INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 13, Issue 3, (2024) E-ISSN: 2226-6348

Students' Academic Performance: Prediction using Machine Learning Approaches

Syarifah Adilah Mohamed Yusoff, Jamal Othman, Azlina Mohd Mydin, Wan Anisha Wan Mohamad, Elly Johana Johan, Ahmad Fkrudin Mohamed Yusoff

Department of Computer Sciences and Mathematics, UiTM Cawangan Pulau Pinang, 13500 Permatang Pauh, Pulau Pinang, Malaysia, General Studies Department, Ungku Omar Polytechnic, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia

To Link this Article: http://dx.doi.org/10.6007/IJARPED/v13-i3/22164 DOI:10.6007/IJARPED/v13-i3/22164

Published Online: 15 July 2024

Abstract

The discipline of higher education is seeing rapid growth and is closely intertwined with the advancements in technology. Utilisation of machine learning (ML) to predict students' academic achievement has demonstrated promising results and has been advantageous for educational institutions. The challenges associated with making predictions reside in the ability to accurately identify potential attributes within multi-class projections, while also considering the varying quantities of distinct attribute categories. Therefore, this study has examined multiple classes and variations of attributes from various categories, including demographic, academic, personal, and parental profiles. The implementation of five distinct machine learning models for prediction exploited a dataset sourced from the Kaggle repository. In order to mitigate attribute complexity across several categories, two approaches for attribute selection or reduction were employed. Furthermore, eight distinct metrics were employed for the examination of the models. The findings indicate that the classification model's performance in terms of accuracy was only average when considering multi-class predictions and variations of categorical attributes. This was observed after using attribute reduction approaches for 50% and 100% of the attributes.

Keywords: Students Performance, Prediction, Machine Learning, Multi-Class Prediction, Classification

Introduction

Education and learning are inherent processes that seek to empower successive cohorts from early infancy to emerging adulthood through the cultivation of knowledge, beliefs, attitudes, and behaviours. The acquisition of knowledge and skills, spanning from primary school to higher education, constitutes a systematic endeavour aimed at cultivating competent individuals capable of addressing practical challenges within society (Tadese et al., 2022). Higher education institutions play a crucial role in the realisation of a nation's vision, with students being expected to dedicate a significant portion of their time to studying and achieving favourable academic outcomes (Shahiri et al., 2015).

The assessment of a student's academic performance serves as a crucial measure of productivity and the development of skilled human capital, which are considered valuable assets for the nation. One of the primary concerns for universities is to effectively monitor the academic progress of their students, with the ultimate goal of cultivating highly skilled graduates who can successfully compete in the job market (Mamoon-Al-Bashir, 2016). Taking into consideration the present and future challenges and requirements of students might result in more effective administration of their well-being.

Therefore, the identification of dependable factors that influence student performance holds potential benefits for admissions, students, and educators in facilitating further enhancements. The admissions process is capable of recognising potential students who may require more support, and it incorporates certain characteristics that can enhance the efficiency of the system. In order to enhance the efficacy of teaching and monitoring methods employed by lecturers, it is important to identify the most commonly dedicated mistakes and afterwards choose the most successful courses of action. It is advisable to provide students with recommendations for supplementary activities, instructional resources, and assignments that might enhance and facilitate their learning process.

The subsequent section, denoted as Section 2, will centre its attention on the literature review. This review will comprehensively examine previous studies, organising them based on the categories of attributes utilised in the analysis. Additionally, it will emphasise the significance of conducting multi-class investigations for target attributes and will explore the Machine Learning algorithms associated with their implementations. In Section 3, a comprehensive examination of the methodology employed and the constituent elements of the study is presented. Section 4 of the paper delves into a comprehensive analysis of the outcomes obtained from the prevailing prediction methodologies. Finally, the conclusion and future research directions are presented in Section 5.

Literature Review

One of the emerging challenges in the field of data mining is the endeavour to predict students' academic performance by uncovering the underlying patterns that contribute to their success or failure during their educational journey in university. The utilisation of descriptive and predictive analytics has been extensively investigated in various research domains, including but not limited to medical research Yusoff et al (2014), fraud detection, social behaviour analysis, and engineering applications (Salleh et al., 2020). In their comprehensive review, Baashar et al (2021) have identified seven distinct categories of attributes that are commonly employed in predicting student performance. These categories include demographic factors, academic indicators, internal assessments, communication skills, behavioural traits, psychological characteristics, and familial or personal factors. The study arrived at this conclusion after examining a total of 68 research studies. Among the various categories, the attributes most frequently used in the academic category are CGPA and attendance. Following this, demographic factors such as age, gender, and nationality. The third most widely used are personal or family-related characteristics, including parent's status, education, and income.

Nedeva & Pehlivanova (2020) has investigate the key variables that effect the educational success for effective machine learning analysis and reap benefit from the all collections of data in educational institutions. Instead of analysing all available variables or attributes, the study claims that reducing attributes while keeping accuracy close to initial is more effective than running all available attributes. The 12 prominent attributes for student's performance highlighted by this study are; 1. Age; 2. Gender; 3. Course by year; 4. Stress; 5. High school; 6. Assessment; 7. Fail exam; 8. Num Exam Fail; 9. Satisfaction with qualification; 10. Edu status; 11. Job satisfaction; 12. Marital status.

