



Forecasting Public Healthcare Services in Jammu & Kashmir Using Time Series Data Mining

Anand Sharma*, Sourabh Shastri, Vibhakar Mansotra
Department of Computer Science & IT, University of Jammu,
Jammu and Kashmir, India

Abstract— In this paper, we have used time series data mining on public healthcare services data. We have specifically applied exponential smoothening methods in order to predict the future requirement of various types of services in the state of Jammu & Kashmir. Exponential Smoothening model is one of the most popular forecasting methods that we have used to predict the immediate future from time series data. The available data are monthly utilization of healthcare services in terms of OPD and IPD of J& K state from National – HMIS portal and cover the period 2010-15, i.e. five years. The various statistical parameters are also being used for evaluating forecasting models. Based on the accuracy of the forecasts, the most efficient model is taken into consideration for forecasting purpose. The aim of this paper is to forecast the number of patients in advance for Indoor Patient Department (IPD) and Outdoor Patient Department (OPD) so that the proper planning of these could be made to provide good facilities for the forthcoming patients.

Keywords—OPD, IPD, Time Series, HMIS, Data Mining, Exponential Smoothening, Winters Additive,

I. INTRODUCTION

Public healthcare agencies have a long history of forecasting the utilization of various healthcare services. Trends in utilization of services provide invaluable information for needs assessment, resource planning, facilities evaluation and policy formulations [1]. Forecasting also helps in making prediction about future requirements and rates of utilization of services. With the rapid application of ICT, the public healthcare facilities have at their disposal vast amounts of data which on thorough analysis can lead to more efficient decision making [2]. Thus, the challenge is to extract relevant knowledge from this data and act upon it in a timely manner. The research paper is focussed on the services provided to the patients in different public healthcare institutions like Sub Centres, Primary Health Centres (PHCs), Community Health Centres (CHCs), Sub-District/Sub-Divisional Hospitals and District Hospitals. The patients are mainly classified as either Indoor Patient (IPD) or Outdoor Patient (OPD). Various methods are being employed to predict the number of patients at particular facilities in advance so that the IPD and OPD could be well prepared in advance to provide good facilities to them.

To carry out such tasks, the time series analysis [3] is the best tool to forecast the trend. A time series requires a database that consists of sequence of values or events those changes with time [4]. In this paper, we focus on the monthly IPD and OPD data of Jammu & Kashmir state from April 2010-11 to February 2014-15. A forecasting for the next twelve months has been carried out so that the public healthcare institutions could prepare in advance to provide high-quality facilities to the patients.

II. PUBLIC HEALTHCARE IN J & K

Jammu & Kashmir state has 2,22,236 sq. km. geographical area and 10.14 million population. There are 22 districts, 107 blocks and 6652 villages in J&K state. The state has population density of 45 per sq. km. as against the national average of 312. The decadal growth rate of the state is 31.42% against 21.54% for the country and the population of the state continues to increase at a much faster rate than the national rate [5]. The Ministry of Health and Family Welfare (MoHFW), Government of India has implemented the population norms for all the public health facilities under the NRHM are as under in Table 1[6].

Table 1: Population Norms implemented by MoHFW.

Health Institution	Population Norm(GOI)		Norms considered for J&K
	Plain area	Hilly/Tribal area	
Sub-Centre	5,000	3,000	4,000
PHC	30,000	20,000	25,000
CHC	1,20,000	80,000	1,00,000

A wide series of indicators are there to measure the public health facilities in a state. In a state, there are different kinds of health institutions like Health Sub-Centres, Primary Health Centres (PHCs), Community Health

Centres (CHCs), Sub-District/Sub-Divisional Hospitals, District Hospitals and Medical College Hospitals etc. The list of health institutions in the J&K state are given in Table 2[5].

Table 2: List of Public Healthcare Institutions in J&K

PHIs	DHs	CHC	PHC	AD	MAC	SC	New SCs
Total	22	84	398	610	311	1679	826

The comparative facts of important health and demographic indicators of Jammu and Kashmir state vs. India are as follows in table 3[5].

Table 3: Health and Demographic Indicators of J&K Vs. India.

S.No	Indicator	Jammu and Kashmir	India
1	Total Population (In Crore) (Census 2011)	10.14	1028.61
2	Decadal Growth (%) (Census 2011)	31.42	21.54
3	Crude Birth Rate (SRS 2013)	17.5	21.4
4	Crude Death Rate (SRS 2013)	5.3	7
5	Natural Growth Rate (SRS 2013)	12.1	14.4
6	Infant Mortality Rate (SRS 2013)	37	40
7	Maternal Mortality Rate (SRS 2010-12)	NA	178
8	Total Fertility Rate (SRS 2012)	1.9	2.4

III. DATA

This paper is mainly based on secondary data. The data used in this paper were collected from NRHM portal from April 2010-11 to February 2014-15. The analysis carried out in this paper is based on the previous 5 years data to predict the data for the next 12 months by using Time Series analysis.

