UNIVERSITY OF ECONOMICS AND LAW
FACULTY OF INFORMATION SYSTEMS

COURSE: DATA MINING

Detection of how performances affect students' math grades in secondary school by using classification techniques on weka

Supervisor:

Nguyen Thon Da Ph.D

Ho Chi Minh City, 2022

Table of contents

List of figures

List of tables

Abstract

Although the statistics keep Portugal at Europe's tail end due to its high student failure rates, the educational level of the Portuguese population has improved in the last decades. That was a great evolution in teaching work in secondary school. This research aims at investigating the interrelation among factors influencing skills of students in secondary education of two Portuguese schools and their impact on mathematical performance. The identified factors are student grades, demographic, social and school related features. The data was collected by using school reports and questionnaires. Dataset was provided regarding the performance in mathematics. The Decision Tree, Naïve Bayes, Support vector machine, Random Forest, Stochastic Gradient Descent were implemented on the dataset to clarify factors having impact on mathematics performance and classification model was used to accurately predict the classification in student academic performance as ‘Pass’ or ‘Fail’ using identified above factors.

Keywords – Data Mining, Classification, Decision Tree, Naïve Bayes, Support vector machine, Random Forest, Stochastic Gradient Descent

1. Introduction

Education is a very important issue regarding the development of a country [1]. One way to accomplish the higher level of quality in higher education scheme is by predicting student’s academic performance and thereby taking early actions to improve student’s performance and teaching quality. The relevant knowledge is hidden with the educational data set and it is extractable during data mining techniques. [2] In 2003, online student grades from the Michigan State University were modeled using three classification approaches (i.e. binary: pass/fail; 3-level: low, middle, high; and 9-level: from 1 - lowest grade to 9 - highest score). The database included 227 samples with online features (e.g. number of corrected answers or tries for homework) and the best results were obtained by a classifier ensemble (e.g. Decision Tree and Neural Network) with accuracy rates of 94% (binary), 72% (3-classes) and 62% (9- classes). [3]

Data mining is a step in the knowledge discovery process consisting of particular data mining algorithms that, under some acceptable computational efficiency limitations, finds patterns or models in data. In other words, the goal of knowledge discovery and data mining is to find interesting patterns and/or models that exist in databases but are hidden among the volumes of data [4]. Education data mining is the application of data mining to different human undertakings, including education. A field of study called "educational data mining" focuses on creating tools for examining the distinctive kinds and patterns of data that come from educational institutions and on using such tools to learn more about students and the environments in which they learn.

Data mining methods or approaches include Classifications, Clustering, Naïve Bayesian, decision trees, neural networks and Fuzzy rules [5]. Prediction of student performance is helpful in order to provide a student with the necessary assistance in the learning process [6]. An institution may be able to alter the outcomes of a new group of students by changing the components that contributed to past success by having knowledge of the potential outcomes of the learning process based on prediction findings.

The first model we use is Decision Tree - Modeling decisions and results using decision trees is one technique to map decisions in a branching structure. Decision trees are used to estimate the likelihood that various iterations of decisions will be successful in achieving a particular goal.

Second, we approach the Naïve Bayes study – a classification algorithm for (binary) and multi-class classification problems [7]. When the method is explained using binary or category input values, it is simplest to understand.

Next, we have Support Vector Machine or SVM which is one of the most popular Supervised Learning algorithms and is used for Classification as well as Regression problems. The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future. A hyperplane is the name given to this optimal decision boundary.

We also use Random Forest Algorithm, it is a classifier that uses many decision trees on different subsets of the input dataset and averages the results to increase the dataset's predicted accuracy [8]. Instead of depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions.

Another approach of our work is the Stochastic Gradient Descent (SGD). This is an iterative technique for optimizing an objective function with acceptable smoothness qualities is stochastic gradient descent. Since it uses an estimate of the gradient instead of the actual gradient (derived from the whole data set), it can be thought of as a stochastic approximation of gradient descent optimization (calculated from a randomly selected subset of the data). This lessens the extremely high computational cost, especially in high-dimensional optimization problems, allowing faster iteration, saving processing memory and processing time compared to other Gradient Descent methods.

By approaching patterns or models in data mining, we hope to be able to solve the above mentioned correlation question.

