Hotel review analysis for the prediction of business using deep learning approach

6 authors, including:

Md. Sagar Hossen
Sepuluh Nopember Institute of Technology
22 PUBLICATIONS **488** CITATIONS

SEE PROFILE

Tasfia Tabassum
Daffodil International University
5 PUBLICATIONS **88** CITATIONS

SEE PROFILE

Md. Mahfujur Rahman
Queensland University of Technology
28 PUBLICATIONS **617** CITATIONS

SEE PROFILE

Anik Hassan Jony
Daffodil International University
2 PUBLICATIONS **34** CITATIONS

SEE PROFILE

Hotel review analysis for the prediction of business using deep learning approach

Md. Sagar Hossen [1]
Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
sagar15-1504@diu.edu.bd

Tasfia Tabassum [2]
Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
tabassum.tasfiaahmed@gmail.com

Anik Hassan Jony [3]
Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
anik15-1501@diu.edu.bd

Md. Tanvir Islam [4]
Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
mdtanvirislam001@gmail.com

Md Mahfujur Rahman [5]
Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
mrrajuiit@gmail.com

Tania Khatun [6]
Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
tania.cse@gmail.com

Abstract— **Sentiment analysis is a widely used topic in Natural Language Processing that allows identifying the opinions or sentiments from a given text. Social media is the scope for the customers to share their opinion over the products or services as part of customer reviews. Dissect this review has become an important factor for business analysis since online business is exponentially growing in today's techno-friendly competitive market. A large number of algorithms have been found in recent articles. Among those deep learning is an important approach. In the proposed methodology, long short-term memory (LSTM) and Gated recurrent units (GRUs) have been used to train the hotel review data where the accuracy rate of identifying customer opinion is 86%, and 84% respectively. The dataset is also tested by using Naïve Bayes, Decision Tree, Random Forest, and SVM. For Naïve Bayes obtains an accuracy of 75%, for Decision Tree obtains an accuracy of 71%, for Random Forest the accuracy is 82% and for SVM our accuracy result is 71%. Deep learning is used to obtain better business performance and also get the review from customers and also to predict the sentiment about customer review. Our algorithm works properly and gives better accuracy.**

Keywords— Natural Language Processing, Machine Learning, Deep Learning, Artificial Intelligent, LSTM, GRU

I. INTRODUCTION

In the age of modern science, everything is based on online and on the internet. Internet-based shopping has become easier and more popular because of better quality, and fast logistic systems. Internet-based shopping and booking are very comfortable. People can easily make a booking without going outside. The most effective side part of online-based work is that people can give a review. Recognizing reviews allows others to easily understand the emotions of others and obtain the rationality result of different products [10].

In the hotel review, the prediction of business using Deep Learning was analyzed [24]. Many start-up businesses became failure due to lack of analysis and the sentiment of the customer. Sentiment Analysis is the most significant to improve a business site. Here, different type of data from social media as well as from the Hotel Management Website was collected using Unamo tools. And also some supervised and unsupervised data is used to predict the best result. This article will help to improve the business.

At present, online-based opinions can easily analysis with the help of Sentiment Analysis (SA). It is the management of sentiments, different opinions, subjective text, and different emoji used for giving reviews. People can easily get the comprehension information related to people reviews. Mainly Sentiment analysis is one kind of tool that helps to get the public sentiment. By capturing reviews of product or location or person might be found from a different internet-based site like Facebook, Amazon. Sentiment Analysis is used to increase the requirement of analyzing and structuring hidden information which comes from social media in the form of unstructured data. A huge amount of data is used due to the capability of automation and can handle a huge amount of data. A different type of font [23] of review are further classified.

