Jurnal Teknik Informatika (JUTIF)
Vol. 5, No. 6, December 2024, pp. 1823-1831
DOI: https://doi.org/10.52436/1.jutif.2024.5.6.3442
p-ISSN: 2723-3863
e-ISSN: 2723-3871

IMPLEMENTATION OF MACHINE LEARNING ON EMPLOYEE ATTRITION BASED ON PERFORMANCE PARAMETERS USING PARTICLE SWARM OPTIMIZATION AND ENSEMBLE CLASSIFER METHODS

Difari Afreyna Fauziah[*1], Agung Muliawan[2], Muhaimin Dimyati[3]

[1]System and Information Technology, Faculty Of Science, Technology and Industry, Institute Technology and Science Mandala, Indonesia
[2]Software Engineering, Faculty Of Science, Technology and Industry, Institute Technology and Science Mandala, Indonesia
[3]Management, Faculty of Economics and Business, Institute Technology and Science Mandala, Indonesia
Email: [1]difariafreyna@itsm.ac.id, [2]agung.muliawan@itsm.ac.id, [3]dimyati@itsm.ac.id

(Article received: September 7, 2024; Revision: October 6, 2024; published: December 29, 2024)

Abstract

This research aims to apply machine learning to predict the start of employee attrition by considering performance parameters and other related factors in the company environment. Employee attrition refers to employee turnover in an organization for various reasons such as resignation, moving, retirement, and so on. This research uses a dataset originating from the IBM HR Analytics Employee Attrition dataset available on Kaggle (https://www.kaggle.com/) which consists of 35 attributes. Particle Swarm Optimization (PSO) method is a dimension reduction method to improve the efficiency and performance of machine learning models by reducing unnecessary data. The machine learning approaches used in the early prediction of employee attrition in this research are Support Vector Machine, Deep Learning and Neural Network methods. This research will combine the dimensionality reduction process with machine learning to obtain employee attrition prediction results that are optimized using the Ensemble method, namely Bagging and Boosting to increase the accuracy value of the prediction results. The results of this research show that applying dimensionality reduction using the PSO method can improve the accuracy of results on the IBM HR Analytics Employee Attrition dataset. The best accuracy in attrition prediction was obtained by the Deep Learning method with an accuracy value of 86.94%, a precision value of 88.90%, and a recall value of 96.40% after combining it with PSO and optimizing with Bagging.

Keywords*: Employee Attrition; Employee Retention; Employee Performance; Highest Accuracy; Machine learning*

IMPLEMENTASI *MACHINE LEARNING* PADA ATRISI KARYAWAN BERDASARKAN PARAMETER KINERJA DENGAN MENGGUNAKAN METODE *PARTICLE SWARM OPTIMIZATION* DAN *ENSEMBLE CLASSIFER*

Abstrak

Penelitian ini bertujuan untuk menerapkan *Machine Learning* dalam memprediksi awal atrisi karyawan dengan mempertimbangkan parameter kinerja dan faktor terkait lainnya yang ada di lingkungan perusahaan. Atrisi karyawan merujuk pada pergantian karyawan dalam organisasi dengan berbagai alasan seperti resignasi, pindah, pensiun dan sebagainya. Penelitian ini menggunakan dataset yang berasal dari *datasets IBM HR Analytics Employee Attrition* yang tersedia di Kaggle (*https://www.kaggle.com/)* yang terdiri dari 35 atribut. Metode *Particle Swarm Optimization* (PSO) merupakan metode pengurangan dimensi untuk meningkatkan efisiensi dan kinerja model *machine learning* dengan melakukan pengurangan data yang tidak dibutuhkan. Pendekatan *Machine Learning* yang digunakan dalam prediksi awal atrisi karyawan pada penelitian ini yaitu metode *Support Vector Machine*, *Deep Learning* dan *Neural Network*. Penelitian ini akan melakukan kombinasi dari proses reduksi dimensi dengan *Machine Learning* untuk mendapatkan hasil prediksi atrisi karyawan yang dioptimasi dengan metode *Ensemble* yaitu *Bagging* dan *Boosting* untuk meningkatkan nilai akurasi dari hasil prediksi. Hasil penelitian ini menunjukkan bahwa penerapan reduksi dimensi menggunakan metode PSO dapat meningkatkan hasil akurasi pada *dataset IBM HR Analytics Employee Attrition*. Akurasi terbaik dalam prediksi atrisi ini diperoleh oleh metode Deep Learning dengan nilai akurasi sebesar 86,94%, nilai presisi sebesar 88,90% dan

nilai *recall* sebesar 96,40% setelah dikombinasikan dengan *Particle Swarm Optimization* (PSO) dan dioptimasi dengan *Bagging*.

Kata kunci: Atrisi Karyawan; Akurasi Tertinggi; Kinerja Karyawan; *Machine learning*; Retensi Karyawan;

1. PENDAHULUAN

Karyawan merupakan aset terpenting dalam sebuah perusahaan karena mereka adalah penggerak utama yang menentukan keberhasilan dan kelangsungan operasional. Karyawan membawa keahlian, pengetahuan, dan keterampilan yang dibutuhkan untuk menjalankan berbagai fungsi bisnis, mulai dari produksi hingga layanan pelanggan. Selain itu, karyawan yang berkomitmen dan termotivasi dapat meningkatkan produktivitas dan kualitas kerja, sehingga berkontribusi langsung terhadap pencapaian tujuan perusahaan. Kepuasan dan kesejahteraan karyawan juga memainkan peran penting dalam menciptakan lingkungan kerja yang positif dan kondusif, yang pada gilirannya dapat meningkatkan retensi karyawan dan mengurangi tingkat atrisi. Oleh karena itu, memahami dan memenuhi kebutuhan serta aspirasi karyawan menjadi hal yang sangat penting bagi manajemen perusahaan dalam upaya membangun tim yang solid dan mendukung pertumbuhan bisnis yang berkelanjutan.

