OPEN ACCESS
Saudi Journal of Engineering and Technology
Abbreviated Key Title: Saudi J Eng Technol
ISSN 2415-6272 (Print) |ISSN 2415-6264 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Review Article

Application of Machine Learning Techniques to Evaluate the Performance of Students in an Academic Environment

Chandana, D. A[1], Dr. S. Meenakshi Sundaram[2*], Bhumika, S[1], Meghana, B. N[1]

[1]Department of Computer Science & Engineering, GSSS Institute of Engineering & Technology for Women, Mysuru, Affiliated to VTU, Belagavi, Karnataka, India
[2]Professor and Head, Department of Computer Science & Engineering, GSSS Institute of Engineering & Technology for Women, Mysuru, Affiliated to VTU, Belagavi, Karnataka, India

DOI: 10.36348/sjet.2022.v07i06.008 | **Received:** 06.06.2022 | **Accepted:** 20.07.2022 | **Published:** 23.07.2022

***Corresponding author:** Dr. S. Meenakshi Sundaram
Professor and Head, Department of Computer Science & Engineering, GSSS Institute of Engineering & Technology for Women, Mysuru, Affiliated to VTU, Belagavi, Karnataka, India

Abstract

Identifying the most influential factors affecting the student's performance plays a vital role in improvising student's academic results. The conventional counseling is a time consuming process to understand students' performance. Machine Learning techniques play a major role in educational institutions to estimate the students, performance leading to better performance in placements. The major objective is to find behavior patterns of students in a timely and accurate manner. We find out the groups of students who need to be counseled in time. The system uses parameters such as attendance status, extra circular activities, grade, technical skills, previous semester results, grasping capability, aptitude grade, interaction with lecturers etc. This also helps faculty members to identify the most influential factors affecting the students' performance. Analyzing student mental issues for low academic performances is a complex task in the current education sector. The system uses data science technique called as "Association Learning" to find the patterns. The Eclat algorithm is used to find patterns. The proposed system builds as real time application useful to educational institutions to understand students' behavioral patterns. The system helps faculty to identify the most influential factors affecting the students' performance. A web browser in the future can be developed and used as an application. More number of parameters can be added to predict the students' performance.

Keywords: Machine Learning, students' performance, association learning, eclat algorithm, behavioral patterns.

1. INTRODUCTION

The data, information and statistics available through Internet are increasing exponentially at a very rapid pace it is no longer possible to manage so much of information in the traditional way. Also, correct handling of global data, which can be available at the touch of a button, can open avenues for business, research and education as never before. But the question is how to correctly utilize the data without getting drowned in sheer numbers. The answer lies in analytical sciences or Big data, which has become a buzzword. Machine learning has become an integral part of many commercial application and research projects, but this field is not exclusive to large companies with extensive research items. The prediction of students' performance with high accuracy is more beneficial for identifying low academic achievements students at the beginning. To improve their performance the teacher will monitor the students' performance carefully. Student retention is an indicator of academic performance and enrolment management of University.

The prediction of students' performance with high accuracy is more beneficial for identifying low academic achievements students at the beginning. To improve their performance the teacher will monitor the students' performance carefully. Students' retention is an indicator of academic performance and enrollment management of University. To assist the low academic achievers in higher education and they are:

a) Generation of data source of predictive variables
b) Identification of different factors, which affects a student's learning behavior and performance during academic career.

c) Construction of a prediction model using classification data mining techniques on the basis of identified predictive variables.
d) Validation of the developed model for higher education students studying in Indian Universities or Institutions.

The major objective is to find behavior patterns of students in a timely and accurate manner. The main aim is especially to find out the groups of students that need to be focused on in time. Without campus behavior analysis, achievements and other aspects of performance of these students are affected. System uses parameters such as attendance status, extra circular activities, grade, technical skills, previous semester results, grasping capability, aptitude grade, interaction with lecturers etc.. The system helps lecturers to identify the most influential factors affecting the students' performance. The system uses data science technique called as "Association Learning" to find the patterns. We use either "apriori algorithm" or "apriori TID algorithm" or "Eclat algorithm" to find patterns. The proposed system builds as real time application useful to Colleges and lecturers to know the students behavior patterns.

Identification of different factors which affects a student's learning behavior and performance during academic career. Analyzing student mental issues and low academic performances is a complex task in the current education sector. In current system it is difficult to track the students' behavior and characteristics. There is no automation or tool which predicts or shows how to improvise the students' academic performances. Finding the correlation between factors that affects students' performance and academic results pays a vital role in the current education sector.

