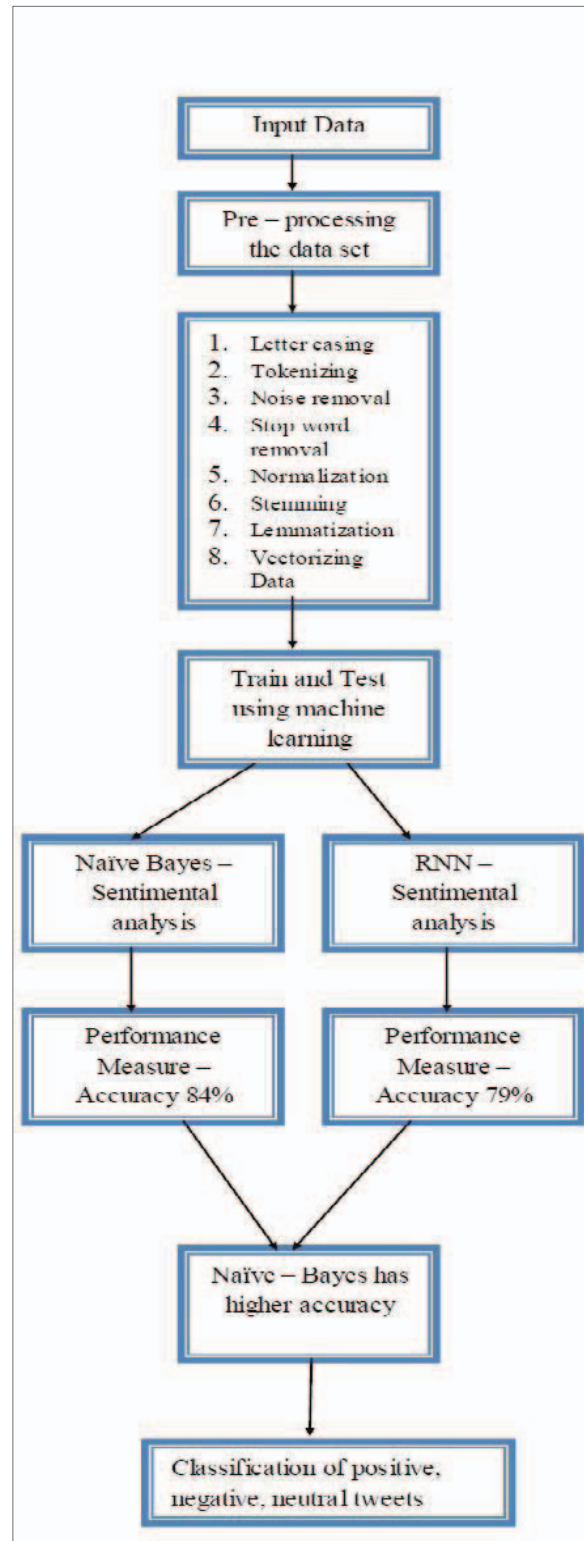

Fig. 2 - Proposed Framework

ANALYSIS:

Table - 1 (NB)	
Performance Measure	
Positive Recall	81.3
Negative Recall	66.06
Positive Precision	84.5
Negative Precision	62.69

Table 1 – Performance Measure of NB

Table - 2 (RNN)	
Performance Measure	
Positive Recall	77.42
Negative Recall	68.23
Positive Precision	79.82
Negative Precision	65.32

Table 2 – Performance Measure of RNN

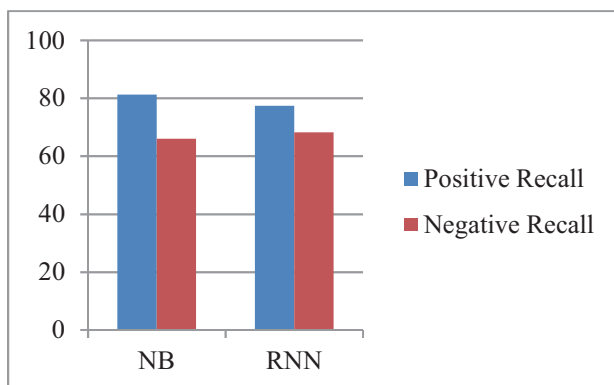| Table - 3	
Methods	
Methods	Accuracy
NB	83
RNN	79

