



# A SURVEY ON RURAL SCHOOL STUDENTS EDUCATION PERFORMANCE USING DATA MINING TECHNIQUES

Suresh Solomon. G<sup>1</sup>, Nancy Jasmine Goldina<sup>2</sup>

<sup>1</sup>Research Scholar, Manonmaniam Sundaranar University, Tirunelveli,

<sup>2</sup>Associate Professor/MCA, Sarah Tucker College, Tirunelveli

Article DOI: <https://doi.org/10.36713/epra8149>

DOI No: 10.36713/epra8149

## ABSTRACT

*In India there exists a lot of Rural areas in which the educational performance of the rural school students are inferior when compared it to the performance of the urban areas due to the lack of facilities, environment, income, employment opportunities and exposure. Equality is one among the basic principle of our country, so it's a mere responsibility of any research study to perform a detailed analysis towards the performance of rural school students by focusing on to the factors to be monitored and improved so that the Rural areas also raise to the equitant level of competition with the Urban areas. For this goal Data mining plays a vital role in order to handle the data in proper way for analysis and prediction of performances for the improvement of rural school student's education domain results. This paper presents a survey on Data Mining strategies used for prediction and performance analysis of rural school students education improvements.*

**KEYWORDS**—Data Mining, Rural, Urban, Prediction, Performance

## 1. INTRODUCTION

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing [7].



Fig-1: Data Mining Architecture

## 2. INDIAN EDUCATION COMPARISON

The figure 2 shows the rural school students performance is nearly half of the mark when compared it to the urban student's education performance during Jan 2016 [6].

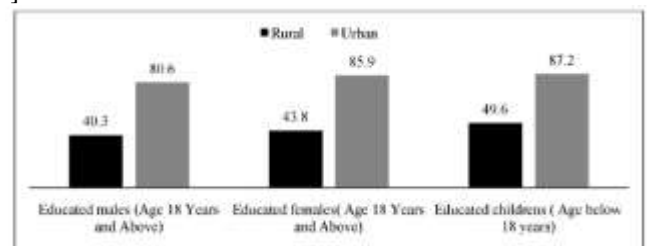


Fig-2: Sample Education comparison Jan 2016[6]

But after Jio mobile implementation nowadays the information technology details are reaching the rural people in a faster way and their impact will also so good in near future which may increase the motivation of learning and educational job opportunity awareness. May be the 2026 statistics will be an entirely appreciable one, if we perform a proper data mining approach towards their education improvements.

## 3. APPROACHES OF DATA MINING

- Data cleaning and preparation  
Education data contains lot of irrelevant and unusable information's which must be cleaned.
- Tracking patterns



The past and present performance can be traced by using the tracking patterns

- c) Classification  
The process of classifying the data as Boy, Girl, Rural and Urban are used for efficient analysis.
- d) Association  
The association rule is formed for the condition based grouping of educational environment of rural and urban students.
- e) Outlier detection  
The extra ordinary information's if needed for the educational analysis are taken into consideration.
- f) Clustering  
The grouping of same performance level students is used for further sub analysis and report formations.
- g) Regression  
Regression analysis is used for the parameter wise sub structuring of huge amount of school data students.
- h) Prediction  
This is the main approach for the further actions to be taken for the betterment of the results of school students in both the rural and urban areas.
- i) Sequential patterns  
Students are categorized as gifted, average and under achievers (normally we don't use the term worst), are gathered from the sequential data information patterns.
- j) Decision trees  
The tree formation is the optimal visible report of student's performance in education towards rural and urban variations.
- k) Statistical techniques  
The statistical mean, median and mode along with the analytical approaches play the vital role in the study of school student's education system.
- l) Visualization  
A single visual representation dominates the collection of table data in order to reduce time and space consumption.
- m) Neural networks  
The improvement process of rural area student education is entirely developed through the proper utilization of neural network approach training.
- n) Data warehousing  
The art of storing and recording the data for further processing can be easily handled by data warehousing.
- o) Long-term memory processing  
Historical data plays the vital role in the improvement of rural education by maintain the proper record of data.
- p) Machine learning and artificial intelligence  
These two advanced technologies are used for the effective improvement based education development through proper algorithmic approaches with needed hypothesis [8].

