



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 931–942

Pengujian Kinerja Model Neural Network untuk Klasifikasi
Status Merokok dengan Metrik *Root Mean Squared Error* (RMSE)

Muhamad Resta Ardhiansyah, Hasbi Firmansyah, Wahyu Asriyani,
Rizki Prasetyo Tulodo

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

DOI : <https://doi.org/10.32672/mister.v3i1.4021>

How to Cite this Article

APA : Ardhiansyah, M. R., Firmansyah, H., Asriyani, W., Prasetyo Tulodo, R., & Prasetyo Tulodo, R. (2025). Pengujian Kinerja Model Neural Network untuk Klasifikasi Status Merokok dengan Metrik Root Mean Squared Error (RMSE). *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 3(1), 931 – 942. <https://doi.org/10.32672/mister.v3i1.4021>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



9 773032 710001 9 773032 601002

Pengujian Kinerja Model *Neural Network* untuk Klasifikasi Status Merokok dengan Metrik *Root Mean Squared Error* (RMSE)

Muhamad Resta Ardhiyansyah^{1*}, Hasbi Firmansyah², Wahyu Asriyani³, Rizki Prasetyo Tulodo⁴

Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal,
Kota Tegal, Indonesia^{1,2,4}

Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Pendidikan,
Universitas Pancasakti Tegal, Kota Tegal, Indonesia³

*Email: ardhiyansyahresta@gmail.com, hasbifirmansyah@upstegal.ac.id, asriyani1409@gmail.com,
rizki.prasetyo.tulodo@gmail.com

Diterima: 10-12-2025

| Disetujui: 20-12-2025

| Diterbitkan: 22-12-2025

ABSTRACT

Smoking is a high-risk behavior that significantly contributes to increased morbidity and mortality worldwide, highlighting the need for more accurate approaches to predict the likelihood of successful smoking cessation for individual smokers. This study aims to develop a prediction model for smoking cessation status using a neural network approach as part of machine learning techniques. The dataset used in this study consists of several attributes representing smoker characteristics, including nicotine dependence level, exhaled carbon monoxide (CO) concentration, number of cigarettes consumed per day, age at smoking initiation, previous quit attempts, emotional condition, and motivation to quit smoking. The research process begins with data preprocessing steps, such as data cleaning, normalization of numerical attributes, and splitting the dataset into training and testing sets. A neural network model based on a multilayer perceptron architecture with the backpropagation learning algorithm is implemented using RapidMiner software. Model performance is evaluated using the Root Mean Squared Error (RMSE) metric to measure the prediction error between actual values and predicted outcomes. The experimental results indicate that the proposed model achieves an RMSE value of 0.409, suggesting that the model is able to perform predictions with a relatively low and stable error rate. This result demonstrates that the neural network method is effective in capturing complex and nonlinear relationships among variables in the dataset. Therefore, the developed prediction model can serve as a supportive decision-making tool for healthcare professionals in identifying smokers with higher probabilities of successful smoking cessation and in designing more targeted and data-driven intervention programs.

Keywords: prediction, machine learning, neural network, RMSE.

ABSTRAK

Merokok merupakan perilaku berisiko yang berkontribusi signifikan terhadap meningkatnya angka morbiditas dan mortalitas, sehingga diperlukan pendekatan yang lebih akurat untuk memprediksi peluang keberhasilan berhenti merokok pada setiap individu. Penelitian ini bertujuan untuk membangun model prediksi status berhenti merokok dengan memanfaatkan metode neural network sebagai bagian dari pendekatan machine learning. Dataset yang