2. Related work

College students' achievement in Algebra was examined by Josiah and Adejoke (2014) [9] with regard to the effects of gender, age, and mathematics anxiety. The study's dependent variable is Algebra achievement, whereas its independent factors are gender, age, and mathematics anxiety.

The mean, standard deviation, independent t-test, and one-way ANOVA were used to examine the results. Findings show that students did well in the Algebra course. Additionally, there were no obvious differences in performance among genders, ages, or levels of mathematics anxiety (low, medium, and high). New students should be properly oriented on how to be top achievers in the program because they are in their first semester of college and often have average marks. Moreover, their professors ought to give them a straightforward set of directions.

Kusum Singh, Monique Granville, and Sandra Dika (2002) [10] used structural equation modeling to estimate and analyze the correlations between two motivational factors, an attitude component, and an academic engagement factor on math and science accomplishment. The outcomes showed how attitude and academic time, two motivational factors, had a favorable impact on math and science achievement. The biggest influence came from how much time was spent on homework in class.

According to Assari, Boyce, Bazargan, and Caldwell (2020) [11], there is a statistically significant relationship between ethnicity (Asian American) and parental educational attainment and student math test scores. This finding suggests that the advantage of high parental educational attainment on youth math function is higher for Non-Hispanic White youth than for Asian-American youth. Even while having educated parents is linked to better educational achievements for kids, Asian-American kids are less likely to benefit from this association than non-Hispanic White kids are. Prior research has shown that parental education has diminishing returns (a reduced ability to produce outcomes for ethnic minorities) for Blacks and Hispanics. In addition, according to Gubbins and Otero (2016) [12], parental engagement type, household income, and the parents' education level are all associated with and significantly predict higher scores on the Language and Math tests.

Al-Radaideh, et al [13] applied mainly decision tree model and other three different classification methods: ID3, C4.5 và the NaïveBayes to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Their final output clearly indicated that Decision Tree model had better prediction than other models.

Another research from Brijesh Kumar Bhardwaj [14] also showed that in present time, academic performance of students was not only affected by their own effort but also potential external factors such as Mother's Qualification, Students Other Habit, Family annual income and students' family status by using Bayesian classification to predict the students' division on the basis of previous year database.

3. Methodology

3.1 Description of dataset

The dataset consists of 396 instances, each instance is described by 24 attributes about factors that can affect student learning outcomes such as: family factors (father or mother's education level; level of family happiness,...) learning factors (whether students take extra lessons, time spent studying,…) or environmental factors (picnic, romantic,...) etc. The data set contains variables including: continuous variable, binary variable, hierarchical variable.

Table 1. Data set description

Attributes	Description
sex	Student's gender
age	Age of student
address	Student's home address type (U: urban; R: rural)

famsize	Family size (LE3: less or equal to 3; GT3: greater than 3)
Medu	Mother's education
Fedu	Father's education
traveltime	Time to go from home to school
studytime	Weekly study time
failures	Number of past class failures
schoolsup	Extra educational support
famsup	Family educational support
paid	Extra paid classes within the course subject
activities	Extra-curricular activities
higher	Wants to take higher education
internet	Internet access at home
romantic	Into a romantic relationship
famrel	Quality of family relationships
freetime	Free time after school
goout	Going out with friends
health	Current health status
absences	Number of school absences
Average	Grade point average in math

Result	Is the math score up to the standard?

3.2 Decision Tree

The J48 Decision Algorithm is a predictive machine learning model that the dependent variables also known as target value of a new sample based on various attribute values of the data available [17]. The node of a J48 decision tree denotes the different utilized attributes [18]. The fundamental distribution of data is made easier to grasp and more flexible to use with the help of the tree classification technique. J48 is an extension of ID3 that creates a decision node using the class's anticipated estimations.

J48 algorithm deals with decision trees pruning, lost or missing attribute estimations of the data and varying attribute costs [19]. The J48 algorithm can be generated via the following three stages [20]:

- Stage 1: If an instance belongs to similar class, the leaves are labeled with a similar class;
- Stage 2: For each attribute, the potential data will be figured and the gain in this data will be attained from the test conducted on attribute;
- Stage 3: Finally, the best attribute will be selected in regard to the current selection parameter.

