

Article Identity

Volume 1 Issue 1 Pages 27–38

October 2026

E-ISSN xxxx-xxxx

Article History

Received 1 August 2026

Revised 30 August 2026

Accepted 08 September 2026

Published 14 September 2026

Keywords

Student Graduation;
Classification; Linear
Discriminant Analysis; Extreme
Learning Machine; K-Fold Cross
Validation

Copyright © 2026 Author(s).

Editorial office: Journal of Statistics and
Data Analytics, PT. Gammabeta Edu and
Tech.

To Cite this Article: Najwa Azzahra,
Dian Candra Rini Novitasari, Adelia Rova
Chumairo. Prediksi Status Kelulusan
Mahasiswa Menggunakan Linear
Discriminant Analysis (LDA) dan Extreme
Learning Machine (ELM). Journal of
Statistics and Data Analytics.
2026;1(1):27–38.

RESEARCH ARTICLE

Prediksi Status Kelulusan Mahasiswa Menggunakan *Linear Discriminant Analysis* dan *Extreme Learning Machine*

Najwa Azzahra¹, Dian Candra Rini Novitasari^{2*}, Adelia Rova Chumairo³

^{1,2,3}Program Studi Matematika, Fakultas Sains dan Teknologi, UIN Sunan Ampel Surabaya, Indonesia

Corresponding author: diancrini@uinsa.ac.id

Abstract. Student graduation plays a crucial role in assessing the quality of higher education. Early prediction of student graduation can assist universities in taking appropriate and precise steps to improve student graduation rates. This study aimed to predict student graduation rates using a combination of Linear Discriminant Analysis (LDA) and Extreme Learning Machine (ELM) methods. The research steps included handling missing values to address incomplete data, Min–Max Scaler normalization to equalize the range of values in data attributes, dimensionality reduction using Linear Discriminant Analysis (LDA), and classification using Extreme Learning Machine (ELM). Model evaluation was conducted using K-Fold Cross Validation to obtain more objective results and reduce data distribution bias. Model performance was measured using accuracy, precision, recall, and F1-score. The test results showed that the best results were achieved with an accuracy of 91.34%, precision of 86.44%, recall of 79.05%, and F1-score of 79.33%. These results indicate that the combination of Linear Discriminant Analysis (LDA) and Extreme Learning Machine (ELM) is capable of producing good classification performance for predicting student graduation.

Abstrak. Kelulusan mahasiswa memiliki peran yang penting dalam menilai kualitas pelaksanaan pendidikan di perguruan tinggi. Kemampuan untuk memprediksi kelulusan mahasiswa sejak dini dapat membantu perguruan tinggi dalam mengambil langkah yang sesuai dan tepat guna meningkatkan tingkat kelulusan studi pada mahasiswa. Penelitian ini dilakukan dengan tujuan memprediksi tingkat kelulusan mahasiswa menggunakan kombinasi metode Linear Discriminant Analysis (LDA) dan Extreme Learning Machine (ELM). Tahapan penelitian meliputi penanganan missing value, normalisasi Min–Max Scaler, reduksi dimensi menggunakan LDA, serta klasifikasi menggunakan ELM. Evaluasi model dilakukan menggunakan K-Fold Cross Validation. Performa model diukur menggunakan accuracy, precision, recall, dan F1-score. Output pada abstrak melaporkan accuracy 91,34%, precision 86,44%, recall 79,05%, dan F1-score 79,33%.

1. PENDAHULUAN

Pendidikan merupakan salah satu pendukung kuat dan berperan besar dalam penentuan masa depan yang cemerlang [1]. Pendidikan di perguruan tinggi merupakan salah satu tahap akhir dari proses menempuhnya pendidikan formal dan dapat menjadi penentu masa depan bagi generasi muda [2]. Berdasarkan data pada Universitas Lampung dari total 1.712 mahasiswa angkatan 2021, yang lulus hanya 1.018 mahasiswa (59%), sedangkan yang tidak lulus sekitar 41% mahasiswa (694 orang) dari 19 program studi [3]. Kondisi tersebut diperlukan tindak lanjut agar jumlah kelulusan mahasiswa sesuai kebutuhan pemantauan Badan Akreditasi. Di mana masa studi tepat waktu mahasiswa adalah empat tahun [4].

Berdasarkan data Pangkalan Data Pendidikan Tinggi (PDDIKTI) tahun 2025 yang dipublikasikan oleh Kementerian Pendidikan Tinggi, Sains, dan Teknologi, permasalahan keberhasilan studi mahasiswa masih menjadi perhatian penting dalam penyelenggaraan pendidikan tinggi di Indonesia. Pada tahun 2025 tercatat sebanyak 289.670 mahasiswa mengalami putus kuliah (*drop out*), yang menunjukkan adanya peningkatan dibandingkan tahun sebelumnya. Kondisi tersebut berpotensi memengaruhi jumlah mahasiswa yang berhasil menyelesaikan studinya sehingga berdampak pada capaian kelulusan perguruan tinggi [5]. Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) merupakan lembaga yang berwenang dalam mengukur kualitas dan keberlanjutan penyelenggaraan perguruan tinggi di Indonesia melalui proses akreditasi [6]. Akreditasi digunakan sebagai sarana evaluasi untuk mengukur kualitas perguruan tinggi maupun program studi. Menurut peraturan penilaian BAN-PT, aspek mahasiswa dan lulusan menjadi salah satu komponen penting yang dinilai, termasuk tingkat kelulusan tepat waktu. Beberapa ketidaksetaraan antara banyak mahasiswa yang terdaftar dengan banyak mahasiswa yang berhasil lulus pada suatu studi dapat memengaruhi hasil penilaian akreditasi program studi maupun perguruan tinggi, dan perguruan tinggi masih belum memiliki cara yang akurat untuk memprediksi kelulusan mahasiswa sejak dini [7].

Negara di Indonesia, akreditasi termasuk salah satu kepastian kualitas dalam sebuah perguruan tinggi dan khususnya di program studi. Permasalahan utama yang akan dialami oleh perguruan tinggi maupun program studi yaitu disaat waktu penanganannya tidak tepat waktu

akibatnya bisa lebih sulit untuk dicapainya target kelulusan. Kondisi tersebut menyebabkan kelulusan tepat waktu mencakup setidaknya tingkat kuliah selama delapan semester atau 4 tahun [8]. Jika tidak dilakukan uji prediksi kelulusan maka pengajuan akreditasi perguruan tinggi akan terhambat, perguruan tinggi yang tidak mengajukan akreditasi pada masa transisi dikarenakan banyaknya ketidaklulusan mahasiswa, akan diberikan status tidak terakreditasi oleh BAN-PT/LAM yang membuat kampus akan kesulitan mendapatkan mahasiswa baru, sulit meluluskan mahasiswa, dan lambat laun juga kalah populer dengan kampus lain [9]. Kondisi tersebut menyebabkan diperlukannya suatu pendekatan yang mampu mengelompokkan atau mengklasifikasikan mahasiswa berdasarkan karakteristik akademiknya sehingga perguruan tinggi dapat lebih mudah mengidentifikasi mahasiswa yang berpeluang lulus tepat waktu serta tidak tepat waktu. Keterangan tersebut bisa digunakan untuk acuan pengambilan keputusan dan penyusunan strategi pendampingan yang lebih efektif untuk meningkatkan keberhasilan studi mahasiswa [10].

Penelitian terdahulu [11] menganalisis data prediksi mahasiswa yang dapat lulus tepat waktu. Penelitiannya menerapkan metode algoritma *Decision Tree* terhadap data klasifikasi, dan mendapatkan hasil nilai akurasi yang tinggi sebesar 87.93% [11]. Penelitian lain [12] juga menganalisis data prediksi kelulusan mahasiswa. Penelitiannya menerapkan metode *Naive Bayes*, dan mendapatkan hasil akurasi tinggi sebesar 90% [12]. Kedua hasil penelitian tersebut sama-sama menggunakan metode klasifikasi dan keduanya memiliki hasil nilai akurasi yang tinggi. Namun *Decision Tree* rentan overfitting pada data yang kompleks yang menyebabkan model terlalu menyesuaikan diri dengan data, serta *Naive Bayes* yang memiliki kelemahan asumsi independensi antar variabel yang tidak terpenuhi dalam konteks pendidikan [13], [14]. Hal tersebut dapat diatasi menggunakan *Extreme Learning Machine* (ELM) yang unggul pada kecepatan komputasi dan kestabilan output, sesuatu yang tidak dijamin oleh *Decision Tree* maupun *Naive Bayes* [15].

