Forecasting of sales by using Machine Learning techniques

Mohammed Nazim Uddin, PhD
School of Science, Engineering and Technology
East Delta University
Chattagram, Bangladesh
nazim@eastdelta.edu.bd

Md Shafayet Hossen Chy
School of Science, Engineering and Technology
East Delta University
Chattagram , Bangladesh
mdshafayet.safu@gmail.com

Jewel Datta
School of Science, Engineering and Technology
East Delta University
Chattagram , Bangladesh
Dattajewel7@gmail.com

Abstract:

Currently, all the organization doesn't take any risk in their business. So, they want a precise decision and sales forecasting of their product using their sales data. The amount of data has increased rapidly in recent years and the gravity of receiving benefits of this data is incessantly rising. The data yielded by the purchasers and it has become trendy to try and gain business advantages from exploring this data. Within the data, there might be patterns that could be used to guide a company in how to make decisions regarding marketing, organization, and sales. Forecasting is an important part for any organization for their decision-making process that helps an organization to modify and improve their model further. For sales forecasting various machine learning approach are used. Such as Multiple Linear Regression, ARIMA, SVM method etc. In this paper, we try to develop a forecasting sales model for time series data for the Russian one of largest IT Company named as 1C Company using ARIMA and Fb-Prophet. Time series data can be used to understand the past behavior of the series and to forecast the future behavior on the basis of past behavior

Keywords—ARIMA, SARIMA, Fb-Prophet, Time Series data, Sale forecasting.

I. INTRODUCTION

Time series forecasting is the basic study to analysis data process over period of time. This is a series of statistical observations recorded over time series. It can be used to realize past behavior of the series and based on past behavior it can forecast future behavior of the series. The target of sales forecasting is to help the organization to determine demands of products and improve their strategy for the future.

ARIMA model has been used in [10], [1] for prediction of Infant Mortality Rate (IMR) and automobile demand prediction. ARIMA work better with linearity.

STL decomposition has been used as a part of [8] to detect abnormal event detection using spatiotemporal social media data. Additive and Multiplicative model has been used in [9] to forecast electricity load. Dataset reveals that seasonal variation is roughly constant.

ARIMA methodology is used to predict next-day electricity prices in [6]. ARIMA techniques are used to analyze time series and have been mainly used for load forecasting because of their accuracy and mathematical soundness.

A hybrid methodology that combines both autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) models is used for predicting short-term electricity prices in [7]. Empirical results indicate that a hybrid ARIMA-ANN model can forecast better price accuracy.

ARIMA: ARIMA stands for Autoregressive Integrated Moving Average. It's basically used for time series forecasting. Its work more accurately when data is quite long and between past observations is stable. Its consist of Three parts where AR stands for lags of differenced series and MA stands for lags of errors and I is the number of difference applied to build time series stationary.

Fb-Prophet: Fb-Prophet is open source tools developed by Facebook research team. It's basically decompose time series data in its components. It is established on an additive model where non-linear trends are matched with yearly and weekly seasonality, plus holidays. It works better with daily periodic data with at least one year of historical data. Prophet is robust to missing data, changes in the trend, and big outliers.

The paper is structured as follows. Section I describes various machine learning model. Section II contains methodology. In section III contains experiment and evaluation part. Machine learning model result is analyzed in section IV. Final conclusions are included in section V.

II. METHODOLOGY

Different kind of machine learning models have been developed for foresting of time series data. In this section we describes how our forecasting model's algorithm work for time series data which are ARIMA and Fb-Prophet.

A)ARIMA: The Autoregressive Moving Average ARMA develops a model to forecast future. It is worthy to exploit with a stationary time series. Stationary is tasted by using dickey-fuller method. ARMA (p, q) has two parts where p stands for the order of the autoregressive part and q stands for the order of the moving average part. The Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) are generally employed to assess the model. High order p and q can cause over-fitting, but increasing the orders of p and q generally improves the fitting suitability. However, substantial forecasting errors may arise in this model. Akaike (1974) proposed the Akaike Information Criterion (AIC). The AIC penalty disarm overfitting and construct a worthy fitting measure for ARMA. The Auto regressive Integrated Moving Average (ARIMA) model was developed to solve this problem. The model is generally referred to as an ARIMA (p, d, q) model, where p, d, and q are integers greater than or equal to zero that refer to the order of autoregressive, integrated, and moving-average parts of the model respectively.