3.3 Naïve Bayes

By determining the frequency and combinations of values in a given dataset, the Naive Bayes algorithm is a basic probabilistic algorithm that estimates a number of probabilities [15]. Uncertainty is quantified in terms of probability. In order to build models with predictive skills that can study and interpret data, the Naive Bayes approach works by assuming that all qualities are independent of another and do not depend on one another [16]. A plus of applying the Naive Bayes approach is that it takes a minimal quantity of training data in order to predict the parameters required for the classification process.

3.4 Stochastic Gradient Descent

In this case, we will use this method to find the minimum math score that the student can achieve based on the influencing factors of 22 attributes of the dataset table. The calculated grade will be the student's average math score. Based on the scores of all students included in the training dataset, the model will take the median of the scale as the standard to see if the student has met or failed to meet the math score standard.

3.5 Random Forest

The random forest classifier is defined as combination of tree classifiers. Aim of random forests algorithm is increasing the classification value. The algorithm achieves this aim by generating more than one decision tree while performing the classification process. Decision trees which are created individually come together to form a decision forest. Each classifier is generated with a random vector sampled independently from input vector [24]. In this algorithm, newly created training sets are built with replacement from the original ones. The tree is created by using a random attribute selection and a new subset. The best split on the random attributes selected is used to split the node.

3.6 Support Vector Machine

Support vector machine (SVM) is a bi-classification algorithm which is used for classifying class based on decision boundary. SVM can also be defined as a vector space-based ML algorithm that finds a decision boundary between two classes that are the furthest from any point in the training dataset. SVM is mainly used to separate data that consists of two classes (binary classification), However, the data can sometimes belong to more than two classes, in such cases the basic SVM algorithm becomes dysfunctional. The objective of SVM algorithm is finding a hyperplane in N number of attributes which distinctly classifies the data points. SVM finds a space with the maximum margin, which means the maximum distance between data points of both classes. Maximizing the margin distance provides reinforcement. As a result, future data points can be classified with higher accuracy.

