

KLASIFIKASI KARYAWAN BERDASARKAN KESEHATAN MENTAL DAN DAMPAKNYA TERHADAP TURNOVER MENGGUNAKAN EXPLORATORY DATA ANALYSIS DAN MACHINE LEARNING

Fadhil Muhammad Basysyar, Asep Yusapra Salim, Divi Adiffia Freza Alana,

Rizqy Amalia Putri, Elliarini Isabela, Mardiman

Sistem Informasi, Universitas Kartamulia Purwakarta

Sukatani, Purwakarta, Jawa Barat, Indonesia

fadhil.m.basysyar@gmail.com

ABSTRAK

Turnover karyawan merupakan tantangan utama dalam manajemen sumber daya manusia karena berdampak langsung pada stabilitas organisasi, biaya rekrutmen, serta keberlanjutan produktivitas tenaga kerja. Penelitian ini bertujuan untuk menganalisis hubungan antara faktor kesehatan mental dan kecenderungan *turnover* serta mengembangkan model klasifikasi berbasis *machine learning* untuk memprediksi risiko berhenti kerja secara lebih akurat. Dataset yang digunakan berasal dari survei terhadap 1.251 responden yang terdiri dari pekerja dan wirausaha di Kabupaten Purwakarta, yang merepresentasikan kondisi psikososial tenaga kerja pada sektor jasa dan teknologi. Analisis diawali dengan *Exploratory Data Analysis* (EDA) untuk memahami distribusi dan hubungan antarvariabel. Ketidakseimbangan kelas pada variabel target diatasi menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) guna meningkatkan kualitas pembelajaran model. Empat algoritma *machine learning* diuji, yaitu *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *Support Vector Classifier* (SVC). Hasil evaluasi menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan akurasi 0.7102, precision 0.725, recall 0.7117, dan F1-score 0.7183. Temuan ini menunjukkan bahwa tekanan pekerjaan, kelelahan, dan kualitas hubungan *interpersonal* merupakan prediktor utama terbentuknya *turnover intention*. Penelitian ini berkontribusi pada pengembangan analitik sumber daya manusia berbasis data dengan memanfaatkan pendekatan *machine learning* untuk mengidentifikasi risiko *turnover* secara lebih dini, serta memberikan dasar bagi organisasi dalam merancang strategi retensi karyawan berbasis kesejahteraan mental.

Kata kunci: Klasifikasi, Kesehatan Mental, Turnover, Machine Learning

1. PENDAHULUAN

Kesehatan mental karyawan merupakan isu krusial dalam manajemen sumber daya manusia karena memengaruhi produktivitas, keterlibatan, dan retensi tenaga kerja. Dalam era pasca-pandemi dan digitalisasi, tekanan psikologis berdampak langsung terhadap *turnover* yang menimbulkan konsekuensi finansial dan sosial bagi organisasi [1]. Perkembangan *machine learning* memungkinkan analisis yang lebih komprehensif dalam memahami hubungan antara kesehatan mental dan perilaku *turnover* melalui berbagai algoritma seperti *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *Support Vector Classifier* (SVC) [2], [3].

Kesehatan mental dalam konteks pekerjaan tidak hanya berhubungan dengan kesejahteraan individu tetapi juga dengan efektivitas organisasi secara keseluruhan. Bahwa karyawan dengan kondisi psikologis yang stabil memiliki tingkat produktivitas dan loyalitas yang lebih tinggi [4]. Namun, tekanan pekerjaan, ketidakpastian karier, dan kurangnya dukungan organisasi dapat memicu stres kronis yang meningkatkan risiko *turnover*. Penelitian terdahulu menggarisbawahi pentingnya pendekatan berbasis data untuk mengidentifikasi pola kesehatan mental dan hubungannya dengan *turnover* [5]. Oleh karena itu, latar belakang penelitian ini berfokus pada kebutuhan akan model analitik prediktif berbasis *machine learning* guna mengoptimalkan strategi retensi

karyawan melalui pemahaman mendalam atas faktor psikologis.

Pentingnya kesehatan mental di tempat kerja, masih terdapat kesenjangan dalam mengintegrasikan pendekatan kuantitatif berbasis *machine learning* untuk menganalisis hubungan antara kesehatan mental dan *turnover*. Studi sebelumnya menggunakan metode survei atau regresi linier tradisional yang memiliki keterbatasan dalam menangkap kompleksitas pola data multidimensional [6]. Kurangnya model klasifikasi yang akurat menyebabkan organisasi kesulitan dalam mengidentifikasi kelompok karyawan dengan risiko *turnover* tinggi.

Penelitian ini mengajukan sejumlah masalah utama untuk menjawab kesenjangan empiris tersebut.

- Bagaimana hubungan antara variabel kesehatan mental dan niat *turnover* karyawan?
- Sejauh mana algoritma *Logistic Regression*, *Decision Tree*, *Random Forest*, dan SVC mampu mengklasifikasikan karyawan berdasarkan kondisi kesehatan mental mereka?
- Algoritma mana yang memberikan kinerja prediksi terbaik dalam konteks ini?

Secara spesifik, penelitian ini bertujuan:

- Menganalisis hubungan antara kesehatan mental dan *turnover*;
- Membangun model klasifikasi menggunakan *Logistic Regression*, *Decision Tree*, *Random Forest*, dan SVC;

- c. Mengevaluasi kinerja model dalam konteks organisasi modern.

Penelitian terdahulu menunjukkan efektivitas pendekatan *hibrid machine learning* untuk memprediksi perilaku tenaga kerja dan mendukung keputusan strategis HR [7]. Dengan demikian, penelitian ini berkontribusi pada pengembangan sistem analitik berbasis bukti untuk kebijakan retensi yang berkelanjutan.

