This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093563, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Multi-class Prediction Model for Student Grade Prediction using Machine Learning

SITI DIANAH ABDUL BUJANG[1], ALI SELAMAT[1,2,3], (Member, IEEE), ROLIANA IBRAHIM[2], ONDREJ KREJCAR[3], ENRIQUE HERRERA-VIEDMA[4,5], (Fellow, IEEE) HAMIDO FUJITA[4,6,7], (Life Senior Member, IEEE), AND NOR AZURA MD GHANI[8]

[1] Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia Kuala Lumpur, 54100, Malaysia
[2] School of Computing, Faculty of Engineering, Malaysia & Media and Games Center of Excellence (MagicX) Universiti Teknologi Malaysia, Johor Baharu 81310, Malaysia
[3] Faculty of Informatics and Management, University of Hradec Kralove, Rokitanskeho 62, 50003 Hradec Kralove, Czech Republic
[4] Andalusian Research Institute DaSCI Data Science and Computational Intelligence, University of Granada, 18071 Granada, Spain
[5] Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[6] i-SOMET incorporated Association, Morioka, Japan
[7] Regional Research Center, Iwate Prefectural University, Iwate, Japan
[8] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

Corresponding author: Ali Selamat (aselamat@utm.my)

ABSTRACT. Today, predictive analytics applications became an urgent desire in higher educational institutions. Predictive analytics used advanced analytics that encompasses machine learning implementation to derive high-quality performance and meaningful information for all education levels. Mostly know that student grade is one of the key performance indicators that can help educators monitor their academic performance. During the past decade, researchers have proposed many variants of machine learning techniques in student grade prediction. However, there is a lack of studies in identifying the effective predictive model, especially in addressing imbalanced multi-classification for student grade prediction. Therefore, this paper presents a comprehensive analysis of machine learning techniques to predict the final student grades in the first semester courses by providing better prediction accuracy. Two modules will be highlighted in this paper. First, we compare the accuracy performance of five well-known machine learning techniques, namely Decision Tree (J48), Support Vector Machine (SVM), Naïve Bayes (N.B.), K-Nearest Neighbor (kNN), and Logistic Regression (L.R.), to our real dataset. Second, we proposed a multi-class prediction model for an imbalanced multi-class dataset using two types of data-level solutions to improve the prediction accuracy performance. The obtained results indicate that the proposed Synthetic Minority Oversampling Technique (SMOTE) and wrapper-based feature selection (F.S.) integrates with kNN shows significant improvement with the highest accuracy of 99.6%. In contrast, all F.S. algorithms perform 99.4% robust performance when working with SVM independently. These findings indicate that the proposed model generally discovers various F.S. algorithms and SMOTE in addressing imbalanced multi-classification, which will help the researcher to improve the performance for student grade prediction.

INDEX TERMS Machine learning; predictive model; imbalanced problem; student grade prediction; multi-class classification;

I. INTRODUCTION

In higher education institutions (HEI), every institution has its student academic management system to record all student data containing information about student academic results in final examination marks and grades in different courses and programs. All student marks and grades have been recorded and used to generate a student academic performance report to evaluate the course achievement every semester. The data keep in the repository can be used to discover insightful information related to student academic performance. Solomon et al. [1] indicated that determining student academic performance is a crucial challenge in HEI. Due to this, many previous researchers have well-defined the influence factors that can highly affect student academic performance [2]. However, the most common factors are relying on socioeconomic background, demographics [3], and learning activities [4] compared to final student grades in the final examination [5]. For this reason, we observe that predicting student grades can be one of the solutions that are applicable to improve student academic performance [6].

Predictive analytics has shown a successful benefit in the HEI. It can potentially benefit the competitive educational domain to find hidden patterns and make predictions trends in a vast database [7]. It has been used to solve several educational areas: student performance, dropout prediction, academic early warning systems, and course selection [8]. Moreover, predictive analytics in predicting student academic performance has increased over the years [9].

The ability to predict student grades is one of the important areas that can help to improve student academic performance. Many previous research has found variant machine learning techniques performed in predicting student academic performance. However, the related works on improving the imbalanced multi-classification problem in predicting students' grade prediction are difficult to find [10-11]. Therefore, in this study, a comparative analysis of well-known machine learning techniques was performed for the student grade prediction by addressing the following questions:

RQ1: Which predictive model among the selected machine learning algorithms performs high accuracy performance to predict student's final course grades?

RQ2: What is the impact of Synthetic Minority Oversampling Technique (SMOTE) and feature selection (F.S.) algorithms on the performance of selected machine learning algorithms in handling imbalanced multi-classification dataset?

To address the above-mentioned questions, we collect the student final course grades from two core courses in the first semester of the final examination result. We present a descriptive analysis of student datasets to visualize student grade trends, which can lead to strategic planning in decision-making for the lecturers to help students more effectively. Then, we conduct a comprehensive analysis using five well-known machine learning algorithms, including L.R., NB, J48, SVM and kNN, on the real student data of Diploma in Information Technology (Digital Technology) at one of Malaysia Polytechnic. As for addressing the imbalanced multi-classification, we endeavor to enhance each predictive model's performance using two types of data- level solutions with SMOTE and F.S. The novel contribution of this paper is, we proposed a multi-class prediction model that could perform very best of accuracy performance up to 99.6% for predicting student final course grades are based on the following categories; 'exceptional', 'excellent', 'distinction', 'pass' and 'fail'. Our work tried to bring out the efficiency of how the accuracy for imbalanced multi-classification can be improved with a data-level solution using SMOTE and F.S. in student grade prediction.

This paper is organized as follows. Section II describes the related research work that has been conducted for student grade prediction. Section III illustrates the methodology of developing predictive models to predict final student grades by phases. Section IV and Section V present the descriptive analysis and prediction results of this study's findings, respectively. Section VI discusses the findings result. Lastly, the paper is highlighted with the main conclusions with some future directions in Section VII.

II. RELATED WORKS

Several studies have been conducted in HEI for predicting student grades using various machine learning techniques. It involves the analytical process of many attributes and sample data from various sources for student grade prediction in different outcomes. However, the contribution to improving the accuracy for imbalanced classes in student grade prediction is still rarely discussed. Related to these issues, a study from [12] used discretization and oversampling SMOTE methods to improve the accuracy of students' final grade prediction. Several classification algorithms have been applied such as N.B., D.T. and Neural Network (N.N.) for classifying students' final grade into five categories; A, B, C, D and F. They showed that N.N. and N.B. applied with SMOTE and optimal equal width binning outperformed other methods with the similar highest accuracy of 75%. However, NB was found better compared to N.N. as the optimal time to utilize the prediction models are faster than N.N. Research conducted by [13], has developed a method for predicting future course grades obtained from the Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE) programs at the University of Minnesota. Based on the proposed methods, the results indicated that Matrix Factorization (M.F.) and Linear Regression (LinReg) performed more accurate predictions than the existing traditional methods. The author also found that the use of a course-specific subset of data can improve prediction accuracy for predicting future course grades. Another study in [14], applied M.F., Collaborative Filtering (C.F.) and Restricted Boltzmann Machines (RBM) techniques on 225 real data of undergraduate students to predict student grades in different courses. They observe that using C.F. does not indicate good performance, especially when there found many sparsity in the dataset compared to M.F. However, their overall findings show that the proposed RBM provides efficient learning and better prediction accuracy compared to C.F. and M.F. with minimum Root Mean Squared Error (RMSE) 0.3, especially for modeling tabular data. A study in [15] has developed a predictive model that can predict student's final grades in introductory courses at an early stage of the semester. They have compared eleven machine learning algorithms in five different categories consist of Bayes, Function, Lazy (IBK), Rules-Based (R.B.) and Decision Tree (D.T.) using WEKA.