IV. METHODOLOGY

Many forecasting techniques use past or historical data in the form of time series [7]. A time series is a set of evenly spaced evenly spaced, continuous, numerical data obtained at regular time periods. In the time series forecasting methods, the forecast is based only on past values and assumes that factors that influence the past and the present will continue influence the future. Time series methods of modelling believe that history repeats itself. By analysing the historical data, time series exhibit the characteristics like trends, seasonal and nonseasonal cycles, pulses and steps, outliers etc. The time series modelling techniques like exponential smoothing and autoregressive integrated moving average (ARIMA) are used for time series data for the purpose of prediction [8]. The main objective of time series analysis is to find out a pattern in the historical data or time series data and then extrapolate the pattern into the future; the forecast is based exclusively on past values of the variable. A time series is a series of observations on a variable measured at successive periods of time. The measurements may be taken every hour, day, week, month, or year, or at any other regular interval [9].

Exponential Smoothing model is one of the most accepted forecasting methods that are used to forecast the future time for a time series data that have no obvious trend or seasonality [10]. Exponential smoothing uses weighted values of previous series observations to estimate future values. As such, exponential smoothing is not based on a theoretical understanding of the data. It forecasts one point at a time, adjusting its forecasts as new data comes in. The technique is useful for forecasting series that exhibit trend, seasonality, or both. There is a variety of exponential smoothing models that differ in their treatment of trend and seasonality [11]. Exponential Smoothing models are classified as either seasonal or nonseasonal. Seasonal models are only available if the periodicity defined using the Time Intervals node is seasonal. The seasonal periodicities are: cyclic periods, years, quarters, months, days per week, hours per day, minutes per day, and seconds per day. In Time Series Exponential Smoothing Criteria, the different model types are: Simple, Holt's Linear trend, Brown's Linear trend, Damped trend, Simple Seasonal, Winters' additive, Winters' multiplicative etc. In this paper, we choose Winters' additive model.

Winters' additive model is suitable for a sequence in which there is a linear trend and a seasonal effect that is constant over time. Its relevant smoothing parameters are level, trend, and season. Winters' additive exponential smoothing is most similar to an ARIMA with zero orders of auto regression; one order of differencing; one order of seasonal differencing; and p+1 orders of moving average, where p is the number of periods in a seasonal interval. For monthly data, p=12[11].

V. DATA ANALYSIS

The historical data has been collected from April 2010-11 to February 2014-15 from HMIS portal for forecasting the future values. The historical data set for IPD and OPD gives the trends and seasonality patterns that help us to decide the accurate model for predicting the future values and thus helps the IPD and OPD to make better decisions for the patients. The data analysis has been carried out by using IBM SPSS Modeler. The data stream to predict the patients for Outdoor Patient Department (OPD) for next four months is shown in Fig 1. The data file is in Excel format and the time interval is selected for 12 months for carrying out prediction of OPD patients services.

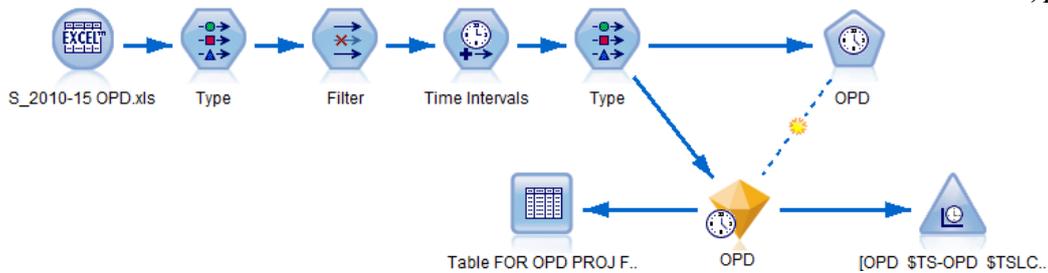


Fig. 1 Data Stream for predicting the patients for OPD for next 12 months.

The data stream to predict the patients for Indoor Patient Department (IPD) for next twelve months is shown in Fig. 2. The data file is in excel format and the time interval selected for prediction is 12 months. The statistical parameters are being evaluated later on.

