

PREDIKSI KESEHATAN MENTAL MENGGUNAKAN RANDOM FOREST BERDASARKAN FAKTOR DEMOGRAFI DAN LINGKUNGAN KERJA

Ririn, Ade Rizki Rinaldi, Fadhil M Basysyar

Sistem Informasi, STMIK IKMI CIREBON

Jl. Perjuangan No.1B Cirebon

Rirykim22@gmail.com

ABSTRAK

Kesehatan mental karyawan di sektor teknologi menjadi isu yang semakin penting dengan meningkatnya tuntutan kerja, tekanan lingkungan, dan kurangnya dukungan perusahaan terhadap kesejahteraan psikologis. Banyak pekerja di industri ini mengalami stres, kecemasan, bahkan depresi akibat beban kerja yang tinggi dan budaya kerja yang kompetitif. Namun, masih minim penelitian yang secara khusus menganalisis bagaimana faktor demografi dan lingkungan kerja mempengaruhi risiko gangguan kesehatan mental. Oleh karena itu, penelitian ini berfokus pada prediksi kesehatan mental karyawan di sektor teknologi dengan menggunakan dataset dari survei "Mental Health in Tech" oleh OSMI periode 2017-2022 yang diambil dari Kaggle. Metodologi yang digunakan adalah *Knowledge Discovery in Database* (KDD), dengan model prediksi berbasis *Random Forest* yang diimplementasikan menggunakan Google Colab Python. Penelitian ini bertujuan untuk menganalisis faktor-faktor utama yang berkontribusi terhadap kesehatan mental pekerja teknologi dan meningkatkan akurasi deteksi dini individu yang berisiko mengalami gangguan mental. Hasil model menunjukkan akurasi prediksi sebesar 93,8%, dengan *precision* untuk kelas terdiagnosis (kelas 0) mencapai 89% dan untuk kelas tak terdiagnosis (kelas 1) mencapai 99%. *Recall* untuk kelas 0 adalah 99%, sementara untuk kelas 1 adalah 89%. *F1-Score* menunjukkan performa yang sangat baik dengan nilai 94% untuk kedua kelas. Analisis AUC yang mendekati 1 (0.99) menegaskan bahwa model *Random Forest* memiliki kapabilitas tinggi dalam membedakan individu yang mengalami gangguan kesehatan mental dengan yang tidak. Temuan penelitian ini memberikan wawasan penting bagi perusahaan teknologi dalam memahami faktor-faktor yang mempengaruhi kesehatan mental karyawan serta sebagai dasar untuk merancang kebijakan yang lebih mendukung kesejahteraan pekerja.

Kata kunci : *Random Forest, Prediksi, Mental Health, Kesehatan Mental, Demografi, Lingkungan Kerja.*

1. PENDAHULUAN

Dalam beberapa tahun terakhir, kesadaran akan pentingnya kesehatan mental semakin meningkat, terutama dalam konteks lingkungan kerja. Beban kerja yang tinggi, kurangnya dukungan sosial, serta tekanan dari lingkungan kerja dapat menyebabkan gangguan kesehatan mental, seperti stres, kecemasan, dan depresi [1].

Berbagai penelitian menunjukkan bahwa faktor demografi dan lingkungan kerja memiliki pengaruh signifikan terhadap kesehatan mental karyawan. Oleh karena itu, diperlukan metode yang dapat mengidentifikasi individu yang berisiko mengalami gangguan kesehatan mental sejak dini, sehingga memungkinkan intervensi yang lebih cepat dan tepat sasaran [2][3].

Di era digital, penerapan *machine learning* dalam bidang kesehatan mental semakin berkembang. Teknologi ini mampu mengolah data dalam jumlah besar dan mengidentifikasi pola yang sulit ditemukan melalui metode konvensional [4][5].

Beberapa algoritma telah diterapkan untuk memprediksi kesehatan mental, seperti *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, dan *Random Forest* [6].

