A Comprehensive Review of Explainable AI Models for Reliable Decision Support in GST Revenue Prediction

[1]Prashant Chandra, [2]Adarsh Raushan, [3]Dr. Vivek Richhariya
[1]M.Tech. Scholar, [2]Assistant Professor, HOD & Professor
CSE Department, LNCT, Bhopal, India

Abstract: Goods and Services Tax (GST) revenue prediction is crucial for fiscal planning, policy formulation, and resource allocation. Traditional econometric forecasting methods often lack flexibility and interpretability when modeling complex multivariate economic data. With the advent of machine learning (ML), predictive accuracy has improved substantially; however, many high-performance models behave as "black-boxes," rendering their internal decision logic opaque to stakeholders. Explainable Artificial Intelligence (XAI) aims to bridge this gap by providing transparent, interpretable models that support reliable decision support systems. This review explores recent developments in explainable AI models applied to GST revenue prediction, with a focus on model interpretability, trustworthiness, and performance. We summarize feature-attribution techniques such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and rule-based approaches in conjunction with predictive algorithms including XGBoost, Random Forests, and Neural Networks. Challenges, limitations, and future directions are also discussed.

Keywords: GST Revenue Prediction, Explainable Artificial Intelligence (XAI), Machine Learning (ML), SHAP (SHapley Additive ExPlanations), Predictive Algorithms.

How to cite this article: *Prashant Chandra, Adarsh Raushan, Dr. Vivek Richhariya. (2026). A Comprehensive Review of Explainable AI Models for Reliable Decision Support in GST Revenue Prediction. International Journal of Scientific Modern Research and Technology, ISSN: 2582-8150, Volume-21, Issue-03, Number-02, Jan-2026, pp.08-12, URL: https://www.ijsmrt.com/wp-content/uploads/2026/05/IJSMRT-26010102.pdf*

IJSMRT-26010102

I. INTRODUCTION

GST, introduced as a unified indirect tax regime, has fundamentally reshaped India's tax structure. Accurate revenue prediction assists governments in budgeting, managing deficits, and evaluating the impact of policy changes. Although traditional statistical models (ARIMA, VAR, etc.) have been widely used, they struggle with nonlinear dependencies, high dimensionality, and interactions between economic indicators (Thayyib et al., 2023).

Machine learning models, such as gradient boosting, random forests, and neural networks, have demonstrated superior performance in forecasting tasks. However, policymakers are increasingly demanding explainability to build trust in automated recommendations. The field of explainable AI (XAI) provides tools that interpret model behavior and produce human-understandable explanations. This review consolidates research on explainable AI models specifically tailored to GST revenue forecasting and decision support (Ranjan et al., 2023).

The main contributions of this paper are as follows:

- The paper reviews Explainable AI (XAI) techniques, such as SHAP and LIME, for transparent GST revenue prediction.
- It compares various machine learning models (e.g., XGBoost, Random Forests) for their predictive accuracy and interpretability.
- Discusses the trade-off between model complexity and interpretability in forecasting models.
- Highlights challenges like data quality and regional heterogeneity in GST revenue prediction.
- Proposes future research into hybrid models, real-time decision support, and advanced SHAP applications.

This paper reviews the integration of Explainable AI (XAI) models, particularly SHAP**,** for GST revenue prediction**.** It explores machine learning techniques such as XGBoost**,** compares their predictive performance, and highlights the importance of model interpretability for decision support systems**.** Challenges, future research directions, and practical applications are also discussed.

II. LITERATURE REVIEW

Traditional Forecasting Methods

Traditional forecasting methods for GST revenue prediction primarily rely on statistical models that capture historical data patterns and economic trends. Commonly used techniques include: ARIMA models are widely used for time series forecasting. They assume that future GST revenue is a linear function of past values and errors. ARIMA is particularly useful for univariate time series but may struggle with nonlinear relationships between features (Liu et al., 2020). The vector autoregression (VAR) method extends ARIMA to handle multivariate time series. It captures the interdependencies between multiple variables, such as GDP growth, inflation, and tax rates, making it suitable for modeling the impact of various economic indicators on GST revenue (Uduak-Obong & Usoro, 2025). Exponential smoothing (ETS) models focus on forecasting by weighting recent observations more heavily. These models are useful for short-term forecasts, particularly when trends and seasonality are present (Ostertagová & Ostertag, 2012). Although effective, these methods often struggle with the complexity and nonlinearities present in large dynamic datasets, such as those used for GST forecasting, prompting a shift to machine learning models for more accurate predictions.