Table 3 – Accuracy of methods

Fig. 3 – Measurement of positive and negative recall of the technique
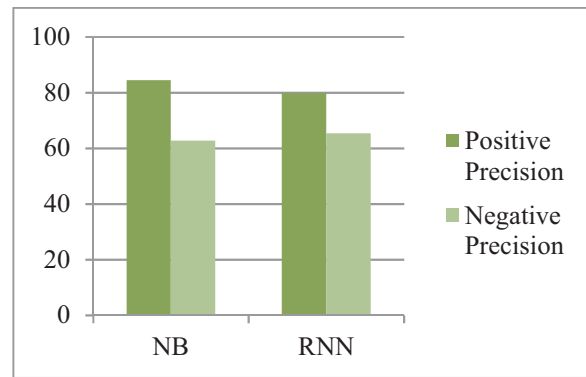

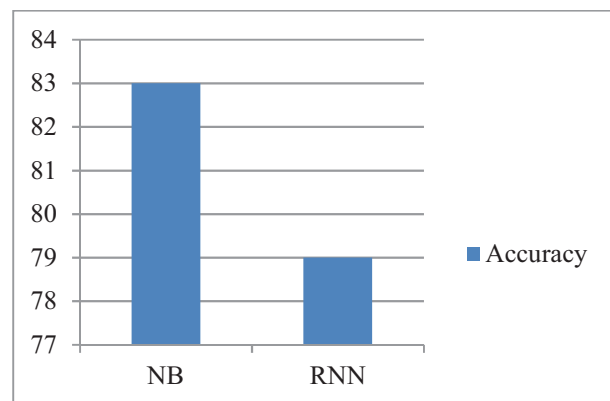**Fig. 4 – Measurement of positive and negative precision of the technique**

Fig. 5 - Accuracy of the methods RNN and NB

From the above Table 1 and Table 2 shows the performance measure of Naive Bayes, RNN in the terms of Recall and Precision. Along with that, Table 3 shows the performance classifier terms of accuracy. Likewise from the Figure 3 and Figure 4 shows the positive, negative recall and positive, negative precision respectively. Figure 5 shows the Accuracy of the methods RNN and NB.

The accuracy of Naive Bayes (NB) was greater than RNN, hence the positive, negative and neutral analysis of NB shown in Fig 6.
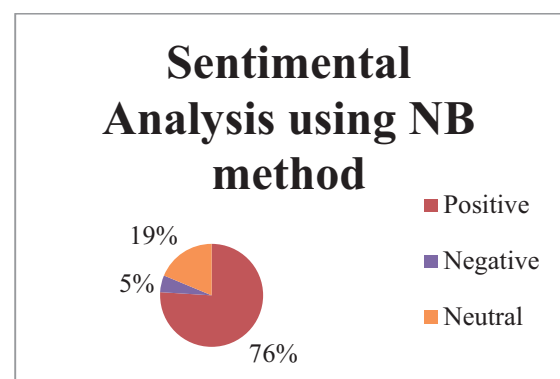

Fig. 6 – NB Sentiment breakdown

Fig. 7 - Word cloud of twitter user profiles under analysis (Positive, Negative and Neutral respectively)

From the analysis, it has been found that the usage of Internet calling was found to be predominant for performing the day-to-day operations in this pandemic scenario. In addition, the results reveal that the maximum number of positive responses has been recorded followed by negative and neutral response (Fig.6) from the tweets and on the other hand, it can be concluded that this culture is mostly agreed by most of the individuals in UAE.

