Identifying The Attributes That Affect Student Performance using Machine Learning

Manasi S[1], Tushar R Bharadwaj[2], Nuthan PM[3], Naganischay M[4]

[1] Student, [2] Student, [3] Student, [4] Student
Department of Computer Science and Engineering, JSS STU, Mysuru.

DOI: 10.29322/IJSRP.13.10.2023.p14238
https://dx.doi.org/10.29322/IJSRP.13.10.2023.p14238

Paper Received Date: 11th September 2023
Paper Acceptance Date: 17th October 2023
Paper Publication Date: 22nd October 2023

Abstract- The field of education has been revolutionized by advancements in big data techniques, enabling educators to gain precise and timely insights into students' behavioral patterns. This newfound capability is invaluable for identifying specific student groups that require targeted attention and transitioning from relying solely on qualitative empirical knowledge to incorporating scientific quantitative analysis in student affairs management.

To fully harness the potential of this revolution, a meticulously developed system was implemented to apply data mining's clustering method in analyzing the campus network behavior of 3,245 students in a particular grade at B University. Over a careful four-year period, a comprehensive dataset comprising 23.843 million Internet access records was collected. Through thorough analysis, it was discovered that the students could be categorized into four distinct groups based on their unique patterns of Internet access. Notably, the study successfully identified 350 students who exhibited remarkably high levels of network usage, allowing for a comprehensive examination of how their academic performance and other achievements were influenced.

This research, driven by the power of data, provides a practical and tangible demonstration of effectively utilizing data science in student affairs management. It presents compelling evidence that supports the accurate and scientific development of student affairs management practices by providing robust data support and generating invaluable insights.

Index Terms- Machine Learning , Naïve Baiyes , Eclat Algorithm

I. INTRODUCTION

Problem statement

The education sector faces a complex challenge in understanding how various factors impact students' learning behavior and academic performance. Within the current system, monitoring and comprehending students' characteristics and behavior present significant obstacles. Unfortunately, there is currently a scarcity of automated solutions or tools capable of accurately predicting or offeringsuggestions to enhance students' academic performance.

However, it remains vital to establish correlations between the factors influencing student performance and their academic outcomes to meet the demands of the modern educationsector.

Aim and Objectives

The main goal is to swiftly and accurately identify behavior patterns among students, facilitating the timely recognition of specific student groups that require focused attention. Neglecting the analysis of on-campus behaviorcan have negative consequences on students' academic achievements and overall performance. The system integrates various parameters, such as attendance, extracurricular activities, grades, technical skills, previous semester results, grasping capability, aptitude grade, and interaction with lecturers. By leveraging the data science technique called "Association Learning," thesystem aids lecturers in determining the most influential factors that impact student performance. It utilizes algorithms like "Apriori", "Apriori TID" or "Eclat" to uncover relevant patterns. Designed as a real-time application, the proposed system offers valuable support to colleges and lecturers in understanding students' behavioral patterns.

Existing methods

While the Student Management System efficiently manages essential student academic information, such as marks, attendance, admissions, fees, sports, andplacement records, it lacks the capacity to provide insights into students' health concerns. In contrast, the General Counseling system relies on manual counseling sessions conducted by college staff to address students' issues. However, this process can be time-consuming and heavily reliant on human intervention Additionally, Human Advisors, who possess expertise in offering guidance andsuggestions, rely on a manual system that necessitates time, experience, and financial resources.

These systems have limitations, including being time- consuming, dependent on experienced individuals, and requiring expert involvement. Furthermore, manual advising may introduce inaccuracies, leading to reduced overall efficiency in the process.

Limitations of existing methods

There are some unclear points that need further study. Forinstance, cluster 2 (20 students) showed “low” level on related self-efficacy (sef01, sef05, and sef08 = 2) but wasa “pass” group.

Cluster 6 (14 students) possessed “middle” level on the same attributes (sef01, sef05, and sef08 = 3) while belonging to “fail” group. One possible explanation may be due to the small size of the sample. For ML models, collecting more participants to improve the performance of the model shouldbe meaningful for future work. Also, adding more valuable parameters such as the learning style of individual students tothe model is worth exploring.
Algorithms generates graphical outputs, not suitable for real time. Requires very huge amount of data as it uses data mining techniques v Accepts only 2 labels.
Takes more time for processing data.

Proposed solution

The main goal is to accurately and promptly identify patternsof behavior among students, with a specific emphasis on identifying groups of students that require timely attention. Neglecting the analysis of campus behavior can have negative consequences for students' achievements and overall performance. The system incorporates a wide range of parameters, including attendance status, extracurricular activities, grades, technical skills, previous semester results, grasping capability, aptitude grade, and interaction with lecturers. Through the implementation of the data science technique known as "Association Learning," the system supports lecturers in identifying the key factors that have a significant impact on student performance. It employs algorithms such as "Apriori", "Apriori TID" or "Eclat" to uncover relevant patterns. Designed as a real-time application, the proposed system provides valuable insights tocolleges and lecturers regarding students' behavior patterns.

II. LITERATURE STUDIES

An Intelligent Student Advising System Using CollaborativeFiltering
-- Kathiravelu Ganeshan Department of
Computing Unitec Institute of Technology Auckland, NewZealand kganeshan@unitec.ac.nz

The intelligent student advising online application we propose utilizes collaborative filtering, a widely used technique in recommendation systems. Collaborative filtering works on the assumption that users who have similar characteristics and behaviors tend to share similar preferences. In our system, students are grouped together based on their similarities, enabling personalized advice that caters to the characteristicsand preferences of their respective groups. For example, if a student shares similarities with a particular group of students,the system can suggest a course that is preferred by that specific group.
However, it is important to acknowledge certain limitations of our system. Firstly, the system primarily focuses on predicting suitable courses for students and may not accurately predict student outcomes. Secondly, not all student behaviors are directly linked to course advising, which implies that there may be additional factors influencing a student's academic success that our system does not consider.

Lastly, the grouping process may encounter limitations due to inadequate data, potentially leading to constraints in the course recommendations provided by the system.

