Prediction of Students' Academic Performance Using Machine Learning: The Case of Wachemo University

Girma Asefa and Fikadu Wayesa, Wachemo University

Abstract

An educational institution must always have an estimated previous knowledge of enrolled students to predict their performance in future academics. The comparative analysis of results helped the weaker students to improve their passing which has eventually led to increased overall passing average of the course. In this research, the data was analyzed using classification models with mixing the internal and external data sources, and then compared the model to select the best prediction model that produced highest accuracy, which helped the university to identify the students likely to fail, and work on their academics accordingly in order to achieve better results. Accuracy of this classification algorithm is compared in order to check best performance. K-NN achieved the best prediction results, which are very satisfactory compared to those of similar approaches in both classification.

Keywords: Prediction, Machine Learning, Classification and Performance

1. Introduction

Recently, every educational institution handles and deals with a large amount of student data which can be useful for a number of reasons. One of the important applications of such data is predicting student performance. As Jabeen al et (2019) described, nowadays students can access education via various ways such as online education system, web-based education system, seminars and workshops. Education gained through these platforms are more challenging to predict students' performance because of the huge bulk of data stored in the environments of educational databases and learning management databases. On the other hand, Witten et al (2016), explained in their study the significance of machine learning covers a wide range of application areas, such as search engines use machine learning to better construct relations between search phrases and web pages by analyzing the content of the websites; search engines can define which words and phrases are the most important in defining a certain web page, and they can use this information to return the most relevant results for a given search phrase. As Murat (2017) explained, computer processes data and learns to identify this data, and then uses this knowledge to make decisions about future data. Machine learning is a field that focuses on the prediction of patterns and hidden knowledge detection in massive data and gives the information in a reasonable form.

1.1 Statement of the Problem

High quality of education is something that everyone expects from lower level class to higher education institutions, both in public and private schools, because those who failed to perform better cannot contribute well to the human resources, as they do not have the needed skill. In educational environment, the amount of data collected in electronic format has seen a dramatic increase currently. The task to manage large amount of data and determine the relationships among variables in the data is not an easy task to be done manually. Machine learning is a field that focuses on the prediction of patterns and hidden knowledge detection in massive data and gives the information in a reasonable form. Because of its importance, machine learning initiate many data quarrying project to predict the effectiveness of students in education performance.

The prediction of success in tertiary institutions is one of the most vital issues in higher education (Golding and Donaldson, 2006), and no common predictors accurately determine whether a student will be an academic genius, a drop out, or an average performer. The task to develop effective predictors of academic success is a critical issue for educators (Golding and Donaldson, 2006). Previous studies revealed that various factors are responsible for scholastic failure of students, such as low socio-economic background, students' cognitive abilities, school related factors, home environment, or the support given by the parents and other family members (Chohan and Khan, 2010)

Most of the studies considered factors like demographic factors, academic marks as their basis of prediction to build different models by using academic and non-academic data, and build the model for both, and finally compare the two models to see the effect of the factors on the students' academic performance. However, in this study a special focus has been given to a mixed approach i.e., from academic and non-academic related students' data to improve the performance accuracy model.

Accordingly, we tried to answer the following research questions:

- ☑ Which factors and group of factors are significant in achieving the most significant prediction analysis that would help to improve course performance of the students in Wachemo University?
- ☑ Which Machine-learning model predicts students' academic performance with precise and accurate classification?
- ☑ How can the discovered knowledge from academic data aid decision makers to improve decision-making processes in wachemo university?

The general objective of the study is to predict the performance of students by using machine learning approach known as classification techniques in order to assist and improve performance of the low academic achievers in a given course as well as to increase the passing average in the newly introduced courses. According to (Baradwaj & Pal, 2012), in any educational institution, it is authoritative to record the past data of students which help the institutions to make a quick and automatic predictions of the level of students and also to identify the drop out students who are in need of extra academic attention and training from their respective teachers in any subjects.

2. Literature Review

Classification of Machine Learning

According to D. Michie et al (1994), machine learning is generally taken to encompass automatic computing procedures based on logical or binary operations that learn a task from a series of dataset. Here we are just concerned with classification, and it is arguable what should come under the machine learning umbrella. As per MI. Jordan (2014), the ideas of machine learning, from methodological principles to theoretical tools, have had a long pre-history in statistics. He also suggested the term data science as a placeholder to call the overall field data model and algorithmic model, where in algorithmic model, means more or less the machine learning algorithms like Random forest. Some statisticians have adopted methods from machine learning, leading to a combined field that they call statistical learning (Gareth, at .et, 2013). It appears in many machine learning applications such as fraud detection, network intrusion detection, oil-spill detection, disease diagnosis, and many other areas (Chawla et al., 2004; He & Garcia, 2009). Most classifiers in supervised machine learning are designed to maximize the accuracy of their models. Thus, when learning from imbalanced data, they are usually overwhelmed by the majority class examples. This is the main cause for the performance degradation of such classifiers, and is also considered as one of the ten challenging problems in data mining research (Yang & Wu, 2006)**.** From an educational machine learning point of view, an accurate and reliable model in predicting student performance may replace some current standardized tests and thus, reducing the pressure, time, as well as effort on teaching and learning for examinations (Feng, 2009).

