BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1803–1814
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1803-1814

# DESIGN OF KIP KULIAH SELECTION SYSTEM AND RECIPIENT DETERMINATION USING SUPPORT VECTOR MACHINE (SVM)

**Mozart Winston Talakua[1*], Berny Pebo Tomasouw[2], Venn Y. I. Ilwaru[3]**

[1,2,3]*Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Pattimura, Ir. M. Putuhena Street, Poka, Ambon, Maluku,97233, Indonesia*

*Corresponding author's e-mail: * ocat1615@yahoo.com*

***Article History:***

*Received: 01st June 2023*
*Revised: 21st August 2023*
*Accepted: 29th August 2023*

***Keywords:***

*App;*
*KIP Kuliah;*
*Selection;*
*System;*
*SVM.*

*ABSTRACT*

*KIP Kuliah is tuition assistance from the government for high school graduates or equivalent with good academic potential but has economic limitations. In recent years it has been seen that the Indonesian government has always tried to increase the quota for KIP Kuliah recipients. In this study, the Support Vector Machine (SVM) method was applied to create a system for selecting and determining KIP Kuliah recipients. To obtain the best model to be used in the system, the training and testing data are divided into three data distribution schemes, namely 60/40, 70/30, and 80/20. After the training and testing process was carried out using the SVM method with various parameter variations, then the best accuracy rate of 94.59% is obtained in the 80/20 data sharing scheme for the nonlinear SVM model with the RBF kernel. With this system, it is hoped that the KIP Kuliah selection process at the tertiary level can run effectively, efficiently and the results of the determination are more targeted.*



***How to cite this article:***

M. W. Talakua and B. P. Tomasouw., "DESIGN OF KIP KULIAH SELECTION SYSTEM AND RECIPIENT DETERMINATION USING SUPPORT VECTOR MACHINE (SVM)," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1803-1814, September, 2023.


Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

**Research Article • Open Access**

## 1. INTRODUCTION

The national education system has a strategic role in educating the nation's life and advancing science and technology. Every child, regardless of their economic background, must have the same rights and opportunities in obtaining education so that efforts to develop Indonesian human resources must be fair, quality, inclusive and equitable. The Indonesian government will continue to intervene in education through various affirmation programs needed to increase access to higher education for all Indonesian children. Based on Law No. 12/2012 concerning Higher Education, the Government of Indonesia is obliged to improve access and opportunities to study at tertiary institutions and prepare intelligent and competitive Indonesians. Therefore the Government will always try to ensure that underprivileged Indonesian children, especially those with achievements, will be able to continue their education up to the university level through Program Indonesia Pintar (PIP) [1].

PIP is an assistance in the form of cash, expanding access, and learning opportunities from the government which are given to students and students who come from poor or vulnerable families to finance education. This is the basis of the government's commitment to placing access to higher education for the whole community as one of the development priorities. Higher education PIP for college students is given in the form of Kartu Indonesia Pintar Kuliah or KIP Kuliah. In recent years it has been seen that the Indonesian government has always tried to increase the quota for KIP Kuliah recipients. In 2018, the quota amount is 90,000 [2] and continued to increase until 2022 to 200,000 [1]. In accordance with the technical guidelines for managing KIP Kuliah, tertiary institutions need to form a team of verifiers who are required to verify the eligibility of prospective Bidikmisi recipient students who have been declared accepted at tertiary institutions. The verification aspect consists of economic disability, academic potential, region of origin, evaluation of supporting documents, and other special considerations [2]. After going through this stage, tertiary institutions can determine KIP Kuliah recipients through a decree and the determination data is entered into the KIP Kuliah system.

On the other hand, universities as KIP Kuliah managers are still evaluating files, processing data, and determining scholarship recipients manually. This can have an impact on the time it takes longer if more and more participants register, it is less effective and the determination results can tend to be objective. Therefore, it is necessary to create a system that can make the right decisions regarding the determination of KIP Kuliah recipients based on the data held by applicants. Previous research conducted to create a system for selecting and determining KIP Kuliah recipients can be seen in ( [3] [4] [5] [6]) with the methods used including Fuzzy C-Means Clustering, Binary Logistic Regression, Fuzzy Multiple Attribute Decision Making, and Analytical Hierarchy Process (AHP). In this study, the method proposed to be used in system design is the Support Vector Machine (SVM) method, which will classify a person as a KIP recipient or not.

SVM itself has been successfully applied to real-world problems and provides better solutions for certain cases than conventional methods such as artificial neural networks. In [7] it is explained that the basic concept of SVM stems from the problem of classifying two classes as linear problems. SVM tries to find the best hyperplane (separator) in the input space to separate the two classes. Furthermore, SVM was developed to work on non-linear problems by incorporating the kernel concept in high-dimensional workspaces. In finding the best hyperplane, SVM will be faced with finding solutions to quadratic programming problems. Studies ( [8] [9] [10] [11] [12] [13] [14] [15] [16]) show that SVM has better results in classification problems.

## 2. RESEARCH METHODS

The data used in this study were KIP Kuliah registrant data obtained from the Student Affairs Section of the Faculty of Mathematics and Natural Sciences. This data consists of applicants who have passed or failed in 2019-2022. Based on the data obtained, the number of input variables used is 18 variables, with the values of all variables set between 0 to 1. In this study, it is assumed that the higher the value of the variable, the higher the KIP Kuliah requirements, and vice versa, the lower the value. a variable then does not meet the requirements to obtain KIP Kuliah. Table 1 below shows in detail all the variables used, along with their respective values.

