Students' Academic Performance Prediction Model Using Machine Learning

Muhammad Sudais (✉ sudaismsm@gmail.com)

FAST - National University of Computer and Emerging Sciences <https://orcid.org/0000-0002-2209-8104>

Muhammad Safwan

FAST - National University of Computer and Emerging Sciences

Maryam Aisha Khalid

FAST - National University of Computer and Emerging Sciences

Shaheer Ahmed

FAST - National University of Computer and Emerging Sciences

Research Article

Keywords: Student academic performance, Academic performance prediction, Machine Learning

Posted Date: January 25th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1296035/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Providing quality education to students is the main objective of higher education institutions. The need of identifying students with weak performances has been a rising problem and most teachers have relied on calculating the average of exam grades. The main objective of our project is to predict and identify the students who might fail in semester examinations. This would prove helpful for teachers in providing additional assistance to such students. The data which was analyzed consisted of students' transcript data that included their CGPA and grades in all courses which were taken from a university. The machine learning algorithms which were used in this research include; Naïve Bayes classifier, Neural Network, Support Vector Machine, and Decision Tree classifier. A comparative analysis has been performed on the obtained accuracy results of the algorithms used. This research shows that machine learning proves useful in predictions, but there is a lot more work to be done using this technology.

1. Introduction

1.1 Background Information

Prediction by analogy is so pervasive, that we normally don't notice it. Higher education institutions always aim to provide best education and tutoring to their students. The alarming increment in drop-out rates in many institutes has gone unnoticed, and therefore there has been a need to identify students with weak performances in courses. This prediction technique would prove advantageous for teachers in providing additional assistance to weak students and also promoting the dedicated ones. The data of students is collected and utilized to meet the needs of students. Other approaches fail to notice the students' performance pattern over the course of passing semesters.

Machine learning algorithms have proven to be a helpful tool in predicting students' performance based on various factors for foreseeing poor performances over the course of their semesters. The at-risk students can be detected using their demographic data. Applying data mining algorithms on datasets could prove beneficial to all participants in educational institutes. [1]. It has been proved that most accurate machine learning algorithm for such prediction is Naïve Bayes classifier. [1] [5]. The variables used for prediction included academic achievements from high school, entrance exam, and attitude toward studying, including marks obtained in assignments. Social and school related features were also noted. [1] [2] [3] [6] Incremental approach in machine learning is important for real world predictions for reasons; the learning approaches must do some changes on the trained system so that unlearned knowledge can be proved useful. It is proven for the fact that student academic achievements are highly influenced by past grades and scores, but there are also other relevant variables contributing for accurate predictions. [3]

1.2 Problem Statement

To predict and identify the students' academic performance to guide them to better results and provide quality education.

1.3 Project Objective

The main objective of higher education institutions is to provide quality education to its students. A good prediction of the students' performance is helpful to identify the low performance students at the beginning. The intention behind this research is identification and extraction of knowledge for foreseeing poor and good performances.

1.4 Overview

In this research, we have described the methodology and presented the results of analyzing performance of different classification algorithms. The dataset has been collected from computer science department of FAST-NUCES which consists of students' transcript data that included their grades in all courses. Different machine learning algorithms Naïve Bayes, Neural Networks, Support Vector Machine and – have been employed to predict students' grades efficiently.

The paper is organized in seven sections. The project statement is presented and explained in the Introduction. A literature review and background study of related research work is mentioned in Section 2, the proposed work is explained in Section 3, the experimental setup comprising of software tools used are described in Section 4, the obtained results are explained in Section 5, the comparative analysis of the results is described in Section 6, and this research is concluded in Section 7.

2. Literature Review

There are different methods of data mining for the prediction of students' success and Edin Osmanbegović, Mirza Suljić (2012) [1] performed a study regarding those methods. The data for was collected from the surveys conducted during the summer semester at the University of Tuzla, the Faculty of Economics, academic year 2010-2011, among first year students. The success was evaluated with the passing grade at the exam. The variables included students' socio-demographic variables, achieved results from high school and from the entrance exam, and attitudes towards studying. The three algorithms were compared (i.e., Naive Bayes, Multilayer Perceptron MLP, J48) In conclusion, Naïve Bayes algorithm provided better prediction accuracy (i.e., 76.65 %). Predicting a student's performance is useful in different ways related with university-level distance learning. Incremental version of algorithms (Naive Bayes, 1-MM and WINNOW) is proposed by S. Kotsiantis, K. Patriarcheas, M. Xenos (2010) for improvement of the classification accuracy using a dataset (i.e., university students' marks in written assignments). This model reaches precision 73%.