digunakan terdiri dari beberapa variabel yang merepresentasikan karakteristik perokok, meliputi tingkat ketergantungan nikotin, kadar karbon monoksida (CO) dalam napas, jumlah konsumsi rokok per hari, usia mulai merokok, riwayat upaya berhenti sebelumnya, kondisi emosional, serta tingkat motivasi untuk berhenti merokok. Tahapan penelitian diawali dengan proses prapemrosesan data, termasuk pembersihan data, normalisasi atribut numerik, dan pembagian data ke dalam data pelatihan dan data pengujian. Model neural network dikembangkan menggunakan arsitektur multilayer perceptron dengan algoritma pembelajaran backpropagation dan diimplementasikan menggunakan perangkat lunak RapidMiner. Evaluasi kinerja model dilakukan dengan menggunakan metrik Root Mean Squared Error (RMSE) untuk mengukur tingkat kesalahan prediksi antara nilai aktual dan nilai hasil prediksi. Hasil pengujian menunjukkan bahwa model yang dibangun menghasilkan nilai RMSE sebesar 0,409, yang mengindikasikan bahwa model mampu melakukan prediksi dengan tingkat kesalahan yang relatif rendah dan cukup stabil. Nilai RMSE tersebut menunjukkan bahwa metode neural network efektif dalam memodelkan hubungan nonlinier antara variabel pada dataset yang digunakan, sehingga model prediksi ini diharapkan dapat menjadi alat bantu dalam pengambilan keputusan terkait program intervensi berhenti merokok secara lebih tepat dan berbasis data.

Katakunci: prediksi, pembelajaran mesin, neural network, RMSE.

PENDAHULUAN

Merokok merupakan salah satu perilaku berisiko utama yang terbukti memberikan dampak negatif yang luas terhadap kesehatan manusia. Konsumsi tembakau diketahui dapat merusak hampir seluruh organ tubuh, meningkatkan risiko berbagai penyakit kronis seperti kanker, penyakit kardiovaskular, dan gangguan pernapasan, serta secara signifikan menurunkan angka harapan hidup perokok (Yapar & İlhan, 2025). Secara global, merokok masih menjadi penyebab utama morbiditas dan mortalitas yang sebenarnya dapat dicegah, sehingga tetap menjadi tantangan besar bagi sistem kesehatan dunia (MASTROBATTISTA et al., 2023). WHO memperkirakan bahwa apabila tidak dilakukan intervensi pengendalian tembakau yang efektif, jumlah kematian akibat rokok dapat mendekati 10 juta jiwa per tahun dalam beberapa dekade mendatang (Herath, 2025).

Berbagai metode perawatan berbasis bukti telah dikembangkan dan direkomendasikan untuk membantu individu berhenti merokok, termasuk konseling perilaku, terapi farmakologis, dan pendekatan kombinasi. Namun demikian, tingkat keberhasilan program berhenti merokok masih tergolong rendah, di mana hanya sebagian kecil peserta yang mampu mempertahankan kondisi pantang merokok dalam jangka panjang (Hersi et al., 2024). Selain itu, banyak tenaga medis menilai bahwa konseling berhenti merokok memerlukan waktu yang cukup besar dan hasilnya tidak selalu konsisten, sehingga penerapannya belum dilakukan secara optimal dalam praktik klinis sehari-hari (Hock et al., 2023).

Untuk meningkatkan efektivitas intervensi, berbagai penelitian telah mengkaji sejumlah faktor yang berpotensi memengaruhi keberhasilan berhenti merokok. Faktor-faktor tersebut meliputi tingkat ketergantungan nikotin, kadar karbon monoksida (CO) dalam napas, jumlah konsumsi rokok per hari, usia awal merokok, riwayat usaha berhenti sebelumnya, kondisi psikososial, serta tingkat motivasi individu untuk berhenti merokok (Pfeifer & Schimek, 2021). Namun, penggunaan faktor-faktor ini secara terpisah sering kali menghasilkan prediksi yang tidak konsisten dan sulit ditafsirkan, sehingga kurang praktis bagi tenaga medis maupun pasien dalam pengambilan keputusan.

