

## PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI UNTUK PEMILIHAN TEMPAT PROMOSI UPAYA MENINGKATKAN JUMLAH MAHASISWA BARU

Nurul Fadilah<sup>\*1</sup>, Sri Yulianto Joko Prasetyo<sup>2</sup>, Budhi Kristianto<sup>3</sup>

<sup>1,2,3</sup>Universitas Kristen Satya Wacana, Salatiga  
Email: <sup>1</sup>972021014@gmail.com, <sup>2</sup>sri.yulianto@uksw.edu, <sup>3</sup>budhik@uksw.edu  
<sup>\*</sup>Penulis Korespondensi

(Naskah masuk: 23 September 2024, diterima untuk diterbitkan: 10 April 2025)

### Abstrak

Efektivitas strategi promosi menjadi kunci keberhasilan institusi pendidikan tinggi dalam menarik minat calon mahasiswa baru. Namun, banyak institusi menghadapi masalah serius dalam pemilihan lokasi promosi yang tepat, yang mengakibatkan penurunan minat calon mahasiswa dan pemborosan sumber daya. Untuk mengatasi masalah ini, penelitian ini mengembangkan model klasifikasi berbasis data mining yang mampu mengidentifikasi lokasi promosi paling efektif. Tiga algoritma klasifikasi yang digunakan dalam penelitian ini adalah Logistic Regression, Support Vector Machine (SVM), dan Decision Tree (C4.5). Data pendaftaran mahasiswa dikumpulkan dan diproses melalui tahapan pra pemrosesan yang meliputi penggantian nilai hilang, normalisasi data, dan transformasi atribut nominal menjadi numerik. Data kemudian dibagi menjadi subset pelatihan dan pengujian menggunakan metode split data dengan rasio 70:30. Hasil evaluasi menunjukkan bahwa model Decision Tree (C4.5) memberikan performa terbaik dengan accuracy 93.75%, precision 97.37%, dan recall 90.24%. Logistic Regression juga menunjukkan hasil yang memuaskan dengan accuracy 90.00%, precision 92.31%, dan recall 87.80%. Sementara itu, SVM menunjukkan performa yang lebih rendah dengan accuracy 72.50%, precision 80.65%, dan recall 60.98%. Kesimpulannya, model Decision Tree (C4.5) dan Logistic Regression dapat diandalkan untuk mengoptimalkan strategi promosi institusi pendidikan tinggi, memastikan alokasi sumber daya yang lebih efisien dan efektif, serta meningkatkan jumlah pendaftar baru. Penelitian ini juga memberikan kontribusi signifikan dalam literatur terkait penggunaan data mining untuk strategi promosi di sektor pendidikan tinggi.

**Kata kunci:** Data Mining, Klasifikasi, Decision Tree, Logistic Regression, Promosi Pendidikan

### **PERFORMANCE COMPARISON OF CLASSIFICATION ALGORITHMS FOR PROMOTION SITE SELECTION IN AN EFFORT TO INCREASE THE NUMBER OF NEW STUDENTS**

#### Abstract

The effectiveness of promotional strategies is crucial for higher education institutions in attracting new student enrollments. However, many institutions face serious issues in selecting the appropriate promotional locations, leading to decreased student interest and resource wastage. To address this issue, this study develops a data mining-based prediction model capable of identifying the most effective promotional locations. The three classification algorithms used in this study are Logistic Regression, Support Vector Machine (SVM), and Decision Tree (C4.5). Student enrollment data were collected and processed through pre-processing stages, including missing value replacement, data normalization, and transformation of nominal attributes to numerical. The data were then split into training and testing subsets using a 70:30 split ratio. Evaluation results indicate that the Decision Tree (C4.5) model performed the best with an accuracy of 93.75%, precision of 97.37%, and recall of 90.24%. Logistic Regression also showed satisfactory results with an accuracy of 90.00%, precision of 92.31%, and recall of 87.80%. Meanwhile, SVM demonstrated lower performance with an accuracy of 72.50%, precision of 80.65%, and recall of 60.98%. In conclusion, the Decision Tree (C4.5) and Logistic Regression models are reliable for optimizing promotional strategies of higher education institutions, ensuring more efficient and effective resource allocation, and increasing new student enrollments. This study also makes a significant contribution to the literature related to the use of data mining for promotional strategies in the higher education sector.

**Keywords:** Data Mining, Prediction, Decision Tree, Logistic Regression, Educational Promotion

## 1. PENDAHULUAN

Di era globalisasi pendidikan tinggi, efektivitas strategi promosi menjadi kunci keberhasilan institusi dalam menarik minat calon mahasiswa baru. Namun, banyak institusi menghadapi masalah serius dalam pemilihan lokasi promosi yang tepat, mengakibatkan penurunan minat calon mahasiswa dan pemborosan sumber daya. Masalah ini tidak hanya berdampak pada aspek finansial, tetapi juga mengancam keberlanjutan program akademik dan reputasi institusi (Manimaran, Priyadharshini, & Yamuna Devi, 2022).

Ketidakefektifan dalam pemilihan lokasi promosi menjadi masalah krusial karena dapat menyebabkan institusi gagal mencapai target penerimaan mahasiswa baru. Hal ini berpotensi mengancam keberlangsungan operasional, menurunkan prestasi akademik, dan bahkan mengganggu stabilitas keuangan institusi (Wan Yaacob et al., 2020). Oleh karena itu, diperlukan pendekatan baru yang lebih efektif dan efisien dalam strategi promosi pendidikan tinggi.

Penelitian ini dilakukan untuk mengatasi masalah tersebut dengan memanfaatkan teknologi data mining dalam mengoptimalkan pemilihan lokasi promosi. Tujuan utamanya adalah memecahkan permasalahan ketidakefektifan promosi melalui analisis data yang lebih strategis dan berbasis bukti. Pendekatan ini diharapkan dapat meningkatkan jumlah pendaftar baru secara signifikan sambil mengoptimalkan penggunaan sumber daya yang tersedia (Gumba & Paragas, 2022).

Pentingnya membahas masalah ini terletak pada potensinya untuk meningkatkan efektivitas dan efisiensi strategi promosi institusi pendidikan tinggi. Dengan optimalisasi pemilihan lokasi promosi, institusi dapat meningkatkan *Return On Investment* (ROI), mencapai target penerimaan mahasiswa, dan memperkuat posisi kompetitif mereka (Xiao, Zhang, Zhang, Li, & Shen, 2021).

