JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis Perbandingan Seleksi Fitur dalam Memprediksi Kelulusan Mahasiswa dengan Menngunakan Artificial Neural Network

M. Khoirul Risqi, Ifnu Wisma Dwi Prastya, Mula Agung Barata*

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
Email: [1]risqisgb@email.com, [2]ifnuprastya@unugiri.ac.id, [3,*]mula.ab26@email.com
Email Penulis Korespondensi: risqisgb@email.com
Submitted **16-12-2025;** Accepted **27-01-2026;** Published **28-02-2026**

Abstrak

Fenomena mahasiswa yang menghentikan studi di perguruan tinggi merupakan persoalan serius dalam dunia pendidikan tinggi karena secara langsung mempengaruhi mutu akademik serta rasio kelulusan. Oleh sebab itu, identifikasi dini terhadap mahasiswa yang memiliki potensi untuk menghentikan kuliah menjadi sangat krusial agar tindakan pencegahan dapat dilakukan secara tepat. Studi ini bermaksud mengevaluasi dan membandingkan efektivitas tiga pendekatan seleksi fitur, yakni Information Gain, Chi-Square, dan ANOVA, dalam mengoptimalkan performa model Jaringan Saraf Tiruan (ANN) untuk klasifikasi status mahasiswa. Dataset yang digunakan mencakup data akademik dan administratif yang telah melalui tahap praproses dan normalisasi menggunakan teknik Min-Max Scaling. Berdasarkan hasil uji coba, seluruh metode menunjukkan performa memadai dengan akurasi berkisar antara 88,71% hingga 91,37%. Di antara ketiganya, Information Gain memperlihatkan performa paling unggul, mencapai akurasi 91,37% dan nilai recall sebesar 97,29%, karena kemampuannya menyaring atribut yang paling signifikan melalui pengurangan entropi. Metode ANOVA juga tampil stabil dengan akurasi 90,82%, sedangkan Chi-Square cenderung kurang optimal karena sensitivitas terhadap variabel kategorikal. Temuan ini menegaskan bahwa proses seleksi atribut memainkan peranan vital dalam meningkatkan akurasi dan efisiensi model ANN untuk mendeteksi risiko mahasiswa yang berpotensi keluar dari pendidikan tinggi. Studi ini memberikan implikasi praktis dalam pembangunan sistem prediktif berbasis data untuk mendukung strategi retensi mahasiswa.

Kata Kunci: ANOVA; Artificial Neural Network; Chi-Square; Dropout; Information Gain

Abstract

Student attrition presents a major challenge in higher education due to its direct impact on academic quality and institutional graduation rates. Detecting students who are likely to withdraw at an early stage is therefore essential to ensure that timely interventions can be made. This study investigates how three distinct feature selection techniques—Chi-Square, Information Gain, and ANOVA—affect the performance of Artificial Neural Networks (ANN) in classifying student outcomes. The data used in the experiment were drawn from academic and administrative records, which had been standardized through Min-Max normalization. The results demonstrate that each method contributes positively, with classification accuracies ranging from 88.71% to 91.37%. Information Gain emerged as the most effective approach, yielding the highest accuracy at 91.37% and a recall score of 97.29%, largely due to its capability to reduce entropy and isolate the most informative variables. ANOVA also performed consistently well with 90.82% accuracy, while Chi-Square was comparatively less effective, potentially due to its reliance on categorical variables that may not capture predictive nuances. These findings emphasize the strategic importance of applying robust feature selection to improve ANN-based prediction models. Ultimately, this research supports the design of data-driven systems aimed at reducing student dropout rates and strengthening academic retention strategies across higher education institutions.

Keywords: ANOVA; Artificial Neural Network; Chi-Square; Dropout; Information Gain

1. PENDAHULUAN

Peningkatan jumlah mahasiswa di perguruan tinggi dari tahun ke tahun menunjukkan kemajuan akses terhadap pendidikan tinggi di Indonesia. Namun, peningkatan tersebut tidak selalu diiringi dengan keberhasilan penyelesaian studi secara tepat waktu. Banyak perguruan tinggi masih menghadapi tantangan serius terkait fenomena mahasiswa yang tidak menuntaskan studinya atau yang dikenal sebagai *Dropout* / putus studi. Kondisi ini menjadi perhatian karena berdampak langsung pada efisiensi internal perguruan tinggi, citra institusi, serta tingkat kepercayaan masyarakat terhadap mutu penyelenggaraan pendidikan tinggi [1]. Selain itu, persoalan *Dropout* juga mencerminkan efektivitas sistem pembelajaran dan kebijakan akademik, yang secara tidak langsung mempengaruhi kinerja lembaga dalam memenuhi target Indikator Kinerja Utama (IKU) perguruan tinggi.

Secara konseptual, *Dropout* diartikan sebagai kondisi ketika mahasiswa menghentikan aktivitas akademiknya sebelum menyelesaikan program studi dan memperoleh gelar akademik. Beberapa literatur membedakan pendekatan makro dan mikro dalam mendefinisikan *Dropout*. Pendekatan makro memandang fenomena ini sebagai penghentian total studi tanpa mempertimbangkan kemungkinan mahasiswa berpindah program atau universitas [2]. Sebaliknya, pendekatan mikro mengkategorikan mahasiswa yang berpindah program studi atau institusi lain sebagai bagian dari kelompok *Dropout*, karena secara administratif mahasiswa tersebut tidak menyelesaikan program pada institusi awalnya. Perbedaan pendekatan ini penting, sebab berdampak pada pengukuran tingkat *retention rate* dan penentuan strategi pencegahan yang tepat di lingkungan perguruan tinggi.

Dalam konteks manajemen mutu pendidikan, upaya meminimalkan tingkat *Dropout* merupakan langkah strategis untuk menjaga kualitas lulusan dan efektivitas sistem akademik. Salah satu pendekatan modern yang kini banyak digunakan adalah analisis berbasis data [3]. Dari perspektif ilmu komputer, permasalahan ini dapat ditangani melalui pendekatan komputasional yang mampu memproses data dalam jumlah besar secara efisien serta mengekstraksi informasi bermakna dari pola-pola yang tidak tampak secara langsung. Pendekatan tersebut memberikan landasan metodologis yang

kuat untuk merumuskan model analitik yang tidak hanya bersifat diagnostik, tetapi juga prediktif, sehingga memungkinkan institusi melakukan intervensi secara lebih dini dan terukur. Melalui penerapan *data mining* / penambangan data, institusi pendidikan dapat mengidentifikasi pola tersembunyi dari data akademik mahasiswa, seperti nilai akademik, kehadiran, aktivitas pembelajaran daring, maupun faktor sosial ekonomi. *Data mining* memungkinkan proses analisis yang lebih objektif dan sistematis dalam menemukan keterkaitan antara variabel-variabel penentu keberhasilan studi. Dengan demikian, hasil analisis tersebut dapat dijadikan dasar untuk pengambilan keputusan dalam merancang kebijakan pencegahan *Dropout* [4].