Penelitian sebelumnya menunjukkan bahwa model berbasis *machine learning* dapat meningkatkan akurasi prediksi gangguan kesehatan mental, namun masih menghadapi tantangan seperti

ketidakseimbangan data dan keterbatasan dalam menangkap faktor lingkungan kerja [7][8].

Salah satu metode yang cukup menjanjikan adalah *Random Forest*, yang dikenal mampu menangani data dengan banyak variabel serta memiliki tingkat akurasi yang tinggi dalam klasifikasi [9].

Model ini bekerja dengan membangun banyak *decision tree* untuk meningkatkan stabilitas prediksi dan mengurangi kemungkinan *overfitting* [6].

Beberapa penelitian menunjukkan bahwa *Random Forest* dapat memberikan hasil yang lebih akurat dibandingkan metode lain dalam memprediksi kondisi kesehatan mental individu, terutama ketika digunakan dengan teknik peningkatan data seperti *Synthetic Minority Over-sampling Technique (SMOTE)* ([10].

Penelitian ini bertujuan untuk menerapkan metode *Random Forest* dalam memprediksi kesehatan mental berdasarkan faktor demografi dan lingkungan kerja. Dengan menggunakan dataset dari *OSMI Mental Health in Tech Survey* (2017–2022), penelitian ini dapat memberikan wawasan lebih dalam mengenai faktor-faktor yang berkontribusi terhadap gangguan kesehatan mental karyawan di industri teknologi. Selain itu, penelitian ini juga bertujuan untuk mengatasi keterbatasan penelitian sebelumnya dengan menggunakan data yang lebih luas dan teknik pemrosesan data yang lebih canggih [2].

Hasil dari penelitian ini berpotensi memberikan kontribusi yang signifikan dalam memahami hubungan antara faktor demografi dan lingkungan kerja dengan kesehatan mental. Selain itu, temuan penelitian ini dapat dimanfaatkan oleh perusahaan dalam merancang kebijakan kesehatan mental yang lebih efektif, serta membantu menciptakan lingkungan kerja yang lebih sehat dan mendukung kesejahteraan karyawan [1].

Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam bidang akademik tetapi juga memiliki dampak praktis dalam dunia kerja dan kebijakan organisasi.

2. TINJAUAN PUSTAKA

Penelitian mengenai prediksi kesehatan mental berbasis *machine learning* telah banyak dilakukan dengan berbagai pendekatan algoritma. Seperti penelitian oleh Alghamdi et al (2020) dimana metode *Support Vector Machine (SVM)* digunakan untuk memprediksi gejala depresi berdasarkan teks yang diambil dari forum psikologi berbahasa Arab. Model berbasis *machine learning* ini menunjukkan akurasi lebih dari 80% dalam mengidentifikasi pola bahasa yang mengindikasikan depresi. Namun, penelitian ini hanya berfokus pada analisis teks dan tidak mempertimbangkan faktor demografi atau lingkungan kerja sebagai variabel prediktif [11].

Penelitian oleh Saha et al (2022) meneliti bagaimana diskusi di media sosial, terutama Reddit, dapat digunakan untuk memprediksi kebutuhan konsultasi kesehatan mental di lingkungan kampus. Model *time series SARIMA* yang digunakan berhasil meningkatkan akurasi prediksi sebesar 41%. Namun, penelitian ini tidak mempertimbangkan data terstruktur seperti lingkungan kerja atau faktor demografi sebagai variabel prediktif [12].

Penelitian oleh Meylani et al (2022) mengembangkan aplikasi berbasis *Naïve Bayes Classifier* untuk memprediksi kondisi kesehatan pengguna berdasarkan parameter vital seperti detak jantung dan tekanan darah. Meskipun model ini berhasil mengklasifikasikan kondisi kesehatan dengan akurasi tinggi, penelitian ini lebih berfokus pada kesehatan fisik daripada kesehatan mental, sehingga tidak relevan dalam konteks lingkungan kerja [13].

