**Building of Informatics, Technology and Science (BITS)**
Volume 3, No 3, Desember 2021 Page: 325−330
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1041

# Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur

**Ahmad Ari Aldino*, Alvin Saputra, Andi Nurkholis, Setiawansyah**

Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email Penulis Korespondensi: aldino@teknorkat.ac.id
Submitted: **16/12/2021**; Accepted: **28/12/2021**; Published**: 31/12/2021**

**Abstrak**−Classification is a technique for grouping and categorizing specific standards as material for compiling information, making conclusions, or making decisions. This paper discusses data classification for underprivileged communities in Tanjung Inten, Purbolinggo, East Lampung using the Support Vector Machine (SVM) algorithm, then grouped into two label classes, namely the less fortunate and capable label classes. From the data that has been collected, 1154 data. The data goes through processing, scoring, labeling, and testing, producing two classes of results, namely less fortunate and capable. From the test data using the Support Vector Machine (SVM) method, the accuracy score is 97%, the precision score is 97%, the Recall score is 100%, and the F1-Score is 98%. This test resulted in a proportion of classification with the capable label is 87% and less fortunate label is 13%.

**Kata Kunci:** Classification; Support Vector Machine (SVM); Linear Kernel

## 1. INTRODUCTION

The problem of the less fortunate is a problem that always exists in every country; even though the era has entered the age of globalization, it cannot be denied that the issue of poverty has always been an obstacle to the progress of each country. The issues of underprivileged communities are not only found in developing countries; even developed countries also have issues with underprivileged communities and poverty [1].

Tanjung Inten is a village located in the Purbolinggo District, East Lampung. The residents in Tanjung Inten village have quite a variety of livelihoods, but most of the residents in Tanjung Inten village make a living as farmers, farm laborers, and traders. Differences in these professions will cause differences in income levels and food demand in the community. Each region certainly has a social assistance program for people considered economically disadvantaged. But of course, terms and conditions apply[2]–[4].

Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts that aim to predict the type of objects whose class label is unknown[5]. Classification algorithms that are widely used are Decision/classification trees, Bayesian classifiers/ Naïve Bayes classifiers, Neural networks, Statistical Analysis, Genetic Algorithms, Rough sets, k-nearest neighbor, Rule-Based Method, Memory-based reasoning, and Support vector machines (SVM) [6]–[8].

Based on what has been described, research will be carried out on the Application of the Support Vector Machine (SVM) Algorithm in Classifying the Less fortunate in Tanjung Inten Village, Purbolinggo, East Lampung. Using data mining techniques, the research will process the data for classification based on population data obtained from the Tanjung Inten village. The input variables that will be used in classifying the underprivileged are Dependents, Jobs, Income, Land Area, Houses, Categories according to the data that has been taken and by the variables to be inputted, then the results of the classification will determine the level of underprivileged such as: Able and Less fortunate

## 2. RESEARCH METHODS

### 2.1 Research Stages

The stages of the research are the design of the flow in a study that is structured and conveyed through pictures with sequential steps of what will be done in a study. The following image of the stages of research proposed by the author in this study can be seen in Figure 1:

**Building of Informatics, Technology and Science (BITS)**
Volume 3, No 3, Desember 2021 Page: 325−330
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1041

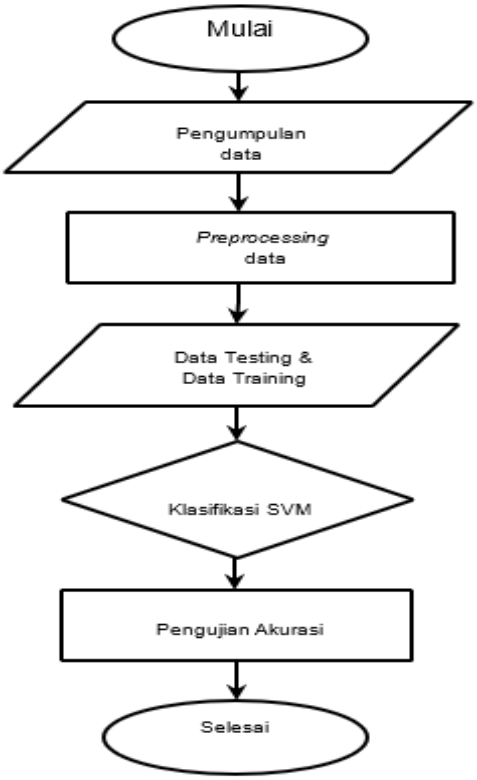


**Figure 1.** Below

Explanation of the research framework to be carried out in this research:

1. **Data Collection Stage**
   The initial stage is the process of collecting *datasets*. In this study, the dataset comes from data from the Tanjung Inten village community, Purbolinggo, East Lampung. The data obtained amounted to 1154; the following is a comparison of the criteria described using a graph on each variable

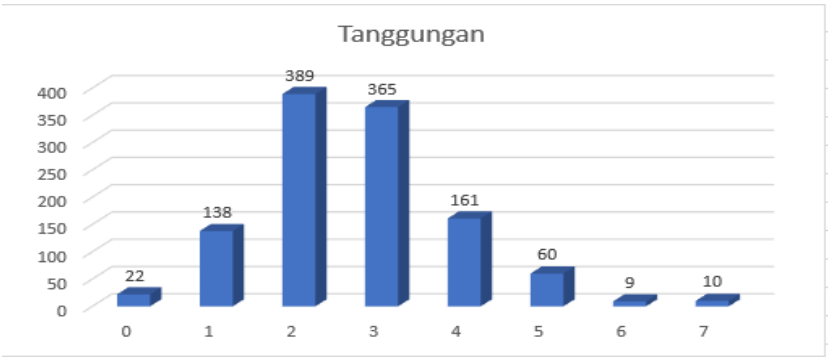


**Figure 2.** Dependent Criteria

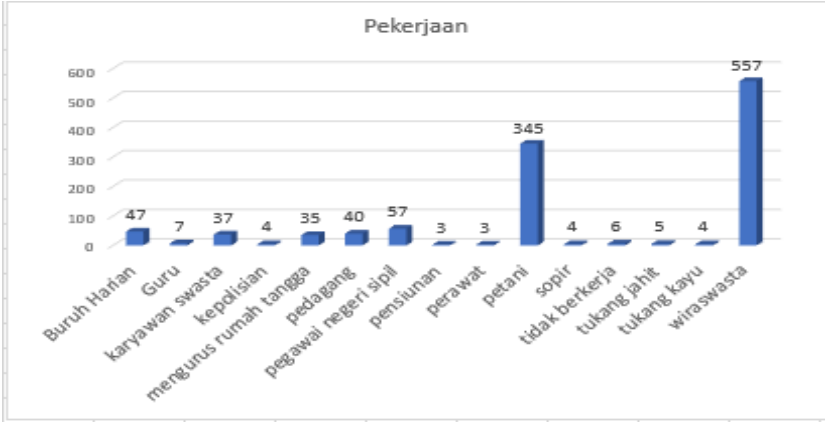


**Figure 3.** Employment Criteria

**Building of Informatics, Technology and Science (BITS)**
Volume 3, No 3, Desember 2021 Page: 325−330
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1041

