

Penerapan Logistic Regression dan SMOTE untuk Memprediksi Atrisi Karyawan pada Imbalanced Data

Viviana Herlita Vidiyari, Hairani Hairani, Heroe Santoso, Farda Milanda Amin

Universitas Bumigora, Mataram, Indonesia

Correspondence : e-mail: herlita.vivi@gmail.com

Abstrak

Atrisi karyawan merupakan tantangan bagi perusahaan karena berdampak langsung pada meningkatnya biaya pelatihan, hilangnya pengetahuan, serta menurunnya produktivitas dan stabilitas tim. Penelitian ini bertujuan untuk membangun model prediksi atrisi karyawan menggunakan pendekatan machine learning. Dataset yang digunakan terdiri dari 1.470 data karyawan, yang telah melalui tahap preprocessing, termasuk penanganan nilai kategorikal, normalisasi, serta seleksi fitur menggunakan Recursive Feature Elimination (RFE). Untuk mengatasi ketidakseimbangan data antara kelas tidak resign dan resign, digunakan teknik SMOTE. Model logistic regression dipilih untuk mendapatkan performa terbaik. Hasil evaluasi menunjukkan bahwa model mampu memberikan hasil yang baik dengan akurasi sebesar 0.74, ROC AUC sebesar 0.82, dan recall 0.79 pada kelas minoritas. Temuan ini diharapkan dapat digunakan oleh perusahaan sebagai alat bantu dalam pengambilan keputusan untuk mengelola risiko atrisi secara lebih efektif.

Kata kunci: Atrisi Karyawan, Klasifikasi, Logistic Regression, SMOTE, Machine Learning

Abstract

Employee attrition poses a significant challenge for companies, as it directly impacts training costs, leads to the loss of organizational knowledge, and reduces team productivity and stability. This study aims to develop a predictive model for employee attrition using a machine learning approach. The dataset consists of 1,470 employee records, which underwent preprocessing steps including categorical value encoding, normalization, and feature selection using Recursive Feature Elimination (RFE). To address the class imbalance between resigned and non-resigned employees, the SMOTE (Synthetic Minority Over-sampling Technique) method was applied. A Logistic Regression model was selected for its interpretability and effectiveness. The evaluation results show that the model performs well, achieving an accuracy of 0.74, a ROC AUC score of 0.82, and a recall of 0.79 for the minority class. These findings suggest that the model can serve as a valuable decision-support tool for organizations in managing and mitigating employee attrition risks.

Keywords: Employee attrition, Classification, Logistic Regression, SMOTE, Machine Learning

1. Pendahuluan

Di tengah era globalisasi dan persaingan bisnis yang semakin ketat, pertumbuhan organisasi yang berkelanjutan menjadi fokus utama perusahaan untuk menjaga kelangsungan dan meraih kesuksesan jangka panjang [1]. Manajemen sumber daya manusia (SDM) memegang peran kunci dalam mendorong keberlanjutan organisasi [2]. Namun, tingginya tingkat *attrition* karyawan menjadi tantangan karena dapat menghambat pencapaian tujuan strategis [3], menimbulkan berbagai kerugian seperti meningkatnya biaya pelatihan, hilangnya pengetahuan, serta menurunnya produktivitas dan stabilitas tim [4]. Oleh sebab itu, kemampuan memprediksi *attrition* menjadi bagian penting dalam strategi manajemen dan retensi karyawan [5], salah satunya melalui analisis prediktif berbasis *machine learning* terhadap data historis [6]. Namun, prediksi *attrition* karyawan menghadapi sejumlah tantangan, yaitu adanya ketidakseimbangan kelas (*imbalanced class*) dalam data, di mana jumlah karyawan yang bertahan jauh lebih besar dibandingkan dengan yang *resign* [6]. Ketidakseimbangan data berdampak negatif terhadap hasil klasifikasi, karena model cenderung salah mengklasifikasikan kelas minoritas sebagai kelas mayoritas [7].

