



# Perbandingan Algoritma Decision Tree dan Support Vector Machine Dalam Pemilihan Calon Mahasiswa Penerima KIP-K

Nayaka Al Syahreal Kanaka<sup>\*</sup>, Rudi Heriansyah, Shinta Puspasari

Fakultas Ilmu Komputer, Prodi Teknik Informatika, Universitas Indo Global Mandiri, Palembang, Indonesia

Email: <sup>1\*</sup>nayakaalsyahrealkanaka@gmail.com, <sup>2</sup>rudi@uigm.ac.id, <sup>3</sup>shinta@uigm.ac.id

Email Penulis Korespondensi: nayakaalsyahrealkanaka@gmail.com

**Abstrak**–KIP Kuliah adalah bantuan biaya pendidikan dari pemerintah bagi lulusan SMA/ sederajat yang memiliki potensi akademik yang baik tetapi memiliki keterbatasan ekonomi. Ada banyak hal yang patut dipertimbangkan oleh Perguruan Tinggi sebelum memilih calon mahasiswa penerima KIP Kuliah sehingga dapat dilakukan seleksi menggunakan machine learning dan algoritma klasifikasi. Pada penelitian ini, akan digunakan dua algoritma machine learning diantaranya: Decision Tree dan Support Vector Machine (SVM). Selanjutnya, kedua algoritma ini akan diuji coba dan dibandingkan hasil akhirnya. Kedua algoritma memiliki hasil yang berbeda-beda. Tingkat accuracy, precision, recall, dan F1 score yang terbesar adalah 100%. Nilai tersebut dapat dicapai oleh algoritma Decision Tree karena dataset yang digunakan cocok untuk diselesaikannya. Oleh sebab itu, algoritma Decision Tree direkomendasikan untuk digunakan dalam memilih calon mahasiswa KIP Kuliah.

**Kata Kunci:** KIP Kuliah; Perbandingan; Machine Learning; Supervised Learning; Seleksi

**Abstract**–KIP Kuliah is tuition assistance from the government for high school / equivalent graduates who have good academic potential but have economic limitations. There are many things that should be considered by universities before selecting prospective students who receive KIP Lecture so that selection can be done using machine learning and classification algorithms. In this research, two machine learning algorithms will be used including: Decision Tree and Support Vector Machine (SVM). Furthermore, these two algorithms will be tested and compared the final results. Both algorithms have different results. The highest level of accuracy, precision, recall, and F1 score is 100%. This value can be achieved by the Decision Tree algorithm because the dataset used is suitable for it to solve. Therefore, the Decision Tree algorithm is recommended to be used in selecting KIP College student candidates.

**Keywords:** College KIP; Comparison; Machine Learning; Supervised Learning; Selection

## 1. PENDAHULUAN

Pada tahun 2022, Kementerian Pendidikan, Kebudayaan & Riset Teknologi Republik Indonesia (Kemendikbud Ristek) memberikan bantuan pendidikan tingkat Perguruan Tinggi (PT) kepada 185.000 calon mahasiswa yang ada Indonesia melalui program Kartu Indonesia Pintar Kuliah (KIP Kuliah). Dilansir dari portal media Kompas.com, “beasiswa ini diperuntukan bagi mahasiswa dari keluarga miskin maupun rentan miskin” (Kasih, 2022).

Berdasarkan website resmi KIP Kuliah, persyaratan untuk mendaftar program tersebut adalah siswa SMA/ sederajat yang akan lulus pada tahun berjalan maksimal 2 tahun sebelumnya, memiliki potensi akademik namun terbatas secara ekonomi, memiliki Kartu Indonesia Pintar (KIP) atau Kartu Keluarga Sejahtera (KKS) atau terdata di Data Terpadu Kesejahteraan Sosial (DTKS) Kementerian Sosial (Kemensos), dan lulus seleksi penerimaan mahasiswa baru di Perguruan Tinggi, baik Perguruan Tinggi Negeri (PTN) atau Perguruan Tinggi Swasta (PTS).

