A Framework for Predicting Academic Success using Classification Method through Filter-Based Feature Selection

Dafid[1], Ermatita[2]*, Samsuryadi[3]
Department of Information System, Universitas Multi Data Palembang, Palembang, Indonesia[1]
Doctoral Program in Engineering Science, Universitas Sriwijaya, Palembang, Indonesia [1, 2, 3]

***Abstract*—Students' academic success is still a serious problem faced by higher education institutions worldwide. A strategy is needed to increase the students' academic performance and prevent students from failing. The need to get early accurate information about poor academic performance is a must and could achieved by constructing a prediction model. Therefore, an effective technique is required to provide the accurate information and improve the accuracy of the prediction model. This study evaluates the filter-based feature selection especially the filter-based feature ranking techniques for predicting academic success. It provides a comparative study of filter-based feature selection techniques for determining the type of features (redundant, irrelevant, relevant) that affect the accuracy of the prediction models. Furthermore, this study proposes a novel feature selection technique based on attribute dependency for improving the performance of the prediction model through a framework. The experimental results show that the proposed technique significantly improved the accuracy of the prediction models from 2-8%, outperforming the existing techniques, and the Decision Tree classifier performs best for predicting with an accuracy score of 92.64%.**

Keywords*—*Academic success; framework; filter-based feature selection; classifier; accuracy

I. INTRODUCTION

Nowadays, students' academic success is still a severe topic for researchers due to its impact on higher education quality. One of the strategies to keep and enhance it is getting early information about the poor students' academic performance through constructing a prediction model. The ability to predict students' academic success accurately will create opportunities to improve educational outcomes, with the result that universities can create academic policies more accurately and effectively [1].

Parallel to this, previous studies have been reported and provided varying results in creating a prediction model [1]–[3]. These reports conclude that there are two main factors in predicting academic success, which are features and prediction methods. The most common prediction method used is the classification [2]. The classification techniques frequently used by researchers to predict student academic success are Decision Tree (44%), Naive Bayes (19%), Artificial Neural Network (10%), K-Nearest Neighbor (5%), and Support Vector Machine (3%) [2]. As shown in this study [2], each technique has given the best result in educational data. However they still suffer from accuracy because the major issue is not all of the features used in building the model hold predictive information (irrelevant and redundant features). To alleviate this limitation, researchers have applied feature selection techniques to remove the unpredicted information features.

Feature selection plays a vital role in increasing the accuracy of the prediction model. It will remove the redundant and irrelevant features and select the relevant features using a specific method [4]. Inclusion of redundant and irrelevant features will reduce the prediction model's accuracy [5]. Each feature selection technique, namely filter, wrapper and embedded when compared to each other; filter technique is the fastest in terms of time complexity but low in performance [6]–[12]. Most researchers use filter-based feature selection on the performance of the academic success classification model due to its cheap computational cost since the academic dataset is large (more than 20 features) [13].

Further, filter-based feature selection can be divided into two types: filter-based feature ranking techniques and filter-based feature subset techniques [5], [13]–[16]. Filter-based feature ranking techniques select the features by choosing only top-ranked features using a ranker search method with important scores such as ChiSquared (Statistics-Based Scores), InfoGain and GainRatio (Probability-Based Scores), Symmetrical Uncertainty (Probability-Based Scores) and Relief (Instance-Based Techniques). In contrast, filter-based feature subset techniques can evaluate the feature subset as a whole instead of evaluating individual features using a feature-subset evaluator. Several previous research have compared and proved that filter-based feature selection gives an impact on increasing the performance of academic success classification [17]–[22]. However, the problem arises when using filter-based feature ranking techniques are computationally cheap but do not deal with redundant features [16]. These techniques tend to select redundant and irrelevant features as they do not consider feature interactions [5], [13]–[15], [23]. In contrast, filter-based feature subset techniques are better for deselecting redundant features but are computationally expensive due to repeating steps in a greedy forward fashion to find the best subset based on certain criteria [11], [16] until the desired subset is reached. Even though in these techniques there are efforts to remove redundant features by finding the relevancy between the individual features, but in the fact the techniques only discard a few redundant variables since the measures evaluate features

*Corresponding Author.

individually [13], [23]. A brief overview of the filter-based feature selection can be seen in Table I.

Due to these above-said problems, the main contribution of this study is a framework that can act as a guide to build the prediction model by using the classification method through filter-based feature selection that fits into the problem to enhance accuracy. This research focuses on improving filter-based feature ranking techniques by addressing redundant and irrelevant problems with attribute dependency techniques and unique attribute technique. These techniques have ability to find the relationship among features not only individually but also in groups that are related to redundancy and irrelevancy problems [24]–[26]. The evaluation measurement of accuracy is then used to test each prediction model that has been constructed.

Through the proposed framework, higher education can accurately identify failed-risk students who need special attention to prevent them from failing and take appropriate action at right time. This finding can be employed to help poor academic performance students in their next semester examination to get better results. Consequently, it will have an academic impact on higher education to enhance students' academic success.

TABLE I. OVERVIEW OF THE STUDIED FILTER-BASED SELECTION TECHNIQUES ON EDUCATIONAL DATA

Family	Method	References	Merit	Demerit
Filter-based feature ranking techniques	ChiSquared (CHI)	[20]–[22]	Redundant features are selected	Low computational cost
	InfoGain (IG)	[17], [18], [20]–[22]		
	GainRatio (GR)	[17], [18], [20]–[22]		
	Symmetrical Uncertainty (SU)	[17], [18], [22]		
	Relief (RF)	[17], [18], [20]		
Filter-based feature subset techniques	Correlation (CFS)	[17], [18], [20], [21]	Reduce redundant features	High computational cost

The following is how this section of the paper is structured: Section II reviews recent similar works on predicting academic success and limitations of them and presents feature selection methods used in this study. In Section III, the research methodology was described. The research results are given in Section IV, along with the study implications. Finally, Section V presents the conclusion in its entirety.

