IAIC International Conference Series
Vol. 4, No. 1, 2023, pp. 49 ~ 57
E-ISSN: 2774-5899 | P-ISSN: 2774-5880
DOI: https://doi.org/10.34306/conferenceseries.v4i1.625

Comparative SVM and Decision Tree Algorithm in Identifying the Eligibility of KIP Scholarship Awardee

[1*]Asriyanik, [2]Agung Pambudi
[1,2]Informatics Engineering, Muhammadiyah University of Sukabumi, Indonesia
e-mail: [1*]asriyanik263@ummi.ac.id, [2]agungpambd@ummi.ac.id
***Corresponding author**

(Received December 13, 2023, Revised December 14, 2023, Accepted December 14, 2023, Available online December 19, 2023)

Abstract

Scholarship selection process has specific rules, but if the number of applicants exceeds the quota, a selection process is needed. Based on the observation of a university in Sukabumi, the selection for KIP scholarship has not yet had a standard method. Several methods can be used to assist the selection process, such as classification based on historical data of applicants. The algorithms used for classification include Decision Tree (DT) and Support Vector Machine (SVM). The research process uses SEMMA (Sample, Explore, Modify, Model, Assess) method. Dataset for KIP scholarship awardee from 2021-2022 consist of 519 samples with 16 attributes. From the exploration results, the most important features for model modeling are Status DTKS, Status P3KE, Father's income, mother's income, combined income, and performance. These attributes are converted into numerical data to facilitate model fitting. The K-Fold Cross-Validation results for the Decision Tree model in the case of KIP Scholarship classification yield an accuracy of 78.44% for the entire test dataset, a precision of 0.73107, indicating that 73.11% of the predictions are true, a recall (sensitivity) of 78.45%, and an F1 score of 73.20%. The results for the SVM model are an accuracy of 80.17%, a precision of 84.44%, and a recall of 80.17%.

Keywords: *Classification, Awardee, Decision Tree, Support Vector Machine*

1. Introduction

Kartu Indonesia Pintar (KIP) “university” is one of the Smart Indonesia Programs as stipulated in the Ministry and Culture Regulation No. 10 of 2020 which is intended for students who are admitted to higher education institutions [1]. KIP university Merdeka aims to increase the economic potential and social mobility of students from poor/vulnerable families to attend college [2].

Each university has a quota for KIP university students. The number of KIP scholarship awardee is calculated based on the accreditation rank. If the number of applicants exceeds the quota, not all applicants can be accommodated, so a re-selection process is needed to ensure that KIP scholarship awardee are truly deserving students [3], so each university has various methods for the final selection process of KIP lecture recipients, including at one of the universities in Sukabumi. Currently, the selection and verification process of KIP university has been conducted through various methods, including interviews, home visits to prospective recipients, and selection of specializations relevant to the intended study program. However, the process of determining final admission eligibility is still not done in a clear and transparent manner [4]. One of them is through machine learning classification methods using data on applicants and recipients of KIP Lecture in previous years [5]. The classification process can be done by utilizing several algorithms such as Naive Bayes, Decision Tree, K-Nearest Neighbor and Support Vector Machine [6].

Previous research includes the classification of college KIP recipients using logistic regression [7]. In the research conducted by Ronny Susetyoko et al, the regression classification method can only produce numerical results or numbers with input variables only numbers. On the other hand, Gagan Suganda conducted research on the classification of college KIP recipients using a simple mathematical formula for conditional probability, Naive Bayes [8]. However, this method has its drawbacks. If the conditional probability is zero, the prediction probability is also zero, and the prediction result will not be optimal [9]. Decision tree was also used in another case study by Sathiyanarayanan to identify breast cancer. The results showed that the decision tree algorithm is simpler and can compare each attribute by assigning a value to each node, and the results showed an accuracy of 99% [10]. In addition, there is also the SVM algorithm. The SVM algorithm has high accuracy and can find subtle patterns in complex data sets [11].

In previous studies, researchers only used one algorithm, so this study will compare two types of algorithms, namely decision tree and support vector machine [12]. The purpose of this research is to compare the accuracy of the classification of college KIP recipients from the two algorithms [13]. The research stages will use the SEMMA (Sample, Explore, Modify, Model, and Assess) method which starts with collecting data sets, understanding data processing, preprocessing data, modeling, and testing the accuracy and precision of the model [14].

