IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-4 E-ISSN: 2347-2693

A Study on Exponential Smoothing Method for Forecasting

Sourabh Shastri[1], Amardeep Sharma[2], Vibhakar Mansotra[3], Anand Sharma[4], Arun Singh Bhadwal[5], Monika Kumari[6]

[1]Dept. of CS&IT, Kathua Campus, University of Jammu, J&K, India
[2]Dept. of CS&IT, Kathua Campus, University of Jammu, J&K, India
[3]Dept. of CS&IT University of Jammu, J&K, India
[4]Dept. of IT & Research, UCCA, Guru Kashi University, Talwandi Sabo, Punjab
[5]Dept. of CS&IT, Ramnagar Campus, University of Jammu, J&K, India
[6]Dept. of CS&IT, Kathua Campus, University of Jammu, J&K, India

[*]*Corresponding Author: sourabhshastri@gmaill.com, Tel.: +919419171476*
Available online at: www.ijcseonline.org

Received: 20/Mar/2018, Revised: 28/Mar/2018, Accepted: 19/Apr/2018, Published: 30/Apr/2018

***Abstract*—**Data mining is one of the most essential steps of Knowledge Discovery process that is required to extract interesting patterns from enormous size of data. In this paper, we have used the BCG coverage data i.e. Percentage of live births who received Bacillus Calmette Guerin (BCG) a vaccine against tuberculosis and forecast the BCG coverage percentage for the next five years based on historical yearly data of BCG coverage in India by using the exponential smoothing technique of forecasting. Exponential Smoothing is a well-liked forecast technique that uses weighted values of previous series observations to predict the immediate future for time series data. The aim of this paper is to study the exponential smoothing method of time series for forecasting purpose.

***Keywords*---**Data Mining, BCG, Time Series data, Exponential Smoothing

I. INTRODUCTION

Organizations collect and store massive amounts of data in their databases. This archived data can be effectively used to find helpful patterns that can improve business. Data Mining is the non-trivial process of mining potentially useful information and discovery of previously unidentified patterns from vast amounts of operational data that should be based on some interestingness criteria. Once the patterns are found, they can additionally be used to build firm decisions for growth of the organizations. Data mining is a decision making process that includes algorithms like classification, association rules, clustering techniques etc. for knowledge extraction from databases. The aim of data mining is to classify valid, novel, potentially useful, and understandable correlations and patterns in data by combining through copious data sets to sniff out patterns that are too subtle or complex for humans to detect [1]. Data mining is a multi-disciplinary field and adopts its techniques from many research areas including statistics, database systems, machine learning, neural networks, rough sets, visualization etc.

Data Mining is the core part of Knowledge Discovery in Database (KDD) process. The KDD process consist of following steps: data selection, data cleaning, data transformation, data mining, finding presentation, finding interpretation and finding evaluation. Data mining and KDD are often used interchangeably because data mining is the key part of KDD process. Data mining consists of vital functional elements that transform data stored in multidimensional data warehouse, facilitates data access to analysts using application tools and techniques and meaningful presentation to managers for quick decisions and strategy achievement [2].

Data mining is popular with many businesses because it allows them to learn more about their customers and make elegant marketing decisions [3] and particularly become accepted in the fields of banking, insurance, marketing, sales, healthcare, forensic science, fraud analysis and many others [4]. As it reduces costs in time and money. Child immunization the core of healthcare field is the procedure of stimulating the body's immunity against certain infectious diseases by administering vaccines [5]. In the present paper, exponential smoothing technique of time series is applied on BCG coverage percentage for the purpose of forecasting.

II. PRESENT SCENARIO OF CHILD IMMUNIZATION IN INDIA

Immunization plays an important responsibility in the lives of children by protecting them against infectious diseases. As per WHO in 2015, 5.9 million children under age five and

696,000 neonatal deaths were recorded. Universal Immunization Programme (UIP) was launched in 1985 by Government of India to protect all infants against six diseases namely tuberculosis, diphtheria, pertussis, tetanus, poliomyelitis and measles.

Mission Indradhanush launched on December 25, 2014 by Ministry of Health and Family Welfare (MoHFW) as an agenda all over the country to vaccinate children who are partially vaccinated or unvaccinated. The main purpose of the mission Indradhanush is to fully immunized all children under the age of two years against seven life-threatening but vaccine-preventable diseases namely: Diphtheria, Pertussis (Whooping Cough), Tetanus, Tuberculosis, Polio, Hepatitis B and Measles which increased by only 1% a year from 2009 to 2013, from 61% to 65%. Besides, vaccines for Japanese Encephalitis (JE) and Haemophilus influenzae type B (HIB) are also being provided in selected states. The purpose of mission Indradhanush is to enhance full immunization coverage from 65% in 2013 to 90% children of the country in next five years [6, 7].
Various reasons of partially vaccinated or unvaccinated children are lack of awareness of immunization to parents, low education status of mothers, paper based vaccination records misplaced by the parents, unnoticed vaccination schedule due to hectic lifestyle of parents etc.

