



Penerapan Seleksi Fitur Information Gain dan Metode Backpropagation Neural Network Untuk Klasifikasi Atrisi Karyawan

Dinyah Fithara, Elvia Budianita*, Iis Afrianty, Siska Kurnia Gusti

Fakultas Sain dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹12050123134@students.uin-suska.ac.id, ^{2,*}elvia.budianita@uin-suska.ac.id, ³iis.afrianty@uin-suska.ac.id, ⁴siskakurniagusti@uin-suska.ac.id

Email Penulis Korespondensi: elvia.budianita@uin-suska.ac.id

Abstrak—Pengelolaan atrisi karyawan menjadi tantangan penting bagi perusahaan karena melibatkan faktor biaya, waktu, dan risiko kesalahan dalam pengambilan keputusan. Permasalahan ini memerlukan strategi bisnis data untuk melakukan prediksi yang lebih akurat terhadap karyawan yang berpotensi mengalami pemberhentian. Penelitian ini menerapkan metode seleksi fitur Information Gain dan algoritma Backpropagation Neural Network (BPNN) dalam proses klasifikasi atrisi karyawan dengan tujuan meningkatkan akurasi dan efisiensi model prediksi. BPNN dipilih karena arsitekturnya yang lebih sederhana, waktu pelatihan yang lebih cepat dan lebih stabil untuk dataset berukuran kecil hingga menengah. Dengan bantuan seleksi fitur Information Gain, BPNN dapat mencapai performa terbaik tanpa memerlukan arsitektur yang kompleks. Dataset yang digunakan terdiri dari 35 atribut dan 1.470 data karyawan yang mencakup berbagai faktor seperti usia, tingkat pendapatan, dan status pekerjaan. Tahapan penelitian meliputi seleksi fitur berdasarkan nilai Information Gain dengan ambang batas tertentu, pembagian data menggunakan k-fold cross validation, serta pelatihan model menggunakan BPNN dengan variasi parameter seperti learning rate dan jumlah neuron tersembunyi. Hasil menunjukkan bahwa kombinasi Information Gain dan BPNN mampu meningkatkan akurasi klasifikasi dibandingkan tanpa seleksi fitur, dengan capaian akurasi rata-rata tertinggi sebesar 87.28% ketika menggunakan 25 atribut terpilih, dengan konfigurasi BPNN berupa learning rate 0.001, 35 neuron tersembunyi, dan 50 epoch. Fitur dengan nilai Information Gain tertinggi adalah JobLevel, OverTime, MaritalStatus, dan MonthlyIncome. Penelitian ini menunjukkan bahwa pendekatan yang dilakukan berhasil meningkatkan performa prediksi atrisi karyawan, sehingga dapat menjadi acuan dalam membangun model berbasis data yang membantu meningkatkan upaya retensi karyawan.

Kata Kunci: Pengurangan Karyawan; Data Mining; Information Gain; Backpropagation Neural Network; Klasifikasi

Abstract—Employee attrition management is a critical challenge for organizations as it involves costs, time, and the risk of decision-making errors. This problem requires a data-driven business strategy to achieve more accurate predictions of employees who are potentially at risk of termination. This study applies the Information Gain feature selection method and the Backpropagation Neural Network (BPNN) algorithm in the employee attrition classification process with the aim of increasing the accuracy and efficiency of the prediction model. BPNN is chosen due to its simpler architecture, faster training time, and greater stability for small to medium sized datasets. With the assistance of Information Gain feature selection, BPNN is able to achieve optimal performance without requiring a complex architecture. The dataset used consist of 35 attributes and 1.470 employee records covering various factor such as age, income level, and employment status. The research stages include feature selection based on information gain values with specific thresholds, data partitioning using k-fold cross validation, and model training using BPNN with variations of learning rates and hidden neuron counts. The results show that the combination of Information Gain and BPNN improves classification accuracy compared to models without feature selection, achieving the highest average accuracy of 87.28% when using 25 selected attributes, with a BPNN configuration of learning rate 0.001, 35 hidden neurons, and 50 epochs. The attributes with the highest Information Gain score include JobLevel, OverTime, MaritalStatus, and MonthlyIncome. This study demonstrates that the proposed approach successfully enhances the prediction performance of employee attrition and can serve as a foundation for developing data-driven models that support employee retention efforts.

Keywords: Employee Attrition; Data Mining; Information Gain; Backpropagation Neural Network; Classification

1. PENDAHULUAN

Salah satu tantangan besar yang dihadapi perusahaan adalah pengelolaan kelebihan karyawan atau keputusan terkait pengurangan karyawan. Proses seleksi ini sangat penting karena tidak hanya memakan waktu dan biaya tetapi juga berisiko menimbulkan kerugian jika keputusan yang diambil kurang tepat [1]. Tidak hanya memengaruhi stabilitas organisasi, tetapi juga menimbulkan bebas biaya rekrutmen, pelatihan, dan penurunan produktivitas yang signifikan karena pengambilan keputusan yang tidak berbasis data sering kali menyebabkan kesalahan strategis, seperti kehilangan karyawan potensial dan meningkatnya *turnover* jangka panjang [2][3]. Pengurangan atau atrisi karyawan merupakan hilangnya karyawan akibat faktor-faktor seperti pensiun, pengunduran diri, penghapusan jabatan, kesehatan pribadi, atau pemutusan hubungan kerja (PHK), menjadi tantangan tersendiri bagi manajemen. Hal ini menuntut perusahaan untuk secara aktif memantau dan memprediksi masalah terkait SDM [4].

Menghadapi tantangan ini, pengambilan keputusan berbasis data menjadi pendekatan penting untuk memprediksi kemungkinan atrisi karyawan [5]. Maka memprediksi kemungkinan atrisi karyawan dengan menggunakan teknik data mining dan machine learning menjadi solusi yang dapat diterapkan. Beberapa penelitian telah memanfaatkan algoritma seperti *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Logistic Regression*, hingga *Random Forest* telah digunakan untuk klasifikasi pengurangan karyawan. Seperti pada penelitian Benabou di tahun 2025 [6] yang menguji dengan algoritma *Logistic Regression* menghasilkan akurasi sebesar 87.7%. Penelitian lain juga dilakukan oleh Raza [7] yang menghasilkan akurasi sebesar 88% menggunakan algoritma *Support Vector Machine* (SVM). Namun demikian,



hasil prediksi sangat bergantung pada relevansi dan fitur yang digunakan, sehingga diperlukan teknik seleksi fitur seperti *Information Gain*, *Select K-Best*, dan *Recursive Feature Elimination* [8].