Atrisi karyawan, atau pergantian karyawan, merupakan salah satu tantangan terbesar yang dihadapi oleh perusahaan dalam mengelola sumber daya manusia. Atrisi karyawan merujuk pada pergantian karyawan dalam organisasi dengan berbagai alasan seperti resignasi, pindah, pensiun dan sebagainya. Atrisi karyawan yang tinggi dapat menyebabkan gangguan operasional, peningkatan biaya perekrutan dan pelatihan, serta hilangnya pengetahuan dan pengalaman berharga [1]. pergantian karyawan yang tinggi dapat mencerminkan masalah mendasar dalam lingkungan kerja, seperti kurangnya kepuasan kerja, kompensasi yang tidak memadai, peluang pengembangan karier yang terbatas, atau budaya kerja yang tidak sehat [2], [3]. Semua faktor ini menekankan pentingnya memahami dan mengelola atrisi karyawan secara efektif untuk memastikan keberlanjutan dan keberhasilan jangka panjang perusahaan.

Perusahaan harus mulai mempertimbangkan pendekatan yang lebih inovatif dan terarah menggunakan teknologi. Salah satu teknologi yang dapat dimanfaatkan yaitu *Machine learning*. *Machine learning* adalah cabang dari kecerdasan buatan (artificial intelligence) yang berkaitan dengan pengembangan sistem komputer yang mampu belajar dari data yang diberikan atau pengalaman yang diperoleh [4]–[6]. Dengan memanfaatkan *Machine learning*, perusahaan dapat menggali informasi penting dari data internal mereka untuk memahami pola-pola atrisi karyawan dan mengidentifikasi faktor-faktor yang mempengaruhinya [7], [8]. Dalam konteks manajemen sumber daya manusia, *machine learning* dapat digunakan untuk memahami faktor-faktor yang mempengaruhi keputusan karyawan untuk pergi dari perusahaan (atrisi karyawan) berdasarkan parameter kinerja dan faktor-faktor terkait lainnya.

Beberapa *literature* telah melakukan penelitian mengenai pendeteksi awal atrisi karyawan dengan menerapkan *machine learning* dengan beberapa metode seperti *Support Vector Machine* (SVM), *Decision Tree*, *Neural Network*, *Random Forest* dan sebagainya. Penelitian yang dilakukan oleh Ozdemir, Coskun, Gezer dan Gungor dalam prediksi atrisi karyawan menggunakan metode *Support Vector Machine* (SVM), *Random Forest*, J48, LogitBoost, Multilayer Perceptron (MLP), *K-Nearest Neighbors* (KNN), *Linear Discriminant Analysis* (LDA), Naive Bayes, *Bagging*, AdaBoost, Logistic Regression. Rekomendasi teknik klasifikasi yang menghasilkan nilai akurasi tertinggi pada penelitian tersebut adalah metode *Support Vector Machine* dengan nilai 85,92 % [9].

Reduksi dimensional adalah proses mengurangi jumlah variabel acak atau fitur dalam dataset dengan tujuan menyederhanakan data sambil tetap mempertahankan informasi yang relevan sebanyak mungkin [10], [11]. Pengurangan jumlah fitur atau dimensi dari dataset dapat menghilangkan kelebihan informasi atau noise, meningkatkan efisiensi komputasi, dan memperbaiki performa model [12]. Penelitian yang dilakukan oleh Sepriadi Bumbungan, Kusrini dan Kusnawi dalam penerapan reduksi dimensional menggunakan metode *Particle Swarm Optimization* (PSO) pada metode *Support Vector Machine* (SVM) dalam prediksi kelulusan mahasiswa Politeknik Amamapare Timika. Hasil penelitian tersebut menjelaskan bahwa *Particle Swarm Optimization* dapat meningkatkan hasil akurasi *Support Vector Machine* dari 93,33% menjadi 98,02% [13].

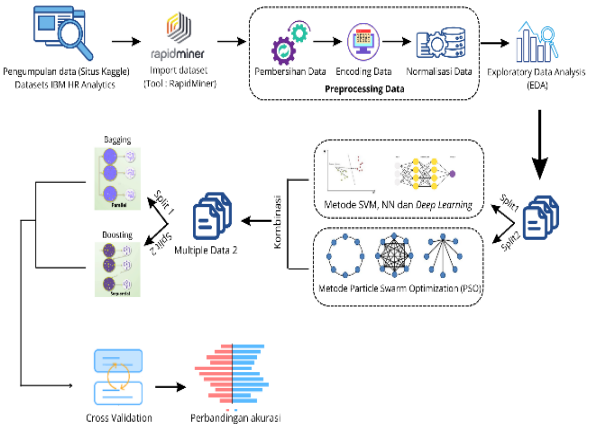
Metode *Ensemble* merupakan teknik yang menggabungkan hasil dari beberapa model prediktif untuk meningkatkan kinerja keseluruhan [14], [15]. Metode Ensemble yang digunakan pada penelitian ini adalah *Bagging* dan *Boosting*. *Bagging* adalah Teknik metode *Ensemble* ini melibatkan pembuatan beberapa model yang serupa dan menggabungkan hasil prediksi dari setiap model tersebut. Setiap model dibuat dengan menggunakan sampel acak dari data pelatihan. *Bagging* memiliki memiliki keuntungan dalam mengurangi varians sekaligus mempertahankan bias yang rendah [16]. Sedangkan *Boosting* membangun model secara berurutan, di mana setiap model baru mencoba untuk memperbaiki kesalahan dari model sebelumnya [17].

Penelitian yang dilakukan oleh Muneera Saad Alshiddy, Bader Nasser Aljaber terkait dengan prediksi atrisi menggunakan metode *Support Vector Machine* (SVM), *Naive Bayes* dan *Random Forest* yang dikombinasikan dengan metode *Ensemble*. Penelitian tersebut menjelaskan bahwa metode *Ensemble* dapat meningkatkan nilai akurasi. Akurasi tertinggi diperoleh oleh metode *Random Forest* dengan nilai akurasi sebesar 94,2%. Penerapan metode *Ensemble* pada penilitian ini dapat meningkatkan nilai akurasi dari metode *Random Forest* menjadi 94,5%. Oleh karena itu, hasil penelitian tersebut menunjukkan bahwa metode *Ensemble* dapat membantu proses prediksi atrisi karyawan [18].