Mining Students’ Data for Performance Prediction
---Sangeeta Gupta Professor, Management and IT Bhagwan Parushram Institute of Technology, Delhi

Accurately predicting student performance holdsimmense importance in educational settings, as itencompasses a wide array of factors including personal,social, psychological, and environmental variables. Datamining has emerged as a promising tool to achieve thisgoal by uncovering hidden patterns and relationshipswithin extensive datasets, thereby aiding decision-makingprocesses. The education sector possesses vast amounts ofdata containing valuable information, making it necessaryto leverage information and communication technologyfor efficient and cost-effective data capture andcompilation. As educational databases continue to growrapidly, the management and processing of student data become crucial for understanding loyal students andestablishing effective connections with prospectivestudents.
However, it is crucial to recognize certain limitations. Firstly, the system primarily relies on student behaviors to predict student performance, which may not be suitable for accurately predicting class results. Secondly, the inclusion of various irrelevant parameters, such as the father's income, mother's income, and qualifications, in student performance prediction may lead to less precise outcomes.

III.SYSTEM REQUIREMENT SPECIFICATION

Functional Requirements:

The admin should be able to log in to the application usingtheir admin ID and password.
The admin should have the ability to add students from different departments to the system.
The admin should be able to set unique IDs and passwordsfor individual students.
The admin should be able to manage the training datasetsused in the project.
The core module of the system should find the relationship between student parameters and performance using

unsupervised learning techniques.
The admin should be able to view students' queries and send replies.
Users should be able to log in by specifying their student ID and password.
Users should have the ability to post queries to the adminif needed.

Non-Functional Requirements:

Availability: The system should be browser-based andavailable 24/7, allowing access from different locations.

Reliability: The application should provide services according to users' satisfaction and interests, meeting their requirements and being user-friendly, ensuring reliability compared to other applications.
Scalability: The system should be scalable, capable of handling dynamic data using datascience techniques. Any changes in data should not require modifications to the coding, and the system should generate output based on the data provided.
Security: The system should be browser-based and deployed on a secure server. Only authorized users should have access to the servers, and the data should be stored securely in SQL server with appropriate authentication measures.
Performance: The system should use data science techniquesand efficient coding practices, such as advanced concepts in C#, to ensure high performance and efficiency.
Quality of Service: Regular updates and easy maintenance arevital elements of the software. It is essential to design the application with a focus on flexibility, allowing for seamless integration of future modifications and enhancements.

Hardware Requirements :

Processor - Pentium IV onwardsProcessor Speed - 2.4GHz RAM -2GB+
Hard disk space -40GB+
Standard PC Configuration to carry challenging computationSystem Design

Software Requirements :

Operating System – Windows XP Version and higherDesign Tool – Visual Studio 2010
Front End – ASP.NET 4.0Language – C#
SQL Server
Data Access Technology - ADO.NET

IV. TOOLS AND TECHNOLOGIES

Microsoft has built a strong reputation for introducing state- of-the-art technologies, often accompanied by jargon that can be perplexing. One such technology is .NET, which is not a standalone application but rather a collection of technologies bundled together under a unified term.The .NET Frameworkis composed of multiple components, which encompass well-known programming languages like C# and VB .NET. Additionally, it incorporates ASP.NET, which serves as a hosting engine for the creation of dynamic web pages and web services. Furthermore, ADO.NET offers a sturdy model for interacting with databases, while developers can takeadvantage of a comprehensive class library that provides a diverse set of tools for tasks such as file manipulation and XML parsing.

The primary objective of the .NET Framework is to streamline the deployment of applications across enterprises and facilitate scalability to cater to varying requirements. It plays a pivotal role in contemporaryprogramming environments, offering versatility for a wide range of development purposes. Similar to other Microsoft products, the .NET framework is renowned for its user-friendly nature, simplifying application development and enhancement when leveraging .NET technologies.
Microsoft has introduced C# (pronounced as C Sharp) as a new programming language within the comprehensive .NET Framework. C# represents a significant advancement in programming languages, characterized by its modernity, object-oriented approach, and emphasison type safety. Developed by Microsoft specifically for the .NET Framework, C# combines established features with cutting-edge innovations, providing an efficient and user-friendly solution for programming across diversecomputing environments, including Windows and the Internet.
C# is meticulously designed to leverage the capabilities ofthe Microsoft .NET platform, which encompasses a widerange of tools and services for computing and communication. Evolving from the foundations of C and C++, C# brings platform independence, enabling seamless execution on major hardware and software platforms.
In contrast to its predecessors, C# not only offers supportfor object-oriented principles but also enforces them at the language level. Moreover, C# places a strong emphasis on security by integrating security measures directly into the language, ensuring a robust programming framework. Simplicity is a key feature of C#. Building upon the syntax and features of C++, C# eliminates elements that contribute to errors and program complexity. It introducesnumerous language features that surpass the capabilities of C and C++, including built-in support formultithreading, which was absent in the older

languages. Let's explore ASP.NET, a powerful server-side technology renowned for its ability to develop dynamic web pages. Itplays a vital role as a key component of the comprehensive .NET framework, serving as a versatile toolkit for creating various applications, particularly thoseintended for the web.

ASP.NET excels in enhancing developer productivity by combining programming ease, language flexibility, and a comprehensive class framework. It ensures exceptional performance and scalability through compiled execution and output caching. Additionally, ASP.NET places a strong emphasis on reliability, actively preventing memory leaks, deadlocks, and offering crash protection. This technology streamlines deployment and enables dynamic application updates.

There are numerous advantages to using ASP.NET. It seamlessly integrates with the .NET framework, grantingaccess to a wide range of features such as multi-language support, compiled code, automatic memory management, and an extensive.NET base class library.

By leveraging ASP.NET, developers can create robust functionality driven by databases, enhance web application performance through compiled code and caching, and benefit from the flexibility of supporting multiple programming languages.

In the realm of internet applications and database access, Microsoft provides ADO.NET as the data access model. ADO.NET facilitates smooth interaction with relational databases and other data sources. ASP.NET applications utilize ADO.NET to establish communication with databases, utilizing optimized namespaces tailored for various data providers, including Microsoft SQL Server, OLEDB, ODBC,and Oracle.