Untuk mengatasi masalah ini, penelitian mengusulkan penggunaan teknik data mining yang inovatif, khususnya dalam analisis klasifikasi untuk pemilihan lokasi promosi. Metode ini meliputi pengumpulan data, pra-pemrosesan, transformasi data, split data, klasifikasi, dan evaluasi (Wan Yaacob et al., 2020). Inovasi utama penelitian ini terletak pada penggunaan kombinasi algoritma klasifikasi yang belum banyak dieksplorasi dalam konteks promosi pendidikan tinggi.

Penelitian ini dilakukan untuk mengisi kesenjangan dalam literatur terkait penggunaan analisis data dalam pemilihan lokasi promosi di sektor pendidikan tinggi. Sementara penelitian sebelumnya lebih fokus pada pendekatan tradisional, penelitian ini menawarkan perspektif baru dengan mengintegrasikan teknik data mining terkini (Gumba & Paragas, 2022).

Konteks dalam penelitian ini meliputi penggunaan data mining di bidang pendidikan,

beberapa penelitian telah menunjukkan efektivitas berbagai algoritma seperti *Logistic Regression*, *Support Vector Machine* (SVM), dan C4.5. Sebuah studi (Manimaran et al., 2022) menggunakan algoritma tersebut untuk mengklasifikasi peluang penempatan kerja mahasiswa, mendemonstrasikan manfaat analisis data historis dalam meningkatkan persentase penempatan kerja.

Penerapan algoritma *Logistic Regression* untuk menganalisis data fisik mahasiswa memungkinkan universitas untuk menentukan kondisi fisik mahasiswa dan memberikan saran latihan yang efektif. Studi komparatif (Rahman & Raihan, 2022) membandingkan berbagai metode data mining untuk mengklasifikasi prestasi akademik mahasiswa, memberikan rekomendasi untuk peningkatan kinerja. Penelitian lainnya (Wan Yaacob et al., 2020) menemukan bahwa *Logistic Regression* merupakan algoritma terbaik untuk mengklasifikasi *dropout* mahasiswa dengan akurasi tinggi.

Untuk pengoptimalan lokasi promosi, sebuah penelitian (Gumba & Paragas, 2022) menggunakan algoritma klasifikasi untuk menganalisis data penerimaan mahasiswa, membantu universitas dalam meningkatkan kebijakan penerimaan berdasarkan data historis. Studi lain (Waheed et al., 2020) membandingkan *Logistic Regression* dengan model *deep learning* untuk mengklasifikasi kinerja akademik mahasiswa, menunjukkan keunggulan model *deep learning* dalam akurasi klasifikasi.

Penelitian-penelitian terdahulu telah mendemonstrasikan efektivitas teknik data mining, terutama algoritma *Logistic Regression*, SVM, dan C4.5, dalam berbagai aspek manajemen pendidikan tinggi. Algoritma-algoritma ini telah berhasil diterapkan untuk klasifikasi kinerja akademik (Rahman & Raihan, 2022), optimalisasi penerimaan mahasiswa (Gumba & Paragas, 2022), dan analisis *dropout* (Wan Yaacob et al., 2020). *Logistic Regression*, khususnya, menunjukkan keunggulan dalam mengklasifikasi peluang penempatan kerja dan potensi *dropout* mahasiswa (Manimaran et al., 2022; Wan Yaacob et al., 2020). Temuan-temuan ini menyediakan dasar yang kuat untuk pengembangan model klasifikasi lokasi promosi yang optimal, yang menjadi fokus utama penelitian ini. Dengan berlandaskan pada hasil-hasil tersebut, penelitian ini bertujuan untuk mengoptimalkan penggunaan algoritma-algoritma ini dalam konteks pemilihan lokasi promosi pendidikan tinggi, guna meningkatkan efektivitas strategi promosi dan alokasi sumber daya (Manimaran et al., 2022; Rahman & Raihan, 2022).

Inovasi yang diusulkan dalam penelitian ini adalah penggunaan kombinasi algoritma klasifikasi yang dioptimalkan untuk konteks promosi pendidikan tinggi. Penelitian akan mengevaluasi kinerja pendekatan ini terhadap data pendaftaran mahasiswa untuk menentukan metode yang paling efektif dalam mengklasifikasi lokasi promosi potensial.

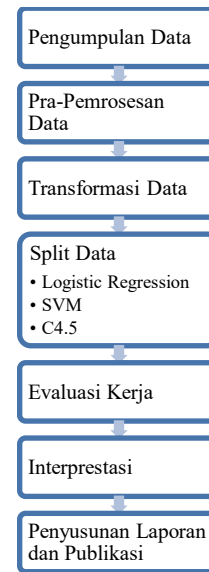
Dengan melakukan analisis mendalam dan perbandingan kinerja algoritma, penelitian ini bertujuan untuk menghasilkan model prediktif yang lebih akurat untuk pemilihan lokasi promosi. Hasil yang diharapkan adalah rekomendasi yang lebih tepat sasaran dalam strategi promosi, yang pada akhirnya akan membantu institusi pendidikan meningkatkan jumlah pendaftar baru dan mengoptimalkan alokasi sumber daya mereka (Manimaran et al., 2022; Rahman & Raihan, 2022).

Manfaat penelitian ini mencakup tiga aspek utama yaitu praktis, teoritis, dan metodologis. Secara praktis, penelitian ini memberikan informasi tentang algoritma *data mining* klasifikasi yang paling efektif dalam mengklasifikasi lokasi promosi yang akurat, membantu perguruan tinggi mengoptimalkan strategi dan alokasi sumber daya promosi mereka, serta berpotensi meningkatkan jumlah pendaftar baru secara signifikan. Dari segi teoritis, penelitian ini memperkaya wawasan dan pengetahuan mengenai penggunaan algoritma klasifikasi data mining (khususnya *Logistic Regression*, SVM, dan C4.5) dalam bidang pemasaran dan promosi pendidikan tinggi, serta memberikan kontribusi pada pengembangan literatur tentang aplikasi data mining dalam optimalisasi strategi pemasaran pendidikan. Secara metodologis, penelitian ini mengembangkan kerangka kerja yang dapat diadaptasi oleh institusi pendidikan lain untuk membandingkan dan mengevaluasi kinerja berbagai algoritma klasifikasi dalam konteks promosi, sehingga menyediakan dasar untuk penelitian lebih lanjut tentang integrasi teknik data mining dalam pengambilan keputusan strategis di institusi pendidikan.