II. RELATED WORK

A. *Previous Research*

This section discusses the previous works related to the academic success field that have used filter-based feature selection techniques to enhance the performance of the prediction models on education data.

A research by Khasanah [17] discovered that filter-based feature selection may be used to carefully choose high influence features to predict student performance. The filter-based feature selections used are correlation-based attribute evaluation, gain-ratio attribute evaluation, information-gain attribute evaluation, relief attribute evaluation, and symmetrical uncertainty attribute evaluation. Student attendance and first-semester GPA were shown to be the most important factors. They employed Bayesian Network and Decision Tree algorithms to classify and predict student performance, where the Bayesian Network outperformed the Decision Tree classification in the accuracy case. Hussain [18] conducted a study to find high influence attributes using correlation-based attribute evaluation, gain-ratio attribute evaluation, information-gain attribute evaluation, relief attribute evaluation, and symmetrical uncertainty attribute evaluation. The techniques showed that internal assessment attributes as highly influential attributes, where random forest outperformed the other classifiers in the accuracy case. A study was carried out by Priyasadie (19) to forecast junior high school students' grades using feature selection to enhance classification performance. The outcome demonstrated that Decision Tree outperforms other classification algorithms in general when parameter optimization and feature selection are used and First Semester Natural Science and First Semester Social Science are the most important attribute. Zaffar [20] conducted a study of feature selection techniques to improve the accuracy of academic achievement prediction. The correlation-based feature evaluator (CFS), the chi-squared test, the filtered, gain ratio (GR), the principal component analysis (PC), and the Relief method are the feature selection approaches used in the study. The study used fifteen prediction models and compared them to each other in order to verify the effectiveness of the proposed feature selection techniques. The experiment shows that using feature selection increases accuracy. Sokkhey [22] developed the CHIMI feature selection technique, a combination of the ChiSquare and Mutual Information ranking algorithms to identify the most relevant features affecting classification performance. The results presented that the technique improved the prediction model accuracy.

According to the review previous research above, each filter-based techniques have its advantage, but in general the disadvantages of them are inability or not optimal to remove the redundant and irrelevant features since they choose only top-ranked features without considering feature interactions [5], [8], [15], [23], [27]–[29].

B. *Current Research*

This section presents the current work to overcome the limitations in the previous research by proposing a framework through a novel filter-based feature ranking techniques utilizing attribute dependency theory. It presents the central concepts and definitions related to feature selection.

This study presents an analysis of filter-based feature selection techniques on a number of classifiers and then determines each classifier's performance in terms of accuracy. It focuses on the existing filter-based feature ranking techniques compared to this proposed technique to justify its better performance in terms of removal of redundant and irrelevant features and improved classification accuracy, while filter-based feature subset techniques are still used to complete the result of this research.

As known, the existing technique failed to remove the redundant and irrelevant attributes because they are guided by a statistical evaluation metric that ranks the attributes in decreasing order of importance without considering the relationship between the features. Thus, highly relevant features with the class, but redundant with others, will be on the top of the ranking. Meanwhile, the attribute dependency theory has the ability to determine relationship among two or more (group) features based on dependency measures [24]–[26]. Thus, there is a chance and suitable for this theory to address the limitation of removing redundant and irrelevant attributes in the previous technique. For an effective process, the technique requires domain expert/domain knowledge that will be helpful to identify the suspect redundant features for next determining them as redundant features or not using dependency measure.

In general, features in a dataset are classified into redundant, irrelevant and relevant features (strong and or weak features) [30]. Redundant features are those that contain predictive information but it has already been conveyed through any of another features or subset of features. Irrelevant features are those that do not convey any predictive information whereas the relevant features are those that hold predictive information in building learning models. Redundancy among the features or group of features is identified based on the relationship among the features, whereas the relevance is identified by finding the relationship between the feature (or group of features) and the class.

Thus, this study considers attribute dependency to remove redundant features. Using this technique, redundant features are identified and removed by ensuring the strong dependency among the features or group of features. In contrast, the relevant features are identified and selected between the features or group of features and the class. Irrelevant attributes are identified and removed by ensuring the uniqueness values or single value (or dominant only one value).

The proposed feature selection technique is rooted on the concepts of relevance, redundancy and features interaction analysis. In this regard, preliminary concepts related to attribute dependency theory are presented here. Further, the rationale for incorporating these concepts in the proposed framework is explained. Aside from how attributes are evaluated, another important aspect is how the feature selection technique relates to the classification method that will be used to further build the model containing patterns from the selected features. In this sense, five main classifiers may be mentioned. The proposed framework of this research is shown in Fig. 1.

C. Unique Attribute to Process the Feature Selection as Proposed Method

An attribute that has a different value for each row or record in dataset [24]–[26]. In addition, the attributes (features) that only have one value or one value dominates other values, they also are considered as unique attribute.

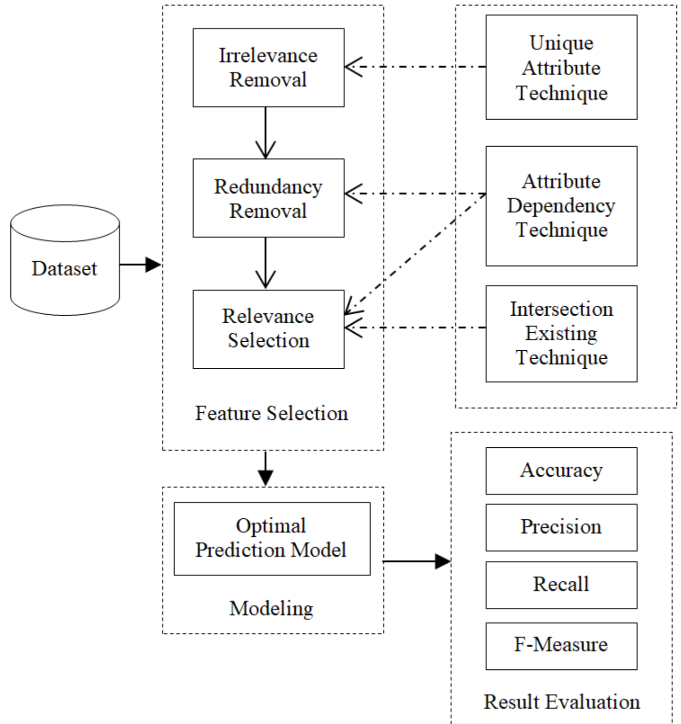

Fig. 1. The proposed framework.

