

Prediksi Faktor-Faktor Yang Mempengaruhi Keputusan Karyawan Meninggalkan Perusahaan Menggunakan Algoritma Random Forest

Rama Aziz Wibowo*¹, Dedi Gunawan²

^{1,2}Teknik Informatika, Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta, Indonesia
Email: ¹1200210263@student.ums.ac.id, ²dg163@ums.ac.id

Abstrak

Keputusan karyawan untuk mengundurkan diri dari perusahaan merupakan masalah penting yang dapat mempengaruhi kinerja dan keberlangsungan operasional perusahaan. Memahami faktor-faktor yang mendorong pengunduran diri karyawan sangat penting untuk merancang strategi retensi yang efektif, mengurangi tingkat turnover, dan meningkatkan kepuasan kerja. Penelitian ini bertujuan untuk memprediksi faktor-faktor yang mempengaruhi keputusan karyawan dalam mengundurkan diri dari perusahaan dengan menggunakan algoritma Random Forest, serta mengembangkan model prediksi yang dapat memberikan wawasan bagi perusahaan dalam meningkatkan strategi retensi karyawan. Metode yang digunakan dalam penelitian ini adalah algoritma Random Forest, yang diterapkan pada dataset publik yang diambil dari Kaggle. Proses evaluasi model dilakukan dengan menggunakan metrik akurasi, precision, recall, dan F1-score untuk mengukur kinerja model dalam memprediksi pengunduran diri karyawan. Hasil penelitian menunjukkan bahwa algoritma Random Forest berhasil mencapai akurasi prediksi sebesar 85%. Faktor utama yang mempengaruhi pengunduran diri karyawan adalah tahun bergabung, usia, dan kota tempat bekerja. Dampak penelitian ini diharapkan dapat membantu perusahaan dalam merancang strategi retensi karyawan yang lebih efektif dan berbasis data.

Kata Kunci: *Machine Learning, Prediksi Retensi Karyawan, Random Forest.*

Prediction of Factors Influencing Employee Decisions to Leave the Company Using the Random Forest Algorithm

Abstract

Employee decisions to resign from a company are an important issue that can affect the performance and sustainability of the company's operations. Understanding the factors that drive employee resignation is essential to designing effective retention strategies, reducing turnover rates, and increasing job satisfaction. This study aims to predict the factors that influence employee decisions to resign from the company using the Random Forest algorithm, as well as to develop a prediction model that can provide insight for companies in improving employee retention strategies. The method used in this study is the Random Forest algorithm, which is applied to a public dataset taken from Kaggle. The model evaluation process is carried out using accuracy, precision, recall, and F1-score metrics to measure the performance of the model in predicting employee resignation. The results of the study show that the Random Forest algorithm managed to achieve a prediction accuracy of 85%. The main factors influencing employee resignation are year of joining, age, and city of work. The impact of this study is expected to help companies design more effective and data-based employee retention strategies.

Keywords: *Employee Retention Prediction, Machine Learning, Random Forest.*

1. PENDAHULUAN

Pergantian karyawan merupakan topik yang diteliti secara kritis dalam sebuah perusahaan, yang menarik perhatian besar atas perannya dalam menentukan alasan karyawan meninggalkan perusahaan. Faktor utama dalam pergantian karyawan dan kepuasan kerja, yang penting untuk memahami masalah ini. Pengaruh budaya organisasi dan praktik sumber daya manusia dalam mengurangi niat untuk berpindah juga telah diteliti, yang menggarisbawahi peran vital mereka dalam manajemen pergantian[1], [2], [3]. Jumlah karyawan mengundurkan diri yang tinggi pada perusahaan dapat menimbulkan kemunduran bisnis, oleh karena itu penting untuk perusahaan mengetahui alasan karyawan *resign* dari perusahaan[4], [5]. Dengan mengetahui faktor-faktor pengunduran diri

karyawan, perusahaan dapat mencegah terjadinya pengunduran diri karyawan dan mengatur strategi bagaimana mengatur SDM di perusahaan[6], [7].

