JETRAIL FORECASTING USING TIME SERIES MODEL

Sreeram K[1], Poornima V[2], Hemanth V M[3], Shivakumara K S[4], Pramod Sunagar[5]

Student, Department of CSE, MSRIT, Bangalore, Karnataka, India[1, 2, 3, 4]

Assistant Professor, Department of CSE, MSRIT, Bangalore, Karnataka, India[5]

Email: sreeram.k0104@gmail.com, poornimavhsn@gmail.com, hemanth.1095@gmail.com, shivakumarks321@gmail.com, pramods@msrit.edu

Abstract— In this modern era, Data mining plays a key important role in this competitive business world and its profit where it predicts the future data based on the past collected records. With high number of passengers travelling in jetrail has increased, so forecasting the future passenger's count and maintain the traffic count is challenging task. In the existing system forecasting is made upon the fixed constant values for all the days.so the future collecting data will be compared with only with the fixed constant.so the existing system does not have the intelligence to predict the future data. The proposed system is based on the time series techniques and model. This paper gives us a broad knowledge about the methodology and final results about the jetrail forecasting with the time series models using R studio analysis program. The collected time series dataset is fed to this time series models as input and predicted values comes out as output based on the past collected data. This paper also includes comparison between techniques and the models. ARIMA and PROPHET are the main models used and PROPHET is compared to best from other models and techniques based upon the predicted data and error values.

Keywords— Jetrail forecasting, ARIMA, PROPHET, traffic prediction, Time series data.

I. INTRODUCTION

Jetrail is one of the public rail transport system where it uses the propulsion technology to move the passengers at a high speed. As usage of jetrail has been increased, where the forecasting about the transport vehicle is a great demand in the transport industry. The exactness of predicting the future data is an important task and one of the interesting areas for data analysts and transport investors.

Time series data has always been in increased trend in different domains. Time series data is ordered sequence of real time value data in an indicated temporal periods. Time series data will also depend on time itself. The collected time series data is used in forecasting the sales of home for the next year, traffic count, number of calls count where a person received and analyzing the air temperature in atmospheric science. Time Series Data Mining (TSDM) mainly depends on the valuable data and the prediction can only be analyzed through discovery of pictorial patterns.

The objective of the time series forecasting depends on the description of the analysis of the time series to get the outcome based on the statistical measure, explaining about the time variants and related time periodic series, the ability to predict the exactness future values based on the time series models, finally the control over the models and predictive data and maintaining it will result in analyzing and operating the models accurately.

The selection of the time series data model is decided purely on the time series dataset and the collected type of the attributes. So we choose the ARIMA and PROPHET model. So the main aim is to find which is the best model to predict the future values?

In the next sections we will discuss about the related works. In the next section III we will learn about the dataset, preprocessing and splitting of the data. In section III we will learn about the methodology used in this paper. In section IV we have the comparison study on the time series techniques and models and finally in section V we have the conclusion of this paper.

II. RELATED WORK

LI LI.et.al [1], describes about the Principal Component Analysis (PCA) patterns where he conveys about the few traffic time arrangement examination. He also discuss about the functionalities of the Intelligent Transportation System (ITS) and proposed the solution for abnormal data detection, missing data and

imputation in the dataset, data compression and the traffic prediction they have handled the missing values and imputed the correct time based value it totally depended on the missing period of time .

Soheila Mehrmolaei.et.al [2], describes about the enthusiasm for estimating time arrangement databases in different domains and also describe about the time series mining objectives and control over the forecasting method and increase the accuracy in predicting the values.

Mahalakshmi.G.et.al [3], tells us about the time arrangement comparison (analysis) and boosting future qualities to the prescribed research center. They also described about the major temporal data's forecasting where and when to apply those methods and also a study on the performance and the accuracy of the temporal data predictions.

Priyamvada.et.al [4], describes about the prediction using the Holt-winter model in different domains and forecasting the data based on this model weather to decide it is a short term or long term forecasting.

Zhangyu.el.al [5], explains about the straight linear slopes and curves and obtained in the results by applying the Artificial Neural Networks. He also describes about the time window, dynamic time wrapping, fitting time straight line and fast pattern match and others linear slopes and he also describes about the high ordered neural network and the multi-liner layer perceptron using the artificial neural network.

Mustafa Akpinar.el.al [6], tells about the natural gas prediction where he has used ARIMA model for his prediction. The main idea of his prediction was to improve ARIMA by removing the cyclic components of the structure and even removing the merged data value he also tested the data values with the prediction of the house sales for the next few years and the results of it was satisfying .

