Student Performance Classification using Support Vector Machine (SVM) with Polynomical Kernel on Online Student Activities

Muhammad Hareez Mohd Zaki, Mohd Azri Abdul Aziz*, Suhana Sulaiman and Najidah Hambali

Abstract—**The increasing usage of classification algorithms has encouraged researchers to explore many topics, including academic-related topics. In addition, the availability of data from various academic information management systems in recent years has been increasing, causing classification to become a technique that is in demand by educational institutes. Thereby, having a classification technique is important in researching the data on students' performance. The purpose of this study is to classify students' performance by using a polynomial kernel of Support Vector Machine (SVM) on online students' activities. A new dataset is proposed in this study, which consists of academic and student online behaviours that influence the students' performance. The proposed dataset also undergoes pre-processing stage to improve the accuracy and identify the significance of the proposed features. The experiment for SVM-POLY classification performance was set with a range of values on the parameters to be optimised by an optimisation algorithm, Grid Search. Classification accuracy, Precision, Recall and f1-score were applied to observe the result and determine the best classifier performance. The experimental results show that SVM – POLY, with a gamma value of 0.005, regularisation value of 0.1 and degree value of 1, come out with the best performance compared to a default value of SVM – POLY. The study is significant towards educational data mining in analysing the students' performance during online students' activities.**

Index Terms—**Student performance classification, SVM, polynomial**

I. INTRODUCTION

SUPPORT Vector Machine (SVM) is a supervised learning algorithm that is used for classification, regression and outlier detection [1]–[3]. The SVM classifier's goal is to find a linear hyperplane or"thickest hyperplane" that is also known as a decision boundary that separates the data in such a way that the margin is maximised [4], [5], [6]. Choosing the boundary that maximises the margin will decrease the chances for misclassification of unknown items in the future [5], [7]. SVM has been used as a classification technique in several applications, including educational purposes [1][2][8]. They include early detection of at-risk students [9], factors influencing student achievement [10] and causes of poor academic performance among students [11]. Research such as predicting student performance is important in every educational institutes in order to maintain the student's good performance.

Student performance has been a widely explored research topic in the past few years [1]–[10] due to exciting information that can help educators and students, mainly when sophisticated algorithms are applied [22]. The growing research on student performance has involved a variety of features in analysing the performance. Different authors used different features in conducting their student performance research. Some research used demographical features and students' personal information to study student performance [12], [19], [20], [23], [24]. Other research also used features that related to students' behaviour and mental development during their studies [1], [14], [25]–[27]. Commonly, most of the research used academic features such as CGPA, internal assessment, External assessment, Examination final score and extra co-circular activities of the student as prediction criteria [8], [12], [15], [28]–[30]. Thus, previous research showed that many categories of features could be used for analysing students' performance. There were no specific features to study their performance as students would be influenced by socioeconomic, psychological, and environmental factors [31].

In past research, many researchers conducted experiments related to students' performance classification. Most of the research used demographical factors and academic factors in conducting their studies. Demographical factors are factors involving the background of a student, such as gender, age and nationality. Academic factors are factors that are mostly used by researchers to evaluate the student's performance, for instance, grades, scores, exercises and assignments. Some of the demographical features also used parents-related information as examples, parents' responsible [1], parents' occupation [12], family size [12] and parental status [32]. Another factors that were used when conducting students' performance study were students' interest in a subject, engaging time and students' behaviour. These features were the examples of psychometric features [33].

Muhammad Hareez Mohd Zaki was with the School of Electrical Engineering, College of Engineering, UniversitiTeknologi MARA. (email: 2020480662@isiswa.uitm.edu.my).

*Corresponding author
Email address: azriaziz@uitm.edu.my

Students' performance based on demographical factors, such as parents' information and previous school performance, are not able to get an accurate result as the factors are not directly related to their performance, and student's performance can be improved and differ from their past performance.Research also found that students' demographic characteristics (i.e., gender, year of curriculum) and the external environmental features (i.e., type of admission, home location) were not determinant factors of their intrinsic and extrinsic motivation towards medical students' performance [34].Thus, the need to have a dataset with features that emphasize the students' performance is significant. Rather than the standard demographical features used for the evaluation, features such as academic and students behaviours should be emphasized. The use of these features highlights the performance of the students academically and also their mental development during their study journey.

This paper focuses on Support Vector Machine (SVM) model and a new proposed dataset. The core contribution is to suggest a new approach to evaluating students' performance by using a combination of academic and features related to students behavior.The proposed dataset consists of a few categories. Some features, such as accessing notes, exercises and assignments, have sub-categories to determine the students'behaviour and the outcomes of the assignments during the class. The difference in the behaviour may indicate students' attitudes, whether they are well-prepared students, in-class students or others. The academic factors will show the student's understanding and proficiency towards the course from the assignment results. In addition, the dataset needs to be pre-processed due as it is raw data. Data standardisation and feature selection processes are executed to pre-process the dataset. For data standardisation, the Z-score method is used to standardise data, while Chi-squared Test is used as a feature selection method to select significant features.

Support Vector Machine (SVM) are a type of binary linear classifier that aim to find the optimal decision boundary that separates two classes with the largest margin. There are some advantages of SVM compared to other classifiers. SVM works well with high-dimensional data. SVM can efficiently handle high-dimensional data because they only require a subset of the training data to compute the decision boundary, known as support vectors [35]. SVM can handle non-linear decision boundaries. SVM can use kernel functions to transform the input space into a higher-dimensional space, where a linear decision boundary can be found [36]. This makes SVM suitable for applications where the data is not linearly separable [25]. SVM are less prone to overfitting. SVM aims to find the decision boundary with the largest margin, which can reduce the risk of overfitting the training data [5]. Additionally, SVM use regularization parameters to control the complexity of the decision boundary and prevent overfitting [37]. SVM have several advantages over other classifiers, such as their ability to handle high-dimensional and non-linear data and their reduced risk of overfitting. This is suitable to be used with the proposed dataset. The objective of this research is to find out the effectiveness of the significant features of the proposed dataset and to study the effect of using optimization method, Grid Search towardsSupport Vector Machine (SVM). The strength of having the Grid Search method is it improves the classification model by finding the best parameter for the

model. The scope is limited to enhancing the classification model and using the proposed dataset as data for this research. The shortcomings of having the proposed dataset are the imbalance in the number of failed and passed students.

In this paper, a classification of the student's performance using the proposed dataset has been performed. SVM with a polynomial kernel was chosen as a classifier to be used during the classification process. The method applied was discussed thoroughly in this paper. This research makes contributions to researchers and practitioners in students' performance classification by introducing a new dataset that focuses on academic factors and student behaviour factors rather than demographical factors to analyse the students' performance.

The rest of this paper is organised as follows. Section II presented the literature review of other research on the features and methods used to study students' performance. Section III elaborated on the theoretical background of the methods used in this research, data collection and flowchart of the process. Section IV was on the result of the research. The conclusion of the paper is presented in section V.

II. RELATED STUDIES

Studies on students' performance are still progressing until the present day. It is important for universities or any educational sector to adapt to the students' growing development throughout their studies. Many data are used for student performance evaluation. However, no specific criteria or features can study students' performance as each student has varieties of personalities and backgrounds with a different history that may impact their future studies [15]. Commonly, researchers used classification, a part of the data mining function, to study and gain more insights regarding student performance classification.

Students' online activities already exist in many research. In [24], the researchers used three different feature categories to analyse students' performance; demographic, engagement and past performance. The results showed that engagement and past performance had the highest accuracy in affecting the students' performance. Other research also used a few features that belong to the academic and psychometric categories, such as the number of view course content [38], the number of each access to learning dashboards [38], total time spent online [38], student engagement in the course [12][39], online session assessment [39], students' activity log [40] and more. The researchers widely use both these categories to find the factors that can determine the students' performance.

