Student Pass Rates Prediction Using Optimized Support Vector Machine and Decision Tree

1st Xiaofeng Ma
SOUTHWEST UNIVERSITY
COLLEGE OF COMPUTER & INFORMATION SCIENCE
Chongqing, CHINA
aa112299@email.swu.edu.cn

2nd Zhurong Zhou
SOUTHWEST UNIVERSITY
COLLEGE OF COMPUTER & INFORMATION SCIENCE
Chongqing, CHINA
zhouzr@swu.edu.cn

***Abstract*—Since student performance and pass rates in school reflect teaching level of the school and even all education system, it is critical to improve student pass rates and reduce dropout rates. Decision Tree (DT) algorithm and Support Vector Machine (SVM) algorithm in data mining, have been used by researchers to find important student features and predict the student pass rates, however they did not consider the coefficient of initialization, and whether there is a dependency between student features. Therefore, in this study, we propose a new concept: features dependencies, and use the grid search algorithm to optimize DT and SVM, in order to improve the accuracy of the algorithm. Furthermore, we added 10-fold cross-validation to DT and SVM algorithm. The results show the experiment can achieve better results in this work. The purpose of this study is providing assistance to students who have greater difficulties in their studies, and students who are at risk of graduating through data mining techniques.**

***Index Terms*—Features Dependencies, Initialization Coefficient Rules, Grid Search Algorithm, Decision Tree, Support Vector Machine**

I. Introduction

Data mining, also known as Knowledge Discovery, is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis [1]. Howevre, educational data mining is the new concept that applying data mining methods and ideas to the education areas, which is different from data mining concepts. Over the past few decades, it has been difficult to identify potentially valuable information from the huge amount of data stored in the education database using traditional methods, and it is a daunting task to identify and build models. Because education data mining experts and scholars are from various fields of research [2], including the realization of personalized education guidance, computer collaboration learning support, and student failure rate or non-retention in courses, the application of educational data mining technology changes this current situation [3].

The most important application of educational data mining technology is to early warn students about their academic performance and even find the impact of different features on the students. The establishment of a high-accuracy prediction of student achievement or behavior is useful for all educational institutions, because as soon as possible to identify the high failure rate of student learning, poor student achievement and low learning efficiency of students can improve the quality of teaching. Cortez and Silva [4] used the data mining algorithm to model the two data sets (mathematics and Portuguese language) of Portuguese students, mainly for two questions: Is it possible for us to predict student performance in Portuguese school using data mining algorithm? What are the important features that affect student achievement? Among them, the DT algorithm model and the NV algorithm model are higher accurate to predict student behavior. In addition, several factors that are most relevant to student achievement are identified by random forest algorithms. Random forest results show that failure, absences, first period grade and second period grade have a great impact on student achievement.

Quadri and Kalyankar [5] applied DT techniques to excavate drop out features of student data for academic performance. The research is a measure of the student's academic performance by cumulative grade point average (CGPA). The research modeled the relationship between students dropout and potential risk factors such as gender, attendance, last semester, parental education, parental income, scholarships, and so on. Experimental results show that parental income features are the most important factor in all the features that affect dropout rates in this research.

Bhardwaj and Pal [6] selected 300 students from five different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India, as the data set for the study, using the classification algorithm for student achievement studies. The study uses Bayesian algorithm to model, and the results show that the features like student's grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status are strongly related to student performance.

Yadav and Pal [7] used the classification method (e.g. DT) in data mining to predict the behavior of engineering students. In this study, the data collected from engineering students were processed using C4.5, ID3 and CART DT algorithms to predict their final grades. This study uses a data set of size 90 from VBS Purvanchal University, including features such as Branch, Sex, Students category, Students grade in High

978-1-5386-4649-6/18/$31.00 ©2018 IEEE.

School, Medium of Teaching, Mother's Qualification and so on. The main purpose of the study is to predict the number of students who may pass, fail and find key factors that affect student behavior to help improve disadvantaged students or students with learning disabilities.