TABLE 1. The taxonomy of related studies on student grade prediction

Paper	Sample Size	Data Source	Attributes	Algorithm	Best Performance	Limitation
Jishan et al. [12]	180	Student Core Course offered at North South University, Bangladesh	CGPA, Quiz, Midterm, Lab, Attendance, Final grade	N.B., D.T., Neural Network Backpropagation with oversampling (SMOTE) and optimal binning	NB (optimal binning+SMOTE) Accuracy 75.28%	Small size of attributes that lead to high missclasification error
Polyzou et al. [13]	76,748	Student-course grade from 2002 to 2014	Historical student course grade information	LinReg, MF	LinReg	Not support a large number of latent factors
Iqbal et al. [14]	225	Undergraduate students of the Electrical Engineering Program from 2013 to 2015	Grades, GPA	CF, M.F., RBM	RBM	Use limited attribute for analysis
Khan et al [15]	50	Student of Buraimi University College, Oman	Test1_marks, CGPA, Attendance, Major, Gender, Year	NB, MLP, SVM, Lazy (IBK), Rules-Based (Decision Table, JRIP, OneR, PART and ZeroR) D.T. ((J48), R.F., RT SimpleCART)	DT (J48) (Feature Selection+SMOTE) Accuracy 88%	Small number of dataset
Barrak et al. [16]	236	A female student from Computer Sciences College at King Saud University 2012	Student name, student I.D., final GPA, the semester of graduation, major, nationality, campus, courses are taken and course grade	D.T. (J48)	D.T. (J48)	Lack of experimental techniques for prediction
Abana [17]	133	Students of Computer Engineering program in 4 years	Research Method (R.M.) grade, Research Project (R.P.) grade, gender, backlog, programming proficiency	RT, RepTree DT (J48)	R.T. Accuracy 75.2%	Lack of experimental techniques for prediction
Ahmad et al. [18]	399	First-year bachelor students in Computer Science at UniSZA from 2006/2007 to 2013/2014	GPA, race, gender, family income, university entry mode, Malaysia Certificate of Education (SPM) grade in 3 subjects	D.T., NB and Rule Based (PART)	R.B. Accuracy 71.3%	A small number of dataset due to incomplete and missing value
Anderson et al. [19]	683	Students of Craig School of Business at California State University, Fresno from 2006 to 2015	Historical grade data from 18 semester	N.B., KNN, SVM	SVM	Some of the dataset are not available due to significant changes.

C.F. – Collaborative Filtering, MF-Matric Factorization, D.T.- Decision Tree, LinReg- Linear Regression, RBM - Restricted Boltzmann Machines, MLP- Multilayer Perceptron, NB – Naïve Bayes, RT – Random Tree, Lazy (IBK) – KNN – K-Nearest Neighbor, RF – Random Forest, SVM – Support Vector Machine

To reduce high dimensionality and unbalanced data, they have performed feature selection correlation-based and information-gain for data-preprocessing. The author also applied SMOTE to balance the distribution instances of three different classes. Among the 11 algorithms, they indicated that the Decision Tree classifier (J48) has the highest accuracy of 88% compared to other categories of algorithms. Al-Barrak [16] used D.T. (J48) algorithm to discover classification rules for predicting students' final Grade Point Average (GPA) based on student grades in previous courses. They have used 236 students who graduated from Computer Science College at King Saud University in 2012. They found that the classification rule produced from J48 can detect early predictors and can extract useful knowledge for final student GPA based on their grades in all mandatory courses to improve students' performance. Another study in [17] have predicted the student's grade performance using three different D.T. algorithms; Random Tree (R.T.), RepTree and J48. In this context, cross-validation is used to measure the performance of the predictive model. From the findings, the results indicated that R.T. obtained the highest accuracy of 75.188% better than the other algorithms. The accuracy of the predictive models can be improved by adding more samples and attributes in the dataset. [18] has proposed a framework for predicting student academic performance at University Sultan Zainal Abidin (UniSZA), Malaysia. The study applied 399 student records from the academic department database in the eight years' intakes that contained student demographics, previous academic records and family background information. The results indicated that the Rule-Based (PART) is the best model with 71.3% accuracy compared to D.T. and N.B. However, using the small sample size has affected accuracy performance due to incomplete and missing value found in the dataset. Anderson et al. [19] performed an experimental study on 683 students at the Craig School of Business at California State University from 2006 to 2015 by applying three machine learning algorithms to predict student grades. The study found that SVM is the best classifier. It consistently outperforms a simple average approach that obtained the lowest error rate to optimize each data class. The result could be different for the large data set due to significant changes in the historical grade dataset's structure and format. We have summarized related studies composed of sample size, data source, attributes, algorithm, best performance and limitation in Table 1.

III. PROPOSED MULTI-CLASS PREDICTION MODEL FOR STUDENT GRADE PREDICTION

This paper aims to identify the most effective predictive model, especially addressing imbalanced multi-class for student grade prediction. The framework consists of four main phases are shown in Figure 1. The input of our framework contains student's final course grade that we extract from student's academic spreadsheet documents and student academic repository. We applied two data-level solutions using oversampling and F.S. to enhance the performance of imbalanced multi-class prediction. Then, we design our proposed model by integrating both techniques into the selected machine learning classifier to evaluate performance metrics. Finally, data visualization is used to visualize the trend of the dataset and final classification results. The description of each phase is given in the following subsection.

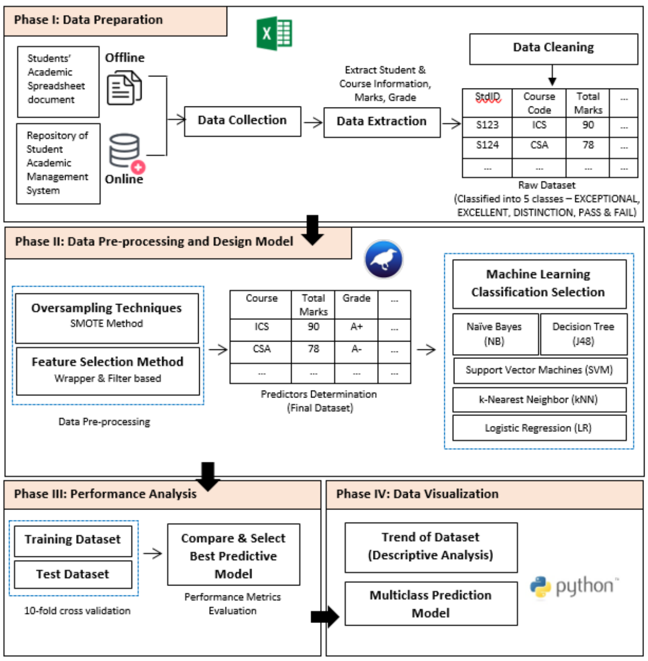


FIGURE 1. **The framework of the proposed multi-class prediction model for predicting final student grades**

A. DATA PREPARATION

The dataset we used was collected by the Department of Information and Communication Technology at one of the Malaysian Polytechnics. The dataset contains 1282 instances of the total course grades of the first-semester students taken from the final examination during the June 2016 to December 2019 session. Students need to take some compulsory, specialization and core course modules to qualify them for the next academic semester. However, in this study, we selected only two core courses that contained the percentage of final examination and course assessment marks. All attributes which are used for prediction are listed in Table 2.

B. DATA PRE-PROCESSING AND DESIGN MODEL

In this phase, we applied data pre-processing for the collected dataset. For the convenience of data pre-processing, we have ranked and grouped the students into five categories of grades: Exceptional (A+), Excellent (A), Distinction (A-, B+, B), Pass (B-, C+, C, C-, D+, D) and Fail (E, E-, F). The group was created to be the output of the prediction class. However, the class distribution of the dataset indicated imbalanced class instances containing a number of (63) exceptional, (377) excellent, (635) distinction, (186) pass and (21) fail that can lead to overfitting.

TABLE 2. **The information of the input attribute**

Attribute	Type	Values	Description
StudID	Nominal	S1-S641	Student Identification
Year	Numeric	[2016,2019]	Year of student intake
Class	Nominal	DDT1A, DDT1B, DDT1C, DDT1D	Class of student
Session	Nominal	DEC, JUNE	Session of student intake per year
Credit Hour	Numeric	[3]	Credit hour of each course
Course Code	Nominal	[CSA, ICS]	Course ID of 2 courses
Total Marks	Numeric	[38,91]	Student Final Marks obtained from final exam and courses assessment
Grade Pointer Average	Numeric	[0.00,4.00]	Student course grade pointer
Grade	Nominal	[A+, A, A-, B+, B, B-, C+, C, C-, D+, D, E, F]	Student Final Grade of each course
Group	Nominal	EXCEPTIONAL, EXCELLENT, DISTINCTION, PASS, FAIL	Category of student academic performance

Therefore, to avoid this problem, we proposed a data-level solution using oversampling (SMOTE) and F.S. (Wrapper and Filter) to overcome the imbalanced multi-classification dataset. The experiment used the open-source tool Waikato Environment for Knowledge Analysis (Weka) version 3.8.3 and explorer application. WEKA provides many machine learning algorithms using accessible graphical user interfaces for simple visualization [20-21].