II. RELATED WORK

In our modern science, many authors work on sentiment analysis. R. K. Bakshi addresses his article about sentiment analysis and how to do it on the opinion of humans [1] [22]. Later on, L. Yang analysis on e-commerce review using the deep learning method. The author shows the CNN and GRU technology [2] in that case accuracy was excellent but another e-commerce review is not at good analysis in CNN algorithm. Hemalatha S., the author describes the notion investigation are the audits on eateries about food, administration, cost, and feeling. Machine Learning calculations in the nltk library of python can end up being exceptionally valuable in any such exploration of Natural Language Processing and the library has been utilized broadly in this work [3]. Zeenia Singla also analysis on e-commerce review [8] and she demonstrates her methodology portrayed characterization of surveys as useful to assess the item comprehensively, empowering better-dynamic for customers [4]. These days, social sites like Facebook, Twitter are generally utilized for posting the client's audits about various things, for example, films, news, food, style, governmental issues, and considerably more. Charu Nanda writes in her research sentiment Analysis on film audits in the Hindi language is examined [5]. Online audits received familiarity as individuals are making choices with the assistance of them. In the future most of the choices are based on Artificial Intelligence (AI). Similarly, many creator examinations on client audit in the various cycles. In this manner, the business future can be anticipated [6] [7] [9]. Hui Yuan, the author designed a novel social media analytics framework on top of Apache Spark for predicting and visualizing consumers' opinion orientations based on their relationships with other consumers whose opinion orientations are known. For analysis of customer opinion, they use state-of-the-art collective classification (CC) algorithms. This algorithm considers not only user's local features but also their relational features [10]. Some authors contribute to this analysis system about many other language comments and reviews [14] [15]. They are delectable about e-commerce sentiment analysis [16] [18].

III. METHODOLOGY

In this article, different types of an algorithm are used for the analysis of sentiment of customer. In research work, the analytical part depends on evaluation or developing an algorithm. Though the work, a dataset of the business sector and the dataset from the different website along with some procedure can be developed. In the modern era, natural Language Processing is mostly effective in the machine learning part. In that case sentiment analysis [22] [23] is most important in any business future. In this method, initially collect the data then process the data for our algorithm purpose. These data are trained by deep learning algorithms [19] such as LSTM and GRU where the prediction accuracy in up to 86% in different epoch. Finally, classify the reviews in machine learning algorithms like Naive Bayes, Decision Trees, Random Forest, and SVM and compare their accuracy level. Sentiment Analysis refers to the use of natural language processing, text, and emoji analysis to identify, extract, qualify, and study affective states. It is mainly applied to the

voice of the customer like the customer opinion, survey response which will give in a different type of social media or sites. Initially, collect some data from the dataset which is unsupervised through Unamo tools from social media. Later on, some unsupervised algorithms were used on those dataset for classification, and some supervised algorithms are utilized.
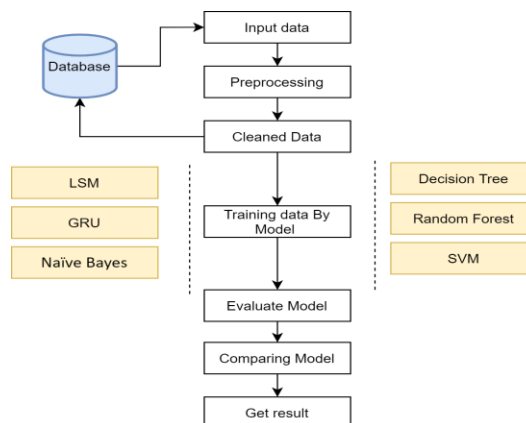

Fig. 1. Block Diagram of Proposed Scheme.

Deep learning aproach is used to calculate the sentinemt and machine learning algorithm is used to calculate better performance analysis.