The Chi-squared test is a univariate feature selection algorithm used to test independence and estimate whether the class label is independent of a feature [47]–[50]. This test has two main phases of the algorithm. In the first phase, consistency checking is performed as the stopping criteria, whereas in phase two, the results of phase one are checked. It continues until there remain no attributes for merging [49]. According to [51], Chi-squared Test is utilised to obtain the significant connection between two categorical variables. In addition, this method belongs to the filter method category. It uses a 'proxy measure' calculated from the general characteristics of the training data to score features or feature subsets as a processing step prior to

modelling [52]. The advantages of having the filter method are that it can run faster, and any features chosen by the method can be passed to any modelling algorithm [52].

The chi-squared test has certain limitations and assumptions when interpreting the results. The first limitationn is sample size [41]. The chi-squared test requires a sufficiently large sample size to produce reliable results. If the sample size is small, the test may not be able to detect significant associations, even if they exist. By increasing the sample size can improve the power of the test and make it more reliable. Other than that, the chi-squared test assumes that the observations in each category are independent of each other . This means that each observation must be independent of all other observations [41]. The chi-squared test assumes that the data are independent, categorical, and randomly selected. If these assumptions are not met, the test may not produce valid results.

The chi-squared test evaluates each of the features by the value of chi-squared. The higher the chi-squared value, the higher the contribution to the students' performance. Chi-squared Test was also used to determine if the two or more classifications of the samples were independent or not [41]. To confirm whether the chi-squared value is high enough to reject the null hypothesis, p-values are considered. The null hypothesis of the Chi-squared test claimed that by having a p-value higher than 0.05, there is no relationship between the feature and the target feature (independent). Otherwise, the feature is considered dependent on the target variable. The null hypothesis is rejected if and only if p is less than α [42]. Due to that, the Chi-squared method was used to find out the Chi-squared value and its p-value. For the Chi-squared test, the red dotted line was also created in the scatter plot in the result section to show the p-value equal to 0.05.

In education, it has been used to analyze student performance, investigate the effectiveness of educational interventions [43], and explore the relationship between different demographic factors and academic outcomes [44]. One example of the Chi-squared test being used in education is in analyzing student performance in exams. By comparing the distribution of correct and incorrect answers for different groups of students, the Chi-squared test can help identify potential biases or disparities in the exam, such as the exam being more difficult for certain groups of students. Another example is in evaluating the effectiveness of educational interventions. For example, the Chi-squared test can be used to analyze the impact of different teaching methods on student learning outcomes. By comparing the distribution of grades before and after the intervention, the Chi-squared test can help assess the effectiveness of the intervention.

A research study used a few feature selection algorithms in predicting students' academic performance [48]. The result of the study recommended the use of the Chi-squared Test for the feature selection process as it is an important technique in predicting student performance. Another research in identifying predicting factors of student retention had shown that the use of feature selection algorithms such as the Chi-squared Test, Info Gain Ratio and Correlation feature selection had significantly improved the accuracy of the result from 80.50% to 92.09% by

reducing the number of features used in the research [53]. In addition, research on the comparison of feature selection methods for academic data analysis also concluded that the best feature selection method before creating predictive models was the Chi-squared statistic method [54]. Besides, a study stated that Chi-squared Test extracted the most appropriate high-ranked features having a significant role in student grade prediction [55].

SVM is considered a good classifier because of its high generalisation performance without adding a priori knowledge, even when the input space dimension is very high [36]. If the data are not linearly separable, the researcher can choose a few kernels function to adapt to the situation, consisting of kernel functions from an array of linear, radial basis, sigmoid second-order multiple, polynomial and reverse second-order kernel [25]. The idea of SVM is to find the points of data known as support vectors, which define the widest linear margin between two classes. Two tricks can be performed to the non-linear class boundaries; first, mapping the data to a higher dimension, where there is a linear boundary and second, allowing misclassification by defining a soft margin. A compromise of these two approaches will avoid overfitting and preserve good classification accuracy [26], [37], [45].

The application of the SVM model in the educational field was widely used due to its good performance [46][47][48][49]. A finding was declared by a systematic review [46] stating that SVM had minimum accuracy of 80% and maximum accuracy of 98%. There are few applications of SVM kernels in the classification area. For instance, linear SVM was used for the prediction of student performance in school-based education [47], and SVM with Radial Basis Function (RBF) was used to investigate students' pre-admission academic profile and final academic performance by predicting students' performance [48], and Polynomial SVM was also used in fostering students' performance by accessing learning quality [49].

SVM-POLY as SVM with the polynomial kernel is one of the classification algorithms that is frequently used in the classification area, and several findings from existing studies have shown the quality of having SVM-POLY as a classification algorithm. For instance, SVM-POLY has been used for accessing learning quality by estimating students' performance at examinations[49]. The use of SVM-POLY can indicate whether the dataset had linear or non-linear characteristics. In [49], the high accuracy achievement, 82.6% by a polynomial of degree 3, showed that the dataset had non-linear characteristics. A study [50] that used the same classification algorithm stated that the use of a UCI dataset, such as the Iris Dataset resulted in 96.64% accuracy. Besides, an experiment on students' performance measures had been successfully executed with an accuracy of 97.62%, and the researchers stated that the methodology could be implemented to help teachers and students to improve learning quality and student performance by taking a significant decision at the right time [51].

To improve the performance of the SVM, the hyperparameter needed to be optimised. One of the optimisation methods that are often used by researchers is theGrid Search method. Grid

Search applies the brute-force method to generate candidates from the grid of parameter values specified with the parameter [52]. Another research [53] stated that the method works by searching the combination of parameters that are specified and then declaring the best parameter value based on the minimum classification error to build the classification model. There are few applications of the Grid Search method to improve the classifier model in previous research. A study on student pass rates[54] using optimised SVM and decision tree showed promising results when the model could achieve above 90% accuracy value. Grid Searchwas also applied in E-commerce [55], which resulted in an improvement in the accuracy of the SVM model by 26.4%. Other than that, a study on coronary artery diseases [56] also improved the SVM model and resulted in 88% accuracy. Previous research showed that theapplication of the Grid Search method and SVM were often used by researchers. The improvement of the SVM model when using the Grid Search method can guarantee the effectiveness of the Grid Search method in tuning the parameter.

III. METHODOLOGY

This section is separated into several sections, beginning with the theoretical background of the feature selection techniques used in the research and followed by the flow of the process. The intelligence technique was then conducted in Python in Anaconda Navigator (Anaconda3) platform.

A. Theoretical Background

1) Chi-squared Test

Pearson's chi-square test of independence is a statistical method used to identify the degree of association between variables [57]. This technique is applied to analyse the dependency of all attributes (factors) on the outcome attribute. So chi-square method proves helpful here. For a contingency table with 'r' rows and 'c' columns, the formula for finding the chi-square is given in equation (1).

$$x^2 = \frac{\Sigma(observed - expected)^2}{expected} \tag{1}$$

The predetermined level of significance is taken as 5%, and P-values are identified using the chi-square values for each attribute.

According to [58], Chi-square is used for assessing two kinds of comparing: tests of independence and tests of goodness of fit. In feature selection, the test of independence is assessed by chi-square and estimates whether the class label is independent of a feature. The Chi-square score with C class and r values is defined as (2):

$$x^2 = -\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-\mu_{ij})}{\mu_{ij}} \tag{2}$$

n_{ij} is the amount of samples value with the ith value of the feature.

$$\mu_{ij} = \frac{(n_{*j}n_{i*})}{n} \tag{3}$$

n_{i*} is the amount of samples with the ith the feature value.
n_{*j} is the amount of samples in class j.
n is the number for samples.

