STUDENT BEHAVIORAL DATA ANALYSIS IDENTIFY THE EDUCATIONAL FACTORS WHICH IMPACTS STUDENTS' ACADEMIC PERFORMANCE USING ML TECHNIQUES

Sakshi[1], Srushty G S[2], Yashaswini[3], Yashaswini G[4], Mr. B N Kiran[5]

[1] *Student, Information Science & Engineering, NIE Mysuru, Karnataka, India*
[2] *Student, Information Science & Engineering, NIE Mysuru, Karnataka, India*
[3] *Student, Information Science & Engineering, NIE Mysuru, Karnataka, India*
[4] *Student, Information Science & Engineering, NIE Mysuru, Karnataka, India*
[5] *Assistant Professor, Information Science & Engineering, NIE Mysuru, Karnataka, India*

ABSTRACT

Identifying the various elements that have an impact on a student's academic performance and learning style. In the modern educational system, analysing students' mental health problems and poor academic performance is a challenging challenge. It is challenging to monitor student behavior and characteristics under the existing system. No technology or tool exists that forecasts or provides guidance on how to adjust a student's academic performance. In the modern educational environment, determining the relationship between the variables influencing students' performance and their academic outcomes is essential. The "Association Learning" data science approach is used by the system to identify patterns. The "apriori algorithm," "apriori TID algorithm," or "Eclat algorithm" are the methods we employ to identify patterns. The suggested system is designed to be a real-time tool that helps instructors and universities understand the behavioral patterns of their students. The system also attempts to forecast each person's performance. The "naive Bayes" machine learning method is used by the system to forecast each student's performance independently.

Keywords: - *Associate Learning Algorithm, Student Behavioral Analysis, Student Performance Prediction, and Predictive modelling in Education.*

1. INTRODUCTION

A major problem for student affairs management is the contradiction between the limited energy of student counsellors and the diversity of student behaviors, which results in many potential problem students losing the opportunity for early intervention. Since the beginning of the 21st century, the rapid development of information technology in education and the construction of digital campuses has made it possible for student counsellors to conduct quantitative analysis of student school behaviors, especially to provide early warning to students who may have problems, so that the contradiction could be alleviated by applying the analysis and early warning methods. As contemporary college students who grew up in the Internet era, their daily lives, learning, and thinking are deeply influenced by the Internet. This provides us with the possibility to understand their campus network behavioral characteristics through big data.

A challenge for today's student counsellors is extracting useful information from massive volumes of data in ever-expanding data categories and scales. This is also a huge opportunity to work with new approaches.

This study began with real work problems and was carried out using data from B college students' network behavior, combining big data science techniques with big data thinking. It also looked into the characteristics of college students' network behavior rules and identified the students who needed special attention due to the high volume of network usage on campus. This research might serve as a useful example of student data science work in practice.

2. RELATED WORK

In article [1], we explore Educational Data Mining (EDM), a field of study concerned with applying data mining methods and algorithms to educational data for a range of applications. The outcomes of three experiments on predicting student performance are presented and discussed in this work. The studies used actual data that was taken from student performance records that were held for a university course. Regression, SVM, and AdaBoost are a few of the classification algorithms used in the methodology. By using these strategies, which don't involve word removal, the entire approach to predicting students' final performance early in the term is explained, allowing for prompt intervention with educational recovery strategies to avoid student failures.

In the subsequent research [2], it is recognized that preliminary education contributes to a nation's literacy rate, while higher education directly impacts the workforce provided to industries, thus influencing the economy. Across India, numerous institutions of higher learning have been established. However, the quality of education is often evaluated based on the success rate of students and the institute's ability to retain them. Predicting students' performance can serve as a vital tool in identifying those at risk of failure, enabling management to offer timely assistance and implement necessary measures to coach students and enhance their performance. In this study, the methodology employed focuses on the implementation of regression analysis to facilitate this predictive process.

