Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

STUDENT PERFORMANCE ANALYSIS USING MACHINE LEARNING

MS.J.LIKHITA SRI [1], MS.S.KEERTHI [2], MS. SK.BEE JAN BEE [3], MS.SD.MOTHI BEGUM [4], Mrs. D. VIJAYA KUMARI[5]

1. BTECH, VIJAYA INSTITUTE OF TECHNOLOGY FOR WOMEN, ENIKEPADU, VIJAYAWADA, ANDHRA PRADESH,– 521108
2. BTECH, VIJAYA INSTITUTE OF TECHNOLOGY FOR WOMEN, ENIKEPADU, VIJAYAWADA, ANDHRA PRADESH,– 521108BTECH, MIC COLLEGE OF ENGINEERING, KANCHIKACHERLA, NTR DISTRICT, ANDHRA PRADESH, INDIA – 521180
3. BTECH, VIJAYA INSTITUTE OF TECHNOLOGY FOR WOMEN, ENIKEPADU, VIJAYAWADA, ANDHRA PRADESH,– 521108
4. BTECH, VIJAYA INSTITUTE OF TECHNOLOGY FOR WOMEN, ENIKEPADU, VIJAYAWADA, ANDHRA PRADESH,– 521108
5. Assistant Professor DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, VIJAYA INSTITUTE OF TECHNOLOGY FOR WOMEN, ENIKEPADU, VIJAYAWADA, ANDHRA PRADESH,– 521108

ABSTRACT

Student performance analysis is one of the most important aspects of any university for assisting and assisting DE scions in implementing new concepts in the university. In this paper, we propose to predict the student performance analysis using previous data and implement machine learning algorithms like svm and random forest to analyse and predict the future performance so that we can assist the student in terms of their efficiency. The suggested technique includes two key components: a bilayered For making predictions based on students' changing performance states, a structure with multiple base predictors and a cascade of ensemble predictors is developed. A data-driven approach based on latent factor models and probabilistic matrix factorization is suggested to find course relevance, which is crucial for creating effective base predictors. We demonstrate that the proposed strategy outperforms benchmark approaches through in-depth simulations on a dataset of undergraduate student data

1 INTRODUCTION

Data mining, which aims to extract useful information from data stored in enormous datasets, is currently one of the most significant fields of research and plays a crucial role in educational institutions. An extremely significant area of research is educational data mining (EDM), which helps to forecast meaningful information from educational databases to enhance educational performance, better understanding, and to have a better assessment of the students' learning process. Data mining, also known as knowledge discovery, has grown in importance as a result of its ability to analyse data from a variety of angles and synthesise it into practical information.

Data mining may be a potential and thriving area for data analysis, and the findings from this study have a wide range of uses. Data mining is another name for KDD, or Knowledge Discovery from Data.With the help of this system, knowledge that is implicitly stored or recorded in sizable databases, data warehouses, the Web, data repositories, and information streams may be easily and quickly extracted. Information technology, machine learning, statistics, pattern recognition, data retrieval, neural networks, information based systems, artificial intelligence, and data visualisation are all included in the multidisciplinary discipline of data mining. The project's major goal is to examine and evaluate school kids' performance using data mining approaches. Numerous activities that can be utilised to examine student performance are offered by data mining. This research examines the accuracy, confusion matrices, and execution times of the various classification data used in the classification assignment used to assess students' performance. mining software.

2. LITERATURE SURVEY AND RELATED WORK

Some of the components of the current system include a user login creator for the user interface, a student

performance analyzer, a student development card, accomplished credit, a passing criterion card, and a smart student performance attribute card.One of the most crucial elements in helping DE scions adopt new ideas in universities is student performance analysis. In this study, we suggest using historical data to predict student performance and applying machine learning techniques like svm and random forest to analyse and forecast future performance. We can help students be more effective by doing this. We ultimately predict the outcome of the execution of the result utilising the coefficient correlation matrix with machine learning methods. the way they performed. In order to determine course relevance, which is essential for creating efficient base predictors, the proposed method proposes a data-driven approach based on latent factor models and probabilistic matrix factorization. Additionally, a bilayered structure made up of multiple base predictors and a cascade of ensemble predictors is developed for making predictions based on students' changing performance states. On a dataset of undergraduate student data collected over three years at UCLA, we use in-depth simulations to show that the suggested strategy outperforms benchmark approaches.