2) Support Vector Machine

Support Vector Machine (SVM) is a powerful machine-learning tool which was proposed by [59] and has become more attractiveto machine-learning researchers and the community. The algorithm of SVM has been proven effective to be used in regression and classification methods. Based on previous studies from [51], the study reported that the SVM is generally able to result in the best accuracy of classification compared to other methods. Also, SVM can perform linear and non-linear classification with high efficiency. However, the challenge of using SVM in classification or regression is to find the best penalty term parameters and kernel parameters. It is because SVM is very sensitive to the parameter used [60].

The main concept of SVM is illustrated in Fig. 1. SVM techniques can be explained as a process of separating two different classes in feature space, such as positive and negative. Discovering the hyperplane that separates those classes is the main issue in the classification process, where this hyperplane depends on the maximal margin [61]. Several optimisation problems can be tackled by SVM, such as regression issues and data classification issues. Distinguishing the positive and negative data points requires identifying a hyper-plane to separate the data points into two classes. Fig. 1 shows the hyperplane that is considered asa decision boundary for linear SVM.
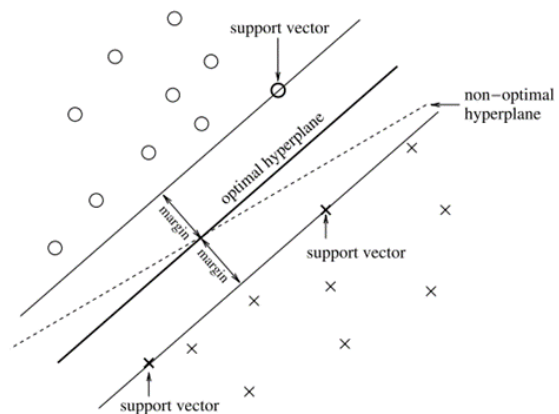

Fig. 1. Illustration of SVM hyperplane.[62]
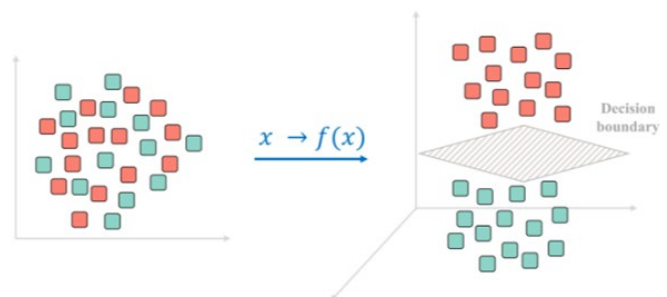
a) Polynomial Kernel of SVM

Fig. 2. An SVM example for non-linearly separable data with the kernel trick

The basic function of the polynomial is not featuring input samples to regulate their resemblance yet combinations of these. Such mergers were named as interaction features for the regression analysis. The polynomial regression will be similar to the feature space of the kernel itself if the combinatorial blow-up number of parameter to be recognised [63]. The mathematical equation for a polynomial is represented in (4) below. The features will correlate with the logical conjunctionof the input if and only if the input is binary-valued, as in Booleans. Despite its limited district performance, the polynomial kernel is a universal kernel function and directional.

$$K(x, x_i) = [\gamma * (x \bullet x_i) + coef]^d \qquad (4)$$

where K(x, xi) is the kernel function for polynomial and d = dimension of mapping function grows. By default, the degree of d is set to 3, γ is 1/k (is the number of class), and coef is 0.

3) Grid Search

The working principle of this method was to search the combination of parameters in the given length of a region that was set up. The method will then issue the best parameter based on the minimum classification error to build the classification model. There were some advantages of having this method; (1) this method did not perform extensive parameter search like any approximation methods or heuristics;(2) The computational time to find the optimal parameter values by the Grid Search is not much more than those by advanced methods [53]; (3) The Grid Search can be easily parallelised because each pair is independent [53]. In addition, the Grid Search is an iterative method, and many advanced algorithms are based on iterative processes.

4) Performance Criteria

A. Confusion Matrix

The matrix in which the matrix row designated the target/real class and the matrix column shows the targeted class referring to the confusion matrix as in TABLE I [63]. The matrix comprises four parameters which are false negatives (fn), false positives (fp), true negatives (Tn) and true positives (Tp) or, the matrix form can be indicated as $\begin{bmatrix} Tn & Fp \\ Fn & Tp \end{bmatrix}$. $Tp$is the data exactly grouped data to passed students, Tn is the data exactly correctly data to failed students. Fp is the data misclassified to passed students and Fn is the data misclassified to failed students. The example of a confusion matrix is as follows:

TABLE I
A CONFUSION MATRIX [64]

Data Class	Classified as High	Classified as Low
High	True High (Tp)	False Low (Fn)
Low	False High (Fn)	True Low (Tn)

The result from the SVM model using a polynomial kernel was tested into the model and was accepted once it passed the performance criteria as a confusion matrix, precision, recall, f1-score and ROC.

i) Accuracy

The overall capability of a classifier is evaluated through accuracy[63]. It can be written as shown in (5).

$$Accuracy = \frac{Tp+Tn}{Tp+ Tn+ Fp+ Fn} \qquad (5)$$

ii) Precision

Precision represents the positive test cases that were correctly classified as predicted[65]. The mathematical equation for precision as shown in (6).

$$Precision = \frac{Tp}{Tp+ Fp} \qquad (6)$$

iii) Recall

Recall represents the fraction of actual positive test cases that were properly classified[65]. The mathematical equation for the recall is as shown in (7):

$$Recall = \frac{Tp}{Tp+ Fn} \qquad (7)$$

iv) F1-score

F1-score represents the precision and recalls harmonic mean[65]. The mathematical equation for f1-score as shown in (8).

$$f1 - score = \frac{2Tp}{2Tp+ Fp+Fn} \qquad (8)$$

B. Data Collection

The new data is extracted from a website that performs online learning for computer courses. The collected students' data consisted of 101 students from semester three of the School of Electrical Engineering of UniversitiTeknologi Mara. Most of the data were based on students' activities during online learning sessions in a course. TABLE II depicts the inputs as classification features for this work.

TABLE II
FEATURES USED IN THE RESEARCH

No.	Features	
1.	Accessing course materials	Before class (Before Notes)
2.		In class (In Notes)
3.		During other class (Other Notes)
4.		After class (After Notes)
5.	Note length	
6.	Accessing exercises materials	Before class (Before Exercise)
7.		In class (In Exercise)
8.		During other class (Other Exercise)
9.		After class (After Exercise)
10.	Tutorials Sections	Correct 3 and above (C3AA)
11.		Answered all questions (AAQ)
12.		Wrong before correct (WBC)
13.	Test 1 (Pass/Fail)	

The features for the classification of the data are examined based on the students' online activities during their online classes. As exhibited in TABLE II, accessing course materials,

accessing exercises materials and tutorial sections are considered as the behaviour and learning outcomes of the students. Both courses and exercises feature datasets that record the students' behaviourin accessing the materials. In addition, these behaviours are divided into four groups; before class, in class, during other class and after class. These groupings are to isolate whether the students accessed or revised the materials before class, in class, during other classes or after the class. For note length, students' notes were recorded on the learning website. Also, there are three categories for tutorial sections: 'correct 3 and above', 'answer all questions' and 'wrong before correct'. These categories are to demonstrate the course's understanding and learning outcomes of the students when the assignments are assigned by the tutor. The output variable for this research is 'test 1'. This variable determines whether the students pass or fail.