Meanwhile, Deepika & Sathyanarayana (2018) has come out with different set of attributes that commonly influence student's performance. The study implements 2 different datasets, the first one performance of secondary school students from UCI machine learning repository; and the second one is e-learning achievement from Kaggle. A list of demographical attributes, including parent status, mother education, mother job, farther education and farther job, demonstrates the impact on student's performance. The analysis only able to perform good results from single category of attributes that was supposed to influence students' performance.

Predictive analysis in data mining is a confluence of artificial intelligence, machine learning, and database techniques, currently being implemented in the context of "big data" environments. The utilization of heuristic algorithms, which integrate advanced mathematical and statistical analysis, has yielded positive outcomes across diverse domains of knowledge throughout the life of humanity (Adilah et al., 2014). Algorithms have emerged as a potent tool in the field of data mining, as they mimic biological processes observed in nature to effectively tackle intricate optimisation problems. This has made algorithms a fundamental component in predictive analysis. The incorporation of comprehensive taxonomies into algorithmic behaviour has significantly enhanced the ability to generate exceptional models in predictive analysis (Molina et al., 2020).

The performance of first-year students at the Faculty of Economics in Tuzla was examined by Osmanbegovic and Suljic (2012), who collected data on 12 distinct qualities or attributes. Three algorithms were used for the prediction models: C4.5, Naïve Bayes, and Multilayer Perceptron. The target attribute, which represents the grade, was evaluated using two different methods. Firstly, it was categorised into six classes, namely A, B, C, D, E, and F. Secondly, it was categorised into two classes, A and B. However, the initial approach was not documented and it was asserted that the analysis contained numerous inaccuracies. In the meantime, the two designated analyses have purportedly yielded statistically significant findings. The study does not include information regarding the imbalance issues that arise when the grade class is divided into two classes, with one class designated as 24.12 percent and the other as 75.88 percent. Bydžovská (2016) has investigated the performance of the higher-education students based on the grades from all the courses taken to predict the final grade for the students. The prediction was generated through the classification of grades as either "easy" if they were less than or equal to 2.4, or "difficult" if they were greater than 2.4. Al-Barrak & Al-Razgan (2016) has studied the impact of grades from all mandatory courses to predict student's final GPA. The attributes taken by each semester that consist about 5 mandatories courses and modelled using decision tree to come out with strong rules for prediction. The final GPA was used as target attribute and labelled into five classes which are Excellent, Very Good, Good, Average and Fail. The study discussed the classification rules of decision tree instead of reporting the accuracies of the model. Hence, the good rules might be generated from five level of class label from small academics attributes. Yohannes & Ahmed (2018) has studied the performance of students focusing to academic attributes that consist of grade of courses taken by student for 2 years and 3 years of studies. All numeric attribute of grade ranging from 0.00 to 4.00 were normalize to 0 and 1 for better coefficient

measures. The study report they yields good accuracies result for 2 years grades consist of 23 attributes using Support Vector Regression and Linear Regression for 3 years grades consist of 35 attributes for prediction. Anyhow the study does not elaborate further regarding on how they construct the target attribute. The target was considered final grade that consist of continuous data from 0.0 to 4.0. Since the target class was in continuous format, thus the prediction only available for regression-based algorithm. Further extension for other types of attribute such as demographic, personal or financial may face difficulties.
Acquiring accurate predictions can be a complex endeavour, despite the effective application of data mining in educational contexts. However, the dependability of these methods is still in its infancy, and the extraction of novel and valuable knowledge remains imperfect. The aforementioned research demonstrates positive outcomes when the analysis includes one or two types of data, typically pertaining to demographics and academic performance. However, the multiclass scenario presents additional complexities as the classifier is required to differentiate among a large number of classes in order to generate accurate predictions. The term "multi-class" pertains to situations when predictions involve more than two classes. Typically, predictions involve a positive class (labelled as 1) and a complementary class (labelled as 0). Yet in multi-class scenarios, there are two or more classes, and each occurrence is associated with only one class. When occurrences are associated with more than one class, the dataset is referred to as having multi-label classes (Agrawal & Sah, 2022). Figure 1 depicts an infographic that serves to distinguish between the three concepts of classes in the target attribute for prediction.

Figure 1. Infographics for 3 concepts of classes in data mining (Projectpro, 2023).

The above researches have identified that the primary target attribute for performance prediction is the grade. The conventional approach of assessing students' performance in higher education typically involves a grading system that encompasses a range of letter grades, including A+, A, A-, B+, B, B-, C+, C, C-, D, and Fail. Therefore, the number of classes for the target attribute will be 10, potentially resulting in significant complexity for prediction. In spite of that, most of data mining algorithms were designed to efficiently run binary or two classes prediction and do not support more than two class prediction such as Logistic regression and Support Vector Machine (SVM). Ishfaq et al (2022) assert that a meticulous algorithm selection process is crucial for multiclass prediction. This process should consider the algorithms' behaviour in relation to the dataset's size, characteristics, and attribute kinds.

