Application of Data Mining Techniques in Students' Performance Prediction and Analysis

[1]Abubakarsidiq Makame Rajab, [2] Ramadhan Mzee Ramadhan

[1] Department of Electronic Information Communication
Huazhong University of Science and Technology, 1037Louyu Road,
Hongshan District, Wuhan 430074, P.R. China

[2] College of Public Administration,
Huazhong University of Science and Technology, 1037Louyu Road,
Hongshan District, Wuhan 430074, P.R. China
[1]Email: mrgovery@hotmail.com [2]Email: ramabwamzee@yahoo.com

Abstract *⸺This study aimed to analyze academic records specifically students' performance using data mining algorithms which can assist to explore more than one factors theoretically assumed to have an effect on students' performance in higher education, and finds a qualitative model which quality classifies the students' performance based totally on associated personal and social factors. The study used educational information mining to predict students' Overall GPA classification for the first year primarily based on their grades in First Semester and Second Semester results, non-public and social factors such as Living Apartment place and attendance. The study uses Data mining; Educational Data Mining (EDM), Students Performance, ID3 Algorithm, C4.5 algorithm, ID3 algorithm, CART and CHAID algorithm as tools to analyze academic records and performance of college students. The accrued data set of students' effects from the College of Health Science at the State University of Zanzibar. The study found that the discretization of the category attribute was no longer appropriate enough to capture the differences in other attributes, or, the attributes themselves was once not clear enough to capture such differences, in other words, the classes used in this study was not absolutely independent, for instance, an "Excellent" student can have the identical characteristics (attributes) as a "Very Good" student, and hence, this can confuse the classification algorithm and have large effects on its overall performance and accuracy.*
This paper can be used to assist instructors with managing their class, recognize their students' getting to know and reflect on their instructing and to assist learner reflection and provide proactive comments to learners.

Keywords: Data mining, Educational Data Mining (EDM), Students Performance, ID3 Algorithm, C4.5 algorithm, ID3 algorithm, CART and CHAID algorithm.

INTRODUCTION

Student's performance prediction in University courses is of the remarkable subject to the higher education where numerous elements might also affect the performance. We are dwelling in a data era, and we have more and more facts been generated in each and every issue you can suppose of. That's statistics being downloaded into a database, and most transaction you do, there is some type of data download. Universities are storing, processing and examining facts more than any time in records and this style will proceed to grow.

Nowadays the quantity of data saved in an academic database increasing rapidly. These databases comprise hidden information for the improvement of students' performance. Educational statistics mining is used to learn about the information accessible in the academic subject and carry out the hidden information from it. Classification methods like decision trees, Bayesian community so on can be utilized on the educational records for predicting the student's overall performance in the examination [1]. This prediction will assist to identify the susceptible students and help them to rate better marks. The ID3, C4.5, and CART decision tree algorithms are utilized on student's information to predict their performance in the last exam.

The consequence of the decision tree predicted the wide variety of students who are in all likelihood to bypass at a higher category (First Class) to the subsequent year. The results provide steps to improve the overall performance of college students who have been envisioned to fail. After the statement of the outcomes in the final examination, the marks received by means of the students are fed into the system and the effects had been analyzed for the subsequent session. The comparative analysis of the results states that the prediction has helped the weaker students to improve and introduced out betterment in the result.

Data Mining is the incorporation of mathematical methods that may consist of mathematical equations, algorithms, regular logistic regression, neural networks, segmentation, classification, clustering and so on. Those are all methods that make use of

mathematics. Data Mining is applicable across industry sectors [2]. Generally wherever we have processes, and wherever we have data, it is the application of these effective mathematical techniques that will extract tendencies patterns. All algorithms require some approach to search the data. The model can be both predictive and descriptive in nature. A predictive mannequin makes a prediction about values of statistics the use of regarded results determined from extraordinary data. Predictive modeling may also be made based on the use of other historical data. Predictive mannequin data mining tasks include classification, regression, time sequence analysis, and prediction. A descriptive model identifies patterns or relationships in data. It serves as a way to discover the homes of the facts examined. Clustering, summarization, association rules, and sequence discovery are typically descriptive in nature [3]. Data mining is the precept of examining the large database and choosing out relevant information. Figure below exhibit absolutely how data mining related to educational system;
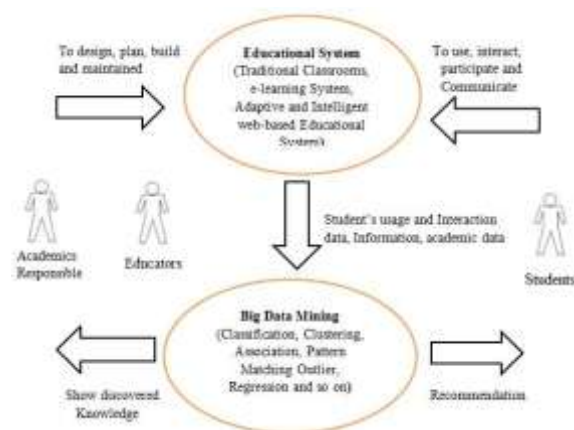

Figure 1: The cycle of applying data mining in educational systems

So with data mining techniques, the cycle is built in an educational machine which consists of forming hypotheses, checking out and training, for this reason, utility of data mining in academic systems, can be directed to support the specific wants of every one of the members in the academic process. The college students are required to recommend additional activities, instructing substances and tasks that would desire and enhance his/her mastering professors would have the remarks possibilities to classify college students into corporations primarily based on their needs for training and monitoring to find the most made mistakes locate the advantageous moves Administration, directors workforce will

obtain the parameters that will enhance gadget performance. The predominant motivation at the back of in this study are applying more than a few techniques and algorithms in order to discover and extract patterns of stored data and know-how discovery applications have obtained a wealthy focal point due to its importance in choice making and it has come to be a fundamental thing in quite a number of organizations. The study will supply a deeper perception to analyze the algorithms on actual data set and supply the prediction on the outcomes received an issue on the overall performance of scholar for futures.

The purpose of conducting this case study is to accumulate basic heritage data about the students' in order to set the following desires such as identification of different factors, which influences a student's two learning behavior and overall performance in the course of academic profession and producing of an information supply of predictive variables, Construction of a prediction model the use of classification data mining strategies on the basis of recognized predictive variables, validation of the developed mannequin for the college students analyzing in Universities, Colleges, and Schools, to find out about and discover the gaps in current prediction techniques and to study the existing prediction approaches for predicting student's performance.

