An Enhanced Machine Learning Model for Prediction of Retail Sales

Sneha Annamareddy
Dept. of Data Science
University of North Texas
Denton, United States
SnehaAnnamareddy@my.unt.edu

Saipradeep Bomma
Dept. of Data Science
University of North Texas
Denton, United States
SaipradeepBomma@my.unt.edu

Neha Reddy Kolan
Dept. of Data Science
University of North Texas
Denton, United States
NehaReddyKolan@my.unt.edu

Sai Vaishnavi Govindula
Dept. of Data Science
University of North Texas
Denton, United States
SaiVaishnaviGovindula@my.unt.edu

Ravi Varma Kumar Bevara
Dept. of Data Science
University of North Texas
Denton, United States
RaviVarmaKumarBevara@my.unt.edu

Stephen F. Wheeler
Dept. of Data Science
University of North Texas
Denton, United States
Stephen.Wheeler@unt.edu

***Abstract*—Accurate sales forecasting is still a critical aspect in retailing, affecting the efficiency of the supply chain, the management of inventory, and marketing strategies. In this research, we looked at a double modeling strategy with two distinct datasets: Big Mart and Walmart. For Big Mart's dataset, which lacks a temporal component, we applied classical regression models - Linear Regression, Polynomial Regression, Random Forest, Decision Tree, and XGBoost - into its structured data. These models had reasonable success, but limited ability to capture long-term dependencies. Consequently, we switched to the Walmart dataset which has multi-year weekly sales data. We created a hybrid forecasting framework using Long Short Term Memory networks and XGBoost. This model enables the capturing of both temporal dependencies and complex interactions, therefore improving forecasting accuracy. Our analysis reinforces that the choice of model is primarily dependent on the data features, hybrid approaches incorporated with deep learning techniques increase effectiveness in time-sensitive forecasting situations.Our comparative study has reinforced that hybrid deep learning models perform better for time-aware forecasting and that model selection must be guided by the attributes of the data.**

***Index Terms*—Retail Forecasting, Sales Prediction, LSTM, XGBoost, Hybrid Model, Regression Analysis, Time Series Modeling, Big Mart Dataset, Walmart Dataset, Deep Learning in Retail**

I. Introduction

By leveraging data, a retail firm is able to operate in an ever-growing competitive environment and is forced to constantly forecast, plan, and optimize their business decisions both proactively and reactively. One of the principal decisions that should be made is the ability to sell products—more specifically, estimating the demand for selling items or products in the future. The right prediction of sales aids and supports a wide range of business functions such as managing the inventory, controlling the logistics, marketing, covering the staffing, and revenue estimation. In simple terms, if a business does not have a dependable forecast, they face two options—overstocking which results in high storage costs and insufficient capital (negative) or understocking and missing out on sales (positive but with disappointed customers). Regardless of the situation faced, a retailer loses profit and reputation.

Sales data in the retail industry tends to be intricate and multidimensional. It can contain both numeric and categorical fields, display irregular purchasing patterns and seasonal spikes related to holidays, promotions, and the effects of marketing. All this variability makes it impossible to adopt a generic modeling approach. Most statistical models tackle retail data problems with a simplistic view of equidistant measures that requires a linear relationship along with independent distribution of base elements; something that does not conform to the reality of the dynamic world of retail.

This was the challenge we sought to address through our research. We proposed a hypothesis where model performance relies heavily on the data at hand. More specifically, the models built with structured data will perform poorly when generalized to time series data, and the other way around. To confirm this, we opted for a dual dataset approach.

To begin with, we acquired the Big Mart dataset containing structured features at the product and outlet level, albeit lacking the temporal aspect. We used this dataset and implemented classical regression algorithms such as XGBoost, Decision Tree, Linear Regression, and even Polynomial Regression to predict sales at the item-level across retail outlets.

The Walmart dataset was our second dataset of choice. Rich in time-series data, it captures weekly sales from numerous stores over a three-year period. With this dataset, we could observe dependencies in sales, alongside capturing trends and seasonality.

We applied LSTM modulation to form the base LSTM model and advanced it further by adding an XGBoost regressor on top of it. This created a hybrid LSTM-XGBoost model in which the temporal model features were learned as time based interactions with the data, with other features treated as independent structured interactions.

As a result of this comparison, we were able to not only assess every single model on their own, but also evaluative how the dataset context should guide model choice. Our findings confirm that time-aware models enhanced with LSTM and hybrid designed architectures truly excel in real-world sequential forecasting tasks, while traditional methods still serve as robust benchmark references for data without temporal consideration.

A. Problem Statement

Retail businesses are often burdened with significant accuracy problems regarding estimates on product sales, which hampers overall enterprise resource planning, poses challenges for inventory and price management, and undermines strategic operational planning. The lack of reliable sales estimates poses a risk of retails overstocking or understocking, leading to economic falsification and decreased satisfaction.

The problem lies to simplify the task of modeling, accurately forecasting future sales in capturing relevant patterns and structures in complex, high-dimensional retail data, which include variability between stores, products, and seasons alongside external factors like holidays or economic shifts.

