Facebook’s Prophet

Applying prophet’s model for successful forecasting of avocado sales

Authors:
Tomás Costa - 89016
João Silva - 88813

Paper analysis

The initial step was analysing the given paper, as if provides a use case of the prophet algorithm, applied to a dataset of retail sales of a store in Bosnia and Herzegovina.

- Time-series sales forecasting
- Performance metrics
- Classify products according to level of forecastability

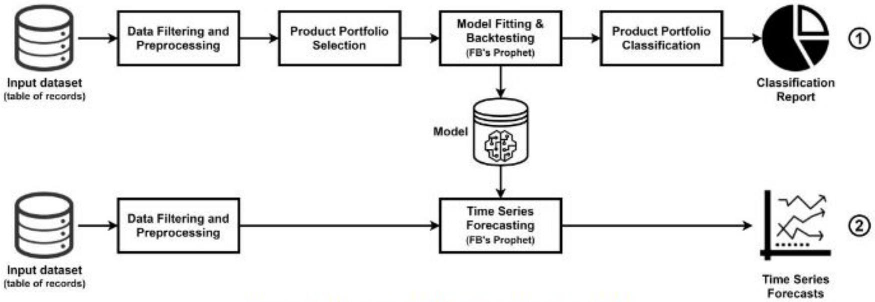


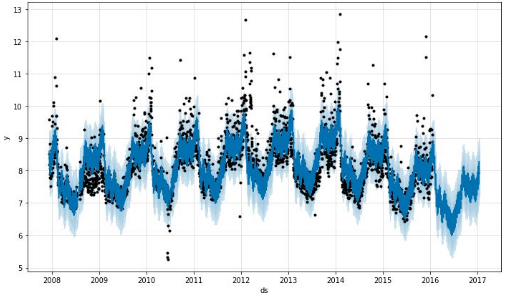
Figure 1. Proposed sales forecasting model.

Prophet Tool - Time-series forecasting

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

Benefits to using prophet:

- Accurate and fast
- Fully automatic
- Tunable forecasts
- Simplicity of use



The datasets

Together with the project, three distinct datasets were provided: Car sales, retail store sales and the avocado dataset. The authors tested the prophet tool with all datasets and they all provided really interesting but expect predictions.
For the purpose of the work, the dataset chosen for the notebook was the avocado dataset, since if provided the most information and seemed like the most complete dataset.

These were the most useful variables in the dataset for our problem:

- Date - The date of the observation
- AveragePrice - the average price of a single avocado
- Total Volume - Total number of avocados sold

Pre-processing

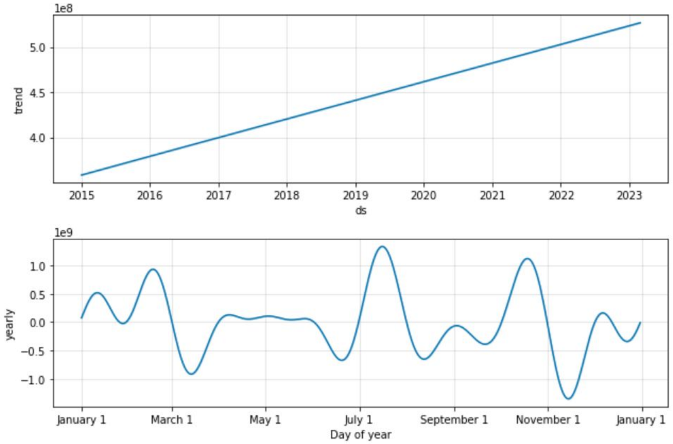
Depending on what we try to predict, the dataset we train the model with has to be prepared accordingly.

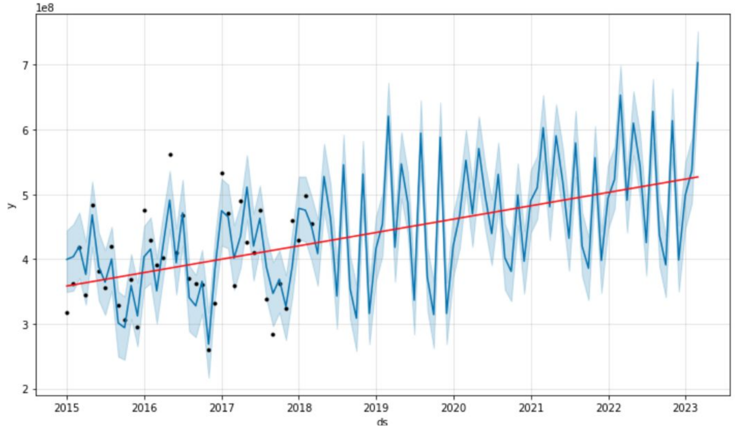
Since the product in question is a fruit, we found that the monthly sales with be a good parameter. To achieve this with our dataset we needed to:

- Aggregate data of the same month, summing the number of units sold
- Change the name of columns **‘Date’** and ‘**Total volume**’ to **‘ds’** and **‘y’**

Forecasts

Obviously that on a real life scenario, when trying to forecast the sales of a product, we should aim for reduced time periods like 3 months. In this case, for a better understanding of the tool, we tried to predict the number of sales for the next 5 years.





Testing and Evaluation

We divided the data into train and test selecting the last 3 month of our dataset for testing and the rest for training.

For the main evaluation parameter we used the mean absolute percentage error (MAPE).

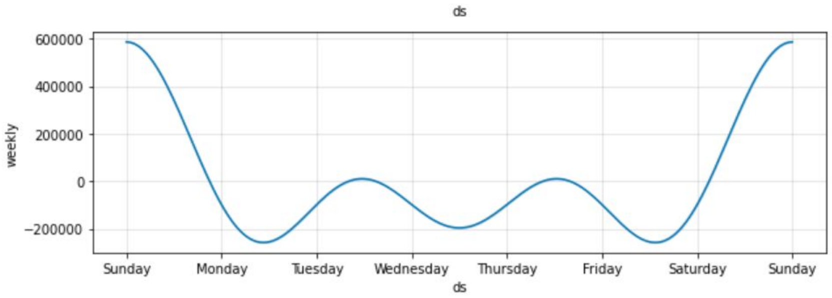
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y^i{}_{forecast} - y^i{}_{true}}{y^i{}_{true}}\right| \cdot 100\%$$

MAPE is used to quantify the overall accuracy of the forecasting framework and calculate the expected level of reliability. Approximately 50% of the products (with a sufficiently long historical data) can be forecasted with MAPE < 30% on a monthly basis.

The MAPE value of our testing was **7.6%.**

Other analysis

Since in our dataset we actually had the information about the sails per day but never got to use them, we decided to train the model with the data organized per day to try and obtain different results.



This time, due to the way the model was trained, we can obtain data relative to the distribution of sales throughout the week.

References

- Prophet Paper - https://peerj.com/preprints/3190/
- Prophet Docs - https://facebook.github.io/prophet/
- Time-series analysis Paper - https://arxiv.org/ftp/arxiv/papers/2005/2005.07575.pdf
- Avocado Prices - https://www.kaggle.com/neuromusic/avocado-prices