

Sentiment analysis with hotel customer reviews using FNet

Shovan Bhowmik¹, Rifat Sadik¹, Wahiduzzaman Akanda¹, Juboraj Roy Pavel²

¹Department of Computer and Information Sciences, University of Delaware, Newark, USA

²School of Marine Science and Policy, University of Delaware, Newark, USA

Article Info

Article history:

Received Mar 23, 2023

Revised Sep 8, 2023

Accepted Oct 3, 2023

Keywords:

Bidirectional encoder representations from transformers
FNet
Fourier transform
Hotel reviews
Sentiment analysis
Transformer-based models

ABSTRACT

Recent research has focused on opinion mining from public sentiments using natural language processing (NLP) and machine learning (ML) techniques. Transformer-based models, such as bidirectional encoder representations from transformers (BERT), excel in extracting semantic information but are resource-intensive. Google's new research, mixing tokens with fourier transform, also known as FNet, replaced BERT's attention mechanism with a non-parameterized fourier transform, aiming to reduce training time without compromising performance. This study fine-tuned the FNet model with a publicly available Kaggle hotel review dataset and investigated the performance of this dataset in both FNet and BERT architectures along with conventional machine learning models such as long short-term memory (LSTM) and support vector machine (SVM). Results revealed that FNet significantly reduces the training time by almost 20% and memory utilization by nearly 60% compared to BERT. The highest test accuracy observed in this experiment by FNet was 80.27% which is nearly 97.85% of BERT's performance with identical parameters.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Shovan Bhowmik

Department of Computer and Information Sciences, University of Delaware
Newark, USA

Email: bhowmik.sshovon5795@gmail.com

1. INTRODUCTION

Blogs, forums, and online social networks such as Facebook and Twitter have expanded due to the advancement of technology, providing opportunities for users to discuss any subject and express their ideas. They might share their opinion on current events, argue issues, or provide satisfactory comments for a product, for instance. This is why sentiment analysis, known as opinion mining, has become an important topic in machine learning research. Despite the significance of sentiment analysis and the range of applications it already has, there are a lot of challenges with natural language processing (NLP) that must be resolved.

Over time, many machine learning (ML) and deep learning (DL) models have been used to improve the robustness of sentiment analysis tasks, including long short-term memory (LSTM), convolutional neural network (CNN), support vector machine (SVM) [1], [2]. Even though these conventional models have shown significant improvements while performing sentiment analysis and handling long texts and sequences, context-associated information extraction is a major challenge. To overcome these constraints, a novel approach named bidirectional encoder representations from transformers (BERT) [3] has been introduced by researchers, which has been successfully integrated into many NLP research projects to acquire remarkable performance improvement. BERT is a transformer-based pre-trained model that utilizes the self-attention mechanism and encoding feature of transformers to perform several tasks, e.g., text classification, question-answering, and relation

extraction. Moreover, this self-attention mechanism also facilitates the context-based approach while working with text data and reduces the steps of preprocessing the data. However, this approach greatly increases training time and memory consumption heavily [4]. Several studies conducted previously have indicated that fourier transforms can be integrated in order to increase the speed of computation of different DL models such as convolutional neural networks [5], recurrent neural networks [6], and transformers [7]. Therefore, Google researchers have recently proposed a new model named FNet: mixing tokens with fourier transforms [8], which replaced the self-attention mechanism with fourier transforms. According to them, FNet achieves almost state-of-the-art performances of BERT with much faster training speed and memory utilization.

Although the researchers claimed the effectiveness of this model by comparing the performance of FNet with other transformer-based models, it was benchmarked only for the pretraining part. In this paper, we applied this model as a fine-tuning task in sentiment analysis to validate their work on a sufficiently large-scale hotel review dataset. The summary of our project is listed here:

- Applying the FNet model to analyze sentiments from hotel reviews.
- Presenting detailed comparison between FNet and BERT models regarding execution time and memory usage.
- Performing a comparative study between FNet, BERT, LSTM, and SVM models.
- Conducting an ablation study with different learning rates to identify the best parameters for our FNet model for the hotel review dataset.