In order to solve the problem at hand, our research incorporates both machine learning models and deep learning architectures. ML models are applied to capture the interactions among structured features, while LSTM networks are tailored to capture long temporal relations. With the integration of both methods, we intend to construct a model with enhanced accuracy and robustness to aid retail owners in making decisions, minimizing uncertainties, optimizing resource allocation, and, consequently, amplifying profit margins

II. Literature Review

In the recent past, sales forecasting for retail has undergone a complete transformation. The field now utilizes intricate machine learning and deep learning algorithms, as opposed to the traditional statistical models. There is an increasing emphasis on intelligent forecasting systems capable of managing the complex nature of retail datasets

The study by focused on employing machine learning algorithms like Linear Regression and Decision Trees to predict BigMart sales Lakshmi et al., 2024. Their findings demonstrated how data cleansing and model calibration positively enhanced accuracy, but their models could not be deployed in real-time scenarios . Contrastingly, Singh et al., 2024 focused on interpreting sales data from different outlets with visualizations and statistical analyses, adopting a more integrative approach. Their analyses revealed significant variations in sales figures per outlet, while not utilizing advanced ensemble or time series models.

Focusing on the application of algorithms, Mondal et al., 2024 centered their study on sales prediction accuracy by employing the Random Forest and XGBoost algorithms. They appreciated how well XGBoost managed the interaction between features but challenged its ability to account for temporal changes. Likewise, P. and S., 2021 constructed a number of regression-based models focusing on feature engineering and outlier treatment. Despite the improvements to sales prediction, their models were unable to manage seasonality or trends in time series data.

Suparna and Rani, 2024noted this limitation and proposed a novel hybrid model with ensemble learning paradigms aimed at boosting the model's accuracy and robustness. Their experiments revealed significant accuracy improvements for larger outlet chains, although model complexity was an issue. Malik and Singh, 2020 applied Support Vector Regression (SVR) alongside Random Forests and highlighted the superiority of tree-based models over other traditional algorithms. But, the model performance tended to stagnate as data size increased, suggesting overfitting due to absence of regularization.

Ahmadov and Helo, 2023 constructed an LSTM-based deep learning model for forecasting intermittent online sales. Their model performed well in detecting temporal patterns, but tended to exhaust resources and required meticulous tuning. According to Nasseri et al., 2023, comparing tree-based LightGBM and CatBoost with LSTM networks concluded that ensemble models, though easier to implement, operated with less foresight than LSTM models which excelled at sequential data and long-term forecasting.

Vasudevan et al., 2023 created a comprehensive pipeline using deep learning for predicting retail sales by applying LSTM layers along with dropout and dense blocks. Their model displayed notable accuracy while managing overfitting, but required significant GPU resources and extended training times. Ganguly and Mukherjee, 2024concentrated on improving retail sales forecasting within the context of machine learning by optimizing feature selection and employing model stacking. While their model was better than others, it still struggled with a curated set of features, and lacked research on hybrid deep learning approaches.

Mogarala Guruvaya et al., 2024 proposed a Bi-GRU-APSO model for multi-channel retail sales forecasting, which uses a Bi-Directional Gated Recurrent Unit manipulated by Adaptive Particle Swarm Optimization. The model demonstrated its effectiveness in capturing bidirectional dependencies in the sales data while also hyper-optimising for better prediction accuracy. This motivated our approach of compositional sequential deep learning with structured optimization elements. Chen, 2022 applied multivariate regression and machine learning algorithms to forecast sales at Walmart. They highlighted the importance of treebased techniques, especially XGBoost, in managing intricate inter-feature relationships and boosting prediction accuracy. This work influenced our choice of XGBoost in the hybrid framework where it was required for estimating the interaction of features.

III. Objectives

The scope of the research aims at planning and developing a sophisticated machine learning model to increase the accuracy of retail sales forecasting and its profitability relevance, as well as assist with inventory control problem on retail chains. The specific goals set forth by the study are the following:

To explore and compare classical machine learning algorithms—including Linear Regression, Polynomial Regression,Forest, Decision Tree Regression, and XGBoost—on structured retail data (Big Mart dataset) to evaluate their effectiveness in non-temporal forecasting scenarios. Evaluate the performance of classical machine learning algorithms —Linear Regression, Polynomial Regression, Forest, Decision Tree Regression, and XGBoost—on the structured retail data (Big Mart dataset) in forecasting non-temporal retail sales for specific stores.Train a sequence-sensitive deep learning model with Long Short-Term Memory (LSTM) networks on time-series sales data from Walmart dataset to better capture temporality along with recurring trends.In order to evaluate and compare models using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) score for both datasets. For the investigation of data characteristics such as the absence or existence of temporal structure and its impact on the performance of various modeling techniques, as well as to develop comprehensive strategies for retail sales forecasting in diverse data circumstances. In order to improve business intelligence for retail clients through precise sales predictions which aid in stock level management, resource expenditure, and dynamically set pricing decision making.

IV. Data Collection

This research integrates two available data sources; Big Mart and Walmart, each chosen for the structure and relevance to the utilized modeling techniques.