Oleh karena itu, pengembangan model prediksi yang mampu mengintegrasikan berbagai faktor tersebut secara komprehensif dinilai sebagai pendekatan yang menjanjikan untuk memperkirakan peluang keberhasilan berhenti merokok pada tingkat individu. Dalam beberapa tahun terakhir, pemanfaatan metode pembelajaran mesin (*machine learning*) dalam bidang kesehatan telah menunjukkan potensi yang besar dalam mengolah data kompleks dan menghasilkan prediksi yang lebih akurat dibandingkan pendekatan konvensional (An et al., 2023).

Menurut (Alhumaidi et al., 2025) data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu:

Dalam pengembangan model prediksi status perokok, berbagai konsep dalam ilmu data modern seperti *data mining*, *machine learning*, *neural network*, dan pengukuran akurasi model menjadi sangat penting. *Data mining* merupakan proses sistematis untuk menemukan pola, hubungan, dan pengetahuan baru dari kumpulan data yang besar. Teknik ini banyak digunakan dalam bidang kesehatan untuk memprediksi risiko penyakit, perilaku kesehatan, serta efektivitas intervensi (Habehh & Gohel, 2021). Melalui pendekatan ini, berbagai variabel seperti tingkat konsumsi rokok, usia mulai merokok, kadar karbon monoksida, dan kondisi emosional dapat diolah untuk menemukan pola yang relevan dalam memprediksi peluang keberhasilan berhenti merokok.

Salah satu algoritma *machine learning* yang populer adalah *artificial neural network* (ANN). ANN terinspirasi dari cara kerja jaringan saraf manusia dan mampu menangani hubungan non-linear antara variabel input dan output. ANN sangat efektif untuk memodelkan data yang kompleks seperti perilaku perokok karena mampu belajar dari pola tersembunyi dan melakukan generalisasi dari data pelatihan (Nuryunarsih et al., 2023). Dengan demikian, ANN menjadi pilihan tepat dalam membangun model prediksi status perokok yang akurat. Gambar 1 dibawah menunjukkan arsitektur dasar dari Artificial Neural Network (ANN), yaitu feed-forward neural network. ANN bekerja dengan meniru cara kerja otak manusia

dalam memproses informasi melalui neuron-neuron yang saling terhubung.

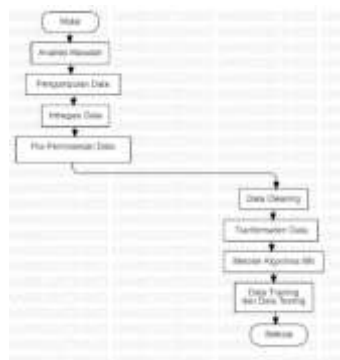
Seiring berkembangnya teknologi, *machine learning* menjadi metode yang banyak digunakan dalam penelitian kesehatan karena kemampuannya mempelajari pola kompleks secara otomatis. *Machine learning* memungkinkan model untuk melakukan pembelajaran dari data historis dan memberikan prediksi yang lebih akurat dibandingkan metode statistik tradisional (Issabakhsh et al., 2023). Dalam konteks prediksi perilaku merokok, *machine learning* terbukti mampu mengintegrasikan banyak variabel sekaligus, sehingga memberikan gambaran prediksi yang lebih komprehensif mengenai peluang keberhasilan berhenti merokok.

Proses *prediksi* sendiri merupakan upaya memperkirakan nilai atau status suatu variabel berdasarkan data historis. Dalam penelitian ini, prediksi difokuskan untuk menentukan status perokok (berhasil atau tidak berhasil berhenti). Keakuratan prediksi sangat penting agar model dapat digunakan untuk mendukung keputusan tenaga kesehatan dalam memberikan intervensi berhenti merokok yang lebih personal.

Untuk menilai kualitas model prediksi, salah satu metrik yang paling umum digunakan adalah *Root Mean Squared Error* (RMSE). RMSE mengukur rata-rata tingkat kesalahan antara nilai prediksi dan nilai aktual, di mana nilai yang lebih rendah menunjukkan model yang lebih akurat (Hidayatulloh et al., 2025). Berdasarkan hasil pengujian menggunakan RapidMiner, model neural network pada penelitian ini menghasilkan nilai RMSE sebesar 0.409, yang menunjukkan bahwa model memiliki tingkat kesalahan yang relatif rendah dan cukup baik dalam memprediksi status perokok.

