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

Sakshi¹, Srushty G S², Yashaswini³, Yashaswini G⁴, Mr. B N Kiran⁵

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

⁵Assistant Professor, Information Science & Engineering, NIE Mysuru, Karnataka, India

ABSTRACT

Recognizing the several components that influence a student's learning style and academic achievement. Analyzing kids' low academic performance and mental health issues is a difficult task in the current educational system. Under the current system, keeping an eye on the traits and conduct of students is difficult. There isn't a piece of technology or a program available that can predict a student's academic performance or indicate how to modify it. Understanding how the factors affecting students' performance relate to their academic results is crucial in the current educational climate. 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 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

The conflict between the limited resources of student counsellors and the variety of student behaviors is a significant issue for student affairs management since it denies many potentially troublesome students the chance for early intervention. Student counsellors can now conduct quantitative analyses of students' school behaviors because of the rapid development of information technology in education since the turn of the twenty-first century and the creation of digital campuses. This is especially useful for early warning students who may have problems, as it allows for the application of analysis and early warning methods to alleviate contradictions. As modern college students who were raised in the Internet era, the Internet has a significant impact on their everyday life, education, and way of thinking.. This gives us the chance to use big data to analyze the behavioral features of their campus networks. The extraction

of valuable information from vast amounts of data in ever growing data categories and scales presents a problem for today's student counsellors. Working with novel techniques is also greatly facilitated by this chance.

Using a combination of big data science methods and big data thinking, this study started with actual work challenges and evaluated network behavior data from B college students. Due to the heavy network usage on campus, it also examined the traits of network behavior regulations among college students and pinpointed the students who need more care. The practical use of student data science work might be demonstrated by this study.

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 those users who possess comparable traits and actions will have comparable choices 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 capacity to forecast a pupil's achievement is crucial [4]. Students' academic achievements are influenced by various elements, such as social, environmental, psychological in nature and personal aspects. Mining data emerges as a useful instrument in achieving the following objective. Data mining techniques enable the discovery of hidden information Connections and trends among large datasets, facilitating informed procedures for determining decisions. 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 processing and administration of student databases. The method of mining information 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 with promptitude and accuracy. The primary goal is to determine which groupings within kids that need particular focus. With no school assessments of behavior, as well as these pupils' academic and other achievements variables suffer. The platform makes advantage of features like presence status, extracurricular actions, grades, abilities, and outcomes from the last semester, 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" for identify trends. We employ the "Eclat Algorithm" to discover patterns and "Naive Bayes Algorithm" to predict achievement by pupils on a particular

basis. The suggested framework would be built like a web-based program for assisting institutions as well as teachers in understanding student behavior trends.

3.1 Objectives

- The system being proposed is an evaluation, forecasting, and administration system for student behavior designed for use in schools and colleges.
- The suggested project analyzes student behavior using machine learning (ML) or data mining methods..
- To determine many elements that impact a the academic conduct of students and academic success during their academic career.
- Finding the various components which influence a the way a learner learns and achievement during the course of their time in college; providing reliable data from current students to manage connections with prospective students.
- The process of building a prediction model based on specified predictive factors through the use of machine learning techniques, specifically unsupervised learning.

4. METHODOLOGY

The best-known, recognizable, and a fundamental data science methodology is most likely affiliation (or link). To look for trends, we commonly establish a fundamental relationship between more than two identical objects of 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 following time they purchase strawberries.



Fig -1 Predictive Analytics Workflow

4.1 Pattern Prediction Process

Step 1: Data Collection

We're building a new system that involves data servers., which are utilized to keep information, 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: Preparing Data

Here, we retrieve and review server data. complete 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

The confidence level of a set is determined by comparing the entire number of interactions containing the component in the set to the entire number of interactions containing LHS..

Step 4: Rules of Association (Eclat Algorithm)

Probably Association data mining is the most widely used, widely recognized, and fundamental method. In order to find patterns, we simply correlate more than two, frequently identical things here.

