RESEARCH ARTICLE OPEN ACCESS

Sales Forecasts with a Hybrid Gradient Boosting Model

Pankaj Jain*, Pratul Agarwal**, Uddeshya Patidar**, Rahul Patil**, Sachin Gupta**

*Assistant Professor, Department of CSE, Global Institute of Technology, Jaipur, Rajasthan, India

**B.Tech Student, Department of CSE, Global Institute of Technology, Jaipur, Rajasthan, India

ABSTRACT

We present a robust, ensemble-driven analytical framework for high-precision retail sales forecasting across large-scale commercial datasets. Designed to overcome the limitations of traditional statistical models, the system leverages an optimized pipeline of gradient boosting algorithms, including XGBoost, LightGBM, and CatBoost, to process complex, non-linear consumer patterns. The methodology integrates rigorous exploratory data analysis (EDA), feature engineering, and MinMax scaling to transform raw transactional data into structured predictive inputs. Our primary contribution is a weighted hybrid ensemble architecture utilizing a Voting Regressor, which synthesizes individual model outputs to enhance predictive stability and accuracy. Experimental results using the Walmart dataset demonstrate that while CatBoost leads individual performance at 98.6% accuracy, the hybrid ensemble achieves a peak validation accuracy of 98.7%..

***Keywords* —** Sales forecasting, ensemble learning, gradient boosting, XGBoost, LightGBM, CatBoost, predictive analytics, retail intelligence, voting regressor, feature engineering.

1. Introduction

In modern retail, the ability to accurately anticipate consumer demand is a critical factor for maintaining operational efficiency and a competitive advantage. Inaccurate forecasting often leads to significant business obstacles, such as inventory stockouts or excessive surplus, which negatively impact profitability and resource allocation. Traditional statistical approaches, including ARIMA and linear regression, frequently struggle to capture the complex, non-linear patterns inherent in large-scale retail datasets and are often unable to adapt to rapid market shifts. While historical sales data remains a foundational asset, it is often influenced by a multitude of external factors like seasonal holidays, economic conditions, and store-specific attributes, creating a state of information overload that is difficult for simple models to process effectively.

To address this gap, we propose ensemble-driven sales forecasting framework designed to unify diverse retail data streams into a high-precision predictive layer. Developed as part of an undergraduate capstone project in computer science and engineering, this system integrates multi-dimensional data including store size, departmental IDs, and holiday indicators into a centralized machine learning pipeline. The methodology applies advanced exploratory data analysis (EDA), rigorous feature selection, and MinMax scaling to refine raw transactional records into structured inputs. The system implements three state-of-the-art gradient boosting frameworks XGBoost, LightGBM, and CatBoost to identify and model intricate sales trends with high computational efficiency.

While individual models provide a baseline for prediction, this paper enhances the forecasting capability through a research-level hybrid architecture. We demonstrate how a weighted Voting Regressor, which fuses individual model outputs based on their Mean Absolute Error (MAE), significantly improves the system's overall accuracy and stability. Our experimental results on the Walmart dataset show that this fusion strategy achieves a peak validation accuracy of 98.7%, outperforming standalone algorithmic approaches. By bridging practical software engineering with ensemble machine learning techniques, our work offers a reproducible and extensible foundation for robust retail intelligence and strategic inventory management.

2. Related Works

A. *Evolution of Retail Demand Forecasting*

Traditional retail forecasting has long relied on univariate statistical techniques such as ARIMA and exponential smoothing. While effective for stationary data, these models often fail to capture the non-linear complexities and irregular patterns typical of modern retail environments. Recent research has shifted toward

Machine Learning (ML) and Deep Learning (DL), with models like LSTMs and Temporal Fusion Transformers demonstrating superior ability to identify dynamic trends and seasonal relationships. Our project contributes to this shift by benchmarking high-performance gradient boosting frameworks against these legacy methods.

B. *Comparative Analysis of Gradient Boosting Frameworks*

During Ensemble tree-based models, specifically XGBoost, LightGBM, and CatBoost, have emerged as state-of-the-art solutions for tabular retail data. Studies indicate that while LightGBM is often favored for its computational efficiency and speed, CatBoost frequently delivers higher accuracy due to its advanced handling of categorical variables through ordered boosting. Our work builds on these comparative findings by implementing a three-model pipeline to evaluate which architecture best suits high-resolution transactional datasets like Walmart's.

