

Analisa Prediksi *Turnover* Karyawan menggunakan *Machine Learning*

Arief Maehendrayuga^{1)*}, Arief Setyanto²⁾, Kusnawi³⁾

¹⁾²⁾³⁾Magister Teknik Informatika, Universitas Amikom
Yogyakarta, Indonesia

¹⁾ariefmaehendrayuga@student.amikom.ac.id

²⁾arief_s@amikom.ac.id

³⁾khusnawi@amikom.ac.id

Article history:

Received 26 Des 2024;
Revised 26 Des 2024;
Accepted 26 Des 2024;
Available online 27 Des 2024

Keywords:

Data Imbalance
Employee Attrition
Machine Learning
Random Forest
SVM

Abstrak

Penelitian ini membahas penerapan machine learning untuk memprediksi turnover karyawan, yang merupakan tantangan utama dalam manajemen Sumber Daya Manusia (SDM). Turnover karyawan sering kali disebabkan oleh berbagai faktor, termasuk ketidakseimbangan kehidupan kerja, ketidakpuasan kerja, dan minimnya peluang pengembangan karier. Dalam penelitian ini, digunakan dataset IBM HR Analytics untuk menganalisis faktor-faktor yang memengaruhi turnover karyawan. Algoritma yang diterapkan meliputi Support Vector Machine (SVM) dan Random Forest. Proses penelitian dimulai dengan pengumpulan data, eksplorasi awal, praproses data, seleksi fitur, dan penyeimbangan data menggunakan teknik Synthetic Minority Over-sampling Technique (SMOTE). Evaluasi kinerja model dilakukan menggunakan confusion matrix untuk mengukur akurasi, presisi, recall, dan f1-score. Hasil analisis menunjukkan bahwa algoritma Random Forest memberikan kinerja yang lebih baik dibandingkan SVM. Random Forest mencapai akurasi 97,72%, sedangkan SVM memperoleh akurasi 92,51%. Setelah menerapkan SMOTE, akurasi meningkat menjadi 97% untuk Random Forest dan 93% untuk SVM. Selain akurasi, Random Forest juga unggul dalam metrik presisi, recall, dan f1-score, membuktikan keandalannya dalam memprediksi turnover karyawan. Temuan ini menegaskan bahwa pendekatan machine learning dapat digunakan untuk memahami pola turnover secara lebih mendalam. Dengan prediksi yang lebih akurat, perusahaan dapat merancang strategi retensi karyawan yang lebih efektif dan berbasis data, menciptakan lingkungan kerja yang mendukung produktivitas serta meningkatkan stabilitas tenaga kerja secara keseluruhan.

I. PENDAHULUAN

Sumber daya manusia (SDM) merupakan aset berharga yang dimiliki oleh setiap perusahaan. SDM harus dikelola untuk memenuhi tujuan organisasi perusahaan. Mulai dari sistem seleksi dan penempatan karyawan hendaknya perlu mendapatkan perhatian lebih dari perusahaan agar satu kesatuan dalam organisasi tersebut dapat bersinergi, berkolaborasi, dan berkontribusi satu sama lain [1]

Persaingan bisnis yang dihadapi oleh perusahaan untuk tumbuh dan berkembang sangatlah kompleks. Perusahaan harus sadar dengan tantangan, peluang, dan kekuatan yang dimilikinya. Dengan semua daya yang dimiliki, perusahaan dapat memanfaatkannya agar tetap berjalan dan berkembang. Berbagai fokus permasalahan menjadi tantangan yang dihadapi perusahaan [2]

Berbagai upaya dilakukan untuk mempertahankan keberlangsungan perusahaan, salah satunya adalah merekrut tenaga kerja yang kompeten. Namun, penelitian menunjukkan bahwa penarikan karyawan baru bukanlah satu-satunya upaya efektif untuk mengoptimalkan produktivitas perusahaan. Karyawan baru harus mampu menyesuaikan diri dengan budaya dan beban kerja yang dihadapi nantinya, sehingga hal ini menjadi tantangan sekaligus ancaman bagi perusahaan. Langkah yang lebih efisien adalah mengelola sumber daya manusia yang tersedia agar mereka dapat berkontribusi secara maksimal [3].

Karyawan direkrut berdasarkan kompetensi dan keahlian yang dimilikinya. Langkah yang dilakukan oleh perusahaan dalam mengoptimalkan kinerja sering kali melibatkan pemindahan karyawan. Beban kerja yang tinggi

*Corresponding author

dan perpindahan karyawan ke unit kerja tertentu dengan beban kerja berbeda dapat mengakibatkan hilangnya kompetensi dan keahlian. Hal ini menjadi permasalahan baru bagi perusahaan [4]

On-job training dalam menyelesaikan perpindahan karyawan membutuhkan biaya yang cukup besar, serta keahlian tertentu tidak dapat dipelajari dalam waktu singkat. Oleh karena itu, sering kali on-job training tidak menjadi solusi optimal dalam perpindahan karyawan. Perubahan mindset seseorang dalam bekerja perlu dipertimbangkan dalam proses perpindahan unit kerja. Deskripsi pekerjaan yang kurang dipahami sering kali menjadi kendala bagi karyawan yang mengalami rotasi atau perpindahan [5]

Turnover karyawan atau perpindahan karyawan dipicu oleh beragam faktor seperti ketidakseimbangan antara pekerjaan dan kehidupan pribadi, ketidakpuasan terhadap pekerjaan, dan kurangnya peluang pengembangan karier. Berdasarkan survei JobStreet.com, sekitar 73% pekerja merasa tidak puas dengan kondisi kerja mereka, menunjukkan perlunya pendekatan manajemen SDM yang lebih efektif dan strategis [6]. Penelitian lain juga menunjukkan bahwa stres kerja, komitmen organisasi, dan kepuasan kerja memiliki pengaruh signifikan terhadap niat karyawan untuk meninggalkan perusahaan [7]. Selain itu, faktor seperti beban kerja yang berlebihan, gaya kepemimpinan yang tidak suportif, dan kompensasi yang tidak memadai turut berkontribusi terhadap turnover karyawan [8]. Oleh karena itu, organisasi perlu mempertimbangkan faktor-faktor tersebut dalam merancang strategi retensi karyawan yang efektif.