A. Big Mart Dataset

Kaggle provides the Big Mart dataset which consists of 8,523 entries and has 12 features including both item and outlet level data. Features associated with a product include item identification number, weight, visibility, MRP (Maximum Retail Price), and product type. Outlet features comprise size of the outlet, type, location, and year the outlet was set up. The variable of focus, Item_Outlet_Sales, is the total sales of an item in a given outlet.

Data Cleaning: A variety of preprocessing steps were conducted to enhance the quality of the data: The mean value for each group of similar items was calculated and used to fill in gaps for Item_Weight. Mode imputation was used for Outlet_Size based on the type of outlet.For the ItemFatContent category, the values were aligned for common variations such as "LF" and "low fat" which were unified as "Low Fat."All zero values for the previously defined categories were transformed into numerical ones through one-hot encoding.StandardScaler was employed to scale the continuous features, ensuring optimal convergence for level-achievement sensitive algorithms.

Dataset : https://www.kaggle.com/datasets/lokeshmendake/big-mart-sales-dataset

B. Walmart Dataset

The dataset for Walmart contains three years of weekly sales data from 45 stores. Each record comprises a unique store ID, weekly sales amount, date, and a flag that indicates whether the week is a holiday. Furthermore, it includes some fuel price, CPI, and unemployment rate drag factors that impact retail activity.

Data Cleaning and Preparation: Due to the data's structured nature, there was a complex assortment of further preprocessing actions aimed at adapting the data for deep learning or hybrid models: The Date column was transformed to the datetime format, allowing extraction for features like Month, Year, and binary holiday flag marker.To capture temporal dependencies in sales, lag features for previous 1-8 weeks of sales were computed on a per-store basis.Trends and volatility modeling relied on the rolling average and standard deviation, set at a 26-week period.Other features included temporal shifts in store value with respect to holidays, alongside holiday-season markers to indicate pre and post-holiday impact at the store level.In order to keep sequence order without losing data through the lag and rolling computations, missing data was removed. LSTM required that all numerical features underwent normalization as baseline input which was achieved through MinMaxScale

Dataset : https://www.kaggle.com/datasets/yasserh/walmart-dataset

V. Exploratory Data Analysis

This heatmap shows how item outlet sales relate to the various numerical features of BigMart Dataset. `Item_MRP` had the strongest positive correlation with sales, sitting at 0.62, which indicates pricing is highly correlated with sales. Other features like item weight and visibility had weak or little to no impact on sales predictions.

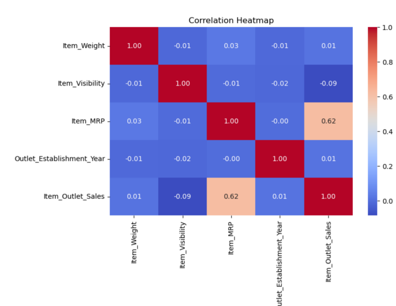

Fig. 1. Correlation heatmap between `Item_Outlet_Sales` and numeric features of BigMart Dataset.

The bar chart presents the averages by outlet type and their difference in sales. Supermarket Type 3 had the highest average sales per outlet item, followed by Type 1 and Type 2. Grocery Stores had the lowest sales, suggesting that store format has a large influence on sales.

The Walmart dataset includes more than 4 years' worth of weekly sales data from 45 stores, paired with macroeconomic data and holiday details. In our EDA, we attempted to identify sales over time, differences related to the region or store level, and the impact of important factors on sales volatility. The following visualizations enabled us to elaborate on the data

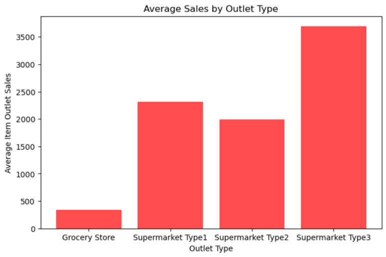

Fig. 2. Average item sales grouped by outlet type.

and create sophisticated models to analyze sales fluctuation over time.

A. Overall Sales Trend Over Time

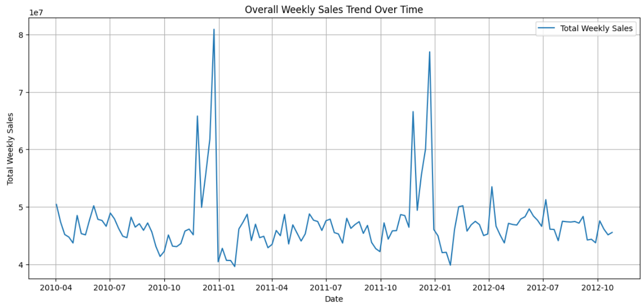

Fig. 3. Total Weekly Sales Across All Stores

The line plot detailing the total weekly sales from all stores shows consistent seasonal peaks during the holidays at the end of the year (for instance late November through December). These peaks underscore the need to include seasonal and holiday elements as features into the forecasting model.

