p–ISSN: 2723 - 6609 e-ISSN: 2745-5254
Vol. 5, No. 11 November 2024 http://jist.publikasiindonesia.id/

Application of Data Mining for Prediction of High School Student Graduation Rates

Muhamad Kurniawan[1*], Sani Muhamad Isa[2]
Universitas Bina Nusantara, Indonesia
Email: Muhamad.kurniawan002@binus.ac.id[1*], sani.m.isa@binus.ac.id[2]

*Correspondence

Keywords: Educational Data Mining; Mean Squared Error; student grades; classification model; feature selection

ABSTRACT

The implementation of Data Mining in the education sector aims to develop methods that are able to discover valuable knowledge from data generated in the educational environment. This can be used to increase learning efficiency by paying more attention to students who are predicted to have low grades. However, in its application, each algorithm shows different performance depending on the attributes and dataset used. In this study, a dataset of semester grades and final school exam scores was used. Some of the prediction techniques used are decision trees, support vector machines, and neural networks. Of the four scenarios for the science major at SMAN 2 and SMAN 3 Pangkalpinang with 3 different models, the Mean Squared Error value shows that the test results are in accordance with the testing dataset and can be used as predictions of students' final grades, namely the decision tree model and support vector machine. For the Social Sciences major at SMAN 2 and SMAN 3 Pangkalpinang with 3 different models, the Mean Squared Error value shows that the test results are in accordance with the testing dataset and can be used as a prediction of students' final grades, namely the support vector machine model.

Introduction

Education is one of the most essential parts of people's lives. It is used to enhance an individual's academic and financial growth (Akour & Alenezi, 2022; Li, 2024). An educated individual must contribute not only to his family but also to society and community. This is all achieved through proper learning. To be able to produce human resources who are capable, insightful, competitive, and creative, educational institutions are required to provide high-quality education for their students. The implementation of this system is achieved through the administration of school or national examinations (Akala, 2021; Wajdi et al., 2020), which serve as a criterion for graduation at the respective educational level and provide a standard for evaluation within the educational system. In the Implementation of Examinations Organized by Education and National Examination Units, participants who are taught or educated can be declared to have passed if they have completed the study period and obtained attitude or behaviour scores. with a minimum good category score and can take part in tests held by the specified agency or educational unit (Kementerian Pendidikan dan Kebudayaan, 2019). SMA Negeri 2 Pangkalpinang City and SMA Negeri 3 Pangkalpinang City are the high school education units located in the Bangka Belitung Islands Province where this educational unit annually carries out School Examinations; SMA Negeri 2 Pangkalpinang city and SMA Negeri 3 Pangkalpinang city experienced several cases of students who got School Examination (US) scores that were below average.

This is due to students' lack of preparation in the subjects that will be tested in the School Examination. The method that will be used in data mining is to produce information that will be useful so that decisions are made appropriately by the school to understand what strategies need to be applied to students and later can place more emphasis on areas of deficiency. This is done so that students can pursue fields that suit their respective abilities and continue to study other subjects with as little stress as possible. The benefit in the future is that it will be easier for students to determine their career path so that they do not fall into fields that are not suitable. This is an anticipatory step for the school to help students who have the potential to face obstacles in learning and lack achievement in the learning process.

The implementation of Data Mining in the education sector, known as Educational Data Mining (EDM) (Du et al., 2020; Fernandes et al., 2019), aims to develop methods that are able to discover valuable knowledge from data generated in the educational environment. Data mining techniques are activities that include collecting and using historical data to find regularities, patterns and relationships in large data sets. EDM uses a computational approach to analyze educational data to study questions about education. One form of application of EDM is to predict student academic performance (Altabrawee et al., 2019; Xiao et al., 2022).

This research aims to determine predictions of students' level of success in graduating based on students' semester grades as well as knowing the working process of the Decision tree, Neural Network, Support Vector Machine method in classifying students' level of success in graduating and it is hoped that there will be early treatment so that it can be predicted which students need it. Early handling displays data related to learning and predicts student grades in the form of a dashboard. The benefits of this research make it easier to analyze graduation data so that factors that influence student graduation rates can be identified, making it easier for schools to make decisions to help improve the quality of student graduation.

The scope of research in this field is limited by the following: The prediction data used by researchers was taken from class 12 of SMA Negeri 2 and SMA Negeri 3 in Pangkalpinang city. This research only concerns the prediction of graduation rates for class 12 students of SMA Negeri 2 and SMA Negeri. 3 cities of Pangkalpinang, all calculation processes provided by the system use data mining techniques with the Decision tree method, Neural Network, and Support Vector Machine. The information displayed is in the form of pattern analysis reports and data mining dashboards for student graduation rates.

Methods

Data Collection

In this stage, the researcher collected the required data, which consisted of the scores of all subjects on the mid-term exam, final semester exam, and school exams according to each major. The data taken comes from three different generations. The data mining process requires training data, which will become a model for testing other data. The data required comes from alumni data from the classes of 2021 and 2022, which will act as training data, and student data from the class of 2023, which will act as testing data. The data required also spans the time period from semester 1 to semester 4. Data such as knowledge and attitude scores will also be collected so that the problems studied can be explored more deeply. The data source comes from the school's database.

The methods used for this data collection stage were Field observations at SMA Negeri 2 and SMA Negeri 3 Pangkalpinang city to obtain data directly, make observations, and record the necessary data; and interviews to obtain deeper information, which will be in the form of research data attributes and knowledge of student attitudes at SMA Negeri 2 and SMA Negeri 3 Pangkalpinang city. Interviews will be conducted with the school principal or guidance and counselling teacher from the school.

The data used in this research is data from SMA Negeri 2 Pangkalpinang City and SMA Negeri 3 Pangkalpinang City in 2021-2023. The attributes used are based on DAPODIK data, with the number of attributes in accordance with Table 1.

Table 1. Dataset Attributes.

