Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.20 : 2024
ISSN : **1906-9685**

SIGNIFICANCE OF DATA MINING TECHNIQUES TO ASSESS STUDENTS PERFORMANCE

Dr.S.Sharmila Assistant Professor, Department of Computer Science, NGM College, Pollachi-642001, Coimbatore, TamilNadu, India Email: mcasharmi2007@gmail.com, sharmila@ngmc.org

Abstract:
Data mining in education is a recently developed field that deals with using data mining techniques and algorithms on data stored in educational system repositories. It can be used to gain important information about the learning environment, student performance, and the conditions that lead to student dropout or poor performance risk detection. The researchers have worked extremely hard to use data mining tools to analyze students' performance. This study highlights the value of several data mining techniques in predicting students' academic success, which, because of early projection, aids teachers, mentors, and management in improving students' abilities and academic performance. Additionally, the management can create unique programs using this for the course's "low" and "outstanding" achievers. Students who are expected to perform well could be motivated to achieve the exceptional level in this way. Conversely, pupils who are predicted to perform poorly could receive support in order to graduate with higher grades. This guarantees the caliber of graduates and other advancements in a constructive manner.
Keywords: Educational data mining, student performance, prediction, classification, clustering, Outlier Detection.

INTRODUCTION

Over the past few decades, the importance of higher education has increased significantly. Because of the private involvement, universities are compelled to alter their goals and scope. A regulating body's controller has implemented criteria concerning faculty, infrastructure, and other resources. Because there is a vast amount of data available in many businesses, both public and private, new technologies are being developed in the fields of data management and analysis. The primary goal of data mining techniques is to find obscure and meaningless connections among data with a variety of attributes [3]. Numerous data mining techniques are employed in a variety of industries, including education. Although data mining is still relatively new, it has been implemented successfully in business applications and is utilized in higher education and learning institutions. Educational data mining demonstrates that this is a relatively new approach in the education industry, with the goal of identifying and extracting useful new knowledge from data.

The technique of data mining is booming in the field of education. It can be applied to data analytics and visualization, student modeling, performance prediction, student grouping, and other applications [2]. When working with educational datasets that come from educational sources, data mining techniques can be used to extract information that will be useful to students, parents, educators, educational developers, and researchers. The primary focus of every institution is the academic achievement of its students. Therefore, if we could forecast students' academic achievement in advance, identify the key markers for their success, and learn more about their perseverance, we could increase kids' academic performance [13]. Advance access to student academic performance data allows universities to better plan for maintaining and improving student performance, as well as directing management to provide more informed admission options. This data is also critical for the long-term growth and development of universities.

Achieving excellence in academic achievement is the primary goal shared by all students and institutions. Academic competence is the primary metric used to assess an individual's educational progress. In the higher education system, a student's CGPA is a crucial indicator of their academic success. It primarily serves to gauge the student's aptitude for the specific program or course they have chosen. Students can succeed academically in a course or program if they use efficient and methodical learning strategies, have the highest levels of self-esteem, and are highly motivated to study. Measuring and storing academic achievements of students is the main function of indicators. If a student drops the course, the university will be impacted financially, emotionally, and otherwise. A higher student failure rate has the potential to harm the university's reputation. The hardest difficulty for any higher education system is predicting the students' concert and improving the educational process[1, 12].

A student's academic success in a higher education system can be evaluated in a variety of ways, including exam scores, performance in practical exercises, participation in curriculum activities, regularity in turning in assignments and doing homework at home, etc [9]. A review of the literature reveals that numerous academics have attempted to investigate the variables influencing students' academic achievement. Cumulative grade point is a widely used metric in most higher education systems to assess students' academic performance. The majority of researchers demonstrate the connection between variables influencing students' performance and their exam-derived CGPA. The subjects being taught, the essential characteristics of the student, the student's family background, the student's learning style, and the mentor's influence on the student's development are the five areas that have an impact on the success of the student. Numerous data mining approaches are utilized on student datasets in order to derive significant patterns from them [4]. To anticipate a student's academic success, these produced patterns which might take the form of rules or models can be applied to newly generated data sets or tuples.