METODE PENELITIAN

Secara garis besar tahapan jalannya penelitian dapat dilihat pada gambar dibawah ini.



Gambar 1 Jalan Penelitian

Analisis Masalah

Tahap analisis masalah menetapkan fokus penelitian pada pengembangan model prediksi hasil berhenti merokok. Masalahnya adalah kurangnya alat klinis yang efektif untuk memprediksi probabilitas keberhasilan berhenti merokok secara individu berdasarkan berbagai faktor risiko. Tujuan spesifiknya adalah menciptakan model yang mampu mengklasifikasikan hasil akhir perokok (berhasil berhenti atau relaps) untuk periode waktu tertentu, sehingga dapat membantu tenaga medis dalam pengambilan keputusan intervensi (Hock et al., 2023).

Pengumpulan Data

Data yang dikumpulkan bersifat kuantitatif, mencakup variabel prediktor dan variabel hasil. Variabel ini berasal dari data klinis dan data perilaku perokok, seperti yang ditunjukkan pada tabel di bawah. Data ini berfungsi sebagai input bagi model *Machine Learning* yang akan dikembangkan.

Table 1 . Variabel (Kolom) Dalam Dataset

No	Nama Variabel (Kolom)	Kategori	Keterangan Singkat
1	age	Demografi	Usia subjek.
2	height(cm)	Antropometri	Tinggi badan dalam sentimeter (cm).
3	weight(kg)	Antropometri	Berat badan dalam kilogram (kg).
4	waist(cm)	Antropometri	Lingkar pinggang (cm).
5	eyesight(left)	Kesehatan Organ	Ketajaman penglihatan mata kiri.
6	eyesight(right)	Kesehatan Organ	Ketajaman penglihatan mata kanan.
7	hearing(left)	Kesehatan Organ	"Kondisi pendengaran kiri (1=Normal, 2=Abnormal)."
8	hearing(right)	Kesehatan Organ	"Kondisi pendengaran kanan (1=Normal, 2=Abnormal)."
9	systolic	Klinis Vital	Tekanan darah sistolik (mmHg).
10	relaxation	Klinis Vital	Tekanan darah diastolik (mmHg).
11	fasting blood sugar	Klinis Metabolik	Kadar gula darah puasa.
12	Cholesterol	Klinis Metabolik	Kadar kolesterol total.
13	triglyceride	Klinis Metabolik	Kadar trigliserida.
14	HDL	Klinis Metabolik	Highdensity lipoprotein (Kolesterol baik).
15	LDL	Klinis Metabolik	Lowdensity lipoprotein (Kolesterol jahat).
16	hemoglobin	Klinis Darah	Kadar hemoglobin.
17	Urine protein	Klinis Fungsi Ginjal	"Kadar protein dalam urin (skala 16, menunjukkan gangguan ginjal)."
18	serum creatinine	Klinis Fungsi Ginjal	Kadar kreatinin serum.

19	AST	Klinis Fungsi Hati	Aspartate Aminotransferase (enzim hati).
20	ALT,Klinis Fungsi Hati	Alanine Aminotransferase (enzim hati).	
21	Gtp	Klinis Fungsi Hati	"Gammaglutamyl transpeptidase (enzim hati, sering terkait konsumsi alkohol/rokok)."
22	dental caries	Kesehatan Gigi	"Status karies gigi (0=Tidak ada, 1=Ada)."
23	smoking	Variabel Hasil (Target)	"Status merokok (0=Non Perokok/Berhenti, 1=Perokok/Gagal Berhenti)."