Furthermore, multi-class datasets often encounter the issue of imbalanced data, characterised by an unequal distribution of occurrences or instances across different classes. This imbalance can result in statistically inaccurate predictions due to significant disparities in the number of instances between the classes. This subject presents significant hurdles as real-world situations often involve imbalanced data, and the majority of studies concentrate on enhancing the prediction of imbalanced two-class scenarios, which often involve a single majority class and a minority class (Buda et al., 2018). There has been a limited amount of research dedicated to examining and comprehending the intrinsic attributes of unbalanced data. However, it has been observed that the disparity among classes is frequently accompanied with supplementary challenges in data analysis. These challenges include the presence of infrequent sub-concepts inside the minority classes, overlapping regions between different classes, and the occurrence of uncommon minority cases situated within the region dominated by the majority class (Lango & Stefanowski, 2022).
Despite the numerous challenges, this study aims to examine the performance of students across various attribute categories, such as demographics, academics, administration, personal, and financial factors. This endeavour is driven by the widely acknowledged reality, as highlighted by Tadese et al (2022); Idris et al (2012), that the evaluation of students' performance should not be confined to a narrow set of criteria. This pilot study aims to identify appropriate algorithms for the classification of multi-class target attributes in predicting the academic performance of higher-education students.

Methodology

Prediction in the field of data mining can be achieved through various techniques, including but not limited to classification, clustering, association analysis, and text mining. However, the selected techniques are contingent upon the purpose of the investigation or predictions. After acquiring the data, the implementation of machine learning through classification involves several subsequent phases. These phases include pre-processing, which is also referred to as data cleaning. The purpose of pre-processing is to ensure that all relevant attributes are appropriate for the selected machine learning algorithm.

In this study a classification approach also known as supervised learning has being utilized in order to develop an appropriate model to predict students' performance. This approach involves dividing the data into training and testing sets to construct and validate the model. Figure 2 illustrates the architectural framework employed in the research, encompassing the entire process from data acquisition to performance evaluation.

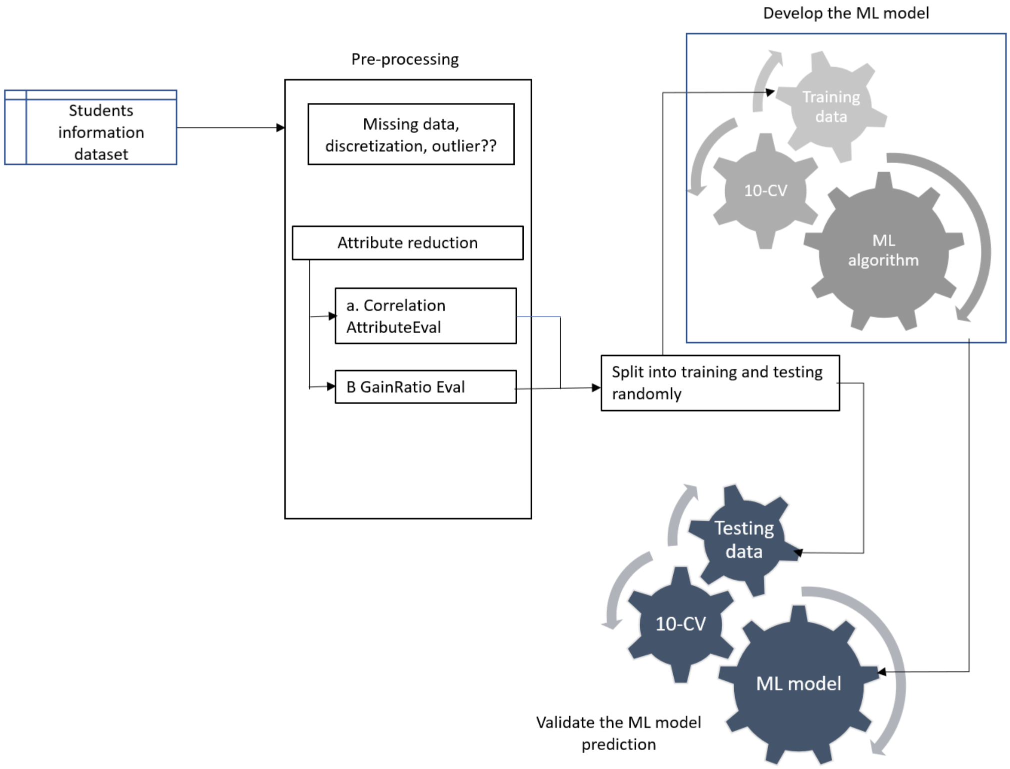

Figure 2. Framework of Research Architecture Diagram.

