



## **PREDIKSI PENERIMAAN MAHASISWA MENGGUNAKAN NEURAL NETWORK BERBASIS RAPIDMINER PADA DATA GRADUATE ADMISSION**

**Ayu Elisya Natama Sianturi<sup>a</sup>, Arron Mosses Jhon Hadi<sup>b</sup>, Andri Wijaya<sup>c\*</sup>**

<sup>a</sup> Fakultas Sains dan Teknologi / Sistem Informasi; [ayuelisyans@gmail.com](mailto:ayuelisyans@gmail.com), Universitas Katolik Musi Charitas; Jalan Bangau No. 60 Palembang, Sumatera Selatan

<sup>b</sup> Fakultas Sains dan Teknologi / Sistem Informasi; [arronmosses31@gmail.com](mailto:arronmosses31@gmail.com), Universitas Katolik Musi Charitas; Jalan Bangau No. 60 Palembang, Sumatera Selatan

<sup>c</sup> Fakultas Sains dan Teknologi / Sistem Informasi; [andri\\_wijaya@ukmc.ac.id](mailto:andri_wijaya@ukmc.ac.id), Universitas Katolik Musi Charitas; Jalan Bangau No. 60 Palembang, Sumatera Selatan

\* Penulis Korespondensi: Andri Wijaya

### **ABSTRACT**

*This study aims to predict graduate admission outcomes using a Neural Network approach implemented in RapidMiner. The dataset was processed through a series of stages, including data cleaning, normalization, and model training, to ensure optimal learning quality. Model performance was assessed using the Root Mean Square Error (RMSE) metric. The resulting RMSE score of 0.054 indicates a low level of prediction error and demonstrates that the constructed model performs reliably. These findings highlight the potential of Neural Networks as an effective analytical tool for estimating student admission likelihood with higher accuracy and supporting data-driven decision-making in the selection process.*

**Keywords:** *Neural Network; Prediction; RapidMiner; RMSE*

### **Abstrak**

Penelitian ini bertujuan memprediksi hasil penerimaan mahasiswa pascasarjana menggunakan pendekatan *Neural Network* yang diimplementasikan pada aplikasi RapidMiner. Dataset diproses melalui beberapa tahapan, mulai dari pembersihan data, normalisasi, hingga pelatihan model untuk memastikan kualitas pembelajaran yang optimal. Kinerja model dievaluasi menggunakan metrik *Root Mean Square Error* (RMSE). Nilai RMSE sebesar 0,054 menunjukkan tingkat kesalahan prediksi yang rendah dan mengindikasikan bahwa model yang dibangun memiliki performa yang andal. Temuan ini menegaskan bahwa *Neural Network* dapat menjadi alat analitik yang efektif dalam memperkirakan peluang penerimaan mahasiswa dengan tingkat akurasi yang lebih baik serta mendukung pengambilan keputusan berbasis data dalam proses seleksi.

**Kata Kunci:** *Neural Network, Prediksi, RapidMiner, RMSE*

### **1. PENDAHULUAN**

Proses penerimaan mahasiswa pascasarjana merupakan tahap penting yang menentukan kualitas input sumber daya manusia pada institusi pendidikan tinggi. Proses ini melibatkan penilaian terhadap berbagai aspek akademik dan non-akademik calon mahasiswa, seperti nilai ujian standar, indeks prestasi kumulatif, kemampuan bahasa, pengalaman riset, serta rekomendasi. Banyaknya variabel yang harus dipertimbangkan sering kali membuat proses seleksi menjadi kompleks dan berpotensi menimbulkan subjektivitas apabila hanya mengandalkan penilaian manual. Oleh karena itu, pemanfaatan pendekatan analitik berbasis data menjadi semakin relevan untuk membantu institusi menghasilkan keputusan seleksi yang lebih objektif dan terukur.

Perkembangan *machine learning* memberikan alternatif yang semakin banyak digunakan dalam proses prediksi admisi. Sejumlah penelitian sebelumnya menunjukkan bahwa model prediksi mampu memberikan estimasi peluang penerimaan dengan tingkat kesalahan tertentu, tergantung pada kualitas data dan metode yang digunakan [1][2]. Di sisi lain, pendekatan berbasis *Neural Network* mulai banyak diterapkan karena kemampuannya mempelajari hubungan non-linier antarvariabel, sehingga mampu menghasilkan performa prediksi yang lebih stabil dibandingkan metode regresi atau model konvensional lainnya [3]. Penggunaan *Neural Network* juga telah diuji dalam berbagai konteks pendidikan untuk memperkirakan performa, kelulusan, maupun kecenderungan akademik mahasiswa. Hasil-hasil tersebut menunjukkan bahwa kualitas prediksi sangat dipengaruhi oleh proses *preprocessing*, pemilihan parameter, serta struktur jaringan yang digunakan [4][5]. Konsistensi performa serupa juga terlihat pada penerapannya dalam kasus peramalan lain yang melibatkan pola data kompleks, sehingga memperkuat relevansi algoritma ini sebagai pendekatan prediktif yang andal [6]. Bahkan, pengembangan berbasis *deep learning* menawarkan performa yang lebih kompetitif dengan nilai galat yang rendah pada masalah prediksi admisi [7].

Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada perbandingan performa antar algoritma dan capaian nilai evaluasi, tanpa mengulas secara mendalam peran tahapan *preprocessing* data serta penentuan parameter model terhadap kinerja prediksi secara sistematis. Selain itu, kajian yang secara khusus membahas implementasi algoritma *Neural Network* menggunakan *platform* RapidMiner sebagai *tools* visual *data mining* pada dataset *Graduate Admission 2* masih relatif terbatas. Padahal, RapidMiner menyediakan lingkungan pemodelan yang terstruktur dan intuitif untuk mengeksplorasi pengaruh tahapan *data cleaning*, normalisasi, serta parameter pelatihan terhadap stabilitas performa model.

*Neural Network* dipilih dalam penelitian ini karena kemampuannya mempelajari hubungan non-linier antar variabel secara lebih adaptif dibandingkan metode regresi konvensional. Melalui mekanisme pembaruan bobot menggunakan algoritma *backpropagation*, *Neural Network* mampu menyesuaikan diri terhadap kompleksitas pola data multivariabel yang umum ditemukan pada kasus prediksi penerimaan mahasiswa. Beberapa studi juga menunjukkan bahwa performa *Neural Network* sangat dipengaruhi oleh kualitas *preprocessing* dan pemilihan parameter pelatihan yang tepat, sehingga diperlukan pengujian yang terstruktur untuk memperoleh model dengan kesalahan prediksi yang rendah dan stabil [3][4].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk memprediksi peluang penerimaan mahasiswa pascasarjana menggunakan algoritma *Neural Network* yang diimplementasikan pada aplikasi RapidMiner. *Dataset Graduate Admission 2* yang bersifat *open-access* diproses melalui tahapan *data mining* yang sistematis, meliputi pembersihan data, normalisasi, pembagian data latih dan uji, serta pelatihan model. Kinerja model dievaluasi menggunakan metrik *Root Mean Square Error* (RMSE) untuk menilai tingkat kesalahan prediksi yang dihasilkan. Melalui pendekatan ini, penelitian diharapkan dapat memberikan gambaran yang lebih jelas mengenai efektivitas *Neural Network* berbasis RapidMiner sebagai alat bantu pengambilan keputusan dalam proses seleksi mahasiswa pascasarjana.

## 2. TINJAUAN PUSTAKA

### 2.1. *Data Mining*

*Data mining* merupakan proses sistematis untuk mengekstraksi pola, pengetahuan, dan informasi bernilai dari kumpulan data yang besar dan kompleks. Proses ini bekerja melalui serangkaian tahapan mulai dari pemahaman tujuan analisis, pengumpulan dan eksplorasi data, *preprocessing* untuk meningkatkan kualitas data, pemodelan menggunakan berbagai teknik analitis, hingga evaluasi serta interpretasi hasil. Dalam praktiknya, *data mining* memanfaatkan pendekatan statistik, matematika, dan komputasi guna mengidentifikasi hubungan serta tren yang tidak tampak melalui analisis konvensional. Teknik yang digunakan antara lain klasifikasi, klustering, asosiasi, dan regresi, yang masing-masing membantu memahami struktur data dan mendukung pengambilan keputusan berbasis bukti [8]. Konsep-konsep ini menjadi dasar dalam penelitian ini terutama pada tahap *preprocessing*, karena kualitas data yang baik secara langsung berpengaruh terhadap stabilitas nilai RMSE pada model *Neural Network*.

Aplikasi *data mining* meluas pada berbagai sektor seperti pemasaran, keuangan, kesehatan, dan pendidikan, misalnya dalam segmentasi pelanggan, deteksi anomali transaksi, prediksi risiko kesehatan, hingga analisis capaian akademik. Seiring meningkatnya kebutuhan akan pemrosesan data, aspek etika dan privasi data juga menjadi bagian penting dalam penerapan *data mining*, termasuk melalui proses anonimisasi serta perlindungan informasi sensitif agar tetap sesuai regulasi dan kaidah etis [8].

## 2.2. Machine Learning

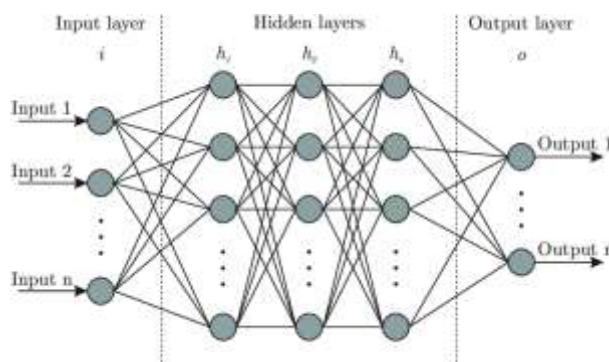
*Machine learning* merupakan cabang ilmu komputer yang berkembang dari pengenalan pola dan teori pembelajaran pada kecerdasan buatan, yang berfokus pada pembangunan model yang mampu belajar dari data dan membuat prediksi atau keputusan secara otomatis. *Machine learning* menjadi teknologi penting dua dekade terakhir seiring meningkatnya ketersediaan data dan kebutuhan analisis cerdas dalam berbagai aplikasi [9].