RELATED WORK

The work of [1, 4] performed a research on a group of students enrolled in a particular direction program throughout a length of 4 years (2007-2010), with multiple overall performance indicators, together with "Previous Semester Marks", "Class Test Grades", "Seminar Performance", "Assignments", "General Proficiency", "Attendance", "Lab Work", and "End Semester Marks". They used ID3 selection tree algorithm to sooner or later construct a choice tree, and if-then rules which will subsequently help the instructors as nicely as the students to better understand and predict students' performance at the end of the semester. Furthermore, they defined their objective of this study as: "This study will additionally work to identify those students which wanted distinct interest to limit fail ration and taking the fabulous motion for the next semester examination" [5]. [6, 7] selected ID3 decision tree as their information mining technique to analyze the students' overall performance in the selected route program; due to the fact it is a "simple" decision tree getting to know the algorithm.

[8] carried out a study on the scholar overall

performance by using selecting a pattern of 300 college students (225 males, 75 females) from a team of colleges affiliated to Punjab University of Pakistan. The hypothesis that was noted as "Student's mindset in the direction of attendance in class, hours spent in find out about on daily basis after college, students' household income, students' mother's age, and mother's education are drastically associated with pupil performance" was framed. By means of easy linear regression analysis, it used to be observed that the elements like mother's training and student's household profits had been distinctly correlated with the scholar educational performance.

[9], in his study on public tutoring and its implications, observed that the proportion of college students receiving public tutoring in Tanzania was once incredibly higher than in Kenya, Uganda, Rwanda, and Burundi. It was once additionally located that there used to be an enhancement of educational overall performance with the intensity of personal tutoring and this version of the depth of personal tutoring depends on the collective thing specifically socio-economic prerequisites [10].

[4] performed similar research that commonly focuses on producing classification regulations and predicting students' performance in a chosen route program based totally on before recorded students' behavior and activities. [4, 11] processed and analyzed previously enrolled students' data in a unique course program throughout 6 years, with a couple of attributes collected from the college database. As a result, this study used to be in a position to predict, to a sure extent, the students' closing grades in the selected course program, as properly as, "help the student's to improve the student's performance, to pick out those students which needed special interest to minimize failing ration and taking gorgeous motion at the right time".

[12] performed an overall performance study on 400 college students comprising 200 boys and 200 female selected from the senior secondary faculty of Aligarh Muslim University, Aligarh, India with a predominant objective to establish the prognostic cost of one of a kind measures of cognition, persona and demographic variables for success at higher secondary stage in science stream. The selection was based totally on cluster sampling approach in which the whole populace of hobby was once divided into groups or clusters, and a random pattern of these clusters was selected for further analyses. It used to be discovered that female with excessive socio-economic reputation had fantastically higher tutorial fulfillment in science circulation and boys with low socio-economic popularity had quite higher academic achievement in general.

[13] applied C4.5 selection tree algorithm to the internal marks of the MCA college students and predict their overall performance in terms of a pass by or fail in the last exam. They evaluate the expected consequences and real outcomes which indicates, that there was a large delight in consequences as the prediction helped a lot to identify vulnerable and excellent college students and help them to rate better marks. They also compared the model with ID three selection tree algorithms and prove that the developed mannequin is higher in phrases of efficiency and time taken to build the selection tree.

[14] investigate a robust correlation between the mantel circumstance and the ultimate performance of the college students. They develop a rule mannequin based totally on decision tree and put into effect these regulations through the SSVM algorithm to predict the final grades of students. They additionally grouped the students on the foundation of their similar traits the use of K-means clustering.

[15] investigate the use of the emerging discipline of Educational Data Mining as a preventative measure alternatively than reiterate factors that have an effect on success at the University of the Witwatersrand. The study used the students' first semester/midyear mark to predict success/failure at the end of the academic year. The research findings indicated that the midyear mark can be viewed as an aspect which correctly predicts the Computer Science I final year marks. After adding investigation with large pattern sizes, the device can be used virtually in the college of Computer Science to perceive students a chance of failing.

[7] in contrast the C4.5, ID3 and CART choice tree algorithms to predict the overall performance of the first year engineering college students. It used to be three types of predictions. Students had been labeled as omitting to fail and promoted. This model used to be suitable for figuring out the college students that are most likely to fail.

[16] proposed a hybrid method of clustering and classification to improve the pupil educational performance in their last examinations. Initially, college students were categorized into three class's high, medium and low requirements and then applied a selection tree algorithm to take excellent selections for the students.

[17] performed a study of two extraordinary crew of college students of undergraduate and postgraduate stage to predict the performance of the students and compared the efficacy of two classifiers particularly decision tree and Bayesian networks using WEKA tool. In this research, the performance of Decision tree was once 3-12% greater correct than Bayesian

networks. This was beneficial for identifying the weak students for in additional training and to selecting the desirable college students for the scholarship.

[18] performed a study to reveal the development of students' academic performance in the educational community of higher learning. A gadget for examining college students results based totally on cluster analysis and makes use of standard statistical algorithms to arrange their scores data according to the degree of their performance is described. The model was combined with the deterministic mannequin to analyze the student's effects of a personal Institution in Nigeria which is a precise benchmark to reveal the development of the tutorial performance of college students in greater Institution for the purpose of making a nice selection by using the tutorial planners.

Most of the previous research centered on the use of the classification for prediction based totally on enrollment data and performance of college students in a sure course. Along the same line of focus, in this research, we will be building selection tree classification fashions to predict students Overall GPA Classification based on their grades in their study plan.

DATA MINING TOOLS AND TECHNIQUES

Competitive advantage requires abilities. Abilities are built through knowledge. Knowledge comes from data. The method of extracting expertise from data is referred to as Data Mining. Data mining is the extraction of hidden predictive facts from massive databases, is strengthen technique to help corporations to spotlight the most essential data in their statistics warehouses. Data mining tools predict future trends and behaviors. Data mining tools can answer commercial enterprise questions that traditionally had been too overwhelming to resolve. Data mining methods can be applied rapidly on current software and hardware systems to enhance the value of existing data resources [19], and can be built-in with new merchandise and system as they are introduced online.