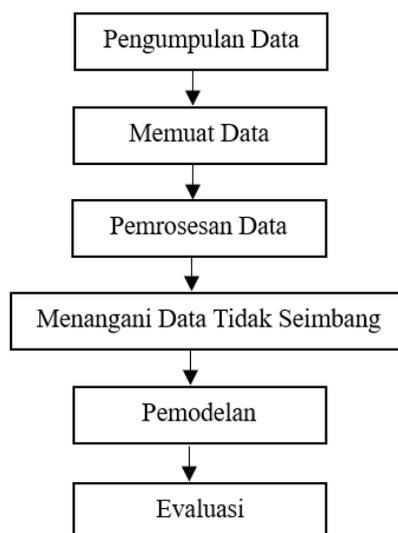
Untuk mengatasi masalah ini, digunakan pendekatan SMOTE sebagai metode *oversampling* yang bertujuan menyeimbangkan distribusi kelas pada data pelatihan [8]. SMOTE menghasilkan sampel baru dengan menginterpolasi data minoritas, sehingga membantu model mengenali pola penting yang sebelumnya tersembunyi [9]. *Logistic regression* banyak digunakan sebagai algoritma klasifikasi karena memberikan hasil yang baik pada data linier dan mudah untuk dianalisis secara interpretatif [10].

Beberapa penelitian yang terkait yaitu penelitian yang dilakukan oleh [11] membandingkan kinerja *naive bayes* dan C4.5 dalam memprediksi keputusan resign karyawan, menggunakan data wawancara CV. Dunia Pustaka dan data sekunder dari Kaggle, lain halnya dengan [12] memprediksi *attrition* karyawan PT. XYZ dengan *soft voting classifier* yang menggabungkan SVM, *decision tree*, dan *logistic regression*, dengan data historis karyawan tahun 2018–2023. Sementara [13] menerapkan metode PSO untuk reduksi dimensi data *attrition* karyawan, kemudian algoritma SVM dan model *neural network* berbasis deep learning untuk mengoptimalkan akurasi. Penelitian [14] memprediksi kemungkinan karyawan keluar pada PT. HCI menggunakan metode *ensemble*, yaitu *stacking*, *random forest*, dan *AdaBoost*, sementara [15] mengembangkan sistem prediksi *employee attrition* berbasis *web* menggunakan algoritma *random forest* di UD. Mas Achiad. Namun belum banyak penelitian yang secara spesifik mengkaji pemanfaatan *logistic regression* yang dikombinasikan dengan SMOTE dalam menangani *imbalanced* data untuk meningkatkan akurasi prediksi terhadap karyawan yang berpotensi resign.

Penelitian ini bertujuan untuk menerapkan model *logistic regression* yang dikombinasikan dengan SMOTE guna memprediksi atrisi karyawan pada *imbalanced* data *HR Employee Attrition*. Pendekatan ini diharapkan dapat meningkatkan sensitivitas model terhadap kelas minoritas, khususnya karyawan yang berpotensi mengundurkan diri. Dengan optimasi yang tepat, penelitian ini menunjukkan bahwa model sederhana seperti *logistic regression* tetap dapat memberikan hasil prediksi yang efektif untuk mendukung pengambilan keputusan strategis dalam manajemen SDM.

2. Metode Penelitian

Berikut merupakan tahapan penelitian yang dilakukan yaitu, identifikasi jenis penelitian, pengumpulan data, *preprocessing* data, penanganan ketidakseimbangan data dengan SMOTE, pemodelan menggunakan *logistic regression*, hingga tahap evaluasi model



Gambar 1. Diagram alur penelitian.

