

Mining Educational Data To Analyze The Performance Of Infrastructure Facilities Of Gujarat State

Divya R. Jariwala¹, Heta Desai², Samiksha H. Zaveri³

¹Research Scholar, JJT University, Chudela, District-Jhunjhunu, Rajasthan State, India

²Assistant Professor, S.D. Jain, Surat, Gujarat, India

³Research Scholar, Parul University, Baroda, Gujarat, India

Abstract

School plays major role in our society of Economics, Education and Environment. A School infrastructure represents good School. Good school infrastructure includes with building in good shape including an adequate number of well-organized classrooms, sufficient blackboards, access to adequate clean drinking water, common toilets, girl's toilets etc. The School which has got sophisticated and produces the good result then it's attracted many students. The schools were analyzed with reference to staff, infrastructure, amenities and achievements. The data was analyzed by using WEKA software. Educational officers to take appropriate steps to improve the quality of education in Gujarat State. An attempt has been made for broad mapping and analysis of existing infrastructures in the context of planning scheme in Gujarat State district wise, and to delineate the development zones of educational infrastructure facilities. The thematic layers considered in this study are infrastructure accessibility, type and condition of classroom and number of classroom allocated for the educational system at primary and upper primary level. Data mining has emerged as one of the major research domain in the recent decades in order to extract implicit and useful knowledge. Data Mining is an emerging technique with the help of this one can efficiently learn with historical data and use that knowledge for predicting future behavior of concern areas. The work aims to use data mining techniques which mines required information, so that the present education system may adopt this as a strategic management tool.

Index Terms—Education Data Mining (EDM), WEKA Tool, Classification Algorithm: Multilayer Perception.

I. INTRODUCTION

Education is a process which imparting or acquiring knowledge, developing the powers of reasoning and judgment and generally of preparing oneself or others intellectually for matured life. It makes an individual civilized, refined, cultured and educated and prepares them for battle of life so as to enable them to shoulder their responsibility towards the progress of the society and country [1]. The process of education based on large amount of category containing the infrastructure factors available to a school. School infrastructure, such as availability of classrooms, furniture, drinking water and equipment like blackboard etc. gives to a learning environment. Educational data mining is an interesting research area which extracts useful, previously unknown patterns from educational database for better understanding, improved educational performance and assessment of the student learning process [2].

A. Importance of data mining in Education

Data Mining is a process that consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data [3]. It is useful whenever a system is dealing with large data sets. In any education system, Infrastructure including i.e. Classrooms, Common toilets, girls toilets, drinking water and many equipment like blackboard, textbook and uniforms may be an important Consideration to analyze various trends since all the systems are now computer based information system so data Availability, modification and updating are a common process

now. [4]. It is data analysis methodology used to identify hidden patterns in a large data set. It has been successfully used in different areas including the educational environment. Educational data mining is an interesting research area which extracts useful, previously unknown patterns from educational database for better understanding, improved educational performance and assessment of the student learning process [5].

II. OBJECTIVES

- To study about the available infrastructure facilities in secondary schools
- To study about the condition of the schools infrastructure in secondary schools.
- To make a comparative study of the available infrastructure of Government and private schools.
- To provide some recommendation for developing the infrastructure in secondary schools.
- To find out the difference between government and private schools, rural and urban schools, run by the Gujarat State.

III. RELATED WORK

Here in the present research work classification data mining technique is applied on the training data set to identify the problem. The work is based on data classification by using only a multilayer perceptron algorithm from different classifier.

A. WEKA as Tool

Researchers select WEKA (Waikato Environment for

Knowledge Analysis) software that was developed at the University of Waikato in New Zealand. WEKA is open source software issued under the GNU General Public License. It contains tools for data preprocessing, classification, regression, clustering, association rules, and visualization. It is portable & platform independent because it is fully implemented in the Java programming language and thus runs on almost any modern computing platform and is now used in many different application areas, in particular for education & research [6]. The software WEKA is suitable because it is open source. WEKA can efficiently work with limited data [11]. WEKA also provides convenient data preprocessing, cleaning and handling missing values. It takes data from excel file in Comma Separated Values (CSV) format, which is a very common application software to be used in each school for initial collection of data [12]. This software contains tools for a whole range of data mining tasks like Data pre-processing, Classification, Clustering, Association and Visualization.

