iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69359

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

India's GDP Growth Prediction System

Sudhansh Rasam[1], Vrunda Patil[2], Rahul Akula[3], Debarathi Goshal[4]
Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India

Abstract: This paper presents a deep learning approach that is integrated so as to forecast India's GDP growth; it combines the structured economic indicators from the Reserve Bank of India (RBI) with the unstructured financial news data. Customary models depend only on numerical data. This study leverages natural language processing techniques, in particular tokenization, lemmatization, and TF-IDF vectorization, to extract meaningful understandings from news articles, capturing real-time sentiment, with context. The TensorFlow neural network model is trained on the combined dataset, then assessed via Mean Absolute Error (MAE) and R-squared score (R^2), reaching satisfactory predictive accuracy. A Python pipeline that is automated streamlines all of the workflow, which goes from data scraping and preprocessing to prediction and visualization. The results show integrating unstructured news data improves the reliability for GDP forecasting in an important way, offering a more dynamic and thorough comprehension of economic trends.
Keywords: GDP Prediction, Deep Learning, TF-IDF, News Sentiment Analysis, Economic Forecasting, TensorFlow, RBI Data.

I. INTRODUCTION

Like any economy's metric, Gross Domestic Product (GDP) requires a value to be assessed; in this case it has to quantify the economic health of a nation. Keeping in mind that 'over a specific period' is the determining factor. It is necessary to predict GDP growth well in advance to aid in strategic planning, budgeting, investment, and even macroeconomics policy making. India is a country in the developing phase. This juxtaposition makes the economy a mosaic of factors, as it tries to assimilate the industrial output, inflation, political scenarios, and global markets. Due to this diversity, predicting tools become ineffective.

It appears that forecasting relies on time series data and the use of ARIMA, linear regression, and seasonal models. All of which, though functional under stable condition, lack the fundamentalism to adapt to shock changes in the economy, such as new policies, economic shifts, or even global crises. Moreover, it makes use of formal reports rather than defaulted economic news to change predictions. With all of these limitations, it is easy to state that uncluttered data could potentially fill all the voids left open through news articles that focus on economic commentary.

With the increasing digitization of news media, vast amount of financial news and sentiment-rich content on the internet has become available, providing valuable real-time signals into public opinion and market sentiment. It is possible to pull out and quantify these many insights using the techniques of Natural Language Processing (NLP). By combining this unstructured news data with structured economic indicators, we can develop predictive models that are context aware, responsive and robust. In this study we develop a deep learning-based model based on structured data from the Reserve Bank of India (RBI) with unstructured economic news data to predict quarterly GDP growth in India.

The unstructured news text is pre-processed using standard NLP methods including stopword removal, tokenization and lemmatization followed by extracting features using the Term Frequency - Inverse Document Frequency (TF-IDF) approach. The structured data consists of macroeconomic indicators including inflation rate, interest rate, GDP growth history and a industrial production index. A neural network is trained using the merged features with the TensorFlow framework to capture the evolution of both the numerical trends and textual sentiment.

In addition, a Python-based automated pipeline is created in which new data is scraped, the text is pre-processed and vectorized, the dataset is updated, predictions are made, and visualizations are produced. This allows the system to stay updated and the model to improve as the model retrains on the new data. The results show that the hybrid models verify more accurate predictions than models based only structured data. Therefore, news sentiment alongside structured data has great potential for economic forecasting.

II. BACKGROUND AND RELATED WORK

In the context of India, however, there's limited research that blends real-time news sentiment with official economic data for predicting GDP growth. While global studies have explored this integration to some extent, the Indian economy—being complex and influenced by unique domestic and international factors—requires more focused investigation. This project addresses that gap by proposing a deep learning model that combines macroeconomic indicators from the Reserve Bank of India with insights extracted from financial news articles. By merging structured and unstructured data, the system aims to offer a more nuanced and accurate view of the country's economic performance.

However, despite their widespread use, these traditional models have several limitations. They tend to assume that past trends will continue into the future and often struggle to handle the complex, nonlinear relationships between various economic factors. More importantly, they are generally unable to respond quickly to unexpected events such as geopolitical tensions, pandemics, or sudden policy changes—factors that can significantly alter the economic landscape but may not be immediately reflected in structured datasets.

In response to these limitations, researchers have increasingly turned to machine learning and deep learning methods, which offer more flexibility and predictive power. Algorithms such as Decision Trees, Random Forests, Support Vector Machines, and Artificial Neural Networks are capable of learning complex patterns from large datasets. Particularly in time-series forecasting, models like Long Short-Term Memory (LSTM) networks have shown strong performance due to their ability to capture dependencies over time and adapt to dynamic patterns. These methods have opened new possibilities for economic forecasting, allowing for more adaptive and data-driven predictions.