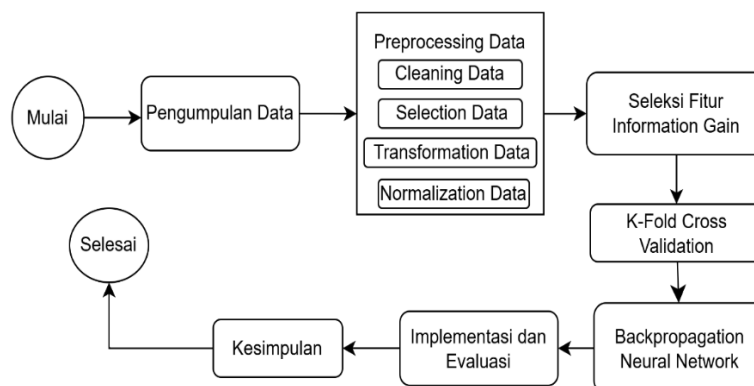
Pada penelitian Sari dan Lhaksamana [9] menggunakan kombinasi *Information Gain* dan *Random Forest* untuk menghasilkan akurasi prediksi pada kasus atrisi karyawan. Hasil penelitian ini menunjukkan bahwa kombinasi kedua metode ini mampu meningkatkan prediksi secara signifikan, dengan capaian akurasi sebesar 89,2%. Sementara itu, penelitian Mardiana[10] mengembangkan pendekatan serupa menggunakan *Naive Bayes* sebagai model klasifikasi serta *Information Gain* sebagai seleksi fitur. Kombinasi ini terbukti secara signifikan meningkatkan prediksi dengan membandingkan hasil akurasi tanpa menggunakan seleksi fitur sebesar 94.15%, sedangkan dengan menggunakan seleksi fitur *Information Gain* akurasi naik menjadi 96.49%. Temuan ini menunjukkan pentingnya tahapan *feature selection* untuk memilih fitur yang paling informatif. Penelitian lain oleh Itqon[11] menggunakan *Random Forest* untuk memprediksi tingkat *employee attrition* pada dataset industri *rolling stock*. Hasil penelitian ini menunjukkan kemampuan algoritma *Random Forest* dalam mengklasifikasi *Attrition* dengan akurasi tertinggi sebesar 93,1%.

Selain gabungan algoritma tersebut, penggunaan jaringan saraf tiruan (*Artificial Neural Network*) juga menjadi perhatian penting karena kemampuannya mengenali pola non-linear yang kompleks antar variabel. Pada penelitian Wahyudi di tahun 2025 menerapkan algoritma *Backpropagation Neural Network* (BPNN) yang dipadukan dengan seleksi fitur *Information Gain* untuk meningkatkan akurasi dalam proses klasifikasi. Proses pelatihan ini menggunakan parameter *learning rate* sebesar 0.01 dan (16,8) hidden neurons pada satu lapisan tersembunyi (hidden layer). Hasil pengujian menunjukkan bahwa model BPNN dengan seleksi fitur *Information Gain* memperoleh akurasi rata-rata 82,15% [12]. Penelitian lain yang dilakukan oleh Fauzan di tahun 2024 membandingkan algoritma BPNN, *Decision Tree*, dan KNN tanpa seleksi fitur dan menemukan bahwa BPNN memberikan akurasi tertinggi sebesar 93,05% [13]. Penelitian ini memperkuat bahwa BPNN efektif dalam klasifikasi berbasis pola, terutama ketika dioptimalkan dengan seleksi fitur yang tepat. Penelitian serupa dilakukan oleh Khair[14] dengan menggunakan *Information Gain* untuk seleksi fitur pada klasifikasi struktur tulang tengkorak yang memperoleh akurasi 93,31% menggunakan algoritma *Backpropagation Neural Network*. Hasil tersebut memperlihatkan efektivitas kombinasi metode seleksi fitur pada klasifikasi dan jaringan saraf tiruan dalam meningkatkan performa klasifikasi.

Berdasarkan literatur tersebut, dapat diidentifikasi adanya celah penelitian dalam penerapan algoritma *Backpropagation Neural Network* dengan metode seleksi fitur *Information Gain* pada dataset atrisi karyawan. Sebagian besar penelitian sebelumnya hanya menganalisis penggunaan algoritma klasik seperti SVM dan *Random Forest*, tanpa eksplorasi mendalam pada jaringan saraf tiruan yang dioptimalkan dengan seleksi atribut. Oleh karena itu, tujuan dari penelitian ini untuk menerapkan kombinasi *Information Gain* dan *Backpropagation Neural Network* dalam proses klasifikasi pengurangan karyawan untuk meningkatkan akurasi model, efisiensi komputasi, serta mendukung pengambilan keputusan strategis berbasis data dalam pengelolaan sumber daya manusia.

2. METODOLOGI PENELITIAN

Berikut Gambar 1 merupakan tahapan penelitian yang dilakukan.



Gambar 1 Tahapan Penelitian

Pada penelitian ini menerapkan seleksi fitur *Information Gain* untuk menemukan atribut yang sangat berpengaruh terhadap variabel target. Fitur-fitur yang telah diseleksi selanjutnya akan digunakan sebagai input menggunakan algoritma *Backpropagation Neural Network* untuk menghasilkan model klasifikasi yang lebih efisien dan akurat. Kombinasi ini diharapkan dapat meningkatkan kinerja prediksi sekaligus meminimalkan kemungkinan terjadinya *overfitting* pada model. Alur lengkap tahapan penelitian ditunjukkan pada Gambar 1, yang menampilkan langkah-langkah penelitian secara keseluruhan.

2.1 Tahapan Penelitian

Data yang digunakan untuk penelitian ini adalah dataset *IBM HR Employee Attrition* yang di dapatkan dari platform *Kaggle* [15]. Dataset ini berisi 1470 data dengan 35 atribut dan menggunakan bahasa Inggris, 1 atribut sebagai variabel



target yang terdiri dari dua kelas yaitu *Yes* dan *No*. Terdapat 1233 pada kelas “*No*” dan 237 pada kelas “*Yes*” pada dataset ini. Deskripsi fitur pada dataset ditunjukkan pada Tabel 1.

Tabel 1. Fitur-fitur dan tipe data

Feature	Data Type	Features	Data Type
Age (Usia)	Numerical Value	MonthlyIncome (Pendapatan Bulanan)	Numerical Value
Attrition (Pengunduran Diri)	Categorical Value	MonthlyRate (Tarif Bulanan)	Numerical Value
BusinessTravel (Perjalanan Bisnis)	Categorical Value	NumCompaniesWorked (Jumlah Perusahaan yang Pernah Dikerjakan)	Numerical Value
DailyRate (Tarif Harian)	Numerical Value	Over18 (Diatas 18 Tahun)	Categorical Value
Department (Departemen)	Categorical Value	OverTime (Lembur)	Categorical Value
DistanceFromHome (Jarak dari Rumah)	Numerical Value	PercentSalaryHike (Presentase Kenaikan Gaji)	Numerical Value
Education (Tingkat Pendidikan)	Numerical Value	PerformanceRating (Penilaian Kinerja)	Numerical Value
EducationField (Bidang Pendidikan)	Categorical Value	RelationshipSatisfaction (Kepuasan Hubungan Inerpersonal)	Numerical Value
EmployeeCount (Jumlah Karyawan)	Numerical Value	StandardHours (Jam Kerja Standar)	Numerical Value
EmployeeNumber (Nomor Karyawan)	Numerical Value	StockOptionLevel (Level Opsi Saham)	Numerical Value
EnvironmentSatisfaction (Kepuasan Lingkungan Kerja)	Numerical Value	TotalWorkingYears (Total Tahun Bekerja)	Numerical Value
Gender (Jenis Kelamin)	Categorical Value	TrainingTimeLastYears (Jumlah Pelatihan Setahun Terakhir)	Numerical Value
HourlyRate (Tarif per Jam)	Numerical Value	WorkLifeBalance (Keseimbangan Kerja-Hidup)	Numerical Value
JobInvlovement (Keterlibatan Kerja)	Numerical Value	YearsAtCompany (Tahun Bekerja di Perusahaan)	Numerical Value
JobLevel (Tingkat Jabatan)	Numerical Value	YearsInCurrentRole (Tahun di Posisi Saat ini)	Numerical Value
JobRole (Jabatan Kerja)	Categorical Value	YearsSinceLastPromotion (Tahun Sejak Promosi Terakhir)	Numerical Value
JobSatisfaction (Kepuasan Kerja)	Numerical Value	YearsWithCurrManager (Tahun Bersama Manejer Saat Ini)	Numerical Value
MaritalStatus (Status Pernikahan)	Categorical Value		

Tabel 1 menyajikan daftar fitur-fitur dan tipe data dari dataset *IBM HR Employee Attrition* yang akan digunakan dalam proses klasifikasi.