2. Research Method

This research will use the SEMMA (Sample, Explore, Modify, Model, and Assess) approach [15]. It starts with the process of determining the dataset, exploring and visualizing the dataset, and modifying the dataset so that it is ready to be modeled [16]. The research also includes modeling with machine teaching algorithms and evaluating the accuracy of the model [17]. To measure more stable model performance and reduce the risk of overfitting on training data, model validation was performed with the K-fold cross-validation method [18][19].

The confusion matrix table will be used to measure the performance of the model. This table can calculate model evaluation metrics such as accuracy, precision, recall, and F1 score [20]. The comparison of a classification result with all classification results is known as accuracy [21]. And recall indicates how successfully the algorithm recognizes the class, while precision indicates how precise the classification result is from all data [22]. The F1 score, which is a combination of recall and precision, shows the overall performance of the method. Figure 1 shows the stages of the research to be carried out.

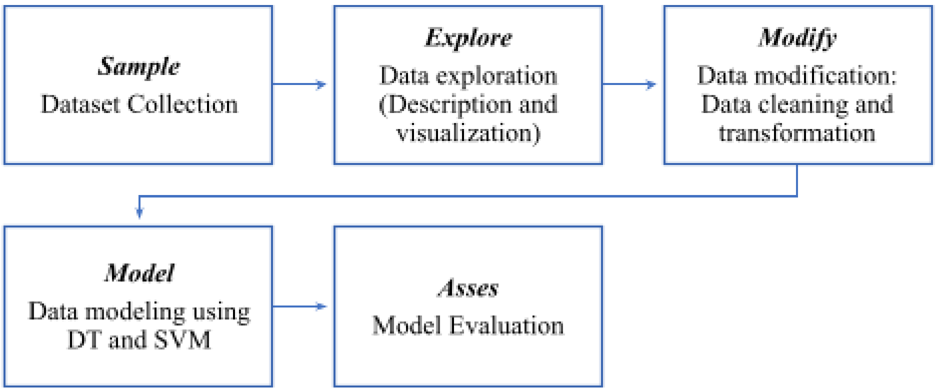

Figure 1. Research Stages

The research was conducted starting from 1) collecting datasets of KIP Lecture recipients for the last 5 years for modeling, 2) Dataset exploration by visualizing and describing the dataset, which is the process of understanding the dataset to sort out the data that suits the modeling needs, 3) Modification: variable selection, cleaning and transformation of the dataset. This process is carried out to ensure that the dataset to be modeled has been verified, the process that will be carried out is variable / feature selection, data cleaning and data transformation, 4) Modeling datasets with decision tree and SVM machine learning algorithms. In classifying data using the SVM method, the kernel function K(xi, xd) is used. The kernel function that will be used as in formula (1) as follows [23]:

$$K\left(x_i, x_d\right) = \left(x_j^T X_j + C)\right)^d, y > 0 \tag{1}$$

Decision tree studies a problem from an independent set of data depicted in a tree chart with a "divide and conquer" approach [24]. Formulas (2) and (3) are equations of data in tuples D

$$Info(D) = \sum_{i=1}^{N} - p_i p_i \tag{2}$$

Info(*D*) = ∑−*pi n i*= 1 *log*2 (*pi*) (3)

Accuracy of the machine learning model is measured by the confusion matrix table. The confusion matrix equation [25] to calculate accuracy, precision, and recall, is by collaborating the system classification results with actual observations grouped into True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Each equation for calculating the confusion matrix is as follows:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \tag{4}$$
$$Precision = TP / (TP + FP) \tag{5}$$
$$Recall = TP / (TP + FN) \tag{6}$$
$$F1\,Score = 2 * Precision * Recall / (Precision + Recall) \tag{7}$$

3. Results and Analysis

3.1. Dataset Collection

The dataset comes from prospective KIP Lecture applicants registered through the KIP website for 2021-2022, including a list of names proposed to get a KIP Lecture scholarship at one of the universities in Sukabumi. The dataset of prospective KIP Lecture scholarship recipients consists of 519 rows of data with 16 columns, namely Student Name (NamaSiswa), DTKS Status (StatusDTKS), P3KE Status (StatusP3KE), Father's Occupation (PekerajaanAyah), Father's Income (PenghasilanAyah), Father's Status (StatusAyah), Mother's Occupation (PekerjaanIbu), Mother's Income (PenghasilanIbu), Mother's Status (StatusIbu), Number of Dependents (JumlahTanggungan), Home Ownership (KepemilikanRumah), Electricity Source (SumberListrik), Land Area (LuasTanah), Building Area (LuasBangunan), Achievement (Prestasi) and Scholarship Status (StatusBeasiswa).