Penelitian ini berlandaskan pada temuan-temuan sebelumnya dan mengisi kesenjangan dalam literatur. Fokus khusus diberikan pada penggunaan kombinasi algoritma *Logistic Regression*, SVM, dan C4.5 untuk optimalisasi lokasi promosi di pendidikan tinggi. Pendekatan ini diproyeksikan memberikan kontribusi signifikan dalam meningkatkan efektivitas strategi promosi institusi pendidikan tinggi.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining untuk mengoptimalkan pemilihan lokasi promosi bagi institusi pendidikan tinggi. Metode yang digunakan melibatkan serangkaian tahapan yang sistematis, dimulai dari pengumpulan data hingga interpretasi hasil dan penyusunan laporan. Pendekatan ini dipilih untuk menganalisis data historis pendaftaran mahasiswa dan menghasilkan model prediktif yang dapat membantu dalam pengambilan keputusan strategis terkait promosi. Berikut adalah diagram alur yang menggambarkan tahapan penelitian secara keseluruhan:



Gambar 1. Tahapan Penelitian

### 2.1 Collecting (Pengumpulan Data)

Tahap ini meliputi pengumpulan data historis pendaftaran mahasiswa STMIK YMI Tegal periode 2017-2021, mencakup 268 data dengan informasi sekolah, wilayah, jenis, status, akreditasi, jumlah lulusan, mahasiswa, dan kompetitor. Data tambahan diperoleh dari [dapo.kemdikbud.go.id](http://dapo.kemdikbud.go.id). Proses deskripsi dan eksplorasi data dilakukan untuk menganalisis sumber data awal dan menentukan sumber data yang paling relevan. Verifikasi kualitas data menjadi langkah penting untuk memastikan data bebas dari kesalahan, konsisten, dan representatif.

### 2.2 Preprocessing (Pra Pemrosesan)

Tahap ini mencakup data *reduction*, *cleaning*, *selection*, dan *balancing* menggunakan *MS Excel*. Proses ini bertujuan untuk memastikan kualitas data yang akan diproses lebih lanjut [8]. Kompetitor dikategorikan menjadi "sedikit" (kurang dari 10 institusi) dan "banyak" (10 atau lebih institusi) berdasarkan jumlah dan kepadatan institusi pendidikan di wilayah tersebut.

### 2.3 Transformation (Transformasi)

Data ditransformasikan untuk mendapatkan variabel relevan dalam pemetaan tempat promosi (Rahman & Raihan, 2022). Ditambahkan kolom target "POTENSI" dan "KURANG POTENSI" berdasarkan analisis historis dan tren data pendaftaran sebelumnya.

### 2.4 Split Data (Data Terpisah)

Data dibagi menjadi subset pelatihan dan pengujian dengan persentase umum seperti 70:30, 80:20, atau 90:10 (Asana & Yanti, 2023; Muningsih, 2022). Metode ini dipilih karena kesederhanaan dan kecepatannya dibandingkan teknik lain seperti *cross-*

validation (Joseph & Vakayil, 2020; Krivzanić, 2020).

Tabel 1. Data Preprocessing

NO	NAMA SEKOLAH	WILAYAH	JENIS SEKOLAH	STATUS	AKRED	JUMLAH LULUSAN	MHS	KOMPET
1	SMKS BHAKTI PRAJA DUKUHWARU	KAB. TEGAL	SMK	SWASTA	UNGGUL	797	ADA	SEDIKIT
2	SMKN 1 ADWIHERNA	KAB. TEGAL	SMK	NEGERI	UNGGUL	764	ADA	SEDIKIT
3	SMKS NUSANTARA 1 COMAL	KAB. PEMALANG	SMK	SWASTA	UNGGUL	654	ADA	SEDIKIT
4	SMKN 1 AMPELGADING	KAB. PEMALANG	SMK	NEGERI	UNGGUL	640	TIDAK	SEDIKIT
5	SMKS BHAKTI PRAJA ADWIHERNA	KAB. TEGAL	SMK	SWASTA	UNGGUL	633	TIDAK	SEDIKIT
6	SMKN 1 SLAWI	KAB. TEGAL	SMK	NEGERI	UNGGUL	617	ADA	SEDIKIT
7	SMKN 1 PEMALANG	KAB. PEMALANG	SMK	NEGERI	UNGGUL	609	TIDAK	SEDIKIT
8	SMK NEGERI 1 PETARUKAN	KAB. PEMALANG	SMK	NEGERI	UNGGUL	566	TIDAK	SEDIKIT
9	SMKN 1 BUMIHAWA	KAB. TEGAL	SMK	NEGERI	UNGGUL	558	ADA	SEDIKIT
10	SMKN 2 SLAWI	KAB. TEGAL	SMK	NEGERI	UNGGUL	527	TIDAK	SEDIKIT
11	SMKS SATYA PRAJA 2 PETARUKAN	KAB. PEMALANG	SMK	SWASTA	UNGGUL	524	TIDAK	SEDIKIT
12	SMKN 1 DUKUHTURI	KAB. TEGAL	SMK	NEGERI	UNGGUL	505	ADA	SEDIKIT
13	SMKS PGRI 2 TAMAN	KAB. PEMALANG	SMK	SWASTA	UNGGUL	504	TIDAK	SEDIKIT
14	SMKN 1 RANDUDONGKAL	KAB. PEMALANG	SMK	NEGERI	UNGGUL	500	TIDAK	SEDIKIT
15	SMKS PGRI 1 TAMAN	KAB. PEMALANG	SMK	SWASTA	UNGGUL	494	TIDAK	SEDIKIT
16	SMK BINA BANGSA KERSANA	KAB. BREBES	SMK	SWASTA	UNGGUL	490	TIDAK	BANYAK
17	SMKN 2 ADWIHERNA	KAB. TEGAL	SMK	NEGERI	UNGGUL	490	ADA	SEDIKIT
18	SMKS BELAM RANDUDONGKAL	KAB. PEMALANG	SMK	SWASTA	UNGGUL	488	ADA	SEDIKIT
19	SMKN 3 TEGAL	KOTA TEGAL	SMK	NEGERI	UNGGUL	484	ADA	BANYAK
20	SMKN 1 TONJONG	KAB. BREBES	SMK	NEGERI	UNGGUL	483	TIDAK	BANYAK

