

## PENERAPAN ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI DEPRESI BERDASARKAN FAKTOR TEKANAN KERJA DAN KEBIASAN HIDUP

**Muhammad Abdul Hamid<sup>✉</sup>, Nor Anisa**

Fakultas Sains dan Teknologi, Universitas Sari Mulia, Banjarmasin, Indonesia

Email: [abdulhamid@student.unism.ac.id](mailto:abdulhamid@student.unism.ac.id)

### ABSTRACT

*Depression represents a significant mental health issue impacting both personal well-being and broader societal dynamics both individuals and society. The Random Forest algorithm will be used in this project to create a depression categorization model based on work pressure and lifestyle characteristics. The Depression Professional Dataset was analyzed using the Knowledge Discovery in Databases (KDD) approach, which included 2,054 data points with 11 factors such as age, work pressure, working hours, sleep habits, and family mental health history. The analysis revealed that the Random Forest algorithm achieved a 91% accuracy in identifying depressive conditions. The investigation found that age was the most important predictor, followed by work pressure, working hours, and job happiness. In contrast, gender and family mental health history had a smaller impact. This study demonstrates that the risk factors for depression are multifaceted, including demography and work pressure. These findings can be used to develop mental health preventive and intervention methods in the workplace. Future model development can include new factors, such as socioeconomic status, to produce a more comprehensive study.*

**Keyword:** Depression, Random Forest, Work Pressure, Knowledge Discovery in Databases (KDD), Classification.

### ABSTRAK

*Depresi adalah kondisi psikologis yang memiliki dampak besar baik pada masyarakat maupun individu. Penelitian ini memanfaatkan algoritma Random Forest untuk membangun model klasifikasi depresi yang mengacu pada faktor tekanan kerja dan gaya hidup. Penelitian ini menggunakan pendekatan Knowledge Discovery in Databases (KDD) untuk menganalisis sampel profesional tentang depresi yang terdiri dari 2.054 data yang mengandung 11 variabel, termasuk usia, tekanan kerja, jam kerja, kebiasaan tidur, dan riwayat kesehatan mental keluarga. Penelitian ini mengungkapkan bahwa Algoritma Random Forest mencapai Tingkat keakuratan 91% dalam mengklasifikasikan kondisi depresi, dengan usia sebagai faktor paling penting, diikuti oleh tekanan kerja, jam kerja, kebiasaan tidur, dan usia. Variabel seperti jenis kelamin dan riwayat kesehatan mental keluarga tidak memberikan pengaruh yang signifikan. Menurut penelitian ini, faktor risiko depresi bersifat multidimensi, termasuk tekanan pekerjaan dan demografi. Temuan ini dapat dimanfaatkan sebagai landasan dalam pengembangan strategi pencegahan dan intervensi kesehatan mental di lingkungan kerja. Untuk melakukan analisis yang lebih mendalam di masa depan, model dapat diperbarui dengan mempertimbangkan faktor lain seperti status sosial ekonomi.*

**Kata Kunci:** Depresi, Random Forest, Tekanan Kerja, The Knowledge Discovery in Databases (KDD), Klasifikasi.

### PENDAHULUAN

Depresi merupakan gangguan emosional yang ditandai dengan rasa sedih berkepanjangan serta kehilangan minat terhadap aktivitas yang biasanya disukai, disertai dengan kesulitan dalam menjalankan aktivitas sehari-hari selama setidaknya dua minggu (Miraza, 2019). Di Indonesia, angka prevalensi depresi cukup tinggi, mencapai 6,1 persen dari populasi usia  $\geq 15$  tahun berdasarkan Riset Kesehatan Dasar (Riskesdas) 2021. Namun, hanya 9 persen dari penderita yang mendapatkan penanganan medis, menunjukkan adanya kesenjangan dalam deteksi dan

intervensi depresi. Depresi yang tidak tertangani dapat menurunkan produktivitas kerja, memengaruhi kualitas hidup, dan meningkatkan risiko kesehatan fisik (Sukamto Jurusan Keperawatan Politeknik Kesehatan Kemenkes Kaltim Jalan Wolter Monginsidi No & Kalimantan Timur, n.d.). Oleh karena itu, diperlukan pendekatan berbasis data untuk memahami faktor risiko depresi secara sistematis guna mendukung upaya pencegahan dan pengobatan (Hakim & Aristawati, 2023).