D. Attribute Dependency to Process the Feature Selection as Proposed Method

Not like the existing technique that selecting features based on statistical measures, attribute dependency selects features based on whether an instance which have the feature value, have the same feature value to others or to the class value. Attribute dependency in the relational model explains the relationship between attributes, or more specifically, the value of an attribute that determines the value of other attributes [24]–[26]. There are several types of dependencies, namely:

1) Functional Dependency (FD): Indicates that if A and B are attributes of relation R, B is functionally dependent on A (denoted A → B), if each value of A is associated with exactly one value of B.

2) Full Functional Dependency (FFD): Indicates that if A and B are attributes of a relation, B is fully functionally dependent on A if B is functionally dependent on A, but not on any proper subset of A.

3) Transitive Dependency (TRD): A condition where A, B, and C are attributes of a relation such that if A → B and B → C, then C is transitively dependent on A via B (provided that A is not functionally dependent on B or C).

4) Total Dependency (TD): A condition where A and B are attributes of a relation such that if B is functionally dependent on A (A → B) and A is functionally dependent on B (B → A).

E. Proposed Technique

In the proposed work, there are three phases to feature selection. The first phase conducts a preliminary search to remove irrelevant features. In this phase, unique attribute analysis is employed by identifying features that have a unique value or not, for each row in the dataset. Under this context, features with a unique value means irrelevant and should be removed from the dataset. Another condition is if the features only have one value or if there are more than one value then one value dominates other values, it also means the irrelevant feature, as shown in Algorithm 1 where dataset D as the parameter for configuring the algorithm.

Algorithm 1: Irrelevance Removal

Input:
{Dataset} $D=\{F_i,C\}$; $i=1,2,\ldots,n \mid F_i = \{f_{i1},f_{i2}, .. f_{im}\}$ }
{ *where F_i : features C: class*
n:total number of features
m:total number of instances
}
Result: D_1 *{the dataset minus irrelevant features}*
Process:
i=1;
$D_1 = D$
while i<=*n* do
Get $F_i = \{f_{i1},f_{i2}, .. f_{im}\}$
if ($f_{i1} \neq f_{i2}\ldots\neq f_{im}$) or ($f_{i1} = f_{i2}\ldots= f_{im}$) or
(one value dominates other values)
then
$D_1 = D_1 - F_i$;
end if
i++;
end
return D_1

The second phase conducts a search to remove redundant features. Attribute dependency concepts especially total dependency (TD), functional dependency (FD) and full functional dependency (FFD) are employed based on dependency measure in this phase. The first step is applying FFD, the second is applying TD, and the last is applying FD. The advantage of FFD is that it can find not only a relationship between two features but also a group of features (hidden relationship among features).

The hidden relationship among features is the problem that makes the features as redundant features and it cannot be identified by the previous techniques. There is an effort by [31] to address this problem, but only between two features, not for a group of features, and it needs high computational cost. Using FD and FFD, redundant features are identified by measuring the dependency among features or a group of features. The suspected redundant features are easier to find with the help of domain knowledge or domain experts of the dataset. It makes the searching process faster. The more domain knowledge you know the more faster the suspect redundant features you find. By using the FD or FFD evaluation measure, the features or the group of features that are functionally or fully functionally dependent on another feature means that the features or the group of features are redundant and should be eliminated from the dataset. The process of removing the features can be shown in Algorithm 2 below where dataset D_1 as the parameter for configuring the algorithm:

Algorithm 2: Redundancy Removal

Input :
{Dataset} $D_1=\{F_i,C\}$; $i=1,2,\ldots,n \mid F_i = \{f_{i1},f_{i2}, .. f_{im}\}$ }
{ *where F_i : features C: class*
n:total number of features
m:total number of instances
}
Result: D_2 *{the dataset minus redundant features}*
Process:
$D_2 = D_1$
Repeat
{*Get the suspected redundant features*}
{ Step 1: utilizing FFD}
Get $FG = \{F_2, F_3, \ldots, F_n\}$ and F_1 *{for example}*
if $FG \rightarrow F_1$ then *{F_1 fully FD on FG}*
$D_2 = D_2 - FG$
end if
{ Step 2: utilizing TD}
Get F_1 and F_2 *{for example}*
if $F_1 \rightarrow F_2$ and $F_2 \rightarrow F_1$ then *{FD each others}*
$D_2 = D_2 - F_1$ or $D_2 = D_2 - F_2$
end if
{ Step 3: utilizing FD}
Get F_1 and F_2 *{for example}*
if $F_1 \rightarrow F_2$ then *{F_2 FD on F_1}*
$D_2 = D_2 - F_1$
end if
Until no more suspected redundant features
return D_2

The third phase conducts a search to select relevant features. The attribute dependency that is used in the second phase is also used in this phase. The difference is only the target that the dependency analysis will be applied between features to the class.

Attribute dependency only yields two values, namely functionally dependent or not, to indicate the dependency between the two or a group of observed features. Related to these, in terms of relevant features, it means there are only two categories, namely relevant features (absolutely strong features) or not relevant features.

