International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

Predicting Academic Student Performance based on e-Learning Platform Engagement using Learning Management System Data

Siti Maisara Murniyati Muin, Fatimah Sidi*, Iskandar Ishak, Hamidah Ibrahim, Lilly Suriani Affendey
Department of Computer Science
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia, 43400, UPM Serdang
*Corresponding author: fatimah@upm.edu.my

Mohd Faizal Daud, Syemsul Bahrim Abdul, Rostam Abu Bakar
Infocomm Development Centre
Universiti Putra Malaysia, 43400, UPM Serdang

Abstract— The identification of at-risk students has become increasingly more significant as these students are in the precarious position of failing their courses. This study aims to achieve the objective of proposing a student performance prediction model to identify the stage of the course where at-risk students (students with the highest potential of failing their courses) can be identified based on student information system and learning management system data. The proposed student performance prediction model leverages machine learning methods to predict at-risk students, combining data from Universiti Putra Malaysia's (UPM) Student Information System (SIS) and learning management system (PutraBlast). Two experiments were conducted to satisfy the objective. The first experiment uses the full semester data to test multiple machine learning models to identify the best model for this dataset. In the second experiment, the dataset was separated into four course stages with four predictive models trained on each stage. Students. Results show that GB outperforms other classifiers when trained on the full semester data. However, classifier performance decreases when trained on data from earlier stages of the course. Hence, based on these results, the earliest stage to predict at-risk students is identified to be the W1—W12 stage.

Keywords- predicting student performance, learning analytic, learning management systems, e-learning

I. INTRODUCTION

The emergence of learning analytics has brought about monumental change in the way data is managed and presented in the perspective of education. Learning analytics has since seen a similar rise in the world of educational research, with researchers finally having the capabilities to gain access to a broad range of data encompassing various aspects of the teaching and learning process. Learning analytics was first defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" [1]. With the widespread use of virtual platforms, educational institutes are given more opportunities to employ learning analytics to measure the progress and effectiveness of students' learning process [2]. The prediction of student performance has been the focus of many studies in the subject of learning analytics. Educational institutes are motivated by learning analytics to unravel the complex relationships between the multiple variables and the circumstances that determine student success, or lack thereof. The prediction of student performance is viewed to identify and support students before it is too late, giving them motivation to perform better. It is also a prerequisite before attempting intervention, as the prediction model built would also allow the instructor to identify variables strongly associated with learning behaviors that would aid in structuring an effective interventions strategy [3]. Interventions have been deemed an important aspect of learning analytics, and its application has been shown to have a positive impact on student performance [4]. With machine learning prediction models reaching their maturity and data collection coming from various modalities, student performance prediction can be even more robust.

Based on the literature, researchers identify at-risk students as students at-risk of failing their courses, depending on the individual country's education system and grading scheme [5,6]. At-risk students are generally a concern for educators, and in learning analytics, prediction of student performance is one of the ways for educators to isolate these students and implement intervention strategies [3]. In the current educational landscape, Learning Management Systems (LMS) play a significant role in the management and organization of university courses, especially since it enables students to access course contents remotely. Universiti Putra Malaysia (UPM) implements its own LMS, called PutraBlast, as a learning hub for both its local and international students. Apart from that, the university also maintains a database, the Student Information System (SIS), for student demographic information, course information, and student academic records. In this study, data from both platforms are combined for the purposes of predicting student performance, making the identification of students at-risk of failing their courses before the end of the semester difficult for lecturers.

Currently, students' final grades are only revealed at the end of the semester. This leads to students being unaware of their performance for a large portion of the semester, resulting in a lack of urgency to seek help or gain motivation to perform better in upcoming exams and assessments. Apart from that, lecturers are also unable to gauge the effectiveness of their teaching methods and are unable to lend aid to struggling students, as they do not have a way to properly identify them until late in the semester. In the work in [3], the authors had mentioned that time is a crucial element in detecting at-risk students, and predictions in student performance should occur at different course lengths so at-risk students can be detected as early as possible. Based on the dataset used in this work, it is possible to divide the course into four stages based on the times when students receive their marks and perform prediction for at-risk students for each of these stages.

