Journal of Theoretical and Applied Information Technology
31st May 2025. Vol.103. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ROBUST QUANTILE REGRESSION-BASED MACHINE LEARNING FRAMEWORK FOR OUTLIER-RESILIENT TIME SERIES ANALYSIS

Mr. TUSHAR MEHTA[1] , DR.DHARMENDRA PATEL[2]

[1,2] Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, Charotar University of Science and Technology Changa, Gujarat, India.

E-mail: [1]tusharmehta.mca@charusat.ac.in, [2]dharmendrapatel.mca@charusat.ac.in

ABSTRACT

Time series analysis is a powerful tool in countless regions, from finance to healthcare, but is often challenged by the presence of outliers that can distort predictions and model output. This article presents a robust quantile regression-based framework for machine learning, which increases the resilience of time series analysis compared to outliers. By using quantile regression, the proposed framework captures the conditional distribution of time series data and provides a more comprehensive understanding of its underlying structure. Machine learning integration improves the ability of models to adapt to complex nonlinear patterns while simultaneously maintaining robustness to anomaly data points. Through extensive research into synthesis and practical datasets, we show that our framework outweighs traditional predictability and trigger resilience methods. The results highlight the possibility of combining quantile regression with machine learning for robust time series analysis, providing promising directions for future research and applications in the environment. Time series forecasts play a key role, especially in financial markets where accurate forecasts are useful for investors and stakeholders. However, traditional models have difficulty recording nonlinear dependencies and are not able to effectively handle outliers. This article presents a robust quantile regression-based machine learning framework for diffusion-preserving time series analysis. It integrates long-term time memory (LSTM), LightGBM, and stacked ensemble models. The proposed ensemble approach uses quantile regression for robust outsourcing processing while combining deep learning strengths to increase base techniques to improve predictive performance. Experimental evaluation of Goldman Sachs BDC, Inc.(GSBD) shared course data demonstrates the advantages of the ensemble model compared to the individual model. The results show that the stacked ensemble model reaches the lowest flipper loss (0.0656), MAE (0.1313), RMSE (0.2185), and the highest R^2 (0.9778), exceeding LSTM and LightGBM. The results highlight the effectiveness of hybrid ensemble learning in financial series forecasting, providing a more accurate and robust approach to dealing with outlier sacrificial data.

Keywords: *Time Series Analysis, Quantile Regression, Outlier Resilience, Machine Learning Framework, Robust Prediction*.

1. INTRODUCTION

In areas such as banking, health, energy, and climate research, time series analysis is the pillar of decision-making. Forecasting and strategic planning rely heavily on the ability to spot trends, patterns, and anomalies in continuous data. Nevertheless, traditional time series models are seriously questioned by the presence of outliers and the presence of extreme or unusual data values. Especially in situations where data integrity is very important, outlier predictions can be skewed, reducing the output of the model and leading to false conclusions. Therefore, there is a high demand for powerful technologies, both in research and practical terms, that can be managed efficiently while maintaining high prediction accuracy at the same time.

Machine learning methods (ML) have been a more effective tool for time series analysis in recent years, as it provides the adaptability of complex, nonlinear interactions. It is known that their long-term memory (LSTM) networks for their ability to recognize temporal dependencies of continuous data are known. LGBM models (light gradient boosting machines) use structured data to demonstrate outstanding skills and produce

interpretable results. These models are powerful, but of course, these models are not intended to solve the problems presented by outliers. This limitation highlights the need for hybrid systems that combine the benefits of modern machine-learning approaches with statistical methods. Quantile regression, a powerful statistical method, provides a possible answer to this question. Quantile regression models the conditional quantum of response variables, in contrast to traditional regression techniques, highlighting the mean of the dependent variable. This feature is particularly suitable for research without research, as it provides a more comprehensive knowledge of data distributions that include tails. The use of the properties of both paradigms, the robustness of quantile regression, and the predictability of machine learning through the combination of quantile regression and machine learning models, including LSTM and LGBM is practical.
This work beats a unique and robust quantile regression-based framework for machine learning before series analysis using outliers. LSTM, LGBM, and ensemble hybrid methods are combined in context to improve the accuracy and robustness of predictions against outliers.
The LGBM model pays for structured functionality management and interpretable knowledge specifications, while the LSTM components record temporary patterns and long-term dependencies of the data. The combined method guarantees resilience to outliers by combining these models with quantile regression to improve general performance. Through thorough experiments on synthesis and actual data records, we demonstrate the advantages of the proposed framework compared to traditional approaches on predictive accuracy and outliers.

In Section 2, relatives work in time series analysis. The entire publication discusses quantile regression and machine learning. Section 3 describes the proposed architecture, including ensemble methods, LSTM, and LGBM integration. Section 4 provides experimental designs, data records, and results. Section 5 concludes the work with a discussion of results and potential research directions.
This work aims to open paths for more reliable and accurate time series analysis in a context that tends to bridge the gap between statistical approaches and advanced technologies for machine learning.

2. LITERATURE REVIEW

Jiang, Z et al. [1] develops Centra-Net to optimize visual locations with a variety of data records, simultaneously maintaining storage efficiency. There is a basic feature extraction unit (BFEU) with parallel branching for local property extraction and attention-based calibration. The Filter Company Sharing Mechanism (FSM) adaptively manages parameter releases, while the Complexity Promotional Algorithm (CPGA) compensates for learning based on task complexity. Large-scale experiments show that Centra-net surpasses cutting-edge models with fewer parameters. Gong, J et al. [2] proposes a hybrid ensemble model combining TCN, LSTM, and LightGBM to combine the model costs using several linear regressions. The ensemble model is evaluated on Australian and Chinese data records and reduces MAE by 4.88% and 28.95% compared to TCN-LSTM or LightGBM, beyond the individual models. The proposed multiple linear regression-ensemble method continues to improve by 3.64% and RMSE. This demonstrates its effectiveness in improving predictive performance. Zhao, T et al. [3] suggest improving prediction accuracy by using the ability of LightGBM to efficiently deal with nonlinear relationships and prevent over-management while also using the ability of TCN to capture seasonal trends. It suggests the LIGHTGB model. Experimental results show that TCN-LightGBM surpassed the conventional model, achieving 91.3°C narration with an MSE of 0.021 and MAE of 0.115. Additionally, parameter sensitivity analysis confirms the robustness of the model and provides a reliable means of predicting agricultural trade and other forecast tasks in time series. Maharina, M et al. [4] address the challenges of flood prediction in Jakarta, Indonesia using machine learning models (ML) to improve the accuracy of forecasts using weather data. As traditional models combat Jakarta's highly diverse climate, this study evaluates logistics regression (LR), LightGBM and XGBoost based on five performance metrics. The proposed approach includes data preprocessing, functional engineering, and overloading with cross-validation. This greatly improves performance. Results show that LightGBM reached 96.81°Curacy, Xgboost 96.67%, and LR 93.82%, indicating the effectiveness of the model in flood risk management and damage limits. Zhang, C et al.[5] proposes a form-supported EE-LightBM model that integrates sub-techniques and hybrid ensemble strategies to improve learning from limited error samples. The SHAP method enhances