Utilizing the Eclat algorithm, we evaluate and detect patterns in electronic commerce data. This is where we create instructive patterns.

The Eclat algorithm was chosen based on the reasons listed below.

- 1. Quicker results (less time spent on forecast)
- 2. Works effectively with a variety of data sets, big and small.
- 3. Just one database scan is necessary.
- 4. Works effectively with a range of limitations

Step 5: Predicting Patterns

The algorithm here forecasts the relationship between student behavior and performance using the educational data set

·

5. RESULT ANALYSIS

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: based on arrays

Memory Usage: varies according to the dataset (lower for smaller datasets).

Number of scans: just one scan is necessary

Execution time - The duration of execution is contingent upon the generation of candidates.

Ducation							
					-		
	Ed	lucational F	Patterns F	Predic	tion		
			EDEODMANCE				
FINDING RELATIONSHIP BET WE	ENSIO	PENT PARAMETERS AND P	ERFORMANCE				
Pattern Predict	tion	Using Eclat	Algorithm				
Fattern Fledic	lion	Using Lotat	Agonum	1			
Rule X	->	Rule Y	Confidence				
APTI_A	->	ExcellentResults	45.79%				
APTI_A APTI_A,FI_Yes	->	ExcellentResults ExcellentResults	45.79% 47.48%	_			
APTI_A APTI_A,FI_Yes APTI_A,FI_Yes,HI_No	-> ->	ExcellentResults ExcellentResults ExcellentResults	45.79% 47.48% 44.35%				
APTI_A APTI_A,FI_Yes APTI_A,FI_Yes,HI_No APTI_A,FI_Yes,SSLC_DISTINCTION	> > ->	ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45.79% 47.48% 44.35% 49.06%				
APTI_A APTI_A,FI_Yes APTI_A,FI_Yes,HI_No APTI_A,FI_Yes,SSLC_DISTINCTION APTI_A,HI_No	> > > >	ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45.79% 47.48% 44.35% 49.06% 43.13%				
APTI_A APTI_A.FI_Yes APTI_A.FI_Yes,HI_No APTI_A.FI_Yes,SSLC_DISTINCTION APTI_A.HI_No APTI_A.HI_No_SSLC_DISTINCTION	> > > > > >	ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45.79% 47.48% 44.35% 49.06% 43.13% 45.30%				
APTI_A APTI_AFI_Yes APTI_AFI_Yes,HI_No APTI_AFI_Yes,SSLC_DISTINCTION APTI_AHI_No APTI_AHI_NO_SSLC_DISTINCTION APTI_ASLC_DISTINCTION	-> -> -> -> -> -> ->	ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45.79% 47.48% 44.35% 49.06% 43.13% 45.30% 47.52%				
APTL, A APTL, AFL, Yes APTL, AFL, Yes, SLC, DISTINCTION APTL, AFL, Yes, SSLC, DISTINCTION APTL, AHL, No. APTL, AHL, NO. APTL, ASSLC, DISTINCTION APTL, SSLC, SGA, Best		ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45.79% 47.48% 44.35% 49.06% 43.13% 45.30% 47.52% 94.83%				
APTL, A APTL, AFL Yes APTL, AFL Yes, SLC, DISTINCTION APTL, AFL Yes, SSLC, DISTINCTION APTL, AHL, No, SSLC, DISTINCTION APTL, ASSLC, DISTINCTION APTL, SSLC, S, GA, Best APTL, SCL, SS, IL No		ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45.79% 47.48% 44.35% 49.06% 43.13% 45.30% 47.52% 94.83% 40.48%				
APTL, A APTL, AFL, Yes, APTL, AFL, Yes, APTL, AFL, Yes, SSLC_DISTINCTION APTL, AFL, Yes, SSLC_DISTINCTION APTL, AFL, No, SSLC_DISTINCTION APTL, ASSLC_DISTINCTION APTL, ASSLC_DISTINCTION APTL, SC, S, GA, Best APTL, SC, S, HIL No, SSLC_DISTINCTION		ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45,79% 47.48% 44.35% 49.06% 43.13% 45.30% 47.52% 94.83% 40.48% 40.13%				