Hamoud [8] applied a variety of DT algorithms (J48, RepTree and Hoeffding Tree (VFD)) to classify and predict student action. Student behavior is influenced by many factors, and the purpose of the study is to find the best algorithm for modeling student data and then evaluating it. In this work, using data set about Portuguese student on two courses (Mathematics and Portuguese) as experimental data set. After the data is preprocessed (data cleansing, exporting and deleting columns), the two source files are merged into a data set. Then we use the WEKA tool to execute the decision tree. Compared with other DT algorithms, the results show that the model built by J48 algorithm has higher accuracy rate. Through the tree diagram of different DT algorithm, we can see that failure, first period grade, second period grade and father'education have a greater influence.

Strang [9] selected a large number of samples from an online university accredited online university course (n = 228), using a mixed approach to finding the relationship between student achievement and major learning engagement factors. The results show that only the course login is found to be significantly related to the course performance.

Kostopoulos et al [10] used data mining algorithms(e.g. J48 decision tree algorithms) to predict the students behavior and even find important features affecting student behavior in higher education. This study used 344 students data from the Hellenic Open University as experimental datasets, including 16 student features. The experimental results show that, the grade of the second written assignment seems to be the important student feature affecting student behavior. Besides that, 10 fold cross-validation method is also used in the algorithm, the experiment achieved better results.

Fonseca et al [11] selected data from on-campus and distance education as the dataset in this study. Student features in the dataset included age, gender, civil status, city, income, student registration, period, class, semeste, etc. This study mainly used the educational data mining technology(e.g. Support vector Machine, Decision Tree via J48 and so on) to predict the students who may fail to avoid the student's failure. The results show that it is effective to use educational data mining techniques to predict students who may fail at an early stage.

Up to now, people are still seeking for the key features that affect the student pass rates, and trying to find ways to improve the accuracy of the student pass rates prediction. So, in this study, we used the DT and SVM algorithms optimized by the grid search algorithm to classify Portuguese students into two categories (pass / fail), in order to find the important features that affect student pass rates and predict the student pass rates. In particular, due to the existence of a certain dependencies between student features, we introduce features dependencies and expert guidance in calculating the relevance of student features. In order to initialize the coefficients and make the algorithm converge faster, we propose the initialization coefficient rule. Furthermore, we applied DT and SVM algorithms using 10-fold cross-validation. In this paper, the data set consists of two Comma Separated Values (CSV) files from data of Portuguese students about Mathematics and Portuguese lessons. Furthermore, results show that the method used in the study is effective in predicting student pass rates and can find important features that affect student' performance.

Next, the second section will give a detailed description of the methods used in this study. In the third section, we will introduce the experimental results in this study. In the fourth section, we will discuss the new findings in the study and summarize the study in the fifth section.

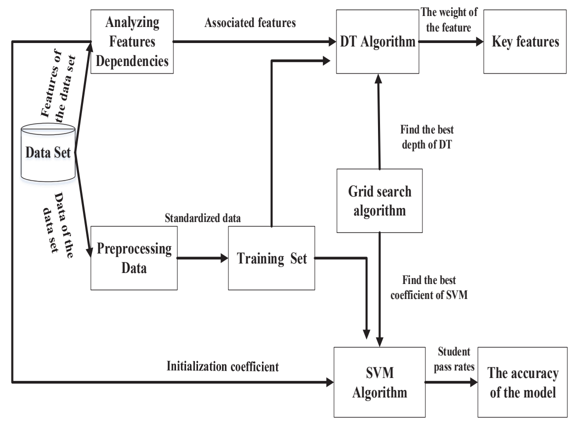

Fig. 1. Algorithm flowchart.

II. METHODS

A. Overview

In the education, data mining technology is mainly used for student data analysis, auxiliary teachers and managers to make right decisions. Data mining includes many algorithms, such as DT, association rules, SVM, clustering analysis, and so on. However, compared with other data mining algorithms, the DT is easier to be understood and can be constructed a decision tree for data sets with many attributes, and SVM can solve nonlinear problems and solve high-dimensional problems.

