Indian Journal of Science and Technology, *Vol 8(15), IPL057, July 2015*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance

C. Anuradha[1*] and T. Velmurugan[2]

[1]Bharathiar University, Coimbatore, India; anumphil14@gmail.com
[2]Department of Computer Science, D.G.Vaishnav College, Chennai, India; velmurugan_dgvc@yahoo.co.in

Abstract

Objectives: Data mining techniques are implemented in many organizations as a standard procedure for analyzing the large volume of available data, extracting useful information and knowledge to support the major decision-making processes. Data mining can be applied to wide variety of applications in the educational sector for the purpose of improving the performance of students as well as the status of the educational institutions. Educational data mining is rapidly developing as a key technique in the analysis of data generated in the educational domain. **Methods:** The aim of this study presents an analysis of final year results of UG degree students using data mining technique, which carried out in three of the private colleges in Tamil Nadu state of India. The primary objective of this research work is to apply the classification techniques to the prediction of the performance of students in end semester university examinations. Particularly, the decision tree algorithm C4.5 (J48), Bayesian classifiers, k Nearest Neighbor algorithm and two rule learner's algorithms namely OneR and JRip are used for classifying the performance of students as well as to develop a model of student performance predictors. **Results:** The result of this study reveals that overall accuracy of the tested classifiers is above 60%. In addition classification accuracy for the different classes reveals that the predictions are worst for distinction class and fairly good for the first class. The JRip produces highest classification accuracy for the Distinction. Classification of the students based on the attributes reveals that prediction rates are not uniform among the classification algorithms. Also shows that selected data attributes have found to be influenced the classification process. The results showed to be satisfactory. **Improvements:** The study can be extended to draw the performance of other classification techniques on an expanded data set with more distinct attributes to get more accurate results.

Keywords: Classification Algorithm, Classifiers, Comparative Analysis, Educational Data Mining (EDM), Predicting Student Performance

1. Introduction

Currently there is an increasing interest in data mining and educational systems, making educational data mining as a new growing research community. In real world, predicting the performance of the students is a challenging task. The growth in information and communication technologies has enabled higher education leaders to gain access to large volume of information that plays very important role in the key decision making process. One of the primary requirements in this process is that high quality and relevant data has to be provided to the educational leaders at the right time.

Traditionally educational institutions are collecting large volumes of data related to students, faculty members, the organization and management of the educational

**Author for correspondence*

process, and other managerial issues. However, the extent to which the available and collected data is being used is not so significant. In general, the data is used for producing simple queries and traditional reports that are not highly significant in contributing to the decisions making process in the institutions. Moreover, the volume and complexity of the data is often very huge that it becomes difficult to the management of the educational institutions to handle the data and hence remains unused. The potentiality of the available volume of data can be exploited only if it transformed into useful information and in turn is used to generate knowledge to support decision making.

Data mining is the process of discovering meaningful patterns in large quantities of data[1]. Data mining is emerging as promising frameworks which provide wide variety of techniques, methods and tools to enables provides for thorough analysis of available data in various fields. Considering the potential application of data mining in educational sector, Educational Data Mining (EDM) was started as a new stream in the data mining research field[2]. EDM concerns with new methods and techniques by inquiring into eccentric type of data from educational settings to understand students learning ability. In educational domain, data mining techniques are very useful for enhancing the current educational standards and managements. These techniques provide a route to a multiple level of ranking, a finding which gives a new perception of how people can become proficient in these educational sectors. As a result of this, EDM has given rise to hypothesis concerned with the scientific study of human sciences[3].

We known that wide range of data is stored in educational databases, so in order to get required data and to find the hidden relationship, for that different data mining techniques are developed and used[4] Anju and Robin[5] have conducted a survey on Decision tree classification algorithms like ID3, C4.5 and CART to predict student academic performance. ID3 chooses the splitting attribute by using information gain measure. It accepts only categorical attributes. An improved version of ID3 is C4.5 which accepts both continuous and categorical attributes in building decision tree. They analyse the efficiency of various decision tree algorithms based on their accuracy and time taken. Ajai and Saurabh[4] have proposed to extract the knowledge discovery from the student database using the data mining techniques including ID3, C4.5 and Bagging. They examine that ID3 does not give accurate result when there is noise and does not support pruning. Whereas C4.5 uses Gain Ratio to build a tree and removes the partial perspective of information gain when there are many upshot values of an attribute. Their result reveals that classifier accuracy shows the true positive rate of the model for the FAIL class is 0.84 for ID3 and C4.5 decision trees that means model is successfully identifying the students who are likely to fail.

