World Journal of Engineering Research and Technology
WJERT
www.wjert.org
SJIF Impact Factor: 5.924

PREDICTING OF STUDENT PERFORMANCE USING DATA MINING CLASSIFICATION TECHNIQUES

Rahul Sharma*[1], Dr. Shiv Shakti Shrivastava[2]

[1]Research Scholar, Department of Computer Science and Engineering, Rabindranath Tagore University, Bhopal (M.P.).

[2]Professor, Department of Computer Science and Engineering, Faculty of Engineering, Rabindranath Tagore University, Bhopal (M.P.).

Article Received on 12/01/2023 **Article Revised on 02/02/2023** **Article Accepted on 22/02/2023**

***Corresponding Author**
Rahul Sharma
Research Scholar, Department of Computer Science and Engineering, Rabindranath Tagore University, Bhopal (M.P.).

ABSTRACT

In this research work, we are analyses and evaluates student performance in the Department of Computer Science, Parul Institute of Technology. The data were collected for two (2) years intake from July 2017 to June 2019 it contains student previous academic records Such as course code, course Name marks obtained for each student by applying the classification technique algorism in Rapid Miner tool. Data mining provides good and powerful methods for education and another different field of study. Due to the vast amount of data of student which is used to find out valuable information which can be used to determine the student success. In this paper a classification task was used for the prediction. A decision tree model is applied during the experiment. The results indicates that it is possible to predict the graduation performance, in addition, a procedure for evaluating the performance for each course have identified.

KEYWORDS: Data Mining, Rapid Miner, Classification, Decision Tree, Random Forest, Graduation Performance.

1. INTRODUCTION

Data mining Data mining is that the knowledge of accomplishing useful information from an enormous data set which normally stored in our college and University Database, by extracting data rule analysis, data dredging, data archeology, information harvesting, and

business intelligence. Education is one among the foremost important aspects/issue for the event of the many societies, and therefore the main purpose implementation of the colleges and college's implementation is to produce good and Quality eduction (Hamsa et-al., 2016).

In nowadays of the tutorial system of upper learning, our university database contains such a lot huge of information of scholars with different courses and subjects upper and different years of intake. The data keeps increasing by times. To realize knowledge's from it by using one among the statistical techniques, and data processing tools is additionally wont to predict the massive data kept in our schools database. The knowledge will be extracted from it.

This research paper is organized into a distinct sections, the related work section is briefly on the previous works of other authors and also the classification technics is employed.

2. RELATED WORKS

The practical application of data mining is spread in educational sectors in many different fields. In educational institutions, many researchers' academics authors have widely discussed the application of data mining in institutions of learning. The authors have gone through the literature to understand the importance of data mining technique and its applications. Data mining is a powerful tool for academic purposes, marketing, and institutional effectiveness. Data Mining is a good and suited technology that can be utilized by lecturers, students, managers, and other supporting staff. And it a useful tool for decision making on their educational activities. (Femina & Sudheep, 2015). Data mining techniques can be used to analyze the student's academic data to promote and enhance the quality of the higher Education system. The Management can use the classification technique using a good tool to enhance the outcomes of the courses. The knowledge can be utilized for better understanding of the student performance in the course faculty performance and it can be used in the faculty and managerial decisions for the necessary action to be taken. Using such knowledge the management can amend their policies and strategy for better decision making. (M.S. Mythili & A.R.Mohamed Shanavas, 2014).

In research a data was collected from the records of forensic laboratory results. And the biological evidence received from the laboratory of the forensic genetic department of molecular medicine university of Padua. The sample of 122 sexual cases are processed by the

laboratory from 1st January 2010 and 31st December 2014 are considered. (Tozzo et al., 2018). The study was conducted in classification trees to anticipate student academic achievement by using student gender, previous school marks, admission type, location of living, a medium of teaching, mother's qualification, father occupation, Mothers occupation, and family annual income, etc. In their work they have achieved around 62.2%, 62.2% and 67.77% is the overall prediction accuracy using CART (classification And Regression trees), ID3m (iterative Dichotomiser 3) and decision tree algorithm respectively. A student-facing dashboards ware developed to show the learning patterns to the students by themself. The intention is to promote and self-reflection. The goal is to motivate them for better performance in their academics. The authors adopt the use of Pearson correlation to compare student academic performance in both homework and exam questions. (Deng et al., 2019).

