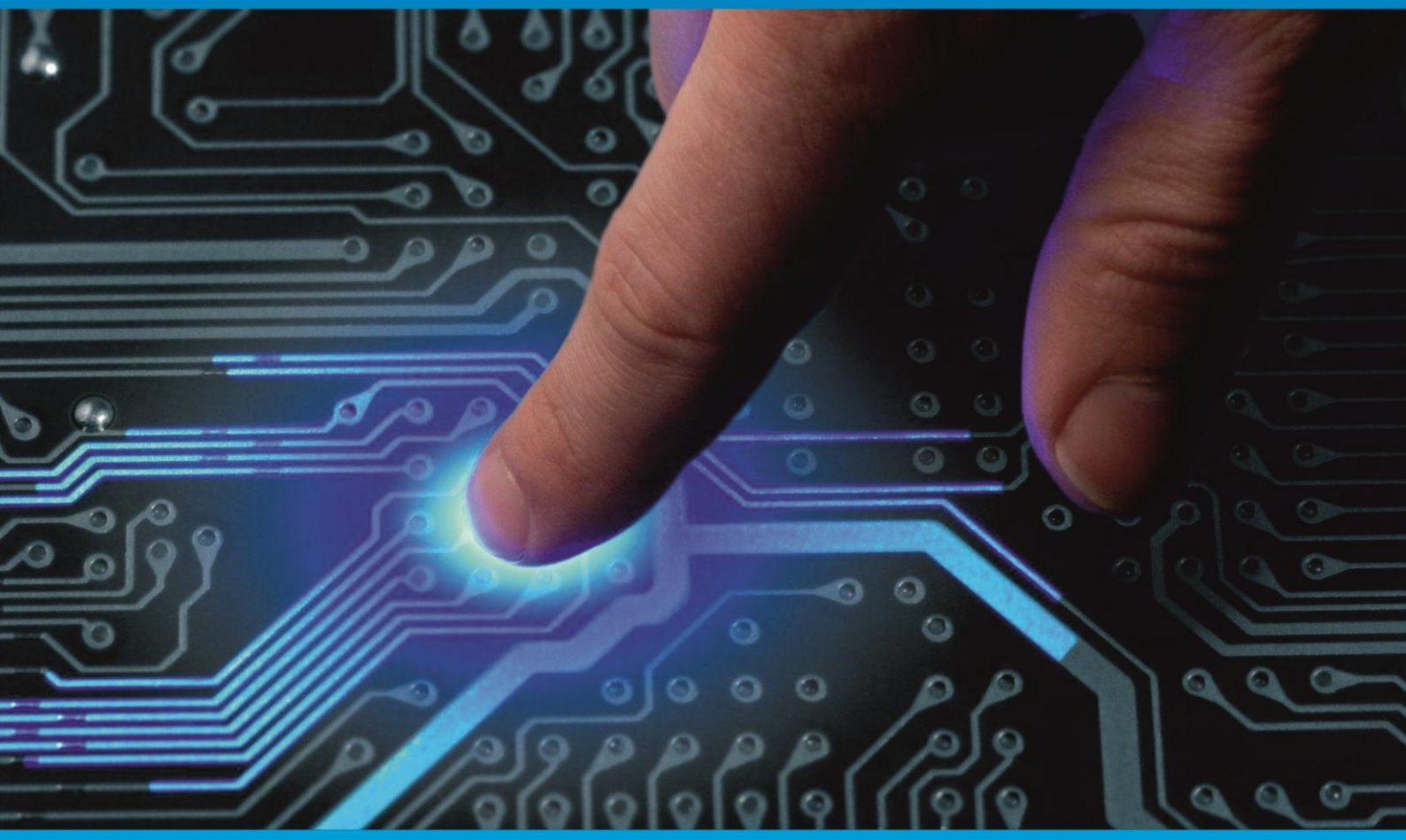
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 10, October 2020

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

Students Performance Analysis System (SPAS) Using Market Basket Analysis and Data Mining

Anila Paul A P, Nisha Raveendran

PG Student, Department of Computer Science, Christ College(Autonomous), Irinjalakuda, Kerala, India

Assistant Professor, Department of Computer Science, Christ College(Autonomous), Irinjalakuda, Kerala, India