Ruang lingkup penelitian ini mencakup analisis data karyawan dengan variabel utama berupa indikator kesehatan mental, perilaku kerja, kepuasan kerja, dan niat *turnover*. Penelitian difokuskan pada penerapan empat algoritma utama yaitu *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *SVC*, yang dibandingkan berdasarkan akurasi untuk menentukan model terbaik.

Penelitian ini memberikan kontribusi ilmiah dalam tiga dimensi utama: teoretis, metodologis, dan praktis. Secara teoretis, penelitian ini memperluas pemahaman tentang hubungan antara kesehatan mental dan *turnover* melalui pendekatan kuantitatif berbasis *machine learning* [8]. Secara metodologis, penelitian ini menghadirkan model klasifikasi baru yang memanfaatkan kombinasi algoritma untuk mengoptimalkan kinerja prediksi. Sementara secara praktis, penelitian ini menyediakan alat bantu analitik bagi organisasi dalam mengidentifikasi risiko *turnover* dini serta merancang strategi retensi berbasis kesejahteraan mental karyawan [9].

2. TINJAUAN PUSTAKA

Kajian literatur terdahulu menunjukkan bahwa penelitian terkait kesehatan mental dan *turnover intention* semakin berkembang dengan memanfaatkan pendekatan *machine learning* untuk meningkatkan akurasi prediksi perilaku karyawan. dalam studi berjudul *A Decade of Research on Machine Learning Techniques for Predicting Employee Turnover: A Systematic Literature Review* meneliti hubungan antara kesehatan mental dan niat *turnover*. Hasil penelitian menunjukkan bahwa dalam mengidentifikasi karyawan dengan risiko *turnover* tinggi berdasarkan indikator stres dan kepuasan kerja [10]. Penelitian ini memberikan kontribusi penting dalam membuktikan kemampuan *machine learning* untuk memetakan pola psikologis dan perilaku kerja secara lebih akurat dibandingkan metode statistik konvensional.

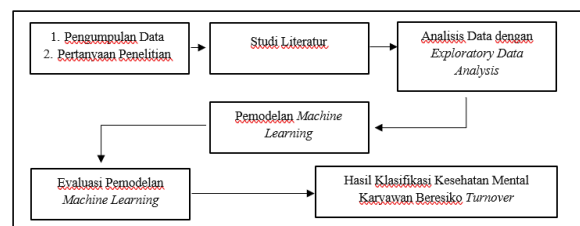
Sementara itu, dalam penelitian berjudul *Employee Well-Being and Turnover Intention* menekankan peran kesejahteraan psikologis sebagai determinan utama niat *turnover*. Penelitian ini menemukan bahwa kesejahteraan psikologis berpengaruh signifikan terhadap niat karyawan untuk keluar melalui variabel mediasi kepuasan kerja ($p < 0.05$). Studi tersebut memperkuat dasar teoretis bahwa kesehatan mental memiliki pengaruh tidak langsung terhadap *turnover intention* dan perlu dianalisis menggunakan pendekatan komputasional yang lebih

adaptif untuk memahami hubungan non-linear antar variabel [11].

Berdasarkan penelitian tersebut, penelitian ini berupaya mengintegrasikan pendekatan eksploratori data *Exploratory Data Analysis* dan perbandingan algoritma *Logistic Regression*, *Decision Tree*, *Random Forest*, serta *SVC* untuk mengidentifikasi model prediktif terbaik dalam mengklasifikasikan kesehatan mental dan mengantisipasi *turnover* karyawan secara lebih presisi.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain *explanatory research* yang bertujuan untuk mengklasifikasikan kondisi kesehatan mental karyawan serta menganalisis dampaknya terhadap potensi *turnover* menggunakan pendekatan *machine learning*. Proses penelitian diawali dengan pengumpulan data survei yang kemudian dianalisis melalui tahapan *Exploratory Data Analysis* (EDA) guna memperoleh pemahaman mendalam terhadap pola, distribusi, dan korelasi antar variabel [12]. Analisis EDA dilakukan untuk mendeteksi keberadaan outlier, menangani nilai hilang (*missing values*), serta menentukan relevansi setiap atribut terhadap variabel target (*turnover*). Setelah tahap eksplorasi, penelitian ini menerapkan empat algoritma pembelajaran mesin, yaitu *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *Support Vector Classifier* (SVC), untuk mengklasifikasikan kondisi kesehatan mental serta memprediksi potensi karyawan yang berisiko tinggi untuk berhenti bekerja. Evaluasi model dilakukan dengan menggunakan metrik akurasi, yang bertujuan untuk menilai kinerja masing-masing algoritma secara objektif. Validasi model dilaksanakan melalui teknik *k-fold cross validation* untuk memastikan reliabilitas hasil dan mencegah *overfitting* [13]. Tahap terakhir adalah menjelaskan hasil klasifikasi kesehatan mental karyawan berisiko berhenti dari pekerjaannya yang dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Dataset penelitian berasal dari survei terhadap 1.251 responden pekerja dan wirausaha di Kabupaten Purwakarta yang dipilih menggunakan *purposive sampling* pada sektor jasa dan teknologi. Data dikumpulkan melalui kuesioner terstruktur yang mengukur faktor demografis, kondisi kerja, dan aspek psikologis seperti tekanan pekerjaan, kelelahan, dukungan sosial, dan hubungan dengan atasan. Instrumen telah diuji validitas dan reliabilitasnya menggunakan *Cronbach's Alpha*.

Variabel penelitian meliputi faktor demografis dan psikososial, dengan variabel dependen berupa “klasifikasi berhenti” yang merepresentasikan turnover. Data diolah secara anonim sesuai prinsip etika penelitian.