The research was carried out in a group of 50 students which are enrolled for the specific program course of four (4) years (2007-2010) performance including: "previous semester marks", "class test", "Summer performance ", "Assignments", "Attendance", "lab work". They have used IDB3 decision tree algorithm for the construction of a decision tree algorithm. The if-then rules will assist the teachers for better understanding which helps to predict the students' performance. They defined their objectives of the study as: the study will also work to figure out those students which needed special attention to reducing the failure rate and taking necessary action for the next semester examination. (Saa, 2016). A broad analysis and comparison of the state of the machine learning technics ware performed for solving the student exam performance prediction. The labeled dataset is to be used for the preparation of analyzing data. The supervised machine learning technic ware considered. (Tomasevic et al., 2020) A fuzzy technic ware used as the mining tool to predict student performance at the end of the semester. Several attributes ware also considered: Previous academic record (PAR), previous semester marks (PSM), Attendance (ATT) and end of semester marks (ESM), are considered as an output variable. (Czibula et al., 2019).

A study was carried out and a multiple tools for analysis ware use, data collection, processing selection and balancing the class. It compared according to t-test with AODE, KNN, (IBK) and j48 decision tree with other classifiers namely Naïve bayes, KSTAR, ZeroR and OneR. The authors proposed (KNAND) which is better than the other classifiers. (Pandey & Taruna, 2016)

The hierarchical fine-grained skill model is presented and a generalized mixed linear effort model was used for generating the statistical formation for describing student learning activities on different skills. (Yang & Li, 2018). Personal information which are stored in the electronic devices such as: tablets, phones laptops and disk etc. the stored document and information that are stored on this devices can be lost due to corruption of this storage devices. (Taskin & Kucuksille, 2018).

A study was also conducted and a combining rules induced by JRIP and PART algorithms. The data is collected from different servers such as Alexa, yahoo, common crawl and a Phis Tank. The preprocessed are taken to extract useful features. (Adewole et al., 2019). A research was conducted and a researcher collected a record of dataset that contains frequency of the involvement of activities for a specific course by the student. The study incorporates academic data and socio-demographic information during student enrolment. The data is obtained from the school LMS (Learning Management System). (Helal et al., 2018).The author used a prediction of student performance model for early identification of student at-risk for attending large classes' that are more than 50 intake. The instructors can use and apply educational interventions to reduce failure rate. (Sandoval et al., 2018) Students' academic records are stored in the offices of the engineering faculties and these records includes the performance of student at different subject as well as the information regarding the student origin, age, previous studies information. All this information should be enough and help to categorize class of the students we are dealing with. (Vilanova et al., 2019).

3. AIMS OF THE PRESENT STUDY

This research paper is intended to answer the following questions: How can we predict the student's performance at the end of their study in the previous year course they undertake? How can we rate the student with passed, pending courses, repeat semester, withdrawn and voluntary withdrawn? Therefore if the prediction is reached it makes the implementation of the performance easier. If the major course's that can make the impact on the student's graduation performance a measure are to be taken at their level.

RAPIDMINER

It is a data Science software that is developed by the Rapid-Miner Company it provides an integrated condition for data preparation, deep learning, predictive analysis, text mining, and machine learning. It is used for business and commercial services. Rapid-Miner also uses in research, education, rapid prototyping, training, application development and it supports all

steps of the machine learning techniques and process including data training, model validation, Result visualization, and optimization. RapidMiner Software initial release in 2006 and stable release Version 9.2/ January-2019. (Markus Hofmann, 2013)

4. DATA MINING TECHNIQUES

This paper uses a Rapid-Miner tool and some of the operators which are useful to carry out this research work. The operators and used are ReadExel, filter the decision tree and Random forest algorism also used. This algorisms have the potentials to yield good result.