2. LITERATURE SURVEY

Machine Learning deals with the development, analysis and study of algorithms that can automatically detect patterns from data and use it to predict future data or perform decision making [1]. Machine learning does its functionality by creating models out of it [2]. Machine Learning has become widespread and has its applications in the field of bio-informatics, computer vision, robot locomotion, computational finance, search engine etc. In real world problems, observations are made on entities associated with a problem so as to make inferences on the target value of those entities. This mapping is encompassed in a predictive model with the help of decision trees. This method of learning is referred to as Decision Tree Learning. This is one of the predictive modeling methods that can be found in the fields of data mining [3], machine learning and statistics. In this model we have made use of classification trees, a typical decision tree in which the predictor variable (target variable) takes on finite set of categorical values only. In this type of trees, the leaves represent the class labels and the branches represent the splitting path through which a decision travels from root to the leaf of the tree.

Hijazi and Naqvi [4] conducted a study to find the factors affecting the academic performance of students. They made use of questionnaires to elicit information from students highlighting factors such as income factor, parents' educational background, size of the family, regularity of teachers, subject interest created by the teachers and student's interest in co-curricular activities. They used Pearson Correlation Coefficient to highlight the important factors and they found that mother's education and family income played an important role in students' academic performance.

Saurabh Pal [5] conducted a study on student data that have information on their academic records and proposed a classification model to find an efficient method to predict student placements. They concluded that Naïve Bayes classifier is the best classification method for use in placements in comparison with Multilayer Perceptron and J48 algorithms. Ramanathan *et al*., [6] conducted a study using sum of difference method for students' placement prediction. They used the attributes such as age, academic records, achievements etc. for the prediction. They concluded that based on their results higher learning institutions can offer its students a superior education.

Fiseha Berhanu & Addisalem Abera [7] conducted a study to predict student placements using data mining. They made predictions on MCA students in Ghaziabad in UP, considering parameters such as MCA result, Communication skills, programming skills, co-curricular activity participation, gender, 12th result and graduation result. They concluded that their model based on decision tree algorithm can assist the placement cell and faculties in identifying set of students that are likely to face problem during final placements.

Elayidom *et al*., [8] designed a generalized data mining framework for placement chance prediction problems. They considered the students' Entrance Rank, Gender, Sector and Reservation Category to predict the branch of study that is Excellent, Good, Average or Poor for him/her using decision trees and neural networks. Naik and Purohit [9] made a study to use prediction technique using data mining for producing knowledge about students of MCA course before admitting them.

3. PROPOSED METHODOLOGY

In this work, three tier architecture is used as shown in Fig 3.1 given below.

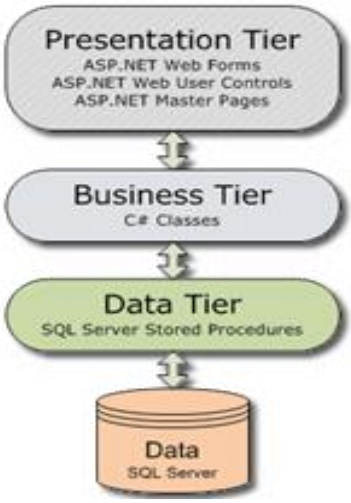


Fig 3.1: Three tier Architecture

This architecture includes the data link tier, business tier and presentation tier. The data layer is a separate component (often setup as a separate single or group of projects in a .NET solution), whose sole purpose is to serve up the data from the database and return it to the caller. Through this approach, data can be logically reused, meaning that a portion of an application reusing the same query can make a call to one data layer method, instead of embedding the query multiple times. This is generally more maintainable. Though a web site could talk to the data access layer directly, it usually goes through another layer called the business layer. The business layer is vital in that it validates the input conditions before calling a method from the data layer. This ensures the data input is correct before proceeding, and can often ensure that the outputs are correct as well. This validation of input is called business rules, meaning the rules that the business layer uses to make "judgments" about the data. One of the best reasons for reusing logic is that applications that start off small usually grow in functionality. The business layer helps move logic to a central layer for "maximum re-usability."