Penelitian terdahulu [16] melakukan analisis data kualitas udara. Penelitiannya menerapkan metode algoritma *Extreme Learning Machine* (ELM) terhadap data klasifikasi serta menghasilkan nilai akurasi yang tinggi yaitu 90% [16]. Penelitian lain [17] melaksanakan uji analisis data jenis kendaraan melalui teknik olah citra digital. Penelitiannya menggunakan metode algoritma *Extreme Learning Machine* (ELM) terhadap data klasifikasi serta menghasilkan nilai akurasi yang tinggi yaitu 80% [17]. Kedua hasil penelitian tersebut sama-sama memanfaatkan metode klasifikasi *Extreme Learning Machine* (ELM) serta sama-sama memiliki nilai akurasi yang tinggi. Hasil kedua penelitian tersebut masih belum menggunakan metode reduksi dimensi untuk meminimaliskan dimensi data yang tinggi, kombinasi dari *Linear Discriminant Analysis* (LDA) dan *Extreme Learning Machine* (ELM) masih belum ada untuk prediksi kelulusan, yang dimana kombinasi kedua metode tersebut digunakan dalam penelitian ini.

Salah satu metode dalam jaringan saraf tiruan adalah *Extreme Learning Machine* (ELM). Keunggulan metode ini yaitu pada kecepatan model dan hasil error yang rendah [18]. *Linear Discriminant Analysis* (LDA) merupakan salah satu metode reduksi dimensi untuk data yang berdimensi tinggi. Keunggulan metode ini yaitu cocok untuk mengurangi dimensi data, yang dimana metode klasifikasi lemah terhadap data berdimensi tinggi [19].

Penelitian ini bertujuan untuk menganalisis tingkat akurasi prediksi kelulusan mahasiswa menggunakan metode *Extreme Learning Machine* (ELM) yang dioptimalkan dengan reduksi dimensi *Linear Discriminant Analysis* (LDA). Secara akademis, penelitian ini sangat berperan dalam pengembangan model klasifikasi yang efisien untuk data pendidikan tinggi. Secara praktis, hasil penelitian ini dapat digunakan oleh perguruan tinggi sebagai sistem pendukung keputusan untuk mengidentifikasi mahasiswa yang berisiko tidak lulus sejak dini, sehingga strategi akademik dapat dilakukan lebih tepat sasaran dan terukur. Dengan kombinasi *Linear Discriminant Analysis* (LDA) dan *Extreme Learning Machine* (ELM), penelitian ini diharapkan menghasilkan model prediksi yang lebih akurat dan efisien dibandingkan pendekatan tanpa reduksi dimensi, sehingga layak dijadikan referensi pengembangan sistem prediksi dini kelulusan di perguruan tinggi.

2. KAJIAN TEORI

2.1. Pendidikan

Pendidikan dapat dipahami sebagai sebuah sistem yang mencakup komponen masukan (input), proses, keluaran (output), umpan balik, serta lingkungan yang turut memberikan pengaruh terhadap keberlangsungan sistem tersebut [20]. Semua komponen tersebut saling berkaitan dan perlu bekerja sama secara baik dan optimal untuk mencapai dan menciptakan hasil yang diinginkan, yaitu pembentukan individu yang cerdas, berakhlak baik, dan memiliki keterampilan hidup yang sesuai dengan berkembangnya zaman [21]. Pendidikan tidak hanya memiliki keterkaitan dengan masa lalu maupun masa kini, tetapi pendidikan lebih penting keterkaitannya dengan masa mendatang. Dengan demikian pendidikan diberikan saat ini dengan merujuk dari ilmu pengetahuan terdahulu, kemudian pengetahuan tersebut diteruskan ke depannya sebagai acuan dalam pengambilan keputusan untuk masalah yang akan datang [22].

2.2. Missing Value

Missing value merupakan sebuah peristiwa ketika terdapat data yang tidak terisi atau bernilai kosong pada suatu dataset [23]. Keberadaan data yang hilang bisa mempengaruhi kualitas data serta meminimaliskan kinerja model yang dibangun. Salah satu tujuan penanganan *missing value* untuk menjaga kelengkapan dan konsistensi data sebelum dilakukan proses analisis atau pemodelan. Penanganan *missing value* dapat dilakukan melalui berbagai metode, seperti menghapus data yang tidak utuh atau mengisi nilai yang hilang dengan memakai pendekatan tertentu. Proses ini bertujuan menghasilkan dataset yang lebih representatif, mengurangi potensi bias, serta meningkatkan keandalan hasil analisis dan pemodelan [24].

2.3. Normalisasi Data

Proses normalisasi dilakukan sebelum tahap klasifikasi untuk menyamaratakan rentang nilai pada setiap atribut data. Normalisasi memiliki perbedaan skala antar atribut bisa dikurangi sehingga data menjadi lebih optimal untuk proses pembelajaran model [25]. Pada penelitian ini, normalisasi diterapkan memakai metode *min max normalisasi*, yang mengubah nilai data ke dalam rentang [0,1] menggunakan persamaan (1) berikut [26]:

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Keterangan:

x' = Data yang sudah dilakukan perubahan.

x_i = Data mula.

$\max(x)$ = Nilai paling besar dari seluruh data.

$\min(x)$ = Nilai paling kecil dari seluruh data.

2.4. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) adalah sebuah metode reduksi dimensi untuk meminimaliskan dimensi dengan mengombinasikan fitur yang memiliki informasi yang sama dan fitur yang saling berkaitan dengan cara mengubah fitur ruang dimensi yang lebih tinggi, menjadi ruang berdimensi yang lebih rendah [19]. *Linear Discriminant Analysis* (LDA) adalah sebuah metode analisis statistik guna mendapatkan kombinasi linier dari fitur-fitur untuk memisahkan data antar kelas ke dalam dua kelas [27].

Rumus *Between-Class Scatter Matrix* menggunakan persamaan (2):

$$S_B = \sum_{i=1}^c n_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (2)$$

Rumus *Within-Class Scatter Matrix* menggunakan persamaan (3):

$$S_W = \sum_{i=1}^c \sum_{j=1}^{n_i} (X_j^i - \mu_i)(X_j^i - \mu_i)^T \quad (3)$$

Rumus *Fisher's criterion* menggunakan persamaan (4):

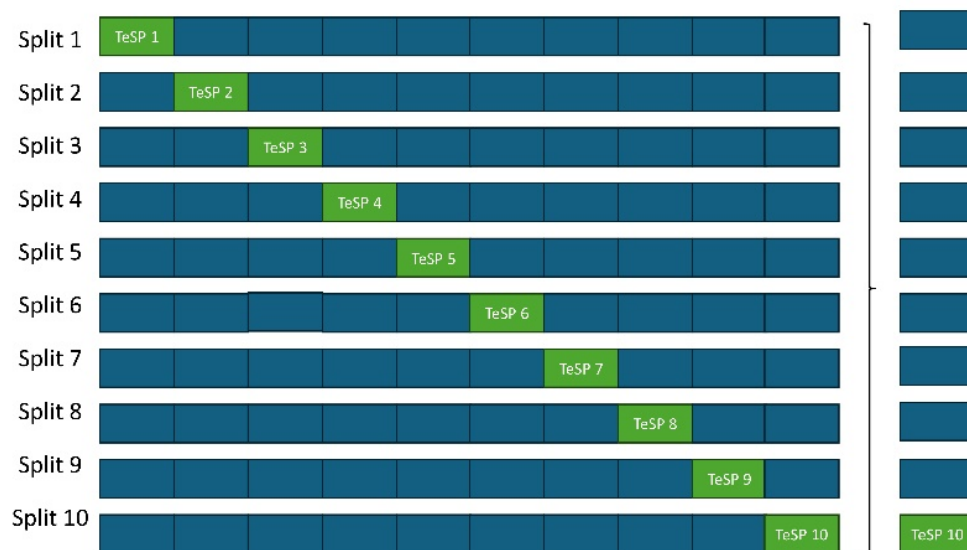
$$J(W) = \arg \max_W \frac{W^T S_B W}{W^T S_W W} \quad (4)$$

