

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

**SKRIPSI**

**Diajukan untuk memenuhi salah satu syarat kelulusan  
di Program Studi Sains Data**



**Disusun Oleh:  
CAGIVA CHAEDAR BEY LIRNA  
20083010020**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI SAINS DATA  
SURABAYA  
2024**

**LEMBAR PENGESAHAN**

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

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data  
pada : Selasa, 23 Juli 2024

**Program Studi S-1 Sains Data  
Fakultas Ilmu Komputer**

**Universitas Pembangunan Nasional Veteran Jawa Timur  
Surabaya**

Oleh

**CAGIVA CHAEDAR BEY LIRNA**

NPM. 20083010020

Disetujui oleh Tim Penguji Skripsi :

Penguji 1

Penguji 2

Kartika Maulida Hindrayani, S.Kom. M.Kom  
NIP. 199209092022032009

Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU  
NIP. 19700619202121109

Pembimbing 1

Pembimbing 2

Triyono, S.Si., M.Si  
NIP. 199509082022031003

Aviolla Terza Damaliana, S.Si., M.Stat  
NIP. 199408022022032015

Fakultas Ilmu Komputer  
Dekan,

Program Studi Sains Data  
Fakultas Ilmu Komputer  
Koordinator,

Prof. Dr. Ir. Novirina Hendrasarie, MT  
NIP. 196811261994032001

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.  
NIP. 198012052005011002

Surabaya, Juli, 2024

## SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Cagiva Chaedar Bey Lirna  
NPM : 20083010020  
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

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

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 05 Juli ..... 2024

Hormat Saya



Cagiva Chaedar Bey Lirna  
NPM. 20083010020

## ABSTRAK

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

**Nama Mahasiswa / NPM** : Cagiva Chaedar Bey Lirna/ 20083010020  
**Program Studi** : Sains Data, FASILKOM, UPN Veteran Jatim  
**Dosen Pembimbing 1** : Trimono, S.Si., M.Si  
**Dosen Pembimbing 2** : Aviolla Terza Damaliana, S.Si., M.Stat

#### **Abstrak**

Penelitian ini merespon kompleksitas tantangan atrisi karyawan di PT.XYZ. Tujuan utamanya adalah mengembangkan sistem prediksi potensi atrisi *voluntary* karyawan dengan memfokuskan pada analisis mendalam faktor-faktor atrisi *voluntary*. Data yang digunakan pada penelitian menggunakan data yang berisi informasi tentang riwayat pekerjaan karyawan PT.XYZ periode 2018 – 2023. Metode yang diterapkan adalah model *ensemble soft voting classifier*, mencakup SVM, *Decision Tree*, dan regresi logistik biner. Faktor penyebab atrisi karyawan diidentifikasi melalui analisis korelasi *pearson* dan uji *chi-square*, mencakup jarak rumah karyawan, gaji karyawan, tunjangan karyawan, tahun karyawan menempati posisi kerja, lama karyawan bekerja di perusahaan, pendidikan terakhir karyawan, status pernikahan karyawan, dan posisi kerja karyawan. Model *soft voting classifier* mampu memprediksi karyawan “Aktif” dan “Atrisi” sebanyak 60 data dari total 61 data uji dengan akurasi mencapai 98%. Berdasarkan temuan – temuan tersebut, diformulasikan rekomendasi strategi retensi karyawan berupa berbagai pertimbangan yang perlu diperhatikan selama proses perekrutan karyawan, program keseimbangan kehidupan kerja, peningkatan gaji dan tunjangan, rencana pengembangan karier, dukungan pendidikan, program penghargaan terhadap karyawan, dan pengembangan budaya perusahaan yang inklusif.

**Kata kunci:** *Atrisi Voluntary, EDA, Model ensemble, Retensi Karyawan, Soft voting.*

## **ABSTRACT**

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

***Student Name / NPM*** : Cagiva Chaedar Bey Lirna/ 20083010020  
***Study Program*** : Sains Data, FASILKOM, UPN Veteran Jatim  
***Advisor 1*** : Trimono, S.Si., M.Si  
***Advisor 2*** : Aviolla Terza Damaliana, S.Si., M.Stat