Secara umum, *data mining* melibatkan serangkaian tahapan seperti prapemrosesan data (*data preprocessing*), seleksi fitur (*feature selection*), pembentukan model klasifikasi (*modeling*), serta evaluasi performa model. Di antara tahapan tersebut, seleksi fitur memegang peran penting karena berfungsi memilih atribut yang paling relevan terhadap target klasifikasi. Penggunaan fitur yang tidak relevan dapat menyebabkan kompleksitas model meningkat, waktu komputasi bertambah, serta akurasi menurun. Beberapa studi terdahulu menunjukkan bahwa penerapan seleksi fitur seperti *Information Gain, Chi-Square*, dan *ANOVA* mampu meningkatkan kinerja algoritma tradisional menunjukkan adanya peningkatan tingkat akurasi rata-rata yang signifikan yaitu 2,45 %. Temuan ini mengindikasikan bahwa penyeleksi fitur bukan hanya sekadar tahap awal, melainkan komponen krusial yang menentukan kualitas prediksi model [5].

Penelitian terkini yang dilakukan oleh Sulfayanti et al. [6] menggunakan dataset yang sama, yaitu *Student Study Status Dataset* dari *UCI Machine Learning Repository*, penerapan metode seleksi fitur *Mutual Information* pada algoritma *Naïve Bayes* terbukti mampu memberikan peningkatan kinerja akurasi model secara bermakna. Dalam penelitian tersebut, *Mutual Information* digunakan untuk mengeliminasi atribut yang kurang relevan sehingga model menjadi lebih sederhana tanpa kehilangan informasi penting. Hasil eksperimen memperlihatkan peningkatan akurasi dari 68,29 % menjadi 72,65 % dengan penggunaan 15 atribut terpilih. Temuan ini menegaskan bahwa proses seleksi fitur berperan penting dalam menyederhanakan dimensi data (*feature dimensionality reduction*) sekaligus mempertahankan relevansi informasi terhadap variabel target, yang secara langsung berkontribusi terhadap peningkatan performa model klasifikas.

Di lain pihak, kemajuan dalam ranah algoritma kecerdasan artifisial turut membuka cakrawala baru bagi peningkatan ketepatan dalam proses pengelompokan data. Salah satu pendekatan komputasional yang menunjukkan performa unggul adalah Jaringan Syaraf Tiruan (JST). JST memiliki kapabilitas dalam menafsirkan relasi yang tidak linear dan bersifat kompleks di dalam sekumpulan data, menjadikannya alat yang efektif untuk menyarikan pola tersembunyi yang sulit ditangkap oleh metode konvensional. Melalui mekanisme pembelajaran berbasis bobot dan aktivasi neuron. Dalam beberapa penelitian, ANN terbukti mampu mengungguli algoritma tradisional seperti *Support Vector Machine* (SVM) atau *Decision Tree*. Misalnya, penelitian pada klasifikasi citra tumor payudara menunjukkan bahwa ANN mampu mencapai akurasi sebesar 87,78 %, melampaui SVM dengan akurasi 86,67 %. Selain itu, ANN juga memiliki tingkat sensitivitas dan stabilitas prediksi yang lebih baik dalam menghadapi data beragam dan ber-noise [7]. Kemampuan tersebut menjadikan ANN relevan untuk diterapkan pada permasalahan kompleks di bidang pendidikan, termasuk prediksi *student Dropout*.

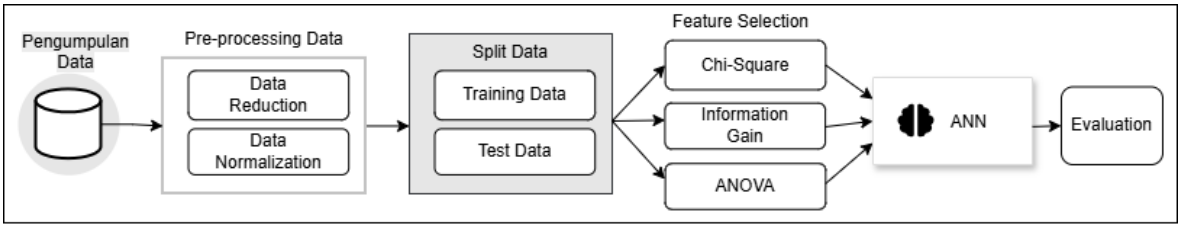
Keberhasilan model prediksi tidak hanya menggantungkan pada pemilihan algoritma, tetapi juga kualitas tahapan prapemrosesan data. Penanganan *missing values*, normalisasi, serta pemilihan fitur berpengaruh besar terhadap hasil akhir model [8]. Ketidaktepatan dalam pengolahan awal data dapat menurunkan akurasi secara signifikan meskipun algoritma yang digunakan tergolong canggih. Selain itu, pada banyak kasus ditemukan adanya fitur dengan nilai korelasi yang hampir serupa, sehingga bila pemilihan fitur dilakukan dengan metode konvensional tanpa pengujian statistik yang kuat, hasil seleksi bisa menjadi tidak stabil dan sulit direproduksi [9]. Oleh karena itu, diperlukan pendekatan penyeleksian fitur yang tidak hanya mempertimbangkan kekuatan hubungan antarvariabel, tetapi juga kestabilan hasil dan kontribusi statistiknya terhadap variabel target. Meskipun penelitian sebelumnya telah membuktikan bahwa seleksi fitur berbasis *Mutual Information* yang dipadukan dengan algoritma *Naïve Bayes* mampu meningkatkan kinerja prediksi, kajian tersebut belum mengkaji apakah metode seleksi fitur tersebut tetap menjadi pilihan paling optimal ketika diterapkan pada algoritma *Artificial Neural Network* (ANN) yang memiliki karakteristik pembelajaran berbeda. Oleh karena itu, masih terdapat celah penelitian untuk melakukan evaluasi komparatif terhadap berbagai metode seleksi fitur guna menentukan pendekatan dengan kinerja paling unggul dalam kerangka pemodelan ANN. Berdasarkan landasan tersebut, penelitian ini difokuskan pada perbandingan tiga metode seleksi fitur yaitu *Information Gain, Chi-Square*, dan *ANOVA*, yang masing-masing mewakili pendekatan statistik dan entropi berbeda. Ketiga pendekatan seleksi fitur tersebut akan dievaluasi dalam kerangka pemodelan Jaringan Syaraf Tiruan (JST) guna meramalkan kemungkinan mahasiswa menghentikan studi di lingkungan perguruan tinggi. Fokus utama dari penelitian ini diarahkan pada konteks pendidikan tinggi, dengan sasaran untuk mengidentifikasi metode seleksi atribut yang paling efektif dalam memaksimalkan kinerja JST. Dengan demikian, model yang dihasilkan diharapkan mampu memberikan keluaran klasifikasi yang lebih presisi dan konsisten. Luaran dari studi ini diharapkan dapat dimanfaatkan sebagai referensi strategis bagi institusi pendidikan tinggi dalam merancang sistem peringatan dini (*early warning system*) berbasis algoritma pembelajaran mesin. Sistem ini bertujuan untuk mengenali secara lebih dini kelompok mahasiswa yang tergolong berisiko tinggi mengalami *Dropout*, sehingga memungkinkan diterapkannya langkah-langkah mitigasi yang lebih tepat, kontekstual, dan berorientasi pada pencegahan kegagalan studi.