Under the scope of this study [3], collaborative filtering a method often employed in recommendation systems on the assumption that users with comparable traits and behaviors will have comparable preferences is offered as the basis for an intelligent student advising system. Students are placed into groups under this advising method, and guidance is offered to them according to how similar they are to the groupings. A course that is favoured by a group of students may be suggested to a student if it is found that the student and that group are comparable.

In the context of educational environments, the ability to predict a student's performance is crucial [4]. Students' academic achievements are influenced by various factors, including personal, social, psychological, and environmental variables. Data mining emerges as a promising tool in achieving this objective. Data mining techniques enable the discovery of hidden information patterns and relationships within vast datasets, facilitating informed decision-making processes. Each dataset contains a wealth of information, and the nature of this information determines the appropriate data processing methods. Within the education sector, there exists a wealth of data holding valuable insights. Information and communication technology (ICT) plays a crucial role in capturing and compiling this low-cost information. The rapid expansion of educational databases is primarily attributed to the substantial volume of data they store. Loyal students serve as key motivators for higher education systems, and understanding their needs requires the effective management and processing of student databases. The data mining approach offers a means to extract valid information from existing students, facilitating the efficient management of relationships with upcoming students.

3. PROPOSED WORK

The primary goal is to identify pupils' behavior patterns in a timely and accurate manner. The primary goal is to identify the groups of kids that require particular attention. Without campus behavior analysis, these students' academics and other performance variables suffer. The system uses characteristics such as attendance status, extracurricular activities, grades, technical skills, previous semester results, comprehension capabilities, aptitude grades, contact with lecturers, and so on. The system assists professors in identifying the most relevant aspects impacting student achievement.

The system employs a data science approach known as "Association Learning" to identify trends. We employ the "Eclat Algorithm" to discover patterns and "Naive Bayes Algorithm" to predict individual student performance. The proposed system would be built as a web application to assist institutions and teachers in understanding student behavior trends.

3.1 Objectives of Proposed System

- Proposed project is a student behavior analysis and prediction and management system that is meant for educational institutes.
- The proposed project makes use of the ML technique or Data Science approach for the students' behavior analysis.
- To identify many elements that impact a student's learning behavior and academic success during their academic career.
- Finding the various elements that influence a student's learning style and performance throughout their academic career; providing reliable data from current students to manage connections with prospective students.
- Construction of a prediction model using ML technique - Unsupervised Learning on the basis of identified predictive variables.

4. METHODOLOGY

The most well-known, recognizable, and basic data science approach is most likely association (or connection). To find patterns, we commonly establish a basic connection between two or more things of the same kind. In a market-basket study, for instance, when we monitor consumers' purchasing patterns, we may note that a customer consistently purchases cream together with strawberries. Based on this information, we might advise them to consider purchasing cream the next time they buy strawberries.
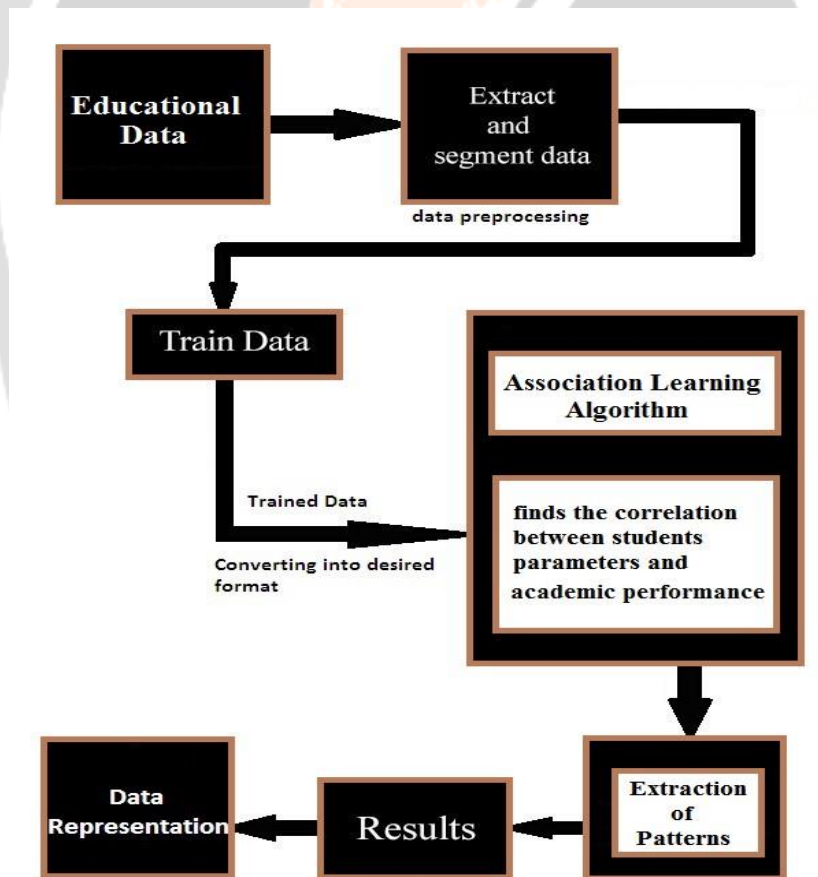

