

Predicting Pilgrimage in Numbers to Shri Mata Vaishno Devi, Katra, J&K using Time Series Analysis

¹Sourabh Shastri*, ²Anand Sharma, ³Vibhakar Mansotra

¹Department of Computer Science & IT, Bhaderwah Campus, University of Jammu, J&K, India

^{2,3}Department of Computer Science & IT, University of Jammu, J&K, India

Abstract-

Data mining is a technique that plays a major role to explore and analyze the data in order to find out valuable information from enormous quantity of data. In this paper, we predict the pilgrimage in numbers to Shri Mata Vaishno Devi, Katra, J&K based on the historical twenty nine years yearly time series data by using ARIMA model and IBM SPSS Modeler 16.0 data mining tool. ARIMA model is one of the popular forecasting methods that we use to predict the immediate future for time series data. The aim of this paper is to forecast the number of pilgrims in advance for Shri Mata Vaishno Devi, Katra, J&K so that the Shri Mata Vaishno Devi Shrine Board (SMVDSB) should be prepared in advance to provide vital amenities for the forthcoming pilgrims.

Keywords--- Data mining, Time series data, ARIMA, IBM SPSS Modeler, Pilgrimage

I. INTRODUCTION

Data mining refers to extracting or mining knowledge from large amounts of data. The data mining is a process that extracts a small set of precious nuggets from a great deal of raw material. Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. Many other terms carry a similar or slightly different meaning to data mining such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging [1]. The main objective of data mining is to improve the decision 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 the core of Knowledge discovery from data (KDD) process that uses algorithms to explore the data, develop the model and discover previously unknown patterns. KDD process consists of an iterative sequence of steps including Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation and Knowledge Presentation.

The applications of data mining are healthcare, tourism, insurance, sales, banking and many others. The challenge is to extract relevant knowledge from these data and act upon it in a timely manner. In this study, tourism field with special focus to pilgrimage has been chosen.

Forecasting is very important in making future decisions such as ordering replenishment for an inventory system or increasing the capacity of the available staff in order to meet expected future service delivery [2]. The tourism now a day's become an important industry that plays a vital role for the economy of any state. In our paper, we predict the number of pilgrims of Holy Shrine of Shri Mata Vaishno Devi, Katra (J&K) for the forthcoming years so that the Shri Mata Vaishno Devi Shrine Board (SMVDSB) adequately prepare in advance for providing better facilities to the pilgrims.

II. PRESENT SCENARIO OF PILGRIMS AT SMVD, KATRA, J&K

The Holy Cave is situated at an altitude of 5200 ft. The pilgrims have to undertake a trek of nearly 14.5 km from the base camp at Katra to Bhawan (Holy Cave). Darshans are open round the clock throughout the year. At the culmination of their pilgrimage, the pilgrims see three natural rock formations known as 'Pindies' inside the holy cave.

The Holy Cave attracts millions of devotees every year. The number of pilgrims visiting the Holy Shrine annually now exceeds ten million [3]. For the benefit of the pilgrims, Shri Mata Vaishno Devi Shrine Board provides clean, well maintained and very economical accommodation. This accommodation is available at Vaishnavi Dham, Kalka Dham and Sarswati Dham near Railway Station in Jammu and at Niharika Yatri Niwas, Shakti Bawan, and Ashirwad Bhawan at Katra near bus stand. Trikuta Bhawan, Katra a dormitory accommodation is also available near the 2nd Yatra counter (YRC 2). Accommodations are also available at Adhkuwari, Sanjichhat and Bhawan. Table 1 [3] shows main accommodations available for the pilgrims.

Table I Accommodations available for the pilgrims during their visit to SMVD, Katra

S.No	Location	Complex
1	Jammu	Vaishnavi Dham Sarswati Dham Kalka Dham
		Niharika

2	Katra	Shakti Bhawan Trikuta Bhawan
3	Adhkuwari	Ashirwad Bhawan Sharda Bhawan Shailputri Bhawan
4	Sanjichhat	Mangla Bhawan
5	Bhawan	Main Bhawan Complex Vaishnavi Bhawan Gouri bhawan

The shrine of Mata Vaishno Devi is one of the most venerated Shrines of India. Since 1986 by an act of legislation; Shrine board has been entrusted with the task of management and development of Shrine. The board has endeavored to provide best facilities to the devotees which have resulted in a tremendous boost to the tourism.

III. DATA AND METHOD

This paper to predict the pilgrims at Shri Mata Vaishno Devi, Katra, J&K is solely based on secondary data. The data we have used in this paper is collected from the official website of Shri Mata Vaishno Devi Shrine Board (SMVDSB) from the year 1986 to 2014 [3]. The analysis carried out in this paper is based on the previous 29 years past data for the prediction of next five years future data using Time Series analysis.