Among the offerings of Microsoft SQL Server 2005, SQL Server Express stands out as a distinct version. It offers a free, user-friendly, and streamlined edition of SQL Server. SQL Server Express caters to the needs of small businesses with limited resources, nonprofessional developers, andindividuals involved in web or client application development. It serves as an excellent training tool for the full version of SQL Server and is particularly well-suited for web application development, especially when combined with Visual Web Developer Express and Visual Basic Express.

V. SYSTEM DESIGN

The design phase is a critical step in planning a solution for the outlined requirements. It acts as a vital link between the problem domain and the solution domain, guiding the fulfillment of the specified needs. The quality of the software depends greatly on the system's design, which significantly impacts subsequent phases like testing and maintenance.

During the design phase, three key deliverables are typically produced: architecture design, high-level design, and detailed design.

Architecture Design: This stage involves understanding the system as a composition of components and their interactions.It includes identifying the major components or subsystems and establishing their connections, with a particular focus onthe essential components.

High-Level Design: In this stage, the necessary modules for system development are identified, along with their specifications. Key aspects such as major data structures, file formats, and output formats are determined. The primary objective is to identify the core modules required.

Detailed Design: This phase entails specifying the internal logic of each module. It provides a detailed understanding of how the logic for each module will be implemented in the software.

Design methodologies are systematic approaches that leverage techniques and guidelines to create effective designs. Many methodologies place significant emphasis on high-level design.

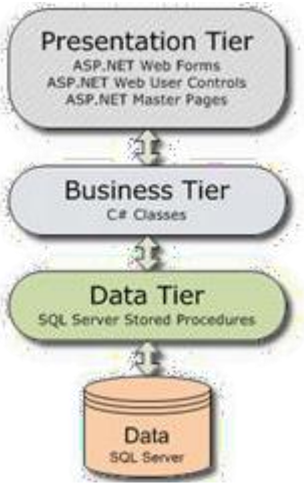
In the mentioned project, a three-tier architecture is employed, consisting of the following layers:

Data Layer: This layer is responsible for retrieving data from the database and serving it to the application. It is often implemented as a separate component to enable logical data reuse and enhance maintainability.

Business Layer: The business layer acts as an intermediary between the data layer and the presentation layer. It validates input conditions, ensures data integrity,and applies business rules to make informed decisions about the data. Centralizing logic in this layer maximizesreusability.

Presentation Layer: This layer encompasses the user interface (UI) of the application, such as an ASP.NET website or a Windows Forms application. It is the layer through which users directly interact with the system andplays a crucial role in delivering a positive userexperience. Effective design of the presentation layer is essential for ensuring user satisfaction.

By adopting a three-tier architecture, the project achieves separation of concerns, improves maintainability, promotes reusability, and ensures a clear division of responsibilities among different layers.

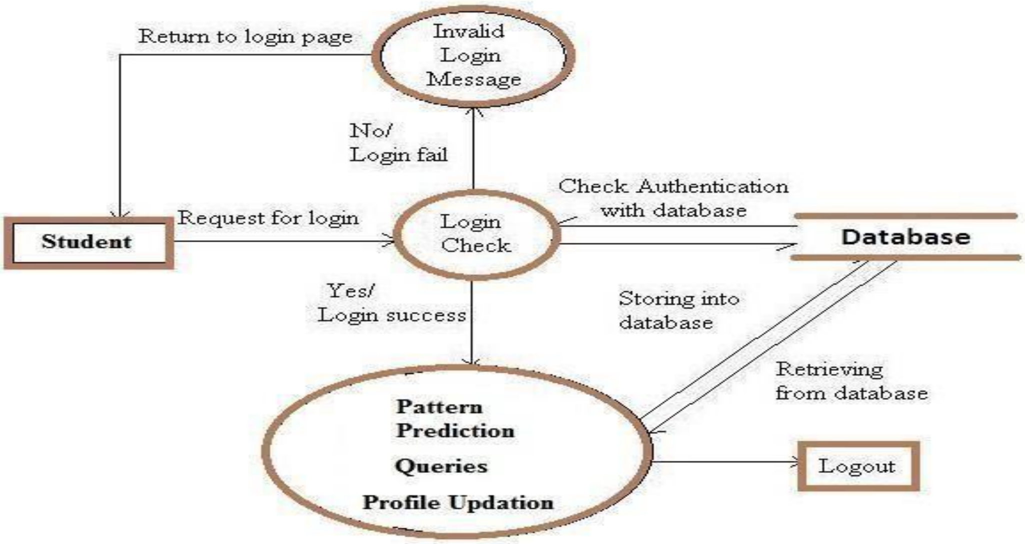


In the system architecture, the presentation tier plays a crucial role in managing the User Interface (UI) elementsof the site and facilitating interactions between visitors and the client's business. It commonly utilizestechnologies like ASP.NET Web Forms, Web User Controls, and ASP.NET Master Pages.

The business tier is responsible for receiving requests from the presentation tier and applying the necessary business logic to process them. It consists of C# classes that handle the requested operations and deliver the resulting outcomes back to the presentation tier.