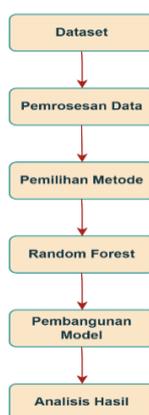
Faktor-faktor yang mempengaruhi keputusan karyawan untuk meninggalkan perusahaan sangat beragam. Faktor demografis seperti usia dan jenis kelamin, faktor organisasi seperti gaji, jabatan, dan lama bekerja, serta faktor lingkungan kerja seperti budaya perusahaan dan hubungan dengan rekan kerja, semuanya berperan dalam mempengaruhi keputusan ini. Dengan menganalisis faktor-faktor ini, perusahaan dapat mengidentifikasi area-area yang memerlukan perhatian khusus dan mengembangkan strategi yang efektif untuk meningkatkan kepuasan dan keterlibatan karyawan.

Perkembangan teknologi saat ini sangat pesat, banyak bidang sudah menggunakan teknologi untuk meringankan pekerjaan ataupun memberikan solusi[8]. Keputusan dalam penggunaan teknologi dalam bidang manajemen SDM, menggunakan teknik analisis *prediktif* yang canggih dipastikan akan sangat meningkatkan kinerja dari manajemen SDM. Analisis yang dilakukan pada data SDM dapat berupa *deskriptif analitik*, *prediktif* dan korelasi yang digabungkan dengan teknik *machine learning* seperti *Random Forest* untuk pengambilan sebuah keputusan[9]. Metode algoritma *Random Forest* memungkinkan perusahaan untuk menganalisis data historis karyawan dan mengidentifikasi pola-pola yang dapat digunakan untuk memprediksi keputusan karyawan[10].

Penelitian ini bertujuan untuk memprediksi faktor-faktor yang mempengaruhi karyawan mengundurkan diri dari perusahaan, mengembangkan analisis data dengan menggunakan algoritma *Machine Learning* seperti *Random Forest*. Algoritma ini dipilih karena kemampuannya dalam menangani data kompleks dan menghasilkan prediksi yang lebih akurat dibandingkan dengan teknik tradisional. Selain itu, penelitian ini juga bertujuan untuk mengembangkan model yang dapat membantu perusahaan memprediksi risiko pengunduran diri karyawan dan merancang strategi *retensi* yang lebih baik.

2. METODE PENELITIAN

Ada berbagai algoritma *machine learning*, pada penelitian ini menggunakan algoritma *random forest* untuk memprediksi *retensi* karyawan [11], [12]. Mengintegrasikan analisis teori sumber daya manusia dengan *Random Forest* untuk mengidentifikasi faktor-faktor utama yang mempengaruhi *turnover* karyawan, dengan menggunakan dataset yang mencakup demografis, seperti usia, jenis kelamin, tingkat gaji, dan faktor lingkungan kerja seperti budaya perusahaan[13]. Penelitian ini menggunakan pendekatan kuantitatif dengan penerapan algoritma *machine learning*, khususnya *Random Forest*, untuk memprediksi pengunduran diri karyawan. Pendekatan ini dipilih karena kemampuan *Random Forest* dalam menangani data kompleks dan memberikan hasil yang akurat dalam klasifikasi data, terutama dalam memprediksi keputusan pengunduran diri karyawan[14].



Gambar 1. *Flowchart* Penelitian

Penelitian ini bertujuan untuk menganalisis faktor yang mempengaruhi karyawan mengundurkan diri dari perusahaan, mengembangkan analisis data dengan menggunakan algoritma *Machine Learning* seperti *Random Forest*. Untuk mencapai tujuan tersebut, metode penelitian akan dibagi menjadi beberapa tahapan seperti pada gambar 1:

2.1. Dataset

Dataset yang digunakan dalam penelitian ini diambil dari sumber publik di Kaggle yang disediakan oleh pengguna Tawfik Elmetwally. Dataset ini mencakup informasi demografis dan pekerjaan karyawan, termasuk variabel seperti:

- **Usia:** Umur karyawan.
- **Tahun Bergabung:** Tahun karyawan mulai bekerja di perusahaan.
- **Kota:** Lokasi tempat karyawan bekerja.
- **Gaji:** Tingkat gaji karyawan.
- **Jenis Kelamin:** Jenis kelamin karyawan.
- **Pengalaman Kerja:** Jumlah tahun pengalaman kerja yang dimiliki karyawan.
- **Status Keaktifan:** Apakah karyawan pernah tidak aktif atau cuti sementara.
- **Keputusan Pengunduran Diri:** Status pengunduran diri karyawan (keluar atau tetap).

