

## **EMPLOYEE VOLUNTARY ATTRITION PREDICTION AT PT.XYZ: ENSEMBLE MACHINE LEARNING APPROACH WITH SOFT VOTING CLASSIFIER**

Cagiva Chaedar Bey Lirna<sup>\*1</sup>, Trimono<sup>2</sup>, Aviolla Terza Damaliana<sup>3</sup>

<sup>1,2,3</sup>Data Science, Faculty of Computer, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia  
Email: <sup>1</sup>[20083010020@student.upnjatim.ac.id](mailto:20083010020@student.upnjatim.ac.id), <sup>2</sup>[trimono.stat@upnjatim.ac.id](mailto:trimono.stat@upnjatim.ac.id), <sup>3</sup>[aviolla.terza.sada@upnjatim.ac.id](mailto:aviolla.terza.sada@upnjatim.ac.id)

(Article received: April 25, 2024; Revision: June 04, 2024; published: October 20, 2024)

### **Abstract**

*This research addresses the complexity of employee attrition challenges at PT.XYZ. The main objective is to develop a predictive system for potential voluntary employee attrition by focusing on an in-depth analysis of the factors contributing to attrition at PT.XYZ. The research utilizes data containing information on the job history of PT.XYZ employees from 2018 to 2023. The method employed in the research is a soft voting ensemble classifier model, incorporating SVM, decision tree, and logistic regression, supported by relevant literature. Analysis and exploration of historical data of PT.XYZ employees are conducted to identify key factors influencing employees' decisions to leave the company. Careful data preprocessing is carried out to ensure dataset quality before applying it to the soft voting classifier model. The results of the soft voting classifier modeling used in this research achieve excellent accuracy in both training and testing datasets with respective accuracy percentages of 99% and 98%. Based on the final results of applying the soft voting classifier model, it is expected to provide deep insights and solutions to enhance employee retention at PT.XYZ.*

**Keywords:** EDA, Employee Retention, Ensemble model, Soft Voting, Voluntary Attrition.

## **PREDIKSI ATRISI VOLUNTARY KARYAWAN PT.XYZ: PENDEKATAN ENSEMBLE MACHINE LEARNING DENGAN SOFT VOTING CLASSIFIER**

### **Abstrak**

Penelitian ini merespon kompleksitas tantangan atrisi karyawan PT.XYZ. Tujuan utamanya adalah mengembangkan sistem prediksi potensi atrisi *voluntary* karyawan dengan memfokuskan pada analisis mendalam faktor-faktor atrisi *voluntary* yang terjadi di PT.XYZ. Data yang digunakan pada penelitian menggunakan data yang berisi informasi tentang riwayat pekerjaan karyawan PT.XYZ periode 2018 – 2023. Metode yang diterapkan dalam penelitian adalah model *ensemble soft voting classifier*, mencakup SVM, *decision tree*, dan regresi logistik, yang didukung oleh literatur terkait. Analisis dan eksplorasi data historis karyawan PT.XYZ dilakukan dalam penelitian untuk mengidentifikasi faktor-faktor kunci yang memengaruhi keputusan karyawan meninggalkan perusahaan. Praproses data yang cermat dilakukan untuk memastikan kualitas dataset sebelum diterapkan pada model *soft voting classifier*. Hasil pemodelan *soft voting classifier* yang digunakan dalam penelitian ini mencapai akurasi sangat baik dalam dataset pelatihan dan pengujian dengan presentase akurasi masing – masing sebesar 99% dan 98%. Berdasarkan hasil akhir dari penerapan model *soft voting classifier*, diharapkan dapat memberikan wawasan mendalam dan solusi untuk meningkatkan retensi karyawan PT.XYZ.

**Kata kunci:** Atrisi Voluntary, EDA, Model Ensemble, Retensi Karyawan, Soft Voting.

### **1. PENDAHULUAN**

Kesuksesan dalam lingkungan bisnis yang kompetitif merupakan tujuan utama setiap perusahaan. Dalam proses mencapai kesuksesan tersebut dibutuhkan strategi dan rencana jangka panjang yang matang dengan mempertimbangkan kekuatan dan kelemahan internal perusahaan serta ancaman eksternal yang mungkin akan dihadapi[1]. Salah satu kendala yang sering dihadapi dalam pencapaian tujuan tersebut adalah masalah atrisi

*voluntary* karyawan atau keluarnya karyawan secara sukarela dari perusahaan[2]. Fenomena atrisi ini terjadi karena berbagai alasan yang meliputi pengunduran diri, keseimbangan kehidupan kerja, masalah keluarga, hubungan buruk dengan rekan kerja, dan ketidakmampuan organisasi dalam memenuhi hak yang seharusnya didapat oleh karyawan[3]. Dampak dari atrisi sangat merugikan bagi perusahaan, terutama kehilangan karyawan dengan kemampuan tinggi yang dapat mengubah

dinamika tim yang produktif dan menyebabkan kerugian finansial, khususnya berupa tingginya biaya pencarian, perekrutan, dan pelatihan karyawan baru[4][5].

Perusahaan yang berhasil adalah yang mampu mengatasi tantangan atrisi dengan cara yang efektif. Manajemen tenaga kerja yang baik dalam pengadaan dan pengembangan sumber daya manusia menjadi kunci untuk meminimalkan atrisi *voluntary* karyawan[6]. Dengan strategi yang tepat, perusahaan dapat mempertahankan karyawan yang berharga, meminimalkan gangguan dalam operasi, dan mengurangi biaya yang terkait dengan perekrutan dan pelatihan karyawan baru[7][8].