**Table 1. Determination of Input Variables Values**

| Input Variable | | | Value |
|---|---|---|---|
| $X_1$ | KIP Ownership | 1. Yes | 1 |
| | | 2. No | 0 |
| $X_2$ | Father's Occupation | 1. Unemployed | 1 |
| | | 2. Farmer/Fisher | 0,75 |
| | | 3. Self-Employed | 0,5 |
| | | 4. Private Employee | 0,25 |
| | | 5. Civil Servant/Military/Police | 0 |
| $X_3$ | Mother's Occupation | 1. Unemployed | 1 |
| | | 2. Farmer/Fisher | 0,75 |
| | | 3. Self-Employed | 0,5 |
| | | 4. Private Employee | 0,25 |
| | | 5. Civil Servant/Military/Police | 0 |
| $X_4$ | Father's Education | 1. Never went to school | 1 |
| | | 2. Elementary School | 0,84 |
| | | 3. Junior High School | 0,7 |
| | | 4. Senior High School | 0,56 |
| | | 5. Diploma 1 | 0,42 |
| | | 6. Diploma 2/Diploma 3 | 0,28 |
| | | 7. Bachelor's Degree | 0,14 |
| | | 8. Master's Degree/Doctoral Degree | 0 |
| $X_5$ | Mother's Education | 1. Never went to school | 1 |
| | | 2. Elementary School | 0,84 |
| | | 3. Junior High School | 0,7 |
| | | 4. Senior High School | 0,56 |
| | | 5. Diploma 1 | 0,42 |
| | | 6. Diploma 2/Diploma 3 | 0,28 |
| | | 7. Bachelor's Degree | 0,14 |
| | | 8. Master's Degree/Doctoral Degree | 0 |
| $X_6$ | Father's Income | 1. $\text{Rp}. 0, -$ | 1 |
| | | 2. $< \text{Rp } 1.000.000$ | 0,8 |
| | | 3. $\text{Rp } 1.000.000 - \text{Rp } 2.000.000$ | 0,6 |
| | | 4. $\text{Rp } 2.000.000 - \text{Rp } 3.000.000$ | 0,4 |
| | | 5. $\text{Rp } 3.000.000 - \text{Rp } 4.000.000$ | 0,2 |
| | | 6. $> \text{Rp } 4.000.000$ | 0 |
| $X_7$ | Mother's Income | 1. $\text{Rp}. 0, -$ | 1 |
| | | 2. $< \text{Rp } 1.000.000$ | 0,8 |
| | | 3. $\text{Rp } 1.000.000 - \text{Rp } 2.000.000$ | 0,6 |
| | | 4. $\text{Rp } 2.000.000 - \text{Rp } 3.000.000$ | 0,4 |
| | | 5. $\text{Rp } 3.000.000 - \text{Rp } 4.000.000$ | 0,2 |
| | | 6. $> \text{Rp } 4.000.000$ | 0 |
| $X_8$ | Number of dependents | 1. $>$ 4 persons | 1 |
| | | 2. 3 persons | 0,66 |
| | | 3. 2 persons | 0,33 |
| | | 4. 1 person | 0 |
| $X_9$ | Academic Grades (Average class report card grades) | 1. 9,6 - 10 | 1 |
| | | 2. 9,1 - 9,5 | 0,8 |
| | | 3. 8,6 – 9,0 | 0,6 |
| | | 4. 8,1 – 8,5 | 0,4 |
| | | 5. 7,6 – 8,0 | 0,2 |
| | | 6. 7,0 – 7,5 | 0 |

| Input Variable | | | Value |
|---|---|---|---|
| $X_{10}$ | Academic Accomplishments (olympiad certificate, etc) | 1. > 2 | 1 |
| | | 2. 1 | 0,5 |
| | | 3. None | 0 |
| $X_{11}$ | Non-Academic Accomplishments (Chess, Swim competition, etc.) | 1. > 2 | 1 |
| | | 2. 1 | 0,5 |
| | | 3. None | 0 |
| $X_{12}$ | Family Residential Home Ownership | 1. None | 1 |
| | | 2. Staying at others | 0,75 |
| | | 3. Monthly Rent | 0,5 |
| | | 4. Annual Rent | 0,25 |
| | | 5. Self-Owned | 0 |
| $X_{13}$ | Surface area | 1. < 25 $m^2$ | 1 |
| | | 2. < 25-50 $m^2$ | 0,75 |
| | | 3. 50-99 $m^2$ | 0,5 |
| | | 4. 100 - 200 $m^2$ | 0,25 |
| | | 5. > 200 $m^2$ | 0 |
| $X_{14}$ | Building area | 1. < 25 $m^2$ | 1 |
| | | 2. < 25-50 $m^2$ | 0,75 |
| | | 3. 50-99 $m^2$ | 0,5 |
| | | 4. 100 - 200 $m^2$ | 0,25 |
| | | 5. > 200 $m^2$ | 0 |
| $X_{15}$ | Physical Condition of the Building | 1. Non-Permanent | 1 |
| | | 2. Semi-Permanent | 0,5 |
| | | 3. Permanent | 0 |
| $X_{16}$ | Power Source | 1. None | 1 |
| | | 2. Solar-Powered | 0,75 |
| | | 3. National Electricity Company (PLN) | 0,5 |
| | | 4. PLN and Generators | 0,25 |
| | | 5. Generators/Private Source | 0 |
| $X_{17}$ | Water Source | 1. River / Water Springs / Mountains | 1 |
| | | 2. Well | 0,66 |
| | | 3. Municipal Waterworks (PDAM) | 0,33 |
| | | 4. Packaged water | 0 |
| $X_{18}$ | Bath, Wash and toilet facilities | 1. Sharing with other households | 1 |
| | | 2. Self-owned outside the house | 0,5 |
| | | 3. Self-owned inside the house | 0 |

While the target variable consists of two classes, namely passed (1) and not passed (-1).

