Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)
IEEE Xplore Part Number: CFP21K25-ART

Time Series Forecasting Model for Supermarket Sales using FB-Prophet

Bineet Kumar Jha
Assistant Professor, Information Science and Engineering
CMR Institute of Technology
Bengaluru, India
yoursbineetjha@gmail.com

Shilpa Pande
Associate Professor, Information Science and Engineering
CMR Institute of Technology
Bengaluru, India
shilpa.p@cmrit.ac.in

2021 5th International Conference on Computing Methodologies and Communication (ICCMC) | 978-1-6654-0360-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCMC51019.2021.9418033

***Abstract*— Forecasting techniques are used in the various problem domains such as- sales, banking, healthcare, stock market, etc. The time-series dataset has time-related information that is useful for prediction and statistical analysis. The supermarket sales prediction helps improve sales in a business environment. The technique helps in decision making in a problem domain. Many tools are available for forecasting such as the regression model, Logistic exponential model. The Facebook (FB) Prophet is the latest tool that has shown an improved performance in terms of accuracy of prediction. This research work has proposed a FB Prophet tool for the sales prediction of the supermarket data. The proposed research work has examined few forecasting models such as- The additive model, the Autoregressive integrated moving average (ARIMA) model, FB Prophet model. From the propsoed research work, it is concluded that, FB Prophet is a better prediction model in terms of low error, better prediction, and better fitting.**

Keywords—additive model;ARIMA; Facebook prophet; time series forecasting

I. INTRODUCTION

Time series analysis of data is useful to get meaningful statistics and other properties of data in the business environment. The time series forecasting model has an important usage in the forecasting model where time plays an important role. This forecasting model has a great impact on determining future sales and managing businesses. It is also important because many predictions involve time-related components that need to be handled carefully to do the prediction when the actual result is unknown. To determine the root cause of a certain event it is required to know the pattern of related data and their time. There are four major components-.

- Level- It is the base value used in the time-series data.
- Trend-It is shown as a curve that may increase or decrease depending on time.
- Seasonability- This is indicated as a cycle or pattern over time.
- Noise- It shows the variation in the observed data.

Facebook has introduced an open-source forecasting tool FB prophet available to use in python and R programming languages as a library.The FB prophet is developed to meet the forecasting need from the business point of view. It has the following characteristics-

- Time series data observed on an hourly, daily, and monthly basis for a year or more.
- It takes care of holiday or break intervals that are known in advance.
- It takes care of trends, outlier detection, missing data, etc.

At its core, the Prophet works on an additive regression model (research.fb.com) with the following trends-

- Modular regressive or liner curve for the growing trend.
- A Fourier series based seasonal component
- A seasonal component every week.
- The user suggested a list of break intervals or holidays

Fig. 1 shows the basic Prophet workflow (Forecasting at Scale, Sean J. Taylor and Benjamin Letham, 2017), it has the sweetest part where the surface problem is being automated and analysts have to inspect the forecasts.

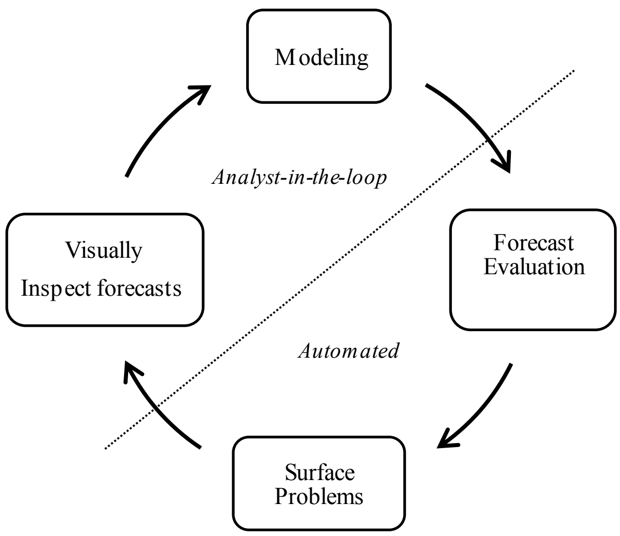


Fig. 1: Prophet workflow

The main objective of our research work is to do the univariate time series analysis of the supermarket sales data using the FB Prophet tool. The collected data is analyzed by fitting them on the Additive model and Auto-Regressive Integrated Moving Average (ARIMA) model which has shown improved future forecasting.