In this work, a student performance prediction model was proposed to identify the stage of the course where at-risk students (students with the highest potential of failing their courses) can be identified based on student information system and learning management system data. The student performance prediction model leverages techniques such as data preprocessing, feature engineering, resampling, and machine learning for at-risk student prediction. The organization of this paper is as follows. Section 2 reviews work related to student performance prediction and identifies two approaches to student performance prediction. Section 3 outlines the steps for the implementation of the proposed student performance prediction model. Section 4 analyzes the results, their implications, and the limitations encountered. Finally, Section 5 concludes this work.

II. LITERATURE REVIEW

Recently, the early prediction of student performance is quickly gaining the interest of researchers. While many researchers have been able to produce robust prediction models by utilizing data from the entire semester, there has been a necessity for early prediction as research in education now does not only encompass the understanding of factors relating to student failure, but also in implementing intervention schemes to prevent said failure [3, 7]. The work in [8] explored prediction techniques to combat early student dropouts through their behaviors and course engagement. The authors' model was trained incrementally on week-by-week data to predict dropout students every week and demonstrated 82% accuracy during as early as week 2 of the course. The work in [9] narrowed their focus on data from only LMS by extracting log data to predict student performance in solving LMS assignments. By dividing their dataset into different percentages of course completion, their best early prediction model using multilayer perceptron (MLP) displayed a steady increase in performance as the course progressed, going from 67% accuracy at 10% of the course stage to 87% accuracy at the halfway mark.

A method for predicting student performance in stages is proposed in [3], where they built multiple models for various percentages of the course delivery. They tested various machine learning and deep learning algorithms and determined the random forest algorithm to be the best-performing algorithm for their dataset. They generated multiple models of the random forest algorithm at 20%, 40%, 60%, 80%, and 100% stages of the course delivery. Their methods were proven to achieve respectable performance at as early as 20% of the course length, with scores of 75% precision, 84% recall, 79% F1- score, and 84% accuracy, which they deem as enough for instructors begin interventions. At 80% of the course length, the scores in various performance metrics see a substantial increase, falling within the range of 88% to 93%.

In [10], ensemble method is employed for student performance prediction at different course stages. The two datasets used by the authors consisted largely of assessments, originating from one Science and one Engineering course held by the University of Genoa. The prediction models they generated in the end consisted of four ensemble learners, trained on two individual datasets at 20% and 50% of the course duration. For the first dataset, their models achieved precision scores of 80% and 86.7% for the 20% and 50% stages respectively. However, their ensemble learner's precision using the second dataset shows better prediction for the earlier course stage than the later course stage. At 20% stage, the precision is 96.6%, while for the 50% stage the precision is 91.7%.

Early prediction for student performance is also proposed in [11] but it focused on enhancing the predictive accuracy of models trained on multiple weeks for 50 weeks through the manipulation of hyperparameters using Genetic Algorithms (GA). They trained several popular classifiers with hyperparameting tuning using data from Moodle. With tuned hyperparameters, their Multilayer Perceptron (MLP) model achieved an AUC of 0.915 during week 25, a significant increase from the AUC achieved by the best-performing classifier with no hyperparameter tuning, which is 0.849. However, they also noted that the model's performance begins to decline in later weeks due to local convergence.

A slightly different approach for early prediction for student performance is applied in [12]. In this work, in order to ascertain whether prediction in earlier weeks is possible with features extracted via Gini index from a dataset of student interactions in a blended learning course. They separated data by weeks and trained classification models for each segment. They proposed prediction model showed promising results, able to predict 74% percent of the unsuccessful students as early as week 3. Likewise, the paper by Adnan et al. (2021) [3] also separated their dataset into different stages of course delivery to detect at-risk students as early as possible during the progression of the course. Their implementation demonstrated promising results by attaining 84% accuracy at as early as 20% of the course length.