Tabel 2 Data Transformation

NO	NAMA SEKOLAH	WILAYAH	JENIS SEKOLAH	STATUS	AKRED	JUMLAH LULUSAN	MHS	KOMPET	LABEL
1	SMKS BHAKTI PRAJA DUKUHWARU	KAB. TEGAL	SMK	SWASTA	UNGGUL	797	ADA	SEDIKIT	POTENSI
2	SMKN 1 ADWIHERNA	KAB. TEGAL	SMK	NEGERI	UNGGUL	764	ADA	SEDIKIT	KURANG POTENSI
3	SMKS NUSANTARA 1 COMAL	KAB. PEMALANG	SMK	SWASTA	UNGGUL	654	ADA	SEDIKIT	KURANG POTENSI
4	SMKN 1 AMPELGADING	KAB. PEMALANG	SMK	NEGERI	UNGGUL	640	TIDAK	SEDIKIT	KURANG POTENSI
5	SMKS BHAKTI PRAJA ADWIHERNA	KAB. TEGAL	SMK	SWASTA	UNGGUL	633	TIDAK	SEDIKIT	KURANG POTENSI
6	SMKN 1 SLAWI	KAB. TEGAL	SMK	NEGERI	UNGGUL	617	ADA	SEDIKIT	KURANG POTENSI
7	SMKN 1 PEMALANG	KAB. PEMALANG	SMK	NEGERI	UNGGUL	609	TIDAK	SEDIKIT	KURANG POTENSI
8	SMK NEGERI 1 PETARUKAN	KAB. PEMALANG	SMK	NEGERI	UNGGUL	566	TIDAK	SEDIKIT	KURANG POTENSI
9	SMKN 1 BUMIHAWA	KAB. TEGAL	SMK	NEGERI	UNGGUL	558	ADA	SEDIKIT	KURANG POTENSI
10	SMKN 2 SLAWI	KAB. TEGAL	SMK	NEGERI	UNGGUL	527	TIDAK	SEDIKIT	KURANG POTENSI
11	SMKS SATYA PRAJA 2 PETARUKAN	KAB. PEMALANG	SMK	SWASTA	UNGGUL	524	TIDAK	SEDIKIT	KURANG POTENSI
12	SMKN 1 DUKUHTURI	KAB. TEGAL	SMK	NEGERI	UNGGUL	505	ADA	SEDIKIT	KURANG POTENSI
13	SMKS PGRI 2 TAMAN	KAB. PEMALANG	SMK	SWASTA	UNGGUL	504	TIDAK	SEDIKIT	KURANG POTENSI
14	SMKN 1 RANDUDONGKAL	KAB. PEMALANG	SMK	NEGERI	UNGGUL	500	TIDAK	SEDIKIT	KURANG POTENSI
15	SMKS PGRI 1 TAMAN	KAB. PEMALANG	SMK	SWASTA	UNGGUL	494	TIDAK	SEDIKIT	KURANG POTENSI
16	SMK BINA BANGSA KERSANA	KAB. BREBES	SMK	SWASTA	UNGGUL	490	TIDAK	BANYAK	KURANG POTENSI
17	SMKN 2 ADWIHERNA	KAB. TEGAL	SMK	NEGERI	UNGGUL	490	ADA	SEDIKIT	KURANG POTENSI
18	SMKS BELAM RANDUDONGKAL	KAB. PEMALANG	SMK	SWASTA	UNGGUL	488	ADA	SEDIKIT	KURANG POTENSI
19	SMKN 3 TEGAL	KOTA TEGAL	SMK	NEGERI	UNGGUL	484	ADA	BANYAK	POTENSI
20	SMKN 1 TONJONG	KAB. BREBES	SMK	NEGERI	UNGGUL	483	TIDAK	BANYAK	KURANG POTENSI

## 2.5 Classification (Klasifikasi)

Penelitian menggunakan algoritma *Logistic Regression*, *Support Vector Machine (SVM)*, dan *C4.5* (Kangéthe & Mburu, 2023; Palacios, Reyes-Suárez, Bearzotti, Leiva, & Marchant, 2021; Rahman & Raihan, 2022). Berikut adalah penjelasan dari masing-masing algoritma:

### a. Algoritma Logistic Regression

*Logistic Regression* digunakan untuk memklasifikasi probabilitas pendaftaran berdasarkan karakteristik demografis dan lokasi promosi. Model ini bekerja dengan menghitung logit, yaitu logaritma dari rasio peluang (*odds ratio*), dan kemudian mengaplikasikan fungsi sigmoid untuk mendapatkan probabilitas (Palacios et al., 2021; Rahman & Raihan, 2022; Sahamony, Terttiaavini, & Rianto, 2024).

Rumus *Logistic Regression*:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (1)$$

Di mana:

$P(y = 1|X)$  adalah probabilitas kejadian  $y = 1$  diberikan variabel prediktor  $X$ .  $\beta_0$  adalah intercept dari model.  $\beta_1, \beta_2, \dots, \beta_n$  adalah koefisien dari variabel prediktor  $X_1, X_2, \dots, X_n$ .

*Pseudocode Logistic Regression*:

Initialize weights  $\beta_0, \beta_1, \dots, \beta_n$

For each iteration:

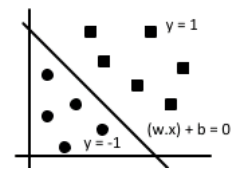
Calculate logit:  $z = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n$

Calculate probability:  $P(y=1|X) = 1 / (1 + e^{(-z)})$

Update weights using gradient descent

### b. Algoritma Support Vector Machine (SVM)

Digunakan untuk menemukan *hyperplane* terbaik yang memisahkan dua kelas data dengan margin maksimum. Jika *input* data dapat dipisahkan secara *linear*, pemisahan *hyperplane* dapat digambarkan sebagai berikut:



Gambar 2 Bidang Pemisah Linier

Model ini sangat efektif untuk menangani data berdimensi tinggi (Kasim & Sudarsono, 2019). Rumus untuk SVM adalah:

$$f(x) = \text{sign}(w \cdot x + b) \quad (2)$$

Di mana:

$w$  adalah vektor bobot.