Fig -1 Predictive Analytics Workflow

4.1 Pattern Prediction Process

Step 1: Data Collection
We are developing a new application that includes data servers, which are utilized to store data, while we work on a real-time application. To collect data, one must gather information from many sources. Student parameters and student outcomes are examples of data.

Step 2: Data Preparation
Here, server data is retrieved and examined. Whole data extraction and analysis, wherein we keep the data needed for processing and eliminate unnecessary data. The project states that to produce outputs, just student parameters and student outcomes are needed.

Step 3: Specify Constraints
SUPPORT COUNT
How many transactions total in the data set relates to how many transactions total that contain that item (A).
CONFIDENCE
A set's confidence is measured by comparing the total number of transactions that contain the item set to the total number of transactions that contain LHS.

Step 4: Association Rules Mining (Eclat Algorithm)
Probably the most popular, well-known, and basic data mining approach is association (or connection). In order to find patterns, we simply correlate two or more items often of the same type here.
To analyse and identify trends in e-commerce data, we employ the Eclat algorithm. Here, we produce educational patterns.
The following factors led to the selection of the Eclat algorithm.
1. Faster outcomes (prediction takes less time)
2. Functions well with both small and large data sets.
3. A single database scan is sufficient.
4. Functions well under various restrictions

Step 5: Patterns Prediction
Here, the educational data set is used by the algorithm to forecast the correlation between student conduct and performance.

5. RESULT ANALYSIS

ECLAT Algorithm

Table -1: Execution Time of Eclat Algorithm

No of Instances (records)	Execution Time (milli Secs)
Around 2k	557
Around 1k	495
Around records 500	465
100 records	445

Performance Factor:

Data Structure – array based

Memory Utilization – depends on the data set [less for small datasets]

Number of scans - single scan required

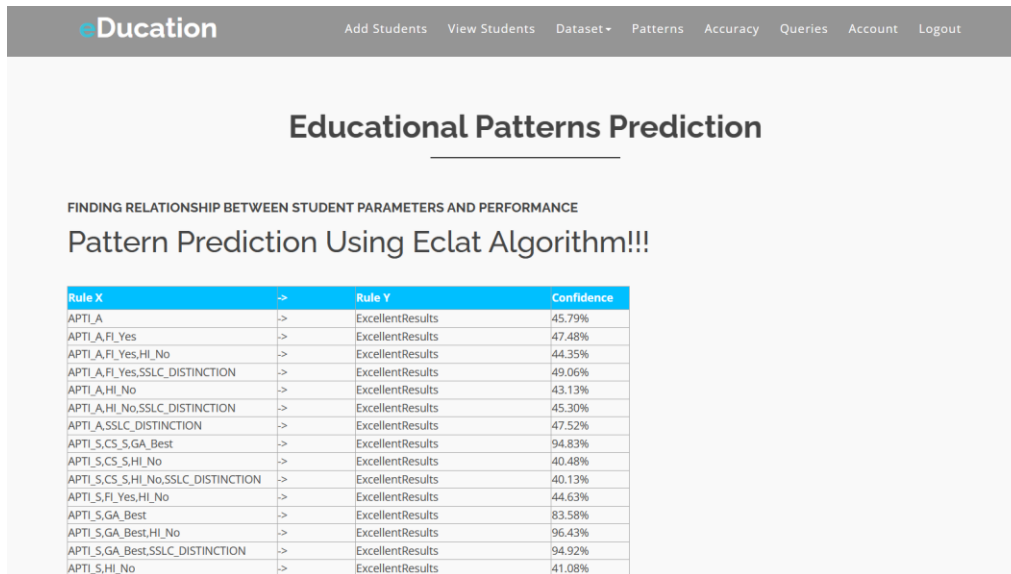Execution time - execution time depends on producing candidates

