

EXPLORING THE IMPACT OF DATA MINING AND MACHINE LEARNING ON STUDENT PERFORMANCE

¹Prasadu Peddi, ²Dr. Akash Saxena

¹Research Scholar, Arni University, Kangra district, Himachal Pradesh

²Research Guide, Arni University, Kangra district, Himachal Pradesh

ABSTRACT

Particularly in various regions of India, it seems that you are emphasising the need of gathering and analysing data to comprehend student performance in all its aspects. The paper highlights predictive modelling as an effective approach for enhancing student performance by identifying the root causes of issues and implementing efficient management strategies. Additionally, you mention the potential for this endeavour to provide outcomes that might inform strategies for improving students' performance and measure the effectiveness of interventions. Data analysis may uncover a vital aspect: the impact of educational tools and resources on student achievement. The use of data-driven insights to inform decision-making and enhance outcomes has seen a recent increase in several sectors, including education.

Keywords: Data Mining, Big data, Prediction.

I. PREAMBLE

This section provides a brief overview of the importance of big data and the obstacles it presents, as well as its relevance to different fields. Now, let's explore the main highlights: Uncovering valuable information from extensive databases: Extracting meaningful insights and patterns from extremely large data sets can be quite challenging. This kind of data goes beyond what algorithms usually used in traditional data mining and machine learning can handle.

Dealing with massive data sets can be quite challenging due to their complexities and limitations. Data in the modern era is massive, complex, and generated at an impressive speed.

Big data has extensive implications and holds value in numerous fields, including bioinformatics, healthcare, marketing, and finance, among others. There is a wealth of raw data in each of these areas that can provide valuable insights into significant questions.

Thanks to the progress in cloud computing, data mining techniques have become more adaptable in tackling the obstacles posed by big data. Cloud platforms offer the ability to easily handle and process massive amounts of data, making it a breeze for those with a data scientist mindset.

This passage highlights the significance of tackling obstacles associated with large amounts of data and harnessing cutting-edge technology to unlock its potential advantages in different sectors.

The Educational field uses data mining techniques widely to get precise diagnosis for diseases including heart disease, renal dialysis, and cancer. Through analysis of a wide range of data from several sources, we use DM concepts to predict useful information. By using methods for data analysis, prediction, and drug development, this study primarily seeks to fight cancer using advanced data mining techniques. Several of the more fundamental issues with data mining methods and their application to healthcare were also addressed in this article. Retrieval of pertinent information from current data resources is the aim of data mining (DM). Data from databases and data warehouses are among the information sources from which evolutionary processes are also retrieved. Using knowledge-based algorithms, the DM measures evaluate the data and identify noteworthy connections and associations.

Big data refers to very large and complicated files that are hard to manage or analyse with standard data processing tools or methods. These systems, which are sometimes called "3Vs," usually give you the following information: Lots of data, sometimes called "big data," can be measured in terabytes, petabytes, or even exabytes. Sensors, social media sites, online shopping, and other similar internet tools could be used to get this information.

A lot of data is being made very quickly, sometimes almost in real time. Things like financial activities, data from monitors that are linked to the internet, and social media posts need to be processed and analysed quickly.

Diversity is very important: The two types of data are organised and unstructured. organised data includes databases, XML, and JSON, while unstructured data includes text, photos, and videos. This difference makes the process of handling and analysing data more difficult.

In addition to the three Vs, some speeches may also talk about value (the possible benefits or insights that can be gained from data analysis), truth (how reliable the data is), and variability (how the data types aren't always the same). "Big data analytics" is the process of getting useful information from very large files in order to make decisions, run businesses better, and gain a competitive edge in many areas. Healthcare, business, banking, and science study are some of the fields that are covered. In big data analytics, different methods are used, such as data mining, natural language processing (NLP), predictive analytics, and machine learning (ML). Big data has changed the way companies receive, store, analyse, and use data in revolutionary ways. This has led to new ideas, higher productivity, and exciting growth potential.

II. CONTEXTUAL WORK

This portion talks about a study that focuses on how machine learning algorithms can be used to combine different kinds of student data. The goal is to accurately find out how well students are doing in school. Let me break down the main points in more detail:

Our main goal is to make it easier to figure out how well a student is doing in school by using machine learning techniques and combining different kinds of student data.

Looking into algorithms for machine learning: Four types of machine learning methods were used for the fusion task: Gaussian Mixture Models Artificial Neural Network Fuzzy Expert System Support Vector Machines

It is very helpful to have expert systems in many areas because they can handle complicated and uncertain data well. Fuzzy logic is used by a system called the Fuzzy Expert System to deal with inaccurate and vague data. The Fuzzy Expert System can handle uncertainty well, just like an information security expert. This lets it give useful insights and suggestions in many different situations.

Vector Machines

The Different Types of Students: Different ways of gathering information from students were used in the study, such as looking at their facial reactions and talking habits. These methods were used to teach and test the machine learning systems. We used a well-known student database to train and test the algorithms. This database has examples of words and faces, along with scores from experts on how close they are.