Pengumpulan Data. Data dikumpulkan dari sumber yang relevan [16]. Dalam penelitian ini, *dataset* yang digunakan berasal dari situs Kaggle.com yang terdiri dari 35 fitur berisi informasi karyawan, termasuk fitur demografis, jabatan, kepuasan kerja, dan lainnya. Mengambil data dari sumber terpercaya seperti Kaggle untuk menjamin *dataset* memiliki keragaman dan representasi yang memadai. Memuat *Data*. Setelah data diperoleh, langkah selanjutnya yaitu memuat dan membaca *dataset* ke dalam proses pemrograman agar dapat diakses dan diolah lebih lanjut [17]. *Dataset* yang diunduh dari Kaggle dimuat ke dalam *DataFrame* menggunakan pustaka Pandas di Python. Format ini memudahkan eksplorasi dan manipulasi data secara tabular [18].

Pemrosesan Data. Pada tahap ini dilakukan pembersihan dan transformasi data untuk meningkatkan kualitas dan konsistensi data sebelum dianalisis lebih lanjut [19]. Diawali dengan menangani nilai-nilai kosong (*missing values*). Selanjutnya, fitur kategorikal dikonversi ke format numerik melalui proses *encoding*. Kemudian tahap *split data* yaitu pembagian data *training* dan data *testing*. Pada penelitian ini menggunakan persentase 80% *training* dan 20% *testing*. Di tahap ini pula dilakukan seleksi fitur untuk menyaring fitur yang paling relevan dengan metode *Recursive Feature Elimination* (RFE).

Menangani *Data* Tidak Seimbang. Tahapan penting untuk menangani ketidakseimbangan data guna meningkatkan kualitas hasil prediksi [20]. Digunakan pendekatan SMOTE yang menghasilkan data sintetis pada kelas minoritas. Pemodelan. Tahap untuk menerapkan teknik data mining yang digunakan [21]. Proses pembuatan model dilakukan dengan menggunakan algoritma *logistic regression*. Algoritma ini dipilih karena memberikan kemudahan dalam interpretasi model melalui hubungan linear antara fitur input dan output yang dihasilkan [22]. Evaluasi. Pada tahap terakhir ini bertujuan untuk menilai keefektifan dan kualitas model [23]. Evaluasi dilakukan menggunakan berbagai metrik evaluasi seperti *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan ROC-AUC. Dalam *imbalanced data*, perhatian khusus diberikan pada nilai *recall* untuk kelas minoritas, karena keberhasilan model dalam mendeteksi karyawan yang berpotensi resign menjadi prioritas utama dibandingkan hanya mengandalkan akurasi keseluruhan.

3. Hasil dan Pembahasan

Setelah melalui tahapan metode penelitian yang telah dijelaskan, diperoleh hasil dan pembahasan sebagai berikut.

3.1. Pengumpulan Data

Dataset yang digunakan terdiri dari 35 fitur yang dikelompokkan ke dalam beberapa kategori, yang dapat dilihat pada Tabel 1, yaitu data identitas karyawan, pendidikan dan pengalaman kerja, jabatan dan status pekerjaan, kompensasi dan gaji, kepuasan kerja, serta jam dan pola kerja. Fitur Attrition digunakan sebagai label target untuk prediksi keputusan resign karyawan.

Tabel 1. Fitur data atrisi karyawan

No	Kategori Fitur	Jumlah Fitur	Fitur
1	Identitas & Demografis	6	Age, Gender, Marital Status, Over 18, Employee Count, EmployeeNumber
2	Pendidikan & Pengalaman	6	Education, Education Field, Num Companies Worked, Total Working Years, Training Times Last Year
3	Jabatan & Pekerjaan	8	Department, Job Level, Job Role, Job Involvement, Years At Company, Years In Current Role, Years With Curr Manager, Years Since Last Promotion
4	Kompensasi & Gaji	5	Monthly Rate, Monthly Income, Hourly Rate, Daily Rate, Stock Option Level
5	Kepuasan & Lingkungan Kerja	5	Environment Satisfaction, Job Satisfaction, Work Life Balance, Relationship Satisfaction, Distance From Home
6	Jam & Pola Kerja	3	Over Time, Standard Hours, Business Travel
7	Label Target	1	Attrition