Skema penyeleksian calon mahasiswa KIP Kuliah di PTN berbeda dari PTS. PTN menyelenggarakan seleksi umum yang diikuti oleh semua calon mahasiswa, setelah itu baru melakukan validasi kepada calon mahasiswa KIP Kuliah. Sementara PTS langsung memisahkan calon mahasiswa reguler dan KIP Kuliah. Kemudian, melakukan seleksi administrasi dan wawancara bersama prodi kepada calon mahasiswa KIP Kuliah (Kasih, 2022). Ada banyak hal yang patut dipertimbangkan sebelum memilih calon mahasiswa penerimanya sehingga calon mahasiswa KIP Kuliah harus diseleksi. Hal ini dilakukan untuk mengetahui apakah calon mahasiswa berpotensi lulus atau tidak? Proses seleksi ini akan menjadi acuan bagi PTS untuk menerima calon mahasiswa KIP Kuliah. Dalam penerapannya, *machine learning* dapat digunakan sebagai alat untuk melakukan seleksi calon mahasiswa (Hakkoum et al., 2023; Salais-Fierro et al., 2020; Tsang & Benoit, 2020).

Sebelum mengimplementasikan *machine learning*, pada tahap awal dilakukan *data collecting* atau pengumpulan data penelitian. Wawancara dilakukan bersama salah satu pegawai BKABK UIGM untuk mengetahui variabel yang akan digunakan pada penelitian. Semua data yang ada dijadikan dalam sebuah *dataset*, kemudian dipersiapkan dalam tahap *data preparation*. Tahap ini mencakup semua kegiatan yang digunakan untuk membuat kumpulan data dalam tahap pemodelan, seperti pembersihan, menggabungkan, dan mengubah data menjadi variabel yang lebih berguna. Tahap ini dimulai dengan data cleansing, ini adalah proses memperbaiki atau menghilangkan data yang keliru, rusak, tidak sesuai format, duplikat atau tidak sempurna dalam *dataset* (Tableu, 2022).

Tahap selanjutnya adalah *data splitting*. Pada tahap ini akan dilakukan pembagian data menjadi dua bagian atau lebih dan memisahkan antara *feature* dan *labels* yang akan didefinisikan sebagai X (*feature*) dan y (*labels*). Setelah melakukan *data splitting*, proses implementasi algoritma dapat dilakukan. Algoritma *machine learning* yang akan diujikan adalah *Decision Tree* dan *Support Vector Machine (SVM)*. Penelitian ini bertujuan untuk mengetahui algoritma *machine learning* yang efektif dalam memilih calon mahasiswa KIP Kuliah.

*Machine learning* atau pembelajaran mesin adalah bagian dari artificial intelligence yang merupakan perpaduan dari beberapa disiplin ilmu pengetahuan yang kompleks. Arthur Samuel memperkenalkan istilah *machine learning* pada

tahun 1959 melalui jurnalnya yang berjudul “Some studies in *Machine Learning* using the Game of Checkers” (Sarno et al., 2022).

*Machine learning* menggunakan beberapa algoritma berbeda yang dijalankan berulang kali agar komputer dapat belajar dari informasi yang diterima. Komputer akan diberikan input berupa data secara terus menerus, sedangkan algoritma akan “berubah” sesuai dengan informasi yang didapatkan. Proses ini meningkatkan kemampuan komputer dalam berpikir dan mampu memahami data input.

*Decision tree* atau pohon keputusan merupakan salah satu metode klasifikasi yang umum digunakan. Metode ini mengubah fakta yang besar menjadi pohon keputusan agar dapat merepresentasikan aturan. Aturan tersebut dapat dengan mudah diinterpretasi oleh manusia. *Decision tree* juga berguna untuk mengeksplorasi data serta menemukan hubungan tersembunyi antara beberapa variabel input dan sebuah variabel target (Berry & Linoff, 2004).