Rumus *Generalized Eigenvalue Problem* menggunakan persamaan (5):

$$S_W W = \lambda S_B W \quad (5)$$

2.5. K-Fold Cross Validation

K-Fold Cross Validation yaitu metode pengujian model yang dilakukan untuk menilai dan mengidentifikasi hasil statistik analisis yang melakukan generalisasi kumpulan data *independen* [28]. *K-Fold Cross Validation* dilakukan sebagai tujuan memperoleh hasil evaluasi model yang lebih optimal. Metode ini membagi data menjadi beberapa bagian, kemudian setiap bagian dilakukan dengan cara bergantian sebagai data uji. Nilai evaluasi dari seluruh pengujian tersebut dirata-ratakan untuk setiap hasil “k” iterasi, sehingga dapat menghasilkan estimasi performa model yang cenderung lebih stabil jika dibandingkan hanya melakukan sekali pembagian pada data menjadi data latih dan data uji [29].



Gambar 1. Simulasi K-Fold Cross Validation

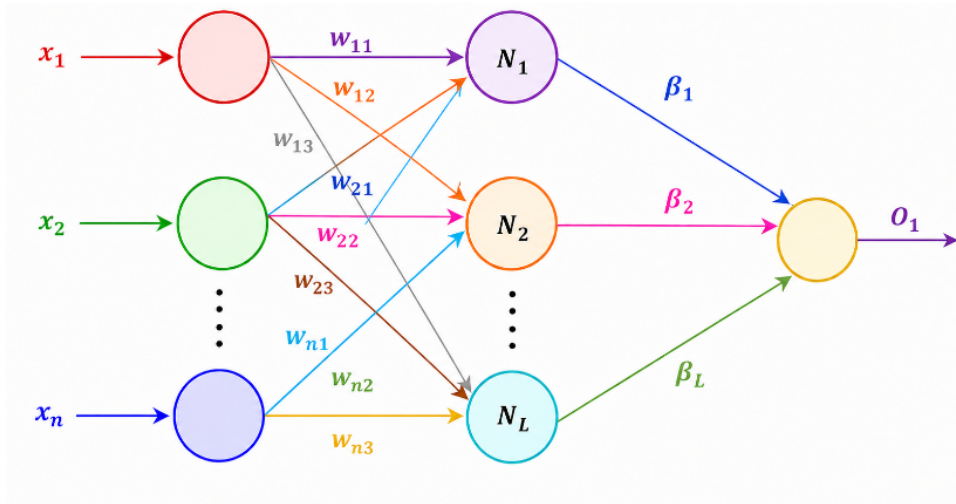
Gambar 1. memperlihatkan susunan Simulasi *K-Fold Cross Validation* yang diadaptasi dari ilustrasi pada [30], yang menunjukkan proses pembagian dataset menjadi sejumlah fold secara proporsional berdasarkan distribusi kelas.

2.6. Extreme Learning Machine (ELM)

Extreme Learning Machine (ELM) merupakan jenis metode jaringan saraf tiruan dengan hanya terdiri dari satu lapisan tersembunyi, *Extreme Learning Machine* (ELM) diterapkan sebagai pendekatan baru dalam pembelajaran jaringan saraf tiruan [16]. *Extreme Learning Machine* (ELM) yaitu metode yang kecepatan kinerja dalam pelatihnannya terbilang cepat dibandingkan dengan metode lainnya, serta memberikan generalisasi yang baik, hal tersebut memberikan hasil error tingkat error yang rendah [31]. Berikut arsitektur dari metode *Extreme Learning Machine* (ELM).

Gambar 2. memperlihatkan alur arsitektur jaringan yang terinspirasi dari ilustrasi pada [32], yang menggambarkan struktur jaringan *Extreme Learning Machine* (ELM) yang terdiri atas lapisan *input*, *hidden layer*, dan *output layer*. Berikut perhitungan metode *Learning Machine* (ELM):

Langkah 1: Menyusun matriks yang berisi data pelatihan, di mana x_i merepresentasikan fitur atau input dari data training, sedangkan t_i menunjukkan label atau target dari data training tersebut, seperti persamaan(6):



Gambar 2. Arsitektur Extreme Learning Machine

$$X_i = \begin{bmatrix} X_{i1} & X_{i2} & \dots & X_{in} \end{bmatrix}, \quad (6)$$

$$t_i = \begin{bmatrix} t_1 & t_2 & \dots & t_i \end{bmatrix}.$$

Langkah 2: Menetapkan banyaknya neuron pada lapisan tersembunyi (j), lalu membangkitkan bobot w_{ij} secara acak dalam bentuk matriks berukuran $i \times j$ dengan rentang nilai antara -1 hingga 1 . Sebagai contoh, jika $i = 3$ dan $j = 3$, maka akan diperoleh matriks berukuran 3×3 seperti persamaan (7):

$$W = \begin{bmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{bmatrix} \quad (7)$$

Langkah 3: Menyusun matriks awal keluaran dari neuron pada hidden layer (H_{init}), di mana x_i merepresentasikan data training yang diproses terhadap j neuron tersembunyi, w_{ij} menunjukkan bobot koneksi dari unit i menuju unit j , serta b_j sebagai nilai bias pada masing-masing *neuron*. Perhitungan ini dilakukan dengan menggunakan persamaan (8):

$$H_{init} = \sum_{i=0}^n (x_i w_{ij} + b_j) \quad (8)$$

Langkah 4: Melakukan perhitungan pada matriks keluaran dari *hidden neuron* (H) dengan menerapkan fungsi aktivasi. Adapun beberapa jenis fungsi aktivasi yang dipakai adalah sebagai berikut:

Rumus **Fungsi Sigmoid Biner** menggunakan persamaan(9):

$$g(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

Rumus **Fungsi Sigmoid Bipolar** menggunakan persamaan(10):

$$g(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (10)$$

Rumus **Fungsi Sin** menggunakan persamaan(11):

$$g(x) = \sin(x) \quad (11)$$

Dalam persamaan fungsi aktivasi tersebut, variabel x berasal dari hasil perhitungan matriks H_{init} .

Langkah 5: Penentuan bobot keluaran (β) dilakukan dengan menggunakan fungsi aktivasi $g(x)$ pada sejumlah j neuron tersembunyi, dimana j menunjukkan jumlah neuron pada hidden layer. Bobot w_{ij} merupakan penghubung antara *input layer* dan *hidden layer*, sedangkan x_i adalah data masukan yang berasal dari data pelatihan. Bias b_j digunakan pada *neuron* tersembunyi ke- j . Selain itu, t_i dan Y merepresentasikan kelas dari data training. Proses penentuan nilai tersebut dapat dilakukan dengan persamaan (12):

$$H_{init} = \sum_{j=0}^n \beta_j g(w_{ij}, x_i + b_j) = t_i = Y \quad (12)$$

Rumus lebih mudah menggunakan persamaan(13):

$$H\beta = Y \quad (13)$$

Melalui tahapan tersebut diperoleh sebuah matriks berukuran $i \times j$, di mana i menyatakan jumlah variabel input, sedangkan j menunjukkan banyaknya *neuron* pada *hidden layer*, dengan persamaan(14):

$$H = \begin{bmatrix} g(w_{11}x_1 + b_1) & \cdots & g(w_{1j}x_1 + b_j) \\ \vdots & \ddots & \vdots \\ g(w_{1d}x_d + b_1) & \cdots & g(w_{dj}x_d + b_j) \end{bmatrix} \quad (14)$$

Karena pada *Extreme Learning Machine* (ELM) jumlah node ditentukan secara acak, maka ketika jumlah node lebih sedikit dibandingkan jumlah data, matriks H yang terbentuk menjadi tidak berbentuk persegi (non-simetris). Untuk menangannya agar dapat diperoleh solusi yang stabil, digunakan konsep *Moore–Penrose Generalized Inverse*. Proses ini dilakukan dengan memakai persamaan matriks *Moore–Penrose inverse* yang dinyatakan dengan persamaan (15):

$$H^+ = (H^T H)^{-1} H \quad (15)$$

Sehingga diperoleh rumus untuk menentukan bobot keluaran dengan menggunakan persamaan(16):

$$\beta = H^+ Y \quad (16)$$

Langkah 6: Perhitungan matriks inisialisasi keluaran hidden neuron (H_{init}). Proses tersebut dilaksanakan dengan metode yang sama seperti pada langkah ke 3 yang diterapkan pada data *training*, namun pada tahap ini digunakan data *testing* (k) sehingga diperoleh persamaan(17):

$$H_{init} = \sum_{k=0}^n x_k w_{ji} \quad (17)$$

Langkah 7: Perhitungan matriks output hidden neuron (H) dilakukan berdasarkan persamaan pada langkah 4 tahap training dengan menggunakan nilai H_{init} hasil dari persamaan (17).