A. System Design

This system initially inputs data from the database then we start our cleaning process of data. Meanwhile, remove the dataset's superfluous Attribute. Remove all white space, punctuation, links, and email addresses, implying the use of the lemmatization method and stemming. The data are trained by the algorithms such as GRU, and LSTM. Then our trained system is evaluated by testing the data. Finally, the customer's satisfaction was assessed.
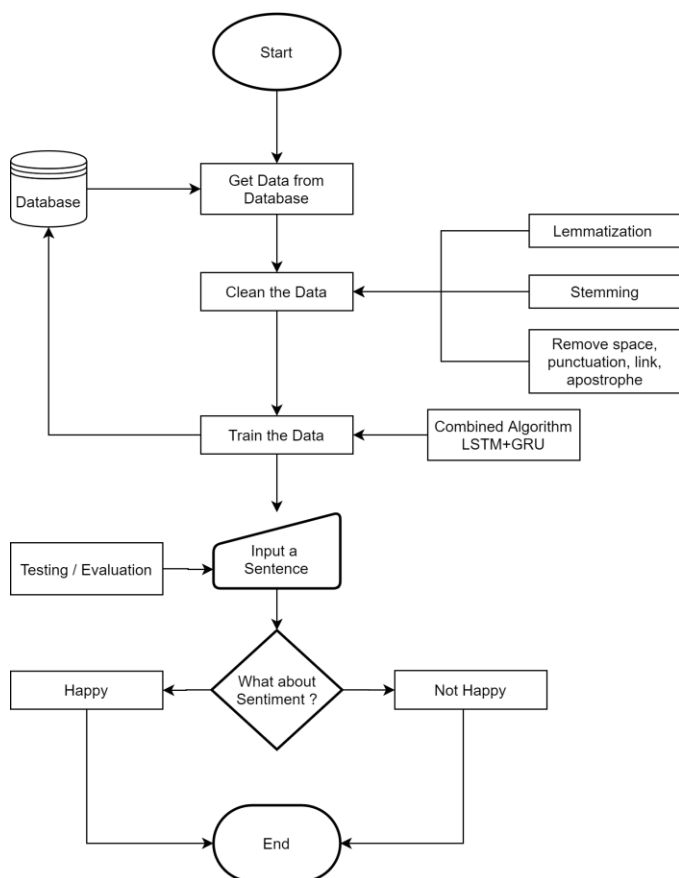
B. *Flow Chart*

Fig. 2.   Unsupervised Algorithm Flowchart

IV.   THE FLOWCHART IN OUR DIAGRAM BEGINS WITH THE DATASET. FOLLOWING THE COLLECTION OF THE DATASET, THE DATA IS SENT FOR CLEANING USING METHODS SUCH AS LEMMATIZATION AND STEMMING. ALL SPACES, PUNCTUATION, LINKS, AND APOSTROPHES WILL BE REMOVED FROM CLEAN DATA. THE CLEAN DATA WILL THEN BE USED TO TRAIN USING A COMBINED ALGORITHM SUCH AS LSTM AND GRU. A SENTENCE WILL BE MANAGED TO INPUT FOR TESTING. THE SENTIMENT OF HAPPY OR NOT HAPPY IS FOUND.DATA DESCRIPTION

The data are collected from the real life hotel booking website review booking.com. It is an Indian Hotel booking data where the customer submit their feedback when they checkout. It's a real data and updated source.

TABLE I.   DATA DESCRIPTION

Attribute	Description
User_ID	Unique Id of Customer
Description	the textual review from customer
Browser_Used	Used browser of Customer
Device_Used	Submitted device
Is_Response	Class Attribute

Initially, deep learning algorithm is used to train the data and to classify our class nltk tools was employed. The actual sentiment analysis is to find the opinions such as positive or happy and negative or unhappy using nltk tool. Machine learning algorithm is used to train the data.

V.   ALGORITHMS

A. *Naïve Bayes Algorithm*

This algorithm is one of the popular classification machine learning algorithms that helps to classify the data, based upon the conditional probability values computation. It is the best algorithm for real-time Prediction,   multi-class prediction, recommendation system, text classification, and sentiment analysis [20] use cases. By using this algorithm, it is easily to implement a huge dataset.

The formula of the algorithm is:

$P(c|x) = (P(x|c) * P(c)) / P(x)$

Here,

$P(c) =$ the prior probability of class.

$P(c|x) =$ the posterior probability.

$P(x) =$ the prior probability of predictor.

$P(x|c) =$ the probability of predictor.

B. *Support vector machine (SVM)*

This algorithm is used for classification problems in Machine Learning. It creates the best line that can convert n-dimensional space into classes that can easily put the new data point in the right categories. It always chooses the extreme point for the best decision boundary. Linear SVM and Non-Linear SVM are two types of SVM. Linear SVM is a dataset that can be classified into two classes using a single straight line. Non-Linear SVM is used when a dataset cannot be classified into two classes using a single.