The FB Prophet is based on the curve fitting technique in the Bayesian model. It has easily understandable parameters and also it doesn't require much time-series data to do prediction. The technique is most suitable when the time-series data has strong seasonal attributes as influencing factors. It also takes care of planned breaks or holidays in the continuous data. FB Prophet deals better in case of missing data, variation in trends, and outlier detection. In real-world scenarios such as sales prediction, such variations need to be addressed. It also has easily usable and interpretable libraries

II. RELATED WORK

Many research works are carried out for the time-series forecasting using the Facebook Prophet for various problem domains. A similar forecast framework is proposed by Emir Žunić*et al.* [1], in that the product portfolio is created based on future demand. The proposed approach also illustrates the time-series curve for the specific product in the product portfolio. In another research work [2] an adaptive Kalman filter is used along with FB Prophet to predict the maximum power demand with improved prediction. In other research work [3] SARIMA and FB Prophet Algorithms are used to predict the power grid failure. The proposed work also explores the ARIMA model and its steps.

Dr. Shikha Gaur *et al.* [4] have proposed an ARIMA and FB Prophet-based forecasting model for COVID-19 cases trend prediction. The proposed model helps detect COVID-19 outbreaks and irregularities in India and overseas. The performance of the proposed approach is compared with traditional linear regression-based forecasting. However, the research is limited to the healthcare domain which can be used for e-commerce and sales domains. Bryan Lima *et al.* [5] have proposed an attention-based DNN architecture-Temporal Fusion Transformer (TFT) which has shown better performance for electricity and traffic datasets. A fuzzy-based regression and time series model is proposed by ErolEgrioglu*et al.* [6]. Liyun Su *et al.* [7] used a polynomial function for the coefficient approximation in autoregressive prediction. The performance of these approaches can be improved using the FB Prophet technique.

A lot of researches have been done in the area of time-series forecasting using neural networks for a period from 2006 to 2016. A study has been by Ahmed Tealab [8] to examine those methods. CemKocak*et al.* [9] have proposed fuzzy autoregressive moving averages (ARMA). In another research word done by WintNyein [10], ARIMA and FB Prophet forecasting model is used to determine the closing price of the stock exchange. The proposed approach works the dataset is divided into two groups- training and validation sets based on daily, weekly, and monthly transactions. The research finding indicates that the ARMA-based model performs well only in short-term predictions such as daily or weekly whereas for the long-term stock prediction such as monthly or yearly FB-Prophet is much better. The research work has mainly been limited to predicting stock price and these models can be used for sales forecasting. A similar approach is used for stock price prediction [11]. The overall performance of the forecasting model needs to be improved. The time series analysis is important to forecast health awareness and precaution related information to the public.

A similar analysis is done for air pollution and health [12, 13, 14] to predict the harmful effect of pollution and mortality in various countries. The S-Plus generalized additive model (GAM) software is examined [12] for the time-series data analytics of the mortality count due to air pollution. The proposed approach is based on Poisson regression and GAM. The GAM is an extension of a linear model with a nonparametric function. The FB based approach can improve prediction accuracy in time-series data analysis.

Retailers use the demand forecasting tools for their operational planning. Many tools are available for doing such forecast are model-based or model free. The model-based approach is more robust in them. One such model-based approach is Cyclic Boosting supervised learning [16]. It is better than the model-free approach- Quantile Regression.The main idea behind the proposed approach is to utilize the contribution of each attribute for predicting the target function. It is also based on the multiplicative regressive mode. It also uses the regularization technique for the fine-tuning of the model whereas the smoothening technique for noise removal. The FB- Prophet is better in terms of simplicity and more accuracy. An ARIMA based model for the environmental deterioration forecasting[17]. The proposed approach uses Microsoft Power BI for data analysis. It is based on NLP and Fuzzy C Means for the dashboard display and customization. The performance of the approach is better as compared to previous techniques. The FB Prophet can be a better approach in terms of accuracy of prediction and ease to use.

For some time-series forecast models, a huge dataset needs to be analyzed which takes few days to analyze such as real-world traffic analysis, which uses huge traffic data in the form of videos. An artificial neural network (ANN) approach can be used for transfer learning to pre-train the voluminous data and apply it to a similar problem. A similar approach is presented in [18] which uses a transfer learning approach to use a pre-trained model to analyze traffic stream to forecast future traffic in a similar situation. A similar transfer learning-based approach is used [19] to analyze the spatiotemporal traffic data for traffic prediction. The Spatial Kernels are used for the image processing. It has shown better performance in terms of edge detection, sharpening, smoothening and other important image processing tasks. The graph convolution neural network

(GCNN) is another technique that can be used for video prediction.