Tugas pembelajaran mesin umumnya terbagi menjadi tiga kategori utama, yaitu *supervised learning* yang memetakan input ke output berdasarkan data berlabel, *unsupervised learning* yang mencari pola dari data tanpa label seperti pengelompokan dan reduksi dimensi, serta *reinforcement learning* yang memungkinkan agen belajar melalui interaksi dengan lingkungan. Selain itu terdapat juga *semi-supervised learning* yang memanfaatkan kombinasi data berlabel dan tidak berlabel. Berbagai algoritma seperti regresi linier, regresi logistik, KNN, *Naive Bayes*, *Random Forest*, *K-Means*, dan *Support Vector Machine* digunakan untuk menyelesaikan tugas klasifikasi, regresi, clustering, dan estimasi kepadatan. *machine learning* menjadi fondasi penting dalam pengembangan sistem cerdas karena kemampuannya untuk beradaptasi dan meningkatkan performa melalui proses pembelajaran dari data [9]. Kerangka *supervised learning* pada penelitian ini digunakan untuk memetakan variabel *input* terhadap *Chance of Admit*, sehingga pemilihan algoritma serta parameter yang tepat menjadi krusial dalam proses prediksi.

## 2.3. Neural Network

*Neural network* merupakan salah satu metode dalam *machine learning* yang dirancang meniru pola kerja sel saraf manusia untuk mengenali hubungan kompleks dalam data. Model ini terdiri atas lapisan-lapisan neuron yang saling terhubung dan bekerja secara bertingkat dalam memproses masukan menjadi keluaran prediktif. Keunggulan utama *Neural Network* terletak pada kemampuannya mempelajari pola non-linier, menangani data berskala besar, serta beradaptasi terhadap variasi nilai masukan. Algoritma ini banyak digunakan pada berbagai tugas seperti prediksi, klasifikasi, pengenalan pola, dan identifikasi perilaku, termasuk dalam riset pendidikan, kesehatan, dan pariwisata.

Dalam proses pengembangannya, arsitektur *Neural Network* perlu ditentukan terlebih dahulu, seperti berapa banyak neuron pada lapisan input, berapa lapisan tersembunyi (*hidden layer*), dan berapa neuron di dalamnya. Menurut penelitian, tidak ada aturan baku untuk menentukan arsitektur terbaik karena setiap kasus memiliki tingkat kerumitan yang berbeda [10] Meskipun secara teori satu *hidden layer* sudah mampu mempelajari pola non-linier praktik di lapangan biasanya tetap melakukan penyesuaian bertahap agar model mencapai performa yang paling stabil. Pada penelitian ini, penentuan arsitektur dilakukan dengan mencoba beberapa konfigurasi secara bertahap hingga diperoleh model yang menghasilkan nilai kesalahan (RMSE) paling rendah.



Gambar 2.1 Arsitektur *Artificial Neural Network*

Sumber: Bre, 2017 [10]

Pada konteks prediksi, *Neural Network* mampu menghasilkan keluaran yang lebih stabil karena proses pelatihannya mengoptimalkan bobot melalui *backpropagation* untuk meminimalkan kesalahan. Beberapa penelitian menunjukkan keberhasilan metode ini dalam memprediksi kedatangan wisatawan, tingkat kelulusan siswa, hingga berbagai fenomena *time series*, khususnya ketika dikombinasikan dengan prosedur *preprocessing* dan evaluasi model yang tepat, misalnya menggunakan metrik RMSE atau akurasi [5][11]. Relevansi dengan penelitian ini adalah pemilihan jumlah neuron, *hidden layer*, serta parameter *training cycles* disesuaikan dengan teori bahwa arsitektur *Neural Network* sangat dipengaruhi karakteristik *dataset* dan tujuan prediksi.

#### 2.4. Root Mean Square Error (RMSE)

*Root Mean Square Error* (RMSE) merupakan salah satu metrik evaluasi yang paling umum digunakan dalam model regresi. RMSE mengukur seberapa jauh hasil prediksi model menyimpang dari nilai aktualnya. Semakin kecil nilai RMSE, semakin dekat prediksi model terhadap data sebenarnya sehingga performanya dapat dianggap lebih baik [12].

Secara konsep, RMSE bekerja dengan menghitung selisih antara nilai aktual dan nilai hasil prediksi, kemudian setiap selisih tersebut dikuadratkan agar tidak ada nilai negatif yang saling meniadakan. Setelah itu, seluruh hasil kuadrat dijumlahkan dan dibagi dengan jumlah data untuk memperoleh nilai *mean*. Tahap terakhir adalah mengambil akar kuadrat dari nilai tersebut sehingga menghasilkan ukuran kesalahan dalam satuan yang sama dengan variabel aslinya [10].

Pendekatan ini membuat RMSE lebih sensitif terhadap kesalahan yang besar (*large errors*), sehingga cocok digunakan ketika peneliti ingin memastikan prediksi model tidak meleset jauh pada beberapa titik data. Secara keseluruhan, RMSE memberikan gambaran yang mudah dipahami terkait tingkat akurasi model dan sering menjadi acuan dalam membandingkan performa beberapa algoritma prediksi.

Rumus RMSE tertera pada persamaan (2.1) berikut:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}} \quad (2.1)$$

Keterangan:

$A_t$  = nilai aktual pada data ke-t

$F_t$  = nilai hasil prediksi pada data ke-t

$n$  = jumlah data pengamatan

$\Sigma$  = penjumlahan seluruh nilai selisih kuadrat

#### 2.5. RapidMiner

RapidMiner merupakan perangkat lunak *data mining* berbasis *open source* yang menyediakan antarmuka grafis (GUI) sehingga memudahkan pengguna dalam membangun alur analisis tanpa memerlukan kemampuan pemrograman. Dikembangkan oleh Markus Hofmann dan Ralf Klinkenberg, RapidMiner memiliki ekosistem operator yang luas mulai dari *data preprocessing*, transformasi, pemodelan, hingga evaluasi, serta mendukung berbagai teknik *machine learning* termasuk *Neural Network*. Perangkat lunak ini terintegrasi dengan banyak sumber data seperti Oracle, MySQL, SQL Server, Excel, maupun format teks sehingga dapat digunakan dalam proses ekstraksi, transformasi, dan pemuatan data (ETL). Dengan sifatnya yang fleksibel dan lengkap, RapidMiner menjadi salah satu *tools* yang banyak digunakan dalam penelitian akademik, karena mampu memfasilitasi eksperimen model secara visual, mempermudah pemilihan parameter, dan mendukung proses validasi model secara terstruktur [8].