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where y_t is the actual value, ε_t is random error at time period respectively, and $\varphi_i (i = 1,2 \ldots . p)$ and $\theta_j = (j = 0,1,2 \ldots . q)$ ARIMA parameters and p, q are integers which are ARIMA parameters. Random errors ε_t are assumed to be independently and identically distributed with a mean of zero and a constant variance σ^2.

Some time series add seasonal and trend characteristics. Seasonal Autoregressive Integrated Moving Avera (SARIMA) methods are famous in this context. Wu (Wu, 2000) defined a SARIMA (p, d, q)(P, D, Q)s model as:

$$\emptyset_p\,(B)\,\Phi_p\,(B^s)(1-B)^d(1-B^s)^D\,y_t = \theta_q\,(B)\,\Theta_Q\,(B^s)\,\varepsilon_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where, θ denotes deterministic trend ,B denotes back-sift operator , y_t actual forecast value , ε_t is the time error term, s is the seasonal cycle, P is the seasonal autoregressive term, D is the seasonal integrated term, Q is the seasonal moving average term.[11]

Overall system architecture describe below by a figure.

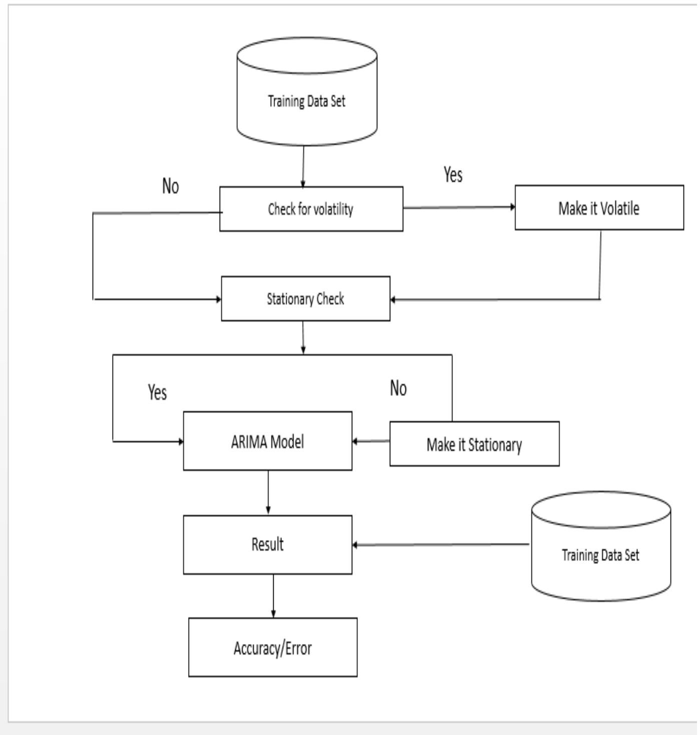

Figure 1: System Architecture

B)Fb-Prophet: The Fb-Prophet method[12] is an additive regression model with four main components:

1. A piecewise linear or logistic growth curve trend. Prophet automatically dig up a change in trends by determining change points from the data.
2. A yearly seasonal component modeled using Fourier series transformation.
3. A weekly seasonal component using dummy variables.
4. A user-provided list of important holidays

It uses a decomposable time series model with three main model components: trend, seasonality, and holidays. Given equation combine this three component:

$y(t)=g(t)+s(t)+h(t)+ \varepsilon(t)$ ……………… (3)

g(t): piecewise linear or logistic growth curve for modeling non-periodic changes in time series
s(t): periodic changes (weekly/yearly seasonality)
h(t): outcome of holidays with irregular schedules
εt: error term accounts for any abnormal changes not makeup by the model.

Trend

Trend is designed by fitting a piecewise linear curve over the trend or the non-periodic part of the time series. The linear matching exercise makes sure that it is least influenced by missing data. The Prophet library appoint two probable trend models for g(t). The first one is Nonlinear, Saturating Growth. It is illustrated in the form of the logistic growth model:

$$g(t) = \frac{C}{1+e^{-k(t-m)}} \quad \text{…………………. (4)}$$

Other trend model is a simple Piecewise linear Model where the growth rate is constant. It works best for problems without saturating growth.