In the given model c is the breakpoint; the two pieces of information are connected at $x = c$; $(x - c)^{+}$is the interaction terms represented as $(x_{i1} - c) * x_{i2}$

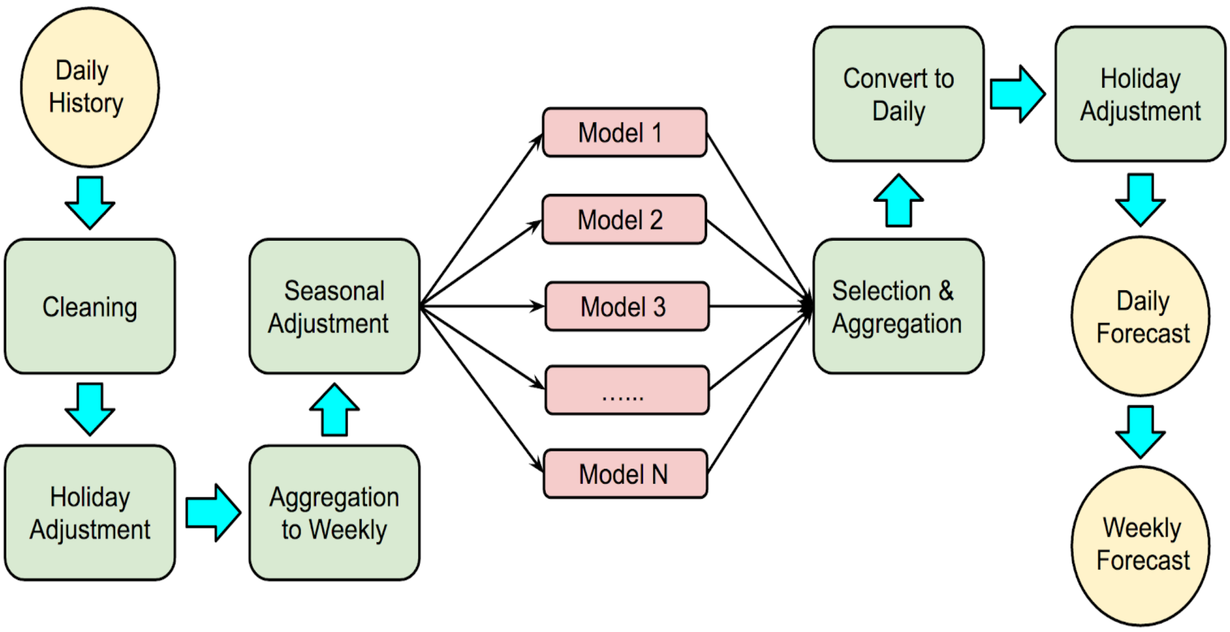


Fig. 2: Time series forecasting flow diagram

III. MODELS AND METHODOLOGY

The FB Prophet forecasting is based on an additive regressive model can be formulated as –

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

In (1), $y(t)$ is the additive regressive model; $g(t)$ is the trend factor; $h(t)$ is the holiday component; $s(t)$ is the seasonality component and et is the error term. The trend factor $g(t)$ can be modeled in two ways-

a) *Logistic growth model*-This model represents the growth in various stages in the first stage the growth is observed approximately exponential after that saturation stage reaches from there it grows linearly. The model can be formulated in (2).

$$f(x) = \frac{L}{1+ e^{-k(x-x0)}} \quad (2)$$

In this mathematical model, L represents the maximum value of the model curve; k is the growth rate; $x0$ is the value of x at the sigmoid point.

b) *Piece-wise linear model*: It is a modified version of the linear model in that the different ranges of x have different linear relationships. The model can be formulated as shown in (3)

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

The methodology used for the time-series forecasting using FB Prophet is described in Fig. 2(Source: *Our quest for robust time series forecasting at scale, Eric Tassone and FarzanRohani, 2017, forecast procedure in Google)*.

A. Data Collection

A sample supermarket sales dataset is collected from kaggle.com. It has a total of 9994 records of sales-related data with 20 attributes related to products (furniture) sold in the supermarket. The dataset used for the experiment is collected between the years 2014-2017. The proposed model is being trained on the dataset.

B. Data Preprocessing

This step cleans the data in terms of removing unwanted columns and including missing values. This step also involves organizing the dataset as per order data or sales. The order dates are set as indices to access and analyze them in the future precisely.

C. Exploratory Data Analysis

This step displays the collected data and predicted values in terms of graphs and matrices. To do the exploratory analysis of the collected data that is plotted towards monthly product sales for the year 2017 as shown in Fig. 3. Variations are observed in the curve as the product sales increases or decreases.

D. Performance Matrices

Three matrices are used to measure the performance accuracy of the proposed models.