Decision Trees

Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclude about the items target value. It is one of the predictive modelling approaches used in statistics, data mining and machine learning. Tree models, where the target variable can take a finite set of values, are called classification trees. In these tree structures, leaves represent class labels, and branches represent conjunctions of features that lead to those class labels. Decision tree learning is a method commonly used in machine learning. The goal is to create a model that predicts the value of a target variable based on several input variables. Each interior node corresponds to one of the input variables; there are edges to children for each of the possible values of that input variable. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf. Decision tree learning is one of the most successful techniques for supervised classification learning. For this section, we assumed that all of the features have finite discrete domains, and there is a single target feature called the classification. Each element of the domain of the classification is called a class.
As explained in Chapman & Hall/CRC book (p.27), one of the biggest advantages of a decision tree is that they are easy to use and, may be even more important; they are also easy to understand even by non-experts. The basic idea behind decision trees is a so-called divide-and-conquer approach. In each step the dataset is divided into different parts while each part should better represent one of the possible classes.
We will get one subset for each branch extending from the node; we will now discuss how we can measure information at all in order to be able to calculate which decision delivers the

highest gain in information. We will see now that information can be easily measured in bits and that a concept exists for calculating this amount, the entropy. The entropy measures the information required in bits (can also mean fractions of bits) to predict the outcome of an event if the probability distribution of this event is given. The formula is as follows: Entropy (p1. . . pn) = -p1 ld (p1) - . . . – pn ld (pn) P1 to pn are the possible probabilities for all outcomes while ld is the logarithm with base

Decision tree Metrics

Algorithms for constructing decision trees usually work top-down, by choosing a variable at each step that best splits the set of items. As per Rokach (2005), different algorithms use different metrics for measuring best. These metrics are applied to each candidate subset, and the resulting values are combined to provide a measure of the quality of the split. Gini impurity can be computed by summing the probability of each item being chosen times the probability of a mistake in categorizing that item. It reaches its minimum (zero) when all cases in the node fall into a single target category. Information gain or IG is a statistical property that measures how well a given attribute separates the training examples according to their target classification. Constructing a decision tree is all about finding an attribute that returns the highest information gain and the smallest entropy.

$$\text{Information Gain}(T,X) = \text{Entropy}(T) - \text{Entropy}(T, X)$$

Where **T→ Current state and X → Selected attribute**

Information gain is used by the ID3, C4.5 and C5.0 tree-generation algorithms. Information gain is based on the concept of entropy from information theory.

Naive Bayes classifier

Naive Bayes classification algorithm and its operator in Rapid Miner.

The use case of this section applies that the Naive Bayes operator on different dataset. The operators explained in this part are: Rename by Replacing, Filter Examples, and Discretize by Binning, and Performance (Binominal Classification) operator. The Naive Bayes algorithm is a simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. In simple words, the Naive Bayes algorithm assumes that the presence of a particular value of an attribute is unrelated to the presence of any other attribute values.

Let X be an example that we want to classify. X is described by measurements made on a set of n attributes. Let H be some hypothesis, such as that the example X belongs to a specified class C. For classification problems, we want to determine P(H/X), the probability that the hypothesis H holds given the example X. In other words, we are looking for the probability that example X belongs to class C, given that we know the attribute description of X.

In contrast, P(H) is the prior probability, or a priori probability, of H. For our example, this is the probability that students will pass in any given course, regardless of mark, and parent background, or any other information. The posterior probability, P(H/X), is based on

more information, i.e., students information, than the prior probability, P (H), which is independent of X. Similarly, P(X/H) is the posterior probability of X conditioned on H. That is, it is the probability that the students X is family income is high and mark 80, given that we know that student will pass.
Bayes' theorem is useful in that it provides a way of calculating the posterior probability

P (H/X), from P (H), P(X/H), and P(X). Bayes' theorem states: P (H/X) = P(X/H). P (H) / P(X).