Taking a look at current methods: The fusion results from the new algorithms were compared to those from other fusion methods that were already available. Also, the success of each programme was compared to each other. SVM did better than the other methods, showing that it is more efficient. When compared to, it showed big improvements in speed.

Students with a unique way of learning. Students with a lot of different skills and abilities Using the most recent and cutting edge methods

The study found that the fusion schemes described in the paper work better than traditional ways and make a big difference in how well students do. The method based on support vector machines works the best out of all the ones that were tried.

RESEARCH GAP

The assessment of student's performance in educational institutions indicates the level of efforts taken by educational institutions should do for enhancing the poor or average learner. The significance of applying EDM models, make use of the previous data of the students for predicting the unseen or upcoming performance of the students. This idea has attracted several researchers to develop classification models to predict the unknown labels of future instances. Several research people and educational institutions started to get attracted to the domain of predicting the performance of the student to classify the educational level of student performance. Though the educational sector uses several techniques for extracting useful information on the features of students undertake the learning process, it is needed to develop a student performance assessment model to assist the students as well as faculties to improve their performance to the next stage. Though some of the existing studies related to the EDM in the educational sector is available, the major goal of this research work concentrates on examining and identifying the useful rules and patterns to motivate the students for handling their education as well as carrier in a good manner, also to improve the and functions academics to supervise the policies for student's benefit.

III. RESEARCH EFFORTS

This Section discusses the adaptation of data mining tools for big data problems, particularly focusing on the MapReduce paradigm and its application in parallel environments. Here's a breakdown of the key points:

1. Adaptation of Data Mining Tools for Big Data:

To solve problems with big data, methods often need to be redesigned and put into multiple settings.

A powerful and reliable way to look at large datasets is the MapReduce model and its distributed file system, which were first created by Google.

MapReduce is better than other parallelization schemes like MPI because it is simple and can handle errors.

2. Parallelization of Machine Learning Tools

Many recent works focus on parallelizing machine learning tools using the MapReduce approach.

Apache Spark, an extension of the MapReduce approach, has been successfully applied in various data mining and machine learning problems.

3. Data Preprocessing and Reduction

Data preprocessing methods, particularly data reduction models, aim to clean and simplify input data to accelerate data mining algorithms and improve their accuracy.

Two main types of data reduction models are described: instance selection/generation and feature selection/extraction.

4. Evolutionary Feature Selection (EFS) Models

Evolutionary approaches have been successfully used for feature selection techniques.

However, traditional evolutionary approaches may face limitations in handling very large problems due to excessive computation time.

5. Objective of the Paper

The main goal of the study is to show how Evolutionary Feature Selection (EFS) models can be used on large amounts of data.

"MapReduce for Evolutionary Feature Selection" (MR-EFS) is the name of the MapReduce method that was made for this reason.

In the map phase, MR-EFS splits the data and runs various EFS processes at the same time. In the reduce phase, it merges the answers to get the most interesting features.

Overall, the paper aims to address the challenge of applying evolutionary feature selection models to big data by developing a MapReduce-based algorithm, MR-EFS, which allows for efficient and parallel processing of large datasets.

This passage provides an overview of the importance of predicting student academic performance and the role of information technology, particularly artificial intelligence (AI) and machine learning (ML), in addressing this challenge. Let's break down the key points:

Significance of Student Academic Performance:

Student academic performance is crucial for educational advancement and institutional reputation.

It influences various aspects such as university rankings and employability prospects for students.

Role of Information Technology in Education:

Information technology, particularly AI, has become increasingly important in education.

AI-based technologies, such as web-based education systems (WBS) and intelligent tutorial systems (ITS), gather vast amounts of student data from multiple sources.

Challenges in Predicting Student Performance:

Despite the availability of student data, effectively applying AI techniques to predict academic performance remains challenging due to data volume and complexity.

Accurately assessing student performance requires appropriate examination of obtained data to identify predictive factors.

Benefits of Predicting Student Performance:

Predicting and analyzing student performance can help educators identify weaknesses, improve learning activities, and enhance administrative operations.

Machine learning techniques in educational data mining aim to model and detect meaningful patterns from educational contexts.

IV. Motivation for the Study

We aimed to address this gap in the existing research by conducting a thorough evaluation of using machine learning models to predict academic success for students. The study conducted a comprehensive survey and analysis of predictive variables and machine learning algorithms used in academic performance prediction. Its aim was to map existing knowledge, identify areas for further research, and offer recommendations for future studies. Through providing a comprehensive analysis of key features and methodologies, and emphasising the need for further exploration, this study aims to contribute to our understanding of the potential applications of ML models in predicting students' academic performance.