#### **Abstract**

*This research responds to the complexity of employee attrition challenges in PT.XYZ. The main goal is to develop a system for predicting the potential for voluntary attrition of employees by focusing on an in-depth analysis of voluntary attrition factors. The data used in the study used data that contained information about the work history of PT.XYZ employees for the period 2018 – 2023. The method applied is an ensemble soft voting classifier model, including SVM, Decision Tree, and binary logistic regression. The factors causing employee attrition were identified through Pearson correlation analysis and chi-square test, including the employee's home distance, employee salary, employee benefits, the year the employee occupied the work position, the length of time the employee worked in the company, the employee's last education, the employee's marital status, and the employee's work position. The soft voting classifier model can predict "Active" and "Attrition" employees as many as 60 data out of a total of 61 test data with an accuracy of 98%. Based on these findings, recommendations for employee retention strategies are formulated in the form of various considerations that need to be considered during the employee recruitment process, work-life balance programs, salary and benefit increases, career development plans, educational support, employee reward programs, and the development of an inclusive company culture.*

***Keywords:*** EDA, Employee Retention, Model ensemble, Soft voting, Voluntary Attrition.

## KATA PENGANTAR

Puji dan syukur kehadirat ALLAH SWT, atas limpahan Rahmat serta Kasih Sayang-Nya sehingga penulis dapat menyelesaikan Skripsi yang merupakan persyaratan dalam menyelesaikan mata kuliah Skripsi pada Program Studi S1 Sains Data di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam penyusunan Skripsi ini tidak terlepas dari bantuan berbagai pihak, dan dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Rama Muchlizar Bakti dan Bunda Ratna Kencana, Terimakasih yang sebesar – besarnya atas segala doa, kasih sayang, dorongan motivasi, dan segala pengorbanan yang telah diberikan kepada penulis tanpa henti. Karena kalian berdua, penulis bisa kuat dan kokoh untuk terus hidup dan berjuang meraih mimpi dan cita – cita. Kesuksesan dan segala hal baik yang kedepannya akan penulis dapatkan adalah karena dan untuk Rama dan Bunda;
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur;
5. Bapak Trimono, S.Si., M.Si selaku dosen pembimbing 1 yang telah meluangkan waktu ditengah kesibukan beliau, memberikan kritik, saran, dedikasi, pengetahuan dan bimbingan yang luar biasa membantu penulis untuk menyelesaikan semua proses penulisan skripsi ini;
6. Ibu Aviolla Terza Damaliana, S.Si., M.Stat, selaku dosen pembimbing 2 yang selalu memberikan dedikasi, masukan berharga, pengetahuan, ilmu dan ketelitian penuh dalam membimbing penulis menyelesaikan penulisan skripsi ini;
7. Seluruh dosen dan staf prodi Sains Data, yang selalu mengusahakan dan menyediakan lingkungan belajar yang mendukung bagi seluruh mahasiswa Sains Data;

8. Almh. Nenek Kusyani, yang tiada hentinya selalu mendoakan, mendukung, memberikan kasih sayang kepada penulis hingga akhir hayatnya. Nasihat yang bijak dari nenek telah memberikan inspirasi, kebijaksanaan, dan kekuatan bagi penulis untuk terus melangkah dan menyelesaikan skripsi ini. Terimakasih atas cinta dan kebaikan yang selalu nenek berikan, segala kenangan indah bersama nenek akan selalu dihati. Semoga nenek diberikan surganya Allah SWT;
9. Adek Dafa Fisabil Bey Lirna, yang selalu memberikan senyuman dan kebahagiaan bagi penulis di saat-saat sulit. Terimakasih telah menjadi adek yang selalu suportif bagi penulis;
10. Keluarga besar yang selalu memberikan perhatian dan dukungan yang selalu mengalir bagi penulis. Kehadiran dan kebersamaan keluarga telah menjadi kekuatan dalam setiap langkah perjalanan penulis;
11. “Dia”, yang selalu ada di setiap langkah perjalanan penulis. Dukungannya yang tak terkatakan telah menjadi penyemangat dalam setiap tantangan yang penulis hadapi. Setiap momen yang kita bagikan selalu menjadi sumber kebahagiaan dan kekuatan. Meskipun namamu tak bisa disebutkan secara langsung di sini, namun peranmu begitu besar bagi penulis.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam Buku Skripsi ini, namun penulis berharap semoga Buku Skripsi ini dapat memberikan kontribusi terhadap perkembangan ilmu pengetahuan, khususnya dalam bidang ilmu sains data.