2.2 Preprocessing Data

Pada tahap preprocessing, data akan proses pembersihan, transformasi, dan penataan ulang agar data sesuai dengan kebutuhan analisis atau permodelan yang akan dilakukan. Langkah-langkah dalam tahap data preprocessing adalah sebagai berikut:

- Cleaning Data** : Tahap ini dilakukan untuk membersihkan data dari berbagai masalah, seperti nilai yang hilang (missing value) dan duplikat [16]. Apabila terdapat data yang sama atau berisi nilai kosong, data tersebut akan dihapus agar dataset menjadi lebih bersih dan akurat untuk dilakukan proses analisis selanjutnya.
- Selection Data** : Pada tahap ini memastikan data yang digunakan tidak memuat informasi yang berulang atau tidak relevan [17]. Pada penelitian ini, ada beberapa fitur seperti ‘StandarHours’, ‘Over18’, ‘EmployeeCount’, dan ‘EmployeeNumber’ yang dihapus karena hanya ID atau hanya berisi nilai yang sama.
- Transformasi Data** : Tahap transformasi dilakukan untuk menyiapkan dataset sebelum masuk ke proses seleksi fitur dan permodelan, yaitu dengan mengubah ke dalam format atau representasi yang sesuai untuk proses analisis lebih lanjut agar dataset menjadi lebih sederhana tanpa kehilangan informasi penting [18]. Penelitian ini menggunakan dua jenis transformasi, yaitu *Label Encoding* dan *One-hot Encoding*, yang diterapkan sesuai dengan tipe data setiap variabel. Pertama, pada variabel target Attrition, dilakukan label encoding untuk mengubah nilai kategorikal menjadi bentuk numerik. Melalui proses ini, nilai “Yes” dikodekan menjadi 1 dan “No” dikodekan menjadi 0 sehingga target dapat diolah oleh algoritma klasifikasi. Selanjutnya, untuk variabel fitur yang bersifat kategorikal seperti



BusinessTravel, Department, EducationField, Gender, JobRole, MaritalStatus, dan OverTime, digunakan teknik One-Hot Encoding. Teknik ini menghasilkan variabel baru dalam bentuk biner (0 atau 1) untuk setiap kategori, sehingga model dapat mempelajari informasi tanpa adanya orde atau tingkatan pada kategori. Tahap ini memastikan bahwa seluruh variabel pada dataset telah berada dalam format numerik untuk dilanjutkan ke tahap Normalisasi.

- d. Normalisasi Data: Pada tahap selanjutnya akan dilakukan proses normalisasi terhadap dataset menggunakan metode *Min-Max Normalization* agar seluruh nilai pada variabel memiliki rentang antara 0 hingga 1[19]. Tujuan dari normalisasi untuk mengatasi nilai dataset agar memiliki skala atau distribusi yang konsisten. *Min-Max Normalization* dihitung menggunakan rumus pada Persamaan 1 berikut.

$$X_{norm} = \left(\frac{X - X_{min}}{X_{max} - X_{min}} \right) \quad (1)$$

Pada persamaan 1, X merupakan nilai asli dari atribut data, seperti nilai dari fitur Age. Selanjutnya, X_{min} adalah nilai minimal fitur dari seluruh data, sedangkan X_{max} merupakan nilai maksimal dari fitur pada seluruh data. Hasil dari proses ini adalah X_{norm} , sebagai nilai yang sudah dinormalisasi.

2.3 Seleksi Fitur Information Gain

Seleksi fitur adalah istilah umum dari bidang pembelajaran mesin dan statistik dan mengacu pada proses pemilihan subset fitur yang dianggap paling relevan sebelum model dibuat.[20]. Information Gain merupakan salah satu metode pemilihan fitur yang banyak digunakan dalam machine learning, memproses data untuk memilih fitur yang paling relevan dengan mengevaluasi kontribusi setiap fitur dalam mengurangi jarak dalam memprediksi target[21]. Perhitungan Information Gain di mulai dengan menghitung nilai Gain tiap fitur yang bisa dilakukan dengan persamaan berikut:

- a. Entropy merupakan ukuran tingkat ketidakpastian pada suatu kelas, yang dihitung berdasarkan peluang terjadinya sebuah peristiwa atau kemunculan suatu fitur tertentu.

$$Entropy(S) = \sum_i^n = p_i \log_2 p_i \quad (2)$$

- b. Melakukan perhitungan *Information Gain* menggunakan Persamaan:

$$Gain(S, A) = Entropy(S) - \sum \left(\frac{|Sv|}{|S|} \cdot Entropy(Sv) \right) \quad (3)$$

Pada persamaan 2 dan 3, Gain (S,A) digunakan untuk mengukur seberapa efektif suatu fitur A dalam mengurangi ketidakpastian (entropy) pada dataset. Entropy(S) merepresentasikan tingkat ketidakpastian sebelum data dibagi berdasarkan fitur A, sedangkan Values (A) adalah kumpulan nilai yang dapat muncul pada fitur tersebut. $|Sv|$ menunjukkan jumlah data yang memiliki nilai fitur v, sementara $|S|$ adalah total keseluruhan data. Entropy(Sv) merupakan nilai entropy dari subset data yang terbentuk sesuai salah satu nilai fitur A [12]. Selanjutnya, karena jumlah data yang tidak seimbang pada variabel target dimana "Yes" lebih

2.4 K-Fold Cross Validation

K-fold cross validation adalah metode data mining yang mencapai akurasi optimal dengan membagi data menjadi data latih dan data uji. Salah satu bentuk validasi silang kfold adalah membagi dataset secara acak menjadi K partisi dan kemudian diuji K kali, dengan satu partisi digunakan sebagai data uji dan partisi yang tersisa digunakan sebagai data latih. [22]. Melalui pengujian berulang dengan setiap bagian data, pengembang mendapat umpan balik menyeluruh tentang kekuatan dan kelemahan model, sehingga mendukung pengambilan keputusan dalam fase penyempurnaan model berikutnya.[23].