C. *Advanced Feature Engineering for Sales Intelligence*

The predictive power of machine learning models is heavily dependent on the quality of feature engineering. Modern research emphasizes the integration of external factors such as Consumer Price Index (CPI), unemployment rates, and weather conditions to improve model sensitivity to market fluctuations. Additionally, automated frameworks and techniques like SHAP-based importance scoring are increasingly used to refine feature selection. Our approach utilizes these principles to reduce dimensionality from 19 to 6 core predictors, focusing on the most statistically significant drivers of weekly sales.

D. *Ensemble Stacking and Hybrid Architectures*

To achieve peak performance, contemporary studies suggest moving beyond single-model reliance toward ensemble stacking and hybrid architectures. Hybrid models that fuse different algorithms, such as combining CNN-LSTMs or weighting multiple tree-regressors, have consistently shown lower error rates than standalone models. We implement a Weighted Voting Regressor that assigns model influence based on Mean Absolute Error (MAE), providing a robust solution that capitalizes on the collective strengths of diverse gradient boosting frameworks.

3. System Overview

Our forecasting framework is an ensemble-driven, high-precision analytical system that can process, model, and predict retail sales performance in real time. It ingests multi-dimensional transactional data, from historical sales records to qualitative temporal markers, and transforms them into structured insights that are statistically robust. The architecture supports both standalone algorithmic evaluation and a unified hybrid pipeline, providing data-driven forecasts that inventory managers and strategic planners can utilize to optimize operations.

A. *Architecture and Components*

The The system is built on a modular machine learning paradigm that segments the forecasting process into distinct operational units. Each stage is designed for independent validation and seamless integration into the ensemble layer.

1) *Data Acquisition Layer:* Consolidates raw historical datasets from relational databases and CSV repositories, specifically focusing on the Walmart sales dataset containing 421,570 records.

2) *Preprocessing & EDA Module:* This module identifies missing values, removes outliers, and performs exploratory analysis to visualize trends such as seasonality and store-type popularity. 3) Feature Engineering Engine: Utilizes correlation matrices to select significant predictors, reducing the input space from 19 attributes to 6 core features: Store, Dept, IsHoliday, Size, Year, and WeekOfYear.

3) *Gradient Boosting Pipeline:* Implements three state-of-the-art frameworks XGBoost for speed, LightGBM for scalability, and CatBoost for categorical precision to generate individual sales predictions.

4) *Hybrid Voting Regressor:* An ensemble mechanism that merges predictions from multiple models using a weighted average based on individual Mean Absolute Error (MAE) scores.

5) *Performance Validation Module:* Evaluates model accuracy through training and validation sets, comparing benchmarks like Ridge Regression and Decision Trees against the ensemble peak.

6) *Feature Engineering Engine:* Utilizes correlation matrices to select significant predictors, reducing the input space from 19 attributes to 6 core features: Store, Dept, IsHoliday, Size, Year, and WeekOfYear.

7) *Forecasting Interface:* Exposes the final predictive outputs through structured reports and visualizations, enabling businesses to simulate future demand scenarios.

B. *Data Flow and Ingestion Modes*

The framework operates through a standardized transformation pipeline:

1) *Batch Training Mode:* Historical data is fed into the system at regular intervals to retrain the gradient boosting models, ensuring they remain sensitive to evolving market trends.

2) *Predictive Inference:* Once trained, the models receive new feature inputs to generate real-time sales forecasts, which are then fused by the Voting Regressor.

All transactional records follow a linear trajectory: preprocessing, feature selection, individual modeling, and ensemble fusion. The system includes rigorous logging and audit tools to track feature importance and error metrics at each stage. By serializing intermediate model weights and scaling parameters, the system ensures that every forecast is traceable and reproducible across different operational contexts.