Berdasarkan penelitian-penelitian terdahulu dapat disimpulkan bahwa *Machine Learning* dapat digunakan dalam prediksi atrisi karyawan. Adapun keterbaruan penelitian ini dari penelitian sebelumnya yaitu penerapan teknik reduksi dimensionalitas menggunakan *Particle Swarm Optimization* (PSO) yang dikombinasikan dengan metode *machine learning* (*Support Vector Machine, Deep Learning* dan *Neural Network*). Hasil dari kombinasi tersebut akan dioptimasi lagi dengan metode *Ensemble* yaitu menggunakan metode *Bagging* dan *Boosting* untuk meningkatan akurasi dalam prediksi atrisi karyawan dalam perusahaan.

2. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pada Gambar 1 terlihat tahapan pertama dalam penelitian ini yaitu pengumpulan dataset. Data yang digunakan dalam penelitian ini bersumber dari datasets *IBM HR Analytics Employee Attrition* yang tersedia di Kaggle (https://www.kaggle.com/) yang terdiri dari 35 atribut dengan total sebanyak 1470 data karyawan. Dataset yang digunakan dalam penelitian ini dapat dilihat pada Gambar 2.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Age	Attrition	BusinessTi	DailyRate	Departme	DistanceFr	Education	EducationI	EmployeeC	EmployeeI	EnvironmE	Gender
2	41	Yes	Travel_Rar	1102	Sales	1	2	Life Scienc	1	1	2	Female
3	49	No	Travel_Fre	279	Research &	8	1	Life Scienc	1	2	3	Male
4	37	Yes	Travel_Rar	1373	Research &	2	2	Other	1	4	4	Male
5	33	No	Travel_Fre	1392	Research &	3	4	Life Scienc	1	5	4	Female
6	27	No	Travel_Rar	591	Research &	2	1	Medical	1	7	1	Male
7	32	No	Travel_Fre	1005	Research &	2	2	Life Scienc	1	8	4	Male
8	59	No	Travel_Rar	1324	Research &	3	3	Medical	1	10	3	Female
9	30	No	Travel_Rar	1358	Research &	24	1	Life Scienc	1	11	4	Male
10	38	No	Travel_Fre	216	Research &	23	3	Life Scienc	1	12	4	Male
11	36	No	Travel_Rar	1299	Research &	27	3	Medical	1	13	3	Male
12	35	No	Travel_Rar	809	Research &	16	3	Medical	1	14	1	Male
13	29	No	Travel_Rar	153	Research &	15	2	Life Scienc	1	15	4	Female
14	31	No	Travel_Rar	670	Research &	26	1	Life Scienc	1	16	1	Male

Gambar 2. Dataset

Tahapan kedua yaitu preprocessing data ini membantu dalam meningkatkan kualitas dan akurasi data. Terdapat beberapa tahapan pada preprocessing data ini yaitu pembersihan data, *encoding* data, normalisasi data. Pembersihan data (*data cleaning*) merupakan proses untuk mendeteksi, memperbaiki, atau menghapus data yang tidak akurat, tidak lengkap, duplikat, atau tidak relevan dari dataset [19].

Hasil dari proses pembersihan data akan masuk ke tahapan selanjutnya yaitu *encoding* data. *Encoding* data yaitu proses mengubah data kategori (non-numerik) menjadi format numerik sehingga dapat digunakan oleh algoritma data mining [20], [21]. Tahapan selanjutnya yaitu normalisasi data yang bertujuan untuk mengubah skala data sehingga fitur-fitur dalam dataset memiliki skala yang sebanding. Pada penelitian ini menggunakan Z-Tranformation sebagai normalisasi data. Berikut merupakan formula Z-Transformation :

$$Z = \frac{X-\mu}{\sigma} \quad (1)$$

Dimana :
Z = Z-score
X = nilai yang ingin dihitung Z-scorenya
μ = rata-rata dari populasi
σ = deviasi standar dari populasi

Tahapan ketiga yaitu analisis data eksploratif (*Exploratory Data Analysis*). Analisis Data Eksploratif (*Exploratory Data Analysis* atau EDA) adalah pendekatan dalam analisis data yang digunakan untuk menganalisis set data dan merangkum karakteristik utamanya. EDA merupakan langkah awal yang sangat penting dalam proses analisis data karena memberikan pemahaman awal yang mendalam tentang data, membantu dalam membuat keputusan yang lebih baik dalam tahap-tahap analisis selanjutnya, dan memastikan bahwa model statistik atau *machine learning* yang digunakan akurat dan relevan [22].

Tahapan selanjutnya adalah implementasi model *machine learning* (SVM, NN dan *Deep learning*) dengan reduksi dimensionalitas metode *Particle Swarm Optimization* (PSO) mengetahui kontribusi reduksi dimensionalitas metode *Particle Swarm Optimization* (PSO) yang digunakan dapat meningkatkan akurasi prediksi atrisi karyawan. Berikut merupakan formula metode PSO [23]:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (p_{best,i} - x_i(t)) + c_2 \cdot r_2 \cdot (g_{best} - x_i(t)) \quad (2)$$

Dimana :
$v_i(t+1)$: Kecepatan partikel ii pada iterasi t+1t+1.
$v_i(t)$: Kecepatan partikel ii pada iterasi tt.
w: Bobot inersia (mengontrol kontribusi kecepatan sebelumnya).
c1: Koefisien pembelajaran untuk menarik partikel menuju posisi terbaik pribadi (pbest).
c2: Koefisien pembelajaran untuk menarik partikel menuju posisi terbaik global (gbest).
r1,r2: Angka acak antara 0 dan 1.

Tahap selanjutnya yaitu optimasi metode *Ensemble* dengan *Bagging* dan *Boosting*. Penggunaan *Bagging* dan *Bosting* pada penelitian ini untuk mengatasi *overfitting*, meningkatkan akurasi, dan membuat prediksi menjadi lebih stabil. Bagging adalah metode yang bekerja dengan cara mengambil sampel acak dari dataset asli dengan pengembalian (*bootstrap*) untuk membuat beberapa subset data. Sedangkan metode *Boosting* adalah metode yang menggabungkan prediksi dari beberapa model sederhana (*weak learners*).