3.2. Memuat Data

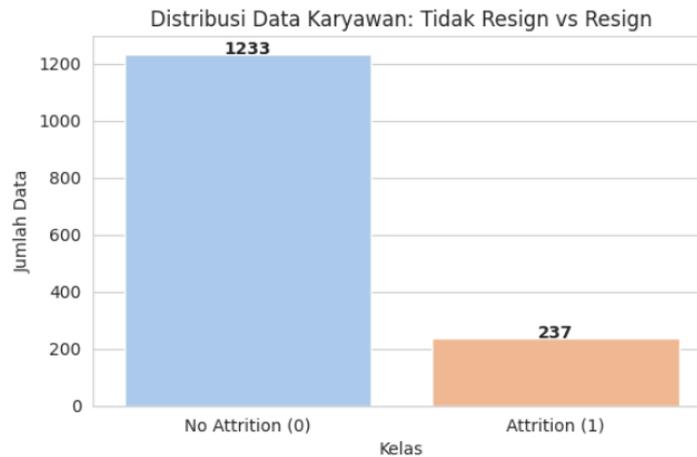
Dataset yang telah diperoleh kemudian di import ke *google drive* untuk diproses menggunakan *google colab*. Pada Table 2 akan ditampilkan 5 data teratas (*head*) untuk memastikan bahwa data yang di *import* sesuai dengan data yang akan diproses. Langkah ini merupakan tahap awal eksplorasi pola terkait keputusan karyawan bertahan atau *resign*, serta dasar untuk analisis dan pemodelan klasifikasi atrisi

Tabel 2. Proses memuat *dataset*

No	Age	Attrition	Business Travel	Daily Rate	Years Since Last Promotion	Years With Curr Manager
1	41	Yes	Travel_Rarely	1102	0	5
2	49	No	Travel_Frequently	279	1	7
3	37	Yes	Travel_Rarely	1373	0	0
4	33	No	Travel_Frequently	1392	3	0
5	27	No	Travel_Rarely	591	2	2

3.3. Pemrosesan Data

Berdasarkan *dataset attrition* yang digunakan, data menunjukkan distribusi kelas *imbalanced*, dengan jumlah karyawan yang tidak mengundurkan diri jauh lebih banyak dibandingkan yang mengundurkan diri. Data tersebut dapat dilihat pada Gambar 2.



Gambar 2. Distribusi data *imbalanced*

Selanjutnya pengecekan *missing value* untuk memeriksa data yang kosong. Berdasarkan hasil pengecekan, semua fitur menghasilkan nilai 0 yang menandakan bahwa tidak ada data yang kosong.

```
Cek Missing Values:
Age 0
Attrition 0
BusinessTravel 0
DailyRate 0
Department 0
DistanceFromHome 0
Education 0
EducationField 0
EmployeeCount 0
EmployeeNumber 0
EnvironmentSatisfaction 0
Gender 0
HourlyRate 0
JobInvolvement 0
JobLevel 0
JobRole 0
JobSatisfaction 0
MaritalStatus 0
MonthlyIncome 0
MonthlyRate 0
NumCompaniesWorked 0
Over18 0
OverTime 0
PercentSalaryHike 0
PerformanceRating 0
RelationshipSatisfaction 0
StandardHours 0
StockOptionLevel 0
TotalWorkingYears 0
TrainingTimesLastYear 0
WorkLifeBalance 0
YearsAtCompany 0
YearsInCurrentRole 0
YearsSinceLastPromotion 0
YearsWithCurrManager 0
dtype: int64
```

Gambar 3. *Missing value*

Selanjutnya dilakukan *encoding* pada fitur kategorikal menggunakan *label encoder* agar dapat dilakukan analisis lebih lanjut ke dalam algoritma pembelajaran mesin. Kemudian pembagian dataset menjadi *data training* dan *data testing*. Dari total 1.470 data, sebanyak 80% atau 1.176 data digunakan untuk *training*, dan 20% atau 294 data untuk *testing*. Proses seleksi fitur dilakukan dengan RFE (*Recursive*

Feature Elimination) untuk mengurangi kompleksitas dan meningkatkan performa model, dengan memilih 15 fitur paling relevan seperti ditampilkan pada Gambar 4.