This expert parameter indicates that the Chapman & Hall/CRC book (p. 27) Laplace correction should be used to prevent high influence of zero probabilities. To avoid zero probabilities, it can be assumed that our training dataset is so large that the addition of one to each count that we need would only make a negligible difference in the estimated probabilities (Chapman & Hall/CRC book (p.54). This technique is known as the Laplace correction

K-Nearest Neighbor Classification

In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. 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 k-NN algorithm is among the simplest of all machine learning algorithms.
Both for classification and regression, it can be useful to assign weight to the contributions of the neighbors so that the nearer neighbors contribute more to the average than the more distant ones. A short coming of the k-NN algorithm is that it is sensitive to the local structure of the data. The algorithm has nothing to do with and is not to be confused with k-means, another popular machine learning technique
Jaskowiak (2014) reported that the classification accuracy of k-NN can be improved significantly if the distance metric is learned with specialized algorithms such as Large Margin Nearest Neighbor or Neighborhood components analysis. The neighbors are taken from a set of examples for which the correct classification or, in the case of regression, the value of the label is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The basic k-Nearest Neighbor algorithm is composed of two steps, (a) Find the k training examples that are closest to the unseen example and (b),take the most commonly occurring classification for these k take the average of these k label values and assign this value as the label of this unseen example.

Predicting Performance for Students

As Maria and Katia (2011) explained the ability to timely predict the academic performance tendency of undergraduate students is very important in degree programs and useful for tutors. Moreover, the computer can render substantial assistance to cognitive psychology, in that it may be used to test the consistency and completeness of learning theories and enforce a commitment to the fine-structure process level detail that precludes meaningless tautological or untestable theories (Bishop, 2006). Apart from this, detecting excellent students can be very useful information for institutions for allocating scholarships. Even from the very beginning of an academic year, by using students' demographic data, the groups that might be at risk can be detected (*L. Rokach*, 2004). Jaime et .al (1983) stated that computer tutoring is starting to incorporate abilities to infer models of student competence from observed performance. Inferring the scope of a student's knowledge and skills in a particular area allows much more effective and individualized tutoring of the student (Sleeman, 1983).

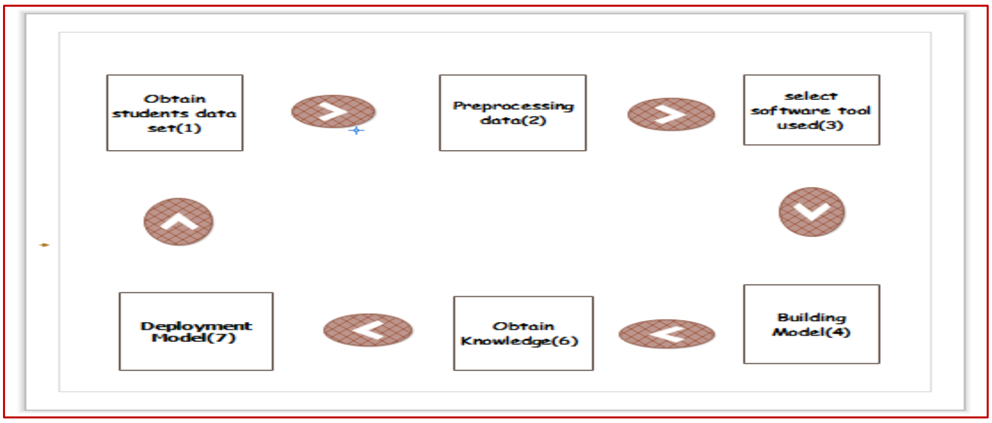
3. Research Design and Method

The knowledge discovery process design consists various strategies as indicated in the literature review and used to provide a good coverage starting from problem domain, data pre-processing, analysis, inferences drawn, and implementation. An academic national education curriculum in an institution generally defines a specialized learning plan and schedule which put some sort of restrictions on how the students should study the courses and many universities adopt the grading system to estimate and decide the performance of students in academics. Hence this study used Hybrid data, and it provides detailed description on the research-orientation, the data quarrying steps, and combines the industrial and academic models. The model steps include: understanding the problem domain, understanding the data, preparation of the data, data quarrying, evaluation of the discovered knowledge, and finally use of the discovered knowledge.

3.1 Research Design

The model steps are understanding the problem domain, understanding the data, preparation of the data, data quarrying, evaluation of the discovered knowledge, and finally use of the discovered knowledge.

Figure 1: Research Design



3.2 Data and Data sources

In this study, we considered two types of variables: **(a)** variables from institutional internal data sources (academic data), and **(b)** variables out of the institutional external (general students) data sources. Wachemo University Registrar was the data source to predict students' academic performance.

Some data which is not obtained from the registrar but very important for the research were collected from the students who were selected by using the questionnaires that were built on Google Forms over the Internet. An open source application was used to conduct a survey on students from the University students. A total of 1986 questionnaires were completed after combining the CSV files from Google Forms and the Lime Survey questionnaire.