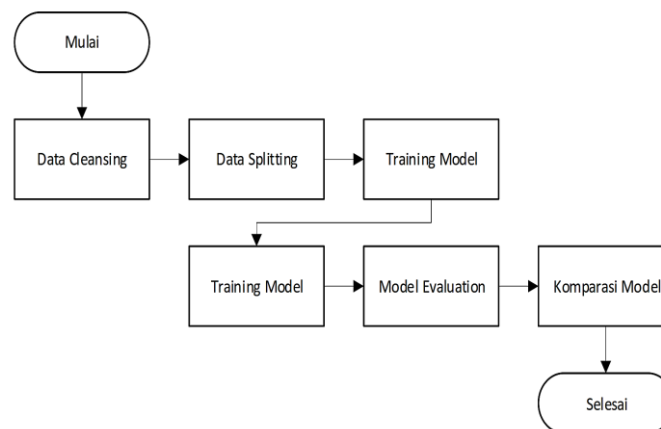
*Support Vector Machine (SVM)* diperkenalkan pertama kali oleh Vapnik tahun 1992 sebagai salah satu metode *machine learning* yang bekerja dengan prinsip Structural Risk Minimization (SRM) (Izzudin et al., 2020; Park et al., 2020; Sueno, 2020). Metode ini bertujuan untuk menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. *SVM* menggunakan hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur yang berdimensi tinggi dengan mengimplementasikan learning bias yang berasal dari teori pembelajaran statistik. *SVM* merupakan salah satu teknik yang relatif baru dibandingkan dengan teknik-teknik lain pada umumnya sehingga mempunyai performa yang lebih baik di berbagai bidang aplikasi. Tingkat akurasi pada model akan dihasilkan oleh proses peralihan dengan *SVM* sangat bergantung terhadap fungsi kernel dan parameter yang digunakan (Siagian, 2011). Penelitian ini bertujuan untuk mengetahui algoritma *machine learning* yang efektif dalam memilih calon mahasiswa KIP Kuliah. Dalam mencari algoritma yang efektif, dilakukan pengujian dengan confusion matrix. Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Pada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)* (Rasywir et al., 2020).

Nilai *True Negative (TN)* merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive (TP)* merupakan data positif yang terdeteksi benar. *False Negative (FN)* merupakan kebalikan dari *True Positive*, dimana hasilnya berupa data positif, namun terdeteksi sebagai data negatif (Visa et al., 2011). Harapannya, algoritma yang efektif dapat menyederhanakan proses seleksi calon mahasiswa KIP Kuliah dan memudahkan proses untuk menentukan calon mahasiswa KIP Kuliah yang diterima.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Dasar Penelitian

Dari gambar 1 berikut merupakan tahapan dari penelitian yang penulis lakukan.



**Gambar 1.** Skema dan tahapan penelitian

#### 2.1.1 Data Collecting

Pada tahap awal, dilakukan *data collecting* atau pengumpulan data penelitian. Proses pengumpulan data dimaksudkan untuk mendapatkan bekal studi pendahuluan tentang inti masalah yang akan diselesaikan (Wali & Ahmad, 2018). Sumber data didapatkan melalui wawancara bersama salah satu pegawai BKABK UIGM untuk mengetahui variabel yang akan digunakan pada penelitian

Penelitian ini menggunakan data KIP Kuliah UIGM tahun 2020-2022. Data penelitian yang diperoleh adalah 500 data dengan 13 variabel. Adapun variabel yang digunakan dapat dilihat pada Tabel 1.

**Tabel 1.** Parameter yang ada dalam *dataset*

No	Variabel	Tipe Data
1	Nilai Rata-Rata	Int64
2	Pendapatan Orang Tua	Int64
3	Jumlah Anggota Keluarga	Int8
4	Biaya per-Kepala	Int64
5	Kondisi Ayah	Boolean
6	Kondisi Ibu	Boolean
7	Kelayakan Rumah	Boolean
8	Kesesuaian PBB	Boolean
9	Prestasi	Int32
10	Organisasi	Int32
11	Nilai Kecocokan Prodi 1	Int32
12	Nilai Kecocokan Prodi 2	Int32
13	Status Kelulusan	Boolean

### 2.1.2 Data Preparation

Data *preparation* adalah tahap mempersiapkan data, mencakup semua kegiatan yang digunakan untuk membuat kumpulan data dalam tahap pemodelan, seperti pembersihan, menggabungkan, dan mengubah data menjadi variabel yang lebih berguna (Abdallah et al., 2016).