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

explainability by identifying key features contributing to faults. Evaluation of real-world network data shows high detection accuracy (0.9968), F1-score (0.9711), and low false negative (0.0033) and false positive (0.0032) rates, demonstrating the model's effectiveness in fault diagnosis. Airlangga, G. [6] explores machine learning techniques for tsunami prediction using seismic data from 1995 to 2023. Various models, including ensemble methods (Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost) and Neural Networks, were evaluated using 10-fold cross-validation. Gradient Boosting achieved the highest ROC-AUC (0.8606) and a balanced F1-score (0.6544), while Random Forest excelled in precision (0.6920). Neural Networks underperformed, highlighting the need for further optimization. The study confirms that ensemble methods, particularly Gradient Boosting and Random Forest, are effective for tsunami prediction, aiding early warning systems. Future improvements will focus on enhancing recall and exploring hybrid models. Kumar, S et al. [7] present a hybrid prediction model that integrates LightGBM, XGBoost, Lasso regression, and random forests. This model surpasses basic line regressors over some valuation metrics, showing excellent accuracy and robustness. The research uncovers the possibilities of hybrid models in predictive analytics and proposes future inspections of additional algorithms and optimization techniques for further improvement. Kumar, M. S et al. [8] introduce an advanced ensemble stacking approach that summarises LightGBM, Plsregrechente, and ElaticNet to improve predictive modeling. LightGBM guarantees efficiency, PLSREGECHEN helps in feature selection, and ElasticNet guarantees regularization to prevent overhanging. Empirical results show superior performance compared to individual models, indicating the potential of this ensemble method to improve predictability in complex, real-world applications. Hola, A et al. [9] examine predictions (PDM) using IoT sensor data and machine learning to predict equipment failures, optimize maintenance plans, and reduce downtime. Compare regression models including LSTM, XGBoost, and LightGBM based on RMSE and CPU usage by analyzing CMAPSS data records from NASA. This study concludes that LSTM represents optimal performance for predicting useful life (RULL) and demonstrates its effectiveness in predictive expectations of aircraft engines. Sui, M et al. [10] study enhances retail investors' stock price predictions by integrating deep learning and time series models. Using techniques such as DNN, LightGBM, LSTM, GRU, and linear regression, the proposed ensemble model improves accuracy by preprocessing data, functional engineering, and adjusting hyperparameters. The results show that the ensemble approach goes beyond the individual models and provides a robust forecasting solution for retail investors. Examines approaches to machine learning to predict harmful algae blooms (HABs). Ahn, J. M et al. [11] propose a significant risk to ecosystems and human health. Compare gradient boosting models (XGBoost, LightGBM, CatBoost) with note-based CNN-LSTM models. It uses Bayesian optimization, and back and stacking techniques to improve accuracy. The results show that ensemble techniques improve predictive power by easing the limitations of individual models and highlighting the possibilities of AI-controlled prediction in HABS management. Gupta, V., & Kumar, E. [12] suggest an optimized LightGBM model.This is improved with the Harris Hawk Optimization (H3O) H3O algorithm for prediction. By using exclusive bundles of features (EFB) and gradient-based single-sided samples (GOSD), the model reduces over-adaptation and improves prediction accuracy. Furthermore, the Markov chain model improves residual errors, leading to improved stock trading and economic prognosis decisions. Experimental results demonstrate the effectiveness of this hybrid approach in predicting market trends and economic growth with high accuracy. Hartanto, A. D. et al. [13] propose a prediction model using LightGBM models (Light Gradient Boosting Machine), known for their accuracy and efficiency. By optimizing hyperparameters with Grid Search Cross Validation (GSCV), LightGBM studies are compared to other boost models such as Xgboost, Adaboost, and CatBoost. The experimental results show that LightGBM exceeds its counterparts, highlighting the potential for stock market forecasts. Future work will show that data preprocessing will be improved, external variables will be improved, and prediction accuracy will be improved.Kumar, S[14] proposes a framework for accurate prediction of the final price of Indian investment funds using an automated Arimax model. This compares the performance of LightGBM with Auto-Arimax and Facebook Prophet on the same data set. This indicates that the proposed model exceeds the alternative. Evaluations based on MSE, RMSE, and MAPE confirm excellent accuracy. Future work will show that additional parameters can be included to improve time series forecasting. Kumar, S et al. [15] present a new hybrid ensemble learning approach that uses a stack of regressors to improve price forecasts for investment funds in Indian financial markets. Traditional forecasting methods struggle to capture complex nonlinear relationships in financial data, which the proposed ensemble model addresses by leveraging multiple base learners. Ridge regression is employed as the meta-regressor, achieving superior predictive accuracy with metrics such as an R^2 score of 0.99998 and a low RMSE of 0.0041. The study demonstrates that the stacking-based ensemble model outperforms individual models, making it valuable for portfolio optimization and risk reduction.

2.1 Objectives

1.Develop a robust quantile regression-based machine learning framework capable of effectively modelling time series data in the presence of outliers.
2.Enhance the accuracy and reliability of forecasts by integrating multiple models and leveraging ensemble techniques.
3.Rigorously evaluate the proposed framework against traditional methods using real-world datasets and comprehensive error metrics.
4.Provide a scalable and interpretable tool to support decision-making in critical applications such as financial forecasting and resource management.