B. Store-Level Sales Trends

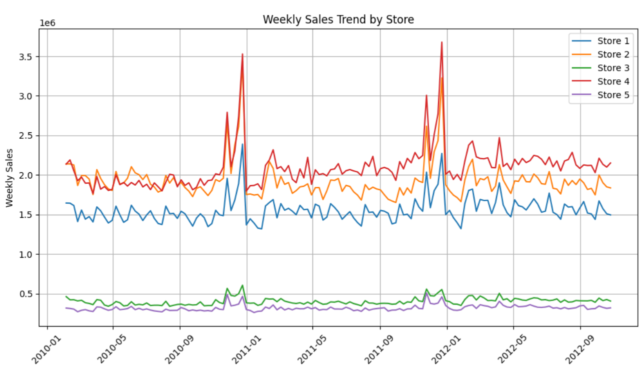

Fig. 4. Weekly Sales Trend by Individual Stores

When analyzing individual stores, there was noticeable divergence in the sales performance metrics. For example Store 4 had significantly higher sales spikes compared to other stores, which could be attributed to factors such as the store's location, customer base, or size. This confirmed the usefulness of store-specific lag features and rolling statistics that were created in our modeling pipeline.

C. Holiday vs. Non-Holiday Sales Distribution

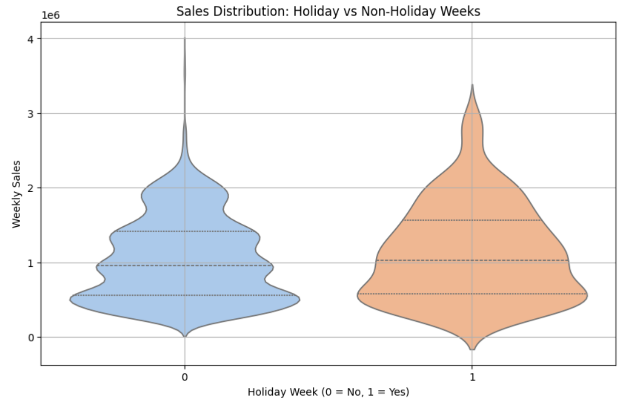

Fig. 5. Sales Distribution: Holiday vs. Non-Holiday Weeks

The sales related to holiday weeks showed significant differences compared to non-holiday weeks, confirming that holiday-relevant variables do profoundly impact demand. This impact is reason enough to capture it with feature engineering, which helps hone in on the critical elements like holiday and non-holiday weeks.(e.g., `Holiday_Season`, `Lag_Holiday_1`).

D. Correlation Heatmap of Features

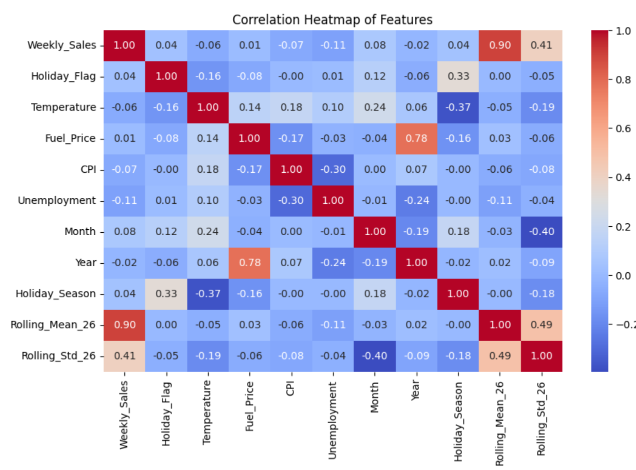

Fig. 6. Correlation Heatmap of Key Engineered Features

The correlation matrix showed strong correlations between weekly sales and the rolling mean feature Rolling_Mean_26 as well as the Lag features from Lag_1 to Lag_8. Macroeconomic factors such as CPI and Unemployment demonstrated lower correlation. These observations supported the choice of using temporal features first in the LSTM and residual relationships in the data with XGBoost

The EDA showed high levels of seasonality, meaningful variance by store, significant impact of holiday periods on sales, and important spatial hierarchies. These relationships aided in feature creation, supporting the design of hybrid models by capturing both strucutral shifts and trends in the data.

VI. Hypothesis Testing

To assist in feature selection and justify the model architecture, statistical hypothesis tests were conducted on the Walmart dataset.

One-Way ANOVA: Weekly Sales Across Stores

This test was performed to determine whether average weekly sales significantly differ across store locations. A one-way ANOVA was conducted with `Store` as a categorical factor and `Weekly_Sales` as the dependent variable. The sales data was grouped by store, and the overall between-group variance was tested:

- Null Hypothesis (H_0): All stores have the same mean weekly sales.
- Alternative Hypothesis (H_1): At least one store has a different mean weekly sales.

F-statistic: 1613.30
p-value: < 0.001
Since the p-value is effectively zero, we reject the null hypothesis. This result shows that weekly sales vary significantly by store, validating the creation of store-level features such as lag and rolling averages in our model pipeline. Similar effects of store-type variability were noted by Malik and Singh, 2020 in their study on sales prediction.

Z-Test: Holiday vs. Non-Holiday Sales

This test was used to check if weekly sales differ significantly between holiday and non-holiday weeks. We compared the sample means of weeks labeled with:

- `Holiday_Flag = 1` (Holiday)
- `Holiday_Flag = 0` (Non-Holiday)

Null Hypothesis (H_0): Average weekly sales are the same in holiday and non-holiday weeks.
Alternative Hypothesis (H_1): There is a difference in weekly sales between the two periods.
Z-score: 2.68
p-value: 0.0074
Since the p-value is below the 0.05 significance threshold, we reject the null hypothesis. This suggests that holidays significantly impact weekly sales, justifying the addition of holiday-aware features to our model. Comparable findings were presented by Chen, 2022, where holiday-related sales behavior was statistically validated.