Penambangan data merupakan proses untuk mengidentifikasi pola serta hubungan tersembunyi

dalam Kumpulan data yang besar dengan memanfaatkan Teknik analisi data Tingkat lanjut. Proses ini melibatkan penggunaan model statistic, algoritma matematika, dan metode pembelajaran mesin sebagai bagian dari pendekatannya (Ningtyas et al., n.d.). Aktivitas ini tidak hanya terbatas pada pengumpulan data, tetapi juga mencangkup analisis dan data secara lebih luas dibandingkan dengan metode regresi (De Wibowo Muhammad Sidik et al., 2020).

Penelitian sebelumnya telah menunjukkan efektivitas algoritma *Machine Learning* dalam memprediksi gangguan mental. Misalnya, penelitian oleh (Sundararajan et al., 2020) memanfaatkan algoritma Support Vector Machine (SVM) untuk melakukan klasifikasi kondisi depresi berdasarkan faktor demografi dan gaya hidup dengan akurasi mencapai 85%. Penelitian lain oleh Lee et al. (2020) menggunakan algoritma *Decision Tree* untuk menganalisis hubungan antara kebiasaan tidur dan depresi, menemukan bahwa durasi tidur memiliki peran signifikan. Sementara itu, Rahman et al. (2021) menggunakan pendekatan Random Forest untuk memprediksi depresi pada tenaga kerja, dengan hasil bahwa tekanan kerja dan riwayat keluarga adalah faktor utama (Dirgayunita, 2016).

Meskipun penelitian tersebut memberikan kontribusi penting, pendekatan mereka masih terbatas dalam pengolahan data yang kompleks dan interpretasi variabel penting. Dalam konteks ini, pendekatan *Knowledge Discovery in Database* (KDD) menawarkan solusi dengan memadukan proses seleksi, praproses, transformasi, penambangan data, dan evaluasi guna memperoleh analisis yang lebih komprehensif dan akurat.

Penelitian ini berfokus pada pengembangan model untuk mengklasifikasikan kondisi depresi pada pekerja menggunakan Algoritma Random Forest yang berasal dari metode CART (Classification and Regression Trees) dan merupakan bagian dari Teknik pohon keputusan (Adrian et al., 2021), mampu mengolah data dalam berbagai skala dan menangkap pola kompleks. Penelitian ini juga bertujuan untuk menentukan faktor utama yang secara signifikan.

State of the Art penelitian ini terletak pada penerapan algoritma Random Forest dalam pendekatan KDD untuk analisis faktor risiko depresi dengan menggunakan dataset komprehensif yang mencakup tekanan kerja, gaya hidup, dan faktor demografi. Dengan demikian, penelitian ini tidak hanya menyumbangkan model prediktif yang akurat, tetapi juga memberikan wawasan yang dapat diterapkan untuk intervensi pencegahan depresi di lingkungan kerja.

Hipotesis Penelitian yang diajukan adalah bahwa faktor usia, tekanan kerja, jam kerja, dan gaya hidup memiliki pengaruh signifikan terhadap risiko depresi pada pekerja.

## TINJAUAN PUSTAKA

### Big Data

Big Data adalah Kumpulan data dengan volume yang sangat besar dan beragam, yang membutuhkan metode pengelolaan serta perangkat bantu dengan kinerja yang memadai. Artikel ini membahas definisi, konsep, dan manfaat Big Data, dilanjutkan dengan Sejarah, penerapan, serta langkah-langkah pengelolaannya. Di bagian akhir, terdapat Kesimpulan mengenai Big Data (Saputra et al., 2022).

### KDD (Knowledge Discovery in Databases)

Knowledge Discovery in Databases (KDD) adalah metode untuk mengekstraksi pengetahuan dari database yang ada. Dalam database, terdapat tabel-tabel yang saling berelasi. Pengetahuan yang dihasilkan dari proses ini dapat dimanfaatkan sebagai dasar pengambilan Keputusan (Mardi, 2017).

### Keunggulan Random Forest

Algoritma Random Forest digunakan untuk klasifikasi Big Data. Berbeda dengan decision tree yang memerlukan proses pemangkasan variabel (pruning), Random Forest memiliki keunggulan dalam menggabungkan banyak pohon untuk melakukan klasifikasi dan prediksi kelas, dibandingkan dengan single tree yang hanya terdiri dari satu pohon (Prasoj & Haryatmi, 2021).

