Comparison and Performance Analysis of SVM and PSO-SVM Algorithms (Case Study Classification of Senior High School)

Basuki Rakhim Setya Permana[1], Andreas Tri Panudju[2]
[1]Information System Department, Bina Bangsa University
[2] Industrial Engineering Departments, Bina Bangsa University

***Abstract:* Attribute Selection is very important for classification process. This research has been done by doing attribute selection using PSO method (Particle Swarm Optimization) on SVM algorithm (Support Vector Machine). The development of the classification model uses three parameters especially data attribute, influence of the transformation of various kernel function and penalty factor (C) toward the performance of SVM and PSO-SVM classification. The analysis uses five kernels in mySVM library that existed in Rapidminer application namely dot, radial, polynomial, neural, and anova kernel. The training data used in the first model classification development is student interest data at ABC high school on 2013-2014 year academic. The first model is evaluated using accuracy, precision, recall, and auc value test. The first result shows that the anova kernel on PSO- SVM is able to work with accuracy level 99.30% using penalty factor 0.1. The second model has been developed to predict student interest in XYZ high school. The second result shows that PSO-SVM with kernel anova is able to classify students interest with 99.29% accuracy level.**

***Keywords*: Optimization, SVM, PSO-SVM, Student Interest.**

I. INTRODUCTION

Guiding students to choose the right majors is very important in any type of learning. Delay in knowing the right direction for students is a loss, both for students and for the institution concerned.

There are already tools to make predictions and classifications with cases in educational institutions, one of which is using the Support Vector Machine method [1]. SVM as a classification method that has a high degree of accuracy in predicting the classification of potential in educational institutions. SVM has the advantage of classifying a pattern accurately despite the limitations of datasets such as the SVM research applied in the scope of the case of educational institutions conducted by Nimit Pattanasri, et al in 2012 [2]. This study aimed to classify the types of presentation slides based on the features of each course using only 102 questionnaire data given to students. However, SVM has limitations when the number of attributes used is relatively large which results in heavy computational burden and accuracy being less accurate.

In previous studies researchers conducted research by applying the SVM algorithm in the case of classifying high school interest pathways, by testing all the kernels contained in mySVM with some C functions (penalty factors) but the results of accuracy, precision, and recall were less than the maximum because they were not using the attribute selection method [3].

One well-known attribute selection method is PSO (particle swarm optimization). PSO is an algorithm used to solve optimization problems. Some of the advantages of PSO are that it is easy to implement and requires only a few parameters, PSO is more efficient in computing, and PSO is more flexible in maintaining a balance between optimal global and local search.

Chung-Jui Tu, et al have conducted research on attribute selection using PSO-Multi Class SVM on several types of datasets. In this study Cheng-Jui Tu, et al [4] have implemented PSO-Multi Class SVM on several types of datasets with many attributes such as vowel dataset, wine, wdbc, ionosphere, and sonar dataset. The results of their study concluded that the use of PSO can be well integrated with SVM Multiclass with a much better level of accuracy compared to SVM without the PSO method. In his research also explained the negative impact is the computing time will be relatively longer than using SVM without PSO [4]. Similar research was also carried out by Fatima Ardjani and Kaddour Sadouni, they used PSO-SVM to optimize multi-class SVM [5].

This research uses a hybrid approach that combines the support vector machine (SVM) classifier with the particle swarm optimization (PSO) method with a total dataset of 281 records and 11 attributes. The purpose and benefit of this research is to do the attribute selection process using PSO on the dataset used so that it will improve the accuracy of the SVM classification model.

II. METHODOLOGY

The methodology and the research flow carried out until the formation of the classification model are listed as in the following figure:

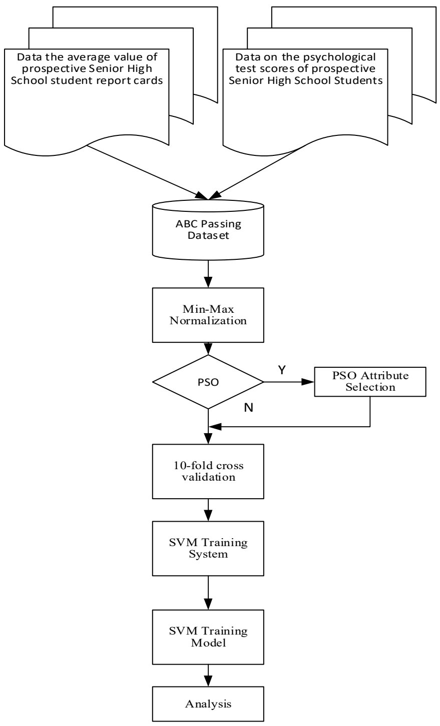

Figure 1. Research Methodology Scheme

The research methodology in Figure 1 starts from the collection of datasets that will be used in the selection attribute, training process and the formation of classification models. The material used is taken from the students' grades when registering at ABC High School in the 2018-209 academic year which includes the name, UN scores in the previous level, average report card grades, and the scores of specialization psychological tests totaling 281 students. Students who are labeled majoring in Natural Sciences are 64 students and students who are labeled majoring in Social Sciences are 216 students.

The dataset that was collected was then made into a single unit called the ABC Majors dataset which was then continued by pre-processing the data in the form of scaling method. Researchers use Min-Max Normalization with a minimum value of -1 and a maximum value of 1 because according to researchers will show good results compared to other scaling methods based on studies of other research libraries [6]. Data that has passed the pre-processing process are then carried out two different treatments to determine the comparison of the performance of the performance models of SVM and PSO-SVM. The model built with SVM will then be directly carried out 10-fold cross validation while the model built with PSO-SVM will be weighted and selected attributes using particle swarm optimization (PSO) techniques. Then the data attributes that have been selected will then be divided into two parts namely training data and testing data using 10-fold cross validation to test the performance of the SVM model to produce the most optimal training model. To test the best model, it is done by comparing the performance or performance models of SVM and PSO-SVM. The testing criteria include accuracy test, precision test, recall test and AUC (area under curve) test. The best model that has been formed will then be applied (testing) into the XYZ SMA dataset which is different from the ABC SMA training data used to produce the model. The XYZ High School dataset is 288 students with 150 Natural Sciences specialization labels, and 138 Social Sciences students. The model that has been tested into the XYZ SMA Dataset will then be analyzed and evaluated the results by matching the labels that have been formed at the time of data classification with the original labels given by the XYZ SMA. The following is an overall picture of the process of forming the model to the process of data classification as in Figure 2.

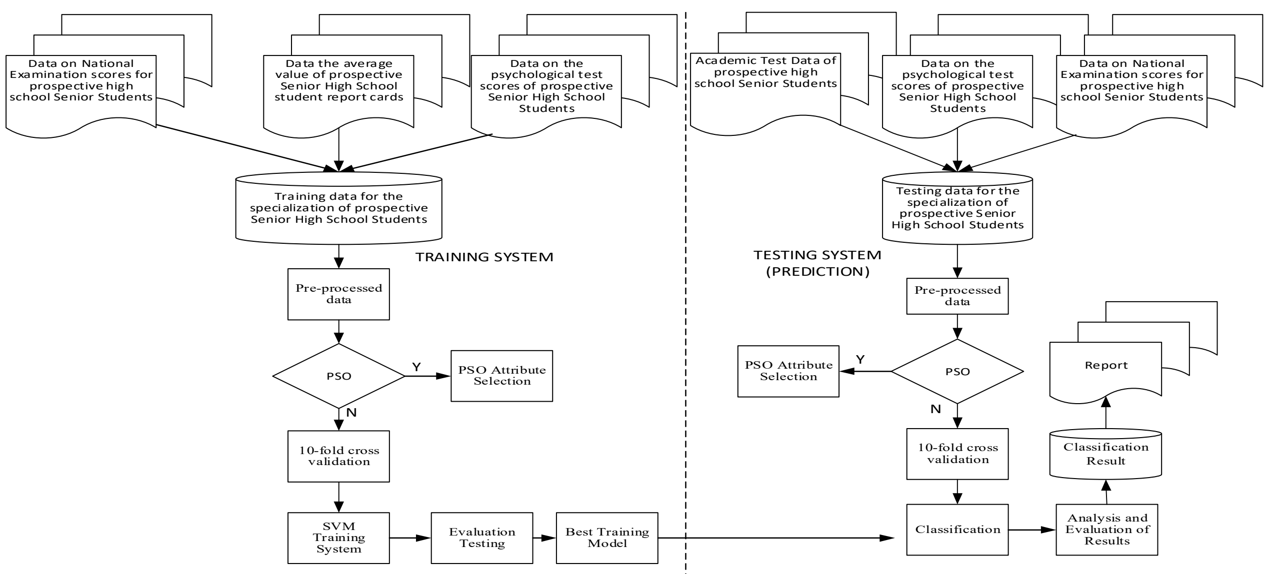

Figure 2. Model Formation Process up to Data Classification

A. Data Mining

Data mining is the process of extracting (extracting) patterns from a set of data that has the potential to obtain the information needed [11]. According to Han and Kamber data mining is developing because of a very basic need that is the presence of large amounts of data that can actually be used to produce useful information. Han and Kamber stated that data mining has seven steps or steps that are generally commonly applied, namely [11].