Perpindahan posisi dan jabatan karyawan merupakan salah satu permasalahan utama yang dihadapi dalam dinamika organisasi. Perubahan tersebut dapat berdampak signifikan terhadap aspek psikologis karyawan, kompetensi, serta keahlian yang dimiliki. Perubahan posisi kerja dapat menyebabkan stres dan kecemasan, karena karyawan harus beradaptasi dengan tuntutan pekerjaan yang baru, yang sering kali melibatkan beban kerja yang lebih berat atau tanggung jawab yang berbeda [9]. Selain itu, menunjukkan bahwa transisi jabatan berisiko mengurangi efisiensi kerja, terutama jika karyawan tidak memiliki keahlian atau pengalaman yang sesuai dengan posisi baru. Oleh karena itu, untuk meminimalisir dampak negatif dari perubahan posisi, penting bagi organisasi untuk mempertimbangkan kesesuaian antara profesi atau keahlian yang dimiliki karyawan dengan jabatan yang diberikan. Hal ini akan menjamin sinergitas dan kontribusi maksimal karyawan dalam melaksanakan tugas-tugas yang baru. Sejalan dengan hal tersebut, penyesuaian yang tepat antara kompetensi karyawan dan pekerjaan baru akan meningkatkan kinerja serta kepuasan kerja, yang pada gilirannya berdampak positif pada keberlanjutan organisasi secara keseluruhan [10].

Data yang diperoleh dapat dianalisis menggunakan machine learning. Perkembangan machine learning saat ini sudah pesat. Dengan menganalisis data historis, algoritma seperti Support Vector Machine (SVM) dan Random Forest dapat mengungkapkan pola serta faktor-faktor utama yang memengaruhi keputusan karyawan untuk tetap bekerja atau keluar dari perusahaan [11]. Penggunaan machine learning berfokus pada pengembangan algoritma yang memungkinkan komputer untuk menganalisis data. Berbagai keputusan dan prediksi dapat dihasilkan menggunakan machine learning. Berdasarkan hal tersebut, penelitian ini bertujuan untuk merancang model prediksi Turnover karyawan menggunakan data dari IBM HR Analytics Employee Turnover and Performance. Penelitian ini diharapkan dapat memberikan wawasan yang mendalam tentang faktor-faktor Turnover karyawan sekaligus membantu perusahaan mengembangkan strategi retensi berbasis data yang lebih efektif.

II. TINJAUAN PUSTAKA

Manajemen Sumber Daya Manusia (SDM) memainkan peran yang sangat penting dalam menunjang keberhasilan suatu organisasi. Manajemen SDM mencakup berbagai aspek mulai dari seleksi, pelatihan, hingga pengembangan karyawan, yang bertujuan agar karyawan dapat memberikan kontribusi yang optimal dalam organisasi [4]. Salah satu aspek utama dalam manajemen SDM adalah sistem seleksi dan penempatan yang efektif. Dengan sistem ini, organisasi dapat membangun sinergi antara karyawan, yang pada gilirannya akan mempercepat pencapaian tujuan bersama, terutama dalam menghadapi persaingan bisnis yang semakin kompleks dan dinamis. Pemilihan karyawan yang tepat tidak hanya berpengaruh pada kinerja individu, tetapi juga pada efektivitas tim secara keseluruhan [1]

Pentingnya bagi perusahaan untuk mengenali tantangan dan peluang yang ada agar tetap relevan dalam industri yang terus berubah [2]. Keberlanjutan organisasi sangat dipengaruhi oleh kemampuannya untuk memanfaatkan kekuatan internal, yakni sumber daya manusia yang ada. Dalam hal ini, perusahaan harus mampu mengoptimalkan potensi karyawan agar dapat bertahan dan berkembang dalam menghadapi berbagai tantangan. Salah satu langkah strategis yang dapat diambil adalah dengan melakukan perekrutan tenaga kerja yang kompeten, yang diharapkan dapat memberikan kontribusi yang signifikan bagi organisasi. Namun, penelitian menunjukkan bahwa rekrutmen bukanlah solusi tunggal yang dapat diandalkan untuk mengatasi tantangan SDM. Pengelolaan karyawan yang sudah ada lebih efisien daripada hanya mengandalkan perekrutan karyawan baru. Dalam hal ini, pengembangan karyawan yang ada menjadi sangat penting, karena hal ini tidak hanya mengurangi biaya rekrutmen, tetapi juga meningkatkan loyalitas dan keterlibatan karyawan dalam jangka panjang [12].

Namun, meskipun pengelolaan karyawan yang ada sudah berjalan dengan baik, tantangan besar yang sering dihadapi oleh perusahaan adalah turnover karyawan yang tinggi. Perpindahan karyawan yang tidak direncanakan

dengan matang dapat mengakibatkan hilangnya kompetensi dan produktivitas, yang pada akhirnya akan merugikan organisasi [4]. Turnover karyawan yang tinggi menambah beban bagi perusahaan dalam hal biaya rekrutmen dan pelatihan karyawan baru. Oleh karena itu, penting bagi perusahaan untuk mengidentifikasi faktor-faktor yang memengaruhi turnover dan merancang kebijakan yang dapat meminimalkan dampaknya.

Pentingnya program pelatihan berbasis pekerjaan (on the job training) dalam mendukung proses adaptasi karyawan terhadap unit kerja baru [5]. Program pelatihan ini dapat mengurangi kecemasan dan meningkatkan rasa percaya diri karyawan, yang pada gilirannya akan menurunkan tingkat turnover. Selain itu, untuk mengoptimalkan pengelolaan turnover, perusahaan dapat memanfaatkan teknologi dan data analitik. Penggunaan algoritma seperti Support Vector Machine (SVM) dan Random Forest, perusahaan dapat mengidentifikasi pola-pola tertentu yang memengaruhi keputusan karyawan untuk bertahan atau meninggalkan organisasi [11]. Pendekatan berbasis data ini memungkinkan perusahaan untuk merancang model prediksi yang lebih akurat dan mengembangkan strategi retensi yang lebih efektif.

Penelitian berbasis data ini memberikan pemahaman yang lebih mendalam tentang dinamika turnover karyawan, sehingga perusahaan dapat mengambil keputusan yang lebih tepat dan berbasis bukti. Pemanfaatan data historis, perusahaan dapat lebih proaktif dalam mengelola sumber daya manusia dan merancang kebijakan yang lebih tepat sasaran [11]. Dalam hal ini, teknologi menjadi alat yang sangat penting dalam meningkatkan efektivitas manajemen SDM.

Penelitian berbasis data memberikan pemahaman mendalam tentang dinamika turnover karyawan, memungkinkan perusahaan membuat keputusan yang lebih tepat dan berbasis bukti. Pemanfaatan data historis, perusahaan dapat lebih proaktif dalam mengelola sumber daya manusia dan merancang kebijakan yang lebih tepat sasaran [3]. Teknologi memainkan peran penting dalam meningkatkan efektivitas manajemen SDM.