Another work carried out by Sunita et al. have put forward to analyse student trends &behaviour towards education and attempt to study the present behavioural pattern of student in a cross section[6]. This paper surveys an application of data mining in education system to analyse the final year UG students and presented their result analysis in WEKA tool. The researchers applied ZeroR algorithm for classification and DBSCAN algorithm for clustering the students to improve the performance of students. A work done by P. Ajith et al. discussed about rule mining framework for students performance evaluation. In this paper, they use Association Rules instead of tree based classification since the result of a tree based classification is complicated to understand and depends on the technical competency of the decision maker. Among sets of items in transaction databases, AssociationRules aims at discovering implicative tendencies that gives valuable information for the decision-maker which is absent in tree based classifications. So they propose a new interactive approach to prune and filter discovered rules[7]. Ogunde and Ajibade have developed a new system for the prediction of students graduation grades based on entry results data. The proposed system uses ID3 algorithm to classify the data and construct the decision tree by employing a top-down, greedy search to test every attributes. The output of their experiment helps management staffs and academic planners to improve overall performance of students and consequently reduce failures rate in most academic institutions[8]. A Work done by Arpit Trivedi has put forward a simple approach for categorizing student data using decision tree based approach. They took database for five subject's marks of 100 students for each four different classes. For taking measures of category of specific student, a frequency measure is used as a feature extraction. The most frequent five subject marks of each of the students are used to develop a trained classifier. With the use of trained classifier, they predicted the class for indefinite student automatically[9].

A work has done by Agrawal and Gurav have done a review on Data Mining Techniques Used for Educational System. They proposed a method to predict student failure by using Data Mining as early as possible so that we can provide some type of assistance for trying to avoid or reduce failures. This paper is based on survey which proposes to apply data mining techniques such as association rule mining, classification techniques[10]. Another research done by Suman and Pooja[11] describe the various approaches and techniques of data mining which can be applied on Educational data to build up a new environment to improve performance of existing data and help to create the new predictions on the data. In this paper, author describes the comparative study of classification techniques are Bayes net, naïve net and decision tree etc. And clustering techniques are k-mean, hierarchical, OPTICS and DBSCAN etc. Dinesh and Radika had done a survey on predicting Student academic Performance in educational environment which is based upon the psychological and environmental factor is predicted using different educational data mining techniques. Researchers also survey the predictive model in data mining and current trends in prediction in data mining[12].

Shanmuga Priya[13] conducted study on improving the student's performance using Educational Data Mining based by selecting 50 students from Hindustan College of Arts and Science, Coimbatore, India. By using decision tree classification on 8 attribute, it was found that the class test, seminar, attendance, lab practicals are used to predict the Student performance. This prediction will help to the teacher to give special attention of students and improve student confidence on their studies. A work done by Dorina Kabakchieva[14] using Data mining Methods for Classification in Predicting Student Performance. This paper presents the initial results from a data mining research project implemented at a Bulgarian university, aimed at revealing the high potential of data mining applications for university management. The specific objective of the proposed research work is to find out if there are any patterns in the available data that could be useful for predicting student's performance and also describe the methodology for the implementation of the initiated data mining projects.

The classification is a data mining technique which includes systematic approach to building the classification models from an input dataset[15]. Some of the popular classifiers used to solve a classification problem are decision tree classifiers, rule-based classifiers, neural networks, support vector machines, and naive Bayes classifiers. The classification techniques uses learning algorithm to identify a model that best fits the relationship between the attribute set and class label of the input data[16]. Therefore, a key objective of the learning algorithm is to build a predictive model that accurately predicts the class labels of previously unknown records.

The aim of classification is to predict the future output based on the available data. Hence, educational institute is looking to predict the future output of their enrolled students based on their available previous and current students' data, which make classification one of the techniques better suited for educational analysis.Most of the previous studies focus on the use of classification for predictions based on enrollment data, Performance of students in certain course, grade inflation, anticipated percentage of failing students, and assist in grading system. Up to our knowledge, there are no studies that use classification to predict a student final outcome based on his/her grades in a program study plan. Analyzing all thecourses that are required in the study plan will identify the listof courses that have a huge impact on final GPAs[17]. Our contribution in this paper examines that various classification algorithms and their performance are compared using WEKA software and results are discussed. The open source data mining tool WEKA, developed at the University of Waikato, New Zealand, which is free software available under the GNU General Public License, was used in the present work. WEKA is a machine learning software written in Java. WEKAis a collection of machine learning algorithms developed for solving real-world data mining problems. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis, predictive modeling. In addition, it contains lot of packages which includes Filters, Classifiers, Clusters, Associations, and Attribute Selection.

The WEKA software was used for the study implementation of the model, since it is freely available to the public and is widely used for research in the data mining field. This would classify the students into the four classes (categories), depending on their pre-college characteristics, performance and other college features. Several types of classification algorithms were selected and the dataset was applied with these algorithms. The classifiers used in this paper consists of common decision tree algorithm C4.5 (J48), two Bayesian classifiers (Naive Bayes and BayesNet), Nearest Neighbor algorithm (IBk) and two rule learners (OneR and JRip). The results obtained from the classification task are presented in the Section 3.