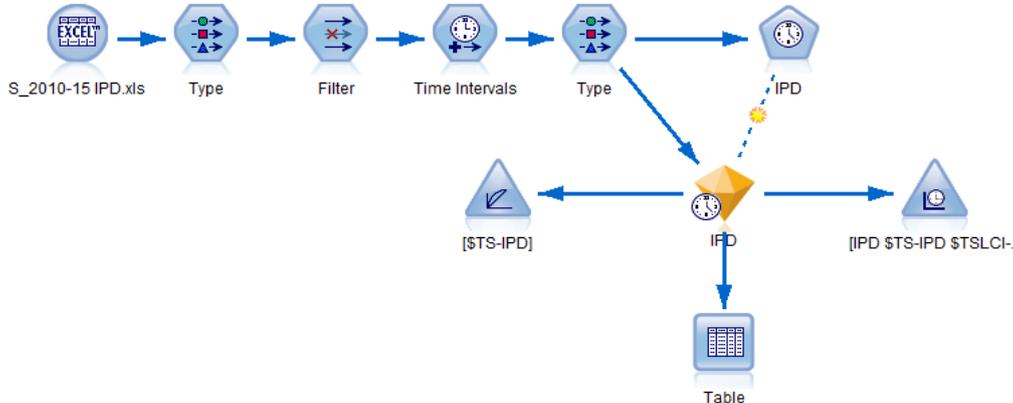


Fig. 2 Data Stream for predicting the patients for IPD for next 12 months.

The data has been taken from April 2010-11 to February 2014-15 and the actual and predicted values for OPD and IPD are shown with the help of time plot in figure 3 and figure 4 respectively. The time plot node allows viewing one or more time series plotted over time. The dots represent the historical data from April 2010-11 to February 2014-15 and the line represents the predicted values for the next twelve months.

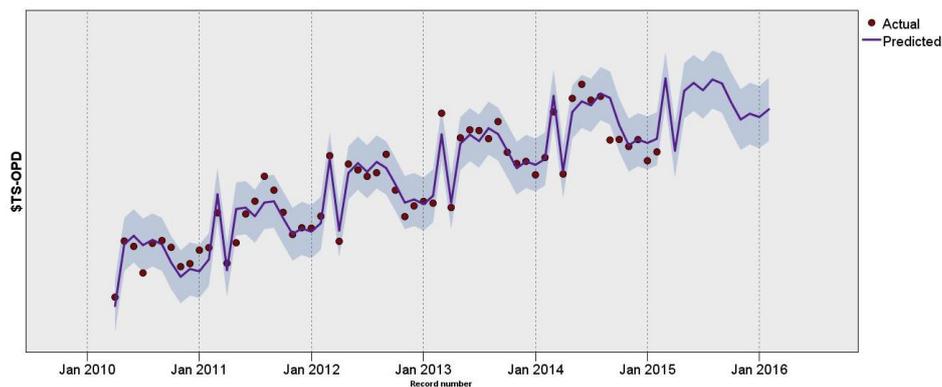


Fig. 3 Time Plot selected time series models for OPD.

The time plot shown above gives an indication that it is trending with the passage of time and also the number of OPD patients is also increasing.

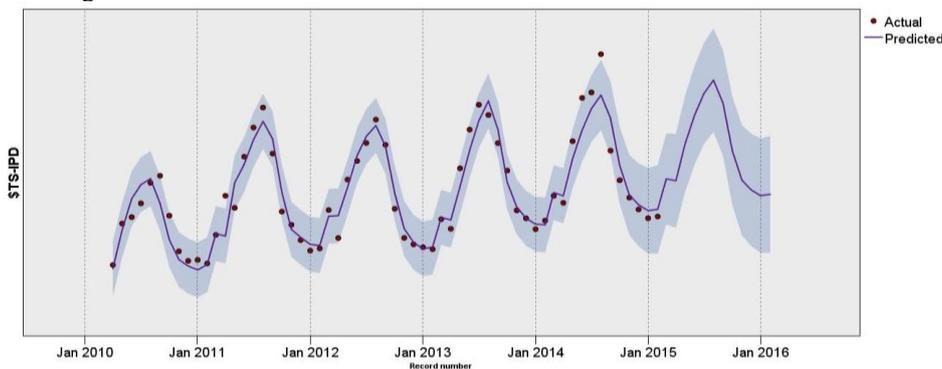


Fig. 4 Time Plot selected time series models for IPD.

The time plot shown in fig 4 provides a uniform pattern on account of seasonality which impacts the rise n fall of number of IPD patients.

Fig 5 represents the actual OPD values, predicted OPD values, Lower Confidence Intervals(LCI) OPD values and Upper Confidence Intervals(UCI) OPD values. The upper and lower limit provide a range of OPD patients that may increase with the passage of time depending upon the historical data. The graph takes into account both trends and seasonality following Winters Additive Exponential smoothening methods.

