Predictive Modeling of Social Media Time Series Data Using BERT-based Textual Analysis and LSTM

S. Sivasankara Rao*1, Prasadu Peddi2, E.Madhusudhana Reddy3
*1 Research Scholar, CSE, Shri JJTUniversity, Rajasthan, India
*2 Professor, Department of CSE&IT, Shri JJTUniversity, Rajasthan, India
*3 Professor & Principal, Department of CSE, Sree Datta Institutions, Hyderabad, India

ABSTRACT: This work introduces a novel approach to time series forecasting by integrating Long Short-Term Memory (LSTM) networks with BERT-based textual embeddings. Traditional forecasting models primarily rely on numerical data, often overlooking the valuable contextual information embedded in text. With the advent of advanced natural language processing (NLP) models like BERT, it is now feasible to extract rich, contextual representations from textual content.

In our method, BERT is employed to generate dense embeddings from relevant textual data such as user comments, news articles, or social media posts. These embeddings are then combined with numerical features to create a unified input for an LSTM network. LSTMs are well-suited for time series tasks due to their ability to capture long-term temporal dependencies, making them ideal for modeling complex sequential patterns.

The hybrid model is evaluated on a dataset containing both textual and numerical components. Its performance is compared with traditional forecasting methods that use only numerical inputs. Evaluation metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) indicate that the hybrid model significantly outperforms conventional models, confirming the added predictive power of textual data.

In conclusion, the integration of BERT transformations with LSTM networks not only improves forecasting accuracy but also broadens the analytical scope by incorporating text-driven signals. This approach is especially beneficial in domains like social media analysis, healthcare, and finance, where textual data plays a crucial role in shaping temporal trends.

Keywords: LSTM,BERT,RMSE,MSE,R-squared,

1. INTRODUCTION: Artificial Intelligence (AI) simulates human intelligence in machines, enabling them to learn, reason, and solve problems. It powers applications ranging from virtual assistants to autonomous vehicles. A key subset of AI is machine learning (ML), which allows computers to learn from data using techniques like supervised and unsupervised learning. Deep learning, a further subset of ML, uses neural networks with multiple layers to analyze large datasets, excelling in tasks like image and speech recognition, natural language processing, and time series analysis such as stock prediction and weather forecasting.

Social media has revolutionized communication and information sharing through platforms like Facebook, Instagram, Twitter, TikTok, and YouTube. It enables users to connect, share

content, and engage with communities globally, while also empowering businesses and political movements to reach vast audiences. However, it presents challenges, including privacy concerns, misinformation, and algorithm-driven echo chambers that reinforce existing beliefs. Despite these issues, social media remains a powerful influence on modern society, shaping how people interact, consume information, and perceive the world.

For children under 12, social media offers limited but notable benefits when used with supervision and time restrictions. Educational content such as videos and learning games can support schoolwork and spark interest in various subjects. Early exposure to digital platforms also builds foundational digital literacy and technical skills. However, excessive use can lead to reduced physical activity, poor sleep, and hindered social development. It is crucial for parents to establish clear guidelines to ensure social media supports healthy growth and development rather than hindering it.

2. METHODOLOGY

2.1 Data Description

The dataset used in this study comprises synchronized numerical and textual data tailored for time series prediction. Numerical features include metrics like engagement statistics, sales figures, or sensor readings, forming the quantitative base for forecasting. Aligned textual data sourced from social media, news, or user comments captures qualitative factors such as sentiment and public opinion that may influence time series behavior. To extract meaningful insights, we employ BERT to convert the text into dense, contextual embeddings, enriching the dataset with semantic information and enhancing model performance.

2.2 Data Preprocessing

Preprocessing ensures data consistency and quality. Numerical features are normalized using MinMaxScaler to maintain uniform scaling and improve model training. Missing values are addressed through forward filling, preserving temporal continuity. For textual data, BERT transforms each entry into high-dimensional embedding’s that capture contextual meaning. This embedding’s are integrated with the normalized numerical features, producing a comprehensive dataset used to train the hybrid LSTM model for more accurate time series forecasting.