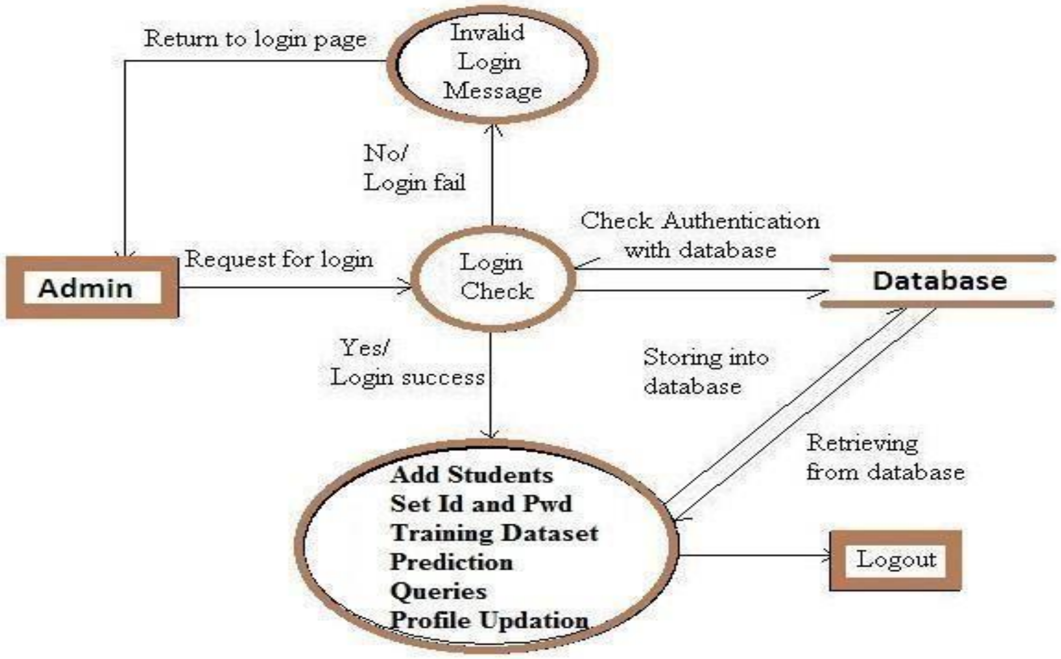
The data tier takes on the responsibility of storing the application's data and providing it to the business tier as required. SQL Server Stored Procedures are frequently employed in this architecture to interact with the database,ensuring efficient data retrieval and updates.

As we transition to the high-level design phase, the use ofa Data Flow Diagram (DFD) becomes instrumental in visually illustrating the flow of data throughout the information system. The DFD serves as a graphical representation, showcasing how data items move betweenexternal data sources, internal data stores, and the system's processes.



https://dx.doi.org/10.29322/IJSRP.13.10.2023.p14238

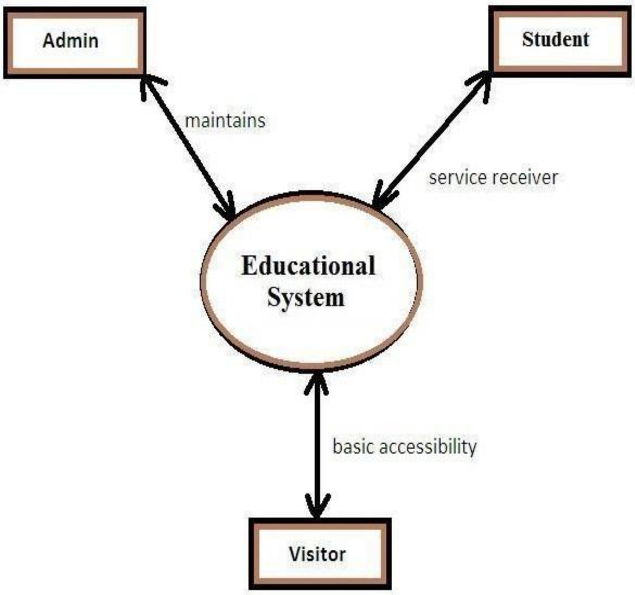
DFDs primarily emphasize visualizing the flow and transformation of data, rather than emphasizing the timing or sequence of operations. They employ symbols to represent distinct components:
Processes: These components represent the transformation ofdata values and can be seen as functions without side effects.
Data Flows: These connections establish links between the output of one object or process and the input of another, visually capturing the intermediate data values involved in the computation. Data flows are represented using labeled arrows.
Actors: Actors are active objects that drive the data flow graphby producing or consuming data values. They are connected to the inputs and outputs of the data flow and serve as sourcesand sinks of data.
Data Store: A data store is a passive object within the DFD that serves as a repository for storing data, ensuring it can be accessed when needed. It does not perform operations autonomously but responds to requests for storing or retrieving data.
To summarize, during the high-level design phase, the creation of a Data Flow Diagram provides a visual representation of the data flow within the system. It encompasses processes, data flows, actors, and data stores, with a focus on illustrating the movement and transformationof data.



In the design process, it is common to begin by creating a context-level Data Flow Diagram (DFD) that illustrates how the system interacts with external agents as data sources or recipients. The context diagram, also known as the 'Level 0 DFD,' presents the system as a single process and focuses onthe flow of data across the system boundary. Its primary purpose is to provide an overview of the system's interactionwith the external environment while maintaining the confidentiality of

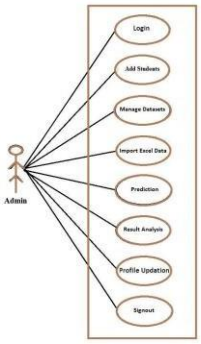
its internal structure.
Subsequently, the design process proceeds to develop a Level 1 DFD, which offers more detailed information about the system being analyzed. The Level 1 DFD decomposes the system into subsystems (processes) responsible for handling specific data flows to or from external agents. These subsystems collectively perform all the necessary functions of the system. Moreover, the Level 1 DFD identifies essential internal data stores required for the system's operation and visually represents the flow of data between different systemcomponents.

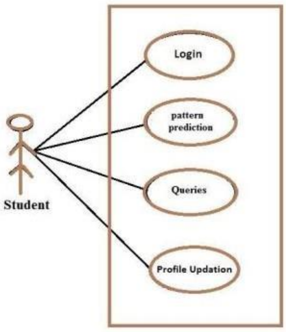
The Level 1 DFD represents the primary processes and their relationships within the system, providing a higher level of detail compared to the context diagram. It ensures consistency with the context diagram in terms of external entities and data flows while allowing for further decomposition and elaboration of those components. This enables a more comprehensive understanding of the system's internal operations.



During the detailed design phase, it is customary to employ a use case diagram to visually depict the system's functionality. This diagram, which belongs to the Unified Modeling Language (UML) and falls under the category of behavioral diagrams, provides an overview of thesystem's features by illustrating actors, their goals(represented as use cases), and the relationships between them. The primary goal of a use case diagram is to present the system functions performed for each actor and illustrate the actors' roles within the system. However, interactions among actors are typically not included in thediagram. If these interactions are essential for acomprehensive description of the desired systembehavior, it may be necessary to reconsider the system oruse case boundaries. Alternatively, assumptions about actor interaction can be incorporated into the use cases. Use cases, depicted as horizontal ellipses, describe a series of actions that deliver value to an actor. Actors can beindividuals, organizations, or external systems that play arole in one or more interactions with the system.