Data Description

The dataset was obtained from online sources and retrieved from the repositories on (Yalmaz & Sekeroglu, 2020). The initial dataset, titled "Higher Education Students Performance Evaluation," was gathered in 2019 from students enrolled in the Faculty of Engineering and the Faculty of Educational Science. The primary objective of this data collection was to predict the academic performance of these students at the end of the term. The compiled data encompassed not only academic aspects, but also encompassed individuals' backgrounds and lifestyles. The questionnaires are divided into three distinct sections. Section 1 pertains to personal inquiries, section 2 focuses on familial matters, and section 3 delves into educational habits. A total of 32 questions were used as attributes for the analysis. Table 1 provides comprehensive descriptions of the attributes that have been taken into consideration. In this study, the 32nd attribute from Table 1 is considered as target attribute and the data consist of 8 distinct types of grade status for each record.

Table 1
Properties of the Dataset from Higher Education Students

No	Attribute	Description
1	Age	1:18-21, 2: 22-25, 3: above 26
2	sex	1:female, 2:male
3	Graduated high school type	1: private, 2:state, 3: other
4	Scholarship type	1: none, 2:25%, 3:50%, 4: 75%, 5: Full
5	Additional work	1:yes, 2:no
6	Activity (regular artistic or sport activity)	1:yes, 2:no
7	partner	1:yes, 2:no
8	Total salary if available (USD)	1:135-200, 2:201-270, 3:271-340, 4:341-410, 5: above 410
9	Transport to university	1:bus, 2:private car/taxi, 3:bicycle
10	Accommodation type in Cyprus	1:rental, 2:dormitory, 3:with family, 4:other
11	Mother’s education	1: primary school, 2: secondary school, 3: high school, 4: university, 5: Msc., 6: Ph.D
12	Father’s education	1: primary school, 2: secondary school, 3: high school, 4: university, 5: Msc., 6: Ph.D
13	Siblings (if available)	1:1, 2:2, 3:3, 4:4, 5: 5 or1 above
14	Parental status	1: married, 2: divorced, 3:died (one of them/both)1
15	Mother occupation	1: retired, 2: housewife, 3: government officer, 4: private sector employee, 5: self-employment, 6: other
16	Father occupation	1: retired, 2: government officer, 3:private sector employee, 4: self-employment, 5: other
17	Weekly study hours	1: none, 2:<5hours, 3: 6-10 hours, 4: 11-20 hours, 5:more than 20 hours
18	Reading non-scientific book/ journals (frequency)	1: none, 2: sometimes, 3: often
19	Reading scientific book/ journals (frequency)	1: none, 2: sometimes, 3: often
20	Attendance seminar/ conference related to department	1: yes, 2:no
21	Impact of project/ activities on your success	1: positive, 2: negative, 3: neutral
22	Attendance to class	1: always, 2: sometimes, 3: never
23	Preparation to midterm exams (accompany)	1: alone, 2: with friends, 3: not applicable
24	Preparation to midterm exams (time)	1: closest date to the exam, 2: regularly during the semester, 3: never
25	Taking notes in classes	1: never, 2: sometimes, 3: always
26	Listening in classes	1: never, 2: sometimes, 3: always

27	**Discussion improves my interest and success in the course**	**1: never, 2: sometimes, 3: always**
28	**Flip class**	**1: not useful, 2: useful, 3: not applicable**
29	**Grade previous (CGPA of last semester)**	**1: <2.00, 2: 2.00-2.49, 3: 2.50-2.99, 4: 3.00-3.49, 5: above 3.49**
30	**Grade expected (for graduation)**	**1: <2.00, 2: 2.00-2.49, 3: 2.50-2.99, 4: 3.00-3.49, 5: above 3.49**
31	**Course id**	
32	**Grade (OUTPUT grade)**	**0: Fail, 1: DD, 2: DC, 3: CC, 4: CB, 5: BB, 6: BA, 7: AA**

Attribute Reduction Techniques

As illustrated in Figure 2, after the pre-processing all attributes except the 32nd were ranked based on the evaluated weightage of their dependency among them. This study has implemented 2 most popular attribute reduction techniques as follows:

1. CorrelationAttributeEval

Evaluates the worth of an attribute by measuring the correlation (Pearson's) between the each of the attribute and the target class [23]. Nominal attributes are considered on a value by value basis by treating each value as an indicator. An overall correlation for a nominal attribute is arrived at via a weighted average.

2. GainRatioAttributeEval

Gain Ratio is an alternative to Information Gain that is used to valuates the worth of an attribute by measuring the gain ratio with respect to the class [24]. It considers both information gain and the number of outcomes of an attribute to determine the best attribute to split on.

$$GainR(Class, Attribute)=(H(Class)-H(Class|Attribute))/ H(Attribute) \quad (1)$$

Where H here represent the entropy.

Evaluation Metrics

The prediction analysis for the above dataset is based on output grade that supposed to be dependent variable for all mentioned attributes (also known as features in data mining). The grade is considered has direct impact for the performance of the students in higher education level and consist of 5 classes. In this study, the grade classes are simplified to as ‘a’ for AA and BA indicate excellent, ‘b’ for BB and CB indicate very good, ‘c’ for CC and DC indicate good, ‘d’ for DD to indicate satisfactory and ‘Fail’ to indicate fail. Therefore, all of these 5 classes of grade will be predicted across all 145 records (also known as instances in data mining) and the performance of accuracy will be recorded to evaluate the performance of classification model selected which are OneR, AttributeSelectedClassifier, J48, MLP and Naïve Bayes algorithms. Table 2 depicts example of one of the confusion matrix table that will construct after prediction from classification models.