Integrasi Data

Data yang mungkin berasal dari sumber yang berbeda (misalnya, data klinis dari rekam medis dan data perilaku dari kuesioner) diintegrasikan ke dalam satu dataset tunggal. Langkah ini sangat penting untuk menyelaraskan semua fitur dan observasi. Proses integrasi memastikan setiap baris data mewakili satu subjek dengan semua variabel yang relevan, serta memverifikasi konsistensi tipe dan format data sebelum memasuki tahap pemrosesan.

Pra-Pemrosesan Data

Tahap ini merupakan fondasi untuk memastikan model *Machine Learning* dapat belajar secara efektif, karena kualitas model sangat bergantung pada kualitas data input (Nisar et al., 2021).

Data Cleaning (Pembersihan Data)

Pembersihan data mencakup penanganan nilai yang hilang (missing values) menggunakan teknik imputasi yang sesuai (misalnya, mengganti dengan nilai median atau rata-rata) serta identifikasi dan penanganan outlier (nilai ekstrem) yang dapat mendistorsi proses pelatihan algoritma *Neural Network*.

Transformation Data (Transformasi Data)

Transformasi data melibatkan penskalaan dan pengkodean fitur. Variabel kontinu akan mengalami Normalisasi Min-Max untuk menskalakan nilainya ke rentang [0, 1], yang krusial untuk stabilitas dan kecepatan konvergensi *Neural Network*. Variabel kategorikal (seperti jenis kelamin atau status sosial) diubah menjadi format numerik melalui One-Hot Encoding agar dapat diolah oleh model.

Metode Algoritma Neural Network

Model prediksi dikembangkan menggunakan Jaringan Saraf Tiruan (Neural Network - NN), khususnya arsitektur Multilayer Perceptron (MLP). NN dipilih karena kemampuannya memodelkan hubungan non-linier yang kompleks antar-variabel, suatu karakteristik umum dalam data hasil kesehatan (Tong et al., 2024).

Arsitektur NN terdiri dari Lapisan Input (berdasarkan jumlah fitur prediktor), satu atau lebih Lapisan Tersembunyi (Hidden Layers) dengan fungsi aktivasi ReLU, dan Lapisan Output tunggal dengan fungsi aktivasi Sigmoid untuk menghasilkan probabilitas prediksi. Model dilatih dengan teknik Backpropagation dan dioptimalkan menggunakan algoritma Adam untuk meminimalkan Loss Function (Binary Cross-Entropy).

Data Training dan Data Testing

Dataset yang telah diproses dibagi menjadi dua himpunan data untuk memastikan pengujian model yang objektif (Furizal et al., 2023):

- **Data Training (Pelatihan):** Sekitar 70% hingga 80% dari total data digunakan untuk melatih model NN, menyesuaikan bobot dan bias secara iteratif.
- **Data Testing (Pengujian):** Sisa 20% hingga 30% data disimpan untuk pengujian akhir. Data ini tidak pernah dilihat oleh model selama pelatihan, sehingga memberikan evaluasi yang akurat mengenai kemampuan generalisasi model terhadap data baru.

HASIL DAN PEMBAHASAN

Hasil pengujian menunjukkan bahwa model yang dibangun menggunakan data *training* mampu mempelajari pola dari variabel input dengan baik, yang terlihat dari nilai error pada training yang relatif rendah. Ketika diterapkan pada data testing, model tetap menunjukkan performa yang stabil, yang berarti kemampuan generalisasinya cukup baik dan tidak terjadi overfitting yang signifikan. Perbandingan antara nilai prediksi dan nilai aktual pada data *testing* juga memperlihatkan selisih yang konsisten, sehingga model dapat dikatakan mampu memprediksi dengan akurat berdasarkan pola yang telah dipelajari dari data *training*. Dengan demikian, model layak digunakan untuk proses prediksi selanjutnya karena telah melalui evaluasi yang menunjukkan hasil memadai pada kedua jenis dataset.

Data Training

Tahap ini merupakan di mana model belajar pola dari sejumlah besar observasi kesehatan untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap status *smoking* (merokok).