This paper is organized as follows. Various Classification algorithms and the research methodology which are taken in this study are discussed in Section 2. The obtained results and the comparative analysis are given in Section 3 and the paper concludes with a summary of the achievements are given in Section 4. Finally, vital references are mentioned.

2. Research Methodology

Classification is a simple process of discovering a prototype (or function) that recognize the salient features of data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. It forecast distinct and unordered labels in huge data sets. As with classification, the test set is used to build apredictor but an independent test set should be used to assess its accuracy[18]. The data classification process involves learning and classification. In learning the training data are analysed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. The classifier training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination[19]. The EDM Classification is used to categorize the students to shape their learning styles and inclination.

2.1 Classification Algorithms

Educational Data mining can be implemented in many techniques such as decision trees, neural networks, k-nearest Neighbor, Naive Bayes, support vector machines and many others. Using these methods many kind of knowledge can be discovered such as association rules, classification, clustering, and pruning the data. Some of the Classification algorithms mentioned here for the proposed work have provided a better understand in educational resources.

2.1.1 Decision Tree Classifier

Decision tree classifiers are one of the popular and powerful tools for classification. Generally, decision tree classifiers have a tree-like structure which starts from root attributes, and ends with leaf nodes. It also has several branches consisting of different attributes, the leaf node on each branch representing a class or a kind of class distribution. Decision tree algorithms describe the relationship among attributes, and the relative importance of attributes. The advantages of decision trees are that they represent rules which could easily be understood and interpreted by users, do not require complex data preparation, and perform well for numerical and categorical variables. In WEKA environment, decision tree classifier is implemented using J48 classification filter in which it is based on the C4.5 decision tree algorithm[14]. VikasChirumamilla et al. have proposed a novel approach to predict student placement chance with Decision tree induction. This paper presents a study on twofold objective model and aims at offering a reliable and predictive tool. The objective is to predict the performance and placement chances of a student by using one of the decision tree algorithms[20].

The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive divide-and-conquer manner. The technique uses Gain Ratio instead of Information Gain for Splitting purpose[21].

$$\text{Gain Ratio}(D,S) = \text{Gain}\,(D,S)\,/\,\text{Split INFO}$$

$$\text{Where, Split INFO} = -\left(\sum_{i-1}^{s} \frac{D_i}{D}\log_2 \frac{D_i}{D}\right)$$

The algorithm, summarized as follows.

Step 1: create a node N;

Step 2: if samples are all of the same class, C then

Step 3: return N as a leaf node labeled with the class C;

Step 4: if attribute-list is empty then

Step 5: return N as a leaf node labeled with the most common class in samples;

Step 6: select test-attribute, the attribute among attribute-list with the highest information gain;

Step 7: label node N with test-attribute;

Step 8: for each known value $a_i$of test-attribute

Step 9: grow a branch from node N for the condition test-attribute= a_i;

Step 10: let $s_i$be the set of samples for which test-attribute= a_i;

Step 11: if $s_i$is empty then

Step 12: attach a leaf labeled with the most common class in samples;

Step 13: else attach the node returned by generate decision tree (si, attribute-list, and test-attribute)

2.1.2 Bayesian Classifiers

Bayesian classifiers are statistical classifiers that predict class membership by probabilities, such as the probability that a given sample belongs to a particular class. Bayesian networks and Naive Bayes are the two popular Bayesian classifiers which are more commonly used in real-world applications because of their simplicity, computational efficiency and very good performance[14]. A work conducted by Praveen Sundar for predicting student's academic performance using Bayesian Network classifier and generates a Model. This model helps earlier in identifying the drop outs and students who need special attention. The objective of this paper is to predict the student performance and make a comparative study on Bayesian network classifiers, through that we compute which classifier predicts more students when compared to other classifiers[22].

A Bayesian classifier is based on the idea that the role of a (natural) class is to predict the values of features for members of that class. Bayesian classifiers are based on Bayes theorem, which says

$$P\left(c_j|d\right)=p\left(d|c_j\right)p\left(c_j\right)p\left(d\right)$$

P (c_j | d) = probability of instance d being in class c_j,

p (d | c_j) = probability of generating instance d given class c_j,

P (c_j) = probability of occurrence of class c_j,

p (d) = probability of instance d occurring

2.1.3 K-Nearest Neighbor Classifier (K-NN)

The k-Nearest Neighbor algorithms (k-NN) classify objects based on the closest training examples in the feature space. K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The major drawback of k-NN algorithm is that its accuracy can be severely degraded by the presence of noisy or irrelevant features. Similarly, its accuracy becomes poor if the feature scales are not consistent with their importance[14]. Pallavikulkarni and Roshani Ade had done a work on Incremental Learning for Predicting Students performance. In this paper, four classifiers that can run incrementally: the Naïve Bayes, KStar, IBK and Nearest neighbor (KNN) have been compared and observed that nearest neighbor algorithm giver better accuracy compared to others if applied on Student Evaluation dataset. In nearest neighbor algorithms, number of training instances is described by n attributes. Each instance is representation of a point in n-dimensional space forming pattern space of training tuples. When unknown instance comes, a KNN algorithm looks the pattern space of the k training tuples that are closest to new instance[23].