3. Research Methodologies

Figure 1. Database Diagram

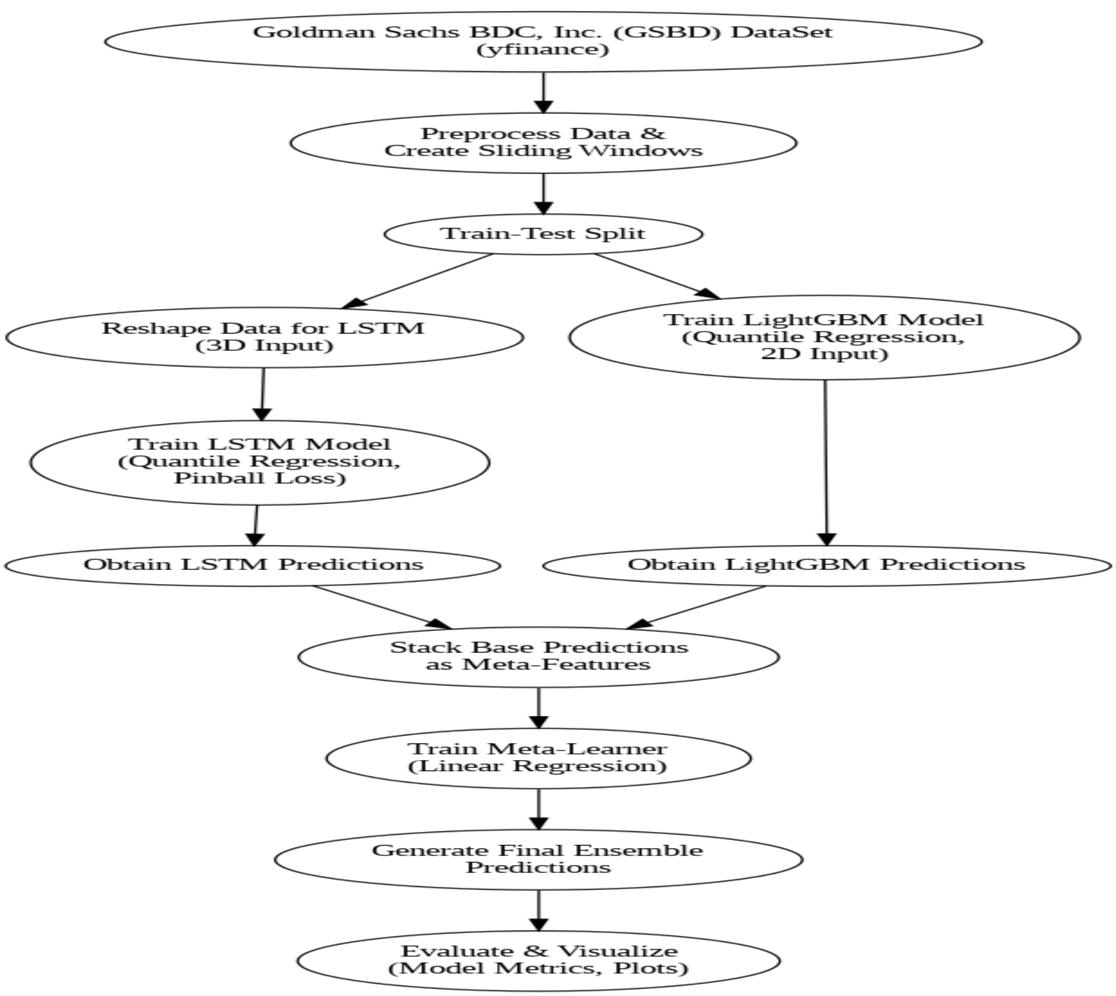
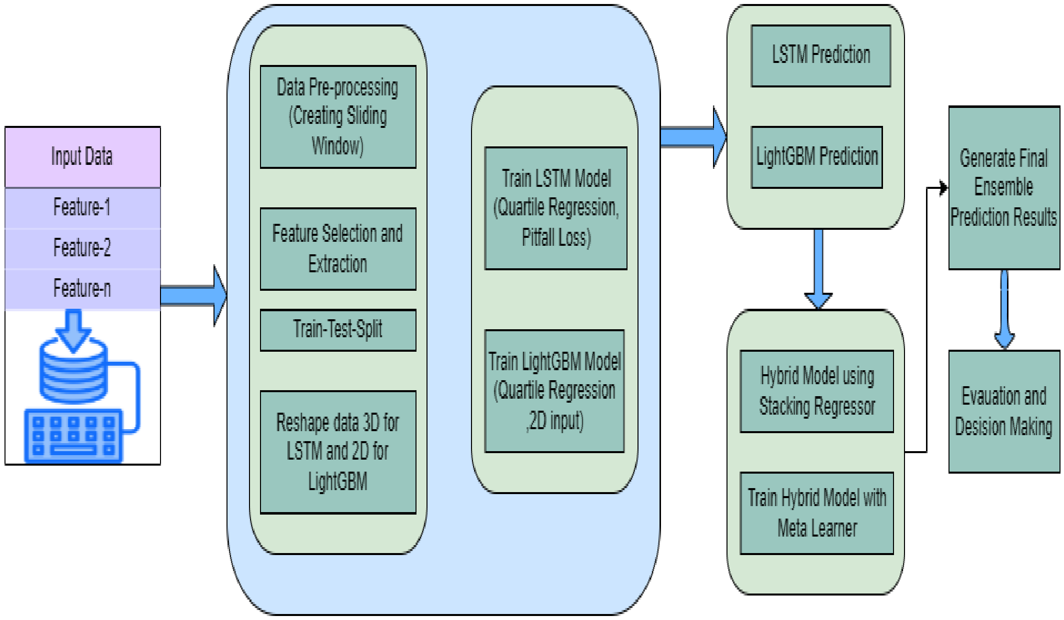


Figure 2. The Architecture of the Proposed Model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



3.1 Data sample GSBG Stock Data

S. N	Date	Close	High	Low	Open	Volume
0	2018-01-02	10.659293174743652	10.67815036816985	10.47543036841897	10.54614731657642	181100
1	2018-01-03	10.536713600158691	10.701718240720147	10.503712851886991	10.631001323905988	111500
2	2018-01-04	10.536713600158691	10.61968665318384	10.5136130763685	10.531999079090884	76900
3	2018-01-05	10.352852821350098	10.541430102697431	10.286851312400403	10.541430102697431	161300
4	2018-01-08	10.272706985473633	10.414139024422864	10.173703838846793	10.414139024422864	204700
1776	2025-01-27	12.8100004196167	12.960000038146973	12.739999771118164	12.800000190734863	1215600
1777	2025-01-28	12.859999656677246	12.970000267028809	12.779999732971191	12.779999732971191	961200
1778	2025-01-29	12.609999656677246	12.89900016784668	12.5600004196167	12.859999656677246	657900
1779	2025-01-30	12.720000267028809	12.770000457763672	12.604999542236328	12.710000038146973	734800
1780	2025-01-31	12.84000015258789	12.850000381469727	12.694999694824219	12.729999542236328	847500

LSTM

Below is an overview of the mathematical foundation underlying Long Short-Term Memory (LSTM) networks. LSTM is a type of recurrent neural network (RNN) specifically designed to address the vanishing (and exploding) gradient problems encountered when learning long-term dependencies. The key innovation in LSTM is the introduction of a memory cell and gating mechanisms that regulate the flow of information. Consider an LSTM unit at the time step t with input vector x_t previous hidden state h_{t-1}, and previous cell state c_{t-1}. The following equations. describe the operations within an LSTM cell:

Forget Gate

The forget gate f_t determines which information to discard from the previous cell state.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

where $\sigma(.)$ is the logistic sigmoid function, W_f is the weight matrix for the forget gate, and b_f is the bias vector.

Input Gate and Candidate Cell State:

The input gate i_t controls which new information to add, and the candidate cell state $\tilde{c}_t$ represents potential new content to be integrated.

$$i_t = \sigma(W_i \cdot [h_{t-1}] + b_i) \quad (2)$$

$$\tilde{c}_i = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

Here, W_i and W_C are weight matrices for the input gate and candidate state, respectively, and b_i and b_c are their corresponding biases.

