



Penerapan Algoritma C4.5 dengan Teknik *Resample* untuk Prediksi Kinerja Pegawai PT X

Wahyu Saputro

Program Studi Teknik Komputer, Program Diploma Tiga, Politeknik Prasetiya Mandiri, Kota Bogor, Jl. Bina Marga No.19, RT.06/RW.08, Baranangsiang, Kec. Bogor Tim., Kota Bogor, Jawa Barat 16143, Indonesia

*Penulis Korespondensi: wahyu.prasetiya.mandiri@gmail.com

Abstract. Human Resource Management (HRM) plays a strategic role in improving organizational competitiveness through proper management of employee placement, training, and performance evaluation. To support the achievement of these goals, a predictive model is needed that can provide an accurate picture of employee performance. This study utilizes a Human Resource Management (HRM) dataset of 1,200 data and applies several classification algorithms to compare their effectiveness, namely J48 or C4.5, Random Forest, Naive Bayes, K-Nearest Neighbor (KNN), Logistic Regression, and Support Vector Machine (SVM). To obtain more optimal results, this study uses resampling techniques and attribute selection methods with a correlation attribute eval approach, so that class distribution can be more balanced and model accuracy increases. From the test results, the Decision Tree J48 algorithm showed the best performance with an accuracy level reaching 95.41%, a kappa value of 0.8925, a mean absolute error (MAE) of 0.0432, a precision of 0.955, a recall of 0.954, and an area under the ROC curve of 0.964. These findings indicate that J48 has excellent predictive capabilities compared to other algorithms. Furthermore, this study also found that the most influential variables in determining employee performance include the percentage of the last salary increase (EmpLast Salary Hike Percent), the level of work environment satisfaction (Emp Environment Satisfaction), the length of time since the last promotion (Years Since Last Promotion), and experience in the current role (Experience Years in Current Role). Overall, the results of the study indicate that the C4.5 algorithm with the application of the resampling technique can be an optimal solution in building an employee performance prediction system. Thus, this model has the potential to be a strong basis for managerial decision-making, particularly in designing HR development strategies and policies to improve organizational performance.

Keywords : C4.5 (J48) Algorithm; Classification; Employee Performance Prediction; Resample; Testing.

Abstrak. Manajemen Sumber Daya Manusia (SDM) berperan strategis dalam meningkatkan daya saing organisasi melalui pengelolaan yang tepat pada aspek penempatan, pelatihan, serta evaluasi kinerja pegawai. Untuk mendukung pencapaian tujuan tersebut, dibutuhkan suatu model prediksi yang mampu memberikan gambaran akurat mengenai performa pegawai. Penelitian ini memanfaatkan dataset Human Resource Management (HRM) sebanyak 1.200 data dan menerapkan beberapa algoritma klasifikasi guna membandingkan efektivitasnya, yaitu J48 atau C4.5, Random Forest, Naive Bayes, K-Nearest Neighbor (KNN), Regresi Logistik, serta Support Vector Machine (SVM). Untuk memperoleh hasil yang lebih optimal, penelitian ini menggunakan teknik resample dan metode pemilihan atribut dengan pendekatan correlation attribute eval, sehingga distribusi kelas dapat lebih seimbang dan akurasi model meningkat. Dari hasil pengujian, algoritma Decision Tree J48 menunjukkan performa paling unggul dengan tingkat akurasi mencapai 95,41%, nilai kappa 0,8925, mean absolute error (MAE) sebesar 0,0432, precision 0,955, recall 0,954, serta area under ROC curve 0,964. Temuan ini mengindikasikan bahwa J48 memiliki kemampuan prediksi yang sangat baik dibandingkan algoritma lainnya. Lebih lanjut, penelitian ini juga menemukan bahwa variabel yang paling berpengaruh dalam menentukan kinerja pegawai meliputi persentase kenaikan gaji terakhir (EmpLast Salary Hike Percent), tingkat kepuasan lingkungan kerja (Emp Environment Satisfaction), lama waktu sejak promosi terakhir (Years Since Last Promotion), serta pengalaman dalam peran saat ini (Experience Years in Current Role). Keseluruhan hasil penelitian menunjukkan bahwa algoritma C4.5 dengan penerapan teknik resample dapat dijadikan solusi optimal dalam membangun sistem prediksi kinerja pegawai. Dengan demikian, model ini berpotensi menjadi dasar yang kuat bagi pengambilan keputusan manajerial, khususnya dalam merancang strategi pengembangan SDM dan kebijakan peningkatan kinerja organisasi.

Kata kunci: Algoritma C4.5 (J48); Klasifikasi; Pengujian; Prediksi Kinerja Pegawai; *Resample*.