2. RELATED WORKS

Machine learning and deep learning are widely used in sentiment analysis tasks. To recognize sentiments from COVID-19 tweet data, Jain *et al.* [9] used different machine and deep learning algorithms. This study also proposed an ensemble-based approach, which outperformed other approaches. In another work, Gandhi *et al.* [10] used deep learning algorithms CNN and LSTM to analyze sentiments from tweeter data. The preprocessing step of this experiment includes Word2Vec embedding for mapping words to vectors. In another study, Rahman *et al.* [11] performed sentiment analysis on customer feedback from six mobile banking applications utilizing the SVM and the Naive Bayes (NB) algorithm. In their experiment, linear SVM generated the best result with 97.17% accuracy. We also studied some research works on sentiment analysis that have been done in different languages, such as Arabic [12]. This research work developed an ensemble-based classifier to detect emotions in an Arabic context. Their ensemble-based model outperformed traditional ML algorithms, such as SVM, AdaBoost classifier, maximum entropy, k-nearest neighbors, decision tree, random forest, logistic regression, and Naive Bayes. In addition, Chiny *et al.* [13] showed that the dimension of word mapping vector on short and long text impacts the sentiment analysis result. Their experiment showed that an increase in the dimension of the word vector for long text resulted in better performances. Some works used ML and DL models in hotel [14]–[16] and social media [17], [18] sentiment reviews. These works include models such as LSTM, decision tree, SVM, and bayesian classifiers.

Recently transformer-based models have become very popular for NLP-based tasks, e.g., sentiment analysis, text or review classification, and, document categorization. A transformer-based BERT model is used to identify harmful news [19], [20]. Additionally, models such as BERT and ELECTRA are also being used to identify abusive comments or texts and hate speech on social media platforms such as Facebook and Twitter [21], [22]. The applications of transformer-based algorithms are expanding in document classification and analysis [23], [24].

There have been some research works that used fourier analysis with convolutional neural network (CNN) and recurrent neural network (RNN) [6], [25]. The motivation for these research works was fast computation and approximation. Fourier transform is also used in a transformer-based model [26]–[28]. All these studies tried to scale or improve the attention mechanism of transformers by introducing a new approximation method or filter. But none of them replaced the entire attention layer and substituted it with some alternatives. This attention mechanism was first replaced with the fourier transform by Lee-Thorp *et al.* [8].

3. METHOD

In this section, we will discuss the overall steps required in our study to predict sentiments from text data using FNet. This will be followed by a description of transformer architecture (BERT) and FNet.

3.1. System architecture

We can show the overall workflow of our project in Figure 1. The dataset for this project was raw text data of customer reviews about hotels with two parts-positive and negative feedback with a review score. As there was no specific label for the texts, we preprocessed the data not only for cleansing but also by assigning labels based on the review score. If the review score was greater than 5 for a particular text, we considered it a positive sentiment; otherwise, it was negative. Then, we split the data into the train-validation-test portion. In this project, we tried to capture the pre-trained weights of the transformer-based models from their original design and fine-tuned them in our dataset for both FNet and BERT. Therefore, we tokenized the text parts of the dataset and prepared an input representation required to feed the data in the FNet architecture. We created tensors of input IDs, types of input IDs, and position embeddings for each text. This was followed by transferring the inputs to the FNet. It needs to be mentioned here that we created attention masks for the BERT and removed them while training in FNet, as the attention was built inside the FNet model through fourier transformation. For predicting the classes and selection of the hyperparameters, we added two classifier layers on top of the FNet with a sigmoid activation function. The classification layers were (768, 256) and (256, 2). 768 was the input shape for the first layer since FNet generates a vector of size 768 for the classification purpose from the input tokens by the computations inside the encoders and decoders of hidden transformer layers. The output shape of 2 in the last layer was for the number of classes. Once the prediction was over, we compared our model with the existing techniques. We also want to state that we generated term-frequency inverse document frequency (TF-IDF) features while training with SVM [29]. As it is not a contextual feature, we made some efforts to preprocess the data before the fitting. Those include punctuation removal, word tokenization, stopword removal, stemming, lower-case conversion, and detokenization.