LITERATURE SURVEY

Student performance can be improved by a variety of strategies, including performance prediction, performance enhancement, investigation of factors influencing performance, comparative examination of data mining methodologies, etc. An overview of the data mining methods that have been applied to forecast student performance can be found in Amirah Mohamed Shahiri et al.'s work [3]. This research also focuses on identifying the most significant qualities in a student's data using the prediction algorithm. Using educational data mining approaches, it truly increases student success and achievement more effectively and efficiently. It might have positive effects on academic institutions, teachers, and students. A study was conducted by Mohammed Abdullah Al-Hagery et al. [7] in order to determine the variables influencing students' academic performance. The data was analyzed using the K-means and X-means clustering algorithms to determine how the students' performance related to these parameters [6]. The study's conclusions include a list of the social and personal variables that have the most effects on pupils' performance, including parents' economic level, occupation, and educational background. Moreover, it is enhancing the quality of education and encouraging educational establishments to reap the rewards of mining their students' gathered data for hidden knowledge.

Mohammed M. Abu Tair and colleagues [8] conducted an analysis and attempted to get valuable insights from graduate student data obtained from Khanyounis College of Science and Technology. The data spans fifteen years, from 1993 to 2007. The authors used data mining approaches to find association, classification, clustering, and outlier detection criteria after preprocessing the data. They give the extracted knowledge and discuss its significance in the educational field in each of these four jobs. In order to determine which students failed the third-year control systems course based on their performance in the first and second years' mathematics courses, Nibras Z. Salih et al. [13] conducted a study. The dataset was subjected to three different algorithms: multilayer perceptrons, support vector machines, and naive bayes. The results showed an overfitting issue due to the imbalanced dataset, which was resolved by using the synthetic minority oversampling technique. Their findings demonstrated that, when the synthetic minority oversampling strategy is used, the support vector machine algorithm demonstrates an effective classification method. With the use of associated metrics

and the Waikato environment for knowledge analysis (WEKA) tool [16], the accuracy of the classifiers was determined using the confusion matrix.

INTRODUCTION TO DATA MINING

Data mining is one of the steps involved in Knowledge Discovery in Databases (KDD). KDD starts with a raw data and finish with extracted knowledge. The steps involved in KDD shows in following figure 1;

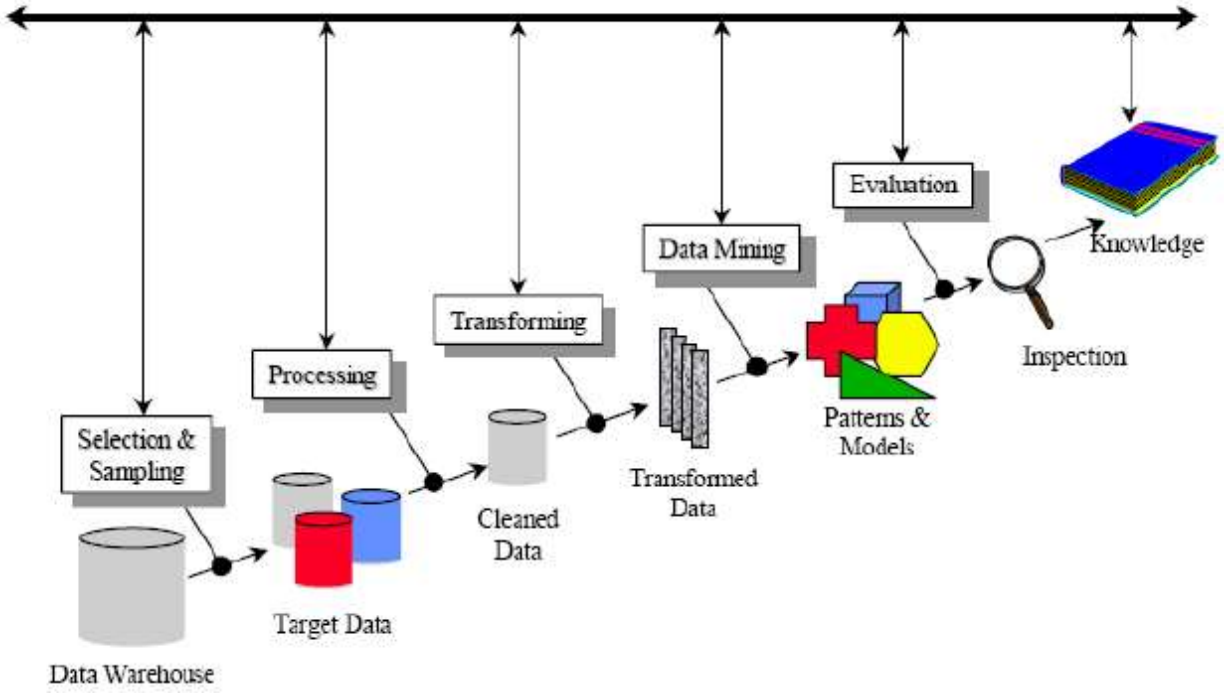

Figure 1: Knowledge Discovery in Databases

3.1 Data Selection (Target Data) - This phase's objective is to extract pertinent data from a bigger data store for the data mining analysis [15]. The procedure is streamlined and expedited in part by this data extraction. In this approach, subsets of the data can be identified, such as all those who own cars.