In k-Nearest Neighbours classification, each of the characteristics in the training set is considered as a different dimension in some space, and take the value an observation has for this characteristic to be its coordinate in that dimension, so getting a set of points in space. The similarity of two points is considered to be the distance between them in this space under some appropriate metric. The way in which the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the k closest data points to the new observation, and to take the most common class among these. This is why it is called the k Nearest Neighbours algorithm.

The k-Nearest Neighbours classification can be summarised as:

- A positive integer k is specified, along with a new sample.
- The k entries in the dataset are selectedin the databases which are closest to the new sample.
- Themost common classification of these entries is identified.
- This is the classification of the new sample.

2.1.4 Rule Learners Classification Algorithm

Rule learners are used to generate classification rules. Two Rule learner classifiers are considered in the study viz. OneR and JRip. The OneR classifier is mostly used for generating one-level decision tree expressed in the form of a set of rules that often test only one particular attribute. It is a simple, cheap method that often produces good rules with high accuracy for characterizing the structure in data. On the other hand, the JRip classifier implements the RIPPER (Repeated Incremental Pruning to Produce Error Reduction) algorithm. JRip classifiers examine the classes in increasing size and an initial set of rules for the class is generated using incremental reduced-error pruning[14].

2.2 Statement of the Problem

In this paper, research work is carried out by using the classification algorithms in order to identify the students' performance. The data pertaining to the students that are often collected includes the demographic details of the students like gender, family size and type, income, parent's educational attainment and locality. In addition, pre-collegiate conditions of the students like their performance in secondary and higher secondary classes are also collected and maintained in the colleges. Thus, it could be useful to the educational leaders and management of the colleges, if the features in the currently available data can be acting as the indicator for predicting the performance of the students. The major objective of this study is to analyze the student's data available in the degree colleges to identify any specific patterns that might be useful in the prediction of their performance in the university exams. The specific objective of the study is to classify students according to their performance in the final examination based on their personal and pre-collegiate characteristics.

2.3 Dataset for Proposed Work

A student's dataset was created based on the demographic and pre-collegiate characteristics of the students along with their performance in the class and university examinations. The dataset was used to evaluate the performance of various classification algorithms in predicting the performance of the students in the final exams. The data mining classification algorithms that are compared in the study includes J48 decision tree algorithm which is an open source Java implementation of C4.5 algorithm[24], Naive Bayes Classifiers[25], *k*-Nearest Neighbours algorithm (K-NN)[26], OneR and JRip algorithm[27,28].

The dataset used in the study consists of primary data generated from the student's admission data available with the college database. In addition, certain aspects of the dataset are collected by administering a structured questionnaire to the concerned students. The target variable or the output variable is Student End Semester Marks (ESM) which is usually available in the numeric form in terms of percentage. Hence categorical target variable was constructed based on the original numeric parameter (percentage score). The target variable has four distinct values as First Class (Score is greater than 60%), Second Class (Score lies between 45 to 60%), Third Class (Score lies between 36 and 45%), Fail (Score less than 36%). The attributes related to the student personal data include gender, category of admission, living location, family size, and family type, annual income of the family, father's qualification and mother's qualification.

The attributes referring to the students' pre-college characteristics include Students Grade in High School and Students Grade in Senior Secondary School. The attributes describing other college features include thebranch of study of the students, place of stay, previous semester mark, class test performance, seminar performance, assignment, general proficiency, class attendance and performance in the laboratory work. The study is limited to student data for three colleges in Tamil Nadu State. The detailed description of the dataset is provided in Table 1.

The domain values for some of the variables were defined for the present investigation as follows:

- GENDER – Gender of the students. It is split into two classes values: Male and Female
- BRANCH – Students branch obtained. Branch is split into four classes: BCA, B.Sc (CS), B.Com, and B.A.
- CAT – Students category obtained. Here Category is split into six classes: BC-Backward class, MBC- Most Backward class, OC-Open category, SBC- Special Backward classes, and SC- Scheduled castes.
- HSG & SSG – Students Grade in High School and Senior Secondary. Here grade is divided into Seven class values: O-905-100%,A-80%-89%,B-70%-79,C-60%-69%,D-50%-59%,E-35%-49%,FAIL- <35%.
- LLoc – Living Location of students is obtained. Location is split into five classes: Village, Taluk, Rural, Town, and District.
- HOS – Students stay in hostel or not. It is split into two Classes: Yes- Students Lives in Hostel, No- Students not lives in hostel.
- FSize – Here Students Family Size is obtained. It is divided into four class values like 1, 2, 3 and > 3.
- Ftype – Family Type of Student. Family Type is split into two classes: Joint family and Individual family.
- FINC–Family Annual Income of student is obtained. Annual Income is divided into Three classes namely Poor, medium and High.
- FQual&MQual – Father and Mother Qualification is obtained. It is split into six classes: no-education, elementary, under graduate, postgraduate and doctorate.
- PSM – Previous Semester marks of Students obtained in BCA, B.Sc (CS), B.Com, B.A course. This response variable is divided in Five Class values: Distinction - >75, First – 60% - 74%, Second – 50%-59%, Third-40%- 9%, Fail-<39.
- CTG – Class test grade is obtained. In each semester three internal tests are conducted and average of three tests are used. CTG is split into three classes: Poor - < 40%, Average - >40% and <60%, Good - >60%.