C. Experiment stage

1) Data Standardization (Z-score method)

The standardised variable is called the z-score. Other terms include z-values, standard scores, and normal scores. These terms can be used interchangeably, as in this manuscript. The z-score measures exactly how many standard deviations are above or below the mean of a data point. The above-average data point has positive standard scores, while others have negative standard scores. Standard deviation and z-score values are shown as representative in normal distribution [66]. According to [67], the Z-score is a form of standardisation used for transforming normal variants into standard score forms. Given a set of raw data Y, the Z-score standardisation formula is defined as (9):

$$X_{ij} = Z(X_{ij}) = \frac{X_{ij} - \bar{X}_j}{\sigma_j} \qquad (9)$$

where, $\bar{X}_j$ and σ_j are the sample mean and standard deviation of the *jth* attribute, respectively. The transformed variable will have a mean of 0 and a variance of 1. The location and scale information of the original variable has been lost, and one important restriction of the Z-score standardisation is that it must be applied in global standardisation and not in within-cluster standardisation [67].

2) Feature Selection Process

The experiment started by the process of data collection. The data collected during online learning were in raw data. The data needed to be standardised for the experiment. The importance of having data standardisation when pre-processing the data was to scale or control the variability of the dataset. Thus, Z-score method was chosen as the method for standardising the data. Feature selection is performed after data standardisation to identify the significant features that affect the students' performance. The method used for feature selection process in this paper was Chi-squared Test from filter method categories [68]. The result of the pre-processing stage will be displayed through a table in results section of this paper.

3) Classification Process

After the feature selection process, the data were divided into 80:20 training and testing ratio respectively. The training data was used to train a classification model to predict the outcome for the experiment prediction. The testing data was used to measure the performance, such as accuracy, of the classification model that was trained.

For this phase, there were two classifications that had been done in the experiment. The differences between the classifications were on the application of feature selection process. One of the classifications did not involve the use of feature selection process while the other classification involved the application of the process. The purpose of having the two different classifications were to test the features of the proposed dataset and compare the results when using all features and only significant features.

The experiment used Support Vector Machine – Polynomial, SVM-POLY as classifiers for the classification process. Polynomial of SVM had a few hyperparameters such as Gamma, C and degree. These hyperparameters needed to be adjusted to maximise the classification model's predictive accuracy. Thus, the Grid Search method was used as an optimisation method to optimise the hyperparameter of the SVM-POLY. The output was then generated, and the data were classified. The output data that did not meet the performance criteria set will go through the classifying process again until it met the criteria. The method used to evaluate the performance criteria was by using confusion matrix, precision, recall and f1-score. The flow of the system is summarised as shown in Fig. 3.
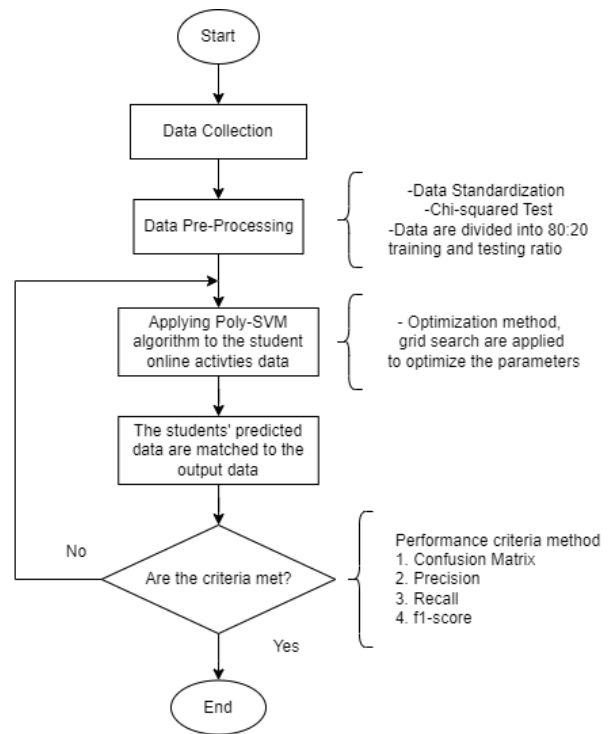

Fig. 3. The Flowchart of the Process

IV. RESULTS AND DISCUSSION

A. Feature Selection Test

The research was performed by using collected data from an online learning website. The data consisted of UniversitiTeknologi MARA (UiTM) students that took a computer class. The main classifier used in this study is SVM with polynomial kernel (SVM-POLY). The proposed data that was used in this research comprised two datasets. The first dataset was the initially proposed dataset without performing any feature selection process, as shown in TABLE II. The second dataset had a few features that were chosen by a feature selection method. The feature selection method to be used for this research is the Chi-squared test. Due to the different datasets used, a comparison can be made by comparing the accuracy of the results. The different accuracy between the original dataset and the dataset that undergoesthe feature selection process would show the significance of those features towards the performance of the students. Before executing the feature selection process, the dataset wasstandardised by using the Z-score method.

1) Chi-squared Test

As mentioned previously, the Chi-squared test evaluates each of the features by the value of chi-squared. The higher the chi-squared value, the higher the contribution to the student's performance. The differences in the features chosen by the Chi-squared test are presented in the bar plot in Fig. 4. The features in descending order consisted of 'After-class notes', 'During-other-class notes', 'In-class notes' and 'Before-class notes'. The test selected 'After-class notes' as the most significant feature as it has the highest chi-squared value.

Chi-squared test also analyses the dependency of all features on the outcome feature [76]. This proposed dataset is suitable for this method as the features in the dataset will be evaluated to investigate the correlation and the contribution of the features towards the outcome feature, 'Test 1'. The larger the chi-squared value, the higher the probability of a significant difference in the feature. As depicted in Fig. 4, the findings infer that the highest value of chi-squared implies a highly influential factor. Thus, 'After-class Notes' is the most significant feature that affects the student's performance compared to other features.
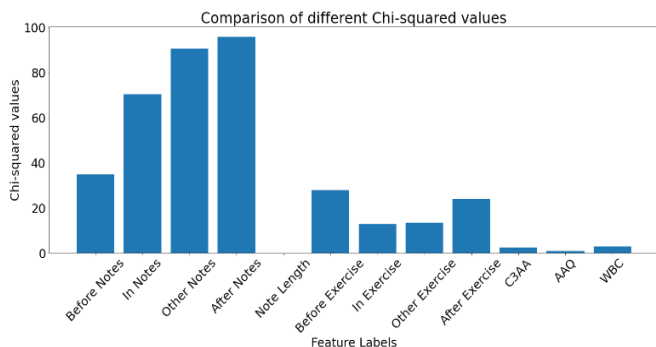

Fig. 4. The Bar Plot of Chi-squared Test

In this study, the feature was on how the students accessed the study materials. The reason behind the accessed study materials were the limitation of this study. It could be because of the students wanted to study, writing additional notes or only printed the notes for upcoming classes.

In past research, there were some studies that used 'After-class Notes' feature to evaluate the students' performance. A study found that students would continue using strategies to finish their notes after class [69]. The strategies of after-class note-taking was frequently used by the students to make elaboration on the notes as there was sufficient time after the class [70]. Another study support that students might consult their lecturers or classmates, compare their notes and adding additional content to notes after class [71].

