Walmart Stores Sales Forecasting

Mohammed Alnajim
University of Delaware
CIS
alnajimm@udel.edu

Karthikheyaa Kurra
University of Delaware
CIS
karth@udel.edu

Srikanth Perumalla
University of Delaware
CIS
srikanth@udel.edu

Ashok Kurapati
University of Delaware
CIS
ashokk@udel.edu

Xi Peng
University of Delaware
CIS
xipeng@udel.edu

Abstract

Walmart operates thousands of stores with varying sales across different departments and locations. This paper aims to build a machine-learning model to forecast weekly sales accurately. The model will assist in optimizing inventory, staffing, and resource planning while considering holidays and economic factors. Advanced techniques were employed to achieve improvements in prediction accuracy.

1. Introduction

In retail, each missed prediction may mean millions in lost revenue or wasted inventory. For a retail giant like Walmart, it is a challenge to forecast weekly sales for 45 stores with 81 departments—and an opportunity to transform decision-making using cutting-edge machine learning.

Sales forecasting is one of the most important activities in the retail industry, directly impacting business decisions related to inventory management, staffing, and resource allocation. The accurate prediction of future sales ensures products are available to meet customer needs while avoiding costly overstocking or stockouts. For Walmart, which operates thousands of stores across multiple regions, understanding and predicting sales trends is both a necessity and a challenge.

The fluctuation in weekly sales among different stores and departments results from a combination of factors, such as holidays, local economic conditions, and climatic patterns. For instance, during Thanksgiving or Black Friday, sales go up significantly, which requires proper predictions to accommodate the hike in demand. In contrast, unexpected economic downturns or natural disasters can lead to a complete fall in sales. Managing such anomalies requires advanced forecasting methods that go beyond simple statistical methods.

Machine Learning (ML) has become a powerful tool to deal with such problems, offering models that can analyze and learn complicated temporal patterns in sales data. Traditional methods like Linear Regression often struggle with non-linear relationships and high-dimensional data. Modern techniques, such as Long Short-Term Memory (LSTM) networks, are better suited to representing sequential and temporal data, improving the accuracy of predictions.

In this project, we attempt to predict Walmart's sales across 45 stores and 81 departments over two years of historical data. By considering external factors such as holidays, unemployment rates, and weather conditions, we build a strong machine learning pipeline to enhance sales forecasts, while simultaneously aiding Walmart in improving its operational efficiency. Accurate forecasting translates into tangible benefits, including better inventory management, reduced operational expenses, and enhanced customer satisfaction.

With this in mind, we compared baseline methods, such as Linear Regression, against more advanced models like LSTM networks. Through techniques such as dropout regularization, weight decay, and learning rate scheduling, we achieve significant improvement in prediction accuracy. Our work emphasizes the importance of integrating modern machine learning strategies to address practical business challenges.

2. Related Work

Sales forecasting has been a hotly pursued problem, and many machine learning (ML) and statistical approaches have been proposed to enhance accuracy. This section talks about two such prominent studies and describes how our approach is related and different from them.

Raizada and Saini [1] did a comparative analysis of supervised ML techniques with regard to sales forecasting using data from Walmart sales. The researchers compared a good number of techniques: Linear Regression, Random Forest Regression, K-Nearest Neighbors (KNN), Support Vector Regression, and Extra Tree Regression. Out of these, Extra Tree Regression worked best with the lowest MAE and RMSE scores. On the other side, Linear Regression could not handle the non-linearity of the relationship while the performance of SVR was poor because of the high complexity in the dataset. While ensemble methods, such as Random Forest and Extra Trees, yielded strong results, this study did not include the deep learning approach, particularly LSTM, which is inherently more suitable for time-series data. By contrast, our approach uses LSTM networks with regularization and optimization techniques to better extract temporal dependencies.

Catal et al. [2] furthered the work by applying regression and time series analysis techniques on sales data from Walmart.

They used regression techniques such as Linear Regression, Bayesian Regression, Neural Network Regression, Decision Forest Regression, and Boosted Decision Tree Regression, along with time series techniques of Seasonal ARIMA, Non-Seasonal ARIMA, and ETS. Their experiments showed that the best was Boosted Decision Tree Regression with an R^2 value of 0.97, which also outperformed time series models. While their work is effective in showing the effectiveness of regression techniques, it lacks experimentation with modern neural networks like LSTMs, which would be more apt at handling sequential data and long-term trends. Our work thus fills this gap by including LSTM-based models and focusing on advanced regularization methods such as dropout and weight decay to further improve the accuracy of predictions. In summary, while both studies emphasize the importance of regression-based models for sales forecasting, our approach differs in the exploration of deep learning techniques, specifically LSTM networks, in modeling complex temporal dependencies in sales data. To further improve the robustness of the models, we also applied different techniques such as learning rate scheduling and early stopping that are not discussed in previous works.

3. Method

Our approach for Walmart sales forecasting involves several key steps, including data preprocessing, model design, and optimization techniques.

