2022 25th International Conference on Computer and Information Technology (ICCIT)
17-19 December 2022, Cox's Bazar, Bangladesh

LSTM-ANN Based Price Hike Sentiment Analysis from Bangla Social Media Comments

Sovon Chakraborty
Department of Computer Scinece and Engineering
School of Data and Sciences
BRAC University
Dhaka-1212, Bangladesh
sovon.chakraborty@g.bracu.ac.bd

Muhammad Borahn Uddin Talukdar
Department of Computer Science and Engineering
Dhaka-1209, Bangladesh
talukder4466@diu.edu.bd

Muhammed Yaseen Morshed Adib
Department of Computer Scinece and Engineering
School of Data and Sciences
BRAC University
Dhaka-1212, Bangladesh
yaseen.morshed.adib@bracu.ac.bd

Sowmen Mitra
Department of Artificial Intelligence
Hebei University of Technology
Beijing-065001, China
sowmenmitra7@gmail.com

Md. Golam Rabiul Alam
Department of Computer Scinece and Engineering
School of Data and Sciences
BRAC University
Dhaka-1212, Bangladesh
rabiul.alam@g.bracu.ac.bd

2022 25th International Conference on Computer and Information Technology (ICCIT) | 979-8-3503-4602-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCIT57492.2022.10055290

***Abstract*— Price hike has always been a substantial concern for people all over the world. The crisis gets more conspicuous, and people find themselves more confounded when even the bare minimum of expenses still exceeds the amount they can get to earn. This tension tends to invite chaos in society as the number of people affected increases. Bangladesh is currently undergoing a formidable wave of price hikes. People have been expressing mixed reactions on social media regarding this issue. Hence, understanding the overall public sentiment can be crucial for policymaking and preventing chaos in society. This study utilizes social media comments for analyzing underlying sentiments. Data were collected from the Facebook pages of some popular Bangladeshi media for this purpose, and thereby a specialized dataset was constructed. The dataset contains 2000 public comments annotated with three polarity values- positive, negative, and neutral. A hybrid LSTM-ANN deep architecture has been exploited in this research. The model outperforms other state-of-the-art models in terms of less trainable parameters along with an F1-score of 88.47%.**

Keywords—Sentiment, Sentiment Analysis, LSTM, Price hike, social media, online news portal

I. INTRODUCTION

The price hike is a significant increase in the expenses related to products or services. It has always been an indispensable concern for people with limited earnings. Crimes such as dacoity, snatching, and other unlawful activities increase when people of lower socio-economic status welter in the meteoric rise in prices [1]. An increased price of different items has different degrees of impact. Recently Bangladesh has experienced a price hike after a sudden increase in fuel prices. Fuel is such an item that can instantaneously contribute to increased prices of anything and everything [2].

Social media has turned out to be a place where people can interact with each other by exchanging what they got to say on different topics [3]. Facebook pages of renowned print and electronic media, along with their respective online portals, can serve as a great compendium of different kinds of public comments, which can further be utilized to analyze public sentiments regarding a matter.

Sentiment analysis has received a lot of research attempts where appropriate Machine Learning and Deep Learning Algorithms have been exploited [4-8]. As the literature does not contain any endeavor to analyze public sentiments regarding price hikes, the authors of this paper resolved to contribute in this area. Data used in this research were gathered from the Facebook pages of some popular print and electronic media or their corresponding online portals. Section I addresses the motivation behind this research and the contributions that the research makes. Section II analyzes what the literature contains and where can we still contribute. The proposed methodology, data preprocessing techniques, and a deep dive into the sentiment analysis process, all these major aspects have been discussed in Section III. Section IV demonstrates the results achieved by the proposed model along with some comparative analysis.

A. Motivation

Bangladesh is currently experiencing a massive price hike which affected thousands of people all over the country. Although it is obvious that some people are necessarily inflicted with a lower quality of life as a result, however, the overall picture is still uncaptured. Given the number of people that express their opinion in the comment sections of different online news articles, a plethora of public reactions can be gathered from there for sentiment analysis and categorization of different kinds of reactions regarding the recent price hike. Hence, we built a dataset

comprising 2000 social media comments that are later used for this study. Considering the immense success of different deep learning architectures in the field of Natural Language Processing, we have employed a deep hybrid model comprising two deep learning architectures.

