ISSN: 2666-2795 Vol. 6 No.2 December, 2021, Netherland

International Journal of Applied Engineering Research

Performance Evaluation to Estimate the Dropout of IUB Students Using Data Mining Techniques

Saira Liaquat[1], Ghulam Gilanie[2, *], Sana Cheema[2], Akkasha Latif[2], Muhammad Ahsan[2]

[1]Department of Data Science, Faculty of Computing, the Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan

[2]Department of Artificial Intelligence, Faculty of Computing, the Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan

[*]Corresponding Author: Ghulam Gilanie. Email: ghulam.gilanie@iub.edu.pk, sairalaiquat7@gmail.com, sanacheema887@gmail.com, akashacheema70@gmail.com, chahsan146@gmail.com

Abstract: Data mining is the process of extracting data from huge database to collect vital and beneficial knowledge. Classification is one of the techniques which is already uses in data mining. The technique was used is Decision tree C4.5. Decision Tree is a method that converts information into decision tree that represents understandable rules. Decision tree is beneficial for discovering data units as well as determine the inherited connections between several input and target variables. When decision tree develops, rules of a case would be obtained, and Data science software Rapid Miner was used for implementation. The purpose of this study was to classify student data at "The Islamia University of Bahawalpur" and to find out the factors of students who experienced dropout. The attributes used consisted of age, school of origin, parents' occupation, Marks obtained in first semester and marks obtained in 2nd semester,SGPA,Grad1 and Grad2,CGPA and cumulative grade .The most influential attribute of the dropout student was the origin of school. The calculation of result and the obtained accuracy is 98.04.

Introduction

Data mining, also known as the discovery of knowledge, is one of the growing the fields due to huge demand of modernworld and extract information from large-scale, database. Data Mining Techniques have been used in many areas of life such as banking, fraud detection and telecommunication. Today'sdata mining techniques were used to improve and evaluate the educational tasks. There ismuchresearch thathas been done to examine student's dropout in Latin America.Most of the researches are about to find out the factors that led to the desertion, calculate the number of dropout students and mechanism to reduce it[1].There are two different proposals for the assessment of student'sdropout: the one is to established ratio of students whose are doing graduation in each time corresponding to time to get the degree; and the second one is just total number of students who dropout their studies.To reduce student dropout, this study proposes to enhance techniques and methods for early detection of student dropout. According to the author[2] data mining is divided into many groups based on the jobs that can be done namely, description, estimation, prediction,classification, clustering and association.The most usedtechnique is the classification technique.

Classification is a process/ technique for finding models that describe or distinguish the idea or class of data, with the objective of being able to quantify the class of an object who is unknown. There are several methods in the classification, like decision treeC4.5, neural network, and fuzzy logic. The Decision tree C4.5 algorithm is one of the most popular algorithms used for data classification that has numerical and categorical attribute.

Studies on Decision Tree C4.5 had been done to find out the client's fulfillment with the results of study that shows accuracy by 91%, with a precise value on the satisfied prediction by 92.21% and the precision value on the displeased side the prediction by 90.91%.class recall for satisfied by 97.71% and class recall for dissatisfied by 75%[2].According to another study in which describes the implementation of Decision tree C4.5 to predict the user service of a mobile operator in an event based on several indicators, weather, distance relative to the event location, as well as the service user in addition to postpaid clients or not. The correctness and functioning of machine learning is particularly depending on the existing data and the interpretation of the technique applied, in this research researchers acquire predictive correctness with the same number in each method used. Additionally, Decision Tree is also used to predict clients credit limit with the decision concluded that the establish application can help the fund section in examine client data to find the target of credit marketing so it is look forward that the operational cost of banking marketing can be keep down.

The decision Tree technique turns huge number of facts into decision tree that describe the principal. The objective of this research is to analyze student's data extracted with data mining decision tree C4.5 so that new knowledge that can know the indicator of dropout students will be obtained.

Literature Review

Students dropout can be defined as the student who are not able to complete their degree within the given time period as recommended by the management of the institute[3].This issue particularly majorly causes students expertise and willingness to drop out in their fields and very much affect the standard of their organization. students dropout matter is not a new problem which occurred in today's world but still a prominent issue that majorly need to be focused on and hence, this matter which is slowly increased in many Universities and other institutions has divert the focus of the research scholars because of its impact on reducing the values of higher education which is majorly affect the social culture, where other expected students lose their opportunity to continue their studies in universities[4].

One of the most powerful and prominent technique used by other research scholars to solve this issue is just implement the data mining technique called Educational Data Mining (EDM) used to prevent the student dropout[5].EDM gives a wide range of algorithmic techniques to elaborate different issues that relate to the educational system as well as initiate new ways for predicting academic results and students behavior especially to prevent variations or indicators that influence dropout in educational institutions[6].

Data mining contains a searching technique of current or model in a large database for future decision making.It is required that data mining devices can identifythese patterns of data with minimum input. These repeated decorative design are recognized by particular devices that can provide a functional and perceptive data analysis that can later be studied more closely, which might be used by other decision-making which might be used by other decision –making support took[7].Data mining covers the sub-area of statics which is called exploratory data analysis, havingthe same purpose and completely dependent on statistical techniques. Data mining is co-related with the artificial intelligence sub-fields called knowledge discovery and machine learning. The most important feature of data mining is the huge volume of data that comes from different related fields can be implemented to data mining problems, the scalability linked with the size of data becomes an important new criterion.

According to the author el all in Anik Andriani, data mining is defined as the process of determining patterns in a data set. This process must be fully automatic or semi-automatic behavior. Thefinal pattern should mean that the pattern gives some advantages. The patterns are to be rectified,validated, and for later use to make a prediction. It can be summaries from the above details that data mining is the process to extract the information from the large-scale database.