2.5 Metode Backpropagation

Salah satu algoritma dari jaringan syaraf tiruan (neural network) adalah Backpropagation [24]. Algoritma pembelajaran yang dikenal sebagai *Backpropagation* ini menggunakan sistem pembelajaran terawasi (Supervised Learning), ini biasanya diterapkan pada perceptron dengan struktur berlapis-lapis untuk mengubah bobot-bobot yang terhubung dengan neuron pada lapisan tersembunyi (*Hidden Layer*) [25]. Ada tiga lapisan di dalamnya yaitu lapisan input, lapisan tersembunyi, dan lapisan output. Input Layer menerima data dari luar, data input yang dimasukkan merupakan gambaran dari suatu masalah. Hidden Layer memproses data masukan, Output Layer menghasilkan keluaran dari data data masukan yang telah diproses [26]. Tahapan perhitungan *Backpropagation Neural Network* adalah sebagai berikut:

- a. Inisialisasi parameter jaringan saraf (neural network), dilakukan dengan menetapkan jumlah neuron pada lapisan input, lapisan tersembunyi, lapisan output, dan menentukan nilai laju pembelajaran. Jumlah neuron pada lapisan tersembunyi dapat dihitung dengan menggunakan persamaan 4.

$$l < m < 2l \quad (4)$$

- b. Inisialisasi bobot dan bias dengan nilai acak (ambil nilai random yang cukup kecil)
 c. Setiap unit input (X_i , $i = 1, 2, 3, \dots, n$) menerima sinyal x_i dan meneruskan sinyal tersebut ke semua unit pada lapisan yang ada di atasnya (Hidden Layer).
 d. Setiap unit yang tersembunyi (Z_j , $j = 1, 2, 3, \dots, p$) menjumlahkan sinyal bobot sinyal input pada persamaan 5 dan akan diterapkan fungsi aktivasi untuk menghitung sinyal outputnya dengan menggunakan persamaan 6.



$$z_{inj} = voj + \sum_{i=1}^n xivi_j = xivi_j \quad (5)$$

Kemudian diproses menggunakan fungsi aktivasi ReLU untuk menghasilkan output neuron tersembunyi :

$$z_j = f(z_{inj}) = \max(0, z_{inj}) \quad (6)$$

- e. Hitung tiap hidden unit (y_k , $k = 1, 2, 3, \dots, m$) yang akan menjumlahkan bobot sinyal input dengan menggunakan persamaan 7.

$$y_{in_k} = w_{ok} + \sum_{i=1}^p z_i w_{jk} \quad (7)$$

- f. Selanjutnya akan diterapkan fungsi aktivasi Sigmoid untuk menghitung sinyal output nya dengan menggunakan persamaan 8.

$$y_k = f(y_{in_k}) = \frac{1}{1+e^{-y_{in_k}}} \quad (8)$$

- g. Setiap unit Output (y_k , $k = 1, 2, 3, \dots, m$) menerima target pola yang berhubungan dengan pola Input pembelajarannya, hitung informasi eror-nya dengan persamaan 9.

$$\delta_k = (t_k - y_k) f'(y_k) \quad (9)$$

- h. Pada setiap unit tersembunyi (Z_j , $j = 1, 2, 3, \dots, p$) akan menjumlahkan delta inputnya (dari unit-unit yang berada pada lapisan kanannya) dengan persamaan 10.

$$\delta_{inj} = \sum_{k=1}^m \delta_k w_{jk} \quad (10)$$

- i. Untuk setiap unit output (y_k , $k = 1, 2, 3, \dots, m$) yang akan dilakukan perubahan bobot dan bias ($j = 0, 1, 2, \dots, p$) dengan persamaan 11, dan setiap unit tersembunyi (Z_j , $j = 1, 2, 3, \dots, p$) akan dilakukan perubahan bobot dan bias ($i = 0, 1, 2, \dots, n$) dengan persamaan 12.

$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk} \quad (11)$$

$$V_{ij}(\text{baru}) = V_{ij}(\text{lama}) + \Delta V_{ij} \quad (12)$$

- j. Uji apakah kondisi berhenti telah terpenuhi.

2.6 Implementasi dan Evaluasi

Implementasi merupakan langkah krusial dalam penelitian ini, karena pada fase inilah desain sistem diimplementasikan secara teknis dan struktural sesuai dengan dokumentasi pengembangan. Dalam penelitian ini, proses implementasi dilakukan untuk menerapkan algoritma *Backpropagation* jaringan saraf tiruan yang dikombinasikan dengan metode pemilihan fitur *Information Gain*. Seluruh proses diimplementasikan menggunakan bahasa pemrograman *Python* melalui platform *Google Colab*. Platform ini dipilih untuk memastikan pelatihan dan pengujian model yang efisien, cepat, dan optimal.

Langkah selanjutnya adalah fase evaluasi, di mana kinerja model yang dikembangkan diukur. Salah satu metode evaluasi yang paling umum untuk menilai kinerja algoritma klasifikasi adalah *Confusion Matrix*. Berbagai proses evaluasi yang dilakukan menggunakan *Confusion Matrix* dapat dijelaskan sebagai berikut:

- a. Akurasi

Akurasi menunjukkan seberapa tepat model dalam membuat prediksi yang benar, baik untuk kasus positif ataupun negative. Nilai akurasi dapat diperoleh dari persamaan 13.

$$Accuracy = \frac{TN+TP}{TP+TN+FP+FN} \quad (13)$$

- b. Precision

Precision merupakan metrik yang digunakan untuk menilai seberapa besar bagian dari hasil prediksi positif yang benar-benar sesuai dengan kelas positif sebenarnya. Nilai dari presisi dapat diperoleh dari persamaan 14.

$$Precision = \frac{TP}{TP+FP} \quad (14)$$

- c. Recall

Recall adalah metrik yang digunakan untuk mengukur seberapa besar jumlah data positif yang berhasil terdeteksi dengan benar oleh model dari seluruh data yang memang termasuk dalam kelas positif. Berikut persamaan yang dapat dilakukan untuk memperoleh nilai recall:

$$Recall = \frac{TP}{TP+FN} \quad (15)$$

- d. F1 Score



F1 Score merupakan metrik yang digunakan untuk menghitung rata-rata harmonis dari nilai precision dan recall, sehingga memberikan keseimbangan antara keduanya dalam mengevaluasi kinerja model. Untuk memperoleh nilai F1 Score menggunakan persamaan 16.

$$F1\ Score = \frac{2(Precision \times Recall)}{(Precision + Recall)} \tag{16}$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Penelitian ini menggunakan dataset yang terdiri dari 1.470 data, yang terdiri dari dua kategori kelas, yaitu karyawan yang mengalami pengurangan (Attrition = Yes) dan karyawan yang tidak mengalami pengurangan (Attrition = No). Setiap data memiliki 35 fitur yang terdiri dari beragam variabel serta faktor-faktor perusahaan yang digunakan sebagai dasar untuk proses klasifikasi.

3.2 Preprocessing Data

- Cleaning Data** : Pada tahap ini memastikan data tidak memiliki duplikat atau hilang (*missing value*). Berdasarkan hasil pemeriksaan, tidak ditemukan adanya data yang duplikat atau data yang memiliki nilai kosong. Maka dari itu, tahap Cleaning Data dapat diselesaikan tanpa memerlukan proses penghapusan atau penambahan data.
- Seleksi Data** : Pada tahap ini dilakukan beberapa penghapusan fitur karena tidak memberikan kontribusi terhadap proses klasifikasi. Fitur-fitur tersebut yaitu, EmployeeCount, Over18, StandardHours, dan EmployeeNumber yang merupakan fitur dengan kolom yang berisi nilai yang sama sehingga tidak relevan. Setelah penghapusan fitur-fitur, jumlah variabel berkurang dari total awal 35 menjadi 31 fitur. Hasil dari seleksi data ditampilkan pada Gambar 2.