Langkah 8: Kelas prediksi (Y_k) ditentukan berdasarkan persamaan berikut dengan menggunakan matriks H yang berasal dari Persamaan (17), sehingga diperoleh persamaan(18):

$$Y_k = \beta H \quad (18)$$

2.7. Confusion Matrix

Confusion Matrix merupakan jenis metode yang dipakai untuk mengevaluasi performa algoritma klasifikasi, baik biner maupun multi-kelas, caranya dengan mengukur perbedaan hasil prediksi sistem dengan data sebenarnya [33]. *Confusion Matrix* berfungsi untuk memastikan tingkat akurasi hasil akhir dalam suatu penelitian. Kinerja suatu model algoritma dapat diukur melalui hasil *Confusion Matrix* yang disajikan dalam bentuk tabel matriks. Berdasarkan data yang digunakan, setiap baris pada matriks merepresentasikan kelas sebenarnya, sedangkan hasil prediksi ditampilkan pada setiap kolomnya [16]. *Confusion Matrix* juga memanfaatkan beberapa parameter evaluasi seperti *Recall*, *F1-Score*, *Precision*, dan *Accuracy*, sebagaimana diperlihatkan pada tabel *Confusion Matrix* berikut [34]:

KELAS A		PREDIKSI	
		A	B
AKTUAL	A	TP	FN
	B	FP	TN

KELAS B		PREDIKSI	
		A	B
AKTUAL	A	TN	FP
	B	FNs	TP

Gambar 3. Confusion Matrix

Confusion Matrix yaitu sebuah teknik evaluasi yang dilakukan guna mengukur performa model klasifikasi dengan memanfaatkan beberapa ukuran evaluasi utama, yaitu *accuracy*, *recall*, *F1-score*, serta *precision*. Dalam studi prediksi masa studi mahasiswa, keberhasilan model dapat diukur berdasarkan tinggi rendahnya nilai dari keempat metrik tersebut yang mencerminkan ketepatan proses klasifikasi yang diterapkan. *Confusion Matrix* menghasilkan empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) sebagai dasar dalam pengukuran performa model klasifikasi [35].

Confusion Matrix juga dapat digunakan untuk melakukan perhitungan terhadap sejumlah metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* yang berfungsi sebagai indikator kinerja model klasifikasi. *Accuracy* menggambarkan tingkat akurasi atau ketepatan suatu model dalam mengklasifikasikan data dengan benar dibandingkan total data yang diuji. *Precision* menggambarkan tingkat ketepatan hasil prediksi yang diklasifikasikan sebagai positif oleh model. *Recall* menunjukkan tingkat kemampuan model dalam menangkap seluruh sampel yang benar-benar berada pada kelas positif. *F1-score* yaitu ukuran gabungan yang menyeimbangkan *precision* dan *recall* serta dapat memberikan model performa yang lebih stabil, terutama pada data yang memiliki ketidakseimbangan kelas [36]:

Rumus *Accuracy* menggunakan persamaan(19):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

Rumus *Recall* menggunakan persamaan(20):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (20)$$

Rumus *Precision* menggunakan persamaan(21):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (21)$$

Rumus *F1 Score* menggunakan persamaan(22):

$$F1\text{-Score} = \frac{2TP}{2TP + FP + FN} \quad (22)$$

3. METODE PENELITIAN

3.1. Pengumpulan Data

Data yang diolah pada penelitian ini berupa dataset yang bersumber dari *Mendeley Data* (Elsevier Inc.) [37]. Terdiri atas 80 data yang merujuk terhadap prediksi status kelulusan mahasiswa. Data tersebut juga memiliki 19 fitur yang diantaranya terdapat persentase kehadiran, aktivitas akademik, tingkat akses pendidikan tinggi daerah, dan lain-lainnya.

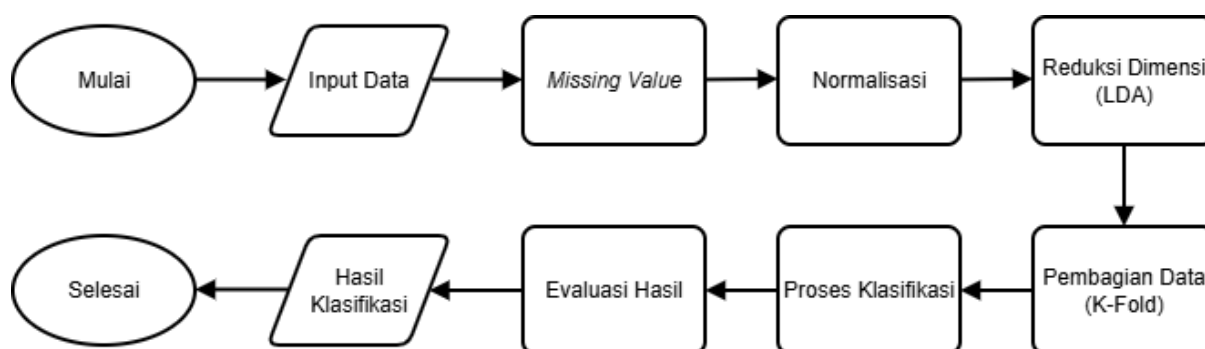
Tabel 1. Data Penelitian

No.	Jenis Kelamin	Usia	Tingkat Akses Pendidikan Tinggi Daerah	...	Persentase Kehadiran	Status Kelulusan
1.	0	38	5.0	⋮	25	1
2.	0	22	4.0	⋮	0	1
3.	0	21	4.0	⋮	41.6	1
4.	0	25	5.0	⋮	0	1
5.	0	20	5.0	⋮	91.6	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
76.	0	22	2.0	⋮	50	0
77.	0	22	4.0	⋮	100	0
78.	0	22	3.0	⋮	91.6	0
79.	1.0	21	5.0	⋮	66.6	0
80.	1.0	20	5.0	⋮	41.6	0

Tabel 1 menunjukkan data penelitian yang digunakan dalam proses analisis. Dataset terdiri dari 80 data dengan beberapa atribut yang menjadi variabel penelitian, seperti jenis kelamin, usia, tingkat akses pendidikan tinggi daerah, persentase kehadiran, dan atribut lainnya. Setiap atribut memiliki nilai numerik yang merepresentasikan karakteristik masing-masing responden. Pada kolom Status Kelulusan terdapat dua kelas, yaitu 1 yang menunjukkan mahasiswa lulus dan 0 yang menunjukkan mahasiswa tidak lulus. Data pada tabel kemudian digunakan sebagai masukan dalam praproses, reduksi dimensi, dan klasifikasi untuk membangun model prediksi status kelulusan mahasiswa.

3.2. Tahap Penelitian

Data pada Tabel 1 di atas akan dilakukan analisis menggunakan metode *Linear Discriminant Analysis* (LDA) sebagai langkah reduksi guna mentransformasi data yang berdimensi tinggi sehingga data lebih optimal untuk melakukan analisis klasifikasi dengan metode *Extreme Learning Machine* (ELM). Alur analisis metode tersebut ditampilkan ke dalam flowchart pada Gambar 4.