Proses ini dimulai dengan data cleansing, ini adalah proses pemeriksaan anomali pada data agar data dapat digunakan dengan baik sebagai *dataset*. Adapun langkah yang dilakukan adalah melakukan penelusuran data yang kosong (*null*) dan data yang terduplikasi. Diawali dengan menentukan terlebih dahulu data modus pada setiap *feature* yang dimiliki oleh *dataset*. Kemudian, apabila modus pada setiap *feature* yang dimiliki oleh *dataset* telah ditemukan, alur proses selanjutnya adalah penelusuran data yang *null*. Penelusuran data dilakukan dengan iterasi pada setiap baris *dataset* dan dilanjutkan iterasi pada setiap *feature* pada baris tersebut. Apabila ditemukan data yang *null* pada *feature* di baris tersebut, *feature* yang *null* akan diisi dengan data modus yang sebelumnya telah ditemukan. Iterasi terus dilakukan hingga baris terakhir pada *dataset* tersebut. Selanjutnya penelusuran pada baris data yang terdapat duplikasi. Hal ini bertujuan untuk menghasilkan data unik pada setiap baris yang akan digunakan sebagai *dataset* (Ridzuan & Zainon, 2019).

### 2.1.3 Data Splitting

Tahap berikutnya adalah *data splitting*. Pada tahap ini akan dilakukan pembagian data menjadi dua bagian. Bagian itu meliputi data *training* dan data *testing*, setelah itu dilakukan pemisahan antara *feature* dan *labels* yang akan didefinisikan sebagai X (*feature*) dan y (*labels*) (Nurhopipah & Hasanah, 2020).

### 2.1.4 Training Model

Algoritma *Decision Tree* dan *Support Vector Machine (SVM)* akan diimplementasikan pada *dataset* yang sebelumnya telah melalui proses *data splitting*.

### 2.1.5 Model Evaluation

Confusion Matrix adalah metode yang umumnya digunakan untuk melakukan pengukuran akurasi pada konsep data mining. Pada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*. Nilai *True Negative (TN)* merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive (TP)* merupakan data positif yang terdeteksi benar. *False Negative (FN)* merupakan kebalikan dari *True Positive*, dimana hasilnya berupa data positif, namun terdeteksi sebagai data negatif (Visa et al., 2011). Pada jenis klasifikasi binary yang hanya memiliki 2 keluaran kelas, *confussion matrix* dapat disajikan seperti pada Gambar 2 berikut.

Reference	Prediction	
	Positive	Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Negative)	TN (True Negative)

**Gambar 2.** Tabel *Confussion Matrix*

Berdasarkan nilai *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), dan *True Positive* (TP) dapat diperoleh nilai akurasi (accuracy), presisi (precision), sensitivitas (*recall*), dan *f1-score*.

Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan persamaan 1.

Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar lalu dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan persamaan 2. Sementara itu, *recall* menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. *Recall* dikenal sebagai sensitivitas, nilai *recall* diperoleh dengan persamaan 3. *F1-Score* merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif. *F1-score* dapat dilihat pada persamaan 4.

$$Akurasi(Accuration) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{1}$$

$$Presisi(Precision) = \frac{TP}{TP+FP} \times 100\% \tag{2}$$

$$Recall(Sensitivitas) = \frac{TP}{TP+FN} \times 100\% \tag{3}$$

$$F1 - Score = \frac{TN}{TN+TP} \times 100\% \tag{4}$$

Keterangan:

1. TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
2. TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
3. FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
4. FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Gambar 3 adalah ilustrasi contoh dari hasil *Confussion Matrix*:

True label	False	TP	FN
	True	FP	TN
		False	True
		Predicted Label	

**Gambar 3.** Ilustrasi *confussion matrix*

### 2.1.6 Komparasi Efektifitas

Setelah mengimplementasikan algoritma, akan dilakukan perbandingan pada kedua hasil dari algoritma tersebut menggunakan *confussion matrix*. Adapun matriks yang diukur meliputi *accuracy*, *precision*, *recall*, dan *F1-score* (Patimah et al., 2021).