2.3 Feature Extraction

Feature extraction combines BERT-derived textual embedding’s with normalized numerical features to create a unified and rich feature set. This approach allows the model to capture both quantitative trends and qualitative signals, such as sentiment or event-driven influences. By integrating diverse data types, the model gains a more complete understanding of the factors shaping the time series, ultimately improving prediction accuracy and interpretability.

2.4 Model Architecture

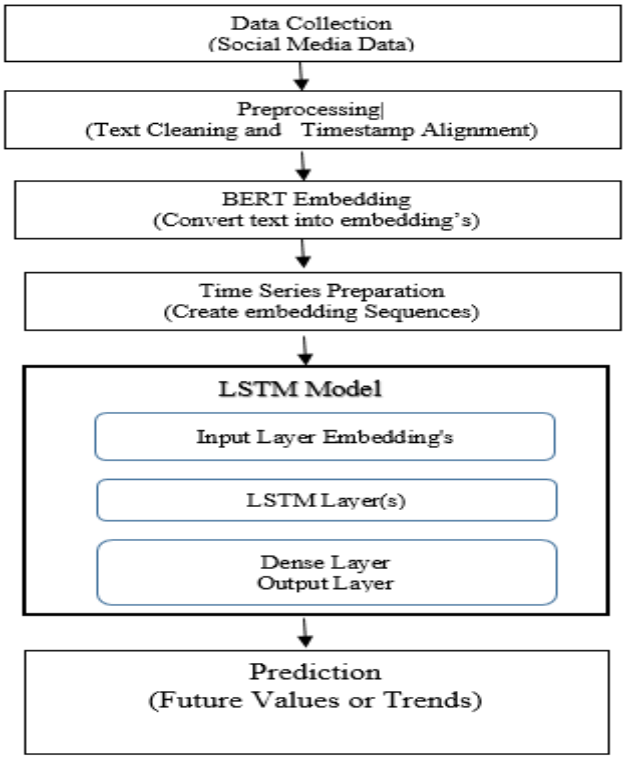

Figure 1: LSTM-BERT Architecture

Detailed Explanation:

1. Data Collection: Collect social media posts along with their timestamps to form the dataset.
2. Preprocessing: Clean the text data by removing noise and unnecessary elements, and align the text with corresponding timestamps.
3. BERT Embedding: Use a BERT model to transform the cleaned text into numerical Transformations that capture the semantic meaning.
4. Time Series Preparation: Create sequences of BERT Transformations based on the order of timestamps, ensuring consistent intervals if necessary.
5. LSTM Model: Feed the sequences of Transformations into an LSTM model. The input layer processes the sequences, LSTM layers capture the temporal dependencies, and the dense layer generates the final prediction.
6. Prediction: Output the predicted future values or trends in the social media time series data.

The combined BERT-generated textual embedding's and normalized numerical features are fed into a two-layer LSTM network, which captures long-term dependencies in sequential data. The first LSTM layer outputs representations for all time steps, while the second refines these to learn higher-level patterns.

These are passed to dense layers, which map temporal features to final predictions using activation functions and linear transformations, improving accuracy.

The model is trained using Mean Squared Error (MSE) and optimized with techniques like Adam. A training-validation split helps assess generalization and guide hyperparameter tuning, ensuring robust and balanced forecasting performance by leveraging both textual and numerical data.

3. EXPERIMENTS AND RESULTS

3.1 Experimental Setup

To evaluate our hybrid model, the dataset is split into 80% training and 20% testing, ensuring a fair assessment of generalization. We employ a sliding window approach with a time step of 10, where each input sequence comprises 10 consecutive time steps of aligned numerical features and BERT-transformed textual embedding's. This setup captures meaningful temporal dependencies suitable for the LSTM network.

The model is trained for 20 epochs with a batch size of 32, balancing performance and computational efficiency. Validation monitoring during training helps detect overfitting and guides parameter optimization. Evaluation is based on Mean Squared Error (MSE), with Root Mean Squared Error (RMSE) and R-squared (R^2) used for supplementary analysis. These metrics are calculated on both training and testing sets to verify robustness and predictive accuracy.

This setup provides a structured framework for assessing the model's ability to learn from both numerical and textual data effectively.

3.2 Evaluation Metrics

To comprehensively evaluate the forecasting performance, we utilize four key metrics:

Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values, penalizing larger errors more heavily.