	Nama Siswa	Status DTKS	Status P3KE	Pekerjaan Ayah	Penghasilan Ayah	Status Ayah	Pekerjaan Ibu	Penghasilan Ibu	Status Ibu	Jumlah Tanggungan	Kepemilikan Rumah	Sumber Listrik	Luas Tanah	Luas Bangunan	Prestasi	Status Beasiswa
0	A. Jamal Assegap Supriadi	Belum Terdata	Belum Terdata	Lainnya	Rp. 250.001 - Rp. 500.000	Hidup	Lainnya	Tidak Berpenghasilan	Hidup	3 Orang	Sendiri	PLN	100 - 200 M2	100 - 200 M2	1. Juara 1 - Lomba volly 2021 - Tingkat Kabupa...	Berhak
1	Abdul Halim	Belum Terdata	Belum Terdata	Petani	Rp. 250.001 - Rp. 500.000	Hidup	Lainnya	Rp. 1.000.001 - Rp. 1.250.000	Hidup	3 Orang	Sendiri	PLN	50-99 M2	50-99 M2	NaN	NaN
2	Abielya Athaya	Belum Terdata	Belum Terdata	Lainnya	Rp. 750.001 - Rp. 1.000.000	Hidup	TIDAK BEKERJA	Tidak Berpenghasilan	Hidup	2 Orang	Menumpang	PLN	25-50 M2	25-50 M2	1. Juara 1 - Lomba film pendek 2019 - Tingkat ...	NaN
3	Adzkya Shaqiela	Belum Terdata	Belum Terdata	Peg. Swasta	Rp. 750.001 - Rp. 1.000.000	Hidup	Peg. Swasta	Tidak Berpenghasilan	Hidup	3 Orang	Sendiri	PLN	100 - 200 M2	50-99 M2	1. Finalis - Empat Pilar 2017 - Tingkat Kabupa...	NaN
4	Agna Mudzkia	Belum Terdata	Belum Terdata	Lainnya	Rp. 250.000	Bercerai	TIDAK BEKERJA	Tidak Berpenghasilan	Hidup	3 Orang	Sendiri	PLN	50-99 M2	50-99 M2	NaN	Berhak
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 2. Dataset of KIP Scholarship Recipients for the Year 2021-2022.

3.2. Data Exploration

Data exploration step to describe and visualize the dataset of KIP Scholarship Recipients. The raw dataset consists of 16 attributes containing non-numerical data objects.

Table 1. Attributes of KIP Scholarship Dataset.

Index	Attribute Name	Condition	Data Type
0	Student Name	Non Null	Object
1	DTKS Status	Non Null	Object
2	P3KE Status	Non Null	Object
3	Father’s Occupation	Non Null	Object
4	Father’s Income	Non Null	Object
5	Father’s Status	Non Null	Object

6	Mother's Occupation	Non Null	Object
7	Mother's Income	Non Null	Object
8	Mother's Status	Non Null	Object
9	Number of Dependents	Non Null	Object
10	Home Ownership	Non Null	Object
11	Electricity Source	Non Null	Object
12	Land Area	Non Null	Object
13	Building Area	Non Null	Object
14	Achievement	Non Null	Object
15	Scholarship Status	Non Null	Object

In the implementation of the decision tree and SVM algorithms, the required data is numeric, so data transformation is needed. The attributes that will form the dataset are DTKS status, P3KE status, combined father and mother's income, and the label is the Scholarship Status.

3.3 Data Modification

Original dataset which was obtained from the KIP Scholarship application process, still contains raw data. Therefore, it is necessary to adjust the data to be able to process it with decision tree and SVM algorithms. Some of the attributes that need to be modified include Status DTKS, Status P3KE, combined father and mother's income, and the label is the Scholarship Status.

