

# **Pemodelan *Turnover* Karyawan PT. Garuda Abdi Satyatama Menggunakan *Feature Importance* Melalui *Logistic Regression* pada Sektor Jasa Keamanan dan Operasional**

**Diajeng Sekar Prameswari<sup>1</sup>, Dwi Arman Prasetya<sup>2</sup>, Amri Muhaimin<sup>3</sup>**

<sup>1,2,3</sup>Universitas Pembangunan Nasional Veteran Jawa Timur

<sup>1,2,3</sup>Surabaya, Indonesia

Email: <sup>1</sup>[Diajeng.sekar11@gmail.com](mailto:Diajeng.sekar11@gmail.com)

## **Abstract**

*Employee turnover remains a major challenge for many companies, including in Indonesia, which recorded a 41% average turnover rate in 2023. This study aims to predict employee turnover using Logistic Regression based on 345 employee records and 24 variables. The process includes data preprocessing, data splitting, modeling, evaluation, and feature importance analysis. Using an 85% training and 15% testing split, the model achieved 90.38% accuracy, 91.67% precision, 73.33% recall, and an F1-score of 81.48%. The confusion matrix shows 36 correct predictions for employees who stayed and 11 for those who left, with 5 misclassifications. Regression coefficient analysis indicates that year of joining (−3.48), length of service (−2.84), job position (−0.99), and age (−0.54) reduce turnover likelihood, while replacement status (0.94), benefits (0.92), and warning status (0.75) increase it. Visualization also shows that employees with shorter tenure are more likely to leave. Logistic Regression is effective and supports retention strategy development in security and operational services.*

**Keywords:** *Classification, employee retention, logistic regression, machine learning, turnover prediction*

## **Abstraksi**

*Turnover karyawan menjadi tantangan bagi banyak perusahaan, termasuk di Indonesia yang mencatat rata-rata 41% pada tahun 2023 dan berdampak pada efektivitas operasional. Penelitian ini bertujuan memprediksi turnover karyawan menggunakan Logistic Regression berdasarkan 345 data karyawan dengan 24 variabel melalui tahapan pra – pemrosesan data, pembagian data, pemodelan, evaluasi, dan analisis feature importance. Performa terbaik diperoleh pada pembagian data training 85% dan testing 15%, dengan akurasi 90,38%, precision 91,67%, recall 73,33%, dan F1-score 81,48%. Confusion matrix menunjukkan 36 prediksi benar untuk kelas bertahan dan 11 untuk kelas keluar, dengan total 5 kesalahan klasifikasi. Analisis koefisien regresi menunjukkan bahwa tahun bergabung (−3,48), lama bekerja (−2,84), jabatan (−0,99), dan usia (−0,54) menurunkan kemungkinan turnover, sedangkan variabel pengganti (0,94), tunjangan (0,92), dan status teguran (0,75) meningkatkan peluang keluarnya karyawan. Visualisasi terlihat bahwa karyawan masa kerja yang lebih singkat lebih rentan keluar. Logistic Regression terbukti efektif untuk prediksi dan informasi feature importance yang mendukung strategi retensi di sektor jasa keamanan dan operasional.*

**Kata Kunci:** *Klasifikasi, logistic regression, machine learning, prediksi turnover, retensi karyawan*

## 1. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat telah mendorong perubahan besar dalam berbagai sektor, termasuk bidang manajemen sumber daya manusia (SDM). Digitalisasi dan penerapan *data analytics* telah memberikan kemudahan bagi perusahaan dalam melakukan pengambilan keputusan berbasis data [1]. Salah satu isu penting yang kini banyak dikaji dalam manajemen SDM adalah fenomena *employee turnover*, yaitu kondisi ketika karyawan meninggalkan perusahaan, baik secara sukarela maupun tidak. *Turnover* yang tinggi dapat berdampak pada meningkatnya biaya rekrutmen, terganggunya stabilitas organisasi, dan menurunnya efektivitas kinerja perusahaan secara keseluruhan [2].

Fenomena *turnover* karyawan terjadi di berbagai perusahaan, termasuk di Indonesia. Menurut data Biro Statistik rata-rata tingkat *turnover* karyawan di Indonesia pada tahun 2023 mencapai 41%. Angka ini menunjukkan bahwa sekitar 41% karyawan di Indonesia meninggalkan pekerjaannya dalam kurun waktu satu tahun. Kondisi tersebut menandakan adanya tantangan serius bagi perusahaan dalam mempertahankan tenaga kerja yang berkualitas serta menciptakan lingkungan kerja yang mampu meningkatkan loyalitas dan kepuasan karyawan [3].