Student achievement is highly affected by previous performances. [3] [7] While considering university level academic predictions, there was a study on students' secondary school data. Paulo Cortez, Alice Silva (2008) [3] performed this study on a dataset (i.e., student grades, demographic, social and school related features). The four models consisted of Decision trees, Random Forest, Neural Networks and Support Vector Machines. The research of J. P. Vandamme, N. Meskens, J. F. Superby (2007) [7] was conducted on a dataset of 533 university students with their demographic information, academic records

and their involvement in studies. The models included algorithms i.e., ID3, Multi-Level Perceptron and Linear Discriminant Analysis with accuracies 40.63%, 51.88% and 57.35% respectively.

Data mining algorithms are used for analyzing available data and extracting information to support in making decisions [4] [5] [6]. A research involved algorithms for data mining (i.e., C4.5 decision tree, Naïve Bayes, BayesNet, k-Nearest Neighbor and JRip rule learner) [4] Dorina Kabakchieva (2013) performed this research involving the dataset from a Bulgarian university. The most reliable algorithms were proved to be C4.5 and JRip. Data mining is used to overcome the problem of low grades of graduate students. A research performed by Mohammed M. Abu Tahir, Alaa M. El-Halees (2012) [5] used data mining to discover association, classification, clustering and outlier detection. The algorithms (i.e., Rule Induction, Naïve Bayes, and K-mean) were used to predict grades of students (Naïve Bayes with an accuracy of 67.50%). The research of Qasem A. Al-Radaideh, Emad Al-Shawakfa, Mustafa I. Al-Najjar (2006) [6] included algorithms (i.e. ID3, C4.5, Naïve Bayes, hold-out and k-cross-validation) to evaluate student data for distinguishing student performance.

3. Proposed Work

The study of the problem and the whole process of the research is implemented following the Data Science Process. It is chosen because it provides all the basic steps needed for predicting the Students' Academic Performance on the basis of their grades.

3.1 Research Goal

To understand the problem more formally, a literature review has been performed to study the existing work and related studies. There are many researchers who analyzed the performance of students. Most of them worked on the effects of variables like students' attendance, personal efforts, time for studies etc. on their performance. Some of them also considered the current CGPA of students but none of them has taken student's grade in individual courses as an attribute for the study.

The primary goal of higher education institutions is to provide quality education to its students. A good prediction of the students' performance is helpful to identify the low performance students at the beginning of the learning process. One way to achieve this objective is by discovering knowledge for prediction regarding academic performance of students. This process will be done by using Machine Learning and Data Science tools and techniques. The intention behind the model is identification and extraction of potentially valuable knowledge for foreseeing poor and good performances in different courses.

The main objective of the research is to predict the future grades of the students using their grades in past and grades of the other graduated students so as to lower the dropout rate and inform students about their expected grades and performance in the course.

3.2 Data Collection

The data required is collected from FAST – National University of Computer and Emerging Sciences. The administration of the university provided the data of 3 years initially which was encoded by them for each student to ensure the anonymity of students' grades and data confidentiality. The data provided has 13,956 rows and 11 columns. Each row has the details of a student registration in a course. It contains details for bachelor's degree in Computer Science and Electrical Engineering.

The attributes of the data collected are the following:

- **Roll No:** The students' identification number in the university record.
- **Program:** The degree program, student is enrolled in.
- **Course:** It is the course code of the course, student is registered in.
- **Course Title:** It is the title of the course, student is enrolled in.
- **Credit Hours:** The credit hour of the course, student has to complete per week.
- **GP:** The grading points, student has scored in the course.
- **Grade:** The grade of the student in the course.
- **Sem.:** The year and semester, the course was studied in.
- **Batch:** The batch (year) of student's enrollment.

The data collected is very redundant and observations of the students are very scattered in raw data. After cleaning and transformation of data, the final data contains only 500 rows for applying the classifiers on. The data is sufficient for Naïve Bayes Algorithm but only 500 observations were not enough for Neural Networks Classifier. Therefore, to apply the classifiers and make better predictions from the past data, more data about the students' past grades and registration information was collected and the new data contains 34,476 rows. The structure of the data is same but it has different records from the previously collected data. After transformation and cleaning this data resulted into the data of 839 students for their grades in different courses.