3.7 Methodology Framework

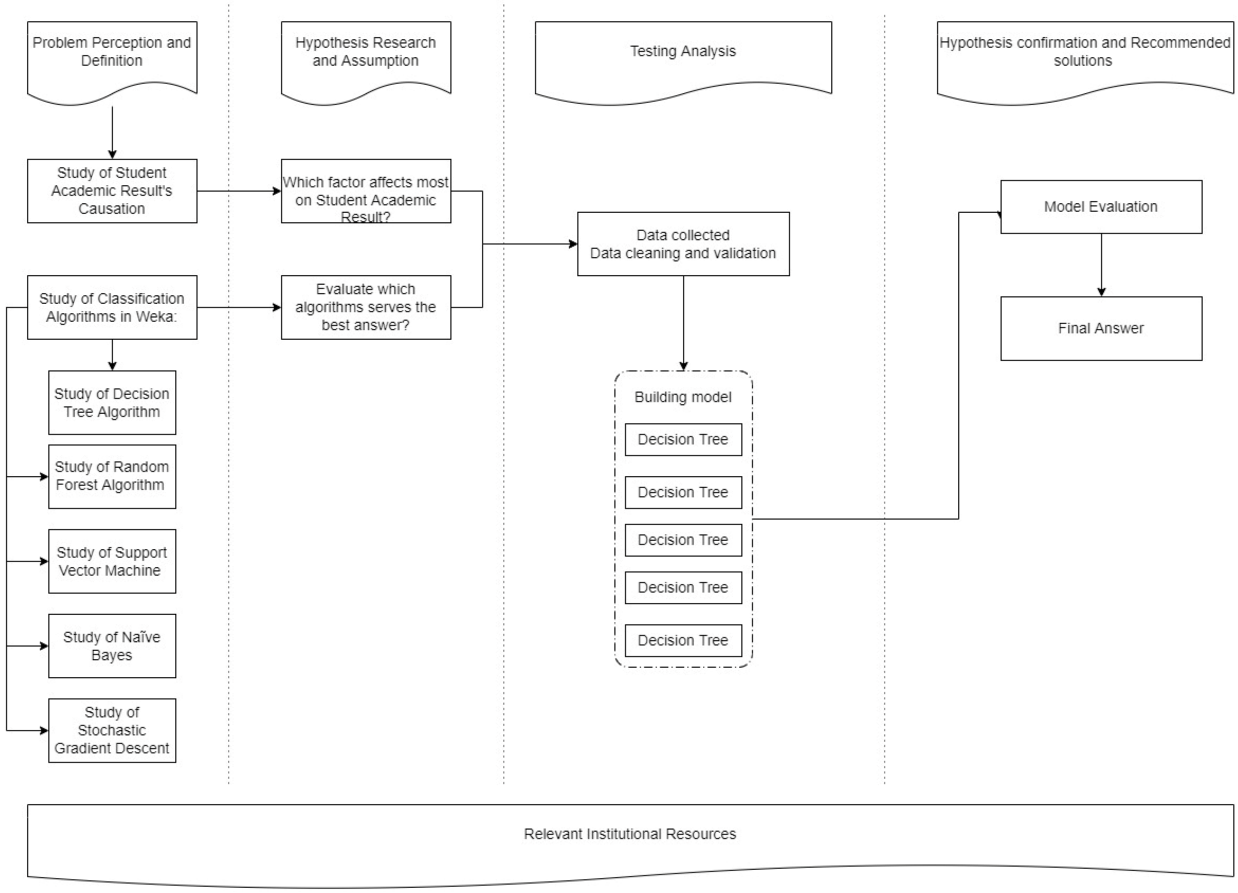

Figure 1. Methodology framework.

The process of conducting data mining analysis consists of four main stages, starting with Problem Perception and Definition when we chose our topic and studied relational resources and decided. The next step is Hypothesis Research and Assumption, we declared our problems and put on a question which algorithms served the best answer. For the Testing Analysis phase, we processed the collected data and built a model. At the final stage, we evaluated the model and stated the answer for the initial question.

4. Result of different classification methods

4.1. Decision Tree (J48)

In the first experiment, Decision Tree (J48 algorithm) classifier was used to classify the Student Performance dataset. In this experiment 10 fold cross-validation technique were used to split training and testing dataset. The J48 algorithm classifier could classify 58.73% of instances correctly.

Table 2. Results of J48 Decision Tree classification

Correctly Classified Instances	232	58.7342%
Incorrectly Classified Instances	163	41.2658%
Total Number of Instances	395	

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         232               58.7342 %
Incorrectly Classified Instances       163               41.2658 %
Kappa statistic                          0.1357
Mean absolute error                      0.4223
Root mean squared error                  0.5828
Relative absolute error                 86.9536 %
Root relative squared error            118.2701 %
Total Number of Instances              395
```

Figure 2. Stratified cross-validation of Decision Tree classification

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.445	0.312	0.503	0.445	0.472	0.136	0.595	0.504	Fail
	0.688	0.555	0.636	0.688	0.661	0.136	0.595	0.645	Pass
Weighted Avg.	0.587	0.454	0.581	0.587	0.583	0.136	0.595	0.586	

Figure 3. Detailed accuracy by class of Decision Tree Classification