IV. TIME SERIES ANALYSIS

A time series is a sequence of observations on a variable measured at successive points in time or other successive periods of time. The measurements may be taken every hour, day, week, month, or year, or at any other regular interval. The pattern of data is an important factor in understanding how the time series has behaved in the past. If such behavior can be expected to continue in the future, we can use the past pattern to guide in selecting and appropriate forecasting method [4].

A time series does not use any mechanism to adopt its values and this makes it very different from others series. Common time series examples are stock market, weekly weather reports, annual precipitation or weekly pizza sales. Real world time series data tend to be continuous, and are usually a sequence of observations or values separated by equal time intervals. Time series predictions/forecasting are the process of studying known past events and extrapolating the results to predict future events.

The commonly used techniques on Time series data are: Exponential smoothing, Regression methods, Autoregressive integrated moving average (ARIMA), Threshold methods, Generalized autoregressive conditionally heteroskedastic (GARCH) etc. [5] In this paper, we use ARIMA method on time series data from 1986 to 2014 to predict the number of pilgrims of SMVD, Katra for the next five years from 2015 to 2019.

V. ARIMA METHOD

Exponential smoothing and ARIMA models are the two most commonly forecasting techniques for time series data and provide complementary approaches to the problem. Exponential smoothing models were based on a description of trend and seasonality in the data and ARIMA models aim to describe the autocorrelations in the data [6].

Autoregressive Integrated Moving Average (ARIMA) are the most general class of models for time series forecasting developed by Box and Jenkins in the 1970s and their approach of identification, estimation and diagnostics is based on principle of parsimony [7].

ARIMA predicts a value in a response time series as a linear combination of its own past values, past errors, and current and past values of other time series. The order of an ARIMA model is usually denoted by the notation ARIMA (p, d, q), where p is the order of the autoregressive part, d is the order of the differencing, q is the order of the moving-average process. Autoregressive orders identifies the previous values from the series that are used to predict current values. Difference identifies the order of differencing applied to the series before estimating models. Moving average orders identifies how deviates from the series mean for previous values are used to predict current values [8].

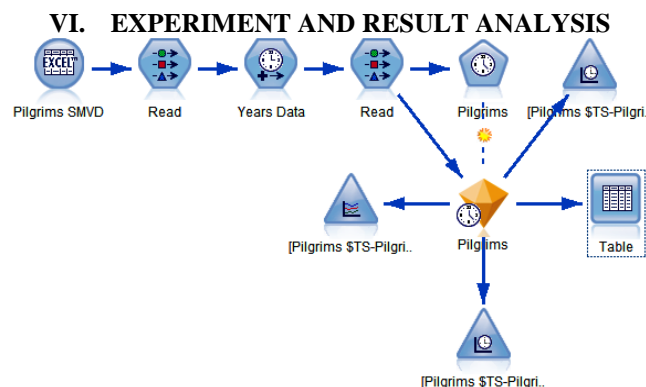


Figure 1: Data Stream for predicting the pilgrims at SMVD Katra for next 5 years.

As previously mentioned, the historical data has been collected from the year 1986 to 2014 from the official website of Shri Mata Vaishno Devi Shrine Board, Katra, J&K, India for forecasting the forthcoming pilgrims from 2015 to 2019. The historical data set for Shri Mata Vaishno Devi Pilgrims gives the trends and seasonality patterns that help us to decide the accurate model for predicting the future values and thus helps the SMVDSB to make better decisions to provide better facilities for the forthcoming pilgrims. The execution works are done through IBM SPSS Modeler. The data stream to predict the pilgrims at SMVD, Katra for the next five years is shown in figure 1.

The data has been taken from the year 1986 to 2014 and the actual and predicted values for pilgrims 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 1986 to year 2014 and the line without dots represents the predicted values for the next five years from the year 2015 to year 2019.