Fig -2: Educational Pattern Prediction Using Eclat Algorithm

Naive Bayes algorithm

In this instance, we develop a real-time application that benefits society. Using Microsoft technology, this project was constructed. We used the Naive Bayes technique to train educational datasets, and the results were excellent. The Naive Bayes method is designed to function with dynamic datasets. Our library contains written logic for the Naive Bayes method. About 99% of the answers are accurate, and the prediction process takes 42331 milliseconds.
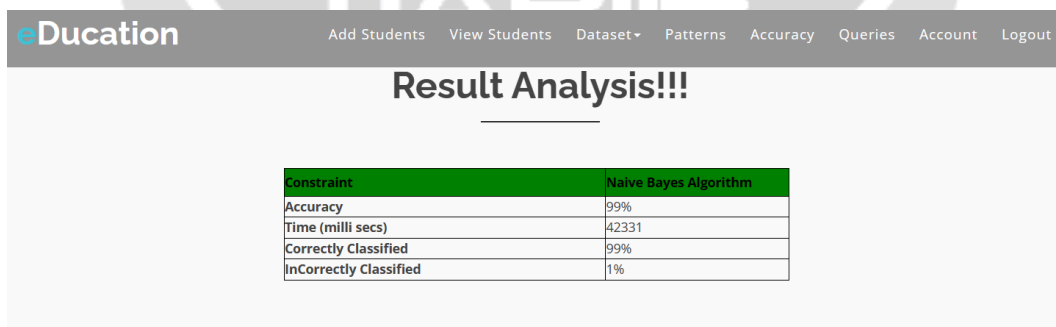

Fig -3: Accuracy Result for Naive Bayes Algorithm

6. CONCLUSION AND FUTURE ENHANCEMENT

6.1 Conclusion

The identification of several elements that influence a student's performance and learning style throughout their academic career. Examining students' mental health problems and poor academic performance is a difficult assignment

in the modern educational system. The data science method known as "Association Learning" is used by the system to identify trends. We employ the "apriori algorithm," "apriori TID algorithm," or "Eclat algorithm" to identify patterns. A real-time program designed to help instructors and universities understand student behavior trends is the proposed system. The system assists instructors in determining which elements have the most impact on students' academic achievement.

6.2 Future Enhancement

To identify further connected patterns, extra training datasets may be utilized. The behavior patterns of the pupils may be discovered using other algorithms, and the algorithms can be compared to see which one produces the best results.

7. REFERENCES

[1]. M. C. Nicoletti, M. Marques and M. P. Guimaraes, "A data mining approach for forecasting students' performance" [2018].
[2]. S. K. Pushpa, T. N. Manjunath, T. V. Mrunal, A. Singh and C. Suhas, "Class result prediction using machine learning"[2017].
[3]. Ganeshan, Kathiravelu & Li, Xiaosong. "An Intelligent Student Advising System Using Collaborative Filtering" [2015].
[4]. T. Mishra, D. Kumar, and S. Gupta, "Mining Students' Data for Prediction Performance" [2014].