```
=== Confusion Matrix ===

  a   b   <-- classified as
 73  91 |   a = Fail
 72 159 |   b = Pass
```

Figure 4. Confusion matrix of Decision Tree Classification

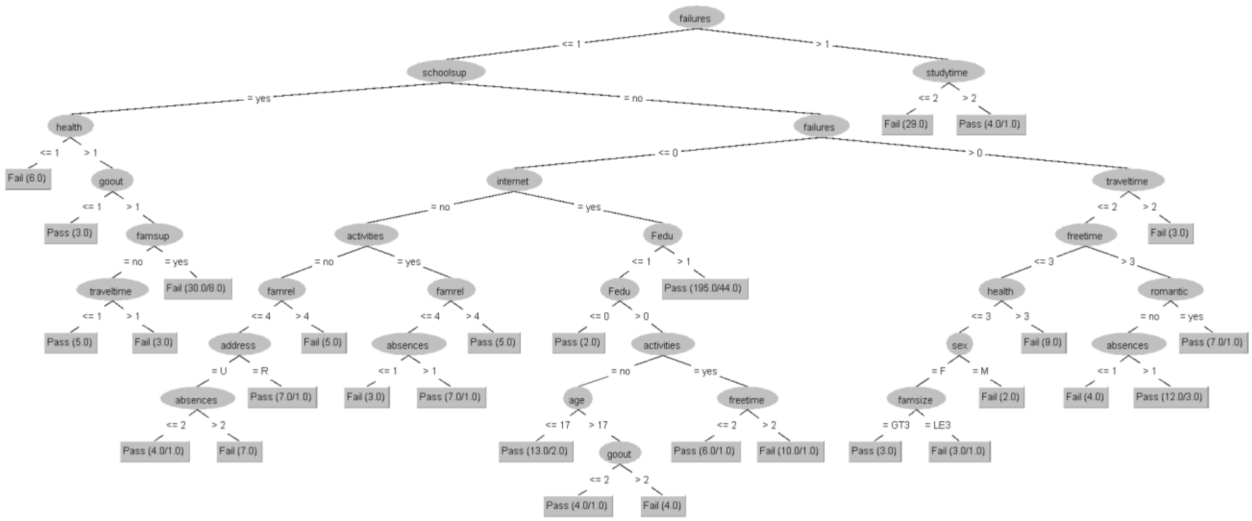

Figure 5. Decision Tree

4.2. Naïve Bayes

In the second experiment, Naïve Bayes classifier was used to classify the Student Performance dataset. In this experiment 10 fold cross-validation technique were used to split training and testing dataset. Naïve Bayes could classify 67.34% of instances correctly.

Table 3. Results of Naïve Bayes classification

Correctly Classified Instances	266	67.3418%
Incorrectly Classified Instances	129	32.6582%
Total Number of Instances	395	

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         266               67.3418 %
Incorrectly Classified Instances       129               32.6582 %
Kappa statistic                          0.2891
Mean absolute error                      0.3477
Root mean squared error                  0.4854
Relative absolute error                 71.5792 %
Root relative squared error             98.5017 %
Total Number of Instances              395
```

Figure 6. Stratified cross-validation of Naïve Bayes classification

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.421	0.147	0.670	0.421	0.517	0.307	0.729	0.675	Fail
	0.853	0.579	0.675	0.853	0.753	0.307	0.729	0.773	Pass
Weighted Avg.	0.673	0.400	0.673	0.673	0.655	0.307	0.729	0.732	

Figure 7. Detailed accuracy by class of Naïve Bayes classification

```
=== Confusion Matrix ===

  a   b   <-- classified as
 69  95 |   a = Fail
 34 197 |   b = Pass
```

Figure 8. Confusion matrix of Naïve Bayes classification

4.3 Stochastic Gradient Descent

In the third experiment, Stochastic Gradient Descent (SGD) was used to classify the Student Performance dataset. In this experiment 10 fold cross-validation technique were used to split training and testing dataset. SGD algorithm could classify 65.32% of instances correctly.

Table 4. Results of Stochastic Gradient Descent classification

Correctly Classified Instances	258	65.3165%
Incorrectly Classified Instances	137	34.6835%
Total Number of Instances	395	

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         258               65.3165 %
Incorrectly Classified Instances       137               34.6835 %
Kappa statistic                          0.2575
Mean absolute error                      0.3468
Root mean squared error                  0.5889
Relative absolute error                 71.4079 %
Root relative squared error            119.513  %
Total Number of Instances              395
```

Figure 9. Stratified cross-validation of Stochastic Gradient Descent classification

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.451    0.203    0.612      0.451    0.519      0.265    0.624     0.504     Fail
                 0.797    0.549    0.672      0.797    0.729      0.265    0.624     0.654     Pass
Weighted Avg.    0.653    0.405    0.647      0.653    0.642      0.265    0.624     0.592
```

Figure 10. Detailed accuracy by class of Stochastic Gradient Descent classification

```
=== Confusion Matrix ===

   a   b   <-- classified as
  74  90 |   a = Fail
  47 184 |   b = Pass
```

Figure 11. Confusion matrix of Stochastic Gradient Descent classification

4.4. Support Vector Machine:

In the fourth experiment, Support Vector Machine (SVM) was used to classify the Student Performance dataset. In this experiment 10 fold cross-validation technique were used to split training and testing dataset. SVM algorithm could classify 69.11% of instances correctly.