Tahap praproses mencakup encoding (label dan one-hot), normalisasi min-max scaling, serta penyeimbangan data menggunakan SMOTE. Model klasifikasi yang digunakan adalah *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *Support Vector Classifier* (SVC). Evaluasi kinerja dilakukan menggunakan accuracy, precision, recall, dan *F1-score*, dengan validasi menggunakan *k-fold cross-validation* untuk meningkatkan keandalan dan mengurangi *overfitting*.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1251 entries, 0 to 1250
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Umur                  1251 non-null   int64
1   Jenis Kelamin         1251 non-null   object
2   Wiraswasta            1251 non-null   object
3   Perawatan Mental      1251 non-null   object
4   Tekanan Pekerjaan     1251 non-null   object
5   Pekerjaan Jarak Jauh  1251 non-null   object
6   Perusahaan Teknologi  1251 non-null   object
7   Fasilitas Karyawan    1251 non-null   object
8   Jaminan Kesehatan    1251 non-null   object
9   Program Kesejahteraan 1251 non-null   object
10  Layanan Konsultasi    1251 non-null   object
11  Dampak Psikologis     1251 non-null   object
12  Dampak Fisik          1251 non-null   object
13  Dampak Rekan Kerja    1251 non-null   object
14  Atasan Kerja          1251 non-null   object
15  Dampak Kelelahan     1251 non-null   object
16  Klasifikasi Berhenti  1251 non-null   object
dtypes: int64(1), object(16)
memory usage: 166.3+ KB
```

Gambar 2. Atribut Dataset

4. HASIL DAN PEMBAHASAN

Memuat hasil, Pengujian dan pembahasan tentang skripsi yang telah dilakukan.

4.1. Jenis Kelamin

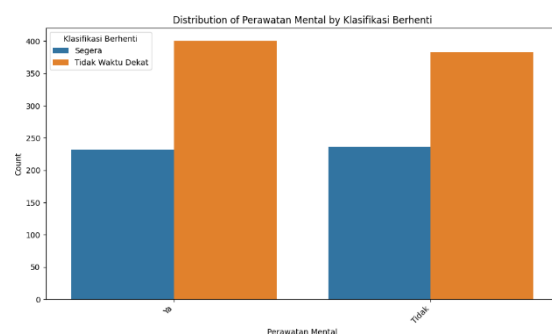
Distribusi jenis kelamin dalam dataset menunjukkan ketimpangan representasi responden, di mana 79,5% merupakan pria dan 20,5% wanita, sebagaimana ditunjukkan pada Gambar 3. Ketidakseimbangan ini mengindikasikan bahwa karakteristik tenaga kerja dalam sampel penelitian didominasi oleh responden pria. Dalam konteks analisis data, distribusi yang tidak merata tersebut berpotensi memengaruhi pola pembelajaran model karena perbedaan karakteristik psikososial antara kelompok gender dapat berkorelasi dengan variabel seperti tekanan pekerjaan, kelelahan, dan hubungan interpersonal di tempat kerja.

Jenis Kelamin	
Pria	995
Wanita	256

Gambar 3. Sebaran Jenis Kelamin

4.2. Melakukan Perawatan Mental terhadap Klasifikasi Berhenti

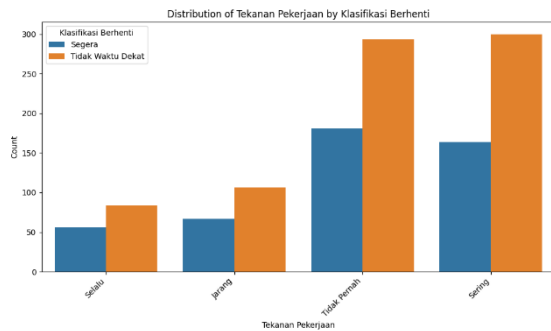
Distribusi variabel perawatan mental terhadap klasifikasi berhenti menunjukkan bahwa responden yang menerima maupun tidak menerima perawatan mental memiliki pola kecenderungan turnover yang relatif serupa, di mana sebagian besar responden berada pada kategori tidak berhenti dalam waktu dekat (Gambar 4). Pola ini mengindikasikan bahwa keberadaan perawatan mental tidak secara langsung berkorelasi kuat dengan keputusan berhenti dalam jangka pendek. Oleh karena itu, dalam proses pemodelan machine learning, variabel ini lebih relevan dianalisis sebagai bagian dari hubungan multidimensional antarvariabel psikososial dibandingkan sebagai determinan tunggal terhadap *turnover*.



Gambar 4. Perawatan Mental terhadap Klasifikasi Berhenti

4.3. Tekanan Pekerjaan terhadap Klasifikasi Berhenti

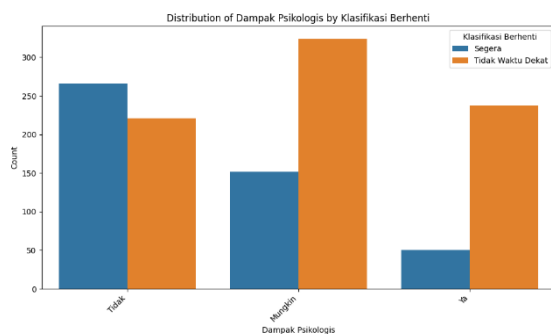
Distribusi tekanan pekerjaan terhadap klasifikasi berhenti menunjukkan bahwa peningkatan tingkat tekanan kerja berkorelasi dengan meningkatnya kecenderungan karyawan untuk berhenti dalam waktu dekat, sebagaimana terlihat pada Gambar 5. Responden dengan tingkat tekanan kerja yang lebih tinggi menunjukkan proporsi yang lebih besar dalam kategori berhenti segera dibandingkan dengan kelompok yang mengalami tekanan kerja lebih rendah. Pola ini mengindikasikan bahwa tekanan pekerjaan merupakan salah satu variabel yang berpotensi menjadi prediktor penting dalam model prediksi turnover. Oleh karena itu, variabel tekanan pekerjaan memiliki potensi kontribusi signifikan dalam proses pembentukan model klasifikasi berbasis *machine learning*.