Attribute dependency is effective only for absolutely strong features but not for ‘not relevant features’ (strongly and weakly relevant features). For this reason, in addition to attribute dependency, this study uses a combination of the existing techniques (CHI, IG, GI, Relief, SU) to get the relevant features by applying intersection to the features produced by each technique. The process of selecting the features can be shown in Algorithm 3 below where dataset D_2 as the parameter for configuring the algorithm:

Algorithm 3: Relevance Selection

Input :
{Dataset} D_2={F_i,C); i=1,2,…,n | $F_i = \{f_{i1}, f_{i2}, .. f_{im}\}$ }
{ *where* F_i *: features C: class*
n:total number of features
m:total number of instances
}
Result: D_3 *{the selected features}*
Process:
D_3 = {}
{ Step 1: utilizing FD}
i = 1
while i<=*n* do
  Get $F_i = \{f_{i1}, f_{i2}, .. f_{im}\}$
  if $F_i \rightarrow C$ then *{ C FD on* F_i *}*
    $D_3 = D_3 + F_i$
  end if
  i++
end
{ Step 2: utilizing the combination of the existing technique
$R, R_1 \ldots R_5 = \{ F_1, F_2, \ldots, F_n\}$
}
$D_2 = D_2 – D_3$
Get F_1 to F_n
R_1 = CHI (F,C) *{ChiSquared}*
R_2 = IG (F,C) *{Information Gain}*
R_3 = GR (F,C) *{Gain Ratio}*
R_4 = RF (F,C) *{Relief}*
R_5 = SU (F,C) *{Symmetrical Uncertainty}*
$R = R_1 \cap R_2 \cap R_3 \cap R_4 \cap R_5$ *{Intersection of ALL}*

$D_3 = D_3 + R$
return D_3

III. Methodology

This research methodology consists of four stages. They are data preparation, data processing, modeling and performance analysis.

A. Data Preparation

Related to the goal, this study used the high dimensional public dataset from higher education that consist of irrelevant, redundant and relevant features. The dataset was from the students of Middle East College (MEC), Muscat, Oman, studying in a computing specialization from the sixth semester and above. The dataset was collected from a variety of learning approaches based on Information and Communications Technology (ICT), namely, the Student Information System (SIS), the Learning Management System (LMS) called Moodle, and video interactions from the mobile application called “eDify”. The dataset comprises five modules of data from Spring 2017 to Spring 2021 that consists of 326 student records with 37 features and 1 class in total, including the students’ academic information from SIS (which has 23 features), the students’ activities performed on Moodle within and outside the campus (comprising 10 features), the students’ video interactions collected from eDify (consisting of 4 features) and the outcome of the student either having passed or failed the module (1 class). The dataset was already clean, so it did not need to clean up the data from the possibility of duplicate data, missing data or other disturbances interfering with the process of classification. It means the data are ready for processing in the next step. An explanation of the data is described in Table II.

TABLE II. Dataset Description

Attribute	Type / Role	Description
ModuleCode	nominal / feature	Code of the module in which the student has been registered, such as “Module 1”
ModuleTitle	nominal / feature	Title of the module in which the student has been registered, such as “Course 2”
SessionName	nominal / feature	Shows the session in which the student has been registered, such as “Session-A”
RollNumber	nominal / feature	Identification number of the student, such as “21S1234”
ApplicantName	nominal / feature	Name of the student
ApplicantMobile	discrete / feature	Mobile number of the student
CGPAPoint	discrete / feature	Cumulative grade point average of the student, such as “4.0”
CGPADesc	nominal / feature	Category of cumulative grade point average of the student, such as “Very Good”
Advisor	nominal / feature	Name of the advisor
AttemptCount	discrete / feature	The number of attempts in the module, such as “1”
AttemptCountCat	nominal / feature	Category of the number of attempts in the module, such as “Low”
RemoteStudent	nominal / feature	Either the student is under remote study mode or not, such as “Yes/No”
Probation	nominal / feature	Either the student has a backlog of modules to clear, such as “Yes/No”
HighRisk	nominal / feature	The high failure rate in a module, such as “Yes/No”
TermExceeded	nominal / feature	Progression rate of the student in the degree plan, such as “Yes/No”
AtRisk	nominal / feature	Previously failed two or more modules, such as “Yes/No”
AtRiskSSC	nominal / feature	Whether the student been registered by the student success center for any educational deficiencies, such as “Yes/No”
SpecialNeed	nominal / feature	Whether the student been registered by the student success center for any special needs, such as “Yes/No”
OtherModules	discrete / feature	A student registered in any other modules in the current semester, such as “1”
OtherModulesCat	nominal / feature	Category of a student registered in any other modules in the current semester, such as “High”
PrerequisiteModules	nominal / feature	Prerequisite module registration, such as “Yes/No”

Attribute	Type	Role	Description
PlagiarismHistory	discrete	feature	The number of onto which modules the student has been booked for academic integrity violation, such as "1"
PlagiarismHistoryCat	nominal	feature	Category of the number of onto which modules the student has been booked for academic integrity violation, such as "Low"
CW1	discrete	feature	Marks obtained by the student in their first coursework, such as "86.5"
CW1Cat	nominal	feature	Category of marks obtained by the student in their first coursework, such as "Adequate"
CW2	discrete	feature	Marks obtained by the student in their second coursework, such as "86.5"
CW2Cat	nominal	feature	Category of marks obtained by the student in their second coursework, such as "Excellent"
ESE	discrete	feature	Marks obtained in the end semester examination, such as "86.5"
ESECat	nominal	feature	Category of marks obtained in the end semester examination, such as "Good"
Online C	discrete	feature	User-performed activities within campus (in minutes), such as "25"
Online Ccat	nominal	feature	Category of user-performed activities within campus, such as "Poor"
Online O	discrete	feature	User-performed activities outside of campus (in minutes), such as "25"
Online Ocat	nominal	feature	Category of user-performed activities outside of campus, such as "Poor"
Played	discrete	feature	The number of times the video has been played
Paused	discrete	feature	The number of times the video has been paused
Likes	discrete	feature	The number of times the student has liked the video
Segment	discrete	feature	The number of times a student has played a specific portion of the video by using the slider
Result	nominal	class	the outcome of the student either having passed or failed the module

B. Data Processing

This phase begins with the identification of the attributes as features and as the class used to build the framework. Based on dataset in the data preparation phase, this study has determined the attribute with label 'result' as the class and the rest of attributes as the features as shown in Table II. This process focuses on selecting relevant or strongly features toward the target attribute (class) and removing redundant and irrelevant features or weakly relevant features from the dataset. The prediction model's performance is negatively impacted by redundant and irrelevant features.