The work in [13], on the other hand, had studied the effectiveness of Deep Artificial Neural Networks (Deep ANN) in early student performance prediction using VLE data. Their proposed model was trained to predict three different cases which are: at-risk students, students with distinction, and students likely to withdraw from their courses with the whole dataset and quarterly clickstream data. Their deep ANN model achieved respectable performances when trained with the whole dataset with the following results: at-risk students were classified with 89% accuracy, possible withdrawals were predicted with 95% accuracy, and 'distinction' students were predicted with 86% accuracy. Their analysis of quarterly clickstream data also revealed that their model was able to predict at-risk students with 77% accuracy as early as the first quartile of the module. On the other hand, prediction using historical or prior academic data has also been tested by researchers. However, these predictions are only limited to the start of the academic year, and do not consider the students' achievements as the course progresses. Hence, they can identify students who are at-risk of failing in the upcoming academic year based on their previous results but are unable to detect at-risk students as the course is in session.

A study done in [14] utilized admission data from multiple academic institutions for their proposed automated machine learning approach for enhanced accuracy in student performance prediction. Their study focused on finding the most optimal model, and thus proposed AutoML as a method to loop through various classification models and their corresponding hyperparameters. They then combined the optimal models into one ensemble model that uses the voting strategy to determine the prediction. However, their automated machine learning approach achieved only an accuracy score of 75.9%, which is lower compared to other works.

In [15], deep neural network performance is tested, along with a range of machine learning methods to generate a predictive model for student performance in an upcoming Data Structures course. Their work placed particular emphasis on resampling methods to handle the imbalance present in their dataset. Their best results were attained with the deep neural network model using SMOTE, obtaining 89% accuracy.

In line with these studies, this work also attempts to identify the earliest stage of the course where at-risk students in UPM can be identified using data from SIS and PutraBlast using machine learning methods. Like all the works listed above, data for the full semester is divided into different course stages. The performance of the predictive model on each stage determines the feasibility of that stage to begin early prediction.

III. RESEARCH METHOD

To achieve the objective of identifying at-risk students, a student performance prediction model was proposed utilizing data from two systems: 1) Student Information System (SIS), and 2) learning management system (LMS) which is dubbed as PutraBlast. The proposed model combines data from both sources and trains a prediction model that predicts student performance at multiple stages of the course.

A. Dataset

The data used is from students in undergraduate courses in the Faculty of Computer Science and Information Technology (FSKTM), UPM, for two semesters of study in one academic session, which is the 2020/2021 session. The full dataset consists of 2416 rows for 705 students and 32 different courses. Data for the proposed student performance prediction model had been extracted from two sources, namely, the two systems utilized by the university:

1) Student Information System (SIS)
2) learning management system (PutraBlast).

Both systems offer different types of data that are significant towards the prediction model.

The data extracted from SIS contains student demographic information, course information, and academic records, with the full overview of the features given in Table 1. The dataset contains no missing values. For the prediction stage, several attributes are removed. Attributes relating to faculty, semester, and student matric number are removed because they have no value in the prediction model. The feature course name was also removed due to it representing the same entity as the feature course code. Finally, grades are removed after generating the target class. Initially, there were a total of 21 features including the target class in the SIS dataset. Following the removal of the features listed above, the remaining features form the SIS dataset numbered up to 16.

TABLE I. Dataset attributes and description from Student information System

Category	Variable	Description	Type
Demographic information	FAKULTI	Faculty the student is under	String
	STUD_MATRIC_NO	Student matric ID number used in UPM	String
	GENDER	Student's gender	String
	AGE	Student's age	Integer
	MARITAL_STATUS	Student's current marital status	String

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *05 July 2023* ***Revised:*** *25 August 2023* ***Accepted:*** *15 September 2023*