TABLE I. FITTINGDATASET USINGARIMA MODEL

Model	Parameters					
	Coef	***Std error***	*z*	***P>*** |*z*|	***95% Confidance Interval [0.025***	***0.975]***
ar.L1	-0.0913	0.286	-0.319	0.750	-0.651	0.469
ma.L1	-0.9982	9.598	-0.104	0.917	-19.810	17.813
ar.S.L12	-0.5598	0.156	-3.594	0.000	-0.865	-0.254
Sigma2	2.79914e+03	1.4577e-05	1.05e-01	9.17e-01	-2.7026e+02	3.0083e+02

- Mean Square Error (MSE) for weekly, monthly, or yearly forecast.
- Root Mean Square Error (RMSE) for the forecast.s
- Mean Absolute Percentage Error (MAPE) for quantifying overall accuracy.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section describes the results obtained from different forecasting models and compared the results.

A. Additive Model

The additive model is used as the core computational component of the FB Prophet. Fig. 4 shows the modular growth curve obtained using the FB Prophet tool on the sales data. The tool automatically detects the trend based on the change points in the gathered data. The trend can be shown on a daily, weekly, monthly, or yearly basis.

B. ARIMA Model

In Python, the ARIMA model is created by calling a method *ARIMA(),* and passing three parameters *p(lag value for autoregression), q(difference order)* and *r(moving average model)*. The training data is provided as input to the model using *fit()* method. Once the model is ready the forecasting is done using *predict()* method. Table 1 shows the result obtained after

fitting the dataset into the training model. The *coef*column shows the weight of each feature and their impact, P>|z shows the importance of feature weight for the various approaches *Autoregressive (AR), Moving Average (MA)*, etc.

Fig. 5 shows the outliers for the ARIMA model. The standardized residual indicates the forecast errors in terms of variations in the curve. The Histogram plus shows kernel density estimation with a partition between [0,1] indicated as N(0,1). The NormalQ-Q is a scatter plot for two sets of quantiles, the graph indicates the standard residual that is the strength of the observed and predicted values. The quantile indicates the number of values above and below a certain limit in that the prediction accuracy is good.

The correlogram indicates the randomness among the dataset.Fig. 6 shows the one-step-ahead sales forecast for the input dataset. The curve indicates that the forecast accuracy is very close to the prediction.