1. Status DTKS and Status P3KE attributes, which contain the value 'Not Registered,' are changed to '0,' and for 'Registered,' they are changed to '1.' The modification process for these attributes is shown in Figure 3.

```
df['Status DTKS'] = df['Status DTKS'].replace({'Terdata': '1', 'Belum Terdata': '0'})
df['Status P3KE'] = df['Status P3KE'].replace({'Terdata': '1', 'Belum Terdata': '0'})
```

Figure 3. Modification of DTKS and P3KE Status Attributes

2. The attributes "Penghasilan Ayah" or Father's Income and "Penghasilan Ibu" or Mother's Income, which contain values such as 'Tidak Berpenghasilan', '-', and a range of salary options, will be combined and a new attribute called 'Family Income' will be created. The process of modifying these attributes is shown in Figure 4.

```
def convert_gaji_to_numeric(gaji_value):
    if gaji_value == 'Tidak Berpenghasilan' or gaji_value == '-':
        return 0
    elif ' - ' in gaji_value:
        range_parts = gaji_value.split(' - ')
        lower = convert_to_numeric(range_parts[0])
        upper = convert_to_numeric(range_parts[1])
        return (lower + upper) / 2
    elif '<' in gaji_value:
        value = gaji_value.replace('<', '').replace(' ', '')
        return convert_to_numeric(value) / 2
    else:
        return convert_to_numeric(gaji_value)

def convert_to_numeric(value):
    try:
        value = value.replace('Rp. ', '').replace('.', '').replace(' ', '')
        return int(value)
    except ValueError:
        return 0

df['Penghasilan Ayah'] = df['Penghasilan Ayah'].apply(convert_gaji_to_numeric)
df['Penghasilan Ibu'] = df['Penghasilan Ibu'].apply(convert_gaji_to_numeric)
df['Penghasilan Orang Tua'] = (df['Penghasilan Ayah'] + df['Penghasilan Ibu'])
```

Figure 4. Parent Income Attribute Modification

3. The "Penghasilan Orangtua" attribute, mentioned in point two, will be categorized into three groups: 1) "Low" for incomes below Rp 2,000,000, 2) "Medium" for incomes between Rp 2,000,000 and Rp 4,000,000, and 3) "High" for incomes above Rp 4,000,000. All values are in monthly income. The process of modifying this attribute is illustrated in Figure 5.

```
df['Rendah'] = df['Penghasilan Orang Tua'] < 2_000_000
df['Sedang'] = (df['Penghasilan Orang Tua'] >= 2_000_000) &
                                        (df['Penghasilan Orang Tua'] < 4_000_000)
df['Tinggi'] = df['Penghasilan Orang Tua'] >= 4_000_000
```

Figure 5. Merging Parent Income Attributes

4. The Achievement or “Prestasi” attribute will be categorized into two groups, "Berprestasi" or achieved and "Tidak Berprestasi" or Not Achieved and will be changed to "1" and "0," respectively. The process of modifying this attribute is illustrated in Figure 6.

```
df['Prestasi'] = df['Prestasi'].replace({'Berprestasi': '1', 'Tidak Berprestasi': '0'})
```

Figure 6. Modification of Student Achievement Attributes

5. From the results of the dataset modification above, a clearer dataset distribution can be obtained as shown in Figure 7.