Another author Cohen [8] mention about the online courses. The chances of students dropout is predicted to rectifying the pattern of the students whether they continue or stop their study is just through the classification technique for instance the decision tree method provides the analysis of the student dropout using most frequent patterns, similarity calculation and correlation[9].Moreover, helpful information based on the students information produced is used to help the management of the institution to establish the stronger approach and strategy with regard to the planning and execution of educationalprogram so that the matter of the student dropout is scale-down conclusively[10].

To estimate the trend of students dropout using EDM, many measurements are widely employ by research scholar[11].The possible variable used as the standard are cumulative grade point average(CGPA), internal evaluation, student demographics factor, external indicators, outside the normal routine, initial schooling, and social environment but the most powerful variables used to forecast students dropout are CGPA and internal assessment factors because of its purpose and usefulness in maximizing the evaluation of the students expertise and skills in the nearby and fortune time[4].

With reference to all the indication that led to the matter of all the student's dropout to take place is based on sector of population which mention the gender that effect the standard of education in the conventionaland also in e-learning way of education. Although, the demographic characteristics such as date of birth,students Active or dropout status,parents' impact, employment chances,maritalstatus, economic restriction can also be one of the major and possible factors to affect the proclivity of students drop out of the university. Data mining can be divided into various stages is illustrated in Figure 1.

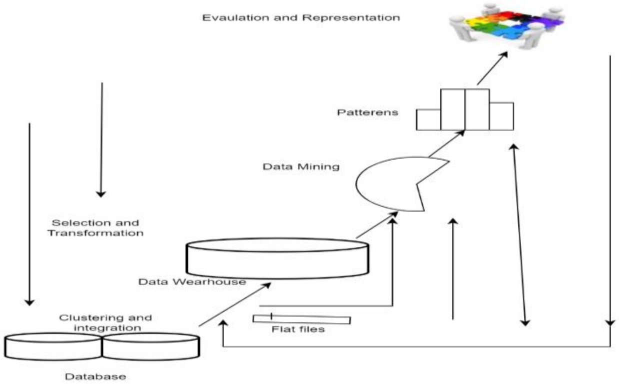

Evaulation and Representation
Patterens
Data Mining
Selection and Transformation
Data Wearhouse
Clustering and integration
Flat files
Database

Figure 1: Data mining stages

Details:

1. Data cleaning (to dispose of conflicting information and commotion).
2. Data incorporation (joining information from numerous sources).
3. Transformation of information (information is change over into the suitable structure of computing).
4. Application of computing techniques.
5. Evaluate design found (discover important one).
6. Presentation of information (with perception strategies).

Information mining bunching comprise of portrayal, estimation, forecast, afflation, grouping and classification. In the classification, there are target classification of factors. For instance, the classification of pay can be isolated into three classes: high pay, medium salary, and low pay. At that time point to discover the pay of a worker, the method for classification is utilized in information mining. All in all, the classification procedure should be possible in two phases, taking in process from information preparation and classification of cases. In the learning procedure, the classification algorithm forms the information preparing to deliver a model. When the model is tried and sufficient, at the classification arrange, the model is utilized to foresee the class of the new case to help the decision-production process. Decision trees are one of the most well-known classification techniques since it is anything but difficult to decipherfor people. The essential idea of the Decision tree algorithm is to change information into decision trees and rules.

The idea of information is the decision tree is as per the following.

1. Information is communicated in table structure with characteristics and records.
2. The Characteristic expresses a parameter made as a measure in the arrangement of a tree. One of the characteristics is a trait that expresses an information arrangement for each thing called the objective property.
3. Trait has values called occurrence. For instance, the weight trait has a case of overweight, normal, and underweight.

Decision Tree Algorithm C4.5 has contribution to the type of preparing tests and tests, preparing test are test information that will be utilized to fabricate a tried tree. What's more, example is information handle that will be utilized as parameter in ordering information. When all is said in done the decision tree C4.5 algorithm for building decision trees is as the accompanying.

In addition, the decision tree C4.5 algorithm is used for developing the decision tree is as defined as described:

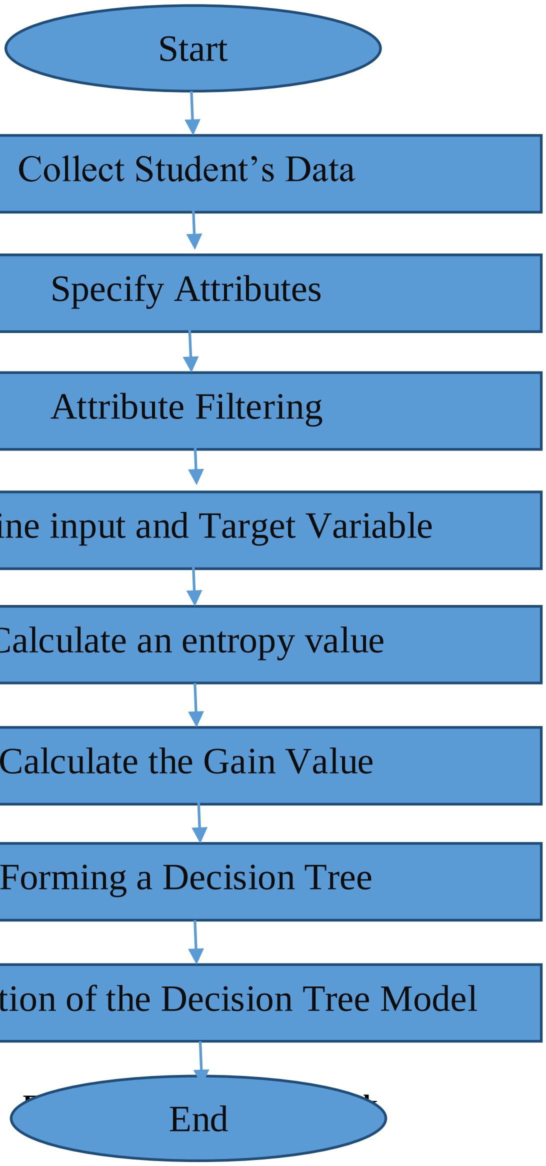
1. Select a property as root.
2. Make a branch for each worth.