Data mining encompasses distinctive algorithms such as Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm; Nearest Neighbor approach and so on.It additionally involves data exploration and visualization to present effects in a convenient way to users [20]. We current right here some algorithms that we have used. A data element will be referred to as an individual. It is characterized by a set of variables. In our context, most of the time a person is a learner and variables can be exercises

tried by way of the learner, marks bought scores, errors made, time spent, and a wide variety of correctly finished exercises and so on. New variables can also be calculated and used in algorithms. Below are the embellishments of data mining tools:-

A.   Classification

Classification is the most typically utilized data mining technique, which employs a set of pre-classified examples to strengthen a model that can classify the population of information at large. Find a mannequin for type attribute as a feature of the values of other attributes, a supervised data mining approach that includes assigning a label to a set of unlabeled enter objects. This strategy regularly employs a decision tree or neural network-based classification algorithms [21]. The data classification process involves mastering and classification. Learning the caching data are analyzed with the aid of the classification algorithm. In classification, test data are used to estimate the accuracy of the classification rules. If the accuracy is suited the regulations can be applied to the new data tuples. [22] argued that the test set is used to decide the accuracy of the model. Usually, the given data set is divided into training and test sets, with a training set used to construct the model and test set used to validate it. The algorithm then encodes these parameters into a model referred to as a classifier.

B.   Clustering Analysis Tools

This is very effective tools for clustering products into groups that naturally fall collectively which are the groups are identified by using the program. By the use of clustering methods, we can further become aware of dense and sparse regions in object space and can find out the usual distribution pattern and correlations amongst data attributes. Given a set of data points, each having a set of attributes, and a similarity measure among them [23], is primarily based on clusters such that data points in one cluster are more comparable to one another while data points in separate clusters are less comparable to one another.

C.   Regression

Regression is a data mining approach used to predict vary of numeric values, given a specific dataset used to model the relationship between one or greater independent variables and dependent variables [24]. In data mining, independent variables are attributes already acknowledged and response variables are what we want to predict. Predict the value of a given continuous-valued variable based totally on the values of different variables, assuming a linear or

nonlinear model of dependency. For example, regression may be used to predict the rankings, one student, given others. This entails superior techniques, such as more than one regression predict a relationship between a couple of variables, for example, is there a correlation between incomes, education and the place one chooses to live. The addition of greater variables appreciably increases the complexity of the prediction.

D. Association Rules Discovery
Association is a data mining feature that discovers the probability of the co-occurrence of gadgets in a collection. The relationships between co-occurring gadgets are expressed as affiliation rules. Association rules are often used to analyze sales transactions [25]. Given a set of archives every of which includes some quantity of gadgets from a given collection, Produce dependency policies which will predict the incidence of an object primarily based on occurrences of other items. Association and correlation are common to find established item set findings among large data sets.

E. Decision Trees
A decision tree is a decision-assist tool that makes use of a tree-like model of decisions and their feasible consequences, such as outcomes, resource costs, and utility. It is one way to display an algorithm that solely incorporates conditional control statements [21]. A decision tree is tree-shaped structures that represent sets of decisions. Given a collection of records (training set), each record contains a set of attributes, one of the attributes is the class. It required finding a model for the class attribute as a function of the values of other attributes hence previously unseen records should be assigned a class as accurately as possible.

F. Artificial neural networks
This method primarily based on non-linear predictive fashions that learn through training and resemble organic neural networks in structure [24]. During the studying phase, the network learns by adjusting weights so as to be in a position to predict the right class labels of the enter tuples. Neural networks have the extremely good capability to derive which means from elaborate or imprecise facts and can be used to extract patterns and discover traits that are too complicated to be noticed by way of both humans and other computer techniques.

G. Genetic algorithms
Genetic algorithms are an optimization approach used to solve nonlinear or non-differentiable optimization problems [19]. They use principles from evolutionary biology to search for a world minimum - an optimization problem. Involves of optimization techniques that use a procedure such as genetic combination, mutation, and herbal determination in a design based on the principles of evolution.

H. Nearest neighbor methods
The nearest-neighbor rule states that a test occasion is labeled according to the classifications of" nearby" education examples from a database of known structures [19, 26]. A method that classifies each report in a dataset based totally on a mixture of the classes of the k record(s) most comparable to it in a historic dataset sometimes called the k-nearest neighbor technique.

I. Rule induction
Rule induction is a data mining method of deducing if-then regulations from a statistics set [26]. These symbolic decision guidelines give an explanation for an inherent relationship between the attributes and category labels in the data set. Involve of extraction of useful if-then policies from data based totally on statistical significance.

BUILDING THE MODEL

A. Data Mining Process
The objective of this study is to find out relations between students' personal and social factors, and their educational performance the use of data mining tasks. Henceforth, their performance could be envisioned in the upcoming semesters. A student's overall performance is decided through the inner assessment and end semester examination. The internal assessment is carried out with the aid of the teacher primarily based upon students' performance in educational activities such as class test, seminar, assignments, accepted proficiency, attendance and lab work. The cease semester examination is one that is scored by the student in the semester examination. Each pupil has to get minimum marks to pass by a semester in inside as properly as end semester examination. The student overall performance is measured and indicated by using the Grade Point Average (GPA), which is an actual variety out of 5.0. The sequences of steps recognized in extracting knowledge from data are proven in Figure 2.
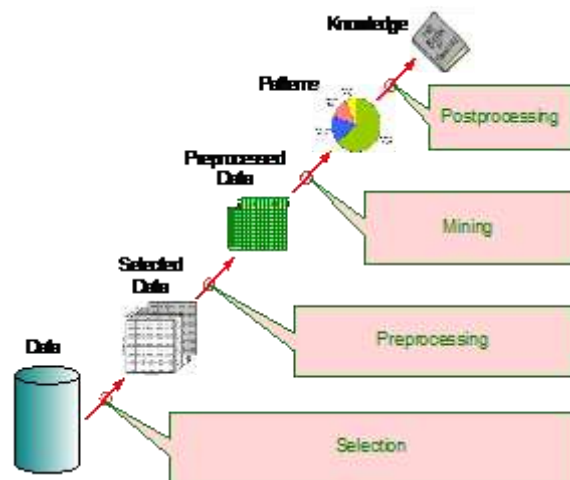
Figure 2: The sequences of steps identified in extracting knowledge from data

B. Dataset

The dataset used in this study was gathered through a case study dispensed of the State University of Zanzibar directed to College of Health Science; the information was once extracted from the database anonymously and without any bias. Unfortunately, we acquired the information very quickly from the State University of Zanzibar after a month. The dataset of 72 archives of college students outcomes used in this find out about was got from the College of Health Science Zanzibar was a first 12 months of outcomes from 2014/2015.

Table 1: Describes the attributes of the data and their possible variables.