Tahapan terakhir pada penelitian ini yaitu evaluasi model yang digunakan untuk mengukur kinerja model. Evaluasi model dilakukan dengan tool RapidMiner menggunakan parameter performance classification yang terdiri dari nilai akurasi, recall dan presisi. Akurasi mengukur proporsi prediksi yang benar dari keseluruhan prediksi yang dibuat oleh model. Recall mengukur kemampuan model untuk mendeteksi semua contoh positif yang sebenarnya dalam data. Sedangkan Presisi mengukur proporsi prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model. Adapun akurasi, recall dan presisi dapat dihitung menggunakan persamaan berikut [24]:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{Presisi} = \frac{TP}{TP+EP} \quad (5)$$

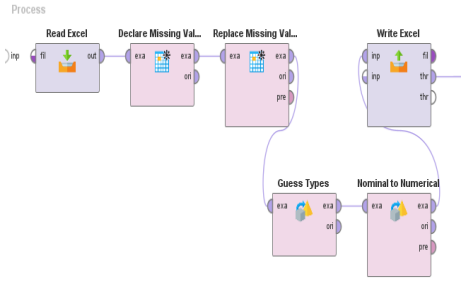
Dimana :
TP = Jumlah prediksi positif yang benar.
TN = Jumlah prediksi negatif yang benar.
FP = Jumlah prediksi positif yang salah.
FN = Jumlah prediksi negatif yang salah.

3. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan *machine learning* dalam memprediksi atrisi karyawan di perusahaan dengan menggunakan teknik reduksi dimensionalitas (PSO) yang dikombinasian dengan metode *Support Vector Machine* (SVM), *Deep Learning* dan *Neural Network* yang dioptimasi dengan metode *Ensemble* (*Bagging* dan *Boosting*). Data yang digunakan pada penelitian ini bersumber dari datasets IBM HR Analytics Employee Attrition. Alat bantu yang digunakan oleh peneliti dalam pembentukan model *machine learning* yaitu aplikasi RapidMiner.

3.1 *Preprocessing* Data

Tahapan ini dilakukan sebelum dataset digunakan dalam analisis atau pemodelan yang diproses melalui aplikasi RapidMiner. Pertama dataset dilakukan proses pembersihan data (data *cleaning*), tahapan ini dilakukan untuk menghapus dan memperbaiki data yang hilang. Proses ini dilanjutkan dengan tahapan *encoding* data, yang mana proses ini dilakukan untuk mengubah data kategori (non-numerik) menjadi format numerik. Tahapan ini dapat dilihat pada Gambar 3.



Gambar 3. Model Pembersihan Data (*Data Cleaning*)

Tahapan selanjutnya yaitu proses normalisasi yang dilakukan bertujuan untuk mengubah nilai data ke dalam skala yang sama atau rentang tertentu menggunakan Z-Transformation. Berikut merupakan *sample* dari proses *Pre-processing* yang dapat dilihat pada Gambar 4.

	A	B	C	D	E	F	G	H	I	J	K
1	Attrition	BusinessT	Departme	Education	Gender	JobRole	MaritalSt	Over18	OverTime	Age	DailyRate
2	0	0	0	0	0	0	0	0	0	41	1102
3	1	1	1	0	1	1	1	0	1	49	279
4	0	0	1	1	1	2	0	0	0	37	1373
5	1	1	1	0	0	1	1	0	0	33	1392
6	1	0	1	2	1	2	1	0	1	27	591
7	1	1	1	0	1	2	0	0	1	32	1005
8	1	0	1	2	0	2	1	0	0	59	1324
9	1	0	1	0	1	2	2	0	1	30	1358
10	1	1	1	0	1	3	0	0	1	38	216
11	1	0	1	2	1	4	1	0	1	36	1299
12	1	0	1	2	1	2	1	0	1	35	809
13	1	0	1	0	0	2	0	0	0	29	153
14	1	0	1	0	1	1	2	0	1	31	670
15	1	0	1	2	1	2	2	0	1	34	1346
16	0	0	1	0	1	2	0	0	0	28	103
17	1	0	1	0	0	3	2	0	1	29	1389
18	1	0	1	0	1	1	2	0	0	32	334
19	1	2	1	2	1	2	2	0	0	22	1123
20	1	0	0	0	0	5	1	0	1	53	1219

Gambar 4. Hasil *Pre-processing* data

Selanjutnya merupakan rincian atribut yang digunakan dalam prediksi atrisi karyawan ini terlihat pada Tabel 1.

Tabel 1. Rincian Atribut Dataset

No	Atribut	Tipe Data	Deskripsi Nilai
1	*Age*	Interger	18-60
2	*Attrition*	Binomial	Yes, No
3	*Bussiness Travel*	Binomial	1 (Travel Rarely), 2 (Travel_Frequently)
4	*Daily Rate*	Interger	102-1499
5	*Department*	Binomial	1 (Sales), 2 (Research), 3 (Human Resources)
6	*Distance From Home*	Interger	1-29
7	*Education*	Binomial	1 (Below), 2

			(College), 3 (Bachelor), 4 (Master), 5 (Doctor)
8	*Education Field*	Binomial	1 (Life Sciences), 2 (Medical), 3 (Marketing), 4 (Technical Degree), 5 (Human Resource), 6 (Other)
9	*Employee Count*	Interger	1
10	*Employee Number*	Interger	1-2068
11	*Enviroment Satisfaction*	Binomial	1 (Low), 2 (Medium), 3 (High), 4 (Very High)
12	*Gender*	Binomial	1 (Female), 2 (Male)
13	*Hourly Rate*	Interger	30-100
14	*Job Involvement*	Binomial	1 (Low), 2 (Medium), 3 (High), 4 (Very High)
15	*Job Level*	Interger	1-5
16	*Job Role*	Binomial	1 (Sales Executive), 2 (Research Scientist), 3 (Laboratory Technician), 4 (Manufacturing Diretor), 5 (Healtcare Representative)
17	*Job Satisfaction*	Binomial	1 (Low), 2 (Medium), 3 (High), 4 (Very High)
18	*Marital Status*	Binomial	1 (Single), 2 (Married), 3 (Divorced)
19	*Monthly Income*	Interger	1009 – 20.000
20	*Monthly Rate*	Interger	2094-27.000
21	*Num Companies Worked*	Interger	0-9
22	*Over 18*	Binomial	Yes, No
23	*Over Time*	Binomial	Yes, No
24	*Percent Salary Hike*	Interger	11-25
25	*Performance Rating*	Binomial	1 (Low), 2 (Good), 3 (Excellent), 4 (Outstanding)
26	*Relationship Satisfaction*	Binomial	1 (Low), 2 (Medium), 3 (High), 4 (Very High)
27	*Standart Hours*	Interger	80
28	*Stock Optional Level*	Binomial	0-3
29	*Total Working Years*	Interger	0-4
30	*Training Times Last Year*	Interger	0-6
31	*Work Life Balance*	Binomial	1 (Bad), 2 (Good), 3 (Better), 4 (Best)
32	*Years at Company*	Interger	0-40
33	*Training Times Last Years*	Interger	0-18
34	*Years Since Last Promotion*	Interger	0-15
35	*Years With Curr Manager*	Interger	0-17