Yet, even with the advancement of machine learning, most approaches still rely heavily on structured data. This means they often overlook a valuable and increasingly accessible source of insight: unstructured data in the form of economic news, expert commentary, and financial articles. News media often serves as a real-time reflection of public sentiment, market expectations, and evolving economic narratives—elements that structured datasets typically fail to capture. With the growth of Natural Language Processing (NLP), it has become possible to analyze these textual sources at scale, extracting sentiment, key topics, and economic signals that can complement traditional data.

Several studies, particularly in the domain of stock market prediction, have demonstrated the effectiveness of using sentiment analysis on news headlines and financial articles to improve forecasting accuracy. For example, news sentiment has been used to anticipate stock price fluctuations, market volatility, and even consumer confidence levels. These efforts have shown that when unstructured textual data is properly processed and integrated, it can provide predictive signals that traditional numerical data alone cannot offer. However, the application of this approach to GDP forecasting—especially within the Indian context—remains relatively underexplored. India's economy is uniquely influenced by a blend of domestic policies, global trade relationships, agricultural performance, and socio-political factors. News coverage often captures these nuances in real time, making it a potentially powerful tool for forecasting. Despite this, few models have been developed that combine structured macroeconomic indicators with unstructured news data to predict India's GDP.

This research aims to bridge that gap by proposing a hybrid system that leverages both data types. Structured indicators from the Reserve Bank of India are combined with financial news articles sourced from reputable online platforms. NLP techniques are applied to clean and analyze the news text, and a deep learning model is trained on the merged data. The ultimate goal is to improve the responsiveness and accuracy of GDP growth predictions by including context-rich, real-time sentiment data. In doing so, this work contributes to a more holistic approach to economic forecasting—one that acknowledges the value of both numbers and narratives in understanding economic trends.

III. PROPOSED ALGORITHM

In the evolving landscape of economic forecasting, relying solely on traditional statistical models is no longer sufficient to capture the complexities and dynamic nature of modern economies. The proposed solution in this research is a hybrid deep learning model that seeks to bridge this gap by intelligently combining both **structured data** (numerical economic indicators) and **unstructured data** (textual content from financial news articles). This dual-input approach is designed to enhance the model's understanding of the economy not just through past patterns but also through real-time economic sentiment, thereby providing a more holistic and timely forecast of India's GDP growth.

The key strength of this approach lies in its ability to integrate diverse sources of data, each offering unique insights into the health of the economy. Structured data sourced from the Reserve Bank of India (RBI) includes critical indicators such as the inflation rate, industrial output, interest rates, and prior GDP performance. These indicators are well-understood and widely used in econometric models, but they can be slow to reflect sudden changes or emerging risks. In contrast, unstructured data—primarily in the form of economic news articles—captures ongoing discussions, market reactions, policy decisions, and public sentiment, often in real time. By combining these two types of inputs, the model aims to compensate for the shortcomings of each, using the strengths of one to balance the limitations of the other. This integration allows the model to make more informed predictions that are both data-rich and context-aware.

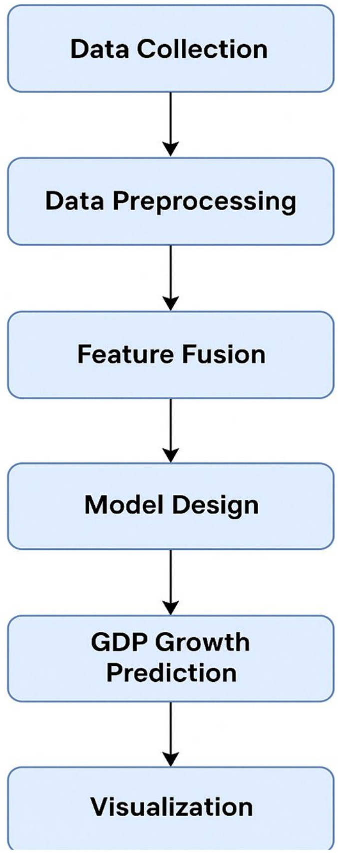

Fig1: Project workflow

A. *Data Collection and Integration*

The first stage involves gathering data from two primary sources:

- Structured Data: Key economic indicators such as inflation rate, repo rate, GDP history, and industrial production index are collected from the Reserve Bank of India (RBI) database.
- Unstructured Data: Economic and financial news articles are scraped from reliable online sources using Python's requests and BeautifulSoup libraries.

Both data sources are stored locally and updated regularly through an automated pipeline.