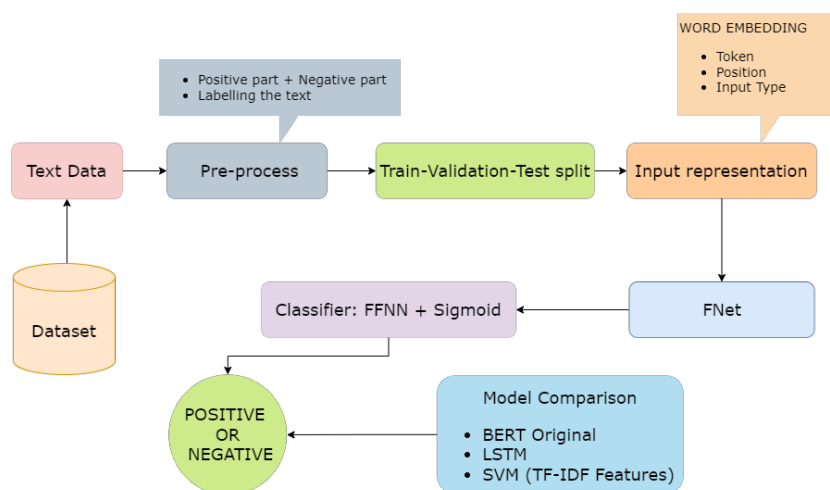


Figure 1. Proposed system architecture for classifying hotel reviews

3.2. Model description

In the following parts, we will describe the internal architecture of our selected FNet model along with its similarity and dissimilarity with the transformer-based BERT architecture.

3.2.1. Transformer based models

Transformer-based models are bringing a revolution in NLP-based tasks. It's designed especially for data with sequential inputs. The attention mechanism is considered the heart of transformer architecture. This is a multi-head self-attention mechanism that relates different positions of the input sequence. This architecture has a decoder and encoder part. The encoder part maps the input sequence to a continuous representation, and the decoder transforms it into an output sequence.

BERT is a popular transformer-based model with a bidirectional transformer encoder [3]. Token sequences generated from single or multiple sentences can be the input for the BERT model. Figure 2 depicts the BERT architecture with required layers. This token sequence consists of three special BERT tokens, namely [CLS], [SEP], and [PAD], representing classification tokens, pair tokens, and padding tokens. Masked language

modeling (MLM) and next sentence prediction (NSP) are two unsupervised sub-tasks for BERT pretraining. These two tasks are done both in a forward and backward direction, and that's why it is called bidirectional. Sub-task MLM is used to mask some of the input tokens on a random basis and tries to predict the mask sentences. NSP is responsible for predicting a sequence available after the [SEP] token. Thus, the prediction through this technique gives context-related information for different tokens. However, the tokenization in the BERT architecture is done with word piece tokenization and has a gigantic vocabulary of 30,522. For the prediction purpose of a particular class, it creates a vector of 768 lengths which can represent the reflection of tokens for any classes in the application where it is applied.

3.2.2. Discrete fourier transform

Using the fourier transform, a function is broken into its constituent frequencies. If we are given an input sequence x_1, x_2, \dots, x_n , where n is in the range between 0 to $N-1$ then DFT can be defined by (1) [8]:

$$X_k = \sum_{n=0}^{N-1} x_n e^{(-2\pi i/N)nk}, 0 \leq k \leq N-1 \quad (1)$$

Here, X_k is the new representation and sum of all input tokens. Convolution from the frequency domain to the time domain is represented by matrix multiplication, a technique by which DFT is computed. This can decrease computational costs and accelerate the convolutional neural network training process. This thought encourages researchers to modify transformer architecture by using fourier transformation, such as FNet architecture.

3.2.3. FNet architecture

FNet is a modification of transformer-based BERT architecture [8] (Figure 2(a)). The modification implies that it is an attention-free transformer. Figure 2(b) depicts the FNet architecture with required sub-layers. This architecture is built in such a way that the self-attention sublayer of each transformer encoder is replaced with a fourier sublayer. In other words, FNet applies a 2D discrete fourier transform to its embedding input, both in sequence (F_{seq}) and hidden (F_h) dimensions. If embedding input x , fourier transform is F , and the real part is \mathbb{R} , then the fourier operation can be defined by (2).

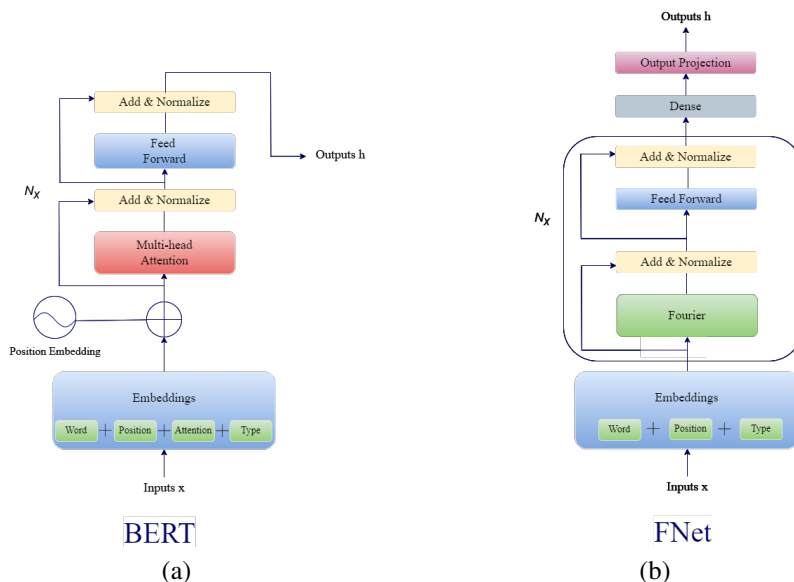


Figure 2. N encoder blocks of: (a) BERT architecture [4] and (b) FNet architecture [8]