2.2. Pemrosesan Data

Sebelum melakukan analisis, dataset harus diproses terlebih dahulu untuk memastikan kualitas data yang optimal. Langkah-langkah yang dilakukan dalam pemrosesan data ini adalah:

a. *Data Cleaning*

Menghapus data yang tidak lengkap atau duplikat untuk memastikan akurasi analisis[15], [16].

b. *Label Encoding*

Mengubah variabel kategorikal (misalnya jenis kelamin dan kota) menjadi format numerik agar bisa digunakan dalam model *machine learning*[17].

c. *Splitting Data*

Membagi dataset menjadi dua bagian: 80% data untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*).

2.3. Pemilihan Metode

Metode yang digunakan dalam penelitian ini adalah Algoritma *Random Forest*. Cara kerja *Random Forest* termasuk dalam kategori *ensemble learning* karena memanfaatkan kombinasi dari beberapa pohon keputusan sebagai *classifier* dasar untuk memprediksi nilai dari suatu data[18]. Algoritma *Random Forest* menunjukkan beberapa kelebihan diantaranya mampu menghasilkan error yang relative rendah, performa yang baik dalam klasifikasi, dapat mengatasi data pelatihan dalam jumlah besar secara efisien, serta metode yang efektif untuk mengestimasi missing data. *Random Forest* menghasilkan banyak pohon independen dengan subset yang dipilih secara acak melalui bootstrap dari sampel pelatihan dan dari variable input disetiap node[19].

Random Forest melakukan serangkaian perhitungan matematis untuk membangun model klasifikasi yang dapat memprediksi apakah seorang karyawan akan mengundurkan diri atau tidak. *Random Forest* bekerja dengan membangun banyak pohon keputusan dan menggabungkan hasilnya menggunakan metode *voting mayoritas* (untuk klasifikasi). Pemilihan fitur dalam setiap node pada pohon didasarkan pada pengukuran *impurity* seperti *Gini Impurity* atau *Entropy*. Untuk pengukuran *Gini Impurity* dapat dilihat pada persamaan (1) dan (2).

$$\text{Gini}(t) = 1 - \sum_{i=1}^c p_i^2 \quad (1)$$

di mana p_i adalah proporsi dari kelas ke- i di node t . Untuk prediksi akhir, hasil voting dari semua pohon digunakan.

$$\hat{y} = \text{mode} (\{h_1(x), h_2(x), \dots, h_n(x)\}) \quad (2)$$

di mana $h_1(x)$ adalah prediksi dari pohon ke- i , dan \hat{y} adalah prediksi final. Selain itu, algoritma juga menghitung *feature importance* berdasarkan penurunan rata-rata impurity pada setiap fitur yang digunakan[20].

2.4. Pembangunan Model

Model prediksi dibangun menggunakan algoritma *Random Forest* dengan parameter:

a. *n_estimators*

Menentukan jumlah pohon keputusan yang akan dibangun dalam model. Nilai yang digunakan adalah 100 pohon keputusan untuk memastikan model memiliki stabilitas yang baik.

b. *random_state*

Digunakan untuk memastikan replikasi hasil yang konsisten setiap kali model dijalankan. Nilai yang digunakan adalah 42 untuk memastikan hasil yang dapat diulang.

c. *Evaluasi Model*

Kinerja model dievaluasi dengan menggunakan beberapa metrik untuk mengukur akurasi prediksi, termasuk:

- **Akurasi:** Persentase prediksi yang benar dari total data.
- **Precision:** Mengukur ketepatan prediksi karyawan yang mengundurkan diri.
- **Recall:** Mengukur seberapa baik model dalam mendeteksi karyawan yang benar-benar mengundurkan diri.
- **F1-Score:** Metrik yang menggabungkan precision dan recall untuk memberikan gambaran yang lebih baik tentang keseimbangan antara keduanya.

2.5. Analisis Hasil

Menganalisis hasil model untuk mengidentifikasi faktor-faktor yang paling berpengaruh dalam keputusan karyawan meninggalkan perusahaan dan untuk mengembangkan model yang dapat membantu perusahaan memprediksi risiko pengunduran diri karyawan dan merancang strategi *retensi* yang lebih baik.