3.2 Data Pre-processing (Cleaned Data) - The data preparation and cleaning operations required to guarantee accurate results are the focus of this KDD phase. During this phase, typical actions include removing missing values from the data, making sure that coded values have a consistent meaning, and making sure that there are no misleading data values. For instance, it is superfluous to record a patient's sex when researching pregnancy.

3.3 Data Transformation (Transformed Data) - In order to ensure that the results are legitimate, this stage of the lifecycle aims to transform the data into a two-dimensional table and remove any fields that are superfluous or highly correlated. Depending on the goal of the work, identifying relevant attributes to represent the data. It is possible to find invariant representations for the data or use dimensionality reduction techniques to decrease the effective number of variables under consideration.

3.4 Data Mining (Patterns & Models) - Using the right collection of algorithms to evaluate the data, the data mining phase seeks to identify relevant patterns and rules and generate predictive models. This is the KDD cycle's central component.

3.5 Interpretation and Evaluation (Knowledge) - Although there is a limitless amount of patterns that data mining algorithms might uncover in the data, many of them might not be significant or helpful. The goal of this last stage is to identify the models that are reliable and practical for usage in upcoming business choices.

OVERVIEW OF DATA MINING IN EDUCATION FIELD

"Educational Data Mining" is the term used whenever mining is applied in the field of education [5]. The primary goal of this expanding field of study is to extract useful hidden knowledge from

educational data sets. Educators have the ability to enhance the decision-making process inside the educational system by utilizing hidden knowledge. It is an application for data mining related to educational data. It is a method of quantitative observation and analytical analysis to determine how students react to the educational system and how that response affects their learning. Its goal is to answer issues in educational research by analyzing educational data [11]. In order to improve the success rate of students, it is also important to examine academic performance, comprehend student learning behaviors, and identify the elements that influence the learning process in a particular educational system.

The education industry has grown rapidly in recent years, which has increased the amount of education data. As a result, it is now crucial to mine education data to understand student behavior during the learning process or to identify problems with students. We often use information from questionnaires, interviews, class activities, and other sources to compile information about students' educational experiences. These methods can't be repeated often because they are usually lengthy. The examination of "structured data" derived from "course management systems (CMS)" was the primary focus of the newly formed fields of learning analysis and educational assessment.

DATA MINING APPROACHES IN EDUCATIONAL FIELD

Too many approaches are followed by data mining that can we applied in educational field. These approaches are fall inside one of the following categories:

1) Prediction: Predicting a variable from the set of predictive variables is the aim of this kind of model. Regression and classification are two prediction methods [17]. Regression techniques can only predict a result on continuous or numeric data; they cannot be applied to categorical or numerical data. Density estimation, which uses probability density functions as projected values, is another data mining prediction techniques.

Regression: It is a commonly employed statistically intrinsic technique in data mining. Establishing a link between "a dependent variable" and "a set of predictors" is the main goal of regression analysis [18]. One method of supervised learning data mining is regression. The database is divided into training and validation data by supervised learning. Regression approaches are divided into two categories. Regressions, both linear and non-linear.

Classification: Classifies a data element into one or more pre-established category classes[19]. Various algorithms are utilized for classification, such as support vector machines, decision trees, naïve based classification, and generalized linear models.

Clustering: Clustering is defined as "partitioning the data into distinct groups." We refer to these collectives as clusters. Similar objects are transferred into the same cluster, while dissimilar objects are passed into other objects, according to the clustering's operating principle. The cluster algorithm can be started using one of two methods: "without prior hypothesis" or "with an earlier hypothesis."

Relationship mining: Its goal is to identify relationships between the variables in the data sets. We can use data mining approaches like "association rules," "sequential pattern mining," "correlation mining," and "causal data mining" to determine the relationship. Relationship mining tools, for instance, can be used to determine a correlation between "students' online activities" and "final grades obtained into the semesters." The second example is that we can build a model that generates "learner's problem solving activity sequences" by applying this technique.