Regardless of all encouraging transformations, there are continuing challenges and deficiencies in the schemes launched by the Government. The vital need of the hour is to make sure that the promotion of complete vaccination is provided to all the children of the country.

III. PRESENT SCENARIO OF BACILLUS CALMETTE GUERIN (BCG) IN INDIA

Bacillus Calmette Guerin (BCG) is a vaccine given to infants which is economical, safe and easily available for the prevention of tuberculosis (TB) since 1921 [8]. The bacillus strain was isolated from bovine tuberculosis and attenuated by French scientists Calmette and Guerin [9]. Tuberculosis (TB) is a serious world-wide problem caused by a bacterium called mycobacterium tuberculosis [10]. The BCG vaccine is usually given to children as it has been shown to provide very good protection against the disseminated forms of TB in children including TB meningitis [11]. The World Health Organization (WHO) examines the estimated BCG coverage in every country. In India, BCG vaccination was introduced in 1948 but still India is the country with the highest burden of tuberculosis. According to World Health Organization (WHO) statistics 2015, 2.2 million cases of tuberculosis are estimated in India out of global occurrence of 9.6 million. In India, 1 dose of BCG is given at birth or up to 1 year if not given earlier. The goal of this study is to forecast BCG coverage in India for the forthcoming years.

IV. DATA AND METHOD

This paper is exclusively based on secondary data collected from Unicef global databases [12] from the year 1980 to the year 2014. The analysis carried out in this paper is based on thirty five years historic BCG coverage data to predict the future BCG coverage data for the next 5 years i.e. 2015 to 2019 by using exponential smoothing method of time series analysis.

V. EXPONENTIAL SMOOTHING METHOD

Exponential smoothing is one of the fixed-model time series prediction methods [13]. It was initially called as "exponentially weighted moving average". Exponential smoothing is a scheme of prediction that uses weighted values of earlier sequence observations to calculate future values. It allocates the utmost weight to the latest observation and the weight decline in an organized way as older and older observations are included [14].

The proposal of exponential smoothing is to smooth the original sequence and then make use of the smoothed sequence to predict upcoming values of the variable of concern [15]. This process is mainly helpful when the parameters relating the time series are varying gradually over time. Exponential smoothing method of prediction uses weighted average of past observations to predict upcoming values. This method is convenient for forecasting series that reveal trend, seasonality or both.

VI. EXPERIMENT AND RESULT ANALYSIS

The chronological data has been collected from the year 1980 to 2014 from Unicef global databases [12] for forecasting the future BCG coverage percentage in India. The historical data set for BCG coverage gives the trends and seasonality patterns that help us to decide the accurate model for predicting the future BCG coverage percentage and thus helps the child immunization wing of healthcare to make better decisions for providing better facilities for infants and their parents. The implementation is prepared through IBM SPSS Modeler. The model to predict the percentage of live births who will receive BCG vaccine against tuberculosis for the next five years is shown in figure 1.

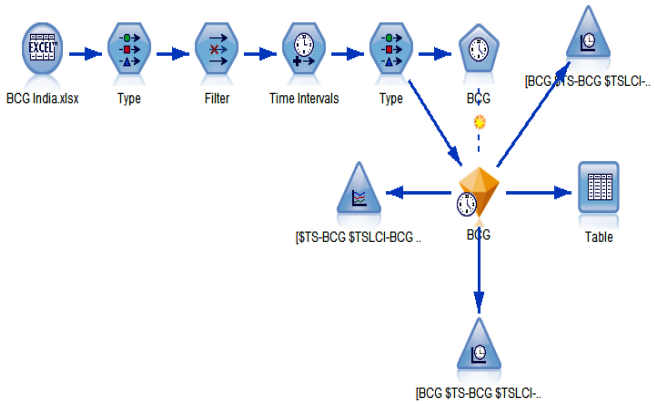

figure 1: Model for predicting the BCG coverage for next 5 years in India.

The data has been taken from the year 1980 to 2014 and the actual and predicted values for BCG coverage are shown with the help of time plot in figure 2. The dots represent the historical data from the year 1980 to year 2014 and the line without dots represents the predicted values for the next five years from the year 2015 to year 2019.