Oleh karena itu, penerapan teknologi analisis data menjadi salah satu solusi strategis untuk memahami pola *turnover*. Melalui analisis prediktif, perusahaan dapat memanfaatkan data historis karyawan seperti usia, masa kerja, jabatan, tingkat kepuasan kerja, dan kompensasi untuk memprediksi kemungkinan seorang karyawan akan keluar. Salah satu metode statistik yang banyak digunakan dalam penelitian prediksi *turnover* adalah *Logistic Regression*, karena kemampuannya dalam memodelkan hubungan antara variabel independen dengan variabel dependen biner seperti “bertahan” atau “keluar.” [4].

*Logistic Regression* merupakan salah satu algoritma klasifikasi yang sederhana namun efektif dalam mengidentifikasi faktor-faktor penting yang mempengaruhi keputusan karyawan [5]. Berbeda dengan algoritma yang lebih kompleks seperti *Random Forest* atau *Neural Network*, *Logistic Regression* memiliki keunggulan dalam interpretabilitas hasil, sehingga mudah dipahami oleh pihak manajemen untuk dijadikan dasar dalam pengambilan keputusan strategis. *Logistic Regression* mampu memberikan interpretasi yang jelas terhadap kontribusi masing-masing variabel terhadap peluang terjadinya turnover [6].

Penelitian oleh Tanujaya, Susanto, dan Saragih (2020) membandingkan performa *Logistic Regression* dan *Random Forest* dalam mengklasifikasikan *mode* audio (*major* dan *minor*) pada data musik *Spotify* tahun 2009–2019 menggunakan *Python* dengan data yang telah dinormalisasi dan dibagi menjadi 70% data latih serta 30% data uji. Hasil penelitian menunjukkan bahwa meskipun *Random Forest* memiliki akurasi sedikit lebih tinggi, *Logistic Regression* unggul dari sisi efisiensi komputasi dan kemudahan interpretasi terhadap hubungan antar variabel. Hasil-hasil ini menunjukkan bahwa *Logistic Regression*

tetap relevan digunakan, terutama untuk penelitian yang menekankan aspek interpretasi dan transparansi hasil [7].

Oleh karena itu, algoritma *Logistic Regression* digunakan dalam memodelkan *turnover* karyawan pada sektor jasa keamanan dan operasional. Perusahaan PT. Garuda Abdi Satyatama, sebagai objek penelitian, merupakan perusahaan *outsourcing* yang menghadapi tingkat *turnover* cukup tinggi dalam dua tahun terakhir. Berdasarkan hasil wawancara dengan pihak manajemen, tingginya tingkat *turnover* diduga dipengaruhi oleh faktor kompensasi (gaji) dan lingkungan kerja. Oleh karena itu, dengan total 345 data yang digunakan mencakup informasi historis karyawan seperti demografi, jabatan, masa kerja, kompensasi, jam kerja, dan pelatihan, dengan status *turnover* sebagai variabel dependen.

Penelitian ini bertujuan untuk mengklasifikasikan dan memprediksi *turnover* karyawan pada sektor jasa keamanan dan operasional menggunakan algoritma *Logistic Regression* untuk mengidentifikasi faktor-faktor signifikan yang memengaruhi keputusan karyawan keluar. [4]. Hasil penelitian diharapkan menjadi dasar bagi divisi Human Resources dalam merancang strategi retensi berbasis data, seperti pengelolaan insentif, pelatihan, dan beban kerja. Pendekatan ini memperkuat penerapan *data science* dalam manajemen sumber daya manusia untuk menekan *turnover* dan meningkatkan stabilitas tenaga kerja secara berkelanjutan.

## 2. TINJAUAN PUSTAKA

### 2.1. Turnover Karyawan

Menurut Stephen dan Timothy (2017), *turnover* karyawan merupakan proses keluarnya individu dari organisasi, baik secara sukarela (*voluntary turnover*) maupun tidak sukarela (*involuntary turnover*). *Turnover* mencerminkan stabilitas tenaga kerja dan menjadi indikator penting dalam menilai efektivitas kebijakan sumber daya manusia [8]. Tingkat *turnover* yang tinggi menunjukkan adanya ketidakseimbangan antara harapan karyawan dan kondisi kerja yang ditawarkan organisasi, yang dapat dipengaruhi oleh faktor kepuasan kerja, kompensasi, serta kesempatan pengembangan karier [9].

