A Model for Forecasting Tourists Arrival in J&K, India

Sourabh Shastri
Department of Computer
Science & IT, Bhaderwah
Campus, University of Jammu,
J&K, India

Anand Sharma
Department of Computer
Science & IT, University of
Jammu, J&K, India

Vibhakar Mansotra
Department of Computer
Science & IT, University of
Jammu, J&K, India

ABSTRACT

Data Mining is a method for extracting patterns from historical data. In this paper, we forecast the number of tourists in J&K state for the next five years that should totally depend upon the historical time series data of tourists in J&K state. For this, Exponential Smoothing model and IBM SPSS Modeler 16.0 data mining tool are used. Exponential Smoothing is a popular forecasting method that is used to predict the immediate future for time series data. The purpose of this paper is to forecast the number of tourists in advance for the J&K state so that the tourism department shall be prepared in advance to provide essential services to the forthcoming tourists.

Keywords

Data Mining, Time series data, Exponential Smoothing, IBM SPSS Modeler

1. INTRODUCTION

Data Mining is defined as the procedure of extracting information from huge sets of data. Data mining is the fundamental stage inside the process of extraction of useful and comprehensible knowledge, patterns and trends from large quantities of historic data stored in different formats. Its objective is to improve the decisions of organizations, companies and institutes by providing tools for automatically and intelligently transforming huge amount of data into information relevant to users in advance. Data mining is a relatively new area of computer science that brings the concept of artificial intelligence, data structures, statistics, and databases together. Data mining models are of two types: Predictive models (supervised) and Descriptive models (unsupervised). Predictive models makes prediction about unknown data values by using the known data values. Examples are: Classification, regression, time series analysis etc. Descriptive models identify the patterns or relationship in data and explore the properties of the data examined. Examples are: clustering, summarization, association rules and Sequence Discovery [1].

Banking, insurance, marketing, finance, sales, healthcare etc. are some of the applications of data mining. But in the present paper tourism field has been selected. Tourism Development Corporation has at their disposal vast amount of data which on thorough analysis can lead to more efficient decision making. There is continues increase in the total tourists arrival including domestic and international. The Tourism Development Corporation has to provide infrastructure like accommodation, transport etc. For this there should be some method to predict how many tourists shall come in the future years so that Tourism Development Corporation is prepared in advance to provide good facilities to the tourists.

To solve these types of problems, the time series analysis will be the best tool to forecast the trend/pattern. A time series database consists of sequence of values changing with time. In this paper, focus is on yearly data from 1989-2014 for forecasting the next 5 years data so that Tourism Development Corporation should prepared in advance to provide high quality facilities to the tourists.

2. PRESENT SCENARIO OF TOURISM IN J&K

Tourism department is one of the most important wings of J&K Government. Its major responsibilities include planning and execution of schemes for the growth, up-gradation and enhancement of the tourism infrastructure in different areas of the State [2].

J&K has three distinct regions, viz. Jammu, Kashmir and Ladakh and all these have huge potential for tourism from both domestic as well as international tourists. There are numbers of spots which are unexplored and can be developed as major tourist spots, having the whole natural as well as the cultural resources for attracting tourists. The state has plenty of large lakes, the largest of the fresh water lake being Wular, Dal and Manasbal lakes. The biggest brackish water lakes are Tso Moriri and Pangong Tso in Ladakh region [3].

Tourism is one of the State's major businesses. It has played a significant job for developing the financial system, particularly in the area of the Valley and Ladakh. The industry has given jobs to a very large number of people and generated economic activities in the primary, secondary and tertiary sectors in the State, which owes dependence to a large extent to this industry.

Jammu & Kashmir is full of colorful rocky mountains, valleys, trekking trail, several high rising chortens, snowy mountains, medows of flowers, lakes, garden and orchards, monuments, mandirs, forts and palaces. Ladakh is famous for its spectacular caves, monasteries, etc. while Jammu is famous for its temples/ religious tourism. Ladakh is the choice of foreign tourists; Jammu attracts pilgrim tourists, while the Valley attracts a larger number of domestic tourists and a fair sprinkling of foreign tourists.

There is a continuous increase in the total tourist's arrival including domestic and foreign in J&K. To handle such a traffic flow, the development of infrastructure has to match this number of tourists. The infrastructure requirements are accommodation, transportation and all essential facilities.