To define the system's scope, a rectangular box called the system boundary is drawn around the use cases. This box helps distinguish the functionality that falls within the system's scope
(inside the box) from the functionality that lies outside the system's scope (outside the box).





https://dx.doi.org/10.29322/IJSRP.13.10.2023.p14238

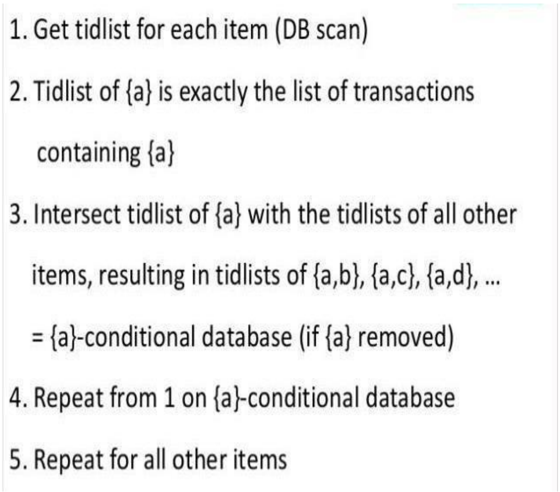
VI. SYSTEM IMPLEMENTATION

This web application is developed using object-oriented programming, which allows for the modularization of programs by separating data and functions into distinct memory areas. This approach facilitates the creation of module copies as needed.

The object-oriented paradigm focuses on data rather than procedures and organizes programs into objects. It designs data structures to define object characteristics and links methods that operate on object data within the data structure. This paradigm enables communication between objects through methods, allows for easy addition of new data and methods, follows a bottom-up approach in program design, and hides data from external functions.

The project follows a three-tier architecture, with ASP.NET employed for the presentation layer, C# classes for the business logic layer, a Table Adapter for the data tier, and MS SQL Server 2005 serving as the backend database. Machine learning (ML) methodology plays a crucial role in developingsystems capable of learning from data. The three main types of ML are supervised machine learning, which relies on labeled input examples; unsupervised machine learning, which focuses on discovering patterns without labels; and semi-supervised machine learning, which combines aspects of both supervised and unsupervised learning. Association learning is a widely recognized data science technique that aims to identify correlations between items. It involves establishing connections between items of the same type to discover patterns. For example, market-basket analysis trackspurchasing habits to uncover relationships between products. In this project, the Apriori Algorithm/Eclat Algorithm is employed as an association learning algorithm to predict the relationship between student behavior and performance usingeducational datasets.

Eclat Algorithm



1. Get tidlist for each item (DB scan)
2. Tidlist of {a} is exactly the list of transactions containing {a}
3. Intersect tidlist of {a} with the tidlists of all other items, resulting in tidlists of {a,b}, {a,c}, {a,d}, … = {a}-conditional database (if {a} removed)
4. Repeat from 1 on {a}-conditional database
5. Repeat for all other items

Sample Example

C1

Items	Support
A	T1,T2,T3
B	T3,T4
C	T1,T2,T3
D	T1
E	T2,T3,T4

L1

Items	Support
A	T1,T2,T3
B	T3,T4
C	T1,T2,T3
E	T2,T3,T4

C2

Items	Support
AB	T3
AC	T1,T2,T3
AE	T2,T3
BC	T3
BE	T3,T4
CE	T2,T3

L2

Item	Support
AC	T1,T2,T3
AE	T2,T3
BE	T3,T4
CE	T2,T3

C3

Items	Support
ACE	T2,T3
ABC	T3
ABE	T3
BCE	T3

L3

Items	Support
ACE	T2, T3

https://dx.doi.org/10.29322/IJSRP.13.10.2023.p14238

Minimum Support Count = 2Minimum Confidence = 80%

TID	Item sets
T1	A,C,D
T2	A,C,E
T3	A,B,C,E
T4	B,E

Item set □A, B, C, D, and E - Student parameters and results

Item	Support
A	T1,T2,T3
B	T3,T4
C	T1,T2,T3
E	T2,T3,T4
AC	T1,T2,T3
AE	T2,T3
BE	T3,T4
CE	T2,T3
ACE	T2,T3

https://dx.doi.org/10.29322/IJSRP.13.10.2023.p14238

GENERATE CONFIDENCE:

RULE X - RULE Y	CONFIDENCE
{A} - {C}	100.00%
{C} - {A}	100.00%
{A} - {E}	66%
{E} - {A}	66%
{B} - {E}	100%
{E} - {B}	66%
{C} - {E}	66%
{E} - {C}	66%
{A} - {CE}	66%
{C} - {AE}	66%
{E} - {AC}	66%
{CE} - {A}	100%
{AE} - {C}	100%
{AC} - {E}	66.00%

STRONG ASSOCIATION RULE:

{B} --> {E}
{CE} --> {A}
{AE} --> {C}
{A} --> {C}
{C} --> {A}

Classification Technique :

Classification is the process of creating a model or function that can differentiate and describe different data classes or concepts. This model is built by analyzing a training dataset with known class labels, enabling the prediction of class labels for new or unfamiliar data objects.

For instance, let's consider a scenario where the sales manager of All Electronics wants to categorize items in the store based on their response to a sales campaign, classifying them as having a good response, mild response, or no response. Descriptive features like price, brand, place of manufacture, type, and category are utilized to construct a model for each class.

The ultimate objective is to develop an effective classification model that accurately separates each class, providing a structured representation of the dataset.

The Naïve Bayes algorithm is a widely employed method forclassification and follows these steps:

1. Retrieving Data: Gather the required data from storageservers, such as databases or cloud storage.
2. Calculating Probabilities: Compute the probabilities of eachattribute value using a specific formula. This formula takes into account the different classes present in the dataset.
3. Probability Formula: Apply the formula P(attribute value(ai) | subject value(vj)) = (n_c + mp) / (n + m), where: n represents the number of training examples where v = vj. nc represents the number of examples where v = vj and a = ai.p is an a priori estimate for P(ai | vj).

m is the equivalent sample size.