- Variabel terpilih oleh RFE:
1. Age
 2. Department
 3. DistanceFromHome
 4. EnvironmentSatisfaction
 5. JobInvolvement
 6. JobSatisfaction
 7. MaritalStatus
 8. MonthlyIncome
 9. NumCompaniesWorked
 10. OverTime
 11. TotalWorkingYears
 12. YearsAtCompany
 13. YearsInCurrentRole
 14. YearsSinceLastPromotion
 15. YearsWithCurrManager

Gambar 4. Seleksi Fitur

3.4. Menangani Data Tidak Seimbang

Berdasarkan hasil SMOTE, jumlah data pada kelas minoritas (*attrition* = 1) meningkat dari 190 menjadi 986. Hal ini bertujuan agar model tidak bias terhadap kelas mayoritas dan mampu mengenali pola pada kelas minoritas secara lebih optimal.

Tabel 3. Hasil SMOTE

Kelas Attrition	Sebelum SMOTE	Sesudah SMOTE
Tidak (0)	986	986
Ya (1)	190	986
Total	1176	1972

3.5. Pemodelan

Pada tahap ini, algoritma *logistic regression* digunakan untuk membangun model klasifikasi *attrition*. *Logistic regression* dipilih karena merupakan salah satu metode statistik yang umum digunakan untuk klasifikasi biner dan memiliki performa yang baik dalam kasus linear. Sebelum pelatihan model, dilakukan *feature selection* untuk memilih fitur yang paling relevan berpengaruh terhadap label target (*attrition*). Selanjutnya dilakukan penyeimbangan data menggunakan SMOTE guna menangani ketidakseimbangan data. Setelah seluruh proses dilakukan, model dilatih menggunakan data yang telah disiapkan dan selanjutnya dievaluasi pada data uji.

3.6. Evaluasi

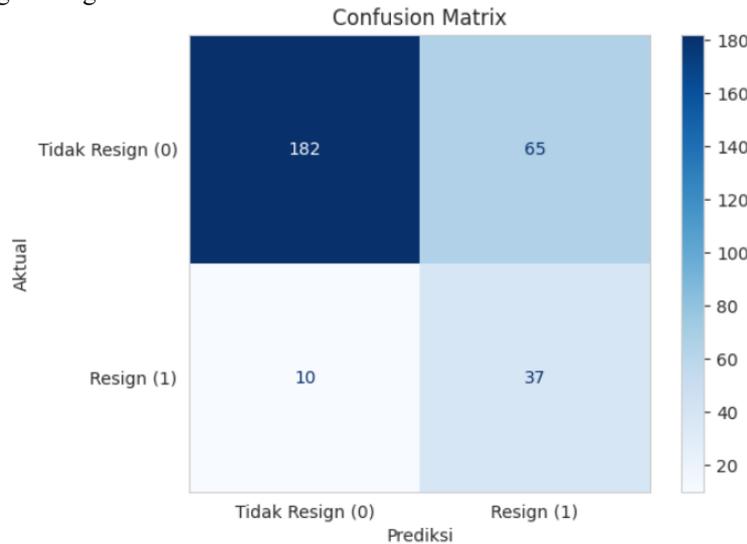
Pada tahap ini akan ditampilkan hasil evaluasi model menggunakan metrik seperti metrik evaluasi, *confusion matrix* dan ROC-AUC untuk menilai kinerja model dalam mengklasifikasikan data *attrition*.