Studi terbaru menunjukkan bahwa algoritma pembelajaran mesin, seperti AdaBoost, SVM, dan RandomForest, dapat digunakan untuk memprediksi tingkat turnover karyawan dengan akurasi tinggi, membantu perusahaan dalam analisis prediktif [13]. Selain itu, integrasi teknologi seperti Kecerdasan Buatan (AI), Realitas Virtual (VR), Realitas Tertambah (AR), dan Metaverse dalam manajemen SDM, yang dapat meningkatkan efisiensi dan efektivitas proses HR. Perbandingan dengan studi sebelumnya menunjukkan bahwa sementara penelitian terdahulu lebih fokus pada faktor-faktor tradisional yang mempengaruhi turnover, seperti kepuasan kerja dan beban kerja, penelitian terkini menekankan pentingnya analisis data dan teknologi canggih dalam memprediksi dan mengelola turnover karyawan. Hal ini menunjukkan pergeseran menuju pendekatan yang lebih proaktif dan berbasis data dalam manajemen SDM. Dengan demikian, integrasi teknologi dan analisis data dalam manajemen SDM tidak hanya memungkinkan perusahaan untuk memprediksi turnover karyawan dengan lebih akurat tetapi juga merancang strategi yang lebih efektif untuk meningkatkan retensi karyawan.

Secara keseluruhan, manajemen SDM yang efektif tidak hanya bergantung pada seleksi dan perekrutan, tetapi juga pada pengelolaan karyawan yang ada, pengurangan turnover, dan penerapan teknologi untuk merancang kebijakan berbasis data. Dengan demikian, perusahaan dapat lebih maksimal dalam memanfaatkan potensi sumber daya manusia untuk mencapai tujuan jangka panjang dan mempertahankan daya saing di pasar yang semakin kompetitif. Seleksi yang tepat dan pengelolaan karyawan yang efektif akan menghasilkan sinergi yang mendukung tercapainya tujuan organisasi [14]. Pendekatan holistik dalam manajemen SDM sangat diperlukan untuk menjaga kelangsungan dan kesuksesan organisasi dalam jangka panjang

III. METODE

Penelitian ini menggunakan beberapa tahapan dimulai dari pengumpulan data, preprocessing data, pemilihan fitur, penyeimbangan data, pelatihan dan evaluasi model menggunakan confusion matrix [15]. berikut adalah gambaran alur penelitian yang ditampilkan di gambar 1.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Dalam penelitian ini, beberapa dataset yang akan digunakan sebagai bahan rujukan untuk menganalisis dan memprediksi tingkat Turnover karyawan diperoleh dari berbagai sumber. Salah satu dataset utama yang digunakan adalah dataset survei Turnover karyawan IBM Human Resource, yang merupakan salah satu dataset yang sering dipakai dalam penelitian terkait manajemen sumber daya manusia dan prediksi Turnover karyawan [16]. Dataset ini berisi informasi mengenai berbagai aspek terkait karyawan, seperti demografi, status pekerjaan, tingkat kepuasan, dan faktor-faktor lain yang dapat memengaruhi keputusan seorang karyawan untuk bertahan atau keluar dari perusahaan. Data ini sangat relevan karena dapat memberikan wawasan yang lebih dalam tentang pola-pola yang terjadi dalam keputusan karyawan.

Dataset IBM HR tersebut telah digunakan dalam berbagai penelitian sebelumnya, seperti yang diteliti oleh Gunawan dan rekan-rekannya pada tahun 2024 [17]. Dalam penelitian tersebut, dataset ini digunakan untuk mengeksplorasi faktor-faktor yang mempengaruhi tingkat Turnover karyawan di perusahaan. Menganalisis hubungan antara variabel-variabel seperti usia, jabatan, durasi bekerja, dan tingkat kepuasan kerja terhadap kemungkinan seorang karyawan untuk keluar dari perusahaan [18]. Dengan memanfaatkan teknik analisis data dan machine learning, hasil penelitian ini memberikan gambaran yang lebih jelas mengenai faktor-faktor yang dapat memprediksi Turnover karyawan, serta bagaimana perusahaan dapat mengelola sumber daya manusia dengan lebih efektif. Dataset ini, oleh karena itu, menjadi sumber data yang sangat penting dan relevan dalam upaya meningkatkan retensi karyawan.

B. Eksplorasi Data

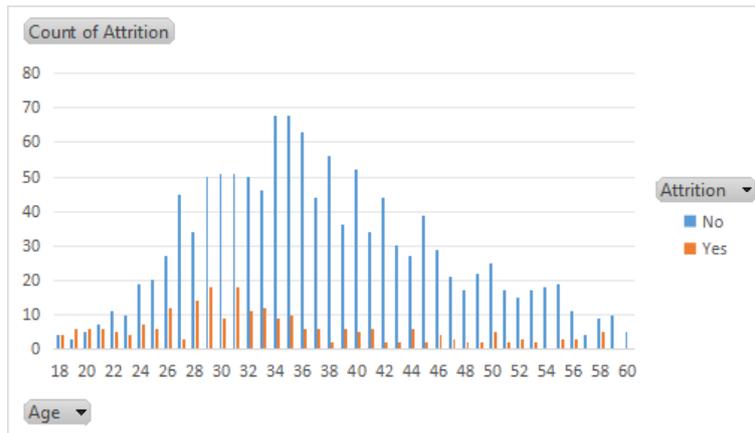
Dataset tersebut terdiri dari 1470 data contoh, 35 fitur atau variabel, dengan 60% jenis kelamin laki-laki, 40% jenis kelamin wanita. Tabel 1 menunjukkan 35 fitur beserta tipe data yang digunakan.