B. Contributions

The contribution of this paper can be summarized as follows:

1. Building a specialized dataset of 2000 instances containing public reactions regarding the recent price hikes in Bangladesh.
2. Proposing a hybrid deep learning architecture for sentiment analysis that outperforms the previous state-of-the-art models.
3. The proposed model can be utilized to analyze how the mass people are reacting to price hikes. The model categorizes public sentiments under three categories: Positive, Negative, and Neutral. The findings can further be useful in understanding the correlation between price hikes and different socio-economic aspects like crime, poverty, quality of life, etc. The analysis can also facilitate studies concerning changes in the pattern of human behavior over time when the price hike is a recurrent phenomenon and the earnings are not commensurate with the expenses.

II. Literature Review

Sentiment analysis has gained the attention of so many researchers in recent years. Although the literature is stuffed with lots of sentiment analysis works based on different languages, the contribution in the Bengali language is still inadequate. The previous research works exploited different machine learning and deep learning algorithms to yield desirable outcomes. Deep learning architectures such as Convolutional Neural Networks (CNN), Transformers, Artificial Neural Networks (ANN), etc are renowned for sentiment analysis tasks.

Sentiment analysis on the tourism sector based on customer reviews has been performed by Luo ***et al.*** [9]. The researchers employed BiLSTM architectures, where two polarities have been considered, namely positive and negative, no neutral sentiment has been considered here. During the Covid-19 pandemic, public sentiment were analyzed based on twitter data where people expressed their opinions regarding Moderna, AstraZeneca and Pfizer [10]. The authors considered three polarities, and KNN algorithm was utilized to render the outcome. Bidirectional Encoder Representations from Transformers (BERT) and Electra are two effective models used for sentiment analysis in Bengali. Text document classification was performed by Rahman **et al**. using three publicly available datasets [11]. The researchers have applied modified BERT models in order to achieve decent accuracy. The model outperforms the state of art models in terms of accuracy on the first two datasets, while yielding unsatisfactory performance on the validation set of the third dataset.

Sentiment analysis based on the Urdu multimodal dataset using LSTM is performed by Sehar ***et al***. [12]. Multimodal data includes text, audio, video, and more. The model shows an accuracy of 84.32% in the validation set. The model achieves lower accuracy when using with unimodal data. Besider other efficient models, some researched utilized Bi-directional LSTM (BiLSTM) for their sentiment analysis research [13,14,15]. But BiLSTM often suffers from poor predictions. Sushmitha *et al.* propose a Bidirectional LSTM model for analyzing aspect-based data, which produces an estimable accuracy [16]. The polarity of the dataset is appropriately understood with a training ratio of 70% and 30% respectively to the training and validation set. This model achieves significant results in terms of accuracy. To improve the accuracy hybrid models are also applied in the literature. CNN-LSTM based model is applied in order to predict the server load where the model shows significant efficiency in order to forecast the future load [17]. ANN-LSTM and CNN-LSTM models also act as good classifiers for predicting electricity consumption [18]. The number of trainable parameters is a matter of concern here.

Despite having numerous sentiment analysis research, the literature shows no endeavors on public sentiment analysis regarding price hikes. Hence the authors intended to contribute to this area with a robust hybrid deep architecture composed of LSTM and ANN.

III. Proposed Methodology

The primary objective of the researchers is to determine the polarity values of each instance of data in terms of Positive, Negative, and Neutral. All data are collected in Bengali text which contains different stopwords, punctuations, and special characters, which need to be taken care of first before feeding the dataset to the proposed model. The data are then tokenized to develop the context. The primary purpose of tokenization is to transform the data into small manageable parts, called tokens.

To extract the necessary features, the authors have used 5000 maximum feature numbers. Tokenizer fits all the sentiment data into a proper sequence.