Table 1. Description of the attributes used for Classification

Variables	Description	Possible Values
Gender	Students Sex	{Male, Female}
Branch	Students Branch	{BCA, B.SC, B.COM, B.A}
Cat	Students category	{BC, MBC, OC, SBC, SC}
HSG	Students grade in High School	{O – 90% -100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 35% - 49%, FAIL - <35%}
SSG	Students grade in Senior Secondary	{O – 90% -100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 35% - 49%, FAIL - <35% }
LLoc	Living Location of Student	{Village, Taluk, Rural, Town, District}
HOS	Student stay in hostel or not	{Yes, No}
FSize	Student's family size	{1, 2, 3, >3}
FType	Students family type	{Joint, Individual}
FINC	Family annual income	{poor, medium, high}
FQual	Fathers qualification	{no-education, elementary, secondary, UG, PG, PhD}
MQual	Mother's Qualification	{no-education, elementary, secondary, UG, PG, Ph.D. NA}
PSM	Previous Semester Mark	{First > 60% Second >45 &<60% Third >36 &<45% Fail < 36%}
CTG	Class Test Grade	{Poor, Average, Good}
SEM_P	Seminar Performance	{Poor , Average, Good}
ASS	Assignment	{Yes, No}
GP	General Proficiency	{Yes, No}
ATT	Attendance	{Poor , Average, Good}
ESM	End Semester Marks	{First > 60% Second >45 &<60% Third >36 &<45% Fail < 36%}

- SEM-P – Seminar Performance of obtained. In each Semester seminar are organized to check the performance of students. It is divided into three Classes: Poor – Presentation and communication skill is low, Average – Either presentation or communication is fine, Good – Both presentation and communication skill is fine.
- ASS – Assignment Performance. In each semester three assignments are given to students. ASS is split into two classes: Yes – Student submitted assignment, No – Student not submitted assignment.
- GP – General Proficiency performance. Like seminar, general proficiency test is conducted in each semester. It is split into two classes: Yes – Student participated in

general proficiency, No – Student not participated in general proficiency.

- ATT– Attendance of Student. Attendance is divided into three classes: Poor - <60%, Average - >60% and <80%, Good - >80%.
- ESM – End Semester Mark of Students obtained in BCA, B.Sc (CS), B.Com.BA course. This response variable is divided in Five Class values: Distinction - >75, First – 60% - 74%, Second – 50%-59%, Third- 40%- 9%, Fail-<39.

3. Experimental Results and Observations

The main objective of the study is to explore if it is possible to predict the performance of the student (output) based on the various explanatory (input) variables which are retained in the model. The classification model was built using several different algorithms and each of them using different classification techniques. The WEKA Explorer application is used at this stage. Each classifier is applied for two testing options - cross validation (using 10 folds and applying the algorithm 10 times - each time 9 of the folds are used for training and 1 fold is used for testing) and percentage split (2/3 of the dataset used for training and 1/3 – for testing). The screen shot of the WEKA preprocessing stage is shown in Figure 1.

3.1 Results of Decision Tree Classifier

In the present study, J48 classification algorithm was implemented on the data and the results of the classification is presented in Table 2. It is inferred from the Table 2, that J48 has correctly classified about 72.51% for the 10-fold cross-validation testing and 69.66%for the percentage split testing. It produces a classification tree with a size of 41 nodes and 30 leaves. The screenshot of decision screen building process in shown in Figure 2.

The results from Table 2 reveal that the True Positive Rate is high for three of the classes – Third (100 %), First (84-98 %). The TP rate is low for the class - Distinction (50 %), while it is very low for the class– Second (40-66 %), Fail (11-16 %). The Precision is high for the First class (67-76 %), Second class (72-85 %), medium for the