Gambar 4. Alur penelitian

Tahap kerja metode ini diawali dengan menginput data. Data kemudian diperiksa dan ditangani terlebih dahulu apabila terdapat *missing value* untuk menjamin bahwa tidak ada nilai yang hilang pada data yang dapat memengaruhi hasil analisis, kemudian data akan dilakukan proses normalisasi dimana pada penelitian ini diubah pada rentang nilai yang lebih kecil [0,1] dengan menggunakan metode *min max scaler*. Hasil data normalisasi yang telah didapatkan kemudian selanjutnya dilakukan reduksi dimensi data dengan menggunakan metode *Linear Discriminant Analysis* (LDA) yang pada dasarnya, metode ini memiliki tiga tahap utama. Pada tahap pertama, menghitung perbedaan antara rata-rata yang terdapat perberbedaan kelas, perbedaan ini dapat dikenal sebagai varians antar kelas. Pada tahap kedua, *Linear Discriminant Analysis* (LDA) menghitung perbedaan antara rata-rata dan nilai sampel dari setiap kelas, perbedaan ini juga dikenal sebagai varians dalam kelas. Pada tahap ketiga, dilakukan reduksi dimensi, penelitian ini memanfaatkan metode *Linear Discriminant Analysis* (LDA) yang berfokus pada optimalisasi pemisahan antar kelas dengan cara memperbesar variasi antar kelas sekaligus memperkecil variasi di dalam kelas. Melalui proses tersebut, sebanyak 19 fitur awal direduksi menjadi satu fitur baru yang dikenal sebagai *Linear Discriminant 1* (LD1). Hasil reduksi ini

selanjutnya dijadikan acuan dalam proses pembagian data menjadi data latih dan data uji memakai teknik *K-Fold Cross Validation*. Pada penelitian ini, nilai K yang dipakai berkisar antara 5 hingga 10 fold. Selain itu, pengujian juga dilakukan dengan variasi jumlah *hidden neuron* mulai dari *hidden neuron* 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 100, dan 500, untuk melihat pengaruhnya terhadap kinerja model. Pendekatan ini diterapkan agar evaluasi model yang dihasilkan menjadi lebih konsisten, lebih stabil, serta mampu mengurangi bias dibandingkan metode pembagian data tunggal. Setelah tahap pembagian pada data selesai diproses, data tersebut kemudian digunakan sebagai input pada tahap klasifikasi menggunakan *Extreme Learning Machine* (ELM).

Extreme Learning Machine (ELM) yaitu metode pembelajaran pada jaringan saraf yang tidak melakukan penyesuaian bobot pada bagian input, melainkan langsung menghitung bobot keluaran menggunakan pendekatan *Moore–Penrose pseudoinverse* pada sistem persamaan linier. Melalui proses ini, model akan menentukan neuron dengan kontribusi keluaran paling optimal berdasarkan hasil perhitungan tersebut. Tahapan dalam algoritma *Extreme Learning Machine* (ELM) terdapat dua tahap utama, yaitu fase pelatihan (*training*) dan fase pengujian (*testing*). Pada tahap pelatihan, model membentuk parameter yang diperlukan, sedangkan pada tahap pengujian model digunakan untuk melakukan prediksi terhadap data baru. Hasil akhir dari seluruh proses ini berupa hasil klasifikasi yang menunjukkan kategori dari data prediksi, dalam hal ini terkait status kelulusan mahasiswa. Dalam penelitian ini, evaluasi model juga menggunakan *K-Fold Cross Validation* untuk mengukur tingkat kesalahan prediksi secara lebih objektif. Teknik ini membagi data menjadi beberapa bagian dengan ukuran yang cenderung seimbang. Proses pelatihan dan pengujian menerapkan secara berulang sebanyak nilai K yang digunakan, di mana setiap bagian data secara bergantian digunakan sebagai data latih dan data uji. Setelah proses evaluasi selesai dilakukan, hasil prediksi kemudian dianalisis menggunakan *Confusion Matrix* untuk mengukur perbedaan hasil klasifikasi model dengan data sebenarnya. *Confusion Matrix* ini sebagai dasar untuk menghitung berbagai metrik evaluasi seperti *precision*, *recall*, *F1-score*, dan *accuracy*. Nilai-nilai tersebut dimanfaatkan guna mengukur besarnya tingkat ketepatan model dalam memprediksi status kelulusan mahasiswa.

4. HASIL DAN PEMBAHASAN

Penelitian ini memakai dataset yang terdiri dari 80 sampel data dengan 19 variabel fitur dan 2 kelas label status kelulusan. Data yang dipakai dalam proses penelitian, diawali dengan melakukan *preprocessing* dengan melakukan tahap pencarian *Missing Value* dengan menggunakan metode median agar tidak terdapat nilai yang hilang atau kosong pada data, sehingga data siap untuk tahap analisis selanjutnya, serta normalisasi data. Langkah pertama mengatasi *Missing Value* menggunakan median, hal tersebut dipilih karena lebih baik terhadap keberadaan nilai ekstrem dibandingkan rata-rata. Median mampu menjaga karakteristik distribusi data dan mengurangi dampak kehilangan informasi akibat data yang tidak lengkap.

Langkah dua yaitu proses normalisasi data menggunakan metode *Min-Max Normalization*. Proses ini bertujuan guna melakukan penskalaan ulang nilai data ke dalam rentang yang lebih kecil, yaitu antara 0 hingga 1. Dengan adanya normalisasi tersebut, setiap fitur memiliki skala yang seragam sehingga tidak terdapat variabel tertentu yang memberikan pengaruh lebih dominan dalam proses pembelajaran model akibat perbedaan satuan maupun rentang nilai. Proses normalisasi yang telah selesai dilakukan, seluruh data dibuat lebih konsisten sehingga siap diproses pada tahap analisis selanjutnya. Hasil dari proses normalisasi tersebut ditampilkan pada Tabel 2.

Tabel 2. Data Normalisasi

No.	Jenis Kelamin	Usia	Tingkat Akses Pendidikan Tinggi Daerah	...	Persentase Kehadiran	Status Kelulusan
1.	0.0	0.76	1.00	⋮	0.25	1
2.	0.0	0.12	0.75	⋮	0.0	1
3.	0.0	0,08	0.75	⋮	0.41	1
4.	0.0	0.24	1.00	⋮	0.00	1
5.	0.0	0,04	1.00	⋮	0.91	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
76.	0.00	0.12	0.25	⋮	0,50	0
77.	0.00	0.12	0.75	⋮	1,00	0
78.	0.00	0.12	0.50	⋮	0.91	0
79.	1.00	0.08	1.00	⋮	0.66	0
80.	1.00	0,04	1.00	⋮	0.41	0

Pertama sebelum dilakukan uji klasifikasi dengan reduksi dimensi menggunakan *Linear Discriminant Analysis* (LDA), akan dilihat terlebih dahulu hasil uji yang hanya menggunakan metode klasifikasi *Extreme Learning Machine* (ELM). Berikut hasil uji yang hanya menggunakan klasifikasi *Extreme Learning Machine* (ELM).