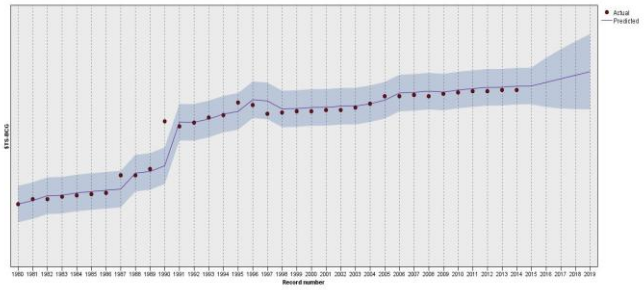

Figure 2: Time Plot for BCG coverage in India.

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

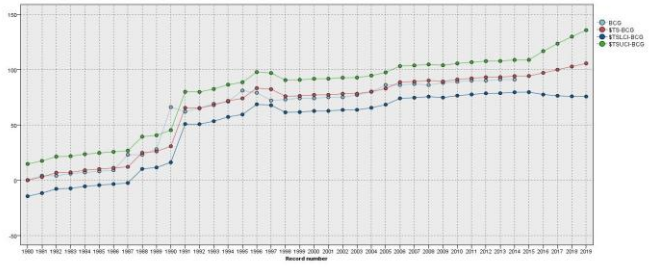

Figure 3: Time Plot representing BCG coverage, TS-BCG coverage, LCI-BCG coverage and UCI-BCG coverage.

Figure 4 represents historical BCG coverage data, predicted BCG coverage data, LCI-BCG coverage data and UCI-coverage data using Multi plot graph.

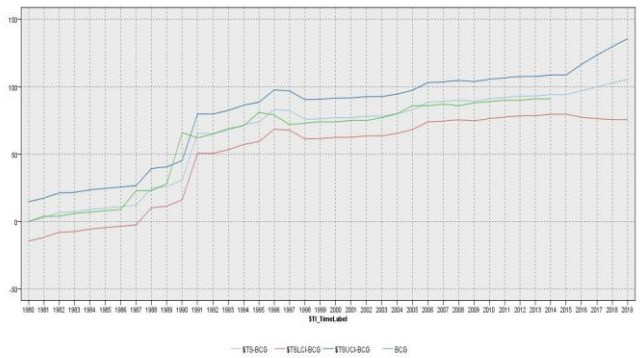

Figure 4: Multi Plot for BCG coverage representing BCG coverage, Predicted BCG coverage, LCI-BCG coverage and UCI-BCG coverage.

Table - 1 represents the original BCG coverage data from year 1980 to year 2014 and predicted BCG coverage values, LCI-BCG coverage values, UCI-BCG coverage values, for the next five years from year 2015 to year 2019.

Year	BCG	$TI_Year	$TI_Future	$TS-BCG	$TSLCI-BCG	$TSUCI-BCG
1980	0	1980	0	0.085	-14.484	14.653
1981	4	1981	0	2.865	-11.703	17.434
1982	4	1982	0	6.744	-7.825	21.312
1983	6	1983	0	7.131	-7.438	21.699
1984	7	1984	0	8.97	-5.599	23.538
1985	8	1985	0	10.053	-4.515	24.622
1986	9	1986	0	11.062	-3.507	25.63
1987	23	1987	0	12.062	-2.506	26.631
1988	23	1988	0	24.765	10.196	39.333
1989	28	1989	0	26.033	11.465	40.601
1990	66	1990	0	30.66	16.092	45.229
1991	62	1991	0	65.33	50.761	79.898
1992	65	1992	0	65.19	50.622	79.759
1993	69	1993	0	67.877	53.309	82.445
1994	71	1994	0	71.746	57.178	86.314
1995	81	1995	0	73.932	59.364	88.501
1996	79	1996	0	83.153	68.584	97.721
1997	72	1997	0	82.273	67.704	96.841
1998	73	1998	0	75.883	61.315	90.452
1999	74	1999	0	76.146	61.577	90.714
2000	74	2000	0	77.072	62.503	91.64
2001	75	2001	0	77.164	62.596	91.733
2002	75	2002	0	78.073	63.505	92.642
2003	77	2003	0	78.164	63.596	92.733
2004	80	2004	0	79.973	65.405	94.542
2005	86	2005	0	82.855	68.286	97.423
2006	86	2006	0	88.543	73.975	103.112
2007	87	2007	0	89.111	74.543	103.68
2008	86	2008	0	90.068	75.5	104.636
2009	88	2009	0	89.263	74.695	103.832
2010	89	2010	0	90.983	76.415	105.552
2011	90	2011	0	92.055	77.487	106.623