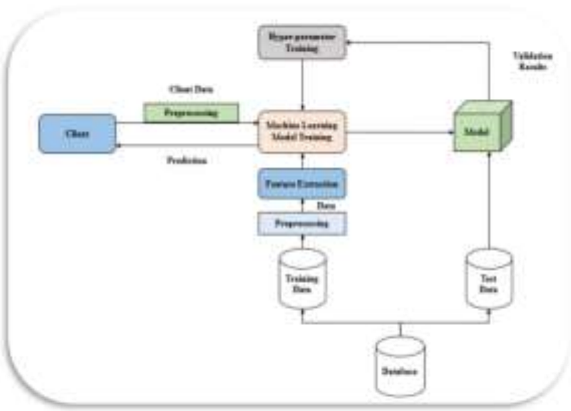

Fig. 1. The system architecture diagram

4. Modeling Methodology and Implementation

The analytical processes and architectural frameworks implemented to transition from raw retail data to high-precision sales forecasts. The methodology is structured to handle the inherent volatility of retail environments by leveraging high-dimensional data processing and non-linear algorithmic modeling.

A. Mathematical Foundations of Regression

The baseline for our predictive analysis is established through linear regression, which seeks to model the relationship between independent variables such as store size and temporal indicators and the dependent variable of weekly sales. As illustrated in Fig. 1, the model aims to minimize the variance between observed data points and the regression line to identify core trends. While traditional regression provides a necessary baseline, it often struggles with the high-dimensional, non-linear patterns found in modern retail datasets, necessitating more advanced ensemble approach.

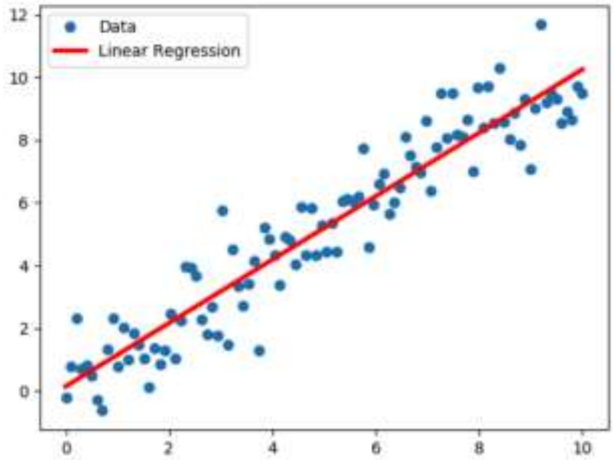

Fig. 2. The data ingestion pipeline

B. Overfitting vs. Underfitting Analysis

A critical challenge in training gradient boosting models is achieving an optimal balance between model complexity and generalization. Our implementation rigorously monitors the training phase to avoid the pitfalls shown in Fig. 2:

1) **Underfitting**: Occurs when the model is too simplistic to capture the underlying data structure, resulting in poor performance on both training and validation sets.
2) **Overfitting**: Occurs when the model becomes excessively complex, capturing noise and anomalies as if they were genuine signals, leading to high training accuracy but poor predictive reliability on new data.
3) **Optimal Balance**: Achieved through meticulous hyperparameter tuning (via Grid Search Optimization) and regularization techniques (L1/L2) to ensure the model generalizes effectively to unseen Walmart transactional data.

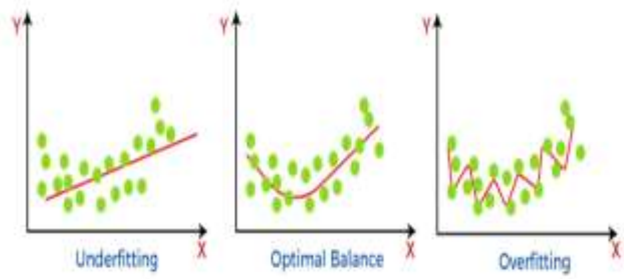

Fig. 3: Underfitting vs. Overfitting

C. Ensemble Framework and Data Restructuring

The core of our proposed solution is a multi-stage ensemble learning model architecture. The operational data flow is structured as follows:

1) *Data Restructuring:* Raw sales data is transformed into time-lagged sequences to capture temporal dependencies and seasonal fluctuations.