$x$  adalah vektor fitur.

$b$  adalah bias.

*Pseudocode SVM*:

Initialize weights  $w$  and bias  $b$

For each iteration:

For each training sample  $(x_i, y_i)$ :

If  $y_i * (w \cdot x_i + b) < 1$ :

Update weights:  $w = w + \eta * (y_i * x_i - 2 * \lambda * w)$

Update bias:  $b = b + \eta * y_i$

Else:

Update weights:  $w = w - 2 * \eta * \lambda * w$

### c. Algoritma C4.5

Algoritma ini digunakan untuk membangun pohon keputusan yang membantu dalam memetakan hubungan antara variabel prediktor dan hasil klasifikasi. Proses ini melibatkan penggunaan *information gain* untuk memilih atribut yang paling informatif dalam membagi dataset (Rahman & Raihan, 2022). Rumus untuk *information gain* adalah:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (3)$$

Di mana:

**Gain(S, A)** adalah informasi yang diperoleh dari pembagian set data **S** dengan atribut **A**.

**Entropy(S)** adalah entropi dari set data **S**.

**S<sub>v</sub>** adalah subset dari **S** yang memiliki nilai **v** untuk atribut **A**.

*Pseudocode C4.5:*

*If all instances are of the same class:*

*Return a leaf node with that class*

*If attribute list is empty:*

*Return a leaf node with the most common class*

*Select attribute A with the highest information gain*

*For each value v in A:*

*Create a branch for value v*

*Split the dataset S into subsets S<sub>v</sub>*

*Recursively apply C4.5 to the subsets*

### 2.6 Evaluation (Evaluasi)

Setelah model diklasifikasikan, kinerjanya dievaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *area under the curve (AUC)*. Hasil dari masing-masing algoritma dikumpulkan dan dibandingkan untuk menentukan algoritma mana yang memberikan hasil paling akurat dalam mengklasifikasi lokasi promosi potensial (Budiman & Niqotaini, 2022; Toro & Lestari, 2023). Tabel 3 menyajikan hasil evaluasi kinerja model.

Tabel 3 Klasifikasi nilai AUC

Nilai AUC	Kualitas AUC
0,90 – 1,00	Excellent Classification
0,80 – 0,90	Good Classification
0,70 – 0,80	Fair Classification
0,60 – 0,70	Poor Classification
0,50 – 0,60	Failuer Classification

### 2.7 Interpretation (Interpretasi)

Interpretasi melibatkan pemahaman hasil dan pemilihan algoritma dengan nilai *accuracy*, *precision*, *recall*, dan *AUC* terbaik untuk rekomendasi keputusan dalam pemilihan tempat promosi (Azizah, Nurdiawan, Dwilestari, Kaslani, & Tohidi, 2022).

### 2.8 Preparation of Reports and Publications (Penyusunan Laporan dan Publikasi)

Laporan penelitian yang komprehensif akan disusun dan dipublikasikan dalam jurnal ilmiah atau konferensi akademik untuk berbagi temuan dengan komunitas akademik.

## 3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengembangkan model klasifikasi berbasis *data mining* yang mampu mengidentifikasi lokasi-lokasi promosi yang paling efektif dalam meningkatkan jumlah pendaftar baru di institusi pendidikan tinggi. Untuk mencapai tujuan ini, tiga algoritma klasifikasi yang berbeda telah

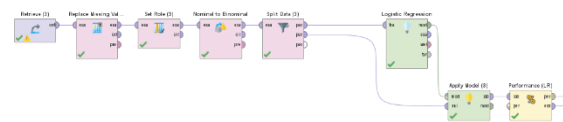
digunakan, yaitu *Logistic Regression*, *SVM*, dan *Decision Tree (C4.5)*. Setiap model dievaluasi berdasarkan beberapa metrik penting termasuk *accuracy*, *precision*, *recall*, dan *AUC*. Berikut adalah hasil dan pembahasan dari masing-masing model.

### 3.1 Hasil Analisis

Proses implementasi model klasifikasi dilakukan menggunakan *RapidMiner*. Diagram proses berikut ini menunjukkan alur kerja yang meliputi pengumpulan data, pra-pemrosesan, transformasi data, validasi silang, dan evaluasi kinerja algoritma. Proses ini memastikan setiap tahapan dilaksanakan dengan teliti untuk menghasilkan model prediktif yang akurat.

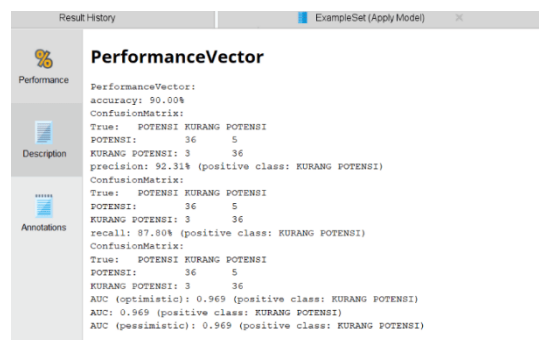
#### A. Logistic Regression

*Logistic Regression* adalah salah satu metode klasifikasi yang paling umum digunakan dalam analisis prediktif. Dalam penelitian ini, *Logistic Regression* menunjukkan performa yang sangat baik. Berikut adalah gambar proses *Logistic Regression*.



Gambar 3 Desain Penerapan *Logistic Regression*

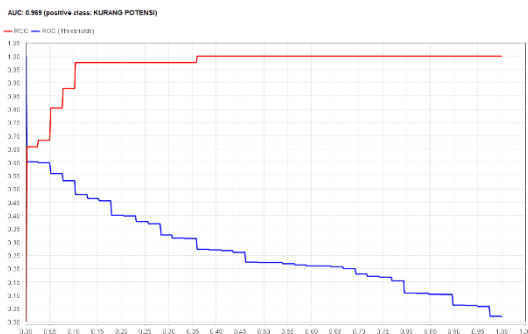
Proses dimulai dengan pengambilan data menggunakan operator "*Retrieve*", diikuti dengan penanganan nilai hilang melalui "*Replace Missing Values*". Selanjutnya, peran atribut ditentukan menggunakan "*Set Role*", dan atribut nominal diubah menjadi binominal dengan "*Nominal to Binominal*". Data kemudian dibagi menjadi subset pelatihan dan pengujian menggunakan "*Split Data*". Model *Logistic Regression* diterapkan pada subset pelatihan, dan hasilnya diuji pada subset pengujian dengan operator "*Apply Model*". Kinerja model dievaluasi menggunakan operator "*Performance*", menghasilkan metrik seperti *accuracy*, *precision*, *recall*, dan *AUC* untuk mengukur efektivitas model dalam memklasifikasi lokasi promosi yang potensial.