Table 2 Data training

No	age	Height(cm)	cholesterol	...	Smoking
1	35	170	239	...	1
2	20	175	211	...	0
...	...	...	...	...	...
149	35	170	259	...	0

Data Testing

Kumpulan data ini memiliki tujuan untuk menguji kemampuan model yang sudah selesai dilatih dalam menghadapi data baru di dunia nyata. Berdasarkan teks, evaluasi hasil dari set data testing ini akan memberikan informasi penting mengenai kemampuan model untuk mengkonfirmasi generalisasinya—artinya, seberapa baik model dapat membuat prediksi yang relevan dan dapat diandalkan pada data yang berbeda dari yang digunakan untuk melatihnya.

Table 3 Data Testing

No	age	Height(cm)	Cholestrol	...	Smoking
1	50	165	188	...	0
2	20	175	211	...	0
...	...	...	...	...	...
149	35	170	259	...	0

Implementasi Algoritma Neural Network

Dalam penelitian ini untuk memprediksi status merokok (smoking) melibatkan penggunaan model Multilayer Perceptron (MLP). Model ini dirancang dengan Lapisan Input yang menerima 20+ fitur klinis terstandarisasi (seperti tekanan darah, Gtp, dan gula darah), beberapa Lapisan Tersembunyi dengan fungsi aktivasi ReLU untuk mempelajari pola non-linier yang kompleks, dan Lapisan Output tunggal dengan fungsi Sigmoid untuk menghasilkan probabilitas klasifikasi biner (perokok/berhenti). Proses pelatihan dijalankan menggunakan optimizer Adam untuk meminimalkan loss Binary Cross-Entropy, di mana bobot jaringan terus disesuaikan melalui backpropagation hingga model mencapai kinerja optimal dalam membedakan subjek perokok dan non-perokok pada data validasi.

Evaluasi dan Pengujian Model



Gambar 2 Tahapan Proses

Pada gambar diatas menggambarkan alur kerja pembelajaran mesin yang menggunakan arsitektur pemodelan prediktif. Alur ini dimulai dengan membaca data (Read CSV) dari dua sumber yang berbeda. Data dari sumber pertama kemudian diteruskan ke operator Jaringan Saraf Tiruan (Neural Net) untuk melatih model berdasarkan data tersebut. Setelah model dilatih, ia kemudian digunakan bersama dengan data dari sumber kedua melalui operator Terapkan Model (Apply Model) untuk menghasilkan prediksi. Terakhir, prediksi yang dihasilkan oleh model dievaluasi oleh operator Kinerja (Performance) untuk menilai seberapa baik model tersebut bekerja dalam membandingkan nilai prediksi dengan nilai yang sebenarnya ada pada data.

probabilitas prediksi (P_i) dan label aktual (Y_i) Semakin kecil nilai RMSE (mendekati nol), semakin baik kinerja model, karena ini berarti probabilitas yang diprediksi oleh *Neural Network* Anda rata-rata sangat dekat dengan nilai hasil sebenarnya—yaitu, probabilitas mendekati 1 untuk perokok dan mendekati 0 untuk non-perokok/berhenti. Oleh karena itu, RMSE berfungsi sebagai metrik ringkas untuk mengukur seberapa jauh, rata-rata, kesalahan prediksi model Anda, meskipun metrik seperti *Binary Cross-Entropy* atau F1-Score biasanya lebih informatif untuk klasifikasi.

Rumus RMSE (Root Mean Square Error) yang standar dan benar adalah:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - Y_i)^2}$$

Keterangan:

- P_i adalah nilai yang diprediksi (dalam kasus ini, probabilitas $\in [0,1]$).
- Y_i adalah nilai actual (nilai target sebenarnya, yaitu 0 atau 1).
- N adalah jumlah sampel data

Dalam konteks penelitian ini:

1. Jika $Y_i = 1$ (Subjek adalah Perokok): Kesalahan dihitung dari seberapa jauh prediksi probabilitas (P_i) model ini dari angka 1.
 - Contoh: Model memprediksi $P_i = 0.6$ Error-nya adalah $(0.6 - 1)^2 = (-0.4)^2 = 0.16$.
2. Jika $Y_i = 0$ (Subjek Berhenti Merokok): Kesalahan dihitung dari seberapa jauh prediksi probabilitas (P_i) model Anda dari angka 0.
 - Contoh: Model memprediksi $P_i = 0.8$. Error-nya adalah $(0.8 - 0)^2 = (0.8)^2 = 0.64$.