Beberapa penelitian menunjukkan bahwa keputusan karyawan untuk keluar dipengaruhi oleh faktor demografis, kompensasi, dan kondisi kerja. Chen (2023) menemukan bahwa usia, gaji, dan status pernikahan berpengaruh signifikan terhadap *turnover* dengan *Logistic Regression* menunjukkan akurasi tinggi [10]. Hasil serupa dilaporkan oleh Wardhani dan Lhaksmana (2022) yang menyoroti peran kepuasan kerja dan faktor demografis terhadap retensi karyawan. Kesimpulannya, faktor penyebab *turnover* bersifat multidimensional sehingga diperlukan analisis berbasis data seperti *Logistic Regression* untuk memahami pola dan memprediksi risiko *turnover* secara lebih akurat [5].

### 2.2. Klasifikasi

Klasifikasi merupakan salah satu teknik dalam *machine learning* dan *data mining* yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan

pola dari data historis. Proses pengelompokan data secara manual sering kali memerlukan waktu dan pengetahuan khusus, sehingga penerapan klasifikasi menjadi penting untuk mengotomatisasi proses secara lebih cepat dan akurat [11]. Proses ini melibatkan dua tahap utama, yaitu *training* untuk mempelajari hubungan antara variabel independen dan dependen, serta *testing* untuk mengevaluasi kemampuan model dalam mengklasifikasikan data baru secara tepat [12].

Berbagai algoritma telah digunakan dalam proses klasifikasi, antara lain *Logistic Regression*, *Decision Tree*, *Random Forest*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN). Dalam penelitian ini, *Logistic Regression* dipilih karena mampu menggambarkan hubungan probabilitas antara variabel prediktor dan variabel dependen yang bersifat biner, seperti keputusan karyawan untuk bertahan atau mengundurkan diri. Selain itu, algoritma ini dikenal efisien, mudah diinterpretasikan, dan cocok untuk analisis faktor-faktor yang memengaruhi *turnover* karyawan [13].

### 2.3. Logistic Regression

*Logistic Regression* merupakan metode statistik yang digunakan untuk memprediksi probabilitas terjadinya suatu peristiwa berdasarkan variabel independen (Hosmer & Lemeshow, 2013). Berbeda dengan regresi linear, metode ini digunakan ketika variabel dependen bersifat kategori atau biner (misalnya, keluar = 1 dan bertahan = 0) [14]. Menurut Naman (2023), keunggulan *Logistic Regression* adalah kemampuannya menganalisis hubungan non-linear antara variabel independen dan peluang terjadinya peristiwa tertentu. Model umum dari regresi logistik berganda dapat dituliskan pada rumus 1 berikut [15].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \varepsilon \quad (1)$$

dengan:

$Y$  = variabel respon,  
 $X_i$  = variabel independen ke- $i$ ,  
 $\beta_0$  = konstanta/intersep,  
 $\beta_i$  = koefisien regresi, dan  
 $\varepsilon$  = galat.

Dalam regresi logistik, asumsi klasik regresi linear seperti *normalitas* galat tidak berlaku secara ketat. Namun, beberapa asumsi yang umum digunakan mencakup: (1) nilai rata-rata galat adalah nol, (2) variansi galat bersifat konstan (*homoskedastik*), (3) tidak terjadi *autokorelasi* antar galat, dan (4) tidak terdapat *multikolinearitas* tinggi antar variabel independen. Parameter  $\varepsilon$  umumnya tidak dituliskan dalam model karena pengaruhnya dianggap tidak signifikan terhadap hasil estimasi probabilitas [16].

### 2.4. Penelitian Terdahulu

Pada penelitian terdahulunya, yang dilakukan oleh Wahyuningsih *et al.* (2024) membandingkan *Logistic Regression*, *Naive Bayes*, dan *Random Forest* untuk memprediksi

kualitas argumentasi siswa, dan menunjukkan bahwa *Logistic Regression* memiliki akurasi terbaik sebesar 94,34%, mengungguli dua metode lainnya [17]. Temuan serupa juga ditunjukkan oleh Denny dan Herwindiati (2024), yang membandingkan *K-Nearest Neighbors*, *Naïve Bayes*, dan *Logistic Regression* dalam klasifikasi kualitas air, di mana *Logistic Regression* kembali menjadi algoritma paling stabil dengan akurasi 89,81% [18]. Temuan-temuan ini menegaskan bahwa *Logistic Regression* konsisten memberikan kinerja unggul pada berbagai jenis data. Namun, penerapan model pada topik *turnover* karyawan khususnya di sektor jasa keamanan dan operasional yang memiliki pola *turnover* yang berbeda dibanding sektor lain masih jarang dilakukan. Oleh karena itu, penelitian ini penting dilakukan untuk memberikan analisis yang lebih sesuai karakteristik sektor tersebut.