TABEL 1
 FITUR DAN TIPE DATA

| Fitur | Tipe Data | Fitur | Tipe Data |
|-------------------------|-----------|--------------------------|-----------|
| Age | Numerik | MonthlyIncome | Numerik |
| Attrition | Kategori | MonthlyRate | Numerik |
| BusinessTravel | Kategori | NumCompaniesWorked | Numerik |
| DailyRate | Numerik | Over18 | Kategori |
| Department | Kategori | OverTime | Kategori |
| DistanceFromHome | Numerik | PercentSalaryHike | Numerik |
| Education | Numerik | PerformanceRating | Numerik |
| EducationField | Kategori | RelationshipSatisfaction | Numerik |
| EmployeeCount | Numerik | StandardHours | Numerik |
| EmployeeNumber | Numerik | StockOptionLevel | Numerik |
| EnvironmentSatisfaction | Numerik | TotalWorkingYears | Numerik |
| Gender | Kategori | TrainingTimesLastYear | Numerik |
| HourlyRate | Numerik | WorkLifeBalance | Numerik |
| JobInvolvement | Numerik | YearsAtCompany | Numerik |
| JobLevel | Numerik | YearsInCurrentRole | Numerik |
| JobRole | Kategori | YearsSinceLastPromotion | Numerik |
| JobSatisfaction | Numerik | YearsWithCurrManager | Numerik |
| MaritalStatus | Kategori | | |

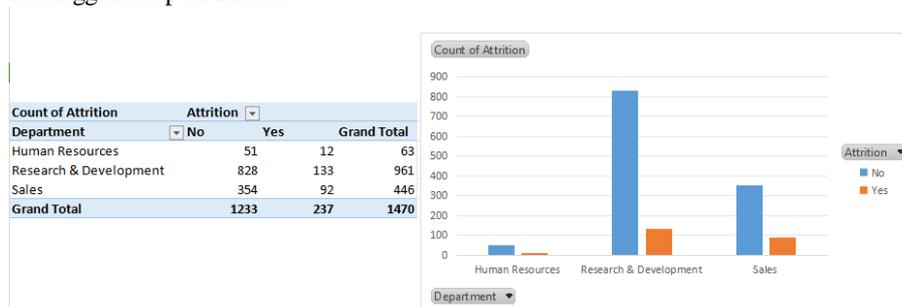
Gambar 2 menyajikan visualisasi data menggunakan pivot tabel untuk memberikan gambaran tingkat Turnover karyawan secara keseluruhan. Pivot tabel ini digunakan untuk menganalisis data secara mendalam, mencakup berbagai aspek yang memengaruhi Turnover, seperti rata-rata usia karyawan yang mengalami Turnover, serta distribusi Turnover berdasarkan departemen dan jabatan tertentu.

Distribusi turnover karyawan berdasarkan departemen dan jabatan. Data menunjukkan bahwa turnover lebih signifikan pada jabatan tertentu, seperti Sales Executive di departemen Sales. Departemen dengan tingkat turnover tinggi, seperti Sales, dapat menjadi fokus analisis untuk memahami penyebabnya, misalnya terkait beban kerja, kepuasan kerja, atau gaji. Selain itu, grafik batang menunjukkan perbandingan jumlah karyawan yang meninggalkan perusahaan (attrition "Yes") dengan yang tetap bertahan (attrition "No"), memberikan gambaran area yang memerlukan perhatian lebih



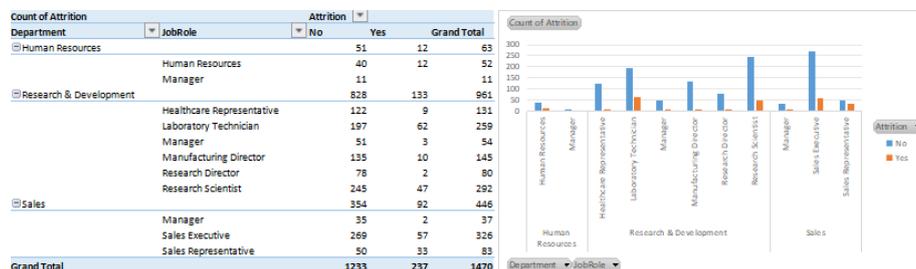
Gambar 2. Turnover berdasarkan Usia Karyawan

Distribusi turnover berdasarkan usia karyawan. Kelompok usia 30-40 tahun tampak memiliki tingkat turnover yang lebih tinggi dibandingkan kelompok usia lainnya. Hal ini menunjukkan bahwa usia karyawan menjadi salah satu faktor penting dalam memprediksi turnover. Pada usia ini, karyawan mungkin menghadapi tuntutan karier yang lebih besar, mencari peluang yang lebih baik, atau merasa stagnan di tempat kerja. Selain itu, usia juga sering dikaitkan dengan tanggung jawab keluarga dan tekanan finansial, yang dapat memengaruhi keputusan untuk bertahan atau meninggalkan perusahaan.



Gambar 3. Turnover Berdasarkan Riset and Development

Data pada gambar 3 menunjukkan bahwa departemen dengan Turnover tertinggi adalah Riset & Pengembangan (R&D), di mana jumlah pegawai yang keluar jauh lebih banyak dibandingkan departemen lainnya. Sebaliknya, departemen SDM memiliki tingkat Turnover yang jauh lebih rendah. Grafik batang yang menyertai data ini memperkuat temuan tersebut dengan memperlihatkan jumlah karyawan yang keluar (Yes) dan yang tetap bekerja (No) di setiap departemen. Proporsi Turnover yang lebih tinggi di R&D menunjukkan bahwa departemen ini memiliki tantangan khusus dalam mempertahankan karyawan, yang bisa menjadi fokus analisis lebih lanjut untuk memahami penyebab Turnover.



Gambar 4 Turnover berdasarkan Jabatan

Analisis lebih rinci tentang Turnover karyawan berdasarkan jabatan seperti pada gambar 4 di dalam masing-masing departemen. Data menunjukkan bahwa jabatan dengan Turnover tertinggi terdapat di departemen R&D, khususnya dalam peran seperti Peneliti dan Teknisi Laboratorium. Di departemen Penjualan, Turnover juga cukup signifikan di antara jabatan seperti Sales Executive. Grafik batang dalam gambar ini menampilkan perbandingan Turnover antarjabatan, yang memperlihatkan bahwa beberapa jabatan memiliki risiko Turnover yang jauh lebih tinggi dibandingkan yang lain. Informasi ini penting untuk mengidentifikasi peran yang memerlukan perhatian khusus dalam strategi retensi karyawan.