	COUNTRY	Student's country of origin	String
	TYPE_SPONSOR	Financial sponsorship the student is	String
		under	
Categorical information	SEM	The semester that the student is enrolled in the course	String
	KOD KURSUS	Course code	String
	NAMA KURSUS	Name of the course	String
	GROUP_NO	Student's course group	String
	JUMLAH JAM KREDIT	Total credit hours of course	Integer
	JAM KREDIT (KULIAH)	Lecture credit hours for course	Integer
Academic records	PERCENTAGE STUDENT W1-7	Marks percentage that the student obtained from week 1 to week 7	Float
	PERCENTAGE STUDENT W8-12	Marks percentage that the student obtained from week 8 to week 12	Float
	PERCENTAGE STUDENT CONTINUOUS	Marks percentage that the student obtained for continuous assessments,	Float
	ASSESSMENT	including assi gnments, qui zzes, exercise, and group studys	
	PERCENTAGE STUDENT FOR CARRY MARKS	Marks percentage that the student obtained for carry marks	Float
	PERCENTAGE STUDENT FOR FINAL	Marks percentage obtained for finals.	Float
	GRADE	The grade achieved by the student in the course	String

The learning management system data was extracted in the form of raw event logs for all interactions that the faculty staff and students of FSKTM have had throughout each semester. The rows are separated by timestamps and each timestamp has a description of the event that occurred and the matric number of the user that incurred the event. An event is marked as course viewed when a student accesses the homepage/dashboard of a course they have enrolled in. The dashboard can be considered as something of a directory for all the course materials pertaining to the course. It displays everything that has been posted by the lecturer of the course, including uploaded presentation slides, assignment materials, exercises, and references.

Submissions can also be posted by the students for assignments that have been uploaded in PutraBlast. Hence, the occurrence of this event indicates that the student is actively accessing the course materials, performing course activities, and uploading submissions. For this study, the cumulative course viewed data for each student is extracted on a weekly basis from Week 1 (the start of the semester) to Week 19 (the end of final exams week). Four variables are created based on the weekly course viewed data to reflect different course lengths. Week 1 to Week 7 course viewed data are aggregated for prediction in the first stage, Week 1 to Week 12 course viewed data for the second stage, Week 1 to Week 14 data for the third stage, and finally Week 1 to Week 19 data for the fourth stage.

In accordance with the approach taken by many similar studies in this field, the prediction of student performance in this study is depicted as a binary classification problem, where there are two classes [3, 5, 13, 14, 15]. For this study, the target class has a value of either pass or at-risk and is generated based on the grade feature in this dataset. The marking scheme used by UPM is depicted in Table 2, where the mappings of marks to alphabetical grades and value points are described. This marking scheme is obtained from Universiti Putra Malaysia (Academic Matters for Undergraduates) Rules 2014.

For this study, students with grades at C, C-, D+, D, and F are assigned to the at-risk class while grades A, A-, B+, B, B-, and C+ are assigned to the pass class. Grades C and below are selected as part of the at-risk class as the main goal of this study is to identify students who are in danger of failing their courses. While the university marks F as the failing grade (save for a few select faculties where the minimum grade is C), at- risk students should be identified before they reach the failing grade for lecturers to plan effective interventions. Also, grades contribute a large portion of the calculation for determining the grade point average (GPA) and cumulative grade point average (CGPA), and students who fail to get the minimum CGPA of 2.000 will be placed in probation. Hence, students who consistently get grades C and below for their courses may have GPAs and CGPAs pulled down until they are in danger of being placed in probation.

B. Data pre-processing

Predictive models are reliant on the quality of the data to generate the best results. As such, data preprocessing and data cleaning are integral steps in predictive modeling. During this stage, normalization, encoding, and oversampling are all applied on the data to prepare for the modeling stage.

- Normalization: The numerical features of the data are made up of various features that fall in different ranges, which may cause bias in the model by skewing towards features with larger values. Normalization transforms data so all the numerical values fall in the same range. In this study, all numerical values are transformed to values within the range of [0, 1] via min max normalization using MinMaxScaler() from the scikit-learn library.
- Encoding categorical variables: Very few machine learning algorithms can accept categorical variables in the form of string values or objects. For most algorithms, encoding categorical values into numerical

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

values is a necessity before training the model. For this study, the one-hot encoding scheme is used to encode all the categorical features into numerical via the OneHotEncoder() function from the scikit-learn library.

