ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE
International Journal of Advanced Research in Computer and Communication Engineering
ICRITCSA
M S Ramaiah Institute of Technology, Bangalore
Vol. 5, Special Issue 2, October 2016

Predictive Analytics on Aviation Data

Arun Kumar R [1], Balaji S [1], Sai Manoj T V[2], Sailaja Kumar K[1], Evangelin Geetha[1]

Department of Computer Applications, M S Ramaiah Institute of Engineering, Bangalore, India [1]

Department of Computer Science and Engineering, M S Ramaiah Institute of Engineering, Bangalore, India [2]

Abstract: The technology developments in Aviation have been and will continue to be at the progressive of human technological and scientific development. The present paper analyses the airline revenue passenger miles on monthly base in the past ten years. Time series analysis is used to predict the increase in revenue passenger miles using the previous years' data from the corresponding time period. The paper also shows the random change of data which will be pre-processed for analysis. Moreover, it presents the seasonal change and the trend that the dataset is following. Based on all these factors, the revenue passenger miles for future years are predicted and the results are shown with graphical output.

Keywords: Airline Revenue Passenger Miles, Prediction, Time Series, Moving Average Model

I. INTRODUCTION

Aerospace technology, in addition to make distances shorter, it facilitates social, economic, scientific, and cultural exchanges. The airline industry is remarkably capital intensive and requires a massive collection of exceedingly expensive and technologically sophisticated equipment and facilities [1].

Revenue passenger miles (RPM) are measures of traffic for an airline flight calculated by multiplying the number of revenue-paying passengers aboard the vehicle by the distance travelled. Revenue passenger miles can be considered the basic amount of "production" that an airline creates.
The revenue passenger miles can be compared to the available seat miles over an airline's system to determine the overall passenger load factor is determined by comparing RPM with available seat miles [2].

The term Time Series is defined as a sequence of data points which are spaced at uniform intervals in time and are measured repeatedly at successive times. Time series analysis includes methods to identify the behaviour of the time series events. The data points collected over time have an internal structure such as autocorrelation, trend or seasonal variation. They help to make a prediction about the future behaviour of the data points using time series analysis [3].

R is a language and widely used environment for statistical analysis and graphics. R provides a wide variety of statistical and graphical techniques, and is highly extensible. In this paper, the data of Revenue passenger miles is considered and time series analysis is carried out using R. We used some of the techniques for smoothing of data, finding moving average and fitting the model. Future behaviour of the data is predicted and presented.

II. LITERATURE REVIEW

The review on the available literature on the analysis of aviation data is presented in this section. Multiple kernel algorithms that are used to analyze the aviation data is presented in. It applies the data mining concepts to identify the hidden patterns in the dataset for prediction [4]. The report given in [5] provides guidance to individuals who prepare airport activity forecasts as well as to those who review the forecasts. The guidance covers the basic steps required for producing forecasts. The report presented in [6] gives the complete idea about the time series and the procedure for applying the time series analysis technique for smoothing the dataset for forecasting. The results of the Oxford Economics studies on the benefits of aviation and the relationships between the performance of specific components of the Travel and Tourism Competitiveness Index (TTCI) are explored in [7].

III.METHODOLOGY

- Download the dataset from the chosen website
- Remove the white noise present in the dataset for analysis using R
- Analyse the data set to determine whether time series analysis can be applied
- Import the dataset to R-Studio and plot it as time series data and analyse the summary.
- Checkout for trends and seasonal variations to know which trend is following every year
- Plot the decomposition of additive time series for analysing the Random, Seasonal, Trends and Observed variation in dataset for each year.
- Make use of ARIMA moving average model for simulating a moving average and plot the result.
- Using predict method perform the forecasting and plot the forecasting graph.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICRITCSA

M S Ramaiah Institute of Technology, Bangalore

Vol. 5, Special Issue 2, October 2016

IV. VALIDATION AND DISCUSSION

A. Case Study

Dataset we have chosen is the Airlines revenue passenger miles. It is extracted from the website https://www.bts.gov[8]. The dataset is collected from 1996 to 2015. Parameters considered in the dataset are Date and Total.

The dataset consist of monthly details of total revenue from the year 1996 so by analysing the dataset we can say that time series analysis would be more suitable for forecasting the future results. The objective of the analysis is to predict future forecast for the revenue passenger miles. The dataset shows the revenue is increasing year by year in a particular trend.

Dataset Operations:

Time series analysis is used for this dataset. We first pre-process the dataset by removing the unwanted data and then used the Smoothing technique and moving average technique in R to predict the future forecast.

Inference of Dataset:

The airline revenue passenger miles are increasing year by year in a linear fashion. It follows a particular trend in every year for each month. Prediction provides knowledgeable information regarding the future of the revenue passenger miles for airlines. Thus this can be utilized in future forecast.

B. Illustration of Methodology

The dataset contains parameters date and total of the airlines revenue passenger miles. As the data contains the periodical information with month and year of past 19 years we choose the time series technique to analyse the dataset for forecasting the future. Sample data set is given in Fig. 1.