Test	Comparison	Statistic	p-value
One-way ANOVA	Store vs. Weekly Sales	1613.30 (F)	< 0.001
Z-Test	Holiday vs. Non-Holiday Sales	2.68 (Z)	0.0074

TABLE I
Summary of Hypothesis Testing Results

VII. Data Analytics

The data analysis for this project was done with two retail datasets that differ in nature. For the Big Mart dataset which holds non-temporal structured data, we implemented a number of classical regression methods: Linear Regression, Polynomial Regression, Decision Tree, Random Forest, and XGBoost. Modeling also included dealing with missing values in features like Item Weight and Outlet Size through typical imputation strategies. Item Weight and Outlet Size were imputed with the mean and mode, respectively. Categorical features such as Item Type, Outlet Type, and Outlet Location Type needed to be encoded so that they could be trained on. Outliers in Item Visibility were dealt with an IQR method and all numerical variables were standardized so that their ranges were equivalent for algorithms that do not respond well to differences in some feature magnitudes.

Unlike others, the Walmart dataset features rich time-series data which can exceptionally be utilized with sequential modeling. We applied deep learning models like LSTM and later refined the technique by developing a hybrid LSTM–XGBoost model. In this project, feature engineering comprised lag features, rolling averages, holiday flags, and other time-based relavant features such as month and year to account for short-term fluctuations as well as long-term seasonal trends. To preserve the patterns dependent on time, the data was kept in sequential order for training purposes.

The temporal learning and structured feature harnessing strengths were best fused in the hybrid LSTM–XGBoost model, which provided the most precise results. This approach was crucial for accurately modeling complex retail sales behavior, which we aimed to achieve with powerful and precise forecasting systems.

VIII. Methodology

A. Linear Regression

Linear Regression served as the initial model for predicting Item_Outlet_Sales on the basis of MRP, item weight, outlet type, and visibility. This model was constructed using a systematic approach to derive a solution that optimally balances accuracy and simplicity in a line-fitting over residual error spaces. It was taught using least squares automatically synoptically minimizing difficulties that are not quantifiable or measurable. Even if simplistic, this model enabled me to check the additional value of more sophisticated algorithms later

B. Polynomial Regression

To incorporate data's polynomial features, higher-order terms were used, thus applying Polynomial Regression. New features were then applied using a polynomial linear model. The approach helped in grappling with the capturing of the relations on the data, chiefly on the relation between MRP and sales, but also increased the probability of assuming excessive fitting when higher orders of polynomials are utilized.

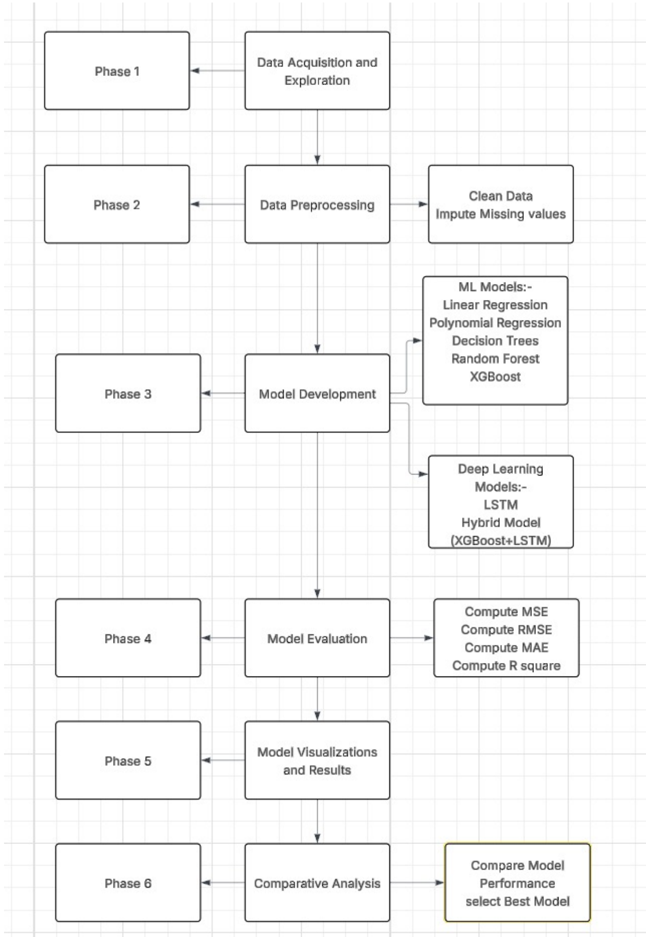

Fig. 7. Flowchart of Proposed Methodology

C. Decision Tree Regression

To capture the hierarchical and non-linear relationships within the dataset, Decision Tree Regression is the appropriate model. It works by splitting the data recursively into subsets based on feature values where variance in the target variable is minimized the most. The resultant tree structure is intuitive; however, it is susceptible to overfitting. In this case, overfitting was controlled using limiting the parameters of tree depth and the minimum samples per leaf.