Penelitian oleh Wardhani dan Lhaksmana (2022) memprediksi *turnover* karyawan menggunakan *Logistic Regression* dengan dan tanpa seleksi fitur melalui RFE, menghasilkan akurasi masing-masing 86,5% dan 85,3% serta AUC 0,925 [5]. Adapun penelitian Chen (2023) memanfaatkan 13 variabel numerik dan kategorik dalam *Logistic Regression* dengan evaluasi Omnibus dan *Hosmer–Lemeshow test*, yang menunjukkan akurasi 85,9% dan mengidentifikasi masa kerja serta kepuasan kerja sebagai faktor signifikan terhadap *turnover* [10]. Meskipun kedua penelitian tersebut menggunakan *Logistic Regression* untuk memprediksi *turnover*, keduanya belum menampilkan analisis *feature importance* secara jelas. Celah ini menjadi dasar penelitian ini, yang tidak hanya memprediksi *turnover* tetapi juga mengidentifikasi fitur paling berpengaruh pada karyawan *outsourcing* di sektor jasa keamanan dan operasional.

Berdasarkan hasil temuan tersebut, terlihat bahwa *Logistic Regression* banyak diterapkan pada berbagai topik prediksi. Namun, penelitian yang fokus pada *turnover* di sektor jasa keamanan dan operasional masih terbatas. Selain itu, sebagian besar penelitian sebelumnya membahas analisis *feature importance* sehingga faktor-faktor yang paling berpengaruh terhadap *turnover* tidak terlihat secara jelas. Oleh karena itu, penelitian ini mengisi kekosongan tersebut dengan menerapkan *Logistic Regression* yang dilengkapi analisis *feature importance*, sehingga mampu memberikan gambaran yang lebih komprehensif dan relevan terhadap pola *turnover* karyawan pada sektor jasa keamanan dan operasional.

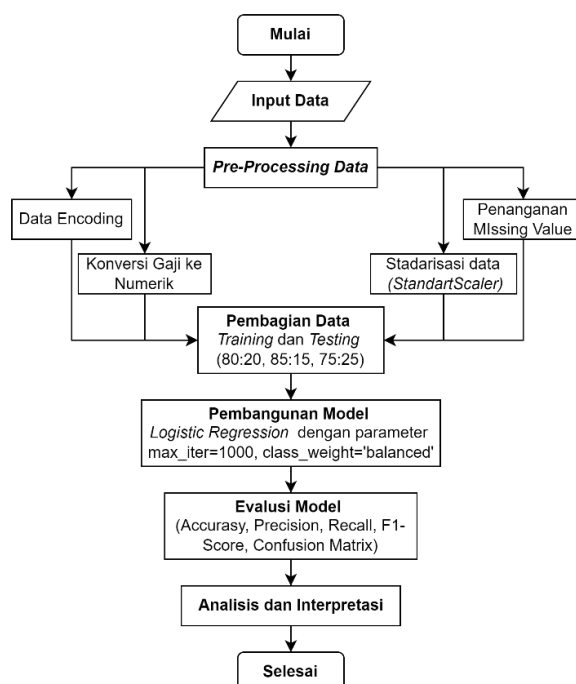
### 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan algoritma *Logistic Regression* untuk memprediksi *employee turnover* pada sektor jasa keamanan dan operasional di PT. Garuda Abdi Satyatama. Pendekatan ini dipilih karena mampu mengukur hubungan *antarvariabel* secara objektif dan menghasilkan model prediksi yang terukur. Data yang digunakan merupakan data sekunder dari PT. Garuda Abdi Satyatama dengan jumlah 345 data yang memuat variabel lokasi kerja, daerah kerja, jenis kelamin, jabatan, tahun *join*, gaji, usia, status pernikahan, pelatihan, pengalaman kerja, tahun pelatihan terakhir, pengganti, menggantikan, pendidikan terakhir, tunjangan, teguran,

hari kerja per minggu, jam kerja per hari, total jam kerja per minggu, jam kerja per bulan, lembur sebulan per jam, total jam kerja, serta lama bekerja (bulan) sebagai variabel independen, serta tanggal *out* sebagai variabel dependen.

Seluruh data diolah menggunakan *Python*, pada tahap awal penelitian dimulai dengan *pre processing* data, seperti penanganan *missing values* menggunakan *SimpleImputer* dengan median untuk variabel numerik dan modus untuk variabel kategorikal. Nilai gaji dikonversi ke format numerik, sedangkan variabel kategorikal diubah menjadi numerik menggunakan *LabelEncoder*. Setelah itu, dilakukan pembagian data menjadi tiga skenario *training:testing* (80:20, 75:25, dan 85:15) untuk melihat pengaruh proporsi data terhadap performa model.