Tidak ada kolom yang relevan untuk dihapus.
Jumlah baris dan kolom setelah penghapusan: (1470, 31)

No	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement	JobLevel	JobRole	JobSatisfaction	MaritalStatus
0	41	Yes	Travel_Rarely	1992	Sales	1	2	Life Sciences	2	Female	94	3	2	Sales Executive	4	Mar
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	3	Male	61	2	2	Research Scientist	2	Mar
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	4	Male	92	2	1	Laboratory Technician	3	Mar
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	4	Female	56	3	1	Research Scientist	3	Mar
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	Male	40	3	1	Laboratory Technician	2	Mar

Jumlah fitur setelah penghapusan: 31

Gambar 2. Hasil Seleksi Data

- Transformasi Data** : Tahap ini memastikan bahwa seluruh variabel pada dataset telah berada dalam format numerik dan siap masuk ke proses seleksi fitur dan permodelan. Pada penelitian ini dilakukan transformasi dengan menggunakan dua jenis transformasi, yaitu *Label Encoding* dan *One Hot Encoding*. *Label Encoding* digunakan pada variabel target Attrition untuk mengubah nilai kategorikal menjadi bentuk numerik. Melalui proses ini, nilai “Yes” dikodekan menjadi 1 dan “No” menjadi 0 sehingga target dapat diolah oleh algoritma klasifikasi. Sedangkan *One Hot Encoding* digunakan untuk variabel fitur yang bersifat kategorikal seperti *BusinessTravel*, *Department*, *EducationField*, *Gender*, *JobRole*, *Marital Status*, dan *Overtime*. Proses ini dilakukan untuk mengubah nilai kategori pada variabel fitur menjadi bentuk numerik biner. Setelah melalui tahap One-Hot Encoding, jumlah fitur mengalami peningkatan dari 31 fitur menjadi 45 fitur, karena setiap kategori pada variabel tersebut dipisah sebagai fitur baru. Hasil dari transformasi data yang telah dilakukan ditampilkan pada Tabel 2 dan 3.

Tabel 2. Hasil Label Encoding

No	Age	Attrition	Education	MonthlyIncome	MonthlyRate	NumCompaniesWorked	PercentSalaryHike
1	41	1	2	5993	19479	8	11
2	49	0	1	5130	24907	1	23
3	37	1	2	2090	2396	6	15
.....
1469	33	0	4	2909	23159	1	11
1470	27	0	1	3468	16632	9	12

Tabel 3. Hasil One-Hot Encoding

No	BusinessTravel_Travel_Frequently	BusinessTravel_Travel_Rarely	Department_Research & Development	OverTime_Yes
1	0	1	0	1
2	1	0	1	0
3	0	1	1	1
.....
1469	1	0	1	1
1470	0	1	1	0



d. Normalisasi Data : Tahap ini dilakukan karena dataset yang tersedia memiliki nilai yang berbeda-beda. Maka diperlukan penyesuaian agar seluruh atribut berada dalam skala yang seragam, yakni berada dalam rentang 0 hingga 1 dengan menggunakan metode Min-Max Normalization pada persamaan 1. Hasil normalisasi data ditampilkan pada Tabel 4.

Tabel 4. Hasil Normalisasi Data

No	Age	DailyRate	Education	MaritalStatus_Single	Gender_Male	OverTime_Yes	JobRole_Manager
1	0.5476	0.7158	0.25	1.0	0.0	1.0	0.0
2	0.7380	0.1267	0.00	0.0	0.1	0.0	0.0
3	0.4523	0.9098	0.25	1.0	0.1	1.0	0.0
....
1469	0.3571	0.9234	0.75	0.0	0.0	1.0	1.0
1470	0.2142	0.3500	0.00	0.0	0.1	0.0	0.0

Hasil pada Tabel 4 menunjukkan bahwa seluruh nilai fitur sudah berada di dalam rentang nilai 0 hingga 1, yang menandakan bahwa proses normalisasi sudah berhasil dilakukan.

3.2 Seleksi Fitur Information Gain

Seleksi fitur information gain diawali dengan menghitung nilai entropy pada fitur *Attrition* dengan menggunakan Persamaan 2. Nilai hasil perhitungan tersebut disajikan pada Gambar 3.

Entropy Attrition: 0.6372308920655037

Gambar 3 Hasil Perhitungan Entropy

Setelah nilai entropy diperoleh, tahap berikutnya adalah menghitung *Information Gain* dengan menggunakan persamaan 3. Proses ini menghasilkan skor information gain untuk setiap fitur yang dianalisis. Nilai perhitungan tersebut ditampilkan pada Tabel 5.

Tabel 5 .Hasil Information Gain

Rangking	Feature	Score	Rangking	Feature	Score
1	JobLevel	0.034396	14	YearsAtCompany	0.013870
2	OverTime_Yes	0.034273	15	JobSatisfaction	0.012349
3	MaritalStatus_Single	0.033621	16	PerformanceRating	0.011490
4	MonthlyIncome	0.032008	17	StockOptionLevel	0.010037
5	YearsInCurrentRole	0.031002	18	EducationField_Technical Degree	0.008807
6	TotalWorkingYears	0.024928	19	DistanceFromHome	0.007305
7	BusinessTravel_Travel_Frequently	0.024651	20	JobRole_Manufacturing Director	0.006959
8	Age	0.020836	21	EducationField_Other	0.006124
9	YearsWithCurrManager	0.019865	22	JobInvolvement	0.005476
10	Education	0.018454	23	EnvironmentSatisfaction	0.005343
11	MaritalStatus_Married	0.017344	24	HourlyRate	0.004728
12	NumCompaniesWorked	0.017258	25	PercentSalary	0.002919
13	TrainingTimesLastYear	0.014880			

Metode seleksi fitur *Information Gain* digunakan dengan mengurutkan fitur berdasarkan nilai tertinggi. Kemudian dikombinasikan dengan metode klasifikasi *Backpropagation Neural Network*. Hasil peringkat fitur berdasarkan nilai *Information Gain* ditampilkan pada Tabel 5, dimana fitur-fitur telah disusun mulai dari skor paling tinggi. Pengujian dilakukan dengan memilih beberapa jumlah fitur yaitu, 10, 15, 20, dan 25.

3.3 Pembagian Data

Proses pembagian data dilakukan menggunakan teknik 5-fold Cross-Validation, dan 10-fold Cross Validation. Pada konfigurasi K=5, sebanyak 1.176 data yang digunakan sebagai data pelatihan untuk membangun model, sedangkan 294 data dialokasikan sebagai data pengujian untuk menilai performanya. Sementara itu, pada konfigurasi K=10, terdapat 1.323 data yang digunakan sebagai data pelatihan dan 147 sebagai data pengujian. Selanjutnya dilakukan pendefinisian Feature Vector dan Target Variable berdasarkan fitur terbaik yang dipilih sesuai dengan nilai ambang (*threshold*) yang telah ditetapkan. Hasil pendefinisian serta pembagian data ditampilkan pada Tabel 6 dan 7.