Narendra Babu C.el.al [7], explains about the forecasting made on the temperature of the global atmosphere based on the ARIMA model using the times series data. They have used different form of ARIMA variant models and predicted the future values of the global temperature. In his paper, we can learn about the basic ARIMA model, the trend based ARIMA model which is a very better mode if u have a trend in the data collection of yours and finally the wavelength based ARIMA model

where it falls into the category of the atmospheric science and deals with prediction of the sine wave. The paper also tell us about the Artificial neural networks in predicting the climatic changes and forecasting its values.

Yan-ming Yang.el.al [8], describes about the forecast in the failure rate in the aircraft where they have describe about the Holt-winter model to predict when the danger and threat is going to happen if don't maintain the proper changes in the aircraft both in term of hardware report and the software report. They mainly concentrated on the seasonal timely data with the Holt-Winters smoothing model and they have proposed about the additive seasonal model for the examination of the results of failure of the aircraft. They have also analysis the multiplicative seasonal model for the amplitude seasonal model with the time series data and it was proportional to it.

III. DATA PREPROCESSING AND FEATURE SELECTION

A. Dataset

JetRail Dataset consists of data from 2012 to 2015 and it is used in our project. The jetrail dataset contains 23400 rows and 3 columns as shown in figure 1. The attributes of this dataset consists of ID, Datetime and Count. The id specifies the row number. The count specifies the number of passengers that are travelled. The Datetime shows the time in which the passengers are travelled in the jetrail. Further the data set is in the form of hourly basis.

	ID	Datetime	Count
1	0	25-08-2012 00:00	8
2	1	25-08-2012 01:00	2
3	2	25-08-2012 02:00	6
4	3	25-08-2012 03:00	2
5	4	25-08-2012 04:00	2
6	5	25-08-2012 05:00	2
7	6	25-08-2012 06:00	2
8	7	25-08-2012 07:00	2
9	8	25-08-2012 08:00	6
10	9	25-08-2012 09:00	2
11	10	25-08-2012 10:00	2
12	11	25-08-2012 11:00	6

Figure 1. An overview of the data set of JetRail.

B. Feature Preprocessing:

Based on our analysis we found our data set attributes Id, count are in the form of integer and the Datetime attribute is in factor form. First we have to convert the datetime into the standard time format using POSIXct or POSIXlt. Transform the converted datetime to XTS format that the time series analysis can be done using that format. We will remove the attribute ID that is not required for our analysis. We will change the dataset from hourly basis to daily basis for better understanding and analysis of the model. We are splitting the train data set into train and validation data set to train the model and test the accuracy of the model by comparing the actual and predicted values by the model in the future sections .We divide the train dataset q from 25-08-2012 to 24-06-2014 and for validation set from 25-06-2014 to 25-09-2014, after splitting we got 18288 for train and 5112 for validation.

IV. PROPOPSED MODELS

A. Auto Regressive Integrated Moving Average (ARIMA) Model

The ARIMA model is one of the famous models which are used for forecasting the financial prediction and other forecasting applications. Seasonal ARIMA and Non-Seasonal ARIMA are the types of ARIMA models. We will not consider the Non-Seasonal ARIMA Model because this model will remove the seasonal and holiday data's from the dataset. Removal of the data will affect the prediction analysis and in our dataset holidays and seasonal period also plays a very important role in forecasting the number of passengers count. We are considering the seasonal ARIMA where we get the seasonal and trend period in count of our dataset. We are constructing this model in R analytics studio program we are calling the inbuilt auto.arima() function to construct the model and train the model for the forecasting purpose the auto.arima() takes three parameters which is the p,d,q values in it.

P value is the total number of autoregressive terms of seasonal period.
D value is the number of difference in the seasonal period.
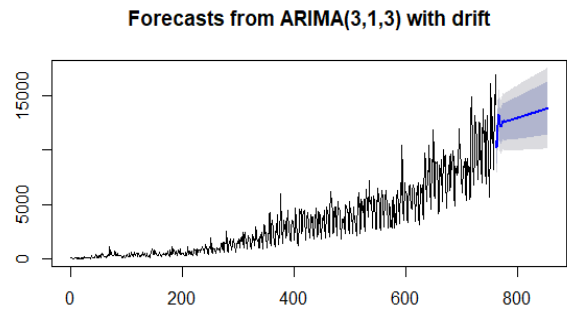Q value is the delaying prediction errors in the forecasting methods.

Fig. 2. Forecast from ARIMA

B. PROPHET Model

PROPHET model is the one of the quick forecasting model which is both available in R studio and Python Analytics platforms. PROPHET model gives strong results even with missing values, longer trends and larger outliers in the dataset.