Feature selection is one of techniques that greatly impact on enhancing model performance mainly increasing accuracy. In this research, feature selection is conducted in three stages: (a) irrelevance removal, (b) redundancy removal, and (c) relevance selection.

In the irrelevance removal stage, unique attribute technique is used to test whether the features have (a) a unique value or not, (b) only have one value and (c) one value dominates other values. Features that meet the condition above are removed from the dataset. From this stage, it will get dataset D_1. The details of the processed are shown in Algorithm 1.

After the irrelevance removal stage, next, it continued to the redundancy removal stage. The total dependency (TD), functional dependency (FD) and full functional dependency (FFD) are used. They are used to test the relationship between two features or features and groups of features. Features or group of features that are functionally (FD) or fully functionally dependent (FFD) on another feature are removed from the dataset. From this stage, it will get dataset D_2. The details of the processed are shown in Algorithm 2.

After the two stages above, it is followed by relevance selection stage as the last stage. Unlike the two stages before, this stage is employed to select the features used for the construct of the prediction model. There are two concepts used in this stage: (a) attribute dependency, (b) combination of the existing techniques (CHI, IG, GI, Relief, and SU) by applying intersection. Unlike the two stages before, in this stage, the relationship is tested between the features and the class (target attribute). From this stage, it will get the final dataset D_3 as selected features. The details of the processed are shown in Algorithm 3.

C. Modeling

After obtaining the selected features (D_3), the next step is constructing a prediction model. The dataset D_3 is divided into 80% training and 20% testing dataset. Training and testing dataset are used as modeling input. A few of classifiers will be used to process the dataset to classify the student's academic success.

The classifier are Decision Tree (DT), Naive Bayes (NB), Artificial Neural Network (ANN), K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) as the most classifier currently used by researchers to predict student academic success [2]. The process started with training the classifier using the training dataset. The testing dataset tests the classifier in classifying student's academic success. Additionally, depending on measures of classification accuracy and effectiveness, the performance of the classifier model is examined and assessed.

D. Performance Analysis

This section described how to measure the performance of the prediction model from a number of views since each classifier employed in the modeling process vary. The primary goal of the proposed framework is to achieve the highest accuracy with less number of features. Related to the goal, the confusion matrix is used to determine the accuracy of the prediction model. Comparing the actual value to the predicted value is the approach to assess the model's performance. The confusion matrix is a combination of four different predicted and actual values. In the confusion matrix, the classification process' outcomes are denoted by four terms: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). According to the confusion matrix, the formula for calculating the accuracy are given in (1)(2)(3)(4) [31].

$$\text{Accuracy} = (TP + TN) / (TP + TN + TP + TN) \quad (1)$$

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

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

$$\text{F-Measure} = (2*\text{Recall}*\text{Precision})/(\text{Recall}+\text{Precision}) \quad (4)$$

IV. Result and Discussion

The dataset was prepared previously, next used in data processing stage using the proposed feature selection technique.

A. Result

Initially, the dataset was processed by applying the Algorithm 1 to remove the irrelevant features and the result is shown in Table III.

TABLE III. IRRELEVANCE REMOVAL

No	Feature	Unique Attribute Measure
1	RollNumber	unique value
2	ApplicantMobile	unique value
3	ApplicantName	unique value
4	RemoteStudent	one value dominated other value (value No : ≥ 99%, value Yes : ≤ 1%)
5	Probation	one value dominated other value (value No : ≥ 96%, value Yes : ≤ 4%)
6	HighRisk	one value dominated other value (value No : ≥ 93%, value Yes : ≤ 7%)
7	TermExceeded	one value dominated other value (value No : ≥ 98%, value Yes : ≤ 2%)
8	AtRiskSSC	one value dominated other value (value No : ≥ 95%, value Yes : ≤ 5%)
9	SpecialNeed	only one value
10	PlagiarismHistory	one value dominated other value (value 0: ≥ 82%, value 1: ≤ 17% value 2: ≤ 1%)
11	PlagiarismHistoryCat	one value dominated other value (value Low : ≥ 99%, value Medium : ≤ 1%)
12	PrerequisiteModules	one value dominated other value (value No : ≤ 8%, value Yes : ≥ 92%)

D_1 = D – Irrelevant Features

D_1 = ModuleCode, ModuleTitle, SessionName, CGPAPoint, CGPADesc, Advisor, AttemptCount, AttemptCountCat, AtRisk, OtherModules, OtherModulesCat, CW1, CW1Cat, CW2, CW2Cat, ESE, ESECat, Online C, Online CCat, Online O, Online OCat, Played, Paused, Likes, Segment

The results in Table III show that there are 12 features that must be removed from the dataset consisting of 1 feature identified as one value (only one value for all instances), 3 features identified as a unique value (different value for each instances), and 8 features identified as one value dominates other value. After removing the irrelevant features, now the rest of the features are: 25 features as a dataset D_1. The dataset D_1 still have the possibility of redundant features, as a solution, the next step, redundancy removal was carried out. The dataset D_1 was produced by previous stage, then processed by applying the Algorithm 2 to remove the redundant features and the result is shown in Table IV.