B. *Data Preprocessing*

Preprocessing is performed separately for each type of data:

- Structured Data is cleaned, normalized, and converted into numeric features ready for machine learning input.
- Unstructured Text Data undergoes Natural Language Processing (NLP), including tokenization, stopword removal, and lemmatization. The clean text is then vectorized using TF-IDF (Term Frequency–Inverse Document Frequency), converting the textual content into a matrix of meaningful numerical values that reflect the importance of terms across the corpus.

C. *Feature Fusion and Model Design*

Once preprocessing is complete, both structured and unstructured features are merged into a single feature matrix. The combined dataset is then passed through a **deep learning regression model** built using **TensorFlow**. The architecture consists of multiple dense (fully connected) layers with ReLU activation, optimized using the Adam optimizer. Dropout layers are included to prevent overfitting.

The model is trained using historical GDP growth values as targets, allowing it to learn complex relationships between macroeconomic indicators and economic sentiment derived from news.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

D. Prediction and Evaluation

Once trained, the model predicts future GDP growth based on the latest inputs. Performance is measured using evaluation metrics such as:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2 Score

These metrics help evaluate both the precision and the generalizability of the model.

E. Workflow Automation

To ensure scalability and real-time updates, a Python-based automation pipeline was built to:

- Scrape new data regularly
- Apply preprocessing steps
- Update feature matrices
- Re-train the model (if required)
- Generate predictions
- Visualize results on dashboards

The workflow follows the sequence:
Scrape → Clean → Feature Engineering → Predict → Visualize
This modular pipeline makes the system easy to maintain and extend, supporting future integration into web-based platforms.

F. Visualization

Final predictions and trends are visualized using Python libraries like Matplotlib and Seaborn. These visual dashboards provide a clear understanding of economic trends and model confidence, offering actionable insights for decision-makers and researchers.

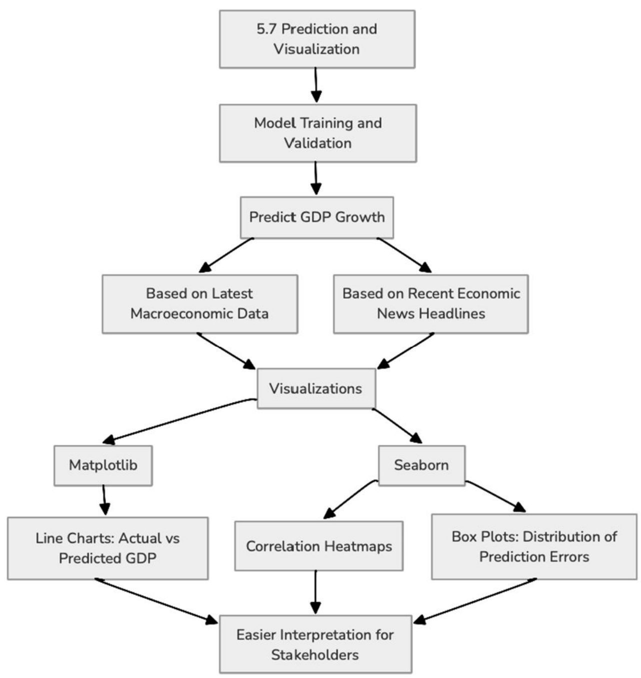

Fig2: Flow Chart

IV. IMPLEMENTATION AND RESULTS

To bring the proposed GDP prediction model to life, we implemented it step-by-step using a combination of data science, machine learning, and natural language processing tools. The entire system was developed using Python because of its robust ecosystem for data handling, model development, and automation. The goal was not only to build a high-performing model but also to create an automated pipeline that could update itself with new data over time, ensuring consistency and scalability.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

We started by gathering data from two sources: structured numerical data from the Reserve Bank of India (RBI) and unstructured textual data from economic news websites. For this, we used web scraping tools like **BeautifulSoup** and **Selenium** to collect recent financial news articles. Simultaneously, we downloaded macroeconomic indicators such as inflation rates, interest rates, GDP history, and industrial outputs from the RBI portal and organized them into structured tables.

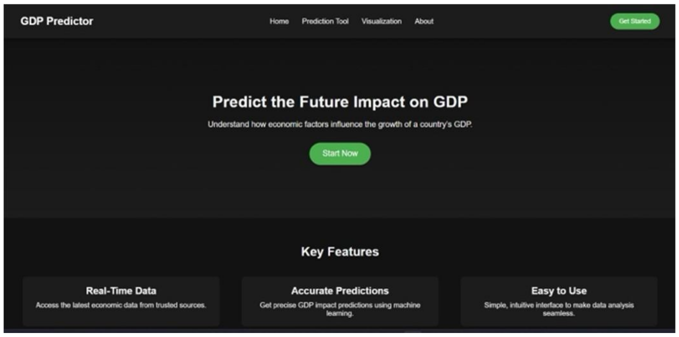

Fig3: Homepage of the GDP Predictor

Once the data was collected, we moved on to the preprocessing phase. The structured data was normalized to ensure consistent scaling across features. For the unstructured text data, we used NLP techniques such as tokenization, stopword removal, and lemmatization to clean the news content. We then applied TF-IDF vectorization to convert this cleaned text into a numerical form that could be processed by the machine learning model.