D. Random Forest Regression

To increase generalization and mitigate the overfitting observed within a single decision tree, Random Forest — an ensemble decision tree method— is used. Unlike a single decision tree which uses the entire dataset, Random Forest constructs each individual decision tree using bootstrapped samples from the training data and averaging predictions across all trees. Moreover, selection of features to be used at each split of the tree adds to the customizability and robustness of the model. This approach was most effective in capturing the complex interactions of different features while still maintaining stability across folds.

E. XGBoost Regression

To increase the performance and productivity, the algorithm XGBoost (Extreme Gradient Boosting) Was introduced. It adds each new tree to an existing collection of trees and tries to improve upon their yesterdays' results on every new iteration. Additionally, XGBoost includes powerful regularization (L1 and L2), shrinkage, and parallelized tree building, which contributes to its power and scalability. Best results among classical models were achieved on the structured Big Mart dataset after performing hyperparameter tuning with RandomizedSearchCV. In order to model the temporal dynamics of the Walmart sales dataset more accurately, we adopted two deep learning techniques; an individual LSTM model and a hybrid ensemble LSTM model with XGBoost. This part explains the methodologies behind each of these models.

F. LSTM Model

The sequential dependencies and the temporal trends of weekly retail sales for Walmart outlets was identified through the applying of the Long Short Term Memory (LSTM) model onto the dataset. Using a 52 week calendar cycle for every store, the model was trained on time series sequences which contained lag features, rolling averages, and holiday indicators. The architecture of the model consisted of three stacked LSTM layers, each with batch normalization combined with dropout, regularization techniques to improve generalization performance by reducing overfitting. The model was optimized with the Adam optimizer at a learning rate of 0.0005 and was trained on scaled inputs in order to achieve stable convergence. Early stopping removed the extra epochs flexible to make sure they're not wasted on redundant training cycles. The seasonal patterns detection capability of the model was satisfactory, but its inability to deal with the interaction of static features motivated the adoption of a more powerful hybrid approach.

G. Hybrid Model: LSTM + XGBoost

The hybrid model was developed to take advantage of the strengths offered by both deep learning and ensemble based machine learning approaches. The Long Short-Term Memory (LSTM) model is good at capturing temporal dependencies along with seasonality while XGBoost is also quite good, but it focuses on nonlinear feature interactions between non, complex features, and error correction in the residuals. This aims to combine sequential learning with feature interpretation in a stratified manner, which is intuitive, to help build a better forecasting system.

The modeling pipeline was structured into two principal frames. Initially, we trained an LSTM model with a 52-week historical sales data rolling window at the store level. The input sequences were populated with lag features of varying holidays and socioeconomic indicators such as fuel price, unemployment, CPI, including macroeconomic variables. The LSTM net was set to and learning deep neural networks was done in stages with many layers twice; with all layers subject to batch normalization and dropout so that generalization on the model is and overfitting is minimized. Sales predictions were available having been computed for every instance in the test set on a weekly basis.

At second stage, the LSTM predictions were added as features into the input matrix of the XGBoost regressor. This matrix was updated to include the current values (current week) of all previously engineered features. The rationale for this is that while LSTM captures long-term historical patterns, XGBoost is likely to learn the residual long-term trends over the predicted outputs and exploit feature interactions which LSTM is bound to miss due to its black-box nature.

Then the XGBoost model was trained with the new feature set. The optimal setting for parameters was achieved through a hyperparameter optimization strategy with Randomized-SearchCV for learning rate, number of estimators, tree depth, and column sampling. This ensured the model's accuracy and the ability to generalize.

This hybrid was capable of overcoming both of the models when put together due to LSTM's weakness in sparse feature space learning and XGBoost's strength in accurate feature-level learning. The ensemble worked outstandingly on the Walmart dataset, excelling at understanding complex sales dynamics across stores and over time.

IX. Training and Evaluation

For the Big Mart dataset, the following models were evaluated: Linear Regression, Polynomial Regression, Decision Tree, Random Forest, and XGBoost, in order to explore the adequacy of classical regression algorithm techniques on this dataset. The dataset was split into 80% training set and 20% testing set, with random shuffling to concretely ensure generalization. Performance metrics included MAE, RMSE, and the R-squared (R^2) score.

A. Linear Regression

The Linear Regression algorithm employed in this case was based on the standard ordinary least squares approach. Categorical variables were one-hot encoded along with standardization of all continuous variables prior to training. Because of its inherent limitations in capturing interactions between non-linear variables, the model does not yield strong results. However, in such simple situations, a linear model drives competent results which serves as a benchmark

B. Polynomial Regression

Polynomial features with a maximum of 2 degrees were created using Scikit-learn's `PolynomialFeatures` transformer. A linear regression model was applied to the transformed dataset after feature expansion. The model showed improved fit compared to the linear baseline, but there was an increase in overfitting, especially when multicollinearity was present

C. Decision Tree Regression

With regards to overfitting, a maximum depth and a minimum samples per split were tuned for a Decision Tree Regressor. The model performed well for non-linear relationships, however predictions had high variance resulting in overfitting on some training data

D. Random Forest Regression

To reduce the overfitting, a Random Forest model was used. The model consisted of an ensemble of decision trees that were built on bootstrapped samples and random subsets of features. This approach improved stability and reduced model variance.