3. Partition the case in the branch.
4. Rehash the procedure for each branch to have a similar class.

Rapid miner is one of the products utilized in data mining process, with C4.5 algorithm technique. What's more, there are other programming, for example Weka, Sipina, Math lab, etc. The Upside of Rapid Miner is to have the option to apply different algorithms and pooling information representation highlights. Rapid Miner is simple and effective for figuring with quick relative time contrasted with other programming.

3. Research Method

When all is said in done there are few phases right now following the general example of logical research appeared in Figure 2.



3.2 Selections of Variables

Students' data used in this research is from the Islamia University of Bahawalpur from 2013 to 2017.The Variable used are as follows.

1. The origin of school
2. Students Age
3. Parent's Occupation
4. Student Status
5. 1st Marks Obtained
6. 1st Grade
7. SGPA1
8. 2ndGPA
9. 2nd marks obtained.
10. SGPA2
11. CGPA
12. Grade

In the classification procedure, the yield of each focused-on information or class must be a number or a discrete. Understudies' information that utilized as target parameter or decision variable above are understudies' status which contain Active and Dropout parameter esteem. By taking a gender at the yield or record of the information in thestatuses filed a number or discrete, Active and Dropout, at that point the classification system can be applied to perform information extracting on the information.Directed information or class must be number or discrete[12]. Understudies' information that utilized as target parameter or decision variable above are understudies' status which contain Active and Dropout parameter esteems. Dynamic parameter esteem implies that understudies finished the examination well, while the dropout parameter esteems are the dropout understudies. By taking a gender at the yield or record of the information in the 'Status' field to be specific a whole number or discrete, Active and Dropout, at that point classification procedure can be applied to perform information Extracting on the information.

3.3 Implementation of Decision Tree C4.5

The framework used to decide and estimate students with dropout status is the Decision tree C4.5 algorithm and entropy calculated in equation1. The steps of the decision tree C4.5 algorithm are as described.

1. Arrange the data for training
2. Establish the root of the tree
3. Estimate the Gain value

$$Entropy(s) = \sum_{i=1^n} - \text{pi} * \text{log2pi} \qquad (1)$$

4.Revise the step2 until all the fields are split up

5.The decision tree division process will end up when all the fields in the n1 take the same class and or no element in the fields being further sub-divided and no fields in the empty branch.

Evaluation of the Decision Tree C4.5 Model

The Calculation is done by measuring the accuracy. The final output accuracy is evaluated by using confusion matrix.

3.4 Evaluation of the Decision Tree C4.5 Model

The calculation is done by measuring the accuracy. The final output accuracy is evaluated by using confusion matrix. The Evaluation of confusion matrix is measured based on true positive predictions(True positive), False positive predictions(False positive), exact negative predictions(True Negative),and false prediction(False negative).

	Detected	
	Positive	Negative
Positive	A: True Positive	C: False Negative
Negative	B: True Positive	D: True Negative

The Final output and result accuracy depend on the better model will be.The formula of measuring the accuracy is as defined in equation 2:

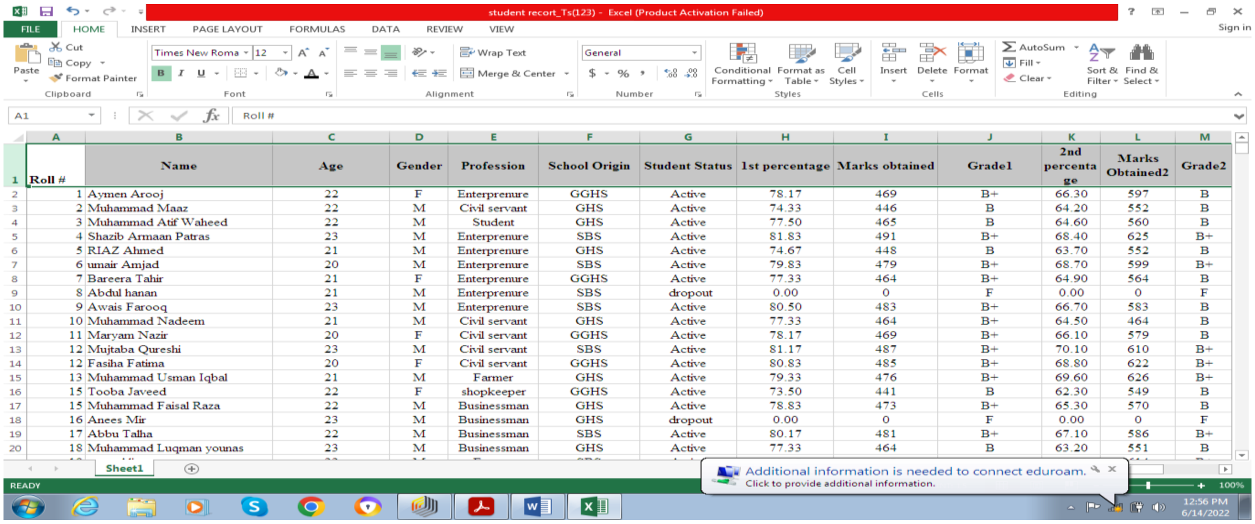
$$\text{Akurasi} = \frac{A+D}{A+B+C+D} \qquad (2)$$

Where A represent True Positive, B represent True Positive, C represent False Negative, D represent True Negative.