Cell State Update:

The new cell state c_t is computed by combining the previous cell state and the candidate cell state, modulated by the forget and input gates.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_i \quad (4)$$

Output Gate and Hidden State:

Finally, the output gate o_t decides what part of the cell state should be output. The hidden state h_t is then computed as:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

where W_o and b_o are the weight matrix and bias vector for the output gate.

LightGBM

LightGBM's mathematical foundation is rooted in the gradient boosting decision tree (GBDT) framework, and it leverages several key innovations to enhance both efficiency and accuracy:

Gradient Boosting and Second-Order Approximation:

LightGBM minimizes an overall objective function that combines a loss function with a regularization term. At each iteration, it fits a new decision tree to the negative gradients (first-order derivatives) of the loss function. To improve convergence, it uses a second-order Taylor expansion:

$$L^{(t)} \approx \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_i(x_i)^2] + \Omega(f_t) \quad (7)$$

Where $g_i = \frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}$ and $h_i = \frac{\partial^2 L(y_i, f(x_i))}{\partial f(x_i)^2}$ are the gradient and Hessian for instance iii, and is a regularization term for the tree.

Optimal Split Gain Calculation:

When constructing each decision tree, LightGBM determines the best splits by maximizing the gain. For a split that partitions the data into left and right children, the gain is computed as:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_L+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \quad (8)$$

Here, G_L and H_L (and , G_R, G_H) are the sums of gradients and Hessians for the left (and right) child nodes, while λ and γ are regularization parameters controlling overfitting.

Histogram-Based Learning:

To speed up computation and reduce memory usage, continuous feature values are bucketed into discrete bins (histograms). This approximation allows LightGBM to quickly compute split gains and perform efficient tree learning without scanning all individual data points.

Advanced Sampling and Feature Bundling:

LightGBM incorporates Gradient-based One-Side Sampling (GOSS) to retain instances with large gradients (which are more informative) while subsampling those with small gradients. It also uses Exclusive Feature Bundling (EFB) to combine mutually exclusive features, reducing dimensionality without significant information loss.

Ensemble Stacking Regressor

Below are several key equations that mathematically describe an ensemble stacking hybrid model. Assume that we have an input feature vector x and a target variable y. The stacking ensemble uses multiple base models to generate predictions, which are then combined by a meta-learner.

Base Model Predictions:

Suppose we have M base models denoted as $f_1, f_2 \ldots \ldots f_n$. Each model produces a prediction for the input x:

$\hat{y}_m = f_m(x) \quad for\ m = 1,2, \ldots\ldots, M$

Meta-Learner (Stacking)

A meta-learner g is then trained to map the meta-feature vector z to a final prediction $\hat{y}$.

$$\hat{y} = g(z) = \beta_0 + \sum_{m=1}^{M} \beta_m \ \hat{y}_m \tag{9}$$

where β_0 is the intercept and β_m are the weights learned for each base model's prediction.

Loss Minimization:

Meta-learner parameters are preserved by minimizing the appropriate loss function through the training set. For a typical regression scenario using Mean Squared Error (MSE), the objective is:

$$\frac{min}{\beta_0, \beta_{1,\ldots\ldots,}\beta_M} \sum_{n=1}^{M} (y^{(i)} - \beta_0 - \sum_{m=1}^{M} \beta_m f_m(x^{(i)}))^2 \tag{10}$$

For quantile regression, where the goal is to estimate a particular quantile q (e.g., the median when q=0.5q=0.5q=0.5), the pinball loss function is used:

$$L_{quantile}(y, \hat{y}\,; q) = \frac{1}{N} \sum_{i=1}^{N} max\{q(y^{(i)} - \hat{y}^{(i)}), (q-1)(y^{(i)} - \hat{y}^{)(i)})\} \tag{11}$$

4. Result Analysis

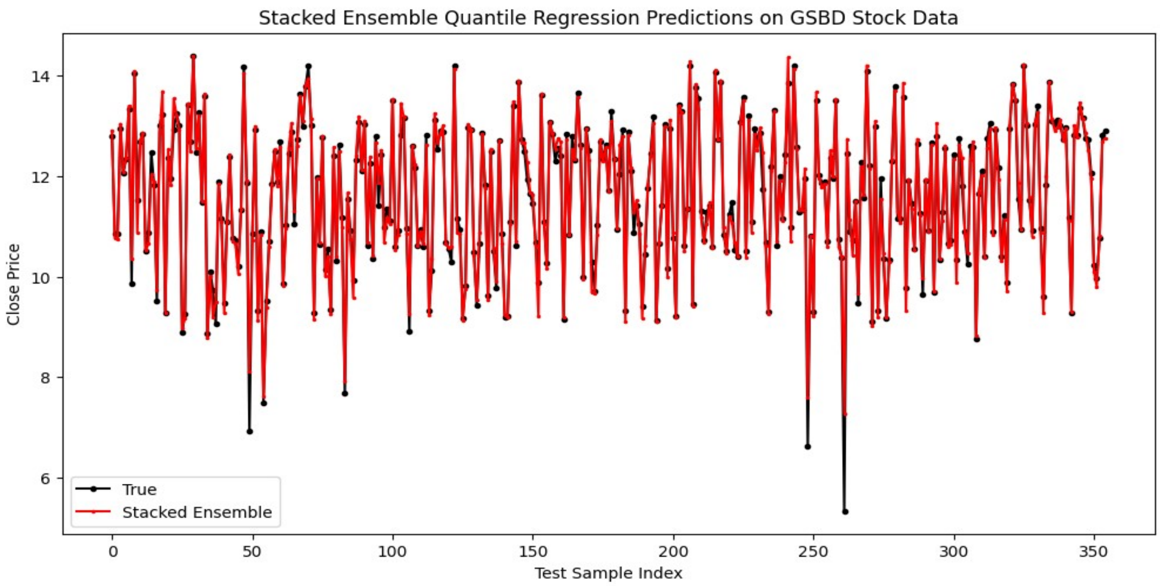


Figure 3. Staked Ensemble Quantile Regression Prediction ON Gsbd STOCK Data

The graphics show the performance of the stacked ensemble quantile regression model when predicting GSBD stock prices and comparing actual values (black) with predicted values (red). This model accurately follows the actual trends that exhibit strong predictive capabilities, especially in areas with high variance. Although well documented, occasional underestimation and overestimation indicate potential improvements through further optimization. Overall, the ensemble model exhibits robustness of share prognosis with space for fine tuning to improve accuracy in warm areas.