TABLE IV. REDUNDANCY REMOVAL

No	Feature	Attribute Dependency Measure	Type
1	ModuleTitle	ModuleCode ←→ ModuleTitle	TD
2	CGPAPoint	CGPAPoint → CGPADesc	FD
3	AttemptCount	AttemptCount → AttemptCountCat	FD
4	OtherModules	OtherModules → OtherModulesCat	FD
5	CW1	CW1 → CW1Cat	FD
6	CW2	CW2 → CW2Cat	FD
7	ESE	ESE → ESECat	FD
8	Online C	Online C → Online Ccat	FD
9	Online O	Online O → Online Ocat	FD

D2 = D1 – Redundant Features

D1 = ModuleCode, SessionName, CGPADesc, Advisor, AttemptCountCat, AtRisk, OtherModulesCat, CW1Cat, CW2Cat, ESECat, Online CCat, Online OCat, Played, Paused, Likes, Segment

The results (Table IV) show that there are 9 features that must be removed from the dataset consisting of 1 feature identified as Total Dependency (TD), 8 features identified as a Functional Dependency (FD). No one features identified as Full Functional Dependency (FFD). After removing the redundant features, the rest of the features now there are 16 features as a dataset D_2. The dataset D_2 need further processed to select the relevant or strongly relevant features. It also still have the possibility of weakly or irrelevant features. The last step, relevance selection was carried out by applying the Algorithm 3 based on dataset D_2. Using the first step in Algorithm 3, no one features functionally dependent on the target attribute (class). It means no feature is selected from this step. Next step, the process continued using the intersection of the combination of the existing technique, and the result is shown in Table V. According to the result as shown in Table V, each technique (CHI, IG, GI, SU) have exactly the same results where the features Paused and Likes are in the lowest rank with the value 0. These two features are irrelevant to the class, so they are removed from $D_{2.}$

While Relief technique have slightly different result where the lowest rank is the feature Paused. Because the value is greater than zero, the feature is still used for processing. After getting the results (R1, R2, R3, R4, R5) of each technique, by applying intersection to the results, the proposed technique got selected features R (Table V) as the final result. The result R consists of 14 features after removing the two lowest ranks. The result R then adds to the result D_3 from the first step. Because in the first step there are no features selected, then D_3 is equal to R that is used for building the model in the step.

TABLE V. RELEVANCE SELECTION

No	Feature Selection Technique	Selected Features	Name of the Result
1	ChiSquared	Advisor, ESECat, Played, ModuleCode, CW1Cat, Segment, CW2Cat, Online Ccat, OtherModulesCat, CGPADesc, Online Ocat, AttemptCountCat, AtRisk, SessionName	R1
2	InfoGain	Advisor, ESECat, ModuleCode, Played, CW1Cat, CW2Cat, Segment, Online Ccat, OtherModulesCat, CGPADesc, Online Ocat, AttemptCountCat, AtRisk, SessionName	R2
3	GainRatio	ESECat, Played, Segment, ModuleCode, Advisor, CW1Cat, CW2Cat, AttemptCountCat, OtherModulesCat, Online Ccat, CGPADesc, AtRisk, SessionName, Online Ocat	R3
4	Symmetrical Uncertainty	ESECat, Played, Segment, Advisor, ModuleCode, CW1Cat, CW2Cat, AttemptCountCat, OtherModulesCat, Online Ccat, CGPADesc, Online Ocat, AtRisk, SessionName	R4
5	Relief	ESECat, ModuleCode, CW1Cat, CW2Cat, Played, Segment, SessionName, Likes, CGPADesc, Online Ccat, Online Ocat, OtherModulesCat, AttemptCountCat, Advisor, AtRisk, Paused	R5
6	Intersection of All	ModuleCode, SessionName, CGPADesc, Advisor, AttemptCountCat, AtRisk, OtherModulesCat, CW1Cat, CW2Cat, ESECat, Online Ccat, Online Ocat, Played, Segment	R= $R_1 \cap R_2 \cap R_3 \cap R_4 \cap R_5$ D_3 = R

Evaluation of the prediction model is used to select the best model raised the highest accuracy. Using a dataset split into training and testing data, the value of each model's cross-validation accuracy as an indicator of model performance is evaluated. Table VI shows the evaluation results based on the cross-validation accuracy and F-measure indicators.

TABLE VI. THE PREDICTION MODEL PERFORMANCE FOR CLASSIFICATION OF STUDENT'S ACADEMIC SUCCESS (%)

Classifier	Accuracy	Precision	Recall	F-Measure
DT	92.64	92.6	92.6	92.2
NB	83.13	83.9	83.1	83.5
ANN	84.62	84.1	85.1	82.1
KNN	85.44	85.9	85.4	85.6
SVM	89.57	89.1	89.6	89.1

The result in Table VI shows the proposed framework using the proposed filter selection technique can achieve high accuracy scores for each classifier. The highest accuracy score is 92.64% on the DT classifier, followed by SVM (89.57%), ANN (84.62%), KNN (83.44%) and the last is NB (83.13%).

When compared with the existing technique that is frequently used in predicting academic success, the result is shown in Table VIII. The selected features (dataset) produced by each technique that will be used by classifier is presented in Table VII. The features with the score ≤ 0 eliminated from dataset because of the features do not have a relationship to target attribute (class).

TABLE VII. THE NUMBER OF SELECTED FEATURES OF FEATURE SELECTION TECHNIQUE

Feature Selection Technique	Features Category	Sub Total Features	Total Selected Features
CHI, IG, GI, SU	Redundant Features : CW2, CW1, ESE, ModuleTitle, AttemptCountCat, OtherModulesCat, Online Ccat, AtRiskSSC, CGPADesc, Online Ocat, AtRisk, SessionName, RemoteStudent	13	28
	Irrelevant Features : RollNumber, ApplicantMobile, ApplicantName,Probation, HighRisk, TermExceeded, PlagiarismHistoryCat, PrerequisiteModules	8	
	Relevant Features : ESECat, Played, Segment, Advisor, ModuleCode, CW1Cat, CW2Cat	7	
RF	Redundant Features: ModuleTitle, OtherModules, Online O, Online C, CW1, CW2, AttemptCount, CGPAPoint	8	31
	Irrelevant Features: ApplicantMobile, RollNumber, PrerequisiteModules, HighRisk, PlagiarismHistory	5	
	Relevant Features: ESECat, ModuleCode, CW1Cat, CW2Cat, ESE, Played, SessionName, Segment, Likes, Online Ocat, AttemptCountCat, OtherModulesCat, Advisor, AtRisk, CGPADesc, Online Ccat, TermExceeded, Paused	18	
CFS	Redundant Features : CW1, CW2	2	3
	Irrelevant Features: ApplicantName	1	
	Relevant Features: -	0	
Proposed Technique	Redundant Features : -	0	14