3. DATA AND METHOD

This paper is exclusively based on secondary data. The data used in this paper was collected from the official website of Incredible India [4] from the year 1989 to the year 2014. The analysis carried out in this paper is based on 25 years historic data to predict the future data for the next 5 years w.e.f 2015 to 2019 by using Time Series analysis.

4. TIME SERIES ANALYSIS

A time series is an ordered set of observations on a variable taken at regular intervals. The intervals can represent any unit of time that is hour, day, week, month or year, but the interval

must be same for all the measurements. Moreover, any interval for which there is no measurement must be set to missing value. The aim of time series analysis is to find out a pattern in the historical data and then extrapolate the pattern into the future [5].

Some of the general data patterns are Horizontal pattern, Trend Pattern, Seasonal pattern, cyclical Pattern etc. A horizontal pattern exists when the data fluctuate around constant mean. A trend pattern is a gradual upward or downward shift in the level of the series or the tendency of the series values to increase or decrease over time [6]. Trends can be local or global, linear or non linear. A seasonal pattern is repetitive, predictable pattern due to seasonal influence in the series value. If the time series plot shows an alternating sequence of points below and above the trend line lasting more than one year then the pattern that exists is called seasonal pattern [7].

In order to make forecasting based on the time series data, the time series node estimate Exponential Smoothing, Autoregressive Integrated Moving Average (ARIMA) models for time series data.

Exponential smoothing is a method of forecasting that uses weighted values of previous series observations to predict future value. It assigns the maximum weight to the most recent observation and the weight decline in a systematic manner as older and older observations are included [8]. This technique is useful for forecasting series that exhibit trend, seasonality or both. ARIMA model provide methods for modeling trend and seasonal components and allow independent (predictor) variable in the model. This involves explicitly specifying autoregressive and moving average orders as well as degree of differencing and the predictors being added helps us to explain the behavior of the series being forecast [6].

5. EXPONENTIAL SMOOTHING AND HOLTS LINEAR TREND

Exponential smoothing is one of the fixed-model time series forecasting techniques. It was originally called an "exponentially weighted moving average". This method of forecasting uses weighted average of past observations to predict future value. It assigns the maximum weight to the most recent observation and the weight decline in a systematic

manner as older and older observations are included [9]. This technique is useful for forecasting series that exhibit trend, seasonality or both.

The idea of exponential smoothing is to smooth the original series and use the smoothed series in forecasting future values of the variable of interest. This method is most effective when the parameters describing the time series are changing slowly over time.

Simple exponential (SES) model is widely used of all forecasting techniques which require little computation. It is used when data pattern is approximately horizontal (i.e., there is neither cyclic variation nor pronounced trend in historical data). The equation for the simple exponential smoothing method is: $F_{t+1} = \alpha Y_t + (1-\alpha) F_{t-1}$, where: F_{t+1} = the new smoothing value or the forecast value for the next year. α= the smoothing constant ($0 < a < 1$), Y_t = the new observation or actual value of the series in period t, F_t = the oldest smoothed value or forecast for period t [8].

Holt (1957) extended simple exponential smoothing to allow forecasting of data with a trend. If a time series is increasing or decreasing approximately at a fixed rate, then it may be described by the linear trend model. Holt's Linear Trend computes an evolving trend equation through the data using a special weighting function that places the greatest emphasis on the most recent time periods. Instead of the global trend equation of the least squares trend algorithm, this technique uses a local trend equation. The trend equation is modified from period to period and the forecasting equation also changes from period to period.