Attribute	Description	Possible Value
Final score	Students' final average grade	Numerical
Religious grades for semesters 1 - 4	Grades in religious studies	Numerical
Civics lessons grades for semesters 1 - 4	Grades in Civics lessons	Numerical
Indonesian language grades for semesters 1 - 4	Grades in Indonesian language lessons	Numerical
Mathematics grades for semesters 1 - 4	Grades in mathematics	Numerical
Chemistry grades for semesters 1 - 4	Grades in chemistry class	Numerical
Physics grades for semesters 1 - 4	Grades in Physics	Numerical
Biology grades for semesters 1 - 4	Grades in biology lessons	Numerical
Geography grades for semesters 1 - 4	Grades in geography lessons	Numerical
Economic grades for semesters 1 - 4	Grades in economics lessons	Numerical
History grades for semesters 1 - 4	Grades in history lessons	Numerical
Sociology grades for semesters 1 - 4	Grades in sociology lessons	Numerical
Local content grades for semesters 1 - 4	Grades for local content lessons	Numerical
English grades for semesters 1 - 4	Grades in English lessons	Numerical

Source: Data processed, 2024

Data Preprocessing

Clean data is needed so that the data used for the analysis process meets the researchers' needs. The Data Preprocessing stages consist of Data Cleaning, Data Integration, Data Selection/Reduction, and Data Transformation processes. Data Preprocessing aims to form clean data from raw data obtained from observations and interviews with Pangkalpinang City Public High Schools. The samples that will be taken in this research can be seen in the following table:

Table 2. Sample Data Taken.

School Name	The class year 2021	Class year 2022	The class year 2023	Total
SMAN 2 PANGKALPINANG	306	295	274	875
SMAN 3 PANGKALPINANG	0	163	151	314

Source: Data processed, 2024

In Table 2, the totals from SMA Negeri 2 Pangkalpinang City and SMA Negeri 3 Pangkalpinang City, the researchers took a total of data samples with a total of 1,189 students. The data that has been processed is then divided into two data sets, namely training data and testing data, with a ratio of 80% for training data and 20% for testing data. The purpose of the training data is to help the algorithmic model to learn patterns from the data that has been collected. , while the training data is used to evaluate and measure the accuracy of the prediction results from the previously created model.

Results and Discussion

Preprocessing Data

At this stage, data preprocessing is carried out before the data can be used to avoid data that makes the model inaccurate. There are several stages of this preprocessing, such as Removing Missing Values, Removing Duplicate Values, and Sampling.

1. Remove Missing Values

At this stage, the first stage in pre-processing, data cleaning is carried out not only by selecting attributes appropriate to the subjects of each class of students who graduate with subjects that will be tested in the final school exam but also by cleaning up any missing power that exists in training data and testing data.

In the SMA Negeri 2 Pangkalpinang dataset, data is missing in both available datasets. Missing data occurs due to many possibilities. The possibility of what happens is as follows:

Table 3. Possible Causes of Missing Data.

Missing Data	Possible Occurrence
There is no data on student grades in semesters 3 and 4	Students change schools when they are in the 11th grade of high school
There is no data on student grades in semesters 1, 2, 3 and 4	Students change schools when they enter the 10th grade of high school
There is no data on student grades in semesters 2, 3 and 4	Students move schools when entering the 2nd semester of grade 10 high school
There is no data on student grades in semesters 1,2	New students enter the school in grade 11 of high school

Source: Data processed, 2024

Table 3 shows that four possibilities caused missing data in these semesters; the possibility is that the student in question has gone to another school, or the possibility that the missing is also a student who just entered the school when the student was in grade 11 of high school.

Table 4. Number of Missing Data for Each Department.

Missing Data	Data Training		Data Testing		Possible Occurrence
	IPS	IPA	IPS	IPA	
Semester 1	11 Record	9 Record	0 Record	0 Record	Students continue at the school
Semester 1 and 2	13 Record	14 Record	0 Record	0 Record	Students change schools when they enter the 11th grade of high school
Semester 1, 2, and 3	8 Record	12 Record	0 Record	0 Record	New transfer student
Semesters 1, 2, 3, and 4	10 Record	15 Record	0 Record	0 Record	New transfer student entering 12th grade of high school

Source: Data processed, 2024

Table. 4 provides the amount of missing data obtained in the training data and data set in each scenario, and also for each department and the possible causes of the missing data. In the science department training data for the first scenario for semester 1, there are 11 missing data; for the second scenario for semesters 1 and 2, there are 13 missing data; for the third scenario for semesters 1, 2 and 3, there are eight records missing data, and for the fourth-semester scenario 1, 2, 3, and 4 contain missing data of 10 records.

Meanwhile, in the IPS department training data for the first scenario for semester 1, there were nine records missing data; for the second scenario for semesters 1 and 2, there were 14 missing data; for the third scenario for semesters 1, 2 and 3 there were 12 records missing data, and for the fourth scenario Semesters 1, 2, 3, and 4 contained 15 missing data. There were no missing data in testing data in any scenario for either science or social studies at SMAN Negeri 2 Pangkalpinang; for testing data and training data at SMA Negeri 3 Pangkalpinang from scenarios 1 to scenario 4, there were no missing data in the science and social studies majors. Missing records can disrupt the data mining process later, so the names of students who have missing records must be cleaned or removed from the dataset table.

Data mining

At this stage, data mining processing is carried out, which aims to find and build a model from a dataset and then use the model in another dataset where you want to predict a result. In this research, data from students from the class of 2021, 2022 from semester 1 to semester 4 of

SMAN 2 and data from students from the class of 2022 from semester 1 to semester 4 from SMAN 3 Pangkal Pinang will act as training data, which will produce a classification model (Courtiol et al., 2019). Later, this model will be used in grade data from the class of 2023 at SMAN 2 and SMAN 3 Pangkalpinang as testing data. In this research, 4 test scenarios were carried out in order to determine scenarios that accurately determine predictions of students' final grades using three models, namely Decision Tree (DT), Support Vector Machine (SVM), and Neural Network (NN).

1. Decision Models tree

In this research, the Decision model tree is used to predict student grades from semester 1 to semester 4 by using two datasets, namely dataset training and dataset testing, to find out whether the Decision model This tree has accurate predictions to predict the final grades of students at SMAN 2 and SMAN 3 Pangkalpinang. This model is used to determine the scenario that has the most accurate final grade predictions so that conclusions can be drawn to find out which students need to do extra learning in order to graduate. Final exam (Mienye et al., 2019).

a. MSE Model Decision Value tree Scenario 1 Science Department of SMAN 2 and SMAN 3 Pangkalpinang

The first scenario where the value to be predicted is the average value from semester 1 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 1 is 4.959515007849291, with calculated differences and percentage differences, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table .

Table 5. Scenario 1 Science Value actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
85.93	85.71	0.22	0.25
82.86	85.93	-3.07	-3.70
84.14	83.71	0.43	0.51
79.36	83.50	-4.14	-5.21
81.00	84.57	-3.57	-4.40
80.57	80.00	0.57	0.70
83.00	81.79	1.21	1.45
79.42	79.57	-0.14	-0.17
82.00	82.14	-0.14	-0.17
80.93	84.71	-3.78	-4.67

Source: Data processed, 2024

b. MSE Model Decision Value tree Scenario 2 Science Department of SMAN 2 and SMAN 3 Pangkalpinang

The second scenario where the value to be predicted is the average value from the combination of semester 1 and semester 2 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 2 is 5.905944629014399, with calculated differences and percentage differences, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage value differences can be seen in Table 6.