4. Multiplying Probabilities: Multiply the calculatedprobabilities by p. The results for each attribute are multiplied by p, yielding final values that are utilized for classification. 5.Classification: Compare the obtained values and assign the attribute values to predefined classes.

In summary, classification involves developing a model basedon known class labels and using it to predict the class labels of unknown data objects. The Naïve Bayes algorithm follows these steps to calculate probabilities and perform classifications based on them.

Sample Example Parameters [m=3]Gender Family_Issues Financial_Issues

Outcomes – Bad, Good [p=1/2=0.5]New Student data – Akash

Parameters (Gender - Male, Family_Issues - Yes,Financial_Issues - No) Outcome = ?

Naive Bayes works based on probability calculation, It uses

Name	Gender	Family Issues	Financial Issues	Result
Anil	Male	Yes	Yes	Bad
Ajay	Male	No	No	Bad
Aruni	Female	Yes	No	Good
Kumari	Female	Yes	Yes	Bad
Naveen	Male	No	No	Good

Formula; P=[n_c + (m*p)]/(n+m)

BAD	GOOD

Male
P=[n_c + (m*p)]/(n+m)n=3,
n_c=2,m=3,p=0.5
p=[2+(3*0.5)]/(3+3) p=0.58
	Male
	P=[n_c+(m*p)]/(n+m)
	n=3,n_c=1,m=3,p=0.5
	p=[1+(3*0.5)]/(3+3) p=0.41

Yes
P=[n_c + (m*p)]/(n+m)n=3,
n_c=2,m=3,p=0.5
p=[2+(3*0.5)]/(3+3)
p=0.58
	Yes
	P=[n_c+(m*p)]/(n+m)
	n=3,n_c=1,m=3,p=0.5
	p=[1+(3*0.5)]/(3+3) p=0.41

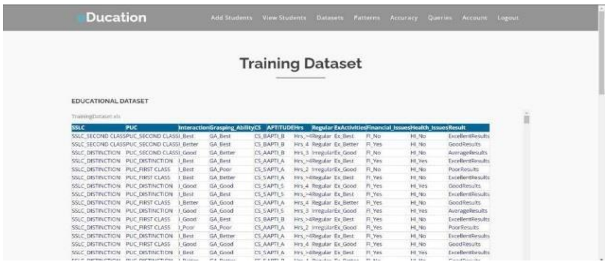
No P=[n_c+(m*p)]/(n+m)
n=3,n_c=1,m=3,p=0.5
p=[1+(3*0.5)]/(3+3)
p=0.41
	No
	P=[n_c + (m*p)]/(n+m) n=3, n_c=2,m=3,p=0.5
	p=[2+(3*0.5)]/(3+3)p=0.58

Bad – 0.58 * 0.58 * 0.41 * 0.5 (p) =0.068
Good– 0.41 * 0.41 * 0.58 * 0.5 (p)=0.048

So this new student is classified to BadSince Bad > Good

https://dx.doi.org/10.29322/IJSRP.13.10.2023.p14238

VII RESULTS

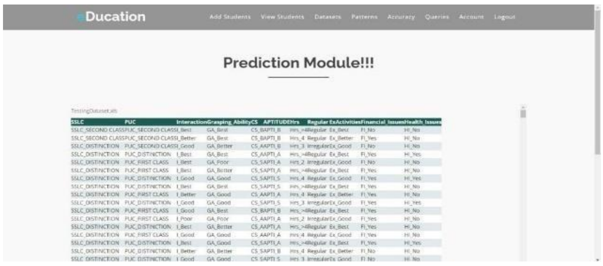


Training Dataset

EDUCATIONAL DATASET

TrainingDataset.xls

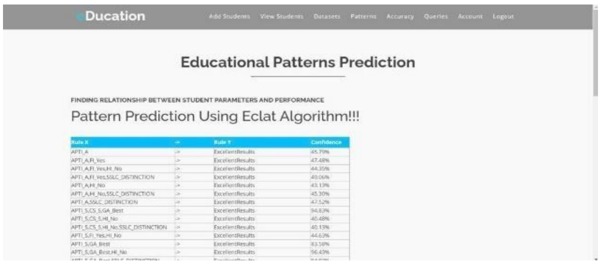
SSLC	PUC	Interaction	Grasping_Ability	CS	APTITUDE	Hrs	Regular	ExActivities	Financial_Issues	Health_Issues	Result
SSLC_SECOND CLASS	PUC_SECOND CLASS	I_Best	GA_Best	CS_B	APTI_B	Hrs_>4	Regular	Ex_Best	FI_No	HI_No	ExcellentResults
SSLC_SECOND CLASS	PUC_SECOND CLASS	I_Better	GA_Best	CS_B	APTI_B	Hrs_4	Regular	Ex_Better	FI_Yes	HI_No	GoodResults
SSLC_DISTINCTION	PUC_SECOND CLASS	I_Good	GA_Better	CS_A	APTI_B	Hrs_3	Irregular	Ex_Good	FI_No	HI_No	AverageResults
SSLC_DISTINCTION	PUC_DISTINCTION	I_Best	GA_Best	CS_A	APTI_A	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_Yes	ExcellentResults
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Best	GA_Poor	CS_S	APTI_A	Hrs_2	Irregular	Ex_Good	FI_No	HI_No	PoorResults
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Best	GA_Better	CS_S	APTI_A	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_No	ExcellentResults
SSLC_DISTINCTION	PUC_DISTINCTION	I_Good	GA_Good	CS_S	APTI_S	Hrs_4	Regular	Ex_Good	FI_Yes	HI_Yes	GoodResults
SSLC_DISTINCTION	PUC_DISTINCTION	I_Best	GA_Best	CS_S	APTI_S	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_No	ExcellentResults
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Better	GA_Good	CS_A	APTI_A	Hrs_4	Regular	Ex_Better	FI_Yes	HI_No	GoodResults
SSLC_DISTINCTION	PUC_DISTINCTION	I_Good	GA_Good	CS_S	APTI_S	Hrs_3	Irregular	Ex_Good	FI_Yes	HI_Yes	AverageResults
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Good	GA_Best	CS_B	APTI_B	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_No	ExcellentResults
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Poor	GA_Poor	CS_A	APTI_A	Hrs_2	Irregular	Ex_Good	FI_Yes	HI_No	PoorResults
SSLC_DISTINCTION	PUC_DISTINCTION	I_Best	GA_Better	CS_A	APTI_A	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_No	ExcellentResults
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Good	GA_Good	CS_A	APTI_A	Hrs_4	Regular	Ex_Good	FI_Yes	HI_No	GoodResults
SSLC_DISTINCTION	PUC_DISTINCTION	I_Best	GA_Good	CS_S	APTI_A	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_Yes	ExcellentResults



