UEF
UNIVERSITY OF ECONOMICS & FINANCE

QS STARS RATING SYSTEM

HUTECH
Đại học Công nghệ Tp.HCM

CỤC CÔNG TÁC PHÍA NAM

THE SECOND INTERNATIONAL CONFERENCE ON SCIENTIFIC, ECONOMIC AND SOCIAL ISSUES

DIGITAL TRANSFORMATION, COOPERATION AND GLOBAL INTEGRATION IN THE NEW NORMAL

FINANCIAL PUBLISHING HOUSE

TABLE OF CONTENT

APPLICATION OF TECHNOLOGY AND BIG DATA IN THE FIELDS OF FINANCE, ACCOUNTING AND AUDITING IN THE CONTEXT OF GLOBALIZATION

DIGITAL ECONOMY IN VIETNAM, TRENDS AND POTENTIABILITY

DIGITAL TRANSFORMATION – COOPERATION – GLOBAL INTEGRATION IN BUSINESS

RECOVERY COMMUNICATIONS IN THE TOURISM AND HOSPITALITY INDUSTRY AFTER THE COVID-19 PANDEMIC

DIGITAL VERSUS NON- DIGITAL

CHALLENGES FACED BY TEACHERS IN NON-TRADITIONAL EDUCATION

DIGITAL TRANSFORMATION
AN INDISPENSABLE EVOLUTION FOR SUSTAINABLE CORPORATES

INFORMATION TECHNOLOGY AND APPLICATIONS

TRENDS AND ISSUES IN ENGLISH LANGUAGE EDUCATION AND RESEARCH

LAW IN THE CONTEXT OF INTERNATIONAL INTEGRATION

FB-PROPHET MODEL FOR TIME SERIES FORECASTING IN SALES

Thanh Cong Tran

Ho Chi Minh City University of Economics and Finance
congtt@uef.edu.vn

Abstract

Forecasting techniques based on time series analysis are applied in various fields such as banking, sales, stock market, and healthcare. Forecasting chain of ten stores in sales using time series can improve sales in the business environments. Recently, there are many tools to use for forecasting, but the Facebook Prophet (FB) tool is the latest and has demonstrated its effectiveness in term of accuracy. This paper proposes the FB Prophet tool to forecast the data of chain of ten stores. Based on the experiment, it is concluded that FB Prophet is a highly accurate predictive model and shows positive outcomes.

Keywords: *FB-Prophet, Prediction, Time series.*

1. Introduction

Time series data analysis is helpful to achieve worthwhile statistics and many other attributes of data in an overall business environment. Time series forecasting models play a crucial role in forecasting models, where time is considered as a key factor. These forecasting models have a great influence on predicting sales in the future and managing the operation of business. It is also vital since various forecasts involve time of many components which should be addressed with care in order to make predictions when the actual outcome is unidentified. In order to identify the primary element of an event in time series, it is necessary to clearly understand the data pattern involved to time. There are four main components in time series analysis of data, namely level, trend, seasonality, and noise. The level component is considered the fundamental value employed in time series data, the seasonality component is represented as a curve that can rise or decline over the period of time. Moreover, seasonality is expressed as a cycle or pattern over the period of time and noise reveals variation in data observed.

Recently, an open source, named forecasting tool FB prophet which is popular for use in the library of python and R programming languages was developed by Facebook. From the perspective of business, FB prophet was generated to meet forecasting needs in sales. FB prophet has attributes related to time series data which is observed hourly, daily and monthly from year to year. FB prophet also considers break intervals or holidays which are known in advance. Additionally, FB prophet takes into account trends, outlier detection, and missing data.

Furthermore, Prophet works on a regression model with trends related to a regression curve or a modular curve for an increasing trend, a seasonal component based on Fourier series, a seasonal component weekly.