Where C is the carrying capacity (which is the curve's maximum value),k is the growth rate is an offset parameter.

Seasonality

To match and forecast the impact of seasonality, prophet depends on Fourier series to provide a suitable model. Seasonal effects s(t) is defined by the following function:

$$s(t) = \sum_{n=1}^{N}\left(a_n cos\left(\frac{2\pi nt}{P}\right) + b_n sin\left(\frac{2\pi nt}{P}\right)\right)$$
……………..………… (5)

P is the period (365.25 for yearly data and 7 for weekly data)
Parameters [a1, b1 . . . aN,bN]need to be calculated for a given N to model seasonality. The Fourier order N that defines whether high-frequency changes are allowed to be modeled is a significant parameter to set here.

The error term *ϵ(t)* describe information that was not reflected in the model. Usually it is modeled as normally distributed noise.

III. EXPERIMENT & EVALUATION

A)ARIMA

For fitting dataset in ARIMA first have to check data is stationary or not. We can check it using dicky-fuller test where stationary depends on P(lag order) value. Dicky-fuller model:

$$\Delta Y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + . + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad \text{.................................} \ (5)$$

Where α is a constant, β the coefficient on a time trend and p the lag order of the autoregressive process

If data is non-stationary that indicate trend, seasonality, cyclist are present on dataset. Most common way to make a non-stationary time series in stationary time series is differencing. Dataset shown yearly seasonality, so frequency of 12 has been used in the prediction process.

Model	AICc value
ARIMA(1,0,1)(0,1,0)12	173.33
ARIMA(1,1,0)(0,1,0)12	156.83
ARIMA(0,1,0)(0,10)12	153.20

Table 1: ARIMA sequence measurement

ARIMA model is measurement with lowest AIC value. ARIMA model (0,1,0)(0,10)12 gave the lowest AIC value .So this model will use to predict forecast.

B)Fb-Prophet Dataset has yearly seasonality. So here regular period P = 365.25. For yearly and weekly seasonality it has found N = 10 and N = 3 respectively to work better for most problems. As dataset has yearly seasonality need to use N = 10 for dataset.

Evaluation Method:

The mean absolute percentage error (MAPE) is a measurement of prediction accuracy of a forecasting method.

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Where A_t is the actual value and F_t is the forecast value.

Correlation coefficients are used to measure how strong a relationship present is between two variables.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2]\,[n\sum y^2 - (\sum y)^2]}}$$

IV. Result Analysis

ARIMA: In figure 2 where horizontal axis represents month and vertical axis represents monthly sales. By analyzing the figure 3, it can be concluded that ARIMA model has successfully covered most linear component of data set but seasonal component hasn't cover successfully.

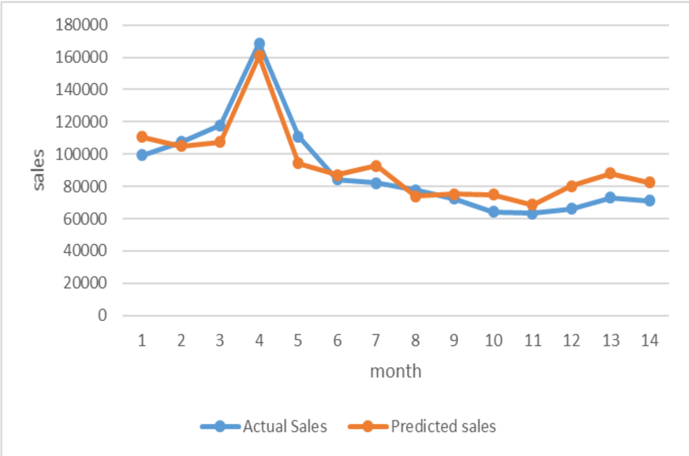

Figure 2: Actual vs ARIMA prediction

Fb-Prophet: In figure 3 where horizontal axis represents month and vertical axis represents monthly sales. By analyzing the figure 3, it can be concluded that Fb-Prophet has successfully covered linear and on-linear component.