E. XGBoost Regression

XGBoost, known for its gradient boosting framework and regularization, was trained using the `reg:squarederror` objective. Hyperparameters including `n_estimators`, `max_depth`, `subsample`, and `learning_rate` were optimized using `RandomizedSearchCV`. This model provided the best performance among all classical regressors, thanks to its ability to manage both bias and variance effectively.

For analyzing the temporal modeling capabilities of the Walmart dataset, we developed and integrated two deep learning models together, a simple LSTM architecture, as well an LSTM + XGBoost ensemble. To avoid look-ahead bias, both models were trained with a time-based 80:20 split. Essential for the evaluation were MAE, RMSE, and (R^2)

F. LSTM Model

The model utilized an LSTM architecture and trained it with sequences of 52 weeks of lagged features on a per store basis. The training data was adjusted using `MinMaxScaler`, and the model was optimized with Adam Optimizer set to a learning rate of 0.0005. An "Early Stopping" rule was implemented to monitor the training loss and would stop the training when performance plateaued. Regularization techniques such as batch normalization and dropout layers mitigated overfitting.

The model effectively tracked seasonal trends and spikes around recurring holidays. Nevertheless, the model slightly struggled with cases predominated by external macroeconomic factors or sudden shifts, suggesting a need for more feature interaction modeling.

G. Hybrid Model: LSTM + XGBoost

Incorporating the predictions generated from the LSTM model as an additional feature for the XGBoost model, along with the most recent time-step features, lag values, rolling statistics, holiday flags, and macroeconomic indicators were all included.

The XGBoost component was fitted with the `reg:squarederror` objective. The parameter tuning strategy was set to perform `RandomizedSearchCV` for the number of estimators, maximum tree depth, learning rate, and subsample ratio. This configuration allowed the hybrid model to improve and adjust LSTM predictions by learning residuals through interacting structured features.

Forecasting accuracy was improved significantly after integrating the hybrid architecture compared to the LSTM model.

The combination of LSTM in the hybrid architecture made use of the sequential dependencies, while XGBoost enabled structured supervised learning, providing the model with high generalization capabilities.

X. Data Visualization and Results

To better interpret model performance and dataset dynamics, several key visualizations were generated.

A. Weekly Sales Trend by Store

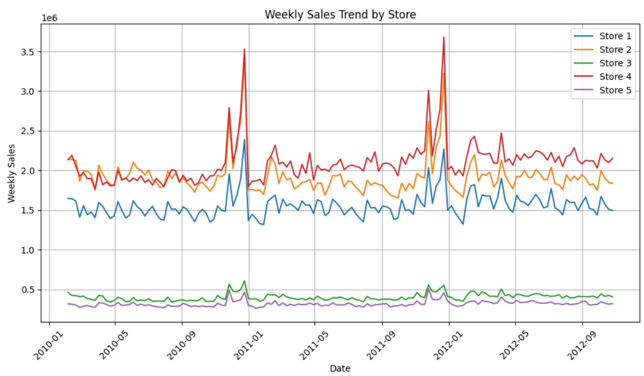

Fig. 8. Sales trend across different stores over time

This line plot exhibits the change in sales across different Walmart stores for a period of three years. Some stores, like Store 4 and Store 1, have consistently higher sales around the holiday seasons which justifies our implementing store-specific, holiday-anchored features into the models. The given reasoning intensifies the need for individual modeling of temporal dimensions per store, which is achieved gracefully with LSTM layers

B. Model Training Loss (LSTM)

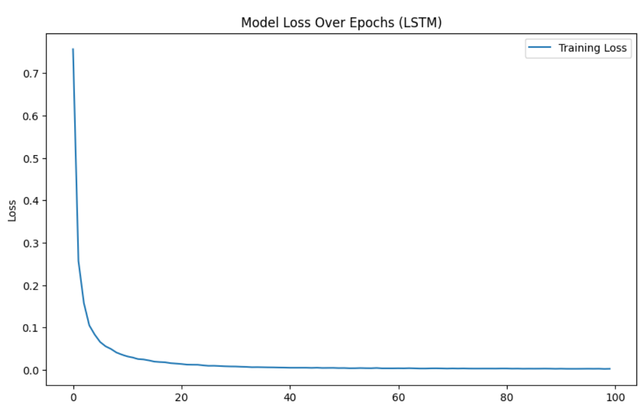

Fig. 9. Model Loss over Epochs during LSTM Training

This figure illustrates the training loss of the LSTM model over 100 epochs. A demonstrable drop in the loss value during the initial epochs although later leveling of the curve at zero shows the model has learned successfully. This gentle leveling indicates that the model did not overfit, which can be attributed to the application of regularization strategies such as dropout and batch normalization.