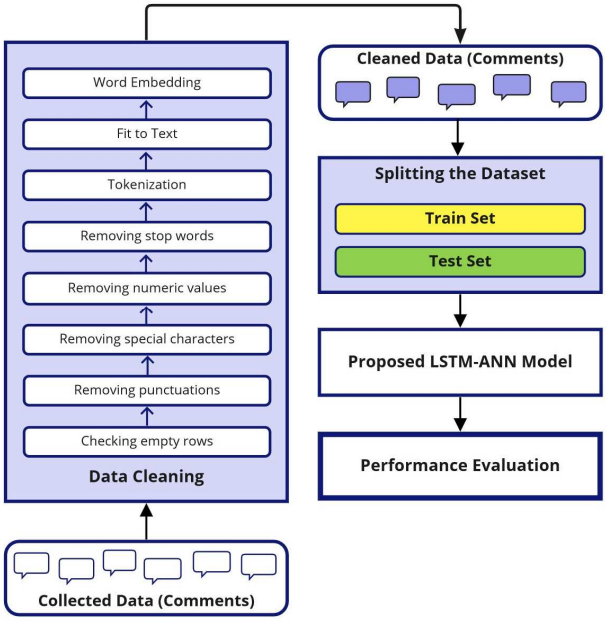

Fig. 1. Proposed methodology

Later the converted sequence is added into a pad sequence. Once the data are cleaned and the dataset is ready for action, we divided the instances into a train-test ratio of 70% to 30%. Refined data is now to be fed to the proposed LSTM-ANN model, and the performance is analyzed to get a picture of how the model performs. The proposed methodology has been demonstrated in Fig. 1.

A. Dataset Description

The data used in this research are mostly outsourced to the comment sections of the Facebook pages of some renowned print and electronic media. We targeted the comment sections of the recently published articles related to the price hike. Table I represents some data instances along with their respective sources.

While collecting data, we encountered some data that are transliterated, which means the comments are written using the English alphabet, to represent Bengali phonetics. An example of such a sentence could be "Lobon er dam eto kno hbe !!!". This group contains approximately 30% of the comments and is excluded from the dataset.

TABLE I: Data samples with Sources

Comments	Sources
নিত্য প্রয়োজনীয় জিনিসপত্রের যে দাম না খেয়ে মরা ছাড়া উপায় নেই।	Collected from the comment section of BBC News Bangla Facebook page.
এতো দাম হতে পারে কখনও কল্পনাও করিনি।	Collected from the comment section of The Business Standard Facebook page.
জনগণের ক্রয়ক্ষমতা দিনদিন কমছেই	Collected from the News article of The Prothom Alo.
অমানুষ মজুদদারদেরকে দৃষ্টান্তমূলক শাস্তির আওতায় আনা হোক।	Collected from the comment section of The Ekattor TV web portal.
না খেয়ে মরা ছাড়া উপায় দেখছিনা।	Collected from the comment section of The Somoy TV Facebook Page.

B. Annotation of the collected Dataset

The primary purpose of annotations is to help the NLP models learn about some key phrases that each comment comprises. This also helps in determining the parts of speech of a particular comment. The dataset has a column that contains respective polarity values associated with each comment. Three polarity values were considered. These are Positive, Negative, and Neutral. With a view to assigning appropriate polarity value for each comment, the authors involved 4 students from the European University of Bangladesh and 3 students from Stamford University of Bangladesh. These 7 students participated in voting for determining the ultimate polarity values for the comments. The following comment is taken from the comments section of a shared article regarding the rise in fuel prices on the BBC News Bangla Facebook page-

"গাড়ীভাড়া বৃদ্ধি পাওয়া দেখে মনে হচ্ছে সাইকেল কিনতে হবে"

This comment above was provided to that 7 students. The authors asked them to assign the polarity of this comment. The result is given in Table III. Out of the 7 participants, 5 votes were cast in favor of Negative polarity, and 2 votes were cast in favor of Neutral polarity. The final polarity value determined for this comment is Negative, as the majority of the votes were cast in the favor of Negative polarity. The odd number of participants eliminates the possibility of a tie. The whole dataset was annotated first using this method. Once the annotation is completed, the dataset is rechecked by the authors to make sure that the assigned polarities make sense.

TABLE II: Polarity counts regarding each polarity type

Polarity	Data Instances
Positive	253
Negative	1359
Neutral	388
Total	**2000**

The final dataset contains 2000 comments annotated by three different polarity values. Table II demonstrates the polarity counts regarding each polarity type.

C. Data Preprocessing

To make the best out of the dataset, some preprocessing tasks are required. All the special characters, punctuations, digits, and emojis are removed from the dataset first.

"সংসার চালানো অনেক কষ্টের হয়ে গেছে।"

The above sentence contained punctuation and stopwords. They need to be removed first.

"সংসার চালানো অনেক কষ্টের হয়ে গেছে"

Stopwords also need to be discarded. The resultant sentence looks like this-

"সংসার চালানো কষ্টের"

To convert the text into tokens, Keras.tokenizer is applied. Following the tokenization, to convert the token into a sequence, Keras Pad_sequence is used. Based on the maximum feature number, the tokenizer tokenizes the text. After converting it into a sequence, it is added to the pad_sequence. The sentiment level is also factorized to make it more understandable. As a result, Positive becomes 0, Negative becomes 1 and Neutral becomes 2. So, after factoring the sentiment level looks like-

[1,1,1,1,0,0,1,2……..]