4 Results and Discussion

4.1 Data Collection

This research used students' data at "The Islamia University of Bahawalpur" with 1275 students. Student's data was taken from 2013 to 2017 with Active and dropout status. The student's dropout issue is not well-identified in the Islamia University of Bahawalpur, but student loses. The data taken was divided into data testing and data testing with a rate of 80% and 20%, so 1020 data used for training and 255 data testing were taken. Data training was used to obtain the classification of dropout students in the form of decision tree, while data testing was used to calculate the level of accuracy of classification result. Examples of data before preprocessing taken from IUB, BWP can be viewed in Figure 3.



student recort_Ts(123) - Excel (Product Activation Failed)

Roll #

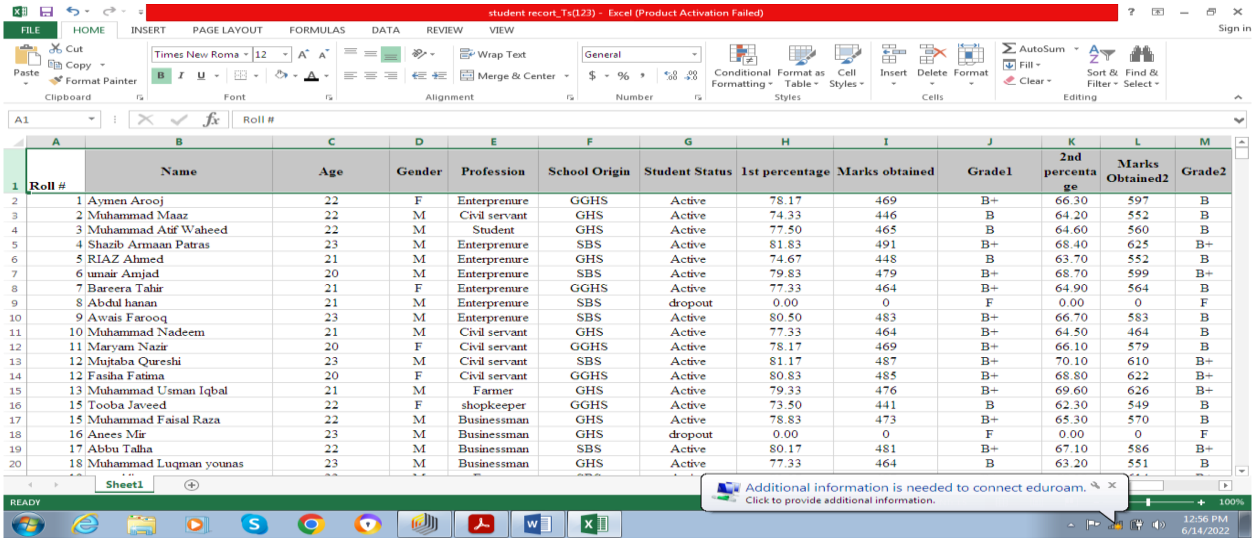
Roll #	Name	Age	Gender	Profession	School Origin	Student Status	1st percentage	Marks obtained	Grade1	2nd percentage	Marks Obtained2	Grade2
1	Aymen Arooj	22	F	Enterprenure	GGHS	Active	78.17	469	B+	66.30	597	B
2	Muhammad Maaz	22	M	Civil servant	GHS	Active	74.33	446	B	64.20	552	B
3	Muhammad Atif Waheed	22	M	Student	GHS	Active	77.50	465	B	64.60	560	B
4	Shazib Armaan Patras	23	M	Enterprenure	SBS	Active	81.83	491	B+	68.40	625	B+
5	RIAZ Ahmed	21	M	Enterprenure	GHS	Active	74.67	448	B	63.70	552	B
6	umair Amjad	20	M	Enterprenure	SBS	Active	79.83	479	B+	68.70	599	B+
7	Bareera Tahir	21	F	Enterprenure	GGHS	Active	77.33	464	B+	64.90	564	B
8	Abdul hanan	21	M	Enterprenure	SBS	dropout	0.00	0	F	0.00	0	F
9	Awais Farooq	23	M	Enterprenure	SBS	Active	80.50	483	B+	66.70	583	B
10	Muhammad Nadeem	21	M	Civil servant	GHS	Active	77.33	464	B+	64.50	464	B
11	Maryam Nazir	20	F	Civil servant	GGHS	Active	78.17	469	B+	66.10	579	B
12	Mujtaba Qureshi	23	M	Civil servant	SBS	Active	81.17	487	B+	70.10	610	B+
12	Fasiha Fatima	20	F	Civil servant	GGHS	Active	80.83	485	B+	68.80	622	B+
13	Muhammad Usman Iqbal	21	M	Farmer	GHS	Active	79.33	476	B+	69.60	626	B+
15	Tooba Javeed	22	F	shopkeeper	GGHS	Active	73.50	441	B	62.30	549	B
15	Muhammad Faisal Raza	22	M	Businessman	GHS	Active	78.83	473	B+	65.30	570	B
16	Anees Mir	23	M	Businessman	GHS	dropout	0.00	0	F	0.00	0	F
17	Abbu Talha	22	M	Businessman	SBS	Active	80.17	481	B+	67.10	586	B+
18	Muhammad Luqman younas	23	M	Businessman	GHS	Active	77.33	464	B	63.20	551	B

Additional information is needed to connect eduroam.
Click to provide additional information.