Fourier transform ensures that the feed-forward sub-layer has sufficient access to all tokens. The input is transformed repeatedly between the temporal and frequency domains using the fourier transform sub-layer. Convolution in the time domain is identical to the result of a fourier transformation in the frequency domain. FNet may operate in this manner by alternately performing multiplications and convolutions.

4. RESULTS AND DISCUSSION

In this section, we will describe the dataset used in our project, the environmental and experimental setup to run the investigation, the performance evaluation metrics, observed results, and the discussion of the findings of our experiment with multiple subsections.

4.1. Dataset description

The source of our dataset was the Kaggle website, and the name of our dataset is 515 K hotel reviews in Europe. This data was collected from the Booking.com website and is publicly available to everyone. This dataset consisted of a total of 515,738 customer reviews with 17 attributes. Our target attributes were positive review, negative review, and reviewer score. The range of the reviewer score is between 0 to 10. In our dataset, we found that the lowest reviewer score was 2.5, and the highest score was 10. Table 1 will present examples of positive and negative reviews. After preprocessing the data and labeling them, we split the dataset into train, validation, and test. We've also presented a textual statistical analysis of the dataset in Table 2.

Table 1. Example of review text

Review type	Review text
Positive review	No real complaints the hotel was great great location surroundings rooms amenities and service...
Negative review	Rooms are nice but for elderly a bit difficult as most rooms are two story with narrow steps...

Table 2. Statistical analysis of dataset details

Label	No. of instances	Train (60%)	Validation (25%)	Test (15%)
Positive	485,035	291,072	121,220	72,743
Negative	30,703	18,370	7,715	4,618
Total	515,738	309,442	128,935	77,361

4.2. Experiment setup

We performed all our investigations on Google Colab with a built-in Tesla k-80 GPU that has a memory of 12 GB. For our experiment on this large corpora with unbalanced classes, we set up the parameters as maximum length of a sentence: 100, batch size: 128, optimizer: Adam, learning rate: $2e - 5$, regularizer: $L2 (1e - 5)$.

For the ablation study on the performance of FNet, the learning rates of $3e - 5$ and $4e - 5$ were chosen, respectively. This setup was the same for BERT and LSTM to make a fair comparison. The dataset was trained for 4 epochs in all the models. However, as SVM used TF-IDF for fitting the features, we used standard parameters for the SVM model.

As FNet is a modified architecture of BERT, we followed the approach shared for classifying texts using BERT and modified this with the FNet weights published in the huggingface website. We removed the attention weights by ourselves while preparing the input representation for the FNet fine-tuning. For the other baseline models (LSTM and SVM), we followed the general approach publicly available in the research community.

4.3. Performance evaluation

As FNet is almost similar to the BERT architecture with a change in the configuration of the self-attention mechanism [8], the focus on selecting performance measurement metrics was solely based on the original FNet paper. So, we considered training time and memory to compare with BERT. We also measured accuracy, precision, recall, and F1-score for the BERT, LSTM, and SVM models selected in this work. These models are selected because of their extensive usage in the field of NLP [30].

4.4. Results

In this section, we showed our obtained results for various performance measurement metrics. To validate the FNet researcher's assertion, we have compared the training time and memory utilization comparison between FNet and BERT in Table 3. The FNet model took around 160 minutes for overall training, validation, and test execution for our dataset, whereas BERT took 199 minutes. Therefore, 19.6% of the time was saved for our model. Moreover, memory consumption was 60.4% less than the BERT (base) model. In Table 4, we portrayed the performance comparison between our selected models.

Table 3. Time and memory comparison between FNet and BERT

Model name	Training time	Memory utilization
BERT	119.4 minutes	11.62 GB
FNet	96 minutes	4.60 GB

Table 4. Comparison of classification results in different models

Model name	Accuracy (%)	Precision (%)	Recall	F1-score (%)
FNet	80.27	79.31	80.47	80.93
BERT	82.03	82.29	82.43	81.37
LSTM	93.96	77.13	49.68	48.16
SVM	90.19	49.36	49.10	49.17

BERT is always the first choice for state-of-the-art applications in NLP [3], achieving better precision, recall, and F1-score values. Although LSTM and SVM with TF-IDF had very good testing accuracy, the performance was poor for the other performance measurement metrics. We also observed that FNet performed as stated in the paper, and the results were close to the BERT performance with a test accuracy of 80.27%.