Source code in R used for prediction is as follows:

Plotting the data as time series:

```
plot(rpm$Total~rpm$YYYYMM)
rpm.ts = ts(as.numeric(rpm$Total),start =
                        c(1996,01),freq=12)
plot(rpm.ts,ylab='Revenue Passenger Miles')
```

Check for trends and seasonal variations:

```
layout(1:2)
plot(aggregate(rpm.ts))
boxplot(rpm.ts~cycle(rpm.ts))
```

Time series decomposition:

```
rpm.decom=decompose(rpm.ts)
plot(rpm.decom)
```

Moving average model:

```
rpm.arma=arima(rpm.ts,order=c(1,0,1))
acf((rpm.arma$resid))
```

Future forecast:

```
ts.plot( cbind( window(rpm.ts,start = c(2008,1)),
redict(rpm.arima,70)$pred ), lty = 1:2)
```

	YYYYMM	Total
1	Jan-96	41,972,194
2	Feb-96	42,054,796
3	Mar-96	50,443,045
4	Apr-96	47,112,397
5	May-96	49,118,248
6	Jun-96	52,880,510
7	Jul-96	55,664,750
8	Aug-96	57,723,208
9	Sep-96	47,035,464
10	Oct-96	49,263,120
11	Nov-96	43,937,074

Fig. 1. Sample data set

C. Results and Discussion

Thetrend that the dataset follows is depicted using a boxplot in Fig. 3. Thefuture forecast for the years 2016 and 2017 are presented using graph in Fig.2. According to the obtained results, the airlines revenue passenger miles will be increasing in coming years with the same trend as followed in previous years also.

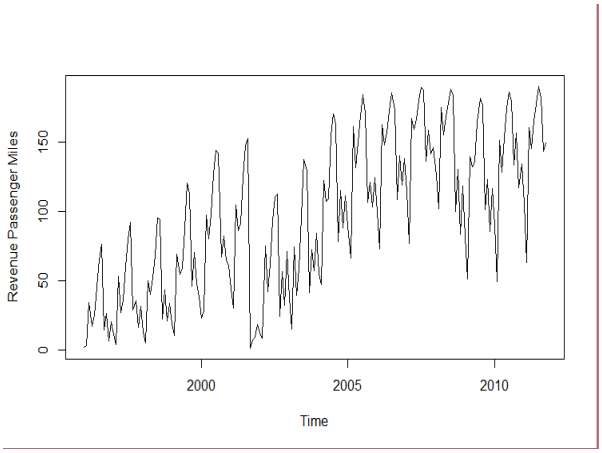

Fig. 2Plotting data as time series

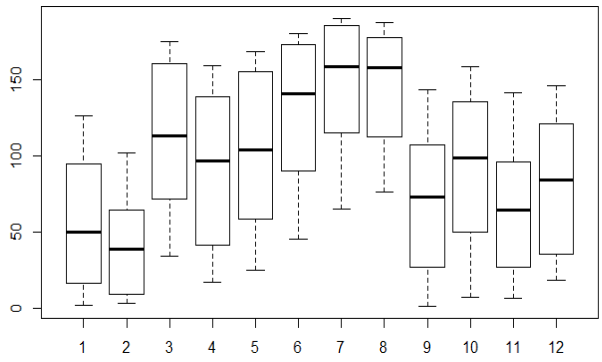

Fig. 3. Trend followed by the data

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICRITCSA

M S Ramaiah Institute of Technology, Bangalore

Vol. 5, Special Issue 2, October 2016

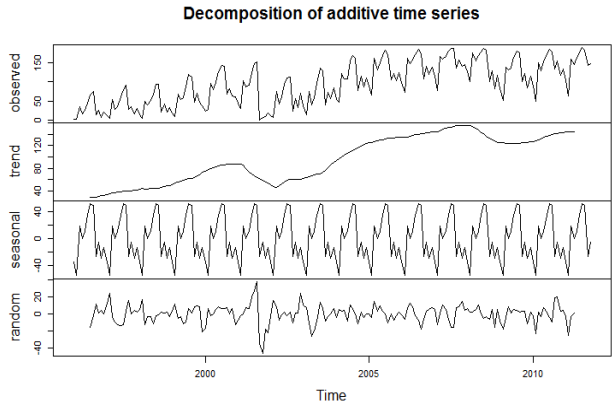

Fig. 3. Decomposition of Time series

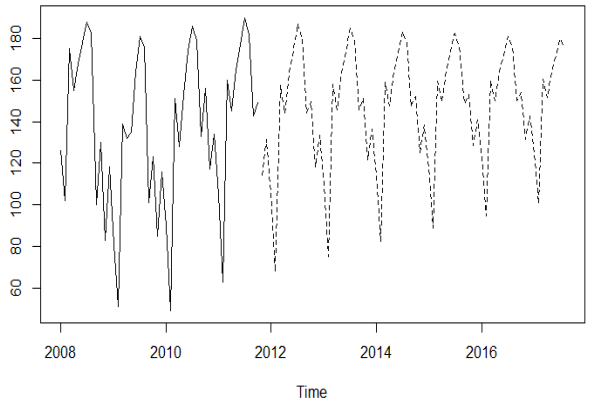

Fig. 4. Future forecast

V. CONCLUSION AND FUTURE WORK

The airlines revenue passenger miles falls under the aviation domain and time series analysis technique is applied to model this dataset. With a particular fixed trend the revenue passenger miles are increasing linearly. According to our results the model specifies that each and every year it follows the same trend and in upcoming years also it increases in same fashion. The future work can be done by considering the change in trends and using different smoothing technique for analysis.

REFERENCES

[1] www.sba.uconn.edu/users/jgolec/aviation.pdf
https://en.wikipedia.org/wiki/Revenue_passenger_mile
[2] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.
[3] Somaye Gharibvand, Lida Gharibvand, "Financial Analysis Using SAS ® PROCS", SAS Global Forum 2010.
[4] www.digital.morris.umn.edu/horizons/vol2/iss1/3/
[5] http://www.faa.gov/af1.doc
[6] http://www.abs.gov.abv
[7] Julie Perovic,"The Economic Benefits of Aviation and Performance in the Travel & Tourism Competitiveness Index, The Travel & Tourism Competitiveness Report 2013.
[8] https://www.bts.gov