C. Data Preprocessing

Fitur yang redundan seperti Employee Count, Standard Hours, Over18 dan yang tidak relevan seperti Employee Number dihapus. Fungsi IS NULL digunakan untuk mengecek apakah suatu nilai dalam kolom adalah null atau tidak dan sum adalah jumlah nilai yang null, dan Fungsi Duplicate digunakan untuk mengecek apakah suatu nilai dalam fitur atau kolom ada duplikasi atau tidak.

```
df_copy['EducationField'] = df_copy['EducationField'].replace('Life Sciences',1)
df_copy['EducationField'] = df_copy['EducationField'].replace('Medical',2)
df_copy['EducationField'] = df_copy['EducationField'].replace('Marketing',3)
df_copy['EducationField'] = df_copy['EducationField'].replace('Technical Degree',4)
df_copy['EducationField'] = df_copy['EducationField'].replace('Human Resources',5)
df_copy['EducationField'] = df_copy['EducationField'].replace('Other',6)

df_copy['JobRole'] = df_copy['JobRole'].replace('Sales Executive',1)
df_copy['JobRole'] = df_copy['JobRole'].replace('Manufacturing Director',2)
df_copy['JobRole'] = df_copy['JobRole'].replace('Healthcare Representative',3)
df_copy['JobRole'] = df_copy['JobRole'].replace('Manager',4)
df_copy['JobRole'] = df_copy['JobRole'].replace('Research Director',5)
df_copy['JobRole'] = df_copy['JobRole'].replace('Laboratory Technician',6)
df_copy['JobRole'] = df_copy['JobRole'].replace('Sales Representative',7)
df_copy['JobRole'] = df_copy['JobRole'].replace('Research Scientist',8)
df_copy['JobRole'] = df_copy['JobRole'].replace('Human Resources',9)

df_copy['BusinessTravel'] = df_copy['BusinessTravel'].replace('Travel_Rarely',1)
df_copy['BusinessTravel'] = df_copy['BusinessTravel'].replace('Travel_Frequently',2)
df_copy['BusinessTravel'] = df_copy['BusinessTravel'].replace('Non-Travel',3)

df_copy['Attrition'] = df_copy['Attrition'].replace('Yes',1)
df_copy['Attrition'] = df_copy['Attrition'].replace('No',2)

df_copy['OverTime'] = df_copy['OverTime'].replace('Yes',1)
df_copy['OverTime'] = df_copy['OverTime'].replace('No',2)

df_copy['Gender'] = df_copy['Gender'].replace('Male',1)
df_copy['Gender'] = df_copy['Gender'].replace('Female',2)

df_copy['MaritalStatus'] = df_copy['MaritalStatus'].replace('Single',1)
df_copy['MaritalStatus'] = df_copy['MaritalStatus'].replace('Married',2)
df_copy['MaritalStatus'] = df_copy['MaritalStatus'].replace('Divorced',3)

df_copy['Department'] = df_copy['Department'].replace('Sales',1)
df_copy['Department'] = df_copy['Department'].replace('Human Resources',2)
df_copy['Department'] = df_copy['Department'].replace('Research & Development',3)
```

Gambar 5. Konversi data kategori menjadi numerik.

Kode tersebut menggunakan metode replace dari pandas untuk mengkonversikan data kategori menjadi numerik yang masing-masing memiliki nilai tertentu, sehingga setelah eksekusi kode tersebut nilai-nilai pada kolom dengan tipe data kategori akan diganti sesuai dengan mapping yang angka yang diberikan. Hasil setelah di konversi menjadi numerik sebagai ditampilkan gambar 6 berikut :

```
print(df_copy[['BusinessTravel', 'Attrition', 'OverTime', 'Gender', 'MaritalStatus', 'Department', 'EducationField', 'JobRole']])
```

| | BusinessTravel | Attrition | OverTime | Gender | MaritalStatus | Department | EducationField | JobRole |
|------|----------------|-----------|----------|--------|---------------|------------|----------------|---------|
| 0 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 |
| 1 | 2 | 2 | 2 | 1 | 2 | 3 | 1 | 8 |
| 2 | 1 | 1 | 1 | 1 | 1 | 3 | 6 | 6 |
| 3 | 2 | 2 | 1 | 2 | 2 | 3 | 1 | 8 |
| 4 | 1 | 2 | 2 | 1 | 2 | 3 | 2 | 6 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1465 | 2 | 2 | 2 | 1 | 2 | 3 | 2 | 6 |
| 1466 | 1 | 2 | 2 | 1 | 2 | 3 | 2 | 3 |
| 1467 | 1 | 2 | 1 | 1 | 2 | 3 | 1 | 2 |
| 1468 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 |
| 1469 | 1 | 2 | 2 | 1 | 2 | 3 | 2 | 6 |

[1470 rows x 8 columns]

Gambar 6. Hasil Konversi

Bedasarkan beberapa penelitian 15 dari 30 atribut teratas telah dipilih berdasarkan beberapa metode seleksi atribut, yaitu Correlation Attribute, Gain Ratio Attribute, dan Symmetrical Uncertainty Attributes. 15 atribut yang dipilih untuk fase pemodelan meliputi Attrition, Stock Option Level, Job Level, Marital Status, Years At Company, Monthly Income, Years With Curr Manager, Total Working Years, Business Travel, Age, Years In Current Role, Job Role, Job Involvement, Environment Satisfaction, Work Life Balance

Teknik yang digunakan untuk mengatasi masalah data yang tidak seimbang. SMOTE membuat data sintetis baru yang mirip dengan data asli dari kelas minoritas, sehingga jumlah data dari kelas minoritas dan mayoritas menjadi seimbang.

D. Pemilihan Model

Support Vector Machine (SVM) memang merupakan teknik yang digunakan untuk klasifikasi. Teknik ini membangun hyperplane untuk memisahkan dua kelas dan bertujuan untuk mencari batas hyperplane terbaik dan jarak antar kelas. Titik data ini direpresentasikan sebagai vektor dalam ruang berdimensi tinggi, dan SVM menemukan hyperplane yang paling baik memisahkannya. [19].

Random forest adalah klasifikasi ensemble yang menggabungkan beberapa algoritma seperti Decision Tree, Support Vector Machines, dan Naive Bayes untuk membuat keputusan kolektif. Algoritma Random Forest telah menunjukkan akurasi dan presisi tinggi dibandingkan dengan algoritma lain seperti Gradient Boosting dan Decision Tree

Random Forest membangun ensemble pohon keputusan dari himpunan bagian acak dari dataset training, memanfaatkan konsensus mayoritas di antara pohon-pohon ini untuk klasifikasi. Manju Nanda, etc [19].