FB prophet is relied on the technique of curve fitting in Bayesian model. It is easy to understand parameters and it also does not request a lot of time series data for forecasting. This technique is best suited when seasonal properties are indicated as strong influencing factors in time series data. FB prophet handles better in case of missing data, variating trends and detecting outliers. In real-world situations, such as sales

predictions, such variations must be tackled. In this paper, we perform the analysis of univariate time series for chain store sales data employing FB Prophet tool.

2. Related works

Many studies have been done to forecast time series employing FB prophet in different areas. In [2], the forecasting framework is used to predict future demand for product categories. The developed method also indicates the time series curve for each product in the product category. In [3], an adaptive Kalman filter is employed together with FB Prophet so as to forecast maximum power demand with improved predictability. In [4], SARIMA and FB oracle algorithms are employed to forecast power grid failures. ARIMA model and its steps are discovered in this proposed model.

Shikha Gaur et al. [5] proposes a prediction model based on prophet ARIMA and FB to predict the trend of COVID-19 infections. This proposed model helps identify outbreaks and irregularities of COVID-19 in India and abroad. Nevertheless, the research is constraint on the healthcare sector and could not be used for the areas of sales and e-commerce. In [6], Liyun Su et al. uses the polynomial function to approximate coefficients in autoregressive prediction. These models can perform better based on the technique of FB prophet.

A number of research relied on neural networks has been implemented in the field of time series forecasting between 2006 and 2016. In [7], these methods related to neural network are reviewed by Ahmed Tealab. In [8], CemKocaket et al. suggests a fuzzy autoregressive moving average (ARMA). In [9], forecasting models ARIMA and FB Prophet are employed to forecast closing prices of stock markets. This method works with the dataset which is separated into two phases, namely training and validation set which depend on daily, weekly and monthly transactions. The results show that ARMA-based model only works efficiently in short-term predictions like daily or weekly while prophet FB gives good results for long-term stock predictions like monthly or yearly. Research work is primarily limited to predict the prices of stock, and these models are implemented to sales prediction. In [10], a similar predicting approach is developed to predict stock prices, but it should be improved the overall performance of the predictive model. Time series analysis plays a key factor to predict health perception and prevention-involved information to the society.

A similar study was performed for the sectors of health and air pollution [11, 12, 13] to predict pollution harm and mortality rates in different nations. According to Poisson regression and GAM, S-Plus' generalized additive modeling (GAM) software is tested [11] to analyze time series data on the number of air pollution deaths. GAM is considered as expansion of a linear model based on a non-parametric function. FB-based method can make an improvement to predict accuracy in time series analysis of data.

Merchandisers and entrepreneurs employ tools involving demand forecasting to plan the operations of business. Various tools are available to perform such forecasting either model-based or model-free. Among them, the model-based method is more powerful. Cyclic Boosting supervised learning which is an effective approach is one of model-based approaches [14]. The primary concept of the proposed method is to take advantage of the benefaction of each attribute to predict the objective function. The Cyclic Boosting

supervised learning method is also based on the kernel regression model and uses normalization technique to refine the model while smoothing technique removes noise. However, FB prophet performs better in case of simplicity and gives more accurate results. In [15], an ARIMA model predicts environmental degradation. According to NLP and Fuzzy C Means, this method uses Microsoft Power BI to analyze data, display and customize the dashboard. The performance of this method is better than preceding approaches, but FB prophet is much more accurate and ease of use when predicting issues related to time series.

3. FB prophet model

Based on an additive regression model, FB prophet [1] prediction can be expressed as (1):

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

Where: y(t) is the additive regression model, g(t) is the trend factor, h(t) is the holiday component, s(t) is the seasonal component, and e(t) is the error term. The trend factor g(t) can be modeled as Logistic Growth Model and Piece-wise linear model. Logistic Growth Model which is expressed as (2) indicates growth in different stages. The growth in the first stage grows exponentially, then reaches a period of saturation, from which it is linear growth.

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} \quad (2)$$

Where: L depicts the maximum value of the model curve; k is denoted as the growth rate; x_0 is expressed as the value of x at sigmoid point.