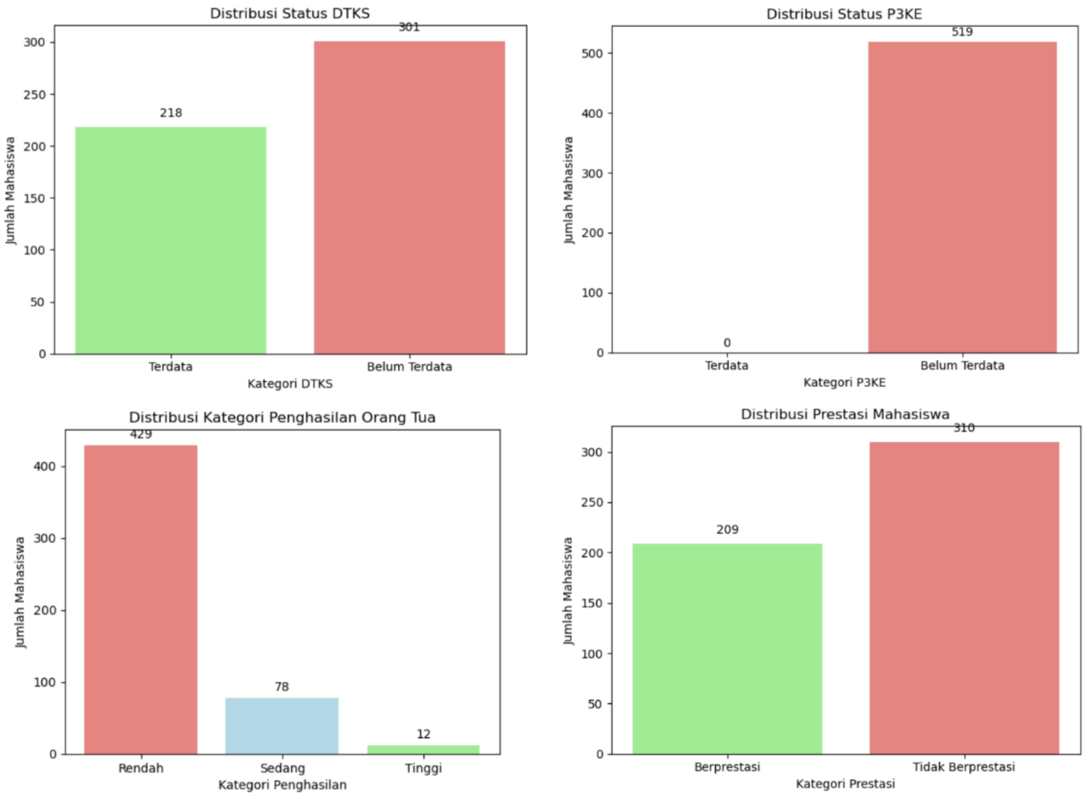

Figure 7. Dataset Distribution of Each Course KIP Attribute

So that the KIP "university" dataset used is DTKS Status, P3KE Status, Achievement, Scholarship Status, Parent Income, Low, Medium, and High.

	Nama Siswa	Status DTKS	Status P3KE	Prestasi	Status Beasiswa	Penghasilan Orang Tua	Rendah	Sedang	Tinggi
0	A. Jamal Assegap Supriadi	0	0	1	1	375000.5	True	False	False
1	Abdul Halim	0	0	0	0	1500001.0	True	False	False
2	Abielya Athaya	0	0	1	0	875000.5	True	False	False
3	Adzkya Shaqiela	0	0	1	0	875000.5	True	False	False
4	Agna Mudzkia	0	0	0	1	250000.0	True	False	False
...	...	...	...	...	...	...	...	...	...
514	YUSLAN PEBRIAN	1	0	1	0	0.0	True	False	False
515	YUSUF RAMDAN LUBIS	1	0	0	1	375000.5	True	False	False
516	Zetira Jelita Fani	1	0	0	0	0.0	True	False	False
517	ZIDANE DWI PUTRA ARIANTO	1	0	0	0	0.0	True	False	False
518	Zulfa Safinatun Naja Medina	1	0	1	0	0.0	True	False	False

519 rows × 9 columns

Figure 8. Final Attributes used in Research

3.3 Data Modelling, Evaluation and Validation

In this study, data modeling techniques, specifically Decision Tree and Support Vector Machine (SVM), are utilized for the classification of KIP university scholarship program recipients. The Decision Tree method constructs a tree-shaped structure to systematically divide the dataset based on attributes such as DTKS status, P3KE status, academic achievement, and parental income, ultimately predicting a student's eligibility for the scholarship. SVM seeks to maximize the margin between KIP university grantees and non-recipients by identifying the best hyperplane to effectively divide the two classes. To evaluate the effectiveness of the decision tree and SVM models in predicting or categorizing KIP university awardee, additional model assessment and validation are carried out. The confusion matrix approach will be used for model evaluation, and the K-Fold cross-validation technique will be used for validation.

```
Iteration        Accuracy        Precision        Recall          F1 score
---------------------------------------------------------------------------
Iteration-1      0.77885         0.73451          0.77885         0.74648
Iteration-2      0.69231         0.62012          0.69231         0.64201
Iteration-3      0.75000         0.69118          0.75000         0.65952
Iteration-4      0.81731         0.76442          0.81731         0.75167
Iteration-5      0.88350         0.84510          0.88350         0.86035

Decision Tree Model Evaluation Scores
--------------------------------------
Accuracy  : 0.78439
Precision : 0.73107
Recall    : 0.78439
F1 score  : 0.73201
```