Table 6 Scenario 2 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
81.36	83.86	-2.50	-3.07
82.71	79.86	2.85	3.44
82.57	88.71	-6.14	-7.43
80.21	76.86	3.35	4.18
85.79	84.86	0.93	1.08
82.50	80.79	1.71	2,072
82.43	82.50	-0.07	-0.08
83.00	84.50	-1.50	-1.80
82.29	84.14	-1.85	-2.24
81.43	82.29	-0.86	-1.05

Source: Data processed, 2024

c. MSE Model Decision Value tree Scenario 3 Science Department of SMAN 2 and SMAN 3 Pangkalpinang

The third scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, and semester 3 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 3 is 4.689620471281302 with calculated differences And calculated percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 7

Table 7. Scenario 3 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
82.71	79.71	2.99	3.62
82.21	82.07	0.14	0.17
83.36	80.71	2.65	3.17
81.43	83.71	-2.28	-2.79
85.86	85.29	0.57	0.66
82.00	84.35	-2.35	-2.87
84.64	82.43	2,210	2.61
82.00	81.36	0.64	0.78
73.64	76.93	-3.29	-4.46
83.71	80.57	3.14	3.75

Source: Data processed, 2024

d. MSE Model Decision Value tree Scenario 4 Science Department of SMAN 2 and SMAN 3 Pangkalpinang

The fourth scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, semester 3, and semester 4 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 4 is 5.013858037676613 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the

Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 8

Table 8. Scenario 4 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
81.21	80.93	0.28	0.34
82.21	81.21	1.00	1.21
80.29	81.50	-1.21	-1.50
81.64	82.21	-0.57	-0.69
81.00	78.92	2.07	2.55
82.50	85.79	-3.29	-3.98
81.14	82.43	-1.29	-1.58
82.64	83.71	-1.07	-1.29
86.36	88.71	-2.35	-2.72
85.79	85.93	-0.14	-0.16

Source: Data processed, 2024

e. MSE Model Decision Value tree Scenario 1 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The first scenario where the value to be predicted is the average value from semester 1 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 1 is 8.531357587650442 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 9

Table 9. Scenario 1 Social Studies Score actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
75.86	78.86	-3.00	-3.95
81.64	83.00	-1.36	-1.66
81.21	84.64	-3.43	-4.22
80.93	79.00	1.93	2.38
83.07	77.50	5.57	6.70
82.07	81.29	0.78	0.95
80.07	79.86	0.21	0.26
80.0	78.07	1.93	2.41
79.50	74.71	4.79	6.02
82.07	82.79	-0.72	-0.87

Source: Data processed, 2024

f. MSE Model Decision Value tree Scenario 2 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The second scenario where the value to be predicted is the average value from the combination of semester 1 and semester 2 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 2 is 7.767002838827839 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage

The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 10

Table 10. Scenario 2 IPS scores actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
76.36	82.64	-6.28	-8.22
78.21	77.00	1.21	1.54
75.86	77.71	-1.85	-2.43
81.79	78.79	3.00	3.66
78.07	77.85	0.21	0.27
75.57	75.07	0.50	0.66
79.71	78.57	1.14	1.43
81.93	79.00	2.93	3.57
76.57	82.00	-5.43	-7.09
83.50	79.21	4.29	5.13

Source: Data processed, 2024

g. MSE Model Decision Value tree Scenario 3 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The third scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, and semester 3 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 3 is 6.318856035234607 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage value differences can be seen in Table 11.

Table 11. Scenario 3 IPS Score actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
78.43	79.21	-0.78	-0.99
79.79	74.71	5.08	6.36
74.71	78.07	-3.36	-4.49
77.79	77.57	0.22	0.28
82.43	81.86	0.57	0.69
78.29	80.00	-1.71	-2.18
84.14	85.21	-1.07	-1.27
78.79	79.21	-0.42	-0.53
79.50	77.00	2.50	3.14
83.07	85.79	-2.72	-3.27

Source: Data processed, 2024

h. MSE Model Decision Value tree Scenario 4 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The fourth scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, semester 3, and semester 4 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 4 is 5.337855004238116 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE

management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 12.

Table 12. Scenario 4 IPS Score actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
78.21	79.21	-0.99	-1.27
82.07	81.86	0.21	0.25
78.50	76.93	1.57	2.00
76.57	77.64	-1.07	-1.39
83.07	81.86	1.21	1.45
76.35	74.50	1,857	2.43
78.29	81.43	-3.14	-4.01
78.50	77.71	0.78	1.00
77.07	77.93	-0.85	-1.11
76.79	79.71	-2.92	-3.80

Source: Data processed, 2024

2. Support Models Vector Machine

In this research, the Support model was also used Vector Machine was used to predict student grades from semester 1 to semester 4 by having two datasets, namely dataset training and dataset testing, to find out whether the model supports Vector. This machine has accurate predictions to predict the final grades of students at SMAN 2 and SMAN 3 Pangkalpinang; this model is used to determine the scenario that has the most accurate final grade predictions so that conclusions can be drawn to find out which students need to do extra learning in order to graduate in Final exam (Dou et al., 2020).

a. MSE Model Support Value Vector Machine Scenario 1 Science Department of SMAN 2 and SMAN 3 Pangkalpinang

The first scenario where the value to be predicted is the average value from semester 1 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 1 is 41.21334272134476 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage value differences can be seen in Table 13.

Table 13. Scenario 1 Science Value actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
85.93	85.85	0.07	0.08
82.86	82.34	0.511	0.61
84.14	82.60	1.53	1.81
79.36	82.13	-2.77	-3.50
81.00	81.14	-0.14	-0.17
80.57	80.3	0.25	0.32
83.00	81.18	1.81	2.19
79.42	80.10	-0.67	-0.84
82.00	81.15	0.84	1.02
80.93	81.73	-0.80	-0.99

Source: Data processed, 2024

b. MSE Model Support Value Vector Machine Scenario 2 Science Department of SMAN 2 and SMAN 3 Pangkalpinang

The second scenario where the value to be predicted is the average value from the combination of semester 1 and semester 2 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 2 is 3.494705329823478 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage value differences can be seen in Table 14.