B. Classification Stage

The kernel used in this research was polynomial kernel. The configuration of the polynomial parameter was optimised by using the optimisation method, Grid Search. The method were capable of tuning the parameter to give improved accuracy result. To tune the parameter, some parameters needed to apply the value range first as guidelines for the Grid Search method to search for the best parameter. For polynomial, there are four parameters that can be configured to improve the classifier;C, γ, r and d [53]. The coefficient parameter, r was kept constant with the default value, 0.0. For the degree parameter, three different values were set (1, 2 and 3). These values are commonly used by some researchers [53], [72], [73]. In addition, the regularisation parameter, C and gamma parameter, γ were configured to a specific range of values according to research [74]. The configurations were done as follows: C (0.1, 0.2 ...1.0) and γ (0.005, 0.010...0.060). Each parameter had its own role in improving the classifier. Parameter γ had an influence on classification outcomes as it affects the partitioning in the feature space [53]. An excessive and disproportionately small value of parameter γ would result in over-fitting and under-fitting, respectively.The degree value was defined as the dimension of the mapping function growth [75]. The dimension of the mapping function grows with the value of the dimensionality of a kernel. Parameter C acts as a free parameter in the polynomial that trading-off the impact of higher-order versus lower-order terms [76]–[78]. If C is too large, then error minimisation is predominant.Otherwise, the margin maximisation is emphasised when C is too small.

The results showed that the classification accuracy for both datasets was the same. Before the tuning process, the classifier in both datasets had an accuracy score of 90.5%. After the tuning was performed by the Grid Search method, the accuracy score significantly increased from 90.5% to 95.2%. Hence, it was proven that by tuning the parameter of the polynomial kernel, the performance of the classifier could be improved. During the tuning process, the Grid Search method finalised the search for the best parameter value for the polynomial as follows: C = 0.1, γ = 0.005 and degree = 1. Hence, Fig. 5 shows the result of the classification accuracy before and after the tuning process.
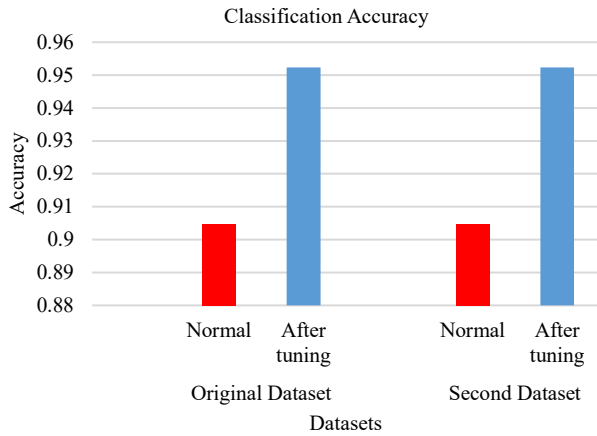
Fig. 5. Classification Accuracy

Fig. 6 showed the other performance criteria that were used to evaluate the performance of the classifier. The figure also tabulated the performance results obtained by the classifier below the bar chart. Both datasets had the same results before and after the tuning process in terms of precision, recall and f1-score despite having different numbers of features used. Precision was used to evaluate the positive test cases that were classified correctly by the classifier [65]. Classifiers in both datasets had the same precision value of 0.95, which indicated that they were able to make most of the classification correctly. For the recall, there was an increase after the tuning process for both datasets from 0.95 to 1. Recall was defined as the capability of a classifier to classify low-quality [63], [65]. For the f1-score, there was a slight increase from 0.95 to 0.98 for both classifiers after the tuning process. The f1-score represented the precision and recalled harmonic mean [65].

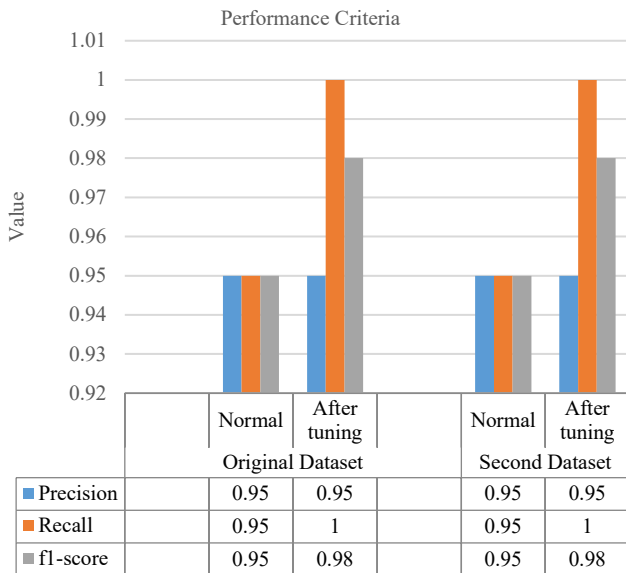|		Normal	After tuning		Normal	After tuning
			Original Dataset			Second Dataset
Precision		0.95	0.95		0.95	0.95
Recall		0.95	1		0.95	1
f1-score		0.95	0.98		0.95	0.98

Fig. 6. Performance Criteria

Another method of displaying the summary of prediction results on the classification was by using confusion matrix.

From the confusion matrix, there were two labels; true label and predicted label. The labels acted as indicators for the researcher to observe the number of correct and incorrect predictions made by the classifier. The results of confusion matrix for classifiers in both datasets were slightly different. For the original datasets as shown in Fig. 7 a), the true positive value was 19 out of 20 actual positive which indicated that the classifier was able to correctly classify 19 passed students and misclassify 1 failed students. Thus, the TP value was 19 and FP value was 1. The model also misclassified 1 passed students which made FN equal to 1. For TN, it was 0 as there was no classification of failed students occurred. After the tuning process, the original dataset had true positive value of 20 out of 21 actual positive. This showed that there was 20 passed students that managed to be classified correctly by the classifier and only misclassify 1 failed students. Thus, the TP value was 20 and FP was 1. For TN and FN, the value was 0 as there was no classification and misclassification of failed students occurred respectively. For the second datasets, both confusion matrix in Fig. 7 c) and d) showed the same result as original dataset after tuning process.

The performance of using original dataset was better when the method involved the use of optimization method. For the second dataset, the outcome showed that the classifier achieved same results before and after the optimization process. The use of significant features in this dataset had improved the performance of the SVM model. In addition, the use of optimization process was supposed to give more improvement on the model. The unaffected result was due to the optimization process that already achieved the near-optimal in terms of their parameter configurations, and no further improvement could be achieved through parameter tuning.

A few findings can be concluded from the result. The Grid Search method improved the accuracy of the classification model. The absence of the Grid Search method caused the hyperparameter of SVM-POLY to be in default configuration of the software used. The use of the Grid Search method optimised the hyperparameter of SVM-POLY and thus improving themodel's accuracy from 90.5% to 95.2%. In addition, the second dataset that used Chi-squared Test also managed to choose important features that affect student performance which resulted in the same accuracy as the original dataset after the tuning process.