2. METODOLOGI PENELITIAN

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.1 Alur Penelitian

Rangkaian prosedur yang digunakan dalam penelitian ini disusun secara sistematis untuk merepresentasikan secara menyeluruh setiap fase yang dilalui dalam rangka mencapai sasaran studi. Setiap tahapan metodologis yang telah dirancang secara komprehensif dijabarkan secara visual dan terstruktur pada Gambar 1, guna memberikan pemahaman yang lebih mendalam mengenai alur pelaksanaan riset secara keseluruhan.



Gambar 1. Alur Penelitian

2.2 Pengumpulan Data

Sumber data yang dimanfaatkan dalam riset ini berasal dari dataset terbuka berjudul *Predict Students' Dropout and Academic Success*, yang tersedia melalui repositori *UCI Machine Learning*. Dataset tersebut memuat informasi akademik mahasiswa dari salah satu institusi pendidikan tinggi di Portugal, dan digunakan sebagai dasar untuk menganalisis berbagai variabel yang berperan dalam keberhasilan studi serta kemungkinan mahasiswa mengalami Dropout. Pemanfaatan dataset ini memungkinkan peneliti untuk menguji hubungan antar faktor internal dan eksternal yang memengaruhi kelangsungan pendidikan mahasiswa secara lebih terukur dan berbasis data nyata. Total terdapat 4.424 entri data dengan 36 variabel fitur dan satu variabel target yang terdiri atas tiga kategori, yaitu *Graduate*, *Dropout*, dan *enrolled* [10]. Pemilihan dataset ini didasarkan pada ketersediaannya secara publik dan dokumentasi yang lengkap, sehingga memungkinkan penelitian ini untuk direplikasi dan dibandingkan dengan studi-studi selanjutnya [11].

Setiap variabel dalam dataset merepresentasikan karakteristik demografis, sosial, ekonomi, dan akademik mahasiswa, serta indikator makroekonomi yang berkaitan dengan kondisi eksternal pada saat mahasiswa menjalani masa studi.

2.3 Pra-Pemrosesan Data

Pre-processing proses awal dalam kegiatan *data mining* berfokus pada transformasi data mentah menjadi informasi yang siap diolah hasil pengumpulan dari berbagai sumber menjadi data yang lebih terstruktur, bersih, dan siap digunakan pada tahapan analisis berikutnya [12]. Tahapan ini mencakup sejumlah proses utama, antara lain pereduksian data guna menyederhanakan informasi tanpa menghilangkan esensi pentingnya, transformasi data untuk menyesuaikan bentuk maupun struktur data dengan kebutuhan analisis, integrasi data untuk menghimpun informasi yang berasal dari berbagai sumber, serta pembersihan data untuk menangani ketidakkonsistenan maupun nilai yang hilang [13].

2.3.1 *Data Reduction*

Metode *data reduction* pada *big data* dikembangkan untuk menyederhanakan volume dan kompleksitas data [14]. Pendekatan ini memungkinkan proses analisis menjadi lebih efisien tanpa mengurangi informasi penting yang dibutuhkan model prediksi. Dalam konteks penelitian ini, data dengan label *enrolled* dihapus karena tidak merepresentasikan kondisi akhir mahasiswa, sehingga fokus analisis hanya diarahkan pada dua kategori utama, yaitu *Graduate* dan *Dropout*, guna memperoleh hasil klasifikasi yang lebih jelas dan interpretatif.

2.3.2 Normalisasi Data

Beberapa atribut dalam dataset menunjukkan keragaman skala nilai yang cukup signifikan antar fitur, yang berpotensi menyebabkan distorsi dalam proses analisis data apabila dibiarkan tanpa penyesuaian. Perbedaan rentang ini dapat memunculkan bias terhadap algoritma yang sensitif terhadap skala numerik. Untuk mengatasi permasalahan tersebut, diperlukan tahap normalisasi guna menyeragamkan skala masing-masing fitur, sehingga seluruh variabel memiliki tingkat keterbandingan yang proporsional dalam proses pemodelan dan analisis selanjutnya [15]. Normalisasi data merupakan prosedur yang sangat penting dalam pembelajaran mesin maupun analisis deret waktu [16]. Normalisasi sendiri merupakan metode yang bertujuan untuk menyesuaikan distribusi data agar lebih seragam [17]. Proses penyetaraan skala data dilakukan dengan menerapkan pendekatan *Min-Max normalization*, yang bekerja dengan mentransformasikan data asli secara linier ke dalam skala antara 0 hingga 1 [18]. Formula *Min-Max normalization* bisa dilihat pada Persamaan (1). Pada penelitian ini, penelitian ini menerapkan metode *Min-Max* dalam proses penskalaan data agar seluruh fitur memiliki skala yang proporsional..

$$X_{new} = \frac{(X_{old} - X_{min})}{(X_{max} - X_{min})} \quad (1)$$

Keterangan:
X_{new} = Nilai hasil normalisasi

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

X_{old} = Data asli sebelum normalisasi
X_{min} = Besaran terkecil dari variabel
X_{max} = Besaran terbesar dari variabel

2.4 *Split Data*

Pendekatan *Hold-Out* merupakan salah satu teknik validasi yang bekerja dengan cara memisahkan himpunan data menjadi dua subset utama. Subset pertama, yaitu data latih (*train set*), dimanfaatkan untuk membangun model prediktif berdasarkan algoritma tertentu, sedangkan subset kedua, yaitu data uji (*test set*), digunakan untuk mengevaluasi kinerja model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya. Pembagian ini memungkinkan pengujian objektif terhadap kemampuan generalisasi model terhadap data di luar sampel pelatihan [18]. Proses pemisahan tersebut dilakukan menggunakan proporsi tertentu guna menjamin bahwa model mampu diterapkan secara optimal pada data yang belum pernah dilihat sebelumnya [19]. Pada penelitian ini, proporsi pembagian data yang digunakan adalah 70% sebagai data latih dan 30% sebagai data uji. Dengan demikian, hasil pengukuran kinerja model diharapkan mencerminkan kemampuan sesungguhnya dalam menggeneralisasi data baru, sekaligus memenuhi prinsip-prinsip akuntabilitas ilmiah.

2.5 Seleksi Fitur

Tahapan ini merupakan bagian krusial dalam ranah *machine learning* maupun statistika, karena bertujuan untuk mengenali sekumpulan atribut yang memiliki kontribusi signifikan terhadap variabel target dalam analisis. Fokus utama dari proses ini adalah menyaring fitur-fitur yang relevan dan berpengaruh, sehingga hanya atribut yang memiliki nilai informatif tinggi yang dilibatkan dalam pembangunan model. Seleksi fitur yang tepat tidak hanya menyederhanakan kompleksitas data, tetapi juga meningkatkan efisiensi komputasi serta akurasi prediksi dari model yang dikembangkan. Pada data berdimensi tinggi, keberadaan atribut yang tidak penting atau berulang dapat menurunkan efektivitas pembelajaran dan kinerja model, oleh karena itu, fitur-fitur tersebut perlu dieliminasi agar kompleksitas model berkurang dan kinerjanya menjadi lebih optimal [20].