Figure 9. Model Evaluation Results with Decision Tree

Figure 9 shows the results of modeling and validation of the decision tree model using the K-Fold cross-validation method. The results of K-Fold Cross-Validation for the Decision Tree model in the case of KIP Lecture recipient classification can be explained as follows:

1. Accuracy: The accuracy value of 0.78439 indicates that the Decision Tree model is able to correctly predict about 78.44% of all cases in the test dataset. Accuracy measures the extent to which the model can correctly classify data on KIP Lecture recipients.
2. Precision: The precision of 0.73107 reflects the model's ability to correctly identify the College KIP recipients from all predicted positives (True Positives + False Positives). This indicates that about 73.11% of the model predictions are correct.

3. Recall (sensitivity level): The recall value of 0.78439 or about 78.45% indicates that the model has the ability to identify a large number of KIP university recipients out of all predicted positives (True Positives + False Negatives).
4. F1 Score: F1 score of 0.73201 is the harmonic mean of precision and recall. This gives an overall picture of the model's performance of 73.20%.

Iteration	Accuracy	Precision	Recall	F1 score
Iteration-1	0.79808	0.83885	0.79808	0.70845
Iteration-2	0.74038	0.80778	0.74038	0.62994
Iteration-3	0.75000	0.81250	0.75000	0.64286
Iteration-4	0.81731	0.85068	0.81731	0.73514
Iteration-5	0.90291	0.91234	0.90291	0.85685

SVM Model Evaluation Scores

Accuracy : 0.80174
Precision : 0.84443
Recall : 0.80174
F1 score : 0.71465

Figure 10. SVM Model Evaluation Result

Figure 10 shows the results of modeling and validation of the SVM model using the K-Fold cross-validation method. The results of K-Fold Cross-Validation for the SVM model in the case of KIP Lecture recipient classification can be explained as follows:

1. Accuracy: The accuracy value of 0.80174 indicates that the SVM model is able to correctly predict about 80.17% of all cases in the test dataset. Accuracy measures the extent to which the model can correctly classify data on KIP university recipients.
2. Precision: The precision of 0.84443 reflects the model's ability to correctly identify the College KIP recipients from all predicted positives (True Positives + False Positives). This indicates that about 84.44% of the model predictions are correct.
3. Recall (sensitivity level): The recall value of 0.80174 or about 80.17% indicates that the model has the ability to identify a large number of KIP university recipients out of all predicted positives (True Positives + False Negatives).
4. F1 Score: F1 score of 0.71465 is the harmonic mean of precision and recall. This gives an overall picture of the model's performance of 71.46%.