Serial Number	Sex	Attendance	O/A Level Secondary School Completed	Living Apartments	GPA for Second Semester	Overall Classification for First Year Results
1	M	Good	Excellent	Yes	4	First Class
2	M	V.Good	V.Good	Yes	3.6	Upper Class
3	M	Good	V.Good	Yes	3.6	Upper Class
4	M	V.Good	Good	Yes	3.1	Lower Class
5	F	Good	Excellent	Yes	4	First Class
6	M	V.Good	Excellent	Yes	4	First Class
7	F	Good	Good	Yes	3.4	Lower Class
8	M	V.Good	V.Good	Yes	3.7	Upper Class
9	M	Good	Excellent	Yes	4.1	First Class
10	M	Poor	Good	Yes	3.4	Lower Class
11	M	Good	V.Good	No	3.6	Upper Class
12	M	V.Good	V.Good	No	3.6	Upper Class
13	F	V.Good	Excellent	No	4	First Class
14	F	Good	V.Good	No	3.6	Upper Class
15	M	Good	V.Good	No	3.6	Upper Class
16	M	Poor	Excellent	Yes	3.4	Lower Class
17	F	Poor	Good	Yes	3.2	Lower Class
18	F	V.Good	V.Good	No	4	First Class
19	F	V.Good	Excellent	No	3.4	Lower Class
20	F	V.Good	Good	No	3.3	Lower Class
21	M	V.Good	V.Good	No	3.2	Lower Class
22	F	V.Good	V.Good	Yes	3.4	Lower Class
23	M	V.Good	Excellent	Yes	3.2	Lower Class
24	F	V.Good	V.Good	Yes	4	First Class
25	M	Good	V.Good	Yes	3.7	Upper Class
26	M	Poor	Excellent	Yes	3.6	Upper Class
27	M	Good	Good	Yes	3.5	Upper Class
28	F	V.Good	Good	No	3.3	Lower Class

No	Gender		Rating	Y/N	GPA	Class
29	M	V.Good	Excellent	No	3.6	Upper Class
30	M	V.Good	V.Good	No	3.4	Lower Class
31	M	V.Good	V.Good	No	3.8	Upper Class
32	F	V.Good	V.Good	No	3.3	Lower Class
33	M	V.Good	Good	Yes	3.2	Lower Class
34	F	V.Good	V.Good	No	3.3	Lower Class
35	F	V.Good	Good	Yes	4	First Class
36	F	V.Good	V.Good	Yes	3.7	Upper Class
37	F	V.Good	Good	Yes	4.2	First Class
38	M	V.Good	Good	Yes	3.2	Lower Class
39	F	V.Good	Good	Yes	4.1	First Class
40	M	V.Good	Excellent	Yes	3.6	Upper Class
41	M	V.Good	V.Good	Yes	4	First Class
42	M	V.Good	Excellent	Yes	3.6	Upper Class
43	M	V.Good	Good	Yes	3.7	Upper Class
44	F	V.Good	Excellent	Yes	3.4	Lower Class
45	F	V.Good	V.Good	Yes	3.7	Upper Class
46	F	V.Good	Excellent	Yes	3.4	Lower Class
47	M	V.Good	V.Good	No	3.6	Upper Class
48	M	V.Good	Good	Yes	3.3	Lower Class
49	F	V.Good	Good	Yes	2.9	Pass
50	M	V.Good	V.Good	No	3.3	Lower Class
51	F	V.Good	V.Good	No	3.2	Lower Class
52	M	V.Good	V.Good	Yes	3.1	Lower Class
53	M	V.Good	Good	Yes	3.6	Upper Class
54	M	V.Good	V.Good	Yes	3.6	Upper Class
55	F	V.Good	Good	Yes	3	Lower Class
56	M	V.Good	V.Good	Yes	3.7	Upper Class
57	M	V.Good	Good	Yes	3.3	Lower Class
58	F	V.Good	Good	Yes	3.5	Upper Class
59	F	V.Good	Good	Yes	3.3	Lower Class
60	F	V.Good	Excellent	Yes	3.8	Upper Class
61	F	V.Good	V.Good	Yes	3.3	Lower Class
62	M	V.Good	Excellent	Yes	3	Lower Class
63	M	V.Good	Good	Yes	4	First Class
64	M	V.Good	Excellent	Yes	3	Lower Class
65	F	V.Good	V.Good	Yes	4.4	First Class
66	F	V.Good	Excellent	Yes	3.4	Lower Class
67	F	V.Good	V.Good	Yes	2.9	Pass
68	M	V.Good	V.Good	Yes	3.9	First Class
69	M	V.Good	V.Good	Yes	3.2	Lower Class
70	M	V.Good	V.Good	Yes	3.6	First Class
71	M	V.Good	V.Good	Yes	3.1	Upper Class
72	F	V.Good	Good	Yes	3.2	Upper

						Class

C.   Data selection

Table 2: Describes the attributes of the data and their possible values.

Attributes	Possible Outcomes	Explanations /Meaning Full
Gender (G)	{Female, Male}	Student's Gender
O/A Level Secondary School Completed (O/A)	{Excellent (75%-100%), Very Good (70% - 74.9%), Average Good (65%-69.9%), Good (55% - 64.9%), Poor (0%-54.9%)	Grading Assigned before Join to Colleges.
Living location	{Yes, No.}	Apartments
Attendance (ATT)	{ Poor, Good, Very Good}	Continuous Assessment
GPA for Second Semester.(GPA)	{ >=4.0 (Excellent), 3.5-3.9 (Very Good), 3.0-3.4 (Good), 2.0- 2.9 (Pass) }	End Semester GPA
Overall GPA Classification for First Year (OGPAFY)	{ >=4.0 (First Class), 3.5-3.9 (Upper Class), 3.0-3.4 (Lower Class), 2.0- 2.9 (Pass) }	Average GPA for 1 and 2 Semester.

D.   Data Preprocessing

In this step, only those fields were selected which were required for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database.

Table 3: Show the ranges of the data in the dataset according to their attributes

Attributes	The range of Possible Outcomes
Gender (G)	Female {31 }, Male {41}
O/A Level Secondary School Completed (O/A)	Excellent {18 }, Very Good { 32}, Good { 22 }, Pass {0 }
Living location (LC)	Yes {54} , No.{18 }
Attendance (ATT)	Poor { 4 }, Good {10}, Very Good { 54},
Scholarship (S)	Yes {0}, No {72}
Practical Experiment	Poor {0}, Average{0}, Good{72}
GPA for Second Semester	Excellent {13 }, Very Good {25}, Good {32 }, Pass { 2}
Overall GPA Classification for First Year (OGPAFY)	{ >=4.0 (First Class), 3.5-3.9 (Upper Class), 3.0-3.4 (Lower Class), 2.0- 2.9 (Pass) }

DATA MINING IMPLEMENTATION AND RESULTS

After the data preparation, the data selection and transformation process were performed. The organized data was then put through the data mining process. In this case study, four decision tree algorithms were used on this report, namely, C4.5 decision tree, ID3 selection tree, CART decision Tree and CHAID.

