JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

OPTIMALISASI PREDIKSI KEHILANGAN KARYAWAN MENGGUNAKAN TEKNIK RFE, SMOTE, DAN ADABOOST

Prambudi Setiyadi[*1)], Muhamad Nur Prayogi[2)], Achmad Solichin[3)]

1. Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia
2. Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia
3. Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Article Info
Kata Kunci: Adaboost; Kehilangan karyawan; RFE; SMOTE;

Keywords: *Adaboost;Employee Churn; RFE; SMOTE;*

Article history:
Received 29 September 2024
Revised 13 Oktober 2024
Accepted 4 November 2024
Available online 4 December 2024

DOI :
https://doi.org/10.29100/jipi.v9i4.5642

* Corresponding author.
Corresponding Author
E-mail address:
Prambudi.setiyadi@gmail.com

ABSTRAK
Kehilangan karyawan menjadi isu vital dalam dinamika organisasi karena dampaknya yang signifikan terhadap produktivitas dan stabilitas tenaga kerja. Penelitian ini menerapkan teknik *machine learning* untuk mengantisipasi pergantian karyawan dengan menggabungkan seleksi fitur, *oversampling*, dan algoritma *ensemble*. Empat pendekatan yang dibandingkan adalah RFE-SMOTE-ADABOOST, RFE-ADABOOST, SMOTE-ADABOOST, dan SMOTE-ADABOOST dengan Hyperparameter. Evaluasi dilakukan menggunakan metrik akurasi, presisi, *recall*, dan F1-*score*. Hasil menunjukkan bahwa SMOTE-ADABOOST dengan *Hyperparameter* mencapai kinerja tertinggi, dengan akurasi 0,907, presisi 0,912, recall 0,898, dan F1-score 0,905. Model ini mengidentifikasi 10 faktor kunci yang mempengaruhi prediksi pergantian karyawan, seperti *Education Field*, *Business Travel*, *dan Monthly Income*. Kesimpulannya, model SMOTE-ADABOOST dengan Hyperparameter terbukti paling efektif dalam memprediksi kehilangan karyawan. Implikasi dari hasil evaluasi ini menunjukkan bahwa organisasi dapat secara proaktif mengidentifikasi dan mengelola faktor-faktor kunci yang mempengaruhi retensi karyawan, sehingga meningkatkan stabilitas tenaga kerja dan produktivitas keseluruhan.

ABSTRACT
Employee turnover is a vital issue in organizational dynamics due to its significant impact on productivity and workforce stability. This study applied machine learning techniques to anticipate employee turnover by combining feature selection, oversampling, and ensemble algorithms. Four approaches were compared: RFE-SMOTE-ADABOOST, RFE-ADABOOST, SMOTE-ADABOOST, and SMOTE-ADABOOST with Hyperparameters. Evaluation metrics included accuracy, precision, recall, and F1-score. Results indicated that SMOTE-ADABOOST with Hyperparameters achieved the highest performance, with an accuracy of 0.907, precision of 0.912, recall of 0.898, and F1-score of 0.905. The model identified 10 key factors influencing employee turnover predictions, such as Education Field, Business Travel, and Monthly Income. In conclusion, the SMOTE-ADABOOST with Hyperparameters model proved most effective in predicting employee turnover. These evaluation results imply that organizations can proactively identify and manage key factors affecting employee retention, thus enhancing workforce stability and overall productivity

I. PENDAHULUAN

KEHILANGAN karyawan merupakan masalah yang sangat penting dalam konteks dinamika organisasi karena dampaknya yang signifikan terhadap produktivitas dan stabilitas tenaga kerja. Dampak kehilangan karyawan tidak hanya terbatas pada biaya penggantian, tetapi juga dapat berakibat pada kerusakan reputasi organisasi, mengganggu pencapaian tujuan, dan mengganggu adaptasi organisasi terhadap perubahan lingkungan, yang dapat berakibat pada kehilangan kompetitivitas. Dalam era digital saat ini, hal tersebut dapat berakibat pada kehilangan

pengetahuan dan keterampilan yang telah diakumulasikan, serta mengganggu proses operasional yang telah terintegrasi dengan teknologi. Oleh karena itu, prediksi kehilangan karyawan sangat penting untuk mengantisipasi dan mengurangi risiko kehilangan karyawan, serta mengoptimalkan strategi pengembangan sumber daya manusia.

Di tengah kompleksitas tantangan tersebut, penelitian ini bertujuan untuk menggunakan teknik *machine learning* guna mengantisipasi pergantian karyawan dengan memadukan berbagai strategi seperti seleksi fitur, *oversampling*, dan algoritma *ensemble*. Faktor-faktor awal yang menjadi pertimbangan dalam analisis ini meliputi *age, Business Travel, Daily Rate, Department, Distance From Home, education, Education Field, Employee Count, Employee Number, Environment Satisfaction, Gender, Hourly Rate, Job Involvement, Job Level, Job Role, Job Satisfaction, Marital Status, Monthly Income, Monthly Rate, Num Companies Worked, Over Time, Percent SalaryHike, Performance Rating, Relationship Satisfaction, Standard Hours, Stock OptionLevel, Total Working Years, Training Times Last Year, Work Life Balance, Years At Company, Years In Current Role, Years Since Last Promotion,* dan *Years With Curr Manager,*. Dalam kerangka penelitian ini membandingkan empat pendekatan yang berbeda, yakni kombinasi metode RFE-SMOTE-ADABOOST, RFE-ADABOOST, SMOTE-ADABOOST, dan SMOTE-ADABOOSTcdengan *Hyperparameter*. Evaluasi kinerja dilakukan dengan menggunakan metrik standar seperti akurasi, presisi, *recall*, dan F1-*score*. Temuan penelitian menunjukkan bahwa pendekatan SMOTE-ADABOOST dengan *Hyperparameter* memberikan kinerja tertinggi dengan nilai akurasi, presisi, *recall*, dan F1-*score* yang optimal. Penelitian ini juga memperkaya pemahaman dengan mencantumkan metode-metode sebelumnya yang pernah digunakan dalam analisis prediksi pergantian karyawan, termasuk *Random Forest,* SVM (*Support Vector Machine*), Regresi Logistik, *Stacking*, *Decision Tree*, dan XG *Boost*. Namun, fokus penelitian kami terletak pada pengembangan dan pembandingan kombinasi empat pendekatan *machine learning* tersebut yang dirancang khusus untuk menangani masalah kehilangan karyawan dengan lebih efektif.

Metode yang dipilih dengan cermat untuk penelitian ini adalah *Recursive Feature Elimination (*RFE), *Synthetic Minority Over-sampling Technique* (SMOTE), dan *Adaptive Boosting* (AdaBoost). RFE digunakan untuk mengidentifikasi subset fitur yang paling relevan dalam model, sementara SMOTE digunakan untuk menangani ketidakseimbangan kelas dalam data pelatihan. Selain itu, AdaBoost dipilih untuk meningkatkan kinerja model dengan menggabungkan hasil dari beberapa model yang lebih lemah.

Pendekatan dalam penelitian ini akan memberikan wawasan yang berharga dan solusi yang praktis bagi organisasi yang menghadapi tantangan dalam memahami dan mengelola kehilangan karyawan. Dengan demikian, penelitian ini memiliki potensi untuk memberikan kontribusi yang signifikan dalam pemahaman dan penanganan masalah kehilangan karyawan di berbagai konteks organisasi.

Penelitian-penelitian sebelumnya tentang kehilangan karyawan telah dilakukan dengan berbagai metode dan algoritma machine learning untuk meningkatkan akurasi prediksi. Hasil penelitian [1] menunjukkan bahwa metode XGB memiliki tingkat akurasi tertinggi, sedangkan penelitian [2] [3] [4] [5][6] menemukan bahwa Random Forest memiliki akurasi tertinggi dibanding model lainnya. Penelitian [7] juga menunjukkan bahwa model kombinasi KS-IFCM-SMOTE-SVM memiliki hasil yang stabil. Selain itu, penelitian [8] memfokuskan pada penggunaan regresi logistik dan LightGBM, sementara penelitian [9]menggunakan Extra Trees Classifier, dan penelitian [10] menggunakan Chi-squared . Penelitian[11] menjelaskan penggunaan GB dan NN yang optimal, dan penelitian [12]menemukan bahwa Information Gain memiliki tingkat tertinggi.