With both datasets ready, we merged the structured indicators and the TF-IDF vectors into a unified feature set. This combined dataset was fed into a deep learning regression model developed using TensorFlow and Keras. The model architecture included multiple dense layers with ReLU activations and dropout for regularization, trained using the Adam optimizer. We tuned various parameters to improve accuracy and reduce error during training.

After training, the model was able to predict future GDP growth based on the most recent economic indicators and news sentiment. The results were promising—when compared with models trained on only structured or unstructured data, our hybrid model consistently performed better. It was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score, all of which indicated strong predictive capability.

To help interpret the results, we created visualizations using Matplotlib and Seaborn. These included graphs showing actual versus predicted GDP trends and the influence of sentiment changes over time. A local SQLite database was used to store predictions and processed inputs, with the whole pipeline designed to run automatically on a scheduled basis—making it easy to update the system with fresh data whenever needed.

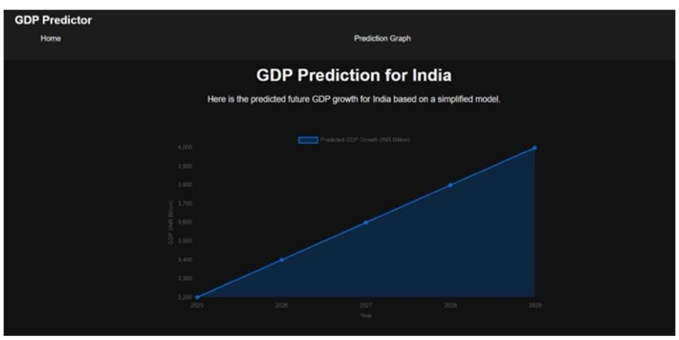

Fig3: Graph displaying predicted GDP growth in India

V. CONCLUSION AND FUTURE SCOPE

This project aimed to bridge the gap between traditional economic forecasting methods and modern data-driven approaches by creating a hybrid system that predicts India's GDP growth using both structured data (like official RBI indicators) and unstructured data (news articles from reliable financial sources). The system effectively combines numerical economic variables with textual sentiment extracted from current events to provide a more holistic and real-time view of the country's economic health.

The development process included building a robust data pipeline to collect and preprocess diverse data types, applying TF-IDF and sentiment analysis techniques to extract relevant insights from news, and feeding this data into machine learning models such as

Linear Regression and Random Forest for GDP prediction. The results were further visualized using dynamic graphs and a lightweight web interface, making the entire prediction workflow both accessible and interpretable to end users. This integration of technical depth with user experience highlights the system's practicality and real-world value.

However, the journey does not end here. There is significant scope to enhance the system's performance and reach. For instance, incorporating deep learning models like LSTM (for time series analysis) or transformer-based models like BERT (for better understanding of news semantics) could substantially improve prediction accuracy. The system can also be expanded to support multilingual news sources, covering a wider range of sentiments from diverse regions of India. Additionally, integrating real-time streaming data (e.g., live economic feeds or news APIs) would make the predictions more up-to-date and relevant for policymakers or investors.

From a deployment perspective, converting this prototype into a fully functional, cloud-hosted application with role-based access and data security features would enable institutions such as research bodies, government departments, or financial firms to utilize it in decision-making processes. Furthermore, incorporating predictive confidence scores and "what-if" scenario testing could make the tool more interactive and powerful for economic planning.

In conclusion, this project not only showcases the potential of hybrid machine learning models in economic forecasting but also opens doors to intelligent, scalable, and real-time GDP prediction systems that adapt to changing economic landscapes.

REFERENCES

[1] Reserve Bank of India, "Handbook of Statistics on Indian Economy," [Online]. Available: https://www.rbi.org.in
[2] Ministry of Statistics and Programme Implementation, Government of India, "National Accounts Statistics," [Online]. Available: https://mospi.gov.in
[3] A. S. Aslam, S. J. Khan, and R. M. Noor, "GDP prediction using machine learning and sentiment analysis of economic news," Procedia Computer Science, vol. 199, pp. 345-352, 2022.
[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
[6] M. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.
[7] Google Developers, "TF-IDF Text Feature Extraction," [Online]. Available: https://developers.google.com/machine-learning/guides/text-classification/step-3
[8] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," arXiv preprint arXiv:1908.10084, 2019.
[9] K. Clark, M. Luong, Q. V. Le, and C. D. Manning, "ELECT