Penelitian oleh Adler et al (2022) menggunakan data dari sensor ponsel untuk memprediksi gejala kesehatan mental seperti stres dan kualitas tidur. Dengan menggunakan model *Gradient Boosting Regression Trees (GBRT)* dan teknik *Leave-One-Subject-Out Cross-Validation (LOSO-CV)*, penelitian ini menunjukkan bahwa penggabungan data dari berbagai sumber dapat meningkatkan akurasi prediksi. Namun, model ini kurang mempertimbangkan faktor lingkungan kerja yang juga dapat memengaruhi kesehatan mental individu [14].

Penelitian oleh Alkahtani et al (2024) mengembangkan sistem berbasis kecerdasan buatan untuk membantu diagnosis dini gangguan kesehatan

mental. Algoritma yang digunakan meliputi *Random Forest (RF)*, *K-Nearest Neighbors (KNN)*, dan *Long Short-Term Memory (LSTM)*, dengan *Random Forest* mencapai akurasi tertinggi (100%). Namun, penelitian ini tidak membahas bagaimana lingkungan kerja dapat berkontribusi terhadap kesehatan mental individu [7].

Penelitian oleh Adeniji et al (2022) mengembangkan model hibrida *Random Forest – Artificial Neural Network (ANN)* untuk memprediksi gangguan kesehatan mental di lingkungan kerja sektor teknologi. Model ini menunjukkan peningkatan performa dengan recall mencapai 97%. Namun, penelitian ini tidak membahas metode penyeimbangan data untuk menangani bias dalam data kesehatan mental, yang dapat menyebabkan prediksi yang kurang akurat untuk kelompok minoritas [9].

Penelitian oleh Khurana et al (2022) memanfaatkan berbagai algoritma *machine learning* termasuk *Random Forest*, *SVM*, dan *Boosting* untuk memprediksi masalah kesehatan mental di lingkungan kerja industri teknologi. Hasil penelitian menunjukkan bahwa algoritma *Boosting* memiliki kinerja optimal dengan akurasi 81,7%. Meskipun demikian, dataset yang digunakan dalam penelitian ini relatif lama (tahun 2014), sehingga kurang relevan dengan kondisi saat ini [10].

Penelitian oleh Pandey et al (2021) berfokus pada prediksi kesehatan mental remaja menggunakan algoritma *Random Forest*, *SVM*, dan *Decision Tree*. Hasil penelitian menunjukkan bahwa *Random Forest* memiliki akurasi tertinggi dalam mengklasifikasikan gangguan kecemasan dan depresi. Namun, penelitian ini terbatas pada populasi remaja dan tidak mempertimbangkan faktor lingkungan kerja sebagai elemen yang dapat mempengaruhi kesehatan mental [15].

Kemudian ada juga penelitian Mohamed et al (2023) yang bertujuan untuk meramalkan atau memprediksi gangguan kecemasan di daerah konflik menggunakan model hibrida yang menggabungkan *Support Vector Machine (SVM)*, *Multilayer Perceptron (MLP)*, dan *Random Forest (RF)*. Model *Random Forest* menunjukkan performa terbaik dengan nilai kappa 0,97. Namun, penelitian ini tidak membahas faktor lingkungan kerja yang juga dapat menjadi pemicu stres dan gangguan kesehatan mental [6].

Terakhir ada penelitian oleh Tate et al (2020) mengembangkan model prediksi kesehatan mental remaja menggunakan *Random Forest*, *SVM*, dan *XGBoost*. Teknik *SMOTE* digunakan untuk mengatasi ketidakseimbangan data. Hasil penelitian menunjukkan bahwa *Random Forest* memiliki kinerja prediksi terbaik dengan AUC sebesar 0.739. Namun, penelitian ini berfokus pada populasi remaja, sehingga hasilnya tidak dapat langsung diterapkan dalam konteks lingkungan kerja.