3.3 Data Preparation

As discussed above, the data acquired contains redundancy and useless information that is not needed for the predictions and modelling. Furthermore, the structure of the data is also required to change because it can't be used as an input for classification models.

3.3.1 Data Integration

Combining the data was required for Neural Network and SVM Classifier. Microsoft Excel and its tools were used to simply integrate two datasets into a single dataset.

3.3.2 Data Transformation

As discussed above, the data acquired is not in the required form and is not structurally valid for the modelling. Let's discuss it more formally, the information required for the single row of the student is

scattered in the whole data set. For the modelling process, each student must be in a single row with all its courses in the same record and in the data acquired each student, course pair (i.e., Registration) is a single separate record. This transformation is done using PowerPivot, an external library of Microsoft Excel for data analysis, modelling, cleaning and transformation. The Tables - 1 & 2 will further illustrate the process of transformation.

Table 1- Data before Transformation

Roll No.	Course Title	Grade
1	Computer Programming	3.33
2	Data Structures	2.33
1	Discrete Structures	2.00
1	Data Structures	2.67
2	Computer Programming	3.33

Table 2- Data after Transformation

Roll No	Computer Programming	Data Structures	Discrete Structures
1	3.33	2.67	2.00
2	3.33	2.33	

After this transformation, the appropriate course columns are selected for the modelling process.

3.3.3 Data Cleaning

The data we now have after transformation has many missing values and some other errors. Given below is the checklist of the whole cleansing process step by step.

Missing Values

Missing Values are found in the data set after the data transformation (you can see missing values in Table 3). The causes of these students with the missing grade in a course are:

1. They are from the other Degree Program (i.e., Electrical Engineering)
2. They haven't completed their degree or they are university dropouts.
3. The course selected is renamed or is discarded from the list of offered courses in the later semesters

For the case 3, above the observations were merged with the renamed courses. For Case 1 and 2 the observations are removed, if they haven't completed the chain of courses or if they are from another degree program.

3.3.4 Exploratory Data Analysis

Exploratory data analysis is necessary because information becomes much easier to grasp when shown in a picture. This phase is about exploring data and the goal is not just to clean the data but to discover the irregularity or exceptions that were missed before.

Graphs, Plots and Charts are used to explore data and are combined so that they can provide even more insights.

3.4 Building Model and Evaluation

3.4.1 Model Selection

The methods are selected on the basis that they classify the students into four categories; depending on the previous university academic performances.

The models selected are:

1. Naïve Bayes:

In machine learning, Naïve Bayes is a classifier based on probabilities, applying Bayes' theorem with strong (naïve) independence assumptions between the features. Naïve Bayes Classifier is selected because the future grades of the particular course chain are required to be predicted if the details of previous courses/subjects are provided. Naïve Bayes Classifier works best in these "given that" and conditional situations.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

The diagram shows the formula with arrows pointing to its components:
 - $P(c|x)$ is labeled "Posterior Probability"
 - $P(x|c)$ is labeled "Likelihood"
 - $P(c)$ is labeled "Class Prior Probability"
 - $P(x)$ is labeled "Predictor Prior Probability"

Chain predictions for Database course can be solved by applying Naïve Bayes as follows;

$$P(DB|ITC,CPDS) = \frac{P(ITC,CPDS|DB)P(DB)}{P(ITC,CPDS)}$$

$$P(ITC,CPDS)$$

$$P(DB|ITC,CPDS) = \frac{P(ITC|DB)P(CP|DB)P(DS|DB)P(DB)}{P(ITC,CPDS)}$$

2. Neural Networks:

Neural networks in machine learning are based upon perceptron, which takes multiple binary inputs and produces a single binary output. The inputs have weights which represent the importance of the respective inputs. This leads to how to weigh different kind of attributes or knowledge in order to make decisions. Neural Networks is selected because the input grades can be weighed and the importance of the pre-requisite courses could be defined easily which would affect the course output to be precise.

Input layer of neurons, there are activation functions; Sigmoid, Tanh – hyperbolic tangent, ReLu - rectified linear units. ReLu is used; it can only be applied within the hidden layers of the neural network model, so the output layer uses Softmax function for computing probabilities of classes.