Table 2
Table of confusion matrix after prediction from one of the classifier models

Classified/ Predicted as					**N=145**
a	**b**	**c**	**d**	**Fail**	**Actual**
TP_a	x	x	x	x	a
x	TP_b	x	x	x	b
x	x	TP_c	x	x	c
x	x	x	TP_d	x	d
x	x	x	x	TP_{Fail}	Fail
FP_a	FP_b	FP_c	FP_d	FP_{Fail}	

The prediction will be based on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values from the confusion matrix table. In general, the TP indicate that model is correctly predicted or classified of positive class (a/b/c/d/Fail) as positive, the TN indicate that model is correctly predicted or classified of negative class (a/b/c/d/Fail) as negative, the FP indicate that model is wrongly predicted or classified of negative class (a/b/c/d/Fail) as positive and lastly the FN indicate that model is wrongly predicted the positive class (a/b/c/d/Fail) as negative.

In this study, the performance of the generated classification model was evaluated using micro-averaging of multi-class metrics. The count of TP, FP and FN across all classes were aggregated and then calculates the performance of all models based on the total counts. Total of TP represent the sum of TP count across all classes, FP is the sum of false positive counts across all classes and FN is sum of false negative count across all class. The total number of FP and FN in multi-class dataset will be equal since the FP in a class is considered FN in the actual class(https://www.evidentlyai.com/classification-metrics/multi-class-metrics). All of the considered parameter for analysis are calculated and explained as follows:

1. True Positive Rate (TPR)

This rate is referring to proportion of correctly predicted for positive class (given class). Also known as sensitivity or recall.

$$TPR = TP/(TP+FN) \tag{2}$$

2. Precision(P)

Precision is the number of correct positive prediction from the total of positive prediction or classification.

$$P = TP_A + TP_B + \ldots\ldots\ldots + TP_N /(TP_A + FP_A + TP_B + FP_B + \ldots\ldots + TP_N + FP_N) \tag{3}$$

3. Recall (R)

Recall measure model's ability to detect positive sample. The higher the recall, the more positive samples predicted.

$$R = TP_A + TP_B + \ldots\ldots\ldots + TP_N /(TP_A + FN_A + TP_B + FN_B + \ldots\ldots + TP_N + FN_N) \tag{4}$$

4. F-Measure(Fm)

F-measure is calculated in a way to combine both precision (P) and recall (R) in order to express both concerns with a single score. Thus, the Fm is considered the harmonic mean of two fraction.

$$Fm = 2PR/(P+R) \tag{5}$$

5. Accuracy

Accuracy is the first metric used to assess how well a model predicts. The calculation is based on the number of correctly predicted from all the prediction.

$$\text{Accuracy} = \text{Correct Prediction/All predictions} \tag{6}$$

Result and Discussion

Selected Attribute

The prediction analysis for the dataset stated above is based on the output grade, which is assumed to be the dependent variable or feature for all the other attributes mentioned.

CorrelationAttributeEval

Combination of CorrelationAttributeEval as attribute evaluator and Ranking method of search is applied to the dataset explained in previous section. Figure 3 shows the ranking attributes with respect to CorrelationAttributeEval method.

GainRatioAttributeEval

Combination of CorrelationAttribute Eval as attribute evaluator and Ranking method of search is applied to the dataset explained in previous section. Figure 4 shows the ranking attributes with respect to CorrelationAttributeEval method.

```
Attribute selection output

=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

average merit      average rank  attribute
 0.159 +- 0.011      1.2 +- 0.4     29 grade_previous
 0.133 +- 0.016      3.2 +- 2.04    21 impact_of_projects
 0.129 +- 0.008      3.5 +- 1.36    31 course_id
 0.115 +- 0.012      5.7 +- 2.41    18 reading_non_scientific
 0.115 +- 0.016      5.9 +- 2.66    30 grade_expected
 0.11  +- 0.009      6.6 +- 1.91     1 age
 0.105 +- 0.012      7   +- 2.57     2 sex
 0.105 +- 0.012      7.6 +- 2.97     6 activity
 0.103 +- 0.014      8.8 +- 3.03     8 total_salary
 0.092 +- 0.018     11.7 +- 5.1     22 attendances_classes
 0.088 +- 0.011     12.2 +- 4.4     11 mother_ed
 0.088 +- 0.011     12.8 +- 3.46    25 taking_notes
 0.086 +- 0.01      13.4 +- 3.8     10 accomodation
 0.082 +- 0.014     15.1 +- 4.44    20 attendance_seminars_dep
 0.079 +- 0.01      16.3 +- 3.1     14 parental_status
 0.074 +- 0.015     17.7 +- 5.44     3 graduated_h_school_type
```