3.3 Problem Domain Identification

In this study, discussion was made with domain experts to understand the problem domain and to have a good understanding about the initial datasets. The domain experts were performance prediction expert, data analysts, computer programmer and statisticians who processed the prediction process to provide statistical information for information of student's performances.

3.4 Preparation of the Data

Data preparation tasks were performed in iterative ways. The major tasks include: description of data sources, carrying out statistical summary measure, finding out distinct value, filling missing values, outlier and noisy data and data transformation/ or reduction activities were also undertaken in this phase. To support the preprocessing tasks, the researcher used Rapid Miner to normalize and to fill missing value with mode value.

3.5 Attribute Description for the Study

Table 1: Questionnaire Description

No.	Attribute	Description	Possible Values
1	Address	Student's home address type	{Urban or Rural}
2	Famsize	Family size	LE3 or GT3
3	Pstatus	Parent's cohabitation status	{ living together or Apart}
4	Medu	Mother's education	{PhD , Diploma , MSc , Fdegree , None }
5	Fedu	Father's education	{ PhD , Diploma , MSc , Fdegree , None }
6	SportA	You participate on any sport activities	{YES,NO}
7	Mjob	Mother's job	{ Teacher , AnyotherService , Cservice , Health , Home , health }
8	Fjob	Father's job	{ Teacher , AnyotherService,Cservice , Health , Home }
9	Reason	Reason to choose this your dept	{Cpreference, Fpressure , Spressure }
10	Guardian	Student's guardian	{ 'Mother', 'Father' or 'other'}
11	studyhours	Weekly study hours	{2 to 5 Hr, 5 to 10 Hr , <Less-2Hr , GR-10Hr }

12	Failures	Number of past class failures	{n if 1<=N<3, else above }
13	Pactivity	you support any political activities	{YES OR NO}
14	Famsup	Family educational support	{YES OR NO}
15	Expaid	Extra paid classes within the others courses subject	{YES OR NO}
16	Activities	Extra-curricular activities	{YES OR NO}
17	Internet	Internet access at University	{YES OR NO}
18	Romantic	With a romantic relationship	{YES OR NO}
19	Famrel	Quality of family relationships	{ Verygood, Excellent, Good , Verybad, Bad }
20	Freetime	Free time after class	{ Medium, Low, High, VeryLow, VeryHigh }
21	Goout	Going out with friends	{ Once, Threetimes, Twice, Fourtimes, Fivetimes }
22	Walc	Weekend alcohol consumption	{ Once, Threetimes, Twice, Fourtimes, Fivetimes }
23	Health	Current health status	{From 1 - very bad to 5 - very good}
24	Absences	Number of class absences	{>=0}
25	***SSTATU***	***Students status Descriptions***	***{Pass ,Fail}***

3.6 Reliability

A measure is said to have a high level of reliability if it gives similar results under consistent conditions. In order to measure the consistency of the variables, there are several methods for calculating a correlation value. The most popular one is weight, the overall reliability of constructs that it is measuring. A correlation test was carried out based on weight of each attributes. Statisticians and data analysts' measured correlation of two numerical variables to find an insight about their relationships. Therefore, the reliability test by using Correlation Matrix accomplished that some the items of the pilot variables have been reliable since the scores of the test were higher than 0.7 as in the table below; so the variables are concluded more related.

Table 2: Correlation matrix on the variables

No	Attributes	Weight
1	**Address of the students**	**0.947**
2	**Family Size of the students**	**0.962**
3	**Parent status**	**0.959**
4	**Mother education**	**0.959**
5	**Father education**	**0.959**
6	**Mother Job**	**0.968**
7	**Father job**	**0.954**
8	**Reason of choice Dept.**	**0.981**

9	**Guardian of student**	**0.944**
10	**Study time per day**	**0.932**
11	**Numbers of in the previous failure**	**0.924**
12	**School supporting**	**0.954**
13	**Family supporting**	**0.966**
14	**Any extra Paid**	**0.953**
15	**Political activity**	**0.946**
16	**Sport Activities**	**0.958**
17	**Attending first class**	**0.916**
18	**Internet using**	**0.954**
19	**Romantic following**	**0.9462**
20	**Family relationship**	**0.959**
21	**Free time students have per a day**	**0.944**
22	**Going out With Friends**	**0.934**
23	**Health**	**0.959**
24	**Absence**	**0.927**

Source: from the Authors experiment Result 19 Nov 2020

4. Results and Analysis

Powerful Machine Learning Algorithms such as decision tree, Naive Bayes and K-Nearest Neighbor algorithms are the classification techniques used to learn a classification function, which decide the value of dependent variable (class) based on the given values of independent variables.

Accuracy Decision Tree Model (Target=>Sstatus)

"Which Machine-learning model predicts students' academic performance most precise and accurate classification?"