Table 5. Results of Support Vector Machine classification

Correctly Classified Instances	273	69.1139%
Incorrectly Classified Instances	122	30.8861%
Total Number of Instances	395	

```
Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         273               69.1139 %
Incorrectly Classified Instances       122               30.8861 %
Kappa statistic                          0.3333
Mean absolute error                      0.3089
Root mean squared error                  0.5558
Relative absolute error                 63.5895 %
Root relative squared error            112.7807 %
Total Number of Instances              395

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.470    0.152    0.688      0.470    0.558      0.348    0.659     0.543     Fail
                 0.848    0.530    0.693      0.848    0.763      0.348    0.659     0.676     Pass
Weighted Avg.    0.691    0.373    0.690      0.691    0.678      0.348    0.659     0.621

=== Confusion Matrix ===

   a   b   <-- classified as
  77  87 |   a = Fail
  35 196 |   b = Pass
```

Figure 12. Summary result of Support Vector Machine classification in Weka

4.5 Random Forest

In the fifth experiment, Random Forest was used to classify the Student Performance dataset. In this experiment 10 fold cross-validation technique were used to split training and testing dataset. Random Forest algorithm could classify 66.33% of instances correctly.

Table 6. Results of Random Forest classification

Correctly Classified Instances	262	66.3291%
Incorrectly Classified Instances	133	33.6709%
Total Number of Instances	395	

```
Time taken to build model: 0.32 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         262               66.3291 %
Incorrectly Classified Instances       133               33.6709 %
Kappa statistic                          0.2832
Mean absolute error                      0.4189
Root mean squared error                  0.461
Relative absolute error                 86.2524 %
Root relative squared error             93.5422 %
Total Number of Instances              395

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.482    0.208    0.622      0.482    0.543      0.289    0.701     0.630     Fail
                 0.792    0.518    0.683      0.792    0.733      0.289    0.701     0.764     Pass
Weighted Avg.    0.663    0.389    0.658      0.663    0.654      0.289    0.701     0.708

=== Confusion Matrix ===

   a   b   <-- classified as
  79  85 |   a = Fail
  48 183 |   b = Pass
```

Figure 13. Summary results of Random Forest classification in Weka

5. Results

Recall: The proportion of positive samples is calculated according to the total number of positive samples in the correct classification used. Recall = TP / (TP+FN) F1-Score: It is the harmonic mean of Recall and Precision values. The purpose here is to measure the performance value shown by the classifiers. It is mostly used to compare classifiers. F1-Score = 2 * Precision * Recall/ (Precision + Recall)

Table 7. Performance metrics

Algorithm	Class	Precision	Recall	F1_Score
Decision Tree (J48)	Fail	0.503	0.445	0.472
	Pass	0.636	0.688	0.661
Naive bayes	Fail	0.67	0.421	0.517
	Pass	0.675	0.853	0.753
Support vector machine	Fail	0.688	0.47	0.558
	Pass	0.693	0.848	0.763
Random Forest	Fail	0.622	0.482	0.543
	Pass	0.683	0.792	0.733

Stochastic Gradient Descent	Fail	0.612	0.451	0.519
	Pass	0.672	0.797	0.729

Table 8. Results of Classification Algorithms

Algorithm	Correctly Classified	Incorrectly Classified
Decision Tree (J48)	58.7342%	41.2658%
Naive bayes	67.3418%	32.6582%
Support vector machine	69.1139%	30.8861%
Random Forest	66.3291%	33.6709%
Stochastic Gradient Descent	65.3165%	34.6835%