Table 14. Scenario 2 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
81.36	82.06	-0.70	-0.86
82.71	81.27	1.43	1.73
82.57	83.80	-1.23	-1.49
80.21	78.69	1.52	1.89
85.79	84.7	1.01	1.18
82.50	82.21	0.28	0.33
82.43	81.25	1.17	1.42
83.00	82.89	0.10	0.12
82.29	81.8 4	0.44	0.53
81.43	80.97	0.45	0.56

Source: Data processed, 2024

c. MSE Model Support Value Vector Machine Scenario 3 Science Department of SMAN 2 and SMAN 3 Pangkalpinang

The third scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, and semester 3 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 3 is 14.182750048882673 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 15

Table 15. Scenario 3 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
82.71	81.02	1.68	2.03
82.21	80.88	1.32	1.61
83.36	82.12	1.23	1.48
81.43	81.36	0.06	0.08
85.86	81.20	4.65	5.42
82.00	78.96	3.03	3.70
84.64	83.70	0.93	1.09
82.00	80.77	1.22	1.49
73.64	80.57	-6.93	-9.41
83.71	81.26	2.44	2.92

Source: Data processed, 2024

d. MSE Model Support Value Vector Machine Scenario 4 Science Department of SMAN 2 and SMAN 3 Pangkalpinang

The fourth scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, semester 3, and semester 4 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 4 is 3.029739424194293 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 16

Table 16. Scenario 4 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
81.21	83.33	-2.12	-2.61
82.21	80.68	1.52	1.85
80.29	80.05	0.23	0.28
81.64	82.74	-1.10	-1.35
81.00	79.85	1.14	1.41
82.50	81.70	0.79	0.96
81.14	81.57	-0.43	-0.53
82.64	81.70	0.93	1.13
86.36	86.81	-0.45	-0.52
85.79	86.45	-0.66	-0.77

Source: Data processed, 2024

e. MSE Model Support Value Vector Machine Scenario 1 Social Sciences Department, SMAN 2 and SMAN 3 Pangkalpinang

The first scenario where the value to be predicted is the average value from semester 1 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 1 is 5.130493760927023 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage value differences can be seen in Table 17.

Table 17. Scenario 1 Social Studies Score actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
75.86	78.19	-2.33	-3.07
81.64	79.19	2.44	2.99
81.21	79.11	2.09	2.58
80.93	79.59	1.33	1.64
83.07	77.48	5.58	6.72
82.07	80.20	1.86	2.27
80.07	80.32	-0.25	-0.32
80.00	79.87	0.12	0.15

79.50	80.54	-1.04	-1.31
82.07	81.63	0.43	0.53

Source: Data processed, 2024

f. MSE Model Support Value Vector Machine Scenario 2 Social Sciences Department, SMAN 2 and SMAN 3 Pangkalpinang

The second scenario where the value to be predicted is the average value from the combination of semester 1 and semester 2 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 2 is 5.1600334873366345 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 18

Table 18. Scenario 2 IPS scores actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
76.36	79.68	-3.32	-4.35
78.21	77.76	0.44	0.57
75.86	78.94	-3.08	-4.0 6
81.79	78.54	3.24	3.96
78.07	79.44	-1,379	-1.76
75.57	78.30	-2,733	-3.61
79.71	77.82	1.88	2.36
81.93	79.9 9	1.93	2.35
76.57	82.28	-5.71	-7.46
83.50	81.40	2,098	2.51

Source: Data processed, 2024

g. MSE Model Support Value Vector Machine Scenario 3 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The third scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, and semester 3 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 3 is 4.220445766997599 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted , difference and percentage values differences can be seen in Table 19

Table 19. Scenario 3 IPS Score actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
78.43	80.16	-1.73	-2.21
79.79	79.11	0.677	0.84
74.71	77.69	-2.98	-4.00
77.79	77.53	0.255	0.32
82.43	82.11	0.31	0.38
78.29	80.10	-1.81	-2.31
84.14	82.93	1.20	1.42
78.79	79.96	-1.17	-1.48

79.50	79.67	-0.17	-0.21
83.07	83.36	-0.29	-0.35

Source: Data processed, 2024

h. MSE Model Support Value Vector Machine Scenario 4 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The fourth scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, semester 3, and semester 4 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 4 is 4.665472066573534 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data, the predicted value is obtained from the results of MSE management, the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 20

Table 20. Scenario 4 IPS Score actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
78.21	78.70	-0.48	-0.62
82.07	81.64	0.42	0.51
78.50	79.81	-1.31	-1.68
76.57	79.09	-2.52	-3.29
83,070	84.36	-1.29	-1.55
76.35	77.70	-1.34	-1.75
78.29	79.90	-1.61	-2.06
78.50	79.49	-0.99	-1.27
77.07	78.50	-1.42	-1.85
76.79	78.93	-2.14	-2.79

Source: Data processed, 2024

3. Neural Network Models

In this research, a Neural Network model was also used to predict student grades from semester 1 to semester 4 by having two datasets, namely the dataset training and dataset testing, to find out whether this Neural Network model has accurate predictions to predict the final grades of students at SMAN 2 and SMAN 3 Pangkalpinang, this model is used to determine the scenario that has the most accurate final grade predictions so that conclusions can be drawn to find out what students should do extra learning in order to pass the Final Exam (Sharifzadeh et al., 2019).

a. MSE Value of Neural Network Model Scenario 1 Science Department at SMAN 2 and SMAN 3 Pangkalpinang

The first scenario where the value to be predicted is the average value from semester 1 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 1 is 24660.441532902347 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted , difference and percentage values differences can be seen in Table 21.

Table 21. Scenario 1 Science Value actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
85.93	106.71	-20.78	-24.19
82.86	93.09	-10.23	-12.35
84.14	61.40	22.73	27.01
79.36	64.17	15.18	19.13
81.00	62.72	18.27	22.55
80.57	71.39	9.17	11.38
83.00	84.78	-1.78	-2.15
79.42	98.70	-19.27	-24.26
82.0	74.01	7.98	9.73
80.93	80.99	-0.06	-0.076

Source: Data processed, 2024

b. MSE Value of Neural Network Model Scenario 2 Science Department at SMAN 2 and SMAN 3 Pangkalpinang

The second scenario where the value to be predicted is the average value from the combination of semester 1 and semester 2 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 2 is 55.90357441058514 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 4.22

Table 22. Scenario 2 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
81.36	75.85	5.50	6.76
82.71	78.36	4.34	5.24
82.57	80.41	2.15	2.60
80.21	74.18	6.03	7.51
85.79	81.13	4.65	5.42
82.50	77.38	5.11	6.19
82.43	72.27	10.15	12.31
83.00	89.92	-6.92	-8.33
82.29	82.84	-0.55	-0.67
81.43	86.50	-5.07	-6.22