The performance metrics used for the experiment is given below. To do this, the data was pre-processed, and gain ratio was used as partition method to build a decision tree. Accuracy is how close a measured value is to the true value and it expresses the correctness of a measurement and determines by absolute and comparative way. Some of the terms used in accuracy calculation are given below:

- ☑ True positives (TP)
- ☑ False positives

- ☑ True negatives (TN)
- ☑ False negatives (FN)

Table 3: Performance table for model with mixed data features using Decision Tree

Status	True FAIL	True PASS	Class Precision
pred. FAIL	201	14	93.49%
pred. PASS	297	1474	83.23%
Class Recall	40.36%	99.06%	

Source: Computation based on data from authors field work, 19 Nov 2020

In this study, the experiment was done to determine how many students would pass or fail within the University in the given courses. The value 'pass' is positive class and 'fail' is negative class. The data set consists of 1986 records of new students, which were used by decision tree algorithm in classifying the results by using the students general attributes. In the first row, in the true pass column of the confusion matrix, 201 students were classified correctly as fail (predicted to fail in the course) and 14 students were classified incorrect; second row in the true column of the confusion conditions 1474 students were classified as positive (predicted to pass) as well as 297 are classified incorrect.

The performance metric of decision tree in the above table shows only an accuracy of 84.34%, where 1474 students were classified as pass, and 297 fail which implies that the model predicts approximately 15.66% of incorrect results when a model is generated based on only students' general data. In addition, class precision for true pass is 83.23% and 93.49% for true fail as well as recall of true pass is 99.06% and 40.36% is true fail.

Accurate

$$\boxed{Accuracy = \frac{Sum\ of\ true\ positives + Sum\ of\ true\ negatives}{Total\ population}}$$

Accuracy = $(\frac{201+1474}{1986})*100$=**<u>84.34%</u>**

As seen in the table, the model shows an accuracy of accuracy: 84.34% with a margin of error (15.66%).

Classification Error

Relative number of misclassified examples or, in other words, percentage of incorrect predictions:

$$\boxed{Classification\ Error = \frac{Sum\ of\ false\ positives + Sum\ of\ false\ negatives}{Total\ population}}$$

$$\textit{Classification Error} = \frac{14+297}{1986} * 100 = 15.66\%$$

Weighted Mean Recall

The weighted mean of all per class recall measurements is calculated through class recalls for individual classes.

The percentage of cases in which Rapid Miner predicted pass (99.06%), with a margin error on 0.94% is acceptable. The recall measure answers the question, "If a student is going to pass, what is the likelihood that the model will sense it?" The answer is (99.06%). On the other side, the percentage of cases in which Rapid Miner predicted fail is (40.36%). Therefore, the model made incorrect prediction of (about 297 students) who actually passed but were predicted as fail. Similarly, cases in which Rapid miner predicted as fail is (about 297 students) and the model is incorrect prediction of (about 14 students) As showed in the following, the weighted mean recall is 69.71%.

$$Recall = \frac{Sum\ of\ true\ positives}{Sum\ of\ true\ positives + Sum\ of\ false\ negatives}$$

Weight mean Recall $= \frac{99.06+40.36}{2} * 100 =$ **69.71%**

Weighted mean precision

The weighted mean of all per class precision measurements. It is calculated through class precisions for individual classes.

$$Precision = \frac{Sum\ of\ true\ positives}{Sum\ of\ true\ positives + Sum\ of\ false\ positives}$$

$$\textbf{weight mean precision} = \frac{\mathbf{93.49\% + 83.23\%}}{\mathbf{2}} = \mathbf{88.36\%}$$

From the above report result, the percentage of cases which rapid miner classified correctly (pass) is 83.23%, and similarly 93.49% correctly classification as fail, respectively. The Weighted mean precision of the correct classified is 88.36%.

Accuracy Calculation of the Naïve Bayes Model

The Naïve Bayes classification algorithm is employed by the administrator to predict the performance of the students considering their general behavior or attributes. A Naïve Bayes classifier is an easy probabilistic classifier based on applying Bayes theorem from Bayesian statistics with strong (naive) independence assumptions.

Depending on the precise nature of the probability model, Naïve Bayes classifiers can be trained very efficiently during a supervised learning setting. The following classification result would be obtained after the experiment was made.

Table 4: Performance table for model with mixed data features using NB

	True FAIL	True PASS	class precision
pred. FAIL	295	168	63.71%
pred. PASS	203	1320	86.67%
class recall	59.24%	88.71%	

Source: Computation based on data from authors field work, 19 Nov 2020

The value 'pass' is positive class and 'fail' is negative class. The data set consists of 1986 records of new students which are used by naïve Bayes algorithm in classifying the results by using the students' general attributes. In the first row, in the true fail column of the confusion matrix, 295 students are classified correctly as negative (predicted to fail) and 168 students are classified incorrect. In the second row in the true column of the confusion matrix 1320 students are classified as positive (predicted to pass) as well as 203 are classified incorrect.