2.5.1 *Chi-Square*

Uji *Chi-Square* adalah cara statistik yang berfungsi mengukur tingkat hubungan atau keterkaitan yang signifikan antara suatu atribut dengan kategori kelas tertentu. dengan cara mengukur seberapa relevan variabel tersebut dalam merepresentasikan hubungan antar interval yang saling berdekatan [21]. Persamaan matematis untuk menghitung nilai *Chi-Square* ditunjukkan pada (2).

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad (2)$$

Keterangan:
O_i = Observed frequency (frekuensi dari data nyata)
E_i = Expected frequency (frekuensi teoritis/harapan)
n = Jumlah kategori

2.5.2 *Information Gain*

Information Gain sebuah cara penyeleksian fitur berbasis entropi, yang bekerja dengan menghitung seberapa besar informasi atau manfaat yang diberikan suatu fitur terhadap prediksi variabel target. Secara spesifik, metode ini menghitung tingkat pengurangan ketidakpastian dari hasil prediksi yang dikelompokkan berdasarkan sebuah fitur tertentu (dilambangkan sebagai gain(y, A)). Dengan kata lain, semakin tinggi nilai IG sebuah fitur, maka semakin baik fitur tersebut dalam membantu mengurangi ketidakpastian prediksi [22]. Persamaan matematis untuk menghitung nilai *Information Gain* ditunjukkan pada (3).

$$IG(S, A) = Entropy(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \qquad (3)$$

Keterangan:
S = Himpunan kasus
A = Atribut
$|S_i|$ = Jumlah kasus pada partisi ke-i
$|S|$ = Jumlah kasus dalam S

Perhitungan entropi dari suatu himpunan dirumuskan pada (4).

$$Entropy(S) = \sum_{i=1}^{n} -p_i \log_2 p_i \qquad (4)$$

Keterangan:
S = Himpunan kasus
c = Total partisi

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.5.3 *ANOVA*

ANOVA merupakan metode seleksi fitur berbasis analisis varians metode ini diterapkan untuk mengevaluasi variasi nilai tengah di antara sejumlah kelompok data yang diamati, fitur yang memiliki perbedaan signifikan dianggap relevan sehingga dapat meningkatkan performa model prediksi [23]. Persamaan matematis untuk menghitung nilai *ANOVA* ditunjukkan pada (5).

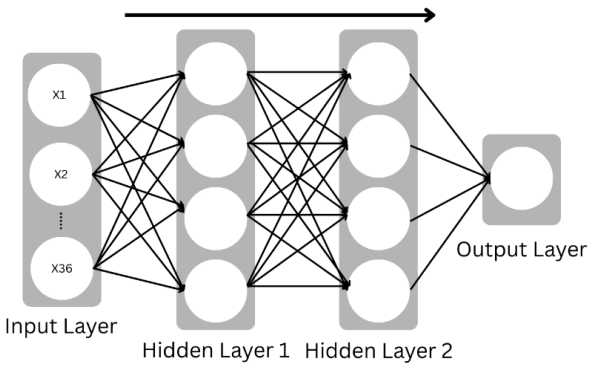
$$F(\lambda) = \frac{S_B^2(\lambda)}{S_W^2(\lambda)} \quad (5)$$

Keterangan:
$S_B^2(\lambda)$ = Sampel varian antar grup (Mean Square Between, MSB)
$S_W^2(\lambda)$ = Sampel varian dalam grup (Mean Square Within, MSW)

2.6 *Artificial Neural Network*

Artificial Neural Network (ANN) merupakan suatu pendekatan komputasional yang meniru prinsip kerja sistem saraf biologis dalam menjalankan fungsi pengenalan pola serta pengelompokan data. Model ini dirancang untuk mempelajari keterkaitan antar data melalui proses pembobotan dan penyesuaian parameter internal, sehingga mampu menangkap relasi kompleks yang tersembunyi di dalam struktur data yang bersifat non-linier [24]. Proses pembelajaran dilakukan melalui penyesuaian bobot sinaptik antar neuron, dengan setiap jaringan terdapat dari *input layer*, *hidden layer*, dan *output layer*.

Setiap unit *neuron* yang berada pada *hidden layer* melakukan perhitungan berupa kombinasi linier dari data masukan yang diterima, kemudian hasil tersebut diproses melalui suatu fungsi aktivasi tertentu. Mekanisme ini dikenal sebagai *forward propagation*, yaitu proses aliran informasi secara bertahap dari lapisan input menuju lapisan output dalam rangka menghasilkan prediksi akhir. Pada penelitian ini, fungsi aktivasi *Rectified Linear Unit* (*ReLU*) diterapkan pada lapisan tersembunyi guna mempercepat proses pelatihan model serta mengatasi permasalahan *vanishing gradient* yang sering muncul pada jaringan yang dalam. Adapun fungsi aktivasi *Sigmoid* digunakan pada lapisan keluaran karena kemampuannya dalam memetakan nilai output ke dalam rentang antara 0 hingga 1, sehingga sangat sesuai untuk menyelesaikan kasus klasifikasi biner, seperti dalam konteks ini: *Dropout* dikodekan sebagai 0 dan *Graduate* sebagai 1. Rancang bangun jaringan saraf tiruan yang digunakan dalam penelitian ini ditunjukkan pada Gambar 2, dengan arsitektur yang terdiri atas 36 neuron pada lapisan input, dua lapisan tersembunyi bertingkat.



Gambar 2. Arsitektur jaringan Artificial Neural Network (ANN)

2.7 Evaluasi

Proses evaluasi dilakukan setelah seluruh pengujian algoritma diselesaikan, dengan tujuan memastikan bahwa model yang dibangun sesuai dengan rancangan serta mampu menunjukkan kinerja secara objektif [25]. Dalam penelitian ini, *Confusion Matrix* dimanfaatkan sebagai acuan utama untuk mengevaluasi tingkat akurasi dari hasil klasifikasi model. *Confusion Matrix* merupakan perangkat evaluasi berbentuk tabel yang digunakan untuk menilai ketepatan serta kinerja algoritma klasifikasi, baik dalam proses pengelompokan maupun dalam memprediksi atribut pada data uji [26]. Evaluasi performa dilakukan dengan menghitung *Precision, Recall, Accuracy, dan F1-Score*, dimana setiap metrik diekspresikan melalui persamaan numerik pada lampiran 6 - 9.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (8)$$

$$F1-Score = 2 \times \frac{Recall \times Precision}{Recall+Precesion} \quad (9)$$

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3. HASIL DAN PEMBAHASAN

Penelitian ini membandingkan kinerja *Artificial Neural Network* yang dipadukan dengan tiga teknik penyeleksian fitur, yaitu *Information Gain, Chi-Square, dan ANOVA*, untuk memprediksi kemungkinan mahasiswa mengalami *Dropout*. Hasil evaluasi ditampilkan dalam bentuk tabel dan grafik, dengan empat metode yaitu *Accuracy, Recall, Precision, dan F1-Score*.