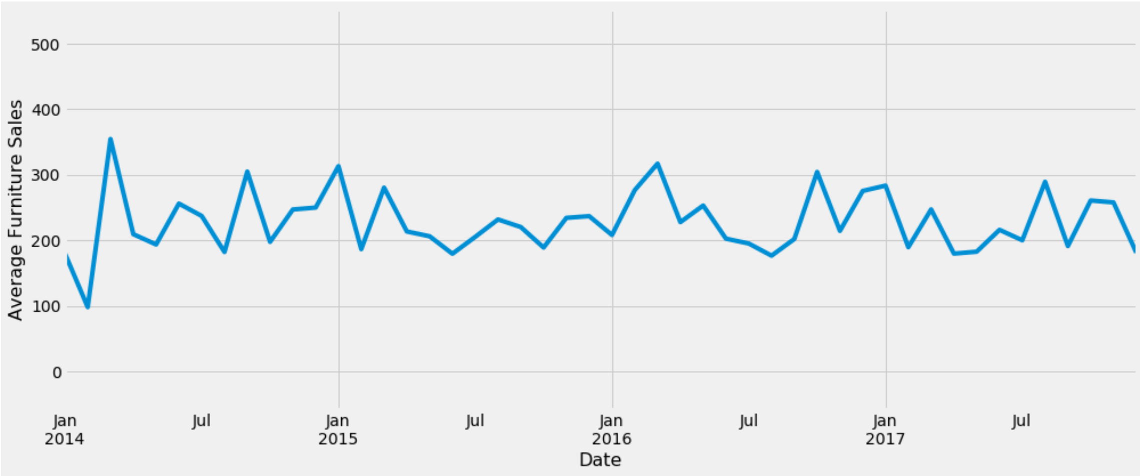


Fig. 3: Visualization of time series data

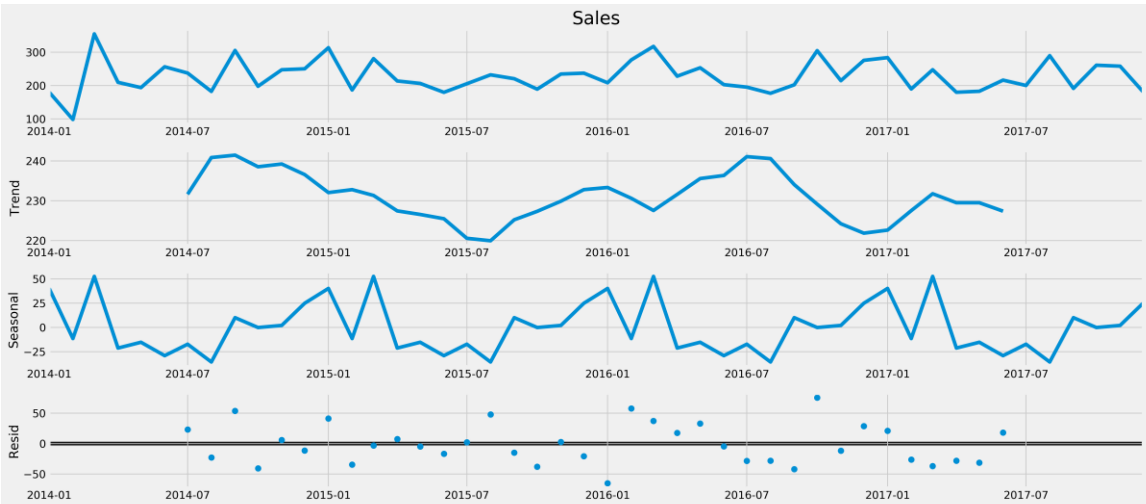


Fig. 4: Decomposition using an additive model

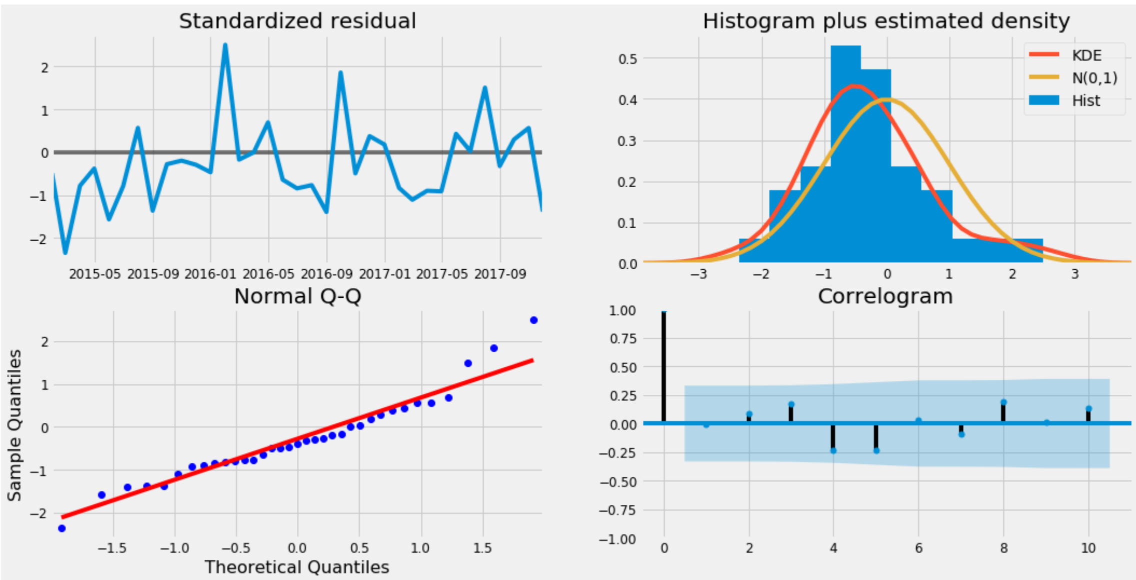


Fig. 5: ARIMA based diagnostic plot outliers' detection

Fig. 7 indicates the sales forecast for the next years.