ABSTRACT: Educational systems need innovative ways to improve quality of education to achieve the best results and decrease the failure rate. Educational Data Mining (EDM) has boomed in the educational systems recently as it enables to analyze and predict student performance so that measures can be taken in advance. Students in higher educational institutions interact with many systems, such as registration and learning management. Mining student data collected from these systems helps discover useful insights with respect to student behavior. Researchers in the field of educational Data Mining and learning analytics have already demonstrated that such data can be used to predict student academic performance. Many educational institutions enforce attendance policies, where students are expected to have their absences below a certain percentage in each class. We apply a Data Mining technique, the Market Basket Analysis, on student attendance data. The contribution of this analysis is the identification of student groups who share highly similar absence records. Such similarity may indicate that the students are not attending their classes due to invalid reasons. The results of this project helped to identify students who would have a tendency to skip their classes due to absence of their friends. In this paper, the system predicts the students performance on the basis of marks and attendance

KEYWORDS: Market Basket Analysis, Data Mining, Educational Data Mining

I. INTRODUCTION

Student administrator and counselors strive to ensure the success of their students. Apart from advising students with respect to courses they need to register for next the semester, it is expected that administrator and counselors encourage students to put more effort in their courses, motivate students to engage in extracurricular activities and help students reflect on their behavior. Teachers may help administrator by providing feedback with regards to the performance of the advisees. The institution uses a system to keep track of student attendance. The institution uses the system to enforce attendance, where students are withdrawn from a course if they exceed a certain number of absences. The system notifies administrator once their students are in danger of being withdrawn from a course. The system fails short from reporting the following behavior. Alice and Bob are two students and they are classmate in a few courses. Bob and Alice are also friends. Should Bob, skip a class, Alice may feel pressured to skip that class. The same may be true the other way around. This behavior may be repeated a few times in each course, will amount to many absences. However, their behavior will fly under the attendance system radar, since they are skipping classes from difference courses. Neither the teachers nor the administrator will suspect that one of the students is skipping the classes because of peer pressure. Such behavior may result in students failing their courses. This is a problem that remains unsolved and it constitutes the problem addressed in this paper. Data Mining has boomed in the educational systems recently as it enables to analyze and predict student performance so that measures can be taken in advance. The system mainly created for students, teachers and administrator. In this paper, paper predicts the student's performance on the basis of attendance and marks. So that measures can be taken in advance. The contribution of Market basket analysis is the identification of student groups who share highly similar absence records. Such similarity may indicate that the students are not attending their classes due to invalid reasons.

II. PROPOSED SYSTEM

Almost every university has their own management system to manage the student's record. Currently even through there is a student management system that manages the students record in University Malaysia Sarawak(UNIMAS), no permission is provided for lectures to access the system. This is because the access permission is only to top

management such as Deans and Deputy Deans of Undergraduate and student development due to its privacy setting. In the existing Student's performance analysis system keeps a track of students result and attendance. The system predicts student's marks and analyzes the student's performance. The system, also analyzes all class performance on the basis of mark, by using technique is that Educational data mining. In this paper, analysis is the identification of student groups who share highly similar absence records. This indicates that the students are not attending their classes due to invalid reasons. The presented method helps teachers and admin discover this behavior, which is more efficient than relying on instructors, who may teach many classes. The results of this project helped to identify students who would have a tendency to skip their classes due to absence of their friends. In my paper also predict the student's performance on the basis of marks and attendance, so that measures can be taken in advance.

Working Diagram

There are mainly three types of users-Admin, teacher and students. Dataset is the collection of the data and Data Mining is the process of analyzing the data. Admin and teacher two main users, teacher predicts students mark and attendance. That predicted result can view to the students. Admin do the Market Basket Analysis. That results help to finding the group of students who share highly similar absence record.