#### 2.6. Penelitian Terdahulu

Berbagai studi telah menerapkan teknik *machine learning* dan *Neural Network* untuk tugas prediksi. Penelitian pada kasus penerimaan mahasiswa pascasarjana menunjukkan bahwa model regresi seperti *Linear Regression*, SVR, *Decision Tree*, dan *Random Forest* mampu menghasilkan variasi error yang berbeda untuk menentukan algoritma paling optimal [1]. Pada analisis kelulusan mahasiswa, model *Neural Network* berbasis MLP dilaporkan mencapai akurasi tinggi hingga 98,27% [4]. Penerapan metode serupa pada prediksi harga saham menghasilkan nilai RMSE sebesar 19,734 sebagai indikator performa model [13]. Studi lain mengenai peramalan kedatangan wisatawan menggunakan RapidMiner memperoleh nilai RMSE 40.439,085 [11]. Penelitian terkait klasifikasi kondisi fisik mobil bekas menggunakan *Neural Network* menunjukkan akurasi 98,52% dengan AUC 0,992 [14].

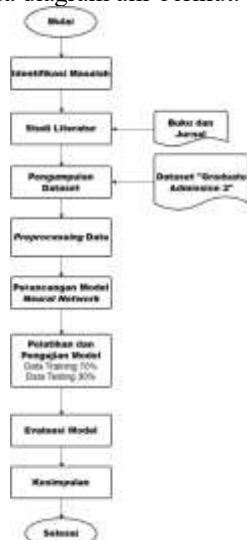
Selain itu, komparasi beberapa algoritma untuk prediksi gaya belajar menemukan bahwa KNN menjadi model paling akurat dengan capaian 78% [15]. Pada konteks klasifikasi tingkat kesejahteraan keluarga, tiga algoritma, yaitu *Naive Bayes*, *Logistic Regression*, dan *Neural Network*, yang mampu mencapai akurasi 100% [16]. Sementara itu, penelitian mengenai prediksi mobilitas masyarakat menggunakan *Backpropagation* menghasilkan akurasi 95,49% dengan nilai kappa 0,908 [6]. Temuan-temuan tersebut

menunjukkan bahwa *machine learning* dan *Neural Network* memiliki kemampuan prediktif yang kuat pada berbagai jenis data.

### 3. METODOLOGI PENELITIAN

#### 3.1. Kerangka Penelitian

Kerangka penelitian menggambarkan langkah-langkah utama yang digunakan untuk mencapai tujuan penelitian. Alur dimulai dari pengumpulan *dataset* hingga hasil evaluasi yang digunakan untuk menarik kesimpulan mengenai kemampuan model dalam memprediksi peluang penerimaan mahasiswa. Gambaran mengenai alur penelitian dapat dilihat pada diagram alir berikut.



Gambar 3.1 Kerangka Penelitian

##### 3.1.1. Identifikasi Masalah

Tahap ini dilakukan untuk memahami permasalahan utama yang ingin diselesaikan, yaitu bagaimana memprediksi peluang penerimaan mahasiswa berdasarkan *dataset* historis. Identifikasi masalah membantu merumuskan tujuan penelitian serta ruang lingkup penelitian.

##### 3.1.2. Studi Literatur

Pada tahap ini dikumpulkan teori, konsep, dan hasil penelitian terdahulu yang berkaitan dengan topik penelitian, seperti *data mining*, *machine learning*, *Neural Network*, serta metode evaluasi model. Studi literatur digunakan sebagai dasar dalam merancang model yang tepat.

##### 3.1.3. Pengumpulan Dataset

*Dataset* diperoleh dari sumber *Kaggle* dengan nama *Graduate Admission 2* yang memuat variabel-variabel terkait penerimaan mahasiswa. Data ini menjadi bahan utama untuk proses pelatihan dan pengujian model prediksi

##### 3.1.4. Preprocessing Data

Tahap preprocessing mencakup pembersihan data (*data cleaning*), pemilihan atribut, transformasi data, normalisasi, serta konversi atribut nominal menjadi numerik. Tahap ini dilakukan untuk memastikan data berada dalam kondisi optimal sebelum masuk ke proses pemodelan.

##### 3.1.5. Perancangan Model Neural Network

Pada tahap ini dilakukan penyusunan arsitektur model, seperti penentuan input, jumlah neuron, jumlah *hidden layer*, serta parameter pelatihan. Model dirancang menggunakan *RapidMiner* dengan mempertimbangkan karakteristik *dataset* dan tujuan prediksi.

##### 3.1.6. Pelatihan dan Pengujian Model

Model dilatih menggunakan sebagian data (*training set*) sebanyak 70% dan diuji menggunakan data lainnya (*testing set*) sebanyak 30%. Proses ini dilakukan untuk menilai kemampuan model dalam mempelajari pola serta memprediksi nilai keluaran secara akurat.