Performance Vector



Gambar 5 . Performancevector

hasil dari proses Machine Learning yang dijalankan menggunakan perangkat lunak RapidMiner Studio, di mana model Neural Network digunakan untuk memprediksi status smoking. Secara keseluruhan, alur proses menunjukkan pembacaan data, pelatihan model Neural Network, penerapan model (Apply Model) pada data uji, dan evaluasi kinerja (Performance). Metrik kunci yang disajikan adalah Root Mean Squared Error (RMSE) sebesar 0.409, yang mengukur rata-rata besarnya kesalahan antara probabilitas prediksi model dan nilai aktual status merokok (0 atau 1). Data statistik menunjukkan bahwa nilai label aktual smoking memiliki rata-rata 0.367, sedangkan rata-rata prediksi probabilitas oleh model adalah 0.297, mengindikasikan model cenderung sedikit meremehkan (underestimate) proporsi kasus perokok dibandingkan dengan data aktual, meskipun nilai RMSE 0.409 sudah cukup rendah untuk menunjukkan kinerja prediksi yang wajar dalam tugas klasifikasi biner ini.

KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi status berhenti merokok dengan memanfaatkan metode Neural Network (Jaringan Saraf Tiruan) yang diimplementasikan melalui arsitektur Multilayer Perceptron (MLP) dalam perangkat lunak RapidMiner. Model ini dibangun untuk mengatasi tantangan klinis terkait rendahnya tingkat keberhasilan dan kesulitan dalam memprediksi secara individual peluang seseorang untuk berhasil berhenti merokok, dengan mengolah data yang kompleks, termasuk variabel ketergantungan nikotin, kadar CO dalam napas, hingga kondisi emosional. Melalui serangkaian tahapan pra-pemrosesan data seperti normalisasi dan pembagian data training dan testing, model diuji dan menghasilkan nilai Root Mean Squared Error (RMSE) sebesar 0.409. Angka RMSE ini mengindikasikan bahwa model dapat melakukan prediksi risiko merokok dengan tingkat kesalahan yang relatif rendah dan stabil, menunjukkan efektivitas metode Neural Network dalam memodelkan hubungan nonlinier di antara variabel-variabel klinis. Dengan demikian, model prediksi yang dikembangkan ini menawarkan solusi berbasis data yang potensial untuk menjadi alat bantu dalam pengambilan keputusan klinis, guna merancang program intervensi berhenti merokok yang lebih tepat dan personal di masa depan.

DAFTAR PUSTAKA

Alhumaidi, N. H., Dermawan, D., Kamaruzaman, H. F., & Alotaqi, N. (2025). The Use of Machine Learning for Analyzing Real-World Data in Disease Prediction and Management: Systematic Review.