Source: Data processed, 2024

c. MSE Value of Neural Network Model Scenario 3 Science Department at SMAN 2 and SMAN 3 Pangkalpinang

The third scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, and semester 3 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 3 is 824.1271515136524 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value,

calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 23

Table 23. Scenario 3 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
82.71	90.32	-7.61	-9.21
82.21	83.09	-0.88	-1.07
83.36	82.87	0.48	0.58
81.43	79.09	2.33	2.87
85.86	78.97	6.88	8.01
82.00	85.53	-3.53	-4.31
84.64	83.17	1.46	1.73
82.00	80.92	1,073	1.30
73.64	86.27	-12.63	-17.15
83.71	78.86	4.84	5.79

Source: Data processed, 2024

d. MSE Value of Neural Network Model Scenario 4 Science Department at SMAN 2 and SMAN 3 Pangkalpinang

The fourth scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, semester 3, and semester 4 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 4 is 27.31309375459551 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 24

Table 24. Scenario 4 Science Values actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
81.21	83.33	-2.12	-2.61
82.21	80.68	1.52	1.85
80.29	80.05	0.23	0.28
81.64	82.74	-1.10	-1.35
81.00	79.85	1.14	1.41
82.50	81.70	0.79	0.96
81.14	81.57	-0.43	-0.53
82.64	81.70	0.93	1.13
86.36	86.81	-0.45	-0.523
85.79	86.45	-0.66	-0.77

Source: Data processed, 2024

e. MSE Value of Neural Network Model Scenario 1 Social Sciences Department at SMAN 2 and SMAN 3 Pangkalpinang

The first scenario where the value to be predicted is the average value from semester 1 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 1 is 277.72135373958025 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data

and testing data; the actual value is obtained from testing data; the predicted value is obtained from the results of MSE management; the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value, Actual, predicted , difference and percentage values differences can be seen in Table 25

Table 25. Scenario 1 Social Studies Score actual, predicted , difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
75.86	58.44	17.41	22.96
81.64	80.15	1.48	1.82
81.21	71.77	9.43	11.61
80.93	66.02	14.90	18.41
83.07	75.11	7.95	9.57
82.07	68.16	13.9 0	16.94
80.07	71.10	8.96	11.19
80.00	80.22	-0.22	-0.27
79.50	90.93	-11.43	-14.3
82.07	70.88	11.18	13.62

Source: Data processed, 2024

f. MSE Value of Neural Network Model Scenario 2 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The second scenario where the value to be predicted is the average value from the combination of semester 1 and semester 2 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 2 is 76.35980830234871 with calculate differences And calculate percentage difference, the difference between the actual and predicted values . The MSE value is obtained from training data and testing data, the actual value is obtained from testing data , the predicted value is obtained from the results of MSE management, the calculated value The difference is obtained from the actual value results with predicted value , calculate percentage The difference is obtained from the actual value with the Difference value . Actual , predicted , difference and percentage values differences can be seen in Table 26

Table 26. Scenario 2 IPS scores actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
76.36	73.20	3.15	4.12
78.21	77.19	1.01	1.29
75.86	82.54	-6.68	-8.81
81.79	72.97	8.81	10.77
78.07	77.01	1.05	1.34
75.57	78.49	-2.92	-3.87
79.71	78.62	1,089	1.36
81.93	79.83	2.09	2.55
76.57	78.39	-1.82	-2.38
83.50	78.10	5.39	6.45

Source: Data processed, 2024

g. MSE Value of Neural Network Model Scenario 3 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The third scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, and semester 3 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 3 is 122.88905701875385 with calculate differences And calculate percentage difference, the difference between the actual and predicted

values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data, the predicted value is obtained from the results of MSE management, the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 27.

Table 27. Scenario 3 IPS Score actual, predicted , difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
78.43	75.01	3.41	4.35
79.79	77.95	1.83	2.30
74.71	81.64	-6.93	-9.27
77.79	76.13	1.65	2.12
82.43	80.75	1.67	2.02
78.29	79.10	-0.81	-1.04
84.14	84.87	-0.73	-0.86
78.79	78.98	-0.19	-0.25
79.50	80.69	-1.19	-1.50
83.07	92.76	-9.69	-11.66

Source: Data processed, 2024

h. MSE Value of Neural Network Model Scenario 4 Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

The fourth scenario where the value to be predicted is the average value from the combination of semester 1, semester 2, semester 3, and semester 4 with testing data, namely the class of 2023. The MSE evaluation value obtained in scenario 4 is 15.625493356140588 with calculate differences And calculate percentage difference, the difference between the actual and predicted values. The MSE value is obtained from training data and testing data; the actual value is obtained from testing data, the predicted value is obtained from the results of MSE management, the calculated value The difference is obtained from the actual value results with predicted value, calculate percentage The difference is obtained from the actual value with the Difference value. Actual, predicted, difference and percentage values differences can be seen in Table 28.

Table 28. Scenario 4 IPS Score actual, predicted, difference, and percentage differences

Actual	Predicted	Difference	Percentage Difference
78.21	78.70	-0.48	-0.62
82.07	81.64	0.42	0.51
78.50	79.81	-1.31	-1.68
76.57	79.09	-2.52	-3.29
83.07	84.36	-1.29	-1.55
76.35	77.70	-1.34	-1.75
78.29	79.90	-1.61	-2.06
78.50	79.49	-0.99	-1.27
77.07	78.50	-1.42	-1.85
76.79	78.93	-2.14	-2.79

Source: Data processed, 2024

Model Evaluation

The evaluation model used is Mean Squared Error (MSE) with objectives For each prediction made by the model (Chicco et al., 2021), MSE measures the squared difference between the predicted value and the actual value; MSE also provides information about how accurate the model is in predicting the data, but can also be used as a tool to improve model performance and understand the characteristics of the data being processed. Once calculated for each prediction, MSE then takes the average of these squared differences to provide an overall picture of the model's performance. The scenarios tested were as follows: Scenario 1, namely

predicting the first semester grades of students at SMAN 2 and SMAN 3 Pangkalpinang; Scenario 2 is predicting the combined 1st and 2nd-semester grades of students at SMAN 2 and SMAN 3 Pangkalpinang; Scenario 3 is predicting the combined 1st, 2nd and 3rd-semester grades of students at SMAN 2 and SMAN 3 Pangkalpinang; Scenario 4 is predicting the combined grades of semesters 1, 2, 3, and 4 of students at SMAN 2 and SMAN 3 Pangkalpinang