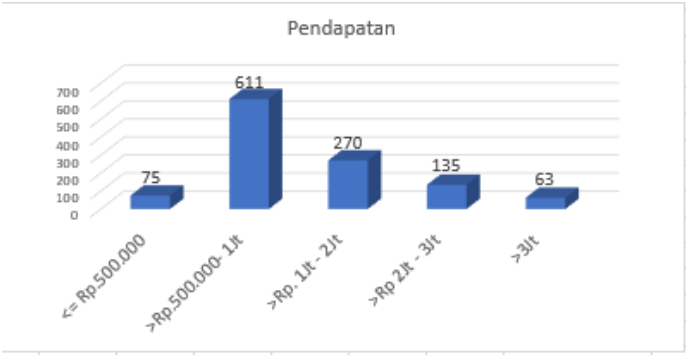


**Figure 4.** Income Criteria

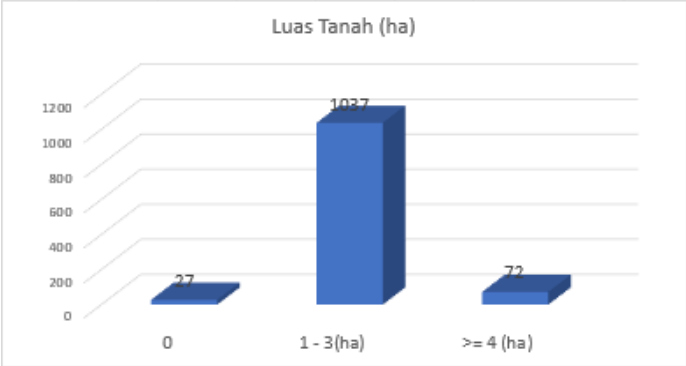


**Figure 5.** Criteria for Land Area (ha)

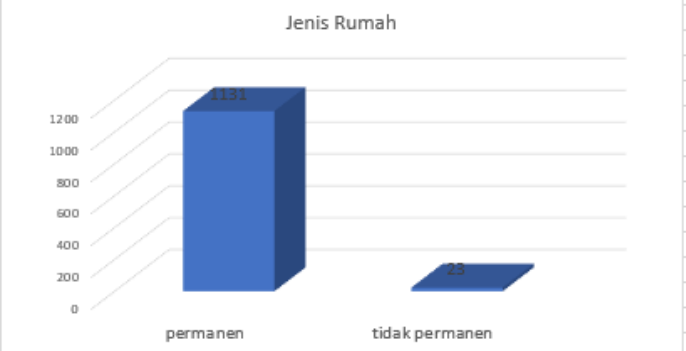


**Figure 6.** Criteria for Type of House

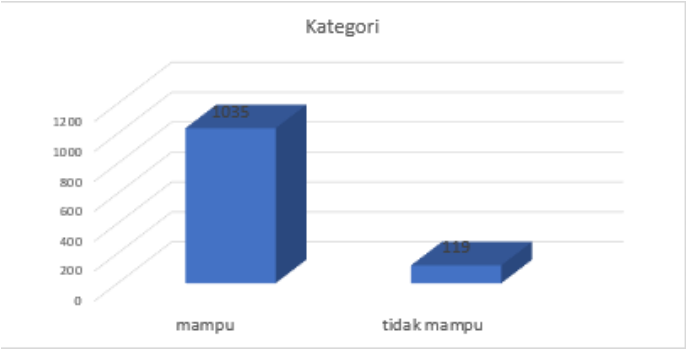


**Figure 7.** Category Criteria

**2. Preprocessing process**

The initial stage of the process is needed to make it easier when classifying later, using the *LabelEncoder* module, a module in the *sci-kit-learn library*. The module changes the Employment and Home type values and categories. The data is converted into numbers and sorted from 0 onwards in alphabetical order.
Day Laborer "0"
Teacher "1"
Private Employee "2"

**Building of Informatics, Technology and Science (BITS)**
Volume 3, No 3, Desember 2021 Page: 325−330
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1041

Indonesian Police "3"
Taking care of household "4"
Merchant "5"
Civil servant "6"
Retired "7"
Nurse "8"
Farmer "9"
Driver "10"
Not working "11"
Tailor "12"
Carpenter "13"
Entrepreneur "14"
As for the house variable, namely:
Permanent "0"
Not permanent "1"
As for the categorical variables, namely:
Capable of "0"
Less fortunate "1"

**Table 1.** *Preprocessing* results

| Dependent | Work | Income | Land Area (ha) | House | Category |
|---|---|---|---|---|---|
| 4 | 14 | 3500000 | 3 | 0 | 0 |
| 2 | 6 | 2500000 | 1.5 | 0 | 0 |
| 3 | 14 | 3000000 | 2.75 | 0 | 0 |
| 4 | 0 | 300000 | 0 | 1 | 1 |
| 3 | 14 | 2500000 | 2 | 0 | 0 |

### 2.2 Scoring Process

The Scoring process at this stage, the initial step is to create an Average column, which will later be used to calculate the Score to be processed into the SVM calculation[9].

### 2.3 Feature Scaling

*Feature Scaling* or normalization is when you change the *numerical* values in the dataset to a standard scale. The goal at this stage is to help speed up the calculation mode process using the Algorithm[10]. *Feature Scaling* used by the author is *StandardScaler* which is part of the *sklearn* module or usually called Standardization Normalization / *Z-Score*. Normalization Standardization or Z-Score has the following formula:

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

With the Mean Formula:

$$\mu = \frac{1}{N-1} \sum_{i=1}^{N}(xi) \quad (2)$$

and the standard Deviation Formula:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N}(x_i - \mu)^2} \quad (3)$$

## 3. RESULTS AND DISCUSSION

### 1. SVM *Linear* Kernel

The author includes the Support Vector Machine Algorithm, a module in the sci-kit-learn library, using a linear kernel. At this stage, the data classification process[11]. Here is the program code used:

```
# Membuat model SVM terhadap Training set
from sklearn.svm import SVC
classifier = SVC(kernel = 'linear', random_state = 0)
classifier.fit(X_train, y_train)
```

**Figure 8.** Linear Kernel SVM program code

### 2. Testing and Evaluation

The classification results will be tested using SVM Kernel Linear at this stage. Among them are Accuracy, Precision, Recall, and F1-Score. Then for testing precision, recall and F1-score are done using the confusion matrix module.