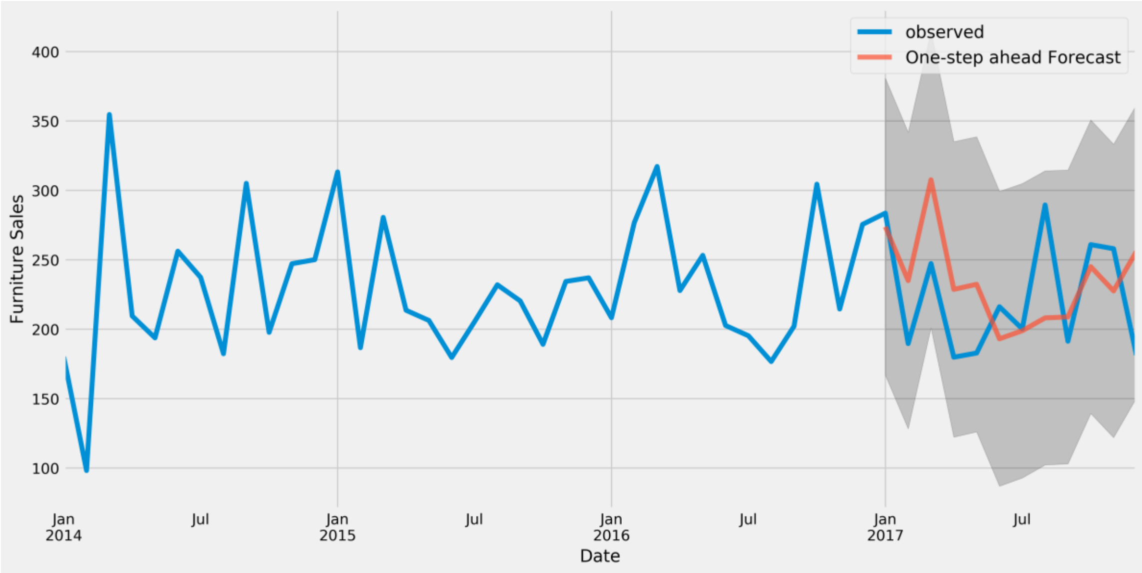


Fig. 6: One-step-ahead furniture sales forecast

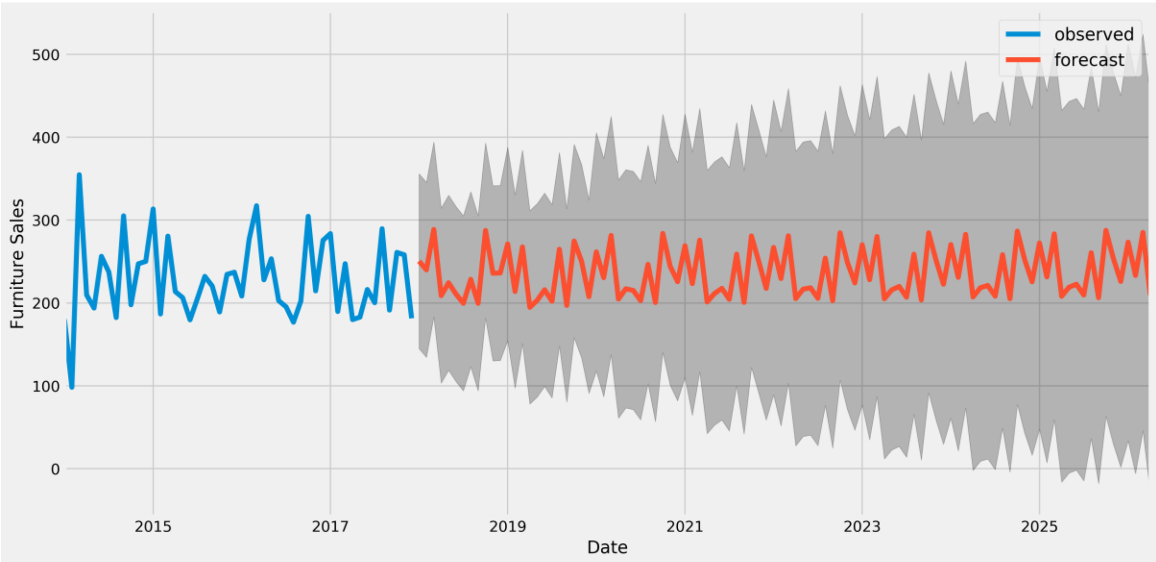


Fig. 7: Forecasting future furniture sales

The Fig.7 shows the future prediction of sales for upcoming years. The predicted trends are similar to the observed trend.