Surabaya, 3 Mei 2024

Penulis

## DAFTAR ISI

LEMBAR PENGESAHAN .....	ii
SURAT PERNYATAAN.....	iii
ABSTRAK .....	iv
<i>ABSTRACT</i> .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xii
DAFTAR LAMPIRAN .....	xiv
BAB I PENDAHULUAN.....	1
1.1    Latar Belakang .....	1
1.2    Rumusan Masalah .....	5
1.3    Batasan Masalah.....	5
1.4    Tujuan Penelitian .....	6
1.5    Manfaat Penelitian .....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1    Atrisi <i>Voluntary</i> Karyawan .....	8
2.2    Retensi Karyawan .....	9
2.3    Analisis Eksploratif Data (EDA) .....	10
2.4    Seleksi Fitur .....	10
2.4.1    Korelasi <i>Pearson</i> .....	10
2.4.2    Analisis <i>Chi-square</i> .....	12
2.5    Metode <i>Ensemble</i> .....	13
2.5.1 <i>Soft Voting Classifier</i> .....	13
2.5.2 <i>Support Vector Machine</i> .....	15
2.5.3 <i>Decision Tree</i> .....	19
2.5.4    Regresi Logistik Biner .....	21
2.6 <i>Hyperparameter Tuning</i> .....	25
2.7    Evaluasi Performa Model.....	26
2.8 <i>Review</i> Penelitian Sebelumnya .....	28
BAB III METODOLOGI PENELITIAN.....	32



3.1	Sumber Data.....	32
3.2	Diagram Alir Penelitian .....	34
3.3	Jadwal Penelitian.....	39
BAB IV HASIL DAN PEMBAHASAN .....		40
4.1	Prediksi Atrisi <i>Voluntary</i> Karyawan dengan <i>Soft Voting Classifier</i> .....	40
4.1.1	Data Historis Karyawan PT.XYZ .....	40
4.1.2	Analisis dan Eksplorasi Data .....	45
4.1.3	Seleksi Fitur .....	56
4.1.4	Praproses Data.....	59
4.1.5	Pemodelan <i>Soft Voting Classifier</i> .....	63
4.1.6	Penyetelan <i>Hyperparameter Tuning</i> .....	80
4.2	Uji Performa Model <i>Soft Voting Classifier</i> .....	84
4.2.1	<i>Confusion Matrix</i> .....	85
4.2.2	<i>Classification Report</i> .....	87
4.3	Rekomendasi Strategi Retensi Karyawan .....	88
BAB V PENUTUP.....		92
5.1	Kesimpulan .....	92
5.2	Saran.....	93
DAFTAR PUSTAKA .....		94
LAMPIRAN .....		101
BIODATA PENULIS .....		104

## DAFTAR TABEL

Tabel 2.1 Fungsi Kernel SVM .....	18
Tabel 2.2 Peluang Kejadian Regresi Logistik Biner .....	22
Tabel 2.3 Tabel <i>Confusion Matrix</i> .....	27
Tabel 2.4 Studi Literatur yang Relevan .....	28
Tabel 3.1 Tabel Karakteristik Personal Karyawan .....	32
Tabel 3.2 Tabel Kinerja Karyawan .....	33
Tabel 3.3 Tabel Riwayat Pekerjaan .....	33
Tabel 3.4 Jadwal Penelitian.....	39
Tabel 4.1 Data pada Tabel Karakteristik Personal Karyawan .....	40
Tabel 4.2 Data pada Tabel Kinerja Karyawan .....	41
Tabel 4.3 Data pada Tabel Pekerjaan Karyawan .....	41
Tabel 4.4 Deskripsi Data Penelitian.....	42
Tabel 4. 5 Tampilan Data Penelitian Tanpa "ID_Karyawan" (1) .....	46
Tabel 4.6 Tampilan Data Penelitian Tanpa "ID_Karyawan" (2) .....	46
Tabel 4.7 Identifikasi Nilai Null pada Variabel Numerik.....	47
Tabel 4. 8 Identifikasi Nilai Null pada Variabel Kategorik.....	53
Tabel 4.9 Hasil Analisis Korelasi <i>Pearson</i> .....	56
Tabel 4.10 Hasil Uji <i>Chi-square</i> .....	58
Tabel 4.11 Data untuk Tahap Praproses Data.....	59
Tabel 4.12 Hasil <i>Scaling Data</i> .....	61
Tabel 4.13 Hasil <i>Splitting Data</i> .....	62
Tabel 4.14 Hasil Perhitungan Parameter untuk Kernel RBF .....	64
Tabel 4.15 <i>Preview</i> Hasil Prediksi Model SVM.....	66
Tabel 4.16 <i>Preview</i> Hasil Prediksi Model <i>Decision Tree</i> .....	69
Tabel 4.17 Hasil Uji <i>Likelihood Ratio</i> .....	71
Tabel 4.18 Hasil Uji <i>Hosmer-Lemeshow</i> .....	72
Tabel 4.19 Hasil Perhitungan <i>Intercept</i> dan Koefisien untuk Variabel Prediktor	73
Tabel 4.20 Hasil Logit.....	74
Tabel 4.21 Probabilitas Kejadian Regresi Logistik Biner.....	74
Tabel 4.22 <i>Preview</i> Hasil Prediksi Model Regresi Logistik Biner .....	75