Gambar 4 Hasil *Performance Vector Logistic Regression*

Model ini memiliki *accuracy* sebesar 90.00%, yang berarti bahwa 90% dari klasifikasi yang dibuat oleh model ini adalah benar. *Precision* dari model ini mencapai 92.31% untuk kelas "KURANG POTENSI", yang menunjukkan bahwa dari semua klasifikasi yang diklasifikasikan sebagai "KURANG POTENSI", 92.31% di antaranya benar-benar merupakan kelas tersebut.

Lebih lanjut, *recall* untuk kelas "KURANG POTENSI" adalah 80.49%, yang berarti bahwa dari semua *instance* yang sebenarnya adalah "KURANG POTENSI", model berhasil mengidentifikasi sekitar 80.49% dari mereka.

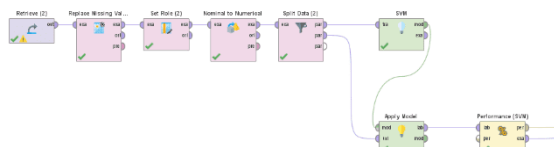


Gambar 5 Grafik AUC Logistic Regression

Nilai AUC yang mencapai 0.969 menunjukkan bahwa model *Logistic Regression* memiliki kemampuan diskriminatif yang sangat baik dalam membedakan antara kelas "POTENSI" dan "KURANG POTENSI". Dengan kata lain, model ini sangat efektif dalam memklasifikasi potensi lokasi promosi dengan tingkat kesalahan yang sangat rendah.

### B. Support Vector Machine (SVM)

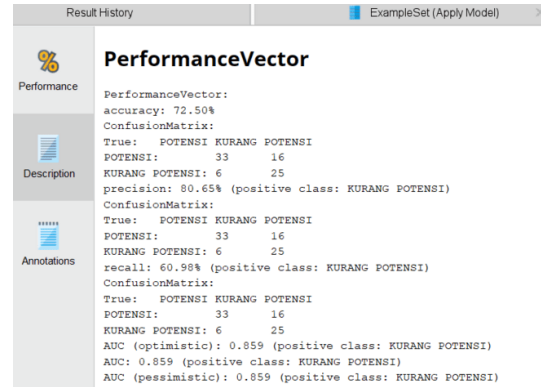
*Support Vector Machine* (SVM) adalah algoritma klasifikasi yang kuat terutama dalam kasus di mana data memiliki dimensi yang tinggi. Berikut adalah gambar proses *Support Vector Machine* (SVM).



Gambar 6 Desain Penerapan Support Vector Machine (SVM)

Proses dimulai dengan pengambilan data menggunakan operator "Retrieve", diikuti dengan penanganan nilai hilang melalui "Replace Missing Values". Selanjutnya, peran atribut ditentukan menggunakan "Set Role", dan atribut nominal diubah menjadi numerik dengan operator "Nominal to Numerical". Data kemudian dibagi menjadi subset pelatihan dan pengujian menggunakan "Split Data". Model SVM diterapkan pada subset pelatihan, dan hasilnya diuji pada subset pengujian dengan operator "Apply Model". Kinerja model dievaluasi

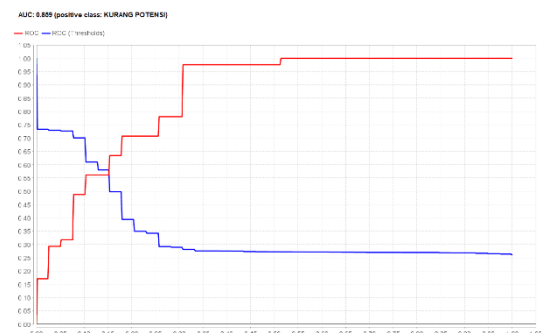
menggunakan operator "Performance", menghasilkan metrik seperti *accuracy*, *precision*, *recall*, dan AUC untuk mengukur efektivitas model dalam memklasifikasi lokasi promosi yang potensial.



Gambar 7 Hasil Performance Vector Support Vector Machine (SVM)

Model SVM menghasilkan *accuracy* sebesar 72.50%, yang berarti sekitar 73% dari klasifikasi yang dibuat adalah benar. *Precision* untuk kelas "KURANG POTENSI" adalah 80.65%, menunjukkan bahwa dari semua klasifikasi yang diklasifikasikan sebagai "KURANG POTENSI", sekitar 81% di antaranya benar.

Namun, nilai *recall* untuk kelas "KURANG POTENSI" hanya mencapai 60.98%. Hal ini menunjukkan bahwa model SVM kurang efektif dalam mengidentifikasi semua instance dari kelas "KURANG POTENSI". Dengan kata lain, ada banyak instance dari kelas "KURANG POTENSI" yang tidak berhasil diidentifikasi oleh model.

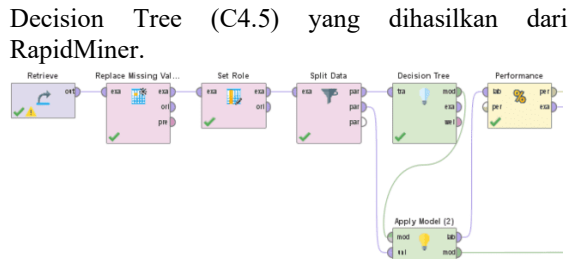


Gambar 8 Grafik AUC Support Vector Machine (SVM)

Nilai AUC sebesar 0.859 menunjukkan bahwa kemampuan diskriminatif model SVM masih cukup baik, namun tidak sebaik *Logistic Regression* atau *Decision Tree*.