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data Cleansing

Pada tahap ini, dilakukan pemeriksaan apakah terdapat data yang kosong (*null*) pada *dataset*. Dari pemeriksaan yang telah dilakukan, tidak ditemukan data kosong dalam *dataset*. Namun jika ada data yang kosong, maka akan diisi dengan data modus yang terdapat pada kolom tersebut.

**Tabel 2.** Jumlah data kosong pada *dataset*

Jumlah Data yang Kosong	Jumlah
Nilai Rata-Rata	0
Pendapatan Orang Tua	0
Jumlah Anggota Keluarga	0
Biaya per-Kepala	0
Kondisi Ayah	0
Kondisi Ibu	0
Kelayakan Rumah	0
Kesesuaian PBB	0
Prestasi	0



Jumlah Data yang Kosong	Jumlah
Organisasi	0
Nilai Kecocokan Prodi 1	0
Nilai Kecocokan Prodi 2	0
Status Kelulusan	0

Selanjutnya dilakukan pemeriksaan duplikasi pada data, hasilnya 0 yang berarti tidak ditemukan data duplikat dalam *dataset*. Jika ditemukan data yang duplikat, maka salah satu dari datanya akan dihapus Berikut adalah hasil dari pencarian data duplikat pada *dataset*.

### 3.2 Data Splitting

Proses *data splitting* dilakukan pembagian data menjadi dua bagian. Bagian itu meliputi data *training* dan data *testing*. Data *testing* berisi 20% data acak dari keseluruhan *dataset*. Setelah itu dilakukan pemisahan antara *feature* dan *labels* yang akan didefinisikan sebagai X (*feature*) dan y (*labels*). *Data splitting* menghasilkan *dataset* dengan *feature* tanpa label sebagai berikut.

**Tabel 3.** Tampilan *feature dataset training*

	Rata-rata	Pendapatan Orang Tua	Jumlah Anggota Keluarga	Biaya per-kepala	Kon-disi Ayah	Kon-disi Ibu	Kelaa-yakan Rumah	Kese-suaian PBB	Pres-tasi	Orga-nisasi	Prodi 1	Prodi 2
100	84	3900000	4	980000	1	0	1	0	4	2	7	7
101	94	2900000	4	730000	0	0	1	0	3	5	10	5
102	80	2600000	5	520000	1	1	1	1	1	0	7	7
103	93	3500000	4	880000	0	1	0	1	3	3	7	8
104	90	3700000	6	620000	1	0	0	1	7	0	9	10
...	...	...	...	...	...	...	...	...	...	...	...	...
495	90	2100000	6	350000	1	1	1	1	8	9	10	6
496	95	2300000	5	460000	0	0	0	0	8	0	5	10
497	78	3100000	3	1040000	0	1	1	1	7	8	5	9
498	90	3000000	4	750000	0	1	1	0	3	3	9	6
499	88	3000000	4	750000	1	0	0	1	10	10	8	7

### 3.3 Training Model

Data yang telah melalui proses *splitting*, selanjutnya akan diimplementasikan pada model *decision tree* dan *SVM*. Pada model *decision tree*, digunakan beberapa parameter dalam tabel berikut.

**Tabel 4.** Parameter pada model *decision tree*

Parameter	Input
Criterion	“gini”
Random State	12
Max Depth	2
Min Sample Leaf	13

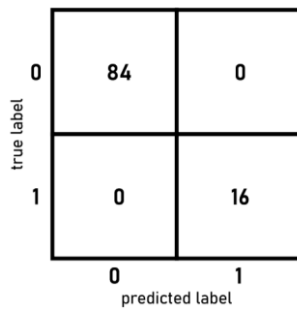
Kemudian, setelah mengimplementasikan algoritma *decision tree*, dilanjutkan dengan mengimplementasikan algoritma *SVM* dengan beberapa parameter sebagai berikut.

