VN1 Forecasting Challenge

5th Place Submission

An Hoang | 13 November 2024

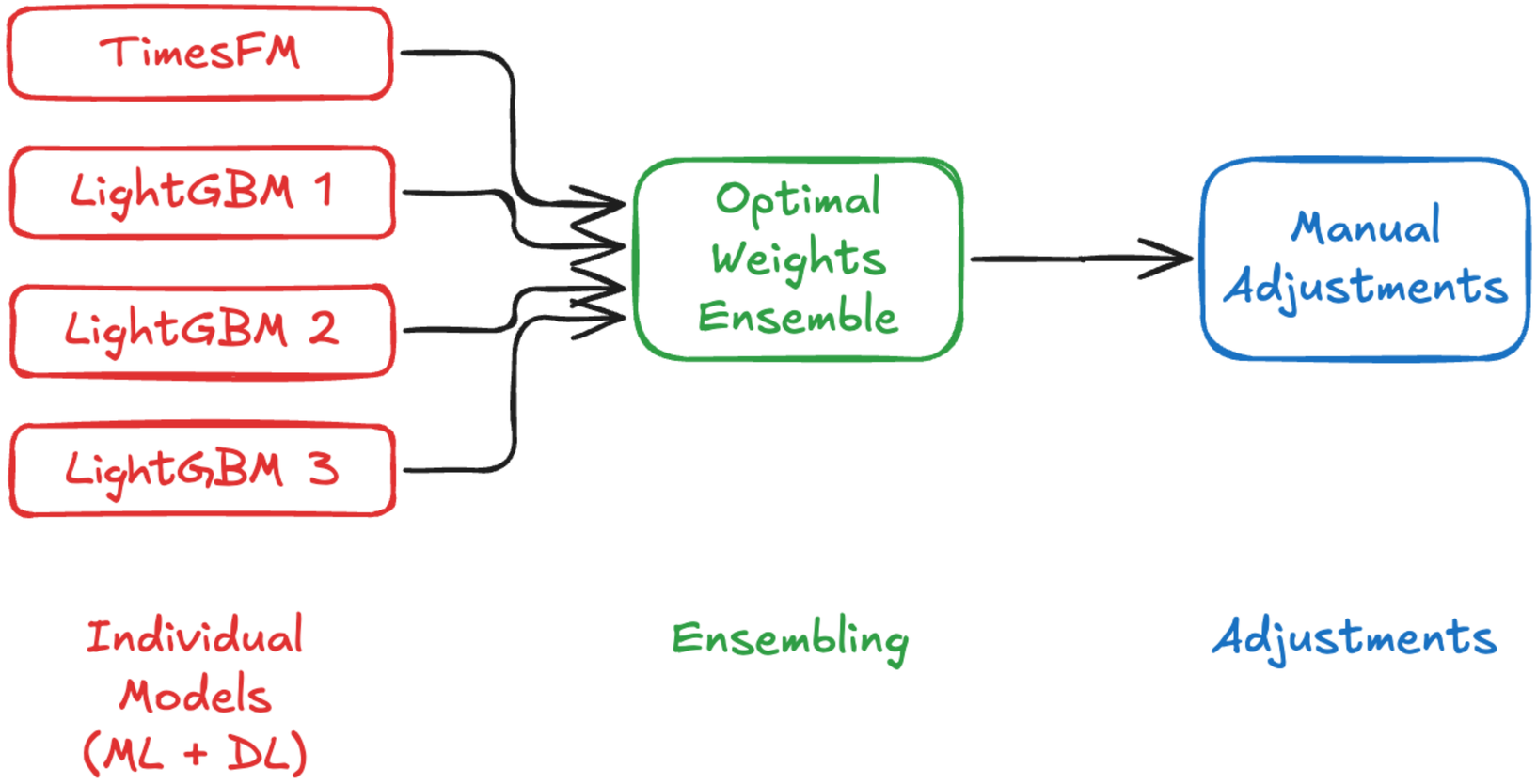
Hey, my name is An 👋

Introduction

- 🎓 M.Sc. in Mathematics
- 🛒 Data Scientist @ dm-drogerie markt
- 👨‍🍳 Amateur Chef
- 👶 1st time participating in forecasting competition