Tabel 4.23 <i>Preview Hasil Prediksi Model Soft Voting Classifier</i> .....	78
Tabel 4.24 <i>Percobaan Penyetelan Parameter Decision Tree</i> .....	81
Tabel 4. 25 <i>Preview Hasil Prediksi Model Decision Tree Setelah Hyperparameter Tuning</i> .....	82
Tabel 4. 26 <i>Preview Hasil Prediksi Model Soft Voting Classifier Terbaru</i> .....	85
Tabel 4.27 <i>Hasil Evaluasi Classification Report</i> .....	87

## DAFTAR GAMBAR

Gambar 2.1 Strategi Umum Retensi menurut Dr.P.Sujendra Swami .....	9
Gambar 2.2 Alur Kerja <i>Soft Voting Classifier</i> .....	14
Gambar 2.3 Ilustrasi <i>Support Vector Machine</i> .....	15
Gambar 2.4 Representasi Pohon Keputusan .....	19
Gambar 3.1 <i>Flowchart</i> Penelitian .....	34
Gambar 3.2 <i>Flowchart</i> Pemodelan <i>Soft Voting Classifier</i> .....	37
Gambar 4.1 Identifikasi Nilai <i>Null</i> pada Variabel Numerik .....	48
Gambar 4.2 Visualisasi Hubungan “Tunjangan(Rp)” dengan “Tahun_Posisi” ....	50
Gambar 4.3 Visualisasi Hubungan “Jarak_Rumah(Km)” dengan “Tahun_Posisi” .....	51
Gambar 4.4 Visualisasi Hubungan "jml_tanggung" Dengan "Usia" .....	52
Gambar 4.5 Identifikasi Nilai <i>Null</i> pada Variabel Kategorik .....	54
Gambar 4.6 Implementasi Kode <i>Scaling Data</i> .....	61
Gambar 4.7 Implementasi Kode <i>Splitting Data</i> .....	62
Gambar 4.8 Import Pustaka Pemodelan.....	63
Gambar 4.9 Implementasi Inisialisasi dan Pelatihan Model SVM .....	64
Gambar 4. 10 Implementasi <i>Code</i> Ekstraksi Parameter dual coefficients, x, dan b .....	64
Gambar 4. 11 Code Menghitung Kernel RBF dan Nilai Keputusan Prediksi .....	65
Gambar 4. 12 Kode Prediksi Kelas dan Perhitungan Nilai Probabilitas SVM ....	66
Gambar 4.13 Hasil Akurasi Model SVM.....	67
Gambar 4. 14 Implementasi Inisialisasi dan Pelatihan Model <i>Decision Tree</i> .....	68
Gambar 4. 15 Kode Prediksi Kelas dan Perhitungan Nilai Probabilitas <i>Decision Tree</i> .....	69
Gambar 4.16 Hasil Akurasi Model <i>Decision Tree</i> .....	70
Gambar 4.17 Implementasi Inisialisasi Model Regresi Logistik Biner .....	71
Gambar 4.18 Implementasi Kode Pelatihan Model Regresi Logistik Biner.....	72
Gambar 4.19 Implementasi Kode Estimasi Parameter .....	73
Gambar 4.20 Kode Prediksi Kelas Regresi Logistik Biner.....	75
Gambar 4.21 Hasil Akurasi Model Regresi Logistik Biner.....	76
Gambar 4.22 Kode Inisialisasi dan Pelatihan Model <i>Soft Voting Classifier</i> .....	77

Gambar 4.23 Kode Menghitung Nilai Probabilitas <i>Soft Voting Classifier</i> .....	78
Gambar 4.24 Hasil Akurasi Model <i>Soft Voting Classifier</i> .....	79
Gambar 4.25 Perbandingan Hasil Evaluasi Model dengan Metriks Akurasi.....	80
Gambar 4.26 Implementasi Kode <i>Hyperparameter Tuning</i> .....	81
Gambar 4.27 Pemodelan Setelah Penyetelan <i>Hyperparameter Tuning</i> .....	83
Gambar 4.28 Perbandingan Akurasi Model Setelah <i>Hyperparameter Tuning</i> .....	84
Gambar 4.29 Hasil Evaluasi <i>Confusion Matrix</i> .....	86
Gambar 4. 30 Perbandingan Akurasi Masing – Masing Model.....	88

## **DAFTAR LAMPIRAN**

Lampiran 1 .....	101
Lampiran 2 .....	102
Lampiran 3 .....	103