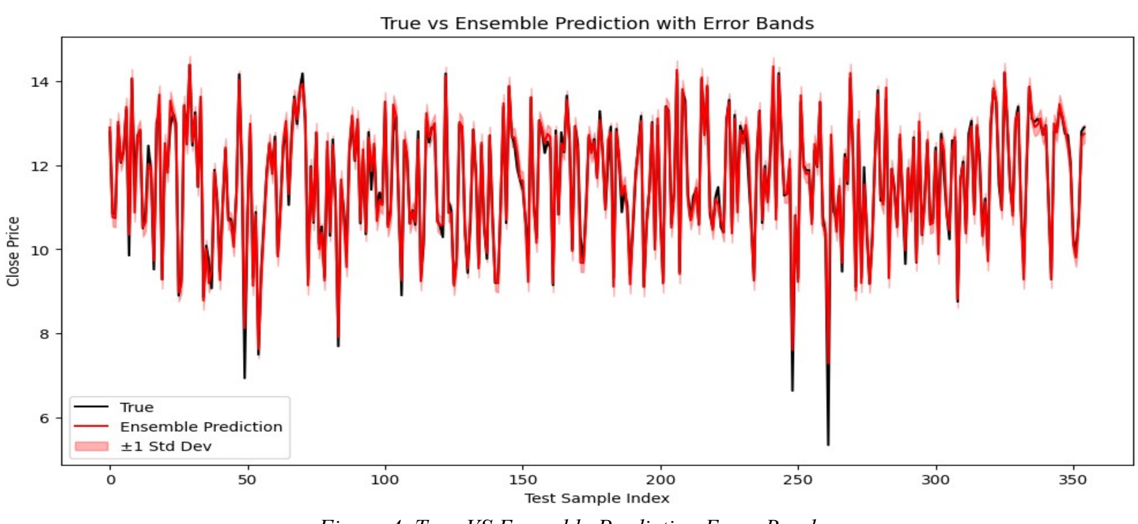


Figure 4. True VS Ensemble Prediction Error Bands

The graphics show a comparison between true stock prices (black) and ensemble model (red) predictions. The close direction of the ensemble prediction using actual values indicates strong prediction accuracy. The inclusion of fault zones provides insight into model uncertainty, and slight deviations are observed in minor areas. Although the model appears to have a good grasp of variability, occasional underestimation and overestimation highlight areas for possible improvement. Overall, the ensemble model exhibits robust predictive performance with controlled uncertainty.

Journal of Theoretical and Applied Information Technology
31st May 2025. Vol.103. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

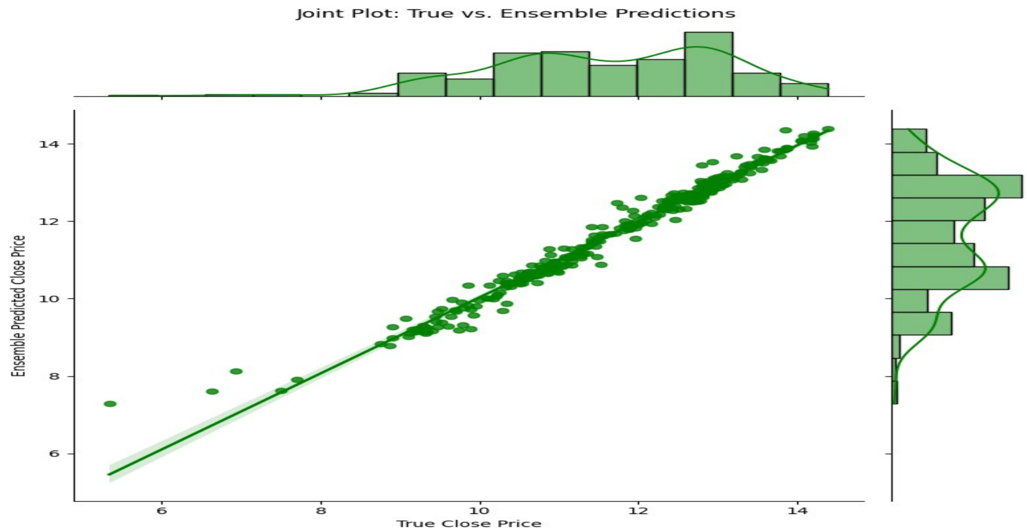


Figure 5. Joint Plot: True Vs. Ensemble Predictions

The general diagram shows the relationship between true and shared final prices for ensemble plans, showing strong linear correlations. The scatterplot shows that most points are near the ID line, indicating an accurate prediction with minimal deviation. The top and right histograms provide insight into the distribution of actual and predicted values and show similar spreads. Although slight variance is displayed at some points, overall the ensemble model shows high prediction accuracy and minimal distortion.

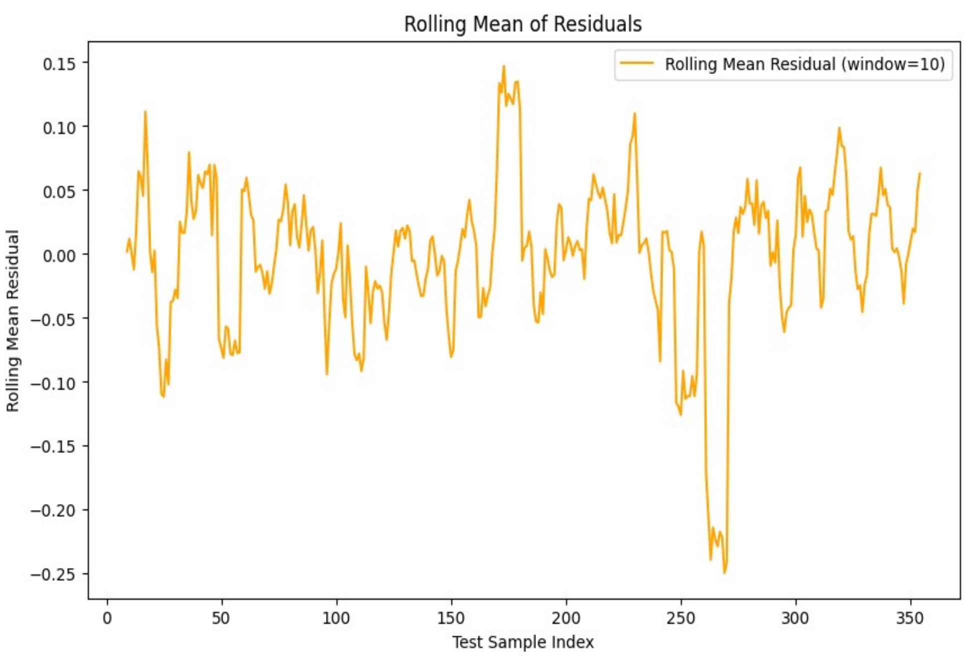


Figure 6. Rolling Mean Of Residuals

This diagram represents the rolling average of residuals with a window size of 10, showing variation in prediction errors for the test sample. Residues vary from zero. This indicates that the model does not have strong systematic distortions. However, there are several notable deviations, particularly in relation to specific test indexes where residues are significantly reduced, indicating potential sub-performance periods. Despite these variations, the residuals remain almost zero. That is, the predictions of the models are generally calibrated over time.