A. Classification

Classification is one of the data mining technique which is used to classify each data items to a predefined set of class label. Classification technique is the method of mathematical technique such as decision tree, linear programming and statistics. Each tuple that constitutes the training set is referred to as a class. Theses tuple can also be referred as sample, data point or object.(Zentut, 2019).

B. Decision Tree

Decision tree is an aggregation of nodes which is meant to produce a decision a decision of values to a class on any estimated numerical values. Each node correspond a specific splitting attribute. The rule separate values belonging to a different class, it separate them in order to reduce the error. The new node building is repeated until the ending criteria is satisfied. The prediction of a class marks are determined depending on the absolute majority of examples which reach this leaf during the generation. The estimation of numerical value is fund through averaging out values in a leaf. (Rapid-Miner, 2019).

5. DATA MINING PROCESS

Data Set Used

The data set used in this research paper was taken from a student's database by the institution. The sample of the student's records were collected for two years intake from July 2017 session to June 2019 from the department of computer science.

6. PROPOSED DATA MINING MODEL

This model will present the proposed framework. The classification model is to predict student academic progress. It uses initial academic information's and records of a specific period. The model shows the steps taken in developing a model to predict student academic

performance. Figure: 1 illustrates the four stages involved in the research. Academic potential of the students, integration and data transformation extraction, classification and predictions.

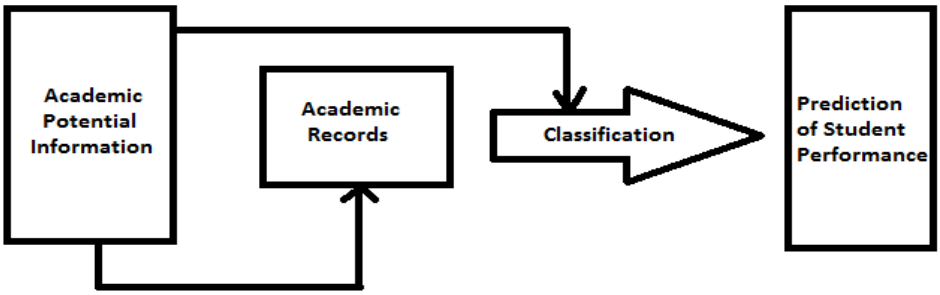

Figure: 1 Proposed Data Mining Model.

7. EXPERIMENTAL DESIGN

Data is stored and process using the model the marks are converted to grades so that the result can easily be classified. The student's names are also excluded because the names cannot be classified. Read excel operator was used to loading data from MS Excel spreadsheet. The filter is used here to filter various data or information. A decision tree operator was also used for the classification. These are the pre-processing steps that are carried out for the study. When it is fully done the classification is carried out accordingly but only the required and necessary fields were selected.