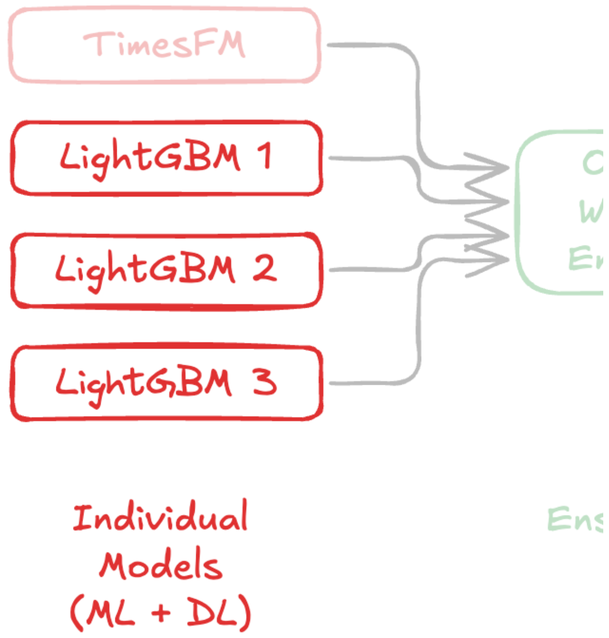
The forecast is an ensemble of diverse models

Modelling Overview



Three recursive LGBM models were trained

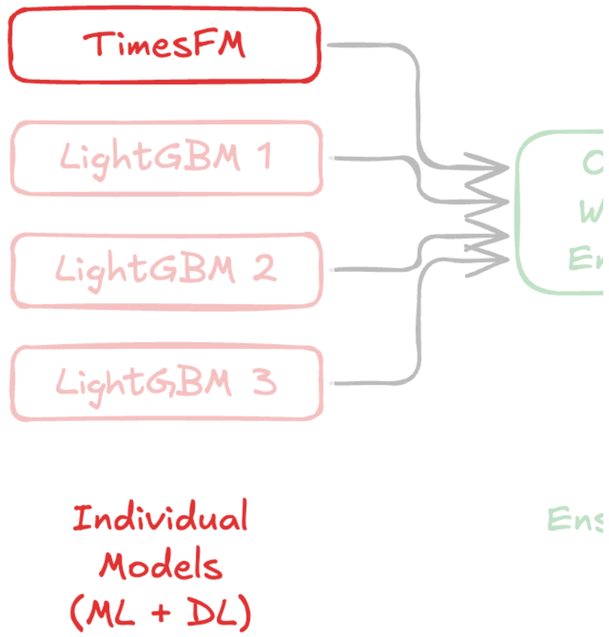
Individual Model (LGBM)



- Trained **LightGBM Models** with MLForecast
- **Recursive** Forecasting strategy
- **Min-Max Scaling** applied to target variable
- Three models: **L2 / L1 / Tweedie** loss functions
- **Feature Engineering**:
  - Date: Month, Fourier Features
  - Target: Lags, Rolling Min/Max/Mean/Std/Quantiles

A pretrained foundation model was included

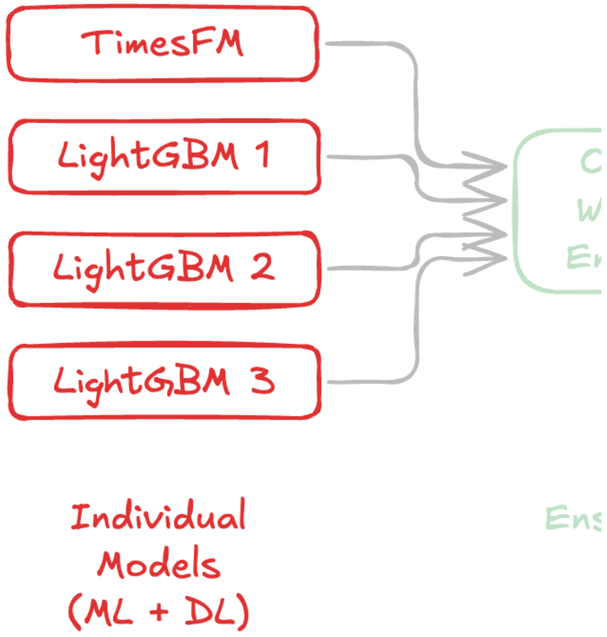
Individual Model (TimesFM)