Algoritma berbasis pohon keputusan Random Forest memiliki keunggulan dalam menangani data dengan banyak fitur, mengurangi overfitting dengan metode pembelajaran kelompok, dan menangkap pola yang kompleks dalam data. Keunggulan utamanya dibandingkan dengan algoritma lain adalah kemampuan untuk menangani nilai yang tidak ada dan menemukan fitur yang paling penting untuk klasifikasi.

### Penelitian Sebelumnya

Algoritma pembelajaran mesin untuk klasifikasi depresi telah digunakan dalam beberapa penelitian sebelumnya. Sundararajan et al. (2020) menggunakan Support Vector Machine (SVM) untuk mengkategorikan depresi berdasarkan metrik demografi dan gaya hidup, dengan akurasi 85%. Di sisi lain, Lee et al. (2020) menggunakan Decision Tree untuk memeriksa hubungan antara kebiasaan tidur dan depresi, menunjukkan bahwa durasi tidur memiliki

dampak signifikan terhadap tingkat depresi (Breiman, 2001).

Dalam penelitian lain yang dilakukan oleh Rahman et al. (2021), Random Forest digunakan untuk memprediksi depresi pada tenaga kerja. Studi mereka menunjukkan bahwa tekanan kerja dan pengalaman keluarga adalah penyebab utama depresi. Namun demikian, penelitian ini tidak memberikan detail tentang proses prapemrosesan data dan parameter model yang digunakan. Akibatnya, hasilnya kurang dapat diulang.

## **METODE PENELITIAN**

### **Proses Pengumpulan Data**

Proses pengumpulan data dalam penelitian ini menggunakan data sekunder. Data sekunder merujuk informasi yang diambil dari sumber yang telah tersedia atau telah dikumpulkan oleh pihak lain. Penggunaan data sekunder memungkinkan peneliti untuk menggunakan data yang telah tersedia untuk dianalisis lanjutan. Dalam penelitian ini, data diambil melalui platform Kaggle, yang menyajikan dataset berjudul Depression Profesional Dataset. Dataset tersebut berisi beragam variabel termasuk usia, beban kerja, waktu tidur, pola makan, dan faktor risiko lain yang terkait dengan kondisi depresi.

Metode pengumpulan data yang diterapkan dalam penelitian ini adalah metode dokumentasi, yang dilakukan dengan mengakses dataset yang telah tersedia di platform Kaggle, kemudian mengunduh dan mengolah data tersebut sesuai kebutuhan penelitian. Dataset yang digunakan memiliki total 10 fitur dan satu label target yang menjelaskan kondisi depresi seorang pekerja.

### **Tahapan Penelitian**

Penelitian ini menggunakan tahapan dalam proses Knowledge Discovery in Databases (KDD). KDD merupakan berharga dari data melalui penerapan metodologi tertentu (Ardiantoro et al., 2019). KDD adalah pendekatan penting dalam mengidentifikasi pengetahuan baru yang tersembunyi di dalam data mentah (Hamidah et al., 2013). Proses KDD bertujuan untuk menemukan pola berharga atau wawasan baru dari data dengan melalui serangkaian langkah yang dirancang untuk mengubah data mentah menjadi informasi yang lebih bermakna dan bermanfaat (Yoga Pratama et al., 2021).

#### **1. Selection**

Tahap Pertama, data yang diperlukan untuk analisis dipilih dari kumpulan data yang tersedia dan dipisahkan dari data lainnya untuk menjaga fokus pada informasi yang penting. Proses ini termasuk

mengidentifikasi sumber data yang tepat dan mengekstrak elemen tertentu dari kumpulan data tersebut yang diperlukan untuk penelitian. Langkah ini membantu mengurangi kompleksitas dan memungkinkan penggunaan data yang lebih terarah dalam proses berikutnya.

#### **2. Preprocessing**

Praproses mencakup pembersihan data untuk menyelesaikan masalah misalnya, data yang tidak lengkap, nilai yang kurang konsisten, serta kesalahan entri. Ini termasuk mencakup penghapusan data duplikat dan koreksi kesalahan pengetikan. Pastikan bahwa data yang digunakan tidak memiliki kesalahan yang dapat memengaruhi hasil analisis adalah tujuan utama tahap ini, agar hasilnya lebih tepat, terpercaya, dan data harus bersih.

