

SALES FORECASTING

Introduction :

This document presents a comprehensive analysis of the sales forecasting model developed for predicting department-wide sales across 45 retail stores. The model leverages historical sales data, store information, and macroeconomic factors to forecast future sales.

DATA OVERVIEW :

1. stores.csv
 - Contains anonymized details about the 45 stores.
 - Includes information on the type and size of each store.
2. features.csv
 - Provides data on promotional markdown events and macroeconomic factors.
 - Markdown events are linked to major holidays like the Super Bowl, Labour Day, Thanksgiving, and Christmas.
3. train.csv
 - Historical training data that includes past sales figures for various departments across all stores.

MODEL PERFORMANCE:

1. **SARIMAX Model Analysis:** This section evaluates the SARIMAX model's efficacy in forecasting weekly sales for Store 1 - Department 1. The model integrates both seasonal and non-seasonal elements to anticipate sales patterns.
 - **Model Specification:** The SARIMAX model is configured with parameters (1, 1, 1) for non-seasonal and (1, 1, 1, 52) for seasonal components, reflecting annual seasonality, complemented by a constant trend element.
 - **Model Assessment:** The model's fit was gauged using the L-BFGS-B optimization method, with key points summarized below:
 - Iteration reached the preset cap of 50 cycles.
 - Log-likelihood at convergence stood at 6.5023.
 - The final gradient norm was a minimal 0.006958, indicating a negligible gradient at the solution.
 - A ConvergenceWarning highlighted the need for additional parameter refinement due to non-convergence.
 - **Forecasting Insights:** The model projected weekly sales for a year ahead, noting that:
 - Predictions mirrored historical sales trends and seasonality.
 - Confidence intervals narrowed over time, signifying greater short-term forecast certainty.

- **Visual Assessment:** The graph juxtaposes actual sales with projected figures and confidence bands, showcasing the model's adeptness at capturing sales patterns.
 - **Conclusion:** The SARIMAX model's proficiency in recognizing seasonal trends is evident, yet parameter optimization is essential to bolster forecast precision and dependability.
2. **KNN vs. Random Forest Comparative Analysis:** This analysis contrasts the K Nearest Neighbors (KNN) and Random Forest models in sales prediction, trained on a feature-rich dataset, excluding certain variables.
- **Random Forest Insights:**
 - Configured with 250 trees, a depth of 15, and a minimum split of 5 samples.
 - Predictions ranged from 0 to 6554.03, with some exact zeros hinting at potential data gaps or predictive challenges.
 - The Mean Absolute Error (MAE) was 469.45, indicating the average deviation from actual sales.
 - **KNN Insights:**
 - The KNN model utilized 3 neighbours for its forecasts.
 - Predictions spanned from 0 to roughly 5696.83, including zeros.
 - A slightly lower MAE of 445.63 suggests marginally enhanced prediction accuracy.
 - **Conclusion:** Both models demonstrate reasonable predictive accuracy, with KNN marginally outperforming in terms of MAE. Nonetheless, the zero-value predictions call for further scrutiny to refine the models' predictive power.

KEY OBSERVATIONS:

- **Random Forest Model:** Its ensemble nature adeptly captures complex data patterns, though the MAE indicates room for improvement.
- **KNN Model:** Its lower MAE reflects a better fit, and its simplicity offers a computationally efficient prediction method.
- **SARIMAX Model:** Its ability to factor in seasonal trends is advantageous, despite convergence issues necessitating parameter adjustments.
- **Comparative Overview:** Model selection should be tailored to the dataset's characteristics, available computational resources, and forecasting objectives, emphasizing the need for model customization for optimal accuracy and reliability.

CONCLUSION :

The examination of SARIMAX, KNN, and Random Forest models for sales forecasting highlights their unique strengths and optimization needs. While all models are capable of reasonably accurate forecasts, fine-tuning is imperative for enhanced precision and reliability. The choice of model should be informed by the dataset's nuances, computational capacity, and specific forecasting aims, with the comparative analysis reinforcing the significance of tailored model selection for the most effective sales predictions.