## 2.1 Linear SVM

Let there be $m$ training data $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ where $x_i \in \mathbb{R}^n$ is the data sample and$y_i \in \{1, -1\}$ is the target or class of the sample data. Suppose also that the data for both classes is linearly separable, then we find the separator function

$$f(x) = xw + b = 0 \quad (1)$$

where $w \in \mathbb{R}^{n\times 1}$ is the weight parameter and $b \in \mathbb{R}$ is the bias parameter, and it applies that

$$\begin{aligned} x_i w + b > 0 &\ \text{ for } y_i = 1 \\ x_i w + b < 0 &\ \text{ for } y_i = -1 \end{aligned} \quad (2)$$

Let $H: xw + b = 0$ is the separator to be determined, while $H_1: xw + b = 1$ and $H_2: xw + b = -1$ are the separators from 1 class and -1 class. To obtain the optimal value of $H$, then de distance of$H_1$ and$H_2$ to $H$

must be the same, under the condition that there are no data samples between $H_1$ and$H_2$, and the distance between$H_1$ and$H_2$ is the maximal distance [17].

To maximize the distance between $H_1$ and $H_2$ then we use a positive data sample located at $H_1$ and negative data sample located at $H_2$. These data samples are called support vector because of its function as a determinant in obtaining the optimal separator. Meanwhile, other data samples can be discarded or moved towards $H_1$ and $H_2$ as long as they do not pass through each separator.

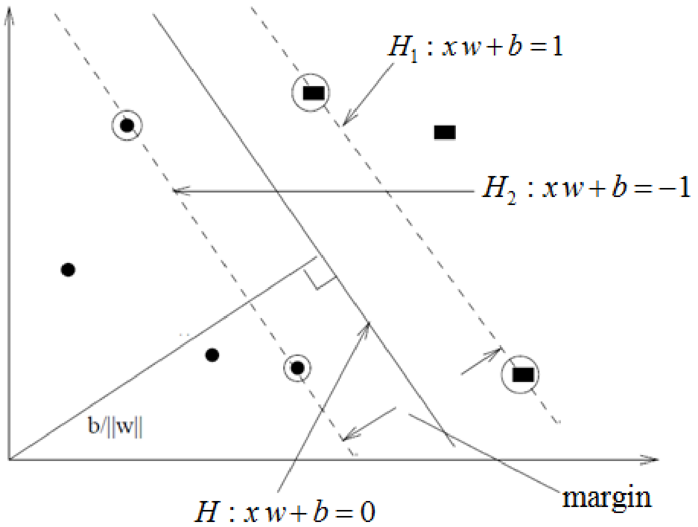


**Figure 1. The optimal separator that separates the two classes.**

Let $(x_0, y_0) \in \mathbb{R}^2$be any arbitrary point, then the distance of this point to the line$Ax + by + C = 0$ is

$$\frac{|Ax_0+By_0+C|}{\sqrt{A^2+B^2}} \tag{3}$$

such that the distance of the sample data $x$ located between$H_1$ and $H$ is

$$\frac{|xw+b|}{\sqrt{w^T w}} = \frac{1}{\|w\|} \tag{4}$$

because the distance of $H_1$ and $H_2$ to $H$ is the same, then the distance between $H_1$ and$H_2$ is$\frac{2}{\|w\|}$.

Thus, the problem of finding the optimal parameters $w$ and $b$ in order to obtain the optimal separator is a quadratic programming problem

$$\min_{w,b} \frac{1}{2} w^T w$$

with the constraints of (5)

$$y_i(x_i w + b) \geq 1, i = 1, \dots, m$$

Usually, it is quite difficult to solve the primal form above, so the primal form is changed to its dual form by introducing a Lagrange multiplier.

Let $\alpha \in \mathbb{R}^{m\times 1}$ is a Lagrange multiplier, then the quadratic programming problem (5) above turns into

$$L(w,b,\alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{m} \alpha_i \big(y_i(x_i w + b)\big) + \sum_{i=1}^{m} \alpha_i \tag{6}$$

By applying the Karush-Kuhn-Tucker (KKT) condition, the dual form obtained is

$$\text{Max} \quad L(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=j=1}^{m} \alpha_i \alpha_j y_i y_j x_i x_j^T$$

with the constraints of (7)

$$\sum_{i=1}^{m} \alpha_i y_i = 0 \text{ and } \alpha_i \geq 0 \text{ where} i = 1,2, \dots, m$$

The weight and bias parameters can be calculated by the equation

$$w = \sum_{i=1}^{N_{sv}} \alpha_i y_i x_i \text{ and } \quad b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (y_i - x_i w) \tag{8}$$

$SV$ is a collection of support vectors and $i \in SV$ if $\alpha_i \neq 0$. $N_{SV}$ is the number of support vectors. By using the equation

$$f(x) = xw + b$$

Then the new input data $x \in \mathbb{R}^n$ is classified into

$$\begin{cases} +1 \text{Class,if} f(x) > 0 \\ -1 \text{Class,if} f(x) < 0 \end{cases} \tag{9}$$

Furthermore, if there is a case of imperfect separation, in this case there is data between $H_1$ and $H_2$ as shown in the figure below.

