JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Study of Educational Data Mining for Predicting Student's Performance in Higher Education

**Birinder Singh Sarao**

Assistant Professor, Mata Gujri College, Fatehgarh Sahib

birsarao@gmail.com,

## Abstract

As there has been a tremendous growth in the field of higher education where various educational institutes have set up from last five to seven years. So, the possible impact of Data Mining analytics is the most emerging trend in higher education. Data mining in education is an interdisciplinary approach which focuses on student's academic performance and learning behaviour. Educational data mining plays a vital role in not only providing inherent knowledge about education system but also stake holders of education sector. This knowledge is further used in enhancing learning process and decision making. This paper discusses the role of Educational Data Mining(EDM) tools and techniques used in it and also the application of various Data mining tools.

**Keywords:** Data mining, Educational Data Mining, EDM Techniques, EDM Tools

## 1. Introduction

Educational Data Mining (EDM) is a growing field that brings together experts from different areas to explore and analyse the wealth of data generated in educational settings. The goal is to better understand how students learn and to improve educational experiences by uncovering patterns and insights from this data [1]. By doing so, educators can make informed decisions that enhance learning outcomes and tailor teaching methods to meet the needs of each student.

Despite a lot progress over many years, Indian higher education system faces many challenges like:

**A gap in supply and demand**: In India, students rate of enrolment in higher studies is very low i.e. it is 20% approximately as compared to other countries. Like in Punjab, most of the youth migrate to other countries after graduation. By 2025, Government plans to achieve 40% gross enrolment.

**Faculty Shortage**: Recruitment of faculty is not on regular basis. Part time faculty is appointed on temporary basis.

**Poor quality teaching methods:** It gives negative impact on student's learning process. All this results due to poor classroom management, inadequate resources and not use of latest teaching methodology.

All these challenges have to be handled in order to result in the growth of education in India. Many techniques and tools are available that are helpful in the improvement of education system in India [2,3].

## 2. Educational Data Mining Techniques

Educational Data Mining uses various techniques for analysis, data processing, to identify patterns, model construction and predicting results. Some of the techniques are:

**2.1. Statistics and Visualization**: Statistics and Visualization are crucial tools in Education Data Mining (EDM) that make it easier to understand and use the wealth of data generated in educational settings.

- **Statistics** is about using numbers to tell a story. By applying statistical methods, educators can spot trends in how students are performing, figure out what factors are influencing learning, and assess whether certain teaching strategies are actually working [4].
- **Visualization** is all about turning data into pictures. Charts, graphs, heat maps, and dashboards transform complex data into clear visuals that highlight key patterns and relationships. This makes it easier for teachers and administrators to see what's happening and make decisions based on real insights [3,4].

When combined, statistics and visualization help turn raw educational data into actionable knowledge, making it possible to improve how we teach and how students learn.

**2.2. Prediction:** One of the most commonly used techniques for making predictions in Education Data Mining is **regression analysis**. This method helps to understand and forecast outcomes by analysing the relationship between different factors, known as predictor variables [5].

In simple terms, regression looks at how one or more known factors (like a student's domain knowledge or communication skills) can be used to predict an unknown outcome (such as their chances of getting placed in a job). For example, if we know a student's level of expertise in their field and how well they communicate, we can use regression to estimate their likelihood of securing a job placement.

In this context, the placement possibility is what we call the **dependent variable** (denoted by **y**), because it depends on other factors. The domain knowledge and communication skills are **independent variables** (denoted by **x**), because they influence the outcome but are not influenced by it [6,7].
So, regression analysis helps us connect the dots between what we know and what we want to predict, making it a powerful tool for decision-making in education. [6]

**2.3. Classification** is a way of sorting data into different categories based on certain traits or characteristics. Think of it as a process where a model, known as a **classifier**, helps predict which group a new piece of data should belong to. For example, it might categorize students based on their performance levels or identify at-risk students who might need extra support [8].

In machine learning, there are many methods for classification. Traditionally, these methods work best with smaller datasets that can be easily processed in memory. But when it comes to handling large datasets, they can run into challenges.

This is where data mining classification techniques shine. They're built to scale, meaning they can handle large amounts of data stored on disk without getting bogged down. This scalability makes them ideal for analysing vast amounts of educational data.

In Education Data Mining (EDM), classification plays a crucial role. It helps educators and administrators sort through educational data to find patterns, make informed decisions, and ultimately improve learning outcomes. Since classification is a type of **supervised learning**, it involves training a model using data where the categories are already known, so that the model can accurately classify new, unseen data [9]

## 3. Education Data Mining Techniques used for improving Student's Performance

**3.1.Prediction:** Most commonly used prediction technique is regression analysis. It consists of one or more than one predictor variables. Regression can be used for continuous as well as attribute variables [5,7]. Prediction is based on the relationship between a thing that is known and a thing need to be predicted that is if certain attributes like domain knowledge and communication level of a student is known than his/her placement possibility can be predicted using multiple regression. Here placement possibility is dependent variable generally denoted by y and domain knowledge and communication level are independent variable generally denoted by x.

**3.2. Classification:** is the form of data analysis that extracts relevant data through classification models. Such models, called classifiers which predicts categorical class labels. In machine learning, many classification methods have been proposed. Most of the algorithms are memory resident typically assume small data size [3,6,7]. Data mining classification technique has strongest feature-scalability. These algorithms are capable of handling large amount of disk-resident data. In EDM also it plays vital role for data analysis and classification. It is supervised learning.

**3.3. Clustering:** Clustering is the process of grouping a set of data objects into multiple groups called as clusters. Objects within a cluster have high similarity. But objects are very dissimilar with other clusters [4,8,9]. For the objects involve in distance measure, similarities and dissimilarities are assessed. This assessment is based on the attribute values of the objects. Scalability and Incremental clusters are the striking feature of clustering. It is an example of unsupervised learning.