There was a study that showed an SVM model with and without feature selection performed better than grid search in the study's cases [79]. Despite that, our study showed that by using Grid Search method, the SVM model can be improved even without having feature selection process as shown in the original dataset. This proved that an optimization method is important in improving and optimizing the classifier model.
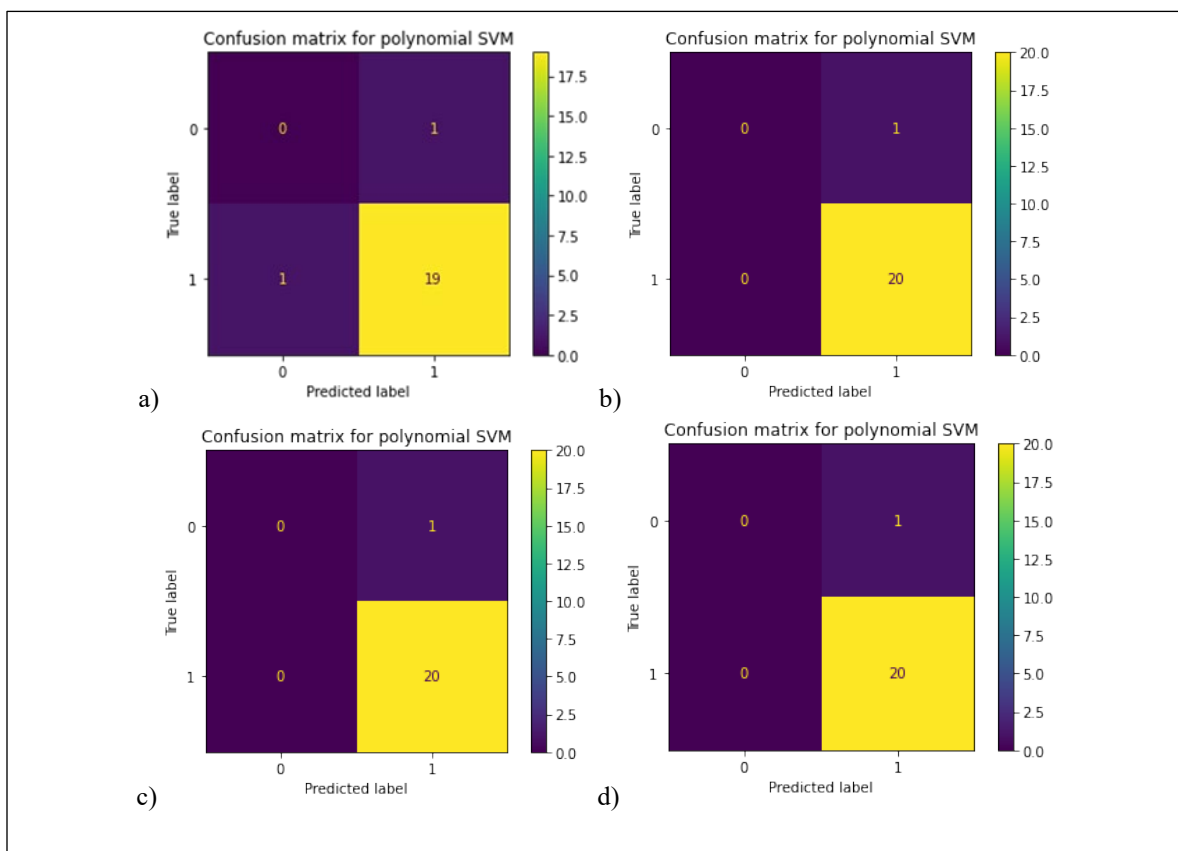
Fig. 7. Confusion matrix plotted for each dataset; a) Original Dataset before tuning, b) Original Dataset after Tuning, c) Second Dataset before Tuning, d) Second Dataset after Tuning

V. CONCLUSION

In this paper, a study on student performance classification by using Support Vector Machine (SVM) with polynomial kernel is presented. Experiments for SVM – POLY were conducted with the original dataset and the dataset without the feature selection process (second dataset). Both datasets were also experimented with using an optimisation algorithm to make comparison of the effect of the algorithm towards the classification performance. During the optimisation process, the value of C and γ were varied; (0.1, 0.2 ...1.0) and (0.005, 0.010...0.060), respectively. The results showed that the optimisation done by the Grid Search towards the parameters C (0.1), γ (0.005) and degree (1) of the SVM-POLY model improved significantly from 90.5% to 95.2% accuracy. Hence, it was concluded SVM-POLY with C value of 0.1, γ value of 0.005 and a degree value of 1 can give the best performance in classifying the student's performance during online learning activities.

Research which involves the use of academic features and student online behaviours, Chi-squared test and SVM-POLY with Grid Search algorithm is recommended to be used in future research. In addition, it is better to have a bigger range containing useful tunings for the optimisation process. The study can also be enhanced by providing data that have balanced failed and passed students so that the classification model can be further improved by classifying both failed and passed students correctly.

REFERENCES

[1] V. Vijayalakshmi and K. Venkatachalapathy, "Comparison of Predicting Student's Performance using Machine Learning Algorithms," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 12, pp. 34–45, 2019, doi: 10.5815/ijisa.2019.12.04.

[2] S. A. Oloruntoba, "Student Academic Performance Prediction Using Support Vector Machine," *Ijesrt Int. J. Eng. Sci. Res. Technol.*, vol. 6, no. 12, pp. 588–598, 2017.

[3] M. Kumar and Y. K. Salal, "Systematic review of predicting student's performance in academics," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 3, pp. 54–61, 2019, doi: 10.13140/RG.2.2.26667.69923.

[4] F. D. Kentli and Y. Sahin, "An SVM approach to predict student performance in manufacturing processes course," *Energy Educ. Sci. Technol. Part B Soc. Educ. Stud.*, vol. 3, no. 4, pp. 535–544, 2011.

[5] J. Khairnar and M. Kinikar, "Machine Learning Algorithms for Opinion Mining and Sentiment Classification," *Int. J. Sci. Res. Publ.*, vol. 3, no. 6, pp. 1–6, 2013.

[6] A. Dinesh Kumar, R. Pandi Selvam, and K. Sathesh Kumar, "Review on prediction algorithms in educational data mining," *Int. J. Pure Appl. Math.*, vol. 118, no. Special Issue 8, pp. 531–537, 2018.

[7] M. Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education," *IEEE Access*, vol. 4, pp. 2379–2387, 2016, doi: 10.1109/ACCESS.2016.2568756.

[8] M. Laxmi, I. Raju, G. Latha Pavani, and K. Vijaya Kumar, "ANALYSIS

OF STUDENT ACADEMIC PERFORMANCE LEADING TECHNOLOGY USING CLASSIFICATION ALGORITHMS," *J. Crit. Rev.*, vol. 7, no. 19, pp. 7251–7259, 2020.

[9] Y. Cui, F. Chen, and A. Shiri, "Scale up predictive models for early detection of at-risk students : a feasibility study," *Inf. Learn. Sci.*, vol. 121, no. 3/4, pp. 97–116, 2020, doi: 10.1108/ILS-05-2019-0041.

[10] J. Superby, J. P. Vandamme, and N. Meskens, "Determination of factors influencing the achievement of the first-year university students using data mining methods," *IEEE Trans. Syst. Man. Cybern.*, no. May, pp. 1–8, 2006.

[11] M. Alami, "Causes of Poor Academic Performance among Omani Students," *Int. J. Soc. Sci. Res.*, vol. 4, no. 1, pp. 126–136, 2016, doi:10.5296/ijssr.v4i1.8948.

[12] A. Abu Saa, M. Al-Emran, and K. Shaalan, "Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques," *Technol. Knowl. Learn.*, vol. 24, pp. 567–598, 2019, doi: 10.1007/s10758-019-09408-7.

[13] D. Kabakchieva, "Student performance prediction by using data mining classification algorithms," *Int. J. Comput. Sci. Manag. Res.*, vol. 1, no. 4, pp. 686–690, 2012.

[14] B. Kapur, N. Ahluwalia, and S. R, "Comparative Study on Marks Prediction using Data Mining and Classification Algorithms," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 3, pp. 632–636, 2017, doi: 10.26483/ijarcs.v8i3.3066.

[15] M. Kumar, A. J. Singh, and D. Handa, "Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques," *Int. J. Educ. Manag. Eng.*, vol. 7, no. 6, pp. 40–49, 2017, doi: 10.5815/ijeme.2017.06.05.