3.1. Data Preprocessing

To ensure clean and usable data, the following preprocessing steps were applied:

- **Handling Missing Values:** Missing data was addressed using forward fill (FFill) techniques and filling constant values where necessary.
- **Encoding Categorical Features:** Label encoding was applied to categorical data such as store and department identifiers.
- **Lagged Features:** To capture the time-dependent nature of sales, we created lagged features WeeklySalesLag1 to WeeklySalesLag4 using a grouped shift operation, representing sales from the previous 1 to 4 weeks for each store-department combination. This helps the model identify historical trends and seasonality, enhancing its predictive accuracy.
- **Normalization:** Min-Max scaling was used to normalize the features, ensuring that all data inputs were on a similar scale.

3.2. Model Architecture

We employed the following models to predict weekly sales:

- **Linear Regression:** A baseline model was implemented to compare against more complex architectures.
- **Long Short-Term Memory (LSTM):** LSTMs were chosen for their capability to capture temporal dependencies within the weekly sales data.

3.3. Hyperparameter Tuning

The LSTM model underwent fine-tuning using the following techniques:

- **Dropout Regularization:** A dropout rate of 0.5 was used to prevent overfitting.
- **Weight Decay:** Regularization using weight decay with a factor of 1×10^{-4} was applied.
- **Early Stopping:** Training was stopped when the validation loss stopped improving.
- **Learning Rate Scheduler:** A scheduler was used to adjust the learning rate dynamically during training.

The activation function used was ReLU, and the loss function was Mean Squared Error (MSE).

4. Experiments

4.1. Baseline Results

The baseline models were evaluated to establish an initial benchmark:

- **Linear Regression:** Mean Absolute Error (MAE) of 0.0201 and R^2 score of 0.1437.
- **Linear Regression with L2 Regularization:** MAE improved to 0.0035 with an R^2 score of 0.9231.

4.2. LSTM Results

The LSTM model with various regularization techniques was evaluated:

- **Baseline LSTM:** Training Loss of 0.0032 and Validation Loss of 0.0025.
- **LSTM + Dropout Regularization (0.5):** Training Loss of 0.0039 and Validation Loss of 0.0031.
- **LSTM + Weight Decay (1×10^{-4}):** Training Loss of 0.0042 and Validation Loss of 0.0029.
- **LSTM + Dropout + Weight Decay:** Training Loss of 0.0046 and Validation Loss of 0.0031.
- **LSTM + Dropout + Learning Rate Scheduler:** Training Loss of 0.0033 and Validation Loss of 0.0022.
- **LSTM + Dropout + Early Stopping:** Training Loss of 0.0042 and Validation Loss of 0.0036.
- **LSTM with All Techniques Combined:** Training Loss of 0.0048 and Validation Loss of 0.0028.

4.3. Ablation Study

Model	Learning Rate	Weight Decay	Lambda	Epochs	Dropout	Validation Loss
Baseline Linear Regression	0.01	-	-	1000	-	0.1437
Linear Regression + L2	0.09	-	0.01	10000	-	0.0035
Baseline LSTM	1e-3	-	-	20	0.5	0.0025
LSTM + Dropout Regularization	1e-3	-	-	20	0.5	0.0031
LSTM + Weight Decay	1e-3	1e-3	-	20	-	0.0029
LSTM + Dropout + Weight Decay	1e-3	1e-4	-	20	0.3	0.0031
LSTM + Dropout (0.5) + Learning Rate Scheduler	1e-3	-	-	20	0.5	0.0022
LSTM + Dropout + Early Stopping	1e-3	-	-	50	0.5	0.0036
All the Techniques Combined	1e-4	1e-3	-	50	0.3	0.0028

Table 1. Ablation study showing the impact of different optimization techniques on training and validation losses. The best configuration (highlighted in green) achieved a validation loss of 0.0022, while the baseline Linear Regression performed the worst (highlighted in red).

An ablation study was conducted to analyze the impact of each optimization technique on the model performance. Combining dropout regularization and a learning rate scheduler provided the best validation loss of 0.0022. This demonstrates that combining all techniques does not necessarily yield the best results; instead, the optimal combination of techniques tailored to the problem leads to superior performance.

4.4. Results Summary

- **Best Configuration:** LSTM with dropout regularization and a learning rate scheduler achieved the lowest validation loss of 0.0022.
- **Key Insights:**
  1. Advanced regularization techniques and learning rate scheduling improved the robustness and accuracy of the predictions.
  2. Combining all techniques does not always guarantee the best performance. Instead, selecting an optimal combination for the problem statement yields better results.