**3.4. Association Rules and Sequential Pattern Mining:** Frequent patterns are patterns that appear frequently in a data set. Finding frequent patterns plays an essential role in mining associations, correlations and many other interesting relationships among data. After finding association among data sets, certain rules are generated. It works on massive amount of data in retail community and industries. Association rule mining finds strong associations and relationships among large data sets [10]. This rule shows association among frequent items in item set. Sequence mining discovers interesting patterns in data with respect to some subjective or objective measure of interest. To find correlations, frequent patterns lot of tools are available [3].

**3.5. Text Mining:** Text mining is an interdisciplinary field that draws on information retrieval, data mining, machine learning, statistics and computational linguistics. It derives high quality information from text. This is typically done through the discovery of patterns and trends by means such as statistical pattern learning, topic modelling and statistical language modelling. Extract meaningful numeric indices of the text from unstructured information is the main purpose of text mining [6].

**3.6. Correlation Analysis:** Correlation Analysis is statistical method that is used to discover if there is a relationship between two variables; and how strong that variables are related to each other [10,11]. For nominal data, we use chi-square test. Commonly we use correlation coefficient and covariance for numeric data, both methods access how one attribute's values vary from those of another.

**3.7. Neural Network:** It is a set of connected units in which each connection has a weight associated with it. To predict the correct output, learner adjust the weights during learning phase. The main advantage of neural network is that it provides high tolerance of noisy data [7,12]. It is a technique to improve the interpretability of the trained network by using extracted rules for learning networks. To predict academic performance based on residency, ethnicity.

## 4. Important Tools used in EDM

There are many frameworks and algorithms can be used in educational Data mining. And to work on these algorithms and the techniques, certain tools are available which are used by the researchers. Table 1 shows some of the effective tools used in EDM

**Table 1: Important EDM Tools**

| Tools | Developer | Details | Environment |
|---|---|---|---|
| SPSS(Statistical Package for Social Science) | IBM | Mainly a statistical package which offers many statistical tests, correlations, regressions etc | Windows, LINUX, SOLARIS |
| DBMiner | DBMiner Technology Inc | Provides Data Mining algorithms used for classification, association and OLAP analysis. | LINUX, Windows |
| D3js | Mike Bostock | Data visualization tool that require data handling for complex data visualization | Windows, LINUX |

| | | | |
|---|---|---|---|
| Rapid Miner | RapidMiner Inc | Software used for text mining, data preparation and predictive analysis | Windows, LINUX |
| Oracle Data Mining | Oracle Corporation | Classification, Prediction, Regression, Clustering, Association Mining | LINUX, Windows, Mac |
| WEKA(Waikato Environment for Knowledge Analysis) | University of Waikato, New Zealand | Software tool used for Data Pre-processing and machine learning algorithms for Data Mining | Windows, LINUX |
| H2O | H2O.ai | Accomplish data analysis on data in cloud computing application systems | Windows, LINUX, SOLARIS |
| Informatica Master Data Management | Informatica Corporation | Flexible deployment options, Managing Quality, Integrating Data | Windows, LINUX |
| Talend Master Data Management | Talend S.A. | Integrating Data, Real Time Data Synchronization, Data Transformation | Windows, SOLARIS,LINUX |

## 5. Conclusion

In today's era, Education Data mining is an important analytical tool used to generate new patterns and prescribe new actions. Data mining techniques helps in extracting useful information for building effortful strategies for the youngsters in education system. This paper covers Data mining techniques used in higher education system and also covers various tools used in EDM for understanding students, accurate predictions, improving teaching learning process and applying new pedagogy in teaching learning which further helps in improving the education standards and meeting many challenges faced earlier.

## 6. References


1. B.M. Monjurul Alom et al. 2018.Educational Data Mining Perspective from Primary to University Education in Australia. Information Technology and Computer Science.
2. Hanan Aldowah et al. 2019. Educational data mining and learning analytics for 21st century higher education: A review and Synthesis. Telematics and Informatics.
3. https://towardsdatascience.com/why-iseducational-data-mining-importantin-the-researche78ed1a17908
4. Eduardo Fernandes et al. 2018. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. Journal of Business Research.
5. Slater, S. et al. 2017. Tools for educational data mining: A Review.
6. Hooshyar, D., Pedaste, M., & Yang, Y. (2020). Mining educational data to predict students' performance through procrastination behavior. Entropy, 22(1)
7. Al-Azawei, A., & Al-Masoudy, M. (2020). Predicting Learners' Performance in Virtual Learning Environment (VLE) based on Demographic, Behavioral and Engagement Antecedents. International Journal of Emerging Technologies in Learning (iJET), 15(9)
8. Navamani, J. M. A., & Kannammal, A. (2015). Predicting performance of schools by applying data mining techniques on public examination results. Research Journal of Applied Sciences, Engineering and Technology, 9(4)
9. Aggarwal, D., Mittal, S., & Bali, V. (2019). Prediction Model for Classifying Students Based on Performance using Machine Learning Techniques. International Journal of Recent Technology and Engineering, 8, 496-503.
10. Lu, H., & Yuan, J. (2018). Student performance prediction model based on discriminative feature selection. International Journal of Emerging Technologies in Learning (iJET), 13(10).
11. Wang, X., Yu, X., Guo, L., Liu, F., & Xu, L. (2021). Student performance prediction with short-term sequential campus behaviors. Information, 11(4), 201.
12. Polinar, E. L., Delima, A. J. P., & Vilchez, R. N. (2021). Students performance in board examination analysis using naïve bayes and C4. 5 algorithms. International Journal of Advanced Trends in Computer Science and Engineering, 9(1), 753-758.