3.2 *Exploratory Data Analysis* (EDA)

Analisis Data Eksploratif (*Exploratory Data Analysis* atau EDA) adalah pendekatan analisis yang bertujuan untuk memahami struktur, karakteristik, dan pola dalam data tanpa asumsi awal. Melalui EDA, analis dapat menggunakan statistik deskriptif dan visualisasi data seperti *histogram*, *boxplot*, dan *scatter plot* untuk mengidentifikasi pola, mendeteksi anomali, dan menemukan hubungan antar variabel.

Pada penelitian ini analisis data eksploratif dilakukan untuk memperoleh wawasan yang penting dari dataset atrisi karyawan yang divisualisasikan dengan *scatter plot*. *Scatter plot* merupakan jenis grafik yang digunakan untuk menampilkan hubungan antara dua variabel numerik. Setiap titik pada *scatter plot* mewakili satu pengamatan dari data, dengan posisi pada sumbu x dan y menunjukkan nilai dari dua variabel yang berbeda. Visualisasi *scatter plot* data atrisi karyawan dapat dilihat pada Gambar 7-10.

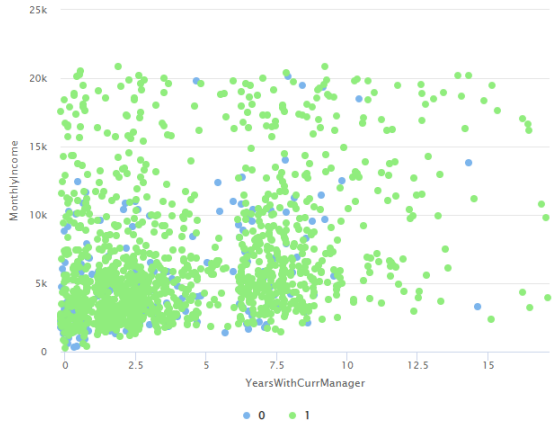
Gambar 5. Analisis distribusi data total tahun bekerja dan pendapatan bulanan berdasarkan atrisi karyawan

Berdasarkan pada Gambar 5 dapat dijelaskan bahwa atribut pendapatan bulanan dan total tahun bekerja berpengaruh pada keputusan atrisi karyawan. Analisis menunjukkan bahwa total tahun bekerja rentang 0-10 tahun dengan pendapatan bulanan ≤10.000 tingkat atrisi karyawan tinggi. Ketika tahun bekerja >10 tahun dan pendapatan bulanan meningkat >10.000 daapat dilihat untuk tingkat atrisi karyawan rendah. Analisis ini menunjukkan bahwa tingkat atrisi karyawan tinggi ketika pendapatan bulanan rendah dan total tahun

kerja pendek. Pendapatan bulanan ini merupakan faktor yang mempengaruhi atrisi karyawan.

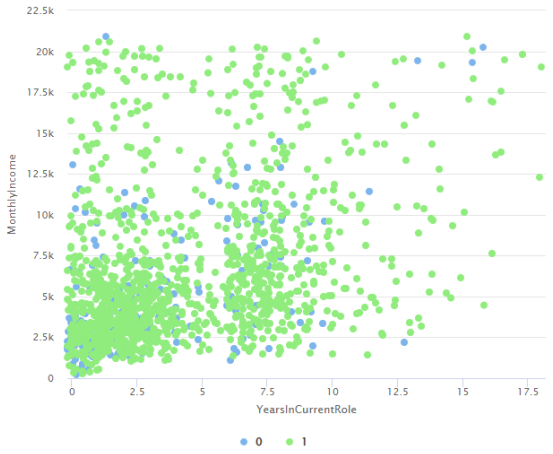
Gambar 6. Analisis distribusi data umur dan pendapatan bulanan berdasarkan atrisi karyawan

Berdasarkan Gambar 6 dapat dianalisis bahwa tingkat atrisi tinggi ketika karyawan berumur ≤40 tahun dengan pendapatan bulannya ≤5000. Ketika pendapatan bulanan meningkat >5000 tingkat atrisi karyawan semakin rendah. Hal ini menunjukkan bahwa pendapatan bulanan sangat berpengaruh sebagai faktor yang paling berpengaruh yang mendasari karyawan untuk keluar dari perusahaan.



Gambar 7. Analisis distribusi data total tahun dengan manajer saat ini dan pendapatan bulanan berdasarkan atrisi karyawan

Gambar 7 menggambarkan total tahun bekerja bersama manajer saat ini dengan pendapatan bulanan. Berdasarkan gambar tersebut dapat dijelaskan bahwa peran manajer saat ini mempengaruhi tingkat atrisi karyawan, yang mana pada rentang 0-5 tahun dengan pendapatan bulanan <5000 menunjukkan tingkat atrisi yang lumayan tinggi. Sehingga dapat menjadi masukan bagi perusahaan bahwa kedekatan dengan manajer yang menjabat menjadi faktor penentu dalam pengurangan karyawan.