Unique numbers are assigned based on the tokenizer function where max_word is 5000. So at max, 5000 unique numbers can be assigned against words. To create associations of words and numbers, the fit_on_text function is used. After completion of all the preprocessing techniques, the sentence looks like the following. To represent the word in the lower dimension finally the word_embedding function has been used.

" সংসার চালানো কষ্টের "

[220, 3212, 1372]

Where 220 is a unique number assigned against 'সংসার'. Before feeding the dataset to the model, all the data need to be preprocessed.

D. The Proposed Model

Fig. 2 shows the detailed workflow of the proposed model. After the completion of all data preprocessing, the dataset is passed to the model.

Sequential() initiates the model creation and the words are embedded in the next phase. To drop the entire feature map instead of individual elements of the network, SpartialDropout1D () has been applied. The Dataset is then fed to an LSTM layer. LSTM is a model that provides feedback connection. In this case, LSTM works like a Recurrent neural network (RNN).

LSTM is comprised of the following elements: Cell, Forget gate, Input gate, Output gate. The input gate has the capability of controlling the information that will be passed inside the cell of the memory based on previous output and measurement of current data. The update of the memory cell is controlled by the forget gate. The output gate is the decision maker that will decide which information will be carried out to the next step. LSTM block is mainly built for maintaining the text sequence. The weight matrix of the LSTM is randomly initialized. All the gates and cells get updated after each information processing cycle.

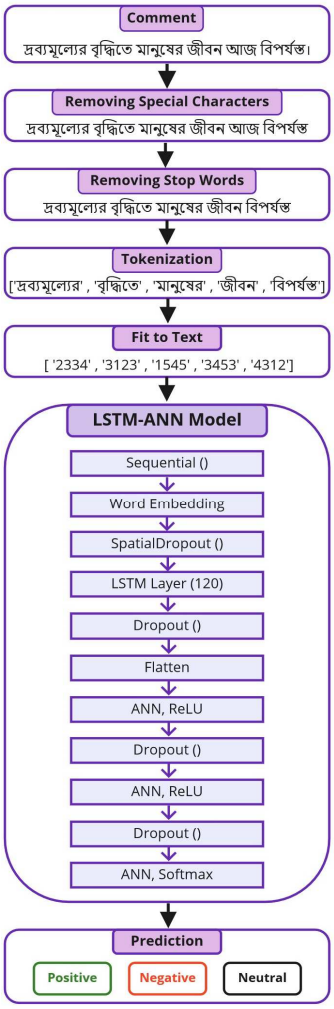

Fig. 2. Details of the proposed model

Deep learning models have a common tendency of getting overfitted. To ignore overfitting, the authors applied Dropout(). After getting processed by the LSTM and Dropout layer, the dataset is passed to a dense layer where ReLU activation function has been used. Table III shows the parametric details of the proposed model.

TABLE III: Parametric Details of the Proposed Architecture

Name of the Parameters	Values
Embedding vector length	32
Spatial Dropout	0.3
Number of Epochs	10
Activation Function	ReLU and Softmax
Optimizer	Adam Stochastic Gradient Descent
Loss Function	Binary Cross-Entropy
Recurrent dropout	0.32
Number of LSTM blocks	120
Number of trainable parameters	228,392

IV. Result Analysis

To evaluate the performance of the proposed model authors have observed the accuracy shown by the proposed LSTM-ANN model for each epoch. Fig. 3 shows the accuracy percentage of the LSTM-ANN model in the training and validation phases.

We examined different deep learning architectures to find out which performs best in this scenario. Fig. 4 represents the F1-score comparison of the proposed model with three other efficient deep learning architectures. It shows that, the proposed LSTM-ANN model outperforms other models in terms of F1-score.

The model is then compared with some previous approaches attempted by researchers [20-22]. Fig. 5 constitutes a summary of the model's accuracy with other Machine learning and deep learning approaches. It has been observed that- the proposed model has the highest accuracy when compared to other efficient models.

Fig. 6 shows a comparative pictorial analysis of the trainable parameters for different models. This time the same set of models was considered as well to compare the number of trainable parameters that each model has. The proposed architecture secured the lowest number of trainable parameters.

Despite having appreciable success, the model suffers from some apparent shortcomings as well. While the model can classify comments with Positive polarity pretty accurately, it however struggles a bit to label some comments as either Negative or Neutral.