6. EXPERIMENT AND RESULT ANALYSIS

As previously mentioned, the historical data has been collected from the year 1989 to 2014 from the official website of Incredible India for forecasting the future values. The historical data set for J&K tourism gives the trends and seasonality patterns that help us to decide the accurate model for predicting the future values and thus helps the J&K tourism department to make better decisions to provide better facilities for the forthcoming tourists. The implementation works are done through IBM SPSS Modeler. The model to predict the tourists for J&K state for the next five years is shown in figure 1.
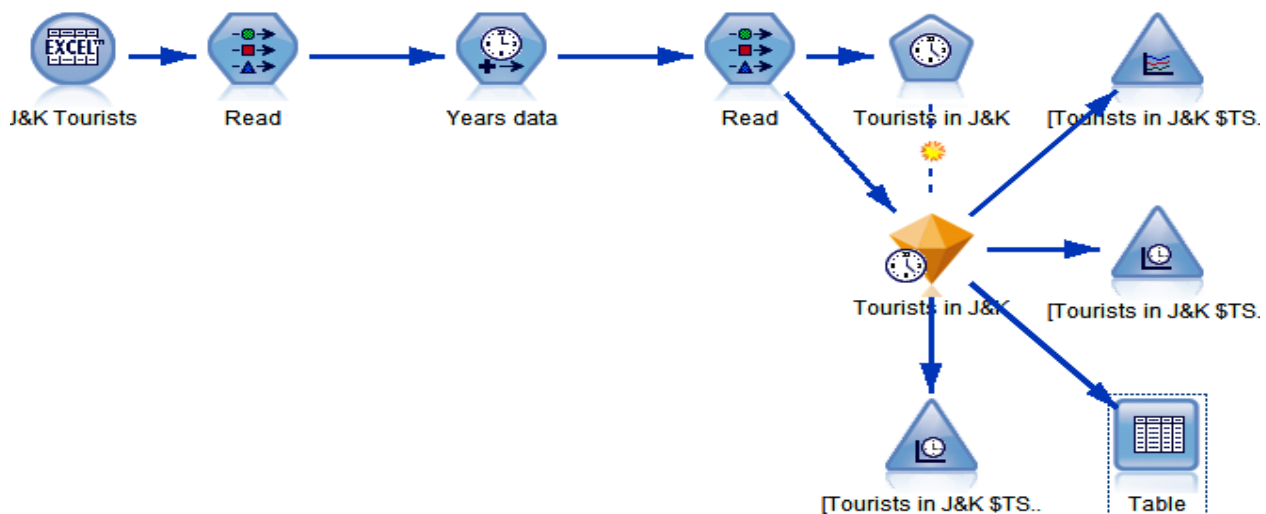

Figure 1: Model for predicting the tourists for J&K state for next 5 years.

The data has been taken from the year 1989 to 2014 and the actual and predicted values for tourists are shown with the help of time plot in figure 2. The time plot node allows viewing one or more time series plotted over time. The dots represent the historical data from the year 1989 to year 2014 and the line without dots represents the predicted values for the next five years from the year 2015 to year 2019.
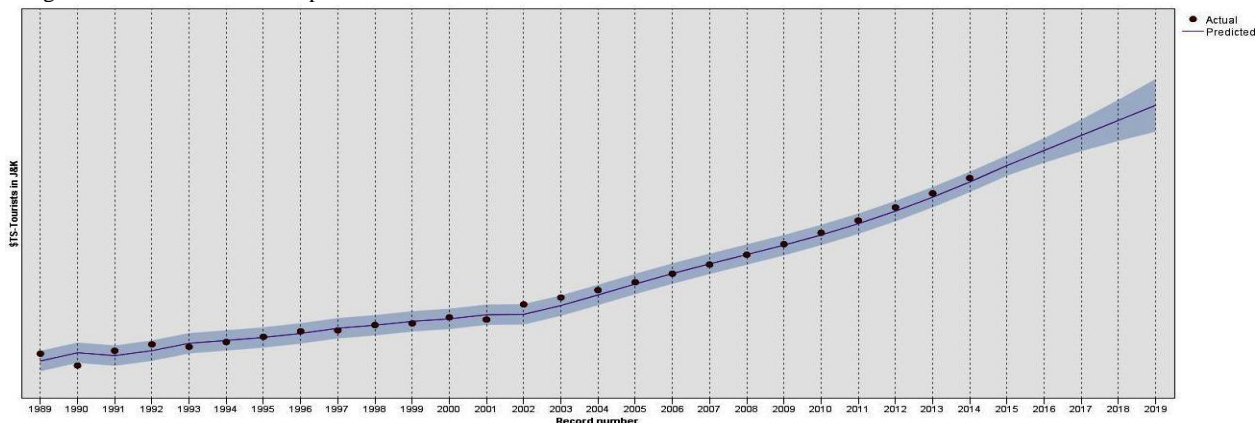

Figure 2: Time Plot selected time series models for tourists in J&K.