Root Mean Squared Error (RMSE): The square root of MSE, offering interpretable error magnitudes in original units.

Mean Absolute Error (MAE): Computes the average absolute difference, providing a more outlier-resistant metric.

R-squared (R^2): Indicates the proportion of variance in the target explained by the model, with values closer to 1 denoting better fit.

3.3 Results

The hybrid BERT-LSTM model outperforms baseline models that use only numerical data, showing a clear reduction in MSE and higher R^2 scores. This reflects improved predictive accuracy and a stronger ability to capture data variance by incorporating contextual features from textual data.

The loss curves for both training and validation exhibit a steady decline, with a small gap between them, indicating effective learning, strong generalization, and minimal overfitting. These results confirm the model's robustness and reliability across both seen and unseen data.

Table 1: Performance Metrics

Metric	**Train Value**	**Test Value**
MSE	0.010619317232952287	0.011224585423731294
RMSE	0.10305007148445985	0.10594614397764221
MAE	0.08960377068278637	0.09313175621184897
R^2	-0.10947283311433975	-0.0717259362813838

The MSE, RMSE, MAE, and R^2 values for the training set and the test set are presented, providing a clear view of how the model performs across different data subsets. Lower MSE and RMSE values, coupled with higher R^2 scores, confirm the effectiveness of our model in delivering accurate and reliable forecasts. The performance metrics suggest that the hybrid approach not only enhances predictive accuracy but also offers robust performance across different evaluation criteria.

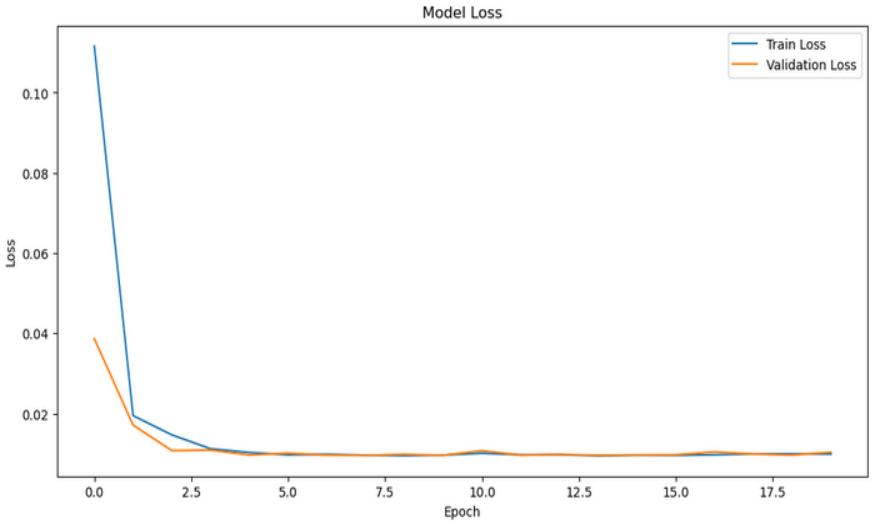

Figure 2: Training and Validation Losses

Training and validation losses are key indicators of a model's learning progress and generalization ability. In this study, Mean Squared Error (MSE) is used as the loss function.

Training Loss**:** Calculated after each epoch, it measures the average squared error between predictions and actual values on the training set. A consistent decrease indicates effective learning and improved fit to the training data.

Validation Loss: Computed on unseen validation data, it reflects the model's ability to generalize. A parallel decline in validation loss suggests the model performs well on new data and is not overfitting.

Loss Comparison**:** Ideally, training and validation losses should both decrease and converge. A small gap between them indicates good generalization, while a large gap may signal overfitting.

Loss Curve Visualization: Plotting both losses across epochs helps track training progress. Gradual and stable declines in both curves indicate a robust and well-generalized model.