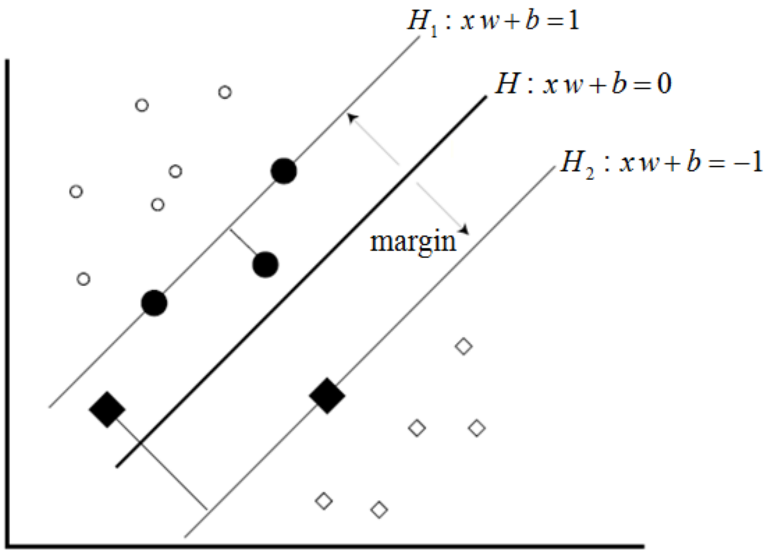


**Figure 2. Examples of cases of imperfect data separation**

To solve this problem, a non-negative slack variable $\mu$ ($\mu \geq 0$) is introduced and entered the constraint function (5) so that it becomes

$$y_i(x_i w + b) \geq 1 - \mu_i, i = 1, \ldots, m$$

Whereas a positive parameter $C$ is added to the objective function until it becomes

$$\frac{1}{2} w^T w + C \sum_{i=1}^{m} \mu_i$$

In full, the quadratic programming problem (5) turns into

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \mu_i$$

with the constraints of (10)

$$y_i(x_i w + b) \geq 1 - \mu_i, \; \mu_i \geq 0, i = 1, \ldots, m$$

Using the Lagrange multiplier $\alpha \in \mathbb{R}^{m \times 1}$ then the primal form (10) can be changed to the dual form as follows

$$\text{Max} \qquad L(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=j=1}^{m} \alpha_i \alpha_j y_i y_j x_i x_j^T$$

with the constraints of (11)

$$\sum_{i=1}^{m} \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C , \quad i = 1,2, \ldots, m$$

### 2.2 Non-linear SVM

In the classification problem, most data samples are not linearly separated so that if linear SVM is used, the results obtained are not optimal and result in poor classification results. One of the advantages of SVM lies in this section, namely SVM can be extended to solve non-linear problems. Linear SVM can be changed to non-linear SVM using kernel methods [17]. This method works by mapping the input data to a higher dimensional feature space using a function $\phi$. for example, let $u = (u_1, u_2)$ is the input data on $\mathbb{R}^2$ and $\phi(u) = \left(1, \sqrt{2}u_1, \sqrt{2}u_2, u_1^2, u_2^2, \sqrt{u_1}u_2\right)$ is the input data in the higher dimensional feature space $\mathbb{R}^5$. It is expected that the input data resulting from the mapping into the feature space will be separated linearly so that the optimal separator can be found.

Let $x \to \phi(x)$ Then the equations (2.11) can be written as

$$\text{Max } \psi(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=j=1}^{m} \alpha_i \alpha_j y_i y_j \phi^T(x_i)\phi(x_j)$$

with the constraints of (12)

$$\sum_{i=1}^{m} \alpha_i y_i = 0 \text{ dan } 0 \le \alpha_i \le C, \text{ where } i = 1,2,\dots,m$$

The weight and bias parameters can be calculated by the equation

$$w = \sum_{i=1}^{N_{SV}} \alpha_i y_i \phi(x_i) \text{ and } b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} \left(y_i - w^T \phi(x_i)\right) \quad (13)$$

Meanwhile, the optimal separator in Equation (1) changes to

$$f(x) = w^T \phi(x) + b = 0 \quad (14)$$

The problem that arises is if the sample input data for the training stage is in large numbers, then the dot product is calculated by $\phi^T(x_i)\phi(x_j)$ on **Equation (12)** will make the calculation time longer. Therefore, we need a way to calculate $\phi^T(x_i)\phi(x_j)$ without knowing the function $\phi$.

Let $K$ be a function with properties

$$K(u, v) = \phi^T(u)\phi(v)$$

where $u, v \in \mathbb{R}^n$ and$\phi: \mathbb{R}^n \to \mathbb{R}^m$ , $n < m$. These K functions are known as kernel functions. According to **[17]**, the frequently used kernel functions are as follows:

a. linear kernel: $K(u, v) = u^T v$.
b. polynomial kernel: $K(u, v) = (1 + u^T v)^d, d \ge 2$.
c. RBF kernel (Radial Basis Function) : $K(u, v) = exp(-\gamma \|u - v\|^2), \gamma > 0$.

By using the kernel function concept above, **Equation (12)** changes to

$$\text{Max } \psi(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

with the constraints of (15)

$$\sum_{i=1}^{m} \alpha_i y_i = 0 \text{ and } 0 \le \alpha_i \le C \text{ where } i = 1,2,\dots,m$$

the bias parameter can be calculated by the equation

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} \left(y_i - \sum_{i=1}^{N_{SV}} \alpha_i y_i K(x_i, x_j)\right) \quad (16)$$

with the optimal separator of

$$f(x) = \sum_{i=1}^{N_{SV}} \alpha_i y_i K(x_i, x) + b \quad (17)$$

the new input data $x \in \mathbb{R}^n$ is still classified based on the conditions in (9).