B. Data set Used

The data has been accessed by the site of DISE and NUEPA where the data is collected on annual basis. The researchers used 26 instances in WEKA as Data Mining Tool. Table 1 describes district wise data for the Gujarat state of the year 2008-2009, 2009-2010 and 2010-2011 respectively.

School Type	Primary	Primary with Upper primary	Primary with Upper Primary sec/higher sec.	Upper Primary Only	Upper Primary with sec./higher sec.
Government	11426	22844	53	741	20
Private	882	5391	735	265	96
Rural-Gov.	10727	20576	36	648	10
Rural-Pri.	505	2163	322	196	73

Table 1 District wise total no. of schools in year 2008-09, 2009-10 and 2010-11.

C. Preprocessing Tool

Data Preprocessing is required to make dataset appropriate for classification algorithms. Here numerical data sets are converted as nominal datasets. It is essential to have a suitable Data Mining tool for the purpose of carrying out Data Mining analysis of the available data.

IV. DESCRIPTION OF AREA

The area of the study is situated in Gujarat State. Gujarat State is situated on north axis 23.00 latitude and east axis 72.00 longitudes. It is a state in Western India, sometimes referred to as the Jewel of Western India [8]. It has an area of 196,204 km² (75,755 sq. mi) with a coastline of 1,600 km (990 mi) and a population in excess of 60 million. Its capital city is Gandhinagar, while its largest city is Ahmedabad. Gujarat is home to the Gujarati-speaking people of India [8]. It is also famous for jewelry and cotton cloths trading. Sex ratio is 918 females per 1,000 males [9]. Literacy rate is 79.31% as per census on 2011. Male literacy stands at 87.23% while female literacy is at 70.73% [10]. There are total number of 23540 total no. of primary schools, 50974 total no. of upper primary schools and 1164 total no. of higher secondary schools in the district of year 2011.



FIG. 1 Gujarat Population Map

V. VALUES OF THE STUDY

If the infrastructure of a school is strong then only the students can get education true sense and teachers can also give their full cooperation and efforts towards the education No

2008-2009					
School Type	Primary	Primary with Upper primary	Primary with Upper Primary sec/higher sec.	Upper Primary Only	Upper Primary with sec./higher sec.
Government	10524	22466	98	80	14
Private	838	3965	688	240	193
Rural-Gov.	9815	20204	75	29	13
Rural-Pri.	469	1697	316	170	137
2009-2010					
School Type	Primary	Primary with Upper primary	Primary with Upper Primary sec/higher sec.	Upper Primary Only	Upper Primary with sec./higher sec.
Government	10177	23076	88	75	9
Private	885	4555	716	248	109
Rural-Gov.	9508	20789	63	23	7
Rural-Pri.	497	1898	331	179	81
2010-2011					

doubt the Government is giving their full cooperation in developing the infrastructure of Government schools, but even after receiving such funds the committee members fails to prepare or give a good layout to their schools. The local people are also not aware to developing school infrastructure. The present condition of the infrastructure of Government schools is so neglected and poor that the schools are not developing at the rate they should develop. For this condition time has arrived to carry out a detailed study of the infrastructure of Government schools in comparison to the private schools of Gujarat. Therefore, the topic is selected to study the infrastructure conditions in schools of Gujarat district [13].

VI. METHODOLOGY

There are 26 instances used with 120 attributes in 2008-2009, 126 attributes in 2009-2010 and 118 attributes in 2010-2011. The present study area covers the primary, upper primary and higher secondary schools of Gujarat District, Western Province of India. There are numbers of schools were found in this area but the high schools are limited.

A. Multilayer Perceptron

Multilayer Perceptron (MLP) is a non-linear classifier based on the perceptron. The learning rule for the Multi-layer Perceptron is named as Back Propagation Rule also known as Generalized Delta Rule. MLP is a back propagation neural network with one or more layers. The following diagram illustrates a perceptron network with three layers.

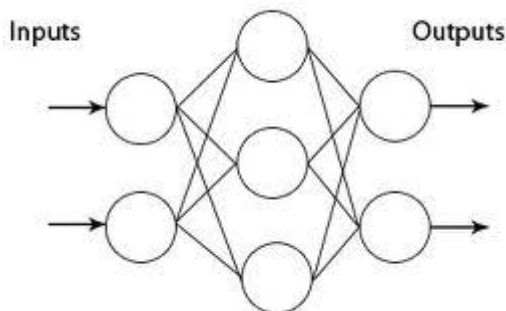


FIG. 2 A sample model of Multi-layer Perceptron network with three layers – One input, one output and one hidden layer in the middle.