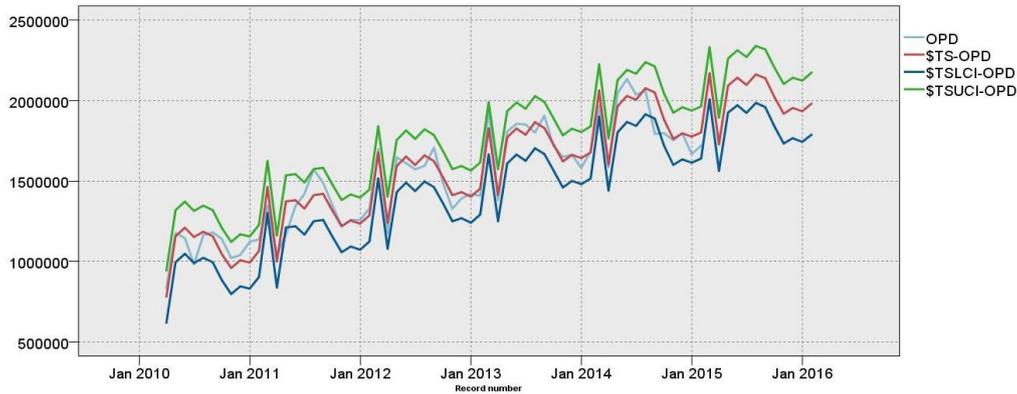


Fig. 5 Time Plot representing OPD, TS-OPD, LCL-OPD and UCL-OPD.

Figure 6 represents the IPD values, Predicted IPD values, Lower Confidence Intervals(LCI) IPD values and Upper Confidence Intervals(UCL) IPD values using Time plot graph. The IPD prediction provide a glimpse of seasonality index in prediction. The trends shows that July and August month has high number of IPD patients on account of rainy season . Such seasonal indications provide a helpful tool to the healthcare planner to plan the resources and facilities accordingly.

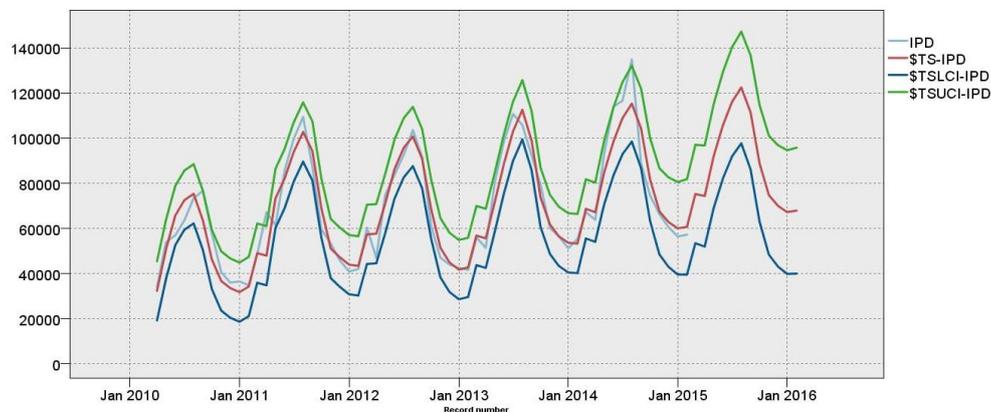


Fig. 6 Time Plot representing IPD, TS-IPD, LCL-IPD and UCL-IPD.

The Table 4 and Table 5 provides an account of predicted number of OPD and IPD patients for next complete year i.e. from Mar 2015 to Feb 2016. The upper and lower limit is also provided so that public healthcare institutions should provide minimal level of resources and facilities to the patients affected.

Table - 4 Predicted Opd Values with Lower & Upper Limit (From March 2015 to Feb 2016)

Forecasting Period	Predicted-OPD	LCI-OPD	UCI-OPD
Mar-2015	2169617	2007548	2331685
Apr-2015	1728879	1563747	1894010
May-2015	2093909	1925770	2262048
Jun-2015	2142326	1971232	2313420
Jul-2015	2097473	1923474	2271472
Aug-2015	2163133	1986276	2339989
Sep-2015	2139543	1959875	2319211
Oct-2015	2023960	1841523	2206397
Nov-2015	1918755	1733591	2103919
Dec-2015	1954613	1766761	2142465
Jan-2016	1933954	1743452	2124455
Feb-2016	1981194	1788079	2174310