Model *Logistic Regression* dibangun dengan parameter *max\_iter=1000* dan *class\_weight='balanced'* untuk menyesuaikan distribusi kelas yang tidak seimbang. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix* untuk menilai performa model. Selanjutnya, analisis koefisien regresi digunakan untuk menentukan variabel yang paling berpengaruh terhadap *turnover*. Berikut diagram alir penelitian yang terdapat pada gambar 1 berikut.



Gambar 1. Flowchart Pemodelan Logistic Regression

## 4. HASIL DAN PEMBAHASAN

### 4.1. Hasil Pemodelan Logistic Regression

Penelitian ini dilakukan untuk memodelkan prediksi *turnover* karyawan pada sektor jasa keamanan dan operasional menggunakan algoritma *Logistic Regression*. Data yang digunakan merupakan data sekunder dari perusahaan. Proses analisis dilakukan melalui beberapa tahapan, tahap awal data melalui proses *pre – processing* yang meliputi

penanganan *missing values*, konversi variabel *kategorikal* menjadi numerik, serta penyesuaian format gaji agar seluruh variabel dapat diolah oleh model. Setelah data dibersihkan, dilakukan pembagian data ke dalam tiga skenario pelatihan dan pengujian (80:20, 75:25, dan 85:15) untuk membandingkan pengaruh proporsi data terhadap performa model.

Pemodelan *Logistic Regression* kemudian dibangun menggunakan parameter *max\_iter* 1000 dan *class\_weight* "balanced" untuk mengatasi ketidakseimbangan kelas. Pada proses pelatihan inilah rumus *Logistic Regression* digunakan, yaitu dengan menghitung koefisien  $\beta$  untuk setiap variabel independen. Koefisien tersebut menunjukkan arah dan besar pengaruh masing-masing variabel terhadap probabilitas *turnover*. Nilai koefisien inilah yang kemudian dijadikan dasar analisis *feature importance* untuk mengidentifikasi variabel yang paling berpengaruh dalam memengaruhi keputusan karyawan keluar.

Setelah model selesai dilatih, dilakukan evaluasi dengan tiga skenario pembagian data, untuk melihat stabilitas kinerja model pada proporsi data yang berbeda. Hasil pengujian performa model dapat dilihat pada Tabel 1 berikut.

Tabel 1. Perbandingan Metrik Evaluasi Model *Logistic Regression*

<i>Split</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
80:20	0.898551	0.866667	0.722222	0.787879
75:25	0.862069	0.833333	0.625000	0.714286
<b>85:15</b>	<b>0.903846</b>	<b>0.916667</b>	<b>0.733333</b>	<b>0.814815</b>

Berdasarkan hasil pada Tabel 1, model dengan proporsi 85:15 memberikan performa terbaik dengan nilai *accuracy* sebesar 90,38%, *precision* sebesar 91,67%, dan *F1-score* sebesar 0,8148. Hal ini menunjukkan bahwa model cukup baik dalam mengklasifikasikan *turnover* dan yang bertahan. Untuk melihat secara lebih rinci hasil prediksi antara kelas aktual dan hasil model, digunakan *Confusion Matrix* pada masing-masing skenario pembagian data seperti pada Tabel 2, Tabel 3, dan Tabel 4 berikut.