**Algorithm 1**. Classification Algorithm with SVM **[17]**

i. Normalization of training data $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ where $x_i \in \mathbb{R}^n$ is the data sample $y_i \in \{1, -1\}$.
ii. Set the $C$ parameters and the kernel to be used.
iii. Solve quadratic programming problem in **Equation (15)** to find the value of$\alpha_i$.
iv. Obtain the bias parameter with **Equation (16)** and optimal separator using **Equation (17)**.
v. The new input data $x \in \mathbb{R}^n$ is classified based on the conditions in **Equation (9)**.

## 3. RESULTS AND DISCUSSION

The data used in the research will be divided into two parts, namely training data and testing data. The training data will be used in the training process using SVM to produce the optimal separator. While the test data will be used to see how much accuracy is obtained from the training results. This data sharing process will be made in three data sharing schemes namely, 60/40, 70/30, and 80/20. For example, the 60/40 data division means that the training data is taken as much as 60% of the total data while the remaining 40% is used as test data. Furthermore, these three data sharing schemes will be processed using linear SVM and non-linear SVM with various parameter variations.

### 3.1 Results Using Linear SVM

The results of data processing and testing using a linear Support Vector Machine with $C = 1$ for the three data distribution schemes can be seen in **Table 2** below.

**Table 2. Accuracy Level with Linear SVM Method**

| No | Sharing Scheme | Accuracy Rate |
|---|---|---|
| 1 | 60/40 | 68,92% |
| 2 | 70/30 | 73,21% |
| 3 | 80/20 | 89,19% |

From **Table 2** it can be seen that the best rate was obtained at 89.19% in the 80/20 data sharing scheme.

### 3.2 Results Using Non-linear SVM

For the nonlinear SVM method, two kernels will be used as a comparison, namely the polynomial kernel and the RBF (Radial Basis Function) kernel. The parameter value to be varied in the polynomial kernel is the value of $d$ (degree) while in the RBF kernel it is the value of $\gamma$ (sigma). This aims to get the best level of accuracy. The results of training and data testing using nonlinear SVM with $C = 1$ for the three data sharing schemes can be seen in **Table 3**, **Table 4** and **Table 5** below.

**Table 3. Accuracy Level with 60/40 Data Sharing Scheme**

| Kernel type: polynomial | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Kernel Parameter | $d$ = 2 | $d$ = 3 | $d$ = 4 | $d$ = 5 | $d$ = 6 | $d$ = 7 | $d$ = 8 | $d$ = 9 |
| Accuracy Rate | 56,76% | 51,35% | 54,05% | 54,05% | 54,05% | 54,05% | 51,35 % | 28,38 % |
| Kernel type: RBF | | | | | | | | |
| Kernel Parameter | $\gamma$ = 0,1 | $\gamma$ = 0,5 | $\gamma$ = 2,1 | $\gamma$ = 2,5 | $\gamma$ = 3,1 | $\gamma$ = 3,5 | $\gamma$ = 4,1 | $\gamma$ = 4,5 |
| Accuracy Rate | 81,08 % | 79,73% | 82,43% | 83,78% | 83,78% | 82,43% | 82,43% | 82,43% |

From **Table 3** it can be seen that the results of testing with the polynomial kernel obtained the best results on the parameter value $d = 2$ with an accuracy rate of 56.76%, while for the RBF kernel an accuracy rate of 83.78% was obtained with $\gamma = 2,5$ and $\gamma = 3,1$.

**Table 4. Accuracy Level with 70/30 Data Sharing Scheme**

| Kernel type: polynomial | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Kernel Parameter | $d$ = 2 | $d$ = 3 | $d$ = 4 | $d$ = 5 | $d$ = 6 | $d$ = 7 | $d$ = 8 | $d$ = 9 |
| Accuracy Rate | 69,64% | 64,29% | 62,5% | 64,29% | 64,29% | 66,07% | 57,14% | 25 % |
| Kernel type: RBF | | | | | | | | |
| Kernel Parameter | $\gamma$ = 0,1 | $\gamma$ = 0,5 | $\gamma$ = 2,1 | $\gamma$ = 2,5 | $\gamma$ = 3,1 | $\gamma$ = 3,5 | $\gamma$ = 4,1 | $\gamma$ = 4,5 |
| Accuracy Rate | 87,5% | 83,93% | 91,07% | 89,29% | 89,29% | 87,5% | 87,5% | 87,5% |

From **Table 4** it can be seen that the test results with the polynomial kernel obtained the best results on the parameter value $d = 2$ with an accuracy rate of 69.64%, while for the RBF kernel the accuracy rate was 91.07% with $\gamma = 2,1$.

**Table 5. Accuracy Level with 80/20 Data Sharing Scheme**

| Kernel type: polynomial | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Kernel Parameter | $d$ = 2 | $d$ = 3 | $d$ = 4 | $d$ = 5 | $d$ = 6 | $d$ = 7 | $d$ = 8 | $d$ = 9 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Accuracy Rate | 72,97% | 70,27% | 75,68% | 72,97% | 75,68% | 72,97% | 43,24% | 16,22 % |
| Kernel type: RBF | | | | | | | | |
| Kernel Parameter | $\gamma = 0,1$ | $\gamma = 0,5$ | $\gamma = 2,1$ | $\gamma = 2,5$ | $\gamma = 3.1$ | $\gamma = 3,5$ | $\gamma = 4.1$ | $\gamma = 4,5$ |
| Accuracy Rate | 94,59% | 94,59% | 94,59% | 94,59% | 94,59% | 94,59% | 94,59% | 94,59% |

From **Table 5** it can be seen that the test results with the polynomial kernel obtained the best results on the parameter values $d = 4$ and $d = 6$ with an accuracy rate of 75.68%, while for the RBF kernel the accuracy rate was 94.59% with all variations of the sigma values. used.