**Tabel 5.** Parameter pada model *SVM*

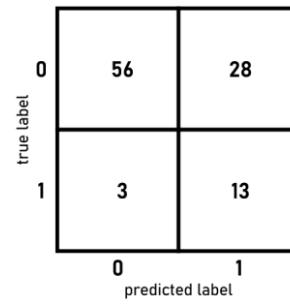
Parameter	Input
Kernel	“rbg”

### 3.4 Model Evaluation

Setelah semua model telah melalui proses *training*, selanjutnya akan dilakukan uji coba menggunakan data *testing* yang sebelumnya telah disiapkan. Adapun tahap uji coba menggunakan metode *confussion matrix*. Berdasarkan hasil uji coba, diketahui model *decision tree* dan *SVM* menghasilkan nilai *confussion matrix* seperti pada Gambar 2.



a) Hasil *confussion matrix* model *decision tree*



b) Hasil *confussion matrix* model *SVM*

**Gambar 4.** *Confussion matrix* pada masing-masing model *machine learning*

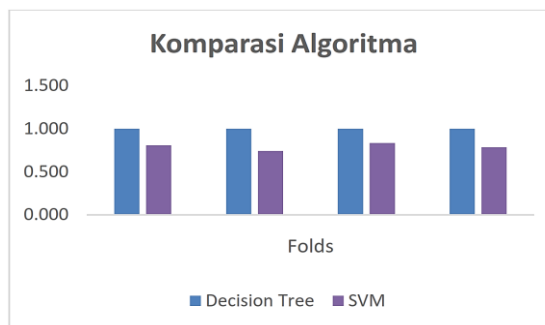
Hasil *confussion matrix* sebelumnya akan dilakukan perhitungan untuk menghasilkan nilai-nilai yang menjadi perbandingan yaitu *accuracy*, *recall*, *precision*, dan *F1-Score*.

### 3.5 Komparasi Model

Setelah dilakukan pengujian pada masing-masing algoritma. Selanjutnya akan dilakukan komparasi model terbaik dari masing-masing algoritma sebelumnya yang telah diujikan. Berikut adalah tabel 6 perbandingan model algoritma.

**Tabel 6.** Tabel perbandingan model algoritma

Indikator	Decision Tree	SVM
Accuracy	1.000	0.810
Precision	1.000	0.744
Recall	1.000	0.833
F1-Score	1.000	0.786



**Gambar 5.** Grafik perbandingan model algoritma *decision tree* & *SVM*

Pada uji coba *confussion matrix*, diketahui apabila sebuah skor menghasilkan nilai mendekati atau sama dengan 1, maka dapat diketahui bahwa nilai tersebut adalah nilai terbaik. Dari hasil pengujian sebelumnya, diketahui.

## 4. KESIMPULAN

Berdasarkan hasil uji coba dan perbandingan sebelumnya, diketahui terdapat dua algoritma yang telah dibandingkan, yaitu Decision Tree dan SVM. Dalam proses perbandingan tersebut, berbagai metrik telah dievaluasi, yaitu akurasi, presisi, recall, dan skor F1. Hasil perbandingan menunjukkan bahwa algoritma Decision Tree telah mencapai hasil paling efektif, dengan mencapai akurasi, presisi, recall, dan F1-Score sebesar 100%. Pencapaian ini menunjukkan kemampuan algoritma Decision Tree dalam menyelesaikan studi kasus seleksi calon mahasiswa penerima KIP Kuliah. Dengan mencapai nilai sebesar 100%, hal ini menunjukkan bahwa algoritma Decision Tree mampu melakukan prediksi calon mahasiswa yang layak menerima KIP Kuliah dengan sempurna. Namun, perlu diketahui bahwa hasil ini mungkin dapat dipengaruhi oleh berbagai faktor, termasuk data pelatihan, parameter yang disesuaikan, maupun karakteristik dari studi kasus ini. Oleh karena itu, meskipun algoritma Decision Tree telah menunjukkan hasil yang sangat baik dalam uji coba ini, kedepannya dapat dilakukan penelitian pada kasus serupa namun dengan kondisi berbeda untuk mendapatkan perspektif yang lebih luas.