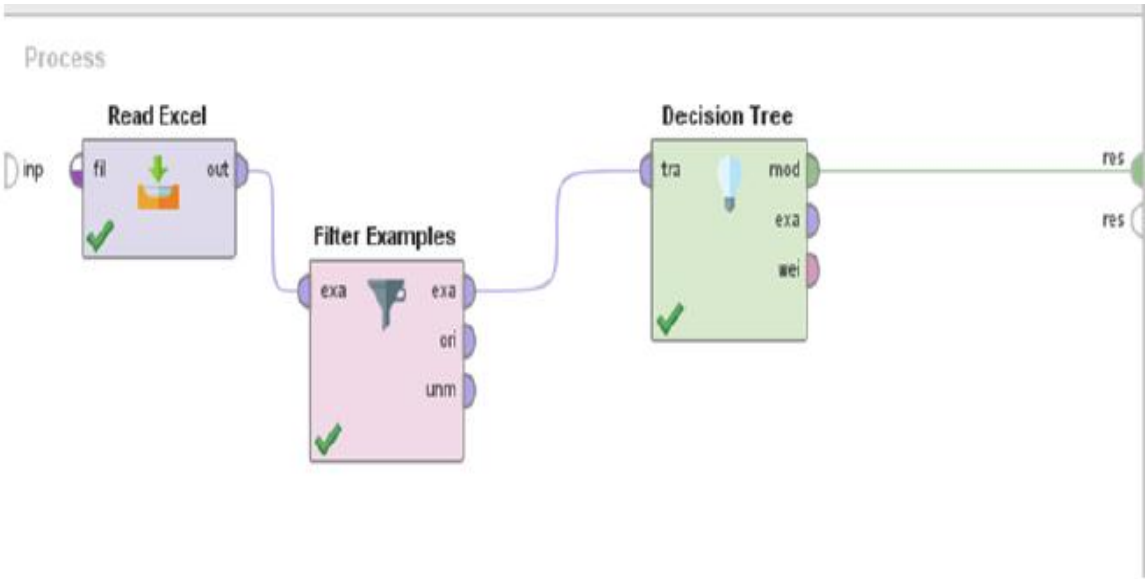

Figure 2: Classification Model.

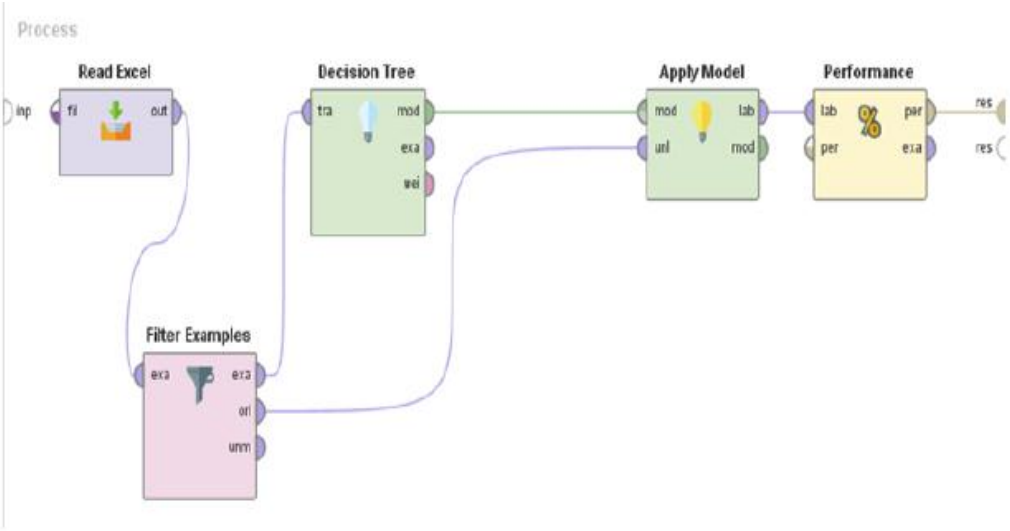

Figure 3: Performance Model.

In this model the additional two operators were added it is use for the performance evaluation it list the performance standard and the values are automatically determined the order that will fit the learning task.(Ahmed & Elaraby, 2014) The operator apply model it applies a model on an example set. A model is first trained on an example set by another operators which is a learning algorithms. The goal is to predict on unseen data or transformation of data by applying a pre-processing model. (Rathipriya, n.d.).

8. RESULTS AND DISCUSSION

The data set used in this research study contains student gender, courses, marks, CGPA, and remarks. The final year examinations scores of students in individual subjects. The knowledge here is transform into the rules. The decision tree shows that some students with CGPA greater than 2.625 are the students who passed all their courses without re-appearing. And the students with CGPA less than or equal to 2.625 they have pending courses to reappear again or rewrite. Those with CGPA greater than 1.385 are those with repeat semester because of very low marks in most of the courses undertake. Students with less than or equal to 1.385 withdrawn because they fail with low mark and grade, absentees in the examination and voluntarily withdraw. These findings can be used to implements some good policies. A lecturers can report the students with less CGPA. Courses with a higher number of failures lead to having pending courses at a final year students can be observed by the decision-makers or Head of Departmental. The lecturer can focus on such a category of students that needs more academics assistance.