3 PROPOSED WORK

Our initiative focuses on the topic of educational data mining (EDM). It has a variety of uses. A strong technique for improving academic achievement is data mining. The goal of educational data mining is to create innovative techniques for extracting knowledge from educational databases that can be applied to decision-making in the educational system. When data mining methods like clustering, decision trees, and association are used in higher education processes, student performance will be enhanced.

The fields of EDM include

Data analysis and visualization using the Random Forest algorithm with SVM Giving suggestions to help instructors

1) Student recommendations
2) Student performance predictions
3) Student modelling
4) Spotting problematic student conduct
5)Student grouping
6)social network analysis

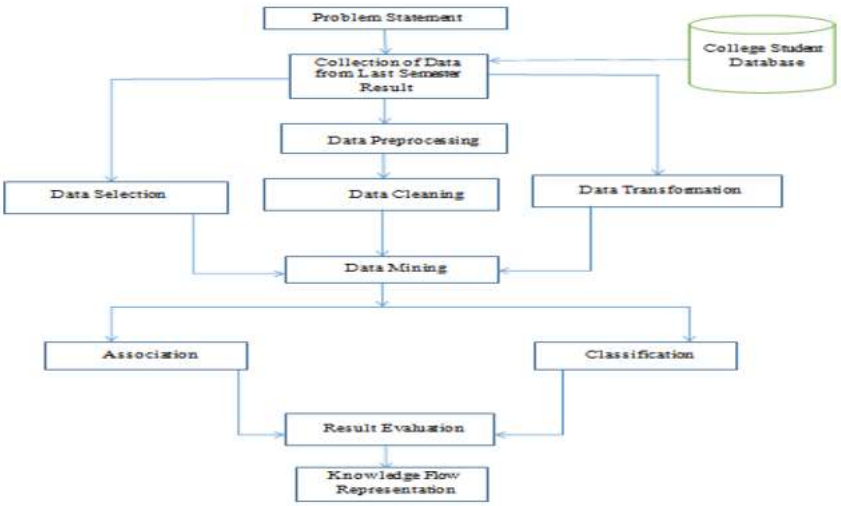

Fig 1:proposed model

4 Implementation

During the implementation phase, a dataset of 637 students' records in course "TMC1013 System Analysis and

Design" is collected and analysed by using data mining technique to generate IF-THEN rules for prediction of students' result in course "TMC1013 System Analysis and Design". The generation of IF-THEN rules is performed by using an open software tool, named WEKA. The dataset is divided into training set and test set. 80% of the dataset is used for the training set and the remaining 20% is for the test set. The training set is used to train the classification model while the test set is used to test the classification model build for its prediction's accuracy. A comparison of accuracy between different decision trees classifications' techniques are tested to ensure the highest prediction of accuracy could be achieved. Table 2 shows the accuracy comparison between five different decision trees'

4..1 System analysis and design

In this phase, the overall flow of the system is planned, analysed and designed. The system and user requirements are analysed and listed in table format. Data flow diagram is used to chart the input, processes and output of the system. Data flow diagram from the context diagram up to the first level is analysed and drawn. Besides, logical design of the proposed system is drawn to ensure the developed system is functioning as expected. The logical design is designed by drawing entity-relationship diagram (ERD). The ERD illustrates the data objects, attributes and relation between tables in the database as it is a graphical representative of theentity-relationship data model. Furthermore, the design of the proposed system includes the design of database and user interface. The hardware requirement in this phase is a computer for analysis and design. Other than that, Microsoft Office Visio 2007 is needed to draw the ERD and data flow diagrams.