Penelitian lainnya juga menyoroti hasil yang baik dari Algoritma CatBoost pada penelitian [13] , serta Regresi logitistik tanpa pemilihan fitur yang mendapatkan hasil terbaik pada penelitian [14] Penelitian [15] juga menunjukkan bahwa Deep Neural Network dapat mendapatkan akurasi prediksi terbaik. Selain itu, penelitian [16]menemukan bahwa penggabungan Chi-squared dan Gradient Boosting Tree dapat meningkatkan akurasi , dan penelitian [17] menunjukkan bahwa Logistik Regression memiliki hasil yang cukup baik. Dengan demikian, hasil penelitian sebelumnya menunjukkan bahwa berbagai metode dan algoritma machine learning dapat digunakan untuk meningkatkan akurasi prediksi kehilangan karyawan.

Berbeda dengan penelitian-penelitian sebelumnya, Penelitian ini memberikan kontribusi yang signifikan dibandingkan dengan penelitian sebelumnya dalam bidang prediksi kehilangan karyawan. Penelitian ini memperkenalkan kombinasi teknik yang inovatif yaitu *Recursive Feature Elimination* (RFE), *Synthetic Minority Over-sampling Technique* (SMOTE), dan algoritma AdaBoost, yang jarang digunakan bersamaan dalam penelitian sebelumnya. Kombinasi ini memungkinkan pemilihan fitur yang lebih relevan, penyeimbangan kelas data yang tidak seimbang, dan peningkatan akurasi prediksi secara signifikan. Dengan menggunakan SMOTE, penelitian ini secara

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

efektif mengatasi masalah data yang tidak seimbang yang sering diabaikan dalam studi terdahulu, sehingga model menjadi lebih sensitif dan adil dalam memprediksi kehilangan karyawan. Penelitian ini juga menonjol dalam evaluasi kinerjanya yang lebih komprehensif, menggunakan metrik seperti *precision*, *recall*, F1-*score* yang memberikan gambaran lebih lengkap mengenai kinerja model. Dengan kontribusi-kontribusi tersebut, penelitian ini menawarkan solusi yang efektif untuk masalah prediksi kehilangan karyawan, memberikan pendekatan yang lebih holistik dan hasil yang lebih dapat diandalkan bagi organisasi yang ingin mengurangi *turnover* karyawan dan mempertahankan talenta terbaik.

Penelitian ini menggunakan metodologi yang terstruktur dengan mengintegrasikan tiga teknik utama yaitu *Recursive Feature Elimination* (RFE), *Synthetic Minority Over-sampling Technique* (SMOTE), dan *Adaptive Boosting* (AdaBoost). Langkah pertama, RFE digunakan untuk memilih fitur-fitur yang paling relevan dalam dataset, mengurangi kompleksitas model dan meningkatkan interpretabilitas dengan hanya mempertahankan variabel yang signifikan. Selanjutnya, SMOTE diterapkan untuk menangani masalah ketidakseimbangan kelas dengan membuat sampel sintetis dari kelas minoritas (karyawan yang keluar), sehingga dataset menjadi lebih seimbang dan model dapat mendeteksi karyawan yang berisiko tinggi dengan lebih adil. Terakhir, AdaBoost digunakan untuk meningkatkan akurasi prediksi dengan menggabungkan beberapa model dasar yang lemah. Setiap model dasar difokuskan pada instance yang sulit diprediksi oleh model sebelumnya, sehingga secara bertahap meningkatkan kinerja keseluruhan.

Pemilihan metode-metode ini didasarkan pada alasan spesifik untuk mengatasi tantangan dalam prediksi kehilangan karyawan. RFE dipilih karena efektif dalam mengidentifikasi dan mempertahankan fitur yang paling relevan, sehingga mengurangi kompleksitas dan risiko *overfitting*, serta meningkatkan interpretabilitas model. SMOTE diterapkan untuk menangani masalah ketidakseimbangan kelas dengan membuat sampel sintetis dari kelas minoritas, meningkatkan sensitivitas model terhadap karyawan yang berisiko tinggi untuk keluar. AdaBoost digunakan untuk meningkatkan akurasi prediksi dengan menggabungkan beberapa model dasar yang lemah menjadi satu model yang kuat, serta menargetkan *instance* yang sulit diprediksi guna mengurangi kesalahan prediksi secara bertahap. Dengan mengintegrasikan ketiga metode ini, penelitian ini mampu mengatasi tantangan utama seperti kompleksitas data, ketidakseimbangan kelas, dan rendahnya akurasi prediksi, sehingga menghasilkan model yang lebih robust, akurat, dan dapat diandalkan untuk membantu organisasi mengurangi *turnover* karyawan dan mempertahankan talenta terbaik.

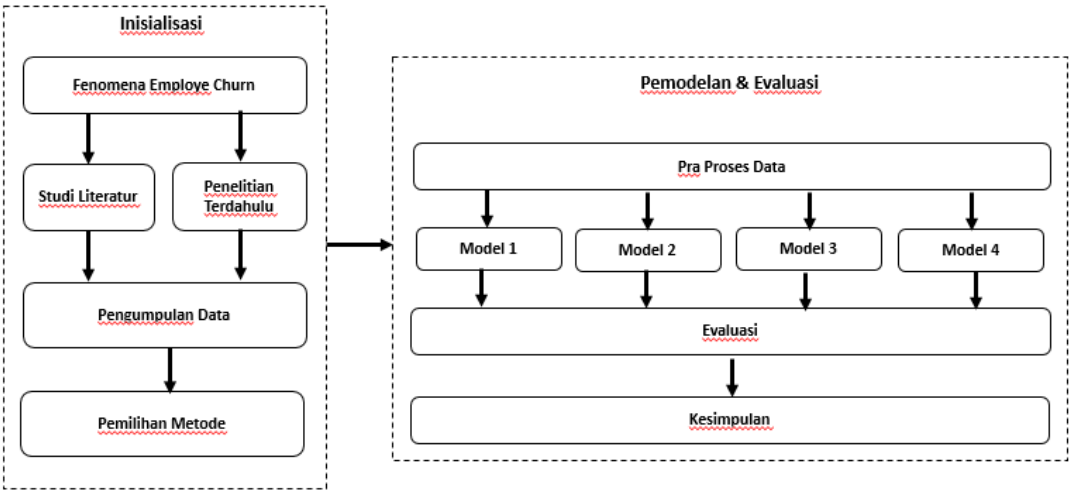
II. METODOLOGI PENELITIAN

Peneliti telah melakukan eksplorasi literatur dari berbagai sumber pengetahuan yang relevan, memperoleh landasan yang kokoh untuk penelitian ini. Melalui penelitian ini tingkat keakuratan yang tinggi dapat dicapai, didukung oleh dasar penjelasan yang kuat dan dapat dipertanggungjawabkan. Sejumlah penelitian sebelumnya telah menginvestigasi prediksi kehilangan karyawan dengan memanfaatkan metode data mining dan machine learning di berbagai sektor. Dalam penelitian ini, pendekatan yang diadopsi adalah penerapan teknik machine learning, dengan fokus pada seleksi fitur, oversampling, dan penggunaan algoritma ensemble Adaboost. Metode yang diujicobakan meliputi kombinasi RFE, SMOTE, dan ADABOOST.

A. Kerangka Pemikiran

Kerangka pemikiran adalah suatu struktur konseptual yang memberikan pandangan yang terperinci dan sistematis mengenai proses analisis data, mulai dari langkah-langkah awal pengumpulan, pengolahan data hingga tahap akhir yaitu pencapaian kesimpulan dari data yang telah dianalisis. Gambar 1 menampilkan secara visual tahapan-tahapan yang tercakup dalam kerangka pemikiran dalam penelitian ini, yang membantu dalam memandu jalannya proses analisis dengan lebih terarah dan efektif.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145



Gambar 1. Kerangka Pemikiran

Dalam penelitian ini, fenomena kehilangan pegawai, atau employee churn, dipelajari melalui langkah-langkah yang sistematis. Studi literatur dilakukan untuk memahami faktor-faktor yang mempengaruhi keputusan pegawai untuk berhenti, diikuti dengan pengumpulan data yang relevan. Data yang telah diproses kemudian diuji menggunakan 4 kombinasi metode yang dipilih, dan evaluasi dilakukan untuk membandingkan kinerja mereka dalam memprediksi employee churn. Hasil evaluasi memberikan pemahaman yang lebih dalam tentang faktor-faktor yang mempengaruhi churn dan efektivitas berbagai metode dalam memprediksinya, yang kemudian digunakan untuk menyimpulkan temuan dan memberikan rekomendasi untuk penelitian dan praktik di masa depan.