1. Data cleaning is the process of removing a piece of data that is inconsistent or containing noise.
2. Data integration is the process of combining or integrating data sources from various data sources.
3. Data selection is the process of selecting or taking data according to need.
4. Data transformation is the process of converting data into a form in accordance with the objectives for decision making.
5. Data mining is a process of processing data using a particular method or algorithm to produce a data pattern.
6. Pattern evaluation of data is a process to test the truth of data patterns that represent the knowledge that is in the data itself.
7. Knowledge representation is the process of presenting knowledge to display the results of processing into information to the user.

B. Data Classification

Data classification is a process of separating and grouping a set of data with other data sets [11]. According to Han and Kamber the classification of data has two stages of the process. First is to build a model or pattern based on a series of data classes which will then be classified. The first stage is also often called the training stage or the learning process, the process of building this model by analyzing existing training data [11].

In general, the process of building this model is translated into classification rules, decision trees, or other mathematical models. Next in the second stage, which is the classification process using testing data, the predetermined model will be used to predict the data that the label is not yet known [11].

C. SVM Algorithm

Support Vector Machine is a method or algorithm for classification and prediction [7]. The working principle of this method is to find the most optimal separation space of a dataset in different classes. In everyday life, we are often faced with problems that are not linear / data that cannot really be separated linearly, namely a condition where there is no line or plane that can be made to be a separator between classes of data. In this problem there are 2 steps that can be done, namely:

1) Using hyperplane soft margin.

The purpose of hyperplane soft margins is to convert non-linear data into a linear form while maintaining a flexible boundary plane.

Formulation on hyperplane soft margin that uses slack variables (formulated with Equation (1) [6].

$$\begin{aligned} x_i.w + b &\geq 1 - \xi \ \mathit{untuk}\ y_i = \mathit{kelas}\ 1 \\ x_i.w + b &\leq -1 + \xi \ \mathit{untuk}\ y_i = \mathit{kelas}\ 2 \end{aligned} \quad (1)$$

With Equation (1), the search for the best separator field can be formulated into Equation) (2) [6].

$$\min \frac{1}{2}\|w\|^2 + C\ (\textstyle\sum_{i=1}^{n} . \xi_i)$$
$$\text{Dengan fungsi pemisah} \quad y_i(x_i.w + b) \geq 1 - \xi\ \forall i = 1, \ldots, n \quad (2)$$

C is the parameter that determines the amount of penalty due to data classification errors.

2) Finding a linear separating hyperplane in the new dimension space (feature space).

Changing the input space (dot product) into the form of feature space is often known as the kernel trick technique which then develops into the Kernel function

$$\Phi(\bar{x}_i).\Phi(\bar{x}_j)$$
$$K\ (x_i.x_j \quad [6].$$

Changes from input space to feature space result in very large computations, because there is a possibility that the dimensions of feature space are very numerous and even infinite. Therefore SVM bridge it with the Kernel function.

In this research, five kernels are used, including kernel dot, radial, polynomial, neural, and ANOVA (analysis of variables).

D. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a swarm intelligence algorithm that is the study of computational systems inspired by collective intelligence [8]. Collective intelligence arises based on population or homogeneous cooperation in an environment. In PSO, a flock is assumed to have a certain size with each particle having a random initial position at a location in one space. Each particle is assumed to have two characters: position and speed [9]. Each particle moves in a certain space and remembers the best position that has ever been traversed or found against a food source. Each particle conveys its best information or position to other particles. In the PSO algorithm, the search for solutions is carried out by a population consisting of several particles.

The population is generated randomly with the smallest and largest limit values (lower and upper limits). Each particle seeks a solution by crossing the search space by making adjustments to its best position (local best) and adjusting to the best particle position of the whole herd (global best) while crossing the search space. A number of iterations are performed to find the best position of each particle until a relatively fixed position is reached or it reaches a specified iteration limit. In each iteration, each solution (particle position) is evaluated by inserting the solution into the fitness function.