C. *Random Forest Algorithm*

D. *It is one of the algorithms used for solving the complex problem of classification and regression problems. It combines multiple trees to predict the class of the dataset and the output with a high accuracy which is used to maintain the accuracy when a large proportion of data is missing.Decision tree :*

It is a supervised Machine Learning Algorithm that divided data at each row based on certain rules until the outcome is generated. This algorithm is adopted to solve the problem easily and it is easy to understand.

E. *Short Term Memory:*

F.   *LSTM is one kind of recurrent neural network which used in the field of deep learning. To work with data for a long period, LSTM is used to retain the information. Processing, predicting, and classifying based on time series data is performed by LSTM.Gated recurrent unit*

It is one kind of Long Term Memory Algorithm with forget Gate. GRU gives the best performance for the less frequent

dataset. In recurrent neural networks when vanishing gradient problems are shown GRU is used to solve a huge dataset.

VI. ALGORITHM COMPARISION

S. N o	Long Short Term Memory	Gated recurren t unit	Random forest	Decision tree	Naïve Byes
1.	It used for the vanishing gradient problem and learn long term	The GRU is like a long short – term memory with forget Gate	Set of multiple decision trees make random forests	They are computat ionally faster	Variable is independ ent from the others
2.	Hidden layers are:30	Hidden layers:2.5	It prevents overfittin g by creating trees on random subsets	It may suffer from overfittin g	It can work with low amount of training work
3.	The LSTM cell maintains a cell state that is read from and written	The LSTM cell maintains a cell state that is read from and written	It is difficult to interpret	It is easy to interpret.	Categoric al features cannot work with this algorithm

TABLE II. COMPARISION AMONG ALGORITHMS

VII. RESULT AND ANALYSIS

TABLE III. RESULT

Algorithm	Accuracy
LSTM	86 %
GRU	84 %
Naive Bayes	75%
Decision Tree	71%
Random Forest	82%
SVM	71%

In this analysis, the results are obtained from Four Supervised Algorithm such as Naive Bayes, Random Forest, Decision Tree, and SVM which are regulated calculations. The Jupiter Notebook and Python Programming Language are used for our classification and achieves the accuracy of 86% and these models are such a great amount of proficient for this dataset. The informational collection was 63%. Finally, testing aids to achieve an accuracy of 37%. Furthermore, Deep Learning method is used for testing new data set and the LSTM and GRU unsupervised algorithm was used so there was no decision class.
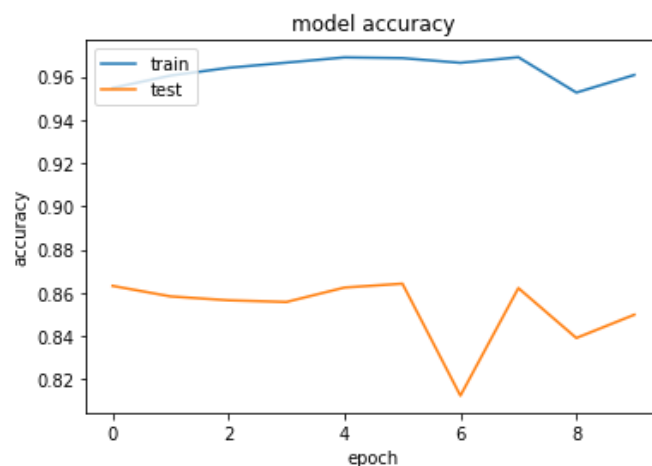

Fig. 3. Model Accuracy

In the training method, our system accuracy is 96%, and in the testing method, our system accuracy is 86%.
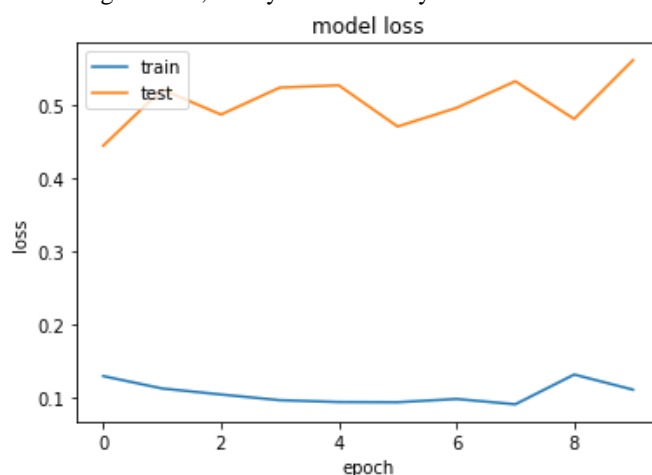

Fig. 4. Model loss- valid graph.