Tabel 4. Metrik evaluasi

	Precision	Recall	F1-Score	Support
0	0.95	0.74	0.83	247
1	0.36	0.79	0.50	47
Accuracy			0.74	294
Macro avg	0.66	0.76	0.66	294
Weighted avg	0.85	0.74	0.78	294

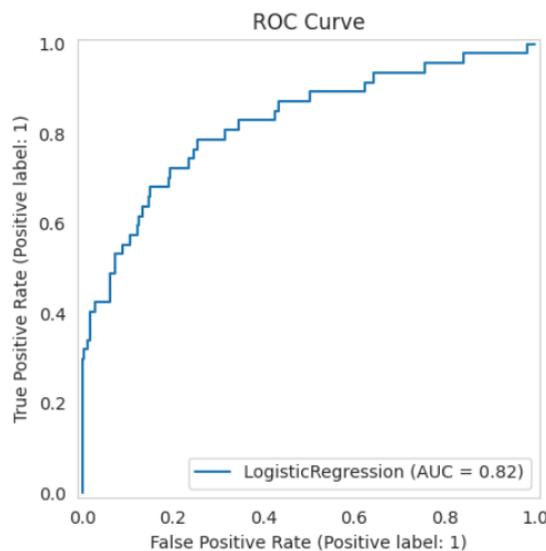
Hasil metrik evaluasi menunjukkan untuk kelas 0, memiliki nilai *precision* 0.95, *recall* 0.74, dan *f1-score* 0.83. Hal ini menunjukkan bahwa model mampu mengenali sebagian besar karyawan yang tidak *resign*, dengan jumlah prediksi positif yang tepat juga tinggi. Sementara, untuk kelas 1 (karyawan *resign*), menunjukkan *precision* 0.36, *recall* 0.79, dan *f1-score* 0.50. Meskipun nilai *recall* cukup tinggi, yang berarti model berhasil mengidentifikasi sebagian besar kasus *resign*, nilai *precision* yang rendah menunjukkan bahwa masih banyak prediksi *false positive*. Secara keseluruhan, model menghasilkan *accuracy* 0.74, dengan *macro average* untuk *precision*, *recall*, dan *f1-score* masing-masing sebesar 0.66, 0.76, dan 0.66. Nilai *weighted average* ditunjukkan dengan *precision* 0.85, *recall* 0.74, dan *f1-score* 0.78,

yang memberikan gambaran umum terhadap kinerja model dengan mempertimbangkan proporsi jumlah sampel di masing-masing kelas.



Gambar 5. Confusion matrix

Berdasarkan Gambar 5 hasil *confusion matrix*, model berhasil mengklasifikasikan: 182 data karyawan yang benar-benar tidak *resign* (*true negative*), 37 data karyawan yang benar-benar *resign* (*true positive*), 65 data salah diklasifikasikan sebagai *resign* (*false positive*), dan 10 data *resign* yang salah diklasifikasikan sebagai tidak *resign* (*false negative*).



Gambar 6. Grafik ROC-AUC

Pada Gambar 6, ditampilkan hasil grafik ROC dengan nilai AUC sebesar 0.82, yang menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan antara karyawan yang *resign* dan tidak. Berdasarkan hasil ini, dapat disimpulkan bahwa *logistic regression* mampu memberikan prediksi yang cukup akurat, meskipun masih dapat dilakukan perbaikan pada aspek *precision* untuk kelas minoritas.

4. Kesimpulan

Penelitian ini berhasil membangun model klasifikasi untuk memprediksi atrisi karyawan dengan menggunakan algoritma *logistic regression* dan SMOTE untuk mengatasi ketidakseimbangan data. Model menghasilkan performa yang cukup baik, dengan nilai akurasi sebesar 0.74, ROC AUC sebesar 0.82, dan nilai *recall* yang tinggi untuk kelas atrisi yaitu 0.79. Hasil ini menunjukkan bahwa model dapat menjadi alat bantu prediktif yang efektif bagi perusahaan dalam mengidentifikasi risiko atrisi sejak dini dan

mendukung pengambilan keputusan strategis dalam manajemen SDM. Sebagai solusi pengembangan ke depan, model dapat ditingkatkan dengan mengeksplorasi algoritma lain seperti *random forest*, *XGBoost*, atau *neural network* untuk memperoleh perbandingan performa yang lebih baik.