3. Support Vector Machine:

Support Vector Machine comprises of the concept of planes. A plane will have sets of different instances belonging to different classes. It has different classifying techniques. The simplest one is linear, which separates sets of instances into their groups with just a line. For more accurate classification, more complex structures are used. In our classification model, we have used different kernel functions; which are linear and RBF.

4. Decision tree classifier:

Decision tree is defined same as the name describes; it is a tree-like structure, with attributes distributed as root till leaf nodes. It has several branches with varying attributes. Each leaf node (end node) represents a class (category or output variable). The model is based on C4.5 decision tree algorithm, and builds a decision tree from the training dataset. It uses the concept of information entropy.

3.4.2 Variable Selection

There are several techniques used in variable selection for checking different criteria. The most used technique is course chain technique. These techniques are:

1. Course Chain Technique:

Course Chain is a chain of courses that are arranged in a series where a previous course is pre-requisite of next course(s) or predominately related to it. In this case, studying the next course highly recommends passing the previous course of the chain and having good knowledge of the field related to the course. Hence, there is a dependency between those courses and it is very probable that the grade obtained in CS Course-2 would be close to the grade obtained in CS Course-1.

So, the variables selected in this case are grades of all the member courses in the class which will be classified into grades of the last course of chain.

2. Previous Semester to Next Course Technique:

This technique is used for predicting the grade of any course from the next semester having the grades of the courses in the next semester as predictors. It is used because we don't have enough courses in the chain for the fresh students because they have not enough courses of a chain to get better predictions.

3. Prerequisite Pass Check:

This technique is used to predict whether a student will pass a course or not based on the grades of the prerequisite course of the said course. Grade of prerequisite course is used as attribute while current course is the class.

4. Experimental Setup

4.1 Software Tools

The software tools used for this research consist of Microsoft Excel is used for data preparation and cleaning. The models are created by using Python 3.0 with NumPy and Pandas. The platform used is Anaconda3. TensorFlow and Keras is used while implementing Neural Networks model.

4.2 Dataset

4.2.1 Description

The data is collected from computer science department of FAST-NUCES which consists of students' transcript data that included their grades in all courses. The dataset consisted of students from the 3 academic years. After eliminating incomplete data, the samples consisted of 839 students. Then after Students' assessments in the academic year are taken as output variable are grouped as:

Students' final grades divided into four classes:

Table 3- Class Labels

Class	Grade
1	A+, A, A-
2	B+, B, B-
3	C+, C, C-
4	D

The data required is collected from FAST-National University of Computer and Emerging Sciences. The administration of the university provided the data of 3 years. The data provided has 13,956 rows and 11 columns. Each row has the details of a student registration in a course.

The attributes of the data collected are the following:

- **Roll No:** The students' identification number in the university record.
- **Program:** The degree program, student is enrolled in.
- **Course:** It is the course code of the course, student is registered in.
- **Course Title:** It is the title of the course, student is enrolled in.
- **Credit Hours:** The credit hour of the course, student has to complete per week.
- **GP:** The grading points, student has scored in the course.
- **Grade:** The grade of the student in the course.
- **Sem.:** The year and semester, the course was studied in.
- **Batch:** The batch (year) of student's enrollment.

To apply the classifiers and make better predictions from the past data, more data about the students' past grades and registration information was collected and the new data contains 34,476 rows. After transformation and cleaning this data resulted into the data of 839 students for their grades in different courses.

The total instances in each course chain categorized by the class are shown in the figure below;

5. Results

The main objective is to predict the class variable by using the input variables. In this study, the input variables consist of grades in different individual courses. Several models have been created using different classification techniques. To get better results and predictions, we analyzed the impact of individual courses on courses being offered in next semesters. Chains of different courses are used to get improved predictions. Different algorithms have given different varying results. Each classifier is applied for testing methods – k-fold cross validation; (using 10–20 folds applying the algorithm k times) and – percentage split (70% training, 30% testing data).

Furthermore, the classifiers are evaluated by certain measures which are; recall, precision and f-measure score for better knowledge about the performance of models created. In the end, the performances of the models are compared and evaluated on the basis of these measures which are illustrated in the next section.

The results obtained with those algorithms are summarized below;

5.1 Naïve Bayes classifier

Naïve Bayes classifier predicts classes by probabilities, it applies Bayes' theorem with strong (naïve) independence assumptions between the features. Naïve Bayes is applied with considering joint probabilities and is tested with k-fold cross validation method.