C. PERFORMANCE ANALYSIS

This paper aims to predict students' final grades based on their previous course performance records in the first semester's final examination. The proposed model applied different machine learning algorithms to evaluate which algorithms provide the highest accuracy for predicting students' final grades. Several experiments were conducted in four distinct phases depending on the five different classes based on Group defined. The accuracy is evaluated using ten-fold cross-validation which our dataset is divided into a training set (90%) and a testing set (10%) on the same data [22]. Figure 2 illustrates the flowchart of the proposed multi-class prediction model applied in this study.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093563, IEEE Access

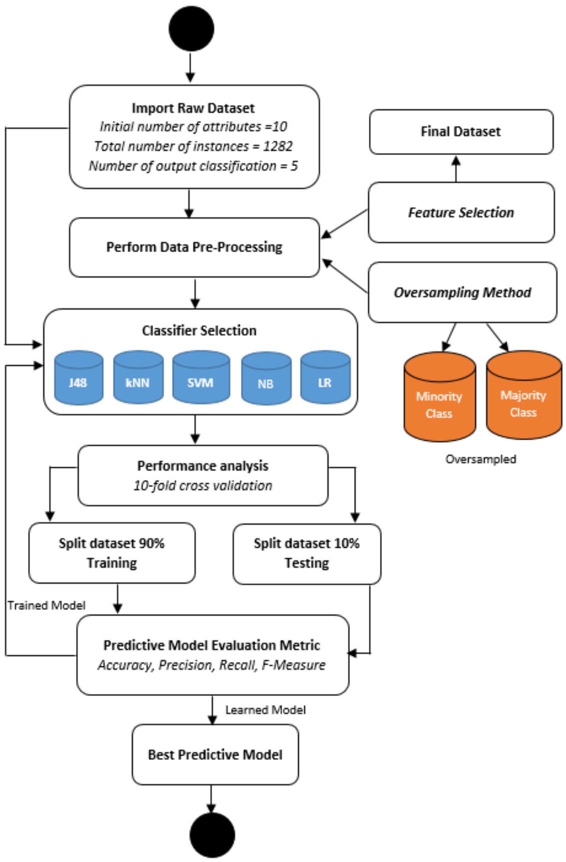


FIGURE 2. **Flowchart of the proposed multi-class prediction model**

Algorithm for multi-class prediction model

Output: Predicted Student's Grade for Testing Dataset
Input: Training Data with Output Labels

1 Begin
2 Import necessary library packages and dataset
3 Perform data preprocessing, including feature selection and oversampling for imbalanced dataset
4 Use classification models to predict the results
   4.1. Splitting data into training and testing dataset using 10-fold cross-validation
   4.2. Using well-known classification models (J48, kNN, SVM, L.R., NB) to predict the output label (Exceptional, Excellent, Distinction, Pass, Fail)
5 Evaluate the accuracy of well-known classification models

In particular, the following are the theoretical model used as the basis to construct our multi-class prediction model:

- Logistic Regression (L.R.), known as cost function, used logistic function to represent mathematical modeling to solve classification problems. The model performs great contextual analysis for categorical data to understand the relationship between variables [23].
- Naïve Bayes (N.B.) is based on the Bayesian theorem that is widely used as it is simple and able to make fast predictions. It is suitable for small datasets that combine complexity with a flexible probabilistic model [24].
- Decision Tree (J48) a widely used in several multi-class classification that can handle missing values with high Dimensional data. It has been implemented effectively to give an optimum accuracy results with minimum number of features [25].
- Support Vector Machine (SVM) is based on decision planes that state decision boundaries that handle classification problems successfully [11]. It takes a sorted dataset and predicts which of two conceivable classes includes the information, making the SVM a non-probabilistic binary linear classifier.
- K-Nearest Neighbor (kNN) is a non-parametric algorithm that classifies and calculates the difference between instances in the dataset based on their nearest vectors, where *k* refers to the distance in the *n*- dimensional space. It uses a distance function to suitability performs in small features of dataset [11].

A confusion matrix helps to visualize the classification performance of each predictive model. Table 3 presents the confusion matrix used for student grade prediction. A, B, C, D and E represent the classes for student grade (S.G.) level as 'exceptional', 'excellent', 'distinction', 'pass' and 'pass failure'. The class label represents in a form an expression:

$$SG \in \{A, B, C, D, E\} \tag{1}$$

TABLE 3. **Confusion matrix for student grade prediction classification**

		Predicted				
		A	**B**	**C**	**D**	**E**
Actual Label	**A**	AA	AB	AC	AD	AE
	B	BA	BB	BC	BD	BE
	C	CA	CB	CC	CD	CE
	D	DA	DB	DC	DD	DE
	E	E.A.	EB	EC	ED	EE

The performance metrics of the confusion matrix is determined using accuracy, precision, recall and f-measure in the following equation:

$$\boldsymbol{Accuracy\ (A)} = \frac{(AA + BB + CC + DD + EE)}{\sum N} \tag{2}$$

where N is the number of samples

$$\boldsymbol{Precision\ (P)} = \frac{1}{5}\left(\frac{AA}{AA + BA + CA + DA + EA} + \frac{BB}{AB + BB + CB + DB + EB} + \frac{CC}{AC + BC + CC + DC + EC} + \frac{DD}{AD + BD + CD + DD + ED} + \frac{EE}{AE + BE + CE + DE + EE}\right) \tag{3}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093563, IEEE Access

$$\boldsymbol{Recall/Sensitivity\ (R)} = \frac{1}{5}\left(\frac{AA}{AA+AB+AC+AD+AE} + \frac{BB}{BA+BB+BC+BD+BE} + \frac{CC}{CA+CB+CC+CD+CE} + \frac{DD}{DA+BD+DC+DD+DE} + \frac{EE}{EA+EB+EC+ED+EE}\right) \quad (4)$$

$$\boldsymbol{F - Measure} = 2\frac{PR}{P+R} \quad (5)$$

where the f-measure is the weighted harmonic mean of precision and recall

D. DATA VISUALIZATION

In this phase, after performed the data analysis, we extracted and visualized our findings to view the useful information and student grade performance trends in different courses using Python. Data visualization allows discovering all the features and insights of the student dataset to help lecturers improve student academic performance for better decision making in the future. We also compare each result of our proposed model using a better graphical approach to better understand the findings' results.

IV. DESCRIPTIVE ANALYSIS OF STUDENT DATASET

Our dataset contains records of 641 students who have taken two core courses, namely Computer System Architecture (CSA) and Introduction to Computer System (ICS). Based on the analysis performed, we found 362 students obtained distinction grade (A-, B+, B) in CSA course, followed by the pass grade (B-, C+, C, C-, D+, D) with 176 students, the excellent grade (A) with 80 students, failed grade (E, E-, F) with 19 students and finally exceptional grade (A+) with four students. On the other hand, for the ICS course, the highest grades obtained by the students were in excellent grade (A) with 297 students, followed by distinction grade (A-, B+, B) with 273 students, exceptional grade (A+) with 59 students, pass grade (B-, C+, C, C-, D+, D) and failed grade (E, E-, F) with 10 and 2 students respectively. Correspondingly, we have investigated the mean and standard deviation of the CSA course's final student grades, respectively 68.95 and 9.189, whereas for ICS courses 79.62 and 7.379. Table 4 shows the number of students in both courses.

TABLE 4. Result of student performance by course

Student Final Grade	No. of Student	
	CSA	ICS
Exceptional	4	59
Excellent	80	297
Distinction	362	273
Pass	176	10
Fail	19	2

The students' overall grades obtained in both courses are presented in Figure 3.