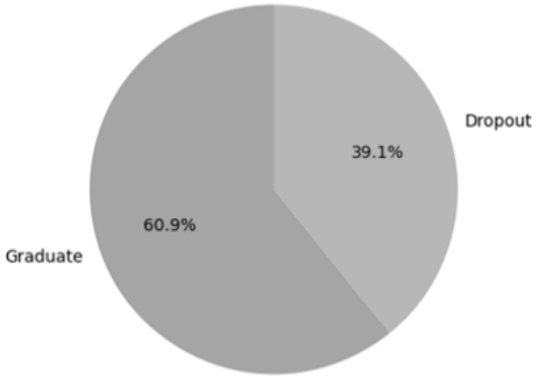
Analisis dilakukan untuk menilai sejauh mana setiap metode seleksi fitur berkontribusi terhadap performa ANN. Perbedaan hasil antar metode didiskusikan secara mendetail, sehingga dapat diidentifikasi teknik yang paling efektif pada dataset yang digunakan. Selain itu, temuan ini juga dibandingkan dengan penelitian sebelumnya guna memperkuat interpretasi dan memberikan konteks terhadap kontribusi penelitian.

3.1 Pra-Pemrosesan Data

Dataset penelitian diperoleh dari *UCI Machine Learning Repository*, yang secara eksplisit mendokumentasikan bahwa seluruh atribut telah bebas dari nilai hilang dan duplikasi. Dengan merujuk pada dokumentasi tersebut, penelitian ini tidak melakukan tahapan verifikasi ulang terhadap *missing values*, melainkan langsung melanjutkan pada tahap pemodelan dan evaluasi kinerja algoritma.

3.1.1 *Data Reduction*

Pada penelitian ini, kategori *Enrolled* dikeluarkan dari analisis karena tidak merepresentasikan kondisi akhir mahasiswa, melainkan menunjukkan status transisional yang berpotensi menimbulkan bias dalam proses prediksi. Penghapusan kategori tersebut dilakukan untuk memastikan bahwa pemodelan klasifikasi dibangun berdasarkan status akademik yang bersifat final, sehingga interpretasi hasil menjadi lebih objektif dan konsisten. Setelah proses penyaringan data dilakukan, dataset yang digunakan terdiri atas dua kelas utama, yaitu *Graduate* dan *Dropout*. Untuk memperjelas komposisi dan proporsi masing-masing kelas setelah penghapusan kategori *Enrolled*, distribusi data mahasiswa disajikan secara visual pada Gambar 3.



Gambar 3. Visualisasi Distribusi Level

Berdasarkan Gambar 3, terlihat bahwa kelas *Graduate* mendominasi dataset dengan jumlah 2.209 data (60,9%), sedangkan kelas *Dropout* berjumlah 1.421 data (39,1%), dengan total keseluruhan 3.630 entri data. Distribusi ini menunjukkan adanya ketidakseimbangan kelas yang masih berada dalam batas wajar untuk penerapan klasifikasi biner. Komposisi tersebut mencerminkan kondisi faktual di lingkungan perguruan tinggi, di mana proporsi mahasiswa yang menyelesaikan studi lebih besar dibandingkan dengan mahasiswa yang menghentikan studi sebelum kelulusan. Pola persebaran kelas ini menjadi dasar penting dalam proses pelatihan dan evaluasi model, karena berpengaruh terhadap kemampuan model dalam mengenali karakteristik mahasiswa yang berisiko mengalami *Dropout*.

3.1.2 Hasil Normalisasi Data

Proses normalisasi data pada penelitian ini diterapkan menggunakan pendekatan *Min-Max Scaling*, yaitu suatu teknik transformasi yang memetakan seluruh nilai atribut numerik ke dalam rentang 0 hingga 1. Penerapan normalisasi ini bertujuan untuk menyetarakan skala antarvariabel sehingga setiap fitur memiliki kontribusi yang seimbang dalam proses pelatihan model, sekaligus mencegah dominasi atribut dengan rentang nilai yang lebih besar. Untuk memperlihatkan hasil dari proses normalisasi yang dilakukan, ringkasan nilai atribut setelah transformasi *Min-Max Scaling* disajikan pada Tabel 1.

Tabel 1. Hasil Normalisasi Data Menggunakan Min-Max Scaling

No	Marital status	Application mode	…	GDP	Target
1	0.0	0.2857142857142857	…	0.7661822985468957	0
2	0.0	0.25	…	0.6406869220607662	1

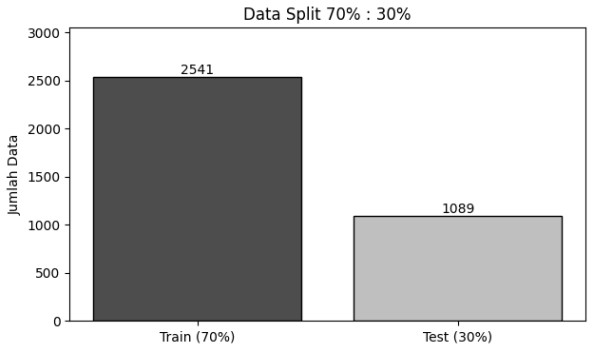
JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3	0.0		0.0	…	0.7661822985468957	0
…		…	…	…	…	…
3.630	0.0		0.1607142857142857	…	0.3117569352708058	1

Berdasarkan Tabel 1, seluruh atribut numerik telah berhasil ditransformasikan ke dalam skala yang seragam, dengan nilai minimum dan maksimum berada pada interval 0 hingga 1. Sebagai ilustrasi, atribut *GDP* yang pada kondisi awal memiliki variasi nilai yang relatif luas, setelah proses normalisasi berada pada rentang yang lebih sempit, yaitu antara 0,3117 hingga 0,7661. Transformasi ini menunjukkan bahwa perbedaan skala antaratribut telah diminimalkan secara efektif. Selain itu, variabel target dalam proses klasifikasi juga telah dikonversi ke dalam bentuk numerik, dengan nilai 0 merepresentasikan kelas *Dropout* dan nilai 1 merepresentasikan kelas *Graduate*, sehingga sesuai dengan karakteristik permasalahan klasifikasi biner. Keseragaman skala fitur ini berperan penting dalam mendukung kestabilan proses pembelajaran jaringan serta meningkatkan efisiensi dan kinerja model *Artificial Neural Network* (ANN) yang diterapkan dalam penelitian ini.

3.2 *Split Data*

Guna mempertegas skema pemisahan data yang diterapkan dalam pendekatan *hold-out*, penyajian visual mengenai distribusi data latih dan data uji disajikan pada Gambar 4.



Gambar 4. Distribusi data latih dan data uji berdasarkan rasio 70:30

Merujuk pada Gambar 4, himpunan data penelitian dipilah ke dalam dua subset utama, yakni data latih sebesar 70% (2541 data) dan data uji sebesar 30% (1089 data). Proporsi tersebut dipilih untuk menjamin kecukupan data pada tahap pembelajaran model, sekaligus menyediakan data uji yang representatif guna menilai kemampuan generalisasi model terhadap data yang berada di luar lingkup pelatihan.

3.3 Hasil Seleksi

3.3.1 Hasil Seleksi Fitur Menggunakan *Chi-Square*

Hasil pengujian dengan metode Chi-Square ditunjukkan pada Tabel 2. Dari seluruh atribut yang dianalisis, hanya fitur dengan nilai signifikansi $p<0.05p < 0.05p<0.05$ yang dipertahankan sebagai kandidat utama. Berdasarkan kriteria tersebut, fitur *Scholarship holder* memperoleh skor tertinggi (180.51), diikuti oleh *Debtor* (167.08) serta *Curricular units 2nd sem (grade)* (157.23).