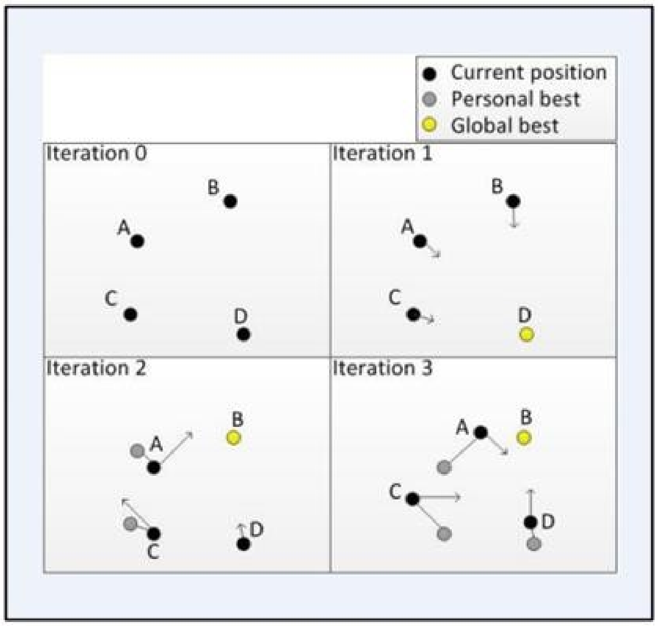

Figure 3. Particle Process Looking for the Best Position

PSO is proven to be a good and effective algorithm for optimization problems because of its ease in implementing code and its consistent performance [10].

E. Min-Max Normalization

Min-Max normalization is a form of data normalization scaling that is used to prevent the dominance of attributes that have a large range of values against attributes with small reach values, in addition min-normalization can also prevent numerical problems during calculations with Equation (3) [11] .

$$D'(i) = \frac{D(i) - \min(D)}{\max(D) - \min(D)}(U - L) + L \quad (3)$$

D '(i) is the value of data i from attribute D that has been normalized, D (i) is the original data value i, U and L are the upper limit and lower limit of normalization. Min (D) is the minimum value of a data attribute D. Max (D) is the maximum value of an attribute D [6]. The use of the Min-Max normalization method is based on a comparative study of input data normalization methods for Support vector machines that have been conducted by Ali and Smith [12] where the results show that the min-max normalization method provides a better level of accuracy and performance compared to normalization which uses zero mean and log scaling methods.

F. K-fold Cross Validation

K-fold cross validation is a technique to estimate the performance of the training model that has been built [11]. This method divides training data and testing data as much as k parts of data. The function of k-fold cross validation is so that there is no overlapping of the testing data. Following is a simple illustration of k-fold cross validation shown in Figure 4.

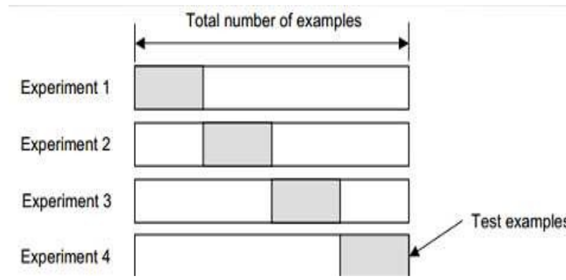

Figure 4. Illustration of k-fold cross validation

In the illustration shown in Figure 4, it can be explained that the experiment uses 4-fold cross validation. This is stated in the number of experiments conducted The gray box is a test set and the rest (the white box) is the training set. For example, there are 40 instances of data in the illustration in Figure 4, then the first gray box experiment from instances 1 to instances in the 10th order which amounts to 10 instances is a test set and the remaining instances are 11th to 30th order which amounts to 30 instances (white boxes) are training sets. From the first experiment we got the average error value. Then proceed the same way for the second experiment as in the first experiment to the fourth experiment. After that the average estimated error of each experiment has been calculated to the end.

III. TESTING PERFORMANCE MODELS

Model performance testing in this study was conducted by comparing the SVM kernel and PSO-SVM in the mySVM library contained in the Rapidminer application to obtain a model with the highest performance. The parameters used in evaluating this kernel comparison are accuracy, precision, recall, ROC curve (AUC), and ANOVA statistical test [6]. The following will discuss the evaluation parameters that will be used as a test of the performance of the SVM model.

1) Accuracy, precision, and recall

Accuracy can be defined as the level of closeness between the predicted value and the actual value. Precision shows the level of accuracy or accuracy in classification. Whereas recall serves to measure the actual positive proportions that are correctly identified. To measure accuracy, precision, and recall, confusion matrix is usually used. Confusion matrix is a matrix measuring tool used to get the amount of class classification accuracy with the algorithm used. The following will be presented in the form of confusion matrix in Table 1.