PT.XYZ merupakan salah satu perusahaan *IT Consultant* di Indonesia yang terdampak adanya fenomena atrisi *voluntary* karyawan. Temuan atrisi yang berhasil tercatat di perusahaan PT.XYZ menunjukkan bahwa jumlah karyawan yang meninggalkan perusahaan tersebut mencapai 26 karyawan pada tahun 2021 dan meningkat menjadi 35 karyawan pada tahun 2022. Gaji dan peluang pengembangan karir menjadi faktor utama yang mendorong karyawan meninggalkan perusahaan. Pada tahun 2021, 18% karyawan meninggalkan perusahaan karena mendapatkan tawaran gaji yang lebih baik di tempat lain, presentase ini meningkat menjadi 48% pada tahun 2022. Selain itu, sebanyak 19% karyawan pada tahun 2021 dan 18% pada tahun 2022 menyatakan bahwa kurangnya peluang pengembangan karir menjadi alasan utama karyawan meninggalkan perusahaan. Tren keluarnya karyawan yang terjadi selama periode 2021-2022 mencerminkan potensi peningkatan atrisi *voluntary* karyawan di masa yang akan datang. Tren ini juga menegaskan pentingnya bagi perusahaan untuk memusatkan perhatian pada strategi retensi dan memahami faktor utama yang memengaruhi keputusan karyawan untuk meninggalkan perusahaan.

Melihat adanya atrisi *voluntary* karyawan yang terjadi di PT.XYZ, perusahaan perlu mengambil langkah proaktif dalam mengelola sumber daya manusia untuk mengatasi potensi atrisi *voluntary* karyawan dan meningkatkan retensi karyawan di masa yang akan datang. Metode *machine learning* telah terbukti efektif dalam memprediksi atrisi di masa yang akan datang. Penelitian sebelumnya oleh [9] menunjukkan akurasi sebesar 92.9% dengan mengeksplorasi metode prediksi *decision tree* dengan algoritma C4.5 untuk memprediksi atrisi karyawan. Temuan lain oleh [10] menunjukkan superioritas model SVM dalam evaluasi dengan akurasi sebesar 86% dalam memprediksi atrisi karyawan. Penelitian lain yang dilakukan oleh [11] melibatkan beberapa metode klasifikasi ZeroR, *decision tree*, KNN, SVM, *random forest*, dan *voting classifier* menunjukkan bahwa metode *voting classifier* muncul sebagai model terbaik dalam memprediksi atrisi dengan nilai

ROC/AUC sebesar 0.89, serta mencapai tingkat akurasi sebesar 88%.

Studi sebelumnya memberikan gambaran respons proaktif terhadap masalah atrisi karyawan di suatu perusahaan. Sebagai kelanjutan dari penelitian sebelumnya, penelitian ini akan menyusun strategi retensi untuk mengatasi atrisi karyawan di perusahaan PT.XYZ di masa yang akan datang dengan pendekatan *machine learning* yang berbeda. Metode *machine learning* dengan pendekatan *ensemble learning* dipilih pada penelitian ini menggunakan model *soft voting classifier*. Metode ini diharapkan dapat membantu perusahaan PT.XYZ dalam memahami faktor penyebab karyawan meninggalkan perusahaan dan strategi dalam menangani dan mencegah atrisi dengan fokus pada strategi retensi yang tepat.

## 2. DASAR TEORI

### 2.1. Atrisi *Voluntary* Karyawan

Atrisi *voluntary* karyawan merupakan keadaan keluarnya karyawan dari perusahaan secara sukarela atas keputusan pribadi mereka. Peristiwa atrisi yang terjadi di perusahaan dapat menimbulkan dampak negatif terhadap operasional bisnis perusahaan, termasuk mempersulit penyelesaian proyek, merusak reputasinya, menaikkan biaya dan upaya dalam melakukan perekrutan.

Menurut [12], dalam merencanakan strategi sumber daya manusia dan mengurangi karyawan penting yang keluar secara sukarela, perlu adanya pemahaman terkait faktor – faktor penyebab terjadinya atrisi *voluntary* karyawan, seperti pemahaman terhadap karakteristik berbagai generasi di pasar tenaga kerja, seperti *baby boomers*, generasi X, dan milenial. Penyebab lain seperti, gaji yang rendah, jarak kantor yang jauh, dan faktor usia juga menjadi penyebab terjadinya fenomena atrisi di suatu perusahaan[10].

Tingkat atrisi dapat dikatakan baik apabila operasional perusahaan dapat berjalan lancar tanpa mengorbankan kinerja karyawan. Oleh karena itu, mencapai keseimbangan yang tepat menjadi kunci untuk memastikan kelangsungan roda bisnis tanpa hambatan yang tidak diinginkan[13].

### 2.2. Retensi Karyawan

Retensi karyawan adalah upaya organisasi berupa strategi untuk mempertahankan karyawan yang dianggap berpotensi agar tetap loyal terhadap perusahaan. Tingkat retensi karyawan yang tinggi dapat membantu perusahaan mengurangi biaya yang diperlukan untuk merekrut, menguji, dan melatih karyawan baru[13]. Beberapa faktor yang memengaruhi retensi karyawan meliputi, kompensasi yang kompetitif, keseimbangan antara kehidupan dan pekerjaan, kualitas karyawan, dan peluang pengembangan karir. Setiap perusahaan perlu memahami dengan baik faktor – faktor tersebut dan

merancang strategi retensi yang efektif untuk menghindari tingkat pergantian karyawan yang tinggi[14].

### 2.3. Analisis Eksplorasi Data (EDA)

Analisis data eksploratif (EDA) adalah proses penting dalam penelitian data untuk mengidentifikasi pola, menemukan anomali seperti pencilan, dan mengidentifikasi karakteristik data menggunakan statistik dan representasi visual dari data[15]. langkah – langkah yang dilakukan pada EDA mencakup data pra-proses, tabulasi, visualisasi, estimasi parameter, penentuan signifikansi parameter, dan pengembangan fitur[16][17]. Analisis data eksploratif memberikan pemahaman yang penting dalam pengambilan keputusan yang informatif dan memberikan wawasan untuk langkah selanjutnya dalam pemrosesan dan pemodelan data.