The architectural diagram for the students' performance evaluation is given in fig 3.2 below:

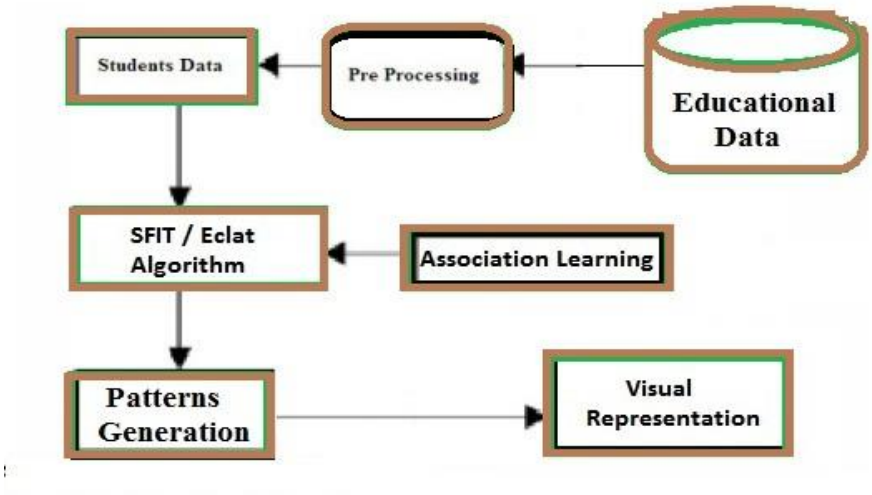


Fig 3.2: Architecture diagram students' performance evaluation

This work is implemented using object oriented programming language using three tier architecture. ASP.NET is used in the presentation layer, C# classes are used in the business logic, table adopter is used in the data tier and MS SQL server 2005 (database) is used as the back-end. Machine Learning (ML) is concerned with the construction and study of system that can learn from data. For example, ML can be used in e-mail message to learn how to distinguish between spam and inbox messages. Association Learning Algorithm "Apriori Algorithm / Eclat Algorithm" is used to predict the relationship between students' behavior and performance using the educational dataset. The Datasets are formed by using blender to create animations of different gestures of alphabets and action words. There are totally 161 datasets which includes the format of Indian sign language which are in the form of alphabets, numbers and action words. These datasets are created using an animation tool named blender which uses sequence of images to form the animated gifs.

Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts. The model is derived based on the analysis of a set of training data (i.e., data objects for which the class labels are known). The model is used to predict the class label of objects for which the class label is unknown.

The data flow diagram for this work for the admin page is given in Fig 3.3 below. This diagram gives admin details like creating login, validations of login and credentials used for logging in validating with

the data base. The various options for adding students successfully, setting IDs and creating passwords, addition of training data sets, predictions, queries and profile updating are shown. The storage and retrieval from databases and logging out options are shown in the Figure 3.3.

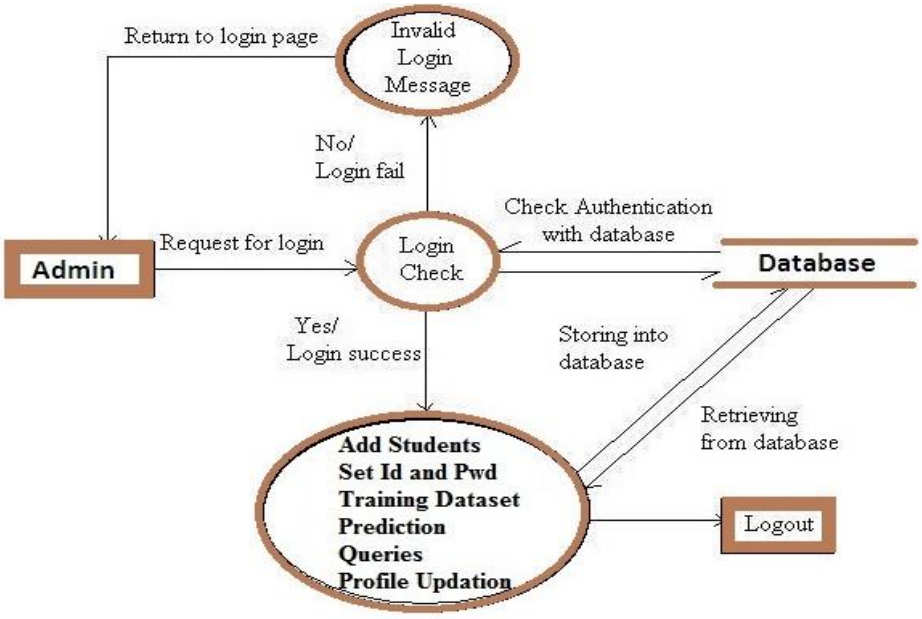


Fig 3.3: Data Flow diagram (Admin)

The data flow diagram for this work for the student page is given in Fig 3.4 below. For the student module options for successful login, validation of login with credentials, pattern prediction, queries and profile updating, storage and retrieval from data bases and logging out options are given.