[16] D. Ph, M. Daud, and D. Ph, "Final Grade Prediction of Secondary School Student using Decision Tree," *Int. J. Comput. Appl.*, vol. 115, no. 21, pp. 32–36, 2015.

[17] S. A. Kumar, "Efficiency of Decision Trees in Predicting Student ' S Academic Performance," *Comput. Sci. Inf. Technol. (CS IT)*, pp. 335–343, 2011, doi: 10.5121/csit.2011.1230.

[18] S. Wiyono and T. Abidin, "Comparative Study of Machine Learning Knn, Svm, and Decision Tree Algorithm To Predict Student'S Performance," *Int. J. Res. -GRANTHAALAYAH*, vol. 7, no. 1, pp. 190–196, 2019, doi: 10.29121/granthaalayah.v7.i1.2019.1048.

[19] Y. K. Salal, S. M. Abdullaev, and M. Kumar, "Educational Data Mining : Student Performance Prediction in Academic," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 4c, 2019.

[20] C. Anuradha and T. Velmurugan, "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance," *Indian J. Sci. Technol.*, vol. 8, no. 15, 2015.

[21] H. Almarabeh, "Analysis of Students' Performance by Using Different Data Mining Classifiers," *Int. J. Mod. Educ. Comput. Sci.*, vol. 9, no. 8, pp. 9–15, 2017, doi: 10.5815/ijmecs.2017.08.02.

[22] S. Natek and M. Zwilling, "Student data mining solution-knowledge management system related to higher education institutions," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6400–6407, 2014, doi: 10.1016/j.eswa.2014.04.024.

[23] A. A. Rimi, "Developing Classifier for the Prediction of Students' Performance Using Data Mining Classification Techniques," *AURUM J. Eng. Syst. Archit.*, vol. 4, no. 1, pp. 73–91, 2020.

[24] B. Kumar Verma, D. N. Srivastava, and H. Kumar Singh, "Prediction of Students' Performance in e-Learning Environment using Data Mining/ Machine Learning Techniques," *J. Univ. Shanghai Sci. Technol.*, vol. 23, no. 05, pp. 596–593, 2021, doi: 10.51201/jusst/21/05179.

[25] E. Filiz and E. Öz, "Finding the best algorithms and effective factors in classification of Turkish science student success," *J. Balt. Sci. Educ.*, vol. 18, no. 2, pp. 239–253, 2019, doi: 10.33225/jbse/19.18.239.

[26] A. O. Ameen, M. A. Alarape, and K. S. Adewole, "Students' Academic Performance and Dropout Prediction," *Malaysian J. Comput.*, vol. 4, no. 2, p. 278, 2019, doi: 10.24191/mjoc.v4i2.6701.

[27] V. G. Karthikeyan, P. Thangaraj, and S. Karthik, "Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation," *Soft Comput.*, vol. 24, no. 24, pp. 18477–18487, 2020, doi: 10.1007/s00500-020-05075-4.

[28] M. Asiah, K. Nik Zulkarnaen, D. Safaai, M. Y. Nik Nurul Hafzan, M. Mohd Saberi, and S. Siti Syuhaida, "A Review on Predictive Modeling Technique for Student Academic Performance Monitoring," in *Engineering Applications of Artificial Intelligence Conference (EAAIC 2018)*, 2019, vol. 255, doi: 10.1051/matecconf/201925503004.

[29] N. Alangari and R. Alturki, "Predicting students final GPA using 15 classification algorithms," *Rom. J. Inf. Sci. Technol.*, vol. 23, no. 3, pp. 238–249, 2020.

[30] R. Tabassum and N. Akhter, "Effect of Demographic Factors on Academic Performance of University Students," *J. Res. Reflections Educ.*, vol. 14, no. 1, pp. 64–80, 2020.

[31] S. Viswanathan and S. Vengatesh Kumar, "Study Of Students ' Performance Prediction Models Using Machine Learning," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 2, pp. 3085–3091, 2021.

[32] V. Ramesh, "Predicting Student Performance : A Statistical and Data Mining Approach," *Int. J. Comput. Appl. (0975 – 8887)*, vol. 63, no. 8, pp. 35–39, 2013.

[33] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015, doi: 10.1016/j.procs.2015.12.157.

[34] H. Wu, S. Li, J. Zheng, and J. Guo, "Medical students' motivation and academic performance: the mediating roles of self-efficacy and learning engagement," *Med. Educ. Online*, vol. 25, no. 1, 2020, doi: 10.1080/10872981.2020.1742964.

[35] J. Wang, P. Neskovic, and L. N. Cooper, "Training data selection for support vector machines," *Lect. Notes Comput. Sci.*, vol. 3610, no. PART I, pp. 554–564, 2005, doi: 10.1007/11539087_71.

[36] N. Krishnaveni and V. Radha, "Feature Selection Algorithms for Data Mining Classification : A Survey," *Indian J. Sci. Technol.*, vol. 12, no. February, 2019, doi: 10.17485/ijst/2018/v12i6/139581.

[37] H. Yu, J. Han, and K. C. C. Chang, "PEBL: Positive example based learning for web page classification using SVM," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 239–248, 2002.

[38] M. Kokoç and A. Altun, "Effects of learner interaction with learning dashboards on academic performance in an e-learning environment," *Behav. Inf. Technol.*, vol. 40, no. 2, pp. 161–175, 2021, doi: 10.1080/0144929X.2019.1680731.

[39] S. M. Aslam, A. K. Jilani, J. Sultana, and L. Almutairi, "Feature Evaluation of Emerging E-Learning Systems Using Machine Learning: An Extensive Survey," *IEEE Access*, vol. 9, pp. 69573–69587, 2021, doi: 10.1109/ACCESS.2021.3077663.

[40] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, 2021, doi: 10.3390/educsci11090552.

[41] M. F. Zibran, "Chi-square Test of Independence," *Concise Encycl. Stat.*, pp. 79–82, 2008, doi: 10.1007/978-0-387-32833-1_58.

[42] C. Shen, S. Panda, and J. T. Vogelstein, "The Chi-Square Test of Distance Correlation," *J. Comput. Graph. Stat.*, vol. 31, no. 1, pp. 254–262, 2022, doi: 10.1080/10618600.2021.1938585.

[43] J. Y. X. Chua and S. Shorey, "Effectiveness of end-of-life educational interventions at improving nurses and nursing students' attitude toward death and care of dying patients: A systematic review and meta-analysis," *Nurse Educ. Today*, vol. 101, no. April, p. 104892, 2021, doi: 10.1016/j.nedt.2021.104892.

[44] W. K. Hoy, C. J. Tarter, and A. W. Hoy, "Academic Optimism of Schools:A Force for Student Achievement," *Am. Educ. Res. J.*, vol. 43, no. 3, pp. 425–446, 2006.

[45] W. Hämäläinen and M. Vinni, "Comparison of machine learning methods for intelligent tutoring systems," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4053 LNCS, no. January, pp. 525–534, 2006, doi: 10.1007/11774303_52.

[46] M. Kumar and Y. K. Salal, "Systematic Review of Predicting Student ' s Performance in Academics," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 3, pp. 54–61, 2019, doi: 10.13140/RG.2.2.26667.69923.

[47] N. Naicker, T. Adeliyi, and J. Wing, "Linear Support Vector Machines for Prediction of Student Performance in School-Based Education," *Math. Probl. Eng.*, vol. 2020, p. 7, 2020.

[48] S. . Oloruntuba and J. L. Akinode, "Student academic performance prediction using support vector machine," *IJESRT Int. J. Eng. Sci. Res. Technol.*, no. May, 2019.