### 2.4. Metode Ensemble

Metode *ensemble* adalah pendekatan model *machine learning* yang digunakan untuk menggabungkan beberapa model klasifikasi untuk membentuk satu model yang memungkinkan kinerja prediksi yang lebih tinggi[18]. Metode *ensemble* memanfaatkan informasi dari masing – masing model yang digabungkan untuk meningkatkan akurasi prediksi dengan menggunakan suatu fungsi dari kemungkinan nilai yang diprediksi. Terdapat berbagai teknik penggabungan model yang dapat dilakukan dengan ciri khasnya masing – masing dalam meningkatkan akurasi prediksi, seperti *bagging* digunakan untuk mengurangi varian, *boosting* mengurangi bias, serta *stacking* dan *voting* meningkatkan kinerja[19][20]. Metode *ensemble* dengan penggunaan teknik yang tepat dan sesuai dengan karakteristik dataset dalam penelitian dapat membuka ruang untuk adaptasi yang lebih baik terhadap berbagai konteks pemodelan. Penggunaan metode *ensemble* menstimulasi keragaman dalam pengambilan keputusan dan menjamin bahwa model yang dihasilkan dapat memberikan prediksi yang lebih baik dan dapat diandalkan. Oleh karena itu, dalam penelitian ini diterapkan metode klasifikasi berbasis *soft voting* sebagai salah satu metode *ensemble*.

### 2.5. Soft Voting Classifier

*Voting classifier* adalah model klasifikasi yang menggunakan hasil prediksi dari berbagai model untuk menentukan keputusan akhir berdasarkan mayoritas suara. Terdapat dua jenis voting yang umum digunakan, yaitu *hard voting* dan *soft voting*. Pada *hard voting*, keputusan akhir ditentukan oleh mayoritas suara dari model – model dasar. Sementara pada *soft voting*, setiap model memberikan prediksi dalam bentuk probabilitas dan keputusan akhir diambil dengan mempertimbangkan bobot proporsional terhadap probabilitas kelas dari masing – masing

model[11]. Perumusan matematis model *soft voting* ditunjukkan pada rumus (1) berikut:

$$S = \frac{1}{N} \sum_{i=1}^N P_i \quad (1)$$

Dengan  $N$  adalah jumlah model dasar,  $P_i$  adalah prediksi probabilitas dari model ke- $i$ , dan  $S$  adalah prediksi akhir dari *soft voting classifier*.

### 2.6. Support Vector Machine

*Support Vector Machine* (SVM) merupakan algoritma *machine learning* yang efektif dalam menyelesaikan masalah klasifikasi dan regresi. Prinsip kerja SVM yaitu membagi dataset menjadi wilayah terpisah untuk mencapai margin maksimum yang memisahkan dua kelompok data, sehingga memungkinkan klasifikasi data baru dengan akurasi tinggi[21]. Keunggulan metode ini terletak pada akurasi tinggi, efisiensi komputasi, dan kemampuannya dalam menangani dataset berdimensi tinggi serta mengelola sumber daya memori dengan optimal[22]. Proses pembagian dataset menjadi dua data yang berbeda dapat dihitung menggunakan persamaan (2) berikut.

$$(w^T \cdot x_i) + b = 0 \quad (2)$$

Dengan  $w$  adalah bobot dari vektor yang tegak lurus terhadap garis pemisah (*hyperlane*),  $x_i$  adalah *support vectors* dari data ke –  $i$ .

### 2.7. Decision Tree

Algoritma *decision tree* merupakan metode *machine learning* yang efektif dalam menangani masalah klasifikasi dan regresi. *Decision tree* menggunakan struktur hirarkis yang terdiri dari akar, simpul internal, dan simpul daun. Pemilihan atribut terbaik dilakukan pada setiap simpul untuk memisahkan data dengan optimal[23]. Algoritma *decision tree* melalui tahapan pemilihan atribut, pembentukan pohon keputusan, pemisahan data, *pruning* (pemangkasan), dan prediksi untuk data baru[24]. Proses penyusunan pohon keputusan dengan menggunakan kriteria *gini impurity* ditulis dalam persamaan 3 berikut.

$$Gini(D) = 1 - \sum_{i=1}^c p_i^2 \quad (3)$$

Dengan  $D$  adalah dataset pada simpul awal yang dipertimbangkan,  $c$  adalah jumlah kelas pada  $D$ , dan  $p_i$  adalah proporsi sampel data.

### 2.8. Regresi Logistik

Regresi logistik menggunakan regresi linear untuk menyajikan variabel biner berdasarkan probabilitas, dengan tujuan utama mengidentifikasi hubungan antara atribut dan probabilitas hasil tertentu[20]. Regresi logistik merupakan bentuk analisis prediktif yang menghitung probabilitas hasil