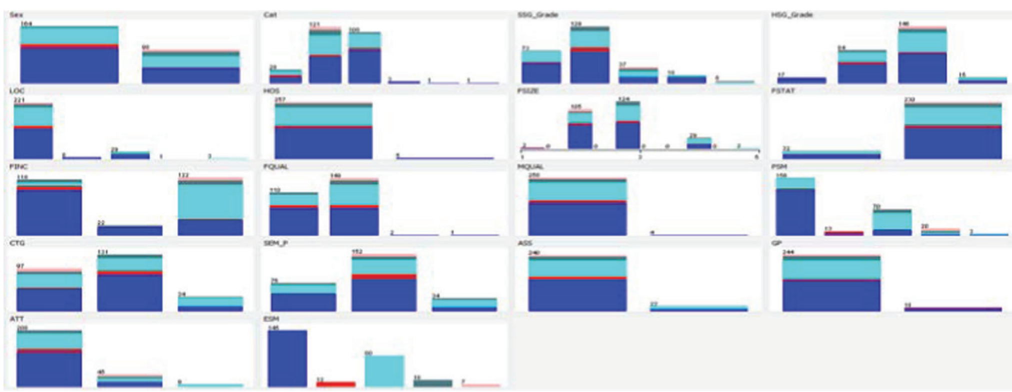

Figure 1. WEKA Screenshot of Data Distribution in the Preprocessing Stage.

Table 2. Classification results for the decision tree algorithm (J48)

Class	J48 – 10-fold Cross validation		J48 – Percentage split	
	TP Rate	Precision	TP Rate	Precision
Distinction	0.5	0.545	0	0
First	0.841	0.758	0.98	0.686
Second	0.663	0.716	0.407	0.846
Third	1	0.778	0.4	1
Fail	0.111	0.286	0.167	0.333
Weighted Avg.	0.725	0.703	0.697	0.713

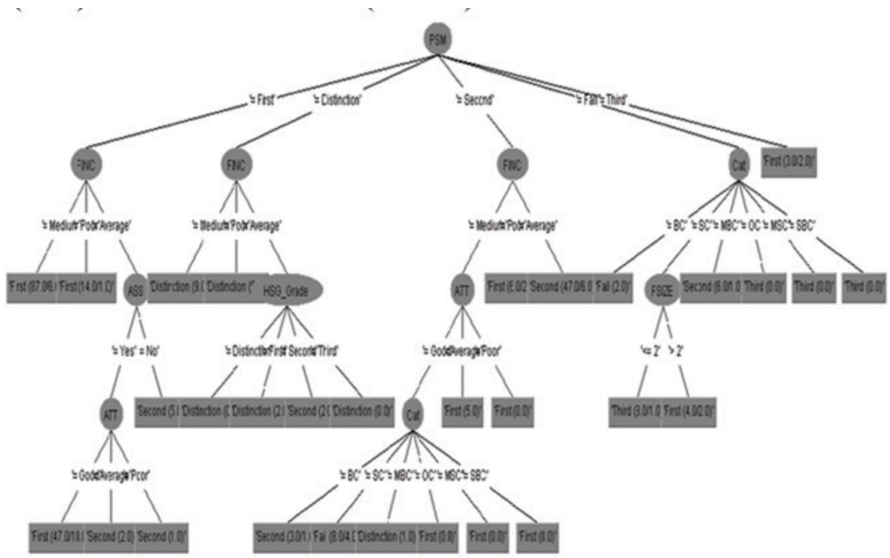

Figure 2. Screenshot of Decision Tree build using J48 Classifier.

Table 3. Classification results for the Naive Bayes Classifiers

Class	Naive Bayes 10-fold Cross validation		Naive Bayes Percentage split	
	TP Rate	Precision	TP Rate	Precision
Distinction	0.333	0.5	0.5	0.333
First	0.841	0.772	0.857	0.778
Second	0.675	0.74	0.741	0.741
Third	1	0.875	0.8	1
Fail	0.167	0.2	0	0
Weighted Avg.	0.725	0.713	0.753	0.717

Table 4. Classification results for the BayesNet Classifiers

Class	BayesNet 10-fold Cross validation		BayesNet Percentage split	
	TP Rate	Precision	TP Rate	Precision
Distinction	0.417	0.5	0.5	0.333
First	0.855	0.785	0.857	0.764
Second	0.675	0.75	0.704	0.731
Third	1	0.875	0.8	1
Fail	0.111	0.143	0	0
Weighted Avg.	0.733	0.719	0.742	0.706

Distinction (54 %) and low for the class Fail (29-33 %) classes.

3.2 Results of Bayesian Classifiers

The present study implements Bayesian classifiers namely Bayesian networks and naive Bayes on the dataset and the results are presented in Table 3 and Table 4. Table 3 presents the classification results for Naive Bayes classifier and it is found that Naive Bayes classifier correctly classifies about 72.5191 % for the 10-fold cross-validation testing and 75.28 % for the percentage split testing.

The results from Table 3 reveal that the True Positive Rate is high for most of the classes – First, second and Third. TP rate is very low for the class Fail (16.7 %). The precision is also high for the classes - First, Second and Third. Table 4 presents results of BayesNet classifier on the dataset. It can be verified that Bayes Net correctly classifies about 73.38 % for the 10-fold cross-validation testing and 74.23 % for the percentage split testing.

The results from Table 4, shows that the True Positive Rate is high for the classes - First and Third. TP rate is very low for the class Distinction (11.1 %). The precision is also high for the classes - Third.