We showed the ablation study for our FNet model with different learning rates. The results of this study can be found in Table 5. $4e-5$ became the best selection of learning rate in our experimental setup compared to the other two learning rates as this provided better performance in accuracy, precision, recall, and F1-score. Figures 3 and 4 present the accuracy and loss values between our FNet model and the BERT for four epochs, respectively.

Table 5. Ablation study for FNet with different learning rates

Learning rate	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
$2e-5$	80.27	79.31	80.47	80.93
$3e-5$	81.28	81.21	81.15	80.97
$4e-5$	81.96	82.32	81.86	82.17

4.5. Discussions

Transformer models have a powerful feature of context-based tokenization and feature selection absent in the traditional NLP algorithms [31]. Although BERT is still the best option in NLP tasks so far, due to its expensive computation, we have tried to apply the recent FNet model in the hotel review dataset for sentiment analysis. As we removed the attention mask for the input texts for the training purpose in the FNet, it utilizes the fourier transform to convert the nonlinear text features to linear features and hence works as an unparameterized attention similar to the masking of BERT. Because of this, we have fewer features to train. This was reflected in our result in Table 3. Besides, we also needed less tokenization time in FNet compared to BERT. Thus, approximately 30% of the overall memory was less used for the tokenization of FNet.

While comparing the performance of the FNet with other models in Table 4, we also found that LSTM and SVM had poor precision, recall, and F1-score values. The possible reason could be the uneven distribution of the classes in the dataset without using cross-validation as well as taking parameters similar to the FNet for a non-discriminative analysis. Nonetheless, we attained generalization for both BERT and FNet architecture. This can be understood from Table 4 performance metrics and Figures 3(a)-(b) and 4(a)-(b). It is because transformer models can distribute the weights of different classes to the optimizer without getting partial to any specific classes [32]. Furthermore, we could only run the training and model selection for 4 epochs. As the curves show generalized patterns, we can say that our model was well-fitted and could be improved with the increase of epochs. Another reason for taking 4 epochs is that BERT pre-trained was performed for 4 epochs, and these models can converge easily in these epochs.

The ablation study shows that a learning rate of $4e-5$ provided a better result in this classification task. However, we can get the best parameters by some hyperparameter tuning operations. Overall, the experiment shows that training time reduction and memory utilization minimization can be helpful in training large corpus, and models such as FNet could achieve performance similar to the sophisticated models even without using attention masks.

Among the transformer-based models that we studied in this research, fine-tuning FNet is more cost-effective compared to the BERT architecture. Because of the unparameterized transformation through fourier

analysis, FNet omits the multi-head attention mechanism, which works with all the tokens generated from the word embedding in the BERT architecture. Therefore, it reduces time and memory consumption simultaneously because of the normalization layers, and it can easily map the nonlinear transformations of token weights accurately. From the results, we can also see that with an increasing learning rate, it can perform similarly to BERT models. As resource optimization is a vital element while working on large amounts of data, FNet can significantly be utilized in NLP tasks such as question answering, name-entity recognition, and relation extraction. Therefore, the goal of this research is also to extend the application of FNet in the text-mining community. In the future, research can also be done on tuning different layers inside the FNet architecture, for example, changing the number of DFT operations for each input representation, changing the number of hidden neurons inside the feed forward and dense layers, to observe the performance variation in the text mining area. This will help us to understand the impact of using fourier transformation in analyzing text features and applications.

