Technical Report - Alki Demand Forecasting Pipeline

Author: Sabarish Ravichandran Mehala

1 Objective

The goal of this project was to design a robust and scalable forecasting pipeline for multiple clients of Alki each exhibiting distinct demand behaviors and data lengths. The pipeline enables

- Consistent data processing and cleaning across all the available clients
- Rolling cross-validation for fair temporal evaluation
- Support for multiple model families, both classical statistical models and machine learning models
- Flexible configuration via YAML for per-customer customization and tuning

2 Methodology Overview

The forecasting process was built as a modular, end to end pipeline with the following stages:

- **Data Loading and Normalization**: Input files are read using a schema-agnostic loader that maps aliases for the present columns.
- **Preprocessing and Cleaning**: Daily re-indexing, hybrid gap filling, inactivity trimming, and causal-safe interpolation ensure data consistency.
- **Feature Engineering**: Lag, rolling, and holiday-based features are created using configurable window sizes. The holiday features are derived from external data.
- **Model Training**: Multiple candidate models are trained per customer, using each customer's individual characteristics and data length
- **Cross-Validation**: Expanding-window rolling CV with customer-specific folds ensures realistic temporal validation
- **Evaluation and Comparison**: Metrics such as MAE, RMSE, and sMAPE quantify performance. The best model is chosen per customer based on the composite metric involving weightages of these metrics. The weight of each metric is decided based on the use case.

3 EDA

3.1 Argalys (Data Period (After Cleaning): 2020 - 2023)

Based on the EDA, it is evident that this customer has continuous daily operations. When it comes to trend and seasonality, there is a clear growth until mid 2022, then mild stabilization. The additive STL fits best based on the analysis. There is a weak monthly seasonality with slight peaks in January and May, dips in June to August. There are also weak weekly activity from Saturday to Monday. From ACF and PACF analysis, we can find there is a short term persistence typical of daily demand. It is stationary. The mean and variance stable after mild detrending. Differencing is not required. In short, it shows continuous but noisy daily shipments, where growth stabilizes after the expansion in 2021–22.

3.2 Les Miraculeux (Data Period (After Cleaning): 2019 - 2023)

Based on EDA, this customer has high volume continuous operations. There is a linear upward trend through 2023 shows a strong growth. This fits best with Multiplicative STL based on the analysis. There are annual seasonality peaks in May and November, and dips in July–August and December. Based on ACF and PACF analysis, this shows a strong yearly periodicity and short lag dependencies. It is non stationary and requires first differencing or trend removal before modelling. This customer represents mature, fast moving product because of their high and growing demand with strong seasonality.

3.3 Minci Delice (Data Period (After Cleaning): 2017 - 2023)

Based on EDA, this customer shows continuous daily activity with moderate volatility. It has a bell shaped trend, where it declines post 2021. This fits best with Multiplicative STL based on the analysis. There is annual seasonality peaks in May and trough in December and a weekly pattern Saturdays to Tuesdays are active periods. Based on ACF and PACF analysis, it shows weekly cycle and there is a short term correlation. It is stationary and variance is stabilized using multiplicative decomposition. In short, there is a steady operational rhythm with clear annual and weekly patterns.

3.4 Nutravance (Data Period (After Cleaning): 2015 - 2023)

This customer has long historical coverage and moderate daily volatility. There is a continuous upward trend through 2023 indicating steady long term growth. This fits best with Multiplicative STL based on the analysis. There is annual seasonality peaks in March and September and a trough in August and a weekly pattern Monday to Thursday are active periods. Based on ACF and PACF analysis, there is medium-term persistence as the ACF decays slowly, suggesting an autoregressive component. It is stationary. This customer has a stable long term growth with consistent repeating seasonality.

4 Model Design And Rationale

Each customer's modelling choices were guided by EDA insights. Each model was wrapped with fault-tolerant 'safe' wrappers ensuring fallback to last-value forecasts in case of convergence failures.

4.1 Argalys

The main models considered for this customer are ARIMA, ETS-Add and LightGBM. The sole reason to choose these models are because of their additive trend and stable variance. This is very well suited for ARIMA/ETS and LightGBM captures short term volatility

4.2 Les Miraculeux

The main models considered for this customer are SARIMA, Prophet and LightGBM. The rationale behind the selection is because of the multiplicative seasonality and SARIMA effectively handles calendar seasonality. LightGBM was chosen because of non linear spikes.