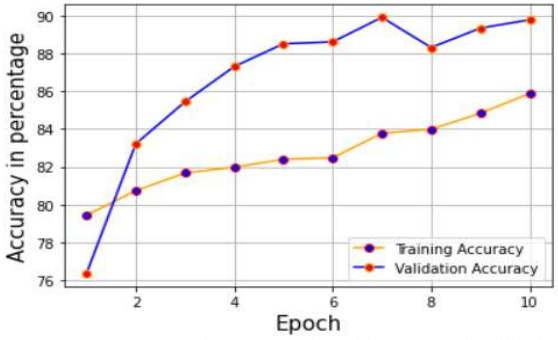

Fig. 3. Accuracy Comparison Graph of Training Set and Validation Set

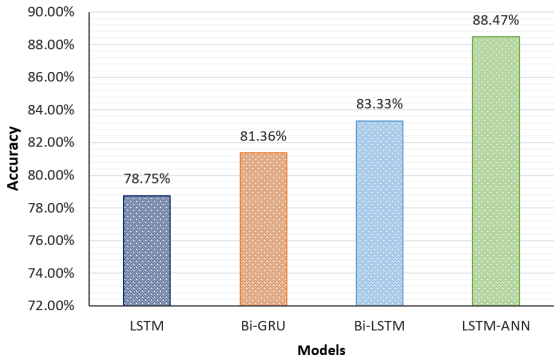

Fig. 4. F1-score comparison of the Proposed model with different other architectures

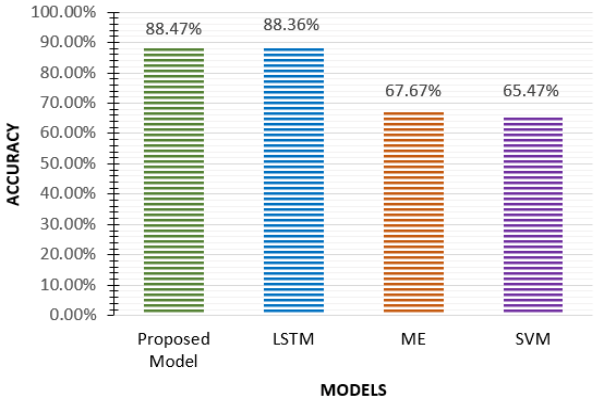

Fig. 5. F1-Score comparison of the proposed model with state-of-the-art models

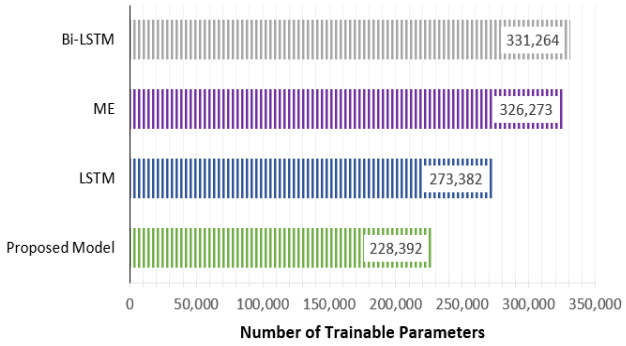

Fig. 6. Trainable parameters of the proposed model compared with that of the previously studied models

The reason why this issue remains is, some comments can be interpreted as both Negative and Neutral. Hence, sometimes the model gets confused and picks up one for the expected other.

V. CONCLUSION

This research utilizes a hybrid LSTM-ANN model to analyze the public sentiment regarding the recent price hikes in Bangladesh. A specialized dataset has been built, validated, and fed to the model. The model outperforms other state-of-the-art models in terms of a higher F1-score and a less number of trainable parameters. In the near future, the authors will work on predicting different kinds of socio-economic crises given some phenomena that they follow.

REFERENCES

[1] A. Naz, Hafeez-ur-Rehamn C., M. Hussain, U. Daraz, and W. Khan, "Inflation: the social monster socio-economic and psychological impacts of inflation and price hike on poor families of district Malakand, Khyber Pakhtunkhwa, Pakistan," in International Journal of Business and Social Science 2, no. 14, 2012.

[2] S. I. Ocheni, "Impact of fuel price increaseon the Nigerian economy," Mediterranean Journal of Social Sciences 6, no. 1 S1: 560-560, 2015.

[3] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das, and T. Mittra, "A deep learning approach to detect abusive bengali text," in 2019 7th International Conference on Smart Computing & Communications (ICSCC), pp. 1-5. IEEE, 2019.

