Point Forecasting on the M5 Walmart Sales Data

Data 695 Group Capstone Project

Gautham Chandra Shekariah

Joshua Quartey

Prince Oloma

Preliminary data

The M5 dataset contains daily unit sales for 3,049 Walmart products organized hierarchically by category (3), department (7), store (10), and state (3), giving 42,840 time series across 12 cross-sectional levels. The competition required submitting point forecasts for the 30,490 bottom-level (item–store) series for a 28-day horizon. Higher-level forecasts are obtained by summing bottom-level predictions. The dataset also included explanatory variables such as prices, promotions, calendar/holiday indicators.

Background

For four decades, the M competitions have benchmarked forecasting methods using large, real-world datasets. M5 focused on retail sales, with a hierarchical structure and shared calendars and exogenous features (price, promos, events). Participants submitted bottom-level point forecasts, which were aggregated to evaluate all levels. Accuracy was assessed with WRMSSE (weighted root mean squared scaled error). For each series, RMSSE was computed (mean-oriented squared-error metric), then averaged using hierarchy-appropriate weights to produce a single score. This metric was chosen to avoid favouring trivial near-zero predictions on intermittent series and keeps evaluation comparable across scales.

Abstract

This project investigates point-forecast modelling for the M5 Walmart hierarchical data. We compare gradient-boosted trees with other ML/statistical approaches across hierarchy levels. We investigate any gains in using calendar, price, promotion, and event features over sales-only lags. Finally, to address intermittent demand at the item–store level, we test a two-stage hurdle approach (occurrence classification + positive-sales regression) versus a single regressor. Our expected contributions are (i) a level-wise, reproducible baseline for point forecasts on M5 and (ii) clear guidance on which features and modelling choices are substantial by aggregation level.

Tools and Modelling

The initial phase establishes a baseline model using Ordinary Least Squares (OLS) linear regression to provide a simple, interpretable benchmark against which more complex methods will be measured. The core of the project involves exploring more sophisticated non-parametric machine learning techniques to capture complex, non-linear relationships and interactions within the time-series data. This will involve evaluating models such as Gradient Boosting Machines and Random Forests.

The analysis will begin by extracting a rich set of features from the pandas datetime object, including cyclical features (e.g., day of the week, month, quarter), temporal features (e.g., time since last holiday, week of the year), and lagged variables of the target demand.

The final models will be selected through hyperparameter tuning. The process will initially utilize Grid Search and Random Search due to their straightforward implementation. However, should time limitations or the high dimensionality of the parameter space become a constraint, the project will transition to more efficient methods like Hyperopt.

Research questions

1. Do gradient boosted models beat other models/statistical approaches at bottom vs higher hierarchy levels?
2. How much do calendar, price, promo, or event features improve accuracy over just sales lags, and does that gain vary by level?
3. What approach is effective for addressing sparse/zero-heavy series at the lower aggregation levels? Does a two-stage model (classify occurrence, then regress positives) beat a single regressor, especially at the bottom level?

References

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting, 38*(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 competition: Background, organization, and implementation. *International Journal of Forecasting, 38*(4), 1325–1336. https://doi.org/10.1016/j.ijforecast.2021.07.007

Kaggle. (2020). *M5 forecasting – accuracy*. Kaggle.
https://www.kaggle.com/competitions/m5-forecasting-accuracy