The above-mentioned diagram depicts the loss and validity in training time. The data loss is less than our validity.

A. Comparision Table

TABLE IV. COMPARISION RESULT

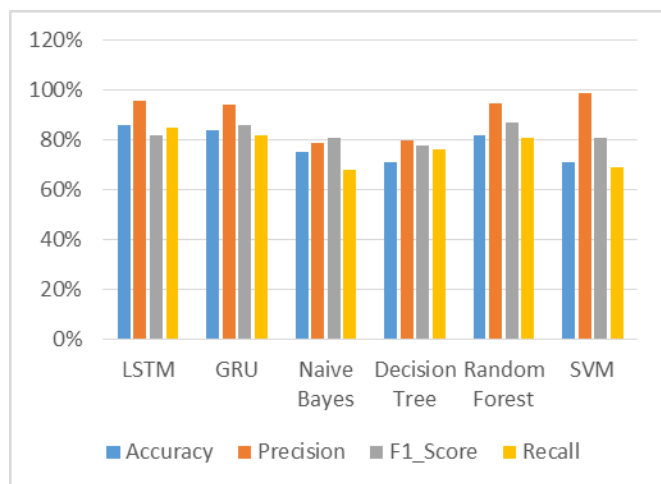| Algorithm	Accuracy	Precision	Recall	F1_Score
LSTM	86%	0.96	0.85	0.82
GRU	84%	0.94	0.82	0.86
Naive Bayes	75%	0.79	0.68	0.81
Decision Tree	71%	0.80	0.76	0.78
Random Forest	82%	0.95	0.81	0.87
SVM	71%	0.99	0.69	0.81

Fig. 5.   Comapision of each model

VIII.   FUTURE WORK

Everyone in today's world knows how far the world has come. Our country is moving forward by maintaining that continuity. New innovations in technology have made human work easier and more dynamic. So, today they can do any work in a short time through technology very easily. The most popular site in this modern age is online marketing. Further, this will be accepted by the people. In the online hotel booking system, people can easily book the hotel of their choice. Based on some algorithms, it is planned to add more features in the future. This will make our website more secure, acceptable, and popular to everyone.

IX.   CONCLUSION

The present age is the modern age. Everything in the age is now technology-dependent and every person in the country is able to familiarize themselves with this technology. With the help of that technology, online marketing has become popular in today's world, which has easily become popular among people. People are now getting a lot of things through their hands very easily. One part of online marketing is the online hotel booking system. With this people can easily pre-book the hotel of their choice and they can easily go to their hotel without having to bother to search for the place. It has become the most popular among people and this led to an increase in the number of people traveling around. And at the same time, they can able to view different beautiful places of the world by taking advantage of this hotel booking. In the future, many more features can be added to the project and ensure more popular things.

References

[1]   R. K. Bakshi, N. Kaur, R. Kaur and G. Kaur, "Opinion mining and sentiment analysis," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 452-455.

[2]   L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, vol. 8, pp. 23522-23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[3]   H. S. and R. Ramathmika, "Sentiment Analysis of Yelp Reviews by Machine Learning," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 700-704, doi: 10.1109/ICCS45141.2019.9065812.

[4]   Z. Singla, S. Randhawa and S. Jain, "Statistical and sentiment analysis of consumer product reviews," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6, doi: 10.1109/ICCCNT.2017.8203960.

[5]   C. Nanda, M. Dua and G. Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 1069-1072, doi: 10.1109/ICCSP.2018.8524223.

[6]   B. Seetharamulu, B. N. K. Reddy and K. B. Naidu, "Deep Learning for Sentiment Analysis Based on Customer Reviews," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-5, doi: 10.1109/ICCCNT49239.2020.9225665.

[7]   Rahul, V. Raj and Monika, "Sentiment Analysis on Product Reviews," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2019, pp. 5-9, doi: 10.1109/ICCCIS48478.2019.8974527.