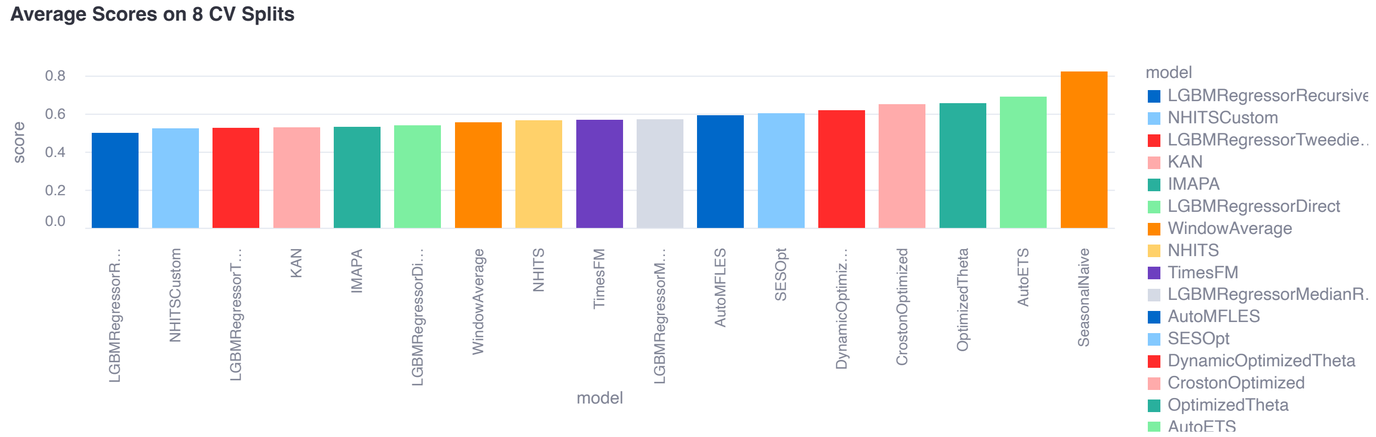
- **TimesFM** is a foundation model for time series developed by Google
- The checkpoint **google/timesfm-1.0-200m-pytorch** was used to generate the predictions

The models were chosen from a large collection

Individual Model

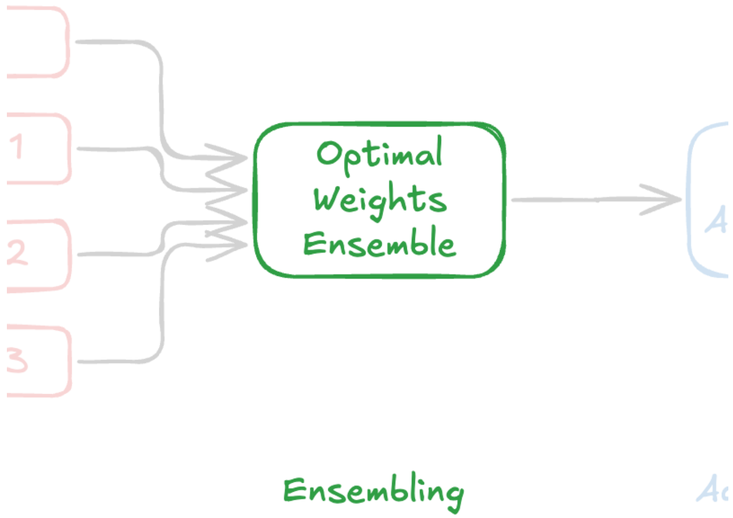


- Many models from deep learning, statistics, and machine learning were tested, but not all were included in the final ensemble
- The final models were selected for their performance and stability across multiple cross-validation splits