Gambar 5. Tekanan Pekerjaan terhadap Klasifikasi Berhenti

4.4. Dampak Psikologis terhadap Klasifikasi Berhenti

Distribusi dampak psikologis terhadap klasifikasi berhenti menunjukkan pola yang tidak linear antar kategori responden (Gambar 6). Responden yang melaporkan tidak mengalami dampak psikologis menunjukkan proporsi yang relatif lebih tinggi pada kategori berhenti segera dibandingkan dengan kelompok lainnya, sementara responden pada kategori “Mungkin” dan “Iya” cenderung lebih banyak berada pada kategori tidak berhenti dalam waktu dekat. Pola ini mengindikasikan bahwa persepsi terhadap dampak psikologis tidak memiliki hubungan langsung yang konsisten dengan keputusan untuk berhenti bekerja. Oleh karena itu, dalam proses pemodelan machine learning, variabel ini lebih relevan dianalisis sebagai bagian dari hubungan multidimensional antarvariabel daripada sebagai determinan tunggal terhadap *turnover*.



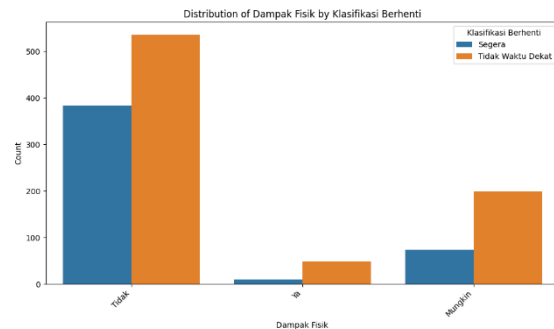
Gambar 6. Dampak Psikologis terhadap Klasifikasi Berhenti

4.5. Dampak Fisik terhadap Klasifikasi Berhenti

Distribusi dampak fisik berdasarkan klasifikasi berhenti menunjukkan bahwa sebagian besar responden yang berhenti segera maupun yang tidak berencana berhenti dalam waktu dekat berada pada kategori “Tidak mengalami dampak fisik”, dengan jumlah yang lebih besar pada kelompok yang tidak berhenti dalam waktu dekat. Pada kategori “Iya”, jumlah responden relatif sangat sedikit pada kedua kelompok, meskipun kelompok tidak berhenti dalam waktu dekat tetap lebih tinggi.

Sementara itu, pada kategori “Mungkin”, responden yang tidak berencana berhenti dalam waktu

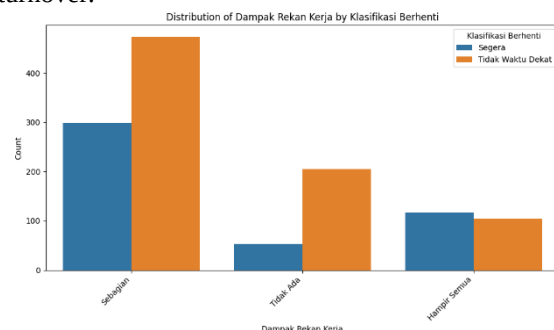
dekat kembali menunjukkan jumlah yang lebih besar dibandingkan mereka yang berhenti segera. Pola ini mengindikasikan bahwa dampak fisik bukan merupakan faktor dominan dalam keputusan untuk berhenti segera, karena mayoritas responden dari kedua kelompok justru tidak mengalami dampak fisik dan perbedaannya terutama terletak pada proporsi jumlah responden yang lebih besar pada kelompok yang tidak berencana berhenti dapat dilihat pada Gambar 7. Dampak Fisik terhadap Klasifikasi Berhenti.



Gambar 7. Dampak Fisik terhadap Klasifikasi Berhenti

4.6. Dampak Rekan Kerja terhadap Klasifikasi Berhenti

Distribusi variabel dampak rekan kerja terhadap klasifikasi berhenti menunjukkan bahwa dukungan sosial di tempat kerja memiliki hubungan dengan kecenderungan *turnover* (Gambar 8). Responden yang melaporkan adanya dukungan rekan kerja pada tingkat tertentu cenderung lebih banyak berada pada kategori tidak berhenti dalam waktu dekat dibandingkan dengan kategori berhenti segera. Pola ini mengindikasikan bahwa hubungan interpersonal di tempat kerja dapat berperan sebagai faktor yang berkaitan dengan stabilitas kerja. Oleh karena itu, variabel ini relevan untuk dipertimbangkan dalam pembentukan model machine learning guna meningkatkan kemampuan prediksi terhadap risiko *turnover*.

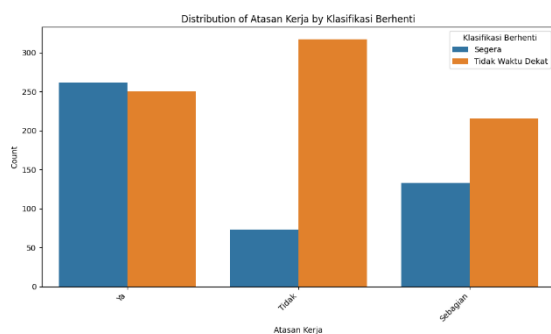


Gambar 8. Dampak Rekan Kerja terhadap Klasifikasi Berhenti

4.7. Dampak Atasan Kerja terhadap Klasifikasi Berhenti

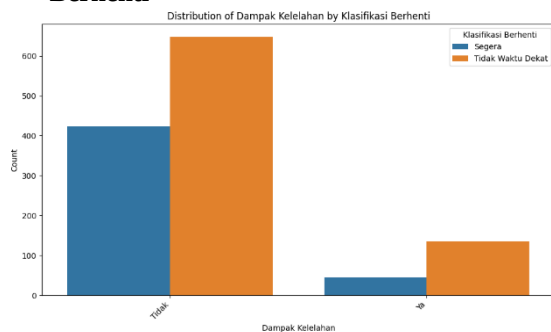
Distribusi variabel hubungan dengan atasan terhadap klasifikasi berhenti menunjukkan adanya