Daftar Pustaka

- [1] N. Basuki, "Mengoptimalkan Modal Manusia: Strategi Manajemen Sumber Daya Manusia Yang Efektif Untuk Pertumbuhan Organisasi Yang Berkelanjutan," *J. Ilm. Manaj.*, vol. 4, no. 2, pp. 182–192, 2023, doi: 10.15575/jim.v4i2.28606.
- [2] G. Sartika, "Peran Strategis Manajemen Sumber Daya Manusia dalam Mendukung Keberlanjutan Organisasi melalui Penerapan Triple Bottom Line," *J. Ekon. dan Bisnis*, vol. 4, no. 1, pp. 177–188, 2024, doi: 10.56145/ekonomibisnis.v4i1.155.
- [3] W. A. Medyanti and M. Faisal, "Sistem Prediksi Awal Pada Atrisi Karyawan Perusahaan 'XYZ' Menggunakan Algoritma Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 8, no. 2, p. 429, 2023, doi: 10.24114/cess.v8i2.46494.
- [4] S. Sumarita and M. Muzakki, "Pengaruh Kepemimpinan Transformasional pada Intensi Karyawan Keluar dari Perusahaan," *Indones. Bus. Rev.*, vol. 6, no. 1, p. 61, 2023, doi: 10.21632/ibr.6.1.61-70.
- [5] A. R. B. Jamroni, W. Hadikristanto, and M. Fatchan, "Analisis Faktor dan Prediksi Atrisi untuk Optimalisasi Retensi Karyawan Menggunakan Machine Learning," *Bit-Tech (Binary Digit. - Technol.)*, vol. 7, no. 3, 2025, doi: 10.32877/bt.v7i3.2301.
- [6] M. Arfah and S. A. Pramono, "Eksplorasi Transformasi Digital dalam MSDM: Dampak Integrasi Artificial Intelligence dan Big Data Analytics terhadap Pengambilan Keputusan Strategis," *J. Minfo Polgan*, vol. 14, pp. 183–192, 2025, doi: 10.33395/jmp.v14i1.14673.
- [7] Louis Madaerdo Sotarjua and Dian Budhi Santoso, "Perbandingan Algoritma Knn, Decision Tree, Dan Random Forest Pada Data Imbalanced Class Untuk Klasifikasi Promosi Karyawan," *J. INSTEK (Informatika Sains dan Teknol.)*, vol. 7, no. 2, pp. 192–200, 2022, doi: 10.24252/instek.v7i2.31385.
- [8] A. Masruriyah, H. Novita, C. Sukmawati, A. Ramadhan, S. Arif, and B. Dermawan, "Pengukuran Kinerja Model Klasifikasi dengan Data Oversampling pada Algoritma Supervised Learning untuk Penyakit Jantung," *Comput. Sci.*, vol. 4, no. 1, pp. 62–70, 2024, doi: 10.31294/coscience.v4i1.2389.
- [9] K. S. Uswatun Hasanah, Agus Mohamad Soleh, "Effect of Random Under sampling, Oversampling, and SMOTE on the Performance of Cardiovascular Disease Prediction Models," *J. Mat. Stat. Dan Komputasi*, vol. 21, no. 1, pp. 88–102, 2024, doi: 10.20956/j.v21i1.35552.
- [10] K. Nisa and S. Kartika Wibisono, "Klasifikasi Penyakit Skizofrenia menggunakan Algoritma Logistic Regresion," *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 8, no. 2, pp. 696–704, 2023, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [11] F. Daewana, "Perbandingan Algoritma Naïve Bayes Dan C4.5 Dalam Prediksi Keputusan Karyawan Untuk Meninggalkan Perusahaan," *Innov. J. Soc. Sci. Res.*, vol. 5, no. 1, pp. 1712–1726, 2025, doi: 10.31004/innovative.v5i1.17664.
- [12] C. Chaedar, B. Lirna, and A. T. Damaliana, "Prediksi Atrisi Voluntary Karyawan Pt.Xyz: Pendekatan Ensemble Machine Learning Dengan Soft Voting Classifier," *J. Tek. Inform.*, vol. 5, no. 5, pp. 1231–1239, 2024, [Online]. Available: <https://repository.upnjatim.ac.id/28002/>
- [13] D. A. Fauziah, A. Muliawan, and M. Dimiyati, "Implementasi Machine Learning Pada Atrisi Karyawan Berdasarkan Parameter Kinerja Dengan Menggunakan Metode Particle Swarm Optimization Dan Ensemble Classifier," *J. Tek. Inform.*, vol. 5, no. 6, pp. 1823–1831, 2024, doi: 10.52436/1.jutif.2024.5.6.3442.
- [14] M. Fadel, Z. Arifin, and G. Triyono, "Application Of Ensemble Method For Employee Turnover Predictions In Financial Services Company," *J. Tek. Inform.*, vol. 5, no. 3, pp. 767–775, 2024, [Online]. Available: <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/1871>
- [15] R. Nurul, S. Hakim, and S. R. Nudin, "Pengembangan Sistem Prediksi Employee Attrition Menggunakan Algoritma Random Forest Berbasis Web Di Ud. Mas Achiad," *J. Manaj. Inform.*, vol. 16, no. 2, 2024, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jurnal-manajemen-informatika/article/view/60074>
- [16] Muhajirin, Risnita, and Asrulla, "Pendekatan Penelitian Kuantitatif Dan Kualitatif Serta Tahapan Penelitian," *J. Genta Mulia*, vol. 15, no. 1, pp. 82–92, 2024, [Online]. Available: <https://ejournal.uncm.ac.id/index.php/gm/article/view/903>
- [17] Y. Rifa'i, "Analisis Metodologi Penelitian Kualitatif dalam Pengumpulan Data di Penelitian Ilmiah pada Penyusunan Mini Riset," *Cendekia Inov. Dan Berbudaya*, vol. 1, no. 1, pp. 31–37, 2023, doi:

- 10.59996/cendib.v1i1.155.
- [18] I. N. Rizki, D. Prayoga, M. L. Puspita, and M. Q. Huda, "Implementasi Exploratory Data Analysis Untuk Analisis Dan Visualisasi Data Penderita Stroke Kalimantan Selatan Menggunakan Platform Tableau," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3856.
- [19] T. M. A. Admira and M. Saefudin, "Model Sentiment Analysis Berbasis Machine Learning Untuk Data Genz-Career Aspiration Menggunakan Flask Dan Naive Bayes," *J. Inf. Syst. Informatics Comput.*, vol. 9, no. 1, pp. 25–39, 2025, doi: 10.52362/jisicom.v9i1.1880.
- [20] F. R. Lumbanraja, E. C. L. Gaol, D. A. Shofiana, and A. Junaidi, "Implementasi SMOTE dan Support Vector Machine Pada Klasifikasi Data Tidak Seimbang Metilasi Arginin," *J. Pepadun*, vol. 5, no. 1, pp. 27–37, 2024, doi: 10.23960/pepadun.v5i1.209.
- [21] A. Rohanah, D. L. Rianti, and B. N. Sari, "Perbandingan Naive Bayes dan Support Vector Machine untuk Klasifikasi Ulasan Pelanggan Indihome," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 1, p. 23, 2021, doi: 10.30998/string.v6i1.9232.
- [22] N. R. Febriyanti and A. D. Hartanto, "Analisis Perbandingan Algoritma SVM, Random Forest dan Logistic Regression untuk Prediksi Stunting Balita," *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 149–158, 2025, doi: 10.29408/edumatic.v9i1.29407.
- [23] R. R. Rahmah, I. R. Setiawan, and F. F. Az-zahra, "Prediksi Keterlambatan Mahasiswa Dalam Membayar Biaya Kuliah Menggunakan Algoritma C4.5 (Studi Kasus : Universitas Muhammadiyah Sukabumi)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 10792–10800, 2024, [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/view/11145>