- Oversampling: There is an imbalance in data of the target class. Out of 2416 rows, there are only 173 rows in the at-risk category, making up only 7.16% of the entire dataset. However, it should be noted the dataset does contain multiple entries with the same student matric numbers as students can be enrolled in multiple courses in the same semester. As the model is liable to be skewed towards the majority class, an attempt is made to handle this imbalance by employing oversampling methods, as this method has been proposed in literature works that have encountered similar problem [14, 15, 16, 17] .The abundance of studies addressing this issue also indicates that dataset imbalance is common in predicting at-risk students, because in real life, at-risk students only make up a small percentage of the entire student population. This study faces a similar issue, and hence will attempt to rectify the data imbalance using Synthetic Minority Oversampling Technique (SMOTE), which has been applied by other researchers in several works [14, 15, 16, 17]. Also, there has been evidence in the form of a research paper shown in [15] proving that SMOTE outperforms other resampling techniques in their student performance prediction model. The Python imbalanced-learn library is used to implement SMOTE for this study.

TABLE II. Grading scheme used in the dataset

Marks	Grades	Value Point
80-100	A	4.000
75-79	A-	3.750
70-74	B+	3.500
65-69	B	3.000
60-64	B-	2.750
55-59	C+	2.500
50-54	C	2.000
47-49	C-	1.750
44-46	D+	1.500
40-43	D	1.000
39 or less	F	0.000

C. *Feature Extraction*

At-risk students run the danger of failing their respective courses when the semester ends. Hence, it is impertinent that they be identified as soon as possible for effective intervention to occur by notifying them of their performance and prompting them to receive additional aid or advice from their lecturers to help them improve. With that in mind, this study follows the procedure recommended in [3] and [10] by dividing the course into separate stages as detailed in Table 3. The percentage marks provided in the SIS dataset are used to determine the division of course stages. Each division combines the features from the SIS dataset with the course viewed data from the PutraBlast dataset.

TABLE III. Features taken from the SIS and Putrablast

Percentage of course length	Weeks	Data from SIS	Data from LMS
37%	W1—W7	Demographic information + Course information + W1-W7 student percentage marks	W1—W7 course viewed
63%	W1—W12	Demographic information + Course information + W1-W7 student percentage marks + W8-W12 student percentage marks	W1—W12 course viewed
74%	W1—W14	Demographic information + Course information + W1-W7 student percentage marks + W8-W12 student percentage marks + student percentage continuous assessment marks + student percentage carry marks	W1—W14 course viewed
100%	W1—W19	Demographic information + Course information + W1-W7 student percentage marks + W8-W12 student percentage marks + student percentage continuous assessment marks + student percentage carry marks + student percentage final marks	W1—W19 course viewed

D. *Modelling*

In this stage, predictive models are trained based on the feature-engineered data from the previous section. K-fold cross-validation technique is employed to validate all the models trained with K set to 10. For each iteration, the model is split into 10-folds, where 9-folds are used for training and 1-fold is used as the test/hold-out set. The scores obtained for each iteration are then averaged and used to evaluate the model performance. The models trained for this study include decision tree (DT), naïve bayes (NB), random forest (RF), K-nearest neighbors (KNN), support vector machine (SVM), logistic regression (LR), and gradient boosting classifier (GB). All their performances are evaluated in terms of accuracy, precision, recall, and F1-score. This study utilizes the Python scikit-learn library for the validation, training, and evaluation of all models.

In this study, a prediction model is proposed to predict student performance and identify at-risk students. Two experiments were conducted with the proposed prediction model, with Experiment 1 being focused on determining the best machine learning model using this dataset, and Experiment 2 being focused on the prediction for different stages of the course. The details for the experiments are as follows:

- Experiment 1 - Identifying the best machine learning model: Seven machine learning models which are Decision Tree (DT), Naive Bayes (NB), Random Forest (RF), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Logistic Regression (LR), and Gradient Boosting Classifier (GB) are trained on the full dataset and validated using K-fold cross-validation technique where K = 10. A comparison is then made between the models in terms of performance. The machine learning method that produces the model with the best performance amongst all those tested will then be used in the following experiment. Figure 1 describes the proposed model for the student performance prediction model in Experiment 1.
- Experiment 2 - Training the machine learning model on data on different stages of the course: The steps in terms of data preprocessing, oversampling, and model validation remain identical with Experiment 1. However, in this experiment, the full course data is divided into four stages: W1—W7 stage (37% course length), W1—W12 stage (63% course length), W1—W14 stage (74% course length), and W1—W19 stage (100% course length). The best machine-learning model identified in Experiment 1 is then trained on each of the four stages.