A.   Decision Tree Induction

A decision tree is a supervised classification approach that builds a top-down tree-like model from a given dataset attributes. The decision tree is a predictive modeling approach used for predicting, classifying, or categorizing a given data objects based on the earlier generated model using a training dataset with the identical features (attributes). The shape of the generated tree consists of a root node, inner nodes, and leaf (terminal) nodes [27]. The root node is the first node in the decision tree which has no incoming edges, and one or extra outgoing edges; an internal node is a middle node in the decision tree which has one incoming edge, and one or extra outgoing edges; the leaf node is the final node in the choice tree structure which represents the final counseled (predicted) type (label) of a facts object.

A decision tree is typically used for gaining data for the motive of decision -making. Decision tree begins with a root node on which it is for users to take actions. From this node, users split each node recursively according to the decision tree getting to know the algorithm. The closing result is a decision tree in which each department represents a viable scenario of choice and its outcome.

B.   ID3 Algorithm

The core algorithm for building decision trees is known as ID3 by way of J. R. Quinlan which employs a top-down, grasping search through the house of viable branches with no backtracking. ID3 makes use of Entropy and Information Gain to assemble a decision tree.

A decision tree is built top-down from a root node and involves partitioning the data into subsets that include cases with comparable values (homogenous). ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is totally homogeneous the entropy is zero and if the pattern is an equally divided it has entropy of one.

To offers with the information gain for every attributes, A relative to target entropy, D is wanted to classify a tuple in D in which the attribute has the best possible information gain, let first to calculate the goal entropy of D.

Table 4: Dataset of target attributes

Overall GPA Classification for First Year Results	First Class	Upper Class	Lower Class	Pass	Total
Corresponding Values	15	25	30	2	72

$$\text{Entropy (D)} = -\sum_{i=1}^{n} P_i \log_2 p_i$$

$$E(D) = - (P_{First})\log_2(P_{First}) - (P_{Upper})\log_2(P_{Upper})$$
$$- (P_{Lower})\log_2(P_{Lower})$$
$$- (P_{Pass})\log_2(P_{Pass})$$
$$Whereby; P_{First} = \frac{15}{72} =$$

$0.20833, P_{Upper} = \frac{25}{72} = 0.34722$

$P_{Lower} = \dfrac{30}{72} = 0.41667 \text{ and } P_{Pass} = \dfrac{2}{72} = 0.02778$

Therefore, the entropy for Overall GPA Classification for First Year Results is given as follows:-

$$\text{Entropy (D)} = - \left(\frac{15}{72}\right)\log_2\left(\frac{15}{72}\right) - \left(\frac{25}{72}\right)\log_2\left(\frac{25}{72}\right)$$
$$- \left(\frac{30}{72}\right)\log_2\left(\frac{30}{72}\right)$$
$$- \left(\frac{2}{72}\right)\log_2\left(\frac{2}{72}\right)$$

$$\text{Entropy(D)} = 0.47147 + 0.52989 + 0.52626 + 0.1436$$
$$\text{Entropy(D)} = 1.67123$$

C. Splitting Criteria based on Information Gain

Constructing a decision tree is all about finding an attribute that returns the very best information gain (the most homogeneous branches). The information gain is based totally on the limit in entropy after a dataset is break up on an attribute Let to calculate the nice attribute for a specific node in the tree we use the measure known as Information Gain. The information gain, Gain (D, A) of an attribute A is described as;

$$GAIN_{Split} = \text{Entropy} - \sum_{i=1}^{k} \frac{D_i}{D} Entropy(D_i)$$

Here we calculate the information gain for every Class Attribute, A relative to Entropy, D where by way of the Parent Node, p is cut up into k partitions; n~i quantity of records in partition i, the purpose of calculating the information gain is simply to measures reduction in Entropy achieved due to the fact of the break up and we choose the split that achieves most reduction (maximizes GAIN), The information gain, Gain (D, A) of an attribute A, can be calculate the use of the following formula;

$$GAIN_{Split} = \text{Entropy (D)} - \sum_{i=1}^{k} \frac{D_i}{D} Entropy(D_i)$$

Again, the dataset is then split into distinctive attributes. The entropy for every branch is calculated. Then it is brought proportionally, to get complete entropy for the split. The resulting entropy is subtracted from the entropy earlier than the split. The result is the Information Gain or reduces in entropy.

Splitting Criteria primarily based on Information Gain is wanted after the use of A to cut up D into k partitions to classify D, This entropy primarily based computations are comparable to the GINI index computations, so the components used are given below;

$$\text{INFO (D, A)} = - \sum_{i=1}^{k} \frac{D_i}{D} Entropy(D_i)$$

The **S**plitting Criteria of Gender, Info $_{Gender}$(G) for the male and girls for each attribute of D;

$$Info_{Gender}(G) = - \sum_{i=1}^{k} \frac{D_i}{D} Entropy(D_i)$$

$$Info_{Gender}(G) = \frac{D_{First}}{|D|}E + \frac{D_{Upper}}{|D|}E(D_{Upper})$$
$$+ \frac{D_{Lower}}{|D|}E(D_{Lower})$$
$$+ \frac{D_{Pass}}{|D|}E(D_{Pass})$$

$$Info_{Gender}(G) = - \frac{15}{72}\left(\left(\frac{8}{15}\right)\log_2\left(\frac{8}{15}\right) + \left(\frac{7}{15}\right)\log_2\left(\frac{7}{15}\right)\right)$$
$$- \frac{30}{72}\left(\left(\frac{15}{30}\right)\log_2\left(\frac{15}{30}\right) + \left(\frac{15}{30}\right)\log_2\left(\frac{15}{30}\right)\right)$$
$$- \frac{2}{72}\left(\left(\frac{2}{2}\right)\log_2\left(\frac{2}{2}\right) + \left(\frac{0}{2}\right)\log_2\left(\frac{0}{2}\right)\right)$$

$$Info_{Gender}(G) = 0.90039$$

Similarly, techniques we applied to the remaining attributes for calculating the splitting criteria primarily based on Information after using A to cut up D into k partitions as shown in the tables below:-

Table 5: The entropy of Gender using the frequency table of two attributes

Gender	Overall GPA Classification for First Year Results				
	First Class	Upper Class	Lower Class	Pass	Total
Female	8	6	15	2	31
Male	7	19	15	0	41
Splitting Information on Gender G (D, G)= 0.90039					