Berdasarkan penelitian sebelumnya, dapat disimpulkan bahwa *Random Forest* merupakan salah satu algoritma yang paling efektif dalam memprediksi

kesehatan mental karena kemampuannya menangani data kompleks dan meningkatkan akurasi klasifikasi. Namun, sebagian besar penelitian masih memiliki keterbatasan dalam memahami hubungan antara faktor demografi dan lingkungan kerja terhadap kesehatan mental. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan menerapkan *Random Forest* pada dataset *OSMI Mental Health in Tech Survey* (2017–2022) serta mengintegrasikan teknik *SMOTE* untuk meningkatkan akurasi model. Dengan pendekatan ini, diharapkan penelitian ini dapat memberikan kontribusi yang lebih signifikan dalam memahami dan mengatasi masalah kesehatan mental di lingkungan kerja.

2.1. Kesehatan Mental

Kesehatan mental merupakan kondisi kesejahteraan seseorang di mana individu dapat menyadari potensinya sendiri, mampu mengatasi tekanan hidup, bekerja secara produktif, dan berkontribusi dalam komunitasnya (*World Health Organization*, 2022). Kesehatan mental yang baik mencerminkan keseimbangan emosional, psikologis, dan sosial dalam kehidupan seseorang, yang mempengaruhi cara berpikir, merasa, serta berinteraksi dengan orang lain [1].

2.2. Random Forest

Random Forest adalah salah satu algoritma *machine learning* berbasis *ensemble learning* yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan membangun sejumlah besar *decision tree* dari sampel dataset yang berbeda, kemudian menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir yang lebih akurat dan stabil.

2.3. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari *OSMI Mental Health in Tech Survey* untuk periode 2017–2022, yang diperoleh dari Kaggle. Dataset ini berisi informasi dari 2000 responden yang bekerja di industri teknologi dan mencakup 83 variabel yang terkait dengan demografi, lingkungan kerja, dan kondisi kesehatan mental individu. Beberapa variabel utama yang digunakan dalam penelitian ini meliputi:

- a. Demografi :
 - *What is your age?* (Usia)
 - *What is your gender?* (Jenis Kelamin)
 - *What country do you live in?* (Negara tempat tinggal)
 - *What country do you work in?* (Negara tempat bekerja)
- b. Lingkungan Kerja
 - *Does your employer provide mental health benefits?* (Apakah perusahaan menyediakan manfaat kesehatan mental?)
 - *Has your employer ever formally discussed mental health?* (Apakah perusahaan pernah membahas kesehatan mental secara resmi?)

- *Would you feel comfortable discussing a mental health issue with your coworkers?* (Apakah Anda merasa nyaman membicarakan kesehatan mental dengan rekan kerja?)

c. Kesehatan Mental

- *Do you currently have a mental health disorder?* (Apakah Anda saat ini memiliki gangguan kesehatan mental?) → Target prediksi
- *Have you ever been diagnosed with a mental health disorder?* (Apakah Anda pernah didiagnosis memiliki gangguan kesehatan mental?)
- *Have you ever sought treatment for a mental health disorder?* (Apakah Anda pernah mencari pengobatan untuk gangguan kesehatan mental?)

2.4. Evaluasi Model Machine Learning

Dalam penelitian ini, evaluasi model dilakukan menggunakan beberapa metrik yang umum digunakan dalam klasifikasi biner. Berikut adalah persamaan yang digunakan:

a. Akurasi

Akurasi mengukur seberapa sering model membuat prediksi yang benar.