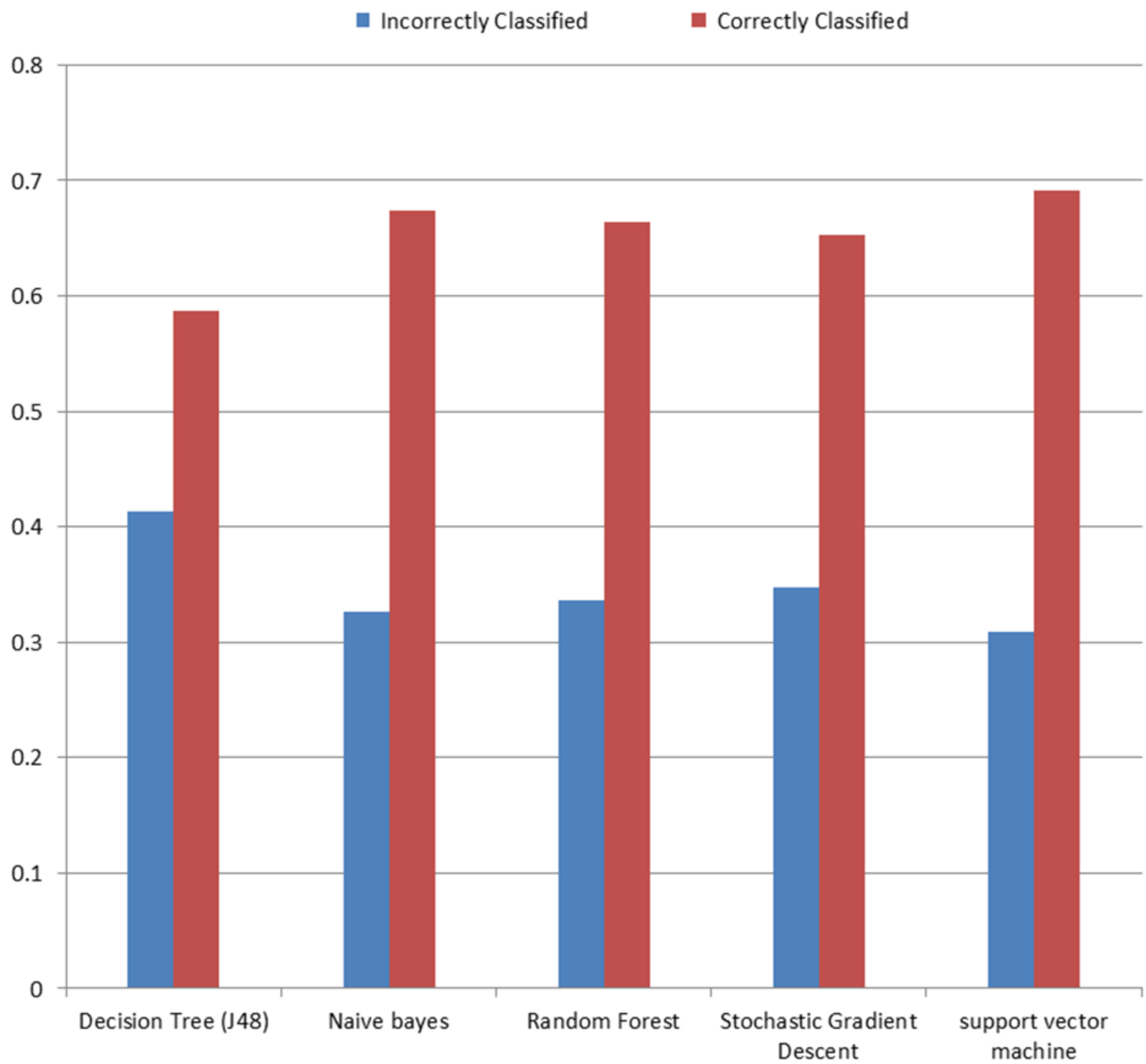

Figure 14. Results of Classification Algorithms

The results of tested classification algorithms on dataset are represented in Table X1 and Figure 1. According to obtained results, support vector machine (SVM) algorithm performed the highest accuracy with 69.11% while Decision Tree (J48) performed the lowest accuracy with 58.73%.

6. Conclusion

In this study, the accurate prediction of mathematical results by different classification techniques is the ultimate goal and we divided our dataset into two main parts as training and test in the first phase. Initially, we tried to find out factors that we can use to predict math performance. Then we make data preprocessing to clear and prepare the data. Then we apply Random Forest, Decision Tree, Naive Bayes, Stochastic Gradient Descent and support vector machine algorithms to reach the most qualified result and compare each one's scores. We observed that addition of relevant factors increased the accuracy score of algorithms. Balanced and enhanced dataset can create better solutions in this case. In addition to all, some deep learning models which showed the efficiency can be adopted to the proposed model in the future work.