Table – 5 Predicted Ipd Values with Lower & Upper Limit
(From March 2015 to Feb 2016)

Forecasting Period	Predicted –IPD	LCI-IPD	UCI-IPD
Mar-2015	75112	53294	96930
Apr-2015	74198	51757	96640
May-2015	91940	68891	114988
Jun-2015	105684	82045	129323
Jul-2015	115952	91736	140168
Aug-2015	122370	97591	147148
Sep-2015	111326	85996	136655
Oct-2015	88663	62795	114532
Nov-2015	74580	48184	100977
Dec-2015	69879	42965	96793
Jan-2016	67075	39653	94497
Feb-2016	67726	39806	95646

Evaluation of Models:

The forecasting models are evaluated on the basis of statistical parameters or goodness of measures. The Table 6 below shows a number of goodness-of-fit measures.

Table 6: Goodness –Of-Fit Measures

S.No.	Target	Model	Stationary R**2	R**2	MAPE	Norm. BIC	Ljung-Box		
							Q	df	Sig.
1	IPD	Winters additive	0.817	0.917	7.534	17.808	11.891	15	0.687
2	OPD	Winters additive	0.689	0.936	4.23	22.81	13.81	15	0.54

R**2 is the R-squared value, an estimation of the total variation in the time series that can be explained by the model. As the maximum value for this statistic is 1.0, both models are fine in this respect. The additional goodness-of-fit measure include the mean absolute percentage errors (MAPE). Absolute percentage error is a measure of how much a target series varies from its model-predicted level, expressed as a percentage value [12]. The MAPE value shows that all models display a mean uncertainty of less than 8%, which is very low. Interesting though these absolute values are, it is the values of the percentage errors (MAPE) which is more useful in this case, as the target series represent the healthcare services of various types. MAPE values represent an acceptable amount of uncertainty with the models. We have found that the goodness-of-fit statistics fall within acceptable bounds.

VI. CONCLUSIONS

In this paper, using time series data, we predict the number of expected patients in the next twelve months by using the historical data of 5 years w.e.f April 2010-11 to February 2014-15 with IBM SPSS Modeler. In addition to this, the lower interval and upper interval range of data is also predicted for the next twelve months from March 2015 to Feb 2016. Similarly, we can predict the future data for 2-5 years based on the historical data. Time series data mining is an integrated solution to forecast correct results that are totally based upon the accurate historical data. So time series data mining offers assurance in helping organizations to uncover hidden patterns in their data. In this paper, we study the behaviour of time series data using IBM SPSS Modeler and predicted the future data for the Indoor Patient Department and Outdoor Patient Department of J&K state. We can predict the patients of IPD and OPD for more than 2-5 years and shall also predict the patients at national level. In addition to IPD and OPD, we can choose other indicators in healthcare to predict the future trends so that the healthcare planners could prepare in advance for providing various facilities to the patients.

REFERENCES

- [1] Somnath R, "Primary Health Care in India", Health and Population- Perspectives & Issues 8(3): 135-167, 1985
- [2] Rajesh Kumar Sinha, " Impact of Health Information Technology in Public Health", *Sri Lanka Journal of Bio-Medical Informatics* 2010;1(4):223-36
- [3] 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.
- [4] Pushpalata Pujari and Jyoti Bala Gupta, "Exploiting Data Mining Techniques for Improving the Efficiency of Time Series data using SPSS-Clementine," *Journal of Arts, Science & Commerce*, vol. 3, issue 2(3), pp.69-80, April 2012.

- [5] The NRHM website. [Online]. Available: <http://nrhm.gov.in/>
- [6] The J&K NRHM website. [Online]. Available: <http://jknrhm.com/>
- [7] Pardeep Kumar Sahu and Rajesh Kumar, "Demand Forecasting for Sales of Milk Product (Paneer) in Chhattisgarh," *International Journal of Inventive Engineering and Sciences*, vol. 1, issue 9, August 2013.
- [8] Albert Orwa Akuno et al., "Statistical Models for Forecasting Tourists' Arrival in Kenya," *Open Journal of Statistics*, vol.5, pp.60-65, Feb. 2015.
- [9] (2012) The CENGAGE Learning website. [Online]. Available: <http://www.cengage.com/>
- [10] Eva Ostertagova and Oskar Ostertag, "The Simple Exponential Smoothing Model," in *Proc. MMaMS*, 2011, p.380.
- [11] IBM SPSS Modeler 16 Modeling Nodes, IBM.
- [12] Labib Arafah, "A Modified Neurofuzzy Based Quality of eLearning Model (Modified SCeLQM)", *International Journal of Computer and Information Technology*, Vol 03, No. 06, November 2014.