Accuracy is a function of Sensitivity and Specificity that offers more information on the accurate prediction of data classification for unknown data classes.

Accuracy:-

$$\boxed{Accuracy = \frac{Sum\ of\ true\ positives + Sum\ of\ true\ negatives}{Total\ population}}$$

Probability of choosing true positives and negatives from all positive and negative tuples, or the probability of correct prediction for data. Following is the function for Accuracy

$$Accuracy = \frac{\mathbf{295 + 1320}}{\mathbf{1986}} * \mathbf{100 = 81.31\%}$$

Classification of Error

When we classify error, we have to apply the above formula.

$$Classification\ Error = \frac{Sum\ of\ false\ positives + Sum\ of\ false\ negatives}{Total\ population}$$

$$\boldsymbol{Classification\ Error} = \frac{\mathbf{168+203}}{\mathbf{1986}}\mathbf{*100=18.68\%}$$

The performance metric of NB in the above table shows only an accuracy of 81.31% where 168 students are classified as *pass* that are actually *fail*, and 203 students are classified as *fail* that are actually *pass* which implies that the model predicts approximately 18.68% of incorrect results ,i.e, when a model is generated based only on the students' general data. In addition, class precision for *true pass* is 86.67% and 63.71% for *true fail* as well as recall of *true pass* is 88.71% and 59.24% is *true fail*.

Weighted Mean Recall

The weighted mean of all per class recall measurements. It is calculated through class recalls for individual class. The percentage of cases in which Rapid Miner predicted pass (88.71%) with a margin error on 11.29% is acceptable. The recall measure answers the question, "If a student is going to pass, what is the likelihood that the model will sense it?" The answer is (88.71%). On the other side, the percentage of cases in which Rapid Miner predicted *fail* is (59.24%). Therefore, the model made incorrect prediction of (about 203 students) who were actually *pass* but were predicted as *fail*. Similarly, cases in which Rapid miner Predicted as *fail* are (about 295 students), and the model made incorrect prediction of (about 168 students). As indicated in the following, the weighted mean recall is 73.97%.

$$Recall = \frac{Sum\ of\ true\ positives}{Sum\ of\ true\ positives + Sum\ of\ false\ negatives}$$

$$\text{Weight mean Recall} = \frac{\mathbf{59.24+88.71}}{\mathbf{2}} * 100 = \mathbf{73.97\%}$$

Weighted mean precision

The weighted mean of all per class precision measurements

It is calculated through class precisions for individual class.

$$Precision = \frac{Sum\ of\ true\ positives}{Sum\ of\ true\ positives + Sum\ of\ false\ positives}$$

$$\textbf{weight mean precision} = \frac{\mathbf{63.71 + 86.67}}{\mathbf{2}} = \mathbf{75.19}\%$$

As we discussed on the above report result, the percentage of precision in the cases which rapid miner classified correctly (pass) is 86.67% and 63.71% correctly classified as fail, respectively. The Weighted mean of the correctly classified is 75.19%.

Accuracy Calculation of the K-NN Model

K-NN is considered as another usual technique used for classification. The input for this Algorithm is the k closest training data set. The classification of data is based on majority vote of its neighbors where the data is assigned to the class which is most common among its k nearest neighbors. K-NN is considered as one of the easiest among all machine algorithms.

Table 5: Performance table for model with mixed data features using Decision Tree

	True FAIL	True PASS	class precision
pred. FAIL	498	0	100.00%
pred. PASS	0	1488	100.00%
class recall	100.00%	100.00%	

Source: Computation based on data from authors field work, 19 Nov 2020

K-NN algorithms used to determine how many students would pass and fail, in any course in order to focus on these students to improve their academic performance. The value ‘pass’ is positive class and ‘fail’ is negative class. The data set consists of 1986 records of new students, which are used by K-NN algorithm in classifying the results by using the students’ general attributes. In the first row, in the *true fail* column of the confusion matrix, 498 students are classified correctly as negative (predicted to fail) and only 0 student are classified incorrect. In the second row in the *true* column of the confusion matrix 1488 students are classified as positive (predicted to pass) as well as only 0 students are classified incorrect.