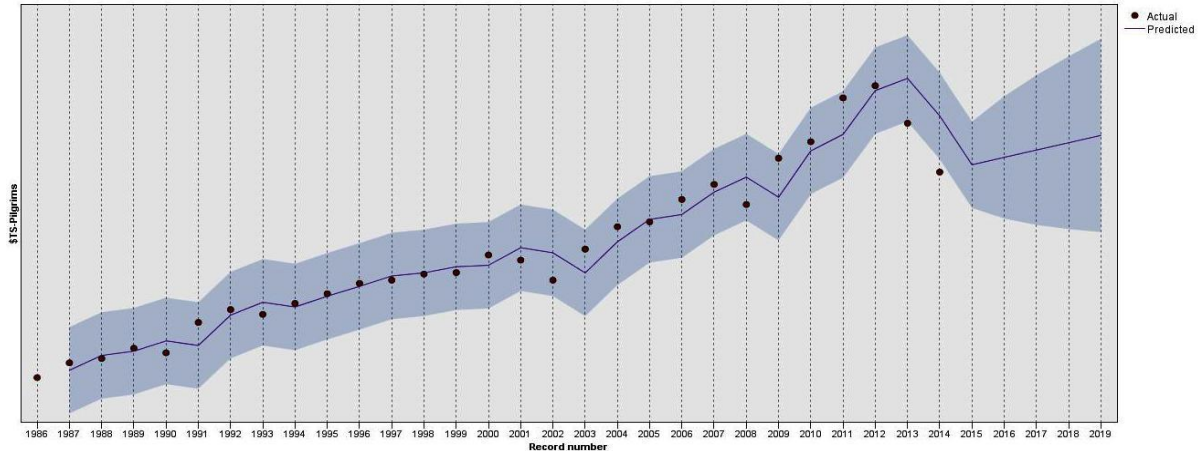


Figure 2: Time Plot selected time series models for pilgrims at SMVD Katra.

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

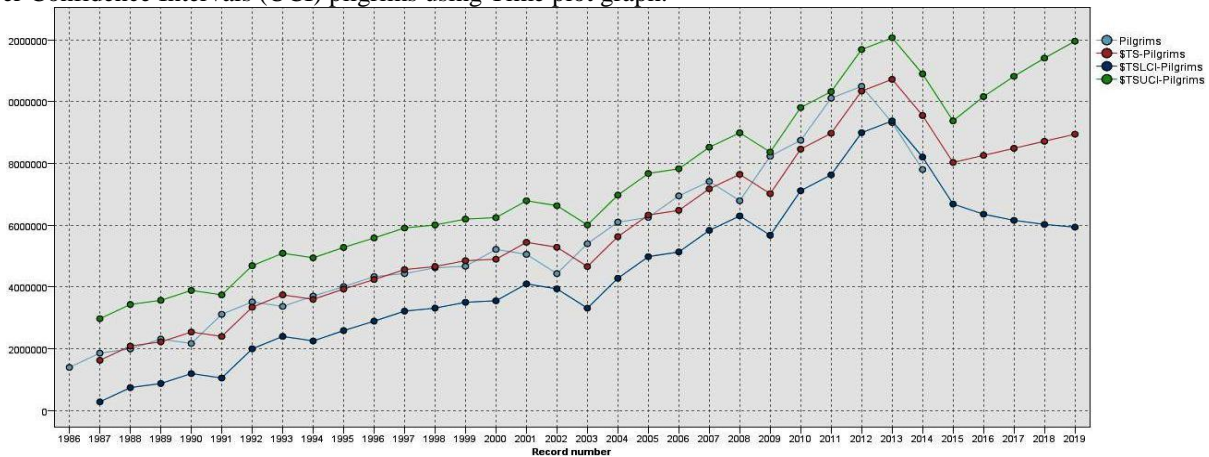


Figure 3: Time Plot representing pilgrims, TS-pilgrims, LCI-pilgrims and UCI-pilgrims.

Figure 4 represents historical pilgrim's data and predicted pilgrims data, LCI-pilgrims data and UCI-pilgrims 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 pilgrims representing pilgrims, Predicted pilgrims, LCI-pilgrims and UCI-pilgrims.

Table - 2 represents the original pilgrim's data from year 1986 to year 2014 and predicted pilgrims values, LCI-pilgrims values, UCI-pilgrims values, for the next five years from year 2015 to year 2019.

Table 2 Representation of historical data and predicted data of pilgrims with lower and upper limit