**Building of Informatics, Technology and Science (BITS)**
Volume 3, No 3, Desember 2021 Page: 325−330
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1041

Which will measure the performance of each class by calculating accuracy, recall, and F1-Score. With the following information:

**Tabel 2.** Confusion Matrix

| Actual Data | Prediction Data | |
|---|---|---|
| | Less fortunate | Capable |
| Less fortunate | TP | FN |
| Capable | FP | TN |

Description:
TP ( *True Positive* ) = data that is predicted to be true to the Less fortunate class
FN ( *False Negative* ) = data that is expected to be incorrectly entered into the Capable class
TN ( *True Negative* ) = data that is expected to be confirmed into the Capable class
FP ( *False Positive* ) = data predicted to be incorrectly entered into the Less fortunate class.

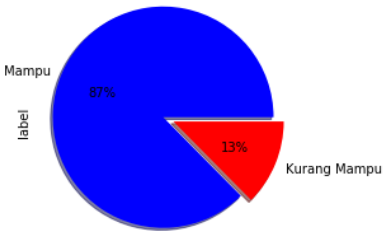


**Figure 9.** Category Criteria

The graph above shows that the results obtained from this study have a significant value in the capable class, which is 87%, and the less fortunate class at 13%. The manual calculations obtained are as follows

**Table 3.** Manual Calculation

| Actual Data | Prediction Data | |
|---|---|---|
| | Less fortunate | Capable |
| Less fortunate | 313 | 0 |
| Capable | 11 | 2 |

Less fortunate:

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{275}{275+12} = \mathbf{0.97}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{275}{275+0} = \mathbf{1.00}$$

$$\text{F1-Score} = 2 \times \frac{Precision \times Recall}{Precision+Recall} = \frac{\mathbf{0.96 \times 1}}{\mathbf{0.96 \times 1}}$$

$$\text{Precision} = \frac{TN}{TN+FN} = \frac{20}{20+0} = \mathbf{1.00}$$

$$\text{Recall} = \frac{TN}{TN+FP} = \frac{20}{20+12} = \mathbf{0.68}$$

$$\text{F1-Score} = 2 \times \frac{Precision \times Recall}{Precision+Recall} = \frac{1 \times 0.62}{01 \times 0.62} = \mathbf{0.81}$$

**Table 4.** Results

| | Precision | Recall | F1-score | support |
|---|---|---|---|---|
| Less fortunate | 0.97 | 1.00 | 0.98 | 313 |
| Capable | 1.00 | 0.68 | 0.81 | 34 |
| | | | | |
| Accuracy | | | 0.97 | 347 |
| Macros avg | 0.98 | 0.84 | 0.89 | 347 |
| Weighted avg | 0.97 | 0.97 | 0.97 | 347 |

## 4. CONCLUSION

Based on the research that has been done, it can be concluded that the classification of the underprivileged community uses the *Support Vector Machine* classification algorithm and the *python* programming language. The data collected

**Building of Informatics, Technology and Science (BITS)**
Volume 3, No 3, Desember 2021 Page: 325−330
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1041

amounted to 1154 data—based on dependent variable, occupation, income, land area, type of house. The data goes through processing, scoring, labeling, and testing. Produces an accuracy value of 97%.

# REFERENCES

[1] M. L. Suyanto, "Tingkat Dasar dan Lanjut," *Inform. Bandung*, 2018.
[2] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
[3] L. LUKMAN, "PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM PEMILIHAN BEASISWA: STUDI KASUS SMK YAPIMDA," *Fakt. Exacta*, vol. 9, no. 1, pp. 49–57, 2016.
[4] A. L. Sukmawati and S. A. Thamrin, "Klasifikasi Penyakit Diabetes Melitus Tipe II Menggunakan Metode Support Vector Machine."
[5] A. P. Kusumaningrum, "Optimasi Parameter Support Vector Machine menggunakan Genetic Algorithm untuk Klasifikasi Microarray Data." Institut Teknologi Sepuluh Nopember, 2017.
[6] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, "Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang," *J. Gaussian*, vol. 3, no. 4, pp. 811–820, 2014.
[7] R. Ilyarisma and L. A. S. Irfan, "Pengklasifikasian Warna Kulit Berdasarkan Ras Menggunakan Algoritma Support Vector Machine (SVM)," *DIELEKTRIKA*, vol. 3, no. 1, pp. 53–59, 2018.
[8] R. R. Saragih, "Pemrograman dan bahasa pemrograman," 2016.
[9] A. M. Puspitasari, D. E. Ratnawati, and A. W. Widodo, "Klasifikasi penyakit gigi dan mulut menggunakan metode Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, p. 964X, 2018.
[10] H. Susanto and S. Sudiyatno, "Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu," *J. Pendidik. vokasi*, vol. 4, no. 2, 2014.
[11] O. Somantri, S. Wiyono, and D. Dairoh, "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)," *Sci. J. Informatics*, vol. 3, no. 1, pp. 34–45, 2016.