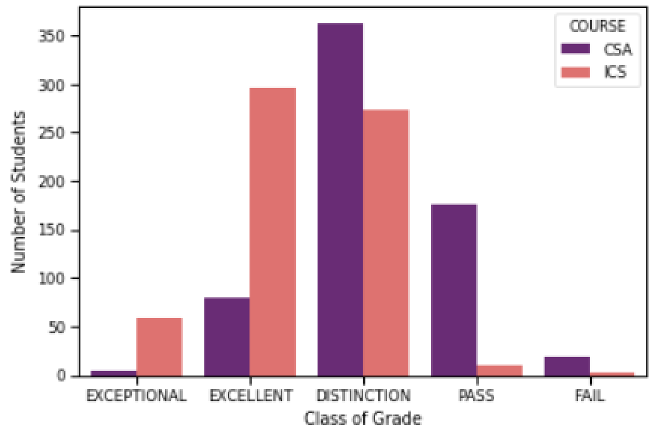


FIGURE 3. Number of students obtained final grades based on five categories of grades in different courses

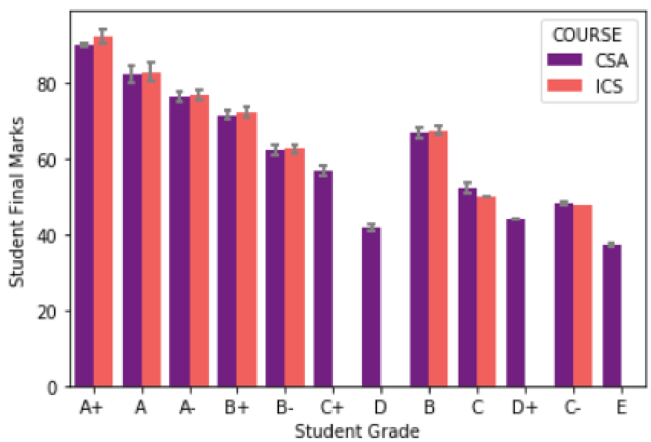


FIGURE 4a. Mean and standard deviation of student's final marks against student's final grades achievement according to the taken courses

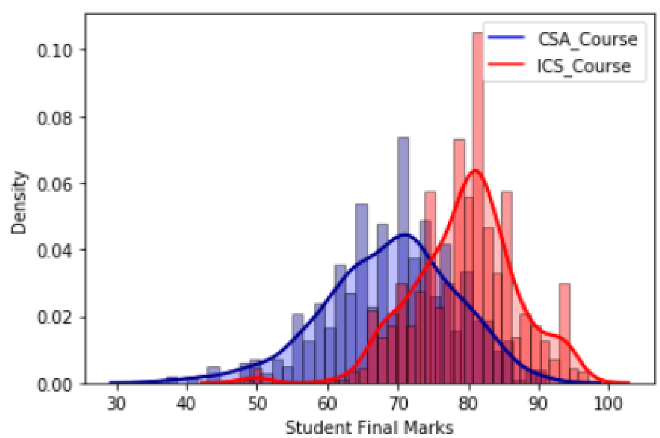


FIGURE 4b. Graph plot of student's final marks distribution

Figure 4a shows the mean and standard deviation of students' final marks and grade achievement according to the taken course. The students' final marks were calculated based on the total percentage from continuous assessment marks evaluated during class and the final test marks in the final exam at the end of the semester. However, the students must earn more than 40 marks for both assessments in order to enable them to pass both courses. From the results, we recognize a difference in student achievement results between the CSA and ICS, where the students obtained higher marks better in the ICS course than CSA. Figure 4b shows the normal trend of final marks distribution achieved by the students. Out of the total number of failed students, we found 3% of them are prominent in CSA compared to the ICS course. From these findings, we indicated that students who failed in both courses were not performed the minimum passing marks of the final examination, although their final marks classified as good and pass grades.

Furthermore, we also visualized the average grade point trend for ICS and CSA courses based on yearly achievement (2016 to 2019) as shown in Figure 5. From the observation, we found that the student's overall academic performance was improved yearly for both ICS and CSA courses. However, it is clearly showing that the grade point obtained from the ICS students is higher than the CSA. Therefore, from these findings, we indicated that the CSA course is more challenging to those students who are weak in mathematics, whereas the ICS course is easier to understand for students who already have basic knowledge of computers before entering the polytechnic.

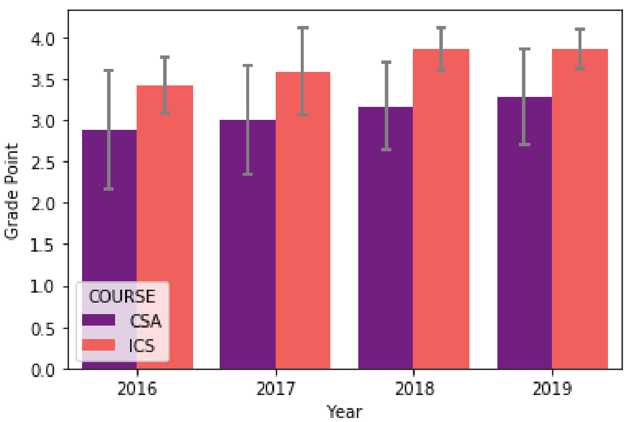


FIGURE 5. **Analysis of average grade point trend for ICS and CSA courses by yearly basis**

V. EXPERIMENTAL RESULTS

In this section, the results of this study are divided into two subsections according to research questions. We have conducted a comprehensive performance study on our experiments that run based on real dataset. The experiments' results of five selected machine learning algorithms (J48, kNN, N.B., SVM and L.R.) were explored and compared. Then, we also compared and measured the effectiveness of using SMOTE and F.S. algorithms with the selected machine learning algorithms in order to improve the imbalanced multi-classification problem.

A. RQ1: COMPARISON OF THE PREDICTIVE MODEL USING MACHINE LEARNING ALGORITHMS

Our main objective is to compare the predictive model based on the accuracy performance in this section. Here, five selected algorithms were used to train the student dataset, and their prediction accuracy was evaluated. In order to analyze the differences, we compare the performance accuracy using the ten-fold cross-validation with stratification as a testing method to derive the best predictive model for optimal results. We measure the performance using various metrics, including classification accuracy, precision, recall (Sensitivity) and f-measure to ensure the predictive model was fit to produce accurate results. Table 5 summarizes the prediction performance measures of the different classifiers on the student dataset.

TABLE 5. **Performance comparison of predictive models**

Metric	J48	kNN	NB	SVM	LR
Accuracy (%)	**0.989**	0.985	0.978	0.984	0.984
Precision	0.989	0.985	0.978	0.981	0.983
Recall (Sensitivity)	0.990	0.985	0.977	0.984	0.984
F-Measure	0.989	0.985	0.978	0.979	0.983

It can be seen from Table 4 that the results indicated J48 achieves the best prediction performance with the precision value of 0.989 whereas followed by kNN with 0.985. Meanwhile, L.R. and SVM obtained precision 0.983 and 0.981 respectively. The lowest model is achieved by NB with 0.978. However, because the classes in our dataset were highly imbalanced, we considered that the prediction results were insignificant due to overfitting and bias issues that may be created while training the dataset. For generalizability purpose, another experiment in dealing with the issues was conducted to rebalance the accuracy of which it is described in the next subsection.

B. RQ2: IMPACT OF OVERSAMPLING AND FEATURE SELECTION FOR IMBALANCED MULTI-CLASS DATASET

Here, we only focus on data-level solutions using oversampling SMOTE and F.S. for addressing imbalanced multi-classification datasets [27-28]. To see the performance of each predictive model, we have performed three experiments on five selected machine learning algorithms to reduce the imbalanced problem. First, we performed SMOTE on our dataset with five selected machine learning algorithms independently. Secondly, the dataset was executed on three F.S. algorithms independently. Thirdly, it was performed on the proposed multi-class prediction model (SFS), which combines SMOTE and F.S. with five selected machine learning algorithms. For a better view of the dimensionality prediction accuracy, other performance metrics on precision, recall and f-measure were used to ensure that our predictive model was fit to produce accurate results.