1. LATAR BELAKANG

Sumber daya manusia merupakan aset paling berharga bagi sebuah organisasi yang perlu dipelihara keberadaannya. Upaya pengembangan SDM memberikan banyak manfaat, seperti meningkatnya produktivitas kerja, terciptanya hubungan yang harmonis antara pimpinan dan bawahan, tersedianya proses pengambilan keputusan yang lebih cepat dan akurat, serta bertambahnya motivasi kerja seluruh anggota organisasi. Oleh karena itu, Manajemen Sumber Daya Manusia (SDM) menjadi faktor krusial bagi para manajer maupun pengambil keputusan di berbagai bidang usaha, karena melalui perencanaan yang baik, perusahaan dapat memperoleh tenaga kerja yang benar-benar kompeten dan sesuai kualifikasi (Putri Latif & Ali, 2025).

HRM memiliki peran strategis dalam meningkatkan daya saing dan efektivitas organisasi secara berkelanjutan. HRM dipandang sebagai “praktisi manusia”, yang bertanggung jawab dalam menempatkan karyawan terbaik pada posisi yang sesuai, memberikan pelatihan dan pengembangan kompetensi, serta membangun sistem evaluasi untuk memantau kinerja sekaligus menjaga keberlangsungan talenta potensial. Dalam hal ini, teknik klasifikasi sering digunakan untuk membangun model yang mampu memprediksi tren di masa depan. Model prediktif melalui klasifikasi memiliki target tertentu, sehingga dapat memperkirakan variabel yang belum diketahui berdasarkan data historis variabel lain yang sudah tersedia. Untuk mendorong kemajuan dan keberhasilan perusahaan, kinerja pegawai yang optimal, berkualitas, dan konsisten sangat diperlukan.

Maka dari itu peneliti melakukan penelitian model untuk memprediksi peringkat kinerja pegawai. Dimana dalam memprediksi peringkat kinerja pegawai tersebut menggunakan beberapa algoritma yaitu *Decision Tree J48*, *Random Forest*, *Naive Bayes*, *Support Vector Machine (SVM)*, *Logistic Regression*, *K-Nearest Neighbour (KNN)* (Diana et al., 2022). Kemudian ditambah dengan *Feature selection algorithm* yaitu *correlation attribute eval* untuk mengetahui variabel apa saja yang menjadi faktor pengaruh dalam prediksi peringkat kinerja pegawai.

Penelitian sebelumnya tentang kinerja pegawai menggunakan algoritma *Decision Tree C.45*, *SVM*, dan *Naive Bayes* (Nur Hafid Ibrahim & Laelatul Khikmah, 2024) melakukan penelitian dengan menggunakan *fitur selection algorithm* yaitu *Correlation-Attributeeval*, *Gainratio-Attributeeval*, *Relieff-Attributeeval* menunjukkan bahwa ketiga teknik memiliki konvergen dan akurasi sedang, yang lebih besar dari 70%. Itu akurasi moderat dapat dianggap dapat diterima akurasi dalam banyak kasus. Dalam ketiga percobaan, dataset menghasilkan model yang memuaskan untuk masing-masing dari ketiganya teknik klasifikasi yang dipilih.

SVM Teknik ditemukan sebagai classifier yang paling cocok untuk membangun model prediksi, di mana ia memiliki yang terbesar akurasi prediksi melalui ketiga percobaan yang telah dieksekusi dengan persentase tertinggi 86,90%.

Merujuk pada paparan sebelumnya, penelitian ini akan menggunakan jumlah data yang lebih besar dibandingkan studi rujukan, yakni sebanyak 1200 data. Selain itu, penulis juga akan menerapkan teknik resampling guna menyeimbangkan distribusi kelas minoritas dan mayoritas sehingga performa model dapat ditingkatkan. Model yang dihasilkan diharapkan mampu memberikan prediksi yang lebih akurat terkait peringkat kinerja pegawai serta mengidentifikasi variabel-variabel dominan yang berpengaruh terhadap penentuan peringkat tersebut.

2. KAJIAN TEORITIS

A. Decision Tree C.45

Algoritma **Decision Tree C4.5**(Ramadani et al., 2024) adalah salah satu metode klasifikasi yang dikembangkan oleh **Ross Quinlan** sebagai pengembangan dari algoritma **ID3**. Algoritma ini bekerja dengan membangun sebuah pohon keputusan (decision tree) yang digunakan untuk memprediksi kelas dari suatu data berdasarkan atribut-atribut yang dimilikinya.(Abdillah et al., 2020)

B. Naive Bayes

Algoritma Naive Bayes(Susanti et al., 2023) merupakan salah satu teknik yang digunakan dalam proses klasifikasi. Metode ini memanfaatkan konsep probabilitas dan statistika untuk menentukan kategori suatu data. Algoritma ini dikembangkan berdasarkan pemikiran Thomas Bayes, seorang ilmuwan asal Inggris, yang merumuskan cara untuk memprediksi kemungkinan kejadian di masa depan berdasarkan data historis. Pendekatan ini dikenal dengan sebutan Teorema Bayes. Dalam penerapannya, teorema ini digabungkan dengan asumsi "Naive" yang menganggap bahwa setiap atribut dalam data bersifat independen satu sama lain.