The achieved results are presented below;

Semester 1 to Semester 2:

In this method, all courses of semester 1 were taken as input and the predicted course is one i.e., Computer Programming.

Table 4 - Naive Bayes classifier results

Method	Prediction	Variations	Accuracy
Semester 1 to Semester 2	Computer Programming	-	22% - 25%

The results in Table – 4 illustrates that Naive Bayes model gives weak accuracy (22%- 25%) for prediction of a single course from all the grades of previous semester performance. Hence, for more accurate prediction, we took the chains of courses which have more dependency on the course to be predicted.

Course Chaining:

To improve the prediction accuracy, certain chains for predicting are made. By using chaining method, we predict different chains of courses with different variations of testing methods. The chains only contain the courses which directly affect the course to be predicted.

Method	Prediction	Accuracy	Precision	Recall
Chain Prediction	Technical & Business Writing	59.4%	41.18	41.608
Chain Prediction	Probability & Statistics	44.20%	-	-
Chain Prediction	Database Systems	60.56%	-	-
Chain Prediction	Computer Architecture	45.47%	-	-

In addition, the model is evaluated by different testing methods which are summarized below;

Database Systems:

The Database Systems chain contains Introduction to Computer Science, Computer Programming, and Data Structures as input variables to predict the class of Database Systems. Naïve Bayes is used with the testing methods; 20-Fold cross validation, weighted featured testing and a combination of both.

Method	Prediction	Variations	Accuracy
Chain Prediction	Database Systems	-	60.56%
Chain Prediction	Database Systems	20-Fold	49.71% (mean)
Chain Prediction	Database Systems	Weighted Features	65% (mean)
Chain Prediction	Database Systems	Weighted Features with 20-Fold	49% (mean)

Computer Architecture:

Computer Architecture chain consists of the following courses; Physics, Digital Logic Design, Comp. Organization & Assembly Lang, and Operating Systems. The grades in these courses are taken as input variables. The testing variations applied are 20-Fold cross validation, weighted featured testing.

Method	Prediction	Variations	Accuracy
Chain Prediction	Computer Architecture	-	45.47%
Chain Prediction	Computer Architecture	20-Fold	53% (mean)
Chain Prediction	Computer Architecture	Weighted Features with 20-Fold	47% (mean)

Probability & Statistics:

Probability & Statistics chain consists of the following courses; Calculus-I, Calculus – II and Linear Algebra. The grades in these courses are taken as input variables. The testing variations applied are 20-Fold cross validation, weighted featured testing.

Method	Prediction	Variations	Accuracy
Chain Prediction	Probability & Statistics	-	44.20%
Chain Prediction	Probability & Statistics	Weighted Features	54% (mean)
Chain Prediction	Probability & Statistics	Weighted Features with 20-Fold	45% (mean)

Technical & Business Writing:

Technical & Business Writing chain consists of the following courses; English-I and English-II. The grades in these courses are taken as input variables. The testing variations applied are 20-Fold cross validation, weighted featured testing.

Method	Prediction	Variations	Accuracy
Chain Prediction	Technical & Business Writing	-	59.4%
Chain Prediction	Technical & Business Writing	Weighted Features	68% (mean)

Evaluation measures:

The evaluation measures (i.e., recall, precision, f-score) are shown in Table- 5, which are obtained by applying Naïve Bayes along with 70-30 cross validation method. They are divided for all the classes for different course chains.

Table 5- Naive Bayes Results

Class	Naïve Bayes					
	TBW Chain			MT Chain		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	0.54	0.57	0.55	0.58	0.50	0.54
2	0.61	0.65	0.63	0.45	0.56	0.50
3	0.49	0.44	0.44	0.46	0.44	0.45
4	0.00	0.00	0.00	0.25	0.16	0.20
Average Accuracy	0.57	0.55	0.56	0.46	0.46	0.46

The detailed results show the recall rate, precision and f-measure distributed by classes. It shows that precision rate for class 1, 2 and 3 are fine whereas class 4 has zero precision and recall rate for TBW Chain. In MT chain, class 1 has the highest precision rate. The TBW chain has better precision than MT Chain.