$$Accuracy = \frac{True\ Positive + True\ Negative}{TP + TN + FP + FN}$$

b. Precision

Precision mengukur proporsi prediksi positif yang benar terhadap semua prediksi.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

c. Recall

Recall mengukur seberapa baik model mengenali semua sampel positif yang sebenarnya.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

d. F1-Score

F1-Score adalah rata-rata harmonik dari *precision* dan *recall*.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

e. AUC-ROC

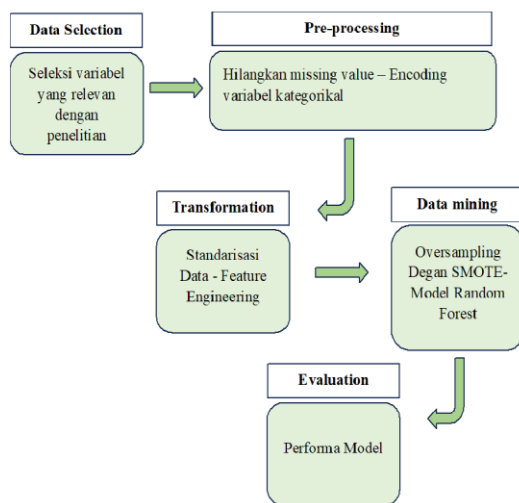
AUC-ROC mengukur kemampuan model dalam membedakan antara kelas positif dan negatif. Nilai AUC mendekati 1 menunjukkan performa model yang sangat baik.

3. METODE PENELITIAN

3.1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database* (KDD) untuk mengekstraksi pola dari data survei "*Mental Health in Tech*" yang diperoleh dari Kaggle. Model prediksi yang diterapkan dalam penelitian ini adalah *Random Forest*, yang merupakan algoritma *machine learning* berbasis *ensemble learning* yang efektif dalam

klasifikasi dan prediksi kesehatan mental berdasarkan faktor demografi dan lingkungan kerja.



Gambar 1. Metodologi Penelitian

3.2. Tahapan Metode KDD

Proses *Knowledge Discovery in Database* (KDD) dalam penelitian ini terdiri dari lima tahapan utama, yaitu:

3.2.1. Seleksi Data (Data Selection)

Pada tahap ini, hanya variabel yang relevan dengan kesehatan mental, demografi, dan lingkungan kerja yang dipilih. Tujuan dari seleksi ini adalah untuk menghindari noise dan memastikan bahwa data yang digunakan benar-benar berdampak pada prediksi kesehatan mental.

3.2.2. Pra-Pemrosesan (Pre-processing)

Pada tahap ini dilakukan beberapa teknik pemrosesan data untuk meningkatkan kualitas data sebelum diterapkan ke model machine learning:

- Mengatasi Missing Values: Data yang hilang diisi dengan metode modus (most frequent value imputation).
- Encoding Data Kategorikal: Variabel seperti jenis kelamin dan negara dikonversi ke dalam format numerik menggunakan Label Encoding.

3.2.3. Transformasi Data (Transformation Data)

Pada tahap ini dilakukan dua proses yaitu :

- Standarisasi Data Numerik: Variabel numerik seperti usia distandarisasi menggunakan *StandardScaler* untuk memastikan distribusi data yang lebih merata.
- Feature Engineering: Dimana pada tahap ini menambahkan kolom baru untuk mengindikasikan apakah seorang karyawan memiliki asuransi yang mencakup perawatan kesehatan mental dari perusahaan atau tidak.

3.2.4. Penambahan Data (Data Mining)

- Oversampling menggunakan SMOTE: untuk menangani ketidakseimbangan kelas dalam dataset.
- Model *Random Forest* : Tahap ini merupakan inti dari penelitian, di mana model *Random Forest Classifier* dilatih untuk memprediksi status kesehatan mental karyawan berdasarkan variabel demografi dan lingkungan kerja. Model ini menggunakan parameter `class_weight='balanced'` untuk menangani data yang tidak seimbang.

3.2.5. Evaluasi (Evaluation)

Performa model dievaluasi menggunakan beberapa metrik berikut:

- Akurasi: Persentase prediksi yang benar dibandingkan total data.
- Precision*: Seberapa tepat model dalam mengklasifikasikan individu yang benar-benar mengalami gangguan mental.
- Recall*: Seberapa baik model dapat mengenali individu yang mengalami gangguan mental.
- F1-Score*: Rata-rata harmonik antara *precision* dan *recall*, memberikan gambaran keseimbangan model dalam mengenali kedua kelas.
- AUC-ROC*: Mengukur kemampuan model dalam membedakan individu yang mengalami gangguan mental dan yang tidak.