C. Support Vector Machine

Bagian ini menjelaskan teori-teori relevan yang menjadi dasar topik penelitian dan memberikan tinjauan beberapa penelitian terdahulu terkait, beserta referensi dan latar belakang yang mendukung penelitian ini. Jika terdapat hipotesis, hipotesis tersebut

dapat dinyatakan tidak secara eksplisit dan tidak perlu diajukan dalam bentuk pertanyaan(Wijiyanto et al., 2024).

D. K-Nearest Neighbor

K-Nearest Neighbor adalah algoritma pembelajaran mesin sederhana. Algoritma ini bekerja berdasarkan gagasan bahwa objek-objek yang berdekatan cenderung memiliki karakteristik yang serupa. Artinya, jika kita mengetahui fitur suatu objek, kita dapat memprediksi fitur objek lain dengan melihat tetangga terdekatnya. K-NN(A'yuniyah & Reza, 2023) adalah versi lanjutan dari teknik klasifikasi Nearest Neighbor(Syihabuddin Azmil Umri, 2021).

E. Random Forest

Random Forest adalah salah satu algoritma machine learning berbasis **ensemble learning** yang digunakan untuk klasifikasi maupun regresi. Algoritma ini dikembangkan oleh **Leo Breiman** dan bekerja dengan menggabungkan banyak pohon keputusan (decision tree) untuk menghasilkan prediksi yang lebih akurat dan stabil.(Hamami & Dahlan, 2022).

F. K-Fold Cross Validation

K-Fold Cross-Validation merupakan metode validasi pada machine learning yang digunakan untuk mengevaluasi performa model secara lebih tepat, meminimalkan risiko overfitting, serta memastikan kemampuan model dalam melakukan generalisasi. Pada teknik ini, data pelatihan dibagi menjadi K bagian (fold) dengan ukuran yang sama. Secara bergantian, satu fold dipakai sebagai data validasi, sedangkan fold lainnya digunakan untuk melatih model. Proses ini dilakukan sebanyak K kali sehingga setiap fold berperan sebagai data validasi tepat satu kali.(Wijiyanto et al., 2024).

G. Preprocessing Data

Preprocessing data merupakan langkah penting dalam menyiapkan data mentah agar lebih bersih, terstruktur, dan siap digunakan dalam analisis maupun pelatihan model machine learning. Tahap ini mencakup pembersihan data (menangani nilai yang hilang), integrasi data (menggabungkan data dari berbagai sumber), transformasi data (mengubah format atau skala data), serta reduksi data (menyederhanakan dimensi atau

ukuran data). Tujuannya adalah menghasilkan data yang berkualitas tinggi sehingga analisis lebih tepat dan pemodelan lebih efisien.(Agung et al., 2023).

H. Confusion Matrix

Confusion Matrix merupakan alat yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan cara mengukur prediksi yang benar maupun salah. Matriks ini membandingkan hasil prediksi dengan kelas asli dari data input, sehingga menyajikan informasi mengenai nilai aktual dan nilai prediksi pada proses klasifikasi(Pahlevi et al., 2023).

I. Klasifikasi

Klasifikasi adalah proses membangun sebuah model atau fungsi yang mampu menggambarkan sekaligus membedakan suatu kelas data atau konsep, dengan tujuan agar dapat digunakan dalam memprediksi kelas dari objek yang labelnya belum diketahui. Penerapan klasifikasi banyak dijumpai dalam berbagai bidang, seperti deteksi kecurangan atau penipuan, penentuan target pemasaran, prediksi kinerja, industri manufaktur, hingga diagnosis kesehatan.Pada proses data umumnya melalui dua tahap. Pertama, tahap pembelajaran (learning phase), di mana algoritma klasifikasi menganalisis data latih (training data) untuk menghasilkan model klasifikasi. Kedua, tahap pengujian klasifikasi, di mana data uji (testing data) digunakan untuk memprediksi serta mengukur tingkat akurasi dari model yang telah dibangun(Gori et al., 2024).

J. Dataset

Dataset adalah kumpulan data yang saling berkaitan dalam satu kesatuan dan umumnya digunakan untuk kasus tertentu, seperti data medis, komentar pengguna media sosial, curah hujan, atau pergerakan harga emas. Dataset berperan penting sebagai sumber informasi valid untuk penelitian maupun pengembangan sistem, misalnya dalam machine learning, klasifikasi, klasterisasi, dan analisis sentimen.Sebuah dataset yang baik ditandai dengan data yang lengkap, konsisten, selalu diperbarui, memiliki variabel yang jelas, bebas dari noise, serta mudah dipahami sehingga dapat mendukung hasil penelitian atau aplikasi yang akurat dan relevan(Ramadani et al., 2024).