E. Evaluasi Data

Evaluasi yang digunakan adalah confusion matrix dimana nilai-nilai yang dihasilkan oleh confusion matrix:

1. True Positives (TP): Jumlah kasus di mana model dengan benar memprediksi kelas positif.
2. True Negatives (TN): Jumlah kasus di mana model dengan benar memprediksi kelas negatif.
3. False Positives (FP): Jumlah kasus di mana model salah memprediksi kelas positif (juga dikenal sebagai "Type I Error").
4. False Negatives (FN): Jumlah kasus di mana model salah memprediksi kelas negatif (juga dikenal sebagai "Type II Error").

| | | Actual | |
|-----------|----------|----------|----------|
| | | Positive | Negative |
| Predicted | Positive | TP | FP |
| | Negative | FN | TN |

TP = True Positif
TN = True Negative
FP = False Positive
FN = False Negative

Gambar 7 Confusion Matrix

Secara ideal, untuk masing-masing kelas dalam klasifikasi biner atau multi-kelas, kita menginginkan nilai-nilai berikut:

- True Positives (TP): Tinggi
- True Negatives (TN): Tinggi
- False Positives (FP): Rendah
- False Negatives (FN): Rendah

Rumusan untuk Accuracy, Precision, Recall dan F1-Score adalah sebagai berikut:

- Accuracy: nilai idealnya mendekati 1 atau 100%
- Precision: nilai idealnya mendekati 1 atau 100%
- Recall: nilai idealnya mendekati 1 atau 100%
- F1-score : nilai idealnya mendekati 1 atau 100%

IV. HASIL

Penelitian ini membahas penerapan machine learning untuk menganalisis turnover karyawan dengan tujuan mengidentifikasi faktor-faktor yang memengaruhi keputusan karyawan untuk keluar dari perusahaan. Setelah melalui tahapan pengumpulan, eksplorasi, dan pemrosesan data, langkah berikutnya adalah melakukan pemilihan model yang dilakukan dengan serangkaian pengujian model machine learning pada dataset dengan menerapkan

algoritma SVM dan Random Forest. Untuk mengevaluasi akurasi model, dataset dibagi menjadi dua bagian, dengan 80% digunakan untuk melatih model (train) dan 20% sisanya untuk menguji model (test).

Support Vector Machine (SVM)

```

accuracy 0.8877551020408163
      precision    recall  f1-score   support

   0.0         1.00      0.15      0.27         39
   1.0         0.89      1.00      0.94        255

 accuracy
macro avg      0.94      0.58      0.60        294
weighted avg   0.90      0.89      0.85        294
    
```

Gambar 8. Hasil SVM sebelum menggunakan SMOTE dan Feature Selection

Akurasi adalah rasio dari jumlah prediksi yang benar (True Positives + True Negatives) terhadap total jumlah sampel. Gambar 8 menunjukkan nilai akurasi 0.8877, atau 88.77%, yang menunjukkan bahwa sekitar 88.77% dari semua prediksi yang dilakukan oleh model tanpa menggunakan SMOTE dan Feature Selection adalah benar.

```

accuracy 0.9251700680272109
      precision    recall  f1-score   support

   0.0         0.72      0.72      0.72         39
   1.0         0.96      0.96      0.96        255

 accuracy
macro avg      0.84      0.84      0.84        294
weighted avg   0.93      0.93      0.93        294
    
```

Gambar 9. Hasil SVM sebelum feature selection dan sesudah menggunakan SMOTE

Gambar 9 menunjukkan nilai akurasi 0.9251, atau 92.51%, yang menunjukkan bahwa sekitar 92.51% dari semua prediksi yang dilakukan oleh model menggunakan SMOTE dan tanpa Feature Selection adalah benar, terlihat peningkatan akurasi, precision, recall jika menggunakan SMOTE

```

accuracy 0.8503401360544217
      precision    recall  f1-score   support

   0.0         0.45      0.62      0.52         39
   1.0         0.94      0.89      0.91        255

 accuracy
macro avg      0.70      0.75      0.72        294
weighted avg   0.87      0.85      0.86        294
    
```

Gambar 10. Hasil sesudah menggunakan feature selection dan SMOTE

Gambar 10 menunjukkan nilai akurasi 0.8503, atau 85.03%, yang menunjukkan bahwa sekitar 85.03% dari semua prediksi yang dilakukan oleh model menggunakan SMOTE dan Feature Selection adalah benar, terlihat penurunan yang signifikan pada akurasi.

Random Forest

```

accuracy 0.8809523809523809
      precision    recall  f1-score   support

   0.0         1.00     0.10     0.19         39
   1.0         0.88     1.00     0.94        255

 accuracy
macro avg      0.94     0.55     0.56        294
weighted avg   0.90     0.88     0.84        294
    
```

Gambar 11. Hasil random forest sebelum feature selection, SMOTE

Gambar 11. menunjukkan nilai akurasi 0.8809, atau 88.09%, yang menunjukkan bahwa sekitar 88.09% dari semua prediksi yang dilakukan oleh model menggunakan tanpa SMOTE dan Feature Selection adalah benar.

```

accuracy 0.9693877551020408
      precision    recall  f1-score   support

   0.0         0.94     0.82     0.88         39
   1.0         0.97     0.99     0.98        255

 accuracy
macro avg      0.96     0.91     0.93        294
weighted avg   0.97     0.97     0.97        294
    
```

Gambar 12.. Hasil algoritma random forest sebelum feature selection dan sesudah menggunakan SMOTE

Gambar 12 menunjukkan nilai akurasi 0.9693, atau 96.93%, yang menunjukkan bahwa sekitar 96.93% dari semua prediksi yang dilakukan oleh model menggunakan SMOTE dan tanpa Feature Selection adalah benar, terlihat peningkatan akurasi,precision,recall jika menggunakan SMOTE

```

accuracy 0.9761904761904762
      precision    recall  f1-score   support

   0.0         0.97     0.85     0.90         39
   1.0         0.98     1.00     0.99        255

 accuracy
macro avg      0.97     0.92     0.95        294
weighted avg   0.98     0.98     0.98        294
    
```

Gambar 13. Hasil random forest sesudah feature selection dan SMOTE

Gambar 13 menunjukkan nilai akurasi 0.9761, atau 97.61%, yang menunjukkan bahwa sekitar 92.51% dari semua prediksi yang dilakukan oleh model menggunakan SMOTE dan Feature Selection adalah benar, hasil penelitian ini diharapkan dapat membantu organisasi dalam mengambil keputusan strategis terkait retensi karyawan dengan mengidentifikasi karakteristik yang berkontribusi terhadap turnover. Dengan demikian, pendekatan berbasis machine learning dapat memberikan solusi yang lebih cerdas dan berbasis data dalam mengelola sumber daya manusia.

Berdasarkan Tabel 2, hasil penelitian menunjukkan bahwa penerapan fitur selection dengan menggunakan 15 atribut teratas secara signifikan mengurangi waktu prediksi pada model SVM, dari 0,0447 detik menjadi 0,0210 detik, meskipun berdampak pada penurunan akurasi dari 92,51% menjadi 85,03%. Sementara itu, pada model Random Forest, perubahan yang terjadi relatif kecil, dengan waktu prediksi berkurang dari 0,0138 detik menjadi 0,0104 detik dan akurasi hanya sedikit menurun dari 97,72% menjadi 97,61%.