Tabel 3. Hasil Uji Coba dengan Menggunakan Extreme Learning Machine

<i>Hidden Neuron</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
3	80.00%	47.05%	51.95%	49.14%
4	85.00%	52.50%	60.00%	55.75%
5	82.50%	52.32%	58.57%	54.99%
6	83.75%	52.41%	59.29%	55.37%
7	85.00%	66.28%	66.90%	65.47%
8	85.00%	52.50%	60.00%	55.75%
9	88.75%	72.23%	73.45%	70.85%
10	83.75%	63.12%	64.29%	62.64%
11	82.50%	52.29%	58.57%	54.93%
12	85.00%	63.36%	65.24%	63.08%
13	86.25%	61.04%	66.67%	63.47%
14	87.50%	69.02%	68.45%	67.19%
15	86.25%	71.83%	76.90%	72.66%
20	85.00%	63.81%	66.79%	63.79%
25	81.25%	70.39%	68.93%	66.82%
30	83.75%	68.57%	70.45%	68.03%
40	77.50%	68.11%	67.95%	64.59%
50	65.00%	56.60%	52.17%	50.38%
60	58.75%	49.36%	46.20%	43.39%
70	55.00%	50.42%	43.15%	43.27%
75	51.25%	47.55%	40.47%	38.72%
80	53.75%	56.33%	56.33%	47.04%
90	51.25%	45.70%	35.77%	38.19%
100	48.75%	40.68%	44.17%	35.36%
500	56.25%	48.75%	45.70%	41.67%

Dari hasil uji yang didapat dengan menggunakan metode klasifikasi *Extreme Learning Machine* (ELM), hasil *accuracy* tertinggi sebesar 88.75%, *precision* 72.23%, *recall* 73.45%, dan *F1-score* 70.85% pada k-fold ke 10 di *hidden neuron* ke 9, hasil tersebut merupakan hasil yang terbilang baik. Uji coba berikutnya akan dilakukan menggunakan klasifikasi *Extreme Learning Machine* (ELM) dengan reduksi dimensi menggunakan *Linear Discriminant Analysis* (LDA).

Data yang telah dilakukan *preprocessing*, tahap selanjutnya yaitu mereduksi dimensi menggunakan *Linear Discriminant Analysis* (LDA), dilakukan untuk mentransformasi seluruh 19 variabel fitur menjadi 1 komponen diskriminan (LD1). *Linear Discriminant Analysis* (LDA) bukan untuk menghilangkan variabel, tetapi membentuk kombinasi linear baru dari seluruh variabel dengan bobot yang dioptimalkan untuk memaksimalkan pemisah antar kelas. Variabel dengan nilai terbesar memiliki kontribusi lebih dominan dalam membentuk komponen tersebut. Berdasarkan hasil analisis koefisien *Linear Discriminant Analysis* (LDA), menjelaskan bahwa komponen LD1 menjelaskan 100% variansi diskriminan. Variabel yang paling berdominan adalah Total waktu belajar pada sistem pembelajaran, Persentase aktivitas belajar dibandingkan rata-rata, Jumlah kehadiran, Persentase kehadiran, dan Jumlah ketidakhadiran. Pembuatan komponen *Linear Discriminant Analysis* (LDA) memiliki aturan dalam menentukan jumlah komponen yaitu sebesar jumlah kelas yang dikurangi 1 ($2 - 1 = 1$). Secara keseluruhan, satu komponen diskriminan mampu menjelaskan 100% informasi diskriminan antar kelas, sehingga tidak ada informasi yang akan hilang dalam proses transformasi ini. Data yang semula berdimensi 19 variabel berhasil direpresentasikan menjadi satu komponen diskriminan yang lebih ringkas namun tetap mempertahankan seluruh informasi pemisah antar kelas.

Tahap klasifikasi dilakukan setelah data melewati *preprocessing*. Penelitian ini memakai metode *Extreme Learning Machine* (ELM) guna memprediksi status kelulusan mahasiswa. Data yang digunakan sebanyak 80 data dibagi memakai *K-Fold Cross Validation* dengan $k = 6$. Langkah awal *Extreme Learning Machine* (ELM) yaitu pendekatan yang menginisialisasi bobot input dan bias *hidden layer* secara acak, serta menetapkan fungsi aktivasi dan jumlah neuron tersembunyi. Hasil percobaan beberapa *hidden neuron* dengan *K-Fold Cross Validation* yang dipakai memperlihatkan bahwa hasil nilai akurasi tertinggi terdapat di *K-Fold* 6 dengan jumlah *hidden neuron* sebanyak 7 neuron, dengan hasil *Accuracy*, *Precision*, *Recall*, dan *F1-Score* secara berturut-turut sebesar 90.11%, 70.54%, 70.08%, dan 69.49%.

Tabel 4. Hasil Uji Coba dengan Menggunakan ELM dan LDA

<i>Hidden Neuron</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
3	90.02%	76.64%	76.64%	73.29%
4	88.74%	76.51%	72.73%	72.86%
5	88.83%	70.41%	69.32%	69.06%
6	87.45%	76.36%	71.97%	72.39%

<i>Hidden Neuron</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
7	90.11%	70.54%	70.08%	69.49%
8	90.11%	70.54%	70.08%	69.49%
9	88.83%	67.76%	69.32%	67.43%
10	84.98%	63.47%	67.05%	63.39%
11	82.42%	59.84%	55.30%	56.03%
12	81.14%	56.37%	57.95%	57.95%
13	79.85%	54.15%	53.79%	52.98%
14	77.29%	53.84%	52.27%	52.04%
15	78.57%	53.96%	53.03%	52.47%
20	79.85%	53.40%	57.20%	54.90%
25	74.91%	47.03%	50.88%	48.14%
30	76.10%	49.67%	51.64%	50.42%
40	82.42%	54.38%	62.25%	57.16%
50	83.70%	58.38%	63.01%	60.26%
60	83.70%	54.49%	63.01%	57.55%
70	77.38%	59.27%	66.10%	60.82%
75	85.07%	71.51%	74.05%	70.53%
80	74.82%	55.03%	61.17%	56.13%
90	78.66%	56.65%	63.45%	58.50%
100	76.19%	76.19%	51.70%	49.62%
500	79.76%	55.95%	60.73%	57.51%

Berdasarkan hasil yang ditampilkan pada Tabel 4. dapat diketahui bahwa nilai akurasi tertinggi diperoleh pada konfigurasi *k-fold* 6 dengan jumlah *hidden neuron* sebanyak 7. Hasil tersebut menunjukkan bahwa kombinasi parameter *k-fold* 6 dan *hidden neuron* ke-7 merupakan konfigurasi yang menghasilkan performa klasifikasi terbaik, sehingga nilai akurasi tertinggi yang dicapai akan dibahas lebih lanjut sebagai berikut.

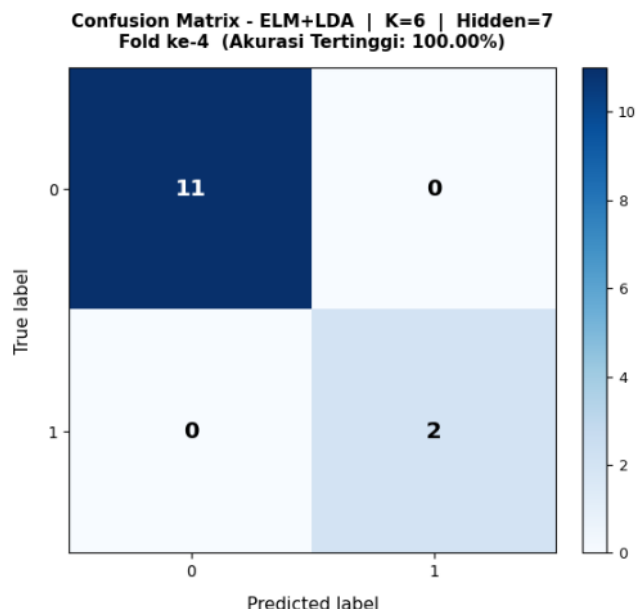
Tabel 5. Hasil Uji Coba dalam Menentukan K-Fold Optimal

<i>K-Fold Cross Validation</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	85.71%	42.86%	50.00%	46.15%
2	85.71%	42.86%	50.00%	46.15%
3	76.92%	41.67%	45.45%	43.48%
4	100.00%	100.00%	100.00%	100.00%
5	100.00%	100.00%	100.00%	100.00%
6	92.31%	95.83%	75.00%	81.16%

Pada Tabel 5 menyajikan hasil pengujian model ELM memakai *K-Fold Cross Validation* menggunakan nilai $K = 7$, dapat terlihat bahwa jumlah *hidden neuron* mempunyai dampak terhadap hasil kinerja klasifikasi. Secara keseluruhan, nilai *accuracy* berkisar 76,92% hingga 100%. Hasil paling optimal diperoleh pada fold ke 4 dan 5, tetapi berdasarkan waktu komputasinya *k-fold* 5 yang paling cepat, sehingga hasil paling optimal terdapat pada *k-fold* 5 dengan *accuracy*, *precision*, *recall*, dan *F1-score* masing-masing sebesar 100%, selain itu fold terendah terdapat pada fold 1 dan 2, tetapi pada waktu komputasinya fold ke 2 lah yang terendah dengan *accuracy* 85,71%, *precision* 42,86%, *recall* 50,00%, dan *F1-score* 46,15%. Berikut *Confusion Matrix* pada *k-fold* 5.