keterkaitan antara kualitas supervisi dan kecenderungan turnover (Gambar 9). Responden yang melaporkan adanya dukungan dari atasan cenderung lebih banyak berada pada kategori tidak berhenti dalam waktu dekat dibandingkan dengan kategori berhenti segera, meskipun perbedaannya tidak terlalu besar. Pola ini mengindikasikan bahwa kualitas hubungan dengan atasan berpotensi memengaruhi stabilitas kerja, namun pengaruhnya kemungkinan bergantung pada interaksi dengan variabel lain seperti tekanan pekerjaan, kelelahan, dan dukungan rekan kerja. Menunjukkan bahwa variabel hubungan dengan atasan dapat berperan sebagai salah satu prediktor dalam model klasifikasi turnover, terutama dalam menangkap dinamika psikososial yang memengaruhi keputusan karyawan untuk bertahan atau meninggalkan pekerjaan.



Gambar 9. Dampak Atasan Kerja terhadap Klasifikasi Berhenti

4.8. Dampak Kelelahan terhadap Klasifikasi Berhenti



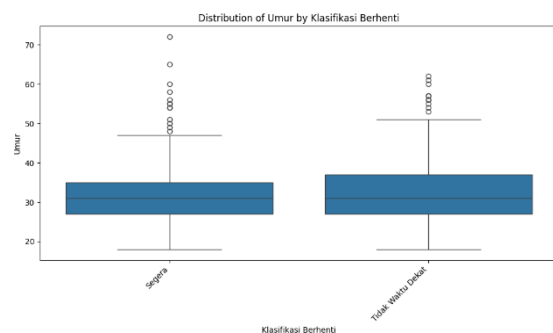
Gambar 10. Dampak Kelelahan terhadap Klasifikasi Berhenti

Distribusi variabel dampak kelelahan terhadap klasifikasi berhenti menunjukkan adanya hubungan antara tingkat kelelahan kerja dan kecenderungan turnover (Gambar 10). Responden yang melaporkan tidak mengalami kelelahan cenderung lebih banyak berada pada kategori tidak berhenti dalam waktu dekat, sementara responden yang mengalami kelelahan menunjukkan proporsi yang lebih besar pada kategori berhenti segera. Pola ini mengindikasikan bahwa kelelahan kerja berpotensi menjadi indikator penting yang berkaitan dengan keputusan karyawan untuk meninggalkan pekerjaan. Menunjukkan bahwa variabel kelelahan memiliki potensi kontribusi

signifikan dalam model klasifikasi turnover, terutama ketika dianalisis bersama variabel psikososial lain seperti tekanan pekerjaan, dukungan rekan kerja, dan hubungan dengan atasan.

4.9. Distribusi Umur Responder terhadap Klasifikasi Berhenti

Distribusi umur berdasarkan klasifikasi berhenti menunjukkan bahwa kedua kelompok responden memiliki rentang usia yang relatif serupa dengan median pada kisaran usia produktif (Gambar 11). Pola ini mengindikasikan bahwa usia tidak menunjukkan perbedaan distribusi yang signifikan antara karyawan yang berhenti segera dan yang tidak berencana berhenti dalam waktu dekat. Menunjukkan bahwa variabel umur kemungkinan memiliki pengaruh yang terbatas sebagai prediktor utama dalam model klasifikasi turnover, namun tetap relevan sebagai variabel demografis pendukung dalam analisis.



Gambar 11. Distribusi Umur Responder terhadap Klasifikasi Berhenti

4.10. Synthetic Minority Over-sampling Technique

Hasil penerapan Synthetic Minority Over-sampling Technique (SMOTE) menunjukkan bahwa dataset awal memiliki ketidakseimbangan kelas yang cukup signifikan. Pada kondisi awal, kelas berhenti (encoded = 1) berjumlah 783 data, sedangkan kelas tidak berhenti (encoded = 0) hanya 468 data. Ketidakseimbangan tersebut berpotensi menimbulkan bias pada model pembelajaran mesin, karena algoritma cenderung lebih mudah mempelajari pola dari kelas mayoritas dan mengabaikan kelas minoritas.

Setelah dilakukan proses resampling menggunakan SMOTE, distribusi kelas menjadi seimbang dengan masing-masing kelas memiliki jumlah data yang sama, yaitu 783 baris. Dengan demikian, total baris dalam dataset meningkat dari 1251 menjadi 1566. Proses ini dilakukan dengan membangkitkan sampel sintesis baru pada kelas minoritas melalui interpolasi antara titik-titik data yang berdekatan, bukan sekadar melakukan duplikasi data.

Penerapan SMOTE secara keseluruhan memberikan kontribusi penting dalam meningkatkan kinerja model pembelajaran mesin, terutama dalam konteks prediksi klasifikasi yang memiliki distribusi target tidak seimbang. Dengan mengurangi bias

terhadap kelas mayoritas, model menjadi lebih mampu mengenali karakteristik kelas minoritas dan meningkatkan generalisasi pada data baru. Visualisasi hasil SMOTE pada Gambar 12. *Synthetic Minority Over-sampling Technique* menunjukkan transformasi distribusi data sebelum dan sesudah oversampling, sekaligus menegaskan efektivitas teknik ini dalam menciptakan kondisi pembelajaran yang lebih stabil dan representatif.