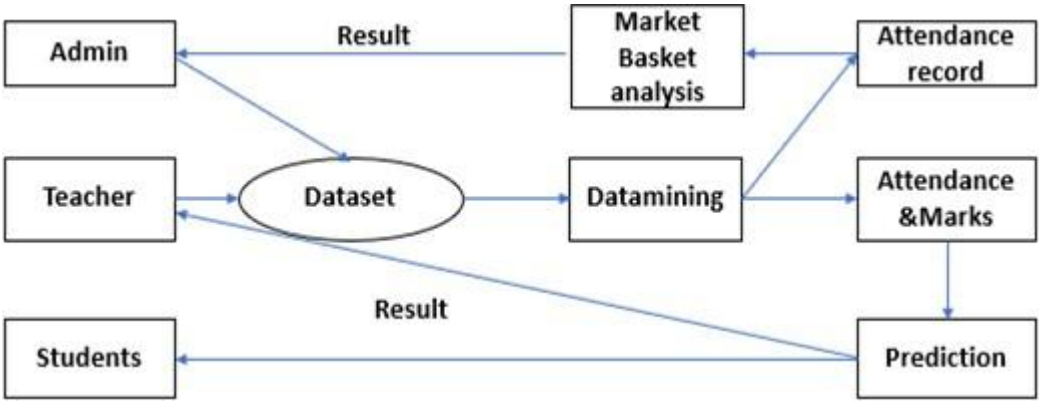


Figure 1: Working Diagram

III. METHOD

We propose a method that utilizes the attendance records for all students. The collected data is then passed to a data mining technique for analysis. Specifically, the method mines the records for association rules, where each rule links the absences of one student to the absences of one or more students. The method makes use of the market basket analysis [15], which is a data mining technique that retailers, such as Amazon and eBay use to find associations between their products. Retailers generate association rules by inspecting their transactions and finding items that frequently appear together. For example, the following rule may be used by a retailer recommender system.

In our work, we use the apriori algorithm to find rules associating the absences of one student to the absences of one or more students. The generated rules help instructors and advisors identify students who are frequently absent together. Instructors and advisors may meet with such students to investigate the reason behind this behavior. Such intervention supports student academic success. Please refer to Fig. 1, which illustrates the proposed method. The next subsection defines the necessary terminology for the apriori algorithm. Please refer to for the exact algorithm.

A. Data Mining

Data mining is the process of discovering patterns in large dataset involving methods at the intersection of machine learning, statistics, and database system. In here, we collecting student's details and collecting wanted dataset from the larger dataset. Market Basket Analysis is a Data Mining Technique used by e-commerce sites. It derives association between two data sets. We have categorical data of transaction record as input to the analysis and the output of the analysis is association rules as new knowledge directly from data.

A common approach among the educational data mining literature described earlier is the use of student attendance, performance in assessment and interaction with learning resources to predict student performance and

success. Student attendance is modeled as a qualitative value, such as low, average and high attendance. This paper utilizes the actual attendance records. Using the actual records allows us to utilize more data mining techniques, such as the market basket analysis used in this paper. Further, existing literature does not help in finding groups of students who share a very similar attendance records across different classes, which may signify missing classes due to peer pressure. The next section describes the presented method for finding patterns within student attendance records.

B. Apriori Algorithm

The apriori algorithm works on transactional datasets where each row represents a transaction consists of a set of items. Consider the set of available items for purchase, gaming consoles, controllers, virtual reality sets, virtual reality games, action games and online passes. There are two important values that need to be calculated for each association rule. They are the rule's support and confidence values. First, let us define the support for a set of items. The support of a set is the number of transactions containing that all of that set items. The support may also be expressed as the ratio of the number transactions containing all the items in that set, to the number of all transactions in the dataset; that is, we may express the support as a percentage.

Support (*S*)= *number of transactions containing S/ number of transactions*

Similarly, the support of a rule is the number of transactions containing the items of that rule left, as well as right-side set items. The confidence of an association rule is calculated as the ratio of the support of the union between the left-side and right-side sets to the support of the left side set. Rules with higher confidence and support signify that if the customer is showing interest in the items of the left-side set, then there is a high probability that the customer will show interest in the items of the right-side set.

C. Mining Attendance Records with Apriori

Products bought together, by one customer, comprise one transaction. Rather than dealing with transactions and products, this paper applies the apriori algorithm on students' absence records. The rules we generate have a structure similar to the example rule below. The rule states that if StudentA and StudentB are absent, then StudentC will likely be absent too. The assumption here is that the three students are missing the same class on the same date. Students who miss the same class on the same date comprise one transaction.

$$(StudentA, studentB) \rightarrow (StudentC)$$

The following is a brief description of the steps to perform the apriori on student absence records and the meaningful use of the algorithm output.