TABEL 2
 PEMILIHAN FITUR

| Model Klasifikasi | Sebelum fitur selection (30 attributtes) | | | |
|-------------------|--|--------|------|------------------|
| | Accuracy (%) | RMSe | ROC | Speed Prediction |
| SVM | 92.51 | 0.2735 | 0.84 | 0.0447 |
| Random Forest | 97.61 | 0.1543 | 0.92 | 0.0138 |
| Model Klasifikasi | Sesudah fitur selection (15 attributtes) | | | |
| | Accuracy (%) | RMSe | ROC | Speed Prediction |
| SVM | 85.03 | 0.2735 | 0.84 | 0.021 |
| Random Forest | 97.61 | 0.1543 | 0.92 | 0.0104 |

Selain itu, algoritma juga diuji dengan menggunakan metode SMOTE dan tanpa menggunakan metode SMOTE untuk membandingkan kinerjanya. Hasil uji akurasi algoritma ditampilkan dalam Tabel 3.

Tabel 3
 SMOTE

| Model Klasifikasi | Sebelum Menggunakan SMOTE | | | |
|-------------------|---------------------------|---------------|------------|--------------|
| | Accuracy (%) | Precision (%) | Recall (%) | f1-score (%) |
| SVM | 89 | 89 | 100 | 94 |
| Random Forest | 88 | 88 | 100 | 93 |
| Model Klasifikasi | Sesudah Menggunakan SMOTE | | | |
| | Accuracy (%) | Precision (%) | Recall (%) | f1-score (%) |
| SVM | 93 | 96 | 96 | 96 |
| Random Forest | 97 | 97 | 99 | 98 |

Setelah penerapan SMOTE, terjadi peningkatan kinerja pada kedua model. Untuk model SVM, akurasi meningkat dari 89% menjadi 93%, sementara precision, recall, dan f1-score masing-masing mencapai 96%. Pada model Random Forest, akurasi meningkat dari 88% menjadi 97%, dengan precision, recall, dan f1-score masing-masing mencapai 97%, 99%, dan 98%.

Perbandingan ini bertujuan untuk menilai apakah penggunaan fitur selection dan SMOTE mampu meningkatkan akurasi algoritma, atau apakah algoritma tetap memberikan hasil yang memadai tanpa pendekatan ini. Melalui analisis skor akurasi di Tabel 3, diperoleh wawasan yang lebih mendalam tentang dampak fitur selection dan SMOTE terhadap kinerja algoritma. Informasi ini dapat dijadikan dasar dalam menentukan apakah fitur selection atau SMOTE perlu diterapkan pada kasus tertentu.

V. PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis turnover karyawan dengan pendekatan machine learning, membandingkan performa algoritma Random Forest dan Support Vector Machine (SVM) dalam memprediksi kecenderungan karyawan untuk keluar dari perusahaan. Berdasarkan hasil pengujian, algoritma Random Forest menunjukkan akurasi sebesar 97,72%, sementara SVM mencatat akurasi 92,51%. Perbedaan ini mencerminkan keunggulan masing-masing algoritma dalam mengatasi karakteristik data yang digunakan dalam analisis turnover.

Random Forest, sebagai model ensemble, terdiri dari beberapa decision tree yang dilatih secara independen menggunakan teknik bagging (Bootstrap Aggregating). Metode ini efektif dalam mengurangi risiko overfitting yang sering ditemukan pada decision tree tunggal, menjadikannya pilihan yang andal untuk dataset dengan pola yang kompleks. Akurasi yang tinggi menunjukkan bahwa Random Forest mampu mengenali pola dan hubungan antarvariabel dengan baik, yang sangat penting dalam memahami faktor-faktor yang memengaruhi turnover karyawan. Selain itu, Random Forest menyediakan interpretasi yang jelas tentang pentingnya fitur, membantu organisasi mengidentifikasi variabel utama yang berkontribusi terhadap keputusan karyawan untuk keluar.

Sebaliknya, SVM menggunakan pendekatan yang berbeda dengan memisahkan kelas-kelas dalam data melalui hiperplane yang memaksimalkan margin. Algoritma ini unggul dalam menangani data berdimensi tinggi dan bekerja dengan baik pada dataset yang memiliki struktur yang jelas. Namun, akurasi 92,51% yang diperoleh menunjukkan bahwa SVM mungkin kurang optimal dalam menangani dataset yang memiliki noise atau pola yang sulit dipisahkan secara linear, seperti yang ditemukan dalam analisis turnover ini. SVM juga bergantung pada pemilihan kernel yang tepat, yang dapat menambah kompleksitas dalam proses tuning model.

Keunggulan Random Forest terlihat dalam kemampuannya menangani data dengan banyak fitur dan variabel yang tidak terstruktur. Selain itu, model ini lebih tangguh terhadap outlier dan variasi data yang besar. Sementara itu, SVM menawarkan kecepatan pelatihan yang lebih baik untuk dataset kecil dan cenderung lebih efisien dalam skenario dengan data yang linier atau berdimensi tinggi.

Dalam konteks analisis turnover karyawan, pemilihan algoritma yang tepat sangat bergantung pada karakteristik dataset dan tujuan penelitian. Jika dataset memiliki pola kompleks dan jumlah data yang besar, Random Forest terbukti lebih efektif. Namun, untuk dataset yang lebih kecil dan terstruktur, SVM tetap menjadi alternatif yang layak dengan efisiensi waktu pelatihan yang lebih tinggi.

Hasil penelitian ini memberikan panduan yang jelas bagi organisasi dalam memilih metode machine learning yang paling sesuai untuk memahami dan memprediksi turnover karyawan. Dengan memanfaatkan algoritma yang tepat, perusahaan dapat mengembangkan strategi retensi yang lebih akurat dan berbasis data, sehingga mengurangi risiko kehilangan talenta dan meningkatkan stabilitas tenaga kerja.

VI. KESIMPULAN

Penelitian ini berfokus pada penggunaan pendekatan machine learning untuk memprediksi tingkat Turnover karyawan dalam suatu organisasi. Turnover karyawan adalah salah satu isu utama dalam manajemen sumber daya manusia yang dapat mempengaruhi kinerja perusahaan secara keseluruhan. Tingkat Turnover yang tinggi sering kali menandakan adanya masalah dalam lingkungan kerja, budaya perusahaan, atau sistem penghargaan yang diterapkan. Oleh karena itu, prediksi tingkat Turnover yang akurat dapat membantu perusahaan dalam mengidentifikasi potensi masalah lebih dini, sehingga langkah-langkah preventif dapat diambil untuk mempertahankan karyawan yang bernilai.