Therefore, in this paper, we use the classification algorithm in data mining technology, also known as supervised learning DT algorithm and SVM algorithm. Particularly, because the grid search algorithm is a method to optimize the model performance by traversing a given combination of parameters [12], the algorithm can be used to optimize the DT and SVM algorithms to improve the accuracy of the algorithm. In addition, because different student features have different effects on student pass rates and there is some dependency between the various features, we introduce the features dependences and expert guidance. This work was supported by 2017WJ033, Southwest University 2017 Annual Network and Continuing Education Educational Research Project: Online Education

Teacher Evaluation Based on Big Data Analysis. In particular, the overall algorithm flowchart is shown in Figure 1.

Next, in this paper, in Section 2.2, we describe the dependency between student features in the Portuguese student data set and propose a new concept: **Features Dependencies (FD)** . Section 2.3 describes the initialization coefficient rule. Section 2.4 describes how to preprocess the data. Section 2.5 describes the DT algorithm used in this experiment. In Section 2.6, we will introduce the SVM algorithm used in this experiment.

B. Features Dependencies

In the machine learning algorithm, the coefficients are generally stochastic initial values, which do not reflect the dependencies of the features and the importance of the features. Actually, there is often a dependency between features. In database theory, the dependencies between features are often used to reduce data redundancy. So we introduce the function dependencies [13], to explore the dependencies between student features and initialization coefficient.

In conjunction with this study, based on function dependencies, we propose a new concept: **Features Dependencies (FD)** . Compared with the function dependency, we do not use the dependency to decompose the features, but use the dependency to influence the initialization of the features coefficient. Of course, the partial dependencies and transfer dependencies still exist in features dependencies, and we name them as partial features dependencies and transfer features dependencies. We refer to the relational schema in the database to define the features schema. The features schema is represented by a triplet:

S: Subject name

D: A collection of features names that make up the subject

F: A collection of features dependencies between features data

We express it as $S(D, F)$. If and only if a subject s on D satisfies F, s is called a subject on the feature schema. Furthermore, we will give the relevant definitions of features dependencies and the coefficient initialization rules.

Definition 1: Features dependencies

Let S be a feature schema and let X and Y be nonempty sets of feature in S. We say that an instance s of S satisfies the FD: $X \rightarrow Y$, if the following holds for every pair of tuples t1 [14] and t2 [14] in s:

If $t1.X = t2.X$**, then** $t1.Y = t2.Y$.

An FD: $X \rightarrow Y$ essentially says that if two tuples agree on the values in feature X, they must also agree on the values in feature Y. So, we say that X is the decisive feature.

Definition 2: Partial features dependencies

A feature dependency $X \rightarrow Y$ is a partial dependency $X \xrightarrow{p} Y$, if some feature $A \varepsilon X$ can be removed from X and the dependency still holds; that is, for some $A \varepsilon X$, $(X - \{A\}) \rightarrow Y$.

Definition 3: Transfer features dependencies

A feature dependency $X \rightarrow Y$ in a feature schema S is a transfer dependency $X \xrightarrow{t} Y$, if there exists a set of feature Z satisfying the condition $(X \bigcup Z) \bigcap Y = \phi$ in S, and both $X \rightarrow Z$ and $X \rightarrow Y$ hold. We can also express transfer features dependencies as: $X \rightarrow Z \rightarrow Y$.

So, in this study, when calculating the information gain using the DT, the results obtained by adding features dependencies and expert guidance are taken as the initial coefficients for calculating the information gain of each features, which will be described in detail in the section 2.5. In addition, the feature dependency is also used to initialize coefficient in the SVM algorithm, which is described in detail in Section 2.6.

C. Initialization coefficient Rules

In this study, in order to make the algorithm converge faster, we propose the initialization coefficient rule. The coefficient initialization rules are shown in Table 1. W represent a collection of the weight of the student's different features in S.