3.3 Results of k-Nearest Neighbor Classifier (K-NN)

In WEKA environment, k-NN classifier is implemented using IBk classification filter on the dataset. The result of this implementation is shown in Table 5. It can be seen that k-NN classifier correctly classifies about 68.32 % for the 10-fold cross-validation testing and 62.92% for the percentage split testing.

The results from Table 5 show that the True Positive Rate is high for the Third and First class (73-87%). TP rate is very low for the classes Distinction and Fail. The precision is found to be high for the classes –First and Third and very low for the classes – Distinction. TP rate is Zero for the class Fail.

3.4 Results of Rule Learners Classification Algorithm

Table 6 shows the classification results for OneR classifier. The OneR classifier correctly classifies about 64.88% for the 10-fold cross-validation testing and 62.92 % for the percentage split testing.

The results from Table 6 show that the True Positive Rate is high for the First (74-87 %) and Second classes. TP rate is zero for the classes - Fail. TP rate is low for the class – Third. The precision is found to be high for the classes – First and Distinction and zero for the classes – Fail. Table 7 shows the results for JRip classifier. It is found that JRip correctly classifies about 71.75 % for the 10-fold cross-validation testing and 73.03% for the percentage split testing. The results also shows that TP rate is high for the majority of the classes like First, Third, Distinction and Second.

3.5 Performance Comparison between the Applied Classifiers

The results for the performance of the selected classification algorithms (TP rate, percentage split test option) are summarized and presented on Figure 3.

Table 5. Classification results for the k-NN Classifiers

Class	k-NN Classifier 10-fold Cross validation		k-NN Classifier Percentage split	
	TP Rate	Precision	TP Rate	Precision
Distinction	0.167	0.5	0	0
First	0.869	0.72	0.735	0.679
Second	0.55	0.657	0.556	0.577
Third	1	0.875	1	1
Fail	0	0	0	0
Weighted Avg.	0.683	0.645	0.629	0.605

Table 6. Classification results for the OneR Classifiers

Class	OneR Classifier 10-fold Cross validation		OneR Classifier Percentage split	
	TP Rate	Precision	TP Rate	Precision
Distinction	0.167	0.5	0.5	0.5
First	0.745	0.761	0.878	0.705
Second	0.738	0.546	0.444	0.462
Third	0.143	0.125	0	0
Fail	0	0	0	0
Weighted Avg.	0.649	0.614	0.629	0.539

Table 7. Classification results for the JRip Classifiers

Class	JRip Classifier 10-fold Cross validation		JRip Classifier Percentage split	
	TP Rate	Precision	TP Rate	Precision
Distinction	0.667	0.615	0.5	0.5
First	0.834	0.752	0.898	0.772
Second	0.65	0.703	0.667	0.643
Third	1	0.7	0.4	1
Fail	0	0	0	0
Weighted Avg.	0.718	0.677	0.73	0.687

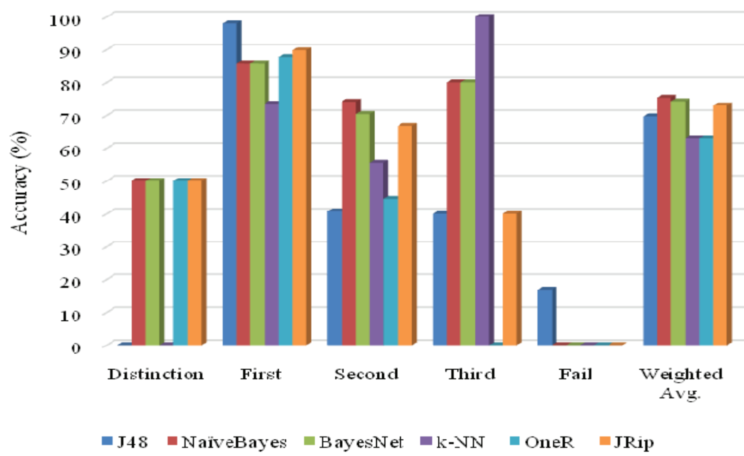

Figure 3. Classification algorithms performance comparison.

The results of the classification reveals that the Bayesian classifiers like Naïve Bayes and BayesNet classifiers performs very well in comparison with other classifiers with the highest overall accuracy, followed by JRip classifier and J48 classifiers. K-NN and OneR performs poorly and are less accurate than the others. The overall accuracy of all the tested classifiers is well above 60%. Naive Bayes and BayesNet registered accuracy higher than 70 %. J48 produces accuracy very near to 70 %. On the other hand, OneR and K-NN classifiers achieved classification accuracy of just 62.9 %. In addition, further detailed analysis of the classification accuracy for the different classes reveals that the predictions are worst for the distinction class and fairly good for the other classifiers. The classification accuracy is very good for first class. The JRip produces highest classification accuracy for the Distinction.