[49] M. Ciolacu, A. F. Tehrani, R. Beer, and H. Popp, "Education 4 . 0 – Fostering Student ' s Performance with Machine Learning Methods," *IEEE 23rd Int. Symp. Des. Technol. Electron. Packag.*, pp. 438–443, 2017.

[50] E. Deepak, G. S. Pooja, R. N. S. Jyothi, S. V. P. Kumar, and K. V Kishore, "SVM Kernel based Predictive Analytics on Faculty Performance Evaluation," *Int. Conf. Inven. Comput. Technol.*, pp. 1–4, 2016.

[51] A. Mishra and N. Chaudhary, "Student Performance Measure by Using Different Classification Methods of Data Mining," *Turkish J. Comput.*

Math. Educ., vol. 12, no. 10, pp. 4063–4069, 2021.

[52] A. Rasool, C. Bunterngchit, L. Tiejian, R. Islam, and Q. Qu, "Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis," *Int. J. Environ. Res. Public Health*, vol. 19, p. 3211, 2022.

[53] M. A. Nanda, U. Padjadjaran, K. B. Seminar, D. Nandika, and A. Maddu, "A Comparison Study of Kernel Functions in the Support Vector Machine and A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection," *Information*, vol. 9, no. 1, 2018, doi: 10.3390/info9010005.

[54] Z. Zhou, "Student Pass Rates Prediction Using Optimized Support Vector Machine and Decision Tree," *IIEEE 8th Annu. Comput. Commun. Work. Conf.*, pp. 209–215, 2018.

[55] M. A. Muslim *et al.*, "Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 8–15, 2020.

[56] K. S. Nugroho, A. Y. Sukmadewa, A. Vidianto, and W. F. Mahmudy, "Effective predictive modelling for coronary artery diseases using support vector machine," *IAES Int. J. Artif. Intell.*, vol. 11, no. 1, pp. 345–355, 2022, doi: 10.11591/ijai.v11.i1.pp345-355.

[57] J. Shana and T. Venkatachalam, "Identifying Key Performance Indicators and Predicting the Result from Student Data," *Int. J. Comput. Appl.*, vol. 25, no. 9, pp. 45–48, 2011, doi: 10.5120/3057-4169.

[58] N. Rachburee, "A Comparison of Feature Selection Approach Between Greedy , IG-ratio , Chi-square , and mRMR in Educational Mining," in *International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2015, pp. 420–424.

[59] O. Yamini and P. S. Ramakrishna, "A Study on Advantages of Data Mining Classification Techniques," *Int. J. Eng. Res. Technol.*, vol. 4, no. 9, 2015.

[60] N. Ain *et al.*, "Modeling Student ' s Academic Performance During Covid -19 Based on Classification in Support Vector Machine," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 5, pp. 1798–1804, 2021.

[61] I. S. Al-Mejibli, J. K. Alwan, and D. H. Abd, "The effect of gamma value on support vector machine performance with different kernels," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 5497–5506, 2020, doi: 10.11591/IJECE.V10I5.PP5497-5506.

[62] F. Lotte and M. Congedo, "TOPICAL REVIEW A review of classification algorithms for EEG-based brain – computer interfaces," *J. Neural Eng.*, vol. 1, 2007, doi: 10.1088/1741-2560/4/2/R01.

[63] M. Azim *et al.*, "Observation on Polynomial and Radial Basis function in Support Vector Machine (SVM) Tuned Parameters for Agarwood Oil Quality Grading," *J. Tomogr. Syst. Sensors Appl.*, vol. 4, no. 1, pp. 16–25, 2021.

[64] N. S. Ismail, N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. M. Ali, and S. N. Tajuddin, "Polynomial tuned kernel parameter in SVM of agarwood oil for quality classification," *Proc. - 2018 IEEE Int. Conf. Autom. Control Intell. Syst. I2CACIS 2018*, no. October, pp. 77–82, 2019, doi: 10.1109/I2CACIS.2018.8603686.

[65] G. Ben Brahim, "Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features," *Arab. J. Sci. Eng.*, vol. 47, no. 8, pp. 10225–10243, 2022, doi: 10.1007/s13369-021-06548-w.

[66] A. Elen and E. Avuçlu, "Standardized Variable Distances : A distance-based machine learning method," *Appl. Soft Comput. J.*, no. xxxx, p. 106855, 2020, doi: 10.1016/j.asoc.2020.106855.

[67] I. Mohamad and D. Usman, "Standardization and Its Effects on K-Means Clustering Algorithm," *Res. J. Appl. Sci. Eng. Technol.*, no. February 2016, 2013, doi: 10.19026/rjaset.6.3638.

[68] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and H. Moore, "Relief-Based Feature Selection : Introduction and Review," *J. Biomed. Inform.*, 2018, doi: 10.1016/j.jbi.2018.07.014.

[69] J. M. Bonner and W. G. Holliday, "How college science students engage in note-taking strategies," *J. Res. Sci. Teach.*, vol. 43, no. 8, pp. 786–818, 2006, doi: 10.1002/tea.20115.

[70] P. H. Chen, "In-class and after-class lecture note-taking strategies," *Act. Learn. High. Educ.*, vol. 22, no. 3, pp. 245–260, 2021, doi: 10.1177/1469787419893490.

[71] P. H. Chen, "The Effects of College Students' In-Class and After-Class Lecture Note-Taking on Academic Performance," *Asia-Pacific Educ. Res.*, vol. 22, no. 2, pp. 173–180, 2013, doi: 10.1007/s40299-012-0010-8.

[72] Y. Aslı and A. C. Mehmet, "The Effects of Kernel Functions and Optimal Hyperparameter Selection on Support Vector Machines," *J. New Theory*, vol. 34, no. 34, pp. 64–71, 2021.

[73] J. Canul-reich, L. Margain, and J. C. Ponce, "Mining : Students Comments about Teacher Performance Assessment using Machine Learning Algorithms Mining : Students Comments about Teacher Performance Assessment using Machine Learning Algorithms," *Int. J. Comb. Optim. Probl. Informatics*, vol. 9, no. January 2020, pp. 26–40, 2018.

[74] N. Lavesson and P. Davidsson, "Quantifying the Impact of Learning Algorithm Parameter Tuning," *Proc. 21st Natl. Conf. Artif. Intell.*, vol. 1, no. 1, pp. 395–400, 2006.

[75] N. S. Ismail and N. Ismail, "Polynomial tuned Kernel Parameter in SVM of Agarwood Oil for Quality Classification," *2018 IEEE Int. Conf. Autom. Control Intell. Syst.*, no. October, pp. 77–82, 2018.

[76] C. A. A. Kaestner, "Support Vector Machines and Kernel Functions for Text Processing," *Rev. Informática Teórica e Apl.*, vol. 20, no. 3, pp. 130–154, 2013.

[77] A. N. Khobragade, M. M. Raghuwanshi, and L. Malik, "Evaluating Kernel Effect on Performance of SVM Classification using Satellite Images," *Int. J. Sci. Eng. Res.*, vol. 7, no. 3, pp. 742–748, 2016.

[78] S. Chettri and G. Science, "Support Vector Machine Classi ers as Applied to AVIRIS Data c Ecosystem Science and Technology Branch," *Summ. Eighth JPL Airborne Earth Sci. Work.*, no. March, pp. 1–10, 1999.

[79] S. W. Lin, Z. J. Lee, S. C. Chen, and T. Y. Tseng, "Parameter determination of support vector machine and feature selection using simulated annealing approach," *Appl. Soft Comput. J.*, vol. 8, no. 4, pp. 1505–1512, 2008, doi: 10.1016/j.asoc.2007.10.012.