#### **3. Transformation**

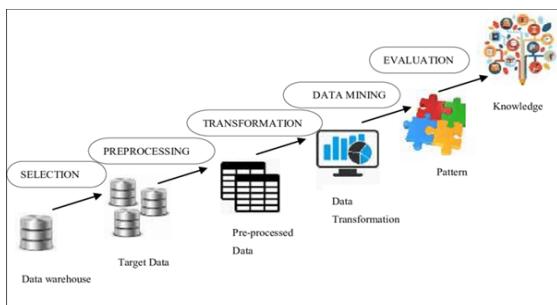
Pada tahap ini, data yang telah dibersihkan diubah untuk memenuhi persyaratan proses analisis. Agregasi, normalisasi, atau pembuatan fitur baru yang relevan dengan tujuan penelitian adalah beberapa contoh transformasi. Proses ini dilakukan agar algoritma data mining yang akan digunakan dapat menggunakan data secara optimal. Contoh untuk memudahkan proses analisis lebih lanjut, data kategori dapat diubah menjadi data numerik.

#### **4. Data Mining**

Pemilihan algoritma untuk klasifikasi, klasterisasi, atau asosiasi, antara lain, bergantung pada jenis data dan tujuan analisis, data mining merupakan komponen utama dalam proses KDD, di mana beragam metode serta algoritma diterapkan untuk mengidentifikasi pola, keterkaitan, atau wawasan yang tidak terlihat secara langsung dalam data. Proses ini menghasilkan pengetahuan baru yang bermanfaat untuk pengambilan keputusan strategis.

#### **5. Evaluation**

Tahap terakhir adalah evaluasi, di mana pola atau hasil dari proses data mining disajikan dalam bentuk yang sederhana dipahami untuk berbagai pihak yang terlibat. Proses interpretasi juga melibatkan verifikasi kesesuaian informasi yang diperoleh dengan membandingkan terhadap hipotesis atau fakta yang telah ada sebelumnya.



**Gambar 1.** Proses KDD

Melalui kelima tahapan ini, penelitian bertujuan untuk mendapatkan hasil yang valid dan berguna, serta memberikan wawasan yang mendalam tentang data yang dianalisis.

## HASIL DAN PEMBAHASAN

### Membuat Model Algoritma Random Forest Untuk Klasifikasi Depresi

#### Dataset

Penelitian ini menggunakan dataset Depression Professional Dataset yang diperoleh melalui platform Kaggle. Dataset ini berisi informasi mengenai faktor-faktor yang memengaruhi kondisi depresi pada pekerja. Dataset terdiri dari 10 fitur prediktor, yang mencakup variabel seperti Usia, Jenis Kelamin, tekanan kerja, Jam kerja, Kepuasan kerja, Stress Finansial, Durasi Tidur, Kebiasaan Makan, Pikiran untuk bunuh diri, serta Riwayat gangguan mental dalam keluarga, serta satu label target yang menggambarkan kondisi depresi, yakni apakah individu tersebut mengalami depresi atau tidak.

Dataset ini mengandung data numerik dan kategoris. Semua nilai kosong yang terdapat dalam dataset telah diatasi dengan metode imputasi menggunakan median, sehingga data siap diproses tanpa ada nilai yang hilang. Tujuan penelitian ini adalah mengungkapkan faktor-faktor kunci yang dapat memprediksi depresi serta memberikan panduan bagi pengembangan strategi pencegahan yang lebih baik.

#### Selection

Pada tahap seleksi, data yang digunakan berasal dari Depression Professional Dataset yang diperoleh dari platform Kaggle. Dataset ini terdiri dari 10 fitur dan satu label target. Fitur-fitur dalam dataset ini seperti Usia, Jenis Kelamin, tekanan kerja, Jam kerja, Kepuasan kerja, Stress Finansial, Durasi Tidur, Kebiasaan Makan, Pikiran untuk bunuh diri, serta Riwayat gangguan mental dalam keluarga. Label target menggambarkan kondisi depresi individu, yaitu Depressed atau Not Depressed. Setelah melalui tahap seleksi, dataset ini sudah siap digunakan untuk analisis lebih lanjut.