Berdasarkan Tabel 4 hasil analisis Linear Discriminant Analysis (LDA) dan Extreme Learning Machine (ELM) dengan variasi jumlah *hidden neuron*, dapat diketahui bahwa performa model mengalami perubahan seiring bertambahnya banyak neuron pada *hidden layer*. Hasil tersebut memperlihatkan bahwa nilai *accuracy*, *precision*, *recall*, dan *F1-score* relatif tinggi pada jumlah *hidden neuron* yang kecil, kemudian mengalami penurunan disaat banyak *hidden neuron* yang bertambah besar.

Hasil tersebut menunjukkan bahwa banyak *hidden neuron* yang terlalu besar tidak selalu meningkatkan kemampuan klasifikasi *Extreme Learning Machine* (ELM). Beberapa data direduksi memakai *Linear Discriminant Analysis* (LDA), informasi penting pada data telah terkonsentrasi dengan baik sehingga *Extreme Learning Machine* (ELM) cukup memakai jumlah neuron yang sedikit untuk membentuk model yang optimal. Beberapa banyak *hidden neuron* bertambah besar, kompleksitas model ikut meningkat. Kondisi ini bisa mengakibatkan model menjadi kurang stabil serta kemampuan generalisasi pada data uji menurun. Akibatnya, hasil *accuracy*, *precision*, *recall*, dan *F1-score* mengalami penurunan meskipun jumlah neuron terus ditambah. Pada penelitian selanjutnya, disarankan agar penggunaan jumlah *hidden neuron* tidak ditetapkan terlalu besar dalam proses pemodelan, serta mempertimbangkan metode ekstraksi fitur seperti *Linear Discriminant Analysis* (LDA) untuk meningkatkan kualitas data masukan. Salah satunya adalah mempertahankan performa klasifikasi, pendekatan tersebut juga mampu mengurangi kompleksitas model dan waktu komputasi [38].



Gambar 5. Hasil Confusion Matrix

5. KESIMPULAN

Pada bagian ini, dari apa yang diharapkan dari penelitian, yaitu mendapatkan hasil akurasi yang tinggi pada hasil analisisnya. Kondisi tersebut bisa ditinjau melalui hasil evaluasi yang didapatkan mempunyai nilai tinggi yaitu *Accuracy* 90.11%, *Precision* 70.54%, *Recall* 70.08%, dan *F1-Score* 69.49%. Penelitian ini menggunakan beberapa tahapan metode, yaitu *Linear Discriminant Analysis* (LDA) untuk reduksi dimensi, pembagian data menggunakan *K-Fold Cross Validation*, proses klasifikasi dengan *Extreme Learning Machine* (ELM), serta *Confusion Matrix* sebagai alat evaluasi kinerja model. Pada tahap analisis, *Linear Discriminant Analysis* (LDA) berperan dalam mereduksi fitur menjadi representasi yang lebih sederhana, sedangkan *Extreme Learning Machine* (ELM) digunakan sebagai teknik klasifikasi data. Kombinasi antara metode *Linear Discriminant Analysis* (LDA) dan *Extreme Learning Machine* (ELM) memiliki hasil akurasi lebih tinggi sebesar 90.11% dibandingkan hanya menggunakan *Extreme Learning Machine* (ELM) sebesar 88,75%. Hasil dari penerapan kedua metode tersebut memperlihatkan bahwa model yang dibangun mempunyai kemampuan yang baik disaat melakukan prediksi pada kasus status kelulusan mahasiswa yang dinyatakan lulus atau tidak lulus, dengan kesalahan yang relatif rendah (9,89%). Perencanaan untuk mengurangi jumlah ketidakkelulusan mahasiswa semoga mendapatkan manfaat dari hasil penelitian ini.

DAFTAR PUSTAKA

- [1] Panoyo, "Pendidikan Sebagai Pilar Pembangunan Bangsa Indonesia Menuju Masa Depan Berkelanjutan," *Jurnal Studi Pendidikan Dasar*, vol. 2, no. 1, pp. 33–44, Jun. 2024, doi: 10.54180/Jsped.V2i1.499.
- [2] M. I. Hasbullah and V. Yasin, "Prediksi Kelulusan Mahasiswa Menggunakan Machine Learning," *Jurnal Teknologi Informasi Dan Komunikasi*, vol. 16, no. 2, pp. 36–39, Sep. 2025, doi: 10.51903/Jtikp.V16i2.1025.
- [3] R. Deviyanti, B. Saputra, and N. P. R. Agustina, "Pelatihan Manajemen Studi Dan Pendampingan Tugas Akhir Untuk Mewujudkan Kelulusan Tepat Waktu Mahasiswa Universitas Lampung," vol. 6, no. 2, 2025.
- [4] "Prediksi Kelulusan Tepat Waktu Mahasiswa Untuk Pemantauan Program Studi Menggunakan Metode Data Mining | Aiti." *Jurnal Teknologi Informasi*, vol. 21 no. 2, 2024, doi: 10.24246/aiti.v21i2.168-182.
- [5] "Data Jumlah Mahasiswa Drop Out Tahun 2025 - Portal Data Pendidikan Tinggi, Sains, Dan Teknologi." Accessed: Jun. 10, 2026. [Online]. Available: <https://Data.Kemdiktisaintek.Go.Id/Dataset/Data-Jumlah-Mahasiswa-Drop-Out-Tahun-2025>.
- [6] Gianto and R. Kurniawan, "Peran Ban-Pt Dalam Penjaminan Mutu Pendidikan Tinggi Di Indonesia," *Jpnm Jurnal Pustaka Nusantara Multidisiplin*, vol. 3, no. 4, 2025, doi: 10.59945/Jpnm.V3i4.859.
- [7] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *Skanika: Sistem Komputer Dan Teknik Informatika*, vol. 6, no. 1, pp. 61–74, Jan. 2023, doi: 10.36080/Skanika.V6i1.2976.
- [8] F. B. Kurniawan and L. Farokhah, "Aplikasi Cerdas Prediksi Kelulusan Mahasiswa Berbasis Website Menggunakan Metode Support Vector Machine (Svm)," *Jurnal Fasilkom*, vol. 15, no. 1, pp. 155–162, Apr. 2025, doi: 10.37859/Jf.V15i1.8767.
- [9] "Jangan Terlewat! Masa Tenggat Transisi Akreditasi Oleh Ban-Pt Pada 2025 Mendatang | Sevima." Accessed: Jun. 25, 2026. [Online]. Available: <https://Sevima.Com/Jangan-Terlewat-Masa-Tenggat-Transisi-Akreditasi-Oleh-Ban-Pt-Pada-2025-Mendatang/>
- [10] S. Imron, A. Faizah, and S. Sugianto, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma XGBoost," *Skanika: Sistem Komputer Dan Teknik Informatika*, vol. 9, no. 1, pp. 76–86, Jan. 2026, doi: 10.36080/Skanika.V9i1.3647.
- [11] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma Decision Tree Dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 1, pp. 21–28, Jan. 2023, doi: 10.33379/Gtech.V7i1.1850.
- [12] J. Jefri and Z. Fatah, "Klasifikasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes," *Jurnal Ilmiah Multidisiplin Ilmu*, vol. 2, no. 1, pp. 29–37, Feb. 2025, doi: 10.69714/Mhjqlv85.