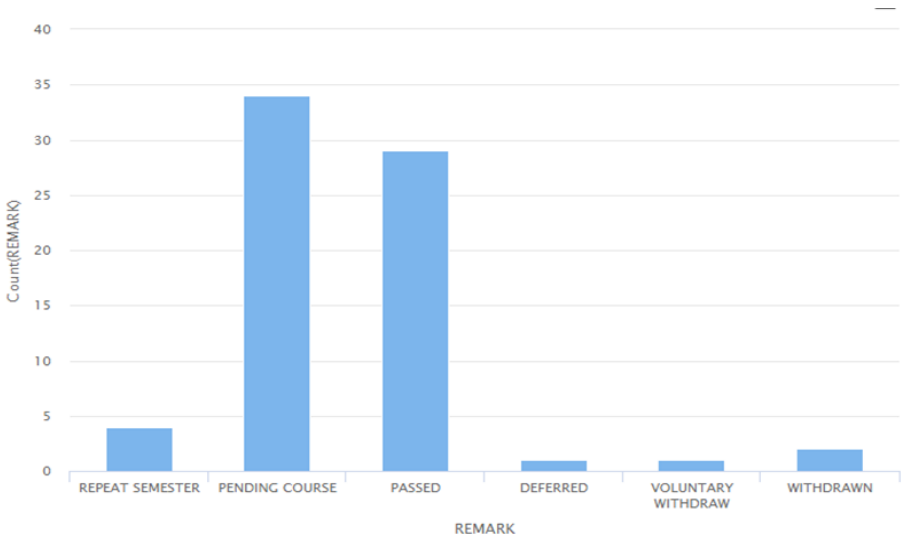

Figure 4: Visual representation of Remarks.

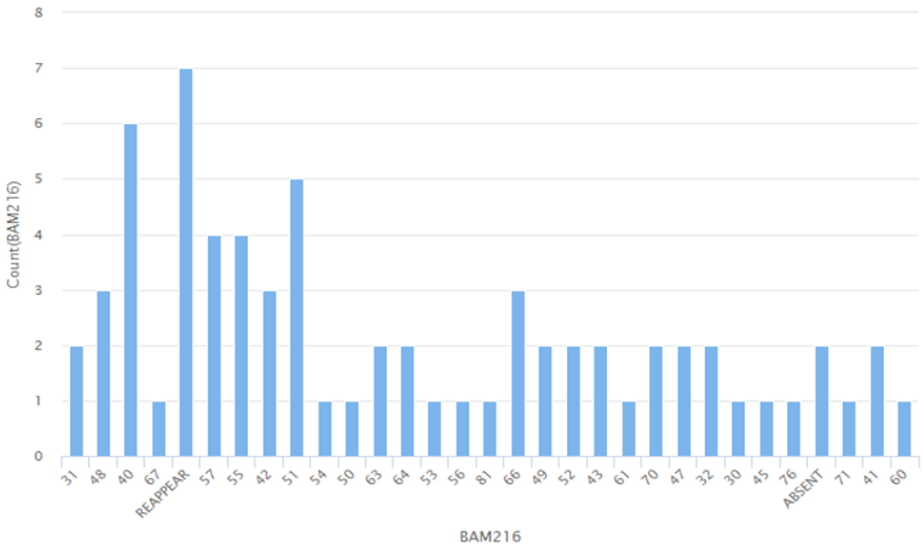

Figure 5: Course with highest number of failure (Reappear).

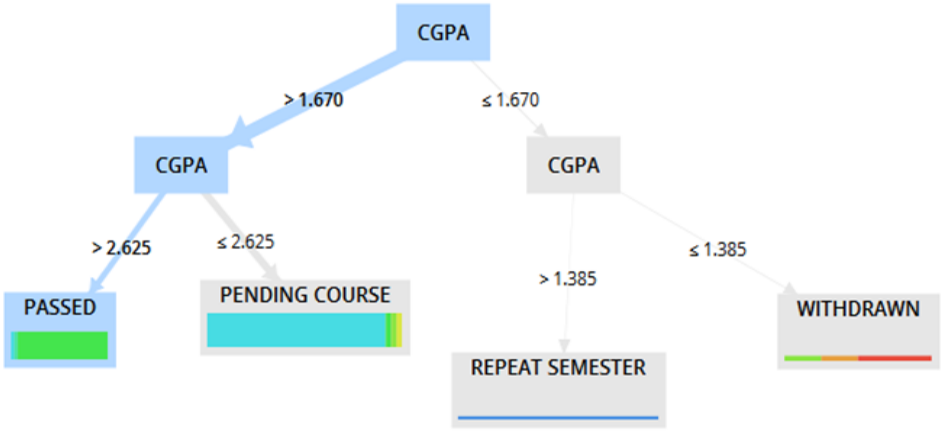

Figure 6: Decision Tree.