Step 1: The first step is to collect the absence records for each student in each class. A list of student_ids needs to be extracted, Stu_IDs. Similarly, a list of classes or sections needs to be extracted, Sec_IDs, such as CS101 3.

Step 2: For each section, extract the dates where that section had a session and the instructor of that section recorded the attendance. Append the Sec ID to each date, creating a list of dates for each section, Sec ID Date IDs. For example, if section CS101 3 had a session on Apr 12th, then CS101 Apr12 is created. This is repeated for each section to create a list of lists, Master SecDateList.

Step 3: Create a matrix where the rows represent the elements of Master Sec Date List ids and the columns represent the student ids. The cell i,j in created matrix represents whether student, j, was present(1)or absent (0) during session i (note that i is composed of a section id and a date id).

Step 4: Fill the matrix created in Step 3 with student absence values based on the records pulled from the attendance system.

Step 5: Run the apriori algorithm on the computed matrix from Step 4. The support and confidence thresholds need to be specified. The support threshold is based on the number of sessions. If a student takes 5 courses in a semester, where each section holds 30 sessions throughout the semester, then the maximum number of sessions is 150. Consider two students having an exact class schedule, the maximum number of times both are absent is 150. However, it is highly unlikely that a student misses all of sessions, in all of the classes. Therefore, one may set the support threshold a reasonable value. For example, a threshold of 10 help generate association rules, where each rule is supported by 10 incidents were the two students were absent together. The confidence threshold is based on the required rule strength. Setting the threshold to 0.7 help generate rules, where each rule associates the absence of one student to another with 85 support and confidence thresholds also depends on the number of rules the user is willing to go through. The lower the thresholds, the more rules the apriori generates.

Step 6: The output of the apriori algorithm is a set of rules that meets the thresholds set in Step 5. If a student appears in a rule, then the advisor of that student may receive a notification with regards to that rule. An example output is displayed in Table III. Table III lists a set of the actual anonymized rules generated based on the attendance records collected from the author institution. The experiment detail is explained the next section.

Step 7: The advisor needs to investigate the rules. Note that some rules may be the result of students coincidentally missing classes, without peer pressure. The advisor may check on the student overall performance to set the priority for meeting students. Students who miss classes and are under probation needs urgent attention, compared to students with good standing status.

Step 8: Finally, advisors provide feedback for the person responsible for generating the rules with regards to the usefulness of the generated rules. The feedback helps tweak the support and confidence thresholds for the next rounds.

D.Time Series Algorithm

For prediction we use **Time Series Algorithm**. Time Series Algorithm is used to predict continues values of data. Once the algorithm is skilled to predict a series of data, it can predict the outcomes of other series. The algorithm generates a model that can predict trends based only the original dataset. New data can also be added that automatically becomes a part of the trend analysis. In here we using Time Series ARIMA method model. Autoregressive Integrated Moving Average is called ARIMA. ARIMA method model the next step in the sequence as a linear function of the differenced observation and residual errors at prior time steps. Here predicting mark and attendance using time series algorithm. **Matplot** method is used to draw the graph.

IV. EXPERIMENT

To demonstrate the presented method, we analyzed the attendance records of more than 150 students, over the period of one semester at the author's institution. The apriori algorithm was applied to perform the basket analysis. The authors implemented their approach using Python programming language, a popular data mining software environment. The following is a summary of the data collection, analysis ,prediction and results of the experiment.

A. Data Collection and Preprocessing

The authors collected more than 5000 absence records, where each record documents the absence of a student with respect to a class. Each record in the collected attendance data consists of the following items. The records were then preprocessed to remove the first and last week of classes. The rationale is to avoid the collection of absence records for the days in which many students are absent. At the first week, students are busy with add and drop, while at the last week, many student miss classes to prepare for assessment. The authors also preprocessed the records for students who change sections. An attendance matrix was generated for each section, where rows represent the dates that section had a session and columns represent the students registered in that section, i.e., the class roaster.

B. Analysis, prediction and Results

The steps 1 to 8 from the Section III.c and d were followed to mine the collected data for association rules. The selected thresholds were support = 0.0001, around 5 transactions, and confidence = 0.25. We used Python visual Studiocode, an open-source IDE for python, to perform the analysis and prediction. To visualize the result in a table and graph.