3. HASIL DAN PEMBAHASAN

3.1. Penjelasan Hasil Tabel

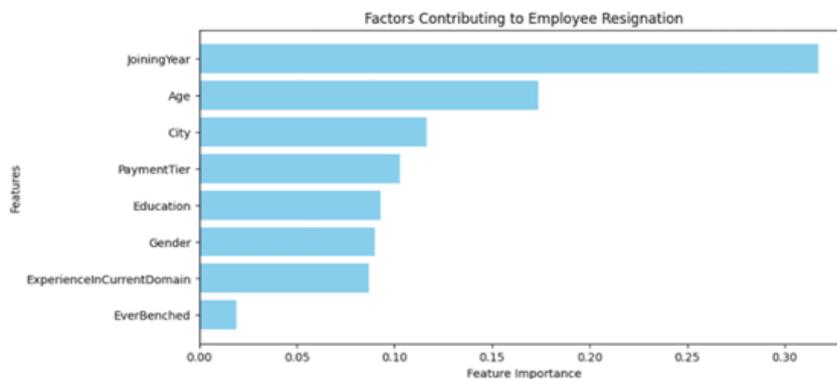
Tabel 1. Kinerja Dataset berdasarkan Algoritma Random Forest dan Decision Tree

Dataset	Algoritma	Keterangan	Presisi	Recall	F1- Score	Akurasi
Classification	Random Forest	Karyawan Resign	82%	72%	77%	85%
		Karyawan Tidak Resign	86%	92%	89%	
Classification	Decision Tree	Karyawan Resgin	79%	73%	76%	84%
		Karyawan Tidak Resign	86%	89%	88%	

Pada tabel diatas menunjukkan perbandingan kinerja dua algoritma, yaitu *Random Forest* dan *Decision Tree*, dalam memprediksi dataset karyawan apakah karyawan akan mengundurkan diri dari perusahaan atau tidak. Berdasarkan hasil perhitungan dua algoritma tersebut, *Random Forest* memiliki kinerja yang lebih baik secara keseluruhan dibandingkan dengan *Decision Tree*. *Random Forest* memiliki akurasi 85%, dengan hasil precision untuk karyawan yang *resign* sebesar 82%, recall sebesar 72%, dan F1-Score sebesar 77%. Untuk karyawan yang tidak *resign*, precision mencapai 86%, recall 92%, dan F1-Score 89%. Sementara itu, *Decision Tree* menunjukkan akurasi sedikit lebih rendah, yaitu 84%, dengan precision untuk karyawan *resign* sebesar 79%, recall 73%, dan F1-Score 76%. Untuk karyawan yang tidak *resign*, precision adalah 86%, recall 89%, dan F1-Score 88%. *Random Forest* memberikan hasil yang lebih seimbang dan lebih akurat, terutama dalam mengidentifikasi karyawan yang mengundurkan diri dari perusahaan

3.2. Penjelasan Hasil Grafik

Pada Gambar 2 menunjukkan hasil dari analisis data karyawan menggunakan algoritma machine learning model Random Forest menghasilkan dari total 4.653 (100%) data karyawan, 1.600 karyawan (34,39%) mengundurkan diri dan 3.053 karyawan (65,61%) tetap bekerja pada perusahaan. Faktor yang paling berpengaruh dalam pengunduran diri karyawan pada perusahaan yang pertama adalah tahun bergabung, kedua umur, ketiga kota tempat bekerja, keempat tingkat gaji, kelima tingkat pendidikan, keenam jenis kelamin, ketujuh pengalaman dibidangnya, dan kedelapan apakah karyawan pernah tidak diberikan tugas atau pekerjaan.



Gambar 2. Grafik Faktor Penyebab Pengunduran diri karyawan dari perusahaan

Faktor tahun bergabung berperan penting karena karyawan yang telah bekerja lebih lama mungkin lebih cenderung merasa jenuh atau mencapai titik keputusan untuk mencari peluang lain. Selain itu, usia juga berhubungan dengan perubahan prioritas dalam hidup, di mana karyawan yang lebih tua mungkin lebih memilih stabilitas atau perubahan karier yang lebih signifikan. Faktor kota tempat bekerja mencerminkan faktor geografis yang dapat mempengaruhi kenyamanan kerja, biaya hidup, dan kesempatan karir di daerah tersebut.