Figure 3: Examples of student's data o of Islamia University of Bahawalpur

4.2 Data Preprocessing

Preliminary processing is done to choose (to filter out) incomplete data and form a group for every attribute of data taken from IUB, BWP. The total collection of data used in this research was 1275 students. Examples of student's data from the preprocessing outcome can be viewed in Figure 4.


student recort_Ts(123) - Excel (Product Activation Failed)
FILE
HOME
INSERT
PAGE LAYOUT
FORMULAS
DATA
REVIEW
VIEW
Sign in
Times New Roma
12
General
A1
Roll #
Sheet1
READY
Additional information is needed to connect eduroam.
Click to provide additional information.
100%
12:56 PM
6/14/2022

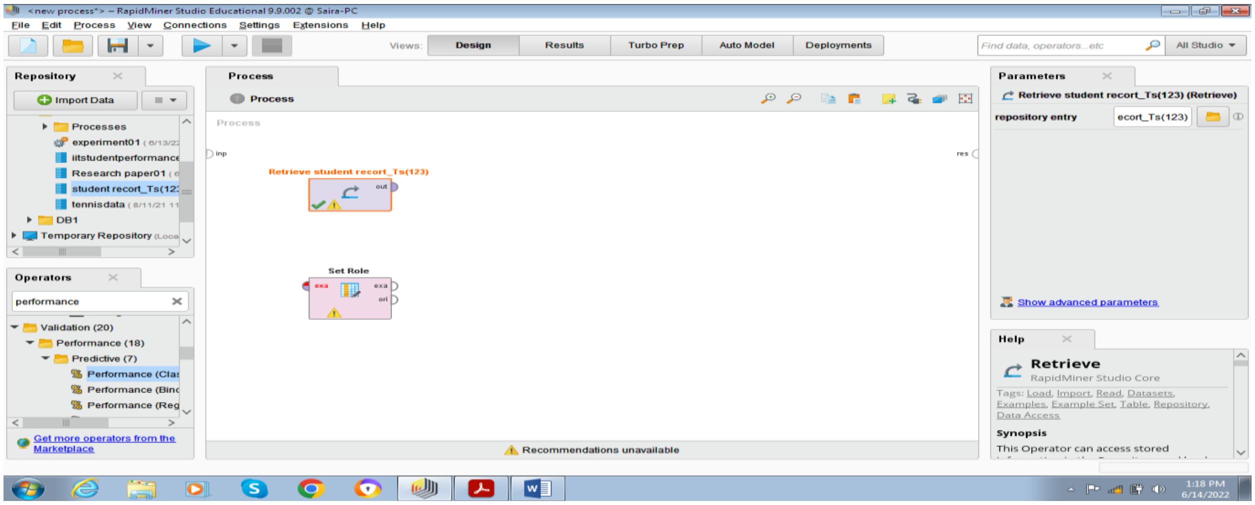
Roll #	Name	Age	Gender	Profession	School Origin	Student Status	1st percentage	Marks obtained	Grade1	2nd percentage	Marks Obtained2	Grade2
1	Aymen Arooj	22	F	Enterprenure	GGHS	Active	78.17	469	B+	66.30	597	B
2	Muhammad Maaz	22	M	Civil servant	GHS	Active	74.33	446	B	64.20	552	B
3	Muhammad Atif Waheed	22	M	Student	GHS	Active	77.50	465	B	64.60	560	B
4	Shazib Armaan Patras	23	M	Enterprenure	SBS	Active	81.83	491	B+	68.40	625	B+
5	RIAZ Ahmed	21	M	Enterprenure	GHS	Active	74.67	448	B	63.70	552	B
6	umair Amjad	20	M	Enterprenure	SBS	Active	79.83	479	B+	68.70	599	B+
7	Bareera Tahir	21	F	Enterprenure	GGHS	Active	77.33	464	B+	64.90	564	B
8	Abdul hanan	21	M	Enterprenure	SBS	dropout	0.00	0	F	0.00	0	F
9	Awais Farooq	23	M	Enterprenure	SBS	Active	80.50	483	B+	66.70	583	B
10	Muhammad Nadeem	21	M	Civil servant	GHS	Active	77.33	464	B+	64.50	464	B
11	Maryam Nazir	20	F	Civil servant	GGHS	Active	78.17	469	B+	66.10	579	B
12	Mujtaba Qureshi	23	M	Civil servant	SBS	Active	81.17	487	B+	70.10	610	B+
12	Fasiha Fatima	20	F	Civil servant	GGHS	Active	80.83	485	B+	68.80	622	B+
13	Muhammad Usman Iqbal	21	M	Farmer	GHS	Active	79.33	476	B+	69.60	626	B+
15	Tooba Javeed	22	F	shopkeeper	GGHS	Active	73.50	441	B	62.30	549	B
15	Muhammad Faisal Raza	22	M	Businessman	GHS	Active	78.83	473	B+	65.30	570	B
16	Anees Mir	23	M	Businessman	GHS	dropout	0.00	0	F	0.00	0	F
17	Abbu Talha	22	M	Businessman	SBS	Active	80.17	481	B+	67.10	586	B+
18	Muhammad Luqman younas	23	M	Businessman	GHS	Active	77.33	464	B	63.20	551	B

Figure4: Example of students' data from preprocessing result

4.3 Implementation of Decision Tree C4.5 using Rapid Miner

Here are the steps to get the Decision tree model using RapidMiner9.9.

1. add the "read excel" operator to the work page as shown in Figure 5. There are 2 "read excel" operators used to store data training and data testing.