##### 3.1.7. Evaluasi Model

---

*Prediksi Penerimaan Mahasiswa Menggunakan Neural Network Berbasis Rapidminer pada Data Graduate Admission (Ayu Elisya Natama Sianturi)*

Model dievaluasi menggunakan metrik RMSE untuk mengukur tingkat kesalahan prediksi. Nilai RMSE yang rendah menunjukkan bahwa model memiliki performa yang baik dalam memprediksi peluang penerimaan mahasiswa.

### 3.1.8. Kesimpulan

Tahap akhir ini berisi penarikan kesimpulan berdasarkan hasil evaluasi model, sejauh mana model *Neural Network* dapat digunakan untuk membantu proses prediksi penerimaan mahasiswa.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Pengumpulan Dataset

*Dataset* yang digunakan dalam penelitian ini berasal dari platform *Kaggle* dengan judul *Graduate Admission 2*, yang berisi sejumlah variabel yang berhubungan dengan proses seleksi penerimaan mahasiswa pascasarjana. *Dataset* ini berfungsi sebagai sumber data utama dalam pembangunan model, mulai dari tahap pelatihan hingga pengujian, sehingga memungkinkan peneliti menganalisis faktor-faktor yang memengaruhi peluang diterimanya seorang calon mahasiswa.

### 4.2 Preprocessing Data

Sebelum data dapat digunakan secara efektif untuk melatih model, data harus melalui serangkaian proses penyiapan yang cermat. Sub-bab 4.2 ini merinci bagaimana data mentah dari sumber awal diolah menjadi format yang bersih dan terstruktur.

#### 4.2.1 Import Data dan Cleaning Data

Tahap pertama adalah mengimpor data mentah dalam format CSV ke dalam RapidMiner menggunakan operator *Read CSV*. Data kemudian dibaca dan dipersiapkan untuk diproses lebih lanjut. Hasil statistik data menunjukkan bahwa tidak ada *missing value*, tetapi ditemukan *outlier* sebanyak 10 dari 500 data. Penanganan *outlier* dilakukan karena keberadaan nilai ekstrem berpotensi memengaruhi proses pembelajaran *Neural Network* secara signifikan. *Neural Network* bersifat sensitif terhadap nilai yang menyimpang karena pembaruan bobot dilakukan berdasarkan kesalahan prediksi, sehingga *outlier* dapat menyebabkan model belajar pola yang tidak representatif. Oleh karena itu, sebanyak 10 data yang teridentifikasi sebagai *outlier* dari total 500 data dihapus menggunakan operator *Filter Examples*. Langkah ini bertujuan meningkatkan kualitas data latih agar model mampu mempelajari pola mayoritas data secara lebih stabil dan menghasilkan performa prediksi yang lebih akurat.

Attribute	Min	Q1	Q2	Q3	Max	StdDev	Skewness	Kurtosis
chance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
gpa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
gmat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
gre	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
toefl	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
toefl_writing	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
total_score	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
total_score_scaled	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
total_score_scaled_min_max	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
total_score_scaled_min_max_scaled	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
total_score_scaled_min_max_scaled_min_max	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
total_score_scaled_min_max_scaled_min_max_scaled	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
total_score_scaled_min_max_scaled_min_max_scaled_scaled	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
total_score_scaled_min_max_scaled_min_max_scaled_scaled_scaled	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Gambar 4.1 Statistik *Cleaning Data*

#### 4.2.2 Seleksi dan Penambahan Atribut

Operator *Select Attributes* digunakan untuk memilih atribut yang relevan saja, agar model lebih fokus dan mengurangi *noise* dari atribut yang tidak diperlukan. Lalu terdapat penambahan atribut baru menggunakan operator *Generate Attributes*, salah satunya adalah konversi status penerimaan menjadi dua label, misalnya *status = Accept* dan *status = Fail*. Penambahan atribut ini bertujuan memperjelas peran setiap variabel dalam proses prediksi.

#### 4.2.3 Set Attribute Role

Operator *Set Role* digunakan untuk menentukan atribut *label*, yaitu variabel yang akan diprediksi oleh model. Dalam penelitian ini, atribut *Chance of Admit* ditetapkan sebagai label untuk melakukan prediksi peluang penerimaan mahasiswa.

**4.2.4 Normalisasi Data**

Proses normalisasi dilakukan untuk menyeragamkan skala seluruh atribut numerik sebelum digunakan dalam pelatihan model Neural Network. Normalisasi diterapkan menggunakan operator *Normalize* dengan metode *Z-transformation*. Setelah normalisasi diterapkan, seluruh atribut berada pada skala yang sama sehingga proses *training* dapat berlangsung lebih optimal dan menghasilkan prediksi yang lebih akurat.

**4.2.5 Analisis Korelasi**

Sebelum data digunakan untuk pelatihan, dilakukan pemeriksaan hubungan antar variabel menggunakan operator *Correlation Matrix*. Analisis ini membantu memvalidasi keterkaitan antar atribut serta memastikan tidak ada multikolinearitas yang berlebihan yang dapat mempengaruhi hasil model.

**4.2.6 Konversi Data**

Semua atribut dalam data dikonversi dari nominal menjadi numerik agar sesuai dan dapat diproses oleh algoritma *Neural Network*.