Tabel 2. Fitur yang Dipilih oleh Chi-Square

No	Fitur	Chi2
1	*Scholarship holder*	180.508351
2	*Debtor*	167.082022
3	*Curricular units 2nd sem (grade)*	157.230504
4	*Curricular units 2nd sem (approved)*	119.993887
5	*Gender*	111.824350
6	*Curricular units 1st sem (grade)*	90.664635
7	*Tuition fees up to date*	66.306945
8	*Curricular units 1st sem (approved)*	63.710762
9	*Application mode*	51.378862
10	*Age at enrollment*	35.431835
11	*Displaced*	18.638353
12	*Marital status*	8.489147
13	*Curricular units 2nd sem (without evaluations)*	6.866561
14	*Previous qualification*	6.122337
15	*Application order*	5.458623
16	*Admission grade*	2.797799

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan Tabel 2, fitur dengan skor tertinggi umumnya berasal dari aspek finansial (*scholarship holder*, *debtor*) dan capaian akademik (*grades* serta *approved units*). Hasil ini mengindikasikan bahwa variabel tersebut memiliki keterkaitan yang signifikan secara statistik terhadap capaian akhir studi mahasiswa.Sebaliknya, atribut demografis seperti usia saat masuk, status pernikahan, atau latar belakang pendidikan sebelumnya memiliki skor lebih rendah, sehingga kontribusinya relatif kecil dalam prediksi *Dropout*.

3.3.2 Hasil Seleksi Fitur Menggunakan Information Gain

Hasil perhitungan skor *Information Gain* terhadap atribut pada dataset ditunjukkan pada Tabel 3. Pemilihan fitur dilakukan dengan mempertimbangkan nilai *MI-score*, di mana sepuluh fitur dengan skor tertinggi dipandang sebagai kandidat paling informatif untuk memprediksi status mahasiswa.

Tabel 3. Hasil Seleksi Fitur Menggunakan Information Gain

No	Fitur	Mi-Skor
1	*Curricular units 2nd sem (approved)*	0.334702
2	*Curricular units 1st sem (approved)*	0.253262
3	*Curricular units 2nd sem (grade)*	0.252415
4	*Curricular units 1st sem (grade)*	0.191798
5	*Tuition fees up to date*	0.105267
6	*Curricular units 1st sem (evaluations)*	0.086447
7	*Curricular units 2nd sem (evaluations)*	0.082229
8	*Application mode*	0.065667
9	*Course*	0.064620
10	*Scholarship holder*	0.058035
11	*Age at enrollment*	0.054797
12	*Curricular units 2nd sem (enrolled)*	0.054242
13	*Previous qualification (grade)*	0.053119
14	*Gender*	0.042087
15	*Displaced*	0.032949
16	*Admission grade*	0.031943
17	*Mother's occupation*	0.031487
18	*Curricular units 1st sem (enrolled)*	0.028126
19	*Debtor*	0.022345
20	*Mother's qualification*	0.020546
21	*Father's occupation*	0.018199
22	*Father's qualification*	0.017658
23	*Curricular units 2nd sem (without evaluations)*	0.013479
24	*Inflation rate*	0.012528
25	*Nacionality*	0.007318
26	*Unemployment rate*	0.007123
27	*Previous qualification*	0.006214
28	*Curricular units 1st sem (credited)*	0.005530
29	*Curricular units 1st sem (without evaluations)*	0.005158
30	*GDP*	0.002989
31	*Educational special needs*	0.001962
32	*Daytime/evening attendance*	0.001909

Berdasarkan Tabel 3, fitur dengan kontribusi terbesar adalah *Curricular units 2nd sem (approved)* (0.3347), *Curricular units 1st sem (approved)* (0.2533), dan *Curricular units 2nd sem (grade)* (0.2524). Dominasi fitur akademik ini menunjukkan bahwa capaian nilai dan kelulusan mata kuliah pada semester awal merupakan indikator yang paling relevan untuk klasifikasi *Dropout*.

Selain itu, faktor administratif seperti *Tuition fees up to date* dan *Scholarship holder* juga masuk ke dalam daftar, meskipun dengan skor yang lebih rendah. Hal ini mengindikasikan bahwa kondisi finansial tetap berpengaruh, namun tidak sekuat variabel akademik.

Berbeda dengan *Chi-Square* yang menilai asosiasi statistik secara langsung, *Information Gain* berbasis pada entropi, sehingga fitur dengan kemampuan lebih tinggi dalam mengurangi ketidakpastian prediksi memperoleh skor yang lebih besar. Oleh karena itu, *Information Gain* menegaskan bahwa performa akademik pada fase awal studi lebih berperan dibandingkan atribut administratif maupun demografis dalam membedakan mahasiswa yang berpotensi lulus dan *Dropout*.

3.3.3 Hasil seleksi fitur dengan metode *ANOVA*

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Hasil seleksi fitur menggunakan metode *ANOVA* ditunjukkan pada Tabel 4. Fitur dengan perbedaan rata-rata yang paling signifikan antar kelompok mahasiswa adalah *Curricular units 2nd sem (approved)* (1891.13), diikuti oleh *Curricular units 2nd sem (grade)* (1522.59), serta *Curricular units 1st sem (approved)* (1097.51).

Tabel 4. Hasil Seleksi Fitur Menggunakan ANOVA

No	Fitur	Skor
1	*Curricular units 2nd sem (approved)*	1891.126562
2	*Curricular units 2nd sem (grade)*	1522.589375
3	*Curricular units 1st sem (approved)*	1097.512441
4	*Curricular units 1st sem (grade)*	962.343101
5	*Tuition fees up to date*	606.256909
6	*Scholarship holder*	272.561985
7	*Age at enrollment*	212.397520
8	*Debtor*	203.569075
9	*Gender*	181.631556
10	*Application mode*	177.333288
11	*Curricular units 2nd sem (enrolled)*	79.930474
12	*Curricular units 1st sem (enrolled)*	56.648517
13	*Displaced*	42.126041
14	*Admission grade*	39.341606
15	*Curricular units 2nd sem (evaluations)*	34.260656
16	*Previous qualification (grade)*	33.235698
17	*Application order*	32.567727
18	*Marital status*	21.539295
19	*Daytime/evening attendance*	21.350098
20	*Curricular units 2nd sem (without evaluations)*	21.108523
21	*Curricular units 1st sem (without evaluations)*	15.947314
22	*Previous qualification*	9.026617
23	*Curricular units 1st sem (evaluations)*	6.927878
24	*GDP*	5.311714
25	*Course*	5.006847
26	*Mother's qualification*	4.989600

Seperti halnya *Information Gain*, hasil *ANOVA* kembali menempatkan fitur akademik semester pertama dan kedua pada peringkat teratas. Hal ini menegaskan bahwa capaian akademik awal mahasiswa memiliki perbedaan rata-rata yang signifikan antara kelompok yang lulus dan yang *Dropout*. Sesuai dengan prinsip *ANOVA*, semakin besar variasi antar kelas, semakin tinggi pula skor yang diperoleh suatu fitur. Oleh karena itu, *approved units* dan *grades* muncul sebagai indikator dominan.