B. *Adaboost*

Adaboost, atau Adaptive Boosting, adalah algoritma ensemble learning yang menggabungkan beberapa model lemah menjadi satu model kuat untuk meningkatkan akurasi prediksi. Dikembangkan oleh Yoav Freund dan Robert Schapire pada tahun 1996, Adaboost bekerja dengan menyesuaikan bobot sampel secara iteratif, memberikan bobot lebih besar pada sampel yang sulit diklasifikasikan. Setiap model lemah, biasanya berupa pohon keputusan sederhana, dilatih pada data berbobot dan digabungkan dengan bobot tertentu berdasarkan akurasinya. AdaBoost memiliki keunggulan sebagai algoritma yang cepat, sederhana, dan mudah diimplementasikan. Namun, algoritma ini juga memiliki beberapa kelemahan, termasuk ketergantungan pada data dan algoritma dasar yang digunakan, serta kepekaan terhadap noise. [18]

C. *Hyperparameter*

Hyperparameter dalam Adaboost adalah aspek penting yang memengaruhi kinerja dan kekuatan model ensemble yang dihasilkan. Salah satu *hyperparameter* utama adalah n_*estimators*, yang mengatur jumlah model lemah yang digunakan dalam ensemble. Sedangkan learning_rate mengontrol seberapa banyak setiap model lemah berkontribusi terhadap model keseluruhan. Selain itu, pemilihan *base_estimator* juga berpengaruh besar, di mana pohon keputusan sering kali menjadi pilihan *default*. Untuk menentukan *hyperparameter* yang optimal, teknik seperti *cross-validation*, *grid search,* atau *random search* sering digunakan. Penyetelan *hyperparameter* yang tepat memainkan peran kunci dalam menemukan keseimbangan antara bias dan varians model, serta meningkatkan kinerja dan generalisasi model *Adaboost*. Dengan memahami dan menyesuaikan *hyperparameter* sesuai dengan karakteristik *dataset* dan kebutuhan aplikasi, *Adaboost* dapat menghasilkan model *ensemble* yang kuat dan efektif dalam menangani berbagai masalah. Dalam penelitian ini parameter *n_estimators* terbaik adalah 150 sedangkan untuk parameter *learning_rate* terbaik adalah 0,5.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

D. Recursive Feature Elimination (RFE)

RFE adalah teknik seleksi fitur dalam machine learning yang bertujuan meningkatkan kinerja model dengan memilih fitur-fitur yang paling relevan. RFE bekerja dengan melatih model pada semua fitur, kemudian mengeliminasi fitur yang paling tidak penting secara berulang hingga hanya tersisa fitur yang paling signifikan. Meskipun efektif dalam mengoptimalkan model dan mengurangi overfitting, RFE dapat memakan waktu komputasi yang signifikan karena memerlukan pelatihan ulang model berulang kali. RFE banyak diterapkan dalam berbagai bidang, termasuk genomik, pengenalan pola, dan analisis data keuangan, untuk memilih fitur-fitur yang memberikan kontribusi terbesar terhadap prediksi model. Recursive Feature Elimination (RFE) adalah algoritma seleksi fitur yang semakin populer dalam berbagai aplikasi machine learning, khususnya untuk klasifikasi dan prediksi. [19].

RFE menggunakan algoritma pembelajaran mesin sebagai estimator, dalam penelitian ini menggunakan model Adaboost Classifier, untuk membantu memilih fitur. Estimator ini digunakan untuk memberikan skor kepentingan pada setiap fitur dan memilih fitur yang paling penting. Hasil RFE kemudian diterbitkan untuk mengetahui fitur yang dipilih dan bobotnya. Fitur yang dipilih bersama dengan bobotnya kemudian disimpan dalam sebuah dataframe untuk menampilkan hasil. Dataframe ini disortir berdasarkan bobotnya untuk menampilkan fitur yang paling relevan.

Langkah pertama adalah mendefinisiakn *instance* RFE dengan menggunakan estimator yang telah dipilih dan jumlah fitur yang ingin dipilih, dalam penelitian ini jumlah fitur yang dipilih adalah 10. RFE kemudian dilakukan *fit* terhadap data pelatihan untuk memilih fitur yang paling relevan. Hasil RFE kemudian diterbitkan untuk mengetahui fitur yang dipilih dan bobotnya. Fitur yang dipilih bersama dengan bobotnya kemudian disimpan dalam sebuah dataframe untuk menampilkan hasil. Dataframe ini disortir berdasarkan bobotnya untuk menampilkan fitur yang paling relevan. Selanjutnya, data yang telah dipilih fiturnya kemudian diubah menjadi bentuk yang sesuai untuk model Adaboost. Model Adaboost kemudian dilakukan *fit* terhadap data pelatihan yang telah dipilih fiturnya atau dilanjutkan proses SMOTE terlebih dahulu sebelum dilakukan poses *fit*. Hasil prediksi kemudian diterbitkan untuk mengetahui performansi model.

E. Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE adalah metode *oversampling* yang digunakan dalam pengolahan data untuk menangani ketidakseimbangan kelas pada dataset. Teknik ini fokus pada kelas minoritas dengan membuat sampel sintetis baru, sehingga mengimbangi perbandingan antara kelas mayoritas dan minoritas.

Tujuannya adalah mengurangi ketidakseimbangan antara kelas dalam masalah klasifikasi. Terutama ketika kelas minoritas memiliki sedikit sampel, SMOTE hadir sebagai pilihan yang sangat berguna. Hal yang menarik, SMOTE dapat digunakan bersama berbagai algoritma pembelajaran mesin. Ini seperti memberikan alat tambahan kepada model untuk lebih baik menangani ketidakseimbangan kelas. Namun, keberhasilan SMOTE bergantung pada konteks data dan permasalahan spesifik. Ada situasi dimana penggunaan SMOTE dapat memberikan peningkatan kinerja model yang signifikan, tetapi ada juga kasus di mana pendekatan ini mungkin tidak sepenuhnya cocok. Penting untuk berhati-hati karena terdapat kemungkinan munculnya kesalahan, terutama jika SMOTE digunakan tanpa mempertimbangkan dengan cermat karakteristik data yang dihadapi. SMOTE dapat meningkatkan akurasi klasifikasi untuk kelas minoritas dengan cara memperluas wilayah keputusan *classifier* dan memperkenalkan bias ke arah kelas minoritas. [20]

Dalam proses SMOTE, langkah pertama adalah menghitung jarak antara setiap sampel dalam kelas minoritas. Setelah jarak dihitung, satu sampel minoritas dipilih sebagai titik awal, dan kemudian k tetangga terdekat dari sampel tersebut diidentifikasi menggunakan algoritma K-NN. Selanjutnya, satu tetangga dipilih secara acak dari k tetangga terdekat. Langkah penting berikutnya adalah menghitung rasio *oversampling*, yang umumnya berupa angka antara 0 dan 1, untuk menentukan seberapa banyak sampel sintetis yang akan dibuat. Sampel sintetis kemudian dibuat dengan menghitung perbedaan antara sampel minoritas yang dipilih dan tetangga yang dipilih, mengalikan perbedaan tersebut dengan rasio oversampling, dan menambahkan hasil perkalian pada sampel minoritas yang dipilih. Proses ini diulang sejumlah iterasi yang diinginkan atau hingga jumlah sampel sintetis mencapai target. Setelah itu, dataset asli digabungkan dengan sampel sintetis yang telah dibuat untuk membentuk dataset baru. Akhirnya, untuk menghilangkan pola tertentu dalam urutan data, dataset diacak ulang. Seluruh proses SMOTE dirancang untuk meningkatkan representasi kelas minoritas dan memperbaiki ketidakseimbangan dalam dataset. Nilai k yang digunakan dalam penelitian ini adalah 5.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