	Age	Work Pressure	Sleep Duration	Dietary Habits
0	37	2.0	7-8 hours	Moderate
1	60	4.0	5-6 hours	Unhealthy
2	42	2.0	5-6 hours	Moderate
3	44	3.0	7-8 hours	Healthy
4	48	4.0	7-8 hours	Moderate

	Family History of Mental Illness	Depression
0	No	No
1	Yes	No
2	No	No
3	Yes	No
4	Yes	No

**Gambar 2.** Data Selection

#### Preprocessing

Pada tahap seleksi, data yang digunakan berasal dari Depression Professional Dataset yang diperoleh dari platform Kaggle. Dataset ini terdiri dari 10 fitur dan satu label target. Fitur-fitur dalam dataset ini seperti Usia, Jenis Kelamin, tekanan kerja, Jam kerja, Kepuasan kerja, Stress Finansial, Durasi Tidur, Kebiasaan Makan, Pikiran untuk bunuh diri, serta Riwayat gangguan mental dalam keluarga. Label target menggambarkan kondisi depresi individu, yaitu Depressed atau Not Depressed. Setelah melalui tahap seleksi, dataset ini sudah siap agar diproses lebih mendalam.

Age	0
Work Pressure	0
Sleep Duration	0
Dietary Habits	0
Family History of Mental Illness	0
Depression	0
<b>dtype: int64</b>	
Age	Age
Work Pressure	Work Pressure
Sleep Duration	Sleep Duration
Dietary Habits	Dietary Habits
Family History of Mental Illness	Family History of Mental Illness
Depression	Depression
0	37
1	60
2	42
3	44
4	48
	2.0
	4.0
	2.0
	3.0
	4.0
	7-8 hours
	5-6 hours
	5-6 hours
	7-8 hours
	7-8 hours
	Moderate
	Unhealthy
	Moderate
	Healthy
	Moderate
	No
	Yes
	No
	Yes
	No

**Gambar 3.** Preprocessing

- Penanganan Data Hilang: Missing values dalam dataset ditangani dengan metode imputasi median untuk data numerik dan modus untuk data kategoris.
- Transformasi Data: Data kategori dikonversi menjadi numerik menggunakan teknik one-hot encoding, sementara fitur numerik dinormalisasi menggunakan MinMaxScaler untuk memastikan distribusi data yang seragam.
- Pengurangan Dimensi: Teknik Principal Component Analysis (PCA) digunakan untuk mengurangi dimensi data guna meningkatkan efisiensi pemrosesan model.

## Transformation

Tahap transformasi melibatkan pembagian data menjadi dua kelompok, yaitu data latih (80%) dan data uji (20%). Tujuan pembagian ini adalah untuk melatih model menggunakan Sebagian data kemudian melakukan pengujian pada bagian lain guna mencegah terjadinya overfitting. Algoritma Random Forest bekerja dengan baik meskipun tanpa normalisasi atau standarisasi data, sehingga pada tahap ini tidak diperlukan langkah tambahan untuk memodifikasi skala data.

Training set: (1643, 5), Testing set: (411, 5)

Gambar 4. Data Training dan Tes

## Data Mining

Pada tahap data mining, algoritma Random Forest diterapkan untuk membuat model klasifikasi yang bertujuan untuk memprediksi apakah seorang individu mengalami depresi atau tidak. Model dilatih menggunakan data latih (80%), kemudian diuji dengan data uji (20%) guna mengevaluasi performa klasifikasi. Hasil evaluasi menunjukkan bahwa model Random Forest mencapai akurasi sebesar 90.51%, yang mengindikasikan efektivitasnya dalam mengklasifikasikan kondisi depresi.

Setelah melatih model dengan data latih lalu model tersebut diuji menggunakan data pengujian guna memprediksi kondisi depresi. Evaluasi menunjukkan bahwa model memiliki akurasi sebesar 90.51%, mengindikasikan efektivitas Random Forest dalam mengklasifikasikan kondisi depresi. Dalam dataset ini.