K. Teknik Resample

Resampling merupakan teknik preprocessing yang digunakan untuk menyeimbangkan kembali distribusi data, khususnya dalam menghadapi permasalahan ketidakseimbangan kelas pada data pelatihan (Kurniawan et al., 2023). Metode ini bertujuan untuk mengurangi dampak negatif dari distribusi kelas yang tidak merata dengan menyesuaikan jumlah sampel dari masing-masing kelas. Proses ini dilakukan melalui berbagai algoritma sampling, sehingga dihasilkan dataset baru yang lebih seimbang, yang kemudian digunakan untuk melatih model klasifikasi secara lebih optimal (Kurniawan et al., 2023).

L. Data Mining

Data mining dan analisis kebutuhan sistem merupakan dua elemen yang saling terkait dalam konteks pengembangan sistem informasi, masing-masing berperan sesuai dengan ruang lingkupnya. Pemanfaatan teknologi serta optimalisasi sumber daya yang tersedia menjadi faktor krusial yang harus diperhatikan oleh pihak yang terlibat dalam pengolahan data. Dalam perjalanannya, data mining telah berkembang dengan berbagai definisi yang beragam, yang secara umum memperkaya khazanah pengetahuan di bidang teknologi informasi. (Luhur et al., 2019) Secara konseptual, data mining didefinisikan sebagai suatu rangkaian proses sistematis untuk menemukan informasi bernilai yang sebelumnya tidak diketahui secara eksplisit dalam basis data. Informasi ini diperoleh melalui proses ekstraksi dan identifikasi pola-pola signifikan yang tersembunyi di dalam kumpulan data. (Kurniawan et al., 2023).

3. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan mulai dari pra-proses data, klasifikasi, validasi dan akurasi

langkah-langkah pemikiran di atas, dapat dijabarkan sebagai berikut :

- a. Dataset terlebih dahulu di uji dengan beberapa model algoritma klasifikasi yaitu Decision Tree J48, Random Forest, Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K- Nearest Neighbour (KNN) menggunakan 10-fold cross validation.
- b. Dataset diuji kembali menggunakan teknik preprocessing data yaitu resample dan diuji kembali menggunakan algoritma klasifikasi yaitu Decision Tree J48, Random

- Forest, Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbour (KNN) menggunakan 10-fold cross validation.
- c. Berdasarkan evaluasi pengujian akan diperoleh nilai akurasi, dimana evaluasi yang digunakan menggunakan enam measurement yaitu berdasarkan evaluasi Confusion matrix, Akurasi, Kappa Statistic, Mean Absolute Error, Precision, Recall dan ROC.
 - d. Langkah selanjutnya yaitu perbandingan algoritma klasifikasi yaitu Decision Tree J48, Random Forest, Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K- Nearest Neighbour (KNN) dengan algoritma klasifikasi yaitu Decision Tree J48, Random Forest, Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K- Nearest Neighbour (KNN) dengan Teknik resample.
 - e. Setelah melakukan perbandingan algoritma maka didapat hasil pengujian Confusion matrix, Akurasi, Kappa Statistic, Mean Absolute Error, Precision, Recall dan ROC.

4. HASIL DAN PEMBAHASAN

A. Dataset

Penelitian ini memanfaatkan dataset HRM. Pada tahap preprocessing dilakukan pemeriksaan terhadap beberapa ketentuan awal yang harus diperhatikan, seperti atribut jumlah, modul jumlah, serta jumlah cacat yang terdapat pada masing-masing dataset. Adapun rincian dataset penelitian yang dipakai adalah sebagai berikut.

Tabel 1. rincian dataset penelitian.

Atribut	Description
<i>Age</i>	Umur
<i>Gender</i>	Male, Female
<i>Education Background</i>	Latar Belakang Pendidikan
<i>Marital Status</i>	Status Pernikahan
<i>Emp Department</i>	-
<i>Emp Job Role</i>	Peran Pekerjaan
<i>Business Travel</i>	Frekuensi Perjalanan Bisnis
<i>Frequency</i>	-
<i>Distance From Home</i>	Jarak Dari Rumah
<i>Emp Education Level</i>	Tingkatan Pendidikan

<i>Emp Environment Satisfaction</i>	Kepuasan Lingkungan
<i>Emp Hourly Rate</i>	Tarif per Jam
<i>Emp Job Involvement</i>	Keterlibatan Pekerjaan

B. Preprocessing Data

Tahap selanjutnya adalah melakukan pengujian dataset dapat menggunakan algoritma klasifikasi yang diantaranya *Random Forest*, *Decision Tree J48*, *KNN*, *Naive Bayes*, *Logistic Regression*, dan *SVM*. Hasil dari keenam algoritma tersebut kemudian dievaluasi dengan menggunakan *Confusion Matrix*, *Akurasi*, *Kappa Statistic*, *Mean Absolute Error*, *Precision*, *Recall*, serta *ROC*. Pengujian algoritma dilakukan dengan bantuan aplikasi **WEKA (Waikato Environment for Knowledge Analysis)** adalah perangkat lunak data mining berbasis Java yang bersifat open source dan banyak digunakan untuk keperluan analisis data serta penerapan berbagai algoritma pembelajaran mesin.