2012	90	2012	0	93.062	78.494	107.63
2013	91	2013	0	93.162	78.594	107.731
2014	91	2014	0	94.073	79.504	108.641
		2015	1	94.163	79.595	108.732
		2016	1	97.02	77.418	116.622
		2017	1	99.876	76.292	123.461
		2018	1	102.733	75.747	129.719
		2019	1	105.59	75.585	135.594

Table 1: Representation of historical data and predicted data of BCG coverage with lower and upper limit in India.

VII. CONCLUSION

In this paper, with the use of time series data, we predict the number of expected BCG coverage in India in the next five years by using the historical data of thirty five years w.e.f year 1980 to year 2014 with exponential smoothing. 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 process that examines large amounts of historical data for the discovery of new and meaningful patterns for the future use. In the nutshell, data mining assures in helping all kind of organizations to expose buried trends in the historical data. In this paper, we study the behaviour of historical data of thirty five years using exponential smoothing and predicted the future data of five years for the BCG coverage in India.

VIII. FUTURE WORK

In the future, we shall predict different vaccination coverage for more than five years and shall also apply other techniques of data mining to child immunization data. In addition to child immunization data, we shall choose other fields to apply data mining so that the concerned departments ought to use meaningful patterns for the benefit of their department and welfare of the society.

REFERENCES

[1] Manaswini Pradhan, "Data Mining & Health Care: Techniques of Application," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 12, pp.7445-7455, Dec. 2014.

[2] V. Priyavadana, A. Sivashankari and R. Senthil Kumar, "A Comparative Study of Data Mining Applications in Diagnosing Diseases," *International Research Journal of Engineering and Technology*, vol. 2, no. 7, pp. 1046-1053, Oct. 2015.

[3] Bharati M. Ramageri, "Data Mining Techniques and Applications," *Indian Journal of Computer Science and Engineering*, vol. 1, no. 4, pp. 301-305.

[4] Sourabh Shastri, Anand Sharma and Vibhakar Mansotra, "Predicting Pilgrimage in Numbers to Shri Mata Vaishno Devi, Katra, J&K using Time Series Analysis," *International Journal of Emerging Research in Management & Technology*, vol. 4, no. 10, pp. 102-106, Oct. 2015.

[5] Mohitulameen Ahmed Mustafi, "Factor Influencing of Child Immunization in Bangladesh," *International Journal of Mathematics and Statistics Studies*, vol. 1, no. 3, pp. 55-65, Sept. 2013.

[6] The NHP India website. [Online]. Available: http://www.nhp.gov.in/

[7] Sourabh Shastri, Anand Sharma and Prof. Vibhakar Mansotra, "Child Immunization Coverage - A Critical Review," *IOSR Journal of Computer Engineering*, vol. 18, no. 5, pp. 48-53, Oct. 2016.

[8] Michael Favorov et. al., "Comparative Tuberculosis (TB) Prevention Effectiveness in Children of Bacillus Calmette-Guerin (BCG) Vaccines from Different Sources, Kazakhstan," *PLoS One*, vol. 7, no. 3, pp. 1-8, Mar. 2012.

[9] Oli, A.N. et. al., "Potency and immunogenicity of bacillus calmette Guerin (BCG) vaccines used in routine immunization programme in South-East, Nigeria," *African Journal of Pharmacy and Pharmacology*, vol. 8, no. 47, pp. 1186-1191, Dec. 2014.

[10] Nagabhushanam D. et. al., "Prediction of Tuberculosis using Data Mining Techniques on Indian Patient's Data," International Journal of Computer Science and Technology, vol. 4, no. spl-4, pp. 262-265, Oct-Dec. 2013.

[11] The TBFACTS website. [Online]. Available: http://www.tbfacts.org/

[12] The UNICEF Global Databases website. [Online]. Available: http://www.data.unicef.org/

[13] Sourabh Shastri, Anand Sharma and Vibhakar Mansotra, "A Model for Forecasting Tourists Arrival in J&K, India," *International Journal of Computer Applications*, vol. 129, no. 15, pp. 32-36, Nov. 2015.

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

[15] Anand Sharma, Sourabh Shastri and Vibhakar Mansotra, "Forecasting Public Healthcare Services in Jammu & Kashmir Using Time Series Data Mining," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 12, pp. 570-575, Dec. 2015.