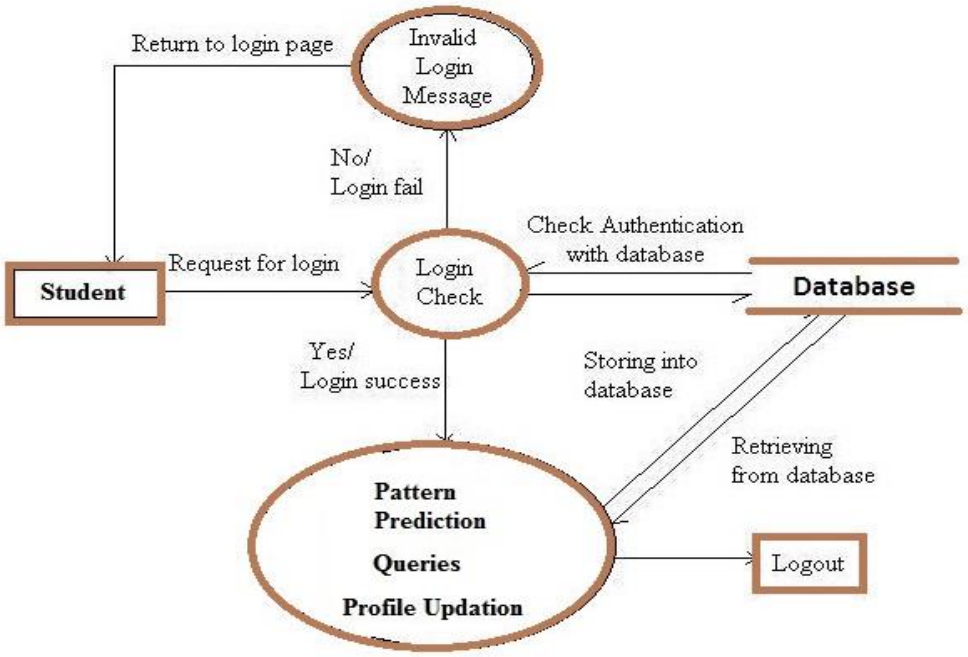


Figure 3.4: Data Flow diagram (Student)

Figure 3.5 given below gives the use case diagram for both the admin and the student. The admin has got the rights for administration to login, adding students, managing databases, importing excel data, prediction, result analysis, profile updating and sign out. The students have the options for logging in, pattern predictions, queries and profile updating.