The MSE value in each scenario is different because each scenario has different data. In determining the MSE value of the data set, extracting features and targets from the variables is carried out so that the training dataset is also tested using regression with the model being tested, namely the decision model. tree (DT), Support Vector Machine (SVM), and Neural Network (NN). The following are the results of the MSE values from scenarios 1 to 4 for the science and social studies majors in the decision model tree (DT), Support Vector Machine (SVM), and Neural Network (NN)

In scenario 1, the science major for the lowest MSE value is the decision model tree MSE results show that the test results are in accordance with actual data and can be used for test calculations in the future period. The results of scenario 1 of the science majors at SMAN 2 and SMAN 3 Pangkalpinang can be seen in Table 29 below:

Table 29. MSE scores for Scenario 1 Science at SMAN 2 and SMAN 3 Pangkalpinang

Model	MSE value
Decision tree	4.959515007849291
Support Vector Machine	41.21334272134476
Neural Networks	24660.441532902347

Source: Data processed, 2024

In scenario 2, the science major for the lowest MSE value is the Support model Vector Machine, and the MSE results show that the test results are in accordance with the actual data and can be used for test calculations in the future period. The results of scenario 2 science majors at SMAN 2 and SMAN 3 Pangkalpinang can be seen in Table 30 below:

Table 30. MSE scores for Scenario 2 Science at SMAN 2 and SMAN 3 Pangkalpinang

Model	MSE value
Decision tree	5.905944629014399
Support Vector Machine	3.494705329823478
Neural Networks	55.90357441058514

Source: Data processed, 2024

In scenario 3, the science major for the lowest MSE value is the decision model tree, and the MSE results show that the test results are in accordance with the actual data and can be used for test calculations in the future period. The results of scenarios for 3 science majors at SMAN 2 and SMAN 3 Pangkalpinang can be seen in Table 31 below:

Table 31. MSE scores for Scenario 3 Science at SMAN 2 and SMAN 3 Pangkalpinang

Model	MSE value
Decision tree	4.689620471281302
Support Vector Machine	14.182750048882673
Neural Networks	824.1271515136524

Source: Data processed, 2024

In scenario 4, the science major for the lowest MSE value is the Support model Vector Machine, and the MSE results show that the test results are in accordance with the actual data and can be used for test calculations in the future period. The results of scenarios for 4 science majors at SMAN 2 and SMAN 3 Pangkalpinang can be seen in Table 32 below:

Table 32. MSE scores for Scenario 4 Science at SMAN 2 and SMAN 3 Pangkalpinang

Model	MSE value
Decision tree	5.013858037676613
Support Vector Machine	3.029739424194293
Neural Networks	27.31309375459551

Source: Data processed, 2024

Of the four scenarios for the science major at SMAN 2 and SMAN 3 Pangkalpinang with three different models, the MSE value shows that the test results are in accordance with the testing dataset and can be used to predict the student's final score, namely the decision model, tree, and support vector machine.

In scenario 1, the Social Sciences major for the lowest MSE value is the Support model Vector Machine. MSE results show that the test results are in accordance with actual data and can be used for test calculations in future periods. The results of scenario 1 of the Social Sciences Department at SMAN 2 and SMAN 3 Pangkalpinang can be seen in Table 33 below:

Table 33. MSE scores for Scenario 1 IPS SMAN 2 and SMAN 3 Pangkalpinang

Model	MSE value
Decision tree	8.531357587650442
Support Vector Machine	5.130493760927023
Neural Networks	277.72135373958025

Source: Data processed, 2024

In scenario 2 of the Social Sciences major, the lowest MSE value is the Support model Vector Machine, and the MSE results show that the test results are in accordance with the actual data and can be used for test calculations in the future period. The results of scenario 2 of the Social Sciences majors at SMAN 2 and SMAN 3 Pangkalpinang can be seen in Table 34 below:

Table 34. MSE scores for Scenario 2 IPS SMAN 2 and SMAN 3 Pangkalpinang

Model	MSE value
Decision tree	7.767002838827839
Support Vector Machine	5.1600334873366345
Neural Networks	76.35980830234871

Source: Data processed, 2024

In scenario 3 of the Social Sciences major, the lowest MSE value is the Support model Vector Machine , that the MSE results show that the test results are in accordance with the actual data and can be used for test calculations in the future period. The results of scenarios for 3 social studies majors at SMAN 2 and SMAN 3 Pangkalpinang can be seen in Table 35 below:

Table 35. MSE scores for Scenario 3 IPS SMAN 2 and SMAN 3 Pangkalpinang

Model	MSE value
Decision tree	6.318856035234607
Support Vector Machine	4.220445766997599
Neural Networks	122.88905701875385

Source: Data processed, 2024

In scenario 4 of the Social Sciences major, the lowest MSE value is the Support model Vector Machine, that the MSE results show that the test results are in accordance with the actual data and can be used for test calculations in the future period. The results of scenarios for 4 social studies majors at SMAN 2 and SMAN 3 Pangkalpinang can be seen in Table 36 below:

Table 36. MSE scores for Scenario 4 IPS SMAN 2 and SMAN 3 Pangkalpinang

Model	MSE value
Decision tree	5.337855004238116
Support Vector Machine	4.665472066573534
Neural Networks	15.625493356140588

Source: Data processed, 2024

Of the four scenarios for the Social Sciences major at SMAN 2 and SMAN 3 Pangkalpinang with three different models, the MSE value shows that the test results are in accordance with the testing dataset and can be used to predict students' final grades, namely the support model vector machine.

1. Prediction results for the science majors at SMAN 2 and SMAN 3 Pangkalpinang

After getting the MSE value, the data will be trained to get the final prediction. In this process, datasets training Final exam score data and semester 1 to semester 4 scores are combined to determine the final results of the decision model predictions tree (DT), Support Vector Machine (SVM), and Neural Network (NN) in scenarios 1 to scenario 4, science majors at SMAN 2 and SMAN 3 Pangkalpinang.