Is a software that enables educators to view the students' performance and keep track of the school's data. The SPA is a tool designed for analysing, displaying, storing, and getting feedback of student assessment data [3]. It is a powerful analyser tool used by schools worldwide to perform analysis and displays the analysis data once raw student data is uploaded to the system. The analysis is done by tracking the student or class to get the overall performance of student or class. It helps to identify the students' performance which is below the expected level, at expected level or above the expected level. This would allow the educators or staffs to identify the current students' performance easily. Other than that, it enables various kinds of students' performance report such as progress report and achievement report to be generated

5.RESULTS AND DISCUSSION SCREENSHOTS

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	free
GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	
GP	F	17	U	GT3	T	1	1	at_home	other	...	5	
GP	F	15	U	LE3	T	1	1	at_home	other	...	4	
GP	F	15	U	GT3	T	4	2	health	services	...	3	
GP	F	16	U	GT3	T	3	3	other	other	...	4	
GP	M	16	U	LE3	T	4	3	services	other	...	5	
GP	M	16	U	LE3	T	2	2	other	other	...	4	
GP	F	17	U	GT3	A	4	4	other	teacher	...	4	
GP	M	15	U	LE3	A	3	2	services	other	...	4	
GP	M	15	U	GT3	T	3	4	other	other	...	5	