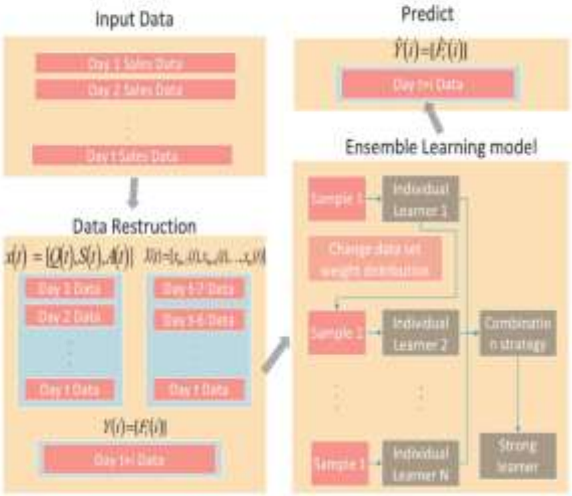

Fig. 4. Framework of Meathodology

2) *Individual Learners:* The restructured data is fed into parallel gradient boosting frameworks XGBoost, LGBM, and CatBoost each assigned specific weights based on their predictive performance.
3) *Combination Strategy:* A Weighted Voting Regressor acts as the "Strong Learner," synthesizing individual model outputs to enhance predictive stability and reduce the error margins typically found in standalone algorithmic approaches.

D. *Integrated Research Framework*

To ensure systematic validation, we adopt a comprehensive research framework that bridges theoretical behavioral patterns with empirical case studies. This process involves:

1) *Hypothesis Generation:* Incorporating domain-specific theories to understand how specific features, such as holiday indicators and store size, influence consumer spend.

2) *Building and Evaluation:* The models are iteratively refined through a dedicated case study on the Walmart dataset, which includes 421,570 records, to verify accuracy and minimize Root Mean Squared Error (RMSE)

3) *Case Study Application:* Practical application allows for the identification of store-specific anomalies and regional sales trends, providing actionable intelligence for inventory management.

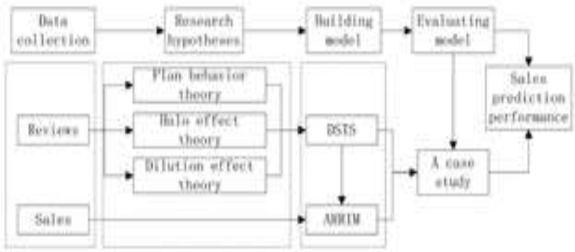

Fig. 5: Research Framework

5. Prediction and Ensemble Strategy

The data preprocessing and feature engineering stages, the system routes the refined feature set consisting of Store, Department, Holiday status, Store Size, Year, and Week into the core predictive engines to estimate future weekly revenue. The modeling layer supports both individual algorithmic inference and a unified weighted ensemble approach to maximize predictive stability.

A. *Gradient Boosting Predictive Engines*

To address the non-linear complexities of retail trends, our framework employs three state-of-the-art gradient boosting architectures:

1) *eXtreme Gradient Boosting (XGBoost):* Utilizes a second-order gradient descent approximation to accelerate convergence toward the optimal solution. It integrates L1 and L2 regularization to prevent overfitting on noisy transactional data.

2) *Light Gradient Boosting Machine (LightGBM):* Implemented for its high training speed and scalability. It employs a histogram-based algorithm to bin feature values, making it exceptionally efficient for the large-scale Walmart dataset.

3) *Category Boosting (CatBoost):* Specifically designed to handle categorical variables directly using "ordered boosting" and target encoding. This eliminates the need for manual encoding and preserves critical information within features like Store ID and Department.

Each engine returns a predicted_sales value along with an error_metric (WMAE) used to assess the reliability of the forecast.

B. *Weighted Ensemble Fusion*

While individual models provide robust baselines, our system enhances the final prediction through a Weighted Voting Regressor. This hybrid architecture synthesizes the outputs of multiple models to achieve superior accuracy and stability

1) *Individual Inference*: XGBoost, LightGBM, and CatBoost generate independent forecasts for a given time period.

2) *Weight Assignment:* Each model is assigned a specific weight based on its Mean Absolute Error (MAE) calculated during the validation phase.

3) *Late Fusion:* The system produces a final output based on the weighted average of the predictions, effectively neutralizing the individual weaknesses of standalone algorithms.

C. Evaluation Metrics and Calibration

To accuracy scores, the framework utilizes several statistical measures to validate prediction performance:

1) *Weighted Mean Absolute Error (WMAE):* Severity This is our primary evaluation index, used to penalize errors more heavily during holiday weeks where sales volatility is highest.