### C. Penerapan Algoritma C.45 (Decision Tree)

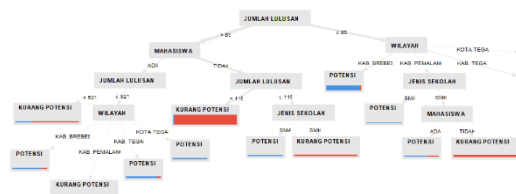
*Decision Tree* (C4.5) adalah algoritma klasifikasi yang sederhana namun sangat efektif dalam banyak kasus. Dalam penelitian ini, *Decision Tree* (C4.5) menunjukkan performa yang sangat baik, bahkan sedikit lebih baik dibandingkan *Logistic Regression*. Berikut adalah gambar proses



Gambar 9 Desain Penerapan Algoritma C.45 (Decision Tree)

Proses dimulai dengan pengambilan data menggunakan operator "Retrieve", diikuti dengan penanganan nilai hilang melalui "Replace Missing Values". Selanjutnya, peran atribut ditentukan menggunakan operator "Set Role". Data kemudian dibagi menjadi subset pelatihan dan pengujian menggunakan operator "Split Data". Model Decision Tree dilatih menggunakan subset pelatihan dan diterapkan pada subset pengujian menggunakan operator "Apply Model". Terakhir, kinerja model dievaluasi menggunakan operator "Performance" untuk mengukur efektivitas model dalam klasifikasi.

Hasil pohon keputusan model Decision Tree (C4.5) yang dihasilkan dari RapidMiner.



Gambar 10 Hasil Pohon Keputusan

Pohon keputusan ini digunakan untuk mengklasifikasikan potensi wilayah promosi berdasarkan jumlah lulusan, status mahasiswa, wilayah, dan jenis sekolah. Hasil pohon keputusan dengan kedalaman maksimal 5 adalah sebagai berikut:

1. Jumlah Lulusan > 89:
  - Jika terdapat mahasiswa:
    - 521 lulusan: KURANG POTENSI (1 POTENSI, 3 KURANG POTENSI).
    - ≤ 521 lulusan:
      - Wilayah KAB. BREBES: POTENSI (5 POTENSI, 1 KURANG POTENSI).
      - Wilayah KAB. PEMALANG: KURANG POTENSI (0 POTENSI, 3 KURANG POTENSI).
      - Wilayah KAB. TEGAL: POTENSI (9 POTENSI, 1 KURANG POTENSI).
      - Wilayah KOTA TEGAL: POTENSI (6 POTENSI, 0 KURANG POTENSI).
  - Jika tidak terdapat mahasiswa:
    - 119 lulusan: KURANG POTENSI (1 POTENSI, 64 KURANG POTENSI).
    - ≤ 119 lulusan:

- Jenis sekolah SMA: POTENSI (5 POTENSI, 0 KURANG POTENSI).
- Jenis sekolah SMK: KURANG POTENSI (0 POTENSI, 7 KURANG POTENSI).

2. Jumlah Lulusan ≤ 89:
  - Wilayah KAB. BREBES: POTENSI (30 POTENSI, 1 KURANG POTENSI).
  - Wilayah KAB. PEMALANG:
    - Jenis sekolah SMA: POTENSI (2 POTENSI, 0 KURANG POTENSI).
    - Jenis sekolah SMK:
      - Jika ada mahasiswa: POTENSI (2 POTENSI, 1 KURANG POTENSI).
      - Jika tidak ada mahasiswa: KURANG POTENSI (0 POTENSI, 12 KURANG POTENSI).
  - Wilayah KAB. TEGAL: POTENSI (22 POTENSI, 3 KURANG POTENSI).
  - Wilayah KOTA TEGAL: POTENSI (9 POTENSI, 0 KURANG POTENSI).

Secara keseluruhan, pohon keputusan ini mengidentifikasi wilayah promosi yang potensial berdasarkan kombinasi variabel jumlah lulusan, status mahasiswa, wilayah, dan jenis sekolah. Wilayah dengan jumlah lulusan tinggi dan beberapa kombinasi spesifik dari variabel ini cenderung diklasifikasikan sebagai "POTENSI", sementara kombinasi lainnya mungkin diklasifikasikan sebagai "KURANG POTENSI". Untuk hasil *performance vector* sebagai berikut:

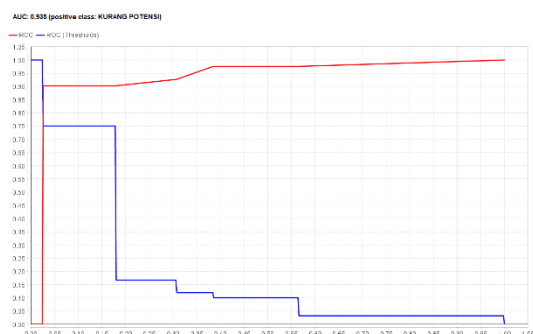
```

istory | Tree (Decision Tree) x
PerformanceVector
PerformanceVector:
accuracy: 93.75%
ConfusionMatrix:
True:  POTENSI KURANG POTENSI
POTENSI:  38    4
KURANG POTENSI: 1    37
precision: 97.37% (positive class: KURANG POTENSI)
ConfusionMatrix:
True:  POTENSI KURANG POTENSI
POTENSI:  38    4
KURANG POTENSI: 1    37
recall: 90.24% (positive class: KURANG POTENSI)
ConfusionMatrix:
True:  POTENSI KURANG POTENSI
POTENSI:  38    4
KURANG POTENSI: 1    37
AUC (optimistic): 0.954 (positive class: KURANG POTENSI)
AUC: 0.935 (positive class: KURANG POTENSI)
AUC (pessimistic): 0.926 (positive class: KURANG POTENSI)
    
```

Gambar 11 Hasil Performance Vector Decision Tree (C4.5)

Decision Tree (C4.5) menunjukkan performa yang sangat baik, dengan *accuracy* sebesar 93.75%. *Precision* untuk kelas "KURANG POTENSI" mencapai 97.37%, yang berarti bahwa dari semua klasifikasi yang diklasifikasikan sebagai "KURANG POTENSI", 97.37% di antaranya benar. Selain itu,

*recall* untuk kelas "KURANG POTENSI" adalah 90.24%, yang berarti model ini berhasil mengidentifikasi lebih dari 90% dari semua *instance* yang sebenarnya adalah "KURANG POTENSI".



Gambar 12 Grafik AUC Decision Tree (C4.5)

Nilai AUC sebesar 0.935 menunjukkan bahwa model *Decision Tree* memiliki kemampuan diskriminatif yang sangat baik. Secara keseluruhan, *Decision Tree* (C4.5) adalah model yang sangat baik dalam penelitian ini dalam hal *accuracy*, *precision*, dan *recall*.