E. *Evaluation*

Metrics allow the user to evaluate the performance of the model based on the data it had been trained on. For this study, the trained models are evaluated based on accuracy, precision, recall, and F1-score [3]:

- Accuracy: Accuracy is calculated by dividing the number of correct predictions with the sum of all samples. While accuracy is one of the most frequent metrics referred to in determining the efficacy of predictive models, it may not be a suitable indicator in some cases, especially when the data is imbalanced, and the model is biased towards the majority class. Accuracy can be calculated using (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- Precision: Precision is the number of positive results correctly classified over the total number of positive results predicted by the model. Precision can be calculated using (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- Recall: Recall is the number correctly identified positive results divided by total actual positive results. Recall can be calculated using (3).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- F1-score: F1-score calculates the harmonic mean between precision and recall. It is a more accurate description of the effectiveness of predictive models, as it balances precision and recall. It is particularly useful in evaluating the model's performance in circumstances where the dataset is imbalanced. F1-score is calculated using (4).

$$F1-score = 2x\frac{Precision\ x\ Recall}{Precision+Recall} \quad (4)$$

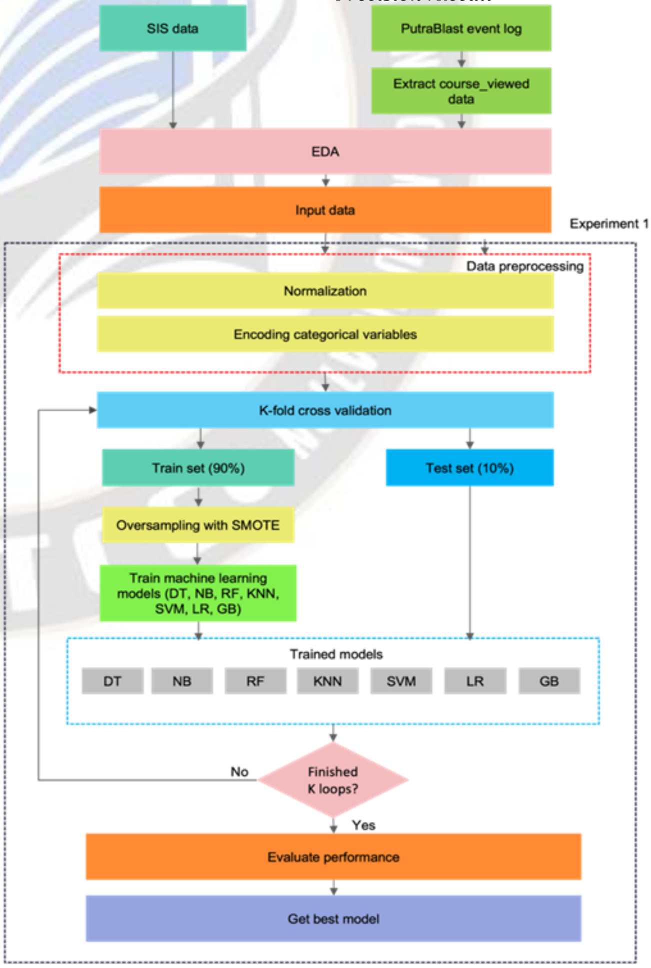

Figure 1. Proposed model for the student performance prediction model in Experiment 1.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

IV. RESULTS AND DISCUSSION

In this study, a student performance prediction model is proposed for predicting the performance of the students in University Putra Malaysia (UPM) and identifying at-risk students using data from both the Student Information System (SIS) and the learning management system (LMS), PutraBlast. Two experiments are conducted during the prediction stage: 1) prediction of student performance for the full course length and 2) prediction of student performance at various stages of the course.