## REFERENCES

Abdallah, Z. S., Du, L., & Webb, G. I. (2016). *Data Preparation* (C. Sammut & G. I. Webb (eds.); p. 318–327 BT–Encyclopedia of Machine Learning and Data Mining). Humana Press . <https://doi.org/10.1007/978-1-4899-7502->



7\_62-1

- Berry, M. J. ., & Linoff, G. S. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*.
- Hakkoum, H., Idri, A., Abnane, I., & Fernades-Aleman, J. (2023). Does Categorical Encoding Affect the Interpretability of a Multilayer Perceptron for Breast Cancer Classification? *Proceedings Ofthe 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, *Data*, 351–358. <https://doi.org/10.5220/0012084800003541>
- Izzudin, Hamzah, Nisrafariza, & Idris. (2020). ANALYSIS OF MULTISPECTRAL IMAGERY FROM UNMANNED AERIAL VEHICLE (UAV) USING OBJECT-BASED IMAGE ANALYSIS FOR DETECTION OF GANODERMA DISEASE IN OIL PALM. *Journal of Oil Palm Research*, *5*(5), 543–546.
- Kasih, A. P. (2022). *KIP Kuliah 2022 Buka 185.000 Kuota, Kondisi Rumah Mahasiswa Akan Dicek*.
- Nurhopipah, A., & Hasanah, U. (2020). Dataset Splitting Techniques Comparison For Face Classification on CCTV Images. *Indonesia Journal of Computing and Cybernetics Systems*, *14*, 341–352.
- Park, K., Hong, J. S., & Kim, W. (2020). A Methodology Combining Cosine Similarity with Classifier for Text Classification. *Applied Artificial Intelligence*, *34*(5), 396–411. <https://doi.org/10.1080/08839514.2020.1723868>
- Patimah, N. F., Abdurrohman, M., Rinaldi, A. R., & Dikananda, A. R. (2021). Implementasi Algoritma Naive Bayes dam Klasifikasi Penyakit Diabetes. *Jurnal Data Science & Informatika*, *1*, 6–10.
- Rasywir, E., Sinaga, R., & Pratama, Y. (2020). Analisis dan Implementasi Diagnosis Penyakit Sawit dengan Metode Convolutional Neural Network ( CNN ). *Jurnal Paradigma UBSI*, *22*(2), 117–123.
- Ridzuan, F., & Zainon, W. M. N. (2019). A Review on Data Cleansing Methods for Big Data. *Procedia Computer Science*, *161*, 731–738. <https://doi.org/10.1016/j.procs.2019.11.177>
- Salais-Fierro, T. E., Saucedo-Martinez, J. A., Rodriguez-Aguilar, R., & Vela-Haro, J. M. (2020). Demand prediction using a soft-computing approach: A case study of automotive industry. *Applied Sciences (Switzerland)*, *10*(3). <https://doi.org/10.3390/app10030829>
- Sarno, R., Sabilla, S. I., Purbawa, D. P., Khah, M., & Ardani, H. S. (2022). *Machine Learning dan Deep Learning- Konsep dan Pemrograman Python*. Penerbit Andi.
- Siagian, R. Y. (2011). *Klasifikasi Parket Kayu Jati Menggunakan Metode Support Vector Machines (SVM)*. Universitas Gunadarma.
- Sueno, H. T. (2020). Multi-class Document Classification using Support Vector Machine (SVM) Based on Improved Naïve Bayes Vectorization Technique. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(3), 3937–3944. <https://doi.org/10.30534/ijatcse/2020/216932020>
- Tableu. (2022). *Guide To Data Cleaning: Definition, Benefits, Components, And How To Clean Your Data*.
- Tsang, W. K., & Benoit, D. F. (2020). Gaussian processes for daily demand prediction in tourism planning. *Journal of Forecasting*, *39*(3), 551–568. <https://doi.org/10.1002/for.2644>
- Visa, S., Ramsay, B., Ralescu, A., & Knaap, E. V. D. (2011). Confusion Matrix-based Feature Selection. *Proceedings of The 22nd Midwest Artificial Intelligence and Cognitive Science*.
- Wali, M., & Ahmad, L. (2018). Perancangan Access Open Journal System (AOJS) dengan menggunakan Framework Codeigniter dan ReactJs. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *2*(1 SE-Articles), 48–56. <https://doi.org/10.35870/jtik.v2i1.53>