C. *Facebook Prophet*

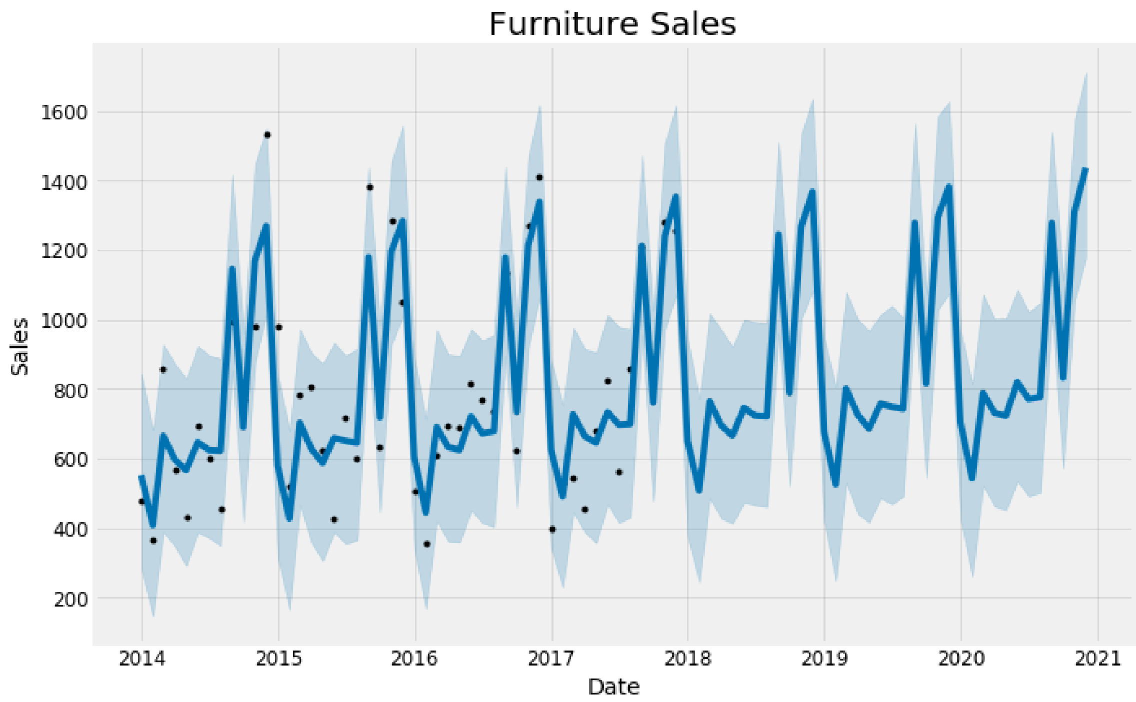
The Facebook Prophet is a comparatively better forecasting model for forecasting than ARIMA and other similar models such as-Tripe Exponential Smoothening also known as Holt-Winter's method. Fig. 8 shows the furniture sales forecast for different years.Fig. 9 shows the trend in the sales in subsequent years and day of the year sales using the FB Prophet Model.

Table 2 shows the performance comparison of various time series forecasting models-

TABLE 2 PERFORMANCE MATRIX OF DIFFERENT MODELS

Model	**MSE**	**RMSE**	**%MAPE**
ARIMA	22993.57	151.64	14.3
Holt-Winter's	7344.49	85.7	11.8
FB Prophet	4329.64	65.8	8.3