Tabel 6. Hasil Pembagian data sesuai K=5

	10	15	20	25
x_train	(1.176, 10)	(1.176, 15)	(1.176, 20)	(1.176, 25)
x_test	(294, 10)	(294, 15)	(294, 20)	(294, 25)

**Tabel 7.** Hasil Pembagian data sesuai K=10

	10	15	20	25
x_train	(1.323, 10)	(1.323, 15)	(1.323, 20)	(1.323, 25)
x_test	(147, 10)	(147, 15)	(147, 20)	(147, 25)

3.4 Backpropagation

Backpropagation dikenal sebagai metode yang sangat efektif dalam proses pelatihan jaringan saraf berlapis banyak atau biasa dikenal *Multi-Layer Neural Network* dan menjadi fondasi penting bagi berbagai perkembangan teknologi kecerdasan buatan. Arsitektur jaringan *Backpropagation* pada penelitian ini dirancang berdasarkan hasil seleksi fitur menggunakan metode *Information Gain* pada masing-masing jumlah atribut dengan nilai tertinggi, yaitu 10, 15, 20, dan 25. Jumlah neuron pada lapisan input disesuaikan dengan jumlah fitur yang lolos seleksi pada setiap *threshold*. *Hidden layer* menggunakan sejumlah neuron yang ditetapkan berdasarkan konfigurasi model pada penelitian ini, sedangkan lapisan output terdiri dari 1 neuron untuk menghasilkan prediksi kelas *Attrition*.

Parameter pelatihan Backpropagation yang digunakan mencakup fungsi aktivasi ReLU pada hidden layer dan fungsi aktivasi sigmoid pada output layer, nilai *learning rate* (0.1, 0.01, 0.001), serta maksimum epoch sebanyak 100 iterasi. Prosedur pelatihan dilakukan sebanyak tiga kali untuk setiap kombinasi atribut dan *learning rate*. Setiap proses pelatihan menghasilkan nilai akurasi yang berbeda, kemudiam dihitung nilai rata-rata dari ketiga hasil tersebut untuk memperoleh performa akhir dari setiap skenario seleksi fitur *Information Gain*.

3.5 Evaluasi dan Pengujian

Pengujian ini dilakukan dengan dua skenario utama, yaitu pelatihan model *Backpropagation Neural Network* tanpa seleksi fitur dan pelatihan model dengan seleksi fitur menggunakan *Information Gain* berdasarkan nilai atribut tertentu. Kedua skenario ini dibandingkan untuk mengetahui sejauh mana proses seleksi fitur dapat meningkatkan performa model dalam mengklasifikasikan data atrisi karyawan. Pada skenario pertama, performa model tanpa Seleksi Fitur *Information Gain*, model BPNN dilatih dengan menggunakan seluruh fitur pada dataset yang menghasilkan akurasi sebesar 83%, meskipun akurasi terlihat cukup tinggi tetapi model cenderung bias terhadap kelas mayoritas karena jumlah fitur yang besar menyebabkan noise dan menghambat kemampuan jaringan dalam mempelajari pola keluarnya karyawan. Setelah di terapkan seleksi fitur, berdasarkan peringkat *Information Gain* tertinggi sebesar 10, 15, 20 dan 25. Setiap model diuji menggunakan tiga *learning rate*, yaitu 0.1, 0.01, 0.001. Pada skenario ini, arsitektur *Backpropagation Neural Network* yang digunakan menyesuaikan jumlah jumlah fitur yang lolos seleksi sebagai neuron pada lapisan input, sedangkan jumlah neuron pada hidden layer dan output layer mengikuti konfigurasi model yang telah ditetapkan dalam penelitian ini. Setelah itu, pengujian nilai *learning rate* dilakukan pada arsitektur jaringan tersebut untuk mengevaluasi pengaruh parameter pelatihan terhadap performa model. Evaluasi performa dilakukan menggunakan beberapa metrik, yaitu akurasi, presisi, recall, dan F1-score, yang masing-masing dihitung menggunakan Perasamaan 13 sampai 16. Hasil pengujian seleksi fitur *Information Gain* menggunakan K=5 disajikan pada Tabel 8 dan K=10 disajikan pada Tabel 9.

Tabel 8. Hasil Pengujian Seleksi Fitur *Information Gin* K=5

No	Jumlah Atribut IG	Learning Rate	Neuron Hidden Layer	Arsitektur BPNN 1 Hidden Layer	Accuracy Rata-Rata
1	10	0,1	11	10:11:1	84,15
2		0,01			84,90
3		0,001			85,17
4	10	0,1	15	10:15:1	84,15
5		0,01			85,10
6		0,001			84,90
7	10	0,1	19	10:19:1	84,22
8		0,01			84,97
9		0,001			85,10
10	15	0,1	16	15:16:1	85,63
11		0,01			84,22
12		0,001			85,31
13	15	0,1	25	15:25:1	84,29
14		0,01			84,56
15		0,001			85,99
16	15	0,1	29	15:29:1	83,67
17		0,01			84,42
18		0,001			84,97
19	20	0,1	21	20:21:1	84,56
20		0,01			84,35
21		0,001			85,92
22	20	0,1	30	20:35:1	85,17
23		0,01			82,72
24		0,001			85,44
25	20	0,1	39	20:39:1	85,10



No	Jumlah Atribut IG	Learning Rate	Neuron Hidden Layer	Arsitektur BPNN 1 Hidden Layer	Accuracy Rata-Rata
26		0,01			84,15
27		0,001			86,05
28	25	0,1	26	25:26:1	85,65
29		0,01			84,01
30		0,001			86,87
31	25	0,1	35	25:35:1	86,46
32		0,01			82,31
33		0,001			86,80
34	25	0,1	49	25:49:1	87,07
35		0,01			84,97
36		0,001			86,46

Tabel 8 menunjukkan bahwa akurasi tertinggi diperoleh dengan menggunakan seleksi fitur *Information Gain* sebesar 87.07% pada jumlah atribut IG 25, dengan konfigurasi *learning rate* 0.1, *neuron hidden* 49, dan arsitektur *Backpropagation* [25:49:1].

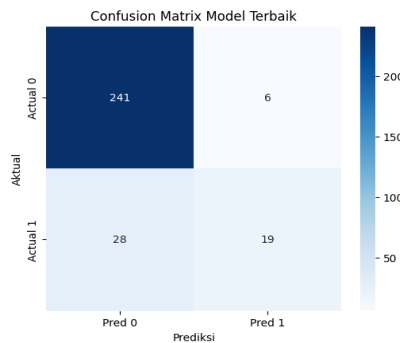
Tabel 9 Hasil Pengujian seleksi fitur *Information Gain* K=10

No	Jumlah Atribut IG	Learning Rate	Neuron Hidden Layer	Arsitektur BPNN 1 Hidden Layer	Accuracy Rata-Rata
1	10	0,1	11	10:11:1	83,95
2		0,01			85,31
3		0,001			85,10
4	10	0,1	15	10:15:1	84,90
5		0,01			84,76
6		0,001			85,24
7	10	0,1	19	10:19:1	84,97
8		0,01			85,37
9		0,001			84,76
10	15	0,1	16	15:16:1	83,67
11		0,01			85,44
12		0,001			85,24
13	15	0,1	25	15:25:1	85,78
14		0,01			84,56
15		0,001			85,24
16	15	0,1	29	15:29:1	85,10
17		0,01			84,56
18		0,001			85,31
19	20	0,1	21	20:21:1	85,24
20		0,01			84,49
21		0,001			86,12
22	20	0,1	30	20:35:1	85,51
23		0,01			84,08
24		0,001			86,12
25	20	0,1	39	20:39:1	86,39
26		0,01			84,29
27		0,001			85,78
28	25	0,1	26	25:26:1	85,65
29		0,01			85,03
30		0,001			86,94
31	25	0,1	35	25:35:1	86,73
32		0,01			84,63
33		0,001			87,28
34	25	0,1	49	25:49:1	86,39
35		0,01			84,15
36		0,001			87,07