The models are combined by optimizing weights

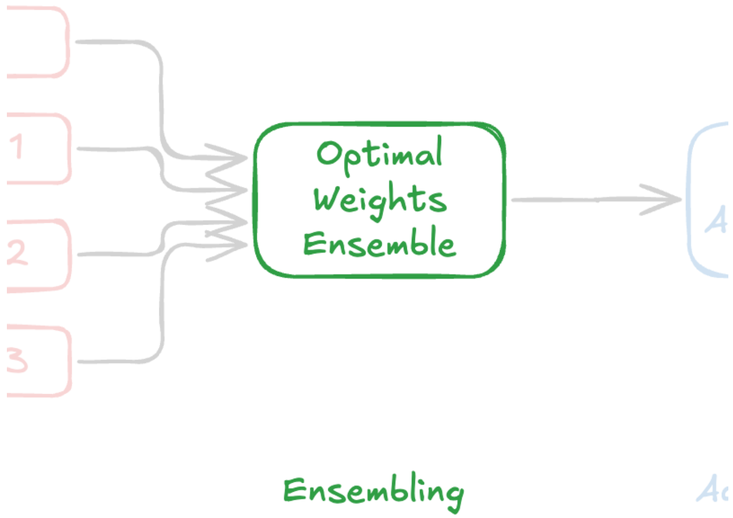
Optional Weights Ensemble



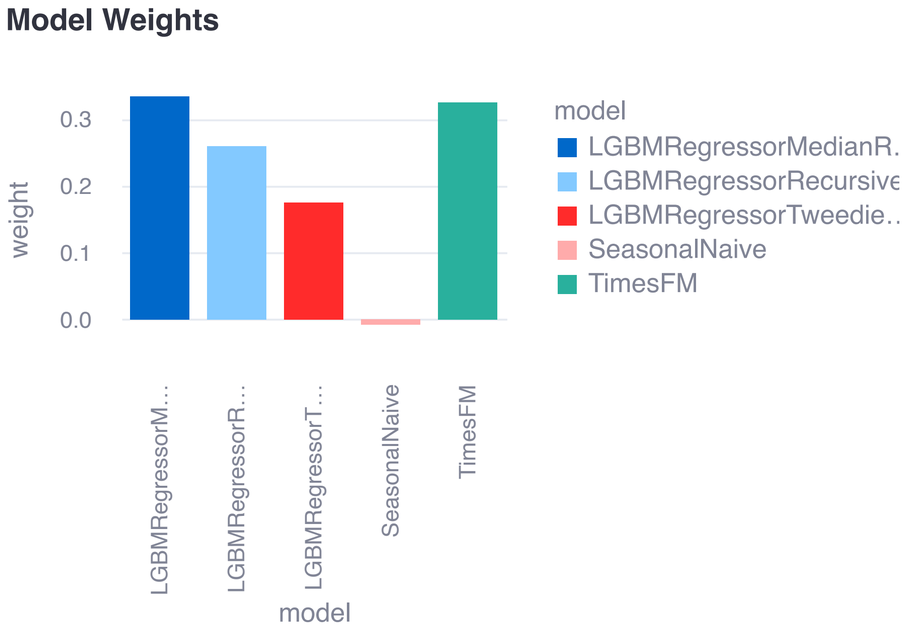
- **Goal:** Combine predictions of different individual models to achieve the most accurate overall forecast.
- **Weighted Models:** Assign a weight to each model's prediction, then combine them to create a single forecast.
- **Minimizing Error:** Choose the weights by minimizing the toal total error (0.5 * Bias + 0.5 * MAE).