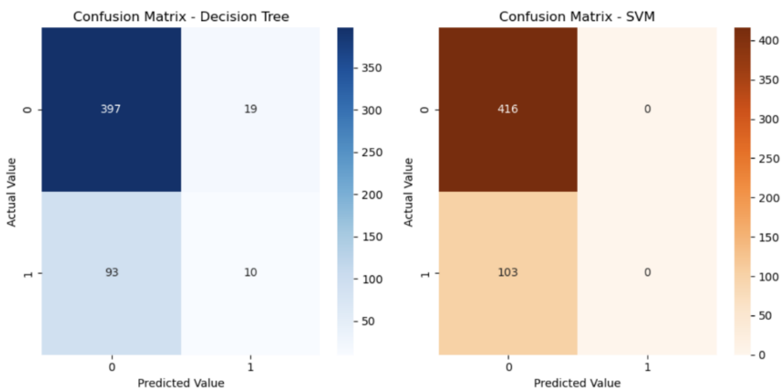

Figure 11. Comparison of Confusion Matrix Results between Decision Tree and SVM

The outcomes of the classification evaluation or the predictions made by the SVM and decision tree models are shown visually in Figure 11 above as the Confusion Matrix findings. The real class or category of the examined data is indicated by the label "Actual" (horizontal). Conversely, the class or category that the model has predicted based on the provided data is indicated by the "Predicted" label (vertical). The decision tree model accurately predicts 397 data as category 0 (True Negative) based on the confusion matrix results for all samples that truly belong to category 0 (Actual 0). However, the model also incorrectly predicts 19 data as category 1 (False Positive). While the model accurately predicts 10 data as category 1 (True Positive) out of all samples that actually belong to category 1 (Actual 1), it mistakenly predicts 93 data as category 0 (False Negative). Conversely, the model accurately predicts all 416 data points as category 0 (True Negative) based on the confusion matrix results for all samples that are actually included in category 0 (Actual 0). The algorithm incorrectly predicts 103 data as category 0 (False Negative) out of all samples that truly belong to category 1 (Actual 1) instead of accurately predicting any data as category 1 (True Positive).

Table 2 Comparison of Decision Tree and SVM Performance

Classification Algorithm	Evaluation using Confusion Matrix				Validation using K-Fold Cross-Validation (5 iterations)			
	TN	FP	FN	TP	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Decision Tree	397	19	93	10	78.44	73.11	78.44	73.20
SVM	416	0	103	0	80.17	84.44	80.17	71.46

4. Conclusion

Overall, the performance of the Support Vector Machine (SVM) model in classifying KIP Lecture scholarship recipients looks slightly superior compared to the decision tree model. As seen in Table 2, the SVM model achieved a higher accuracy rate of 80.17%, compared to the accuracy of the decision tree model which only reached 78.44%. Similarly, SVM is superior in terms of precision with a score of 84.44%, while the decision tree model has a precision of 73.11%. In addition, both models showed similar recall values to their accuracy values, indicating their ability to correctly identify KIP university recipients from the dataset. Overall, when considering the F1 score, which combines precision and recall, the SVM model achieved a score of 71.46%, while the Decision Tree model had an F1 score of 73.20%. Thus, although the SVM model performed slightly better in terms of accuracy and precision, both models showed competitive performance in performing classification for this case study.

References

[1] Kemendikbud. PERATURAN MENTERI PENDIDIKAN DAN KEBUDAYAAN REPUBLIK INDONESIA NOMOR 10 TAHUN 2020 TENTANG PROGRAM INDONESIA PINTAR [Internet]. Jakarta; 2020. Available from: jdih.kemdikbud.go.id

[2] U. Usanto, L. Nurlaela, A. Sopian, and F. Alfiah, "Umrah Registration System Using Extreme Programming Method Towards Worship Tourism," Int. J. Cyber IT Serv. Manag., vol. 3, no. 1, pp. 22–31, 2023.

[3] Iskandar A. Sistem Pendukung Keputusan Kelayakan Penerima Bantuan Dana KIP university Menggunakan Metode ROC-EDAS. Build Informatics, Technol Sci. 2022;4(2):856–64.

[4] D. Jonas, E. Maria, I. R. Widiasari, U. Rahardja, and T. Wellem, "Design of a TAM Framework with Emotional Variables in the Acceptance of Health-based IoT in Indonesia," ADI J. Recent Innov., vol. 5, no. 2, pp. 146–154, 2024.

[5] Firmansyah S, Gaol J, Susilo SB. Perbandingan Klasifikasi SVM dan Decision Tree untuk Pemetaan Mangrove Berbasis Objek Menggunakan Citra Satelit Sentinel-2B di Gili Sulat, Lombok Timur. J Pengelolaan Sumberd Alam dan Lingkung (Journal Nat Resour Environ Manag. 2019;9(3):746–57.

[6] I. Khong, N. A. Yusuf, A. Nuriman, and A. B. Yadila, "Exploring the Impact of Data Quality on Decision-Making Processes in Information Intensive Organizations," APTISI Trans. Manag., vol. 7, no. 3, pp. 253–260, 2023.

[7] Susetyoko R, Wiratmoko Yuwono, Elly Purwantini. Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP university Menggunakan Regresi Logistik Biner. J Inform Polinema. 2022;8(4):31–40.

[8] K. Diantoro, D. Supriyanti, Y. P. A. Sanjaya, and S. Watini, "Implications of Distributed Energy Development in Blockchain-Based Institutional Environment," Aptisi Trans. Technopreneursh., vol. 5, no. 2sp, pp. 209–220, 2023.