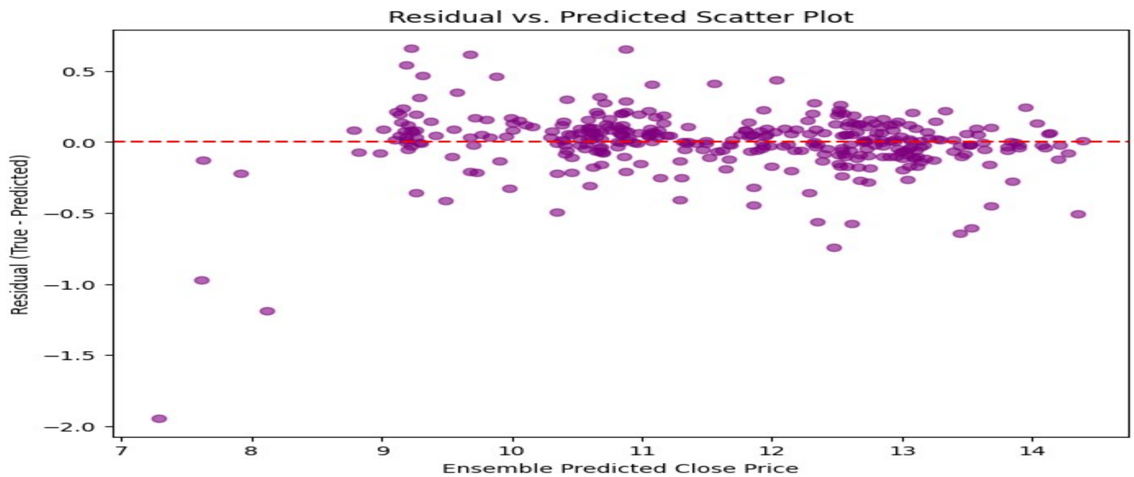


Figure 7. Residual vs. Predicted Scatter Plot

The scatter plot shows the true prediction for the final price predicted by the ensemble. Most of the residues are migrated at zero. This indicates that the model predictions are generally accurate with minimal distortion. However, some points are very different. This may indicate subpar customs or excessive regression. The dashed red lines of zeros act as references and indicate that the residuals are symmetrically distributed around them, but this does not imply a clear pattern of systematic error. This indicates that there is no strong distortion in the model, but there is a constant variation in prediction.

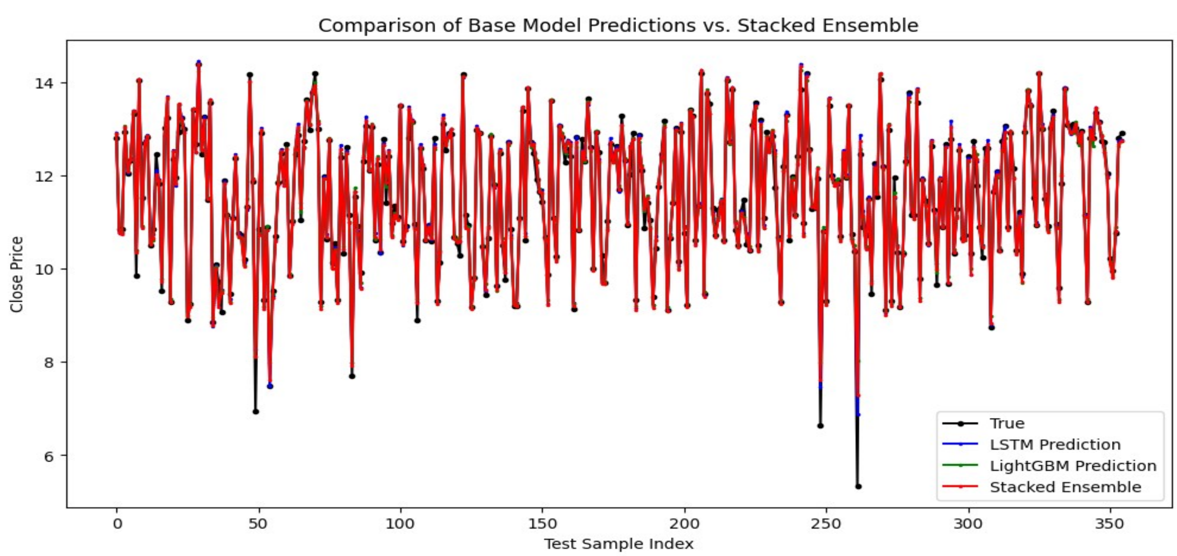


Figure 8. Comparison Of Base Model Predictions Vs. Stacked Ensemble

The graphics compare the actual final price (black) with predictions made with LSTM (blue), LightGBM (green), and stacked ensemble models (red). The stacked ensemble follows the actual values and indicates that the prediction accuracy for the individual models has improved. This variation suggests that the ensemble approach reduces prediction errors and better records trends in the data.