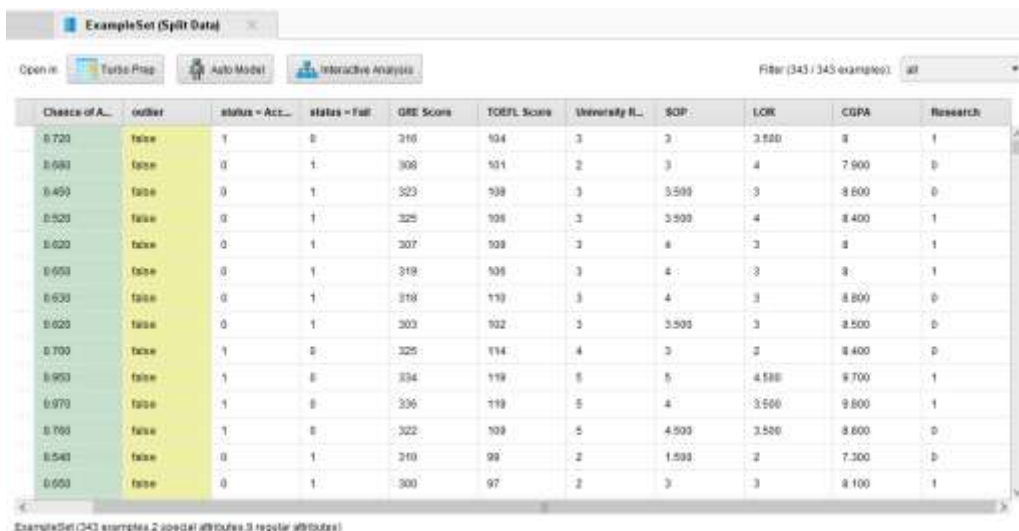
**4.2.7 Pembagian Data**

Pembagian *dataset* menjadi 70% data pelatihan (*training data*) dan 30% data pengujian (*testing data*) dilakukan untuk menjaga keseimbangan antara kemampuan model dalam mempelajari pola data dan kemampuannya dalam melakukan generalisasi. Proporsi ini umum digunakan dalam penelitian *supervised learning* karena memberikan jumlah data latih yang cukup untuk membangun model yang stabil, sekaligus menyediakan data uji yang representatif untuk mengevaluasi performa model secara objektif. Dengan pembagian ini, hasil evaluasi RMSE diharapkan dapat mencerminkan kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya.

Tabel 4.1. Pembagian Data *Training* dan Data *Testing*

<i>Dataset</i>	<i>Data Training</i>	<i>Data Testing</i>
<i>Graduate Admission 2</i>	70%	30%

Berikut ini adalah tampilan data setelah langkah *preprocessing data* dilakukan.



Gambar 4.2 Hasil *Preprocessing Data*

**4.3 Perancangan Model *Neural Network***

Pada tahap ini digunakan operator *Neural Net* dengan parameter sebagai berikut.

Tabel 4.2. Parameter Model *Neural Network*

<i>Algoritma</i>	<i>Training Cycles</i>	<i>Learning Rate</i>	<i>Momentum</i>
<i>Neural Network</i>	200	0.01	0.9

Model dilatih menggunakan data hasil *split* untuk menghasilkan pola prediksi.

Penentuan parameter *Neural Network* dilakukan dengan mempertimbangkan karakteristik *dataset Graduate Admission* yang berukuran relatif kecil hingga menengah serta bersifat numerik. Nilai *training cycles* sebesar 200 dipilih untuk memberikan kesempatan yang cukup bagi model dalam melakukan pembaruan bobot secara bertahap hingga mencapai kondisi konvergen. *Learning rate* sebesar 0.01 digunakan untuk menjaga keseimbangan antara kecepatan pembelajaran dan stabilitas model, sehingga pembaruan bobot tidak terlalu agresif. Sementara itu, nilai *momentum* sebesar 0.9 diterapkan untuk membantu mempercepat konvergensi dan mengurangi risiko terjebak pada minimum lokal. Kombinasi parameter ini dipilih karena mampu menghasilkan nilai RMSE yang rendah dan stabil pada proses pengujian.

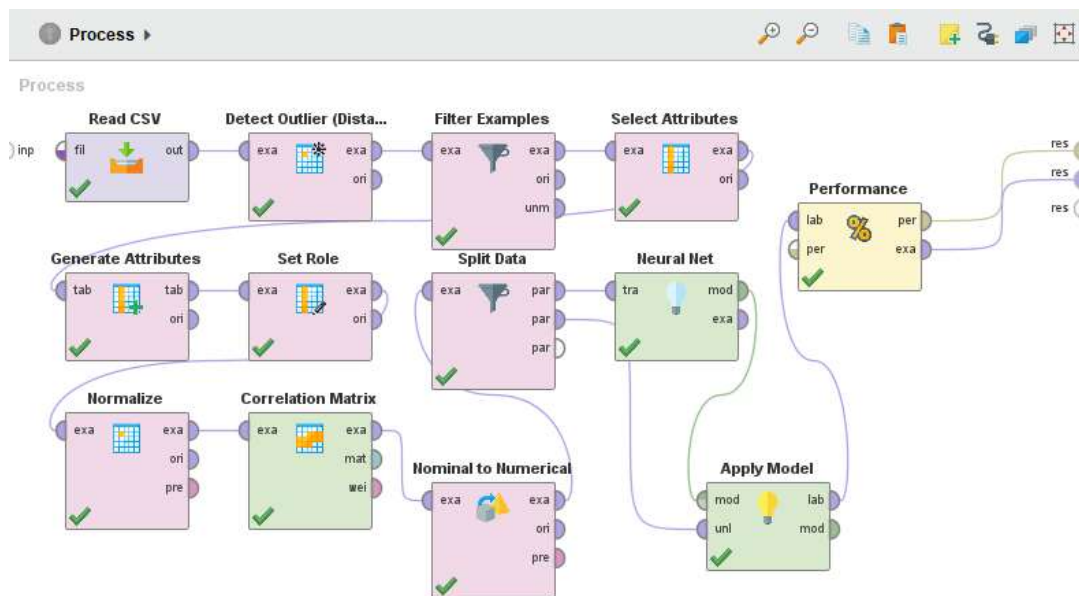
#### 4.4 Pelatihan dan Pengujian Model

Model yang telah dilatih kemudian diaplikasikan pada data pengujian menggunakan operator *Apply Model*, untuk menghasilkan prediksi dari atribut label