Additionally, Piece-wise linear model shown as (3) is customize based on the linear model where x shows various range with distinguished linear relationships.

$$y = \beta_0 + \beta_1 x + \beta_2 (x - c)^+ + \varepsilon \quad (3)$$

The method employed for time series forecasting based on FB Prophet is depicted in Figure 1.

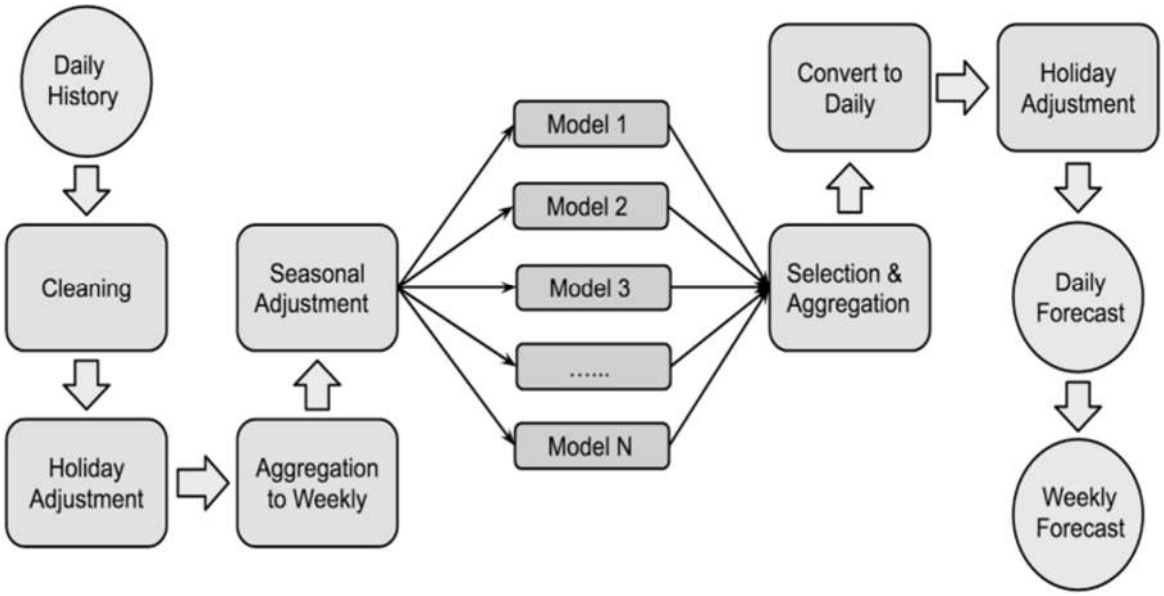

Figure 1. Diagram of time series forecast

4. Data description and evaluation metrics

4.1. Data description

The store chain dataset of ten stores is used in this paper [16]. This dataset was collected over a 5-year period from 2013 to 2017. The data field of the dataset includes: date, store, item, and sales. Figure 2 reveals the total daily sales of all stores. Figure 3 shows the total daily sales for each store. Total daily sales seems to be trending fairly steadily and gives an indication that the FB prophet model may be suitable for predicting store-level sales.