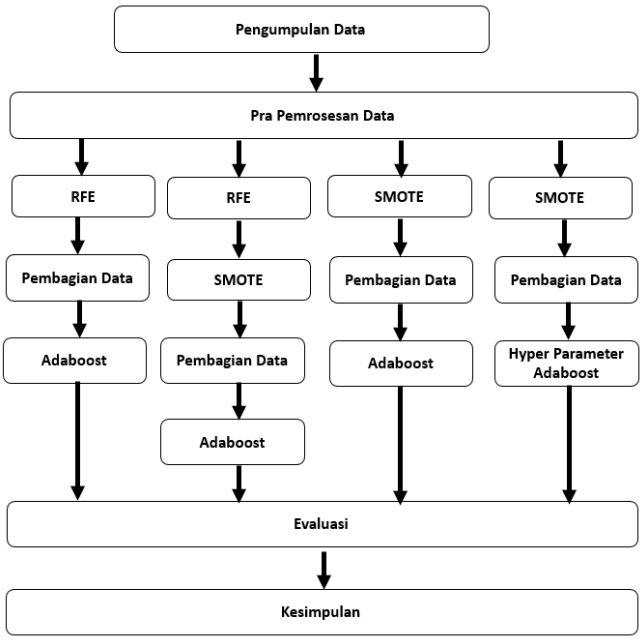
F. Evaluasi Kinerja

Penelitian ini menggunakan evaluasi kinerja *Confusion Matrix*, melibatkan beberapa metrik yang digunakan untuk mengevaluasi kinerja model klasifikasi. Akurasi adalah metrik yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan menghitung jumlah prediksi yang benar dan salah. Akurasi dapat dihitung menggunakan rumus Akurasi = (*True Positive* + *True Negative*) / (*Total Data*). Presisi adalah metrik yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan menghitung jumlah prediksi positif yang benar. Presisi dapat dihitung menggunakan rumus Presisi = (*True Positive*) / (*True Positive* + *False Positive*). Recall adalah metrik yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan menghitung jumlah prediksi positif yang benar. Recall dapat dihitung menggunakan rumus Recall = (*True Positive*) / (*True Positive* + *False Negative*). sedangkan F1-Score adalah metrik yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan menghitung rata-rata dari presisi dan recall. F1-Score dapat dihitung menggunakan rumus berikut: F1-Score = 2 * (*Precision* * *Recall*) / (*Precision* + *Recall*). Metrik evaluasi kinerja menggunakan *Confusion Matrix* dapat dihitung dengan menggunakan data yang diperoleh dari hasil prediksi model klasifikasi.

III. HASIL DAN PEMBAHASAN

Pada bab ini, akan dibahas secara mendetail hasil dari analisis yang telah dilakukan serta interpretasi dari hasil tersebut. Pembahasan dimulai dari proses pengumpulan data, dilanjutkan dengan pra-pemrosesan data menggunakan *Standard Scaler*, dan pemrosesan dengan empat kombinasi model yang berbeda. Selanjutnya, performa setiap model dievaluasi berdasarkan metrik akurasi, presisi, *recall*, dan F1-*Score*. Evaluasi ini bertujuan untuk menilai efektivitas masing-masing pendekatan dan menentukan model yang paling optimal untuk masalah yang dihadapi. Hasil perbandingan tersebut kemudian dianalisis untuk memberikan pemahaman yang lebih mendalam mengenai kinerja setiap model serta implikasi praktisnya dalam konteks penggunaan.

A. Analisis Rancangan Sistem



Gambar 2. Analisis Rancangan Sistem

Pada Gambar 2, setelah proses pengumpulan data selesai, langkah berikutnya adalah melakukan pra-pemrosesan data. Pra-pemrosesan ini sangat penting untuk memastikan data dalam kondisi optimal untuk analisis lebih lanjut. Salah satu teknik yang akan digunakan adalah *Standard Scaler*, yang berfungsi untuk menstandarkan fitur-fitur dalam dataset agar memiliki distribusi dengan nilai rata-rata nol dan standar deviasi satu. Teknik ini membantu mencegah dominasi fitur

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

dengan skala besar terhadap fitur dengan skala kecil, sehingga semua fitur memiliki kontribusi yang seimbang dalam analisis model. Setelah pra-pemrosesan, data akan diproses menggunakan empat kombinasi model yang berbeda.

Model pertama adalah RFE diikuti dengan pembagian data dan kemudian menggunakan *AdaBoost*. RFE digunakan untuk memilih fitur-fitur yang paling relevan dengan mengeliminasi fitur secara berulang berdasarkan pentingnya fitur dalam model. Setelah seleksi fitur, data dibagi menjadi data latih dan data uji, dan kemudian model AdaBoost diterapkan untuk membangun model prediksi yang kuat. Model kedua menambahkan langkah SMOTE setelah RFE dan sebelum pembagian data dan penerapan *AdaBoost*. SMOTE digunakan untuk mengatasi ketidakseimbangan kelas dengan menghasilkan sampel sintetis dari kelas minoritas, sehingga model dapat lebih efektif dalam memprediksi kelas yang kurang terwakili. Model ketiga mengikuti urutan SMOTE diikuti dengan pembagian data dan penerapan *AdaBoost* tanpa melalui proses RFE terlebih dahulu. Terakhir, model keempat adalah kombinasi RFE, SMOTE, pembagian data, dan penerapan AdaBoost dengan penyesuaian *hyperparameter* untuk mengoptimalkan performa model.

AdaBoost memiliki beberapa kelebihan yang membedakannya dari algoritma lain seperti *Random Forest* atau *Gradient Boosting*. Salah satu kelebihan utama adalah toleransi terhadap noise, karena algoritma ini dapat menyesuaikan beratnya terhadap setiap instance dalam dataset, sehingga memungkinkan model untuk lebih fokus pada *pattern* yang lebih signifikan dan mengabaikan noise yang tidak berarti. Selain itu, AdaBoost juga lebih baik dalam menghadapi *dataset* yang *imbalance*, seperti pada kasus di mana *minority class* sangat kecil, dengan menggunakan SMOTE yang membangkitkan data sintetik untuk *minority class.* AdaBoost juga lebih fleksibel dalam pemilihan model base, sehingga memungkinkan pengguna untuk memilih *model base* yang paling sesuai dengan *dataset* yang digunakan. Kinerja AdaBoost juga lebih baik pada data yang berdimensi tinggi dan data yang berkelompok, karena algoritma ini dapat menyesuaikan beratanya terhadap setiap *feature* dan mengabaikan *feature* yang tidak relevan. Dengan demikian, kelebihan-kelebihan ini membuat AdaBoost menjadi pilihan dalam penelitian ini

RFE merupakan algoritma yang digunakan untuk mengurangi jumlah fitur dalam dataset yang digunakan untuk pelatihan model *machine learning*. Dalam RFE, fitur-fitur yang kurang penting dieliminasi secara berurutan, dan model yang dihasilkan memiliki fitur-fitur yang lebih relevan. Proses RFE dimulai dengan memilih model *machine learning* yang akan digunakan untuk pelatihan, seperti Adaboost. Kemudian, setiap fitur dalam dataset dievaluasi menggunakan metrik yang relevan, Fitur-fitur yang memiliki korelasi terendah dengan target variable kemudian dieliminasi, dan proses ini diulangi hingga jumlah fitur yang diinginkan telah dicapai.

SMOTE adalah teknik yang digunakan untuk meningkatkan jumlah data kelas minoritas dalam dataset yang tidak seimbang. Dalam SMOTE, data kelas minoritas digandakan secara sintetis untuk mencapai jumlah yang lebih dekat dengan kelas mayoritas. Proses SMOTE dimulai dengan mengidentifikasi kelas minoritas dalam *dataset* yang tidak seimbang. Kemudian, jarak antara data kelas minoritas dengan data kelas mayoritas yang terdekat dihitung. Data kelas minoritas kemudian digandakan secara sintetis menggunakan jarak yang dihitung, dan data sintetis tersebut ditambahkan ke dataset asli.

Tuning hyperparameter Adaboost dilakukan untuk menemukan kombinasi nilai *hyperparameter* yang optimal untuk model Adaboost. Dalam tuning *hyperparameter*, langkah pertama adalah memilih model Adaboost yang akan digunakan untuk pelatihan. Kemudian, metode tuning hyperparameter yang akan digunakan adalah *grid search*. Setelah itu, jangkauan nilai untuk setiap *hyperparameter* yang akan dioptimalkan ditentukan. Model Adaboost kemudian dilatih dengan setiap kombinasi nilai *hyperparameter* yang ditentukan, dan hasil pelatihan model dievaluasi menggunakan metrik yang relevan, seperti akurasi atau F1-*score*. Kombinasi nilai *hyperparameter* yang memberikan hasil terbaik kemudian dipilih sebagai hasil *tuning hyperparameter*.

Kombinasi RFE, SMOTE, dan hyperparameter tuning Adaboost digunakan untuk meningkatkan akurasi model Adaboost pada dataset yang tidak seimbang. Dalam kombinasi ini, RFE digunakan untuk mengurangi jumlah fitur yang tidak relevan, SMOTE digunakan untuk meningkatkan jumlah data kelas minoritas, dan *hyperparameter tuning* digunakan untuk menemukan kombinasi nilai *hyperparameter* yang optimal. Dengan menggunakan kombinasi ini, model Adaboost dapat meningkatkan akurasi dan performa pada dataset yang tidak seimbang.

