ID: 019

Long-Term Forecasting Despite Data Shortages

MARIA A. ZULUAGA, LISONG GUO, EOIN THOMAS, RODRIGO ACUNA-AGOST
Innovation and Research Division
Amadeus SAS

Forecasting plays a critical role within the travel industry with many applications which require accurate forecasting of future behavior by inferring from observations of the past. A common requirement is these systems is to make accurate long-term forecasts based on a stochastic model of the data at hand, which are often limited. Such scenario is challenging for any prediction algorithm. In this study, we benchmark different methods -from classic statistical approaches to state-of-the-art Long Short-Term memory (LSTM) networks- in the task of long-term price prediction with limited training data. Our ultimate goal is to establish the scenario that best suits each method and to determine empirical limits on the training data requirements and the forecast horizon of each method.

Research Questions

1. Which is the largest forecast horizon of each method in order to achieve an error < 20%?
2. Which is the minimum required sample size for each method?
3. Which method performs best?

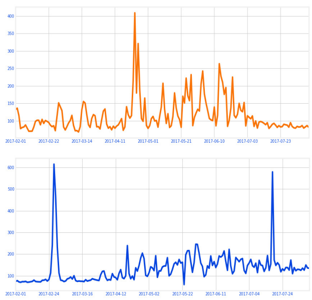
Materials & Methods

Materials: Data

Daily price data collection for 230 *travel products*.
Data presents missing values due to:

- Failure in data collection
- Product unavailability

Min, sample size: 73 days
Max. sample size: 180 days



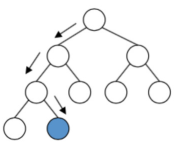
Methods: Algorithms

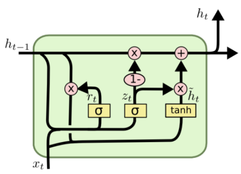
ARIMA. A separate model is created for each product using stationary data. Model selection is done via AIC.

$$\underbrace{(p,d,q)}\ \underbrace{(P,D,Q)_m}$$

Prophet[1]. Each product has an associated model. No data imputation is performed for missing. Holiday-related data is used, when available.

Gradient Boosting[2] uses a single model for regression and classification. Five extra features are used to differentiate products, whereas further 19 are derived from time information.





A stateful **LSTM** is trained for each product using dropout, 16 neurons and a look back window.

Methods: Evaluation Metrics

For regression:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t-\hat{y}_t}{y_t}\right|$$

For classification:

$$BER = 1 - 0.5\left(\frac{TP}{TP+FN}+\frac{TN}{TN+FP}\right)$$

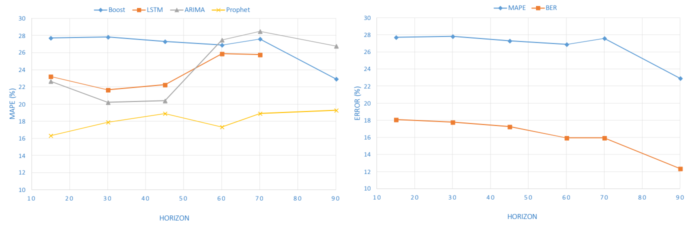
References
[1] Taylor, S.J & Letham, B. Forecasting at Scale. e-pub (2017)
[2] Friedman, J. Stochastic Gradient boosting. Comput. Stat. Data Anal. 38 (2002)

Experiments & Results

The table below summarizes the evaluated training set sizes, forecasting horizons and the number of products that could be used.

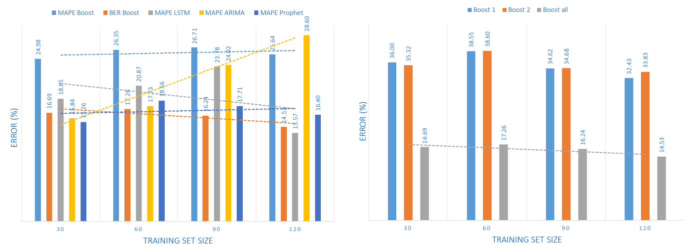
Training set	Horizon (min – max)	No. of cases (max – min)
30	15 - 30	228
60	15 - 60	228 - 201
90	15 - 70	201 - 95
120	15 - 60	157 - 92

Experiment 1: Forecast horizon



Left. Performance at different forecasting horizons. *Right.* Horizon performance in classification and regression (training set: 90 days).

Experiment 2: Training set

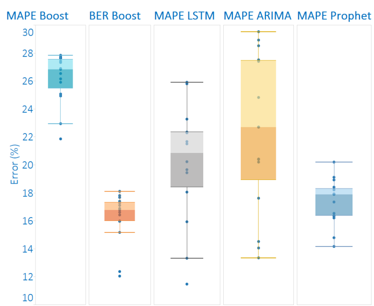


Left. Mean error vs training set size . *Right.* Effect of extra features.

Experiment 3: Overall performance

Prophet and the gradient boosting classifier report an error rate < 20% for all the evaluated cases.

Gradient boosting classifier shows the smaller variance.



Conclusions

- Price range forecast is a robust alternative to overcome high errors due to data shortage.
- Extra features aid performance of models with small training sets.
- Classical methods[1] can perform as good as state-of-the-art techniques.