Table 6: The entropy of Attendance using the frequency table of two attributes

Attendance	Overall GPA Classification for First Year Results				
	First Class	Upper Class	Lower Class	Pass	Total
Poor	0	1	3	0	4
Good	3	6	1	0	10
Very Good	12	18	26	2	58
Splitting Information of Attendance = 0.78606					

Table 7: The entropy of O/A Level Secondary School Completed using the frequency table of two attributes

O/A Level Secondary School Completed	Overall GPA Classification for First Year Results				
	First Class	Upper Class	Lower Class	Pass	Total
Excellent	5	5	8	0	18
V. Good	6	15	10	1	32
Good	4	5	12	1	22

Splitting Info of O/A Level Secondary (D, O/A)= 1.48231

Table 8: The entropy of Living location using the frequency table of two attributes

Living location	Overall GPA Classification for First Year Results				
	First Class	Upper Class	Lower Class	Pass	Total
Yes	13	18	21	2	54
No	2	7	9	0	18
Splitting Information of Living location = 0.78225					

Table 9: The entropy of GPA for Second Semester using the frequency table of two attributes

GPA Second Semester	Overall GPA Classification for First Year Results				
	First Class	Upper Class	Lower Class	Pass	Total
Excellent (>=4.0)	13	0	0	0	13
V. Good (3.5-3.9)	2	23	0	0	25
Good (3.0-3.4)	0	2	30	0	32
Pass (2.0- 2.9)	0	0	0	2	2
Info Gain of GPA of Second Semester = 0.25767					

D. Information Gain

The information gain is primarily based on a reduction in entropy after a dataset is broken upon an attribute. build up a decision tree is all about finding an attribute that returns the absolute best information gain (i.e., the most homogeneous branches).therefore the information gain is given as;

$$INGAIN_{Split} = Entropy(D) - \sum_{i=1}^{k} \frac{D_i}{D} Entropy(D_i)$$

$$INGAIN_{Gender} = Entropy(D) - Info_{Gender}(G)$$
$$INGAIN_{Gender} = 1.67123 - 0.90039$$
$$GAIN_{Gender} = 0.77084$$

Similarly ways the table 5 exhibit certainly the Information Gains for the Gender, Attendance, O/A Level Secondary, Living region and School Completed and GPA of Second Semester respectively;

Table 10: Information Needed After Using A to Split D into k partitions.

Information Gains (D, A) of an attribute A	Corresponding Values
Info Gain of Gender G (D, G)	0.77084
Info Gain of Attendance (D, ATT)	0.88517
Info Gain of O/A Level Secondary (D, O/A)	0.18892
Info Gain of Living location (D, LC)	0.88898

Info Gain of GPA of Second Semester (D, GPA)	1.41356

The selecting attribute with the greatest facts gain is used as the selected node, where by means of the GPA of Second Semester (D, GPA) has the perfect gain, therefore it is used as the root node as shown in figure 3.

Figure 3: GPA of Second Semester as the root node

The ID3 algorithm is run recursively on the non-leaf branches until all data is categorized in such a way that a department with the entropy of 0 is a leaf node and with entropy greater than 0 desires similarly splitting. The ID3 algorithm is run recursively on the non-leaf branches, till all records are classified. The ID3 algorithm was capable to predict that the overall performance of students of Overall GPA classification for the first year is exactly affected by way of the end result of Second Semester.

E. C4.5 Algorithm

The C4.5 decision tree algorithm is an algorithm developed by Ross Quinlan, which was once the successor of the ID3 algorithm. The C4.5 algorithm makes use of pruning in the technology of a decision tree, the place a node may want to be eliminated from the tree if it adds little to no value to the remaining predictive model. Information gain measure is biased closer to attributes with a massive variety of values. C4.5 uses attain ratio to overcome the problem (normalization to data gain). The attribute with the most acquire ratio is chosen as the splitting attribute.

F. Gain Ratio

In order to overcome the drawback of Information Gain, let adjusts Information Gain by way of the calculating the Gain Ratio.

$$GainRatio_{split}(D,A) = \frac{GAIN_{Split}}{SplitINFO(D,A)}$$

Table 11: shows the Gain Ratio after using A to split D into k partitions;

Gain Ratio (D, A) of an attribute A	Corresponding Values
Gain Ratio of Gender G (D, G)	0.85612
Gain Ratio of Attendance (D, ATT)	1.12608

Gain Ratio of O/A Level Secondary (D, O/A)	0.12745
Gain Ratio of Living location (D, LC)	1.13644
Gain Ratio of GPA of Second Semester	5.48593

Therefore the maximum gain ratio attribute is a GPA of Second Semester (D, GPA) is 5.48593 therefore it is selected as the splitting attribute.

G. CART Algorithm

CART algorithm can be used for constructing both Classification and Regression Decision Trees. The impurity (or purity) measure used in building decision tree in CART is Gini Index. The decision tree built by means of CART algorithm is always a binary choice tree (each node will have only two child nodes). If a data set D includes examples from n classes, Gini index, Gini (D) is defined as;

$$\text{Gini Index (D)} = 1 - \sum_{i=1}^{k} P_i^2$$

Where by, P_i is the relative frequency of type i in D, Now let first to calculate the Gini index Gini (D) of the target attribute then accompanied the corresponding attributes in order to analysis nicely the CART algorithm.

$$\text{Gini Index (D)} = 1 - \sum_{i=1}^{k} P_i^2$$
$$D = 1 - \left(\left(\frac{15}{72}\right)^2 + \left(\frac{25}{72}\right)^2 + \left(\frac{30}{72}\right)^2 + \left(\frac{2}{72}\right)^2\right)$$
$$\text{Gini Index (D)} = 0.6616$$

Similarly, let us find the Gini Indexes for all attributes in a way that a data set D is split on A into two or extra subsets D_1, D_2, D_3 and D_4 the Gini index Gini (D) is described as;

$$\text{Gini Index}_A(D) = \frac{D_1}{|D|} Gini(D_1) + \frac{D_2}{|D|} Gini(D_2)$$

To determine the great attribute for a specific node in the tree we use the measure Gini Index.