Machine Learning Approaches

Machine learning approaches have significantly improved GST revenue prediction by capturing complex, nonlinear relationships in large datasets. Commonly used methods include: Random Forest is an ensemble technique based on decision trees; random forest builds multiple trees and aggregates their predictions to reduce overfitting. It handles both regression and classification tasks and is robust to outliers, making it suitable for predicting GST revenue based on numerous economic features (Chen et al., 2025). XGBoost is a highly efficient gradient boosting model that builds trees sequentially, with each correcting the errors of its predecessor. It is known for its high predictive accuracy, scalability, and ability to manage missing data. XGBoost is widely used for time-series forecasting and is ideal for predicting GST revenue (Qin, 2022). Deep learning models, such as long short-term memory (LSTM) networks, can capture temporal dependencies in GST data and are particularly useful for modeling complex trends and seasonality (Sarkar et al., 2025). Machine learning techniques offer superior predictive accuracy; however, they often lack interpretability, necessitating explainable AI (XAI) models such as SHAP to provide transparency (Kedar, 2024).

Explainability in Forecasting

Explainability in forecasting is critical for ensuring that machine learning models, such as those used for GST revenue prediction, are transparent and understandable. Although models like XGBoost and neural networks offer high accuracy, they often operate as "black boxes," making it difficult for users to comprehend how predictions are made. This lack of transparency can lead to mistrust, especially in high-stakes decision-making environments, such as tax policy planning (Joshi, 2025). To address this, explainable AI (XAI) techniques, such as SHapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME), are used to break down complex models into understandable components. These techniques provide feature importance, showing how individual features (e.g., economic indicators and tax rates) influence the model's predictions. SHAP assigns a contribution value to each feature, whereas LIME approximates the decision-making process of the model using simpler and interpretable models for individual predictions (Arunika et al., 2024).

III. EXPLAINABILITY TECHNIQUES

SHAP (SHapley Additive ExPlanations)

SHAP explains how each feature contributes to the prediction. Given features $x_1, x_2, ..., x_n$, SHAP values $\phi_i$are computed using:

$$\hat{y} = \phi_0 + \sum_{i=1}^{n} \phi_i$$

Where ϕ_0 is the expected model output. SHAP provides both local and global interpretation.

LIME

LIME approximates a model locally with a simple interpretable surrogate. For a prediction $f(x)$, LIME finds a linear surrogate $g(x)$minimizing:

$$\text{argmin}_g \ L(f, g, \pi_x)$$

Where π_x represents the proximity of samples to the instance x.

IV. COMPARATIVE ANALYSIS

This is a comparative analysis of traditional and machine learning methods used for GST revenue prediction in terms of methods, conclusions, and limitations:

Table 1: Comparative Analysis

Method	Type	Key Features	Conclusion	Limitations
ARIMA (AutoRegressive Integrated Moving Average)	Traditional Statistical Model	Good for time-series forecasting, assumes linear relationships	Useful for short-term forecasts but struggles with complex, non-linear relationships	Does not handle external predictors well, assumes stationary data, limited to univariate data
VAR (Vector Autoregression)	Traditional Statistical Model	Handles multivariate time series, captures interdependencies	Suitable for modeling relationships between multiple economic variables	Assumes linear relationships, doesn't handle large datasets well, computationally intensive
Exponential Smoothing (ETS)	Traditional Statistical Model	Focuses on smoothing and trend, seasonal data emphasis	Effective for short-term, seasonal forecasting	Struggles with non-linear trends, not suitable for long-term forecasting
Random Forest	Machine Learning Model	Ensemble method using multiple decision trees, robust to overfitting	Can handle complex interactions and large datasets, interpretable to an extent	Computationally expensive, may overfit with too many trees, less interpretable compared to simpler models
XGBoost (Extreme Gradient Boosting)	Machine Learning Model	Gradient boosting, handles missing data, high predictive accuracy	Very effective for large-scale and high-dimensional data, offers excellent prediction accuracy	Lack of interpretability, requires careful tuning to avoid overfitting

Neural Networks (e.g., LSTM)	Machine Learning Model	Deep learning model that captures temporal dependencies	Effective for modeling long-term dependencies, good for time series	Complex to train, require large datasets, lack interpretability, computationally expensive

V. CHALLENGES & LIMITATIONS

Challenges & Limitations in GST revenue prediction using machine learning include:

Data Quality: Missing values, inconsistencies, and incorrect entries hinder accurate modeling.

Interpretability vs. Performance: High-performing models like XGBoost and Neural Networks often lack interpretability, which can reduce stakeholder trust.

Regional Heterogeneity: GST collections vary across regions, requiring models to account for local economic conditions.

Feature Selection: Selecting relevant features for complex models is challenging.

Computational Cost: Machine learning models, especially deep learning, can be computationally expensive and require large datasets to function optimally.

VI. FUTURE DIRECTIONS

Hybrid Models: Combining deep learning models like LSTMs for capturing long-term dependencies with tree-based methods like XGBoost can enhance predictive accuracy while maintaining interpretability. These hybrid models will better handle both non-linear patterns and provide transparent decision-making processes, ensuring reliable forecasts.