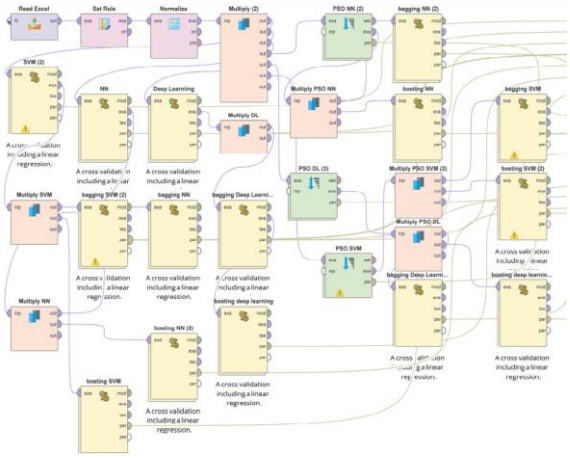


Gambar 8. Analisis distribusi data total tahun bekerja pada jabatan saat ini dan pendapatan bulanan berdasarkan atrisi karyawan

Ilustrasi pada Gambar 8 dapat menjelaskan bahwa total tahun bekerja pada jabatan saat ini ≤5 tahun dengan pendapatan ≤5.000 mengalami tingkat atrisi karyawan yang tinggi. Bertambahnya pendapatan dengan total tahun bekerja mengurangi tingkat pengurangan karyawan.

3.2 Pemodelan *Machine learning*

Data terlebih dahulu diproses reduksi dimensionalitas menggunakan metode *Particle Swarm Optimization*. Tahapan selanjutnya data yang telah direduksi akan dilakukan klasifikasi menggunakan metode *Support Vector Machine*, *Deep Learning* dan *Neural Network*. Hasil dari proses klasifikasi akan dioptimasi menggunakan metode *Ensemble* yaitu *Bagging* dan *Boosting*. Penelitian ini bertujuan untuk melakukan perbandingan akurasi terhadap metode tanpa menggunakan reduksi dimensionalitas dan metode yang diterapkan reduksi dimensionalitas. Pemodelan *machine learning* ini dilakukan pada aplikasi RapidMiner. Berikut model *machine learning* yang dibangun dapat dilihat pada Gambar 9.



Gambar 9. Model Predikasi Atrisi Karyawan

Pengukuran performa dalam penelitian ini peneliti menggunakan validasi 10 *Fold Cross*. Validasi 10 *Fold Cross* ini merupakan metode evaluasi model *machine learning* yang melibatkan pembagian dataset menjadi 10 bagian yang kira-kira sama besar. Dalam setiap iterasi dari total 10 iterasi, satu *fold* digunakan sebagai data uji, sementara sembilan *fold* lainnya digunakan sebagai data pelatihan. Proses ini berulang sehingga setiap *fold* digunakan sekali sebagai data uji.

Penelitian ini bertujuan untuk memperoleh metode dengan performa tertinggi yang dapat digunakan untuk memprediksi awal atrisi karyawan. Percobaan pertama peneliti melakukan proses klasifikasi tanpa menerapkan teknik *Particle Swarm Optimization*, sehingga metode langsung dioptimasikan dengan metode *Ensemble* yaitu *Bagging* dan *Boosting*. Hasil performa percobaan pertama ini dapat dilihat Pada tabel 2.

Tabel 2. Performa Metode Tanpa *Particle Swarm Optimization*

Metode	**Metode Ensemble**	**Performa**		
		Akurasi	Presisi	*Recall*
SVM	-	85,37%	85,44%	99,60%
SVM	*Bagging*	85,58%	85,49%	99,76%
SVM	*Boosting*	83,88%	83,88%	100,00%
DL	-	85,85%	89,59%	94,09%
DL	*Bagging*	86,87%	89,17%	96,03%
DL	*Boosting*	84,08%	88,91%	92,68%
NN	-	84,69%	89,02%	93,26%
NN	*Bagging*	85,10%	88,76%	94,14%
NN	*Boosting*	84,22%	89,94%	91,40%

Percobaan kedua peneliti melakukan proses klasifikasi atrisi karyawan dengan menerapkan teknik reduksi dimensional dan mengoptimasikan dengan metode *Ensemble* yaitu *Bagging* dan *Boosting*. Hasil performa percobaan kedua ini dapat dilihat pada Tabel 3.

Tabel 3. Performa Metode dengan *Particle Swarm Optimization*

Reduksi Dimensionalitas	**Metode**	**Metode *Ensemble***	**Performa**		
			Akurasi	**Presisi**	**Recall**
-	SVM	-	85,37%	85,44%	99,60%
PSO	SVM	-	85,92%	85,76%	99,84%
PSO	SVM	*Bagging*	85,10%	85,18%	99,60%
PSO	SVM	*Boosting*	85,24%	85,24%	99,68%
-	DL	-	85,85%	89,59%	94,09%
PSO	DL	-	86,60%	89,91%	94,75%
PSO	DL	*Bagging*	86,94%	88,90%	96,40%
PSO	DL	*Boosting*	85,24%	89,12%	93,83%
-	NN	-	84,69%	89,02%	93,26%
PSO	NN	-	85,24%	89,57%	93,38%
PSO	NN	*Bagging*	86,80%	89,80%	95,05%
PSO	NN	*Boosting*	83,74%	90,03%	90,54%

Berdasarkan hasil performa setiap metode yang digunakan pada penelitian ini dapat mencapai nilai akurasi, presisi dan *recall* yang baik. Berdasarkan perbandingan dari Tabel 2 dan Tabel 3 dapat diketahui bahwa penerapan teknik reduksi dimensionalitas dapat meningkatkan nilai akurasi meski tidak secara signifikan yaitu $\pm$1%. Peningkatan akurasi ini terjadi dikarenakan *Particle Swarm Optimization* dapat digunakan untuk mencari kombinasi fitur yang optimal dengan mengevaluasi subset fitur yang berbeda secara iteratif. Setiap partikel dalam *Particle Swarm Optimization* mewakili subset fitur, dan posisi partikel diperbarui berdasarkan kinerja klasifikasi pada subset tersebut, sehingga *Particle Swarm Optimization* dapat memberikan kinerja terbaik, mengurangi kompleksitas model, dan meningkatkan akurasi prediksi.

Tabel 2 dan Tabel 3 menjelaskan performa dari penerapan metode *Ensemble*. Berdasarkan tabel tersebut dapat dijelaskan untuk metode *Bagging* dapat meningkatkan nilai akurasi ketiga metode dalam kedua percobaan tersebut, sedangkan metode *Boosting* gagal dalam meningkatkan nilai akurasi ketiga metode dalam kedua percobaan tersebut. Peningkatan akurasi pada metode *Bagging* pun kurang signifikan yaitu $\pm$1%. Dari hasil pengolahan data ini dapat dikatakan bahwa metode *Boosting* tidak sesuai apabila diterapkan pada ketiga metode yang digunakan pada data atrisi karyawan ini.