Based on the results obtained using the linear SVM method and the non-linear SVM method, it can be seen that the non-linear SVM method has a better level of accuracy than the linear SVM. This is because of the three data sharing schemes, the linear SVM only has the best accuracy rate of 89.19% while the nonlinear SVM has an accuracy rate of 94.59%. Especially for nonlinear SVM, the RBF kernel has a better level of accuracy than the polynomial kernel. This can be seen from the results of testing the three data sharing schemes, the RBF kernel has better results on each data sharing scheme. In the 60/40 data distribution scheme an accuracy rate of 83.78% is obtained, the 70/30 distribution scheme obtains an accuracy rate of 91.07%, and the 80/20 distribution scheme obtains an accuracy rate of 94.59%.

### 3.3 System Design and Implementation

In this study, the Matlab 2018b software was used to design and create the interface for the selection and determination of KIP Kuliah recipients. The system will be built based on the results of non-linear SVM training with the RBF kernel. On the initial page of the system, there will be two menu buttons and instructions for using the application. Application details can be seen in **Figure 3**- **Figure 5** below.

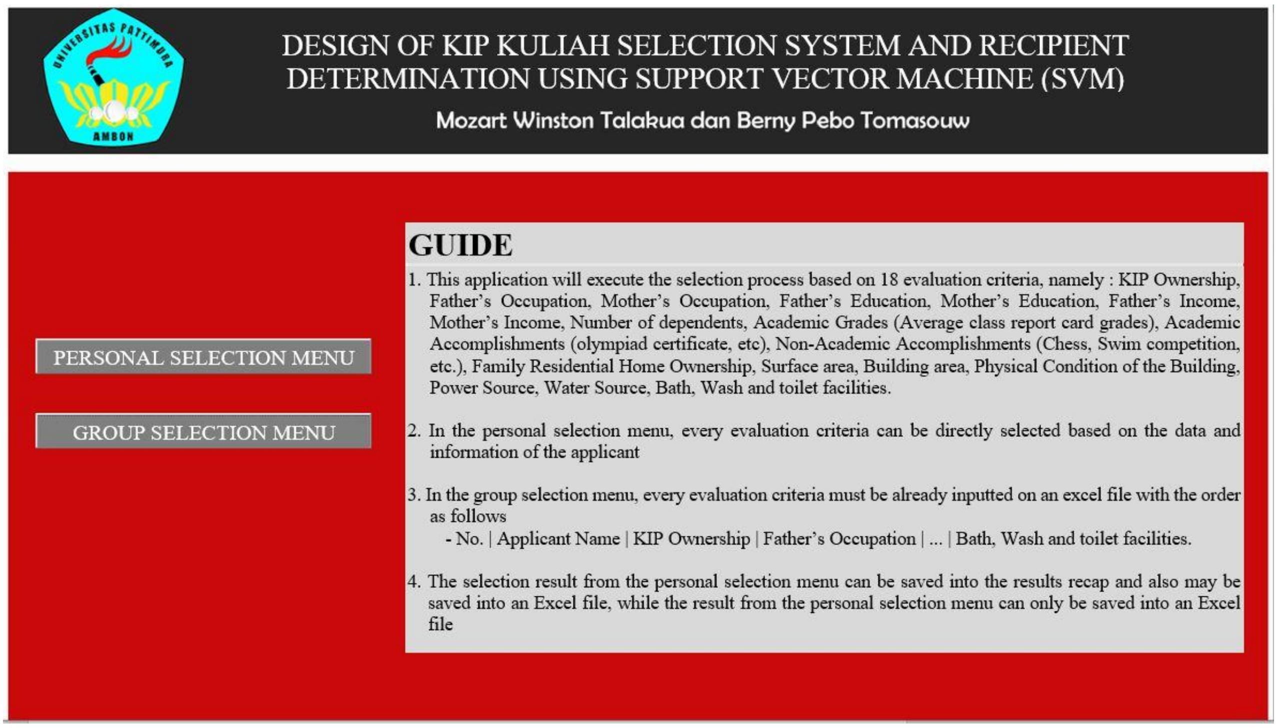


**Figure 3. First Page of the Application**

**Figure 3** shows the first page of the application which consists of two menu options, namely the individual selection menu and the group selection menu. This page also contains instructions for using the application. While **Figure 4** is an individual selection menu. In this menu, the name of the registrant must be filled in completely with all data and information. After that, the selection button can be used to run the selection process and display the results. There is also an option to save the results of the data and information of the registrant and the selection results. A summary of all registrant data that has been processed can be seen by pressing the view summary button.