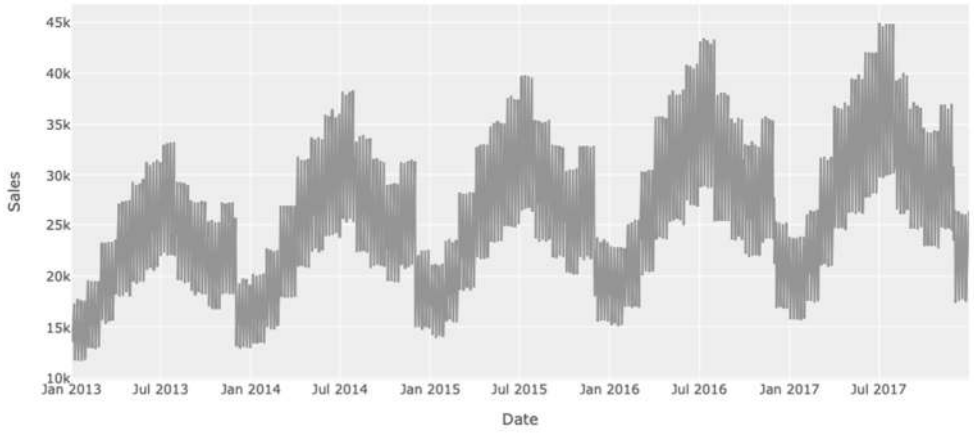

Figure 2. Total daily sales of all stores

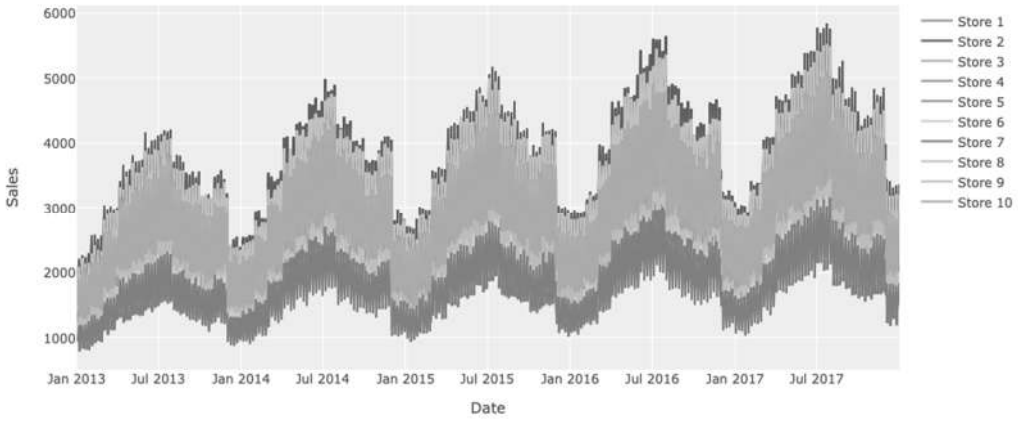

Figure 3. Total sales by day by each store

4.2. Evaluation metrics

In this paper, the evaluation metric to consider the accuracy in FB prophet models is the weighted mean absolute percentage error (WMAPE) [17]. Based on opinions of many business professionals, financial data analysts, and accountants, WMAPE metric has high reliability to evaluate the accuracy of the forecasting models.

WMAPE is a way to measure the accuracy of financial and statistical projections relative to actual or actual results for a sample. The different parts of WMAPE are weight, mean, absolute, percentage, and error. Weighted means that there is a component against the measurement results of the computation. Average means that the result of this calculation is an average of the accuracy of the predictions. Absolute means that regardless of whether the actual result is more or less than predicted, the computation will

provide a positive number. Percentage means the computation results in percentage format for ease of use. The error means that the result of the computation is a measure of the difference between the forecast and the actual result.

Typically, WMAPE is used to compare forecasts over a long period of time, which reveals a general trend as to whether forecast is accurate rather than a very specific date or time. WMAPE can be used in conjunction with other predictive measurements to better understand the accuracy and responsiveness of data models. The formula for WMAPE is expressed as (4):

$$\text{WMAPE} = \frac{1}{n} * \frac{\sum(|Actual - Forecasted|)}{Actual} * 100 \quad (4)$$

Where: *n* is the size of the sample; *Actual* is the actual value for a certain period of time, *Forecasted* is the expected value over a certain period of time.

5. Results

The results of the store sales prediction of 10 stores are shown in Figure 4. The blue line is the forecast value while the black dots are the actual value. We can see that the blue line matches the overall trend of the data with some outliers still present, which is a pretty good fit. In addition, we now need to test how successful the model is on data it hasn't seen yet, specifically 2018.