Fig 2:- sample data in dataset

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

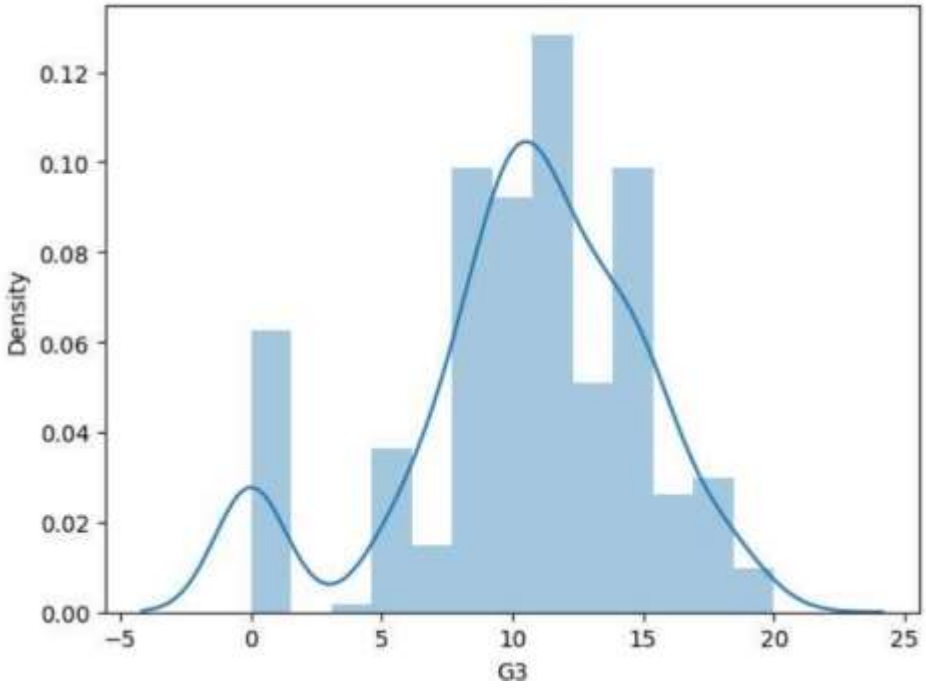

Fig 3:- average marks graph

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

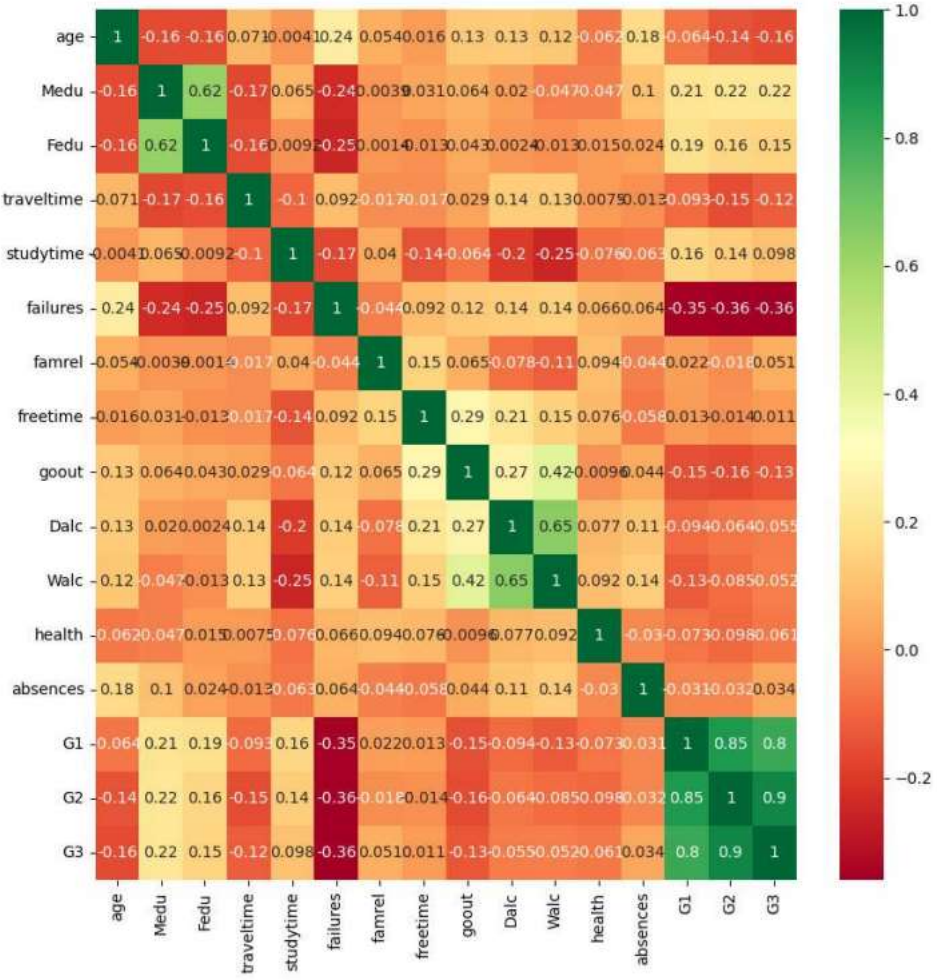

Fig 4 :- correlation matrix of each features

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

```
[18.33356333] [ 6 18 18  4  3  3  1] [18]
[4.72525591] [6 5 6 2 4 4 4] [6]
[10.82503374] [ 2 10 11  2  3  4  4] [12]
[14.87787915] [ 8 15 14  2  5  2  1] [14]
[8.65073109] [ 0 10  9  2  5  3  3] [9]
[12.06519617] [40 13 11  3  4  2  3] [11]
[7.06875815] [20  9  7  2  5  1  3] [8]
[8.80298129] [ 0  8 10  1  3  4  4] [11]
[9.47226275] [ 4 10 10  2  4  2  2] [10]
[9.6220482] [38  8  9  1  4  4  4] [8]
[11.03551924] [12 11 11  2  4  4  3] [11]
[8.59084992] [6 9 9 1 4 1 1] [10]
[10.6147186] [ 8 13 11  3  3  2  2] [11]
[14.72534195] [ 0 13 15  1  4  0  2] [15]
[9.2895703] [ 0  8 10  2  5  3  3] [12]
[12.68651733] [ 0 14 13  4  4  4  3] [14]
[8.97092372] [12 11  9  2  4  2  2] [9]
[5.89980256] [2 7 7 2 4 4 4] [7]
[8.72497588] [ 7 10  9  3  5  4  4] [9]
[3.42821461] [0 6 5 1 3 3 4] [0]
[5.78866586] [18  8  6  2  4  4  4] [7]
[14.440037] [ 2 11 15  1  3  4  3] [15]
[15.56890286] [11 16 15  2  4  3  4] [15]
[15.98297148] [ 4 15 16  1  2  4  2] [15]
[13.1178479] [ 0 13 13  2  5  2  2] [12]
[12.85351285] [ 2 13 13  2  4  3  3] [13]
[9.9680371] [ 0  8 11  2  4  1  1] [11]
[10.63518058] [ 0 11 11  1  4  4  4] [10]
[7.00327025] [0 8 8 2 4 2 2] [0]
[9.08422793] [ 6  8 10  2  4  2  3] [10]
[14.63374054] [ 4 15 14  3  5  4  2] [15]
[6.34887606] [0 8 7 2 5 2 2] [8]
[4.79452078] [2 8 6 2 3 1 1] [5]
[3.84820027] [4 6 5 2 4 3 3] [6]
[5.98993888] [6 7 7 3 4 4 3] [7]
[15.0898536] [ 0 15 15  2  4  3  3] [15]
[8.47082574] [8 8 9 1 4 3 3] [10]
[12.43949157] [ 0 13 13  2  4  1  4] [13]
[10.63518058] [ 0 11 11  1  4  4  4] [12]
```