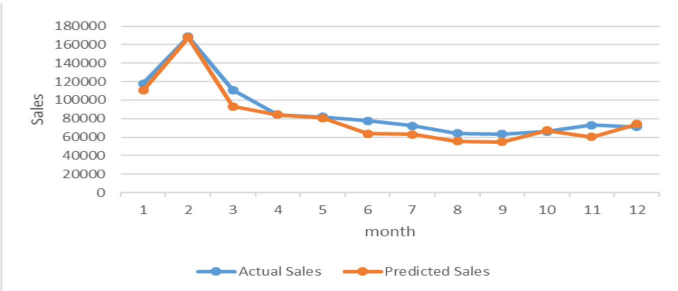

Figure 3: Actual vs Fb-Prophet prediction

Performance measure	RMSE	MAE	MAPE
ARIMA	10153	9033.428571	10.844%
Fb-Prophet	19036	12861.250000	12.623%

Table: Performance Measure of ARIM and Fb-Prophet

Here, We get better RMSE value for ARIMA as tested dataset has large trendy component and less seasonal component as AIRMA works better for trendy data.

V. CONCLUSIONS

In this paper we apply two machine learning technique on time series data for forecast which are ARIMA and Fb-Prophet. Analyzing these two model we can conclude that ARIMA give best prediction for linear data and doesn't cover seasonal component enough. Fb-Prophet kind of decomposition method. It's decompose data into its component using additive regression method and cover all the component successfully for prediction. In future interested to work on STL decomposition where will try to fit different algorithm for different Component.

VI. REFERENCE

[1] D. Chen, "Chinese automobile demand prediction based on ARIMA model," 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), Shanghai, 2011, pp. 2197-2201

[2] J. H. Lo, "A study of applying ARIMA and SVM model to software reliability Prediction ," Uncertainty Reasoning and Knowledge Engineering(URKE), 2011 International Conference on, Bali, 2011, pp. 141-144.

[3] Y. He, Y. Zhu and D. Duan, "Research on Hybrid ARIMA and Support Vector Machine Model in Short Term Load Forecasting," Sixth International Conference on Intelligent Systems Design and Applications, Jinan, 2006, pp. 804-809.

[4] Y. Yan, P. Guo and L. Liu, "A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Forecasting Resource Consumption in an IIS Web Server," Software Reliability Engineering Workshops (ISSREW), 2014 IEEE International Symposium on, Naples, 2014, pp. 437-442.

[5] Ren Ye, P. N. Suganthan, N. Srikanth and S. Sarkar, "A hybrid ARIMA-DENFIS method for wind speed forecasting," Fuzzy Systems (FUZZ), 2013 IEEE International Conference on, Hyderabad, 2013, pp.1-6.

[6] J. Contreras, R. Espinola, F. J. Nogales, A. J. Conejo, "ARIMA models to predict next-day electricity prices", *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014-1020, Aug. 2003.

[7] Phatchakorn Areekul, Tomonobu Senjyu, Hirofumi Toyama, Atsushi Yona, "A Hybrid ARIMA and Neural Network Model for Short-Term Price Forecasting in Deregulated Market", *IEEE TRANSACTIONS ON POWER SYSTEMS*, vol. 25, no. 1, FEBRUARY 2010.

[8] J. Chae et al., "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on, Seattle, WA, 2012, pp. 143-152

[9] H. K. Temraz, M. M. A. Salama and V. H. Quintana, "Application of the decomposition technique for forecasting the load of a large electric power network," in IEE Proceedings - Generation, Transmission and Distribution, vol. 143, no. 1, pp. 13-18, Jan 1996.

[10] D. Purwanto, C. Eswaran and R. Logeswaran, "A Comparison of ARIMA, Neural Network and Linear Regression Models for the Predic-tion of Infant Mortality Rate," 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, Kota Kinabalu, Malaysia, 2010, pp. 34-

[11] Pei Liu, Shih-Huang Chen, Hui-Hua Yang, Ching-Tsung Hung, Mei-Rong Tsai "Artificial Neural Network and SARIMA in Portland Cement Supply Chain to Forecast Demand," in 2008 Fourth International Conference on Natural Computation, 18-20 Oct. 2008.

[12] Forecasting at scale: FBPROHET [online]. Available: https://peerj.com/preprints/3190.pdf