Retrieve student recort_Ts(123) (Retrieve)

repository entry: ecord_Ts(123)

Show advanced parameters

Retrieve
RapidMiner Studio Core

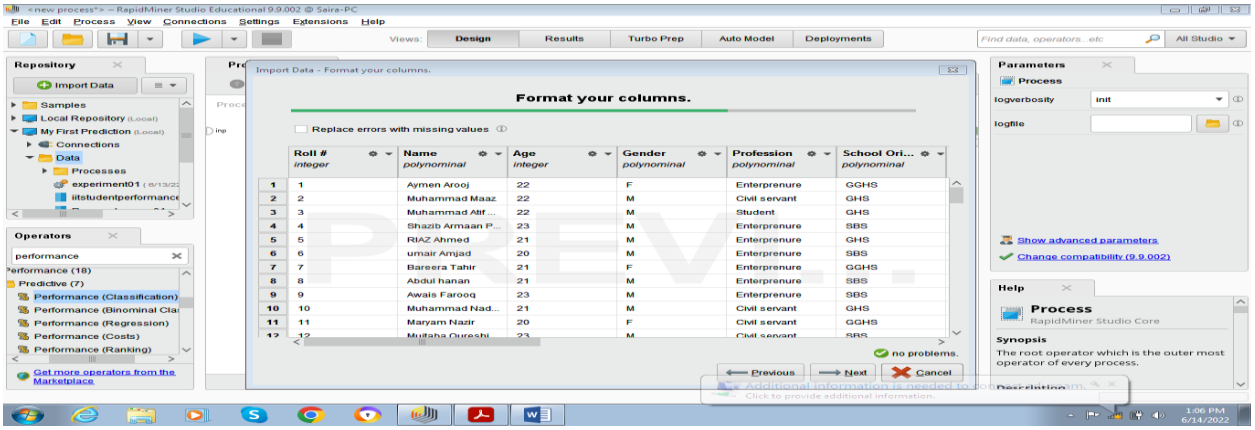
Tags: Load, Import, Read, Datasets, Examples, Example Set, Table, Repository, Data Access

Synopsis

This Operator can access stored

Figure 5: Input training and set role

2. Figure 6 is the process of inputting data by removing the check list on the attribute of student's name (number) because students name is not used. The attributes used are age, obtained marks1,origin of school (as attributes),Grade1 as(attributes),parents'occupation (as attributes),Grade2,Obtained marks 2,CGPA(as attributes) and students'status (as label).


File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Repository
Import Data
Operators
Parameters
Process
Help

Import Data - Format your columns.

Format your columns.

□ Replace errors with missing values

	Roll # *integer*	Name *polynominal*	Age *integer*	Gender *polynominal*	Profession *polynominal*	School Ori... *polynominal*
1	1	Aymen Arooj	22	F	Enterprenure	GGHS
2	2	Muhammad Maaz	22	M	Civil servant	GHS
3	3	Muhammad Atif ...	22	M	Student	GHS
4	4	Shazib Armaan P...	23	M	Enterprenure	SBS
5	5	RIAZ Ahmed	21	M	Enterprenure	GHS
6	6	umair Amjad	20	M	Enterprenure	SBS
7	7	Bareera Tahir	21	F	Enterprenure	GGHS
8	8	Abdul hanan	21	M	Enterprenure	SBS
9	9	Awais Farooq	23	M	Enterprenure	SBS
10	10	Muhammad Nad...	21	M	Civil servant	GHS
11	11	Maryam Nazir	20	F	Civil servant	GGHS
12	12	Mujtaba Qureshi	23	M	Civil servant	SBS

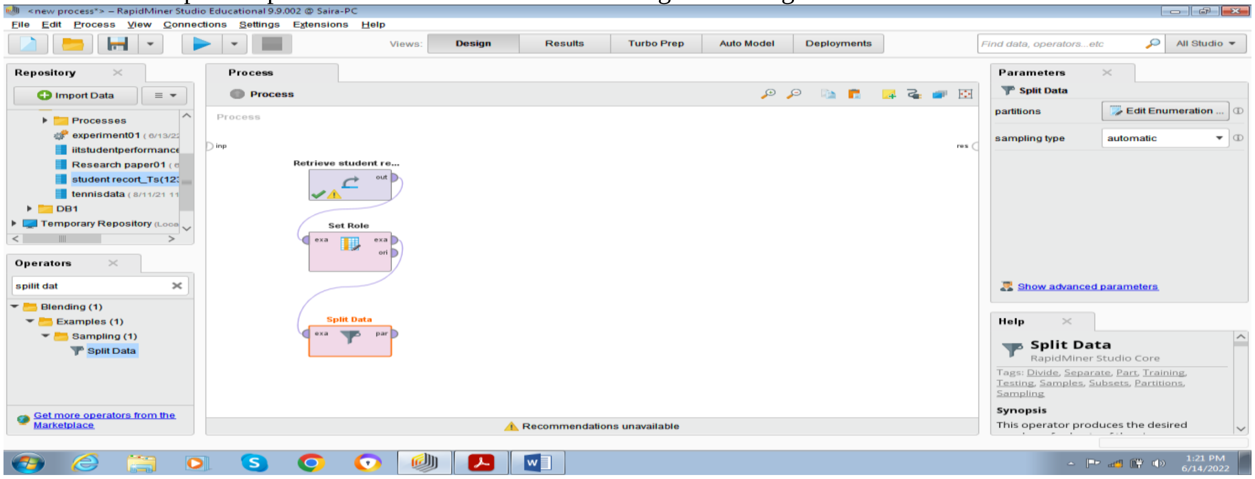
no problems.