PROPHET model is one of the optimized open source software from Facebook to forecast the values. PROPHET works well and forecast better when the dataset consist of the time, daily, weekly, yearly, trend and seasonality .As per our observation we have all these factors in our dataset.

The framework of the PROPHET model is made to work easily with the trend and seasonality. The framework consists of its own data frames and where we need to predict the values based on those data frames. The data frames will utilize two columns that is "ds" and "y" where it indicates storing the date time series. The framework of the PROPHET model itself will handle and give option to take care the seasonality and trend of the dataset. Due to this feature the model is very much popular in the forecasting applications.
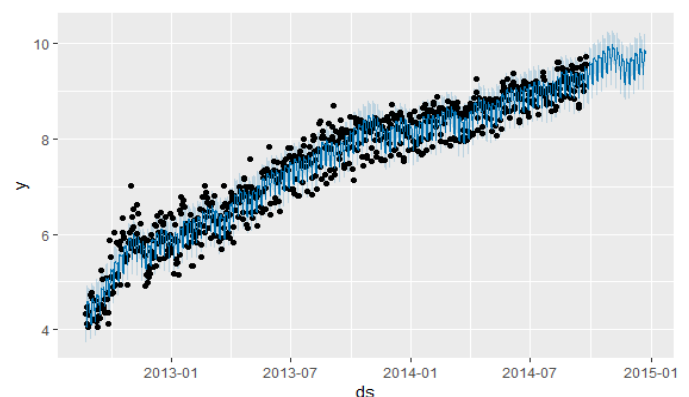

Fig. 3. Forecasting in Prophet Model

V. PERFORMANCE EVALUATION

As we have seen the construction and implementation of the ARIMA and PROPHET model we will compare these two models with other time series techniques and models. So we are considering different metrics for the comparison of these models and techniques. The metrics are Mean Square ERROR (MSE), Root Mean Square ERROR (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percent Error (MAPE)

MODEL	MSE	RMSE	MAE	MAPE
NAVIE approach	34735.15	186.3737	148.731	51.0172
Simple Moving Average	34811.12	186.5774	149.7117	52.3202
Simple Exponential	901474.29	949.46	628.0508	29.86324
ARIMA	958695.55	979.13	624.675	32.03639
PROPHET	4.7932	2.18933	1.8461	1.212

Table I: Comparison between the time series model.

The above table as the comparison between the time series techniques and models we found out that PROPHET model is the best in the prediction and the error values are very low compared with other models and techniques.

VI. CONCLUSION

The main goal of this paper is to predict the forecasting of jetrail passenger count from 2014 to 2015 and we have examined the models as accurate as possible .We have trained the model from 2012 to 2014 dataset and the model should predict the next year values based on the trained values. As we have used the r studio analytics program where we have implemented the ARIMA, PROPHET model and other time series techniques and we have found that the predicted values from the PROPHET model is better and the error values of the PROPHET model in forecasting the future values is very less when compared to other time series techniques and models. The ARIMA model is making the forecasting values quite close to the actual value with 85% confidence and PROPHET model is predicting the future values 97.53% to the actual data.

REFERENCES

[1] Li Li, Xiaonan Su, Yi Zhang, Yuetong Lin, and Zhiheng Li, "Trend Modeling for Traffic Time Series Analysis: An Integrated Study",IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 16, NO. 6, DECEMBER 2015.

[2] Soheila Mehrmolaei and Mohammad Reza Keyvanpour ,"Time series forecasting using improved ARIMA ",978-1-5090-2169-7/16/$31.00 ©20 16 IEEE.

[3] G.Mahalakshmi, Dr.S.Sridevi and Dr.S.Rajaram ,"A Survey on Forecasting of Time Series Data", 978-1-4673-8437-7/16/$31.00 ©2016 IEEE .

[4] Priyamvada and Dr. Rajesh Wadhvani 2,"Review on Various Models for Time Series Forecasting", Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) IEEE Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9.

[5] Zhang Yu," Research Of Time Series Finding Algorithm Based On Artificial Neural Network", 2009 World Congress on Computer Science and Information Engineering.

[6] Mustafa Akpinar, Nejat Yumusak." Forecasting Household Natural Gas Consumption with ARIMA Model: A Case Study of Removing Cycle", IEEE 2013 7th International Conference on Application of Information and Communication Technologies .

[7] Narendra Babu .c, B. Eswara Reddy," Predictive Data Mining On Average Global Temperature Using Variants of ARIMA models",IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012.

[8] Yan-ming Yang', Hui Yu, Zhi Sun," Aircraft Failure Rate Forecasting Method Based on Holt-Winters Seasonal Model", 2017 the 2nd IEEE International Conference on Cloud Computing and Big Data Analysis.