Machine Learning Model Evaluation Report: Sales Forecasting

1. Executive Summary

This report evaluates a machine learning model for predicting supermarket sales using historical transaction data. The goal was to forecast future sales (Total) based on features like Branch, Product line, Quantity, and temporal trends.

Key Findings:

- **Best Model:** XGBoost outperformed Random Forest with lower error rates.
- **Top Influencers:** Quantity and Unit price were the most significant predictors.
- **Error Rates:**
    - **XGBoost:** MAE = $23.12, RMSE = $32.89
    - **Random Forest:** MAE = $25.34, RMSE = $35.67

---

2. Methodology

Data Preprocessing

- **Temporal Features:** Extracted Year, Month, Day, and DayOfWeek from Date.
- **Categorical Encoding:** One-hot encoded Branch, City, Product line, and Payment.
- **Train-Test Split:** Time-based (80% train, 20% test) to preserve chronological order.

Models Tested

MODEL	HYPER PARAMETER
Random forest	**n_estimators = 100, random state = 42**
XGBOOST	**n_estimators = 100, learning rate = 0.1**

---

3. Performance Metrics

Model	MEAN ABSOLUTE ERROR MAE (↓)	ROOT MEAN SQUARED ERROR RMSE (↓)
Random Forest	$25.34	$35.67
XGBoost	$23.12	$32.89

Interpretation:

- **MAE (Mean Absolute Error):** XGBoost's predictions are, on average, **$23.12** away from actual sales.

- **RMSE (Root Mean Squared Error):** Larger errors are penalized more heavily (XGBoost RMSE = **$32.89**).

---

4. Feature Importance

Top 5 Features Impacting Sales (XGBoost):

1. Quantity (Impact: **42%**)
2. Unit price (Impact: **28%**)
3. Product line Fashion accessories (Impact: **8%**)
4. Month (Impact: **6%**)
5. DayOfWeek (Impact: **4%**)

Insight: Sales are highly driven by transaction volume (Quantity) and pricing (Unit price), with minor seasonality effects.

---

5. Limitations & Future Improvements

Limitations

- **Short-Term Focus:** The model predicts daily sales but may not capture long-term trends.
- **External Factors:** Promotions, holidays, or economic shifts are not included.

---

6. Conclusion

The XGBoost model provides a robust baseline for sales forecasting, with room for improvement via time series techniques and feature enrichment.

Appendix:

- Dataset link: *https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales?resource=download*