```
Shape of original data: (1251, 22)
Shape of resampled data: (1566, 22)
Class distribution of original data:
Klasifikasi Berhenti_encoded
1    783
0    468
Name: count, dtype: int64
Class distribution of resampled data:
Klasifikasi Berhenti_encoded
0    783
1    783
Name: count, dtype: int64
```

Gambar 12. Synthetic Minority Over-sampling Technique

4.11. Evaluasi Pemodelan Algoritma

Hasil evaluasi model menunjukkan adanya perbedaan performa antar algoritma dalam memprediksi klasifikasi berhenti. *Random Forest* memperoleh akurasi tertinggi sebesar 0.7102, diikuti oleh *Support Vector Classifier* (SVC) dengan akurasi 0.7038, dan *Logistic Regression* dengan akurasi 0.6815. Sementara itu, *Decision Tree* menunjukkan performa terendah dengan akurasi sebesar 0.6242. Perbedaan performa ini menunjukkan bahwa kemampuan algoritma dalam menangkap pola hubungan antarvariabel sangat dipengaruhi oleh karakteristik model serta kompleksitas struktur data yang dianalisis.

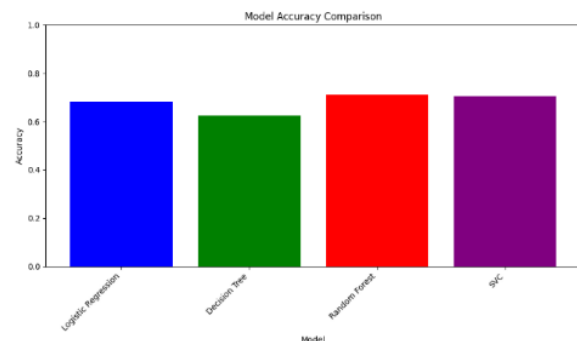
Keunggulan *Random Forest* dapat dijelaskan oleh pendekatan ensemble learning yang menggabungkan banyak pohon keputusan untuk meningkatkan stabilitas prediksi dan mengurangi variansi model. Pendekatan ini memungkinkan *Random Forest* menangkap hubungan non-linear serta interaksi kompleks antarvariabel psikososial, seperti tekanan pekerjaan, kelelahan, hubungan dengan atasan, dan dukungan rekan kerja. Selain itu, mekanisme pemilihan subset fitur secara acak pada setiap pohon juga membantu meningkatkan kemampuan generalisasi model terhadap data baru.

SVC menunjukkan performa yang relatif kompetitif dengan akurasi yang mendekati *Random Forest*. Hal ini menunjukkan bahwa pemisahan kelas dalam dataset memiliki struktur batas keputusan yang dapat ditangkap secara efektif melalui pembentukan hyperplane dalam ruang fitur berdimensi tinggi. Sementara itu, *Logistic Regression* masih mampu menghasilkan performa yang cukup baik meskipun berbasis model linear. Hal ini mengindikasikan bahwa sebagian hubungan antarvariabel dalam dataset masih

memiliki komponen linier yang dapat dijelaskan secara memadai oleh model tersebut.

Sebaliknya, performa *Decision Tree* yang lebih rendah menunjukkan keterbatasan model pohon tunggal dalam menghadapi kompleksitas pola data. Model ini cenderung sensitif terhadap variasi data dan lebih rentan mengalami overfitting dibandingkan metode ensemble. Oleh karena itu, hasil evaluasi menunjukkan bahwa pendekatan berbasis ensemble seperti *Random Forest* lebih efektif dalam meningkatkan stabilitas dan akurasi prediksi pada dataset yang memiliki hubungan non-linear dan distribusi variabel yang beragam.

Secara keseluruhan, hasil perbandingan ini menegaskan bahwa pemilihan algoritma memiliki pengaruh signifikan terhadap kualitas model prediksi turnover. Visualisasi pada Gambar 13 yang menampilkan perbandingan akurasi model.



Gambar 13. Perbandingan Akurasi Model

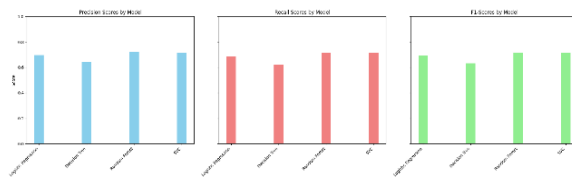
Berdasarkan hasil evaluasi model, *Random Forest* menunjukkan performa paling unggul dibandingkan algoritma lainnya pada seluruh metrik utama yang digunakan dalam penelitian ini. Model ini mencatat nilai *precision* tertinggi sebesar 0.725, yang menunjukkan kemampuan model dalam menghasilkan prediksi positif yang lebih akurat. Nilai *precision* yang tinggi mengindikasikan bahwa *Random Forest* mampu meminimalkan kesalahan klasifikasi positif palsu, sehingga lebih andal dalam mengidentifikasi karyawan yang benar-benar berpotensi berhenti. Dalam konteks manajemen sumber daya manusia, kemampuan ini penting karena kesalahan dalam mengklasifikasikan karyawan sebagai berisiko tinggi dapat menyebabkan keputusan intervensi yang tidak tepat.

Dari sisi *recall*, *Random Forest* dan *Support Vector Classifier* (SVC) menunjukkan performa yang sama dengan nilai 0.7117. Nilai ini mencerminkan kemampuan kedua model dalam mendeteksi sebagian besar kasus positif yang sebenarnya, yaitu individu yang memiliki kecenderungan untuk berhenti dalam waktu dekat. *Recall* yang tinggi menunjukkan bahwa model mampu mengurangi kemungkinan terlewatnya karyawan yang berisiko turnover. Hal ini menjadi aspek penting dalam sistem prediksi berbasis risiko, karena organisasi memerlukan model yang mampu mengidentifikasi potensi permasalahan secara dini

agar dapat merancang strategi retensi yang lebih efektif.