[4] M. T. Akter, M. Begum, and R. Mustafa, "Bengali sentiment analysis of E-commerce product reviews using K-nearest neighbors," in 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), pp. 40-44. IEEE, 2021.

[5] N. R. Bhowmik, M. Arifuzzaman, and M. R. H. Mondal, "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms," Array 13: 100123, 2022.

[6] N. Romim, M. Ahmed, H. Talukder, and S. Islam, "Hate speech detection in the bengali language: A dataset and its baseline evaluation," in International Joint Conference on Advances in Computational Intelligence, pp. 457-468. Springer, Singapore, 2021.

[7] S. K. Rahut, R. Sharmin, and R. Tabassum, "Bengali Abusive Speech Classification: A Transfer Learning Approach Using VGG-16," in 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), pp. 1-6. IEEE, 2020.

[8] A. M. Ishmam, and S. Sharmin, "Hateful speech detection in public facebook pages for the bengali language," in 2019 18th IEEE international conference on machine learning and applications (ICMLA), pp. 555-560. IEEE, 2019.

[9] J. Luo, S. Huang, and R. Wang, "A fine-grained sentiment analysis of online guest reviews of economy hotels in China," Journal of Hospitality Marketing & Management 30, no. 1: 71-95, 2022.

[10] F. M. J. M. Shamrat, Sovon Chakraborty, M. M. Imran, Jannatun Naeem Muna, Md Masum Billah, Protiva Das, and O. M. Rahman, "Sentiment analysis on Twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm," Indonesian Journal of Electrical Engineering and Computer Science 23, no. 1: 463-470, 2021.

[11] M. M. Rahman, M. A. Pramanik, R. Sadik, M. Roy, and P. Chakraborty, "Bangla documents classification using transformer based deep learning models," in 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), pp. 1-5. IEEE, 2020.

[12] U. Sehar, S. Kanwal, K. Dashtipur, U. Mir, U. Abbasi, and F. Khan, "Urdu Sentiment Analysis via Multimodal Data Mining Based on Deep Learning Algorithms," IEEE Access 9: 153072-153082, 2021.

[13] R. Jin, Z. Chen, K. Wu, M. Wu, X. Li, and R. Yan, "Bi-LSTM-Based Two-Stream Network for Machine Remaining Useful Life Prediction," IEEE Transactions on Instrumentation and Measurement 71: 1-10, 2022.

[14] Y. L. Yang, G. C. Wan, and M. S. Tong, "A Novel Wireless Propagation Model Based on Bi-LSTM Algorithm," IEEE Access 10: 43837-43847, 2022.

[15] G. Tian, Q. Wang, Y. Zhao, L. Guo, Z. Sun, and L. Lv, "Smart contract classification with a bi-LSTM based approach," IEEE Access 8: 43806-43816, 2020.

[16] M. Sushmitha, K. Suresh, and K. Vandana, "To Predict Customer Sentimental behavior by using Enhanced Bi-LSTM Technique," in 2022 7th International Conference on Communication and Electronics Systems (ICCES), pp. 969-975. IEEE, 2022.

[17] X. Shao, C. Kim, and P. Sontakke, "Accurate deep model for electricity consumption forecasting using multi-channel and multi-scale feature fusion CNN–LSTM," Energies 13, no. 8: 1881, 2020.

[18] K. Ijaz, Z.Hussain, J. Ahmad, S. F. Ali, M. Adnan, and I. Khosa, "A Novel Temporal Feature Selection Based LSTM Model for Electrical Short-Term Load Forecasting," IEEE Access 10: 82596-82613, 2022.

[19] E. Patel, and D. S. Kushwaha, "A hybrid CNN-LSTM model for predicting server load in cloud computing," The Journal of Supercomputing 78, no. 8: 1-30, 2022.

[20] M. Hoq, P. Haque, and M. N. Uddin, "Sentiment analysis of bangla language using deep learning approaches," In International Conference on Computing Science, Communication and Security, pp. 140-151. Springer, Cham, 2021.

[21] K. Sarkar, and M. Bhowmick, "Sentiment polarity detection in bengali tweets using multinomial Naïve Bayes and support vector machines," in 2017 IEEE Calcutta Conference (CALCON), pp. 31-36. IEEE, 2017.

[22] N. Banik, M. H. H. Rahman, S. Chakraborty, H. Seddiqui, and M. A. Azim, "Survey on text-based sentiment analysis of bengali language," In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1-6. IEEE, 2019.