When forecasting trends in 2018, the WMAPE metric is used to measure the accuracy of the forecast trend in 2018. Table 1 describes the results of the WMAPE metric for the sales forecasting model of each store, from store 1 to store 10. Thereby, we see that the prediction results based on WMAPE metric are quite high, over 94% in all predictive models of stores.

Table 1. Results of WMAPE method for store-by-store forecasting model

Metric	Store 1	Store 2	Store 3	Store 4	Store 5	Store 6	Store 7	Store 8	Store 9	Store 10
WMAPE (%)	94.456	94.50	95.143	94.492	94.347	94.867	94.516	94.870	94.828	94.767

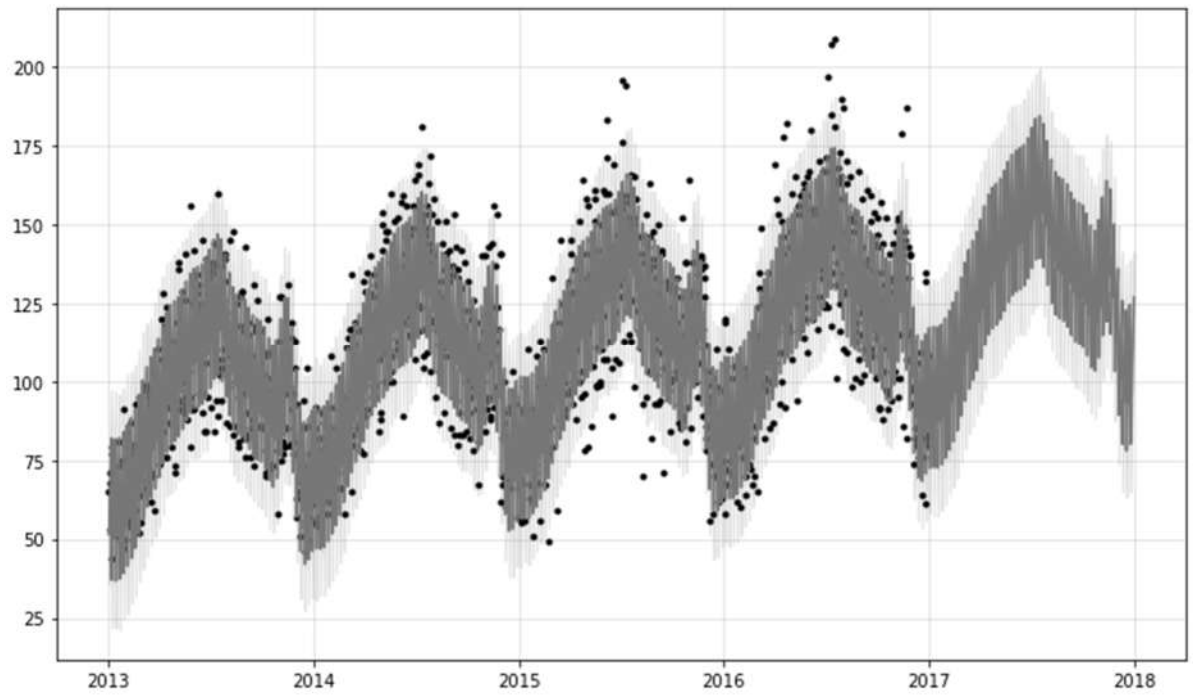

Figure 4. Forecasting results of 10 stores in sales

6. Conclusion

In the proposed research work, the FB Prophet model is implemented for sales forecasting. We utilize the store chain dataset of ten stores collected between 2013 and 2017 for testing. We have achieved that the prediction implemented using FB Prophet is extremely adjacent to reality. The proposed tool is indicating great performance in forecasting of time series data with minor error accuracy. Nevertheless, scalability is another challenge for the analysis of a large dataset. Transformation learning can be used together with FB Prophet in order to enhance scalability and handling of large datasets. Real-time forecasting with high accuracy relies on the model employed for training and validation.

References

[1] B. Kumar Jha and S. Pande, "Time Series Forecasting Model for Supermarket Sales using FB-Prophet," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 547-554, doi: 10.1109/ICCMC51019.2021.9418033.