INITIALS	Rata-Rata	AGM	PPKN	B_IND	MAT_U	SJR_IN	B_ING	SENI	PENJAS	PKR	MAT_P	FIS	BIO	KIM	Semester	Sekolah	y_pred_dt	y_pred_svm	y_pred_nn
AAS	86.820000	85	88	85	89	78	82	90	88	85	90	88	93	86	1	SMAN 2	85.640000	86.400875	86.219490
AS	89.570000	85	93	89	90	80	87	87	92	88	90	92	95	86	1	SMAN 2	86.360000	87.162534	91.455383
AP	83.700000	87	83	85	80	77	74	79	92	84	80	78	74	78	1	SMAN 2	79.570000	79.942578	84.242523
AF	85.840000	85	87	85	84	80	82	80	92	82	75	81	81	83	1	SMAN 2	80.000000	81.649153	78.480461
ANS	87.580000	85	87	85	86	80	84	85	92	85	80	83	90	81	1	SMAN 2	84.570000	83.603597	77.140373
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
RA	84.622857	85	84	81	83	82	90	80	93	83	82	79	85	76	4	SMAN 3	84.570000	82.715473	87.450668
ARF	89.917143	81	78	80	70	71	79	78	73	76	83	71	78	76	1	SMAN 3	73.640000	79.877437	87.500847
ARF	89.917143	82	78	80	80	75	79	83	75	81	75	72	78	70	2	SMAN 3	79.640000	80.420288	75.487770
ARF	89.917143	83	80	80	74	75	80	81	85	79	75	76	78	75	3	SMAN 3	78.000000	79.674212	73.942749
ARF	89.917143	84	80	80	76	75	82	80	82	83	77	75	78	75	4	SMAN 3	78.571429	80.436305	78.029419

Figure 2. Prediction Results for the Science Department of SMAN 2 and SMAN 3 Pangkalpinan

Source: Data processed, 2024

Figure 2 shows the final result of a prediction using the three decision models tree (DT), Support Vector Machine (SVM), and Neural Network (NN). For students with the initials AAS who attend SMAN 2 with an average UAS score compared to student subject scores in each semester, the predicted score from the model is obtained. decision tree is 85.64, Support model Vector Machine is 86.40, and the Neural Network model is 86.22. Likewise, with students with the initials ARF who attend SMAN 3 Pangkalpinang with an average UAS score compared to student scores in each semester, the predicted score from the model is obtained. decision tree is 73.64, the Support model Vector Machine is 79.88, and the Neural Network model is 87.50.

2. Prediction results for the Social Sciences majors at SMAN 2 and SMAN 3 Pangkalpinang

After getting the MSE value, the data will be trained to get the final prediction. In this process, datasets training Final exam score data and semester 1 to semester 4 scores are combined to determine the final results of the decision model predictions tree (DT), Support Vector Machine (SVM), and Neural Network (NN) in scenarios 1 to scenario 4 of the Social Sciences Department at SMAN 2 and SMAN 3 Pangkalpinang.

INITIALS	Rata-Rata	AGM	PPKN	B_IND	MAT_U	SJR_IN	B_ING	SENI	PENJAS	PKR	SEJR	GRO	SOS	EKO	Semester	Sekolah	y_pred_dt	y_pred_svm	y_pred_nn
AYP	79.550000	75	80	75	78	76	76	75	70	78	74	80	76	76	2	SMAN 2	77.790000	78.990726	80.368408
ARR	80.580000	75	80	74	75	73	76	76	85	77	77	77	76	76	2	SMAN 2	78.857143	78.223270	76.180977
CSB	80.740000	80	80	82	75	74	74	76	85	82	78	85	76	76	2	SMAN 2	78.860000	79.230884	90.158905
CW	81.460000	80	85	84	79	78	78	80	85	78	74	75	78	75	2	SMAN 2	80.070000	79.024552	77.139069
DRO	81.140000	75	83	80	78	75	78	76	70	80	81	78	76	76	2	SMAN 2	75.930000	80.012785	83.147644
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
NRN	80.857143	81	75	73	77	75	78	82	76	82	79	81	81	70	4	SMAN 3	78.500000	80.006504	91.027054
NLS	81.000000	82	80	75	70	78	74	79	76	75	78	76	76	78	1	SMAN 3	80.642857	78.580310	75.364365
NLS	81.000000	82	81	75	71	79	79	90	73	83	78	79	79	87	2	SMAN 3	84.710000	79.597442	92.419456
NLS	81.000000	82	73	75	75	80	80	83	80	78	75	84	81	81	3	SMAN 3	79.430000	79.016971	82.732422
NLS	81.000000	83	78	77	76	86	81	86	83	87	82	81	84	82	4	SMAN 3	77.571429	80.870032	93.238365

Figure 3. Prediction Results for the Social Sciences Department of SMAN 2 and SMAN 3 Pangkalpinang

Source: Data processed, 2024

Figure 3 shows the final result of a prediction using the three decision models tree (DT), Support Vector Machine (SVM), and Neural Network (NN). For students with the initials AYP who attend SMAN 2 with an average UAS score compared to student subject scores in each semester, the predicted score from the model is obtained. decision tree is 77.79, the model Support Vector Machine is 78.99, and the Neural Network model is 80.37. Likewise, students with the initials NLS who study at SMAN 3 Pangkalpinang with an average UAS score compared to student subject scores in each semester will get a predicted score from the model. decision tree is 80.64, the Support model Vector Machine is 78.58, and the Neural Network model is 75.36.

Data Visualization & Dashboard Evaluation

In this section, a dashboard will be created that displays the student's initials, school of origin, subject grades, semester, scenarios 1 to 4, and prediction range for each scenario. The application used is Power BI. The results of the predictions that have been made will be submitted to the programmer to create a dashboard. The aim of creating this dashboard is so that the school can easily see the values and become the basis for making decisions regarding the learning strategies that must be implemented.

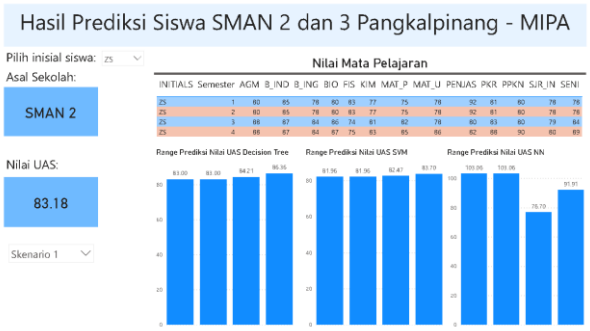

Figure 4 SMAN 2 Pangkalpinang Science department dashboard

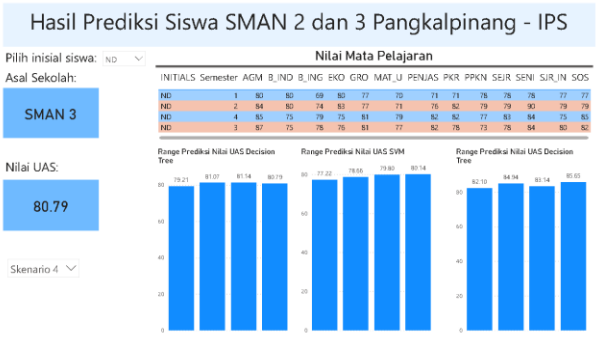

Figure 5 Social Sciences department dashboard at SMAN 3 Pangkalpinang

Figure 4 and Figure 5 are displays of the dashboard that has been created. The dashboard is created with a simple display by using the slicer function to select the student's initials and choose a scenario, a bar chart to display the range of predicted UAS scores for semester 1 to semester 4 from 3 different models, namely Decision tree (DT), Support Vector Machine (SVM), and Neural Network (NN), meanwhile the table is used to view students' subject scores based on their subjects, UAS scores and school origin will automatically appear if you have selected the student's initials and scenario.