Tabel 2. *Confusion Matrix Split 80:20*

	<i>Predicted (0)</i>	<i>Predicted (1)</i>
<b>0</b>	49	2
<b>1</b>	5	13

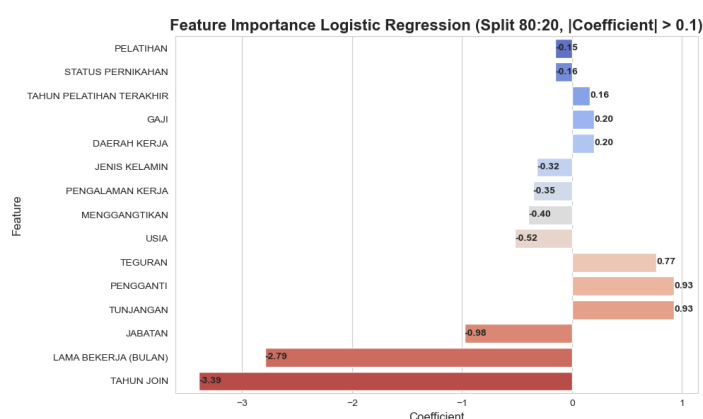
Tabel 3. *Confusion Matrix Split 75:25*

	<i>Predicted (0)</i>	<i>Predicted (1)</i>
<b>0</b>	60	3
<b>1</b>	9	15

Tabel 4. *Confusion Matrix Split 85:15*

	<i>Predicted (0)</i>	<i>Predicted (1)</i>
<b>0</b>	36	1
<b>1</b>	4	11

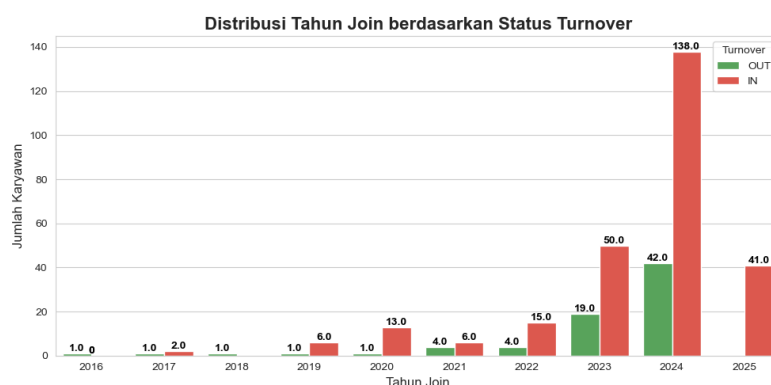
Berdasarkan hasil pada ketiga tabel tersebut, terlihat bahwa pada skenario 85:15, model berhasil mengenali lebih banyak data secara benar dibandingkan dengan dua skenario lainnya. Model mampu mengklasifikasikan 36 karyawan bertahan dan 11 karyawan keluar dengan tepat, hanya menghasilkan sedikit kesalahan klasifikasi (4 *false negative* dan 1 *false positive*). Hal ini menunjukkan kestabilan dan akurasi model yang tinggi. Selanjutnya, dilakukan analisis terhadap persebaran *feature importance* untuk mengetahui variabel yang paling berpengaruh terhadap keputusan *turnover*. Variabel yang paling berpengaruh ditunjukkan pada Gambar 2 berikut.



Gambar 2 Feature Importance Logistic Regression (Split 85:15)

Dari hasil tersebut, variabel dengan koefisien negatif seperti Tahun Join (-3.4850), Lama Bekerja (-2.8463), Jabatan (-0.9915), dan Usia (-0.5494) berperan dalam menurunkan kemungkinan *turnover*, sementara variabel seperti Pengganti (0.9456), Tunjangan (0.9277), Status Teguran (0.7514) memiliki koefisien positif tinggi, menunjukkan peningkatan potensi *turnover* pada karyawan dengan kondisi tersebut.

Selain itu, dilakukan visualisasi hubungan antara Tahun *Join* dengan Status *Turnover* untuk memahami pola distribusi masa kerja terhadap keputusan karyawan pada gambar 3 berikut.



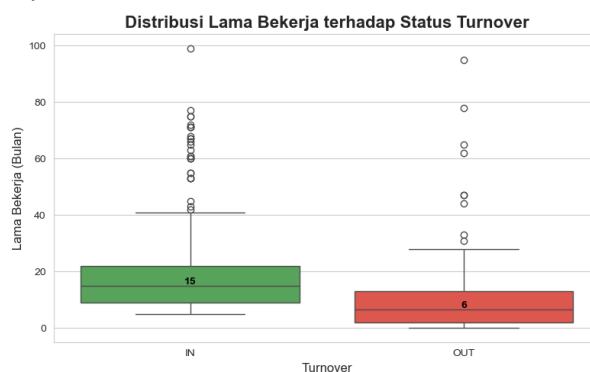
Gambar 3. Distribusi Tahun *Join* berdasarkan Status *Turnover*

Pada Gambar 3 terlihat jika karyawan yang bergabung pada tahun – tahun awal memiliki kecenderungan bertahan lebih lama dibandingkan karyawan yang baru bergabung. Mayoritas karyawan yang keluar (status 1) berasal dari kelompok yang



bergabung pada tahun – tahun lebih baru (2023 – 2024), yang menandakan bahwa masa awal bekerja merupakan periode paling rentan terhadap *turnover*.

Selanjutnya, dilakukan analisis terhadap lama bekerja (bulan) untuk melihat perbedaan durasi kerja antara karyawan yang bertahan dan yang keluar. Visualisasi hasilnya ditunjukkan pada Gambar 4 berikut.