[8]   Y. Saito and V. Klyuev, "Classifying User Reviews at Sentence and Review Levels Utilizing Naïve Bayes," 2019 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang Kwangwoon_Do, Korea (South), 2019, pp. 681-685, doi: 10.23919/ICACT.2019.8702039.

[9]   A. Salinca, "Business Reviews Classification Using Sentiment Analysis," 2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, 2015, pp. 247-250, doi: 10.1109/SYNASC.2015.46.

[10] ChhayaChauhan, SmritiSehgal "SENTIMENT ANALYSIS ON PRODUCT REVIEWS", International Conference on Computing, Communication and Automation (ICCCA2017) ISBN:978-1-5090-6471-7/17/$31.00 ©2017 IEEE

[11] R. Hu, L. Rui, P. Zeng, L. Chen and X. Fan, "Text Sentiment Analysis: A Review," 2018 IEEE 4th International Conference on Computer and Communications (ICCC), Chengdu, China, 2018, pp. 2283-2288, doi: 10.1109/CompComm.2018.8780909.

[12] M. A. Islam, M. S. Islam, M. S. Hossen, M. U. Emon, M. S. Keya and A. Habib, "Machine Learning based Image Classification of Papaya Disease Recognition," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2020, pp. 1353-1360, doi: 10.1109/ICECA49313.2020.9297570.

[13] N. Srivats Athindran, S. Manikandaraj and R. Kamaleshwar, "Comparative Analysis of Customer Sentiments on Competing Brands using Hybrid Model Approach," 2018 3rd International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2018, pp. 348-353, doi: 10.1109/ICICT43934.2018.9034283.

[14] K. S. Sabra, R. N. Zantout, M. A. E. Abed and L. Hamandi, "Sentiment analysis: Arabic sentiment lexicons," 2017 Sensors Networks Smart and Emerging Technologies (SENSET), Beirut, 2017, pp. 1-4, doi: 10.1109/SENSET.2017.8125054.

[15] J. Ding, H. Sun, X. Wang and X. Liu, "Entity-Level Sentiment Analysis of Issue Comments," 2018 IEEE/ACM 3rd International Workshop on Emotion Awareness in Software Engineering (SEmotion), Gothenburg, 2018, pp. 7-13.

[16] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp. 1275-1279, doi: 10.1109/ICIRCA.2018.8597286.

[17] A. Razzaq et al., "Text sentiment analysis using frequency-based vigorous features," in China Communications, vol. 16, no. 12, pp. 145-153, Dec. 2019, doi: 10.23919/JCC.2019.12.011.

[18] E. M. Alshari, A. Azman, S. Doraisamy, N. Mustapha and M. Alkeshr, "Effective Method for Sentiment Lexical Dictionary

Enrichment Based on Word2Vec for Sentiment Analysis," 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), Kota Kinabalu, 2018, pp. 1-5, doi: 10.1109/INFRKM.2018.8464775.

[19] M. S. Hossen, I. Haque, M. S. Islam, M. T. Ahmed, M. J. Nime and M. A. Islam, "Deep Learning based Classification of Papaya Disease Recognition," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 945-951, doi: 10.1109/ICISS49785.2020.9316106.

[20] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985884.

[21] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," in IEEE Access, vol. 7, pp. 43749-43762, 2019, doi: 10.1109/ACCESS.2019.2907772.

[22] K. Zvarevashe and O. O. Olugbara, "A framework for sentiment analysis with opinion mining of hotel reviews," 2018 Conference on Information Communications Technology and Society (ICTAS), Durban, 2018, pp. 1-4, doi: 10.1109/ICTAS.2018.8368746

[23] Vijayakumar, T., and Mr R. Vinothkanna. "Capsule Network on Font Style Classification." Journal of Artificial Intelligence 2, no. 02 (2020): 64-76.

[24] Sathesh, A. (2019). PERFORMANCE ANALYSIS OF GRANULAR COMPUTING MODEL IN SOFT COMPUTING PARADIGM FOR MONITORING OF FETAL ECHOCARDIOGRAPHY. Journal of Soft Computing Paradigm (JSCP), 1(01), 14-23.