Figure 3 represents the actual tourists, predicted tourists, Lower Confidence Intervals (LCI) tourists and Upper Confidence Intervals (UCI) tourists using Time plot graph.
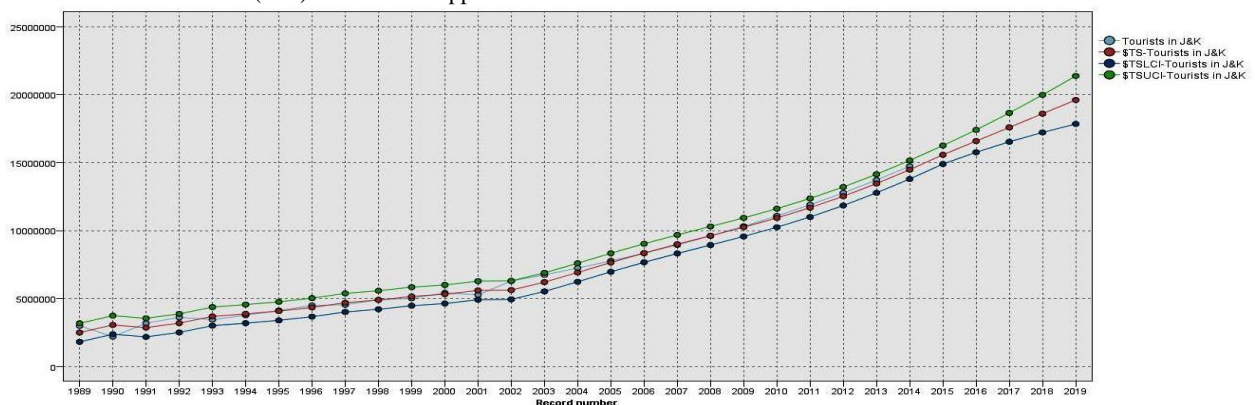

Figure 3: Time Plot representing tourists, TS-tourists, LCI-tourists and UCI-tourists.

Figure 4 represents historical tourists data and predicted tourists data, LCI-tourist data and UCI-tourist data using Multi plot graph. Multi plot is a special type of plot that displays multiple Y fields over a single X field. Multi plots are functional when there is time sequence data and would like to discover the variation of several variables over time.
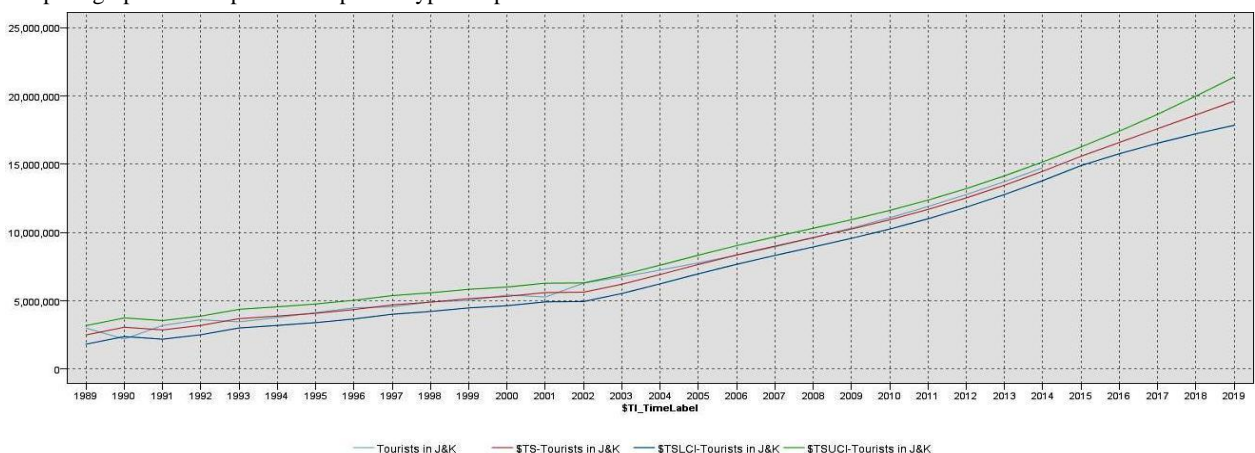

Figure 4: Multi Plot for tourists representing tourists, Predicted tourists, LCI-tourists and UCI-tourists.

Table - 1 represents the original tourists data from year 1989 to year 2014 and predicted tourist values, LCI-tourist values, UCI-tourist values, for the next five years from year 2015 to year 2019.