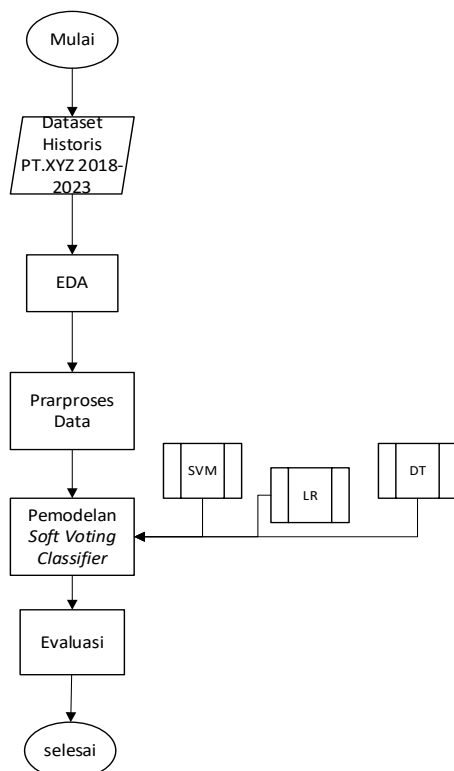
prediksi pada variabel target kategorikal dan regresi linear[25]. Model regresi logistik dapat dirumuskan dengan persamaan 4 berikut.

$$\pi(x_i) = \frac{1}{1+e^{-\beta_0+\beta_1x_1+\dots+\beta_px_p}} \quad (4)$$

Dengan  $\pi(x_i)$  adalah nilai peluang kejadian pada saat  $Y = 1$ .

### 3. METODE PENELITIAN

Pengerjaan penelitian dilaksanakan berdasarkan implementasi yang digambarkan pada gambar 1. secara keseluruhan, penelitian ini terdiri dari 6 tahapan berikut:



Gambar 1. Flowchart Penelitian

#### 3.1. Pengumpulan Data

Data yang digunakan pada penelitian menggunakan data primer yang berasal dari PT.XYZ. Data yang digunakan berisi informasi tentang riwayat pekerjaan karyawan PT.XYZ periode 2018 – 2023. Data tersebut digunakan untuk analisis terkait faktor penyebab utama atrisi karyawan dan prediksi atrisi karyawan di masa mendatang yang spesifik pada konteks perusahaan PT.XYZ.

#### 3.2. Analisis dan Eksplorasi Data (EDA)

Tahap analisis dan eksplorasi data menjadi langkah awal penelitian ini. Dilakukan analisis dan eksplorasi terhadap dataset untuk memahami distribusi dan hubungan antar variabel. Analisis mencakup pemeriksaan statistika deskriptif dan

visualisasi data untuk menemukan pola yang terdapat dalam dataset. Sementara eksplorasi data bertujuan untuk mengidentifikasi fitur penting dan pola relasi antar-variabel.

#### 3.3. Praproses Data

Praproses data adalah serangkaian langkah yang dilakukan pada penelitian ini setelah analisis data eksploratif. Praproses data dilakukan dengan tujuan untuk menyiapkan dataset yang akan digunakan dalam algoritma *machine learning* dengan efisiensi maksimum. Tahapan praproses data mencakup berbagai langkah seperti, *encoding* data, transformasi fitur, normalisasi data, dan *splitting* data. Hasil akhir dari tahapan ini adalah dataset yang telah diolah dan siap untuk dimodelkan dengan algoritma *machine learning*.

#### 3.4. Pemodelan

Pemodelan dilakukan dengan menerapkan metode *ensemble learning* menggunakan model *soft voting classifier* yang menggabungkan SVM, *decision tree*, dan regresi logistik sebagai model dasar. Keputusan menggabungkan model – model ini didasarkan pada keunggulan masing – masing model yang saling melengkapi dan temuan pada penelitian sebelumnya yang menunjukkan bahwa integrasi model tersebut dapat meningkatkan kinerja prediktif, terutama dalam konteks prediksi atrisi karyawan. Pemodelan dengan metode *ensemble* diharapkan mampu menangkap kompleksitas data, mencapai generalisasi yang lebih baik, dan memberikan prediksi atrisi yang lebih akurat.

#### 3.5. Evaluasi model

Evaluasi dilakukan dengan menganalisis kinerja model *soft voting classifier* yang telah digunakan. Evaluasi ini memanfaatkan metrik kinerja *classification report* dan *confusion matrix* untuk menilai seberapa baik model dalam memberikan prediksi yang akurat terkait atrisi karyawan. Hasil evaluasi memberikan landasan untuk menginterpretasi hasil dan merekomendasikan strategi retensi yang tepat guna.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Data Penelitian

Data yang digunakan dalam penelitian berasal dari data internal perusahaan PT.XYZ, yang mencakup informasi tentang karyawan dari tahun 2018 hingga 2023 dengan total *record* data sebanyak 402. Terdapat 16 variabel yang akan diobservasi selama penelitian dengan 10 variabel merupakan variabel numerik yang mencakup usia, jumlah tanggungan, jarak rumah(km), gaji, tunjangan, jam kerja standar, tahun karyawan di posisi kerja, dan lama bekerja di perusahaan dan 6 variabel kategorik yang mencakup jenis kelamin, pendidikan terakhir,

bidang pendidikan, status pernikahan, posisi kerja, dan status karyawan.

## 4.2. Analisis dan Eksplorasi Data

### 4.2.1. Mengatasi Nilai Null

Penanganan nilai *null* dilakukan dengan memeriksa setiap variabel dalam dataset dan mengelompokkannya menjadi variabel numerik dan

kategorik. Pendekatan ini membantu penyesuaian penanganan nilai *null* berdasarkan jenis variabelnya serta mempermudah dalam mengidentifikasi nilai *null*. Tindakan yang dilakukan untuk mengatasi nilai *null* pada masing-masing variabel disajikan dalam tabel 1. berikut.