TABLE I
INITIALIZATION COEFFICIENT RULES

Initialization Coefficient Rules	
Rules	***Description***
1	If $X \rightarrow Y$, then $X.W > Y.W$. That is, the weight of features X is greater than the weight of features Y.
2	If $X \xrightarrow{p} Y$, then $(X - \{A\}).W > A.W$ and $X.W > Y.W$. That is, the features X minus the weight of the features A is greater than the weight of features A, and the weight of the features X is greater than Y.
3	If $X \xrightarrow{t} Y$, then $X.W > Z.W > Y.W$. That is, the weight of features X is greater than the weight of features Z, and the weight of features Z is greater than the weight of features Y.

However, there are many features dependencies between student features, which may form a directed acyclic graph. Furthermore, there are partial order between features, in order to get the total order, we can use the topology sorting method. In this study, each student feature represents a node and is sorted according to the importance of the student features. Next, we will give the pseudo-code from the topological sorting algorithm of Wikipedia shown in Table 2.

TABLE II
TOPOLOGICAL SORTING ALGORITHM

Topological sorting algorithm
1: L →Empty list that will contain the sorted nodes
2: S ←Set of all nodes with no outgoing edges
3: for each node n in S do
4: visit(n)
5: function visit(node n)
6: if n has not been visited yet then
7: mark n as visited
8: for each node m with an edge from m to n do
9: visit(m)
10: add n to L

According to the above description, we can get a total order. In this study, compared to the partial order, we can compare all the features in the data set in the total order.

D. Preprocessing Student Data

In the real world, there are many potentially valuable information in the field of education. So the data must be standardized and preprocessed before we use the data mining to dig information. The datasets used in this paper can be downloaded from the UCI machine learning website. In this section, we mainly elaborate on how to deal with different types of data, such as numeric, nominal, binary and etc., mainly dealing with nominal types.

The nominal type of data were transformed into a 1-of-C encoding [15] and all features were standardized to a zero mean and one standard deviation. This means that instead of normalizing between 0 and 1 each category is given its own slot [15] . For example, there are five types of work, teachers, doctor, civil servants (such as administrative or police), at home or other. If the mother is a teacher, you can use the binary array as [1, 0, 0, 0, 0], if the mother's work is a doctor, you can use the binary array as [0, 1, 0, 0, 0]. Compared with the nominal type, we will directly use the numerical data and binary type of data translated into 1-0.

E. Decision tree method

In the field of education, the student pass rates represent the teaching quality of the school to a certain extent, and even the teaching level of the whole educational institution. Different student features have impact on student pass rates in different degrees, and it is important to find key features that affect student pass rates. The index used to measure degree of impurity is Entropy [16] . So the information entropy in the DT algorithm is often used to calculate the weight of student features. DT is a branch structure composed of rules. Leaf nodes represent the results of classification, the root node represents the key features. The process of constructing a branch structure in a DT is called recursive partitioning [17]. This branch structure can be expressed as IF-THEN form, easily understood by human. Since the data set contains features attributes and tag attributes, we can use classes to distinguish. Information entropy in the set D can be calculated by using the formula:

$$Entropy(D) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{1}$$

The information gain is used as the criterion of the node of the partition tree to determine the best features in the data set, that is, the root node of the tree. For example, the information gain of features B in the set D is expressed as :

$$Info\_gain(D,B) = Entropy(D) - \sum_{m \in Values(B)} \frac{|D_m|}{|D|} Entropy(D_m) \tag{2}$$

In this study, in order to better fit the algorithm to predict s-tudent pass rates, we select the maximum depth of the decision tree to tune its parameters. In addition, due to the existence