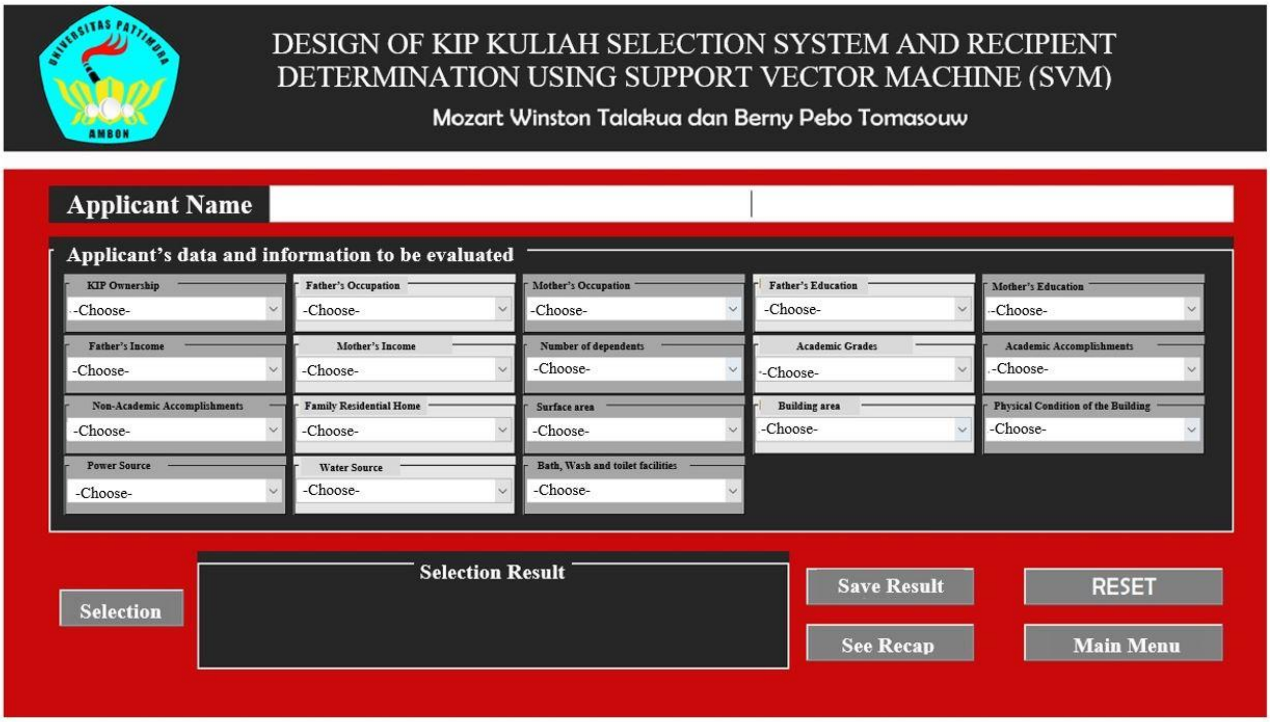


**Figure 4. Individual Selection Menu**

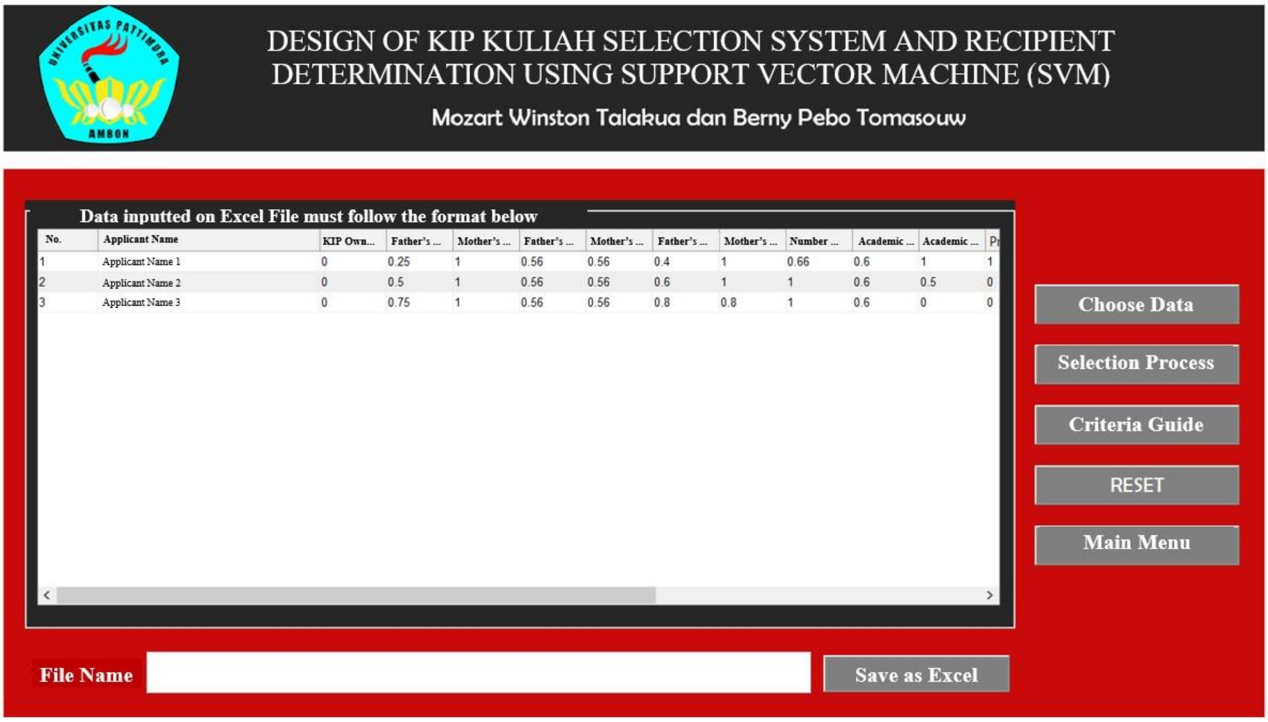


**Figure 5. Group Selection Menu**

**Figure 5** is a menu display for the group selection process. In this section, registrant data must be prepared in an excel file with a predetermined format. If the registrant data has been selected then the selection process can be run by pressing the selection process button.