C. Preprocessing Data Sebelum di Resample

Tampilan awal saat proses preprocessing data sebelum di resample, terlihat pada gambar pada grafik batang dibawah bahwa label luar biasa memiliki *count* tertinggi dibanding dua label lainnya sehingga data tersebut harus dibalance kan terlebih dahulu.

D. Proses Preprocessing Data di Resample

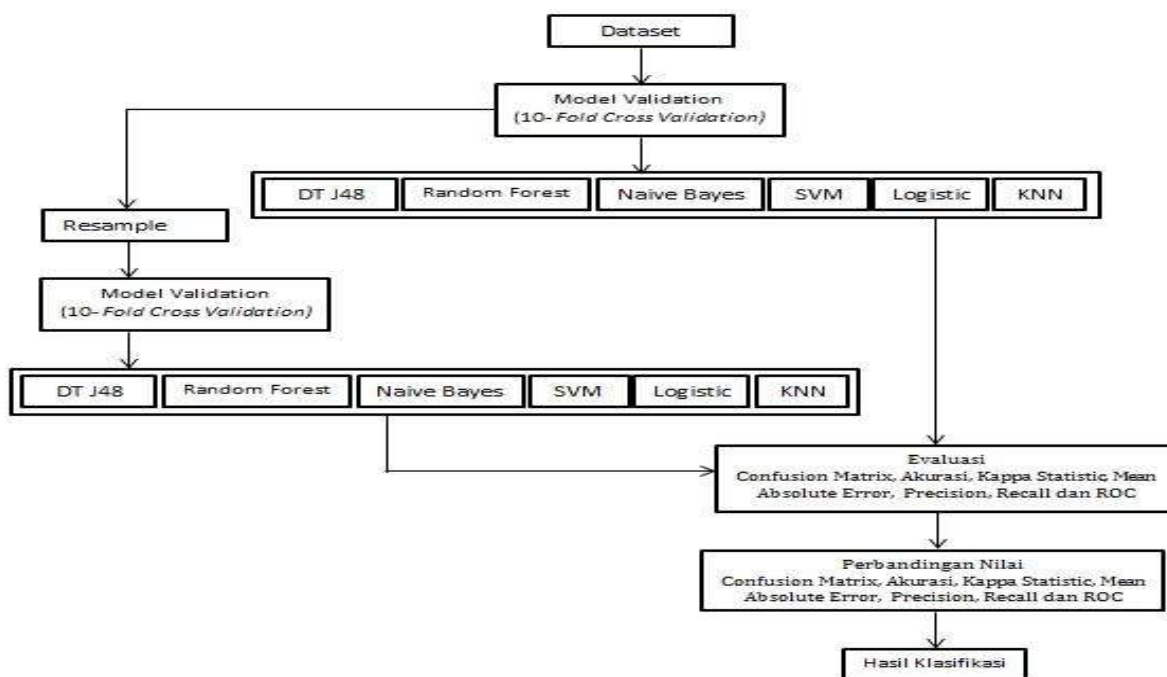
Pada halaman antarmuka di weka langkah yang dilakukan setelah dataset yang akan diuji sudah dibuka selanjutnya adalah di menu filter kita pilih button choose untuk memilih teknik resample didalam folder instance kita klik



Gambar 1. Proses Preprocessing Data di Resample.

E. Eksperimen serta Evaluasi Model

Pada tahap eksperimen dan pengujian model, teknik data mining diterapkan dengan terlebih dahulu menentukan metode yang tepat serta memilih algoritma yang akan digunakan dalam proses klasifikasi. Pada penelitian ini, perangkat lunak yang digunakan adalah WEKA versi 3.8.3. Tujuan dari pengujian ini adalah mengelompokkan data ke dalam tiga kategori, yaitu **Luar Biasa**, **Sangat Luar Biasa**, dan **Baik**, menggunakan beberapa algoritma klasifikasi seperti **Decision Tree J48**, **Random Forest**, **Naive Bayes**, **K-Nearest Neighbors (KNN)**, **Logistic Regression**, serta **Support Vector Machine (SVM)**.



Gambar 2. Eksperimen serta Evaluasi Model.