Selain faktor akademik, variabel administratif seperti *Tuition fees up to date* dan *Scholarship holder* juga muncul dalam daftar dengan nilai lebih rendah, menunjukkan bahwa aspek finansial turut memengaruhi, meskipun kontribusinya tidak sebesar performa akademik. Di sisi lain, atribut demografis seperti usia masuk dan jenis kelamin juga masuk dalam peringkat, namun dengan skor relatif kecil. Dengan demikian, *ANOVA* menegaskan kembali bahwa performa akademik awal merupakan prediktor utama dalam membedakan mahasiswa berisiko *Dropout* dari mereka yang berpotensi menyelesaikan studi.

3.4 Evaluasi Kinerja ANN dengan dan Seleksi Fitur

Hasil evaluasi kinerja *Artificial Neural Network* (ANN) pada model baseline tanpa seleksi fitur serta tiga cara seleksi fitur (*Chi-Square, ANOVA, dan Information Gain*) dan tanpa seleksi fitur divisualisasikan melalui tabel 5.

Tabel 5. Hasil Confusion Matrix Berdasarkan Teknik Seleksi Fitur

Seleksi Fitur	TP	TN	FP	FN	Keterangan
No feature selection	347	644	19	79	Model menunjukkan performa yang baik tanpa reduksi fitur, meskipun masih terdapat beberapa kesalahan dalam mengklasifikasikan mahasiswa *Dropout*.
Chi-Square	342	624	39	84	Model cukup baik, namun masih terdapat 84 mahasiswa *Dropout* yang diprediksi keliru sebagai *Graduate*.
Information Gain	350	645	18	76	Memberikan hasil terbaik dengan kesalahan prediksi paling rendah (FN dan FP minimal).
ANOVA	349	640	23	77	Kinerja stabil dan mendekati Information Gain dengan keseimbangan antara *Dropout* dan *Graduate*.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9420
Hal 210-220
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan informasi yang disajikan pada Tabel 5, dapat diidentifikasi bahwa pendekatan *Information Gain* menghasilkan capaian tertinggi pada metrik *true positive* maupun *true negative*, melampaui dua metode lainnya. Hasil ini mengindikasikan bahwa model yang dibangun dengan seleksi fitur tersebut memiliki kapabilitas lebih unggul dalam mengenali kedua kategori secara akurat. Sebaliknya, pendekatan *Chi-Square* menunjukkan kelemahan signifikan, dengan tingkat *false negative* tertinggi, yang berarti sejumlah individu yang seharusnya diklasifikasikan sebagai *Dropout* justru terdeteksi sebagai *Graduate*. Sementara itu, metode *ANOVA* (*F-Value*) memperlihatkan performa yang cenderung stabil, mendekati akurasi *Information Gain*, sehingga tetap dapat dianggap layak sebagai alternatif seleksi fitur yang efektif.

Untuk memperkuat interpretasi hasil, Tabel 6 menghadirkan perbandingan menyeluruh atas capaian performa model berdasarkan metrik *Accuracy*, *Recall*, *Precision*, dan *F1-Score* setelah penerapan masing-masing teknik seleksi fitur. Penyajian ini memberikan pandangan yang lebih luas dan mendalam terkait sejauh mana setiap metode mampu berkontribusi terhadap optimalisasi performa klasifikasi, khususnya dalam konteks prediksi risiko *Dropout* pada model *Artificial Neural Network*.

Tabel 6. Hasil Evaluasi Kinerja ANN dengan Seleksi Fitur

Metode	Jumlah Fitur	Akurasi	Presisi	Recall	F1-Score
No feature selection	36	0.91	0.8907	0.9713	0.9293
Chi-Square	16	0.8871	0.8814	0.9412	0.9103
ANOVA	26	0.9082	0.8926	0.9653	0.9275
Information Gain	32	0.9137	0.8946	0.9729	0.9321

Berdasarkan Tabel 6, dapat diketahui bahwa model *baseline* tanpa seleksi fitur sudah menunjukkan kinerja yang baik, dengan nilai akurasi sebesar 91% dan *recall* 97,13%. Hasil ini mengindikasikan bahwa meskipun seluruh atribut digunakan secara utuh tanpa proses reduksi, model *Artificial Neural Network* (ANN) tetap mampu melakukan klasifikasi secara akurat. Namun demikian, setelah diterapkan proses seleksi fitur, terlihat adanya peningkatan performa pada beberapa metode, terutama pada *Information Gain*, yang menghasilkan akurasi tertinggi sebesar 91,37% dan *recall* 97,29%.

Metode *ANOVA (F-Value)* menempati posisi berikutnya dengan akurasi 90,82% dan *recall* 96,53%, menunjukkan hasil yang kompetitif dan stabil. Sementara itu, metode *Chi-Square* menghasilkan akurasi paling rendah sebesar 88,71%, meskipun masih tergolong baik dalam mendeteksi mahasiswa yang lulus (*Graduate*).

Perbedaan kinerja antar metode tersebut disebabkan oleh karakteristik masing-masing teknik penyeleksian fitur. Pendekatan *Information Gain* berfokus pada pengurangan entropi sehingga dapat menyoroti fitur-fitur akademik yang paling relevan dengan status akhir mahasiswa. *ANOVA* bekerja dengan membandingkan rata-rata antar kelas untuk menentukan atribut yang signifikan, sedangkan *Chi-Square* lebih sensitif terhadap distribusi frekuensi kategori, yang menyebabkan beberapa atribut non-akademik seperti status beasiswa atau kondisi keuangan mendapat bobot tinggi meskipun tidak selalu meningkatkan akurasi model.

Secara umum, temuan ini mengungkap bahwa penggunaan teknik seleksi fitur, khususnya dengan metode *Information Gain*, mampu meningkatkan performa dibandingkan model yang tidak menggunakan seleksi fitur, sekaligus memperkuat efektivitas model ANN dalam memproyeksikan potensi mahasiswa mengalami *Dropout*.