Setiap model akan dievaluasi berdasarkan metrik akurasi, presisi, *recall*, dan F1-*Score*. Akurasi mengukur persentase prediksi yang benar, sedangkan presisi mengukur kemampuan model dalam mengidentifikasi sampel positif dengan benar dari semua prediksi positif. *Recall* mengukur kemampuan model dalam menemukan semua sampel positif yang sebenarnya dari semua sampel positif, dan F1-*Score* memberikan gambaran keseimbangan antara presisi dan *recall*.

Evaluasi yang komprehensif ini memungkinkan kita untuk menilai performa setiap model secara menyeluruh. Setelah evaluasi, hasil dari keempat model akan dibandingkan untuk menentukan model yang memberikan performa terbaik secara keseluruhan. Perbandingan ini akan didasarkan pada nilai akurasi, presisi, *recall*, dan F1-*Score*. Dengan melakukan analisis dan perbandingan yang teliti, kita dapat memilih model yang paling optimal dan memberikan solusi terbaik untuk masalah yang dihadapi. Proses ini memastikan bahwa model yang dipilih tidak hanya memiliki kinerja yang baik tetapi juga mampu mengatasi tantangan spesifik dari dataset yang digunakan.

B. Batasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan untuk interpretasi hasil dan pengembangan lebih lanjut. Pertama, keterbatasan dataset menjadi salah satu faktor utama. Dataset yang digunakan mungkin tidak cukup representatif atau memiliki variasi yang cukup untuk mencerminkan kondisi yang lebih umum di berbagai industri atau lokasi geografis. Selain itu, *dataset* mungkin memiliki jumlah sampel yang terbatas, terutama untuk kelas minoritas, yang meskipun telah diatasi dengan teknik SMOTE, masih dapat mempengaruhi keakuratan model. Kedua, keterbatasan dalam model yang digunakan juga perlu diperhatikan. Meskipun kombinasi RFE, SMOTE, dan AdaBoost menunjukkan peningkatan kinerja, setiap model memiliki asumsi dan parameter yang dapat mempengaruhi hasil. Terakhir, batasan dalam evaluasi model. Meskipun metrik kinerja seperti akurasi, presisi, *recall*, dan f1-*score* telah digunakan, mereka mungkin tidak cukup untuk menangkap semua aspek dari kinerja model dalam konteks dunia nyata. Model yang baik di satu metrik mungkin tidak selalu unggul di metrik lain, dan penting untuk mempertimbangkan keseimbangan antara berbagai metrik ini saat mengevaluasi hasil.

C. Pengumpulan Data

Pengumpulan data dilakukan melalui situs https://www.kaggle.com/, yang menyediakan dataset berisi 1.470 baris dan 34 kolom. Kolom-kolom dalam dataset ini mencakup berbagai aspek yang meliputi usia (*age*), jenis perjalanan bisnis (*Business Travel*), tarif harian (*Daily Rate*), departemen (*Department*), jarak dari rumah (*Distance From Home*), tingkat pendidikan (*education*), bidang pendidikan (*Education Field*), jumlah karyawan (*Employee Count*), nomor karyawan (*Employee Number*), kepuasan lingkungan kerja (*Environment Satisfaction*), jenis kelamin (*Gender*), tarif per jam (Hourly Rate), keterlibatan kerja (Job Involvement), tingkat pekerjaan (Job Level), peran pekerjaan (Job Role), kepuasan kerja (Job Satisfaction), status pernikahan (Marital Status), pendapatan bulanan (Monthly Income), tarif bulanan (Monthly Rate), jumlah perusahaan yang pernah bekerja (Num Companies Worked), lembur (Over Time), persentase kenaikan gaji (Percent SalaryHike), penilaian kinerja (Performance Rating), kepuasan hubungan (Relationship Satisfaction), jam standar (Standard Hours), tingkat opsi saham (Stock Option Level), total tahun bekerja (Total Working Years), pelatihan tahun terakhir (Training Times Last Year), keseimbangan kehidupan kerja (Work Life Balance), tahun di perusahaan (Years At Company), tahun dalam peran saat ini (Years In Current Role), tahun sejak promosi terakhir (Years Since Last Promotion), dan tahun dengan manajer saat ini (Years With Curr Manager). Dataset ini digunakan untuk menganalisis faktor-faktor yang mempengaruhi loyalitas pegawai, membantu mengidentifikasi pola yang mendahului keputusan pegawai untuk berhenti dari perusahaan.

D. Kondisi Data

Berdasarkan analisis statistik deskriptif terhadap dataset yang terlihat pada Tabel 1, terdapat 1.470 data karyawan dengan berbagai informasi terkait. Usia rata-rata karyawan adalah 36,92 tahun, dengan rentang usia antara 18 hingga 60 tahun. Mayoritas karyawan (1,393) berada pada kategori 1 (perjalanan jarang) dalam perjalanan bisnis, dengan nilai maksimum 3 (perjalanan sering). Tarif harian karyawan bervariasi, dengan rata-rata 802,49 dan rentang nilai dari 102 hingga 1499. Distribusi karyawan di antara beberapa departemen, dengan nilai rata-rata departemen adalah 1,74, menunjukkan sebagian besar karyawan berada di departemen kategori 1 dan 2. Rata-rata lama bekerja dalam peran saat ini adalah 4,23 tahun, dengan lama bekerja berkisar dari 0 hingga 18 tahun. Rata-rata lama sejak promosi terakhir adalah 2,19 tahun, dengan rentang dari 0 hingga 15 tahun. Lamanya karyawan bekerja dengan manajer saat ini bervariasi, dengan rata-rata 4,12 tahun dan rentang dari 0 hingga 17 tahun. Tingkat attrisi menunjukkan bahwa sekitar 16,12% dari total karyawan mengundurkan diri, sementara mayoritas karyawan (83,88%) tetap bekerja di perusahaan.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

Data ini memberikan gambaran menyeluruh mengenai kondisi demografis dan pengalaman kerja karyawan di perusahaan tersebut.

TABEL 1. STATISTIK DESKRIPTIF

	age	Business Travel	Daily Rate	Department	…	YearsIn Current Role	YearsSince Last Promotion	YearsWith Curr Manager	Attrition
count	1.470	1.470	1.470	1.470	…	1.470	1.470	1.470	1.470
mean	36.923	1.393	802.486	1.739	…	4.229	2.188	4.123	0,111805556
std	9.135	0,461805556	403.509	0,366666667	…	3.623	3.222	3.568	0,255555556
min	18	1	102	1	…	0	0	0	0
25%	30	1	465	1	…	2	0	2	0
50%	36	1	802	2	…	3	1	3	0
Max	60	3	1499	3	…	18	15	17	1

TABEL 2. DISTRIBUSI TARGET

Kelas	Jumlah	%
0	1.233	84%
1	237	16%

Pada Tabel 2 memberikan gambaran tentang distribusi kelas target dalam dataset. Terdapat dua kelas yang diamati: kelas 0 dan kelas 1. Jumlah observasi yang tercatat untuk kelas 0 adalah 1.233, yang merupakan sekitar 84% dari total data. Sementara itu, jumlah observasi untuk kelas 1 adalah 237, yang menyumbang sekitar 16% dari total data. Distribusi persentase ini memberikan informasi penting tentang seimbangnya dataset dalam hal kelas target. Dengan 84% data pada kelas 0 dan 16% pada kelas 1, dapat disimpulkan bahwa dataset cenderung memiliki ketidakseimbangan kelas, di mana kelas mayoritas (kelas 0) jauh lebih dominan dibandingkan kelas minoritas (kelas 1). Hal ini penting untuk diperhatikan dalam analisis lebih lanjut, terutama dalam konteks pembuatan model prediksi di mana ketidakseimbangan kelas dapat memengaruhi kinerja model dan interpretasi hasilnya.