Real-Time Decision Support Systems: Developing real-time GST revenue prediction models will allow continuous monitoring of economic indicators, tax policy changes, and other dynamic factors. By integrating these systems into government dashboards, policymakers can adjust fiscal strategies on-the-fly based on up-to-date predictions.

Automated Feature Engineering: Utilizing AutoML tools for feature selection and engineering will streamline the model-building process. Automated methods will adapt to new data inputs and provide optimal feature sets, reducing the manual effort required for model development and improving model performance.

Advanced SHAP Applications: Exploring advanced uses of SHAP for multi-modal models will help interpret complex data types, such as combining numerical data with unstructured inputs like text or sentiment analysis, enhancing model transparency.

Handling Noisy and Imbalanced Data: Future research should focus on developing methods to filter noise and address data imbalance, common in economic data, ensuring more accurate and reliable predictions.

Policy Simulations: Incorporating policy simulations into the model will help predict the potential impacts of changes in tax rates or exemptions, offering valuable insights into the effects of policy adjustments on GST revenue.

Global Comparative Models: Extending GST revenue prediction models to a global level will enable comparisons between countries, offering valuable cross-national insights for global tax forecasting strategies.

These directions will enhance both the accuracy and transparency of GST revenue forecasting models, making them more effective for decision-making and policy formulation in dynamic economic environments.

VII. CONCLUSION

In conclusion, this paper highlights the importance of Explainable AI (XAI) in GST revenue prediction. While machine learning models like XGBoost offer superior predictive accuracy, their lack of transparency poses challenges in real-world applications. By integrating SHAP and other explainable techniques, the study demonstrates how model interpretability can enhance trust and enable informed decision-making for policymakers. Future research should focus on hybrid models, real-time systems, and automated feature engineering to improve both accuracy and transparency. These advancements will ensure that GST revenue forecasting is not only precise but also trustworthy and actionable.

REFERENCES

[1] Arunika, M., Saranya, S., Charulekha, S., Kabilarajan, S., & Kesavan, G. (2024). *A Survey on Explainable AI Using Machine Learning Algorithms Shap and Lime*. 1–6. https://doi.org/10.1109/icccnt61001.2024.10725120

[2] Chen, M., Yu, S., & Deng, Y. (2025). *Research on Large-Scale Data Mining and Machine Learning Model for Accurate Prediction*. 1444–1448. https://doi.org/10.1109/edpee65754.2025.00260

[3] Joshi, S. (2025). Gradient Boosting and Explainable AI for Financial Risk Management: A Comprehensive Review. In *Preprints.org*. Mdpi Ag. https://doi.org/10.20944/preprints202504.0817.v1

[4] Kedar, M. M. M. (2024). Exploring the Effectiveness of SHAP over other Explainable AI Methods. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, *08*(06), 1–5. https://doi.org/10.55041/ijsrem35556

[5] Liu, T., Liu, S., & Shi, L. (2020). *ARIMA Modelling and Forecasting* (pp. 61–85). Springer Singapore. https://doi.org/10.1007/978-981-15-0321-4_4

[6] Ostertagová, E., & Ostertag, O. (2012). Forecasting using simple exponential smoothing method. *Acta Electrotechnica et Informatica*, *12*(3). https://doi.org/10.2478/v10198-012-0034-2

[7] Qin, R. (2022). The Construction of Corporate Financial Management Risk Model Based on XGBoost Algorithm. *Journal of Mathematics*, *2022*(1), 1–8. https://doi.org/10.1155/2022/2043369

[8] Ranjan, M. K., Roy, K., Mahind, A., Bhavsar, D., Nemade, S., & Sattar, A. M. (2023). Unlocking the Potential: Machine Learning Driven Explainable AI Applications Across Diverse Domains. *International Journal of Software Computing and Testing*. https://doi.org/10.37628/ijosct.v9i1.948

[9] Sarkar, A., Bhattacharjee, V., & Pandey, P. (2025). An Empirical Analysis of Deep Learning Models for Temperature Prediction. *International Journal of Experimental Research and Review*, *47*, 12–24. https://doi.org/10.52756/ijerr.2025.v47.002

[10] Thayyib, P. V., Thorakkattle, M. N., Usmani, F., Yahya, A. T., & Farhan, N. H. S. (2023). Forecasting Indian Goods and Services Tax revenue using TBATS, ETS, Neural Networks, and hybrid time series models. *Cogent Economics & Finance*, *11*(2). https://doi.org/10.1080/23322039.2023.2285649

[11] Uduak-Obong, A. B., & Usoro, A. E. (2025). Application of Multivariate Time Series Analysis to Modelling of Total Tax Revenue and Some of Its Components. *International Journal of Mathematics and Statistics Studies*, *13*(3), 1–19. https://doi.org/10.37745/ijmss.13/vol13n3119