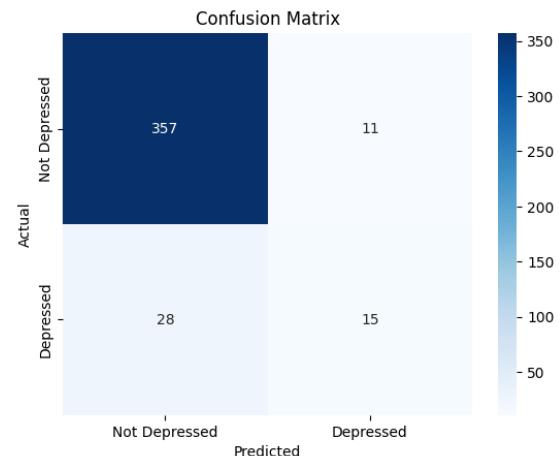
Akurasi Model: 90.51%

Gambar 5. Akurasi Model

## Evaluation

Model dievaluasi menggunakan berbagai metrik evaluasi, termasuk akurasi, presisi, recall dan F1-score. Hasil evaluasi mengindikasikan bahwa model memiliki performa yang sangat, dengan hasil sebagai berikut:

- Akurasi Model: 90.51%
- Confusion Matrix:

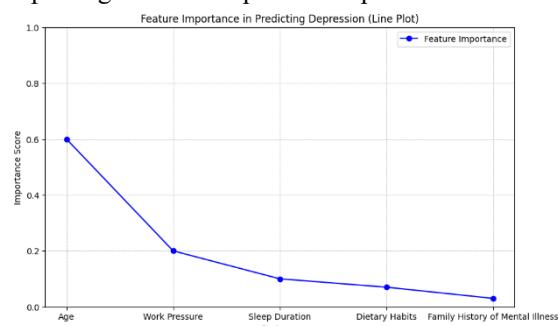


Gambar 6. Confusion Matrix

- *Not Depressed*: 357 prediksi benar, 11 prediksi salah.
- *Depressed*: 15 prediksi benar, 28 prediksi salah.
- Classification Report:

Classification Report:					
	precision	recall	f1-score	support	
0	0.93	0.97	0.95	368	
1	0.58	0.35	0.43	43	
accuracy				0.91	411
macro avg	0.75	0.66	0.69	411	
weighted avg	0.89	0.91	0.89	411	

- Feature Importance: Grafik pentingnya fitur menunjukkan bahwa *Age* merupakan fitur paling penting dalam memprediksi depresi.



Gambar 7. Importance Score

## Interpretasi Hasil

Algoritma Random Forest digunakan dalam penelitian ini untuk mengklasifikasikan kondisi depresi berdasarkan berbagai faktor, seperti usia, tekanan kerja, durasi tidur, dan elemen gaya hidup lainnya. Hasil analisis menunjukkan bahwa variabel usia memiliki kontribusi terbesar terhadap prediksi depresi, dengan nilai feature importance sebesar 39,82%. Ini menunjukkan bahwa usia adalah faktor utama yang berkontribusi terhadap tingkat depresi individu dalam kumpulan data.

Ini sesuai dengan data epidemiologi yang menunjukkan bahwa depresi adalah penyebab utama

perilaku bunuh diri dan disabilitas pada remaja.(Mandasari et al., 2020) Dengan nilai penting 10,53%, faktor kedua yang sangat penting adalah tekanan kerja. Tekanan kerja yang tinggi sering mengganggu keseimbangan antara kehidupan pribadi dan pekerjaan, yang dapat menyebabkan stres jangka panjang dan akhirnya depresi. Pada tahun 2019, data dari Mental Health America menunjukkan bahwa hampir 90% pekerja di seluruh dunia melaporkan bahwa tekanan dan stres di tempat kerja memengaruhi kesehatan mental mereka(Oktavia Soumokil-Mailoa et al., 2022).

Faktor lain, seperti Jam Kerja dan Kepuasan Kerja, juga memiliki kontribusi signifikan, dengan nilai penting 10,04% dan 9,69%, masing-masing. Jam kerja yang panjang (lebih dari 40 jam seminggu) berdampak pada peningkatan depresi, kecemasan, kesulitan tidur, dan penyakit jantung koroner.

Hasil ini memberikan wawasan penting bahwa faktor utama yang memengaruhi depresi bersifat multidimensional, mencakup aspek demografis (usia), tekanan pekerjaan, serta kebiasaan hidup. Temuan ini dapat digunakan sebagai dasar untuk intervensi yang lebih terarah, seperti program kesehatan mental yang difokuskan pada kelompok usia tertentu atau reformasi kebijakan terkait keseimbangan kerja-hidup. Ke depan, penelitian mendatang menggunakan dataset yang lebih besar dan variabel tambahan, seperti status sosial ekonomi atau tingkat pendidikan, dapat dilakukan untuk memperoleh hasil yang lebih komprehensif.