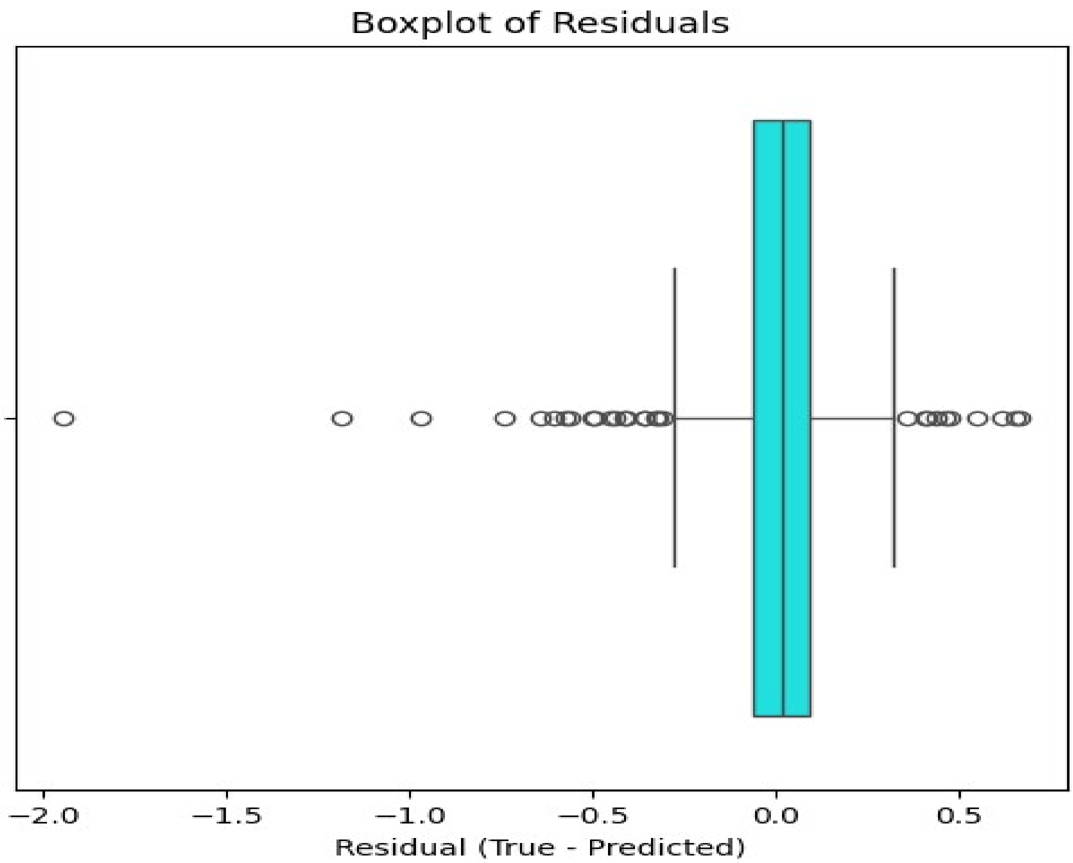


Figure 9. Boxplot of Residuals

The box plot shows the distribution of residues (true -prediction), with most residues tightly packed at zero, indicating a sufficient capacity model. However, there are some outliers, sometimes showing larger prediction errors. Overall, this model exhibits strong accuracy with minimal distortion.

Journal of Theoretical and Applied Information Technology
31st May 2025. Vol.103. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

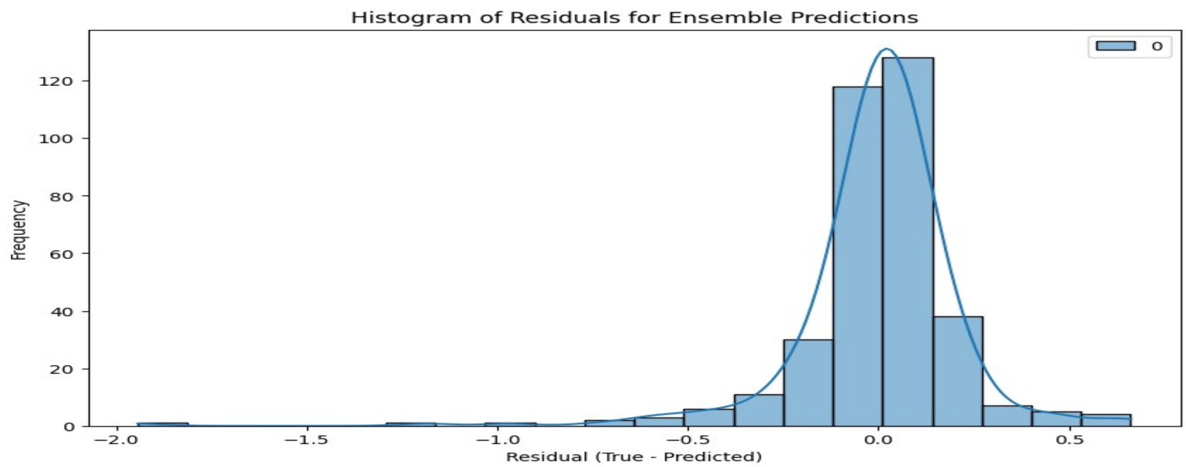


Figure 10. Histogram Of Residuals For Ensemble Predictions

The histogram of residuals shows a right-skewed distribution, with most residuals concentrated around zero, indicating that the ensemble model has a low error rate. However, some negative residuals suggest occasional under predictions. Overall, the model performs well with minimal bias.

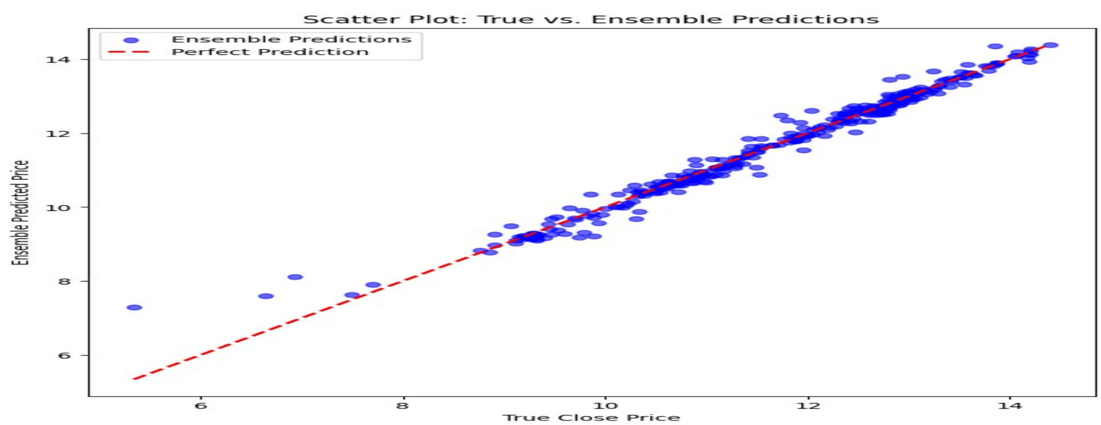


Figure 11. Scatter Plot: True Vs. Ensemble Predictions

The scatter plot shows a strong correlation between the true close prices and the ensemble predicted prices, as most points align closely with the red dashed line (perfect prediction). This indicates that the model performs well with minimal deviation, though slight errors exist in lower price ranges.

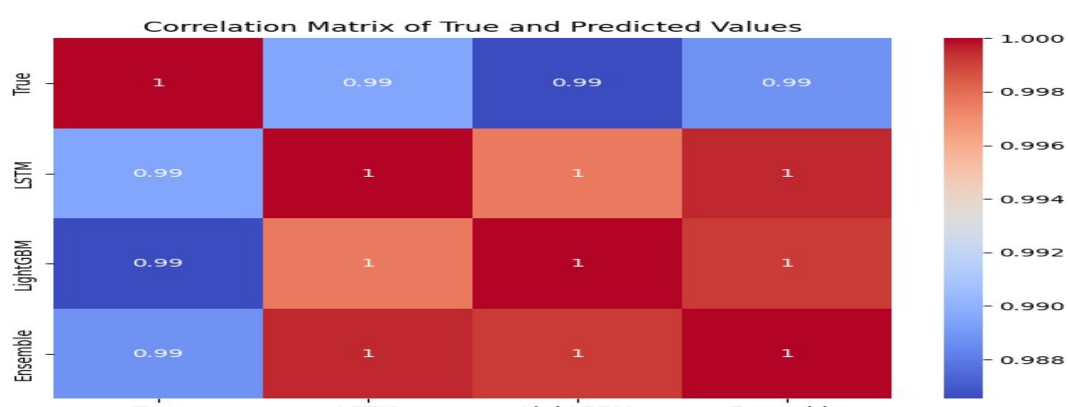


Figure 12. Correlation Matrix Of True And Predicted Values

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The correlation matrix shows that the ensemble model, LSTM, and LightGBM all have a very high correlation (≈0.99) with the true values, indicating strong predictive performance. The near-perfect correlation between LSTM, LightGBM, and the ensemble suggests that the ensemble model effectively integrates both techniques for accurate predictions.