Tabel 1. Penanganan Nilai Null

Variabel	Jumlah null	Penanganan Nilai Null	Tipe Data
Usia	7	Menghapus nilai <i>null</i>	Numerik
Jumlah Tanggungan	14	Imputasi dengan nilai median	Numerik
Tunjangan	173	Imputasi dengan nilai rata-rata	Numerik
Jarak Rumah	34	Imputasi dengan nilai median	Numerik
Pendidikan Terakhir	24	Mengisi nilai <i>null</i> dengan "Tidak Diketahui"	Kategorik
Bidang Pendidikan	25	Mengisi nilai <i>null</i> dengan modus	Kategorik
Status Pernikahan	7	Mengisi nilai <i>null</i> dengan modus	Kategorik

### 4.2.2. Seleksi Fitur

Seleksi fitur dilakukan dengan pendekatan analisis korelasi *pearson* untuk variabel numerik dan uji *chi-square* untuk variabel kategorik. Pendekatan

seleksi fitur yang dilakukan bertujuan untuk mengetahui hubungan antara masing – masing variabel dengan variabel target. Hasil seleksi fitur untuk masing – masing variabel disajikan dalam tabel 2. berikut.

Tabel 2. Hasil Seleksi Fitur

Variabel	Nilai Korelasi Pearson / P-Value	Keterangan	Tipe Data
Jarak Rumah (Km)	0.2699	Korelasi Positif	Numerik
Gaji	-0.1562	Korelasi Negatif	Numerik
Tunjangan	-0.1083	Korelasi Negatif	Numerik
Tahun di Posisi Kerja	-0.5783	Korelasi Negatif	Numerik
Lama Bekerja di Perusahaan	-0.1188	Korelasi Negatif	Numerik
Pendidikan Terakhir	0.0000255	Signifikan	Kategorik
Status Pernikahan	0.0194	Signifikan	Kategorik
Posisi Kerja	0.00003	Signifikan	Kategorik

## 4.3. Praproses Data

Beberapa langkah praproses data telah dilakukan dalam mempersiapkan data untuk tahapan pemodelan. Berikut adalah hasil dan pembahasan dari setiap langkah:

### 4.3.1. One-Hot Encoding

Dilakukan *one-hot encoding* untuk mengolah variabel kategorik dalam pembelajaran mesin. Teknik ini bekerja dengan mengubah variabel kategorik menjadi representasi numerik yang dapat dipahami model.

### 4.3.2. Scaling Data

Dilakukan *scaling data* menggunakan *robustscaler* dari library *sklearn*. *Robustscaler* bekerja dengan mentransformasi data numerik ke dalam rentang tertentu, sehingga dapat meningkatkan akurasi model prediksi. *Robustscaler* dipilih karena tahan terhadap penciliran, tidak mengubah distribusi data, mudah diimplementasikan, dan dapat meningkatkan akurasi dan stabilitas model.

### 4.3.3. Splitting Data

Data dibagi menjadi dua bagian yaitu 20% untuk data latih dan 80% untuk data uji. Pembagian

data ini bertujuan untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya.

## 4.4. Pemodelan Soft Voting Classifier

Pemodelan *soft voting classifier* dilakukan dengan menggabungkan model dasar SVM, *Decision tree*, dan Regresi Logistik tanpa penyesuaian tambahan terhadap parameter – parameter khususnya. Pendekatan seperti ini memungkinkan untuk model *soft voting classifier* dalam mengambil keputusan berdasarkan hasil voting dari model – model tersebut. Algoritma yang dilakukan dalam menggabungkan model dasar ke dalam *soft voting classifier* diuraikan sebagai berikut.

1. Inisialisasi model SVM, *decision tree*, dan Regresi Logistik menggunakan modul *sklearn* dalam pemrograman *python*.
2. Membuat objek model SVM, *decision tree*, dan Regresi Logistik menggunakan konstruktor masing – masing kelas
3. Pelatihan model SVM, *decision tree*, dan Regresi Logistik menggunakan data latih (*X\_train* dan *y\_train*)
4. Inisialisasi model *soft voting classifier* dengan menggunakan kelas '*VotingClassifier*' dari modul *sklearn.ensemble*.

5. Menyediakan daftar tupel yang berisi nama model dan objek model dasar yang dilatih.
6. Mengatur parameter 'voting=soft' untuk menggunakan soft voting.
7. Latih model soft voting classifier menggunakan data latih (X\_train dan y\_train)

Pemodelan soft voting classifier juga dilakukan dengan menggabungkan model dasar SVM serta model decision tree dan regresi logistik yang menggunakan penyetelan hyperparameter tuning melalui metode GridSearch. Penyetelan hyperparameter tuning pada model Decision tree dan Regresi Logistik dilakukan dengan mencari parameter terbaik untuk masing – masing model. Hasil penelusuran parameter terbaik untuk model decision tree dan regresi logistik ditampilkan dalam tabel 3. berikut.

Tabel 3. Hasil Penyetelan Hyperparameter Tuning

Model	Hyperparameter
SVM	SVC(Probability=True)
Decision Tree	'max_depth': 5, 'min_samples_leaf': 1, 'min_samples_split': 8
Regresi Logistik	'C': 10, 'solver': 'liblinear'

Algoritma pemodelan yang menggabungkan SVM dan decision tree serta regresi logistik yang dilakukan penyetelan hyperparameter tuning diuraikan sebagai berikut.