TABLE III
THE PREPROCESSED STUDENT CHARACTERISTICS

The preprocessed student characteristics	
Attributes	**Description**
sex	student's sex (binary: female or male)
age	student's age (numeric: from 15 to 22)
school	student's school (binary: Gabriel Pereira or Mousinho da Silveira)
address	student's home address type (binary: urban or rural)
Pstatus	parent's cohabitation status (binary: living together or apart)
Medu	mother's education (numeric: from 0 to 4)
Mjob	mother's job (nominal)
Fedu	father's education (numeric: from 0 to 4)
Fjob	father's job (nominal)
Guardin	student's guardian (nominal: mother, father or other)
famsize	family size (binary:$\leq 3 or \succ 3$)
famrel	quality of family relationships (numeric: from 1 very bad to 5 excellent)
reason	reason to choose this school (nominal: close to home, school reputation, course preference or other)
traveltime	home to school travel time (numeric: $1- \prec$ 15min, $2- \prec$ 15 to 30 min, $3-$ 30 min to 1 hour or $4- \succ$ 1 hour).
studytime	weekly study time (numeric: $1- \prec 2$ hours, $2-2$ to 5 hours, $3-5$ to 10 hours or $4- \succ$ 10 hours)
failures	number of past class failures (numeric: n if $1 \leq n \prec 3$, else 4)
schoolsup	extra educational school support (binary : yes or no)
famsup	family educational support (binary: yes or no)
activities	extra-curricular activities (binary: yes or no)
paidclass	extra paid classes (binary: yes or no)
internet	Internet access at home (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
romantic	with a romantic relationship (binary: yes or no)
freetime	free time after school (numeric: from $1-$ very low to $5-$ very high)
goout	going out with friends (numeric: from $1-$ very low to $5-$ very high)
Walc	from weekend alcohol consumption (numeric: $1-$ very low to $5-$ very high)
Dalc	workday alcohol consumption (numeric: from $1-$ very low to $5-$ very high)
health	current health status (numeric: from $1-$ very bad to $5-$ very good)
absences	number of school absences (numeric: from 0 to 93)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final period grade (numeric: from 0 to 20)

of a certain dependencies between the various features, we introduce features dependencies and expert guidance when calculating information gain. Therefore, the correlation degree of the features can be symbolized as:

$$DelayInfoGain(D, B) = W_i * Info\_gain(D, B) \quad (3)$$

Here, i represents the position of the selected important features, and W_i represents the weight of the i-th variable.

For this study, the DT is fitted by grid search algorithm. The grid search algorithm is a method of optimizing the performance of a model by traversing a given combination of parameters. From the experimental results, the accuracy of the decision tree model is about 93%, and the accuracy of the model is about 95% when the grid search algorithm and features dependencies are added. So, the experiment achieved good results.

F. Support vector machine method

For this study, compared with other algorithms, support vector machines are very advantageous for binary classification problems [18], so we use the SVM algorithm to calculate the student pass rates and the accuracy of the model. SVM is a nonlinear function of supervised learning. The biggest advantage of the SVM algorithm is that when solving the problem, it is irrelevant to the dimension of the sample, and the small sample can still maintain a strong adaptability to the new sample. The original problem of the standard SVM algorithm can be reduced to the following quadratic programming problem:

$$\min_{w,a,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{m} \xi_i \quad (4)$$

$$y_i\left[(w, x_i) + b\right] - 1 + \xi_i >= 0 \quad (5)$$

Among them, $\xi_i >= 0$, $x_i \in R^m$ is the input index vector of the support vector machine, $y_i \in \{-1, 1\}$ is the category of x_i, i = 1, . . . , w is the normal vector of the hyperplane, b is the offset of the hyperplane, ξ_i is the relaxation variable, and C is the penalty factor. In particular, we use features dependency as a basis for specifying the coefficient. Using the grid search algorithm to search the basic three kernel functions (γ, r, and d are kernel parameters) in SVM in the study, kernel function is expressed as follows [19] :

Polynomial(poly):

$$k(x_i, x_j) = (\gamma {x_i}^T x_j + r)^d, (\gamma > 0).$$

Radial Basis Function (RBF):

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), (\gamma > 0).$$

Sigmoid:

$$k(x_i, x_j) = \tanh(\gamma {x_i}^T x_j + r).$$

In the SVM algorithm, the choice of penalty factor C and kernel function is not appropriate, it will have over-fitting problem [20], so, in this study, we use a grid search algorithm to solve this problem and use this algorithm to optimize the parameters(penalty factor C and kernel parameters). Besides that, the features dependencies are used to initialize the coefficient. Last but not least, we also added 10 fold cross-validation method to verify the accuracy of the algorithm and experiment achieved good results.