F. Pengujian Model Klasifikasi

Penulis melakukan pengujian pada dataset HRM dalam bagian ini dengan menggunakan model klasifikasi

Tabel 2. Hasil *Confusion Matrix* Algoritma *Decision Tree J48*.

a	b	c	Classified as
854	3	17	a = Luar Biasa
18	109	5	b = Sangat Luar Biasa
11	1	182	c = Baik

Tabel 3. Hasil *Confusion Matrix* Algoritma *Random Forest*.

a	b	c	Classified as
861	1	12	a = Luar Biasa
27	104	1	b = Sangat Luar Biasa
19	0	175	c = Baik

Tabel 4. Hasil *Confusion Matrix* Algoritma *Naive Bayes*.

a	b	c	Classified as
786	28	60	a = Luar Biasa
46	81	5	b = Sangat Luar Biasa
97	2	95	c = Baik

Tabel 5. Hasil *Confusion Matrix* Algoritma *Support Vector Machine (SVM)*.

a	b	c	Classified as
792	13	64	a = Luar Biasa
39	91	2	b = Sangat Luar Biasa
91	0	103	c = Baik

Tabel 6. Hasil *Confusion Matrix* Algoritma *Logistic*.

a	b	c	Classified as
786	24	64	a = Luar Biasa
35	93	4	b = Sangat Luar Biasa
87	5	102	c = Baik

Tabel 7. Hasil *Confusion Matrix* Algoritma *K-Nearest Neighbor*.

a	b	c	Classified as
809	27	38	a = Luar Biasa
37	91	4	b = Sangat Luar Biasa
43	7	144	c = Baik

Tabel 8. Perbandingan Hasil Kinerja Klasifikasi (Decision Tree , Random Forest, SVM, Logistic Regression, Naive Bayes serta KNN).

Algoritma	Akurasi	Kappa Statistic	MAE	Precision	Recall	ROC
<i>DT J48</i>	92,83%	0,8316	0,067	0,928	0,93	0,93
<i>Random Forest</i>	90,41%	0,759	0,167	0,906	0,9	0,94
<i>Naive Bayes</i>	79,16%	0,4776	0,169	0,783	0,79	0,86
<i>SVM</i>	82,33%	0,5562	0,266	0,817	0,82	0,78
<i>Logistic</i>	81,25%	0,5522	0,159	0,809	0,81	0,91
<i>KNN</i>	66,91%	0,1755	0,221	0,646	0,67	0,58

G. Pengujian Model Klasifikasi dengan Resample

Tabel 9. Hasil *Confusion Matrix* Algoritma *Decision Tree J48* dengan *Resample*.

a	b	c	Classified as
854	3	17	a = Luar Biasa
18	109	5	b = Sangat Luar Biasa
11	1	182	c = Baik

Tabel 10. Hasil *Confusion Matrix* Algoritma *Random Forest* dengan *Resample*.

a	b	c	Classified as
861	1	12	a = Luar Biasa
27	104	1	b = Sangat Luar Biasa
19	0	175	c = Baik

Tabel 11. Hasil *Confusion Matrix* Algoritma *Naive Bayes* dengan *Resample*.

a	b	c	Classified as
786	28	60	a = Luar Biasa
46	81	5	b = Sangat Luar Biasa
97	2	95	c = Baik

Tabel 12. Hasil *Confusion Matrix* pada Klasifikasi Menggunakan *SVM* dengan *Resample*.

a	b	c	Classified as
792	13	64	a = Luar Biasa
39	91	2	b = Sangat Luar Biasa
91	0	103	c = Baik

Tabel 13. Hasil *Confusion Matrix* Algoritma *Logistic* dengan *Resample*.

a	b	c	Classified as
786	24	64	a = Luar Biasa
35	93	4	b = Sangat Luar Biasa
87	5	102	c = Baik

Tabel 14. Hasil *Confusion Matrix* Algoritma *K-Nearest Neighbor* dengan *Resample*.

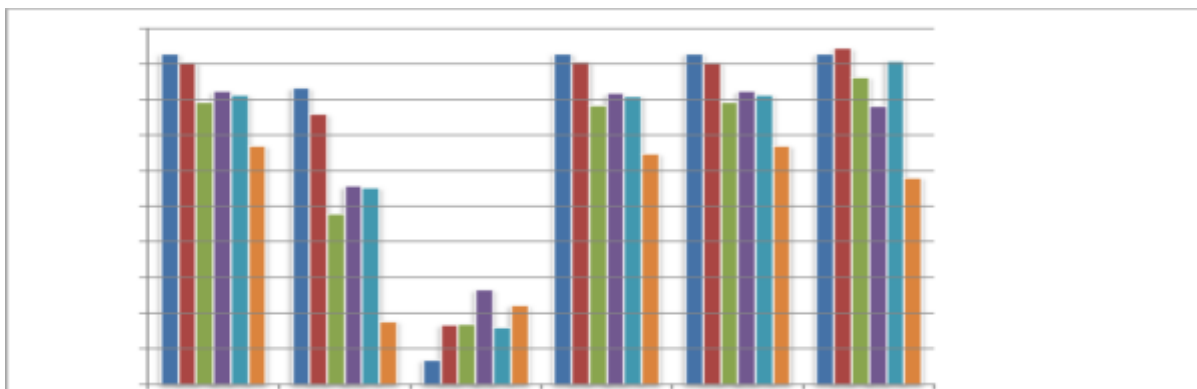
a	b	c	Classified as
809	27	38	a = Luar Biasa
37	91	4	b = Sangat Luar Biasa
43	7	144	c = Baik