- [13] I. A. K. Adi and W. A. E. Prabowo, "Klasifikasi Penyakit Diabetes Menggunakan Pendekatan Pembelajaran Mesin Dengan Model Non-Linier," *Jurikom (Jurnal Riset Komputer)*, vol. 12, no. 3, pp. 262–268, Jun. 2025, doi: 10.30865/Jurikom.V12i3.8586.
- [14] G. H. Hilmawan, "Literatur Review : Efektifitas Penerapan Metode Naïve Bayes Untuk Prediksi Kelulusan," *Jurnal Media Akademik (Jma)*, vol. 3, no. 6, Jun. 2025, doi: 10.62281/V3i6.2434.
- [15] K. Kireyna Cindy Pradhisa and R. Yotenka, "Penerapan Metode Extreme Learning Machine (ELM) Untuk Meramalkan Produksi Air Minum Dan Air Yang Dijual Di Kota Yogyakarta Tahun 2023: Penerapan Metode Extreme Learning Machine (ELM) Untuk Meramalkan Produksi Air Minum Dan Air Yang Dijual Di Kota Yogyakarta Tahun 2023," *Esds*, vol. 2, no. 2, pp. 258–263, Jun. 2024, doi: 10.20885/Esds.Vol2.Iss.2.Art19.
- [16] R. R. Jannah, M. Z. Sholahuddin, D. Z. Haq, and D. C. R. Novitasari, "Klasifikasi Kualitas Udara Menggunakan Metode Extreme Learning Machine (ELM)," *Methodika: Jurnal Teknik Informatika Dan Sistem Informasi*, vol. 10, no. 2, pp. 19–23, Sep. 2024, doi: 10.46880/Mtk.V10i2.3066.
- [17] M. S. Maulana, B. I. Nugroho, and S. Surejo, "Sistem Klasifikasi Jenis Kendaraan Melalui Teknik Olah Citra Digital," *Jurnal Minfo Polgan*, vol. 11, no. 2, pp. 89–99, Sep. 2022, doi: 10.33395/Jmp.V11i2.11793.
- [18] M. Warohma, E. Budianita, F. Syafria, and I. Afrianty, "Prediksi Jumlah Perceraian Menggunakan Metode Extreme Learning Machine (ELM)," *Josh*, vol. 4, no. 4, pp. 1448–1454, Jul. 2023, doi: 10.47065/Josh.V4i4.3581.
- [19] W. Winarnie, K. Kusrini, and A. D. Hartanto, "Pengurangan Dimensi Dengan Metode Linear Discriminant Analist (LDA)," *Infotek*, vol. 6, no. 2, pp. 228–237, Jul. 2023, doi: 10.29408/Jit.V6i2.10069.
- [20] S. Handayani, A. M. Akbar, and N. Septia, "Konsep Pendidikan Sebagai Suatu Sistem Dan Komponen Sistem Pendidikan," *Jurnal Bintang Pendidikan Indonesia*, vol. 3, no. 3, pp. 41–48, Jan. 2025, doi: 10.55606/Jubpi.V3i1.3606.
- [21] N. Aliah, R. N. R. Sari, A. Rosdiana, M. Mustafa, and A. Haling, "Literature Review About Pedagogical Components | Pinisi Journal Of Education," May 2025, Accessed: Jun. 14, 2026. [Online]. Available: <https://Journal.Unm.Ac.Id/Index.Php/Pje/Article/View/8055>
- [22] "Kajian Ilmiah Tentang Ilmu Pendidikan | Jurnal Cakrawala Pendidikan." Accessed: Jun. 08, 2026. [Online]. Available: <https://Journal.Uny.Ac.Id/Index.Php/Cp/Article/View/9125>.
- [23] M. R. A. Prasetya, A. M. Priyatno, and Nurhaeni, "Penanganan Imputasi Missing Values Pada Data Time Series Dengan Menggunakan Metode Data Mining," *Jurnal Informasi Dan Teknologi*, pp. 52–62, Jun. 2023, doi: 10.37034/Jidt.V5i2.324.
- [24] A. Sya'adah, A. A. Fadhillah, and I. Daulay, "Prediksi Penyakit Jantung Dengan Random Forest Melalui Teknik Imputasi Dan Feature Scaling Pada Preprocessing Data," vol. 1, 2024.
- [25] I. Permiana and F. N. S. Salisah, "Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation: The Effect Of Data Normalization On The Performance Of The Classification Results Of The Backpropagation Algorithm," *Indonesian Journal Of Informatic Research and Software Engineering (Ijirse)*, vol. 2, no. 1, pp. 67–72, Mar. 2022, doi: 10.57152/Ijirse.V2i1.311.
- [26] P. p. Allorerung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis Performa Normalisasi Data Untuk Klasifikasi K-Nearest Neighbor Pada Dataset Penyakit," *Jiska (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, Sep. 2024, doi: 10.14421/Jiska.2024.9.3.178-191.
- [27] "Analisis Diskriminan Linier — Dokumentasi Multivariatestats 0.1.0." Accessed: Jun. 10, 2026. [Online]. Available: <https://Multivariatestatsjl.Readthedocs.Io/En/Latest/Lda.Html>
- [28] R. Tuntun, K. Kusrini, and K. Kusnawi, "Analisis Perbandingan Kinerja Algoritma Klasifikasi Dengan Menggunakan Metode K-Fold Cross Validation," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, pp. 2111–2119, Oct. 2022, doi: 10.30865/Mib.V6i4.4681.
- [29] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation Untuk Mengevaluasi Kinerja Mahasiswa | Jurnal Algoritma", Accessed: Jun. 10, 2026. [Online]. Available: <https://Jurnal.Itg.Ac.Id/Index.Php/Algoritma/Article/View/1618>.
- [30] H. T. Santoso, F. A. Felmidi, A. N. Fadhila, A. Ristyawan, and E. Daniati, "Analisis Kinerja Algoritma Data Mining Pada Klasifikasi Tingkat Obesitas Dengan K-Fold Cross Validation Dan Auc," *Prosiding Semnas Inotek (Seminar Nasional Inovasi Teknologi)*, vol. 8, no. 1, pp. 113–122, Jul. 2024, doi: 10.29407/Inotek.V8i1.4917.
- [31] "Perbandingan Metode Extreme Learning Machine (ELM) Dan Kernel Extreme Learning Machine (Kelm) Pada Klasifikasi Penyakit Cedera Panggul | Jurnal Fourier." Accessed: Jun. 10, 2026. [Online]. Available: <https://Mail.Fourier.Or.Id/Index.Php/Fourier/Article/View/184>
- [32] V. Novykov, C. Bilson, A. Gepp, G. Harris, and B. J. Vanstone, "Empirical Validation Of Elm Trained Neural Networks For Financial Modelling," *Neural Comput & Applic*, vol. 35, no. 2, pp. 1581–1605, Jan. 2023, doi: 10.1007/S00521-022-07792-3.
- [33] M. R. U. Pulungan, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Ulasan Aplikasi Pedulilindungi Dengan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 6, no. 9, pp. 4378–4385, Sep. 2022.
- [34] I. Rahmawan, "Verifikasi Klasifikasi Tiga Kategori Curah Hujan Menggunakan Confusion Matrix = Verification Of Three Rainfall Categories Classification Using Confusion Matrix," Other, Universitas Hasanuddin, 2025. Accessed: Jun. 11, 2026. [Online]. Available: <https://Repository.Unhas.Ac.Id/Id/Eprint/50483/>
- [35] "(Pdf) Komparasi Algoritma Machine Learning Dalam Klasifikasi Loyalitas Nasabah Bank Berbasis Particle Swarm Optimization," *Researchgate*, Jun. 2026, doi: 10.36040/Jati.V8i4.10941.
- [36] "Klasifikasi: Akurasi, Perolehan, Presisi, Dan Metrik Terkait | Machine Learning," Google For Developers. Accessed: Jun. 13, 2026. [Online]. Available: <https://Developers.Google.Com/Machine-Learning/Crash-Course/Classification/Accuracy-Precision-Recall?hl=Id>
- [37] N. Nnamoko and J. Barrowclough, "Student Grade Prediction Dataset," vol. 1, Jun. 2022, doi: 10.17632/Wf8568hxb7.1.

- [38] Y. Liu and S. Guo, “Weighted Linear Discriminant Analysis: An Effective Feature Extraction Method For Multi-Class Imbalanced Datasets,” *Symmetry*, vol. 16, no. 12, p. 1656, Dec. 2024, doi: 10.3390/Sym16121656.