1) OVERSAMPLING TECHNIQUE

SMOTE known as Synthetic Minority Oversampling Technique, is the most commonly used to improve the overfitting problem based on random sampling algorithm [29]. It can modify an imbalanced dataset and generates new existing minority class instances by using synthetic sampling technique to create the

TABLE 6. Result of oversampling SMOTE with different predictive models

Predictive Model	Oversampling	Accuracy	Precision	Recall	F-Measure
J48	None	0.989	0.989	0.990	0.989
	SMOTE	0.992	0.991	0.991	0.991
kNN	None	0.985	0.985	0.985	0.985
	SMOTE	**0.993**	**0.993**	**0.993**	**0.993**
NB	None	0.978	0.978	0.977	0.978
	SMOTE	0.983	0.983	0.983	0.983
SVM	None	0.984	0.981	0.984	0.979
	SMOTE	0.989	0.989	0.989	0.989
LR	None	0.984	0.983	0.984	0.983
	SMOTE	0.988	0.988	0.988	0.988

class distribution more balanced. The idea was implemented that based on the nearest neighbors (k) of sample S.G. in the minority class, select N samples randomly and record them as $SG_{i.}$ The new sample SG_{new} is defined by the following expression:

$$SG_{new} = SG_{origin} + rand \times (SG_i - SG_{origin}), i = 1,2,3,..n \quad (6)$$

where $rand$ is a seed used of random sampling within range (0,1) and index of class value 0 with the ratio of generating new samples approximates 100%. In Weka, we implemented *weka.filters.supervised.instance.SMOTE* to insert synthetic instances between minority class samples of neighbors to our dataset. Then, we set classValue=0 to auto-detect the non-empty minority class, k = 10, percentage =100% and apply SMOTE filter ten times of iteration. The oversampled dataset increased the number of instances from 1282 up to 2932 where the class distribution with SMOTE becomes (504) exceptional, (377) excellent, (635) distinction, (744) pass and (672) fail. In Table 6, we present the detailed comparison results of all predictive models with all performance measures. When the classifiers were used with SMOTE, we found that the effectiveness of all models was consistently improved. Among these classifiers, kNN generated the most promising precision of 0.993, followed by J48 with 0.992, SVM with 0.989, L.R. with 0.988 and NB 0.983. This result was statistically significant with a confidence level of 95% using Paired T-Tester (corrected) as showed in Figure 7.

```
Dataset                    (1) bayes.Na | (2) lazy. (3) trees (4) funct (5) funct
---------------------------------------------------------------------------------
new5-weka.filters.supervi(100)   98.35 |   99.35 v   99.16 v   98.86 v   98.82 v
---------------------------------------------------------------------------------
Average                          98.35 |   99.35     99.16     98.86     98.82
---------------------------------------------------------------------------------
                                 (v/ /*) |   (1/0/0)   (1/0/0)   (1/0/0)   (1/0/0)
```

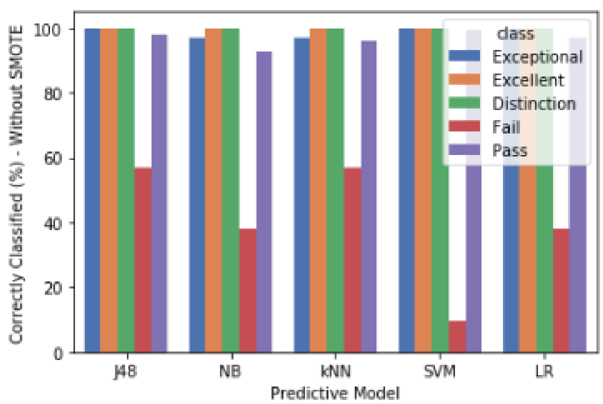
FIGURE 7. Result of predictive model performance with SMOTE

We also observed when SMOTE method was applied, the minority class instance has increased to balance with other classes by a number of iteration and the number of k to our dataset. In order to further performed a detailed analysis of the accuracy gained, we compared the confusion matrix derived via SMOTE as reported in Table 7. It is evident from the confusion matrix that all predictive model derived from J48, NB, kNN, SVM and L.R. shows improvement results of correctly classified for 'Pass' and 'Fail' grades. However, there is little decrease in performance from SVM where the predictive model correctly classified 97.2% of students who obtained 'Pass' grades compared to 99.5% when applied without SMOTE.

TABLE 7. Detailed analysis of confusion matrix over predictive models

Predicted Class	J48	NB	kNN	SVM	LR	SMOTE
Exceptional	100	96.8	96.8	100	100	**Before**
Excellent	100	100	100	100	100	
Distinction	100	100	99.8	100	99.8	
Pass	**97.8**	**92.5**	**96.2**	**99.5**	**96.8**	
Fail	**57.1**	**38.1**	**57.1**	**9.5**	**38.1**	
Exceptional	100	99.6	99.8	100	100	**After**
Excellent	100	100	100	100	100	
Distinction	100	100	99.8	100	99.7	
Pass	**98.7**	**95.3**	**98.8**	**97.2**	**96.9**	
Fail	**97.8**	**98.1**	**98.7**	**98.2**	**98.5**	

All values are measure in percentage (%)



.

FIGURE 8. Comparison of predictive model for correctly classified class

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093563, IEEE Access

without applied SMOTE

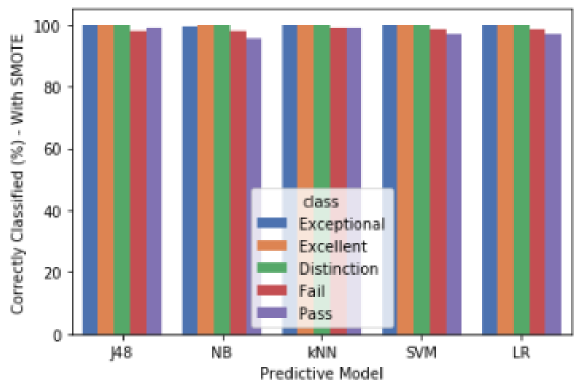


FIGURE 9. Comparison of predictive model for correctly classified class with applied SMOTE

For comparative analysis, Figures 8 and Figure 9 illustrate actual scores and predictions based on five categories of grade before and after applying the SMOTE respectively. Each predictive model performance shows the significant improvement for correctly classified prediction results for all class.

2) FEATURE SELECTION

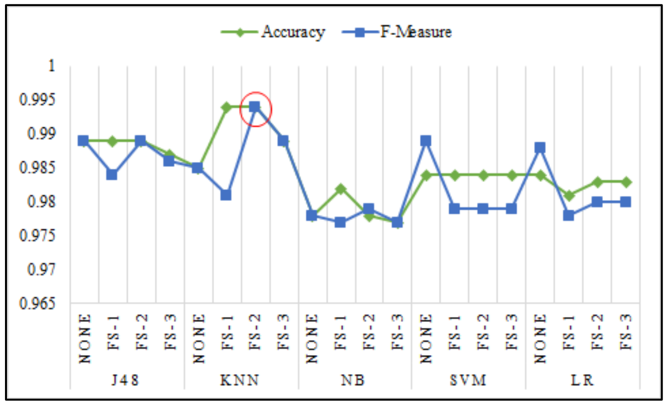
Another technique that we applied is feature selection (F.S.) which is effective in reducing dimensionality, removing irrelevant data and learning accuracy [30-31]. In this experiment, we compared three algorithms consist of wrapper and filter based to maximize the predictive performance of the multi-class prediction model proposed. The F.S. algorithm used to identify the best features in this study include; WrapperSubsetEval (FS-1), ClassifierSubsetEval (FS-2) and InfoGainAttributeEval (FS-3). The selected features are presented in Table 8.

TABLE 8. Detailed selected features over different algorithms

Main Attribute	Wrapper-based		Filter-based
	FS- 1	FS- 2	FS-3
StudID		/	/
Year	/	/	
Class			
Session			
Credit Hour			
Course Code	/		
Total Marks	/	/	/
Grade Pointer Average		/	/
Grade	/	/	/
Group (Class)	/	/	/
Total Features Selected	**5**	**6**	**5**

For the analysis, we have used the same dataset to find the best predictive model that fit with the requirements for giving an optimal result. Table 9 shows overall results of different predictive model with all measurement of F.S. algorithms. From the analysis, the result showed that kNN exhibited a highest performance rate of precision, recall and f-measure with 99.4% when integrate with FS-2 (ClassifierSubsetEval) algorithm compared to others predictive models. As we can see from Table 9, N.B. shows the lowest performance of accuracy but the f-measure for N.B. shows slightly improvement varied from 97.8% to 97.9% when FS-2 (ClassifierSubsetEval) was undertaken. On the other hand, the performance of L.R. showed slightly decreased when compared to without F.S. However, SVM showed a consistent rate of accuracy, precision, recall and f-measure values among all predictive models with 98.4% when three F.S. algorithms were applied. This result indicated that the high imbalance ratio in the multi-class dataset hinders SVM learning performance to predict student grades better. The comparison of highest accuracy and f-measure when applied F.S. are highlighted in Figure 10.

FIGURE 10. Comparison of accuracy and f-measure with F.S.