1. Inisialisasi model SVM, decision tree, dan Regresi Logistik menggunakan modul sklearn dalam pemrograman python.

Tabel 4. Hasil Evaluasi Classification Report

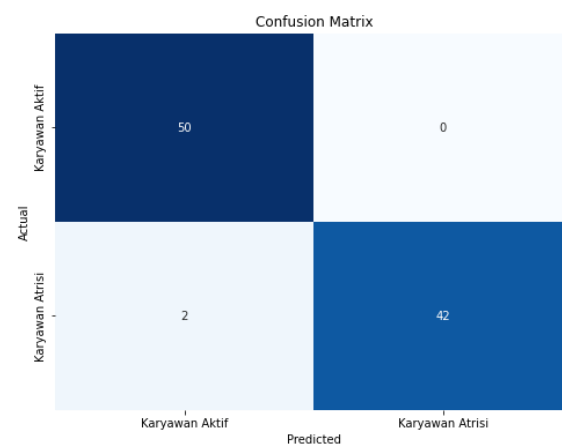
Model	Precision	Recall	F1-Score	Akurasi
SVM(Training)	0.98	0.98	0.98	0.98
SVM(Testing)	0.97	0.97	0.97	0.97
Decision Tree(Training)	0.98	0.98	0.98	0.98
Decision Tree(Testing)	0.95	0.95	0.95	0.95
Regresi Logistik(Training)	0.99	0.99	0.99	0.99
Regresi Logistik(Testing)	0.96	0.96	0.96	0.96
Soft Voting Classifier(Training)	0.99	0.99	0.99	0.99
Soft Voting Classifier(Testing)	0.98	0.98	0.98	0.98

Berdasarkan tabel 3. diketahui bahwa semua model yang dievaluasi menunjukkan kinerja yang baik dalam memprediksi atrisi voluntary karyawan. Nilai precision yang tinggi menunjukkan bahwa model mampu mengidentifikasi dengan tepat karyawan yang akan meninggalkan perusahaan, sementara nilai recall yang tinggi menunjukkan bahwa model dapat menangkap sebagian besar kasus atrisi voluntary karyawan. Nilai F1-Score yang tinggi menunjukkan keseimbangan antara precision dan recall yang mengindikasikan prediksi yang baik secara keseluruhan. Model soft voting classifier menunjukan performa yang paling baik dengan nilai precision, recall, F1-Score, akurasi yang sama sebesar 0.99 untuk data training dan 0.98 untuk data testing.

2. Membuat objek model SVM, decision tree, dan Regresi Logistik menggunakan konstruktor masing – masing kelas
3. Menentukan grid himpunan parameter untuk decision tree dan regresi logistik
4. Menjalankan metode GridSearch untuk mencari parameter terbaik pada model decision tree dan regresi logistik menggunakan data latih (X\_train dan y\_train)
5. Mendapatkan model decision tree dan regresi logistik terbaik dari hasil GridSearch
6. Membuat soft voting classifier dengan model terbaik dari hasil hyperparameter tuning model decision tree dan regresi logistik, serta model SVM tanpa penyetelan parameter tambahan.
7. Latih model soft voting classifier menggunakan data latih (X\_train dan y\_train).

#### 4.5. Evaluasi Model Soft Voting Classifier

Evaluasi dilakukan terhadap model soft voting classifier beserta model dasar SVM tanpa penyetelan hyperparameter tuning, decision tree, dan regresi logistik yang sudah dilakukan penyetelan hyperparameter tuning menggunakan classification report dan confusion matrix. Hal ini dilakukan karena model soft voting classifier dengan model dasar decision tree dan regresi logistik yang disesuaikan parameternya memiliki performa yang lebih dibandingkan dengan model soft voting classifier yang model dasarnya tidak melalui penyesuaian hyperparameter tuning. Hasil evaluasi masing – masing metrik dijabarkan pada tabel 4. berikut.



Gambar 2. Hasil Evaluasi Confusion Matrix

Hasil evaluasi *confusion matrix* pada gambar 1. mempertegas bahwa model *soft voting classifier* memiliki performa yang baik dalam memprediksi atrisi *voluntary* karyawan. Berdasarkan gambar 1. tersebut diketahui bahwa model berhasil mengklasifikasikan 42 sampel sebagai *true positive*, yang berarti model berhasil mengidentifikasi karyawan yang berpotensi tinggi meninggalkan perusahaan. Model juga berhasil mengklasifikasikan 50 sampel sebagai *true negative* yang menunjukkan kemampuan model dalam mengidentifikasi karyawan yang tidak akan meninggalkan perusahaan. Model *soft voting classifier* juga tidak melakukan kesalahan dalam mengklasifikasikan sampel *false positive*. Hal ini menandakan bahwa model *soft voting classifier* mampu memprediksi kasus atrisi dengan baik. Namun, model belum sepenuhnya dapat menangkap semua kasus atrisi yang sebenarnya karena terdapat 2 sampel yang diklasifikasikan sebagai *false negatif*. Meskipun demikian, secara keseluruhan model *soft voting classifier* yang digabungkan dengan model dasar SVM, *decision tree*, dan regresi logistik memungkinkan prediksi yang akurat dan *robust* untuk kasus atrisi *voluntary* karyawan dan memberikan landasan yang kuat bagi perusahaan untuk mengambil tindakan preventif yang tepat untuk memertahankan karyawan yang berpotensi meninggalkan perusahaan.

## 5. DISKUSI

Studi sebelumnya telah melihat upaya untuk mencegah dan mengelola atrisi karyawan dengan fokus pada berbagai algoritma *machine learning* yang berbeda. Seperti halnya penelitian sebelumnya yang berhasil merancang sistem prediksi atrisi karyawan menggunakan metode *decision tree* dengan algoritma C4.5 yang menunjukkan hasil akurasi sangat tinggi mencapai 92.9% [9]. Selanjutnya penelitian yang dilakukan dengan mengevaluasi performa model *machine learning*, terutama *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN) menunjukkan hasil bahwa SVM memiliki evaluasi yang lebih baik dibandingkan KNN, dengan akurasi mencapai 86% dan *geometric-mean* sebesar 75%. Hasil tersebut menunjukkan bahwa SVM lebih efektif dalam memprediksi data ke dalam kelas atrisi dan kelas tak-atrisi dibandingkan model KNN [10].