III. Results

In this work, the input of the experiment is set to A, B, and C. The three input structures are tested in two models:

- **A** - with all features from Table 1 except G3 (the output);
- **B** - similar to **A** but without G2 (the second period grade);
- **C** - similar to **B** but without G1 (the first period grade).

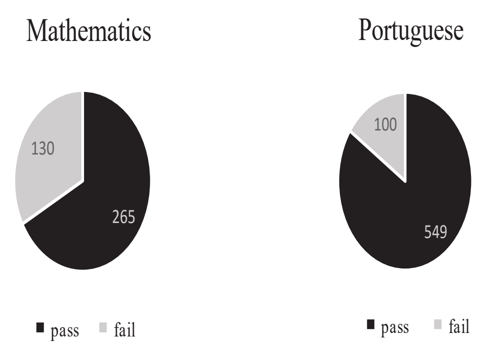

Fig. 2. The number of students passing through the course and failing the course.

Simultaneously, the Mathematics and Portuguese grades (i. e. G3) of Table 1 will be modeled using DT and SVM in this work. If G3 is greater than or equal to 10, then student in this course means pass otherwise fail. Figure 2 shows the situation of students passing through and failing the course.

Next, based on the above conditions, we will describe the experimental results from the key features that affect the student pass rates and the accuracy of the model.

A. Key features of affecting student pass rates

Until now, important student features selection is still critical to student pass rates. Although the DT algorithm has long been used by researchers to find the relevance of student features, unlike previous studies, we considered the dependencies between student features and introduce features dependencies and expert guidance in this study.

From the experimental results of the DT algorithm model, we selected Portuguese (input setup is A) and mathematical data set (input setup is B). As shown in Figure 3 and Figure 4, the important features of the student pass rates are somewhat different in both cases (not using features dependencies and using features dependencies), and the results are closer to reality after using features dependencies and expert guidance. Therefore, we think that the most important characteristics of students can choose six, namely: G1, G2, absences, failures, studytime, guardian. In addition, due to the introduction of features dependencies, the algorithm can be converged faster.

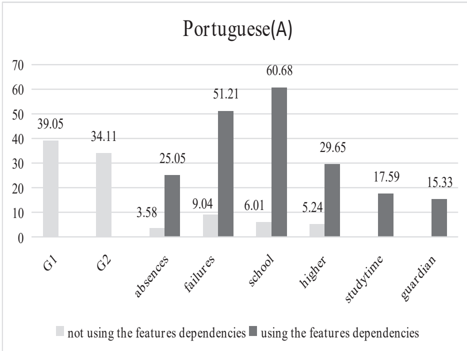

Fig. 3. The importance of the features in both cases in the Portuguese(A).

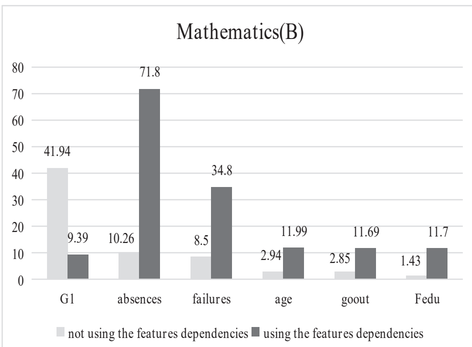

Fig. 4. The importance of the features in both cases in the Mathematics(B)

B. *The accuracy of the model of predicting student pass rates*

Although the experiment has found important features that affects student pass rates, the accuracy of the model still needs to be verified. The experimental results are valuable only if the accuracy of the model is higher. Therefore,we test the accuracy of the two algorithms using a 10-fold cross-validation method and compare the accuracy of the two algorithm models in this study with the accuracy of the model in Document [4] , and the results are shown in Figure 5 and Figure 6.