Figure 3. The first 50% of attribute rank by CorrelationAttributeEval

```
Attribute selection output

=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

average merit      average rank  attribute
 0.4   +- 0.009      1   +- 0       31 course_id
 0     +- 0          2.8 +- 0.6     11 mother_ed
 0     +- 0          3.8 +- 0.6     10 accomodation
 0.096 +- 0.064      4.8 +- 4.07     2 sex
 0     +- 0          4.9 +- 0.83    12 farther_ed
 0     +- 0          5.9 +- 0.83     8 total_salary
 0     +- 0          6.6 +- 0.49    13 siblings
 0     +- 0          7.8 +- 0.6      9 transport
 0     +- 0          8.8 +- 0.6      7 partner
 0     +- 0          9.8 +- 0.6     15 mother_occup
 0     +- 0         11.5 +- 1.02     3 graduated_h_school_type
 0     +- 0         11.7 +- 0.9      6 activity
 0     +- 0         12.8 +- 0.6      4 scholarship_type
 0     +- 0         14.1 +- 0.3      5 additional_work
 0     +- 0         15.1 +- 0.3     14 parental_status
 0     +- 0         16.1 +- 0.3     16 father_occup
```

Figure 4. The first 50% of attribute rank by GainRatioAttributeEval.

Performance Comparison for Difference Classification Models

The OneR, AttributeSelectionClassifier, J48, Naïve Bayes and MLP models were constructed using the classification architecture discussed in the previous section. All of this was done in order to produce predictive analysis using a classification approach to predict student performance from a variety of attributes other than academic attributes. The result of classifications was split based on attribute selection with 50% that consist 16 attributes from both CorrelationAttribut Eval and GainInfoAttribute Eval, meanwhile 100% which 32 attributes (exclude Grade) from original dataset. Table 3 shows the performance of 16 attributes selected from CorrelationAttribute Eval across 5 predictive models. The overall result indicates low performance of accuracies where the highest value is 55.17 percent for both OneR and Attribute selectionClassifier, followed by J48 about 51.12 percent, Naïve Bayes about 48.27 percent and lastly MLP about 31.03 percent.

Table 3

Analysis of different model for the selected 16 attributes (50%) ranking from correlation-based attribute reduction

Model Constructed	TP rate	Precision	F-Measure	Accuracy%	MAE	RMSE	RAE%	RRSE%
OneR	**0.552**	**0.556**	**0.714**	**55.17**	**0.1793**	**0.4235**	**58.7159**	**108.9332**
Attribute SelectedClassifier	**0.522**	**0.556**	**0.714**	**55.17**	**0.2495**	**0.3612**	**81.69**	**92.93**
Naïve Bayes	**0.483**	**0.488**	**0.401**	**48.27**	**0.2532**	**0.3825**	**82.9248**	**98.4102**
MLP	**0.310**	**0.336**	**0.315**	**31.03**	**0.2575**	**0.4604**	**84.308**	**118.5**
J48	**0.512**	**0.548**	**0.513**	**51.1628**	**0.2128**	**0.4116**	**69.2557**	**105.4979**

Next, Table 4 shows the performance of attribute selection of GainRatioAttribute Eval for 50% that consist of 16 top ranking attributes across five different predictive models. The accuracy of the five predictive models much more lower than correlation-based selected attributes. The highest accuracy performances are from OneR and AttributeSelectedClassifier about

55.17 percent, followed by J48 about 48.84 percent, NaiveBayes 27.58 percent and lastly MLP about 20.68 percent.

Further, Table 5 depicts classification for all 32 attributes with five predictive models. The result show that the highest accuracy classification performance models are from OneR and AttributeSelectedClassifier about 55.17 percent, followed by J48 about 41.38 percent, MLP about 39.54 percent and lastly NaiveBayes 27.9 percent.

Table 4
Analysis of different model for the selected 16 attributes (50%) ranking from GainRatio-based attribute selection

Model Constructed	TP rate	Precision	F-Measure	Accuracy%	MAE	RMSE	RAE%	RRSE%
OneR	**0.552**	**0.556**	**0.714**	**55.1724**	**0.1793**	**0.4235**	**58.7159**	**108.9332**
Attribute SelectedClassifier	**0.552**	**0.556**	**0.714**	**55.17**	**0.2495**	**0.3612**	**81.69**	**92.93**
Naïve Bayes	**0.276**	**0.221**	**0.233**	**27.58**	**0.2817**	**0.4213**	**92.2345**	**108.3889**
MLP	**0.207**	**0.325**	**0.23**	**20.68**	**0.3138**	**0.5077**	**102.7637**	**130.6186**
J48	**0.488**	**0.503**	**0.488**	**48.84**	**0.221**	**0.4372**	**71.9325**	**112.0704**

Table 5
Analysis of different model for all 32 attributes (100%)

Model Constructed	**TP rate**	**Precision**	**F-Measure**	**Accuracy%**	**MAE**	**RMSE**	**RAE%**	**RRSE%**
OneR	**0.552**	**0.556**	**0.714**	**55.1724**	**0.1793**	**0.4235**	**58.7159**	**108.9332**
Attribute SelectedClassifier	**0.552**	**0.556**	**0.714**	**55.17**	**0.2495**	**0.3612**	**81.69**	**92.93**
Naïve Bayes	**0.279**	**0.263**	**0.243**	**27.907**	**0.2703**	**0.4361**	**87.987**	**111.7695**
MLP	**0.395**	**0.395**	**0.370**	**39.54**	**0.2402**	**0.435**	**78.18**	**111.5**
J48	**0.414**	**0.453**	**0.405**	**41.38**	**0.2281**	**0.4438**	**74.68**	**114.1773**

Low performance of the classification models happened for several reasons. Firstly, according to [25] multiclass label confusion matrix in predictive model has several challenges due to variety of model types and hyperparameters. Secondly, the variety of attributes used as input features, which includes demographics, social economics, and personal lifestyle, complicates the success of attribute correlations.