4. KESIMPULAN

Penelitian ini menegaskan bahwa penerapan seleksi fitur memiliki peran esensial dalam mengoptimalkan kinerja model *Artificial Neural Network* (ANN) pada prediksi kelulusan mahasiswa. Evaluasi terhadap empat konfigurasi eksperimental menunjukkan bahwa model ANN tanpa penerapan seleksi fitur tetap mampu mencapai tingkat akurasi yang tinggi, yang merefleksikan kapabilitas adaptif ANN dalam mengakomodasi kompleksitas variabel masukan yang bersifat multivariat dan heterogen. Kendati demikian, integrasi mekanisme seleksi fitur terbukti memberikan kontribusi nyata dalam meningkatkan performa model secara lebih terarah, baik dari aspek akurasi klasifikasi maupun efisiensi komputasi. Di antara pendekatan yang diuji, *Information Gain* menunjukkan supremasi kinerja dengan capaian akurasi dan *recall* tertinggi, yang mencerminkan kemampuannya dalam mereduksi ketidakpastian informasi serta menyeleksi atribut akademik dan administratif yang memiliki pengaruh signifikan terhadap hasil prediksi. Metode *ANOVA* menampilkan performa yang relatif sebanding, sedangkan *Chi-Square* menghasilkan capaian yang lebih moderat, yang mengindikasikan adanya perbedaan sensitivitas masing-masing metode terhadap karakteristik distribusi data. Temuan ini mengindikasikan bahwa meskipun ANN memiliki fleksibilitas tinggi dalam memodelkan hubungan nonlinier antarvariabel, penerapan seleksi fitur tetap krusial untuk meningkatkan efektivitas klasifikasi, stabilitas model, dan efisiensi pemrosesan data. Pada tahap lanjutan, eksplorasi metode seleksi fitur alternatif seperti *ReliefF*, *Recursive Feature Elimination* (RFE), serta pendekatan berbasis algoritma evolusioner berpotensi memperluas representativitas pemilahan atribut dan memperdalam kinerja model prediktif berbasis ANN dalam konteks pendidikan tinggi.

REFERENCES

[1] A. Behr, M. Giese, H. D. Teguim Kamdjou, and K. Theune, “Motives for dropping out from higher education—An analysis of

bachelor's degree students in Germany," *Eur. J. Educ.*, vol. 56, no. 2, pp. 325–343, 2021, doi: 10.1111/ejed.12433.
[2] V. Realinho, J. Machado, L. Baptista, and M. V. Martins, "Predicting Student Dropout and Academic Success," *Data*, vol. 7, no. 11, 2022, doi: 10.3390/data7110146.
[3] L. Hakim, A. Sobri, L. Sunardi, and D. Nurdiansyah, "Prediksi Penyakit Jantung Berbasis Mesin Learning Dengan Menggunakan Metode K-NN," *J. Digit. Teknol. Inf.*, vol. 07, no. 02, pp. 14–20, 2024.
[4] E. Haryatmi and S. Pramita Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 386–392, 2021, doi: 10.29207/resti.v5i2.3007.
[5] T. A. Yoga Siswa, "Komparasi Optimasi Chi-Square, CFS, Information Gain dan ANOVA dalam Evaluasi Peningkatan Akurasi Algoritma Klasifikasi Data Performa Akademik Mahasiswa," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 18, no. 1, p. 62, 2023, doi: 10.30872/jim.v18i1.11330.
[6] S. Situju, N. Nur, and N. Halal, "Analisis Penerapan Mutual Information pada Klasifikasi Status Studi Mahasiswa Menggunakan Naïve Bayes," *J. Appl. Comput. Sci. Technol.*, vol. 6, no. 1, pp. 23–28, 2025, doi: 10.52158/jacost.v6i1.1106.
[7] M. O. Abdullah, Y. Altun, and R. M. Ahmed, "Leveraging Artificial Neural Networks and Support Vector Machines for Accurate Classification of Breast Tumors in Ultrasound Images," vol. 16, no. 11, 2024, doi: 10.7759/cureus.73067.
[8] M. A. Jassim and S. N. Abdulwahid, "Data Mining preparation: Process, Techniques and Major Issues in Data Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1090, no. 1, p. 012053, 2021, doi: 10.1088/1757-899x/1090/1/012053.
[9] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1060–1073, 2022, doi: 10.1016/j.jksuci.2019.06.012.
[10] L. Realinho, V., Vieira Martins, M., Machado, J., & Baptista, "Predict Students' Dropout and Academic Success [Dataset]," UCI Machine Learning Repository. [Online]. Available: https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success
[11] M. F. Naufal and S. F. Kusuma, "Comparison Analysis of Machine Learning and Deep Learning Algorithms for Image Classification of Indonesian Language Signing Systems (Sibi)," *Jtiik*, vol. 10, no. 4, pp. 873–882, 2023, doi: 10.25126/jtiik.2023106828.
[12] A. Nurkholis, D. Alita, and A. Munandar, "Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 2, pp. 227–233, 2022, doi: 10.29207/resti.v6i2.3906.
[13] S. D. Amalia, M. A. Barata, and P. E. Yuwita, "Optimization of Random Forest Algorithm with Backward Elimination Method in Classification of Academic Stress Levels," vol. 9, no. 3, 2025.
[14] M. H. ur Rehman, C. S. Liew, A. Abbas, P. P. Jayaraman, T. Y. Wah, and S. U. Khan, "Big Data Reduction Methods: A Survey," *Data Sci. Eng.*, vol. 1, no. 4, pp. 265–284, 2016, doi: 10.1007/s41019-016-0022-0.
[15] G. A. B. Suryanegara, Adiwijaya, and M. D. Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 114–122, 2021.
[16] A. Faqih, M. J. Vikri, and I. Aristia, "Optimizing Gated Recurrent Unit (GRU) for Gold Price Prediction : Hyperparameter Tuning and Model Evaluation on Historical XAU / USD Data," vol. 14, pp. 141–147, 2025.
[17] Saputra, I. and D. A. Kristiyanti, *Machine learning untuk pemula*. Informatika Bandung, 2022.
[18] P. P. Allorerung, A. Erna, and M. Bagussahrir, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," vol. 9, no. 3, pp. 178–191, 2024.
[19] E. F. Laili *et al.*, "KOMPARASI ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE (SVM) DALAM," vol. 8, no. 1, pp. 67–76, 2025.
[20] S. R. Azizah, R. Herteno, A. Farmadi, D. Kartini, and I. Budiman, "Kombinasi Seleksi Fitur Berbasis Filter dan Wrapper Menggunakan Naive Bayes pada Klasifikasi Penyakit Jantung," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 6, pp. 1361–1368, 2023, doi: 10.25126/jtiik.2023107467.
[21] M. A. Barata, Edi Noersasongko, Purwanto, and Moch Arief Soeleman, "Improving the Accuracy of C4.5 Algorithm with Chi-Square Method on Pure Tea Classification Using Electronic Nose," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 2, pp. 226–235, 2023, doi: 10.29207/resti.v7i2.4687.
[22] M. I. Prasetiyowati, N. U. Maulidevi, and K. Surendro, "Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest," *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00472-4.
[23] M. M. Ali, M. S. Islam, M. N. Uddin, and M. A. Uddin, "A conceptual IoT framework based on Anova-F feature selection for chronic kidney disease detection using deep learning approach," *Intell. Med.*, vol. 10, no. October, p. 100170, 2024, doi: 10.1016/j.ibmed.2024.100170.
[24] A. I. Sakti *et al.*, "Implementasi Artificial Neural Network (ANN) dalam Memprediksi Nilai Tukar Rupiah terhadap Dolar Amerika Implementasi Artificial Neural Network (ANN) dalam Memprediksi Nilai Tukar Rupiah terhadap Dolar Amerika," vol. 12, no. 2, pp. 124–130, 2024.
[25] E. Sutoyo and A. Almaarif, "Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 95–101, 2020, doi: 10.29207/RESTI.V4I1.1502.
[26] A. A. Yaqin, M. A. Barata, and N. Mahmudah, "Implementation of the Random Forest Algorithm with Optuna Optimization in Lung Cancer Classification," vol. 14, pp. 561–569, 2025.