After performing several experiments, we investigated and extrapolated a combination of SMOTE and feature selection (SFS) to deal with imbalanced multi-classification. The detailed results of the proposed model SFS are presented in Table 10. Based on that result, kNN outperformed the highest f-measure performance up to 99.7% with SFS-2 (SMOTE-ClassifierSubsetEval), whereas J48 has the highest f-measure of 99.4% with SFS-3 (SMOTE-InfoGainAttributeEval). The other predictive models did not perform well when combined with both SMOTE and F.S. However, NB, SVM and L.R. outperformed when we execute with SMOTE independently. Figure 11 shows the comparative analysis of accuracy and f-measure values on SMOTE, F.S. and proposed model SFS with all predictive models. The highlighted red color represents the highest accuracy and f-measure obtained during the experiments done on each predictive model.

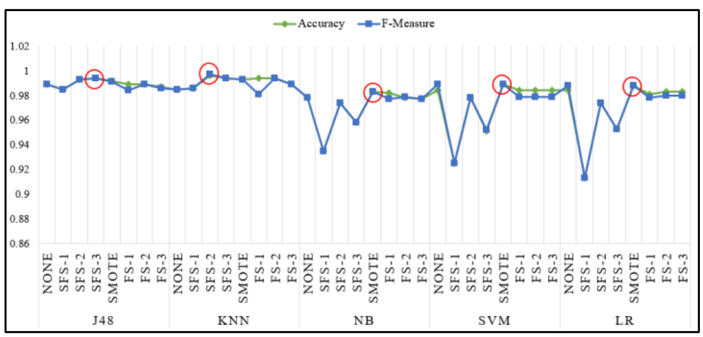


FIGURE 11. Comparison of accuracy and f-measure of SMOTE, F.S. and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093563, IEEE Access

proposed SFS

TABLE 9. Overall result of F.S. with different predictive models

Predictive Model	Feature Selection	Accuracy	Precision	Recall	F-Measure
J48	None	0.989	0.989	0.99	0.989
	FS-1	0.989	0.984	0.986	0.984
	FS-2	0.989	0.988	0.989	0.989
	FS-3	0.987	0.985	0.987	0.986
kNN	None	0.985	0.985	0.985	0.985
	FS-1	0.994	0.981	0.983	0.981
	FS-2	**0.994**	**0.994**	**0.994**	**0.994**
	FS-3	0.989	0.989	0.989	0.989
NB	None	0.978	0.978	0.977	0.978
	FS-1	0.982	0.978	0.976	0.977
	FS-2	**0.978**	**0.979**	**0.978**	**0.979**
	FS-3	0.977	0.978	0.977	0.977
SVM	None	0.984	0.989	0.989	0.989
	FS-1	0.984	0.981	0.984	0.979
	FS-2	0.984	0.981	0.984	0.979
	FS-3	0.984	0.981	0.984	0.979
LR	None	0.984	0.988	0.988	0.988
	FS-1	0.981	0.976	0.983	0.978
	FS-2	0.983	0.978	0.983	0.980
	FS-3	0.983	0.978	0.983	0.980

TABLE 10. Result of proposed SFS with different predictive models

Predictive Model	SMOTE+ Feature Selection (SFS)	Accuracy	Precision	Recall	F-Measure
J48	None	0.989	0.989	0.99	0.989
	SFS-1	0.985	0.985	0.985	0.985
	SFS-2	0.993	0.993	0.993	0.993
	SFS-3	**0.994**	**0.994**	0.994	**0.994**
kNN	None	0.985	0.985	0.985	0.985
	SFS-1	0.986	0.986	0.986	0.986
	SFS-2	**0.996**	**0.997**	0.997	**0.997**
	SFS-3	0.994	0.994	0.994	0.994
NB	None	0.978	0.978	0.977	0.978
	SFS-1	0.935	0.938	0.935	0.935
	SFS-2	0.974	0.975	0.974	0.974
	SFS-3	0.958	0.959	0.958	0.958
SVM	None	0.984	0.989	0.989	0.989
	SFS-1	0.925	0.932	0.925	0.925
	SFS-2	0.978	0.979	0.978	0.978
	SFS-3	0.951	0.959	0.952	0.952
LR	None	0.984	0.988	0.988	0.988
	SFS-1	0.913	0.917	0.913	0.913
	SFS-2	0.974	0.975	0.974	0.974
	SFS-3	0.953	0.955	0.953	0.953

VI. DISCUSSION

The study was conducted to create a predictive model that can predict students' final grades in Computer System Architecture (CSA) and Introduction to Computer System (ICS) for the first semester courses. The results of final student grades from the final examination of these two courses were considered the essential key indicators to determine the students' academic performance. A similar study conducted in [7][32], also justified the significant course grades that can help decision-making in the educational field. To answer our research question, we conducted a comprehensive experiment on our actual student dataset by comparing the accuracy performance of the prediction model in a selected machine learning algorithm. Then, we also applied SMOTE and F.S. algorithms to compare the effectiveness of the predictive model in addressing imbalanced multi-classification datasets for predicting student grades. The evaluation results were based on the performance metrics of accuracy, precision, recall and f-measure to compare the predictive models' performance.

Overall results indicated that all predictive models derived from J48, NB, kNN SVM and L.R. delivered a better performance when we applied SMOTE to the imbalanced dataset. The best accuracy obtained by the kNN with 99.3% slightly higher than J48 with 99.2% shows that the kNN algorithm was the ideal solution that can work with the best value of k = 10. Based on this result, we indicated that the predictive model kNN is an ideal algorithm to predict final student grades depending on the selected number of k values to obtain the best accuracy result as depicted in [33]. On the other hand, N.B. performed lower accuracy with 98.3% than others, indicating it was not highly effective for predicting student grades.

However, after we applied the F.S. method using wrapper-based, only kNN and N.B. showed, whereas SVM remained the same with no changes. Therefore, our result proved that SVM did not support multi-classification independently due to the limitation of computing the best hyperplane for a high-dimensional imbalanced dataset [34]. For NB, the used of F.S. for predicting student's grade is also supported in [30], where the author found N.B. showed the highest accuracy performance when wrapper-based subset feature selection was undertaken.

Then, we attempted to reduce the overfitting and bias of the minority class by combining SMOTE with a selection of appropriate features for all predictive models. The experiment results revealed that the proposed model SFS had the most significant effect on kNN depending on the selected F.S. algorithms. Certainly, these results also similar to the best performance of kNN in handling imbalanced data with different case studies as depicted in [35]. In this context, we also observe that most of the predictive models considered benefit when performing SMOTE, but we identify that integrating the accurate features with different F.S. algorithms can influence the prediction effectiveness as well. Furthermore, we found F.S. enabled the predictive model to be interpreted more quickly, but the improvement was not depending on few features [36]. Despite these findings, we have identified several limitations to this fact; (1) the analysis is based on a defined dataset, but other dataset should be tested for data generalization that could affect the analysis results; (2) the analysis is only carried out with certain well-known algorithms but can be analyzed with an ensemble or advanced machine learning algorithms to compare the effectiveness for the imbalanced multi-classification prediction model. (3) we used only one method of oversampling SMOTE, more methods could be used to analyze whether they can improve the multi-class imbalanced problem. Therefore, this study still needs to be improved in predicting students' final grades by selecting the relevant features for the imbalanced multi-class dataset that might affect the accurate prediction results. In addition, we can also use SVM ensemble to be part of the analysis since it has produced greater accuracy when predicting students' final grades, as mentioned in [37].

VII. CONCLUSION AND FUTURE DIRECTIONS

In this paper, a multi-class prediction model using five machine learning algorithms was compared to predict final students' grades based on the previous student final examination result of the first-semester course. We successfully presented the importance of F.S. and SMOTE methods that can efficiently give different prediction results to improve student grade prediction. Interestingly, we discovered that the prediction accuracy performance result was improved consistently when using SMOTE than F.S. for all predictive models. Performing F.S. with SMOTE significantly impacts the performance of the student grade prediction. We indicated that selected features for small datasets and imbalanced classes could influence prediction accuracy in various predictive analysis models to predict student values. However, our findings contribute to a multi-class prediction model that can be a practical approach for predicting the students' final grades based on final examination results.