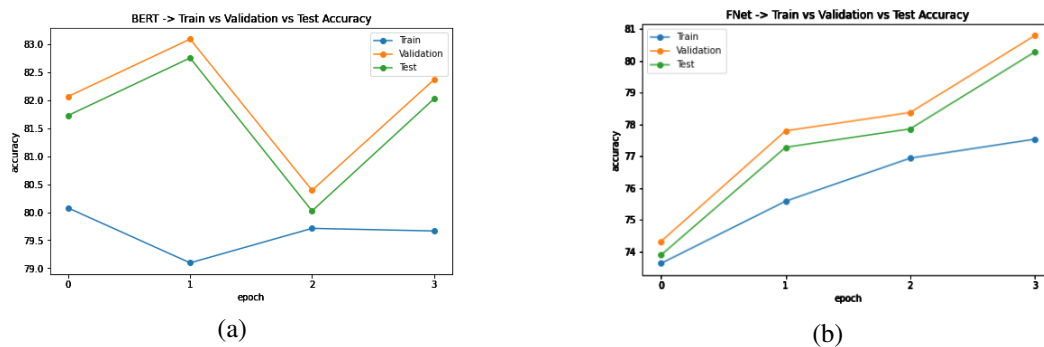


Figure 3. Accuracy curves for (a) BERT and (b) FNet models over four epochs

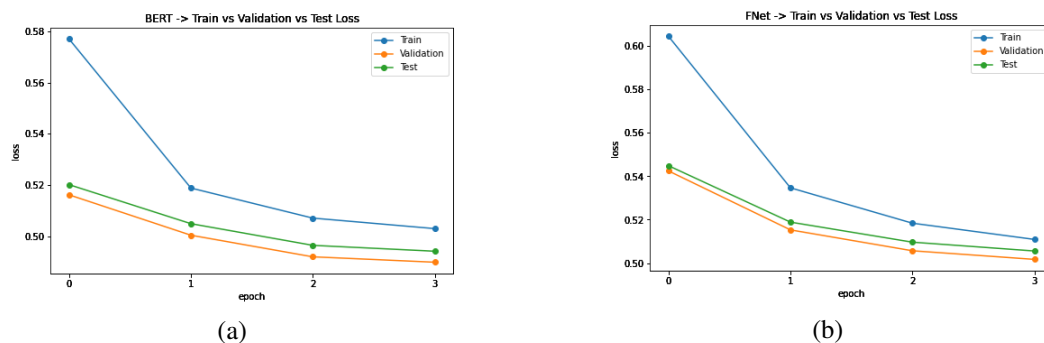


Figure 4. Loss curves for (a) BERT and (b) FNet models over four epochs

5. CONCLUSION

In this research, we implemented the FNet model to classify hotel reviews to extract customer sentiments. Our main focus was to compare the FNet model with the BERT model since FNet is a modification of the BERT model. We have found that the FNet model significantly reduced memory usage (60%) and execution time (20%) compared to BERT. The highest test accuracy observed in this experiment by FNet was 80.27% which is almost 97.85% of BERT with the same parameter setting. We also compared the performance of the FNet model with classic ML models like SVM and LSTM and found that FNet achieved state-of-the-art; we also presented an ablation study in which we demonstrated the performance of the FNet model in terms of different learning rates. For learning rate $4e-5$, we have achieved the best result with a test accuracy of 81.96%. For future studies, we wish to conduct this research with more diverse datasets, besides other transformer-based models like RoBERT, ALBERT, and XLNet.