Table 1. Correlation Matrix

S. N	Methodologies	Values
1	LSTM Model Pinball Loss	0.8238
2	LightGBM Model Pinball Loss	0.8139
3	Stacked Ensemble Pinball Loss	0.0656
4	Stacked Ensemble MAE	0.1313
5	Stacked Ensemble RMSE	0.2185
6	Stacked Ensemble R^2	0.9778

Table 2. Accuracy Matrices

S.N	Accuracy Matrices	Proposed Model	LSTM	LightGBM
1	Mean Absolute Error (MAE)	0.1313	0.2347	0.1359
2	Root Mean Squared Error (RMSE)	0.2185	0.2778	0.2432
3	R-squared (R^2)	0.9778	0.9641	0.9725
4	Mean Squared Error (MSE)	0.0478	0.0771	0.0591
5	Mean Absolute Percentage Error (MAPE)	0.01%	0.02%	0.01%

The proposed stacked ensemble model outperforms both LSTM and LightGBM across all accuracy metrics. It achieves the lowest Mean Absolute Error (MAE) of 0.1313, Root Mean Squared Error (RMSE) of 0.2185, and Mean Squared Error (MSE) of 0.0478, indicating higher precision in predictions. Additionally, it attains the highest R-squared (R^2) value of 0.9778, signifying a better fit to the data. The Mean Absolute Percentage Error (MAPE) is 0.01%, matching LightGBM but outperforming LSTM. These results demonstrate that the stacked ensemble model provides superior predictive accuracy and robustness compared to individual models.

5. CONCLUSION

This study provides a stacked ensemble quantile regression model for robust time-series predictions. This specifically takes into account financial market triggers. By integrating LSTM, LightGBM, and quantile regression, the proposed framework effectively captures temporal patterns and reduces the effect of outliers. Experimental analysis of GSBD stock price data shows that the ensemble model far exceeds individual models and achieves excellent prediction accuracy. The stacked ensemble model reached an R^2 of 0.9778, an RMSE of 0.2185, and a MAAPE of 0.01%.This indicates its effectiveness in real-world applications. Furthermore, analysis of residual distribution and error metrics confirm that the ensemble approach reduces distortion, improves prognosis stability, and maintains accuracy under volatile conditions. The results show that a combination of deep learning, increased methodology, and quantile regression provides a reliable and efficient solution for forecasting financial series. For future research, further study of hybrid profound architectures to improve external macroeconomic indicators, improved hyper parameter adjustments, and generalization via various financial instruments and exploring explainable AI (XAI) in forecasting interpretability.

. The proposed model forms the basis for the development of adaptive and resistant forecasting systems, contributing to better-founded investment decisions in the financial sector.

REFERENCES

[1]. Z. Jiang, K. Dai, K. Wang, T. Xie, Z. Fan, R. Li, and L. Zhao, "Centra-Net: A centralized network for visual localization spanning multiple scenes," *IEEE Transactions on Multimedia*, 2025.

[2]. J. Gong, Z. Qu, Z. Zhu, H. Xu, and Q. Yang, "Ensemble models of TCN-LSTM-LightGBM based on ensemble learning methods for short-term electrical load forecasting," *Energy*, vol. 134757, 2025.

[3]. T. Zhao, G. Chen, T. Gatewongsa, and P. Busababodhin, "Forecasting agricultural trade based on TCN-LightGBM models: A data-driven decision," *Research on World Agricultural Economy*, pp. 207–221, 2025.

[4]. M. Maharina, T. Paryono, A. Fauzi, J. Indra, S. Sihabudin, M. K. Harahap, and L. T. Rizki, "Machine learning models for predicting

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

flood events using weather data: An evaluation of logistic regression, LightGBM, and XGBoost," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 496–507, 2025.

[5]. C. Zhang, Y. Chen, M. Zhang, Z. Liu, and D. Wang, "SHAP-assisted EE-LightGBM model for explainable fault diagnosis in practical optical networks," *Journal of Optical Communications and Networking*, vol. 17, no. 2, pp. 81–94, 2025.

[6]. G. Airlangga, "Machine learning for tsunami prediction: A comparative analysis of ensemble and deep learning models," *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 6, no. 1, pp. 302–311, 2025.

[7]. S. Kumar, M. Srivastava, V. Prakash, and S. M. Islam, "A fusion of LightGBM, PLSRegression, and ElasticNet regressors in advanced ensemble stacking for empowering predictive modeling," *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Nov. 2024, pp. 1432-1437, IEEE.

[8]. Kumar, M. S., Srivastava, D. M., & Prakash, D. V. (2024). Advanced Hybrid Prediction Model: Optimizing LightGBM, XGBoost, Lasso Regression, and Random Forest with Bayesian Optimization. *Journal of Theoretical and Applied Information Technology, 15*(9), 4103-4115.

[9]. A. Chola, R. Rastogi, P. Kaur, A. Chaudhary, and D. Biswas, "Predictive analytics beyond the hype: A comprehensive comparison of LSTM, XGBoost, and LightGBM with emphasis on RMSE and CPU utilization," in *Proceedings of the 2024 Third International Conference on Power, Control and Computing Technologies (ICPC2T)*, Jan. 2024, pp. 179-184, IEEE.

[10]. M. Sui, C. Zhang, L. Zhou, S. Liao, and C. Wei, "An ensemble approach to stock price prediction using deep learning and time series models," in *Proceedings of the 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, July 2024, pp. 793-797, IEEE.

[11]. J. M. Ahn, J. Kim, and K. Kim, "Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting," *Toxins*, vol. 15, no. 10, p. 608, 2023.

[12] .V. Gupta and E. Kumar, "H3O-LGBM: Hybrid Harris Hawk Optimization-based Light Gradient Boosting Machine model for real-time trading," *Artificial Intelligence Review*, vol. 56, no. 8, pp. 8697-8720, 2023.

[13] .A. D. Hartanto, Y. N. Kholik, and Y. Pristyanto, "Stock price time series data forecasting using the Light Gradient Boosting Machine (LightGBM) model," *JOIV: International Journal on Informatics Visualization*, vol. 7, no. 4, pp. 2270-2279, 2023.

[14] . Kumar, S., Srivastava, M., & Prakash, V. (2023). Performance Analysis of Auto ARIMAX Model with Facebook Prophet and LightGBM Forecasting Models: Affirmation from Indian Mutual Fund. *International Journal of Agricultural & Statistical Sciences, 19*(2).

[15] .Kumar, S., Srivastava, M., & Prakash, V. (2024). Enhancing Mutual Fund Price Prediction: A Hybrid Ensemble Approach with Random Forest, SVR, Ridge, and Gradient Boosting Regressors. In *International Conference on Signal, Machines, Automation, and Algorithm* (pp. 551-566). Springer, Singapore.