Discovery with Models: Its primary goal is to use different data mining methods to educational data sets in order to extract some intriguing information from them that has either never been observed before or has been buried [20]. For instance, use prediction data mining to determine the student's outcome based on their academic standing.

Outlier Detection: This strategy aims to identify the disparate data points. When compared to the typical value of another parameter, an outlier is identified by its extremely high or low value. For instance, outlier analysis can be used in educational data mining to pinpoint "scholar's learning incapacities," "asymmetrical education processes," and behavioral inconsistencies in learners.

STUDENT PERFORMANCE MEASURES

Academic performance of students is typically assessed through tests or ongoing evaluation [10]. There are three category of student's academic performance.

Formative Assessment: It provides teachers and students with ongoing feedback. Both teachers and students used it to better their methods of instruction and learning. Formative assessments include quizzes and homework.

Interim Assessment: It happens sporadically or at the end of the course's semester. Teachers can use this evaluation to gauge how well a student is doing in their coursework and whether they should anticipate doing well on subsequent assessments.

Summative Assessment: Its goal is to evaluate a student's learning at the conclusion of a unit of study by contrasting it with a reference point or industry standard.

In addition to the effectiveness of the student throughout the academic year, other factors like the student's demographics, past activities, knowledge level, attendance, gender, and location can also have a significant influence on the forecasting process.

NEED FOR PREDICTING STUDENT PERFORMANCE

Numerous scholarly articles have been reviewed regarding the need to forecast students' academic achievement in postsecondary education. There are other justifications for the necessity of forecasting student success. According to the following, those are.

1) Mentors can improve the performance of week or average students by taking positive activities if we are aware of their performance in advance. Predictive models provide mentors with a number of tools to aid in better understanding and categorizing students. Additionally, it recognizes students who are at danger and in need of additional care. Teachers might approve of many teaching methods, for instance, observation, demonstration, and examination/verification. Additionally, the prediction model's output can alter the best course design and offer unbiased commentary on the assignments, activities, and assessments.

2) Researchers and developers may evaluate the course structure using predictive models, and then decide whether to change the course prospectus to enhance student performance and the learning process.

3) Predictive model results could help students identify their areas of strength and weakness in the subject. Additionally, this model suggests various tasks and gives students the opportunity to reconsider their approach to learning. Classifying course sections that need extra attention can be done with the aid of student performance prediction.

MOTIVATION OF THE RESEARCH

The standard of the academic environment, the caliber and volume of research, and the overall caliber of the system's output (students, research, and development) have all declined during the past few decades in India's higher education. Providing high-quality education is the goal of any institution of higher learning. By providing the most accurate anticipated model, every higher education institution can achieve academic greatness. There are numerous prediction models with various methodologies available. It is impossible to anticipate with any degree of accuracy whether a student would perform well academically, controversially, or averagely. Higher education has a number of difficulties, including growing student populations, a highly competitive global education market, rising student expectations, a need for and need for new technology, government funding, etc. Universities are compelled to consider ways to improve the delivery and sustainability of education in light of these demands and difficulties.

The current systems have just been unable to keep up with the changing needs of higher learning. The field of education makes extensive use of data mining applications. Numerous mining applications have been applied in the field of education by scholars. Using academic and student data, mining educational data facilitates a deeper understanding of learners and learning. As a result, the investigation of educational data aids in the creation of techniques that enhance the educational process in its entirety. By applying data mining techniques to the student's data set, we can find numerous hidden models. Administrators and mentors use these underutilized hidden models to make well-

informed decisions. Numerous writers have contributed to the field of "data mining in higher education" through their study. While data mining techniques have also been utilized in the educational field, in this case they were applied to student datasets in order to provide an early estimate of the academic achievement of the students [14]. The purpose of this study was to determine the factors that have a significant impact on students' academic performance. The aforementioned justifications have motivated me to work in this field.

CONCLUSION

The primary goal of educational institutions is to equip students with the knowledge and abilities they will need to enter the workforce at a given time. Social and economic advancement are significantly influenced by how well educational systems achieve this goal. Massive amounts of data were generated by the technology employed in educational systems, which make it challenging to examine by hand. A range of approaches and strategies are provided by Educational Data Mining (EDM), which aids in assessing student performance and supports teachers in making predictions and taking action to enhance students' performance in a course or program. EDM's primary goal is to examine, resolve, and enhance educational problems in order to enhance educational procedures. As a result, its objective is to analyze educational data in order to address related educational problems. EDM involves taking data from the educational domain and extracting new, intriguing, interpretable, and helpful information.