In this model network, the neurons will be arranged into 3 or more layers which include an input layer, an output layer and one or more hidden layers. The back propagation rule continuously calculates a value based on an error function for each input and back propagates the error from one layer to the previous layer and the weights of the nodes are adjusted accordingly. MLP can be used to solve non-linear problems by connecting a number of neurons in the form of layers. Each of the perceptron identifies linearly separable inputs and the outputs of the perceptions are combined into a new perceptron to get the final output [14].

B. Data analysis

Descriptive statistics were used to describe the sample with reference to the variables taken for the study. Statements were converted in to simple statistical tables. In differential analysis the Significance of difference between groups was studied. Table 2 Represents attribute selection according to the year wise Minimum, Maximum, Mean and Standard Deviation data

are displayed.

Table 2 Year wise Minimum, Maximum, Mean and Standard Deviation.

Year	Minimum	Maximum	Mean	Standard Deviation
2008-2009	57	1275	404.769	346.826
2009-2010	39	1372	391.423	362.931
2010-2011	38	1915	439.462	468.18

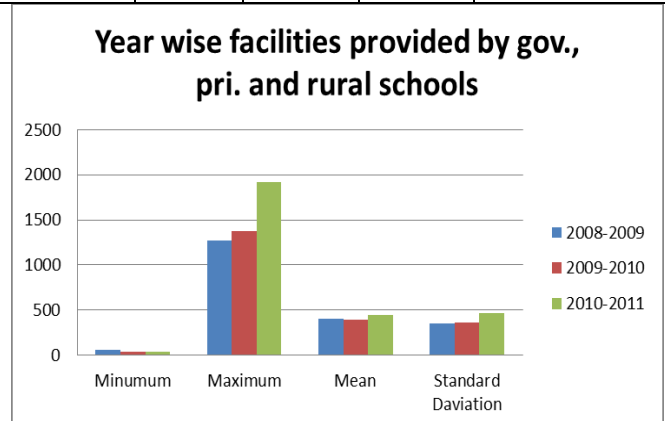


FIG. 3 Year wise Statistical equation applied on data set

C. Comparison of Government and Private Schools

The existing literature highlights the differential characteristics of state, private and rural schools with respect to both school inputs and indicators of school performance. Table 1 summarizes the differences between governments, private, rural government and rural private schools in our sample not only in school inputs, but also in school performance.

VII. RESULT AND ANALYSIS

The research work has chosen multilayer perceptron classifiers for comparative study and to analysis the performance of infrastructure facilities provided by Government, Private, Rural Government and Rural Private schools of Gujarat State as given in Table 3. Researchers have taken data set of year 2008-09, 2009-10 and 2010-11. It provides best result for year 2010-2011 among all these three years. Evolutionary infrastructure facilities have been increased year by year compared to previous year. We compared Government, Private and Rural School infrastructure facilities provided by Gujarat state. In Our research, we found that time taken to build model is 6.34 seconds, Co-relation coefficient is 0.9972, Mean absolute error is 181.6765, Root mean square error is 243.0331, relative absolute error is 9.374% and root relative error is 9.4667% which is the best result.

Year	Time taken to Build Model	Time taken to test module on training data	Correlation coefficient	Mean Absolute Error	Root Mean Square Error	Relative Absolute Error	Root Relative Square Error
2008-2009	6.95 Sec	0.03	0.9641	262.2388	344.682	29.7198%	30.1288%
2009-2010	7.09 Sec	0.02	0.8965	1214.0715	1413.2442	70.1012%	65.7712%
2010-2011	6.34 Sec	0.01	0.9972	181.6765	243.0331	9.374%	9.4567%

Table 3 classification techniques applied on three years dataset

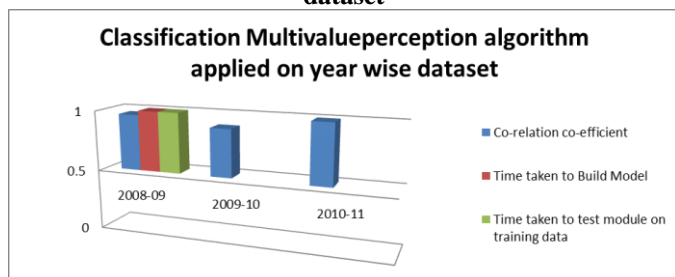


FIG. 4 classification techniques applied on dataset

VIII. CONCLUSION AND FUTURE SCOPE

The above discussion reveals that to analyze the availability of infrastructure facilities provided by government, private and rural schools, in which it determines multilayer perceptron classifier gives optimum solution for year 2010-2011. The government schools are getting more aids from the government to improve their infrastructure facilities, such as adequate numbers of classroom, drinking water, toilet, textbooks, uniforms and stationary, But their also needs more Government support as well as public support and awareness for developing schools. Because it is clear from all aspects that without proper infrastructure facilities none can realize the value of education.