Gambar 4. Distribusi Lama Bekerja terhadap Status *Turnover*

Berdasarkan Gambar 4, terlihat bahwa karyawan dengan status IN (bertahan) memiliki rata-rata lama bekerja sekitar 15 bulan, sedangkan karyawan OUT (keluar) memiliki rata-rata hanya sekitar 6 bulan. Hal ini menunjukkan bahwa durasi kerja menjadi salah satu indikator penting dalam prediksi *turnover*.

Karyawan dengan masa kerja yang lebih panjang cenderung memiliki loyalitas dan adaptasi yang lebih baik terhadap lingkungan kerja, sedangkan mereka yang keluar umumnya belum melewati satu tahun masa kerja. Pola ini sejalan dengan hasil *feature importance* di mana variabel Lama Bekerja (Bulan) memiliki pengaruh negatif terhadap peluang *turnover*.

## 4.2. Pembahasan

Berdasarkan hasil pengujian, algoritma *Logistic Regression* menunjukkan performa klasifikasi yang baik dengan tingkat akurasi di atas 85% pada seluruh skenario pembagian data. Model terbaik diperoleh pada rasio 85:15 dengan akurasi sebesar 90,38% dan nilai F1-score tertinggi, menunjukkan bahwa model mampu mengenali pola *turnover* secara akurat. Proporsi data latih yang lebih besar menjadikan model lebih stabil dalam memprediksi karyawan yang bertahan dan yang keluar. Hasil analisis koefisien menunjukkan bahwa faktor-faktor seperti lama bekerja, tahun *join*, usia, dan pengalaman kerja berpengaruh negatif terhadap *turnover*, artinya semakin lama masa kerja dan semakin berpengalaman karyawan, semakin kecil kemungkinan mereka untuk keluar. Sebaliknya, variabel seperti tunjangan, teguran, dan status pengganti berpengaruh positif terhadap *turnover*, yang dapat menandakan adanya ketidakpuasan, beban kerja lebih, atau lingkungan yang tidak mendukung. Hal ini menunjukkan bahwa faktor finansial dan lingkungan kerja berperan penting dalam keputusan karyawan untuk bertahan atau keluar. Visualisasi pada Gambar 2 dan 3 memperkuat hasil analisis bahwa karyawan dengan masa kerja singkat dan tahun masuk yang lebih baru lebih rentan mengalami

*turnover*. Oleh karena itu, perusahaan perlu menyeimbangkan beban kerja, memperjelas sistem pengganti, serta meninjau kembali kebijakan tunjangan agar lebih adil dan transparan. Pendekatan ini diharapkan dapat menekan *turnover*, terutama pada karyawan baru. Secara keseluruhan, *Logistic Regression* terbukti efektif dalam memprediksi *turnover* dan mendukung HRD dalam menyusun strategi retensi yang berbasis data.

## 5. KESIMPULAN

Berdasarkan hasil penelitian ini, beberapa faktor personal dan pekerjaan terbukti berpengaruh terhadap kemungkinan karyawan keluar. Model *Logistic Regression* mampu mengklasifikasi *turnover* dengan baik, mencapai akurasi 90,38% pada rasio data 85:15. Faktor lama bekerja, tahun *join*, usia, dan pengalaman kerja menurunkan risiko *turnover*, sedangkan tunjangan, teguran, dan status pengganti meningkatkan potensi keluarnya karyawan. Visualisasi juga menunjukkan karyawan dengan masa kerja singkat dan tahun masuk lebih baru (2023–2024) lebih rentan keluar, sehingga diperlukan peningkatan orientasi, pembinaan, dan dukungan adaptasi kerja. Meskipun mudah diinterpretasikan, *Logistic Regression* masih kurang optimal dalam mengenali kelas minoritas karena nilai *recall* yang rendah. Untuk penelitian selanjutnya, disarankan menerapkan teknik *balancing* data, menambahkan variabel seperti kepuasan dan evaluasi kinerja, serta membandingkan dengan algoritma lain untuk meningkatkan performa prediksi.