The Gini Index $_{Gender}$(D) will be;

$$\text{Gi}_{Gender}(D) = \frac{D_{First}}{|D|} Gini(D_{First})$$
$$+ \frac{D_{Upper}}{|D|} Gini(D_{Upper})$$
$$+ \frac{D_{Lower}}{|D|} Gini(D_{Lower})$$
$$+ \frac{D_{Pass}}{|D|} Gini(D_{Pass})$$

$$\text{Gini}_{Gender}(D) = \frac{15}{72}\left(1 - \left(\left(\frac{8}{15}\right)^2 + \left(\frac{7}{15}\right)^2\right)\right)$$
$$+ \frac{25}{72}\left(1 - \left(\left(\frac{6}{25}\right)^2 + \left(\frac{19}{25}\right)^2\right)\right)$$
$$+ \frac{30}{72}\left(1 - \left(\left(\frac{15}{30}\right)^2 + \left(\frac{15}{30}\right)^2\right)\right)$$
$$+ \frac{2}{72}\left(1 - \left(\frac{2}{2}\right)^2\right)$$

The Gini Index $_{Gender}$(D) = 0.4387

H. Reduction in Impurity

The attribute offers the smallest $Gini_{split}$ (D) (or the largest discount in impurity) is chosen to break up the node. The reduction in impurity is given as:

$$\Delta\text{Gini (A)} = \text{Gini Index (D)} - \text{Gini Index}_A(D)$$

Therefore, Reduction in Impurity of Gender will be;

$$\Delta\text{Gini Index (Gender)} = \text{Gini (D)} - \text{Gini}_{Gender}(D)$$
$$\Delta\text{Gini Index (Gender)} = 0.6616 - 0.4387$$
$$\Delta\text{Gini Index (Gender)} = 0.2229$$

Table 12: Shows the Gini Index and reduction in impurity after using A to split D into k partitions;

Attributes A	$Gini\ Index$ (D, A)	Impurity Reduction
$Gini\ Index$ of Gender (D, G)	0.4387	0.2229
$Gini\ Index$ of Attendance (D, ATT)	0.41512	0.24648
$Gini\ Index$ of O/A Level Secondary (D, O/A)	0.61945	0.2984
$Gini\ Index$ of Living location (D, LC)	0.2126	0.2984
$Gini\ Index$ of GPA of Second Semester (D, GPA)	0.3443	0.5623

As a closing analysis, it was of course observed that some algorithms worked better with the dataset than others, in detail, CART had the excellent accuracy which used to be appreciably greater than the predicted (default model) accuracy. Therefore the most Gini Index (D, A) and Impurity Reduction of all attributes is a GPA of Second Semester (D, GPA) which have Gini Index (D, A) of 0.3443 two and Impurity Reduction of 0.5623 therefore it is selected as the splitting attribute. Hence nevertheless is cozy to be a break up the node. This really suggests the overall performance of college students of Overall GPA classification for first year is precisely affected through the result of Second Semester.

EXPERIMENT AND ANALYSIS

In this case study, more than one classification methods was once used in the data mining technique for predicting the students' result at the subsequent year. This approach was once used because it can furnish a broader seem to be and perception of the final results and output, as properly as, it will lead to a comparative conclusion over the results of the study. Furthermore, a 10-fold cross validation was used to confirm and validate the results of the used algorithms and provide accuracy and precision measures. All data mining implementation and processing in this study was executed the use of manuals manipulation and WEKA software. In this study, four decision tree algorithms was used on the simulation, C4.5 decision tree, ID3 choice tree, CART decision Tree, and CHAID.

A. CHAID Decision Tree

Chi-squared Automatic Interaction Detection (CHAID) is every other decision tree algorithm which uses chi-squared based splitting criterion instead of the traditional splitting criterions used in other decision tree algorithms. Following are the settings used with the CART operator to produce the decision tree. Minimal dimension of cut up = 4, Minimal leaf measurement = 2, Minimal acquire = 0.1, Maximal depth = 20 Confidence = 0.5 .After walking the CHAID algorithm with the 10-fold move validation on the dataset, the following confusion matrix was once generated:-

Table 13: Confusion Matrix

		Predicted Class			
Actual Class	First Class	13	2	0	0
	Upper Class	0	23	2	0
	Lower Class	0	0	30	0
	Pass	0	0	0	2
	First Class	13	2	0	0

The CHAID algorithm was able to predict the class of 68 students out of 72, which provides it an Accuracy value of 94.44%.

Similarly, as a last analysis, it was once obviously noticed that some algorithms worked better with the dataset than others, in detail, CART had the exceptional accuracy of 100%, which was considerably more than the predicted accuracy, CHAID and C4.5 was subsequent with 94.44% and 95.56% respectively, and the least correct used to be ID3 with 93.70%. On the different hand, it was once substantive that the classification remembers was continually greater than the expectations, which some might argue with. Furthermore, it have been seen that most of the algorithms have struggled in distinguishing comparable classes objects, and as a result, more than one objects was noticed being classified to their nearest similar class.

B. Decision Tree to Decision Rules

This procedure goes on until all information labeled flawlessly or run out of attributes. The understanding represented by decision tree can be extracted and represented in the shape of IF-THEN rules. Two so let to builds the classification Rules primarily based on C4.5 decision tree.
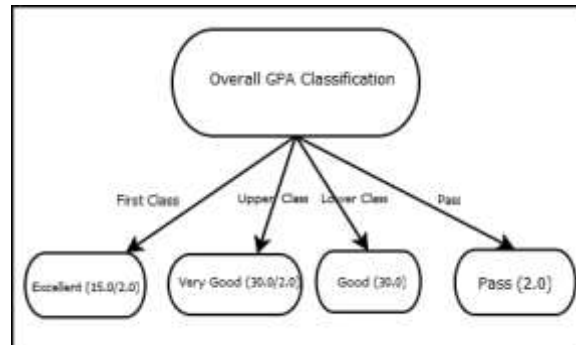

Figure 4: Decision Tree Rule Model

IF GPA of First Semester Result = "Excellent" AND GPA of Second Sec Semester Result = "Excellent" AND Attendance ='V. Good' or "Good" AND O/A Level Secondary = "Excellent" AND Living Location ="Yes" or "No" Then the Overall GPA = "First Class"

Else IF GPA of First Semester Result = "V. Good" AND GPA of Second Sec Semester Result = "Excellent" AND Attendance = "Good" or "V.Good" AND O/A Level Secondary = "V.Good" AND Living Location ="No" Then the Overall GPA = "First Class"

Else IF GPA of First Semester Result = "V. Good" AND GPA of Second Sec Semester Result = "V. Good" AND Attendance = "Good" or "V.Good" AND O/A Level Secondary = "Excellent" AND Living Location = "Yes" or "No" Then the Overall GPA = "Upper Class"