Tabel 9 menunjukkan bahwa akurasi tertinggi diperoleh sebesar 87.28% pada jumlah atribut IG 25 dengan konfigurasi *learning rate* 0.001, *neuron hidden* 35 dan arsitektur *Backpropagation* [25:35:1]. Variasi akurasi yang muncul pada setiap skenario pengujian menunjukkan bahwa performa model sangat dipengaruhi oleh kombinasi antara arsitektur jaringan dan nilai *learning rate* yang digunakan. Pemilihan kombinasi yang tepat ini memberikan dampak langsung terhadap kualitas proses pembelajaran model dalam mengklasifikasikan data *Attrition*. Hasil ini menegaskan bahwa proses tuning parameter seperti *threshold Information Gain*, jumlah neuron pada *hidden layer*, dan nilai *learning rate* merupakan faktor penting yang menentukan tingkat akurasi *Backpropagation Neural Network*. Analisis lanjutan dilakukan evaluasi peforma menggunakan *Confussion Matrix* untuk menganalisis tingkat ketepatan model dalam memprediksi kelas *Attrition*. Berdasarkan hasil *Confussion Matrix*, pada kelas “No” *Attrition* (0), model berhasil mengidentifikasi secara benar sebanyak 241 data sebagai *True Negative*. Sementara itu, pada kelas “Yes” *Attrition* (1), model berhasilkan memprediksi secara tepat sebanyak 19 data sebagai *True Positive*. Namun demikian, masih terdapat



sejumlah kesalahan prediksi yaitu ada 6 data kelas “No” Attrition (0) yang salah diklasifikasikan sebagai Attrition (False Positive) serta 28 data kelas “Yes” Attrition yang keliru di prediksi sebagai “No” Attrition (False Negative). Dari hasil tersebut kita bisa mengetahui jumlah Precision 0.76, Recall 0.40, dan F1-Score 0.53 dengan menggunakan rumus pada persamaan 14, 15, dan 16. Hasil *Confusion Matrix* dilihat pada Gambar 4:



Gambar 4 Hasil Confusion Matrix dengan akurasi tertinggi pada pengujian Backpropagation

Karena dataset yang digunakan tidak seimbang, maka dilakukan pengujian lain dengan menggunakan metode imbalance data yaitu SMOTE. Pengujian dilakukan dengan menggunakan beberapa nilai tertinggi *Information Gain* dan arsitektur *Backpropagation* yang sama, hasil menunjukkan akurasi sebesar 89.79%.

4. KESIMPULAN

Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi pergantian karyawan dengan menerapkan *Information Gain* sebagai kriteria seleksi dan algoritma jaringan saraf Backpropagation sebagai model klasifikasi. Dataset yang digunakan merupakan *IBM HR Employee Attrition* dari Kaggle yang berjumlah 1.470 data dengan 35 fitur. Pada tahap pra-pemrosesan, dilakukan penghapusan atribut dari 35 menjadi 31, selanjutnya juga dilakukan penambahan atribut kategorikal yang dipisah menjadi atribut baru, hasil dari penambahan tersebut dari atribut 31 menjadi 45. Setelah melalui tahap pra-pemrosesan, dilakukan seleksi fitur menggunakan *Information Gain* dengan mengurutkan beberapa nilai tertinggi hasil *Information Gain* yaitu 10, 15, 20, dan 25, yang masing-masing menghasilkan jumlah fitur terpilih berbeda sesuai tingkat informasi yang diberikan terhadap variabel Attrition. Pengujian ini dilakukan menggunakan metode K-Fold Cross Validation dengan nilai K=5 dan K=10. Arsitektur jaringan *Backpropagation* menggunakan satu hidden layer dengan beberapa konfigurasi jumlah neuron, serta diuji menggunakan tiga nilai *learning rate* (0.1, 0.01, 0.001). Hasil pengujian menunjukkan akurasi rata-rata tertinggi 87.28% dengan konfigurasi terbaik diperoleh dari 25 atribut tertinggi *Information Gain*, menggunakan neuron hidden 35, *learning rate* 0.001, dan K=10. Hasil ini membuktikan bahwa penerapan *Information Gain* mampu meningkatkan performa klasifikasi dengan mengurangi fitur yang tidak relevan sehingga model belajar lebih fokus pada variabel yang berkontribusi terhadap Attrition. Performa tersebut menunjukkan peningkatan signifikan dibandingkan model tanpa seleksi fitur yang hanya mencapai akurasi 84%. Evaluasi tambahan dengan menggunakan *Confusion Matrix* menunjukkan model memiliki nilai Precision 0.76, Recall 0.40, dan F1-Score 0.53 yang menandakan bahwa model memiliki tingkat akurasi yang baik tetapi kemampuan dalam mengidentifikasi kelas minoritas Attrition masih perlu ditingkatkan. Maka dilakukan pengujian lain dengan menggunakan metode SMOTE sebagai penyeimbang data dengan arsitektur yang sama, hasil menunjukkan akurasi sebesar 89.79%. Variasi hasil antar skenario menunjukkan bahwa performa model sangat dipengaruhi oleh kombinasi beberapa jumlah atribut tertinggi *Information Gain*, arsitektur jaringan, dan nilai *learning rate*. Penelitian ini memiliki keterbatasan karena hanya menggunakan metode seleksi fitur berbasis hubungan informasi dan ketidakseimbangan data pada dataset yang digunakan. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan metode penyeimbang data lain seperti, ADASYN, atau *Random Sampling* untuk meningkatkan kemampuan dalam mengenali kelas minoritas. Penelitian lanjutan juga disarankan menguji beberapa metode seleksi fitur yang lain seperti *Recursive Feature Elimination* (RFE), Mutual Information atau ReliefF agar dapat membandingkan efektivitas dalam peningkatan performa model. Menguji dengan menggunakan algoritma klasifikasi yang lebih kompleks seperti Random Forest, XGBost atau Deep Neural Network yang memberikan peningkatan akurasi terutama pada dataset dengan ketidakseimbangan kelas juga disarankan. Pengujian juga dapat diterapkan pada kasus perusahaan berbeda dengan data yang seimbang pada masing-masing kelas.

REFERENCES

- [1] N. M. Suindari and N. M. R. Juniariani, “PENGELOLAAN KEUANGAN, KOMPETENSI SUMBER DAYA MANUSIA DAN STRATEGI PEMASARAN DALAM MENGUKUR KINERJA USAHA MIKRO KECIL MENENGAH (UMKM),” *KRISNA: Kumpulan Riset Akuntansi*, vol. 11, no. 2, pp. 148–154, Jan. 2020, doi: 10.22225/kr.11.2.1423.148-154.
- [2] R. Selviasari, “Analisis Faktor-Faktor Penentu Turnover Intention pada Karyawan Generasi Z: Pendekatan Human-Centered Management di Era Kerja Fleksibel,” *Al-Muraqabah: Journal of Management and Sharia Business*, vol. 05, no. 01, pp. 111–120, 2025, doi: 10.30762/al-muraqabah.v5i1.2435.