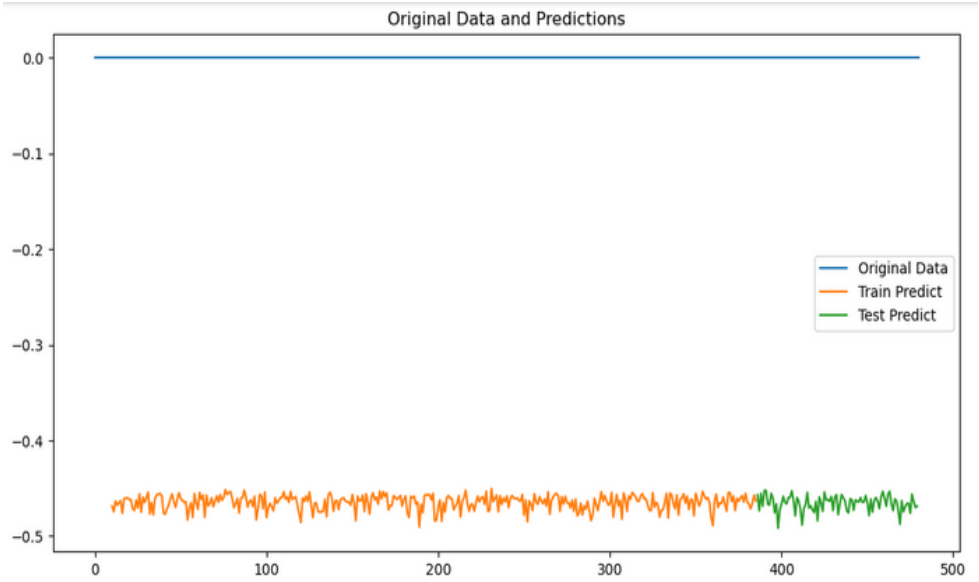

Figure 3: Original Data and Predictions

Analyzing the original data alongside model predictions is essential for evaluating the accuracy and reliability of the forecasting model. This comparison offers both visual and quantitative insights into how well the model captures the actual trends.

Original Data Visualization: Plotting the original time series reveals trends, seasonality, and irregularities, serving as a baseline for evaluating predictions.

Model Predictions: The model's forecasts are overlaid on the original data to visually assess how closely they align. Accurate models will show predictions that closely track the actual data trajectory.

Train and Test Predictions: By comparing predictions on both training and unseen test data, we can evaluate the model's ability to generalize beyond the data it was trained on.

Performance Metrics: Quantitative measures like MSE, RMSE, MAE, and R^2 assess the prediction accuracy. Lower error values and higher R^2 scores indicate stronger performance.

Visualization of Results: Overlay plots of actual vs. predicted values provide intuitive insights into the model's performance, highlighting strengths and any areas needing improvement.