Feature Selection Technique	Features Category	Sub Total Features	Total Selected Features
	Irrelevant Features: -	0	
	Relevant Features: ModuleCode, SessionName, CGPADesc, Advisor, AttemptCountCat, AtRisk, OtherModulesCat, CW1Cat1, CW1Cat2, ESECat, Online Ccat, Online Ocat, Played, Segment	14	

The result in Table VII shows the existing techniques successfully reduced the number of features that have not significant affecting to the target attribute. However, the selected features produced by each existing technique still have irrelevant and redundant features, impacting the accuracy of the prediction model is not optimal. The techniques (CHI, IG, GI, SU) have the same result in a number of selected features (28 features) and the features that have been selected. The difference is only the ranking of the selected features. The CFS technique slightly has a different result in producing selected features (only three features). It is an excellent factor in reducing the computational cost. According to these results, it concludes that selected features that still contain redundant and irrelevant features, making the accuracy result is not optimal.

TABLE VIII. THE ACCURACY COMPARISON OF FEATURE SELECTION TECHNIQUE FOR PREDICTING ACADEMIC SUCCESS (%)

Feature Selection Technique	DT	NB	ANN	KNN	SVM
All (No FS Technique used all dataset)	85.88	71.47	81.54	82.21	86.50
CHI	85.89	73.31	82.82	84.96	86.81
IG	85.89	73.31	82.82	84.96	86.81
GI	85.89	73.31	82.82	84.96	86.81
SU	85.89	73.31	82.82	84.96	86.81
RF	85.89	73.31	81.53	82.51	88.03
CFS	84.62	80.98	61.04	81.60	80.98
Proposed Technique	92.64	83.13	84.62	85.44	89.57

The result in Table VIII shows the proposed framework using the proposed filter selection technique compared to existing technique can achieve high accuracy rates for each classifier. According to Table VIII, it can be seen the accuracy of the existing technique is less than the proposed technique due to the existing the irrelevant and redundant features. Based on these results, the proposed technique outperforms the existing technique (CHI, IG, GI, SU, RF), even with the CFS (the filter-based feature subset techniques). The RF technique produced the most selected features, but the accuracy is not quite different from others (CHI, IG, GI, SU as shown in Table VIII). The RF technique raised the highest accuracy in classifier SVM and the lowest accuracy in ANN and KNN, while in DT or NB have the same result from others (CHI, IG, GI, and SU).

B. Discussion

According to the results (Table VI), it concludes that the best classifier to classify student's academic success is DT. This statement is supported by the previous studies that appropriate classifier is one of main factor besides feature selection in predicting academic success, and DT leads to the highest accuracy and performs well in a large dataset [1], [2]. This happen because the feature CGPADesc is used as the main feature. The results proved that the proposed framework is feasible to implement. According to Tables VII and VIII, it can be concluded that the criteria of good feature selection are not only determined by the number of selected features but also by the kind of selected features. Even though the result of CFS is the least (Table VII), but the accuracy is the lowest result (as shown in Table VIII). This happen because the result still consist of redundant and irrelevant features and have no relevant features, so it affects to the accuracy of the model [5]. The techniques (CHI, IG, GI, SU) have exactly the same results in each classifier because the dataset (the selected features produced by each technique) is the same. The proposed technique raised the highest accuracy rate even though the number of selected features (14) is greater than CFS (3) due to the nonexistence of the redundant and irrelevant features. It proved that feature selection is another main factor besides the classifier in predicting academic success aligned with the previous studies [1], [2], and the proposed feature selection leads to the highest accuracy and performs well in a large dataset. From the whole result, this research has proven that the framework develop using the proposed feature selection technique reached the highest accuracy score over all the existing technique. This research also proved that the feature selection can increase the prediction model accuracy, and it aligned with the previous studies [17]–[22].

V. CONCLUSION

Generating graduates with better academic performance through a prediction model is a crucial factor and challenging task in higher education. This study evaluates filter-based feature ranking techniques for predicting academic success and proposes a novel feature selection technique to improve the performance of the prediction model through a framework. The proposed framework will act as a guide that gives suggestions for creating the prediction model. Based on the accuracy rates used to evaluate performance, the best model was chosen. According to the results, it conclude that the best classifier to classify student's academic success is the DT with accuracy = 92.64%. The results show that the proposed framework utilizing the proposed technique significantly improves the accuracy of the proposed prediction models. The proposed technique increase the accuracy compared to existing technique from 2-8%. These results indicate the proposed framework utilizing the proposed technique effectively used to predict student success for the higher education institution making policies or giving intervention to poor academic performance students.

REFERENCES

[1] A. Hellas *et al.*, "Predicting academic performance: A systematic literature review," *Annu. Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE*, pp. 175–199, 2018, doi: 10.1145/3293881.3295783.

[2] E. Alyahyan and D. Düştegör, "Predicting academic success in higher education: literature review and best practices," *Int. J. Educ. Technol. High. Educ.*, vol. 17, no. 1, 2020, doi: 10.1186/s41239-020-0177-7.

[3] Dafid and Ermatita, "Filter-Based Feature Selection Method for Predicting Students' Academic Performance," *2022 Int. Conf. Data Sci. Its Appl. ICoDSA 2022*, pp. 309–314, 2022, doi: 10.1109/ICoDSA55874.2022.9862883.

[4] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. Cham: Springer International Publishing, 2015.