Hasil ini memiliki hal penting bagi perusahaan dalam merancang strategi *retensi* karyawan. Dengan mengetahui faktor-faktor yang paling berpengaruh, perusahaan dapat fokus pada kebijakan yang menargetkan karyawan yang berada dalam kelompok yang berisiko tinggi untuk keluar, seperti mereka yang baru bergabung atau yang berada di lokasi dengan tingkat *turnover* yang tinggi. Misalnya, perusahaan bisa memperkenalkan program pengembangan karier atau kebijakan kesejahteraan yang lebih baik untuk karyawan yang lebih tua, atau menawarkan insentif khusus untuk karyawan yang bekerja di kota-kota dengan tingkat pengunduran diri yang tinggi.

4. KESIMPULAN

Penelitian ini memprediksi faktor-faktor utama yang mempengaruhi keputusan karyawan untuk mengundurkan diri, yaitu tahun bergabung, usia, dan kota tempat bekerja. Model Random Forest yang diterapkan berhasil mencapai akurasi sebesar 85% dalam memprediksi pengunduran diri karyawan. Hasil ini menunjukkan bahwa perusahaan harus lebih memperhatikan karyawan yang baru bergabung, yang lebih tua, atau yang bekerja di kota-kota dengan tingkat turnover yang tinggi. Berdasarkan temuan ini, perusahaan dapat merancang strategi *retensi* yang lebih tepat sasaran, seperti program pengembangan karier, kebijakan kesejahteraan, atau insentif khusus untuk kelompok-kelompok karyawan tersebut. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi faktor-faktor tambahan seperti budaya perusahaan dan hubungan interpersonal di tempat kerja, yang juga dapat mempengaruhi pengunduran diri karyawan. Penggunaan teknik *machine learning* yang lebih canggih, seperti *ensemble methods* atau *deep learning*, juga dapat meningkatkan akurasi prediksi.

DAFTAR PUSTAKA

- [1] M. R. Shafie, H. Khosravi, S. Farhadpour, S. Das, and I. Ahmed, "A cluster-based human resources analytics for predicting employee turnover using optimized Artificial Neural Networks and data augmentation," *Decision Analytics Journal*, vol. 11, Jun. 2024, doi: 10.1016/j.dajour.2024.100461.
- [2] Luthfiyyah Wasiilah Maahiroh, "Klasifikasi Turnover Karyawan Menggunakan Algoritma XGBoost (Studi kasus: Divisi Engineering, Perusahaan Jasa Pertambangan)," 2024, Accessed: May 10, 2025. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/dspace.uui.ac.id/123456789/49400>
- [3] Y.-S. Lee and W.-K. Liu, "The Moderating Effects of Employee Benefits and Job Burnout among the Employee Loyalty, Corporate Culture and Employee Turnover," *Universal Journal of Management*, vol. 9, no. 2, pp. 62–69, Mar. 2021, doi: 10.13189/ujm.2021.090205.
- [4] S. K. Setianto and D. Jatikusumo, "Employee Turnover Analysis Using Comparison of Decision Tree and Naive Bayes Prediction Algorithms on K-Means Clustering Algorithms at PT. AT," *Jurnal Mantik*, vol. 4, no. 3, 2020, [Online]. Available: <https://iocscience.org/ejournal/index.php/mantik>
- [5] E. Sri Palupi, "EMPLOYEE TURNOVER CLASSIFICATION USING PSO-BASED NAÏVE BAYES AND NAÏVE BAYES ALGORITHM IN PT. MASTERSYSTEM INFOTAMA," 2021. [Online]. Available: www.bsi.ac.id
- [6] A. White, "Organizational Climate Variables and Performance as Predictors of Voluntary and Involuntary Turnover Among Nurses," 2022. Accessed: May 10, 2025. [Online]. Available: <https://search.proquest.com/openview/357445382dde2944873441e1f6dbf794/1?pq-origsite=gscholar&cbl=18750&diss=y>
- [7] G. Anwar and N. N. Abdullah, "The impact of Human resource management practice on Organizational performance," *International journal of Engineering, Business and Management (IJEEM)*, vol. 5, no. 1, pp. 2456–8678, 2021, doi: 10.22161/ijeem.5.1.
- [8] Helmi Imaduddin, Aditya Hermansyah Brian, Aura Salsabilla, and Frischa B, "ARISON OF SUPPORT VECTOR MACHINE AND DECISION TREE METHODS IN THE CLASSIFICATION OF BREAST CANCER," 2021. doi: <http://dx.doi.org/10.22373/cj.v5i1.8805>.
- [9] Daniel Dwi Eryanto Manurung, Fachran Sandi, Fajar Akbardipura, Hashfi Ashfahan, and Desta Sandya Prasvita, "Prediksi Pengunduran Diri Karyawan Perusahaan 'Y' Menggunakan Random Forest," 2021,