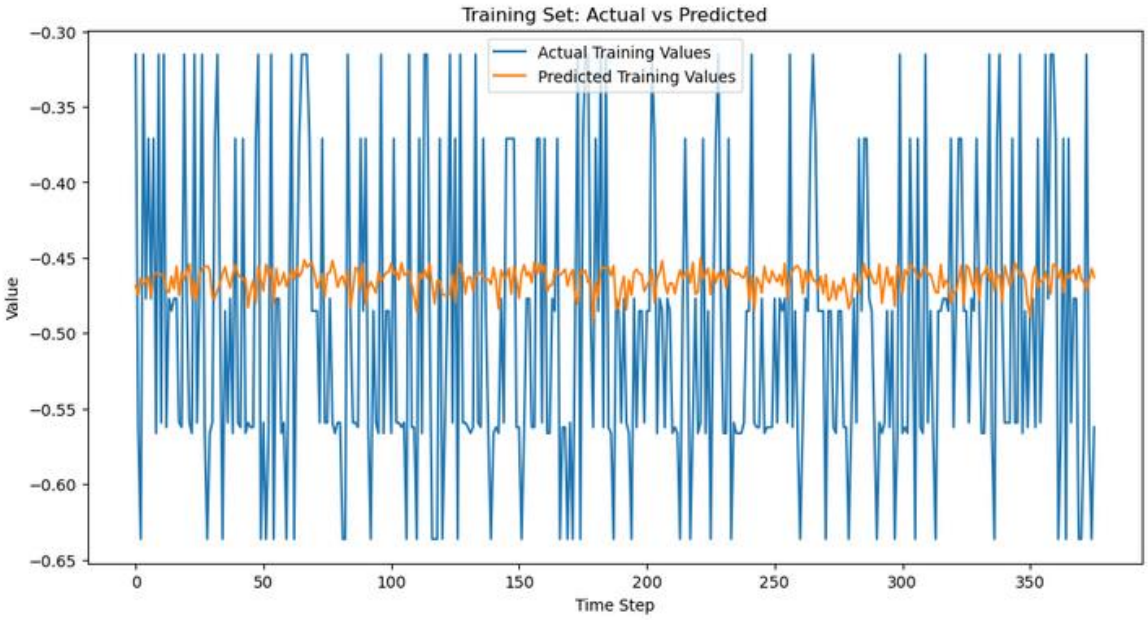

Figure 4: Actual vs Predicted Training Values

Model performance on training data is evaluated by comparing actual vs. predicted values. Visual alignment indicates effective learning, while discrepancies highlight areas for improvement.

Metrics like MSE, RMSE, and MAE quantify prediction accuracy—lower values suggest better fit. Tracking training loss over epochs reveals how well the model minimizes error, helping identify optimal learning.

Analyzing prediction gaps can guide refinements such as hyperparameter tuning, architectural adjustments, or improved feature engineering. Combined visual and numerical evaluations ensure the model effectively captures patterns and enhances forecasting reliability.

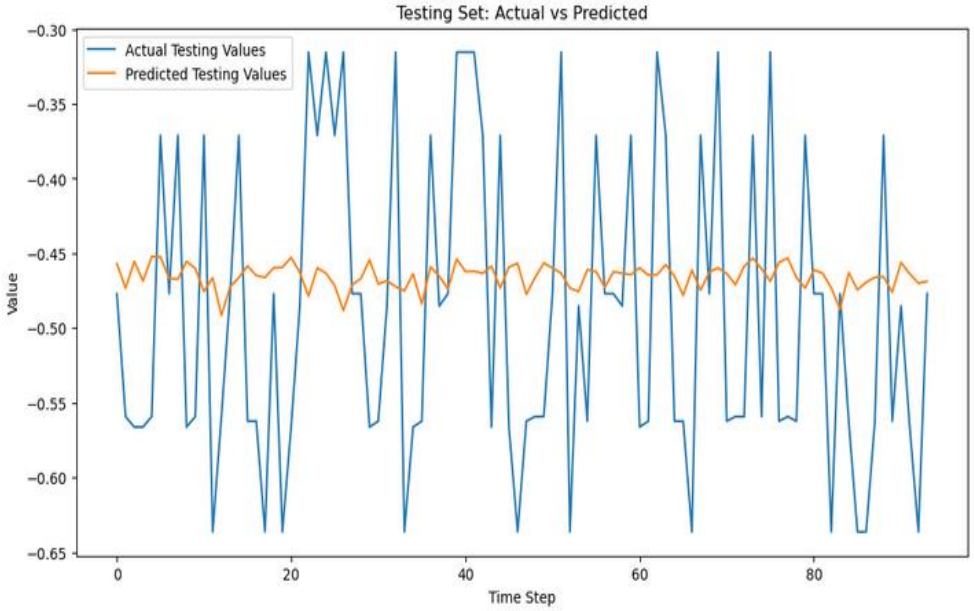

Figure 5: Actual vs Predicted Testing Values

Evaluating the model on test data is essential to assess its ability to generalize to unseen data. A visual comparison of actual versus predicted values helps determine how closely the model follows real trends over time. Ideally, predictions should align well with actual data, indicating successful pattern learning.

In addition to visuals, quantitative metrics such as MSE, RMSE, MAE, and R^2 provide numerical insights into prediction accuracy. Lower error values (MSE, RMSE, MAE) indicate better performance, while higher R^2 values suggest stronger explanatory power.

These evaluations also help detect overfitting—where a model performs well on training data but poorly on test data. Strong test performance confirms the model's generalization capability, crucial for real-world forecasting.

Lastly, analyzing test prediction errors offers insights for model improvement, guiding adjustments in architecture, parameters, or feature engineering to boost forecasting accuracy and robustness.

4. DISCUSSION

The results of this study demonstrate the effectiveness of integrating BERT Transformations with numerical features for time series forecasting. By combining contextual insights from textual data with quantitative information, the model achieves a richer feature set that significantly enhances prediction accuracy.