5.2 Neural Network classifier

A model implementing Neural Networks with multilayer perceptron is applied to the course chains. It takes the courses as input for the first layer and predicts the class of the last course in the chain. The results (i.e., accuracy, precision and recall) of the neural net-work model for the course chains are summarized in the table below;

Method	Prediction	Accuracy	Precision	Recall
Chain Prediction	Technical & Business Writing	58.55%	0.344	0.32
Chain Prediction	Database Systems	49.33%	-	-
Chain Prediction	Probability & Statistics	48.5%	0.6758	0.4345
Chain Prediction	Computer Architecture	47.33%	-	-

Evaluation measures:

The evaluation measures (i.e., recall, precision, f-score) are shown in Table-6, which are obtained by applying Neural Networks along with 70-30 cross validation method.

Table 6- Neural Network Results

Class	Neural Networks					
	TBW Chain			MT Chain		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	0.58	0.39	0.47	0.79	0.38	0.51
2	0.60	0.81	0.69	0.59	0.59	0.55
3	0.57	0.37	0.45	0.67	0.67	0.50
4	0.00	0.00	0.00	1.00	0.11	0.20
Average Accuracy	0.57	0.59	0.56	0.67	0.49	0.47

The detailed results show the recall rate and the precision distributed by classes. It shows that precision rate for class 1, 2 and 3 are fine whereas class 4 has zero precision and recall rate for TBW Chain. In Mt chain, class 4 has the highest precision rate.

5.3 Support Vector Machine

Support Vector Machines is applied along with 70-30 cross validation method, the results obtained for different chains of courses are illustrated in the Table-7;

Table 7- SVM results

Method	Prediction	Variation	Accuracy	Precision	Recall
Chain Prediction	Technical & Business Writing	-	60.78%	0.4464	0.4155
Chain Prediction	Probability & Statistics	-	50.22%	0.383	0.430
Chain Prediction	Database Systems	20-Fold	44.12%	-	-
Chain Prediction	Computer Architecture	20-Fold	42.51%	-	-

Technical & Business Writing chain shows good accuracy of 60.78% whereas Computer Architecture chain has lowest accuracy.

The evaluation measures (i.e., recall, precision, f-score) are shown in Table-8;

Table 8- Support Vector Machine Results

Class	Support Vector Machine					
	TBW Chain			MT Chain		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	0.59	0.54	0.56	0.54	0.48	0.54
2	0.62	0.77	0.69	0.52	0.54	0.48
3	0.58	0.35	0.44	0.48	0.39	0.38
4	0.00	0.00	0.00	0.00	0.11	0.14
Average Accuracy	0.59	0.61	0.59	0.50	0.43	0.42

The detailed results show the recall rate and the precision distributed by classes. It shows that precision rate for class 1, 2 and 3 are fine whereas class 4 has zero precision and recall rate. The TBW chain has better precision and recall than MT Chain.

5.4 Decision Tree classifier

Decision tree classifier is of different kinds, in this research C4.5 classifier is applied on the dataset by using information entropy. The dataset is split into training and testing dataset. To evaluate this classifier, 70-30 cross validation method is executed on it. 70% data is taken as training dataset and 30% data is taken as testing data.

The achieved results are presented below;

Method	Prediction	Accuracy	Precision	Recall
Chain Prediction	Technical & Business Writing	51.63%	0.37	0.52
Chain Prediction	Probability & Statistics	42.66%	0.421	0.377
Chain Prediction	Database Systems	37.36%	-	-

TBW chain has the highest accuracy in the above accuracies.

The evaluation measures are illustrated in Table-9;

Table 9- Decision tree Results

Class	Decision tree					
	TBW Chain			MT Chain		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	0.00	0.00	0.00	0.61	0.48	0.54
2	0.53	0.89	0.66	0.44	0.54	0.48
3	0.45	0.27	0.34	0.37	0.39	0.38
4	0.00	0.00	0.00	0.19	0.11	0.14
Average Accuracy	0.37	0.52	0.41	0.42	0.43	0.42

The detailed results show the recall rate i.e., the proportion of examples which were classified correctly and the precision distributed by classes. It shows that Precision rate for class 2 and 3 is high whereas class 1 and 4 have zero precision rate. The Probability & Statistics chain (MT) has better precision than TBW chain whereas recall rate is much better of technical & Business writing chain.