As a conclusion of this paper, we found that the Support Vector Machine (SVM) algorithm works better than the others with relatively high accuracy rates. Additionally, apart from the relative features, some content-based features can also be used here. Finally, we can also determine factors that have strong impact on math performance to improve the quality of secondary school teaching

7. Future work

Expected to develop in the future, because the accuracy of applied machine learning models is not high, it is necessary to understand and analyze more deeply the impact of variables on student learning outcomes. For example, a closer look at how family factors such as happiness or family size affect academic performance, could lead to a more detailed analysis of how members' attitudes will affect students or not. Not only the family factor, the study also needs more detail in future development on how the importance of each component score column will affect the GPA as well as factors that can affect the component score results.

References

1. Gaviria, A. (2002). Los quesuben y los quebajan: educacion y movilidad social en Colombia. Fedesarrollo, Alfaomega.
2. Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. *Procedia Technology*, *25*, 326-332.
3. Cortez, P., & Silva, A. M. G. (2008). Using Data Mining to Predict Secondary School Student Performance. In A. Brito, & J. Teixeira (Eds.), Proceedings of 5th Annual Future Business Technology Conference, Porto, 5-12.
4. Bae, E., Bailey, J: "COALA: A novel approach for the extraction of an alternate clustering of high quality and high dissimilarity". Proceedings of the Sixth International Conference on Data Mining. Pp. 53 – 62. (2006).
5. Kolo, D. K., & Adepoju, S. A. (2015). A decision tree approach for predicting students' academic performance.
6. Bekele, R., Menzel, W. "A bayesian approach to predict performance of a student (BAPPS): A Case with Ethiopian Students". Journal of Information Science (2013).
7. Shukla, p. (2021). Identifying the potential customers for loans.
8. Kirasich, K., Smith, T., & Sadler, B. (2018). Random forest vs logistic regression: binary classification for heterogeneous datasets. SMU Data Science Review, 1(3),9.
9. Josiah, O., & Adejoke, E.O. (2014). Effect of Gender, Age and Mathematics Anxiety on College Students' Achievement in Algebra. American Journal of Educational Research, 2, 474-476.

10. Singh, K., Granville, M., & Dika, S. (2002). Mathematics and Science Achievement: Effects of Motivation, Interest, and Academic Engagement. Journal of Educational Research, 95, 323-332.

11. Assari, S., Boyce, S., Bazargan, M., Caldwell, C. H., & Zimmerman, M. A. (2020, April). Place-based diminished returns of parental educational attainment on school performance of non-Hispanic White youth. In Frontiers in Education (Vol. 5, p. 30). Frontiers Media SA.

12. Gubbins, V., & Otero, G. (2016). Effect of the parental involvement style perceived by elementary school students at home on Language and Mathematics performance in Chilean schools. Educational Studies, 42(2), 121-136.

13. AI-Radaideh, Q. A., AI-Shawakfa, E.M., and AI-Najjar, M. I., “Mining Student Data using Decision Trees”, International Arab Conference on Information Technology (ACIT 2006), Yarmouk University, Jordan, 2006

14. Brijesh Kumar Bhardwaj (2011) Data Mining: A prediction for performance improvement using classification from (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011

15. N. Ye and N. Ye, “Naïve Bayes Classifier,” in Data Mining, 2020.

16. S. A. Pattekari and A. Parveen, “Prediction System for Heart Disease Using Naive Bayes,” Int. J. Adv. Comput. Math. Sci., vol. 3, no. 3, pp. 290–294, 2012.

17. Farhad, A.; Sanjay, P. Comparative Study of J48, Naive Bayes and One-R Classification Technique for Credit Card Fraud Detection using WEKA. In Advances in Computational Sciences and Technology; Research India Publications: Rohini, India, 2017; Volume 10, pp. 1731–1743. ISSN 0973-6107

18. Ihya, R.; Namir, A.; Sanaa, E.F.; Mohammed, A.D.; Fatima, Z.G. J48 Algorithms of Machine Learning for Predicting User's the Acceptance of an E-Orientation Systems. In Proceedings of the 4th International Conference on Smart City Applications, Casablanca, Morocco, 2–4 October 2019.
19. Kaur, G.; Chhabra, A. Improved J48 Classification Algorithm for the Prediction of Diabetes. Int. J. Comput. Appl. 2014, 98, 13–17.
20. Adhatrao, K.; Gaykar, A.; Dhawan, A.; Jha, R.; Honrao, V. Predicting Students performance using ID3 extension and C4.5 classification algorithms. Int. J. Data Min. Knowl. Manag. Process IJDKP 2013, 3, 39–52.