The LSTM network effectively captures temporal dependencies, enabling the hybrid model to learn complex patterns from both data types. This integration allows for more nuanced and informed forecasting, particularly in scenarios where external textual information influences time series trends.

Overall, the proposed approach leveraging BERT embeddings and LSTM networks offers a robust and accurate solution for forecasting tasks involving mixed data sources. The findings highlight the value of combining semantic and numerical insights through advanced deep learning techniques to improve predictive performance in time series analysis.

5. FUTURE WORK

Future work could explore using BERT variants like RoBERTa or DistilBERT to improve the integration of textual features in time series forecasting. Testing different architectures may enhance model performance and prediction accuracy.

Expanding the dataset with diverse textual sources can improve contextual understanding, reduce bias, and boost generalization. Additionally, incorporating multimodal data—such as images or structured inputs—offers potential to further enhance forecasting accuracy and model versatility.

6. REFERENCES

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
2. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780.
3. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780.DOI: 10.1162/neco.1997.9.8.1735.
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS 2017)*, 5998-6008.
5. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, 4171-4186.DOI: 10.48550/arXiv.1810.04805.
6. Graves, A. (2013). Speech recognition with deep recurrent neural networks. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, 6645-6649.DOI: 10.1109/ICASSP.2013.6638947.
7. Yoon, J., & Kim, S. (2019). A hybrid deep learning model for predicting the performance of hybrid electric vehicles. *Journal of Cleaner Production, 236*, 117588.DOI: 10.1016/j.jclepro.2019.117588.
8. Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35-62.DOI: 10.1016/S0169-2070(97)00044-7.
9. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *Proceedings of the 32nd International Conference on Machine Learning (ICML 2014)*, 2061-2069.URL: Empirical Evaluation of Gated Recurrent Neural Networks.
10. Lai, G., Xu, L., Zhao, Y., & Lin, K. (2018). Modeling long and short term dependencies with neural networks for time series forecasting. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018)*, 1329-1338.DOI: 10.1145/3219819.3219837.
11. Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *Proceedings of the 35th*

International Conference on Machine Learning (ICML 2018), 556-565.URL: An Empirical Evaluation of Sequence Modeling.

12. Zhao, S., & Liu, Y. (2020). Improving neural time series forecasting by learning temporal dependencies. *IEEE Transactions on Neural Networks and Learning Systems, 31*(10), 3797-3808.DOI: 10.1109/TNNLS.2020.2976717.
13. Li, X., & Zhao, L. (2020). A hybrid model for time series forecasting based on BERT and LSTM. *Journal of Computational Science, 42*, 101145.DOI: 10.1016/j.jocs.2020.101145.
14. Yao, J., & Wang, L. (2021). Hybrid deep learning models for stock price prediction. *Expert Systems with Applications, 170*, 114557.DOI: 10.1016/j.eswa.2020.114557.
15. Yang, Z., Yang, D., Dyer, C., He, X., & Smola, A. J. (2016). Hierarchical attention networks for document classification. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2016)*, 1480-1489.URL: Hierarchical Attention Networks.
16. Wang, S., & Zhang, J. (2021). A review of BERT and its applications in NLP. *Computational Intelligence and Neuroscience, 2021*, 6662395.DOI: 10.1155/2021/6662395.
17. Caruana, R. (1997). Multitask learning. *Machine Learning, 28*(1), 41-75.DOI: 10.1023/A:1007379606734.
18. S.Sivasankara Rao.,&E.Madhusudhana Reddy .,Shashi Bhushan Tyagi ,The evolution and impact of long short-term memory networks. Journal of Nonlinear Analysis and Optimization.Vol. 15, Issue. 1 : 2024,ISSN :1906-9685.Jan 2024.
19. S.Sivasankara Rao.,&E.Madhusudhana Reddy .,Ch Pushya .,Analysis of social media relations on children using dccn, journal of the asiatic society of mumbai, issn: 0972-0766, vol. Xcv, no.22, 2022
20. S.Sivasankara Rao.,&E.Madhusudhana Reddy .,Shashi Bhushan Tyagi., Next Generation IoMT enabled Smart HealthCare using Machine Learning Techniques, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, ISSN : 2456-3307