MSE- Mean Square Error
RMSE- Root Mean Square Error



. Fig. 8: Furniture sales forecast

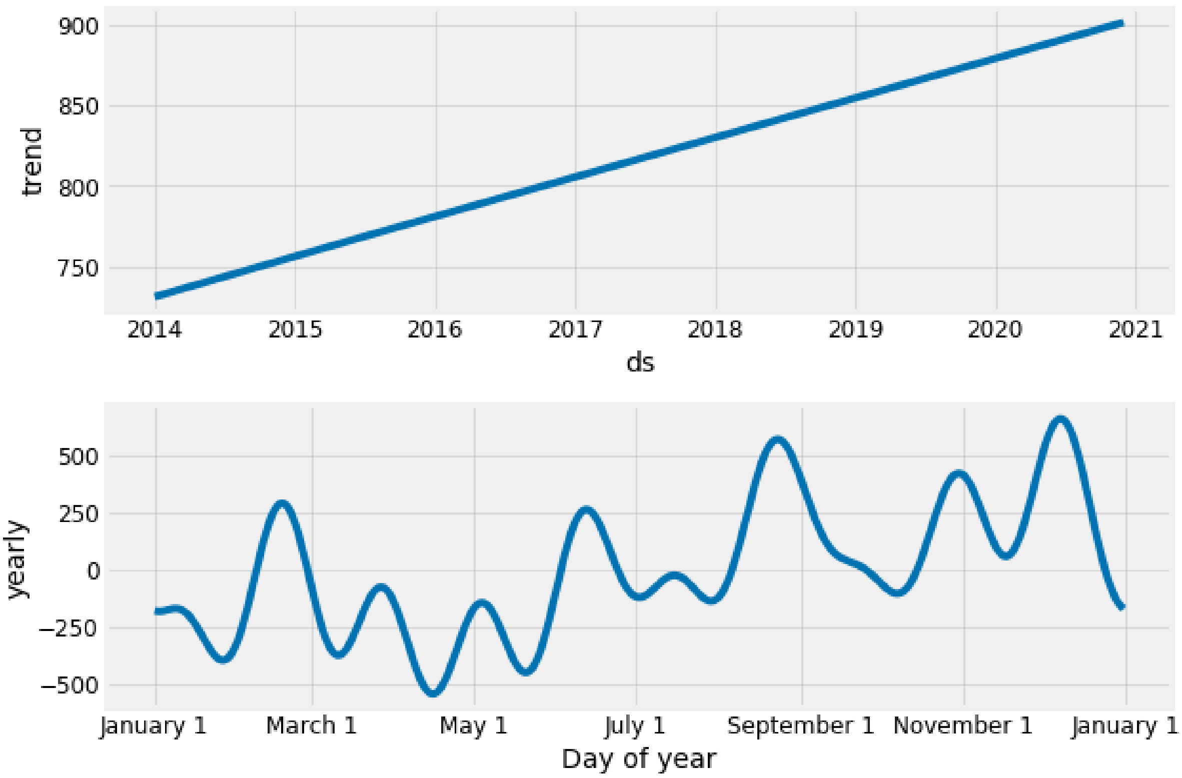


Fig. 9: Yearly and daily furniture sales

FB Prophet is a better forecasting model in terms of low error, better fitting, and better forecasting than others.

V. CONCLUSION

In the proposed research work we have examined the FB Prophet Additive and ARIMA model for sales forecasting. We have used the supermarket furniture dataset to experiment. We have concluded that the prediction made using FB Prophet is very close to the actual. The proposed tool is showing good performance in terms of forecast accuracy for time series data with low error. However, performance can be improved using fusion techniques with FB Prophet. Scalability can be another challenge for analyzing a large dataset. A transfer learning approach can be used along with FB Prophet to improve scalability and handle large datasets. Real-time prediction with high accuracy depends on the model that is used for training and validation.

ACKNOWLEDGMENT

The author is thankful to all authors who have participated to conduct the experiment and research in the proposed research work. We are also thankful to the CMR Institute of Technology, Bengaluru for providing all resources and computing facilities to conduct this research.

REFERENCES

[1] Zunic, Emir &Korjenić, Kemal &Hodžić, Kerim&Donko, and Dzenana.,"Application ofFacebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data", International Journal of Computer Science and Information TechnologyInternational Journal of Computer Science & Information Technology (IJCSIT), Vol 12, No 2, PP. 23-36, April 2020.

[2] Chen Guo, Quanbo Ge, HaoyuJiang,Gang Yao and Qiang Hua, "Maximum Power Demand Prediction UsingFbprophetwith Adaptive Kalman Filtering",IEEE Access, 2020.

[3] 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

[4] 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, ISSN No. 2455-2143, 2020, pp. 463-467.

[5] Bryan Lima, Sercan O. Arık, Nicolas Loeff, and Tomas Pfister," TemporalFusion Transformers for Interpretable Multi-horizon Time Series Forecasting", Preprint submitted to Elsevier. arXiv:1912.09363v3 [stat.ML] ,27 Sep 2020.

[6] ErolEgrioglu, Mehdi Khashei, CagdasHakanAladag, I. Burhan Turksen, and UfukYolcu, "Advanced Time Series Forecasting Methods", Mathematical Problems in Engineering, vol. 2015, Article ID 918045, 2 pages, 2015.

[7] Liyun Su, Chenlong Li, "Local Prediction of Chaotic Time Series Based on Polynomial Coefficient Autoregressive Model", Mathematical Problems in Engineering, vol. 2015, Article ID 901807, 14 pages, 2015.

[8] Ahmed Tealab,"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.

[9] CemKocak, "A New High Order Fuzzy ARMA Time Series Forecasting Method by Using Neural Networks to Define Fuzzy Relations", Mathematical Problems in Engineering, vol. 2015, Article ID 128097, 14 pages, 2015.

[10] 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.

[11] 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.

[12] 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

[13] 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.

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

[15] Zunic, Emir & Korjenić, Kemal & Hodžić, Kerim & Donko, Dzenana. (2020),"Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data", International Journal of Computer Science and Information Technology. 12. 10.5121/ijcsit.2020.12203.

[16] Wick, F. & Kerzel, U. & Hahn, M. & Wolf, M. & Singhal, T. & Feindt, M., "Demand Forecasting of individual Probability Density Functions with Machine Learning", PP. 1-20, 2020.

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

[18] Kumar, T. Senthil. "Video based Traffic Forecasting using Convolution Neural Network Model and Transfer Learning Techniques." Journal of Innovative Image Processing (JIIP) 2, no. 03 (2020): 128-134.

[19] Pavlyuk, "Transfer Learning: Video Prediction and Spatiotemporal Urban Traffic Forecasting," Algorithms, vol. 13, no. 2, p. 39, Feb. 2020.