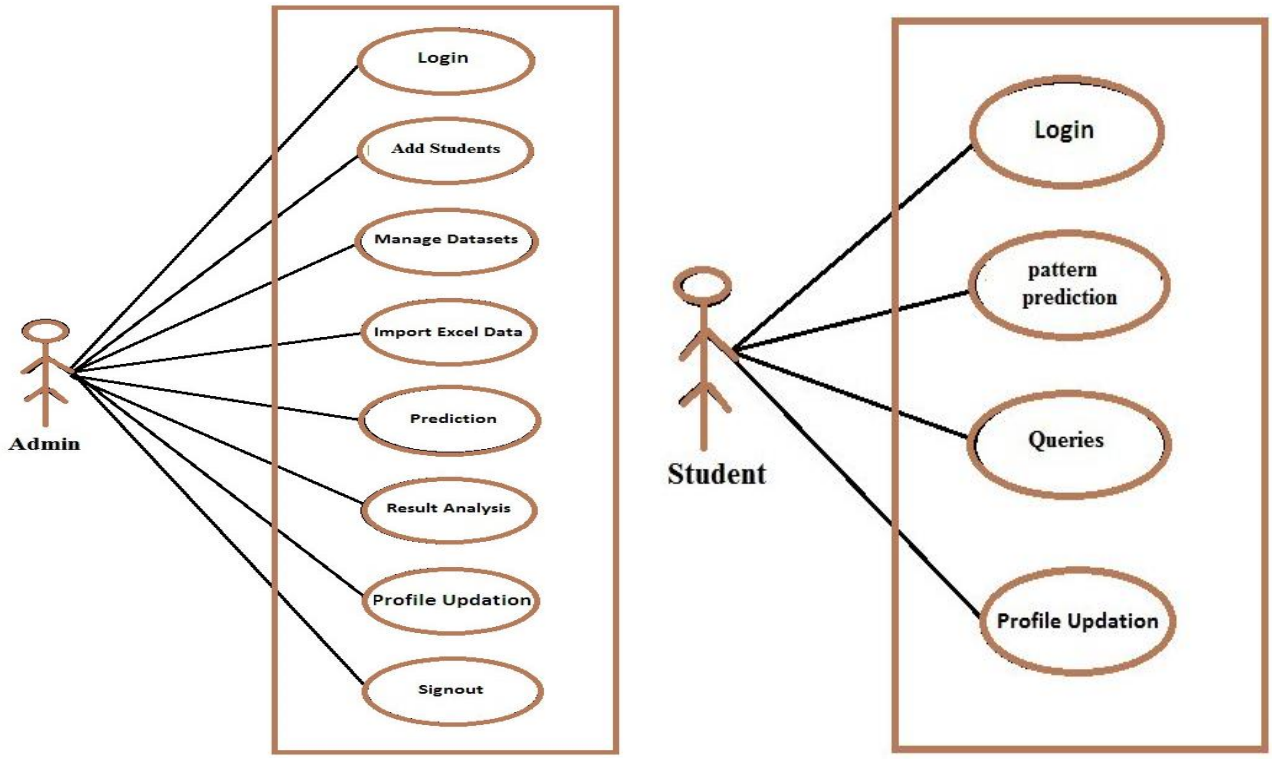


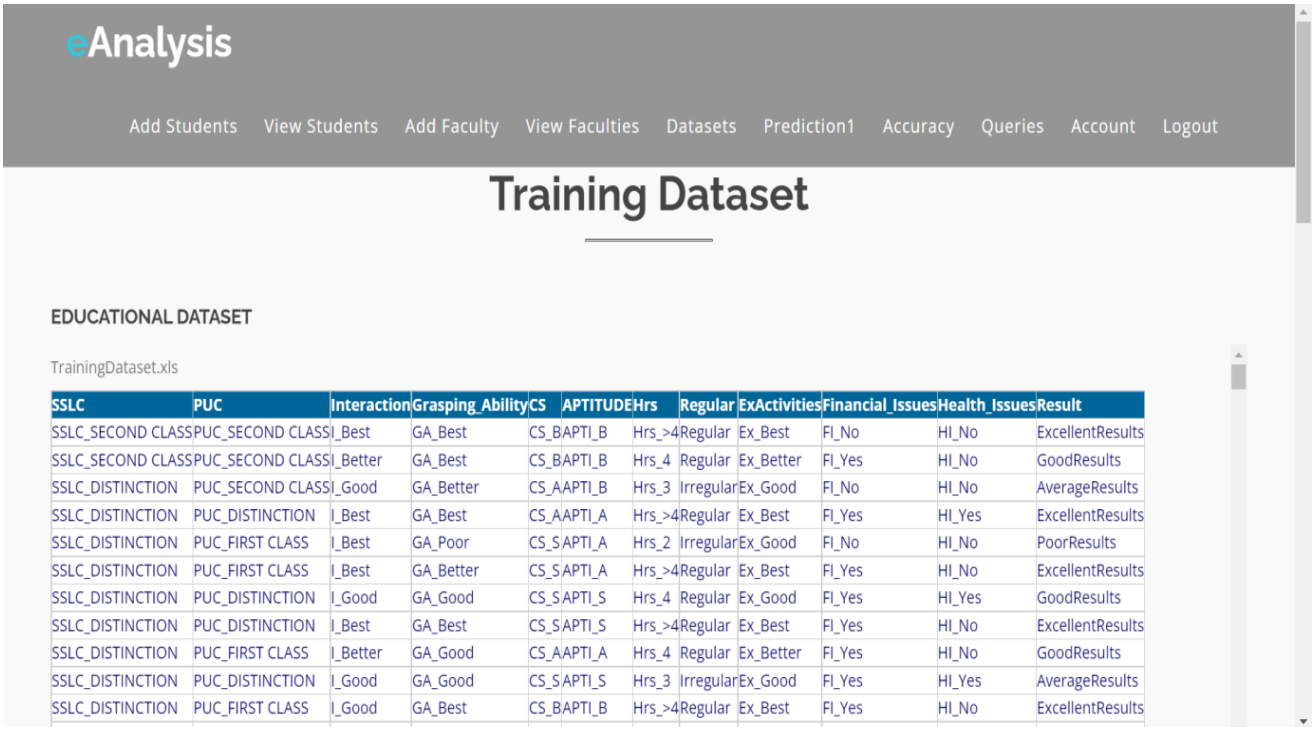
Fig 3.5: Use case diagram (Admin and Student)

4. Experimentation and Results

This web application is implemented using object oriented programming language. Object oriented programming is an approach that provides a way of modularizing programs by creating partitioned memory area for both data and functions that can be used as templates for creating copies of such modules on demand. Association Learning Algorithm "Apriori Algorithm / Eclat Algorithm" is used to predict the relationship between students' behavior and performance using the educational dataset-set.