Accuracy is a function of Sensitivity and Specificity that offers more information on the accurate prediction of data classification for unknown data classes. Accuracy is the probability of choosing *pass* and *fail* tuples, or the probability of correct prediction for data. Following is the function for Accuracy:

$$\boxed{Accuracy = \frac{Sum\ of\ true\ positives + Sum\ of\ true\ negatives}{Total\ population}}$$

$$\boldsymbol{Accuracy} = \frac{\mathbf{498 + 1488}}{\mathbf{1986}} * \mathbf{100 = 100\%}$$

Classification Error

When we Classification error we have to apply the above formula.

$$Classification\ Error = \frac{Sum\ of\ false\ positives + Sum\ of\ false\ negatives}{Total\ population}$$

$$\textbf{Classification Error} = \frac{\mathbf{0+0}}{\mathbf{1986}}\mathbf{*100=0\%}$$

The performance metric of K-NN in the above table shows only an accuracy of 100% where no students that are actually *fail* are classified as *pass,* and no students that are actually *pass* are classified as *fail,* which implies that the model predicts approximately 0% of incorrect results which is when a model is generated based on only on the students' general data.

The performance metric of K-NN in the above table shows only an accuracy of 100% where 168 students that are actually *fail* are classified as *pass*, and 498 students that are actually *fail* are classified as fail, which implies that the model predicts approximately 100% of correct results when a model is generated based on only students' students general data. In addition, class precision for *true pass* is 100% , and 100% for *true fail* as well as recall of *true pass* is 100% and 100% is *true fail.*

Weighted Mean Recall

The weighted mean of all per class recall measurements: It is calculated through class recalls for individual class.

The percentage of cases in which Rapid Miner predicted *pass* (100%) with a margin error on 0% is no error acceptable. The recall measure answers the question, "If a student is going to pass, what is the likelihood that the model will sense it?" The answer is (100%). On the other side, the percentage of cases in which Rapid Miner predicted *fail* is (100%). Therefore, the model made incorrect prediction of (about only 0 students) who were actually *pass* but no one was predicted as *fail*. Similarly, cases in which Rapid miner predicted as *fail* is (about 0 students) and the model is incorrect prediction of (about 0 students). As given in the following, the weighted mean recall is 100%.

$$Recall = \frac{Sum\ of\ true\ positives}{Sum\ of\ true\ positives + Sum\ of\ false\ negatives}$$

$$\textbf{Weight mean Recall} = \frac{100\%+100\%}{2} = 100\%$$

Weighted mean precision

The weighted mean of all per class precision measurements: It is calculated through class precisions for individual classes.

$$Precision = \frac{Sum\ of\ true\ positives}{Sum\ of\ true\ positives + Sum\ of\ false\ positives}$$

$$\textbf{weight mean precision} = \frac{100\% + 100\%}{2} = 100\%$$

As we discussed on the above report, the percentage of precision in the cases which rapid miner classified correctly (pass) is 100% and similarly 100% correctly classified as *fail*, respectively. The Weighted mean precision of the correct classified is 100%.

The factors that affect the students' Academic Performance

"What are the students' performances factors, which affect students' behavior and academic performance?"

It was conformed that the students' internal and external or non-academic factor plays a significant role in predicting their future performance. In this study, it was clearly accepted that the models generated using both mixed data together produced reliable predictions.

To estimate and reduce the numbers of drop outs of students in the new courses, proper analysis done on past data can also be applied for other courses. This can help the institution to rapidly and spontaneously predict students' level and at the same time to recognize dropouts in the new courses.

Table 6: - Comparison of the group factors which affect academic performance

Measure	Using only Academic data(internal data) **Accuracy (%)**	Using external data(Non Academic data) **Accuracy (%)**	Using Mixed data **Accuracy (%)**
DT	80.56%	74.92%	84.34%
NB	81.57%	74.42%	81.32%
K-NN	97.89%	99.70%	100.00%

Average	86.67%	83.01%	88.55%

As we see on the above Table 5, the highest 100% valid accuracy weighted is produced when using the mixed students' general data. The students' academic performances are affected more by internal (reason of choice of department, study time per a day, time in which students spend going out with friends and others). Similarly the students' academic performance is also affected by the external factors (guardian of the students, family relationship, parents' education level). 84.34% accuracy was produced by the decision tree when the data are mixed. Similarly, 81.32% medium result was obtained from the mixed approach by the Naïve Bayes when using the internal and external data.

In the same way, 80.56%, 81.57%, and 97.89% accuracy results were obtained when the researchers used the academic data (internal data) by applying the Decision Tree, Naive Baye and K-NN, respectively.

Correspondingly, when the researchers used the non-academic related data, various results 74.92%, 74.42% and 99.70% were obtained by the Decision Tree, Naïve Bayes and K-NN, in that given order.

The average of the overall model shows that, the different data sets obtained from different field, can be used to predict the students' academic performance. Firstly, 88.55% were more influenced by combination of the internal and external related data. Secondly, the students' academic data used to predict the students' academic performance was about 86.67%. Thus, the external data source gives 83.01% average accuracy.