DECISION TREE

CGPA > 1.670
| CGPA > 2.625: PASSED {
REPEAT SEMESTER=0, PENDING COURSE=1, PASSED =28, REPEAT SEMESTER=0, DEFERRED=0, VOLUNTARY WITHDRAW=0, WITHDRAWN=0}
| CGPA ≤ 2.625: PENDING COURSE {
REPEAT SEMESTER=0, PENDING COURSE=33, PASSED =1, REPEAT SEMESTER=1, DEFERRED=1, VOLUNTARY WITHDRAW=0, WITHDRAWN=0}
CGPA ≤ 1.670
| CGPA > 1.385: REPEAT SEMESTER {
REPEAT SEMESTER=2, PENDING COURSE=0, PASSED =0, REPEAT SEMESTER=0, DEFERRED=0, VOLUNTARY WITHDRAW=0, WITHDRAWN=0}
| CGPA ≤ 1.385: WITHDRAWN {
REPEAT SEMESTER=0, PENDING COURSE=0, PASSED =0, REPEAT SEMESTER=1, DEFERRED=0, VOLUNTARY WITHDRAW=1, WITHDRAWN=2}

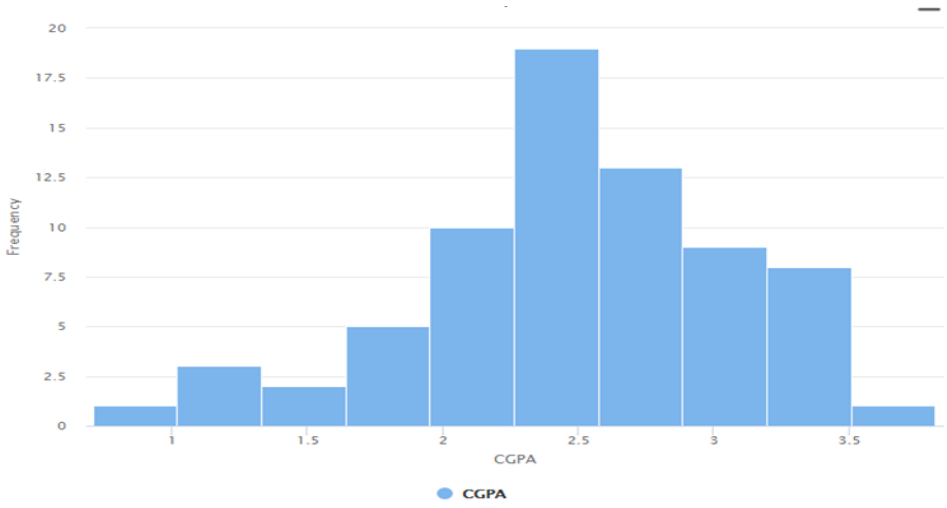

Figure 7: CGPA Histogram.

9. CONCLUSION AND FUTURE WORK

The data stored in our educational database in the institutions is increasing quickly by time. To get knowledge from such data. Student success and improvement can be determined. The classification technique is applied using the Rapid Miner tool. The research result shows the classification accuracy by adding a performance model operator for the evaluation. The model would help the lecturer to take necessary actions to assist the students with problems

in their courses that lead to repeat semester, voluntarily withdraw and withdraw may be due to poor performance by the students or lecturers. The limitation of this research is the small amount of data as the reasons for some missing values in the data collected. This study will be expanded in the future by adding more data to a greater extent from different years to increase the accuracy of the prediction.