The Datasets is formed by using blender to create animations of different gestures of alphabets and action words. There are totally 161 datasets which includes the format of Indian sign language which are in the form of alphabets, numbers and action words. These datasets are created using an animation tool named blender which uses sequence of images to form the animated gifs. Some of the test cases carried out for the work implemented includes: execution of the application, Verification of Admin Login with Input User Name input and Password validations, Verification of Input User Name and Password of Admin, verification of Input User Name & Password of Administration with invalid cases.

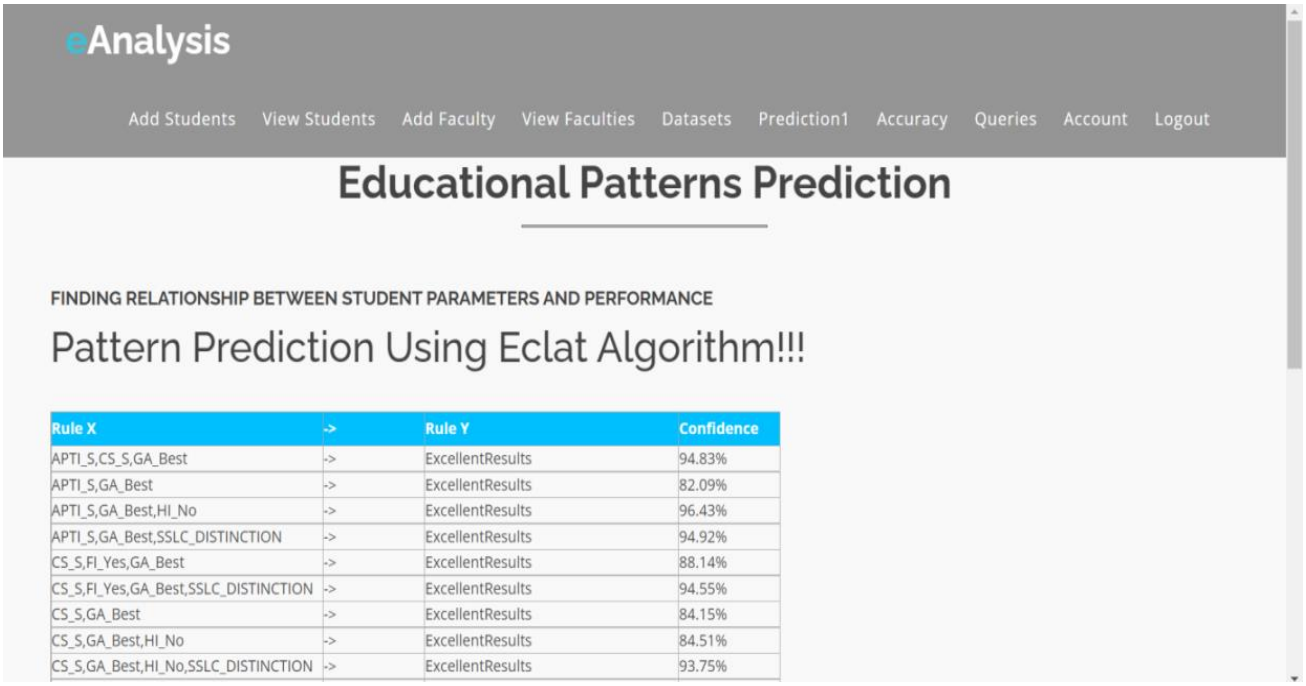
Test Case #	Description	Expected Result	Actual Result	Status of Execution Pass/Fail
TC01	Execute/run the application	Application should run without any interrupts.	Application is executing properly	Pass
TC02	Verification of Admin Login Input User Name and Password then click on Login button.	Admin User Name & Password should be check/verify with database.	Admin User Name & Password successfully checked with database.	Pass
TC03	Verification of Input User Name & Password of Admin.	If Admin User Name & Password is valid then it should navigate to respective Admin home page.	Admin User Name & Password is valid then successfully navigating respective home page.	Pass
TC04	Verification of Input User Name & Password of Admin. (Invalid Case)	If Admin User Name & Password is invalid then show message that Input Username & Password is wrong.	If User Name & Password is not valid or wrong input then message box shown that User Name & Password wrong.	Pass



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

Fig 4.1: Educational Training Dataset

The figure 4.1 given above shows the educational data sets created that gives details of educational qualification, grasping capabilities, interaction levels, aptitude level, extracurricular activities, financial and health issues The prediction of results are classified as excellent, good, average and poor.