### 3.2 Pembahasan

Analisis tiga model klasifikasi menunjukkan *Decision Tree* (C4.5) sebagai model terbaik, dengan *accuracy* tinggi serta *precision* dan *recall* seimbang. Keunggulan ini dikaitkan dengan kemampuannya menangani hubungan *non-linear* dan interpretabilitas yang baik, sejalan dengan temuan Gumba dan Paragas (2022) dalam meningkatkan kebijakan penerimaan mahasiswa (Gumba & Paragas, 2022). *Logistic Regression* menunjukkan performa yang hampir setara, terutama dalam akurasi dan AUC, mendukung hasil Manimaran et al. (2022) tentang klasifikasi peluang penempatan kerja mahasiswa (Manimaran et al., 2022). SVM, meskipun memiliki *precision* tinggi, menunjukkan performa lebih rendah dalam *recall*. *Decision Tree* dianggap paling efektif untuk klasifikasi ini, dengan *Logistic Regression* sebagai alternatif yang andal. Hasil penelitian ini akan dipublikasikan untuk berkontribusi pada pengembangan strategi promosi di pendidikan tinggi menggunakan teknik *data mining*.

## 4. KESIMPULAN

Penelitian ini berhasil mengembangkan dan menguji model klasifikasi berbasis data mining menggunakan *Logistic Regression*, SVM, dan *Decision Tree* (C4.5) untuk mengidentifikasi lokasi promosi yang efektif dalam meningkatkan jumlah pendaftar baru di institusi pendidikan tinggi. Hasil analisis menunjukkan bahwa *Decision Tree* (C4.5) dan *Logistic Regression* adalah algoritma yang paling efektif dengan *accuracy*, *precision*, *recall*, dan AUC yang tinggi, sementara SVM menunjukkan performa yang lebih rendah. Dengan demikian, institusi pendidikan tinggi dapat mengandalkan *Decision Tree*

(C4.5) atau *Logistic Regression* untuk mengoptimalkan strategi promosi mereka, memastikan penggunaan sumber daya yang lebih efisien dan efektif. Penelitian ini juga berkontribusi pada literatur dengan mengkombinasikan algoritma klasifikasi yang belum banyak dieksplorasi dalam konteks promosi pendidikan tinggi.

## DAFTAR PUSTAKA

- ASANA, I. M. D. P., & YANTI, N. P. D. T. 2023. Sistem Klasifikasi Pengajuan Kredit Dengan Metode Support Vector Machine (SVM) I Made Dwi Putra Asana. *Jurnal Sistem Cerdas*, 6(2), 123–133.
- AZIZAH, H. W., NURDIWAN, O., DWILESTARI, G., KASLANI, K., & TOHIDI, E. 2022. Klasifikasi Pemberian Bantuan UMKM Cirebon dengan Menggunakan Algoritma K-Nearest Neighbor. *Journal of Computer System and Informatics (JoSYC)*, 3(3), 110–115.
- BUDIMAN, & NIQOTAINI, Z. 2022. Analisis Perbandingan Kinerja Algoritma Klasifikasi Data Mining Penelusuran Minat Calon Mahasiswa. *SISTEMASI: Jurnal Sistem Informasi*, 11, 271–290.
- GUMBA, G., & PARAGAS, J. R. 2022. Prediction analysis of student admission to information technology education (ite) programs using classification algorithm. *2022 Fifth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 112–117. IEEE.
- JOSEPH, V. R., & VAKAYIL, A. 2020. SPlit: An Optimal Method for Data Splitting. *Technometrics*, 64, 166–176.
- KANGÉTHE, G., & MBURU, L. W. 2023. Evaluating the Efficacy of Elearning in Higher Educational Institutions using Educational Data Mining. *2023 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 1–6. IEEE.
- KASIM, A. A., & SUDARSONO, M. 2019. Algoritma Support Vector Machine ( SVM ) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah. *Seminar Nasional APTIKOM (SEMNASTIK)*, 568–573.
- KRIVZANIĆ, S. 2020. Educational data mining using cluster analysis and decision tree technique: A case study. *International Journal of Engineering Business Management*, 12.
- MANIMARAN, P., PRIYADHARSHINI, G., & YAMUNA DEVI, N. 2022. Predicting the eligibility of placement for students using data mining. *International journal of health sciences*, 6(June), 460–467.

- MUNINGSIH, E. 2022. Kombinasi metode k-means dan decision tree dengan perbandingan kriteria dan split data. *Jurnal Teknoinfo*.
- PALACIOS, C. A., REYES-SUÁREZ, J. A., BEARZOTTI, L. A., LEIVA, V., & MARCHANT, C. 2021. Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile. *Entropy*, 23(4), 1–23.
- RAHMAN, A. T., & RAIHAN, A. 2022. Prediction of Student Performance in Higher Academy Through Different Data Mining Techniques. *2022 4th International Conference on Electrical, Computer \& Telecommunication Engineering (ICECTE)*, 1–4.
- SAHAMONY, N. F., TERTTIAVINI, T., & RIANTO, H. 2024. Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak: Analysis of Performance Comparison of Machine Learning Models for Predicting Stunting Risk in Children's Growth. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 413–422.
- TORO, R., & LESTARI, S. 2023. Perbandingan Algoritma Data Mining Untuk Penentuan Lokasi Promosi Penerimaan Mahasiswa Baru Pada IIB Darmajaya Lampung. *Techno.Com*, 22(1), 223–234.
- WAHEED, H., HASSAN, S.-U., ALJOHANI, N. R., HARDMAN, J., ALELYANI, S., & NAWAZ, R. 2020. Predicting academic performance of students from VLE big data using deep learning models. *Comput. Hum. Behav.*, 104, 106189.
- WAN YAACOB, W. F., MOHD SOBRI, N., NASIR, S. A. M., WAN YAACOB, W. F., NORSHAHIDI, N. D., & WAN HUSIN, W. Z. 2020. Predicting Student Drop-Out in Higher Institution Using Data Mining Techniques. *Journal of Physics: Conference Series*, 1496(1).
- XIAO, T., ZHANG, P., ZHANG, Y., LI, D.-X., & SHEN, J. 2021. *A research on the application of college students ' physique Data Mining based on Logistic Regression Algorithm*.

*Halaman ini sengaja dikosongkan*