E. Pra Pemrosesan Data

Prapemrosesan data dengan normalisasi adalah langkah krusial dalam analisis data yang bertujuan untuk mengubah data ke dalam skala yang seragam sehingga memudahkan dalam analisis dan pembuatan model prediktif. Normalisasi membantu dalam menghilangkan perbedaan skala antar fitur-fitur dalam dataset dengan memastikan bahwa setiap fitur memiliki skala yang sama. Proses normalisasi seringkali melibatkan penskalaan fitur-fitur dalam rentang tertentu, seperti mengubah nilai-nilai fitur sehingga mereka memiliki rata-rata nol dan standar deviasi satu. Dengan melakukan hal ini, normalisasi memungkinkan model yang dibangun dari data tersebut untuk tidak dipengaruhi oleh perbedaan skala dan memperlakukan setiap fitur dengan adil. Hal ini membantu mencegah fitur-fitur dengan skala besar mendominasi proses pembelajaran mesin. Dalam Penelitian ini , normalisasi menggunakan tehnik StandardScaler yang secara otomatis menangani normalisasi data dengan melakukan penskalaan fitur-fitur sesuai dengan mean dan standar deviasi mereka. Dengan melakukan prapemrosesan data seperti ini, kita dapat memastikan bahwa data yang digunakan dalam analisis atau pembuatan model prediktif memiliki distribusi yang seragam dan dapat diinterpretasikan dengan baik. Sebagai contoh pada Tabel 3 di bawah ini

TABEL 3. CONTOH HASIL NORMALISASI

Age	BusinessTravel	DailyRate	Department	DistanceFromHome

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

0,309722222	-0.590	0,515277778	-1.401	-10.109
13.223	0,634027778	-12.977	0,342361111	-0.147
0.008	-0.590	14.143	0,342361111	-0.887
-0.429	0,634027778	14.614	0,342361111	-0.7641
-10.866	-0.590	-0.524	0,342361111	-0.887

Dalam penelitian ini alasan menggunakan *Standard Scaler* karena kelebihannya dalam menghadapi *outliers*, mempertahankan distribusi data asli, meningkatkan kinerja model, memberikan flexibilitas, dan memberikan interpretasi yang lebih baik dan tidak mengubah distribusi data asli

F. Model *Recursive Feature Elimination* (RFE) - *Adaboost*

Model kombinasi Recursive Feature Elimination (RFE) - Adaboost merupakan pendekatan yang menggabungkan dua teknik dalam proses pemodelan.Pertama, RFE digunakan untuk memilih fitur-fitur terbaik dari dataset, Ini dilakukan dengan menghitung pentingnya setiap fitur dan menghapus yang paling tidak penting, kemudian memperbarui model dengan fitur yang tersisa. Proses ini diulangi sampai jumlah fitur yang diinginkan tercapai. RFE membantu mengurangi dimensi data dan meningkatkan interpretasi model dengan mempertahankan fitur-fitur yang paling relevan.

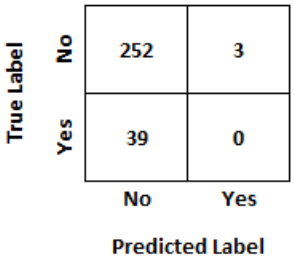
TABEL 4. HASIL PROSES RFE

No	*Feature*	*Weight*
1	*DailyRate*	0,18
2	*EmployeeNumber*	0,18
3	*MonthlyIncome*	0,16
4	*YearsWithCurrManager*	0,12
5	*Age*	0,1
6	*PercentSalaryHike*	0,08
7	*BusinessTravel*	0,06
8	*NumCompaniesWorked*	0,04
9	*OverTime*	0,04
10	*YearsSinceLastPromotion*	0,04

Tabel 4 menyajikan hasil dari proses Recursive Feature Elimination (RFE) yang digunakan untuk memilih fitur-fitur terbaik dari dataset. Tabel ini terdiri dari tiga kolom: No, *Feature*, dan *Weight*. Kolom "No" menunjukkan nomor urutan dari setiap fitur yang dipilih, sedangkan kolom "*Feature*" berisi nama fitur-fitur yang terpilih. Kolom "*Weight*" menunjukkan bobot yang diberikan pada setiap fitur yang dipilih oleh RFE, yang menunjukkan seberapa penting fitur tersebut dalam memprediksi target. Hasilnya menunjukkan bahwa RFE telah memilih 10 fitur terbaik dari dataset. Fitur-fitur tersebut, seperti "*DailyRate*," "*EmployeeNumber*," dan "*MonthlyIncome*," diberi bobot yang cukup tinggi, menunjukkan bahwa fitur-fitur ini memiliki kontribusi yang signifikan dalam mempengaruhi target. Fitur-fitur lainnya, seperti "*YearsWithCurrManager*," "*age*," dan "*PercentSalaryHike*," juga memiliki bobot yang cukup besar, meskipun tidak sebesar fitur-fitur utama.

Proses dilanjutkan dengan prapemrosesan data dengan StandardScaler untuk menormalkan data, diikuti oleh seleksi fitur menggunakan RFE. Setelah itu, model Adaboost dibangun untuk prediksi. Evaluasi dilakukan menggunakan metrik seperti akurasi, presisi, recall, dan f1-score untuk memahami kinerja model dalam memprediksi churn karyawan.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

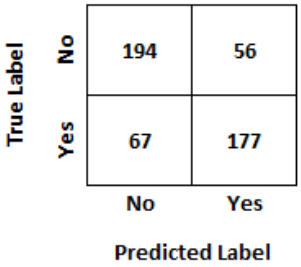


Gambar 1. Confusion Matrix RFE dengan Adaboost

Terlihat Pada Gambar 1 merupakan Confusion Matrix dari hasil evaluasi metode kombinasi RFE dengan AdaBoost, Hasil evaluasi ini menunjukkan metode RFE - ADABOOST memiliki akurasi sebesar 0,857142857. Namun, perlu diperhatikan bahwa presisi, recall, dan f1 score memiliki nilai hasil 0, yang menunjukkan adanya kebutuhan untuk mengevaluasi ulang metode tersebut. Diperlukan perbaikan dalam hal presisi dan recall agar metode ini dapat memberikan hasil yang lebih dapat diandalkan.

G. Kombinasi metode RFE, SMOTE dan Adaboost

Setelah proses normalisasi dan pemilihan fitur pada model kombinasi ini, langkah berikutnya adalah menerapkan proses SMOTE untuk mengatasi ketidakseimbangan antara kelas-kelas data. Setelah berhasil mencapai keseimbangan antara kelas, dilakukan pembagian data menjadi data latih dan data uji dengan perbandingan 80:20. Selanjutnya, model Adaboost dibangun menggunakan data latih. Langkah terakhir adalah melakukan evaluasi kinerja model dengan menghitung akurasi, presisi, recall, dan F-1 Score menggunakan data uji.



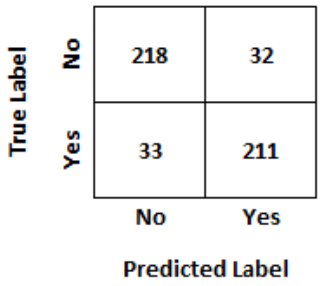
Gambar 2. Confusion Matrix Kombinasi Metode RFE, SMOTE dan Adaboost

Dari *Confusion Matrix* pada Gambar 2, dapat disimpulkan bahwa metode RFE-SMOTE-ADABOOST memiliki tingkat akurasi sebesar 0,771255061. Selain itu, presisi yang dicapai sebesar 0,771784232, recall sebesar 0,762295082, dan f1-score sebesar 0,767010309. Dengan demikian, dapat dikatakan bahwa model ini mampu memberikan prediksi dengan tingkat akurasi yang cukup baik, namun perlu diperhatikan bahwa terdapat ruang untuk peningkatan dalam recall dan f1-score untuk menjadikannya lebih seimbang.

H. Kombinasi metode SMOTE dan Adaboost Tanpa Seleksi Fitur

Metode kombinasi ini menghadirkan pendekatan yang berbeda dengan tidak melibatkan seleksi fitur, melainkan langsung melakukan pemrosesan SMOTE setelah normalisasi. Langkah selanjutnya adalah menggunakan algoritma Adaboost untuk melakukan prediksi. Dengan demikian, pendekatan ini fokus pada penyeimbangan kelas data menggunakan SMOTE dan kemudian memanfaatkan kekuatan algoritma Adaboost untuk melakukan prediksi tanpa proses seleksi fitur yang terpisah.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

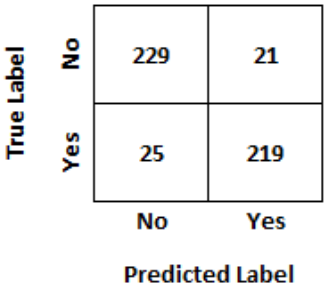


Gambar 3. Confusion Matrix Kombinasi SMOTE dan Adaboost

Dari confusion matrix ini pada Gambar 3, dapat disimpulkan bahwa metode SMOTE-ADABOOST memberikan hasil evaluasi yang sangat baik. Dengan akurasi sebesar 0,868421053, presisi sebesar 0,868312757, recall sebesar 0,864754098, dan f1-score sebesar 0,866529774, metode ini menunjukkan kemampuan prediksi yang tinggi dan seimbang antara presisi dan recall. Dengan demikian, model ini mampu memberikan prediksi yang akurat dan handal dalam memprediksi kelas target.