H. Pengujian klasifikasi dengan Resample dan Feature Selection Algoritm

Tabel 15. Hasil Kinerja Pengujian Klasifikasi dengan *Resample*.

Algoritma	Akurasi	Kappa Statistic	MAE	Precision	Recall	ROC
<i>DT J48</i>	95,42%	0,8925	0,043	0,955	0,95	0,96
<i>Random Forest</i>	95%	0,8794	0,099	0,951	0,95	0,99
<i>Naive Bayes</i>	80,16%	0,5096	0,164	0,792	0,8	0,87
<i>SVM</i>	82,58%	0,57	0,2644	0,821	0,83	0,79
<i>Logistic</i>	81,75%	0,56	0,151	0,811	0,82	0,89
<i>KNN</i>	87%	0,6933	0,087	0,868	0,87	0,85

Hasil pengujian kinerja klasifikasi dengan teknik *Resample* yang ditampilkan pada memperlihatkan performa beberapa algoritma, yaitu *Decision Tree J48*, *Naive Bayes*, *Random Forest*, *Logistic Regression*, *KNN* serta *Support Vector Machine (SVM)*. Dari hasil tersebut, algoritma dengan akurasi tertinggi adalah *Decision Tree J48* dengan capaian sebesar **95,42%**.



Gambar 3. Kinerja Hasil Pengujian Klasifikasi *Decision Tree J48*, *Random Forest*, *Naive Bayes*, *KNN*, *Logistic*, dan *SVM* Menggunakan Teknik *Resample*.

I. Pengujian klasifikasi dengan Resample dan Feature Selection Algoritm

Tabel 16. Pengujian klasifikasi dengan *Resample* dan *Feature Selection Algoritm*.

0.32647	17	EmpLastSalaryHikePercent
0.3014	10	EmpEnvironmentSatisfaction
0.149	24	YearsSinceLastPromotion
0.1307	23	ExperienceYearsInCurrentRole
0.12278	25	YearsWithCurrManager
0.09055	5	EmpDepartment
0.07032	22	ExperienceYearsAtThisCompany
0.05964	6	EmpJobRole
0.05723	21	EmpWorkLifeBalance
0.05131	9	EmpEducationLevel
0.04865	16	OverTime
0.04368	13	EmpJobLevel
0.04188	18	EmpRelationshipSatisfaction
0.04009	14	EmpJobSatisfaction
0.03873	26	Attrition
0.03827	20	TrainingTimesLastYear
0.03801	4	MaritalStatus
0.03622	19	TotalWorkExperienceInYears
0.035	3	EducationBackground
0.02553	15	NumCompaniesWorked
0.02353	7	BusinessTravelFrequency
0.01902	8	DistanceFromHome
0.0177	1	Age
0.01611	12	EmpJobInvolvement
0.01476	11	EmpHourlyRate
0.00854	2	Gender
0.01476	11	EmpHourlyRate
0.00854	2	Gender

Berdasarkan hasil dari pengujian menggunakan correlation attribute eval didapatkan beberapa attribut teratas yang paling berpengaruh dalam memprediksi peringkat kinerja karyawan yaitu :

1. *Emp Last Salary Hike Percent,*
2. *Emp Environment Satisfaction,*
3. *Years Since Last Promotion,*
4. *Experience Years In Current Role.*

5. KESIMPULAN

Penelitian ini berhasil membangun model prediksi kinerja pegawai dengan menerapkan beberapa algoritma klasifikasi, yaitu Decision Tree J48, Random Forest, Naive Bayes, KNN, Logistic Regression, dan Support Vector Machine (SVM). Proses pengujian dilakukan dengan menggunakan teknik Resample serta Feature Selection untuk meningkatkan kualitas data dan menyeimbangkan distribusi kelas. Berdasarkan hasil evaluasi, Decision Tree J48 menunjukkan performa terbaik dengan tingkat akurasi sebesar 95,42%, nilai kappa 0,8925, MAE 0,043, precision 0,955, recall 0,95, serta ROC 0,96. Hasil ini membuktikan bahwa Decision Tree (J48) paling efektif dalam memprediksi kinerja pegawai dibandingkan algoritma lainnya. Selain itu, melalui penerapan Feature Selection, ditemukan bahwa atribut yang paling berpengaruh dalam memprediksi kinerja pegawai adalah Emp Last Salary Hike Percent, Emp Environment Satisfaction, Years Since Last Promotion, dan Experience Years In Current Role. Dengan demikian, model yang dihasilkan tidak hanya mampu memberikan akurasi tinggi, tetapi juga dapat dijadikan sebagai alat bantu pengambilan keputusan manajerial dalam mengevaluasi dan meningkatkan kinerja pegawai di masa mendatang.