As described earlier, the apriori algorithm generates rules in the following structure:

$$(left \rightarrow right : support, confidence)$$

Where left refers to the left-hand side of the rule and right is the right-hand side of the rule. Both left and right denote a set of one or more students, where the arrow between them denotes that if left is absent, then right is absent too. The support is the percentage of records where both students were found to be absent together. The confidence is the percentage of students in the left being absent, with students in the right being absent too. Table III lists a number of rules, where actual student ids were replaced with random letters to ensure anonymity. A support of 0.01078absences in our dataset. A confidence of 1 means that student in the right side of a rule is always absent if student in the left side of that rule is absent.

The results of the analysis helped us identify over 10 pairs of students who share similar absence records. A likely cause is one student in each pair is may be skipping classes because his or her friend is absent. To visualize the

result in a table and graph. Admin can also check the all class performance in the help of plotted attendance and mark graphs. In the result of prediction, teachers are individually predicted mark and attendance of the student and doing the analysis with the help of plotted graph individually.

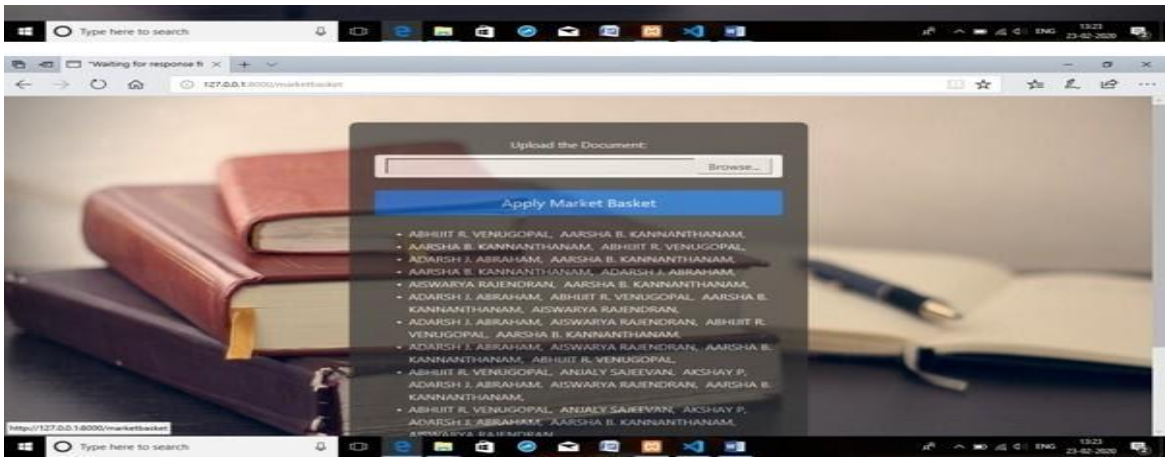


Figure 2: Market Basket Analysis Result

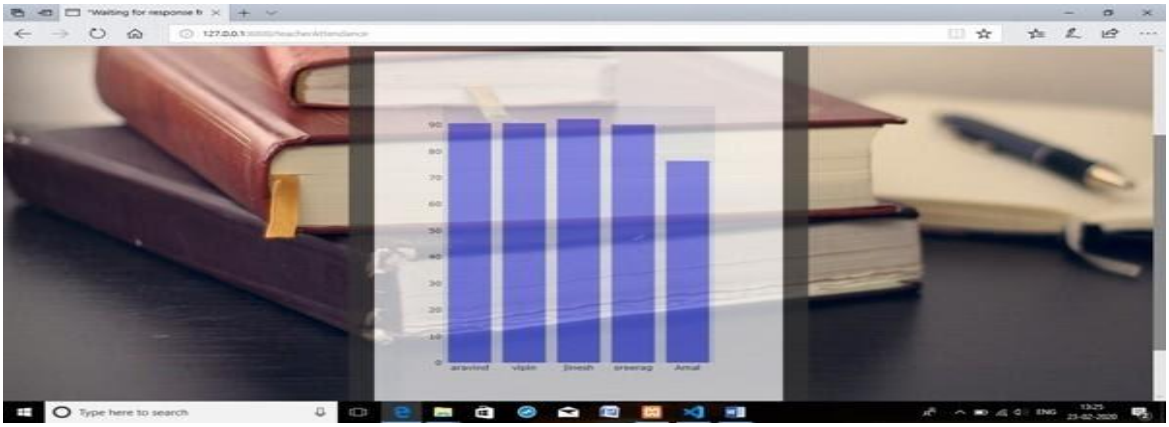


Figure 3: Students Attendance Analysis of all students Result

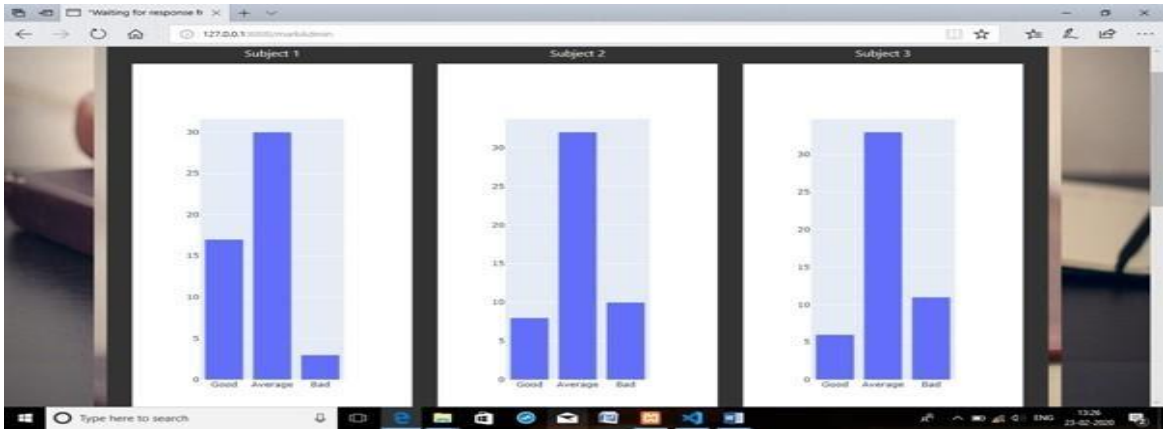


Figure 4: Mark Analysis of all students Result

Figure 5: Predicted attendance Result

Figure 6: Predict Mark Result

V. CONCLUSION