I. Kombinasi metode SMOTE – AdaBoost *Hyperparemeter*

Dalam penelitian ini, metode ini dilakukan dengan menggunakan pendekatan yang mencari optimasi parameter terbaik pada Adaboost setelah sebelumnya dilakukan proses SMOTE tanpa melibatkan seleksi fitur, dan hasilnya ditampilkan dalam konfusi matrix pada Gambar 4 berikut ini.



Gambar 4. Confusion Matrix Kombinasi SMOTE dan Adaboost Hyper Parameter

Dalam evaluasi ini, akurasi model mencapai 0,907, yang berarti sekitar 0,907 dari semua data yang dianalisis dikelasifikasi dengan benar. Selain itu, presisi model juga mencapai 0,912, yang berarti sekitar 0,912 dari semua data yang dikelasifikasi sebagai positif sebenarnya positif. Dalam hal recall, model mencapai 0,898, yang berarti sekitar 0,898 dari semua data yang sebenarnya positif dikelasifikasi dengan benar. F1 score, yang menggabungkan presisi dan recall, mencapai 0,905, menunjukkan bahwa model memiliki keseimbangan yang baik antara presisi dan recall. Dengan demikian, model ini dapat dianggap efektif dalam mendiagnosis dan memprediksi data dengan tingkat akurasi yang tinggi. Selain itu, dari metode ini juga ditemukan 10 fitur yang paling mempengaruhi terjadinya kehilangan pegawai, yang ditampilkan pada Tabel 5 di bawah ini.

TABEL 5. FITUR YANG BEPENGARUH

Feature	Weight
EducationField	0,09
BusinessTravel	0,08
TrainingTimesLastYear	0,06
YearsWithCurrManager	0,05

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 4, Desember 2024, Pp. 2131-2145

MonthlyIncome	0,05
JobInvolvement	0,04
YearsSinceLastPromotion	0,04
EnvironmentSatisfaction	0,04
NumCompaniesWorked	0,04
JobLevel	0,04

J. *Trade-off* Akurasi, Presisi, *Recall* dan F1-*Score*

Dalam hasil penelitian yang ditunjukan pada tabel 6, beberapa metode prediksi pergantian karyawan telah diuji dan dibandingkan untuk mengetahui performa masing-masing. Dalam analisis ini, kita dapat melihat bahwa setiap metode memiliki kelebihan dan kekurangannya, yang dapat dilihat sebagai *trade-off* antara beberapa indikator kinerja. Dalam beberapa kasus, metode yang memiliki akurasi yang lebih tinggi tidak selalu memiliki performa yang lebih baik dalam beberapa indikator lain, seperti presisi, *recall*, dan F1-*score*. Misalnya, metode RFE-SMOTE-ADABOOST memiliki akurasi yang moderat, yaitu 0,771, namun memiliki nilai presisi, *recall*, dan F1-*score* yang relatif rendah. Hal ini menunjukkan bahwa metode ini memiliki kelebihan dalam menghitung akurasi, tetapi memiliki kekurangan dalam menghitung presisi, *recall*, dan F1-*score*.

TABEL 6. HASIL EVALUASI

Metode	Akurasi	Presisi	Recall	F1-Score
RFE-SMOTE-ADABOOST	0,771	0,772	0,762	0,767
RFE-ADABOOST	0,857	0	0	0
SMOTE-ADABOOST	0,868	0,868	0,865	0,867
SMOTE-ADABOOST HYPERPARAMETER	0,907	0,913	0,898	0,905

Sebaliknya, metode RFE-ADABOOST memiliki akurasi yang lebih tinggi, yaitu 0,857, namun gagal dalam mendeteksi kejadian kehilangan karyawan karena memiliki nilai presisi, *recall*, dan F1-*score* yang nol. Hal ini menunjukkan bahwa metode ini memiliki kelebihan dalam menghitung akurasi, tetapi memiliki kekurangan dalam menghitung presisi, *recall*, dan F1-*score*.

Dalam beberapa kasus, metode yang memiliki performa yang lebih baik dalam beberapa indikator tidak selalu memiliki performa yang lebih baik dalam indikator lain. Misalnya, metode SMOTE-ADABOOST memiliki performa yang baik dengan nilai akurasi 0,868, presisi 0,868, recall 0,865, dan F1-score 0,867. Namun, metode ini tidak memiliki performa yang terbaik dalam indikator akurasi jika dibandingkan dengan metode SMOTE-ADABOOST dengan *Hyperparameter*, yang memiliki akurasi 0,907.

Dalam sintesis, analisis *trade-off* dalam metode prediksi pergantian karyawan menunjukkan bahwa setiap metode memiliki kelebihan dan kekurangannya. Oleh karena itu, dalam memilih metode prediksi, perlu mempertimbangkan beberapa indikator kinerja yang relevan dan tidak hanya fokus pada satu indikator saja. Dengan demikian, dapat diperoleh metode yang memiliki performa yang lebih baik secara keseluruhan dan lebih efektif dalam memprediksi pergantian karyawan.

K. Dampak Ketidakseimbangan Kelas

Ketidakseimbangan kelas dalam *dataset* dapat berdampak signifikan pada hasil model prediksi, terutama ketika kelas mayoritas jauh lebih dominan dibandingkan kelas minoritas. Dalam situasi ini, model cenderung lebih fokus pada kelas mayoritas, mengabaikan atau kurang mempelajari kelas minoritas, sehingga menghasilkan akurasi tinggi secara keseluruhan tetapi performa buruk dalam mendeteksi kelas minoritas. Hal ini terlihat dalam hasil metode RFE-ADABOOST, yang meskipun memiliki akurasi tinggi sebesar 0,857, tetapi memiliki nilai presisi, recall, dan f1-score yang nol, menunjukkan kegagalan total dalam mendeteksi kejadian kelas minoritas. Untuk mengatasi masalah ini,

teknik SMOTE digunakan untuk menciptakan sampel sintetis dari kelas minoritas, membuat dataset lebih seimbang dan memungkinkan model untuk belajar lebih baik tentang pola dalam kelas minoritas tersebut. Metode SMOTE-ADABOOST, misalnya, menunjukkan peningkatan signifikan dalam semua metrik: akurasi 0,868, presisi 0,868, recall 0,865, dan f1-score 0,867, yang mengindikasikan kemampuan model yang lebih baik dalam mendeteksi kelas minoritas. Metode SMOTE-ADABOOST *HYPERPARAMETER* bahkan menunjukkan hasil terbaik dengan akurasi 0,907, presisi 0,913, recall 0,898, dan f1-score 0,905, menunjukkan bahwa dengan tuning *hyperparameter*, performa model dapat dioptimalkan lebih lanjut. Dengan demikian, penerapan SMOTE secara jelas membantu mengatasi dampak negatif dari ketidakseimbangan kelas, memungkinkan model untuk memberikan hasil yang lebih akurat dan reliabel dalam mendeteksi kejadian kelas minoritas.

L. Perbandingan Hasil Penelitian sebelumnya

Dalam penelitian ini, berbagai metode prediksi kehilangan karyawan diuji dan dibandingkan dengan hasil yang menunjukkan bahwa metode SMOTE-ADABOOST *HYPERPARAMETER* menghasilkan performa tertinggi dengan akurasi 0,907, presisi 0,913, recall 0,898, dan f1-score 0,905 lebih baik dari sisi akurasi jika dibandingkan dengan penelitian [1] yang menggunakan regresi logistik, KNN, dan XGB. Selain itu, penelitian [4] menggunakan algoritma *Random Forest* dengan *dataset* yang mencakup faktor sosial, budaya, finansial, profesional, dan relasional. Mereka melaporkan bahwa *Random Forest* memiliki akurasi prediksi terbaik. Namun, metode SMOTE-ADABOOST *HYPERPARAMETER* dalam penelitian ini tidak hanya mendapatkan akurasi yang lebih baik hingga 0,907 tetapi juga menunjukkan keseimbangan kinerja yang lebih baik pada metrik presisi, *recall*, dan f1-*score*. Dari perbandingan ini, dapat dilihat bahwa metode SMOTE-ADABOOST *HYPERPARAMETER* tidak hanya unggul dalam hal akurasi tetapi juga menunjukkan kinerja yang lebih seimbang dan andal pada berbagai metrik dibandingkan dengan metode yang digunakan dalam penelitian sebelumnya. Hal ini menunjukkan bahwa penggunaan teknik penyeimbangan data seperti SMOTE dan optimasi parameter dapat secara signifikan meningkatkan performa model prediksi kehilangan karyawan. Dengan demikian, kontribusi penelitian ini terletak pada peningkatan kinerja model secara keseluruhan dan menunjukkan bahwa kombinasi teknik-teknik ini dapat lebih efektif dalam memprediksi pergantian karyawan dibandingkan metode tradisional lainnya.