4.3 Minci Delice

The main models considered for this customer are ETS-Mul, SARIMA and Prophet. The rationale behind the selection is weekly multiplicative patterns which fit prophet and ETS.

4.4 Nutravance

The main models considered for this customer are ETS-Add, SARIMA and LightGBM. The rationale behind the selection is long consistent series. LightGBM adds flexibility.

5 Cross-Validation And Evaluation

The strategy considered here is Rolling Expanding Window Cross validation. In this, each fold trains on all data up to the anchor date and validates the next 25 days (since the forecasting horizon is 21 days). Then the anchor moves forward by 28 days between the folds. There are totally 7 folds per customer to maintain temporal integrity. This strategy balances data efficiency and realistic simulation of rolling forecasts. The Evaluation is based on three key metrics: sMAPE, RMSE, and MAE. I adopted a custom scoring using these metrics that fits the use case because of the model inefficiencies, so the final composite scoring is based on 30% of MAE and 50% of RMSE and 20% sMAPE.

6 Result Summary

Table 1: Composite Model Performance Summary and Observations

Customer	Best Model	Key Observations
ARGALYS	LightGBM	Captured short-term demand volatility and local structural shifts better than statistical baselines. Its feature-driven approach leveraged lag and calendar effects effectively, yielding strong accuracy across folds.
LES MIRACULEUX	SARIMA (1,1,1)(1,1,1,12)	Performed best on this highly seasonal and trend-driven series. The model captured annual and monthly cycles accurately while maintaining smooth residual behavior. LightGBM and Prophet tended to overfit promotional spikes.
MINCI DELICE	LightGBM	Excelled in modeling mixed seasonalities and nonlinear dynamics, particularly after 2021 when demand patterns changed. Outperformed SARIMA by adapting to irregular periodicity and local variations in amplitude.
NUTRAVANCE	SARIMA (1,0,1)(1,1,1,7)	Delivered the most stable forecasts for long, steady historical data. Effectively modeled weekly cycles and trend consistency with minimal forecast drift, outperforming ML alternatives on stability and residual balance.

The plots show good alignment between forecast and actual, except for high-amplitude spikes where statistical models lag slightly.

7 Model Strengths

- The complete workflow is end to end reproducible and keeps every artifacts versioned and traceable
- The per-customer customization lets each client specify its own transformation, CV window and candidate families.
- Robust CV & Diagnostics with composite scoring, analysing peak focused metrics gives a solid view of both average performance and high-demand behavior before promoting a model.
- Classical SARIMA/ETS/Prophet coexist with gradient-boosted GBMs and share deterministic exogenous features (e.g., calendar and holiday effects). That mix covers both interpretable seasonal structure and nonlinear effects in a consistent pipeline.
- Consistent feature engineering generate aligned lag, rolling, trend, calendar, and holiday signals for training and inference, reducing leakage risk and ensuring deterministic models get the same exogenous context they saw in CV.
- Final forecasts now pull model-native intervals (SARIMA/ETS/Prophet) or bootstrap residuals for GBM, giving horizon-aware P10/P90 bands instead of a flat, homoscedastic σ assumption.

8 Model Limitations

- The model has poor performance for spikes which could be because of promotions, or other external events resulting in missed high-performance peaks.
- Prophet occasionally fails due to stan backend issues. So made it smoothly shifts to ETS- Mul if it fails.
- The LightGBM performance is heavily dependent on the feature scaling and lag window choice.
- SARIMA convergence warnings arise for long seasonal periods.
- The residual based quantiles are generated using a simple normal approximation. This assumes gaussian errors, symmetric ignores temporal dependencies in residuals.
- Once the composite winner is chosen no ensembling or model averaging is performed. If multiple models are close in score, averaging might reduce variance.
- For some customers (Argalys & Minci Delice) we have very narrow confidence P10 and P90 involving more risk. This is due to LightGBM quantile objectives.

9 Potential Improvements

- Train a spike-aware model by integrating change point detections or regime switching methods (eg: Bayesian Structural time series)
- Using other external factors like promotion events, weather (if the location of the warehouses and business operations countries are known)
- Combine statistical models with machine learning models to balance the bias and variance.
- Parallelization of the hyperparameter tuning or using Optuna to reduce hyperparameter tuning time.

10 Key Takeaways

- A unified, modular forecasting framework was successfully developed and validated.
- Each client benefits from tailored models aligned with their data structure.
- The system's configuration-driven design (YAML + Make + uv sync) ensures reproducibility and ease of deployment.