Dalam penelitian ini, dua model klasifikasi digunakan untuk memprediksi tingkat Turnover, yaitu Support Vector Machine (SVM) dan Random Forest. SVM adalah salah satu teknik yang sering digunakan dalam klasifikasi data dengan margin pemisah yang jelas antara kelas-kelas yang ada. Sementara itu, Random Forest adalah teknik ensemble yang menggabungkan banyak pohon keputusan untuk menghasilkan prediksi yang lebih stabil dan akurat. Kedua model ini dipilih karena kemampuannya dalam menangani masalah klasifikasi yang kompleks dan ketepatannya dalam memprediksi hasil berdasarkan pola yang ditemukan dalam data.

Hasil evaluasi dari kedua model menunjukkan perbedaan kinerja yang signifikan. Model Random Forest berhasil mencapai tingkat akurasi yang lebih tinggi, yaitu 97,72%, dibandingkan dengan SVM yang hanya mencapai 92,51%. Selain itu, Random Forest juga menawarkan presisi yang lebih unggul pada kedua kelas, yaitu kelas karyawan yang bertahan dan kelas karyawan yang keluar. Hal ini menunjukkan bahwa Random Forest memiliki keunggulan dalam menangani masalah ketidakseimbangan kelas yang sering terjadi dalam dataset Turnover karyawan. Salah satu aspek penting yang memengaruhi hasil dari model machine learning adalah tahap pra-proses data. Pra-proses data mencakup serangkaian langkah seperti pembersihan data, normalisasi, dan transformasi fitur, yang semuanya berperan penting dalam meningkatkan kualitas data yang digunakan untuk pelatihan model. Meskipun tahap ini memakan waktu dan usaha yang signifikan, langkah-langkah pra-proses yang tepat dapat meningkatkan kemampuan model dalam memprediksi dengan akurasi yang lebih tinggi. Oleh karena itu, peneliti harus memastikan bahwa data yang digunakan benar-benar bersih dan siap untuk diproses lebih lanjut. Tingkat Turnover karyawan. Dengan hasil yang lebih akurat, perusahaan dapat melakukan intervensi lebih dini untuk mengurangi tingkat Turnover dan meningkatkan retensi karyawan. Penerapan machine learning ini dapat memberikan wawasan berharga yang membantu perusahaan menciptakan lingkungan kerja yang lebih kondusif, meningkatkan kepuasan karyawan, serta mendukung tercapainya tujuan jangka panjang perusahaan.

DAFTAR PUSTAKA

- [1] Gary Dessler and B. Varrkey, "Human Resource Management by Gary Dessler 15th-ed," 2005, [Online]. Available: <http://www.pearsonmylabandmastering.com>
- [2] A. Wilkinson, N. Bacon, T. Redman, and S. Snell, *The SAGE handbook of human resource management*. 2010. doi: 10.4135/9780857021496.
- [3] S. Najafi-Zangeneh, N. Shams-Ghameh, A. Arjomandi-Nezhad, and S. H. Zolfani, "An improved machine learning-based employees attrition prediction framework with emphasis on feature selection," *Mathematics*, vol. 9, no. 11, 2021, doi: 10.3390/math9111226.
- [4] Dessler, *Human Resource Management : Personnel Human Resource Management*, vol. 6, no. January 2009. 2018. [Online]. Available: <https://open.umn.edu/opentextbooks/BookDetail.aspx?bookId=71>
- [5] N. Edition, *Organizational Behavior*. 2013. doi: 10.4324/9780203763209.
- [6] T. Konten, "73% Karyawan Tidak Puas dengan Pekerjaan Mereka," *Jobstreet*. Accessed: Dec. 02, 2024. [Online]. Available: <https://id.jobstreet.com/id/career-advice/article/73-karyawan-tidak-puas-dengan-pekerjaan-mereka>
- [7] M. Dardiri and M. Sulhan, "Pengaruh Fasilitas Kerja, Lingkungan kerja Dan Kepuasan Kerja Terhadap Turnover Intention pada Karyawan Produksi PT Intim Harmonis Food Sukorejo," *J. Manuhara Pus. Penelit. Ilmu Manaj. dan Bisnis*, vol. 1, no. 4, pp. 198–204, 2023, doi: 10.61132/manuhara.v1i4.218.
- [8] A. Nurfahrani and H. Armaniah, "Pengaruh Kompensasi dan Gaya Kepemimpinan Terhadap Turnover Intention Karyawan," *J. Ilm. Multidisiplin*, vol. 1, no. 7, pp. 2986–6340, 2023, [Online]. Available:

- <https://doi.org/10.5281/zenodo.8263456>
- [9] J. Yarker, R. Lewis, E. Donaldson-Feilder, and P. Flaxman, *Management competencies for preventing and reducing stress at work*. London: HSE, 2007.
 - [10] N. M. Thi Nong, N. Q. Phuong, and H. Duc-Son, "The effect of employee competence and competence – job – fit on business performance through moderating role of social exchange: A study in logistics firms," *Asian J. Shipp. Logist.*, vol. 40, no. 4, pp. 40–42, 2024, doi: 10.1016/j.ajsl.2024.10.001.
 - [11] B. Raharjo, *Deep Learning dengan Python*. 2022.
 - [12] M. Karimi, "Employee Turnover Analysis Using Machine Learning Algorithms".
 - [13] R. Punnoose, "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms," vol. 5, no. 9, pp. 22–26, 2016.
 - [14] B. A. D. V and H. Rajeev, "Employee Attrition Prediction," vol. 9246, no. 1, pp. 26–30, 2024, doi: 10.54105/ijdm.A1636.04010524.
 - [15] C. Series, "Confusion Matrices and Rough Set Data Analysis Confusion Matrices and Rough Set Data Analysis," 2019, doi: 10.1088/1742-6596/1229/1/012055.
 - [16] Y. Zhang, Z. Cai, and H. Fei, *Predicting Employee Turnover in High-Tech Enterprises Using Machine Learning : Based on the Psychological Contract Perspective*, no. Cdems. Atlantis Press International BV, 2024. doi: 10.2991/978-94-6463-488-4.
 - [17] M. A. Commitment, "JBTI : Jurnal Bisnis : Teori dan Implementasi Modeling Affective Commitment and Job Satisfaction as A Mediating Role in Predicting Turnover Intention among Gen Z," vol. 15, no. 2, pp. 143–160, 2024.
 - [18] F. Idayati *et al.*, *Pengembangan Sumber Daya Manusia Berkelanjutan*. 2024
 - [19] E. Vankov and K. Gadhomi, "Supervised Learning," *J. AI & Automation*, 2024, doi: 10.1201/9781003254515-12