[5] G. Manikandan and S. Abirami, "An efficient feature selection framework based on information theory for high dimensional data," *Appl. Soft Comput.*, vol. 111, 2021.

[6] K. P. Shroff and H. H. Maheta, "A comparative study of various feature selection techniques in high-dimensional data set to improve classification accuracy," *IEEE Trans. Knowl. Data Eng.*, 2015, doi: 10.1109/ICCCI.2015.7218098.

[7] N. Rachburee and W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," *Proc. - 2015 7th Int. Conf. Inf. Technol. Electr. Eng. Envisioning Trend Comput. Inf. Eng. ICITEE 2015*, pp. 420–424, 2015, doi: 10.1109/ICITEED.2015.7408983.

[8] M. Malekipirbazari, V. Aksakalli, W. Shafqat, and A. Eberhard, "Performance comparison of feature selection and extraction methods with random instance selection," *Expert Syst. Appl.*, vol. 179, no. February 2020, p. 115072, 2021, doi: 10.1016/j.eswa.2021.115072.

[9] C. Anuradha and T. Velmurugan, "Performance Evaluation of Feature Selection Algorithms in Educational Data Mining," *Int. J. Data Min. Tech. Appl.*, vol. 5, no. 2, pp. 131–139, 2016, doi: 10.20894/ijdmta.102.005.002.007.

[10] W. Zheng, "A Comparative Study of Feature Selection Methods," *Int. J. Nat. Lang. Comput.*, vol. 7, no. 5, pp. 01–09, 2018, doi: 10.5121/ijnlc.2018.7501.

[11] S. Asim, A. Shah, H. M. Shabbir, S. U. Rehman, and M. Waqas, "A Comparative Study of Feature Selection Approaches: 2016-2020," *Int. J. Sci. Eng. Res.*, vol. 11, no. February, pp. 469–478, 2020.

[12] M. B. de Moraes and A. L. S. Gradvohl, "A comparative study of feature selection methods for binary text streams classification," *Evol. Syst.*, vol. 12, no. 4, pp. 997–1013, 2021, doi: 10.1007/s12530-020-09357-y.

[13] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, 2016, doi: 10.1109/TEVC.2015.2504420.

[14] B. Ghotra, S. McIntosh, and A. E. Hassan, "A large-scale study of the impact of feature selection techniques on defect classification models," *IEEE Int. Work. Conf. Min. Softw. Repos.*, no. May 2017, pp. 146–157, 2017, doi: 10.1109/MSR.2017.18.

[15] A. R. S. Parmezan, H. D. Lee, N. Spolaôr, and F. C. Wu, "Automatic recommendation of feature selection algorithms based on dataset characteristics," *Expert Syst. Appl.*, vol. 185, no. July 2020, p. 115589, 2021, doi: 10.1016/j.eswa.2021.115589.

[16] D. A. A. Gnana, "Literature Review on Feature Selection Methods for High-Dimensional Data," vol. 136, no. 1, pp. 9–17, 2016.

[17] A. U. Khasanah and Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 215, no. 1, doi: 10.1088/1757-899X/215/1/012036.

[18] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 2, pp. 447–459, 2018, doi: 10.11591/ijeecs.v9.i2.pp447-459.

[19] N. Priyasadie and S. M. Isa, "Educational Data Mining in Predicting Student Final Grades on Standardized Indonesia Data Pokok Pendidikan Data Set," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 212–216, 2021, doi: 10.14569/IJACSA.2021.0121227.

[20] M. Zaffar, M. A. Hashmani, K. S. Savita, and S. S. H. Rizvi, "A study of feature selection algorithms for predicting students academic performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, pp. 541–549, 2018, doi: 10.14569/IJACSA.2018.090569.

[21] J. D. Febro, "Utilizing feature selection in identifying predicting factors of student retention," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, pp. 269–274, 2019, doi: 10.14569/ijacsa.2019.0100934.

[22] P. Sokkhey and T. Okazaki, "Study on dominant factor for academic performance prediction using feature selection methods," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, pp. 492–502, 2020, doi: 10.14569/IJACSA.2020.0110862.

[23] B. Venkatesh and J. Anuradha, "A Review of Feature Selection and Its Methods," vol. 19, no. 1, pp. 3–26, 2019, doi: 10.2478/cait-2019-0001.

[24] Thomas M. Connolly and C. E. Begg, Database Systems A Practical Approach to Design, Implementation, and Management. Pearson Education, 2015.

[25] A. Silberschatz, H. F. Korth, and S. Sudarshan, *Database System Concepts*, 6th ed., no. 6. McGraw-Hill, 2010.

[26] R. Elmasri and S. B. Navathe, *Fundamentals of Database Systems*, 7th ed. Pearson, 2016.

[27] N. A. Al-thanoon, Z. Yahya, and O. Saber, "Chemometrics and Intelligent Laboratory Systems Feature selection based on a crow search algorithm for big data classi fi cation," *Chemom. Intell. Lab. Syst.*, vol. 212, no. September 2020, p. 104288, 2021, doi: 10.1016/j.chemolab.2021.104288.

[28] S. L. Shiva Darshan and C. D. Jaidhar, "Performance Evaluation of Filter-based Feature Selection Techniques in Classifying Portable Executable Files," in *Procedia Computer Science*, 2018, vol. 125, pp. 346–356, doi: 10.1016/j.procs.2017.12.046.

[29] B. Nouri-moghaddam, M. Ghazanfari, and M. Fathian, "A novel multi-objective forest optimization algorithm for wrapper feature selection," *Expert Syst. Appl.*, vol. 175, no. December 2020, p. 114737, 2021, doi: 10.1016/j.eswa.2021.114737.

[30] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," vol. 9340, pp. 1–13, 2016, doi: 10.1109/TC.2016.2519914.

[31] C. C. Aggarwal, *Data Mining: The Textbook*, no. April. Springer International Publishing Switzerland, 2016.