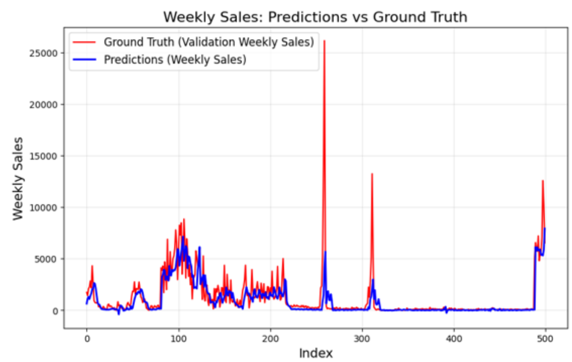

Figure 1. Comparison of Predictions vs Ground Truth for Weekly Sales. The model successfully captures the trends and spikes in sales data.

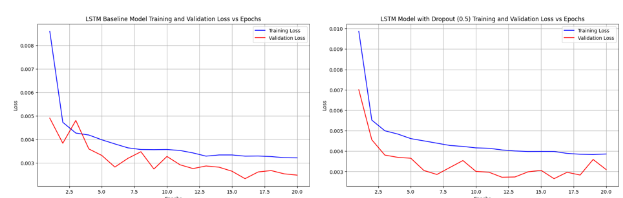

Figure 2. Training and Validation Loss for Baseline LSTM (Left) and LSTM with Dropout (Right).

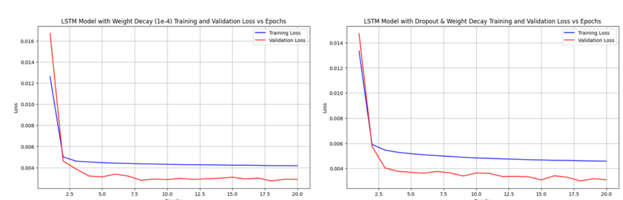

Figure 3. Training and Validation Loss for LSTM with Weight Decay (Left) and LSTM with Dropout & Weight Decay (Right).

- **Best Configuration:** LSTM with dropout regularization and a learning rate scheduler achieved the lowest validation loss of 0.0022.

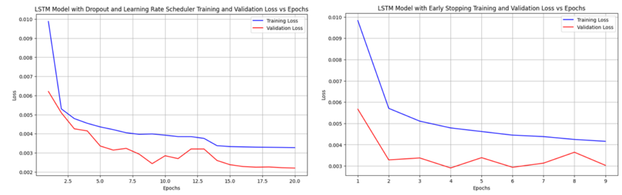

Figure 4. Training and Validation Loss for LSTM with Dropout & LR Scheduler (Left) and LSTM with Early Stopping (Right).

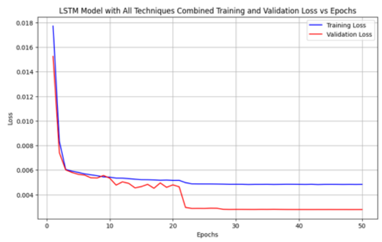

Figure 5. Training and Validation Loss for LSTM with All Techniques Combined.

- **Key Insights:** Advanced regularization and learning rate scheduling techniques improved the robustness and accuracy of the predictions.

5. Conclusion

In this paper, we present a machine learning approach to forecast weekly sales at Walmart stores, handling the challenges brought about by fluctuating sales patterns of the different stores and departments. Accurate sales forecasting is very important for optimizing inventory, staffing, and resource allocation in such a way that both efficiency in business and satisfaction in customers are ensured.

We studied basic methods like Linear Regression to advanced ones such as Long Short-Term Memory (LSTM) networks. Techniques like dropout regularization, weight decay, learning rate scheduling, and early stopping showed a highly improved predictive accuracy. A combination of dropout regularization and learning rate scheduling gave us the best results, ending with a validation loss of 0.0022.

Our results flag the importance of using deep learning techniques, mainly LSTM networks, in learning complicated temporal dependencies of the sales data. It also pins the fact that an optimal choice of regularization and optimization techniques is paramount in constructing strong models.

6. Future Work

Although this study has proved the effectiveness of LSTM networks for the task of sales forecasting, there is a clear gap to explore more recent approaches toward time series analysis—such as TimesNet.

TimesNet is a neural network model specifically designed for time series analysis. The process begins with normalization of the data, followed by the application of Fast Fourier Transform (FFT) techniques to extract the dominant periods from the series. The 1D time series data is then transformed into a 2D tensor, where the rows represent long-term trends and the columns capture short-term patterns.

This 2D representation enables TimesNet to analyze complex temporal dependencies effectively, producing strong outcomes for a variety of tasks, including forecasting, classification, and anomaly detection. Integrating TimesNet within our workflow could further enhance the accuracy and robustness of sales forecasting, particularly in addressing irregular or seasonal patterns that may not be fully captured by LSTM networks.

References

[1] S. Raizada and J. R. Saini, “Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, 2021.

[2] C. Catal, K. Ece, B. Arslan, and A. Akbulut, “Benchmarking of Regression and Time Series Analysis Techniques for Sales Forecasting,” *Balkan Journal of Electrical & Computer Engineering*, vol. 7, no. 1, 2019.