Fig 5 :- predicted average score result

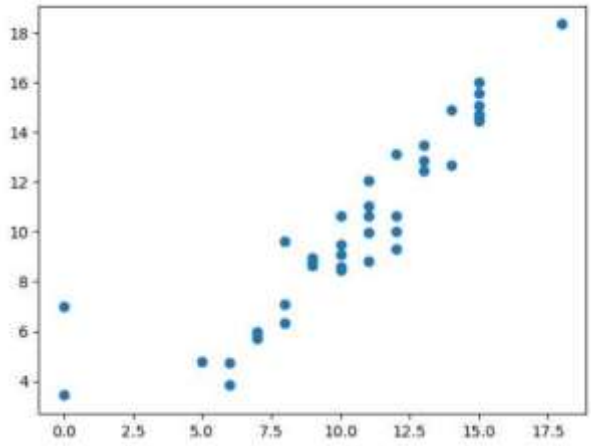

Fig 6 :- predicted score graph

6.CONCLUSION AND FUTURE SCOPE

In this study, we presented a novel strategy for estimating students' degree programme achievement in the future based on their previous and present performance. To find pertinent courses for building base predictors, a latent component model-based course clustering method was created. The forecast was created using an ensemble-based progressive prediction architecture that took into account the pupils' changing performance. These data-driven techniques can be combined with other pedagogical techniques to assess student performance, and they give academic advisors important information they can use to suggest follow-up courses and implement pedagogical

intervention strategies as needed. Additionally, this work will influence how degree programmes' curricula are designed as well as how education policies are designed generally. The performance prediction will be expanded to elective courses in the future, and recommendations will be made using the prediction results. students with coursesThere may be pre-requisite courses for direct prerequisite courses. The sequence of prerequisite courses is regarded in this benchmark

7.REFERENCES

[1] Han, J. and Kamber, M., (2006) Data Mining: Concepts and Techniques, Elsevier.
[2] Dunham, M.H., (2003) Data Mining: Introductory and Advanced Topics, Pearson Education Inc.
[3] Kantardzic, M., (2011) Data Mining: Concepts, Models, Methods and Algorithms, Wiley-IEEE press.
[4] MySQL – The world's most popular open source database, http://www.mysql.com/ [5] J.R.Quinlan, "C4.
[5]: Programs for Machine Learning", Morgan Kaufmann Publishers, Inc, 1992.
[6] Quinlan, "Learning decision tree classifiers". ACM Computing Surveys (CSUR), 28(1):71–72, 1996.
[7] J.R.Quinlan, "Induction of Decision Tree", Journal of Machine learning", Morgan Kaufmann Vol.1, 1986, pp.81-106.
[8] J. Quinlan, "Learning decision tree classifiers". ACM Computing Surveys (CSUR), 28(1):71–72, 1996.
[9] Shreenath Acharya, Madhu N, "Discovery of students' academic patterns using data mining techniques" IJCSE,Vol. 4 No. 06 June 2012