Penelitian yang dilakukan oleh [11] juga memberikan kontribusi dalam memperluas pemahaman tentang prediksi atrisi karyawan melalui pendekatan *machine learning*. Peneliti melakukan implementasi beberapa model klasifikasi yaitu, *ZeroR*, *Decision tree*, *K-Nearest Neighbors*, *Support Vector Machine*, *Random Forest*, dan *Voting Classifier* dalam penelitian yang dilakukan. Hasil penelitian menunjukkan bahwa *voting classifier* dengan pengoptimalan *hyperparameter tuning* menjadi model terbaik dengan nilai AUC/ROC mencapai 0.89. Diikuti dengan hasil evaluasi metrik *presisi*, *recall*, *F1-Score*, dan akurasi masing-masing

mencapai 0.9, 0.88, 0.85, dan 0.88. Temuan ini menegaskan bahwa penggunaan *voting classifier* dalam memprediksi atrisi karyawan menunjukkan kinerja yang cukup baik dan berpotensi untuk diterapkan dalam konteks perusahaan.

Penelitian ini melanjutkan kerangka kerja sebelumnya dengan fokus pada pemahaman yang lebih mendalam terkait masalah atrisi karyawan di PT.XYZ. penelitian sebelumnya memberikan dasar pemahaman untuk penelitian ini terutama dalam penerapan algoritma *machine learning*. Berbeda dengan penelitian sebelumnya, penelitian ini menggunakan metode klasifikasi yang menggabungkan model *Support Vector Machine*, *Decision tree*, dan *Logistic Regression* dalam penerapan *soft voting classifier*. Dengan demikian, penelitian ini tidak hanya melanjutkan kerangka kerja penelitian sebelumnya, tetapi juga menawarkan pendekatan *machine learning* yang berbeda yang berpotensi untuk meningkatkan pemahaman tentang masalah atrisi karyawan dan memberikan kontribusi yang berarti dalam penanganannya.

## 6. KESIMPULAN

Penelitian ini menyelidiki fenomena atrisi *voluntary* karyawan di PT.XYZ dengan mengusulkan pendekatan *machine learning* untuk memprediksi atrisi *voluntary* karyawan di masa mendatang. Penelitian dilakukan dengan menggunakan data historis karyawan PT.XYZ periode 2018 hingga 2023. Peneliti berhasil mengidentifikasi faktor-faktor yang berkontribusi terhadap keputusan karyawan meninggalkan perusahaan yaitu jarak rumah, gaji, tunjangan, tahun di posisi kerja, lama bekerja di perusahaan, pendidikan terakhir, status pernikahan, dan posisi kerja, melalui data historis karyawan tersebut.

Model *soft voting classifier* yang diusulkan berhasil memprediksi atrisi dengan baik yang menandakan efektivitasnya dalam mengatasi tantangan atrisi *voluntary* karyawan. Hasil evaluasi model menunjukkan bahwa model *soft voting classifier* memberikan performa terbaik dalam memprediksi atrisi karyawan, dengan nilai *precision*, *recall*, *F1-Score*, akurasi yang sama besar yaitu 0.99 untuk data pelatihan dan 0.98 untuk data *testing*. Dengan demikian, penelitian diharapkan dapat memberikan wawasan yang berharga bagi manajemen sumber daya manusia PT.XYZ dalam mengembangkan strategi retensi karyawan yang efektif.

Penelitian ini dapat dijadikan landasan bagi penelitian selanjutnya terkait faktor-faktor penyebab dan metode prediksi atrisi *voluntary* karyawan. Penelitian ini dapat diperluas cakupannya dengan menambahkan variabel tambahan seperti kepuasan kerja, hubungan dengan atasan, dan faktor psikologis lainnya. Selain itu, pengambilan data dengan cara wawancara mendalam terhadap karyawan akan memberikan gambaran yang lebih komprehensif

terkait penyebab karyawan ingin meninggalkan perusahaan. Penerapan metode seleksi fitur yang lebih efisien dapat dijadikan sebagai topik penelitian selanjutnya untuk mendapat tingkat akurasi prediksi yang lebih baik. Penelitian juga bisa dikembangkan dengan membandingkan model *ensemble* lain seperti *stacking*, *bagging*, atau *boosting* untuk mengetahui perbedaan performa model dalam melakukan prediksi.