## DAFTAR PUSTAKA

- [1] N. Srivastava dan I. Bhardwaj, "Impact of HR Analytics on Organizational Performance: A Modern Approach in HR," *HR Analytics: Fundamentals and Applications*, hlm. 24–41, Nov 2024, doi: 10.2174/9789815274196124010005.
- [2] M. V. Chudinovskikh, A. N. Tkach, dan D. Y. Korolkov, "Staff turnover management based on HR analytics data," *Digital models and solutions*, vol. 3, no. 3, hlm. 88–97, Okt 2024, doi: 10.29141/2949-477X-2024-3-3-7.
- [3] Awardco Staff, "How Companies with High Turnover Rates Turned Things Around," Awardco.
- [4] I. Setiawan, S. Suprihanto, A. C. Nugraha, dan J. Hutahaeen, "HR analytics: Employee attrition analysis using logistic regression," *IOP Conf Ser Mater Sci Eng*, vol. 830, no. 3, hlm. 032001, Apr 2020, doi: 10.1088/1757-899X/830/3/032001.
- [5] F. H. Wardhani dan K. M. Lhaksmana, "Predicting Employee Attrition Using Logistic Regression With Feature Selection," *Sinkron*, vol. 7, no. 4, hlm. 2214–2222, Okt 2022, doi: 10.33395/sinkron.v7i4.11783.
- [6] A. S. Budiman, Desmulyati, dan Fahrizal, "Uji Akurasi Algoritma Machine Learning Untuk Pemodelan Prediksi Faktor Pendorong Pergantian Karyawan," *JSI: Jurnal Sistem Informasi (E-Journal)*, vol. 15, no. 1, hlm. 3006–3018, Apr 2023, doi: 10.18495/jsi.v15i1.104.

- [7] L. Britanthia, C. Tanujaya, B. Susanto, dan A. Saragih, "Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Fitur Mode Audio Spotify," *Indonesian Journal of Data and Science (IJODAS)*, vol. 1, no. 3, hlm. 68–78, 2020.
- [8] Stephen P. Robbins dan Timothy A. Judge, *Organizational Behavior*, 17th Edition. Pearson, 2017.
- [9] D. A. Sulastri, V. R. Zainal, dan A. Hakim, "Key Success Factors for Organizations in Controlling Employee Turnover," *Jurnal Indonesia Sosial Sains*, vol. 6, no. 4, hlm. 1169–1177, Apr 2025, doi: 10.59141/jiss.v6i4.1698.
- [10] B. Chen, "Factors of Employee Attrition: A Logistic Regression Approach," *Advances in Economics, Management and Political Sciences*, vol. 20, no. 1, hlm. 214–225, Sep 2023, doi: 10.54254/2754-1169/20/20230198.
- [11] A. A. Rahmawati, A. Muhaimin, dan D. A. Prasetya, "CLASSIFICATION OF JAVANESE NGLEGENA SCRIPT USING COMPLEXVALUED NEURAL NETWORK," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 1, hlm. 30–35, Apr 2024, doi: 10.33387/jiko.v7i1.7808.
- [12] A. Majumder, "Classification Models in Machine Learning Techniques," 2023, hlm. 1–16. doi: 10.4018/978-1-6684-8531-6.ch001.
- [13] R. Gupta, A. S. Chand, S. Solanki, T. Gautam, dan N. Garg, "A Comparative Analysis of Logistic Regression and Support Vector Machine for Employee Churn Prediction," dalam *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, IEEE, Okt 2024, hlm. 1821–1827. doi: 10.1109/ICSES63445.2024.10762955.
- [14] D. Banerjee, R. Kumar, S. Tripathi, dan B. Murry, "Application of Binary Logistic Regression in Biological Studies," *Journal of the Practice of Cardiovascular Sciences*, vol. 10, no. 1, hlm. 48–52, Jan 2024, doi: 10.4103/jpcs.jpcs\_82\_23.
- [15] N. Kaur dan Himanshu, "Logistic Regression: A Basic Approach," 2023, hlm. 481–488. doi: 10.1007/978-981-19-9638-2\_41.
- [16] A. F. Lukman, S. Mohammed, O. Olaluwoye, dan R. A. Farghali, "Handling Multicollinearity and Outliers in Logistic Regression Using the Robust Kibria–Lukman Estimator," *Axioms*, vol. 14, no. 1, hlm. 19, Des 2024, doi: 10.3390/axioms14010019.
- [17] T. Wahyuningsih, D. Manongga, I. Sembiring, dan S. Wijono, "Comparison of Effectiveness of Logistic Regression, Naive Bayes, and Random Forest Algorithms in Predicting Student Arguments," *Procedia Comput Sci*, vol. 234, hlm. 349–356, 2024, doi: 10.1016/j.procs.2024.03.014.
- [18] M. S. Denny dan D. E. Herwindiati, "Klasifikasi Kualitas Air Menggunakan K-Nearest Neighbors, Naïve Bayes, Dan Logistic Regression," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 4, hlm. 851–858, Nov 2024, doi: 10.47233/jteksis.v6i4.1649.