

Predictive Analytics for Inventory Management at Gala Groceries

Gala Groceries, a leading grocery store chain, sought to minimize waste and optimize inventory by accurately predicting product stock levels. A predictive model was developed to forecast hourly stock levels using sales and IoT sensor data, allowing for more thoughtful procurement strategies.

The challenge faced by Gala Groceries was to balance the freshness of their produce with the risks of overstocking and understocking. This required a predictive approach to manage inventory levels dynamically. (fig. 1)

Data from sales transactions, storage temperatures, and stock levels were analyzed. Key insights included peak sales times and popular payment methods. Exploratory analysis also highlighted the importance of various product categories. (Table 1)

A two-pronged approach was adopted, starting with a Random Forest model followed by a more sophisticated 1D CNN to capture the temporal patterns in the data. Cross-validation ensured model reliability, while hyperparameter tuning was conducted via grid search. Regularization techniques were applied to refine the model's performance.

The CNN model with regularization demonstrated a promising MAE of 0.2219 and an RMSE of 0.2635. This combination of MAE and RMSE underscores the model's potential for making reliable stock-level predictions. Feature importance analysis further revealed that unit price and temperature were significant predictors. (Fig. 3)

The project achieved a promising model for forecasting stock levels. Future work will focus on integrating additional data sources, refining the model, and exploring deployment strategies for real-time predictions.

The project leveraged several datasets to inform the predictive model:

- sales.csv: Contains transactional data capturing sales figures, which helped identify buying patterns and peak purchasing times.
- sensor_stock_levels.csv: Includes stock level data from in-store sensors, crucial for understanding inventory dynamics.
- sensor_storage_temperature.csv: Provides temperature readings from storage areas, offering insights into the conditions affecting produce freshness.
- sample_sales_data.csv: An additional sales dataset used to corroborate findings and enhance the model's accuracy.

These datasets underwent rigorous exploratory analysis, revealing essential insights such as the busiest transaction hours, popular product categories, and customer purchase behaviors. Visualizations of unit price distributions and payment types offered a deeper understanding of the factors influencing stock levels.

Graphs:

Fig 1

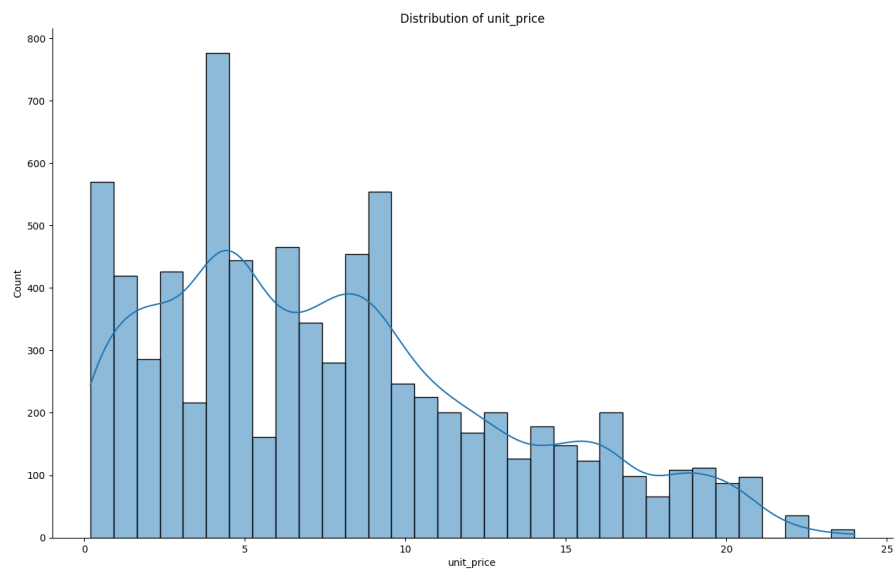


Table 1

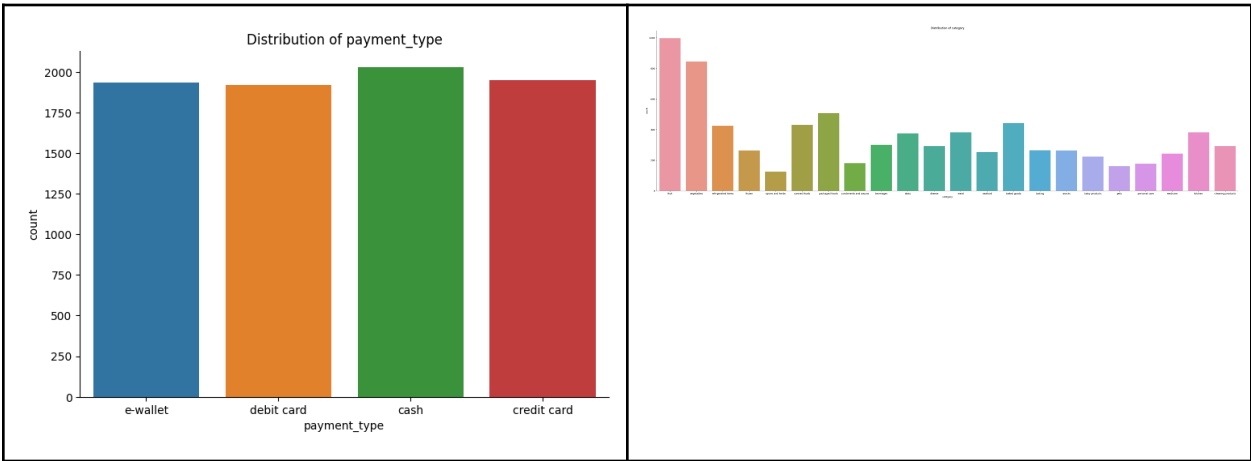
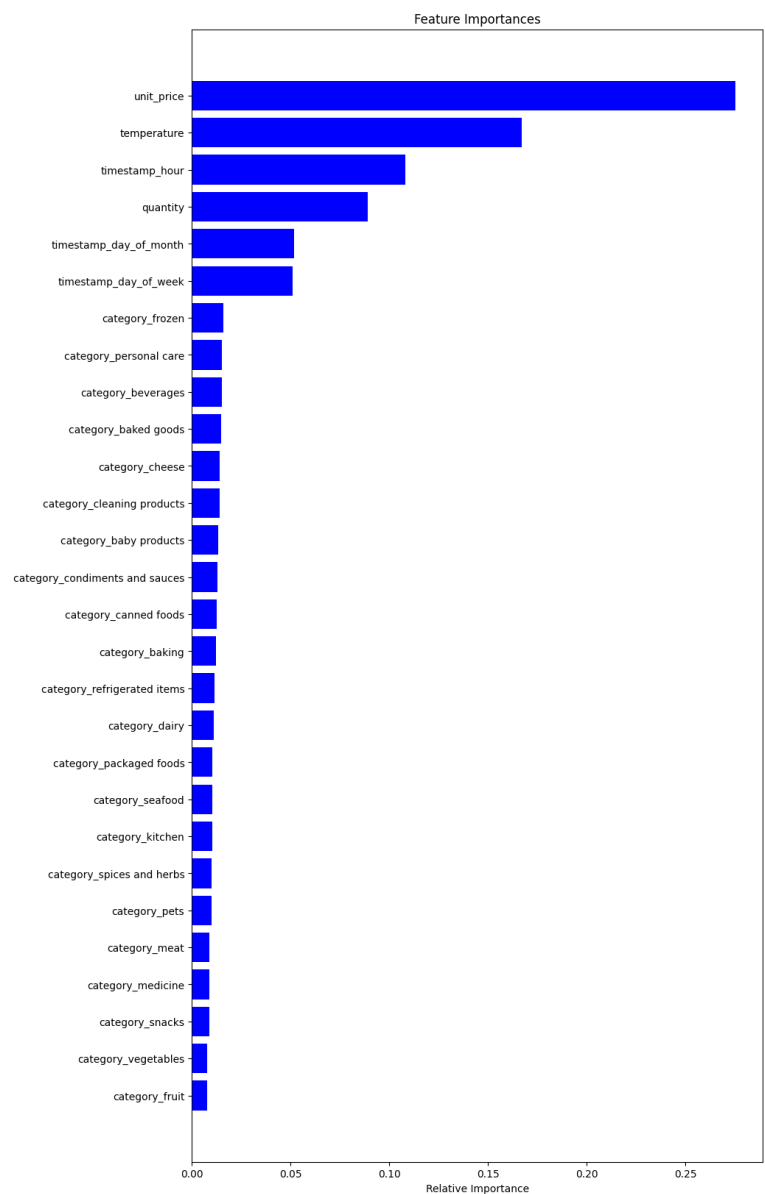


Figure 3



Time log:

- 12/13 - 4 hours - Data preparation, designing, and understanding.
- 12/14 - 3 hours - Continued data preparation, designing, and experimenting.
- 12/15 - 3 hours - ML production, hyperparameter tuning.
- 12/15 - 1 hour - research.
- 12/16 - 4 hours - Testing, terating deep learning-based models.
- 12/16 - 1 hours - Research.
- 12/18 - 1 hour - Analyzing results and debugging.
- 12/19 - 4 hours - 1D CNN Model implementation, training, experimenting.
- 12/19 - 1 hour - Reading.
- 12/20 - 3 hours - Implementation with regularization, training, experimenting.
- 12/21 - 2 hours - Analyzing results, report writing, code finalization.