[9] Astuti LW, Saluza I, Alie MF. Optimalisasi Klasifikasi Kanker Payudara Menggunakan Forward Selection Pada Naive Bayes. J Ilm Inform Glob [Internet]. 2020;11(2):63–7. Available from: https://archive.ics.uci.edu/ml/machine-learning-

[10] T. Handra and V. P. K. Sundram, "The Effect of Human Resource Information Systems (HRIS) and Artificial Intelligence on Defense Industry Performance," IAIC Trans. Sustain. Digit. Innov., vol. 4, no. 2, pp. 155–163, 2023.

[11] Zhao W, Lai X, Liu D, Zhang Z, Ma P, Wang Q, et al. Applications of Support Vector Machine in Genomic Prediction in Pig and Maize Populations. Front Genet. 2020;11(December):1–7.

[12] W. Sejati and V. Melinda, “Education on the Use of Iot Technology for Energy Audit and Management Within the Context of Conservation and Efficiency,” Int. Trans. Educ. Technol., vol. 1, no. 2, pp. 138–143, 2023.
[13] Hozairi H, Anwari A, Alim S. Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. Netw Eng Res Oper. 2021;6(2):133.
[14] R. Widayanti, M. H. R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, “Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering,” J. Appl. Data Sci., vol. 4, no. 3, pp. 289–302, 2023.
[15] Pradnyana GA, Darmawiguna IGM, Suditresna Jaya DKS, Sasmita A. Performance analysis of support vector machines with polynomial kernel for sentiment polarity identification: A case study in lecturer’s performance questionnaire. J Phys Conf Ser. 2021;1810(1):1–9.
[16] Wahyuningsih S, Utari DR. Perbandingan Metode K-Nearest Neighbor , Naive Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit. Konf Nas Sist Inf 2018 STMIK Atma Luhur Pangkalpinang, 8 – 9 Maret 2018. 2018;619–23.
[17] Apriyani H, Kurniati K. Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. J Inf Technol Ampera. 2020;1(3):133–43.
[18] U. Rahardja, Q. Aini, P. A. Sunarya, D. Manongga, and D. Julianingsih, “The Use of TensorFlow in Analyzing Air Quality Artificial Intelligence Predictions PM2. 5,” Aptisi Trans. Technopreneursh., vol. 4, no. 3, pp. 313–324, 2022.
[19] Khotimah K. Teknik Data Mining menggunakan Algoritma Decision Tree (C4.5) untuk Prediksi Seleksi Beasiswa Jalur KIP pada Universitas Muhammadiyah Kotabumi. J SIMADA (Sistem Inf dan Manaj Basis Data). 2022;4(2):145–52.
[20] Puslapdik. PEDOMAN PENDAFTARAN KARTU INDONESIA PINTAR university KIP university MERDEKA 2023 [Internet]. Jakarta; 2023. Available from: https://kip-university.kemdikbud.go.id/uploads/BsImnu09yFOxop5dfJAwkaRleMTUqP_tgl20200412205459.pdf
[21] Suganda G, Asfi M, Subagio RT, Kusuma RP. Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (Kip) university Menggunakan Naïve Bayes Classifier. JSiI (Jurnal Sist Informasi). 2022;9(2):193–9.
[22] Sathiyanarayanan P, Pavithra S, Sai Saranya M, Makeswari M. Identification of breast cancer using the decision tree algorithm. In: 2019 IEEE International Conference on System, Computation, Automation and Networking, ICSCAN 2019. IEEE; 2019. p. 1–6.
[23] Suwitono YA, Kaunang FJ. Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Daun Dengan Metode Data Mining SEMMA Menggunakan Keras. J Komtika (Komputasi dan Inform. 2022;6(2):109–21.
[24] Latifah R, Wulandari ES, Kreshna PE. Model Decision Tree Untuk Prediksi Jadwal Kerja Menggunakan Scikit-Learn. J Univ Muhammadiyah Jakarta [Internet]. 2019;1–6. Available from: https://jurnal.umj.ac.id/index.php/semnastek/article/download/5239/3517
[25] Puslapdik. PEDOMAN PENDAFTARAN KARTU INDONESIA PINTAR university KIP university MERDEKA 2023 [Internet]. Jakarta; 2023. Available from: https://kip-university.kemdikbud.go.id/uploads/BsImnu09yFOxop5dfJAwkaRleMTUqP_tgl20200412205459.pdf