6. Discussion Of Results

The results for the performance prediction accuracy of the selected algorithms are summarized in the figure below;

Table 10- Prediction Accuracies of classifiers

Classifier	Prediction Accuracy
Naïve Bayes	60.01%
Neural Network	50.92%
SVM	49.40%
C4.5	43.88%

The results achieved in the Table-10 and the above figure illustrate that Naïve Bayes classifier performs with the highest overall accuracy, followed by Neural Networks, SVM and then Decision tree C4.5 with the least accuracy. However, all overall accuracies are below 70% and this shows that error rate is too high which results in unreliable predictions. The evaluation criteria based on precision and recall is compared and showed in the table below;

Table 11- Evaluation Measures

	Precision	Recall
Naïve Bayes	0.48	0.475
Neural Network	0.56	0.458
SVM	0.487	0.47
C4.5	0.54	0.39

The Table-11 shows that Neural Networks has highest precision rate among other classifiers, whereas Naïve Bayes has highest recall rate as compared to other models. The lowest precision and recall rate is given by decision tree C4.5. Neural Networks is overall more reliable in light of these measures.

Chains of different courses are used to evaluate predictions; Database Systems chain, Probability & Statistics Chain, Technical and Business Writing chain and Computer Architecture Chain. The total number of instances taken as inputs differ in each chain, due to difference in the number of students enrolled in the courses involved. Each classifier is applied on every chain. The overall results for each chain obtained by each model are shown in the next table;

Table 12- Prediction Accuracy of Chaining

Chains of courses	Prediction Accuracy			
	Naïve Bayes	Neural Network	SVM	C4.5
DB Chain	56.06%	49.33%	44.12%	37.36%
MT Chain	47.73%	48.5%	50.22%	42.66%
TBW Chain	63.7%	58.55%	60.78%	51.63%
CA Chain	48.5%	47.33%	42.51%	46.11%

In Table-12, Naïve Bayes has the highest accuracy for DB chain i.e., 56.06%, then Neural Networks, then SVM and then comes C4.5 with the lowest accuracy. For chain MT, highest accuracy result is obtained by SVM, then Neural Networks, Naïve Bayes and then c4.5 with the lowest. Naive Bayes shows highest accuracy for TBW chain (63.7%) and CA chain, whereas the lowest for both the chains is given by C4.5. CA chain has no accuracy result greater than 50% which shows that the error rate is high.

The algorithm Naïve Bayes is most reliable for all class predictions, with the highest accuracy for all chains except for MT chain. It is also observed that Neural Network gives relatively good results. C4.5 is less accurate than others having the lowest accuracy, recall and precision rates. The performance of the classifying models is evaluated on the basis of prediction accuracy, error, precision and recall rates. More accurate and interesting results will be obtained on further exploration of datasets. These models provide some insight for further student performance predictions, and provide guidance to teachers for choosing a suitable path for students.

7. Conclusion & Future Work

The results achieved reveal that the prediction rates are not very reliable. Furthermore, the results of classes vary from every classifier. The attributes only consist of grades in individual courses and has no knowledge and information about the semester progress, marks and time utilized by student in studying. These factors must have a bigger impact in predictions of future grades and performances. The future work will include possible transformations of dataset, obtaining more information about students' performance and extra activities etc. in order to achieve more accurate results. This will lead to extracting more meaningful information from the available data and providing reliable predictions.

8. References

- [1] Edin Osmanbegović, Mirza Suljić, (May 2012), "Data Mining Approach for Predicting Student Performance", Journal of Economics and Business, Vol. X, Issue 1
- [2] S. Kotsiantis, K. Patriarcheas, M. Xenos, (2010), "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education", Hellenic Open University, School of Sciences and Technology, Computer Science, Greece, Knowledge-based System – Elsevier
- [3] Paulo Cortez, Alice Silva (2008), "Using Data Mining to Predict Secondary School Student Performance", University of Minho, Portugal
- [4] Dorina Kabakchieva, (2013), "Predicting Student Performance by Using Data Mining Methods for Classification", Cybernetics and Information Technologies Volume 13, No 1
- [5] Mohammed M. Abu Tahir, Alaa M. El-Halees, (February 2012), "Mining Educational Data to improve Students' Performance: A Case Study", International Journal of Information and Communication Technology Research Volume 2 NO. 2
- [6] Qasem A. Al-Radaideh, Emad Al-Shawakfa, Mustafa I. Al-Najjar (2006), "Mining Student Data Using Decision Trees", The International Arab Conference on Information Technology ACIT'2006, Jordan
- [7] J. P. Vandamme, N. Meskens, J. F. Superby (December 2007), "Predicting Academic Performance by Data Mining Methods", Belgium, Education Economics Vol. 15, No, 4, pp 405-419

9. Conflict Of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.