From Figure 5 and Figure 6, we can see that the model established by the two algorithms used in this experiment has

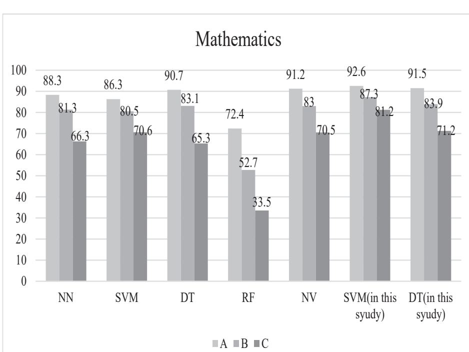

Fig. 5. The accuracy of SVM algorithm model and DT algorithm model in the Mathematics.

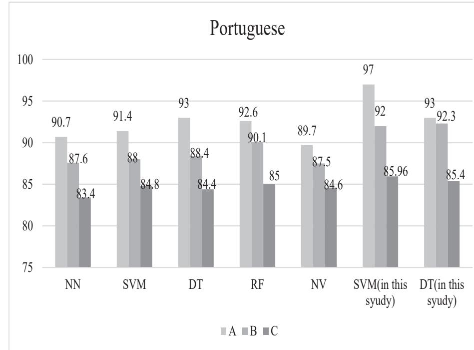

Fig. 6. The accuracy of SVM algorithm model and DT algorithm model in the Portuguese

high accuracy, that is, we can use this model to predict the student pass rates and find out the key features that affect the student pass rates.

IV. Discussion

From the experimental results, we found that there are two sets: the set of important student features obtained by the machine learning algorithm and the set of the decisive student features by the features dependency. The decisive student features are the left features that satisfies the features dependencies and play a decisive role. The important student features are that student features of the larger weight after the sorting, used as a decisive feature of the reference basis, but do not play a decisive role.

Through these two sets, we assume that there are two kinds of relationships between the two sets : intersections and disjoint, so as to further determine the initialization coefficient and make algorithm converge faster.

We use FDA to represent the set of adding features dependencies. We use MLB to represent the set of student features obtained by machine learning algorithm. Next, we discuss these two relationships in detail.

1) Intersection

If there are some features belonging to both the features set FDA and the MLB, the set of these features is called the intersection of FDA and MLB. We express it as $FDA \bigcap MLB$. Obviously, $FDA \bigcap MLB = \{x \in FDA \ and \ x \in MLB\}$, x stands for student features. In this study, once the situation occurs, we regard the intersection of student features set FDA and MLB as the decisive features and change the initial coefficient of calculation of the importance of student features, so that the experimental results are more real and reliable.

2) Disjoint

If there are some student features belonging to the features set FDA or belonging to the features set MLB, and meeting the conditions $FDA \bigcap MLB = \emptyset$, the set, composed of these features, is called the disjoint of FDA and MLB. In this study, once the situation occurs, we use the machine learning algorithm to get the features set MLB as the decisive features.

Furthermore, we find that there is still shortcomings in the above relations if there is the weight of the student feature in the set FDA is greater than the student feature in the set MLB. In the study, when this happens, we first judge what the relationship between the two sets, and then find out the weight of the student features in the set FDA is greater than the student features in the MLB. Finally, we add these features to the set MLB and use the features in the MLB as the decisive features.

Next, in section 5, we will summarize the study.

V. Conclusion

In the education field, many of the previous researchers have shown great interest in student pass rates or even finding important student features. However, they often overlook the links that exist between student features. Different from the previous study, in this study, we consider the dependency between student features, introducing features dependencies and expert guidance to initialize coefficient, so that the algorithm can be converged faster. In particular, in order to further determine the initial values of the coefficients, we propose the initialization coefficients rules.

Furthermore, in order to improve the accuracy of the model, we use the SVM algorithm and the DT algorithm optimized by the grid search algorithm to predict the student pass rates and use the information gain to find out the features that have a great influence on the student performance. Experiments show that the algorithm has achieved good result, so we can use the model to identify the important features of students and student pass rates.