Pada metrik *F1-score*, *Random Forest* kembali menunjukkan performa terbaik dengan nilai 0.7183. *F1-score* merupakan ukuran harmonis yang menggabungkan *precision* dan *recall*, sehingga memberikan gambaran yang lebih seimbang mengenai kualitas model dalam melakukan klasifikasi. Nilai *F1-score* yang lebih tinggi menunjukkan bahwa *Random Forest* memiliki keseimbangan yang baik antara kemampuan mengidentifikasi kasus positif dan meminimalkan kesalahan prediksi. Hal ini mengindikasikan bahwa model tersebut tidak hanya akurat, tetapi juga stabil dalam mendeteksi pola hubungan antara variabel kesehatan mental, kondisi kerja, dan kecenderungan *turnover*.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa pendekatan ensemble learning pada *Random Forest* memberikan keunggulan dalam mengolah dataset yang memiliki karakteristik hubungan non-linear serta interaksi kompleks antarvariabel psikososial. Kemampuan model ini dalam menggabungkan banyak pohon keputusan memungkinkan peningkatan stabilitas prediksi dan mengurangi risiko overfitting yang sering terjadi pada model pohon tunggal. Visualisasi pada Gambar 14. Perbandingan Evaluasi Model.



Gambar 14. Perbandingan Evaluasi Model

4.12. Pembahasan

Kajian literatur terdahulu menunjukkan bahwa penelitian mengenai kesehatan mental dan turnover intention semakin berkembang seiring meningkatnya pemanfaatan pendekatan machine learning untuk meningkatkan akurasi prediksi perilaku karyawan. Penelitian berjudul *A Decade of Research on Machine Learning Techniques for Predicting Employee Turnover: A Systematic Literature Review* menyoroti hubungan antara indikator kesehatan mental, stres kerja, dan niat karyawan untuk berhenti. Studi tersebut menegaskan bahwa variabel seperti stres kerja, kelelahan emosional, dan kepuasan kerja merupakan prediktor signifikan dalam mengidentifikasi individu dengan risiko turnover tinggi. Selain itu, penelitian tersebut menemukan bahwa algoritma berbasis ensemble, khususnya *Random Forest*, menunjukkan performa prediksi yang lebih konsisten dibandingkan model tunggal karena kemampuannya menangkap hubungan non-linear dan interaksi kompleks antarvariabel psikososial [14]. Kemampuan ini menjadi penting dalam konteks analisis kesehatan mental karyawan, di mana variabel seperti tekanan kerja, kelelahan, dan dukungan sosial sering kali saling berinteraksi secara tidak linear. Oleh karena itu,

pendekatan machine learning berbasis ensemble memberikan keunggulan analitis dalam mengidentifikasi pola psikologis yang sulit ditangkap melalui metode statistik tradisional.

Sementara itu, penelitian berjudul *Employee Well-Being and Turnover Intention* menekankan peran kesejahteraan psikologis sebagai determinan utama niat turnover. Studi tersebut menemukan bahwa kesejahteraan psikologis memiliki pengaruh signifikan terhadap keinginan karyawan untuk keluar melalui mekanisme mediasi kepuasan kerja dengan nilai signifikansi $p < 0.05$ [15], [16]. Pola hubungan yang kompleks ini memperkuat kebutuhan akan pendekatan analitik yang mampu memodelkan interaksi antarvariabel secara lebih fleksibel. Dalam konteks penelitian ini, variabel seperti tekanan pekerjaan, kelelahan, dan kualitas hubungan interpersonal di tempat kerja muncul sebagai faktor yang memiliki kontribusi penting dalam model prediksi. Oleh karena itu, penggunaan algoritma *Random Forest* menjadi relevan karena kemampuannya mengidentifikasi variabel yang paling berpengaruh melalui mekanisme pemilihan fitur serta menangkap hubungan non-linear antar indikator kesehatan mental. Secara keseluruhan, temuan dari penelitian terdahulu tersebut memperkuat argumen bahwa analisis turnover berbasis indikator kesehatan mental memerlukan pendekatan machine learning yang mampu mengintegrasikan berbagai variabel psikososial secara komprehensif untuk menghasilkan prediksi yang lebih akurat.

5. KESIMPULAN DAN SARAN

Penelitian ini menganalisis hubungan antara kesehatan mental dan kecenderungan turnover karyawan serta mengembangkan model klasifikasi berbasis machine learning untuk memprediksi risiko tersebut. Hasil penelitian menunjukkan bahwa faktor psikososial seperti tekanan pekerjaan, kelelahan, kualitas hubungan dengan atasan, dan dukungan rekan kerja memiliki peran penting dalam membentuk kecenderungan turnover. Temuan ini menegaskan bahwa kesehatan mental tidak hanya berfungsi sebagai indikator kesejahteraan individu, tetapi juga sebagai faktor strategis dalam pengelolaan sumber daya manusia.

Dari sisi metodologis, penelitian ini menunjukkan bahwa algoritma *Random Forest* memberikan performa prediksi yang paling konsisten dibandingkan algoritma lainnya. Keunggulan ini menunjukkan bahwa pendekatan ensemble learning lebih efektif dalam menangkap hubungan non-linear dan interaksi kompleks antarvariabel psikososial dalam konteks prediksi perilaku karyawan. Dengan demikian, penerapan machine learning dapat memberikan pendekatan analitis yang lebih adaptif dalam memahami dinamika turnover dibandingkan metode statistik konvensional.