#### 4. LITERATURE SURVEY

In the field of education techniques "Data Mining" has also been used to analyze the curriculum and subject of the current research topics, as well as to analyze the student's performance [1]. There have been several investigations made under this proposed study object. For example, Bhardwaj used the Naïve Bayes algorithm to predict student performance based on 13 variables [2]. The results were used to build a model that is used to predefine the students who are at risk of failure and thus activate a guidance and counseling program. Varghese, Tommy and Jacob [3] in their research used the "K means" algorithm to cluster 8000 students based on five variables (input average in the University average scores of the tests / exams, average scores of papers, seminars notes and notes the work by frequency). The results showed a strong relationship between attendance and student performance. Gulati and Sharma [4] claim that knowledge through analysis by "Data Mining" can improve the education system in orientation, student performance and organizations management. Ayesha Mustafa [5] directed a study on evaluation, taking into account the evolution of learning and analysis of tests at the beginning and end of the courses.

#### 5. ISSUES IN DATA MINING IN EDUCATION

- ✓ Unstructured data prediction
- ✓ Education Facility wise comparison searches
- ✓ Multiple alignment and reports construction
- ✓ Graduate sequence analysis and Illiterates findings.
- ✓ Rural urban overlay chart preparations.

#### 6. IMPLEMENTATION IN RURAL URBAN AREAS

- ❖ Study and perform background analysis of locality and students family
- ❖ Investigate integral functional strategies for education environment
- ❖ Bridge the gap between computationally feasible and functionally relevant time scales
- ❖ Improve multiresolution state prediction.
- ❖ Combine classical and modern techniques in education environment.
- ❖ Sample larger sets of dynamical educational events and result matching's
- ❖ Realize interactive modeling
- ❖ Foster the development of data mining based education model.
- ❖ Train computational based practices in learning environments with physical concepts
- ❖ Bring experimental and computational groups in Rural and Urban closer together.

#### 7. TOOLS PERFORMANCE

The following table illustrates Data Mining tools with its implementation scope for the betterment of education in Rural and Urban areas



**Table-2: Tools for Data mining in education**

Tool Name	Description
<u>Rapid Miner</u>	A data science software platform providing an integrated environment for various stages of data modeling
<u>Oracle Data Mining</u>	Oracle, the world leader in database software, combines it's prowess in database technologies with Analytical tools
<u>IBM SPSS Modeler</u>	Modeler is a visual data science and machine learning solution, helping in shortening the time to value by speeding up operational tasks for data scientists.
<u>Knime</u>	Konstanz Information Miner is an open-source data analysis platform, helping you with build, deployment and scale in no time.
<u>Python</u>	Python is a freely available and open-source language that is known to have a quick learning curve.
<u>Orange</u>	Orange is a machine learning and data science suite, using python scripting and visual programming featuring interactive data analysis.
<u>Kaggle</u>	Kaggle although started as a platform for machine learning competitions, is now extending its footprint into the public cloud-based data science platform arena.
<u>Rattle</u>	The rattle is an R language based GUI tool for data mining requirements. The tool is free and open-source and can be used to get statistical and visual summaries of data.
<u>Weka</u>	Waikato Environment for Knowledge Analysis (Weka) is a suite of machine learning tools written in Java.
<u>Teradata</u>	A cloud data analytics platform marketing its no code required tools in a comprehensive package offering enterprise-scale solutions.

## 8. RECENT TRENDS

### i. Multimedia Data Mining

This is one of the latest methods which are catching up because of the growing ability to capture useful data accurately. It involves the extraction of data from different kinds of multimedia sources such as audio, text, hypertext, video, images, etc. and the data is converted into a numerical representation in different formats. This method can be used in clustering and classifications, performing similarity checks, and also to identify associations.