Table 1: Representation of historical data and predicted data of tourists with lower and upper limit

Year	Tourists	TI_TimeLabel	TI_Future	TS-Tourists	TSLCI-Tourists	TSUCI-Tourists
1989	2987964	1989	0	2494549.207	1811882.718	3177215.697
1990	2191486	1990	0	3052936.358	2370269.869	3735602.848
1991	3182251	1991	0	2859398.091	2176731.602	3542064.581
1992	3608269	1992	0	3188368.282	2505701.792	3871034.771
1993	3447162	1993	0	3683414.299	3000747.809	4366080.788
1994	3770208	1994	0	3870993.417	3188326.927	4553659.906
1995	4118632	1995	0	4075196.263	3392529.773	4757862.752
1996	4482072	1996	0	4344319.331	3661652.841	5026985.82
1997	4546200	1997	0	4690154.876	4007488.386	5372821.366
1998	4903930	1998	0	4890662.981	4207996.491	5573329.47
1999	5013571	1999	0	5151265.559	4468599.069	5833932.049
2000	5412863	2000	0	5315020.199	4632353.71	5997686.689
2001	5268246	2001	0	5593229.281	4910562.791	6275895.771
2002	6284029	2002	0	5619189.365	4936522.876	6301855.855
2003	6743282	2003	0	6205601.583	5522935.093	6888268.073
2004	7236634	2004	0	6915216.931	6232550.441	7597883.42
2005	7766684	2005	0	7648581.666	6965915.177	8331248.156
2006	8336228	2006	0	8346608.074	7663941.584	9029274.563
2007	8948294	2007	0	8995669.69	8313003.2	9678336.18
2008	9606150	2008	0	9616726.671	8934060.182	10299393.16
2009	10313328	2009	0	10247562.95	9564896.455	10930229.44
2010	11073650	2010	0	10926163.9	10243497.41	11608830.39
2011	11891251	2011	0	11680045.13	10997378.64	12362711.62
2012	12770613	2012	0	12522764.15	11840097.66	13205430.64
2013	13716589	2013	0	13456166.85	12773500.36	14138833.34
2014	14734445	2014	0	14475628.43	13792961.94	15158294.92
$null$	$null$	2015	1	15575601.17	14892934.68	16258267.66
$null$	$null$	2016	1	16583489.25	15762114.05	17404864.44
$null$	$null$	2017	1	17591377.33	16531005.45	18651749.22
$null$	$null$	2018	1	18599265.42	17218430.37	19980100.46
$null$	$null$	2019	1	19607153.5	17842759.61	21371547.39

7.   CONCLUSIONS

In this paper, with the use of time series data, the expected tourists in the next five years by using the historical data of 25 years w.e.f year 1989 to year 2014 with IBM SPSS Modeler is predicted. In addition to this, the lower limit and upper limit range of data is also predicted for the next five years from year 2015 to year 2019. Similarly, the future data can be predicted for more than five years based on the historical data. Data Mining is a key that works on the historical data to find out patterns for the future use. In this way, data mining recommends guarantee in serving the organizations to expose hidden patterns in the historical data. In this paper, we study the behaviour of time series data using IBM SPSS Modeler and predicted the future data for the tourists of J&K state.

8. FUTURE WORK

In the future, there is a option to predict the tourists of J&K state for more than five years and can also predict the tourists at national level and particular tourist place level. In addition to tourists prediction, there may be other indicators in different fields to predict the future values so that the concerned departments should prepare in advance for providing various facilities to their clients.

9. REFERENCES

[1] Jiawei Han, Micheline Kamber and Jian Pei, *Data Mining Concepts and Techniques*, 3rd ed., Waltham, USA: Elsevier-Morgan Kaufmann, 2014.

[2] The J&K Tourism website. [Online]. Available: http://www.jktourism.org/

[3] *How to Prepare J&K SSB (District/Divisional/State Cadre)*, Vishal International, 2014.

[4] The Incredible India website. [Online]. Available: http://www.incredibleindia.org

[5] (2012) The CENGAGE Learning website. [Online]. Available: http://www.cengage.com/

[6] *IBM SPSS Modeler 16 Modeling Nodes,* IBM.

[7] B. Uma Devi, D. Sundar and Dr. P. Alli, "An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50," *International Journal of Data Mining & Knowledge Management Process,* vol. 3, no. 1, pp. 65-78, January 2013.

[8] Pardeep Kumar Sahu and Rajesh Kumar, "Demand Forecasting For Sales of Milk Product (Paneer) in Chhatttisgarh," *International Journal of Inventive Engineering and Sciences*, vol. 1, no. 9, pp.10-13, August 2013.

[9] Eva Ostertagova and Oskar Ostertag, "The Simple Exponential Smoothing Model," in Proc. MMaMS, 2011, p.380.