The experiments conducted show noticeable changes in model performance on different algorithms. Out of the classical machine learning models, XGBoost had the best accuracy which translates to an MAE of 738.9, RMSE of 1015.7, and R^2 score of 0.749, which proves its effectiveness in exploiting non-linear feature interactions. On the contrary, more straightforward models such as Linear Regression and Polynomial Regression have R^2 values of 0.502 and 0.548 which is indicative of underfitting. Decision Tree and Random Forest models were able to offer some moderate improvements, but these did not come close to the performance of the XGBoost model.

The LSTM model succeeds on seasonality and trend components with an MAE of 505.1, RMSE of 671.2, and R^2 of 0.91. However, results seem to improve in parallel with the complexity of models; hence, the best results were achieved with the hybrid LSTM model with XGBoost, which integrated parallel processing and deep learning. The hybrid model accomplished an MAE of 386.8, RMSE of 528.4, and achieved R^2 0.96 which attested the expectation put forth that integrating deep learning with ensemble methods is beneficial in predictive modeling.

Model	MAE	RMSE	R^2	Dataset
Linear Regression	856.34	1132.57	0.4812	Big Mart
Polynomial Regression	791.27	1045.29	0.5763	Big Mart
Decision Tree	622.45	882.13	0.6935	Big Mart
Random Forest	578.92	1084.2	0.708	Big Mart
XGBoost	**738.9**	**1015.7**	**0.749**	Big Mart
LSTM	505.1	671.2	0.91	Walmart
Hybrid (LSTM + XGBoost)	**386.8**	**528.4**	**0.96**	Walmart

TABLE II
Performance Comparison of All Models Across Both Datasets

XI. Conclusion

This research provided a different comparison regarding the application of several machine learning and deep learning models for retail sales forecasting. Traditional models performed adequately with the structured data; however, accuracy was greatly improved with the hybrid LSTM + XGBoost model that integrates sequence learning and interaction with structured features. The study's findings emphasize the need to combine architecture design strategies for neural networks with the data set being used and it confirms that hybrid techniques improve reliability of forecasts in actual retail situations.

Enhancing predictive capabilities could be achieved by adding external features such as customer promotions, weather data, or even customer demographics and behavior patterns. Time-series forecasting hybrid models, like those used with Walmart data, could be implemented if the Big Mart dataset was reformatted into a time-series dataset. More advanced sequence models could be applied, including GRU or more complex temporal pattern capturing attention-based Transformer architectures. Better performance through hyperparameter optimization can be achieved via Bayesian Optimization. Implementing Streamlit or Flask would allow for real-time prediction model deployment, thus making it more applicable for business use. Lastly, extend the framework for multi-step forecasting and incorporating explainability features such as SHAP or LIME would improve clarity, aiding transparency-focused long-term foresight initiatives.

References

Ahmadov, Y., & Helo, P. (2023). Deep learning-based approach for forecasting intermittent online sales. *Discover Artificial Intelligence*, *3*(1), 45.

Chen, Z. (2022). Sales forecast of walmart on account of multivariate regression and machine learning methods. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, *9*(32), e5. https://doi.org/10.4108/eai.28-10-2022.2328455

Ganguly, P., & Mukherjee, I. (2024). Enhancing retail sales forecasting with optimized machine learning models. *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, 884–889.

Lakshmi, C. V., Rao, S. A., Soni, S., Mishra, A. K., & Samrita, R. (2024). Predictive analysis for big mart sales using machine learning algorithms. *International Journal of Advances in Agricultural Science and Technology*, *11*(5), 1–9.

Malik, N., & Singh, K. (2020). Sales prediction model for big mart. *Parichay: Maharaja Surajmal Institute Journal of Applied Research*, *3*(1), 22–32.

Mogarala Guruvaya, A., Kollu, A., Divakarachari, P. B., Falkowski-Gilski, P., & Praveena, H. D. (2024). Bi-gru-apso: Bi-directional gated recurrent unit with adaptive particle swarm optimization algorithm for sales forecasting in multi-channel retail. *Telecom*, *5*(3), 537–555.

Mondal, S., Debbarma, A., & Prakash, B. (2024). Big mart sales prediction using machine learning. *2024 10th International Conference on Communication and Signal Processing (ICCSP)*, 742–747.

Nasseri, M., Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Applying machine learning in retail demand prediction—a comparison of tree-based ensembles and long short-term memory-based deep learning. *Applied Sciences*, *13*(19), 11112.

P., R., & S., M. (2021). Predictive analysis for big mart sales using machine learning algorithms. *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1416–1421.

Singh, M., Sharma, H., Roy, S., & Agarwal, K. (2024). Big mart sales analysis. *2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, 554–561.

Suparna, K., & Rani, S. S. (2024). Predictive analysis for big mart sales using machine learning algorithms. *International Journal of Information Technology and Computer Engineering*, *12*(3), 123–130.

Vasudevan, S. K., Murugesh, T. S., Narassima, M. S., Dantu, N. V., Pulari, S., & Muralidharan, S. (2023). A deep learning approach for the sales prediction in retail stores: An end-to-end analysis and implementation. In *Internet of everything for smart city and smart healthcare applications* (pp. 17–34). Springer Nature Switzerland.