### ii. Ubiquitous Data Mining

This method involves the mining of data from mobile devices to get information about individuals. In spite of having several challenges in this type such as complexity, privacy, cost, etc.

this method has a lot of opportunities to be enormous in various industries especially in studying human-computer interactions.

### iii. Distributed Data Mining

This type of data mining is gaining popularity as it involves the mining of huge amount of information stored in different company locations or at different organizations. Highly sophisticated algorithms are used to extract data from different locations and provide proper insights and reports based upon them [11].

### iv. Spatial and Geographical Data Mining

This is new trending type of data mining which includes extracting information from environmental, astronomical, and geographical data which also includes images taken from outer space. This type of data mining can reveal various aspects such as distance and topology which is mainly used in geographic information systems and other navigation applications [9].

### v. Time Series and Sequence Data Mining

The primary application of this type of data mining is study of cyclical and seasonal trends. This practice is also helpful in analyzing even random events which occur outside the normal series of events. This method is mainly being use by retail companies to access customer's buying patterns and their behaviors [10].

## 9. APPLICATIONS

- Future Healthcare
- Market Basket Analysis.
- Manufacturing Engineering.
- Customer Relationship management
- Fraud Detection.
- Intrusion Detection.
- Customer Segmentation.
- Financial Banking.
- Lie detection
- Corporate Surveillance
- Research Analysis
- Criminal investigation
- Bioinformatics

## 10. CONCLUSION

The detailed research survey in the field of data mining domain towards educational performance improvement in rural areas when compared to the urban education development shows that the higher level of impact in the field of data mining approaches with the cope up towards latest trends and systematic pathways for the improvement progress of several advanced strategies. The approaches for data mining dealt with the various levels of implications towards the selection strategies for the analysis and prediction of specific key areas in rural education focusing along with the factors in the development of educational environment for the proper teaching learning situation implementations. The tools performance and applications of data mining provides the several directions for the development of different



methodologies to implement the rural education domain in several real-time domains. In future this research will lead the direction of data mining application towards the educational upliftment of several rural areas in India.

## REFERENCES

1. Permata Alfiani and F. Ayu Wulandari, "Mapping Student's Performance Based on Data Mining Approach (A Case Study)," *Ital. Oral Surg.*, vol. 3, pp. 173–177, 2015.
2. V. Kumar, "An Empirical Study of the Applications of Data Mining Techniques in Higher Education," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 3, pp. 80–84, 2011.
3. Varghese, Bindiya M; Tomy, Jose; Unnikrishnan, A; Poulose, "Clustering Student Data to Characterize Performance Patterns," *Int. J. Adv. Comput. Sci. Appl.*, pp. 138–140, 2010.
4. P. Gulati and A. Sharma, "Educational data mining for improving educational quality," *Int. J. Comput. Sci. Inf. Technol. Secur.*, vol. 2, no. 3, pp. 648–650, 2012.
5. S. Ayesha, T. Mustafa, A. Raza Sattar, and M. I. Khan, "Data mining model for higher education system," *Eur. J. Sci. Res.*, vol. 43, no. 1, pp. 24–29, 2010.
6. Shubini A saraf, Yogesh, Sumit, Venkatesh et al " study of factors affecting rural and urban Healthcare system using probit model"- Bridging the SciencePolicy Gap for Inclusive Growth in India-Jan-2016.
7. L. Tang and H. Liu, "Toward predicting collective behavior via social dimension extraction," *IEEE Intelligent Systems*, vol. 25, pp. 19–25, 2013.
8. Baker, Ryan. "Data Mining for education", oxford, UK: Elsevier. Retrieved 9 February 2014.
9. L. Tang and H. Liu, "Scalable learning of collective behavior based on sparse social dimensions," in *CIKM '09: Proceeding of the 18th ACM conference on Information and knowledge management*. New York, NY, USA: ACM, 2015, pp. 1107–1116.
10. Romero, S. Ventura. Educational Data Mining: A Review of the State-of-the-Art. *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. 40(6), 601-618, 2016.
11. Alisa Bilal Zoric, "Benefits of educational data mining", *Journal of international business research and marketing*, Vol-6, Issue-1, Nov-2020.