Table 5 provides a detailed overview of the prediction results using several machine learning algorithms (DT, NB, RF, KNN, SVM, LR, and GB) with K-fold cross-validation where K = 10. The models are evaluated using accuracy, precision, recall, and F1-score. As the dataset used for this study is highly imbalanced in nature, accuracy is not an appropriate measure by which to judge the model's performance. The ability of the model to predict the positive (at-risk) class is significantly more important than its ability to predict the negative (pass) class, as students belonging to the former class are more in need of intervention compared to the latter. Thus precision, recall, and F1-score are more suitable performance metrics by which to judge the model's performance.

The prediction using data at full course length produced mixed results, depending on the algorithm used. It should be noted that the accuracy, precision, and F1-scores for NB are by far the lowest, indicating that the algorithm is unsuitable for prediction using this dataset. Conversely, GB produces the best results with a precision score of 91.79%, recall score of 98.48% and F1- score of 94.84%. The high precision score indicates that this model is good at identifying the at-risk class, while the high F1- score indicates that there is good balance between the precision and recall scores of the model. Its closest adversary is DT, which obtained 90.93% precision score, 95.15% recall score, and 92.90% F1-score. GB is therefore selected to proceed with Experiment 2 due to its scores surpassing those of the other models.

This study also aims to identify, using the available SIS and PutraBlast data, the earliest stage of the course to begin the identification of at-risk students. Table 5 shows the results of the implementation after the full dataset has been separated into several course stages. All different courses stages use GB (identified as the best-performing model in the previous experiment) as the prediction model to classify at-risk students.

Overall, the best prediction achieved by GB is when the model is trained with the full course data, and the worst prediction is during W1—W7. The decrease in performance was initially expected, as the model was trained on fewer features.

Based on Table 6, there is a decline in performance as the dataset is gradually stripped of the features from later weeks. For example, the W1—W14 stage of the prediction does not include the final marks for the course, as the students have yet to sit for the exam during this stage of the course. The values to take the harshest decline are the precision scores. The precision of the gradient boosting classifier has a difference of 33.1% from the W1—W19 to W1—W14 stage. The difference in recall and F1-score for the two stages are also calculated to be 14.99% and 26.38% respectively. The decrease in performance highlights the significance that final marks have on the prediction outcome using this dataset, despite that it only makes up 30% or 40% of the final course marks for the whole semester, depending on the course. The predictions for stages W1—W14 going to W1—W12, and W1—W12 going to W1—W7, see less of a drastic change in terms of the scores. These results further demonstrate the significance of final exam marks on predicting student performance for students in UPM.

In previous studies pertaining to student performance prediction at different course lengths, researchers have not formally identified a baseline for evaluation metric scores to identify the best time frame for early prediction to take place, however, some researchers such as in [3] have declared their early prediction models feasible when accuracy, precision, recall, and F1-scores fall within the range of 75% to 85%. In this study, the stage that achieved the closest results to the scores that have been accepted by the authors is the W1—W14 stage where this model was able to attain a recall score of 83.5%.

V. CONCLUSIONS

In this study, a student performance prediction model was proposed by using machine learning methods, combining data from UPM's Student Information System (SIS) and learning management system (PutraBlast) with the aim of identifying the stage of the course where at-risk students (students with the highest potential of failing their courses) can be identified based on student information system and learning management system data.

Two experiments were conducted for the prediction process. First used the full semester data to identify the machine learning method that produces the best predictive model out of all methods. The second experiment separated the dataset into four stages, W1—W7 stage (37% course length), W1—W12 stage (63% course length), W1—W14 stage (74% course length), and W1—W19 stage (100% course length) and trained four prediction models for each stage to identify the earliest stage of the course where at-risk students could be identified. Results show Gradient Boosting Classifier (GB) produced the best results out of all the machine learning methods on the full dataset with the accuracy score of 99.30%, precision score of 91.79%, recall score of 98.48% and F1-score of 94.84%.