DAFTAR REFERENSI

- A'yuniyah, Q. A., & Reza, M. (2023). Penerapan algoritma K-Nearest Neighbor untuk klasifikasi jurusan siswa di SMA Negeri 15 Pekanbaru. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 3(1), 39–45. <https://doi.org/10.57152/ijirse.v3i1.484>
- Abdillah, M. A., Setyanto, A., & Sudarmawan, S. (2020). Implementasi decision tree algoritma C4.5 untuk memprediksi kesuksesan pendidikan karakter. *Respati*, 15(2), 59. <https://doi.org/10.35842/jtir.v15i2.349>
- Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data preprocessing patterns in the assessment of teacher education program students. *Journal of Informatics and Computer Science*, 5, 97–100.
- Diana, D., Indrajit, R. E., & Dazki, E. (2022). Komparasi algoritma Naïve Bayes, logistic regression dan support vector machine pada klasifikasi file application package kit Android malware. *JUTISI: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 11(1), 109. <https://doi.org/10.35889/jutisi.v11i1.815>
- Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing data dan klasifikasi untuk prediksi kinerja akademik siswa. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- Hamami, F., & Dahlan, I. A. (2022). Klasifikasi cuaca Provinsi DKI Jakarta menggunakan algoritma random forest dengan teknik oversampling. *Jurnal Teknoinfo*, 16(1), 87. <https://doi.org/10.33365/jti.v16i1.1533>

- Ibrahim, N. H., & Khikmah, L. (2024). Perbandingan metode algoritma C4.5, Naïve Bayes, dan logistic regression untuk penentuan kelayakan penerima kredit. *Teknologi*, 14(2), 85–93. <https://doi.org/10.26594/teknologi.v14i2.4650>
- Kurniawan, I., Buani, D. C. P., Abdussomad, A., Apriliah, W., & Saputra, R. A. (2023). Implementasi algoritma random forest untuk menentukan penerima bantuan raskin. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(2), 421–428. <https://doi.org/10.25126/jtiik.20236225>
- Luhur, U. B., Utara, P., Pembaharuan, P., & Asuransi, P. (2019). Data mining klasifikasi untuk memprediksi status keberlanjutan polis asuransi kesehatan dengan algoritme Naïve Bayes. *Jurnal Data Mining*, 3(10), 219–223.
- Pahlevi, O., Amrin, A., & Handrianto, Y. (2023). Implementasi algoritma klasifikasi random forest untuk penilaian kelayakan kredit. *Jurnal Infortech*, 5(1), 71–76. <https://doi.org/10.31294/infortech.v5i1.15829>
- Putri Latif, D., & Ali, H. (2025). Pengaruh pengambilan keputusan, investasi teknologi informasi dan pengembangan SDM terhadap efisiensi operasional. *Jurnal Komunikasi dan Ilmu Sosial*, 3(1), 1–10. <https://doi.org/10.38035/jkis.v3i1.1724>
- Ramadani, P., Fadillah, R., Adawiyah, Q., & Al Ghazali, B. R. (2024). Perbandingan algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk klasifikasi kelayakan program keluarga harapan. *Jurnal Media Informatika*, 6(2), 775–782.
- Susanti, Y., Choyyin, M. G., Priyatna, A., & Lestari, S. (2023). Perbandingan penerapan algoritma decision tree C4.5 dan Naïve Bayes dalam analisa kelulusan siswa pada SMK Swadhipa 2 Natar Kabupaten Lampung Selatan. *Jurnal SIMADA (Sistem Informasi dan Manajemen Basis Data)*, 6(2), 117–123. <https://doi.org/10.30873/simada.v6i2.3772>
- Umri, S. A. (2021). Analisis dan komparasi algoritma klasifikasi dalam indeks pencemaran udara di DKI Jakarta. *JIKO (Jurnal Informatika dan Komputer)*, 4(2), 98–104. <https://doi.org/10.33387/jiko.v4i2.2871>
- Wijiyanto, W., Pradana, A. I., Sopingi, S., & Atina, V. (2024). Teknik K-Fold cross validation untuk mengevaluasi kinerja mahasiswa. *Jurnal Algoritma*, 21(1), 239–248. <https://doi.org/10.33364/algoritma/v.21-1.1618>