REFERENCES

- [1] G. S. N. Murthy, S. R. Allu, B. Andhavarapu, M. Bagadi, and M. Belusonti, "Text based Sentiment Analysis using LSTM," *International Journal of Engineering Research*, vol. V9, no. 05, 2020, doi: 10.17577/ijertv9is050290.
- [2] N. Zainuddin and A. Selamat, "Sentiment analysis using Support Vector Machine," in *International Conference on Computer, Communications, and Control Technology (I4CT)*, Sep. 2014, pp. 333–337, doi: 10.1109/I4CT.2014.6914200.
- [3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL HLT Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 4171–4186, Oct. 2018, [Online]. Available: <http://arxiv.org/abs/1810.04805>.
- [4] Z. Sun, H. Yu, X. Song, R. Liu, Y. Yang, and D. Zhou, "MobileBERT: A compact task-agnostic BERT for resource-limited devices," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 2158–2170, 2020, doi: 10.18653/v1/2020.acl-main.195.
- [5] T. Highlander and A. Rodriguez, "Very Efficient Training of Convolutional Neural Networks using Fast Fourier Transform and Overlap-and-Add," *proceedings of the British Machine Vision Conference*, pp. 160.1–160.9, Jan. 2016, doi: 10.5244/C.29.160.
- [6] J. Zhang, Y. Lin, Z. Song, and I. S. Dhillon, "Learning Long Term Dependencies via Fourier Recurrent Units," *35th International Conference on Machine Learning, ICML 2018*, vol. 13, pp. 9256–9264, Mar. 2018, [Online]. Available: <http://arxiv.org/abs/1803.06585>.
- [7] K. Choromanski et al., "Rethinking Attention With Performers," *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.
- [8] J. Lee-Thorp, J. Ainslie, I. Eckstein, and S. Ontanon, "FNet: Mixing Tokens with Fourier Transforms," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4296–4313, May 2021, doi: 10.18653/v1/2022.naacl-main.319.
- [9] R. Jain, S. Bawa, and S. Sharma, "Sentiment Analysis of COVID-19 Tweets by Machine Learning and Deep Learning Classifiers," *Lecture Notes in Networks and Systems*, vol. 318, pp. 329–339, 2022, doi: 10.1007/978-981-16-5689-7_29.
- [10] U. D. Gandhi, P. Malarvizhi Kumar, G. Chandra Babu, and G. Karthick, "Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)," *Wireless Personal Communications*, 2021, doi: 10.1007/s11277-021-08580-3.
- [11] N. A. Rahman, S. D. Idrus, and N. L. Adam, "Classification of customer feedbacks using sentiment analysis towards mobile banking applications," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 4, pp. 1579–1587, 2022, doi: 10.11591/ijai.v11.i4.pp1579-1587.
- [12] N. Hicham, S. Karim, and N. Habbat, "Customer sentiment analysis for Arabic social media using a novel ensemble machine learning approach," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4504–4515, 2023, doi: 10.11591/ijece.v13i4.pp4504-4515.
- [13] M. Chiny, M. Chihab, A. A. Lahcen, O. Bencharef, and Y. Chihab, "Effect of word embedding vector dimensionality on sentiment analysis through short and long texts," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, pp. 823–830, 2023, doi: 10.11591/ijai.v12.i2.pp823-830.
- [14] J. J. Bonny, N. J. Haque, M. R. Ulla, P. Kanungoe, Z. H. Ome, and M. I. H. Junaid, "Deep Learning Approach for Sentimental Analysis of Hotel Review on Bengali text," in *Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, Apr. 2022, pp. 1–7, doi: 10.1109/ICAECT54875.2022.9808001.
- [15] A. Ishaq et al., "Extensive hotel reviews classification using long short term memory," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 10, pp. 9375–9385, 2021, doi: 10.1007/s12652-020-02654-z.
- [16] L. Mostafa, "Machine Learning-Based Sentiment Analysis for Analyzing the Travelers Reviews on Egyptian Hotels," *Advances in Intelligent Systems and Computing*, vol. 1153 AISC, pp. 405–413, 2020, doi: 10.1007/978-3-030-44289-7_38.
- [17] N. Kewsuwun and S. Kajornkasirat, "A sentiment analysis model of agritech startup on Facebook comments using Naive Bayes classifier," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, pp. 2829–2838, 2022, doi: 10.11591/ijece.v12i3.pp2829-2838.
- [18] V. Pandi, P. Nithiyandam, S. Manickavasagam, I. M. Meerasha, R. Jaganathan, and M. K. Balasubramanian, "A comprehensive analysis of consumer decisions on Twitter dataset using machine learning algorithms," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 3, pp. 1085–1093, 2022, doi: 10.11591/ijai.v11.i3.pp1085-1093.
- [19] S. Y. Lin, Y. C. Kung, and F. Y. Leu, "Predictive intelligence in harmful news identification by BERT-based ensemble learning model with text sentiment analysis," *Information Processing and Management*, vol. 59, no. 2, 2022, doi: 10.1016/j.ipm.2022.102872.
- [20] W. Gao, H. Deng, X. Zhu, and Y. Fang, "Topic-BERT: Detecting harmful information from social media," *Intelligent Decision Technologies*, vol. 15, no. 3, pp. 333–342, 2021, doi: 10.3233/IDT-200094.
- [21] K. Mnassri, P. Rajapaksha, R. Farahbakhsh, and N. Crespi, "BERT-based Ensemble Approaches for Hate Speech Detection," *IEEE Global Communications Conference, GLOBECOM 2022 - Proceedings*, pp. 4649–4654, 2022, doi: 10.1109/GLOBECOM48099.2022.10001325.
- [22] T. T. Aurpa, R. Sadik, and M. S. Ahmed, "Abusive Bangla comments detection on Facebook using transformer-based deep learning models," *Social Network Analysis and Mining*, vol. 12, no. 1, 2022, doi: 10.1007/s13278-021-00852-x.
- [23] M. Ostendorff, T. Ruas, T. Blume, B. Gipp, and G. Rehm, "Aspect-based Document Similarity for Research Papers," *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, pp. 6194–6206, 2020, doi: 10.18653/v1/2020.coling-main.545.
- [24] M. M. Rahman, M. Aktaruzzaman Pramanik, R. Sadik, M. Roy, and P. Chakraborty, "Bangla Documents Classification using Transformer Based Deep Learning Models," *International Conference on Sustainable Technologies for Industry 4.0, STI 2020*, 2020, doi: 10.1109/STI50764.2020.9350394.
- [25] S. Lin et al., "FFT-based deep learning deployment in embedded systems," in *Proceedings of the 2018 Design, Automation and Test in Europe Conference and Exhibition DATE*, vol. 2018-January, pp. 1045–1050, 2018, doi: 10.23919/DATE.2018.8342166.
- [26] A. Tamkin, D. Jurafsky, and N. Goodman, "Language through a prism: A spectral approach for multiscale language representations," *Advances in Neural Information Processing Systems*, vol. 2020-December, 2020.