In term of performance of the models across different number of attributes, Figure 5 shows that stagnant of classification performance for OneR and AttributeSelectedClassifier for all both types of attribute selections; i) CorrelationAttribute Eval; and ii) GainRatioAttribute Eval. In another hand J48 demonstrates consistency evaluation for three different attribute selections and the ability to differentiate between different attributes. J48 claims to be robust and reliable because the prediction is sensitive to small changes. In addition, the relative values are small among attribute reduction (16 selected attributes) and all features indicate that the reduction does not eliminate overall important information from the dataset. Meanwhile MLP and Naïve Bayes unable to give any significant performance across all selected attributes.

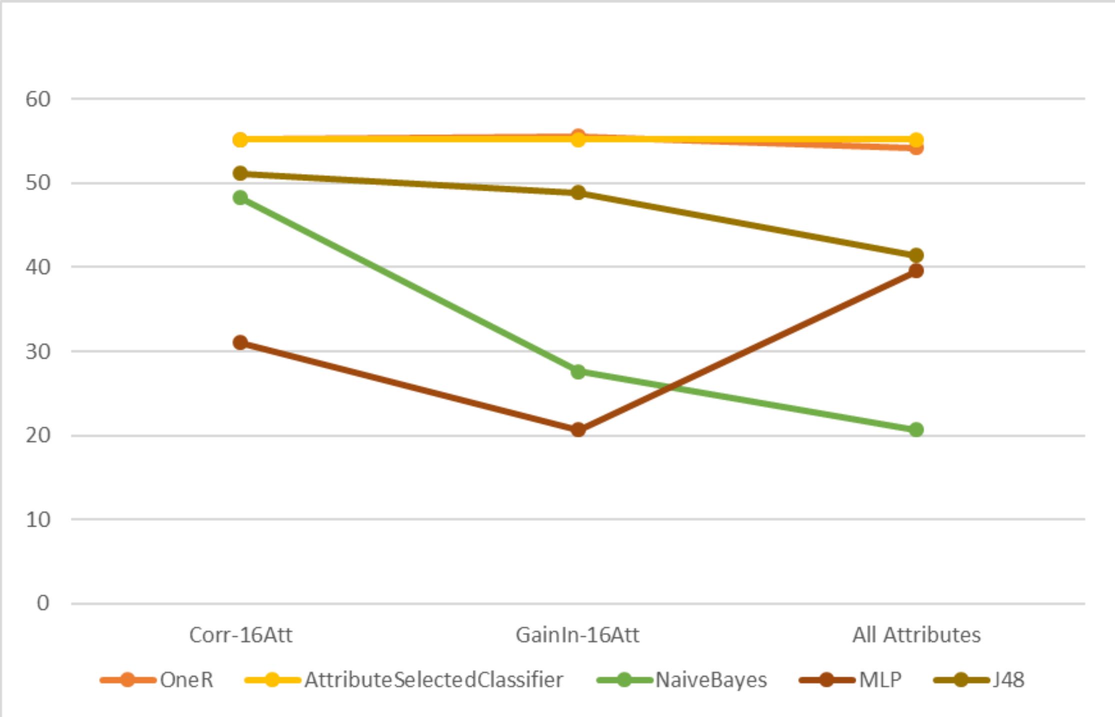

Figure 5. Performance of the predictive models across different number of attributes selection.

Conclusion

The primary objective of this pilot project was to assess the feasibility of predicting students' performance based on a diverse range of attribute categories, extending beyond solely academic attributes. This is because universities necessitate a comprehensive understanding of the factors contributing to a student's success and emphasise the importance of meticulous planning across every aspect of life. This study demonstrates that the successful achievement of this purpose hinges upon the utilisation of a dependable algorithm capable of accurately predicting outcomes based on a wide range of attribute categories, while also being resilient in its ability to analyse grades across several classes. The subsequent investigation will centre on the application of pre-processing techniques to address the issue of imbalanced data while utilising a multiclass grading system for target attributes. Furthermore, there will be a focus on improving the precision of predictive analysis by employing sophisticated learning models, such as hybrid algorithms, which purportedly exhibit strong performance when dealing with complex datasets.

Acknowledgement

The authors would like to thank the Computer Science and Mathematicals staff at Universiti Teknologi MARA, Cawangan Pulau Pinang, Malaysia, for their assistance with this project.