Prediction Module!!!

TestingDataset.xls

SSLC	PUC	Interaction	Grasping_Ability	CS	APTITUDE	Hrs	Regular	ExActivities	Financial_Issues	Health_Issues
SSLC_SECOND CLASS	PUC_SECOND CLASS	I_Best	GA_Best	CS_B	APTI_B	Hrs_>4	Regular	Ex_Best	FI_No	HI_No
SSLC_SECOND CLASS	PUC_SECOND CLASS	I_Better	GA_Best	CS_B	APTI_B	Hrs_4	Regular	Ex_Better	FI_Yes	HI_No
SSLC_DISTINCTION	PUC_SECOND CLASS	I_Good	GA_Better	CS_A	APTI_B	Hrs_3	Irregular	Ex_Good	FI_No	HI_No
SSLC_DISTINCTION	PUC_DISTINCTION	I_Best	GA_Best	CS_A	APTI_A	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_Yes
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Best	GA_Poor	CS_S	APTI_A	Hrs_2	Irregular	Ex_Good	FI_No	HI_No
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Best	GA_Better	CS_S	APTI_A	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_No
SSLC_DISTINCTION	PUC_DISTINCTION	I_Good	GA_Good	CS_S	APTI_S	Hrs_4	Regular	Ex_Good	FI_Yes	HI_Yes
SSLC_DISTINCTION	PUC_DISTINCTION	I_Best	GA_Best	CS_S	APTI_S	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_No
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Better	GA_Good	CS_A	APTI_A	Hrs_4	Regular	Ex_Better	FI_Yes	HI_No
SSLC_DISTINCTION	PUC_DISTINCTION	I_Good	GA_Good	CS_S	APTI_S	Hrs_3	Irregular	Ex_Good	FI_Yes	HI_Yes
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Good	GA_Best	CS_B	APTI_B	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_No
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Poor	GA_Poor	CS_A	APTI_A	Hrs_2	Irregular	Ex_Good	FI_Yes	HI_No
SSLC_DISTINCTION	PUC_DISTINCTION	I_Best	GA_Better	CS_A	APTI_A	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_No
SSLC_DISTINCTION	PUC_FIRST CLASS	I_Good	GA_Good	CS_A	APTI_A	Hrs_4	Regular	Ex_Good	FI_Yes	HI_No
SSLC_DISTINCTION	PUC_DISTINCTION	I_Best	GA_Good	CS_S	APTI_A	Hrs_>4	Regular	Ex_Best	FI_Yes	HI_Yes
SSLC_DISTINCTION	PUC_DISTINCTION	I_Better	GA_Better	CS_S	APTI_B	Hrs_4	Regular	Ex_Better	FI_No	HI_No
SSLC_DISTINCTION	PUC_DISTINCTION	I_Good	GA_Good	CS_S	APTI_S	Hrs_3	Irregular	Ex_Good	FI_No	HI_No

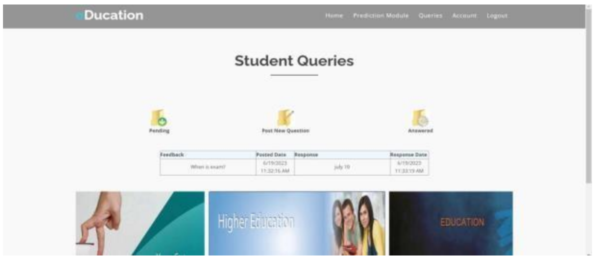


Educational Patterns Prediction

FINDING RELATIONSHIP BETWEEN STUDENT PARAMETERS AND PERFORMANCE

Pattern Prediction Using Eclat Algorithm!!!

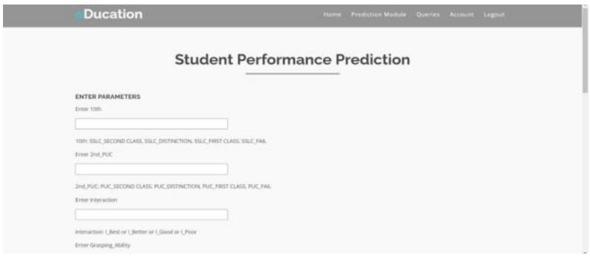
Rule X	->	Rule Y	Confidence
APTI_A	->	ExcellentResults	45.70%
APTI_A,FI_Yes	->	ExcellentResults	47.48%
APTI_A,FI_Yes,HI_No	->	ExcellentResults	44.35%
APTI_A,FI_Yes,SSLC_DISTINCTION	->	ExcellentResults	49.06%
APTI_A,HI_No	->	ExcellentResults	43.13%
APTI_A,HI_No,SSLC_DISTINCTION	->	ExcellentResults	45.30%
APTI_A,SSLC_DISTINCTION	->	ExcellentResults	47.52%
APTI_S,CS_S,GA_Best	->	ExcellentResults	94.83%
APTI_S,CS_S,HI_No	->	ExcellentResults	40.48%
APTI_S,CS_S,HI_No,SSLC_DISTINCTION	->	ExcellentResults	40.13%
APTI_S,FI_Yes,HI_No	->	ExcellentResults	44.63%
APTI_S,GA_Best	->	ExcellentResults	83.58%
APTI_S,GA_Best,HI_No	->	ExcellentResults	96.43%



Student Queries

Feedback	Posted Date	Response	Response Date
When is exam?	6/19/2023 11:32:16 AM	July 19	6/19/2023 11:33:19 AM

The above figure illustrates the snapshots of the student loginpage, where students can log in to the website by giving their details, and student queries, where the students are allowed to put their queries to get the answers from the respective faculties.



