

MARKETING ANALYSIS OF THE GLOBAL MARKET USING MACHINE LEARNING

E.I. Gubin, PhD, Associate Professor
I.B. Soliev, student gr. 8PM9I
Tomsk Polytechnic University
E-mail: ibs2@tpu.ru

Introduction

Forecasting is a common data science task that helps organizations plan capacity, set goals and detect anomalies. Despite its importance, there are serious challenges in making reliable and high-quality forecasts, especially when there are many time series and analysts with experience in time series modeling are relatively rare. This research paper suggests a method for forecasting the total profit of global market, which based on concepts from machine learning that differ from the traditional stock prices and gold prices forecasting methods.

Research methods

The main workflows were focused on selecting an appropriate machine learning algorithm that could predict the future based on historical data. The chosen machine learning algorithm is called Prophet. The prophet is an open-source package for predicting time series data based on an additive model in which nonlinear trends correspond to annual, weekly, and daily seasonality. This works best with time series that have strong seasonal effects and multiple seasons of historical data. The prophet is resilient to missing data and trend changes and usually handles outliers well. The prophet is open-source software released by the core Data Science team at Facebook. And the main approach to forecasting will be implemented by the time series method. The time series method predicts the simple assumption that the future is a function of the past. In other words, they look at what happened over a period of time and use a series of previous data to predict. Technologies used in workflows: Oracle DB and PostgreSQL for cleaning and preparing a dataset, Python for machine learning.

Results

The prophet machine learning algorithm is most appropriate algorithm that can figure it out the business forecasting challenges, typically have any of the following characteristics:

- Hourly, daily or weekly observations with a history of at least several months (preferably a year).
- Strong multiple "human" seasonality: day of the week and season.
- Reasonable number of missing observations or large outliers.
- Historical trend changes, for example due to product launches or logging of changes.

The prophet's procedure is [an additive regression model](#) with four main components:

- Piece wise linear or logistic trend of the growth curve.
- Prophet automatically detects changes in trends by selecting points of change from the data.
- Annual seasonal component modeled using Fourier series.
- Weekly seasonal component using dummy variables.

The basic equation of the prophet's algorithm shown in formula (1):

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t) \quad (1)$$

According to the formula (1), $g(t)$ is the trend function which models non-periodic changes in the value of the time series, $s(t)$ represents periodic changes (e.g., weekly and yearly seasonality), and $h(t)$ represents the effects of holidays which occur on potentially irregular schedules over one or more days. The error term $\epsilon(t)$ represents any idiosyncratic changes, which are not accommodated by the model. Estimation of the parameters of the fitted prophet's algorithm is performed using the principles of Bayesian statistics. For this, the [Stan probabilistic programming](#) platform is used. The package is nothing more than a convenient interface for working with this platform from the Python or R environment. Correspondence of the global market dataset first of all it is worth showing imagination the basic features of dataset using Python. Fig. 1 present the basic features of global market dataset using Python.

	REGION	COUNTRY	ITEM_TYPE	SALES_CHANNEL	QUALITY_PRODUCT	ORDER_DATE	ORDER_ID	SHIP_DATE	UNITS_SOLD	UNIT_PRICE	UNIT_COST
0	Middle East and North Africa	Morocco	Vegetables	Online	M	29.01.16	371943700	12.03.16	5261	154.06	90.93
1	Middle East and North Africa	Somalia	Furniture	Online	H	07.10.12	366155707	23.10.12	2465	668.27	502.54
2	Europe	Azerbaijan	Cosmetics	Offline	L	11.06.15	714965188	26.06.15	2942	437.20	263.33
3	Europe	Moldova	Clothes	Offline	H	21.05.13	290295970	25.05.13	3883	109.28	35.84
4	Europe	Italy	Optical instruments	Online	C	02.02.12	284384381	10.02.12	6513	152.58	97.44

Fig. 7. Imagination basic features of global market dataset

Understanding workflow of prophet's algorithm was important to realize definition, mathematical aspects, and working process of algorithm because that's playing a principal role to understand advantages and disadvantages of prophet's algorithm. The following stage is forecasting procedure apply of machine learning algorithm. Suppose we need to make a forecast of the total profits by sales channel for the next 3 years that worth define total profit by sales channel. Matching with the prophet's algorithm will sort out that problem a much better way. Fig. 2 present forecasting the outcome of the prophet algorithm:

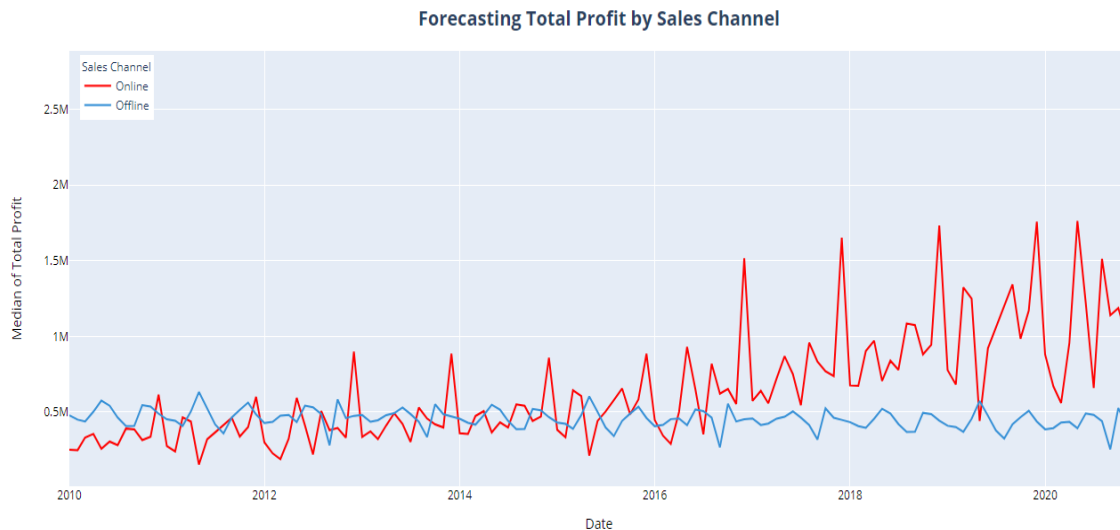


Fig. 8. Forecasting the outcome of the prophet algorithm

Conclusion

In this paper, we used techniques from machine learning to illustrate how a data analyst or trader can forecast total profit. In our simplified forecasting procedure, we subdivided the major feature global market dataset the total profit into two categories of sales channels: Online, offline. The major objective of the work was performing the corresponding action to identify which hypothesis total profit through a sales channel going to grow or reduce in the next 3 years. And outcomes present the hypothesis total profit through online sales channel will grow, however hypothesis total profit through offline sales channel will remain as it was before or reduce.

References

1. Dykstra, R. L. (1981). An isotonic regression algorithm, *Journal of Statistical Planning and Inference*.
2. Banter J., & Booms T. (2000). Marketing of services, *Journal of Marketing*.
3. Daniel Gerona (2019). *Machine Learning with Python*.