In HEI, predictive analytics plays a significant role in governance for improving valuable information and developing trusted decision-making that contributes to data science [38]. Determining the quality of the collected dataset to reduce the imbalance and missing values difficulties is part of the challenging issues that adhere to choosing the relevant and valuable predictive models [39]. Therefore, as for future works, further investigation on the use of appropriate emerging predictive techniques in such advanced machine learning algorithms [40] and ensemble algorithm are recommended to optimize the result for predicting student grade. It is also essential to upgrade the number of features to be analyzed and improve the multi-class imbalanced dataset with appropriate models to improve the effectiveness of the prediction model. Thus, using machine learning in higher learning institutions for student grade prediction will ultimately enhance the decision support system to improve their student academic performance in the future.

ACKNOWLEDGMENT

This work was supported/funded by the Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS/1/2018/ICT04/UTM/01/1). The authors sincerely thank Universiti Teknologi Malaysia (UTM) under Research University Grant Vot-20H04, Malaysia Research University Network (MRUN) Vot 4L876, for the completion of the research.

REFERENCES

[1] D. Solomon, S. Patil and P. Agrawal, “Predicting Performance and Potential Difficulties of University Student using Classification: Survey Paper,” Int. J. Pure Appl. Math, vol. 118, no. 18, pp. 2703–2707, 2018.

[2] E. Alyahyan and D. Düştegör,” Predicting academic success in higher education: literature review and best practices,” Int. J. Educ. Technol. High. Educ., vol. 17, no. 1, 2020.

[3] V. Miguéis, A. Freitas, P.J. Garcia and A. Silva, “Early segmentation of students according to their academic performance: A predictive modelling approach,” Decision. Support Syst, 115, 36–51, 2018.

[4] P. M. Moreno-Marcos, T. C. Pong, P. J. Munoz-Merino, and C. D. Kloos, “Analysis of the Factors Influencing Learners’ Performance Prediction with Learning Analytics,” IEEE Access, vol. 8, pp. 5264–5282, 2020.

[5] A.E. Tatar and D. Düştegör, “Prediction of academic performance at undergraduate graduation: Course grades or grade point average?” Applied Sciences (Switzerland), vol. 10, no. 14, pp. 1–15, 2020.

[6] Y. Zhang, Y. Yun, H. Dai, J. Cui and X. Shang, “Graphs Regularized Robust Matrix Factorization and Its Application on Student Grade Prediction,” Applied Sciences (Switzerland), , pp. 1–19, 2020.

[7] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, ‘‘Educational data mining and learning analytics for 21st century higher education: A review and synthesis,’’ Telematics Informat., vol. 37, pp. 13–49.

[8] K. L. M. Ang, F. L. Ge, and K. P. Seng, “Big Educational Data Analytics: Survey, Architecture and Challenges,” IEEE Access, vol. 8, pp. 116392–116414, 2020.

[9] A. Hellas et al. “Predicting academic performance: A systematic literature review,” Annu. Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE. 2018, pp. 175–199, 2018.

[10] L. M. Abu Zohair, “Prediction of Student’s performance by modelling small dataset size,” Int. J. Educ. Technol. High. Educ., vol. 16, no. 1, 2019, doi: 10.1186/s41239-019-0160-3.

[11] X. Zhang, R. Xue, B. Liu, W. Lu, and Y. Zhang, “Grade prediction of student academic performance with multiple classification models,” ICNC-FSKD 2018 - 14th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov., pp. 1086–1090, 2018.

[12] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, “Improving accuracy of students ’ final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique,” pp. 1– 25, 2015.

[13] A. Polyzou and G. Karypis, “Grade prediction with models specific to students and courses,” Int. J. Data Sci. Anal., vol. 2, no. 3–4, pp. 159–171, 2016.

[14] Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran, “Machine Learning Based Student Grade Prediction: A Case Study,” pp. 1–22, 2017.

[15] I. Khan, A. Al Sadiri, A. R. Ahmad, and N. Jabeur, “Tracking student performance in introductory programming by means of machine learning,” 2019 4th MEC Int. Conf. Big Data Smart City, ICBDSC 2019, pp. 1–6, 2019.

[16] M. A. Al-Barrak and M. Al-Razgan, “Predicting Students Final GPA Using Decision Trees: A Case Study,” Int. J. Inf. Educ. Technol., vol. 6, no. 7, pp. 528–533, 2016.

[17] E. C. Abana, “A decision tree approach for predicting student grades in Research Project using Weka,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 7, pp. 285–289, 2019.

[18] F. Ahmad, N. H. Ismail, and A. A. Aziz, “The prediction of students’ academic performance using classification data mining techniques,” Appl. Math. Sci., vol. 9, no. 129, pp. 6415–6426, 2015.

[19] T. Anderson and R. Anderson, “Applications of Machine Learning to Student Grade Prediction in Quantitative Business Courses,” Glob. J. Bus. Pedagog., vol. 1, no. 3, pp. 13–22, 2017.

[20] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, “Educational data mining and analysis of students’ academic performance using WEKA,” Indones. J. Electr. Eng. Comput. Sci., vol. 9, no. 2, pp. 447–459, 2018.

[21] Verma, “Evaluation of Classification Algorithms with Solutions to Class Imbalance Problem on Bank Marketing Dataset using WEKA,” *Int. Res. J. Eng. Technol.*, vol. 06, no. 03, 2019.

[22] D. Berrar, “Cross-validation,” Encycl. Bioinforma. Comput. Biol. ABC Bioinforma., vol. 1–3, no. January 2018, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.

[23] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, “Using machine learning to predict student difficulties from learning session data,” Artif. Intell. Rev., vol. 52, no. 1, pp. 381–407, 2019.

[24] B. Predić, G. Dimić, D. Rančić, P. Štrbac, N. Maček, and P. Spalević, “Improving final grade prediction accuracy in blended learning environment using voting ensembles,” Comput. Appl. Eng. Educ., vol. 26, no. 6, pp. 2294–2306, 2018, doi: 10.1002/cae.22042.

[25] K. Srivastava, D. Singh, A. S. Pandey, and T. Maini, “A Novel Feature Selection and Short-Term Price,” Energies, pp. 1–18, 2019.

[26] K. Hamoud, A. S. Hashim, and W. A. Awadh, “Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis,” Int. J. Interact. Multimed. Artif. Intell., vol. 5, no. 2, p. 26, 2018.

[27] T. M. Barros, L. A. Guedes, and I. Silva, “education sciences Predictive Models for Imbalanced Data: A School Dropout Perspective,” no. i, 2019.

[28] T. Alam, C. F. Ahmed, S. A. Zahin, M. Asif, H. Khan, and M. T. Islam, “An Effective Recursive Technique for Multi-Class Classification and Regression for Imbalanced Data,” *IEEE Access*, vol. 7, pp. 127615–127630, 2019.

[29] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE : Synthetic Minority Over-sampling Technique,” vol. 16, pp. 321–357, 2002.

[30] C. Jalota and R. Agrawal, Feature Selection Algorithms and Student Academic Performance: A Study, vol. 1165. Springer Singapore, 2021.

[31] G. A. Sharifai and Z. Zainol, “Feature Selection for High-Dimensional and Correlation Based Redundancy and Binary,” pp. 1–26, 2020.

[32] Buenaño-Fernández, D. Gil, and S. Luján-Mora, “Application of machine learning in predicting performance for computer engineering students: A case study,” Sustain., vol. 11, no. 10, pp. 1– 18, 2019.

[33] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, “Learning k for kNN Classification,” *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 3, 2017.

[34] S. C. Gopi, B. Suvarna, and T. M. Padmaja, “High Dimensional Unbalanced Data Classification Vs SVM Feature Selection,” vol. 9, no. August, 2016.

[35] P. Nair and I. Kashyap, “Optimization of kNN Classifier Using Hybrid Preprocessing Model for Handling Imbalanced Data,” vol. 12, no. 5, pp. 697–704, 2019.

[36] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. U. Sattar, “Predicting student performance in higher educational institutions using video learning analytics and data mining techniques,” Appl. Sci., vol. 10, no. 11, 2020.

[37] Brodic, A. Amelio, and R. Jankovic, “Comparison of different classification techniques in predicting a university course final grade,” 2018 41st Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2018 - Proc., pp. 1382–1387, 2018.

[38] P. Brous, “administrative sciences Trusted Decision-Making: Data Governance for Creating Trust in Data Science Decision Outcomes,” 2020.

[39] H. Sun, M. R. Rabbani, M. S. Sial, and S. Yu, “Identifying Big Data ’ s Opportunities , Challenges , and Implications in Finance,” 2020.

[40] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, “Implementing autoML in educational data mining for prediction tasks,” Appl. Sci., vol. 10, no. 1, pp. 1–27, 2020.