M. Interprestasi Hasil Penelitian

Berdasarkan hasil penelitian dapat diinterpretasikan bahwa metode yang digunakan untuk memprediksi pergantian karyawan memiliki performa yang berbeda-beda. Metode RFE-SMOTE-ADABOOST menunjukkan performa yang moderat dengan nilai akurasi 0,771, presisi 0,772, recall 0,762, dan F1-score 0,767. Metode ini memiliki performa yang cukup seimbang namun lebih rendah dibandingkan metode lain. Metode RFE-ADABOOST, meskipun memiliki akurasi 0,857, gagal dalam mendeteksi kejadian pergantian karyawan karena memiliki nilai presisi, recall, dan F1-score yang nol. Ini menunjukkan bahwa metode ini tidak mampu memberikan prediksi yang berguna dalam konteks ini. Metode SMOTE-ADABOOST menunjukkan performa yang baik dengan nilai akurasi 0,868, presisi 0,868, recall 0,865, dan F1-score 0,867. Ini menandakan bahwa metode ini cukup efektif dalam memprediksi pergantian karyawan. Metode SMOTE-ADABOOST dengan Hyperparameter mencapai performa terbaik di antara semua metode yang diuji, dengan akurasi 0,907, presisi 0,913, recall 0,898, dan F1-score 0,905. Hal ini menunjukkan bahwa penyesuaian hyperparameter pada SMOTE-ADABOOST meningkatkan kemampuan prediksi model secara signifikan. Dari hasil evaluasi ini, dapat disimpulkan bahwa metode SMOTE-ADABOOST dengan Hyperparameter adalah yang paling efektif dalam memprediksi pergantian karyawan, dengan kinerja terbaik berdasarkan semua metrik yang dievaluasi. Hasil ini dapat digunakan sebagai referensi dalam memilih model machine learning yang paling sesuai untuk memprediksi kehilangan pelanggan. Dari model SMOTE-ADABOOST dengan hyper parameter telah mengidentifikasi 10 faktor kunci yang berpotensi mempengaruhi prediksi pergantian karyawan, termasuk *EducationField, BusinessTravel, Training Times Last Year, Years With Curr Manager, Monthly Income, Job Involvement, Years Since Last Promotion, Environment Satisfaction, Num Companies Worked, dan JobLevel*.

IV. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa Hasil menunjukkan bahwa SMOTE-ADABOOST dengan Hyper parameter mencapai kinerja tertinggi, dengan akurasi 0,907, presisi 0,912, recall 0,898, dan F1-score 0,905.. Selain itu,

model ini juga mengidentifikasi 10 faktor kunci yang berpotensi mempengaruhi prediksi pergantian karyawan, termasuk *EducationField*, *BusinessTravel*, *Training Times Last Year*, *Years With Curr Manager*, *Monthly Income*, *Job Involvement*, *Years Since Last Promotion*, *Environment Satisfaction*, *Num Companies Worked*, dan *JobLevel*. Dalam sintesis, hasil penelitian ini menunjukkan bahwa model SMOTE-ADABOOST Hyper Parameter memiliki performa yang paling baik dalam evaluasi dan dapat digunakan sebagai referensi dalam memprediksi kehilangan pelanggan. Namun, perlu dilakukan penelitian lebih lanjut untuk memverifikasi hasil ini dan mengembangkan model yang lebih akurat.

DAFTAR PUSTAKA

[1] J. Park, Y. Feng, and S. P. Jeong, 'Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques', *Sci Rep*, vol. 14, no. 1, May 2024, doi: 10.1038/s41598-023-50593-4.

[2] N. Leo, B. Tennisson, K. P. Prasad, E. P. Kumar, K. N. Kumar, and U. G. Scholar, 'ANALYSIS AND PREDICTION OF EMPLOYEE ATTRITION', *International Journal of Creative Research Thoughts*, vol. 11. p. 467, 2023. [Online]. Available: www.ijcrt.org

[3] V. Musanga and C. Chibaya, 'A Predictive Model to Forecast Employee Churn for HR Analytics'. pp. 17–30, 2023.

[4] 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.

[5] U. Students, 'EMPLOYEE ATTRITION PREDICTION USING STACKING AND ITS EVALUATION', *International Research Journal of Engineering and Technology*, 2021, [Online]. Available: www.irjet.net

[6] N. Bandyopadhyay and A. Jadhav, 'Churn Prediction of Employees Using Machine Learning Techniques', *Tehnicki Glasnik*, vol. 15, no. 1, pp. 51–59, 2021, doi: 10.31803/tg-20210204181812.

[7] Z. Li and E. Fox, 'Prediction and optimization of employee turnover intentions in enterprises based on unbalanced data', *PLoS One*, vol. 18, no. 8 AUGUST, May 2023, doi: 10.1371/journal.pone.0290086.

[8] M. Lazzari, J. M. Alvarez, and S. Ruggieri, 'Predicting and explaining employee turnover intention', *Int J Data Sci Anal*, vol. 14, no. 3, pp. 279–292, 2022, doi: 10.1007/s41060-022-00329-w.

[9] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, 'Predicting Employee Attrition Using Machine Learning Approaches', *Applied Sciences (Switzerland)*, vol. 12, no. 13, May 2022, doi: 10.3390/app12136424.

[10] K. Naz, I. F. Siddiqui, J. Koo, M. A. Khan, and N. M. F. Qureshi, 'Predictive Modeling of Employee Churn Analysis for IoT-Enabled Software Industry', *Applied Sciences*, vol. 12, no. 20, 2022, doi: 10.3390/app122010495.

[11] F. K. Alsheref, I. E. Fattoh, and W. Mead, 'Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning Algorithms', *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/7728668.

[12] S. F. Sari and K. M. Lhaksmana, 'Employee Attrition Prediction Using Feature Selection with Information Gain and Random Forest Classification', *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 410–419, May 2022, doi: 10.47065/josyc.v3i4.2099.

[13] M. Chaudhary, L. Gaur, N. Jhanjhi, M. Masud, and S. Aljahdali, 'Envisaging Employee Churn Using MCDM and Machine Learning', *Intelligent Automation and Soft Computing*, vol. Vol.33, p. pp.1009-1024, May 2022, doi: 10.32604/iasc.2022.023417.

[14] F. H. Wardhani and K. M. Lhaksmana, 'Predicting Employee Attrition Using Logistic Regression With Feature Selection', *Sinkron : jurnal dan penelitian teknik informatika*, vol. 7, no. 4, pp. 2214–2222, May 2022, doi: 10.33395/sinkron.v7i4.11783.

[15] P. R. Srivastava and P. Eachempati, 'Intelligent Employee Retention System for Attrition Rate Analysis and Churn Prediction: An Ensemble Machine Learning and Multi- Criteria Decision-Making Approach', *Journal of Global Information Management*, vol. 29, no. 6, May 2021, doi: 10.4018/JGIM.20211101.oa23.

[16] M. Subhashini and R. Gopinath, 'Employee Attrition Prediction in Industry Using Machine Learning Techniques', *International Journal of Advanced Research in Engineering and Technology*, vol. 11, no. 12, pp. 3329–3341, 2020, doi: 10.17605/OSF.IO/9XDWE.

[17] J. Kinoto, J. L. Damanik, E. T. S. Situmorang, J. Siregar, and M. Harahap, 'Prediksi Employee Churn Dengan Uplift Modeling Menggunakan Algoritma Logistic Regression', *JURNAL TEKNOLOGI DAN ILMU KOMPUTER PRIMA (JUTIKOMP)*, vol. 3, no. 2, pp. 503–508, May 2020, doi: 10.34012/jutikomp.v3i2.1645.

[18] T. Chengsheng, L. Huacheng, and X. Bing, 'AdaBoost typical Algorithm and its application research', *MATEC Web of Conferences*, vol. 139, p. 222, May 2017, doi: 10.1051/matecconf/201713900222.

[19] A. Priyatno and T. Widiyaningtyas, 'A SYSTEMATIC LITERATURE REVIEW: RECURSIVE FEATURE ELIMINATION ALGORITHMS', *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 2, pp. 196–207, May 2024, doi: 10.33480/jitk.v9i2.5015.

[20] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 'SMOTE: Synthetic Minority Over-sampling Technique', *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, May 2002, doi: 10.1613/jair.953.