REFERENCES

- Abdallah Namoun, Abdullah Alshanqiti et al., "Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review", MDPI, Appl. Sci. 2021, 11, 237, doi.org/10.3390/app11010237.
- Ahmed Mueen, Bassam Zafar et al., "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques", I.J. Modern Education and Computer Science, 2016, 11, 36-42, DOI: 10.5815/ijmecs.2016.11.05.
- Amirah Mohamed Shahiri, Wahidah Husain et al., "A Review on Predicting Student's Performance using Data Mining Techniques", ScienceDirect, Procedia Computer Science 72 (2015) 414-422.
- Edin Osmanbegovic, Mirza Suljic., "Data Mining Approach for Predicting Student Performance", Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.
- Fatima Alshareef, Hosam Alhakami et al., "Educational Data Mining Applications and Techniques", International Journal of Advanced Computer Science and Applications, Vol. 11, No. 4, 2020.
- M. Durairaj, C. Vijitha., "Educational Data mining for Prediction of Student Performance Using Clustering Algorithms", International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5987-5991, ISSN:0975-9646.
- Mohammed Abdullah Al-Hagery, Maryam Abdullah Alzaid, et al., "Data Mining Methods for Detecting the Most Significant Factors Affecting Students' Performance", I.J. Information Technology and Computer Science, 2020, 5, 1-13, DOI: 10.5815/ijitcs.2020.05.01.
- Mohammed M. Abu Tair, Alaa M. El-Halees et al., "Mining Educational Data to Improve Students' Performance: A Case Study", International Journal of Information and Communication Technology Research, Volume 2 No. 2, February 2012, ISSN 2223-4985.
- Mukesh Kumar, Dr. Disha Handa et al., "Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques", I.J. Education and Management Engineering, 2017, 6, 40-49, DOI: 10.5815/ijeme.2017.06.05.
- Mukesh Kumar, J. Singh, "Evaluation of Data Mining Techniques for Predicting Student's Performance", I.J. Modern Education and Computer Science, 2017, 8, 25-31, DOI: 10.5815/ijmecs.2017.08.04.

- Myla M. Arcinas, Guna Sekhar Sajja et al., "Role of Data Mining in Education for Improving Students Performance for Social Change", Turkish Journal of Physiotherapy and Rehabilitation; 32(3), ISSN 2651-4451 | e-ISSN 2651-446X.
- N D Lynn, A W R Emanuel., "Using Data Mining Techniques to Predict Students' Performance. a Review", IOP Conf. Series: Materials Science and Engineering, 1096 (2021) 012083.
- Nibras Z. Salih, Walaa Khalaf, "Prediction of student's performance through educational data mining techniques", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 22, No. 3, June 2021, pp. 1708-1715, ISSN: 2502-4752, DOI: 10.11591/ijeecs.v22.i3.pp1708-1715.
- Oyebade K. Oyedotun, Sam Nii Tackie et al., "Data Mining of Students' Performance: Turkish Students as a Case Study", I.J. Intelligent Systems and Applications, 2015, 09, 20-27, DOI: 10.5815/ijisa.2015.09.03.
- Reda M. Ahmed, Nahla F. Omran et al., "Predicting and Analysis of Students' Academic Performance using Data Mining Techniques", International Journal of Computer Applications (0975 - 8887), Volume 182 - No.32, December 2018.
- Sadiq Hussain, Neama Abdulaziz Dahan et al., "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 9, No. 2, February 2018, pp. 447-459, ISSN: 2502-4752, DOI: 10.11591/ijeecs.v9.i2.pp447-459.
- Shujat Hussain, Saif Ur Rehman et al., "Significance of Education Data Mining in Student's Academic Performance Prediction and Analysis", International Journal of Innovations in Science & Technology, Vol. 5 Issue.3, pp. 215-231, Sep 2023.
- V.Ramesh, P.Parkavi et al., "Predicting Student Performance: A Statistical and Data Mining Approach", International Journal of Computer Applications (0975 – 8887), Volume 63– No.8, February 2013.
- Y. K. Saheed1, T. O. Oladele et al., "Student Performance Prediction Based on Data Mining Classification Techniques", Nigerian Journal of Technology (NIJOTECH), Vol. 37, No. 4, October 2018, pp. 1087-1091.
- Yupei Zhang, Yue Yun et al., "Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis", Frontiers in Psychology, December 2021, Volume 12, Article 698490.