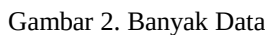
3.3. Algoritma Random Forest

Random Forest adalah algoritma berbasis *ensemble learning* yang terdiri dari banyak *decision tree*. Model ini bekerja dengan membangun beberapa pohon keputusan (decision trees) dari subset data yang berbeda, lalu menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir yang lebih akurat dan stabil [4].

Adapun beberapa keunggulan dari *random forest* adalah dapat mengurangi *overfitting* karena menggabungkan banyak pohon keputusan. Kemudian dapat menangani data dengan banyak variabel, baik numerik maupun kategorikal. Terakhir dapat menangani data yang tidak seimbang, dengan teknik seperti SMOTE (*Synthetic Minority Over-sampling Technique*).

4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi kesehatan mental karyawan di lingkungan kerja, dengan fokus pada faktor demografi dan lingkungan kerja. Dataset yang digunakan mencakup hasil survei mengenai kesehatan mental yang diambil antara tahun 2017 hingga 2022, dengan 83 kolom dan 2000 baris. Karena data untuk tahun 2023 sangat terbatas, hanya data hingga 2022 yang digunakan dalam penelitian ini.



4.7. Evaluasi Model

Accuracy: 0.9383647798742139

Classification Report:

	precision	recall	f1-score	support
0	0.89	0.99	0.94	383
1	0.99	0.89	0.94	412

accuracy			0.94	795
macro avg	0.94	0.94	0.94	795
weighted avg	0.94	0.94	0.94	795

Confusion Matrix:

```
[[381  2]
 [ 47 365]]
```

AUC-ROC: 0.9892868006793581
 Log-Loss: 0.11473276602167563
 Brier Score: 0.0378102476244762

Gambar 8. Evaluasi Model

Akurasi model pada penelitian ini mencapai 93.8%, berbeda dengan penelitian sebelumnya yang memiliki akurasi 99.5%. Meskipun demikian, model dalam penelitian ini lebih seimbang dan tidak mengalami bias terhadap kelas mayoritas. Pada penelitian sebelumnya, precision dan recall untuk kelas 0 (Terdiagnosis) bernilai 0%, yang menunjukkan bahwa model gagal mengenali kelas ini karena jumlah sampel yang terlalu kecil. Sebaliknya, dalam penelitian ini, precision mencapai 89% dan recall sebesar 99% untuk kelas 0, menunjukkan bahwa model mampu mengenali kelas tersebut dengan baik.

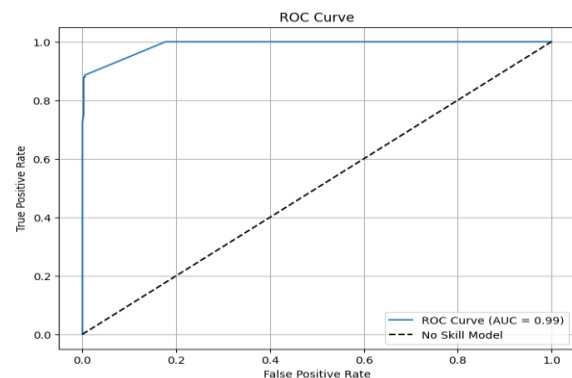
Selain itu, F1-Score dalam penelitian sebelumnya sangat rendah, sedangkan dalam penelitian ini mencapai 94%, yang menunjukkan keseimbangan antara kelas 0 (Terdiagnosis) dan kelas 1 (Tidak Terdiagnosis). Untuk AUC-ROC, model dalam penelitian ini mendapatkan skor 99%, yang berarti model mampu membedakan kedua kelas dengan cukup baik, tanpa bias terhadap kelas mayoritas. Pada metrik Macro Average, penelitian ini memperoleh nilai 94%, yang memberikan gambaran rata-rata performa model tanpa mempertimbangkan ketidakseimbangan data. Namun, nilai ini bisa memberikan gambaran yang kurang akurat jika dataset tidak seimbang. Sementara itu, Weighted Average juga mencapai 94%, yang menunjukkan konsistensi performa model karena mempertimbangkan distribusi data di setiap kelas.