Table 1 Form Matrix Confusion of Two Classes

Confusion Matrix		Value	
		TRUE	FALSE
Prediction Value	TRUE	TP (True Positive) Correct result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

In Table I the values of TP (true positive) and TN (true negative) indicate the level of classification accuracy. Generally the higher the TP and TN values the better the classification level of accuracy, precision, and recall. If the predicted output label is true (true) and the true value is false (false) is called false positive (FP). Whereas if the predicted output label is false (false) and the true value is true (true) then this is referred to as false negative (FN) [11]. The following formulations for calculating accuracy, precision, and recall in the formation of classification models are shown in Equation (4), Equation (5), and Equation (6) [11].

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} x100\% \quad (4)$$

$$Recall = \frac{TP}{TP + FN} x100\% \quad (5) \qquad Presisi = \frac{TP}{TP + FP} x100\% \quad (6)$$

A. ROC curve

The ROC (receiver operating characteristic) curve is one measure to assess the ability of a classification system. This research will use the ROC curve measurement tool to compare SVM kernels with PSO-SVM in the mySVM library contained in the Rapidminer program. The ROC curve was first implemented in signal detection theory. Then developed and used in medicine, radiology, and other fields. ROC curves are often used to evaluate classifications because they have the ability to evaluate algorithms quite well [6].

The ROC curve is a comparison graph between sensitivity (true positive rate / (TPR)) which is translated into a vertical axis or y-axis coordinate with specificity (false positive rate (FPR)) which is translated in the form of a curve. The following formulations of sensitivity and specificity are presented in Equation (7), and Equation (8) [6].

$$Sensitivity = \frac{TP}{TP + FN} x100\% \quad (7)$$

$$Specificity = \frac{FP}{FP + TN} x100\% \quad (8)$$

The ROC curve can be used as a comparison of several methods (classifier) or classifier models that have different parameters to get the best model. Following is an example of applying the performance comparison of two different classifiers in Figure 5.

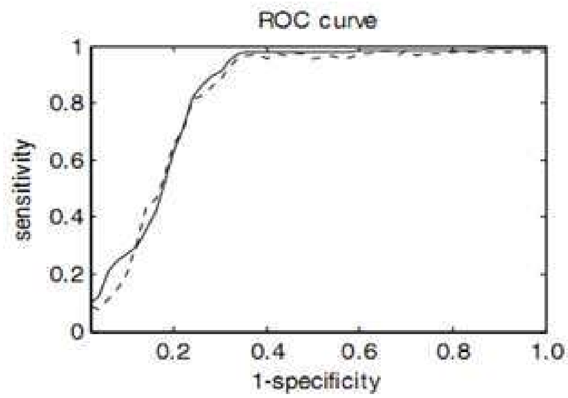

Figure 5 Comparison of Classifiers with ROC Curves

In Figure 5 it can be seen that there are two classifiers symbolized by dashed lines and solid lines. If in Figure 5 shows the location of coordinates (0, 1) it represents sensitivity and specificity of 100%. To calculate and ascertain which classifier is superior then AUC (area under curve) calculation is used.

AUC (area under curve) is the area under the curve. The area of the AUC is always between 0 and 1. The AUC is calculated based on the combined area of the trapezoidal points (sensitivity and specificity). In Figure 5 shows that the solid line has an area under a larger curve than the broken line, this means that the level of performance of the classification of the classifier which is represented by a solid line is better than the level of performance of the classification of the classifier which is represented by the broken line break.

Following are the standard classification class tables based on the AUC values in Table 2

Table 2. Category of Classification based on AUC value

AUC Value	Classification Category
0.90 - 1.00	Excellent
0.80 - 0.90	Good
0.70 - 0.80	Fair
0.60 - 0.70	Poor
0.50 - 0.60	Fail

IV. RESULTS AND DISCUSSION

In this study, training data are used to form a classification model that is the value of students when registering at one of the high schools in Central Java in the 2013-2014 academic year which includes the name, UN scores at the previous level, average report card grades, and psychological test scores in specialization 280 students.

The purpose of this study is to compare and analyze the performance of SVM models that do not use PSO (SVM) and SVM models based on PSO (PSO-SVM) applied to the five SVM kernel functions in the mySVM library contained in the Rapidminer application including kernel dot, radial , polynomial, neural, and anova. Each experiment with various SVM models is then evaluated using an accuracy, precision, recall evaluator to get the best model.