Previous Next Cancel

logverbosity: init

Figure6: Selection of attributes

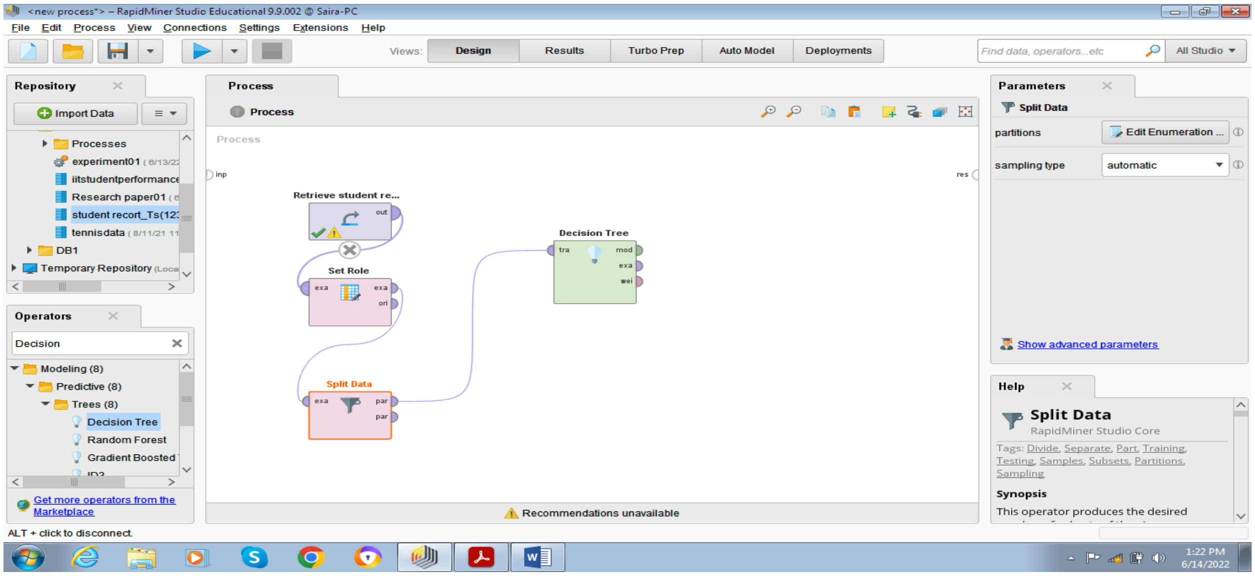
3. Add the data splitter operator to divide the data into training and testing.



Split Data — partitions: Edit Enumeration ...; sampling type: automatic

Figure7: Adding the data splitter

4. Add the “Decision tree “operator to the work page and set the required parameters used.In parameter criterion select information gain and check list no pre-pruning and no pruning.



Split Data — partitions: Edit Enumeration ...; sampling type: automatic

Figure 8: Adding the Decision Tree Operator

5. Add the apply model operator.

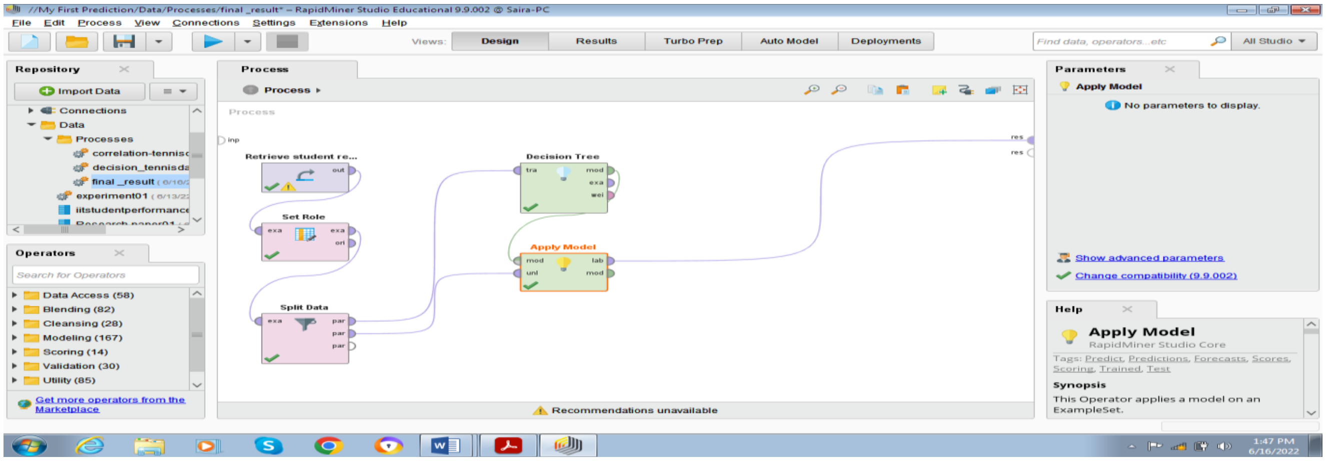

//My First Prediction/Data/Processes/final _result* – RapidMiner Studio Educational 9.9.002 @ Saira-PC
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Repository
Import Data
Connections
Data
Processes
correlation-tennis
decision_tennisda
final _result
experiment01
iitstudentperformance
Operators
Search for Operators
Data Access (58)
Blending (82)
Cleansing (28)
Modeling (167)
Scoring (14)
Validation (30)
Utility (85)
Get more operators from the Marketplace
Process
Retrieve student re...
Set Role
Split Data
Decision Tree
Apply Model
Recommendations unavailable
Parameters
Apply Model
No parameters to display.
Show advanced parameters
Change compatibility (9.9.002)
Help
Apply Model
RapidMiner Studio Core
Tags: Predict, Predictions, Forecasts, Scores, Scoring, Trained, Test
Synopsis
This Operator applies a model on an ExampleSet.
1:47 PM
6/16/2022