#### DAFTAR PUSTAKA

- [1] S. Sumarita dan M. Muzakki, “Pengaruh Kepemimpinan Transformasional pada Intensi Karyawan Keluar dari Perusahaan,” *Indones. Bus. Rev.*, vol. 6, no. 1, hal. 61, 2023, doi: 10.21632/ibr.6.1.61-70.
- [2] W. A. Medyanti dan M. Faisal, “Early Prediction System for Employee Attrition Company ‘XYZ’ Using Support Vector Machine Algorithm,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 8, no. 2, hal. 429, 2023, doi: 10.24114/cess.v8i2.46494.
- [3] N. Mansor, N. S. Sani, dan M. Aliff, “Machine Learning for Predicting Employee Attrition,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, hal. 435–445, 2021, doi: 10.14569/IJACSA.2021.0121149.
- [4] D. Jain dan M. B. Buckley, “Evaluation of Employee Attrition by Effective Feature Selection using Hybrid Model of Ensemble Methods MSc Research Project Data Analytics”.
- [5] F. K. Alsheref, I. E. Fattoh, dan 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.
- [6] P. Wulansari, B. Meilita, dan Y. Ganesan, “The Effect of Employee Retention Company to Turnover Intention Employee—Case Study on Head Office Lampung Bank,” vol. 117, no. Gcbme 2018, hal. 236–239, 2020, doi: 10.2991/aebmr.k.200131.050.
- [7] M. Silalahi, S. Purba, I. K. E. M. Mulyana, M. T. Rambe, dan Abdurohim, *Manajemen Sumber Daya Manusia (Strategi Organisasi Modern)*. 2022.
- [8] A. Qutub, A. Al-Mehmadi, M. Al-Hssan, R. Aljohani, dan H. S. Alghamdi, “Prediction of Employee Attrition Using Machine Learning and Ensemble Methods,” *Int. J. Mach. Learn. Comput.*, vol. 11, no. 2, hal. 110–114, 2021, doi: 10.18178/ijmlc.2021.11.2.1022.
- [9] T. H. L. Tulus, A. Id Hadiana, dan I. Santikarama, “Sistem Prediksi Awal Terhadap Atrisi Karyawan Menggunakan Algoritma C4.5,” *Informatics Digit. Expert*, vol. 4, no. 1, hal. 18–24, 2022, doi: 10.36423/index.v4i1.882.
- [10] M. A. Al Fatih dan K. M. Lhaksana, “Prediksi Employee Attrition menggunakan Algoritma Support Vector Machine (SVM),” *J. Prediksi Empl. Attrition*, vol. 10, no. 2, hal. 1930, 2023.
- [11] S. Bhatta, I. U. Zaman, N. Raisa, S. I. Fahim, dan S. Momen, “Machine Learning Approach to Predicting Attrition Among Employees at Work,” *Lect. Notes Networks Syst.*, vol. 502 LNNS, no. July, hal. 285–294, 2022, doi: 10.1007/978-3-031-09076-9\_27.
- [12] R.-D. Căplescu, M. Ilie, dan V. A. Strat, “Voluntary employee attrition. Descriptive and predictive analysis,” *Proc. Int. Conf. Appl. Stat.*, vol. 1, no. 1, hal. 145–161, 2019, doi: 10.2478/icas-2019-0013.
- [13] M. Maulana, “Pengaruh Kompensasi, Pengembangan Karir, dan Kepuasan Kerja terhadap retensi Karyawan (Studi pada PT Suk Sempurna Furnindo),” no. 2005, hal. 1–90, 2020.
- [14] D. P. S. Swami, “EMPLOYEE RETENTION STRATEGIES IN IT/ITeS SECTOR: EMPLOYER VIEW,” *Interantional J. Sci. Res. Eng. Manag.*, vol. 06, no. 10, hal. 1–10, 2022, doi: 10.55041/ijrsrem16680.
- [15] V. Da Poian *dkk.*, “Exploratory data analysis (EDA) machine learning approaches for ocean world analog mass spectrometry,” *Front. Astron. Sp. Sci.*, vol. 10, no. May, hal. 1–17, 2023, doi: 10.3389/fspas.2023.1134141.
- [16] P. A. Riyantoko, K. M. Hindrayani, T. M. Fahrudin, dan E. M. Safitri, “Southeast Asia Happiness Report in 2020 Using Exploratory Data Analysis,” *Netw. Secur. Inf. Syst.*, vol. 2, no. 1, hal. 16–21, 2020.
- [17] K. Maulida Hindrayani, T. Maulana F, P. Aji R, dan Kartini, “Determining Students Preparation for College Entrance Examinations in Indonesia From Twitter Data Using Exploratory Data Analysis,” *Ijconsist Journals*, vol. 2, no. 02, hal. 66–70, 2021, doi: 10.33005/ijconsist.v2i02.47.
- [18] M. Khairy, T. M. Mahmoud, A. Omar, dan T. Abd El-Hafeez, “Comparative performance of ensemble machine learning for Arabic cyberbullying and offensive language detection,” *Lang. Resour. Eval.*, no. 0123456789, 2023, doi: 10.1007/s10579-023-09683-y.
- [19] A. T. Damaliana, T. Trimono, dan D. A. Prasetya, “Ensemble Tree untuk Memprediksi Level Resiko Maternal Mortality di Bangladesh,” ... *Sains Data*, vol. 2022, no. Senada, hal. 24–30, 2022, [Daring].



- Tersedia pada: <https://prosiding-senada.upnjatim.ac.id/index.php/senada/article/view/36%0Ahttps://prosiding-senada.upnjatim.ac.id/index.php/senada/article/download/36/20>
- [20] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, dan J. Haider, "An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI," *Sensors*, vol. 22, no. 19, 2022, doi: 10.3390/s22197268.
- [21] S. Sumayah, F. Sembiring, dan W. Jatmiko, "Analysis of Sentiment of Indonesian Community on Metaverse Using Support Vector Machine Algorithm," *J. Tek. Inform.*, vol. 4, no. 1, hal. 143–150, 2023, doi: 10.52436/1.jutif.2023.4.1.417.
- [22] Y. B. Lasotte, E. J. Garba, Y. M. Malgwi, dan M. A. Buhari, "An Ensemble Machine Learning Approach for Fake News Detection and Classification Using a Soft Voting Classifier," *Eur. J. Electr. Eng. Comput. Sci.*, vol. 6, no. 2, hal. 1–7, 2022, doi: 10.24018/ejece.2022.6.2.409.
- [23] NAGİHAN TAŞKIRAN, "A RECOMMENDATION APPROACH FOR EMPLOYEE RETENTION BY USING A NEW FEATURE SELECTION STRATEGY," no. September, hal. 31–41, 2023.
- [24] A. Rana dan R. Pandey, "A review of popular decision tree algorithms in data mining," *Asian J. Multidimens. Res.*, vol. 10, no. 10, hal. 230–237, 2021, doi: 10.5958/2278-4853.2021.00837.5.
- [25] F. H. Wardhani dan K. M. Lhaksmana, "Predicting Employee Attrition Using Logistic Regression With Feature Selection," *Sinkron*, vol. 7, no. 4, hal. 2214–2222, 2022, doi: 10.33395/sinkron.v7i4.11783.