- JMIR Medical Informatics*, 13, 1–22. <https://doi.org/10.2196/68898>
- An, Q., Rahman, S., Zhou, J., & Kang, J. J. (2023). A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. *Sensors*, 23(9). <https://doi.org/10.3390/s23094178>
- Furizal, Ma'arif, A., & Rifaldi, D. (2023). Application of Machine Learning in Healthcare and Medicine: A Review. *Journal of Robotics and Control (JRC)*, 4(5), 621–631. <https://doi.org/10.18196/jrc.v4i5.19640>
- Habehh, H., & Gohel, S. (2021). Machine Learning in Healthcare. *Current Genomics*, 22(4), 291–300. <https://doi.org/10.2174/1389202922666210705124359>
- Herath, H. M. P. (2025). Comprehensive Review of Tobacco Control Measures in Sri Lanka: Insights from the WHO Report on the Global Tobacco Epidemic, 2023. *European Journal of Medical and Health Sciences*, 7(1), 29–32. <https://doi.org/10.24018/ejmed.2025.7.1.2183>
- Hersi, M., Beck, A., Hamel, C., Esmaeilisaraji, L., Pussegoda, K., Austin, B., Ahmadzai, N., Pratt, M., Thuku, M., Yazdi, F., Bennett, A., Shaver, N., Vyas, N., Skidmore, B., Hutton, B., Manuel, D., Morrow, M., Pakhale, S., Presseau, J., ... Stevens, A. (2024). Effectiveness of smoking cessation interventions among adults: an overview of systematic reviews. *Systematic Reviews*, 13(1), 1–33. <https://doi.org/10.1186/s13643-024-02570-9>
- Hidayatulloh, R., Aji, W., & Prabowo, E. (2025). Analisis Performansi Model Machine Learning dalam Klasifikasi Penyakit Diabetes Tipe 2. *JURIKOM (Jurnal Riset Komputer)*, 12(4). <https://doi.org/10.30865/jurikom.v12i4.8747>
- Hock, E. S., Franklin, M., Baxter, S., Clowes, M., Chilcott, J., & Gillespie, D. (2023). Covariates of success in quitting smoking: a systematic review of studies from 2008 to 2021 conducted to inform the statistical analyses of quitting outcomes of a hospital-based tobacco dependence treatment service in the United Kingdom. *NIHR Open Research*, 3, 1–19. <https://doi.org/10.3310/nihropenres.13427.2>
- Issabakhsh, M., Sánchez-Romero, L. M., Le, T. T. T., Liber, A. C., Tan, J., Li, Y., Meza, R., Mendez, D., & Levy, D. T. (2023). Machine learning application for predicting smoking cessation among US adults: An analysis of waves 1-3 of the PATH study. *PLoS ONE*, 18(6 JUNE), 1–16. <https://doi.org/10.1371/journal.pone.0286883>
- MASTROBATTISTA, L., MORTALI, C., SOLIMINI, R., PICHINI, S., & PALMI, I. (2023). The national report on tobacco use 2023. *Tabaccologia*, 21(2), 22–26. <https://doi.org/10.53127/tblg-2023-a011>
- Nisar, D. E. M., Amin, R., Shah, N. U. H., Ghamdi, M. A. A., Almotiri, S. H., & Alruily, M. (2021). Healthcare Techniques through Deep Learning: Issues, Challenges and Opportunities. *IEEE Access*, 9, 98523–98541. <https://doi.org/10.1109/ACCESS.2021.3095312>
- Nuryunarsih, D., Okatiranti, O., & Herawati, L. (2023). Artificial neural network machine learning prediction of the smoking behavior and health risks perception of Indonesian health professionals. *Environmental Analysis Health and Toxicology*, 38(1). <https://doi.org/10.5620/eaht.2023003>
- Pfeifer, B., & Schimek, M. G. (2021). A hierarchical clustering and data fusion approach for disease subtype discovery. *Journal of Biomedical Informatics*, 113, 103636. <https://doi.org/10.1016/j.jbi.2020.103636>
- Tong, J., Li, L., Reps, J. M., Lorman, V., Jing, N., Edmondson, M., Lou, X., Jhaveri, R., Kelleher, K. J., Pajor, N. M., Forrest, C. B., Bian, J., Chu, H., & Chen, Y. (2024). Advancing Interpretable Regression Analysis for Binary Data: A Novel Distributed Algorithm Approach. *Statistics in Medicine*, 43(29), 5573–5582. <https://doi.org/10.1002/sim.10250>
- Yapar, D., & İlhan, M. N. (2025). Global Tobacco Industry. *Addicta: The Turkish Journal on Addictions*, 12(1), 1–9. <https://doi.org/10.5152/ADDICTA.2025.23137>