- Accessed: May 10, 2025. [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/1655>
- [10] Safitri Linawati, Siti Nurdian, Kartika Handayan, and Latifah, "PREDIKSI PRESTASI AKADEMIK MAHASISWA MENGGUNAKAN ALGORITMA RANDOM FOREST DAN C4.5," *Jurnal Khatulistiwa Informatika*, vol. VIII, no. 1, 2020, [Online]. Available: www.bsi.ac.id
- [11] O. Iparraguirre-Villanueva, L. Chauca-Huete, R. Prieto-Chavez, and C. Paulino-Moreno, "Employee attrition prediction using machine learning models," in *Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology*, Latin American and Caribbean Consortium of Engineering Institutions, 2024. doi: 10.18687/LACCEI2024.1.1.498.
- [12] M. Pratt, M. Boudhane, and S. Cakula, "Employee attrition estimation using random forest algorithm," *Baltic Journal of Modern Computing*, vol. 9, no. 1, pp. 49–66, 2021, doi: 10.22364/BJMC.2021.9.1.04.
- [13] A. A. Shlash Mohammad *et al.*, "Machine Learning Models for Predicting Employee Attrition: A Data Science Perspective," *Data and Metadata*, vol. 4, p. 669, Feb. 2025, doi: 10.56294/dm2025669.
- [14] A. Maehendrayuga, A. Setyanto, and Kusnawi, "Analisa Prediksi Turnover Karyawan menggunakan Machine Learning," *bit-Tech*, vol. 7, no. 2, pp. 648–659, Dec. 2024, doi: 10.32877/bt.v7i2.1999.
- [15] Dimas Bayu Stiawan and Yusuf Sulistyo Nugroho, "Perbandingan Performa Algoritma Decision Tree," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 4, p. 2108, 2023, doi: <https://doi.org/10.33022/ijcs.v12i4.3339>.
- [16] Febrian Andy Kusuma and Endang Wahyu Pamungkas, "PENDETEKSIAN HATE SPEECH PADA SOSIAL MEDIA INDONESIA DENGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN DECISION TREE," 2023. Accessed: Apr. 19, 2025. [Online]. Available: https://eprints.ums.ac.id/111722/1/Naskah%20Publikasi_Febrian%20Andy%20Kusuma_L200190137.pdf%20%281%292.pdf
- [17] D. Haganta Depari, Y. Widiastiwi, and M. Mega Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *JURNAL INFORMATIK*, vol. 18, p. 2022, 2022, doi: <https://doi.org/10.52958/iftk.v18i3.4694>.
- [18] A. Ferdita Nugraha, R. Faticha, A. Aziza, and Y. Pristyanto, "Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing," *Jurnal InfoMedia*, vol. 7, no. 1, 2022, doi: <http://dx.doi.org/10.30811/jim.v7i1.2959>.
- [19] F. Yulian Pamuji, V. Puspaning Ramadhan, and R. Artikel, "Komparasi Algoritma Random Forest Dan Decision Tree Untuk Memprediksi Keberhasilan Immunotherapy," *Jurnal Teknologi Dan Manajemen Informatika*, vol. 7, no. <https://jurnal.unmer.ac.id/index.php/jtmi/issue/view/451>, pp. 46–50, 2021, doi: <https://doi.org/10.26905/jtmi.v7i1.5982>.
- [20] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," Feb. 01, 2023, *King Saud bin Abdulaziz University*. doi: 10.1016/j.jksuci.2023.01.014.