Rule X	->	Rule Y	Confidence
APTI_S,CS_S,GA_Best	->	ExcellentResults	94.83%
APTI_S,GA_Best	->	ExcellentResults	82.09%
APTI_S,GA_Best,HI_No	->	ExcellentResults	96.43%
APTI_S,GA_Best,SSLC_DISTINCTION	->	ExcellentResults	94.92%
CS_S,FI_Yes,GA_Best	->	ExcellentResults	88.14%
CS_S,FI_Yes,GA_Best,SSLC_DISTINCTION	->	ExcellentResults	94.55%
CS_S,GA_Best	->	ExcellentResults	84.15%
CS_S,GA_Best,HI_No	->	ExcellentResults	84.51%
CS_S,GA_Best,HI_No,SSLC_DISTINCTION	->	ExcellentResults	93.75%

Figure 4.2: Pattern Prediction using Eclat Algorithm

Figure 4.2 given above shows the pattern predictions using the proposed Eclat Algorithm. For Rule X and Rule Y, the predictions for confidence levels are given in terms of percentage are shown in Figure 4.2.

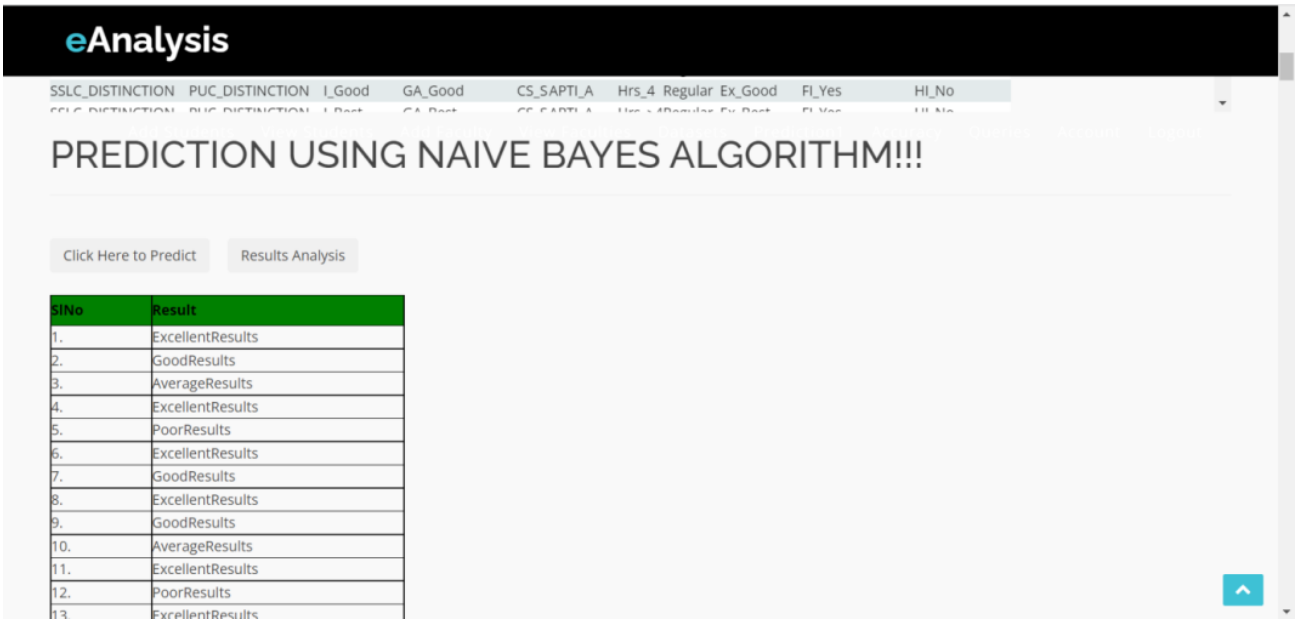


Figure 4.3: Prediction of Overall Class Results using Naive Bayes Algorithm

Figure 4.3 given above shows the prediction of overall Class Results using Naive Bayes Algorithm.