Machine Learning

An ML model is defined as a computer-intensive mechanism and applies re-sampling and iterative methodologies for classification approaches. ML approaches are considered with optimal subset selection and eliminate the issues of classical classifiers like over-fitting as well as distributional demands of parameters. ML technologies that have emerged in computer science with logic and basic mathematics, statistics as ML approaches do not estimate the group features rather it is initialized with an arbitrary group separator and tunes frequently till satisfying the classification groups. ML examines the tuning variables and individual ML functions became unstable, which makes a suitable process. As the non-statistical nature is embedded, these approaches can apply the data in various formats like nominal data that generates maximum classification accuracies.

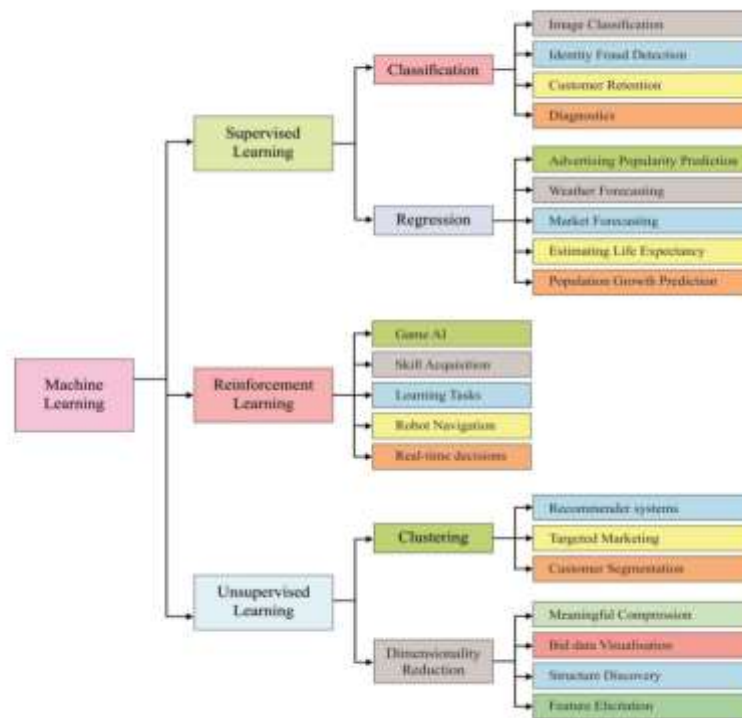


Fig 1: Models of Machine Learning

Bootstrapping involves estimating the classification accuracies, however several techniques are used. Other machine learning techniques include support vector machines (SVMs), random forests (RF), recursive partitioning, and neural networks (NN). Economic and social progress in India mostly depends on its educational system. Giving pupils excellent employable skills and knowledge is the primary goal of an educational system. Most educational institutions are using Educational Data Mining (EDM) now that technology has advanced. Using data mining methods on a large volume of student data, EDM is the process of obtaining insightful information that might enhance the teaching and learning environment. Education has shown great interest in data mining (DM) methods. EDM is fast gaining importance in the educational industry because of its essential capabilities of prediction and decision making. Specifically, data-driven approaches provided by decision-making models enable educational planners to maintain the objective of raising the effectiveness and calibre of the teaching and learning process. From this angle, the global educational system seems to be greatly impacted by the use of DM models. Education institutions might acquire focused answers for certain problems with the help of EDM. As such, the DM tools and processes provide priceless support in predicting student results and making informed decisions.

CONCLUSION

Possible future efforts including this topic are discussed in this publication. To further improve the classifier's performance, the suggested models might be enhanced by using an efficient outlier identification technique. Additionally, the datasets provided may have their missing data replaced by extending the suggested model. Plus, you may use data mining methods to make the categorization process better. In addition, metaheuristic algorithms may be used to handle the optimisation issue associated with the hyperparameters in ML and DM models.

REFERENCES

1. Park, H. S., & Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36(2), 3336-3341.
2. Novak, J.D. and Cañas, A.J. (2008). The theory underlying concept maps and how to construct and use them. Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition
3. Macfadyen, L.P. and Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & education.*, 54(2): 588-599.
4. M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.
5. Parameswaran, A., Venetis, P. and Garcia-Molina, H. (2011). Recommendation systems with complex constraints: A course recommendation perspective. *ACM Transactions on Information Systems (TOIS).*, 29(4):1-33.
6. Romero, C. and Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications.*, 33(1):135-146.
7. R. E. Fan, P. H. Chen, and C. J. Lin, "Working set selection using second order information for training support vector machines," *Journal of Machine Learning Research*, vol. 6, pp. 1889–1918, 2005.
8. S. K. Yadav and S. Pal, "Data mining: a prediction for performance improvement of engineering students using classification," vol. 2, no. 2, pp. 51–56, 2012,

9. Hsia, T.C., Shie, A.J. and Chen, L.C. (2008). Course planning of extension education to meet market demand by using data mining techniques—an example of Chinkuo technology university in Taiwan. *Expert Systems with Applications.*, 34(1): 596-602.
10. Lee, C.H., Lee, G.G. and Leu, Y. (2009). Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning. *Expert Systems with applications.*, 36(2):1675-1684.