Perbandingan akurasi dari kedua percobaan ini didapatkan hasil bahwa akurasi tertinggi dalam prediksi atrisi karyawan ini yaitu pada percobaan kedua dimana menerapkan reduksi dimensionalitas dengan ketiga metode yaitu *Support Vector Machine*, *Deep Learning* dan *Neural Network*, yang kemudian hasil dari keduanya dioptimasi dengan metode *Ensemble*. Metode dengan akurasi tertinggi yaitu metode *Deep Learning* dengan nilai akurasi sebesar 86,94%, nilai presisi sebesar 88,90% dan nilai recall sebesar 96,40%. Penerapan metode *Particle Swarm Optimization* dan metode *Bagging* ini dapat meningkatkan nilai akurasi hingga 1,44%.

4. DISKUSI

Penerapan teknik reduksi dimensionalitas untuk mengurangi jumlah fitur dalam dataset sekaligus mempertahankan informasi yang paling relevan dan memiliki manfaat penting dalam tahapan klasifikasi. Metode yang digunakan pada penelitian ini yaitu metode *Particle Swarm Optimization* (PSO) dapat meningkatkan akurasi metode *Support Vector Machine*, *Deep Learning* dan *Neural Network*. Hasil ini sejalan dengan penelitian terdahulu yang dilakukan oleh Jiayuan Hu, Rui Zhang, Huahong Zhu, dan Lin He dengan judul *A Novel Classification Method Based on Particle Swarm Optimization*. Penelitian ini melakukan

percobaan dengan mengkombinasikan metode *Particle Swarm Optimization* dengan metode *Support Vector Machine* dengan menggunakan dataset publik ISCXVPN2016. Hasil dari penelitian tersebut didapatkan nilai akurasi mencapai 97,71% yang mengalami peningkatan sebesar 27,96% ketika dikombinasikan dengan metode *Particle Swarm Optimization* [10].

Penelitian lainnya dilakukan oleh Mandakini Priyadarshani Behera, Archana Sarangi, Debahuti Mishra dan Shubhendu Kumar Sarangi melakukan sebuah penelitian pada tahun 2023 dengan judul *A Hybrid Machine learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine*. Penelitian tersebut melakukan prediksi terhadap hati dan jantung menggunakan metode *Support Vector Machine* yang dikombinasikan dengan *Particle Swarm Optimization*. Hasil penelitian tersebut menjelaskan bahwa PSO dapat meningkatkan metode *Support Vector Machine*. Akurasi prediksi jantung menggunakan *Support Vector Machine* yaitu 87,04% dan meningkat menjadi 88,89% setelah ditambahkan metode *Particle Swarm Optimization*. Sedangkan akurasi prediksi hati yang didapatkan sebesar 62,93% dan meningkat signifikan menjadi 88,89% [11].

Penelitian ini juga mengoptimasi metode (*Support Vector Machine*, *Deep Learning* dan *Neural Network*,) menggunakan metode Ensemble yaitu *Bagging* dan *Boosting*. *Bagging* dapat meningkatkan nilai akurasi dari metode yang digunakan pada penelitian ini, sedangkan *Boosting* tidak meningkatkan keseluruhan metode. Penelitian ini sejalan dengan penelitian sebelumnya yang dilakukan oleh Prakriti Dhakal, Santosh Khanal, dan Rabindra Bista dengan judul *Prediction of Anemia using Machine learning Algorithms* [15]. Penelitian ini dilakukan untuk merancang model prediksi anemia yang dikumpulkan dari Rumah Sakit Anak. Hasil penelitian ini menjelaskan bahwa penggunaan metode *Ensemble* yaitu *Bagging* dapat meningkatkan nilai akurasi dari beberapa metode yaitu *Logistic Regression*, *Support Vector Machine*, Naive Bayes, Decision Tree dan *Random Forest*. Akurasi yang tertinggi didapatkan oleh metode *Random Forest* dengan nilai akurasi sebesar 93,8% menjadi 98,8% setelah dioptimasi dengan *Bagging*. Sedangkan untuk percobaan menggunakan *Boosting* didapatkan bahwa *Boosting* dapat meningkatkan nilai akurasi metode Naive Bayes dari 90,7% menjadi 93,7%,

5. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan untuk prediksi atrisi karyawan menggunakan metode *Support Vector Machine*, *Deep Learning* dan *Neural Network* yang dikombinasikan dengan reduksi dimensionalitas (*Particle Swarm Optimization*) dan dioptimasi dengan metode *Ensemble* (*Bagging*) dapat meningkatkan nilai akurasi tidak terlalu signifikan. Namun untuk penerapan metode ensembel (*Boosting*) gagal meningkatkan nilai akurasi dari ketiga metode tersebut. Akurasi terbaik dalam prediksi atrisi ini diperoleh oleh metode *Deep Learning* dengan nilai akurasi sebesar 86,94%, nilai presisi sebesar 88,90% dan nilai recall sebesar 96,40% setelah dikombinasikan dengan *Particle Swarm Optimization* dan dioptimasi dengan *Bagging*.

Proses Analisis data eksplorasi (*Exploratory Data Analysis*) ini membantu perusahaan untuk memahami pola-pola kompleks yang mendasari keputusan karyawan untuk keluar perusahaan sehingga perusahaan dapat mengambil langkah-langkah yang lebih proaktif dan tepat sasaran dalam meningkatkan retensi karyawan. Pada penelitian ini faktor utama penentu pengurangan karyawan adalah pendapatan bulanan, umur, tahun bekerja dikantor dan total tahun bekerja dengan manajer saat ini (peran manajer). Dengan adanya model ini, diharapkan dapat meningkatkan efisiensi dan keakuratan dalam prediksi awal prediksi karyawan dalam perusahaan berdasarkan parameter kinerja dan parameter terkait lainnya, sehingga dapat membantu manajer dalam pengambilan keputusan yang lebih tepat sebagai upaya untuk meningkatkan retensi karyawan dalam perusahaan

DAFTAR PUSTAKA

[1] T. H. L. Tulus, A. Id Hadiana, and I. Santikarama, "Sistem Prediksi Awal Terhadap Atrisi Karyawan Menggunakan Algoritma C4.5," *Informatics Digit. Expert*, vol. 4, no. 1, pp. 18–24, 2022, doi: 10.36423/index.v4i1.882.

