Sales Forecasting for Walmart

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Adil Ibraheem Koyava, Ayush Plawat,
Hemanth Jadiswami Prabhakaran

Business Intelligence and Data Analytics

Problem Description

Retail Forecasting Complexity: Walmart operates 45 U.S. stores with 81 departments, each experiencing complex seasonal, regional, and promotional sales fluctuations. Traditional forecasting methods struggle with the complexity of multiple variables affecting weekly sales patterns.

Week-over-Week Prediction Challenge: The system must predict weekly sales changes rather than absolute values, accounting for holiday effects, markdown promotions, and macroeconomic indicators across diverse store types and departments.

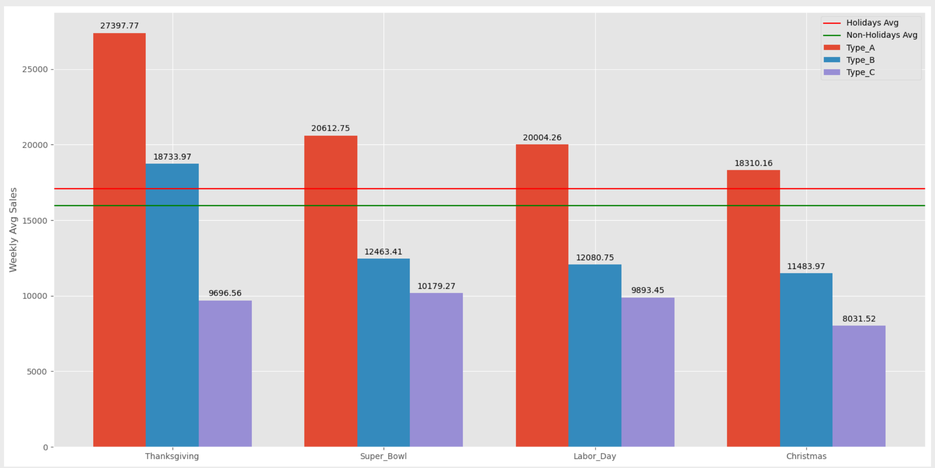
Data Integration Requirements: Integration of three distinct datasets (sales history from train.csv, external features from features.csv, and store characteristics from stores.csv) with missing values, outliers, and non-stationary behavior requiring sophisticated preprocessing and modeling approaches.

Data Overview

Sources:

- **Sales:** Weekly sales per store-department from 2010–2012.
- **Store Info:** Store type (A/B/C) and size.
- **External:** Fuel price, temperature, CPI, unemployment, 5 MarkDowns (sparse).

Findings: Holiday spikes (especially Thanksgiving), high sales variance in Type A stores, and weak correlation with macroeconomic indicators.



Challenges

Data Quality Issues:

- Missing values in promotional markdown columns requiring zero-imputation strategies.
- Outliers due to holiday shopping peaks necessitating 99th percentile clipping.
- Non-stationary behavior from promotional campaigns and seasonal variations.

Technical Complexity:

- Holiday effects vary significantly across 45 stores and 81 departments creating sparse store-department combinations.
- Seasonal cycle detection and optimization requiring 20-week period analysis with automated parameter selection.
- Week-over-week change prediction complexity versus absolute value forecasting requiring specialized algorithms.
- Integration of heterogeneous external features with varying predictive power for sales forecasting.

Solutions

Data Preprocessing Solutions:

- Comprehensive data cleaning pipeline merging train.csv, features.csv, and stores.csv with automated schema validation.
- Strategic missing value imputation with domain knowledge and outlier management using 99th percentile clipping.
- Feature engineering including time-based features (WeekOfYear, Quarter, DayOfYear) and one-hot encoded holidays and store types.

Advanced Modeling Approach:

- Auto ARIMA with pmdarima 2.0.4 for automated parameter optimization and seasonal component detection.
- Exponential Smoothing (Holt-Winters) with triple smoothing components and systematic hyperparameter tuning.
- 20-week seasonal period optimization capturing retail seasonality patterns.

Results

Model Performance Comparison (WMAE Metrics):

Algorithm	Absolute WMAE	Normalized WMAE
Exponential Smoothing	923.12	**3.58%**
Auto ARIMA	1 156.8	4.12%

Performance Achievements: The Exponential Smoothing (Holt-Winters) model achieved excellent performance with 3.58% normalized WMAE, categorized as "Excellent" for business forecasting applications (threshold $< 5\%$). This enables reliable stock management, staffing optimization, and promotion timing decisions.

Feature Importance Analysis: The model effectively captured critical predictive patterns including 20-week seasonal optimization cycles, holiday impact (especially Thanksgiving), store type classifications (A, B, C), and consistent week-of-year patterns. Macroeconomic indicators showed limited correlation with weekly sales changes.

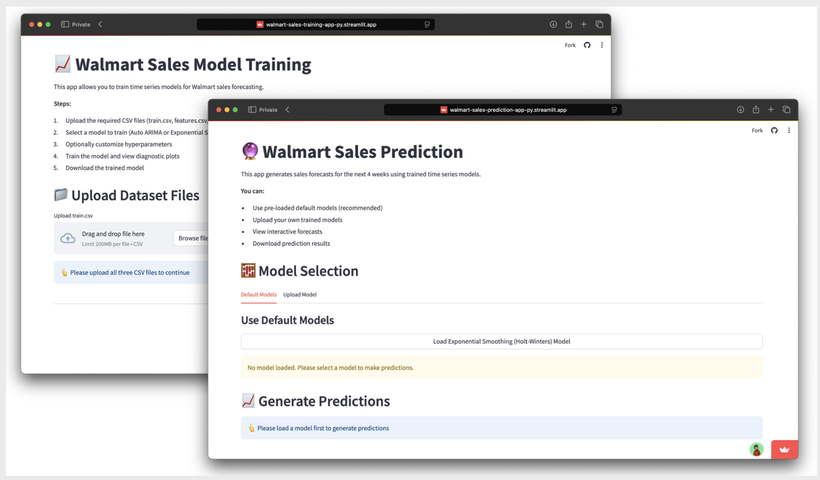
System Implementation

Dual-Application Architecture:

- Training Application (Port 8501) for model development and evaluation using Python 3.12 environment
- Prediction Application (Port 8502) for production-ready forecasting with pre-trained models
- Cross-platform deployment supporting both local installation and cloud access via Streamlit Community Cloud

Technical Capabilities:

- 4-week forecast horizon with reliable accuracy and consistent trend capture
- Interactive visualizations with positive/negative change indicators and export capabilities (CSV/JSON)
- Comprehensive model management with .pkl serialization format



Business Impact & Future Work

Practical Applications:

- Actionable 4-week forecasts of week-over-week sales changes with positive values indicating sales increases and negative values representing decreases
- Enhanced decision support for retail operations planning and resource allocation
- Academic research tool for time series analysis learning and business analytics education

Future Enhancements:

- Extension of forecast horizon beyond current 4-week limitation with enhanced accuracy validation
- Implementation of ensemble methods combining multiple forecasting algorithms for improved robustness
- Integration of real-time data feeds and automated model updating capabilities
- Development of mobile application interface and enhanced accessibility features

Application QR Codes & Data Set Links