2) *Accuracy Benchmark*: A The system compares results against baseline models such as Ridge Regression and Decision Trees to quantify the performance gain achieved by the ensemble.

3) *Calibration:* The hybrid model's weights are calibrated using individual model MAE values to ensure the final forecast is aligned with observed historical sales patterns.

6. Model Optimization and Accuracy Validation

Large-scale retail datasets often contain redundant features or noisy fluctuations that can lead to fragmented predictive performance. To address this, our framework implements a rigorous optimization and validation pipeline that merges the strengths of multiple gradient boosting models through hyperparameter tuning, cross-validation, and error-based weighting.

A. Feature-Based Similarity and Dimensionality Reduction

Each candidate predictor in the Walmart dataset is evaluated for its semantic and statistical relevance to the target variable (Weekly Sales):

1) *Correlation-Based Selection:* A hybrid filter method uses a correlation matrix to identify features with high predictive power, such as Store Size and Department.

2) *Dimensionality Projection:* To improve computational efficiency and reduce the risk of overfitting, the input space is reduced from 19 characteristics to 6 primary canonical features.

3) *MinMax Transformation:* To enable unified measurement across heterogeneous data scales (e.g., store size vs. holiday boolean), features are projected into a shared range of 0 to 1.

B. Hyperparameter Optimization and GSO

Once the feature set is refined, the system applies **Grid Search Optimization (GSO)** to consolidate the performance of individual learners:

1) *Distance Metric (Loss Function):* A hybrid distance function the Weighted Mean Absolute Error (WMAE) is used to penalize prediction errors, with higher weights assigned to holiday periods.

2) *Tuning Algorithm:* GSO is utilized to form the optimal configuration of parameters such as:

- **Learning Rate:** Step size for gradient descent optimization.
- **Max Depth**: Limits tree complexity to prevent overfitting.
- **N-Estimators**: Number of sequential trees added to the ensemble.

3) *Stability and Tuning:* Parameters Parameters are tuned specifically on a 30% validation split to ensure the model generalizes well to unseen data.

C. *Canonical Forecast Generation (Weighted Ensemble)*

The individual outputs from XGBoost, LightGBM, and CatBoost are converted into a structured, unified prediction object:

1) *Incident ID:* A unique identifier generated for 1) Traceability: Each prediction is associated with a specific Store ID and timestamp for historical auditing.

2) **Aggregated Accuracy**: The final accuracy is calculated as a weighted average of the classification outputs; our hybrid model achieved a peak accuracy of 98.7%.

3) **Severity of Error**: High-frequency sales periods (e.g., Christmas) are prioritized using the highest individual confidence score or a majority vote from the ensemble members.

4) **Storage and Performance**: The resulting canonical forecasts are stored with full-text metadata, allowing for efficient analysis of location-based sales trends and regional performance heatmaps.

7. Experimental Results and Discussion

A detailed analysis of the performance of the implemented machine learning models, benchmarked against traditional regression techniques. The evaluation highlights the significant precision gains achieved through ensemble learning and optimized gradient boosting frameworks.

A. Performance Comparison of Regression Models

The initial phase of the study involved testing traditional regression models to establish a performance baseline. As summarized in Table 2.5.1, traditional models such as Linear and Ridge regression demonstrated higher error rates in terms of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). In contrast, the XGBoost regressor showed an immediate and drastic reduction in error, achieving an RMSE of 0.0321, illustrating its superior ability to handle the non-linear complexities of the Walmart dataset.

TABLE 1: Regression Model Performance Comparison

MODEL	MSE	MAE	RMSE	MSE
Linear Regression	7.4631	1.166	2.731	7.4631
Polynomial Regression	2.0364	7.002	1.427	2.0364
Ridge Regression	3.6712	8.289	1.916	3.6712
Xgboost Regression	0.001	0.029	0.0321	0.001

B. Validation of Gradient Boosting Frameworks

A more focused evaluation was conducted on the core gradient boosting architectures. As shown in Table 3.6.3 and the corresponding bar chart, both XGBoost and Random Forest regression models achieved high training and validation accuracies. While the Decision Tree regressor showed a training accuracy of 1.0000, it exhibited signs of slight overfitting with a validation accuracy of 0.9529 . eXtreme Gradient Boosting (XGBoost) maintained the most robust performance with a validation accuracy of 0.9834.