[2] P. K. Jain, M. Jain, and R. Pamula, "Explaining and predicting employees' attrition: a machine learning approach," *SN Appl. Sci.*, vol. 2, no. 4, pp. 1–11, 2020, doi: 10.1007/s42452-020-2519-4.

[3] N. Mansor, N. S. Sani, and M. Aliff, "Machine Learning for Predicting Employee Attrition," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 435–445, 2021, doi: 10.14569/IJACSA.2021.0121149.

[4] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.

[5] D. Afreyna Fauziah, A. Maududie, and I. Nuritha, "Klasifikasi Berita Politik Menggunakan Algoritma K-nearst Neighbor (Classification of Political News Content using K-Nearest Neighbor)," *Berk. Sainstek*, vol. 6, pp. 106–114, 2018, doi: https://doi.org/10.19184/bst.v6i2.9256.

[6] A. Muliawan, A. Rizal, and S. Hadiyoso, "Heart Disease Prediction Based on Physiological Parameters Using Ensemble Classifier and Parameter Optimization," *J. Appl. Eng. Technol. Sci.*, vol. 5, no. 1, pp. 258–267, 2023, doi: 10.37385/jaets.v5i1.2169.

[7] A. Raza, K. Munir, M. Almutairi, F. Younas, M. Muhammad, and S. Fareed, "Predicting Employee Attrition Using Machine Learning Approaches," *Appl. Sci.*, 2022, doi: https://doi.org/10.3390/app12136424.

[8] İ. ERSÖZ KAYA and O. KORKMAZ, "Machine Learning Approach for Predicting Employee Attrition and Factors Leading to Attrition," *Çukurova Üniversitesi Mühendislik Fakültesi Derg.*, vol. 36, no. 4, pp. 913–928, 2021, doi: 10.21605/cukurovaumfd.1040487.

[9] F. Ozdemir, M. Coskun, C. Gezer, and V. Cagri Gungor, "Assessing Employee Attrition Using Classifications Algorithms," *ACM Int. Conf. Proceeding Ser.*, pp. 118–122, 2020, doi: 10.1145/3404663.3404681.

[10] J. Hu, R. Zhang, H. Zhu, and L. He, "A Novel Classification Method Based on Particle Swarm Optimization," *Proc. 2022 5th Int. Conf. Artif. Intell. Pattern Recognit.*, 2022, doi: https://doi.org/10.1145/3573942.3573983.

[11] M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine," *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 818–827, 2022, doi: 10.1016/j.procs.2023.01.062.

[12] T. Cuong-Le, T. Nghia-Nguyen, S. Khatir, P. Trong-Nguyen, S. Mirjalili, and K. D. Nguyen, "An efficient approach for damage identification based on improved machine learning using PSO-SVM," *Eng. Comput.*, vol. 38, pp. 3069–3084, 2022, doi: 10.1007/s00366-021-01299-6.

[13] S. Bumbungan, Kusrini, and Kusnawi, "Penerapan Particle Swarm Optimization (PSO) dalam Pemilihan Parameter Secara Otomatis pada Support Vector Machine (SVM) untuk Prediksi Kelulusan Mahasiswa Politeknik Amamapare Timika," *J. Tek. AMATA*, vol. 4, no. 1, pp. 81–93, 2022, doi: 10.55334/jtam.v4i1.77.

[14] I. Syarif, E. Zaluska, A. Prugel-Bennett, and G. Wills, "Application of bagging, boosting and stacking to intrusion detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7376 LNAI, pp. 593–602, 2012, doi: 10.1007/978-3-642-31537-4_46.

[15] P. Dhakal, S. Khanal, and R. Bista, "Prediction of Anemia using Machine Learning Algorithms," *Int. J. Comput. Sci. Inf. Technol.*, vol. 15, no. 1, pp. 15–30, 2023, doi: 10.5121/ijcsit.2023.15102.

[16] N. H. A. Malek, W. F. W. Yaacob, Y. B. Wah, S. A. Md Nasir, N. Shaadan, and S. W. Indratno, "Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 1, pp. 598–608, 2023, doi: 10.11591/ijeecs.v29.i1.pp598-608.

[17] H. ŞEVGİN, "A comparative study of ensemble methods in the field of education: Bagging and Boosting algorithms," *Int. J. Assess. Tools Educ.*, vol. 10, no. 3, pp. 544–562, 2023, doi: 10.21449/ijate.1167705.

[18] M. S. Alshiddy and B. N. Aljaber, "Employee Attrition Prediction using Nested Ensemble Learning Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 7, pp. 932–938, 2023, doi: 10.14569/IJACSA.2023.01407101.

[19] F. Ridzuan and W. M. N. Wan Zainon, "A review on data cleansing methods for big data," *Procedia Comput. Sci.*, vol. 161, pp. 731–738, 2019, doi: 10.1016/j.procs.2019.11.177.

[20] S. Mittal and S. Nag, "A survey of encoding techniques for reducing data-movement energy," *J. Syst. Archit.*, vol. 97, pp. 373–396, 2019, doi: 10.1016/j.sysarc.2018.11.001.

[21] A. Muliawan, T. Badriyah, and I. Syarif, "Membangun Sistem Rekomendasi Hotel dengan Content Based Filtering Menggunakan K-Nearest Neighbor dan Haversine Formula," *Technomedia J.*, vol. 7, no. 2, pp. 231–247, 2022, doi: 10.33050/tmj.v7i2.1893.

[22] A. Data, E. Eda, S. Peraih, and M. Olimpiade, "Exploratory Data Analysis (EDA): A Study of Olympic Medallist," vol. 11, no. September, pp. 578–587, 2022, doi: https://doi.org/10.32520/stmsi.v11i3.1857.

[23] E. H. Houssein, A. G. Gad, K. Hussain, and P. N. Suganthan, "Major Advances in Particle Swarm Optimization: Theory, Analysis, and Application," *Swarm Evol. Comput.*, vol. 63, no. March, p. 100868, 2021, doi: 10.1016/j.swevo.2021.100868.

[24] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.