Results during the second experiment show that there is a large divide in terms of precision score, recall score, and F1-score when the predictions were made with data from W1—W19 compared to data from W1—W14, where the main difference in the two stages is the absence of final marks in the latter stage.

These results highlight the significance of final marks in predicting student performance using this dataset.

By comparing the results obtained with the range of scores accepted in the work in [3] for early prediction, it is shown that the W1—W14 stage is the earliest stage for predicting students at-risk. However, based on UPM's education system, it is deemed that this stage is already too late to attempt prediction of at-risk students as they have already received their carry marks, which make up a large portion of their final grades. Taking this into account, an earlier stage (W1—W12 stage) is deemed more suitable to begin student performance prediction of at-risk students with the current available dataset.

In comparison with the work in [3], this study was able to achieve similar scores in terms accuracy, precision, recall, and F1-scores for the full course length. However, in the case of early prediction, the authors were able to predict with higher precision, recall, and F1-scores at as early as 20% of the course length, while the model proposed in this study was only able to attain a similar recall score during the W1—W14 stage (74% of the course length). The precision and F1-scores attained by the proposed model during that stage were also much lower than the authors' with the values of 58.69%, 68.46% respectively, while theirs were able to achieve the values of 75% and 79.3%.

ACKNOWLEDGMENT

This work is supported by Universiti Putra Malaysia through the university's Trust Account (6233232). We also would like to thank Infocomm Development Centre and Division of Admission and Division of Academic Governance from Universiti Putra Malaysia for their support.

REFERENCES

[1] Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5), 31–40G.

[2] Alexander, B., Ashford-Rowe, K., Barajas-Murph, N., Dobbin, G., Knott, J., McCormack, M., Pomerantz, J., Seilhamer, R., & Weber, N. (2019). *Horizon Report 2019 Higher Education Edition*.

[3] Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., Bashir, M., & Khan, S. U. (2021). Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. *IEEE Access*, *9*, 7519–7539. https://doi.org/10.1109/ACCESS.2021.3049446

[4] Wong, B. T. M. (2017). Learning analytics in higher education: an analysis of case studies. Asian Association of Open Universities Journal, 12(1), 21–40. https://doi.org/10.1108/AAOUJ-01-2017-0009

[5] Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Developing an early-warning system for spotting at-risk students by using eBook interaction logs. *Smart Learning Environments*, *6*(1), 4.

[6] Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. Computers & Education, 103, 1–15. https://doi.org/https://doi.org/10.1016/j.compedu.2016.09.005

[7] Burgos, C., Campanario, M. L., Peña, D. de la, Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. Computers & Electrical Engineering, 66, 541–556. https://doi.org/https://doi.org/10.1016/j.compeleceng.2017.03.005J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[8] Mubarak, A. A., Cao, H., & Zhang, W. (2020). Prediction of students' early dropout based on their interaction logs in online learning environment. Interactive Learning Environments, 1–20. https://doi.org/10.1080/10494820.2020.1727529

[9] Riestra-González, M., Paule-Ruíz, M. del P., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & Education, 163, 104108. https://doi.org/https://doi.org/10.1016/j.compedu.2020.104108

[10] Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. Knowledge-Based Systems, 200, 105992. https://doi.org/https://doi.org/10.1016/j.knosys.2020.105992

[11] Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. International Journal of Educational Technology in Higher Education, 16(1), 40. https://doi.org/10.1186/s41239-019-0172-z

[12] Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. International Journal of Educational Technology in Higher Education, 16(1), 40. https://doi.org/10.1186/s41239-019-0172-z

[13] Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104, 106189. https://doi.org/https://doi.org/10.1016/j.chb.2019.106189

[14] Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. Computers & Electrical Engineering, 89, 106903. https://doi.org/https://doi.org/10.1016/j.compeleceng.2020.106903.

[15] Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. IEEE Access, 9, 140731–140746. https://doi.org/10.1109/ACCESS.2021.3119596

[16] Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. IEEE Access, 9, 95608–95621. https://doi.org/10.1109/ACCESS.2021.3093563

[17] Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior, 73, 247–256. https://doi.org/https://doi.org/10.1016/j.chb.2017.01.047