Penelitian ini memberikan kontribusi ilmiah pada pengembangan analitik sumber daya manusia berbasis data dengan mengintegrasikan indikator kesehatan

mental dalam model prediksi turnover. Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh organisasi untuk mengembangkan sistem pemantauan karyawan berbasis data guna mengidentifikasi risiko turnover secara lebih dini serta merancang strategi retensi yang berfokus pada pengelolaan stres kerja, peningkatan kualitas supervisi, dan penguatan dukungan sosial di lingkungan kerja.

Penelitian selanjutnya disarankan untuk mengembangkan pendekatan *Explainable Artificial Intelligence* (XAI) guna meningkatkan transparansi dan interpretabilitas model prediktif. Selain itu, perlu dilakukan perluasan dataset lintas sektor industri serta eksplorasi algoritma machine learning yang lebih lanjut untuk meningkatkan kemampuan generalisasi model dalam memprediksi perilaku turnover karyawan.

DAFTAR PUSTAKA

- [1] J. Bin Xu, Q. X. Zheng, X. M. Jiang, Q. Zhuo, J. X. Nian, and J. T. Wang, "Mediating effects of social support, mental health between stress overload, fatigue and turnover intention among operating theatre nurses," *BMC Nurs.*, vol. 22, no. 1, 2023, doi: 10.1186/s12912-023-01518-z.
- [2] R. Wernigg and M. Wernigg, "A case study for assessing the utility of a decision tree based learning algorithm in mental health inpatient care quality management," *Eur. Psychiatry*, vol. 65, no. S1, 2022, doi: 10.1192/j.eurpsy.2022.454.
- [3] S. Fukui *et al.*, "Machine Learning with Human Resources Data: Predicting Turnover among Community Mental Health Center Employees," *J. Ment. Health Policy Econ.*, vol. 26, no. 2, 2023.
- [4] N. A. Andarsari and P. B. Setiadi, "The The Impact of Working Environment on Employee Loyalties with Job Satisfaction as an Intervening Variable," *Ilomata Int. J. Manag.*, vol. 4, no. 1, 2023, doi: 10.52728/ijjm.v4i1.619.
- [5] Y. Yulfanani and R. Roziana, "Dampak Kelelahan Mental (Burnout) Terhadap Kinerja Karyawan PT. Duta Merpati Indonesia," *Etn. J. Ekon. dan Tek.*, vol. 1, no. 4, 2022, doi: 10.54543/etnik.v1i4.42.
- [6] J. Mao *et al.*, "Comprehensive strategies of machine-learning-based quantitative structure-activity relationship models," 2021. doi: 10.1016/j.isci.2021.103052.
- [7] D. A. Al Mudaifer, R. S. Al Qahtani, S. V. Tharayil, A. Almass, and S. Dursun, "Intelligent Course Recommender for Professional Development," in *Society of Petroleum Engineers - Gas and Oil Technology Showcase and Conference, GOTS 2023*, 2023. doi: 10.2118/214116-MS.
- [8] Y. Choi and J. W. Choi, "The prediction of workplace turnover using machine learning technique," *Int. J. Bus. Anal.*, vol. 8, no. 4, 2021, doi: 10.4018/IJBAN.2021100101.
- [9] S. Hussain, S. Qazi, R. R. Ahmed, D. Streimikienė, and J. Vveinhardt, "Employees management: Evidence from gamification techniques," *Montenegrin J. Econ.*, vol. 14, no. 4, 2018, doi: 10.14254/1800-5845/2018.14-4.7.
- [10] M. Al Akasheh, E. F. Malik, O. Hujran, and N. Zaki, "A decade of research on machine learning techniques for predicting employee turnover: A systematic literature review," *Expert Syst. Appl.*, vol. 238, 2024, doi: 10.1016/j.eswa.2023.121794.
- [11] W. Wu, M. Rafiq, and T. Chin, "Employee well-being and turnover intention," *Career Dev. Int.*, vol. 22, no. 7, 2017, doi: 10.1108/cdi-04-2017-0072.
- [12] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, "Predicting Employee Attrition Using Machine Learning Approaches," *Appl. Sci.*, vol. 12, no. 13, 2022, doi: 10.3390/app12136424.
- [13] K. Sahoo, A. K. Samal, J. Pramanik, and S. K. Pani, "Exploratory data analysis using python," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 12, 2019, doi: 10.35940/ijitee.L3591.1081219.
- [14] M. Al Akasheh, E. F. Malik, O. Hujran, and N. Zaki, "A decade of research on machine learning techniques for predicting employee turnover: A systematic literature review," *Expert Systems with Applications*, vol. 238, 2024, doi: 10.1016/j.eswa.2023.121794.
- [15] W. Wu, M. Rafiq, and T. Chin, "Employee well-being and turnover intention," *Career Development International*, vol. 22, no. 7, 2017, doi: 10.1108/cdi-04-2017-0072.
- [16] W. Wu, M. Rafiq, and T. Chin, "Employee well-being and turnover intention: Evidence from a developing country with Muslim culture," *Career Development International*, vol. 22, no. 7, 2017, doi: 10.1108/CDI-04-2017-0072.