[2] Emir Zunic, Kemal Korjenic, Kerim Hodzic, Dzenana Donko, “Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data”, International Journal of Computer Science and Information TechnologyInternational Journal of Computer Science & InformationTechnology (IJCSIT), Vol 12, No 2, PP. 23-36, April 2020.

[3] Chen Guo, Quanbo Ge, HaoyuJiang,Gang Yao and Qiang Hua, “MaximumPower DemandPrediction UsingFbprophetwith Adaptive Kalman Filtering”,IEEE Access, 2020.

[4] Yi Yan, Bo Li, Jianyi Xiao, Yunde Liang, Yanwei Shang and Kaidong Zhou, “Comparative Study on Prediction Algorithms for Power GridSystem Access Failure Times”, IOP Conf. Series: Earth and Environmental Science ,252, 2019032183 IOP Publishing

[5] Dr. Shikha Gaur, “Global Forecasting of COVID-19 Using ARIMA Based FB-Prophet”, International Journal of Engineering Applied Sciences and Technology, Vol. 5, Issue 2, ISSNNo. 2455-2143, 2020, pp. 463-467.

[6] LiyunSu, ChenlongLi,"LocalPredictionofChaoticTimeSeriesBased on Polynomial Coefficient Autoregressive Model", Mathematical Problems in Engineering, vol. 2015, Article ID 901807, 14 pages, 2015.

[7] Ahmed T ealab,"Time series forecasting using artificial neural networks methodologies: A systematic review", Future Computing and Informatics Journal,vol. 3, Issue 2, pp. 334-340, December 2018.

[8] CemKocak,"ANewHighOrderFuzzyARMATimeSeriesForecasting Method by Using Neural Networks to Define Fuzzy Relations", Mathematical Problems in Engineering, vol. 2015, Article ID 128097, 14 pages, 2015.

[9] WintNyein Chan, "Time Series Data Mining: Comparative Study of ARIMA and Prophet", Journal of Computer Applications and Research, Volume 1, No 1, 2020, pp. 75-80.

[10] A. A. Adebiyi, A. O. Adewumi, C. K. Ayo "Stock Price Prediction Using the ARIMA Model", UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp.106-112.

[11] Dominici F. "On the Use of Generalized Additive Models in Time- Series Studies of Air Pollution and Health. American Journal of Epidemiology".vol. 156, Issue 3, 2002, pp. 193-203

[12] Samet,J&Zeger,S&Dominici,Francesca&Curriero,F&Coursac, I & Dockery,Douglas & Schwartz, Joel &Zanobetti, Antonella.(2000).The National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity and mortality from air pollution in the United States. Research report (Health Effects Institute), vol. 94, discussion 71,pp. 5- 70, 2000.

[13] Schwartz, Joel & MARCUS, ALLAN, 'Mortality and air pollution in London: a time series analysis", American Journal of Epidemiology, vol. 131, pp. 185-194, 1990.

[14] Wick, F.&Kerzel, U.&Hahn, M.&Wolf,M.&Singhal,T.&Feindt, M., "DemandForecastingofindividual Probability Density Functions with Machine Learning", PP. 1-20, 2020.

[15] Anand, J.V."AMethodologyofAtmosphericDeteriorationForecasting and Evaluation through Data Mining and Business Intelligence." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 02(2020): 79-87.

[16] Kaggle, https://www.kaggle.com/c/demand-forecasting-kernels-only/overview.

[17] Indeed, https://www.indeed.com/career-advice/career-development/weighted-mean-absolute-percentage-error.

NOT FOR SALE

ISBN: 978-604-79-3782-0

HO CHI MINH CITY UNIVERSITY OF ECONOMICS AND FINANCE

141 - 145 Dien Bien Phu, Ward 15, Binh Thanh District, HCM City

Website: uef.edu.vn - Hotline: (028) 5422 6666 * (028) 5422 5555