In Table 3 and Table 4 shows the comparison of the accuracy of the classification of data using ordinary SVM and PSO-SVM.

Table 3. Accuracy Level of Svm Classification

Par C	Tipe Kernel				
	Dot	Radial	Poli-nomial	Neu-ral	Anova
0.0	84.72	80.42	80.09	66.23	95.37
0.1	84.72	76.87	86.13	73.63	97.87
0.2	85.78	76.87	82.56	69.33	96.08
0.4	84.70	77.23	83.29	68.62	93.23
0.8	86.12	78.65	80.80	67.56	96.08
1.0	86.83	80.42	85.09	66.23	95.73
1.2	87.19	81.83	82,59	66.95	97.16
1.4	87.19	82.19	84.72	65.17	97.16
1.8	87.19	81.85	84.73	66.96	97.16

Table 4 PSO-SVM Accuracy Level

Par C	Tipe Kernel				
	Dot	Radial	Poli-nomial	Neu- ral	Anova
0.0	87.94	93.25	81.86	76.17	98.60
0.1	87.56	85.03	91.10	80.10	99.30
0.2	87.94	82.93	92.21	80.10	98.60
0.4	87.23	93.25	83.80	80.10	97.88
0.8	87.57	95.04	84.20	77.97	98.30
1.0	87.92	93.25	86.21	79.03	98.60
1.2	87.23	93.60	81.86	77.64	98.50
1.4	87.56	92.55	90.10	77.98	98.30
1.8	87.92	92.91	83.80	78.29	98.60

Judging from Table 3 and Table 4 we can compare the accuracy between SVM and PSO-SVM and conclude that optimization by selecting the appropriate attributes (parameters) using PSO (particle swarm optimization) can increase the accuracy of SVM models on all types of kernels tested with several penalty factors from 0.0,0.1,0.2,0.4,0.8,1.0,1.2,1.4, and 1.8.

The greatest accuracy level of 99.30% is obtained when implementing PSO-SVM with a C (penalty) parameter of 0.0 using anova kernel.

The second test is to use a precision evaluator. In Table V and Table VI shows the comparison of the level of precision of classification of data using ordinary SVM with PSO-SVM.

Table 5 SVM Classification Precision Level

Par C	Tipe Kernel				
	Dot	Radial	Poli-nomial	Neural	Anova
0.0	88.03	79.80	81.64	83.56	95.05
0.1	88.03	76.87	89.29	87.47	98.65
0.2	89.81	76.87	87.12	84.11	96.73
0.4	89.96	77.15	89.16	88.10	97.56
0.8	90.38	78.32	85.75	82.57	97.55
1.0	91.12	79.80	89.42	83.60	98.65
1.2	90.62	81.01	88.11	82.80	98.65
1.4	90.88	81.33	90.47	84.93	98.65
1.8	90.88	81.50	92.07	82.96	98.65

Table 6 PSO-SVM Classification Precision Level

Par C	Tipe Kernel				
	Dot	Radial	Poli nomial	Neural	Anova
0.0	90.71	92.04	82.52	84.47	98.86
0.1	88.19	76.87	92.50	86.03	99.57
0.2	90.71	83.16	94.12	90.70	98.66
0.4	89.89	95.55	83.20	88.49	98.18
0.8	89.20	97.76	84.40	85.27	98.40
1.0	90.30	92.04	85.12	84.78	99.57
1.2	91.00	97.71	93.12	86.60	98.86
1.4	89.87	93.11	95.16	84.73	98.23
1.8	91.06	94.74	96.12	84.47	98.86

Judging from Table 5 and Table 6 we can compare the precision values between SVM and PSO-SVM. The conclusion is drawn that optimization by selecting the right attributes (parameters) using PSO (particle swarm optimization) can broadly increase the precision value of the SVM model in almost all types of kernels tested.

The greatest precision level of 98.86% is obtained when implementing PSO-SVM with a C (penalty) parameter of 0.0 using anova kernel.

The third test is to use the recall evaluator. Table 7 and Table 8bshow a comparison of the recall rates of classifying data using the usual SVM with PSO-SVM.Table 7.