These the above groups factors are summarized through graphical to provide descriptive information about the accuracy which would be used for the students' academic performance prediction used models by using internal students data, external students data and mixed data which obtained from the external and internal data sources.

4. Conclusion

This study has been done on factual and real data. From the above analysis, we have concluded that K-NN produces the best prediction results as compared to Bayesian network and Decision tree. K-NN exhibits higher accuracy i.e. 100 % when the target group is *pass* or *fail*, and 99.85% when the target data is of grade average point status. This result indicates that K-NN algorithms predict better performance of students. Two different data sets were mixed and analyzed with three different machine learning methods: as K-NN, Decision Tree and Naïve Bayes, and their results were compared using three evaluation measures. The Accuracy values for the first data set (academic data set) were 86.67%; likewise, the non-academic data set produced 83.01% average. However, the and best accurate was produced from the mixed data set, that is, 88.55% average of the overall accuracy of the classifications for the target attribute *pass* or *fail*. The results of this study indicate that integrating data would provide more improvement to prediction results than method selection.

Future work

In the future, the study can be enhanced by including various demographic, academic and economic related factors. This research has certain limitations that must be noted. Because of Covid-19 pandemic there was limitation in getting data of all students; and the study relied on public data sources. In addition, both data sets were small, having less than two thousand records. A research that has access to more comprehensive data may offer more conclusive results

Other methods, such as hybridization of artificial intelligent and deep learning can be used to have a better understanding of the importance of machine learning, especially to improve the performance of the Decision tree and Naïve Bayes.

5. References

Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417

Bushra Iqbal Chohan and Rehana Masrur Khan (2010,) Impact of Parental Support on the Academic Performance and Self Concept of the Student

Bishop (2006). Pattern Recognition and Machine Learning. Springer ISBN 0 387-31073-8.

Chapman & Hall/CRC Data Mining and Knowledge Discovery Series

Corbett, A.T., Koedinger, K.R. & Anderson, J.R. (1997). Intelligent tutoring systems. In M.G

Michie, D. J. (1994). Machine Learning, Neural and Statistical Classification.Prentice Hall Inc

D. Michiel, D. J. (1994). Machine Learning, Neural and Statistic Classification. Prentice Hall Inc.

Dipshi Shetty, Isha D Shetty,Sneha Roundhal (2019),Student Performance Prediction,Volume 8–Issue 05, 157-160, 2019, ISSN:-2319–865 www.ijcat.com

Feng, M. (2009). Towards Assessing Students' Fine Grained Knowledge: Using an Intelligent

Gareth James; Daniela Witten; Trevor Hastie; Robert Tibshirani (2013). An Introduction to Statistical Learning. Springer

H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis–a general method for cognitive model evaluation and improvement," in International Conference on Intelligent Tutoring Systems. Springer, 2006, pp. 164 –175

Jabeen Sultana, M. Usha Rani, M.A.H. Farquad, Student's Performance Prediction using Deep Learning and Data Mining Methods ,International Journal of Recent Technology and Engineering (Volume-8, Issue-1S4,June 2019

Jaime G. Carbonell, R. S. (1983). Machine Learning: A Historical and Methodological Analysis. Association for the Advancement of Artificial Intelligence , 4(3), 1-10

L. Rokach (2004). Data Mining with Decision Trees: theory and applications, vol. 69, World scientific

Liu, W., Chawla, S., Cieslak, D.A. & Chawla, N.V. (2010). A Robust Decision Tree Algorithm for Imbalanced Data Sets. In SIAM International Conference on Data Mining (SDM'10)

Maria Koutina and Katia Lida Kermanidis (2011), Predicting Postgraduate Students' Performance Using Machine Learning Techniques

MI Jordan (2014-09-10). "Statistics and Machine Learning". reddit. Retrieved 2021-1 01.

Mitchell(2000), T. M. (2006). The Discipline of Machine Learning. Machine Learning Department technical report CMU-ML-06-108, Carnegie#

Murat Pojon (2017). Using Machine Learning to Predict Student Performance

Pablo A Jaskowiak, Ricardo J G B Campello, Ivan G Costa (2014). On the selection of appropriate distances for gene expression data clustering

Rokach &Maimon (2005). Data Mining with Decision Trees: theory and applications. World scientific

Sleeman, D. H. (1983). Inferring Student Models for Intelligent CAI. Machine Learning. Press Polytechnic Institute, USA. 5

Yang, Q. & Wu, X. (2006). 10 challenging problems in data mining research. International Journal of Information Technology and Decision Making , 5,597{604. 25, 102}

Research and Knowledge Management Office
(RaKMO)
P.O. Box 18490, Addis Ababa, Ethiopia
Tel: +251(0)11-558-06-16
Email: rakmo@smuc.edu.et
Website:http://www.smuc.edu.et