## 4. CONCLUSIONS

Based on the results of the study it can be concluded that

a. The nonlinear SVM method has a better level of accuracy than the linear SVM. This is because of the three data sharing schemes, the linear SVM only has the best accuracy rate of 89.19% while the nonlinear SVM has an accuracy rate of 94.59%.

b. Especially for nonlinear SVM, the RBF kernel has a better level of accuracy than the polynomial kernel. This can be seen from the results of testing the three data sharing schemes, the RBF kernel has better results on each data sharing scheme. In the 60/40 data distribution scheme, an accuracy rate of 83.78% was obtained, the 70/30 distribution scheme obtained an accuracy rate of 91.07%, and the 80/20 distribution scheme obtained an accuracy rate of 94.59%.

c. The KIP Kuliah selection and recipient determination system has created based on the results of non-linear SVM training with the RBF kernel.

## REFERENCES

[1] K. R. d. T. Pusat Layanan Pembiayaan Pendidikan Kementerian Pendidikan, "kip-kuliah.kemdikbud.go.id," 2 Februari 2023. [Online]. Available: https://kip-kuliah.kemdikbud.go.id/uploads/Pedoman-Pendaftaran-KIP-Kuliah-2023-update-15Mei2023-_e79a89.pdf. [Accessed 10 Januari 2023].

[2] D. K. D. B. Kemenristekdikti, "kemdikbud.go.id," 14 Juni 2019. [Online]. Available: https://lldikti13.kemdikbud.go.id/2019/06/14/petunjuk-teknis-pengelolaan-bidikmisi-2019/. [Accessed 12 April 2023].

[3] N. N. Rahmawati, M. I. A. Fathoni and Ismanto, "Penentuan Penerima KIP Kuliah Mahasiswa S1 Unugiri Menggunakan Fuzzy C-Means Clustering," *Transformasi : Jurnal Pendidikan Matematika dan Matematika,* vol. 6, no. 2, pp. 121-130, 2022.

[4] R. Susetyoko, W. Yuwono and E. Purwantini, "Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner," *Jurnal Informatika Polinema,* vol. 8, no. 4, pp. 31-40, 2022.

[5] H. Nopriandi, A. Aprizal and S. Chairani, "Sistem Pendukung Keputusan Seleksi Calon Penerima Beasiswa Kartu Indonesia Pintar Kuliah (Kip-K) Di Universitas Islam Kuantan Singingi," *JURNAL TEKNOLOGI DAN OPEN SOURCE,* vol. 6, no. 1, pp. 41-54, 2022.

[6] P. A. Kurniawijaya and I. W. W. Karsana, "Implementasi Metode AHP Dalam Sistem Penunjang Keputusan Penerima KIP Kuliah," *JUKI : Jurnal Komputer Dan Informatika,* vol. 5, no. 1, p. 22–31, 2023.

[7] B. P. Tomasouw and M. I. Irawan, "Multiclass Twin Bounded Support Vector Machine Untuk Pengenalan Ucapan," in *Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA,* Yoygakarta, 2012.

[8] Z. A. Leleury and B. P. Tomasouw, "Diagnosa Penyakit Saluran Pernapasan Dengan Menggunakan Support Vector Machine (SVM)," *BAREKENG: J. Math. & App.,* vol. 9, no. 2, pp. 109-119, 2015.

[9] R. Damasela, B. P. Tomasouw and Z. A. Leleury, "Penerapan Metode Support Vector Machine (SVM) Untuk Mendeteksi Penyalahgunaan Narkoba," *PARAMETER: Jurnal Matematika, Statistika dan Terapannya,* vol. 1, no. 2, pp. -, 2021.

[10] I. Hmeidi, B. Hawashin and E. El-Qawasmeh, "Performance of KNN and SVM classifiers on full word Arabic articles," *Advanced Engineering Informatics,* vol. 22, no. 1, pp. 106-111, 2008.

[11] Y. Qiu, "Towards Prediction of Pancreatic Cancer Using SVM Study Model," *JSM Clinical Oncology And Research,* vol. 2, no. 4, pp. 1-6, 2014.

[12] I. T. Utami, "Perbandingan Kinerja Klasifikasi Support Vector Machine (SVM) Dan Regresi Logistik Biner dalam Mengklasifikasikan Ketepatan Waktu Kelulusan Mahasiswa FMIPA UNTAD," *Jurnal Ilmiah Matematika dan Terapan,* vol. 15, no. 2, pp. 256-267, 2018.

[13] R. Wahyudi and G. Kusumawardhana, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *JURNAL INFORMATIKA,* vol. 8, no. 2, pp. 200-207, 2021.

[14] M. T. Anjasmoros, Istiadi and F. Marisa, "Analisis Sentimen Aplikasi Go-Jek Menggunakan Metode SVM Dan NBC (Studi Kasus: Komentar Pada Play Store)," in *The 3rd Conference on Innovation and Application of Science and Technology (CIASTECH 2020),* Malang, 2020.

[15] S. Dong, "Multi class SVM algorithm with active learning for network traffic classification," *Journal Expert Systems with Applications,* vol. 176, no. -, pp. -, 2021.

[16] M. Onel , C. A. Chris A. Kieslich and E. N. Pistikopoulo, "A nonlinear support vector machine-based feature selection approach for fault detection and diagnosis: Application to the Tennessee Eastman process," *AIChE Journal,* vol. 65, no. 3, p. 992–1005, 2018.

[17] S. Abe, Support Vector Machines for Pattern Classification, 2 ed., London: Springer-Verlag, 2010.