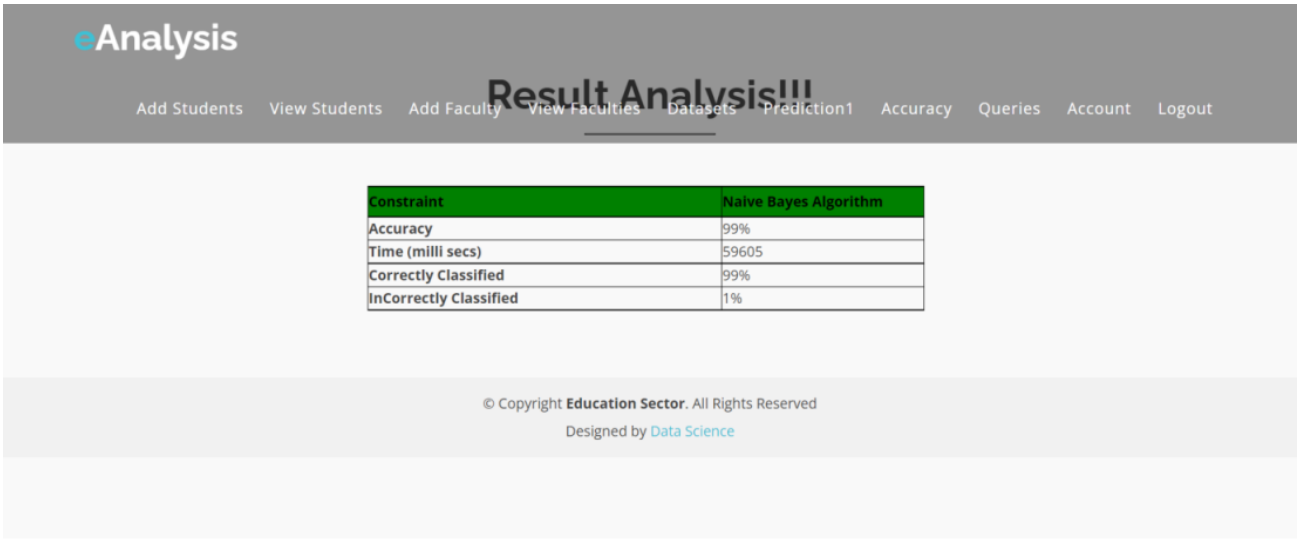


Figure 4.4: Accuracy of Results

Figure 4.4 given above gives the accuracy of results. For the data sets used we see that the accuracy is 99%, the time for execution is 59605 mill seconds Correctly Classified is 99% and incorrectly classified is 1%. The students' performance prediction is given taking input data like qualification details, interaction levels, financial issues and health issues of the students. We are able to classify as excellent, good, average and poor results.

CONCLUSION

Identification of different factors which affects a student's learning behavior and performance during academic career. Analyzing student mental issues and low academic performances is a complex task in the current education sector. System uses data science technique called as "Association Learning" to find the patterns. We used "Eclat algorithm" to find patterns. The proposed system builds as real time application useful to Colleges and lecturers to know the students behavior patterns. The system helps faculty members to identify the most influential factors affecting the students' performance. More training datasets can be used to find more related patterns. More algorithms can be used to find the students behavior patterns and algorithms can be compared to identify the algorithm with better results. The web browser in the future can be developed as an application. More number of

parameters can be used to predict the performance of the students.

REFERENCES

1. Ramsay, J. O., & Silverman, B. W. (1997). Functional data analysis [M]. New York: Springer.
2. Ramsay, J. O., & Silverman, B. W. (2002). Applied functional data analysis: methods and case studies [M]. Vol. 77. New York: Springer.
3. Liu, K., & Wang, S. (2019). Variable selection in regression models including functional data predictors. *Journal of Beijing University of Aeronautics and Astronautics*, *45*(10), 1990-1994.
4. Syed Tahir Hijazi1 & Raza Naqvi, S. M. M. (2006). Factors affecting students' performance: A case of private colleges. *Bangladesh e-Journal of Sociology*, *3*(1), 1-10.
5. Pal, A. K., & Pal, S. (2013). Classification model of prediction for placement of students. *International Journal of Modern Education and Computer Science*, *5*(11), 49. DOI: 10.5815/ijmecs.2013.11.07
6. Ramanathan, L., Geetha, A., Khalid, M., & Swarnalatha, P. (2016). Apply of sum of difference method to predict placement of students' using educational data mining. In *Information Systems Design and Intelligent Applications* (pp. 367-377). Springer, New Delhi. 10.1007/978-81-322-2755-7_39
7. Berhanu, F., & Abera, A. (2015). Students' performance prediction based on their academic record. *International Journal of Computer Applications*, *131*(5), 0975-8887.
8. Elayidom, S., Idikkula, S. M., Alexander, J., & Ojha, A. (2009, December). Applying data mining techniques for placement chance prediction. In *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies* (pp. 669-671). IEEE.
9. Naik, N., & Purohit, S. (2012). Prediction of final result and placement of students using classification algorithm. *International Journal of Computer Applications*, *56*(12), 35-40.