The top model isn't always the most weighted

Optional Weights Ensemble

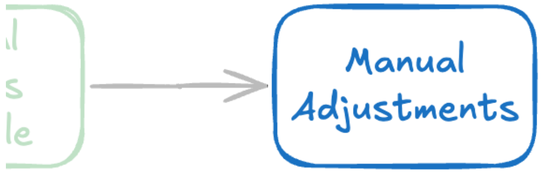


- The model with the highest score is not necessarily the model with the strongest individual performance in the backtesting



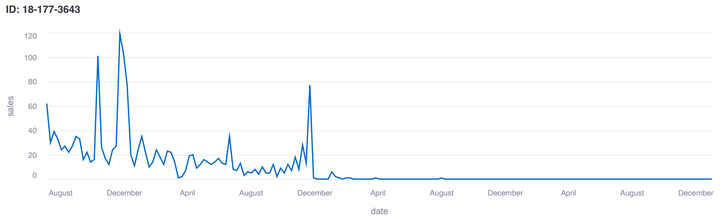
The forecast is adjusted manually afterwards

Manual Adjustments

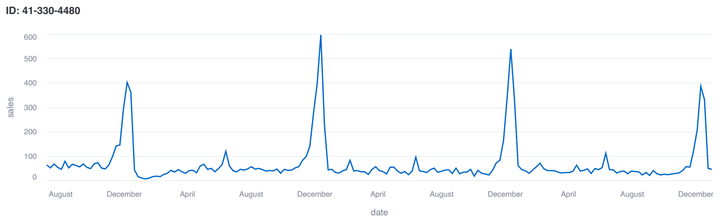


ng Adjustments

- **Zeros:** Time series with no sales in the last 13 days are forecasted with constant zeros

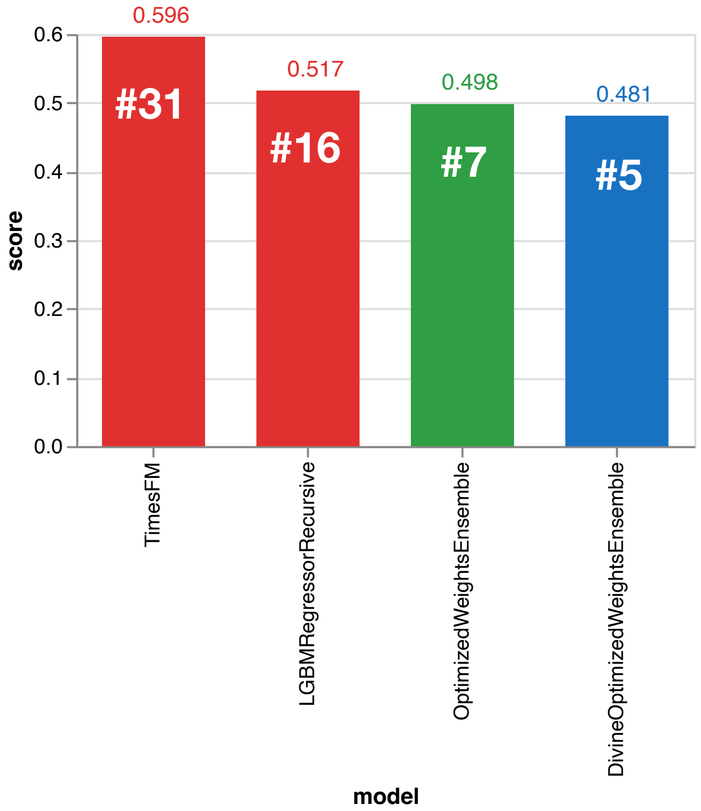


- **Seasonal:** Time series with a high seasonal ACF score are forecasted using a seasonal naive model



Each modeling step improved the final model

Final Result



- The **individual models** alone wouldn't have been able to get into the top five
- Surprisingly, **foundation models** did improve the final ensembling, even if they alone perform worse than simple benchmark models
- **Manual intervention** can proactively prevent the model from making significant mistakes and improve the score

Thank you for your attention 🙏