Figure 9: add the apply model operator

6. Add the performance operator to get the outcome of accuracy calculation and classification error.

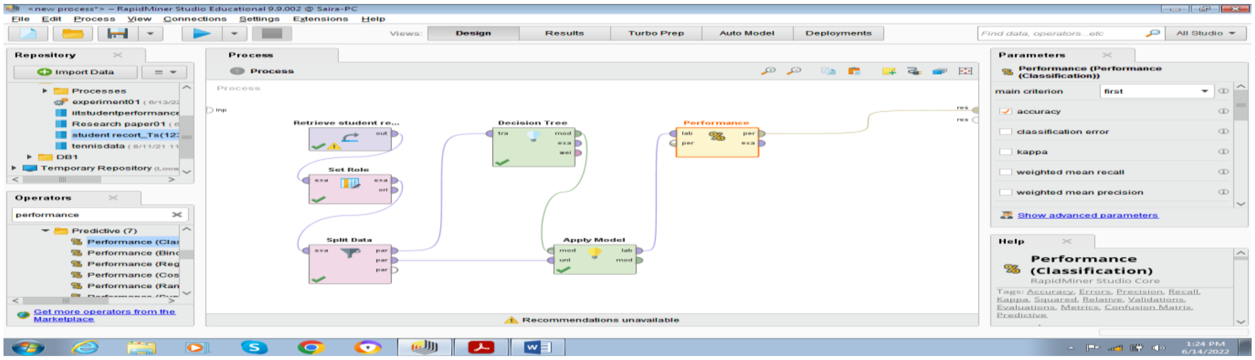


Figure 10: Decision tree series model

4.4 Decision Tree C4.5 Model

The decision tree process was completed by using the Data science software RapidMiner. To obtain the required knowledge on the most critical attributes, it could be focusable in the tree structure .Based on the running process of the training data, the most critical attribute was the origin of school .Studentswho experience dropouts came from high school with less marks.

4.5 Testing of the Decision Tree

To find out the accuracy of the decision tree model are developed, the tests were formed by using the Confusion Matrix as shown in Table 2.

Table 2. Confusion Matrix

	Pred.Active	Pred.Dropout	Class Precision
Pred. Active	229	1	99.57
Pred. Dropout	3	20	83.33
Class recall	98.71	95.24	

Based on Table 2, the measurement of accuracy was 98.04.The resulting accuracy was not very much up to the mark because the data used were not enough.

5.Conclusion

The data taken were 1275 students divided into 1020 data training and 255 for testing data.The attributes used consisted of origin of school, Age, parent's occupation,CGPA, Grad1,Grad2 and students' status. Parent's occupation and age much more contributed to students' dropouts and also origin of school is the most influential attribute in this regard. In the implementation phase Data Science software Rapid Miner is used to analyze the data of dropout students. The calculation of result of the obtained accuracy value was 98.04.

References

[1] A. Viloria, A. S. Naveda, H. H. Palma, and W. N. Núñez, "Using Big Data to Determine Potential Dropouts in Higher Education," 2020, doi: 10.1088/1742-6596/1432/1/012077.

[2] S. Wahyuni, U. Pembangunan, P. Budi, U. Pembangunan, and P. Budi, "THE IMPLEMENTATION OF DECISION TREE ALGORITHM C4 . 5 USING RAPIDMINER IN ANALYZING DROPOUT STUDENTS," no. November, 2017.

[3] M. V. Amazona and A. A. Hernandez, "Modelling student performance using data mining techniques: Inputs for academic program development," *ACM Int. Conf. Proceeding Ser.*, no. May, pp. 36–40, 2019, doi: 10.1145/3330530.3330544.

[4] N. Hutagaol and Suharjito, "Predictive modelling of student dropout using ensemble classifier method in higher education," *Adv. Sci. Technol. Eng. Syst.*, vol. 4, no. 4, pp. 206–211, 2019, doi: 10.25046/aj040425.

[5] M. Alban and D. Mauricio, "Predicting University Dropout trough Data Mining: A systematic Literature," *Indian J. Sci. Technol.*, vol. 12, no. 4, pp. 1–12, 2019, doi: 10.17485/ijst/2019/v12i4/139729.

[6] A. Abu, "Educational Data Mining & Students' Performance Prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 5, pp. 212–220, 2016, doi: 10.14569/ijacsa.2016.070531.

[7] H. Rahman, "Predict Student ' s Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques," no. December, pp. 27–28, 2017.

[8] A. Cohen, "Analysis of student activity in web-supported courses as a tool for predicting dropout," *Educ. Technol. Res. Dev.*, vol. 65, no. 5, pp. 1285–1304, 2017, doi: 10.1007/s11423-017-9524-3.

[9] S. Wahyuni, "Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree," *J. Phys. Conf. Ser.*, vol. 970, no. 1, 2018, doi: 10.1088/1742-6596/970/1/012030.

[10] S. Abdallah, "Predicting Student Retention Among a Homogeneous Population using Data Mining," no. June 2021, 2020, doi: 10.1007/978-3-030-31129-2.

[11] A. C. Study and P. S. Uni-, "The Investigation of Student Dropout Prediction Model in Thai Higher Education Using Educational Data Mining :," no. 1, pp. 356–368, 2019.

[12] I. J. Biomed, D. Min, T. Yang, A. Hee, and H. Ngu, "International Journal of Biomedical Data Mining Implementation of Decision Tree Using Hadoop Map Reduce," vol. 6, no. 1, pp. 1–4, 2017, doi: 10.4172/2090-4924.1000125.