Year	Pilgrims	TimeLabel	Future	TS-Pilgrims	TSLCI-Pilgrims	TSUCI-Pilgrims
1986	1396000	1986	0	\$null\$	\$null\$	\$null\$
1987	1858000	1987	0	1624821.43	278336.231	2971306.626
1988	1992000	1988	0	2086821.43	740336.231	3433306.626
1989	2312000	1989	0	2220821.43	874336.231	3567306.626
1990	2169000	1990	0	2540821.43	1194336.231	3887306.626
1991	3115000	1991	0	2397821.43	1051336.231	3744306.626
1992	3516000	1992	0	3343821.43	1997336.231	4690306.626
1993	3369000	1993	0	3744821.43	2398336.231	5091306.626
1994	3705000	1994	0	3597821.43	2251336.231	4944306.626
1995	4011000	1995	0	3933821.43	2587336.231	5280306.626
1996	4335000	1996	0	4239821.43	2893336.231	5586306.626
1997	4434000	1997	0	4563821.43	3217336.231	5910306.626
1998	4622000	1998	0	4662821.43	3316336.231	6009306.626
1999	4670000	1999	0	4850821.43	3504336.231	6197306.626
2000	5217000	2000	0	4898821.43	3552336.231	6245306.626
2001	5057000	2001	0	5445821.43	4099336.231	6792306.626
2002	4432000	2002	0	5285821.43	3939336.231	6632306.626
2003	5400000	2003	0	4660821.43	3314336.231	6007306.626
2004	6100000	2004	0	5628821.43	4282336.231	6975306.626
2005	6252000	2005	0	6328821.43	4982336.231	7675306.626
2006	6950000	2006	0	6480821.43	5134336.231	7827306.626
2007	7417000	2007	0	7178821.43	5832336.231	8525306.626
2008	6792000	2008	0	7645821.43	6299336.231	8992306.626
2009	8235000	2009	0	7020821.43	5674336.231	8367306.626
2010	8749000	2010	0	8463821.43	7117336.231	9810306.626
2011	10115000	2011	0	8977821.43	7631336.231	10324306.63
2012	10495000	2012	0	10343821.4	8997336.231	11690306.63
2013	9324000	2013	0	10723821.4	9377336.231	12070306.63
2014	7803000	2014	0	9552821.43	8206336.231	10899306.63
\$null\$	\$null\$	2015	1	8031821.43	6685336.231	9378306.626
\$null\$	\$null\$	2016	1	8260642.86	6356425.229	10164860.49
\$null\$	\$null\$	2017	1	8489464.29	6157283.511	10821645.06
\$null\$	\$null\$	2018	1	8718285.71	6025315.319	11411256.11
\$null\$	\$null\$	2019	1	8947107.14	5936274.71	11957939.58

The number of pilgrims at Shri Mata Vaishno Devi, Katra, J&K are increasing year by year since the inception of Shri Mata Vaishno Devi Shrine Board (SMVDSB) in 1986. In the year 2013 and 2014, there was a decline in the number of pilgrims due to floods and elections in the state. From the table 2 mentioned above, it is obvious that in the forthcoming years from 2015 to 2019, there shall be an increase in the number of pilgrims. In the year 2015, there is a probability of 8031821.43 pilgrims and similarly in the year 2019 there are chances of 8947107.14 pilgrims to visit the Holy cave of Shri Mata Vaishno Devi, J&K.

VII. CONCLUSIONS

In this paper, with the use of Time series data, we predict the number of expected pilgrims in the next five years by using the historical data of 29 years w.e.f year 1986 to year 2014 with IBM SPSS Modeler. 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, we can predict the future data for more than five years based on the historical data. Data Mining is a way out to predict

results that are entirely based upon the historical data. So data mining offers guarantee to organizations to discover hidden patterns in their data. In this paper, we study the behavior of time series data using IBM SPSS Modeler and predicted the future data for the pilgrims at SMVD Katra of J&K state.

VIII. FUTURE WORK

In the future, we shall predict the pilgrims at SMVD Katra of J&K state for more than five years and shall also predict the pilgrims at other holy places. In addition to pilgrimage prediction, we shall choose other indicators in different fields to predict the future data so that the concerned organizations should prepare in advance to handle the forthcoming clients and for providing them various facilities.

REFERENCES

- [1] Jiawei Han, Micheline Kamber and Jian Pei, *Data Mining Concepts and Techniques*, 3rd ed., Waltham, USA: Elsevier-Morgan Kaufmann, 2014.
- [2] Albert Orwa Akuno et al., "Statistical Models for Forecasting Tourists' Arrival in Kenya," *Open Journal of Statistics*, vol.5, pp.60-65, Feb. 2015.
- [3] The SMVDSB website. [Online]. Available: <http://smvdsb.nic.in/>
- [4] (2012) The CENGAGE Learning website. [Online]. Available: <http://www.cengage.com/>
- [5] 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.
- [6] The OTexts website. [Online]. Available: <https://www.otexts.org/>
- [7] Elangbam Haridev Singh, "Forecasting Tourist Inflow in Bhutan using Seasonal ARIMA," *International Journal of Science and Research*, vol. 2, no. 9, Sept. 2013.
- [8] *IBM SPSS Modeler 16 Modeling Nodes*, IBM.