Else IF GPA of First Semester Result = "V. Good" AND GPA of Second Sec Semester Result = "Good" AND Attendance = "Good" or "V.Good" AND O/A Level Secondary = "V.Good" AND Living Location = "Yes" or "No" Then the Overall GPA = "Upper Class"

Else IF GPA of First Semester Result = "Good" AND GPA of Second Sec Semester Result = "Good" AND Attendance = "Good" or "V.Good" AND O/A Level Secondary = "Excellent" or "V.Good" AND Living Location = "Yes" or "No" Then the Overall GPA = "Lower Class"

Else IF GPA of First Semester Result = "Pass" AND GPA of Second Sec Semester Result = "Pass" AND Attendance = "Good" or "V.Good" AND O/A Level Secondary = "Excellent" or "V.Good" AND Living Location = "Yes" or "No" Then the Overall GPA = "Pass"
Else the Overall GPA = "Fail"

CONCLUSION

In this section, a couple of decision tree methods and algorithms were reviewed, and their performances and accuracies were tested. As a remaining analysis, it was for sure seen that some algorithms labored better with the dataset than others. The end result of testing of the proposed algorithm proved that the classification accuracy in the generated decision tree is improved. Also by way of the use of the proposed algorithm depth of the tree is reduced. Nevertheless, the proposed algorithm has some weaknesses such as extra time for computation and it is not still able to break out from being trapped into nearby optimum.

Decision trees are so famous because they produce classification rules that are handy to interpret than different classification methods. From the classifiers accuracy it is clear that the authentic high-quality rate of the model for the overall GPA classification for first 12 months class is affected by both semesters results especially the Second Semester for ID3, C4.5 and CART and CHAID decision trees that means model is correctly figuring out the college students who are probable to get lower class and Pass. These students can be considered for suitable counseling so as to enhance their result. This commentary leads to conclude that the discretization of the category attribute was once no longer appropriate enough to capture the differences in other attributes, or, the attributes themselves was once not clear enough to capture such differences, in other words, the classes used in this study was no longer absolutely independent, for instance, an "Excellent" student can have the identical characteristics (attributes) as a "Very Good" student, and hence, this can confuse the classification algorithm and have large effects on its overall performance and accuracy.

REFERENCE

[1] K. Bunkar, U. K. Singh, B. Pandya, and R. Bunkar, "Data mining: Prediction for performance improvement of graduate students using classification," in *2012 Ninth International Conference on Wireless and Optical Communications Networks (WOCN)*, 2012, pp. 1-5.

[2] Q. A. Al-Radaideh, A. Al Ananbeh, and E. Al-Shawakfa, "A classification model for predicting the suitable study track for school students," *Int. J. Res. Rev. Appl. Sci,* vol. 8, pp. 247-252, 2011.

[3] M. Pandey and V. K. Sharma, "A decision tree algorithm pertaining to the student performance analysis and prediction," *International Journal of Computer Applications,* vol. 61, 2013.

[4] A. B. E. D. Ahmed and I. S. Elaraby, "Data mining: A prediction for student's performance using classification method," *World Journal of Computer Application and Technology,* vol. 2, pp. 43-47, 2014.

[5] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," *arXiv preprint arXiv:1201.3417,* 2012.

[6] B. K. Bhardwaj and S. Pal, "Data Mining: A prediction for performance improvement using classification," *arXiv preprint arXiv:1201.3418,* 2012.

[7] S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," *arXiv preprint arXiv:1203.3832,* 2012.

[8] S. T. Hijazi and S. Naqvi, "FACTORS AFFECTING STUDENTS'PERFORMANCE," *Bangladesh e-journal of sociology,* vol. 3, 2006.

[9] T. K. Cheruiyot and L. C. Maru, "Service quality and relative performance of public universities in East Africa," *The TQM Journal,* vol. 25, pp. 533-546, 2013.

[10] V. E. Lee and T. L. Zuze, "School resources and academic performance in Sub-Saharan Africa," *Comparative Education Review,* vol. 55, pp. 369-397, 2011.

[11] A. A. Saa, "Educational data mining & students' performance prediction," *International Journal of Advanced Computer Science and Applications,* vol. 7, pp. 212-220, 2016.

[12] Z. N. Khan, "Scholastic Achievement of Higher Secondary Students in Science Stream," *Online Submission,* vol. 1, pp. 84-87, 2005.

[13] S. A. Kumar and M. Vijayalakshmi, "Mining of student academic evaluation records in higher education," in *2012 International Conference on Recent Advances in Computing and Software Systems*, 2012, pp. 67-70.

[14] S. Khan, S. Gupta, Y. K. Sharma, and R.

Rambola, "A Study on Data Mining Techniques and Genetic Algorithm in Education Sector," 2015.

[15] L. Mashiloane and M. Mchunu, "Mining for marks: a comparison of classification algorithms when predicting academic performance to identify "students at risk"," in *Mining Intelligence and Knowledge Exploration*, ed: Springer, 2013, pp. 541-552.

[16] M. Shovon, H. Islam, and M. Haque, "An Approach of Improving Students Academic Performance by using k means clustering algorithm and Decision tree," *arXiv preprint arXiv:1211.6340,* 2012.

[17] R. Jidagam and N. Rizk, "Evaluation Of Predictive Data Mining Algorithms in Student Academic Performance."

[18] O. Oyelade, O. Oladipupo, and I. Obagbuwa, "Application of k Means Clustering algorithm for prediction of Students Academic Performance," *arXiv preprint arXiv:1002.2425,* 2010.

[19] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*: Elsevier, 2011.

[20] D. A. Keim, "Information visualization and visual data mining," *IEEE transactions on Visualization and Computer Graphics,* vol. 8, pp. 1-8, 2002.

[21] Y. Shao and R. S. Lunetta, "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 70, pp. 78-87, 2012.

[22] A. Golbraikh and A. Tropsha, "Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection," *Molecular diversity,* vol. 5, pp. 231-243, 2000.

[23] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping multidimensional data*, ed: Springer, 2006, pp. 25-71.

[24] E. W. Ngai, L. Xiu, and D. C. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert systems with applications,* vol. 36, pp. 2592-2602, 2009.

[25] S. M. Bridges and R. B. Vaughn, "Fuzzy data mining and genetic algorithms applied to intrusion detection," in *Proceedings of 12th Annual Canadian Information Technology Security Symposium*, 2000, pp. 109-122.

[26] X. Wu and V. Kumar, *The top ten algorithms in data mining*: CRC press, 2009.

[27] C. Apté and S. Weiss, "Data mining with decision trees and decision rules," *Future generation computer systems,* vol. 13, pp. 197-210, 1997.