APTL, A APTL, AFL, Yes APTL, AFL, Yes, SLC, DISTINCTION APTL, AFL, Yes, SSLC, DISTINCTION APTL, AHL, No. SSLC, DISTINCTION APTL, SSLS, CA, Best APTL, SC, SS, AG, Best APTL, SC, SS, AFL, No. SSLC, DISTINCTION APTL, SC, SSLH, NO. APTL, SC, SSLHI, NO. APTL, SFL, NO. APTL, SC, SSLHI, NO. APTL, SFL, SFL, APLL, NO. APTL, APLL, APL		ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45,79% 47,48% 44,35% 49,06% 43,13% 45,30% 47,52% 94,83% 40,48% 40,13% 44,63%				
APTL, A APTL, AFL, Yes APTL, AFL, Yes, SLC, DISTINCTION APTL, AFL, Yes, SSLC, DISTINCTION APTL, AHL, No, SSLC, DISTINCTION APTL, ASLC, DISTINCTION APTL, SC, S, GA, Best APTL, SC, S, SHL, No APTL, S, CS, SHL, NO APTL, S, SHL, SHL, ND		ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45,79% 47,48% 44,35% 49,06% 43,13% 45,30% 47,52% 94,83% 40,43% 40,13% 44,63% 83,58%				
APTL, A APTL, AFL, Yes APTL, AFL, Yes, SLC, DISTINCTION APTL, AFL, Yes, SLC, DISTINCTION APTL, AHL, NO, SSLC, DISTINCTION APTL, ASSLC, DISTINCTION APTL, ASSLC, SLR, NO, APTL, SLC, SLR, MO, APTL, SLC, SLR, NO, SSLC, DISTINCTION APTL, SLC, SLR, NO, SSLC, DISTINCTION APTL, SLR, Yes, HL, NO APTL, SLR, SLR, MO, APTL, SLR, BESLH NO		ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45.79% 47.48% 43.35% 49.06% 43.13% 45.30% 47.52% 94.83% 40.48% 40.13% 40.13% 83.58% 95.43%				
APTL, A APTL, AFL, Yes APTL, AFL, Yes, SLC, DISTINCTION APTL, AFL, Yes, SSLC, DISTINCTION APTL, AHL, No, SSLC, DISTINCTION APTL, ASSLC, DISTINCTION APTL, SC, SA, Best APTL, SC, SS, AG, Best APTL, SC, SS, AG, Best APTL, SC, SA, Best APTL, SC, SA, Best APTL, SCA, Best, HL, No APTL, SA, Best, HL, No APTL, SA, Best, HL, No APTL, SA, Best, SLC, DISTINCTION APTL, SA, Best, SLC, DISTINCTION		ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults ExcellentResults	45,79% 47,48% 43,35% 49,06% 43,13% 45,30% 47,52% 40,48% 40,13% 40,48% 40,13% 83,58% 96,43% 96,43%				



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.

Ducation	Add Students	View Students	Dataset -	Patterns	Accuracy	Queries	Account	Logout			
Result Analysis!!!											
	Constraint		Naivo F	aves Algorith	-						
	Accuracy	uracy 99%									
	Time (milli secs)		42331								
	Correctly Classified		99%								
	InCorrectly Classified		1%								

Fig -3: Accuracy Result for Naive Bayes Algorithm

6. CONCLUSION AND FUTURE ENHANCEMENT

6.1 Conclusion

The identification of several factors that affect an applicant's performance and preferences for learning during their years of study. Examining students' mental health problems and poor academic performance is a difficult assignment in the modern educational system. The field of data science referred to as "Association Learning" is used by the framework 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 can be found by utilizing other algorithms, and the best results may be obtained by comparing the algorithms.

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].