- [27] K. Choromanski *et al.*, “Masked Language Modeling for Proteins via Linearly Scalable Long-Context Transformers,” *Machine Learning (cs.LG)*, 2020, [Online]. Available: <http://arxiv.org/abs/2006.03555>.
- [28] F. Sahinuc and A. Koc, “Fractional Fourier Transform Meets Transformer Encoder,” *IEEE Signal Processing Letters*, vol. 29, pp. 2258–2262, 2022, doi: 10.1109/LSP.2022.3217975.
- [29] S. Bhowmik, P. R. K. Prosun, and K. S. Alam, “A Novel Three-Level Voting Model for Detecting Misleading Information on COVID-19,” *Lecture Notes in Networks and Systems*, vol. 292, pp. 374–383, 2022, doi: 10.1007/978-981-16-4435-1_36.
- [30] R. Karsi, M. Zaim, and J. El Alami, “Assessing naive bayes and support vector machine performance in sentiment classification on a big data platform,” *IAES International Journal of Artificial Intelligence*, vol. 10, no. 4, pp. 990–996, 2021, doi: 10.11591/IJAI.V10.I4.PP990-996.
- [31] C. I. Eke, A. A. Norman, and L. Shuib, “Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model,” *IEEE Access*, vol. 9, pp. 48501–48518, 2021, doi: 10.1109/ACCESS.2021.3068323.
- [32] H. T. Madabushi, E. Kochkina, and M. Castelle, “Cost-Sensitive BERT for Generalisable Sentence Classification with Imbalanced Data,” *Computation and Language (cs.CL)*, Mar. 2020, doi: 10.48550/arXiv.2003.11563.

BIOGRAPHIES OF AUTHORS






Shovan Bhowmik    is a Ph.D. student in the Computer and Information Sciences Department at the University of Delaware, Delaware, USA. He received his bachelor's in Computer Science and Engineering from Khulna University of Engineering & Technology (KUET), Bangladesh in 2019. His research interests include natural language processing, machine learning, text mining, and speech recognition. Currently, his Ph.D. research is focused on the investigation of large language models in the field of computational biology, where he works on various impact relation extraction techniques from biomedical literature to measure the impact of post-translational modifications on proteins and specific sites. He can be contacted at email: bhowmik.sshovon5795@gmail.com.






Rifat Sadik    is a Ph.D. student in the Computer and Information Sciences Department at the University of Delaware, Delaware, USA. He received his B.Sc. and M.Sc. from the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh. He has authored several journals and conference papers. His research interests include deep learning, artificial intelligence, natural language processing, and developing machine learning models for intelligent systems. He received several scholarships, and some significant ones are the national science and technology (NST) Fellowship and the University Merit Scholarship. He can be contacted at email: rifat.sadik.rs@gmail.com.



Wahiduzzaman Akanda    is a Ph.D. student in the Computer and Information Sciences Department at the University of Delaware, Delaware, USA. He received his bachelor's in Computer Science and Engineering from Khulna University of Engineering & Technology (KUET), Bangladesh in 2015. He also received his master's from American International University-Bangladesh in Computer Science major. His research interests include software engineering, software design, software quality, machine learning, and natural language processing. Prior to his doctoral study, he had been working as a software quality assurance engineer for almost seven years. His current work focuses on enhancing the commenting functionality within visual programming languages. He can be contacted at wahiduzzaman.akanda@gmail.com.



Juboraj Roy Pavel    is a Master's in Marine Policy student and working as a graduate research assistant at the University of Delaware. He has earned a B.S.S in Economics with a second major in Computer Science & Engineering from Metropolitan University, Sylhet, Bangladesh. His research interests lie in the fields of machine learning, climate change economics, environmental & natural resource economics, and technology adoption. He can be contacted at email: jroy-pavel@gmail.com.