Clustering

On clustering the positive, negative and neutral analysis it is found that people in UAE feels secured, satisfied and internet calling is very useful for them in the prospect of work, education, etc. Negative analysis shows that people were struggle to adapt to the situation and neutral analysis were people how have the capable to adopt to the culture.

IV. CONCLUSION

Based on the above analysis, the results are summarized as follows:

1. The sentimental analysis has been successfully performed for the real time scenario.

2. It has been found that the Naive Bayes algorithm was found to be more accurate than the RNN

3. The results reveal maximum number of positive responses compared to negative and neutral response from the total number of tweets.

4. The proposed model is found to produce better accuracy and performance than the earliest real time analysis.

Finally, it can be concluded that the individuals are showing positive signs for this sudden shift and with an effective strategic plan, the negative response can be sorted which will pave the way for the implementation of Industrial 5.0. Also, the Naive Bayes algorithm is found to be more accurate and the analysis is performed in a short span. However, it is suggested to carry out the analysis with a larger data capturing the mindsets of the global population outside UAE which will provide a holistic view for preparing a strategy for implementing the changes in the mere future.

REFERENCES

[1] Vinay Chamola, Vikas Hassija, Vatsal Gupta, and Mohsen Guizani, "A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact", IEEE Access, vol. 8, 2020, pp. 90225 – 90265.

[2] Weiming Shen, Chen Yang, and Liang Gao, "Address business crisis caused by COVID-19 with collaborative intelligent manufacturing technologies", IET Collaborative Intelligent Manufacturing, vol.2, 2020, pp. 96-99.

[3] Okechukwu S. Okorie, Ramesh Subramoniam, Fiona Charnley, David Widdifield, John Patsavellas and Konstantinos Salonitis, "Manufacturing in the time of COVID-19: An Assessment of Barriers and Enablers", IEEE Engineering Management Review, 2020, pp. 1-1.

[4] 4. Li, Gang, and Fei Liu, "A clustering-based approach on sentiment analysis.Intelligent Systems and Knowledge Engineering (ISKE)", International Conference on. IEEE, 2010.

[5] M. Yuan, Y. Ouyang, Z. Xiong, & H. Sheng, "Sentiment classification of web review using association rules", In Online communities and social computing, 2013, pp. 442–450.

[6] Ye Wu and Fuji Ren, "Learning Sentimental Influence in Twitter", International Conference on Future Computer Sciences and Application, 2011, pp. 119-122.

[7] Rabia Batool, Asad Masood Khattak, Jahanzeb Maqbool and Sungyoung Lee, "Precise Tweet Classification and Sentiment Analysis", 2013 IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS), IEEE Xplore.

[8] Mohammad A. Hassonah, Rizik Al-Sayyed, Ali Rodana, Ala M. Al-Zoubi, Ibrahim Aljarah and Hossam Faris, "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter", Knowledge-Based Systems, 2019.

[9] P.H. Shahana, and Bini Omman, "Evaluation of Features on Sentimental Analysis", Procedia Computer Science, vol. 46, 2015, pp. 1585-1592.

[10] Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, and Sarthak Mendiratta, "Sentiment Analysis of Tweets using Machine Learning Approach", Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), 2018.

[11] Avinash Chandra Pandey Dharmveer Singh Rajpoot, and Mukesh Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method", Information Processing and Management, vol.53, 2017, pp. 764–779.

[12] Shipra Goel, Muskan Banthia, and Adwitiya Sinha, "Modeling recommendation system for real time analysis of social media dynamics", Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), 2018.

[13] Mestan Fırat, and Eliktug, "Twitter Sentiment Analysis, 3-Way Classification: Positive, Negative or Neutral?", IEEE International Conference on Big Data, 2018, pp.2098-2103.

[14] E. Haddi,X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis", Procedia Computer Science, vol. 17, 2013, pp. 26–32.

[15] Veny Amilia Fitri, Rachmadita Andreswari, and Muhammad Azani Hasibuan, "Sentimental analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree and Random Forest Algorithm", Procedia Computer Science, vol. 161, 2019, pp. 765–772