- [3] S. Barara and U. Soni, "Employee Attrition Prediction using Machine Learning," in *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2023, pp. 1–9. doi: 10.1109/ICECCME57830.2023.10252571.
- [4] L. A. Sutisna, "USING FEATURE ENGINEERING IN LOGISTIC REGRESSION AND RANDOM FOREST METHODS TO IMPROVE EMPLOYEE ATTRITION PREDICTION IN KIMIA FARMA under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)," *Jurnal Ekonomi*, vol. 12, no. 02, p. 2023, 2023, doi: <https://doi.org/10.54209/ekonomi.v12i02>.
- [5] H. Alqahtani, H. Almagrabi, and A. Alharbi, "EMPLOYEE ATTRITION PREDICTION USING MACHINE LEARNING MODELS: A REVIEW PAPER," *International Journal of Artificial Intelligence and Applications (IJAI)*, vol. 15, no. 2, 2024, doi: 10.5121/ijai.2024.1520223.
- [6] A. Benabou, F. Touhami, and M. A. Sabri, "Predicting Employee Turnover Using Machine Learning Techniques," *Acta Informatica Pragensia*, vol. 14, no. 1, pp. 112–127, 2025, doi: 10.18267/j.aip.255.
- [7] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fared, "Predicting Employee Attrition Using Machine Learning Approaches," *Applied Sciences (Switzerland)*, vol. 12, no. 13, Jul. 2022, doi: 10.3390/app12136424.
- [8] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, "Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 172–180, Jan. 2024, doi: 10.57152/malcom.v4i1.1074.
- [9] S. F. Sari and K. M. Lhaksana, "Employee Attrition Prediction Using Feature Selection with Information Gain and Random Forest Classification," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 410–419, Sep. 2022, doi: 10.47065/josyc.v3i4.2099.
- [10] Mardiana, Jasmir, and Sharipuddin, "Peningkatan Performa Naïve Bayes dengan Information Gain Menggunakan Machine Learning untuk Klasifikasi Kanker Payudara," *Jurnal Manajemen Teknologi dan Sistem Informasi (JMS)*, vol. 5, no. 2, 2025, doi: 10.33998/jms.v5i2.
- [11] M. Itqon and J. D. T. Purnomo, "Employee Attrition Prediction using Machine Learning in Rolling Stock Manufacturing Company," *Jurnal Teknobisnis*, vol. 8, no. 1, pp. 74–85, Jul. 2024, doi: 10.12962/j24609463.v8i1.941.
- [12] A. Wahyudi *et al.*, "KLASIFIKASI PENYAKIT JANTUNG MENGGUNAKAN PENDEKATAN HYBRID IINFORMATION GAIN dan BACKPROPAGATION NEURAL NETWORK (BPNN) Heart Disease Classification Using a Hybrid Approach of Information Gain and Backpropagation Neural Network (BPNN)," *Insisiva Dental Journal : Majalah Kedokteran Gigi Insisiva*, doi: <https://doi.org/10.59737/jpi.v17i1.347>.
- [13] Y. R. Fauzan, Y. I. Fajarendra, M. Noor, T. Ridha, and S. ' Uyun, "Klasifikasi Persediaan Stok Darah Menggunakan Algoritma K-NN, Decision Tree, dan JST Backpropagation," *jurnal JUPITER*, vol. 16, no. 2, pp. 623–634, 2024, doi: <https://doi.org/10.5281/zenodo.13755935>.
- [14] N. Tsawaabul Khair, I. Afrianty, F. Syafria, E. Budianita, and S. Kurnia Gusti, "Penerapan Information Gain Untuk Seleksi Fitur Pada Klasifikasi Jenis Kelamin Tulang Tengkorak Menggunakan Backpropagation," *Media Online*, vol. 5, no. 4, pp. 666–678, 2025, doi: 10.47065/bulletincsr.v5i4.637.
- [15] Pavansubhash, "IBM HR analytics employee attrition & performance," kaggle. Accessed: Dec. 21, 2025. [Online]. Available: <https://www.kaggle.com/datasets/pavansubhash/ibm-hr-analytics-attrition-dataset>
- [16] I. I. Indra, U. Rizki, P. M. Jakak, M. B. Prayogi, and M. Rahman, "Penerapan Metode K-Means Clustering Dalam Pengembangan Strategi Promosi Berbasis Data Penerimaan Mahasiswa Baru (Studi Kasus :Universitas Nurul Huda)," *Jurnal Nasional Ilmu Komputer*, vol. 5, no. 1, pp. 25–43, 2024, doi: 10.47747/jurnalnik.v5i1.1656.
- [17] D. S. Soper, "Greed is good: Rapid hyperparameter optimization and model selection using greedy k-fold cross validation," *Electronics (Switzerland)*, vol. 10, no. 16, Aug. 2021, doi: 10.3390/electronics10161973.
- [18] F. H. Wardhani and K. M. Lhaksana, "Predicting Employee Attrition Using Logistic Regression With Feature Selection," *Sinkron*, vol. 7, no. 4, pp. 2214–2222, Oct. 2022, doi: 10.33395/sinkron.v7i4.11783.
- [19] R. G. Whendasoro and J. Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 872, Aug. 2022, doi: 10.30865/jurikom.v9i4.4526.
- [20] R. A. Azizah, F. Bachtiar, and S. Adinugroho, "Klasifikasi Kinerja Akademik Siswa Menggunakan Neighbor Weighted K-Nearest Neighbor dengan Seleksi Fitur Information Gain," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 3, pp. 605–614, 2022, doi: 10.25126/jtiik.2022935751.
- [21] I. K. Hasan, R. Resmawan, and J. Ibrahim, "Perbandingan K-Nearest Neighbor dan Random Forest dengan Seleksi Fitur Information Gain untuk Klasifikasi Lama Studi Mahasiswa," *Indonesian Journal of Applied Statistics*, vol. 5, no. 1, p. 58, May 2022, doi: 10.13057/ijas.v5i1.58056.
- [22] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "Aplikasi Naïve Bayes Classifier (NBC) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.
- [23] Z. R. Tembusai, H. Mawengkang, and M. Zarlis, "K-Nearest Neighbor with K-Fold Cross Validation and Analytic Hierarchy Process on Data Classification," *International Journal of Advances in Data and Information Systems*, vol. 2, no. 1, Jan. 2021, doi: 10.25008/ijadis.v2i1.1204.
- [24] H. Hartati, A. H. Saputra, and I. Saluza, "Optimisasi Backpropagation Neural Network dalam Memprediksi IHSG," *Jurnal Ilmiah Informatika Global*, vol. 13, no. 1, Apr. 2022, doi: 10.36982/jiig.v13i1.2066.
- [25] H. Hambali *et al.*, "JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0) PENERAPAN ALGORITMA BACKPROPAGATION DALAM MEMPREDIKSI JUMLAH JAMAAH HAJI PEMATANG SIANTAR," *Journal Of Information Systems And Informatics Engineering*, vol. 8, no. 1, pp. 135–143, 2024, doi: 10.35145/joisie.v8i1.3882.
- [26] M. Simanjuntak, M. Muljono, G. F. Shidik, and A. Zainul Fanani, "Evaluation Of Feature Selection for Improvement Backpropagation Neural Network in Divorce Predictions," in *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2020, pp. 578–584. doi: 10.1109/iSemantic50169.2020.9234297.