Student Performance Prediction

ENTER PARAMETERS

Enter 10th

10th: SSLC_SECOND CLASS, SSLC_DISTINCTION, SSLC_FIRST CLASS, SSLC_FAIL

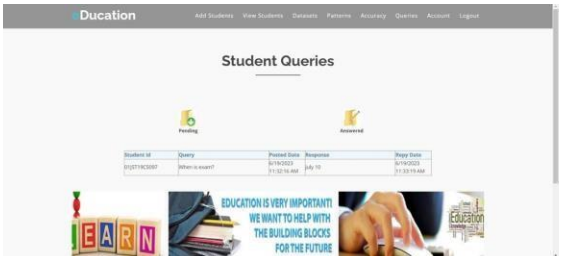
Enter 2nd_PUC

2nd_PUC: PUC_SECOND CLASS, PUC_DISTINCTION, PUC_FIRST CLASS, PUC_FAIL

Enter Interaction

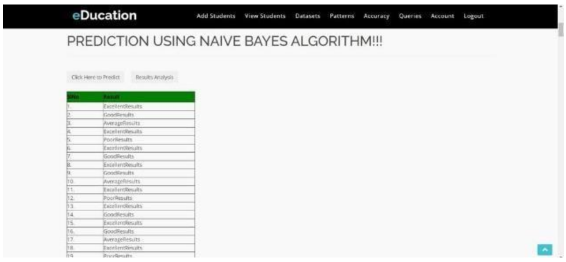
Interaction: I_Best or I_Better or I_Good or I_Poor

Enter Grasping_Ability



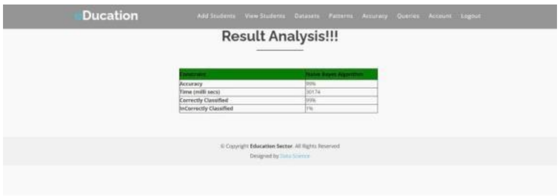
Student Queries

Student Id	Query	Posted Date	Response	Repy Date
01JST19CS097	When is exam?	6/19/2023 11:32:16 AM	July 10	6/19/2023 11:33:19 AM



PREDICTION USING NAIVE BAYES ALGORITHM!!!

SlNo	Result
1.	ExcellentResults
2.	GoodResults
3.	AverageResults
4.	ExcellentResults
5.	PoorResults
6.	ExcellentResults
7.	GoodResults
8.	ExcellentResults
9.	GoodResults
10.	AverageResults
11.	ExcellentResults
12.	PoorResults
13.	ExcellentResults
14.	GoodResults
15.	ExcellentResults
16.	GoodResults
17.	AverageResults
18.	ExcellentResults
19.	PoorResults



Result Analysis!!!

Constraints	Naive Bayes Algorithm
Accuracy	99%
Time (milli secs)	30174
Correctly Classified	99%
InCorrectly Classified	1%

© Copyright **Education Sector**. All Rights Reserved
Designed by Data Science

The above table show the result analysis of the student academic performance using naive bayes algorithm. The accuracy of the algorithm for the performance we performed is 99%.

VIII CONCLUSION AND FUTURE WORK

Analyzing students' learning behavior and academic performance in the education sector is a complex task.To tackle this challenge, the proposed system utilizes"Association Learning," a data science technique. Specifically, the system applies the "Eclat algorithm"to uncover patterns and associations among various factors.

The system serves as a real-time application, providing substantial benefits to colleges and lecturers. It offers valuable insights into students' behavior patterns, enabling the identification of influential factors that affect their academic performance. This knowledge empowers lecturers to gain a deeper understanding of their students and make informed decisions to support their learningjourney.

To further enhance the system, incorporating additional training datasets would unveil more relevant patterns. Additionally, exploring alternative algorithms and comparing their results would help determine the most effective algorithm for identifyingstudents' behavior patterns. These improvements would significantly boost the system's accuracy and effectiveness, delivering valuable insights to educators.

https://dx.doi.org/10.29322/IJSRP.13.10.2023.p14238

REFERENCES

1. Ramsay J. O., Silverman B. W. Functional dataanalysis[M]. New York: Springer, 1997.
2. Ramsay J. O., Silverman B. W. Applied functional data analysis: methods and case studies[M]. Vol. 77. New York: Springer, 2002.
3. Kesheng Liu, Siyang Wang. Variable selection in regression models including functional data predictors. Journal of Beijing University of aeronautics and astronautics, 2019, 45(10):1990- 1994.
4. Romero C., Ventura S., Data mining in education[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2013, 3(1):12-27.
5. Locantore N., Marron J., Simpson D., et al. Robust principal component analysis for functional data[J]. Test, 1999, 8(1):1–73.
6. Yao F., Lee T. Penalized spline models for functional principal component analysis[J]. Journal of the Royal Statistical Society: Series B (StatisticalMethodology), 2006,68(1):3–25..
7. iiMedia. w 2018 Chinese College Students' Online Leisure and Entertainment Behaviour
8. Monitoring Analysis Reportx[EB/OL].2018-11-16.https://www.iimedia.cn/c400/62969.html.

AUTHORS

First Author – Manasi S, Department of Computer Science and Engineering, JSSSTU, Mysuru and
email address: manasimanu1718@gmail.com**Second Author** – Tushar R Bharadwaj, Department of Computer Science and Engineering, JSS STU, Mysuru and email address: rtushar702@gmail.com
Third Author – Nuthan PM, Department ofComputer Science and Engineering, JSS STU, Mysuru and
email address: nuthannayakpm@gmail.com**Fourth Author** – Naganischay M, Department of Computer Science and Engineering, JSS STU, Mysuru and
email address: nischay1600@gmail.com