REFERENCES

1. Adewole, K. S., Akintola, A. G., Salihu, S. A., Faruk, N., & Jimoh, R. G. (2019). Hybrid Rule-Based Model for Phishing URLs Detection. In Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST (Vol. 285). Springer International Publishing. https://doi.org/10.1007/978-3-030-23943-5_9
2. Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data Mining: A prediction for Student's Performance Using Classification Method. World Journal of Computer Application and Technology, 2(2): 43–47. https://doi.org/10.13189/wjcat.2014.020203
3. Czibula, G., Mihai, A., & Crivei, L. M. (2019). S PRAR: A novel relational association rule mining classification model applied for academic performance prediction. Procedia Computer Science, 159: 20–29. https://doi.org/10.1016/j.procs.2019.09.156
4. Deng, H., Wang, X., Guo, Z., Decker, A., Duan, X., Wang, C., Alex Ambrose, G., & Abbott, K. (2019). PerformanceVis: Visual analytics of student performance data from an introductory chemistry course. Visual Informatics, 3(4): 166–176. https://doi.org/10.1016/j.visinf.2019.10.004
5. Femina, B. T., & Sudheep, E. M. (2015). An efficient CRM-data mining framework for the prediction of customer behaviour. Procedia Computer Science, 46(Icict 2014): 725–731. https://doi.org/10.1016/j.procs.2015.02.136
6. Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. Procedia Technology, 25: 326–332. https://doi.org/10.1016/j.protcy.2016.08.114
7. Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018). Predicting academic performance by considering student heterogeneity. Knowledge-Based Systems, 161: 134–146. https://doi.org/10.1016/j.knosys.2018.07.042
8. M.S. Mythili, M., & A.R.Mohamed Shanavas, D. (2014). An Analysis of students' performance using classification algorithms. IOSR Journal of Computer Engineering, 16(1): 63–69. https://doi.org/10.9790/0661-16136369

9. Pandey, M., & Taruna, S. (2016). Towards the integration of multiple classifier pertaining to the Student's performance prediction. Perspectives in Science, 8: 364–366. https://doi.org/10.1016/j.pisc.2016.04.076
10. RapidMiner. (2019). （Decision Tree. https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/parallel_decision_tree.html
11. Rathipriya, R. (n.d.). Predicting the Model for Academic Performance Abstract : VIII(I): 46–54.
12. Sandoval, A., Gonzalez, C., Alarcon, R., Pichara, K., & Montenegro, M. (2018). Centralized student performance prediction in large courses based on low-cost variables in an institutional context. Internet and Higher Education, 37: 76–89. https://doi.org/10.1016/j.iheduc.2018.02.002
13. Saa, A. (2016). Educational Data Mining & Students' Performance Prediction. International Journal of Advanced Computer Science and Applications, 7(5): 212–220. https://doi.org/10.14569/IJACSA.2016.070531
14. Taskin, S. G., & Kucuksille, E. U. (2018). Recovering Data Using MFT Records in NTFS File System. Academic Perspective Procedia, 1(1): 448–457. https://doi.org/10.33793/acperpro.01.01.88
15. Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers and Education, 143(February 2019): 103676. https://doi.org/10.1016/j.compedu.2019.103676
16. Tozzo, P., Ponzano, E., Spigarolo, G., Nespeca, P., & Caenazzo, L. (2018). Collecting sexual assault history and forensic evidence from adult women in the emergency department: A retrospective study. BMC Health Services Research, 18(1): 1–6. https://doi.org/10.1186/s12913-018-3205-8
17. Babandi Usman, Rabi'u Adamu and Sani Salisu, Prediction of Student Performance Using Classification Technique, International Journal of Information Processing and Communication (IJIPC) 8(1): (May, 2020).
18. Vilanova, R., Dominguez, M., Vicario, J., Prada, M. A., Barbu, M., Varanda, M. J., Alves, P., Podpora, M., Spagnolini, U., & Paganoni, A. (2019). Data-driven tool for monitoring of students performance. IFAC-PapersOnLine, 52(9): 190–195. https://doi.org/10.1016/j.ifacol.2019.08.188

19. Yang, F., & Li, F. W. B. (2018). Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. Computers and Education, 123(April): 97–108. https://doi.org/10.1016/j.compedu.2018.04.006
20. Zentut. (2019). Z. ZenTutWebsite. https://www.zentut.com/data-mining/data-mining-techniques/