SITI DIANAH ABDUL BUJANG received the B.S degree in Science (Computer Science) and M.S degree in Science from Universiti Teknologi Malaysia (UTM) in 2006 and 2010, respectively. She is currently pursuing the Ph.D. degree in Software Engineering at Malaysia- Japan International Institute of Technology, UTM in Kuala Lumpur. Her thesis focuses on the application of predictive analytics on student grade prediction in a higher education institution. From 2010 to 2019, she was the senior lecturer of Information and Communication Technology Department at Polytechnic Sultan Idris Shah, Sabak Bernam, Selangor, Malaysia. She has experienced in developing the polytechnic curriculum for Diploma in Information Technology (Technology Digital) 2.5 years' program. She is one of the book authors that contribute for the Department of Polytechnic and Community College Education. Her research interests include data analytics, predictive analytics, learning analytics, educational data mining and machine learning.

ALI SELAMAT (Member, IEEE) has also been the Dean of the Malaysia Japan International Institute of Technology (MJIIT), UTM, since 2018. An academic institution established under the cooperation of the Japanese International Cooperation Agency (JICA) and the Ministry of Education Malaysia (MOE) to provide the Japanese style of education in Malaysia. He is currently a Full Professor with Universiti Teknologi Malaysia (UTM), Malaysia, where he is also a Professor with the Software Engineering Department, Faculty of Computing. He has published more than 60 IF research papers. His H-index is 20, and his number of citations in WoS is more than 800. His research interests include software engineering, software process improvement, software agents, Web engineering, information retrievals, pattern recognition, genetic algorithms, neural networks, soft computing, computational collective intelligence, strategic management, key performance indicator, and knowledge management. He is on the Editorial Board of the Journal Knowledge-Based Systems (Elsevier). He has been serving as the Chair for the IEEE Computer Society Malaysia, since 2018.

ROLIANA IBRAHIM is the Director of Applied Computing at the School of Computing, formerly known as Faculty of Computing. Previously she was the Head of Information Systems Department for three years at the Faculty of Computing, Universiti Teknologi Malaysia (UTM). She received her PhD in the field of Systems Engineering from Loughborough University, MSc Computer Science from Universiti Teknologi Malaysia, BSc (Honours) Computer Studies from Liverpool John Moores University. She has been an academic staff at Information Systems Department since 1999 and was previously a Coordinator of BSc Computer Science (Bioinformatics) Program and Master of Information Technology (I.T. Management).

\

ONDREJ KREJCAR Ondrej Krejcar, is a full professor in systems engineering and informatics at the University of Hradec Kralove, Faculty of Informatics and Management, Center for Basic and Applied Research, Czech Republic; and Research Fellow at Malaysia-Japan International Institute of Technology, University Technology Malaysia, Kuala Lumpur, Malaysia. In 2008 he received his Ph.D. title in technical cybernetics at Technical University of Ostrava, Czech Republic. He is currently a vice-rector for science and creative activities of the University of Hradec Kralove from June 2020. At present, he is also a director of the Center for Basic and Applied Research at the University of Hradec Kralove. In years 2016-2020 he was vice-dean for science and research at Faculty of Informatics and Management, UHK. His h-index is 20, with more than 1300 citations received in the Web of Science, where more than 100 IF journal articles is indexed in JCR index. In 2018, he was the 14th top peer reviewer in Multidisciplinary in the World according to Publons and a Top Reviewer in the Global Peer Review Awards 2019 by Publons. Currently, he is on the editorial board of the MDPI Sensors IF journal (Q1/Q2 at JCR), and several other ESCI indexed journals. He is a Vice-leader and Management Committee member at WG4 at project COST CA17136, since 2018. He has also been a Management Committee member substitute at project COST CA16226 since 2017. Since 2019, he has been Chairman of the Program Committee of the KAPPA Program, Technological Agency of the Czech Republic as a regulator of the EEA/Norwegian Financial Mechanism in the Czech Republic (2019-2024). Since 2020, he has been Chairman of the Panel 1 (Computer, Physical and Chemical Sciences) of the ZETA Program, Technological Agency of the Czech Republic. Since 2014 until 2019, he has been Deputy Chairman of the Panel 7 (Processing Industry, Robotics, and Electrical Engineering) of the Epsilon Program, Technological Agency of the Czech Republic. At the University of Hradec Kralove, he is a guarantee of the doctoral study program in Applied Informatics, where he is focusing on lecturing on Smart Approaches to the Development of Information Systems and Applications in Ubiquitous Computing Environments. His research interests include Control Systems, Smart Sensors, Ubiquitous Computing, Manufacturing, Wireless Technology, Portable Devices, biomedicine, image segmentation and recognition, biometrics, technical cybernetics, and ubiquitous computing. His second area of interest is in Biomedicine (image analysis), as well as Biotelemetric System Architecture (portable device architecture, wireless biosensors), development of applications for mobile devices with use of remote or embedded biomedical sensors.

ENRIQUE HERRERA-VIEDMA (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 1993 and 1996, respectively. He is currently a Professor of computer science and A. I, and the Vice-President of Research and Knowledge Transfer, University of Granada. His H-index is 69, with more than 17 000 citations received in the Web of Science and 85 in Google Scholar, with more than 29 000 cites received. He has been identified as one of the World's most influential researchers by the Shanghai Centre and Thomson Reuters/Clarivate Analytics in both the scientific categories of computer science and engineering, from 2014 to 2018. His current research interests include group decision making, consensus models, linguistic modeling, aggregation of information, information retrieval, bibliometric, digital libraries, Web quality evaluation, recommender systems, block chain, smart cities, and social media. He is the Vice-President of Publications of the SMC Society and an Associate Editor of several JCR journals, such as IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, Information Sciences, Applied Soft Computing, Soft Computing, Fuzzy Optimization and Decision Making, Journal of Intelligent and Fuzzy Systems, International Journal of Fuzzy Systems, Engineering Applications of Artificial Intelligence, Journal of Ambient Intelligence and Humanized Computing, International Journal of Machine Learning and Cybernetics, and Knowledge-Based Systems. He is also the Editor-in-Chief of the Journal Frontiers in Artificial Intelligence (Section Fuzzy Systems).

HAMIDO FUJITA (Life Senior Member, IEEE) is Emeritus Professor of Iwate Prefectural University, Takizawa, Japan. He is currently the Executive Chairman of i-SOMET Incorporated Association, Morioka, Japan. He is Highly Cited Researcher in Cross-Field for the year 2019 and 2020 in Computer Science field, respectively from Clarivate Analytics. He received Doctor Honoris Causa from Óbuda University, Budapest, Hungary, in 2013 and received Doctor Honoris Causa from Timisoara Technical University, Timisoara, Romania, in 2018, and a title of Honorary Professor from Óbuda University, in 2011. He is Distinguished Research Professor at the University of Granada, and Adjunct Professor with Stockholm University, Stockholm, Sweden; University of Technology Sydney, Ultimo, NSW, Australia; National Taiwan Ocean University, Keelung, Taiwan, and others. He has supervised Ph.D. students jointly with the University of Laval, Quebec City, QC, Canada; University of Technology Sydney; Oregon State University, Corvallis, OR, USA; University of Paris 1 Pantheon-Sorbonne, Paris, France; and University of Genoa, Italy. Dr. Fujita is the recipient of the Honorary Scholar Award from the University of Technology Sydney, in 2012. He has four international patents in software systems and several research projects with Japanese industry and partners. He is the Emeritus Editor-in-Chief for Knowledge-Based Systems, and currently Editor-in-Chief of Applied Intelligence (Springer), He headed a number of projects including intelligent HCI, a project related to mental cloning for healthcare systems as an intelligent user interface between human-users and computers, and SCOPE project on virtual doctor systems for medical applications. He collaborated with several research projects in Europe, and recently he is collaborating in OLIMPIA project supported by Tuscany region on Therapeutic monitoring of Parkison disease. He has published more 400 highly cited Papers.

NOR AZURA MD. GHANI (member IEEE) is a Professor in Center for Statistical Studies and Decision Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia and Chair IEEE Computer Society Malaysia Chapter. Currently, she serves as Deputy Director (Research Impact) at Research Management Center, Universiti Teknologi MARA, Malaysia. Her expertise is big data, image processing, artificial neural network, statistical pattern recognition and forensic statistics. She is the author or co-author of many journals and conference proceedings at national and international levels.