Table 7. Level Recall Classification SVM

Par C	Tipe Kernel				
	Dot	Radial	Poli-nomial	Neural	Anova
0.0	93.55	100.00	95.80	72.27	99.52
0.1	93.55	100.00	93.48	80.04	98.59
0.2	92.62	100.00	90.65	77.34	98.59
0.4	90.78	100.00	89.24	71.45	93.38
0.8	92.14	100.00	90.28	75.58	97,66
1.0	92.14	100.00	91.65	72.27	97.66
1.2	88.33	100.00	89.33	74.68	97.66
1.4	93.05	100.00	89.26	70.13	97.66
1.8	93.05	99.07	87.49	74.20	97.66

Table 8. Level of Recall PSO-SVM

Par C	Tipe Kernel				
	Dot	Radial	Poli-nomial	Neu-ral	Anova
0.0	94.03	100.00	92.27	84.46	99.55
0.1	96.80	100.00	96.32	89.37	99.55
0.2	94.03	84.00	97.32	83.96	99.55
0.4	94.05	95.84	90.42	85.69	99.07
0.8	95.41	95.84	80.69	87.06	99.07
1.0	94.48	100.00	84.54	89.35	99.07
1.2	92.64	94.03	92.45	84.04	99.55
1.4	94.48	97.71	95.67	87.47	99.55
1.8	91.06	93.62	92.32	88.44	99.55

Seen from Table VII and Table VIII can be seen and compared the recall value between SVM and PSO-SVM. The conclusion is drawn that optimization by selecting the appropriate attributes (parameters) using PSO (particle swarm optimization) can broadly increase the recall value of the SVM model in almost all types of kernels tested.

The greatest recall level of 99.55% was obtained when implementing PSO-SVM with the parameter C (penalty) of 0.0,0.1,0.2,1.2,1.4 and 1.8 using the ANOVA kernel.

The next test is to measure the value of AUC (area under curve) on the ROC curve. Table 9 and Table 10 show the comparison of AUC level of data classification using SVM compared to PSO-SVM.

Table 9. AREA UNDER CURVE Value of SVM

Par C	Tipe Kernel				
	Dot	Radial	Poli-nomial	Neu- ral	Anova
0.0	0.926	0.969	0.835	0.729	0.991
0.1	0.935	0.894	0.911	0.892	0.995
0.2	0.928	0.935	0.936	0.880	0.996
0.4	0.925	0.952	0.836	0.879	0.994
0.8	0.930	0.952	0.850	0.776	0.995
1.0	0.935	0.969	0.837	0.749	0.996
1.2	0.923	0.948	0.912	0.777	0.994
1.4	0.936	0.947	0.946	0.826	0.994
1.8	0.937	0.946	0.835	0.724	0.995

Table 10. AREA UNDER CURVE Value of PSO-SVM

Par C	Tipe Kernel				
	Dot	Radial	Poli- nomial	Neu- ral	Anova
0.0	0.938	0.916	0.755	0.822	0.987
0.1	0.938	0.915	0.817	0.799	0.994
0.2	0.935	0.915	0.830	0.755	0.996
0.4	0.922	0.915	0.718	0.756	0.983
0.8	0.934	0.916	0.718	0.734	0.996
1.0	0.930	0.916	0.815	0.757	0.995
1.2	0.936	0.916	0.806	0.751	0.994
1.4	0.936	0.916	0.843	0.752	0.985
1.8	0.937	0.914	0.838	0.756	0.994

Seen from Table IX and Table X can be seen and compared the value of the AUC (area under curve) between SVM and PSO-SVM. The conclusion is drawn that optimization by selecting the appropriate attributes (parameters) using PSO (particle swarm optimization) method can broadly increase the AUC (area under curve) value of the SVM model in several types of kernels tested such as radial kernels.

The highest AUC (area under curve) level of 0.996% is obtained when implementing PSO-SVM with a C (penalty) parameter of 0.2 using anova kernel.

Based on the SVM and PSO-SVM testing tables based on accuracy, precision, recall, and AUC values in Table III, Table IV, Table V, Table VI, Table VII, Table VIII, Table IX, and Table X above, conclusions can be drawn. that PSO-SVM is relatively superior compared to SVM with the performance of the ANOVA kernel as the kernel that provides the greatest level of accuracy, precision, recall, and AUC value compared to other kernels.

The next step of this research is to test the best PSO-SVM classification model into different datasets. In this case the XYZ SMA dataset was used with a total data of 288 students. The model used as a predictive reference is the PSO-SVM model with anova kernel with a penalty factor (C) parameter of 0.1.

The results of the model testing are then analyzed by matching the label prediction with the actual specialization label so that the accuracy, precision and recall performance are obtained using a confusion matrix.