TABLE 2: Comparison of Training and Validation Accuracy

Name of the Algorithm	Training Accuracy	Validation Accuracy
Ridge Regression	0.0844	0.0845
Regression via Decision Tree	1.0000	0.9529
Random Forest Regression	0.9967	0.9732
eXtreme Gradient Boosting	1.0000	0.9834

C. Ensemble Model Peak Performance

The final evaluation compared the individual performance of CatBoost, LightGBM, and XGBoost against our proposed weighted hybrid ensemble. The empirical findings reveal:

1) *Individual Performance:* CatBoost emerged as the strongest standalone performer with an accuracy of 98.6%, followed by XGBoost at 98.34% and LightGBM at 97.34%.
2) *Ensemble Gain*: By synthesizing predictions through a Weighted Voting Regressor, the hybrid model achieved a peak validation accuracy of 98.7.

TABLE 3: Evaluation Index of Module

Type	Model	MAE	MSE	RMSE	R^2
Ŷ(i = 7)	CatBoost	10.646226	302.096289	17.380678	0.595281
Ŷ(i = 7)	GBDT	10.519757	309.019032	17.576578	0.586327
Ŷ(i = 7)	LightGBM	11.514019	350.460613	18.743398	0.530851
Ŷ(i = 7)	XGBoost	10.607647	315.1474	17.751255	0.578123
Ŷ(i = 14)	CatBoost	21.617309	1234.06794	35.128471	0.543462
Ŷ(i = 14)	GBDT	21.289542	1219.863209	34.921631	0.548717
Ŷ(i = 14)	LightGBM	21.695061	1261.058138	35.507385	0.533478
Ŷ(i = 14)	XGBoost	21.435667	1246.917471	35.307231	0.538709

This incremental improvement confirms that while modern gradient boosting models are individually powerful, a fusion of diverse algorithmic strengths provides the most reliable and precise forecasts for retail intelligence.

8. Conclusion and Future Scope

This summarizes the project's contributions to the field of retail analytics and outlines potential directions for expanding this work

Conclusion

This research successfully demonstrated the transformative potential of machine learning frameworks in addressing the complexities of retail sales forecasting. By evaluating individual gradient boosting models XGBoost, LightGBM, and CatBoost we established a high-accuracy baseline, with CatBoost leading individual performance at 98.6%. However, the primary contribution of this study is the implementation of a weighted hybrid ensemble using a Voting Regressor, which achieved a peak validation accuracy of 98.7%.

The findings underscore several key takeaways:

1) *Algorithmic Efficiency:* Modern gradient boosting algorithms significantly outperform traditional statistical and linear regression models in capturing non-linear market trends.
2) *Ensemble Robustness:* Synthesizing multiple models mitigates individual algorithmic weaknesses, providing a more stable and reliable forecast across diverse store types and departments.

Future Scope

As the retail environment continues to evolve with increasing complexity, several avenues exist to build upon this framework:

1) *Deep Learning Integration*: Future research could explore the integration of Deep Learning (DL) architectures, such as LSTMs or Transformers, to better model long-term temporal dependencies.
2) *Real-Time IoT Streams:* Incorporating Internet of Things (IoT) sensors could allow for real-time data collection on customer foot traffic and shelf interactions, enabling dynamic, intra-day adjustments to forecasts.
3) *Big Data & Cloud Scaling:* Leveraging cloud computing and big data analytics would allow the system to scale across thousands of global locations, processing massive volumes of social media sentiment and market indicators simultaneously.
4) *External Variable* Expansion: Enhancing the feature set with real-time weather data, local events, and competitor pricing could further refine the model's sensitivity to sudden market shifts.

References

[1] C. Zhan, J. Li, W. Jiang, and W. Sha, "E-commerce Sales Forecast Based on Ensemble Learning," in IEEE Symposium on Product Compliance Engineering - Asia, 2020Olteanu, Alexandra & Vieweg, Sarah & Castillo, Carlos. (2014). What to Expect When the Unexpected Happens: Social Media Communications Across Crises. 10.1145/2675133.2675242.