Initially, schools can select the desired student's initials in the dropdown in the top left corner of the dashboard, and the school must select the desired scenario. Then, automatically, the

UAS and School Origin scores will appear in the dashboard table, the purpose of using the student's initials for privacy reasons. Predicted student final grades will appear on the bar chart at the bottom of the dashboard along with the numbers in each Decision model tree (DT), Support Vector Machine (SVM), and Neural Network (NN), which will move dynamically following student data and the selected scenario (Aydoğdu, 2020). It is hoped that this dashboard will make it easier for schools to see students' learning results each semester with the aim of students receiving further treatment so they can pass their final school exams.

Conclusion

The results of students' exams and their graduation can be shown from the quality and weaknesses of the students in terms of learning, so efforts are always made to support students who have difficulty learning in educational units/programs. Student achievement, student progress and student potential are very important for measuring learning outcomes and selecting learning materials and learning activities. There are various influencing factors in the success rate of graduating students. One of the main factors is the daily score which has an impact on test scores, but that is not the only thing in predicting student performance in increasing graduation. Other factors also play a role, including behaviour and discipline, which also influence it. Increasing the number of students graduating in an educational unit/program can elevate and increase the popularity of that educational unit/program. If an educational unit/program is declining in its rating, this could have an impact on the attractiveness of prospective new students who will enter that educational unit/program. So, of course, it can cause concerns for educational units/programs. The aim is to find and build a model from a dataset and then use the model in another dataset where you want to predict a result.

The MSE value in each scenario is different because each scenario has different data. In determining the MSE value of the data set, extracting features and targets from the variables is carried out so that the training dataset is also tested using regression with the model being tested, namely the decision model. tree (DT), Support Vector Machine (SVM), and Neural Network (NN). From these results, it can be concluded that: Of the four scenarios for the science major at SMAN 2 and SMAN 3 Pangkalpinang with three different models, the MSE value shows that the test results are in accordance with the testing dataset and can be used to predict the student's final score, namely the decision model, tree, and support vector machine. Of the four scenarios for the Social Sciences major at SMAN 2 and SMAN 3 Pangkalpinang with 3 different models, the MSE value shows that the test results are in accordance with the testing dataset and can be used as a prediction of students' final grades, namely the support model vector machine. The dashboard created using Power BI shows that each scenario has a different predicted range for UAS scores for each model. It is hoped that schools will take action so that students whose scores are still below receive special and intensive treatment. The dashboard also shows the scores for each subject, which students are expected to study more actively in subjects with scores that are still quite low.

.

Bibliography

Akala, D. B. M. (2021). Revisiting education reform in Kenya: A case of Competency Based Curriculum (CBC). *Social Sciences & Humanities Open*, *3*(1), 100107. https://doi.org/10.1016/j.ssaho.2021.100107

Akour, M., & Alenezi, M. (2022). Higher Education Future in the Era of Digital Transformation. *Education Sciences*, *12*(11), 784. https://doi.org/10.3390/educsci12110784

Altabrawee, H., Ali, O. A. J., & Ajmi, S. Q. (2019). Predicting Students' Performance Using Machine Learning Techniques. *Journal of University of Babylon, Pure and Applied Sciences, 27*(1), 194–205.

Aydoğdu, Ş. (2020). Predicting student final performance using artificial neural networks in online learning environments. *Education and Information Technologies*, *25*(3), 1913–1927. https://doi.org/10.1007/s10639-019-10053-x

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Courtiol, P., Maussion, C., Moarii, M., Pronier, E., Pilcer, S., Sefta, M., Manceron, P., Toldo, S., Zaslavskiy, M., Le Stang, N., Girard, N., Elemento, O., Nicholson, A. G., Blay, J.-Y., Galateau-Sallé, F., Wainrib, G., & Clozel, T. (2019). Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nature Medicine*, *25*(10), 1519–1525. https://doi.org/10.1038/s41591-019-0583-3

Dou, J., Yunus, A. P., Bui, D. T., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.-W., Han, Z., & Pham, B. T. (2020). Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides*, *17*(3), 641–658. https://doi.org/10.1007/s10346-019-01286-5

Du, X., Yang, J., Hung, J.-L., & Shelton, B. (2020). Educational data mining: a systematic review of research and emerging trends. *Information Discovery and Delivery*, *48*(4), 225–236. https://doi.org/10.1108/IDD-09-2019-0070

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, G. Van. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, *94*, 335–343. https://doi.org/10.1016/j.jbusres.2018.02.012

Kementerian Pendidikan dan Kebudayaan. (2019). *Peraturan Menteri Pendidikan dan Kebudayaan Nomor 43 Tahun 2019 Tentang Penyelenggaraan Ujian yang Diselenggarakan Satuan Pendidikan dan Ujian Nasional*.

Li, L. (2024). Reskilling and Upskilling the Future-ready Workforce for Industry 4.0 and Beyond. *Information Systems Frontiers*, *26*(5), 1697–1712. https://doi.org/10.1007/s10796-022-10308-y

Mienye, I. D., Sun, Y., & Wang, Z. (2019). Prediction performance of improved decision tree-based algorithms: a review. *Procedia Manufacturing*, *35*, 698–703. https://doi.org/10.1016/j.promfg.2019.06.011

Sharifzadeh, M., Sikinioti-Lock, A., & Shah, N. (2019). Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. *Renewable and Sustainable Energy Reviews*, *108*, 513–538. https://doi.org/10.1016/j.rser.2019.03.040

Wajdi, M. B. N., Iwan Kuswandi, Umar Al Faruq, Zulhijra, Z., Khairudin, K., & Khoiriyah, K. (2020). Education Policy Overcome Coronavirus, A Study of Indonesians. *EDUTEC : Journal of Education And Technology*, *3*(2), 96–106. https://doi.org/10.29062/edu.v3i2.42

Xiao, W., Ji, P., & Hu, J. (2022). A survey on educational data mining methods used for predicting students' performance. *Engineering Reports*, *4*(5). https://doi.org/10.1002/eng2.12482