Forecasting Time Series Sales at Favorita Stores

Background:

Within the United States, it's estimated that 30-40% of food production, on average, goes to waste[1]. Grocery stores contribute to much of this waste, as a major supplier for American's food. Sales forecasting plays a crucial role in making sure food waste is kept to a minimum. If grocery stores continue to overshoot their sales estimate, we will stay stagnant at about 30% of food in grocery stores being thrown away[2]. Undershoot a forecast, and shoppers are left frustrated. This project explores how time-series forecasting ML models can play a role in minimizing food waste at grocery stores. Corporación Favorita, a large Ecuadorian-based grocery retailer, has offered up a dataset of sales over time in their grocery store, and includes many features that could point to more accurate estimating grocery sales.

Literature Review:

The use of historical time-series data to predict future sales has been a widely studied and explored topic with many successes. A few notable examples include work by Ensafi, Amin, Zhang, and Shah[3] in the International Journal of Information Management Data Insights as well as Fredén and Larsson's work[4] at the Kth Royal Institute of Technology. The former implemented a Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) leading to a Convolutional Neural Network (CNN). The latter also implemented a LSTM model with XGBoost.

Motivation:

To forecast future sales and analyze the impact of events Prophet will be used. Prophet was developed by Facebook to predict business times series sales data. Prophet breaks the data down into three components[5]:
1. Overall trend of the times series data using piecewise linear regression.
2. Seasonality component using Fourier series
3. Noise which is random and cannot be explained by the previous two points

Prophet is also advantageous because other features can be included in it. This can be things like the oil price in Ecuador, marketing campaigns, or even holidays.

Prophet will be compared against a baseline of (S)ARIMA which is ARIMA with a seasonal element added in. (S)ARIMA can not include special events like holidays or promotions. It will predict as if those do not exist.

Citations:

[1] Center for Food Safety and Applied Nutrition. "Food Loss and Waste." *U.S. Food and Drug Administration*, FDA, www.fda.gov/food/consumers/food-loss-and-waste. Accessed 18 Jan. 2024.
[2] "Food Waste in America in 2024: Statistics & Facts: RTS." *Recycle Track Systems*, www.rts.com/resources/guides/food-waste-america/. Accessed 18 Jan. 2024.
[3] Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, *2*(1), 100058. sciencedirect. https://doi.org/10.1016/j.jjimei.2022.100058
[4] Fredén, D., & Larsson, H. (2020). Forecasting Daily Supermarkets Sales with Machine Learning. *Royal Institute of Technology*.
[5] Khare, P. (2023) *Understanding FB Prophet: A Time Series Forecasting Algorithm*, *medium.com*. Available at: https://medium.com/illumination/understanding-fb-prophet-a-time-series-forecasting-algorithm-c998bc52ca10