However, our results recommend that the availability of infrastructure elements such as electricity, library, computers, playground, type and condition of classroom is very of great significance for improving the learning environment. As a process that produces specific functional products, EDM is fundamentally an educational micro-planning effort focused on increasing school resources efficiency and equity. This demand has a particular importance when it is connected with the area of education which is sensitive and important for the future progress of societies. Moreover, this study may also help the investors and funding bodies, as well as those who are responsible for planning, managing and designing educational

facilities to take necessary action in those areas suffering from the deficiencies of elementary school infrastructure in the near future.

REFERENCES

- [1] S. Tamilenthirai, K. Mohanasundaram and V. Padmini, "Staff, infrastructure, amenities and academic achievements of the high schools of Chipata District, Eastern Province of Zambia" Scholars research library Archives of Applied Science Research, 2011, 3 (6):131-140
- [2] M. Durairaj and C. Vijitha, "Educational Data mining for Prediction of Student Performance Using Clustering Algorithms", International Journal of Computer Science and Information Technologies, Vol. 5, No. 4, 2014, pp. 5984-5991.
- [3] Agnieszka Kubik-Komar and et. al., "application of decision tree algorithms FOR discriminatinG AMONG woody plant taxa BASED on the POLLEN season characteristics", Arch. Biol. Sci., Belgrade, vol. 67, No. 4, 2015, pp. 1127-1135.
- [4] Sonali Agarwal and et. al., "Data Mining in Education: Data Classification and Decision Tree Approach", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2011, pp. 140-144.
- [5] I. H. Witten and E. Frank, "Data mining: Practical Machine Learning Tools and Techniques," 2nd ed, Morgan-Kaufman Series of Data Management Systems San Francisco Elsevier, 2005.
- [6] S. Lakshmi Prabha and Dr.A.R.Mohamed Shanavas (2015) "Application of Educational Data mining techniques in e-Learning- A Case Study", International Journal of Computer Science and Information Technologies, Vol. 6, No. 5, pp.- 4440-4443.
- [7] Sonali Agarwal, G. N. Pandey, and M. D. Tiwari (2012) "Data Mining in Education: Data Classification and Decision Tree Approach", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, pp. 140-144.
- [8] <https://en.wikipedia.org/wiki/Gujarat>
- [9] https://www.google.co.in/?gfe_rd=cr&ei=dyQGV9eNB8308wfBvbzIDw&gws_rd=ssl#q=total+population+by+gender+in+gujarat
- [10] https://www.google.co.in/?gfe_rd=cr&ei=dyQGV9eNB8308wfBvbzIDw&gws_rd=ssl#q=total+literacy+rate+by+gender+in+gujarat
- [11] "WEKA Data Mining Book" (n.d.) <http://www.cs.waikato.ac.nz/~ml/weka/book.html>.
- [12] "WEKA 3: Data Mining Software in Java" (n.d.) Retrieved March 2010 from <http://www.cs.waikato.ac.nz/ml/weka/>.
- [13] Swapna Lahon, "A STUDY OF INFRASTRUCTURE FACILITIES IN SECONDARY SCHOOLS OF ASSAM STATE WITH SPECIAL REFERENCE TO SIVASAGAR DISTRICT", Abhinav National Monthly Refereed Journal of Research in Arts & Education, Volume 4, Issue 8, August, 2015, pp. 1-10.
- [14] Gopala Krishna Murthy Nookala and et. al., "Performance Analysis and Evaluation of Different Data Mining Algorithms used for Cancer Classification", International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No.5, 2013, pp. 49-55
- [15] Sarmistha Pal, "Public Infrastructure, Location of Private Schools and Primary School Attainment in an Emerging Economy", Public Infrastructure, Location of Private Schools and Primary School Attainment in an Emerging Economy, November 2009, pp. 1-37.