#### 4.5 Evaluasi Model

Tahap terakhir adalah mengevaluasi performa model menggunakan operator *Performance (Performance Regression)* dengan metrik yang digunakan ialah *Root Mean Square Error (RMSE)*. Evaluasi dilakukan untuk melihat tingkat akurasi dan metrik lain yang relevan terhadap hasil prediksi.

Keseluruhan proses hingga evaluasi model tertera pada gambar berikut.

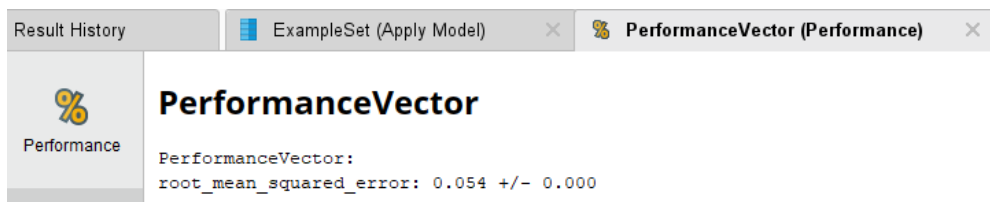


Gambar 4.3 Keseluruhan *Data Mining Workflow*

#### 4.6 Analisis Hasil

Nilai RMSE sebesar  $0.054 \pm 0.000$  menunjukkan bahwa model *Neural Network* memiliki tingkat kesalahan prediksi yang sangat rendah dan konsisten. Rendahnya nilai RMSE ini dipengaruhi oleh beberapa faktor, antara lain kualitas *preprocessing data* yang baik, penerapan normalisasi *Z-transformation*, serta penghapusan *outlier* yang berpotensi mengganggu proses pembelajaran. Selain itu, variabel-variabel utama seperti *GRE Score*, *TOEFL Score*, dan *CGPA* memiliki hubungan yang kuat dengan variabel *Chance of Admit*, sehingga memudahkan model dalam mempelajari pola hubungan antarvariabel secara efektif.

Hasil ini sejalan dengan penelitian serupa yang menerapkan *Neural Network* pada kasus prediksi di bidang pendidikan maupun prediksi lainnya, di mana model cenderung menghasilkan nilai kesalahan yang rendah ketika data telah melalui proses *preprocessing* yang terstruktur dan parameter pelatihan dipilih secara tepat. Dengan demikian, nilai RMSE yang rendah pada penelitian ini tidak hanya mencerminkan keakuratan model, tetapi juga menunjukkan bahwa pendekatan *data mining* yang digunakan telah sesuai dengan karakteristik permasalahan prediksi penerimaan mahasiswa.



Gambar 4.4 Hasil RMSE

Grafik perbandingan antara *prediction (Chance of Admit)* dan *actual Chance of Admit* menunjukkan bahwa pola garis prediksi (biru) bergerak sangat mendekati garis nilai aktual (hijau). Kedua garis mengikuti fluktuasi yang hampir sama pada setiap titik data, yang menandakan bahwa model mampu menangkap pola hubungan antarvariabel dengan baik. Walaupun terdapat sedikit penyimpangan pada beberapa titik, selisih tersebut relatif kecil dan konsisten. Hal ini sejalan dengan nilai RMSE yang rendah, sehingga dapat disimpulkan bahwa model regresi yang digunakan memberikan performa prediksi yang akurat dan stabil dalam memperkirakan peluang diterimanya calon mahasiswa.

Pola kedekatan antara garis prediksi dan nilai aktual juga menunjukkan bahwa *Neural Network* mampu menangkap hubungan non-linier antar variabel input secara efektif. Fluktuasi yang serupa pada kedua garis mengindikasikan bahwa perubahan nilai atribut input diikuti secara proporsional oleh hasil prediksi model. Hal ini memperlihatkan bahwa model tidak hanya menghafal *training data*, tetapi juga mampu melakukan generalisasi dengan baik pada *testing data*.



Gambar 4.5 Grafik Hasil Prediksi

**5. KESIMPULAN DAN SARAN**

Penelitian ini menunjukkan bahwa algoritma *Neural Network* mampu memprediksi peluang penerimaan mahasiswa pascasarjana dengan tingkat kesalahan yang rendah. Melalui proses *data mining* menggunakan RapidMiner, diperoleh nilai RMSE sebesar 0.054 yang menandakan kesesuaian model terhadap pola data aktual. Secara praktis, model ini dapat dimanfaatkan sebagai alat bantu pendukung keputusan bagi institusi pendidikan dalam proses seleksi mahasiswa, khususnya pada tahap penyaringan awal, sehingga keputusan dapat dilakukan secara lebih objektif dan berbasis data.