4. Conclusion

The results of the data mining algorithms for the classification of the students based on the attributes selected reveals that the prediction rates are not uniform among the algorithms. The range of prediction varies from 61-75 %. Moreover, the classifiers perform differently for the five classes. The data attributes that are found to have significantly influenced the classification process are First and Second classes. The study can be further extended to study the performance of other classification techniques with larger sample dataset. In future work we can apply different data mining techniques on an expanded data set with more distinct attributes to get more accurate results. Apart from classification, data mining techniques like Clustering can be applied to the dataset to draw more intelligence from it.

5. References

1. Galit S, Patel NR, Bruce PC. Data mining for business intelligence: concepts, techniques, and applications in Microsoft Office Excel with XLMiner. John Wiley & Sons; 2007.
2. Romero C, Ventura S. Educational data mining: A survey from 1995 to 2005. Expert systems with applications. 2007; 33(1)135–46.

3. Baker RSJD. Data mining for education. International encyclopedia of education. 2010; 7:112–8.
4. Pal AK, Pal S. Analysis and Mining of Educational Data for Predicting the performance of Students. International Journal of Electronics Communication and Computer Engineering. 2013; . 4(5):1560–5.
5. Rathee A, Mathur RP. Survey on Decision Tree Classification algorithm for the Evaluation of Student Performance. International Journal of computers & Technology. 2013; 4(2):244–7.
6. Aher SB, Lobo LMRJ. Data mining in educational system using Weka. IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT). 2011; 3:20–5.
7. Ajith P., Tejaswi B, Sai MSS. Rule Mining Framework for Students Performance Evaluation. International Journal of Soft Computing and Engineering. 2013; 2(6):201–6.
8. Ogunde AO, Ajibade DA. A Data Mining System for Predicting University Students' Graduation Grades Using ID3 Decision Tree Algorithm. Journal of Computer Science and Information Technology. 2014; 2(1):21–46.
9. Trivedi A. Evaluation of Student Classification Based On Decision Tree. Int Journal of Advanced Research in Computer Science and Software Engineering. 2014 Feb; 4(2):111–2.
10. Agrawal BD, GuravBharti B. Review on Data Mining Techniques used For Educational System. Int Journal of Emerging Technology and Advanced Engineering. 2014 Nov; 4(11):325–9.
11. Suman, Pooja Mittal P. A Comparative Study on Role of Data Mining Techniques in Education: A Review. International Journal of Emerging Trends & Technology in Computer Science. 2014 Jun; 3(3):65–9.
12. Dinesh KA, Radhika V, A Survey on Predicting Student Performance. Int Journal of Computer Science and Information Technologies. 2014; 5(5):6147–9.
13. Shanmuga PK. Improving the student's performance using Educational data mining. International Journal of Advanced Networking and Application. 2013; 4(4):1680–5.
14. DorinaKabakchieva. Predicting Student Performance by using Data mining Methods for Classification. Bulgarian Academy of Science, Cybernetics and Information Technologies. 2013; 13(1):61–72.
15. Longbing C. Data mining and multi-agent integration. Springer Science & Business Media; 2009.
16. Tan, Pang-Ning, Steinbach M, Kumar V. Introduction to data mining. Boston: Pearson Addison Wesley; 2006; 1.
17. Al-Barrak MA, Al-Razgan M. Predicting Students Final GPA Using Decision Trees: A Case Study. International Journal of Information and Education Technology. 2016 Jul; 6(7):528–33.
18. Han J, Kamber M. Data Mining: Concepts and Techniques, New Delhi: Morgan Kaufmann Publishers; 2006. ISBN: 978-81-312-0535-8.
19. Baradway BK, Pal S. Mining Educational Data to Analyze Students Performance. Int Journal of Advances in Computer Science and Applications. 2011; 2(6):63–9.
20. Chirumamilla V, BhagyaSruthi T, Velpula S, Sunkara I. A Novel approach to predict Student Placement Chance with Decision Tree Induction. International Journal of Systems and Technologies. 2014; 7(1):78–88.
21. Masethe MA, Masetha HD. Prediction of work integrated Learning placement using Data mining Algorithms. Proceedings of the World Congress on Engineering and Computer Science. 2014 Oct; 1:22–4.
22. Sundar PVP. A Comparative Study For Predicting Students Academic Performance using Bayesian Network Classifiers. IOSR Journal of Engineering. 2013 Feb; 3(2):37–42.
23. Kulkarni P, Ade R. Prediction of Students Performance based on Incremental Learning. International Journal of Computer Applications. 2014 Aug; 99(14):10–6.
24. Ross QJ. C4. 5: programs for machine learning. Elsevier; 2014.
25. Stuart R, Norvig P. Artificial Intelligence: A Modern Approach. EUA: Prentice Hall; 2003.
26. Altman NS. An introduction to kernel and nearest neighbor non-parametric regression. The American Statistician. 1992; 46(3):175–85.
27. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2005.
28. Cohen WW. Fast effective rule induction. Proceedings of the twelfth international conference on machine learning; 1995. p. 115–23.