### **Implikasi Hasil dan Batasan Penelitian**

Hasil penelitian ini memiliki konsekuensi signifikan untuk pengobatan depresi dan pencegahannya. Program kesehatan mental yang difokuskan pada kelompok usia rentan, seperti remaja dan dewasa awal, dapat membantu mengurangi tingkat depresi melalui sesi terapi psikologis dan program edukasi kesehatan mental. Selain itu, peningkatan kebijakan keseimbangan kerja-hidup di tempat kerja, seperti jam kerja yang lebih fleksibel dan dukungan mental bagi karyawan yang menghadapi tekanan kerja tinggi, dapat menjadi strategi efektif dalam mencegah stres kronis. Penerapan sistem evaluasi kepuasan kerja secara berkala juga dapat membantu mendeteksi risiko depresi di lingkungan kerja dan mencegah dampak psikologis negatif pada karyawan.

Namun, penelitian ini memiliki beberapa keterbatasan yang harus diperhatikan. Karena data yang digunakan hanya terbatas pada satu dataset Kaggle, hasilnya mungkin tidak dapat sepenuhnya digeneralisasi ke populasi yang lebih luas. Selain itu, faktor sosial ekonomi tidak dimasukkan dalam model

ini, padahal aspek ini dapat berpengaruh besar terhadap tingkat stres dan kesehatan mental seseorang. Model ini juga belum dibandingkan dengan algoritma lain, seperti deep learning atau SVM, yang mungkin dapat meningkatkan akurasi prediksi. Oleh karena itu, penelitian selanjutnya dapat menggunakan dataset yang lebih beragam serta mempertimbangkan variabel tambahan, seperti status sosial ekonomi dan tingkat pendidikan, untuk memperoleh hasil yang lebih komprehensif dan aplikatif dalam pencegahan serta intervensi depresi.

### **KESIMPULAN**

Studi ini sukses membangun model klasifikasi depresi dengan menerapkan algoritma Random Forest dengan pendekatan Knowledge Discovery in Databases (KDD). Hasil analisis menunjukkan bahwa algoritma Random Forest mampu mengklasifikasikan kondisi depresi dengan baik, dengan faktor usia sebagai variabel yang paling signifikan dalam memprediksi risiko depresi. Variabel-variabel lainnya, seperti tekanan kerja, jam kerja, dan kepuasan kerja, juga memiliki kontribusi yang penting, sementara variabel seperti jenis kelamin dan riwayat kesehatan mental keluarga memiliki dampak yang lebih kecil.

Temuan ini menegaskan bahwa faktor-faktor yang mempengaruhi depresi bersifat multidimensional dan saling terkait, baik dengan aspek demografis maupun tekanan pekerjaan. Pendekatan KDD yang diterapkan dalam studi ini memungkinkan pengolahan data dengan pendekatan sistematis serta menghasilkan wawasan yang bermanfaat dalam merancang strategi pencegahan depresi yang lebih efektif di lingkungan kerja.

### **Saran**

1. Penggunaan Model yang Lebih Kompleks: Untuk meningkatkan akurasi prediksi depresi, disarankan untuk mengembangkan model dengan algoritma yang lebih kompleks, seperti deep learning neural networks (DNN) atau ensemble models lainnya, yang dapat menangkap hubungan non-linear antara variabel-variabel yang mempengaruhi depresi.
2. Pengumpulan Data yang Lebih Luas dan Diversifikasi Variabel: Penelitian ini menggunakan data yang terbatas dalam hal jumlah variabel. Disarankan untuk mengumpulkan data dengan cakupan yang lebih luas, termasuk variabel tambahan seperti status sosial ekonomi, tingkat pendidikan, pola makan, tingkat aktivitas fisik, dan kesehatan fisik secara keseluruhan. Pendekatan ini mendukung mendapatkan temuan yang lebih

- komprehensif dalam memahami faktor-faktor yang mempengaruhi depresi.
3. Implementasi di Lingkungan Kerja Nyata: Hasil penelitian ini bisa diaplikasikan dalam pengembangan program intervensi di lingkungan kerja untuk mengurangi stres dan meningkatkan kesejahteraan karyawan. Misalnya, melalui program kesehatan mental yang terstruktur, sesi konsultasi psikologis, atau penyesuaian kebijakan tempat kerja terkait tekanan kerja dan jam kerja.
4. Evaluasi Berkala pada Model Klasifikasi: Disarankan untuk melakukan evaluasi berkala terhadap model klasifikasi depresi yang dikembangkan, termasuk memperbarui data latih dan memperbaiki metode jika diperlukan. Hal ini bertujuan untuk memastikan model tetap relevan dan akurat seiring dengan perubahan kondisi di lingkungan kerja dan masyarakat.