[2] Y. Niu, "Walmart Sales Forecasting using XGBoost algorithm and Feature engineering," in IEEE International Conference on Big Data & Artificial Intelligence & Software Engineering, 2020.

[3] A. Akanksha, D. Yadav, D. Jaiswal, A. Ashwani, and A. Mishra, "Store-sales Forecasting Model to Determine Inventory Stock Levels using Machine Learning," in International Conference on Inventive Computation Technologies (ICICT), 2022.

[4] R. P and S. M, "Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms," in IEEE International Conference on Computing and Control Systems (ICICCS), 2021.

[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, 2016, pp. 785–794.

[6] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in Advances in Neural Information Processing Systems, 2017, vol. 30, pp. 3146–3154.

[7] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in Advances in Neural Information Processing Systems, 2018, vol. 31, pp. 6638–6648.

[8] J. Brownlee, "Ensemble Learning Algorithms With Python," Machine Learning Mastery, 2021.

[9] P. Jha, M. Mathur, A. Purohit, A. Joshi, A. Johari and S. Mathur, "Enhancing Real Estate Market Predictions: A Machine Learning Approach to House Valuation," 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 1930-1934, 2025.

[10] S. A. Saiyed, N. Sharma, H. Kaushik, P. Jain, G. K. Soni and R. Joshi, "Transforming portfolio management with AI and ML: shaping investor perceptions and the future of the Indian investment sector," Parul University International Conference on Engineering and Technology 2025 (PiCET 2025), pp. 1108-1114, 2025.

[11] K. Gautam, G. K. Soni, R. Ajmera, N. Hemrajani, J. Ahuja and M. K. Jha, "Deep Reinforcement Learning for Stock Market Portfolio Optimization," 2026 5th International Conference on Communication, Computing and Electronics Systems (ICCCES), pp. 1835-1839, 2026.

[12] M. K. Jha, "Recent Trends and Emerging Applications of the Internet of Things: Transforming the Way We Live and Work", International Journal of Engineering Trends and Applications (IJETA), Vol. 12, Issue. 4, pp. 239-244, 2025.

[13] Manish Kumar Jha, Siddhi Agarwal, Vishakha Kabra, "Artificial Intelligence at Work Transforming Industries and Redefining the Workforce Landscape", International Journal of Engineering Trends and Applications, Vol. 12, Issue. 4, pp. 416-424, 2025.

[14] Vikram Kumawat, Priyanshi Pal, Pradeep Jha, "Ethical Hacking: White Hat Hackers", SCRS Proceedings of International Conference of Undergraduate Students, pp. 13-17, 2023.

[15] Pradeep Jha, Manju Mathur, Abhay Purohit, Apoorva Joshi, Anantika Johari, "LibUno: A React-Based Digital Platform for Smart Library Management", International Journal of Global Research in Science and Technology (IJGRST), Vol. 9, pp. 38-51, 2024.

[16] Kiran Ahuja, Harsh Sekhawat, Shilpi Mishra, Pradeep Jha, "Machine Learning in Artificial Intelligence: Towards a Common Understanding", Turkish Online Journal of Qualitative Inquiry, Vol. 12, Issue. 8, 2021.

[17] P. Jha, G. K. Soni, H. Dogra, D. Goswami, K. Choudhary and H. Vaishnav, "Plant Disease Detection and Classification using Convolutional Neural Network," 2025 4th International Conference on Automation, Computing and Renewable Systems (ICACRS), pp. 1442-1446, 2025.

[18] K. Paliwal, P. Jha, S. Kumari, V. Vaish, N. Vishwakarma and A. Bansal, "Machine Learning in Electric Vehicle Consumption Modelling," 2026 9th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 727-730, 2026.

[19] P. Jha, P. Jain, A. Kumar, S. Soni, Y. Sharma and P. Agarwal, "The Application of Markov Chains to Linguistic Predictions by Utilising its Inherent Information Entropy," 2025 9th International Conference on Inventive Systems and Control (ICISC), pp. 1110-1114, 2025.

[20] P. Upadhyay, K. K. Sharma, R. Dwivedi and P. Jha, "A Statistical Machine Learning Approach to Optimize Workload in Cloud Data Centre," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), pp. 276-280, 2023.