Untuk penelitian selanjutnya, disarankan dilakukan eksplorasi lebih lanjut terhadap variasi arsitektur dan parameter *Neural Network*, seperti penambahan *hidden layer* atau penerapan teknik optimasi parameter. Selain itu, penggunaan *dataset* yang lebih besar atau berasal dari beberapa institusi pendidikan dapat meningkatkan kemampuan generalisasi model. Perbandingan dengan algoritma prediksi lain juga dapat dilakukan untuk memperoleh gambaran metode yang paling optimal dalam konteks prediksi penerimaan mahasiswa.

## DAFTAR PUSTAKA

- [1] M. S. Acharya, A. Armaan, and A. S. Antony, "A comparison of regression models for prediction of graduate admissions," in *ICCIDS 2019 - 2nd International Conference on Computational Intelligence in Data Science, Proceedings*, IEEE, 2019. doi: 10.1109/ICCIDS.2019.8862140.
- [2] A. Alghamdi, A. Barsheed, H. Almshjary, and H. Alghamdi, "A Machine Learning Approach for Graduate Admission Prediction," in *ACM International Conference Proceeding Series*, 2020, pp. 155–158. doi: 10.1145/3388818.3393716.
- [3] L. Meilina, I. N. S. Kumara, and N. Setiawan, "Literature Review Klasifikasi Data Menggunakan Metode Cosine Similarity dan Artificial Neural Network," *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, p. 307, 2021. doi: 10.24843/mite.2021.v20i02.p15.
- [4] R. Ridwan, H. Lubis, and P. Kustanto, "Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa," *J. Media Inform. Budidarma*, vol. 4, no. 2, pp. 286–293, 2020, doi: 10.30865/mib.v4i2.2035.
- [5] Z. Zulfikar, E. S. Podungge, M. I. Saleh, and Nurdin, "Penerapan Data Mining Untuk Memprediksi Tingkat Kelulusan Siswa Menggunakan Algoritma Neural Network," *J. Elektron. Sist. Inf. dan Komput.*, vol. 5, no. 1, pp. 7–13, 2019, [Online]. Available: <http://jesik.web.id/index.php/jesik/article/view/91>
- [6] S. R. Dana and E. P. Silmina, "Prediction of Increased Mobility of Yogyakarta Residents in Controlling the Spread of COVID-19 Cases Using the Neural Network Algorithm," in *Prosiding Seminar Nasional Teknologi Informasi dan Kedirgantaraan : Peran Generasi Z dalam Dunia Kedirgantaraan*, Yogyakarta: SENATIK 2021, 2022, pp. 113–120. doi: 10.28989/senatik.v7i0.452.
- [7] B. Zuhri, N. H. Harani, and C. Prianto, "Probability Prediction for Graduate Admission Using CNN-LSTM Hybrid Algorithm," *Indones. J. Comput. Sci.*, vol. 12, no. 3, pp. 1105–1119, 2023, [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- [8] Z. Setiawan *et al.*, *Buku Ajar Data Mining*, 1st ed. Yogyakarta: PT. Sonpedia Publishing Indonesia, 2023.
- [9] A. Aditya Permana *et al.*, *MACHINE LEARNING*, 1st ed. Padang, Sumatera Barat: PT GLOBAL EKSEKUTIF TEKNOLOGI, 2023. [Online]. Available: <https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKEwiodmqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA>
- [10] F. Ramdhani and K. Setiawan, "Penerapan Data Mining untuk Prediksi Pelanggan di PT. XYZ Menggunakan Algoritma Linear Regression," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 490–497, 2024, doi: 10.57152/malcom.v4i2.1217.
- [11] F. T. Julfia, S. Sibuea, and E. Satryawati, "Penerapan Algoritma Neural Network dalam Prediksi Kedatangan Wisatawan Mancanegara di DKI Jakarta Melalui Pintu Masuk Bandara Soekarno-Hatta," *Joined J. (Journal Informatics Educ.*, vol. 8, no. 1, pp. 25–32, 2025, doi: 10.31331/joined.v8i1.3813.
- [12] D. Wonda and I. Pratama, "Implementasi Data Mining Menggunakan Neural Network Untuk Prediksi Penjualan (studi kasus: Burjo Burneo Seturan Raya)," *J. Inf. Syst. dan Artif. Intell.*, vol. 4, no. 1, pp. 48–56, 2023.
- [13] M. Zulfani and A. Dapadeda, "Prediksi harga saham menggunakan metode ensemble," *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 18, no. 1, pp. 1–6, 2024.
- [14] M. Sulaeman and Y. Arie Wijaya, "Klasifikasi Kondisi Fisik Mobil Bekas Menggunakan Algoritma Neural Network (Nn)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1555–1559, 2024, doi: 10.36040/jati.v8i2.9020.
- [15] Sumarlin and D. Anggraini, "Data Mining Pendidikan: Prediksi Gaya Belajar Mahasiswa Teknik Menggunakan Machine Learning," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 3, pp. 563–572, 2025.
- [16] N. R. Simon, A. Lasarudin, and W. Hasyim, "PERBANDINGAN ALGORITMA NAIVE BAYES , LOGISTIC REGRESSION DAN ARTIFICIAL NEURAL NETWORK UNTUK KLASIFIKASI TINGKAT KESEJAHTERAAN KELUARGA," *J. Ilmu Komput.*, vol. 5, no. 3, pp. 196–211, 2025.

## NOMENKLATUR

- $A_t$  nilai aktual pada data ke-t  
 $F_t$  nilai hasil prediksi pada data ke-t  
 $n$  jumlah data pengamatan

$\Sigma$  penjumlahan seluruh nilai selisih kuadrat