I have successfully designed a prototype of Student's Performance Analysis System using Python Visual studio code software. The algorithms we used here are time series for prediction and apriori for market basket analysis Time Series Algorithm is used to predict continues values of data. Once the algorithm is skilled to predict a series of data , it can predict the outcomes of other series. Apriori Algorithm for finding frequency item set. Its analyses a data set to determine which combinations of items occur together frequently. The system so developed was successfully tested, its limitations identified it is very slow.

In future works we can do the same performance analysis system in organizations. So applying this method in employee data we can easily find the performance of an employee, overall performance of employees etc. And we can also find the employees who have similar absence record. The future scope of this project is using new algorithms for solving the problem.

REFERENCES

1. Software Engineering :A Practitioner's Approach by Roger S Pressman
2. Contemporary System Analysis by John Stubbe & Marvin Gore
3. Software Engineering Concepts by Richard E. Farley
4. www.w3schools.com
5. www.djangoproject.com
6. www.mysql.org
7. K¨aser, T., N.R. Hallinen, and D.L. Schwartz. Modeling exploration strategies to predict student performance within a learning environment and beyond. in Proceedings of the Seventh Interna- tional Learning Analytics & Knowledge Conference. 2017. ACM.

8. Muhammad Gufran Khan,Sajid Anwer,Aysha Ashra.A Comparative Study of Predicting Student's Performance by use of Data Mining Techniques Proceedings of the Seventh International Learning Analytics & Knowledge Conference. june 2018

9. H. Alagib, A. Hamza, & P. Kommers, A Review of Educational Data Mining Tools& Techniques, International Journal of Educational Technology and Learning, vol. 3, no. 1, pp. 17-23, 2018.

10. M. Hussain, M. Al-Mourad, S. Mathew& A. Hussein, Mining educational data for academic ac- creditation: aligning assessment with outcomes, Global Journal of Flexible Systems Management, vol. 18, no. 1, pp. 51-60, 201

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details