The following is Table 11 from the confusion matrix results from the PSO-SVM model anova kernel parameter 0.1 which is applied (implementation) to the specialization data of XYZ high school students.

TABLE 11. Confusion Matrix From Kernel Anova Parameter C 0.1 PSO-SVM Models On Senior High School Testing Data

PSO-SVM KERNEL ANOVA (C=0.1)	true IPA	true IPS	class precision
pred. IPA	150	2	98.23%
pred. IPS	0	136	100.00%
class recall	100.00%	98.82%	Accuracy : 99.29%

V. CONCLUSIONS AND SUGGESTIONS

From the results of the research and discussion conducted, several conclusions can be drawn as follows:

1) The study was conducted by comparing the SVM algorithm with PSO-SVM.
2) To conduct an experiment or training process the dataset used for specialization is ABC High School Students involving 280 students.
3) Comparison of algorithms is done by using several kernels, namely dot, radial, polynomial, neural, and anova kernel.
4) Optimization performed using the PSO (particle swarm optimization) algorithm is proven to increase the value of accuracy, precision, recall, AUC of the SVM model that was built. The highest performance value obtained by PSO-SVM is in anova kernel testing with an accuracy value of 9.30%. With an accuracy value of more than 70% it can be said that the construction of the classification model using PSO-SVM was successful.
5) The model that has been formed is used for the prediction / classification process of specialization data in XYZ high school with 288 students.
6) Prediction performance implemented gets an accuracy of 99.29%. With prediction accuracy of more than 70%, the PSO-SVM model is suitable for other school datasets.

Some suggestions for the future to be improved (improve) from this study are:

1) The results of testing the appropriate kernel selection can be replaced with automatic kernel selection so it does not require much time for the selection of the right kernel.
2) The results of testing the selection of penalty factor variations (C) can be replaced by the parameter selection method so that it does not require much time for the selection of penalty factors (C) right.

REFERENCE

[1]. F. Jui-Hsi, C. Jui-Hung, H. Yueh-Min, and C. Han-Chieh, "A Support Vector Regression based Prediction of Student's School Performance", Computer, Consumer and Control (IS3C), International Symposium, pp.84-87, 2012.

[2]. P. Nimit, M. Masayuki, M. Michihiko, "Learning to Estimate Slide Comprehension in Classroom with Support Vector Machines". IEEE Computer Society, vol. 5, pp.52-61, March. 2012.

[3]. B. Theopilus, "Klasifikasi Jalur Minat Siswa Menggunakan Algoritme Support Vector Machine (SVM) (Kasus: SMA Negeri 1 dan SMA Negeri 2 Sragen)", Tesis, Jurusan Teknik Elektro dan Teknologi Informasi, UGM, Yogyakarta, 2014.

[4]. J.T. Chung, Y.C. Li, C.Y. Jun, and Y. Cheng-Hong, "Feature Selection using PSO-SVM", IAENG International Journal of Computer Science, IJCS, February, 2007.

[5]. A. Fatima and S. Kaddour, "Optimization of SVM Multiclass by Particle Swarm (PSO-SVM)". Modern Education and Computer Science (MECS), vol. 2, pp.32-38, 2010.

[6]. I.H. Witten, E. Frank, and M.A. Hall, "Data Mining Practical Machine Learning Tools and Techniques", Third Edition, Elsevier Publisher, USA, 2011.

[7]. J. Agustinus, "Sistem Deteksi Intrusi Jaringan dengan Metode Support Vector Machine", M. Eng, Thesis, Jurusan Ilmu Komputer FMIPA UGM, Yogyakarta, 2012.

[8]. J. Brownlee, "Clever Algorithms," 2013. [Online]. Available: http://www.cleveralgorithms.com/nature-inspired/swarm.html. [Accessed 28 Juni 2020].

[9]. B. Santosa, "Tutorial Particle Swarm Optimization," [Online]. Available: http://oc.its.ac.id/ambilfile.php?i dp=1221. [Accessed 28 Juni 2020].

[10]. N. Dhanasaputra and B. Santosa, "Pengembangan Algoritma Cat Swarm Optimization (CSO) untuk Klasifikasi". Jurnal ITS, p. 3, 2010.

[11]. J. Han, and M. Kamber, "Data mining: Concepts and Techniques", Third Edition, Morgan Kaufmann Publishers, San Fransisco, 2013.

[12]. S. Ali, and A. Smith, "Improved Support Vector Machine Generalization Using Normalized Input Space".AI 2010: Advances in Artificial Intelligence, vol.4304, pp.362-371, 2010.