Dengan mempertimbangkan saran-saran tersebut, diharapkan, studi ini dapat berkontribusi secara signifikan dalam memperdalam pemahaman tentang faktor risiko depresi dan mengembangkan strategi pencegahan yang lebih efektif di lingkungan kerja serta masyarakat pada umumnya.

## **DAFTAR PUSTAKA**

- Adrian, M. R., Putra, M., Rafialdy, M. h, & Rakhmawati4, N. A. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB. *Informatika Upgris*, 7.
- Ardiantoro, L., Zahara, S., & Sunarmi, N. (2019). Pemanfaatan Knowledge Data Discovery(Kdd) Pada Pola Permainan Atlet Bulutangkis. In *Jurnal Explore It!* (Vol. 11, Issue 1). Online. <http://jurnal.yudharta.ac.id/v2/index.php/EXPLORE-IT/>
- Breiman, L. (2001). *Random Forests* (Vol. 45).
- De Wibowo Muhammad Sidik, A., Himawan Kusumah, I., Suryana, A., Artyasa, M., & Praditha Junfithrana, A. (2020). *Gambaran Umum Metode Klasifikasi Data Mining*. 2(2), 34–38.
- Dirgayunita, A. (2016). *Depresi: Ciri, Penyebab dan Penangannya*.
- Hakim, Moh. A., & Aristawati, N. V. (2023). Mengukur depresi, kecemasan, dan stres pada kelompok dewasa awal di Indonesia: Uji validitas dan reliabilitas konstruk DASS-21. *Jurnal Psikologi Ulayat*, 10(2), 232–250. <https://doi.org/10.24854/jpu553>
- Hamidah, K., Voutama, A., (2023). Analisis Faktor Tingkat Kebahagiaan Negara Menggunakan Data World Happiness Report dengan Metode Regresi Linier. *Jurnal Explore IT*. <https://doi.org/10.35891/explorit>
- Mandasari, L., Tobing, D. L., Fakultas, ), & Kesehatan, I. (2020). Bidang ilmu: Keperawatan

- Tingkat Depresi Dengan Ide Bunuh Diri Pada Remaja. In *Indonesian Jurnal of Health Development* (Vol. 2, Issue 1).
- Mardi, Y. (2017). *Data Mining : Klasifikasi Menggunakan Algoritma C4.5*.
- Miraza, B. H. (2019). Seputar Resesi dan Depresi. *Seputar Resesi Dan...{Miraza}| 11 Jurnal Ekonomi KIAT*, 30(2). <https://journal.uir.ac.id/index.php/kiat>
- Ningtyas, A. R., Puspitasari, I. M., & Sinuraya, R. K. (2018). *Review Artikel : Farmakoterapi Depresi Dan Pengaruh Jenis Kelamin Terhadap Efisiensi Antidepresan*.
- Oktavia Soumokil-Mailoa, E., Paulus Hermanto, Y., Hindradjat, J., & Tinggi Teologi Kharisma Bandung, S. (2022). *Orang Tua Sebagai Supporting System: Penanganan Anak Remaja Yang Mengalami Depresi*. <https://jurnal.sttekumene.ac.id>
- Prasodo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(2), 79–89. <https://doi.org/10.25077/teknosi.v7i2.2021.79-89>
- Saputra, A., Firdaus, M. I., Wahyudi, R., Mohdo, L., Gunawan, M. E., Encep, M., & Khaira, M. (2022). Big Data. In *Karimah Tauhid* (Vol. 1).
- Sukamto (2018). Depresi Mempengaruhi Kinerja (Sebuah Best Practice). In *Mahakam Nursing Journal* (Vol. 2, Issue 7).
- Yoga Pratama, A., Umaidah, Y., & Voutama, A. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 5, Issue 2).