Corresponding Author

Syarifah Adilah Mohamed Yusoff
Department of Computer Sciences and Mathematics, UiTM Cawangan Pulau Pinang, 13500 Permatang Pauh, Malaysia.
Email: syarifah.adilah@uitm.edu.my

References

Agrawal, R., & Sah, R. (2022). Multi-label Classification Methods and Challenges. *International Journal of Mechanical Engineering*, *7*(4), 1295–1299.

Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting Students Final GPA Using Decision Trees: A Case Study. *International Journal of Information and Education Technology*, *6*(7), 528–533. https://doi.org/10.7763/ijiet.2016.v6.745

Baashar, Y., Alkawsi, G., Ali, N., Alhussian, H., & Bahbouh, H. T. (2021). Predicting student's performance using machine learning methods: A systematic literature review. *Proceedings - International Conference on Computer and Information Sciences: Sustaining Tomorrow with Digital Innovation, ICCOINS 2021*, 357–362. https://doi.org/10.1109/ICCOINS49721.2021.9497185

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Bydžovská, H. (2016). A Comparative Analysis of Techniques for Predicting Student Performance. *International Educational Data Mining Society*.

Deepika, K., & Sathyanarayana, N. (2018). Comparison of Student Academic Performance on Different Educational Datasets Using Different Data Mining Techniques. *International Journal of Computational Engineering Research (IJCER)*, *8*(9), 28–38.

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. *April*.

Idris, F., Hassan, Z., Ya'acob, A., Gill, S. K., & Awal, N. A. M. (2012). The Role of Education in Shaping Youth's National Identity. *Procedia - Social and Behavioral Sciences*, *59*, 443–450. https://doi.org/10.1016/j.sbspro.2012.09.299

Lango, M., & Stefanowski, J. (2022). What makes multi-class imbalanced problems difficult? An experimental study. *Expert Systems with Applications*, *199*(January), 116962. https://doi.org/10.1016/j.eswa.2022.116962

Shapiro, L. (University of Washington). (2015). *Information Gain Which test is more informative?* https://homes.cs.washington.edu/$~$shapiro/EE596/notes/InfoGain.pdf

Mamoon-Al-Bashir, M., Kabir, M. R., & Rahman, I. (2016). The Value and Effectiveness of Feedback in Improving Students' Learning and Professionalizing Teaching in Higher Education. *Journal of Education and Practice*, *7*(16), 38–41. www.iiste.org

Molina, D., Poyatos, J., Ser, J. Del, García, S., Hussain, A., & Herrera, F. (2020). Comprehensive Taxonomies of Nature- and Bio-inspired Optimization: Inspiration Versus Algorithmic Behavior, Critical Analysis Recommendations. *Cognitive Computation*, *12*(5), 897–939. https://doi.org/10.1007/s12559-020-09730-8

Osmanbegovic, E., & Suljic, M. (2012). Data mining approach for predicting student performance. *Economic Review: Journal of Economics and Business*, *10*(1), 3–12.

ProjectPro (2023). *How to Solve a Multi Class Classification Problem with Phyton.* https://www.projectpro.io/article/multi-class-classification-python-example/547

Saleh, A. Y., & Liansitim, E. (2020). Palm oil classification using deep learning. *Science in Information Technology Letters*, *1*(1), 1–8.

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, *72*, 414–422. https://doi.org/10.1016/j.procs.2015.12.157

Adilah, S. M. Y., Venkat, I., Abdullah, R., & Yusof, U. K. (2011). Mass spectrometry analysis via metaheuristic optimization algorithms. *Proceedings - 2011 6th International Conference* *on Bio-Inspired Computing: Theories and Applications, BIC-TA 2011*, 75–79. https://doi.org/10.1109/BIC-TA.2011.7

Tadese, M., Yeshaneh, A., & Mulu, G. B. (2022). Determinants of good academic performance among university students in Ethiopia: a cross-sectional study. *BMC Medical Education*, *22*(1), 1–9. https://doi.org/10.1186/s12909-022-03461-0

Nedeva, V., & Pehlivanova, T. (2021). Students' Performance Analyses Using Machine Learning Algorithms in WEKA. *IOP Conference Series: Materials Science and Engineering*, *1031*(1), 0–13. https://doi.org/10.1088/1757-899X/1031/1/012061

YÄ±lmaz N., Sekeroglu B. (2020) Student Performance Classification Using Artificial Intelligence Techniques. In: Aliev R., Kacprzyk J., Pedrycz W., Jamshidi M., Babanli M., Sadikoglu F. (eds) 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019. ICSCCW 2019. Advances in Intelligent Systems and Computing, vol 1095. Springer, Cham. https://www.kaggle.com/datasets/csafrit2/higher-education-students-performance-evaluation

Yohannes, E., & Ahmed, S. (2018). Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression: A Case Study. *International Journal of Computer Applications*, *180*(40), 39–47. https://doi.org/10.5120/ijca2018917057

Yusoff, S. A. M., Abdullah, R., & Venkat, I. (2014). Adapted bio-inspired artificial bee colony and differential evolution for feature selection in biomarker discovery analysis. *Advances in Intelligent Systems and Computing*, *287*, 111–120. https://doi.org/10.1007/978-3-319-07692-8_11