In the future work, we intend to study from the following two points:

1). Although we have explored the relationship between the set of student features obtained by adding features dependencies and the set of student features using machine learning at the theoretical level, it is not implemented in the algorithm. Therefore, in the future study, we will focus on the implementation of this part of the algorithm.

2). In this study, we used two data sets from Portuguese students, but with a small amount of data and offline data. Therefore, in future work, we hope to collect online and offline data of students from other schools and use the algorithm in this study to predict the student pass rates in blended learning.

References

[1] A.O.Ogunde, and D.A .Ajibade. Data Mining System for Predicting University Students' Graduation Grades Using ID3 Decision Tree Algorithm. *Journal of Computer Science and Information Technology*, 2015.

[2] R.S.J.D.Baker. Data mining for education. *International encyclopedia of education*, 7, 112-118, 2010.

[3] R.S.J.D.Baker, and Kalina Yacef. The state of educational data mining in 2009: A review and future visions. JEDM-Journal of Educational Data Mining, 3-17, 2009.

[4] P.Cortez and A.Silva. Using Data Mining to Predict Secondary School Student Performance, 2008. In A. Brito and J. Teixeira Eds. Proceedings of 5th FUture BUsiness TEChnology Conference,5-12, Porto, Portugal, EUROSIS, ISBN 978-9077381-39-7, 2008.

[5] Mr.M.N.Quadri and Dr.N.V.Kalyankar. Drop Out characteristics of Student Data for Academic Performance Using Decision Tree Techniques. Global Journal of Computer Science and Technology, 2010.

[6] B. K. Bharadwaj and S. Pal. Data Mining: A prediction for performance improvement using classification. International Journal of Computer Science and Information Security (IJCSIS), 136-140, 2011.

[7] Surjeet Kumar Yadav and Saurabh Pal. Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification. World of Computer Science and Information Technology Journal (WCSIT), 2012.

[8] Alaa Khalaf Hamoud. Selection of Best Decision Tree Algorithm for Prediction and Classification of Students' Action, 2016.

[9] Kenneth David Strang. Beyond engagement analytics: which online mixed-data factors predict student learning outcomes, 2017.

[10] Georgios Kostopoulos, Anastasia-Dimitra Lipitakis, Sotiris Kotsiantis and George Gravvanis. Predicting Student Performance in Distance Higher Education Using Active Learning. Engineering Applications of Neural Networks, 75-86, 2017.

[11] Evandro B. Costa, Baldoino Fonseca, Marcelo Almeida Santana, Fabrłsia Ferreira de Arajo, Joilson Rego. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior, 247-256, 2017.

[12] Surjeet Kumar Yadav, Saurabh Pal and Yong Liu . An Improved Grid Search Algorithm of SVR Parameters Optimization, 2012.

[13] Ramez Elmasri and Shamkant B.Navathe. Fundamentals of database systems, 2011.

[14] Raghu Ramakrishnan and Johannes Gehrke. Database Management Systems, 2000.

[15] James McCaffrey. Neural Network Data Normalization and Encoding, 2013.

[16] Nikhil Rajadhyax, Rudresh Shirwaikar. Data Mining on Educational Domain, 2012.

[17] Carolin Strobl, James Malley, and Gerhard Tutz. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods, 14(4), 323-348, 2009.

[18] Kuo-Ping Wu, Sheng-De Wang. Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. Pattern Recognition, 710- 717, 2009

[19] Zeeshan Gillani, Muhammad Sajid Hamid Akash, MD Matiur Rahaman and Ming Chen. CompareSVM: supervised, Support Vector Machine (SVM) inference of gene regularity networks. BMC Bioinformatics, 2014.

[20] Yongsheng Ding, Xinping Song and Yueming Zen. Forecasting financial condition of Chinese listed companies based on support vector machine. Expert Systems with Applications, 3081-3089, 2008.