Tabel 1. Tabel Perbandingan Sesudah dan Sebelum Menerapkan SMOTE

Metrik	Penelitian Ini	Penelitian Sebelum Menggunakan SMOTE
Akurasi	93.80%	99.50%
Precision (kelas 0)	89%	0%
Recall (kelas 0)	99%	0%
F1-Score	94%	Rendah
AUC-ROC	99%	Tidak disediakan

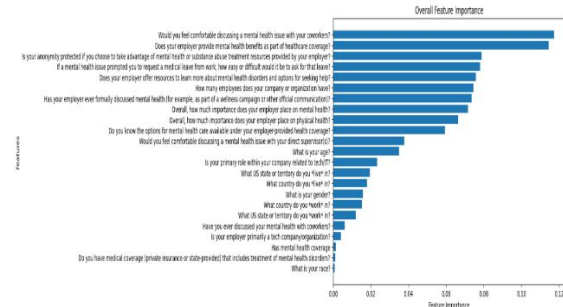
Hasil confusion matrix menunjukkan bahwa True Negative (TN) = 381, False Positive (FP) = 2, False Negative (FN) = 47, dan True Positive (TP) = 365. Hal ini menunjukkan bahwa setelah dilakukan oversampling menggunakan SMOTE, kemampuan

model dalam mengenali kelas 0 (Terdiagnosis) meningkat secara signifikan. Selain itu, model menjadi lebih adil dalam membedakan kedua kelas, sebagaimana dibuktikan oleh F1-Score yang seimbang sebesar 94% untuk kedua kelas.



Gambar 9. ROC Kurva

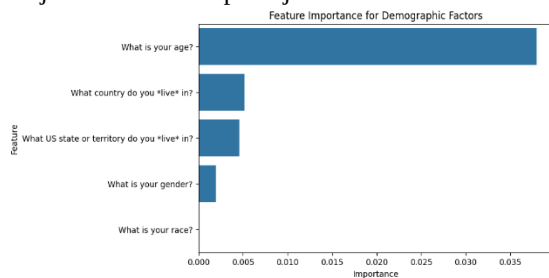
Kurva ROC menggambarkan kinerja model dalam membedakan antara dua kelas. Sumbu X menunjukkan *False Positive Rate* (kesalahan dalam memprediksi positif padahal seharusnya negatif), sementara sumbu Y menunjukkan *True Positive Rate* (kemampuan model memprediksi positif dengan benar). Kurva biru menggambarkan performa model pada berbagai ambang batas, sementara garis putus-putus menunjukkan model acak sebagai acuan. Semakin mendekati sudut kiri atas, semakin baik performa model. Dalam grafik, *False Positive Rate* yang sangat rendah dan *True Positive Rate* yang mendekati 1 menunjukkan bahwa model sangat akurat. AUC (*Area Under Curve*) yang mencapai 99% mengindikasikan bahwa model mampu membedakan kelas positif dan negatif dengan sangat baik.



Gambar 10. Feature Importance Keseluruhan

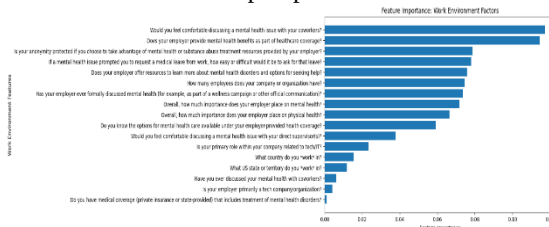
Pada grafik ini menunjukkan feature digunakan untuk mengidentifikasi faktor-faktor yang paling berpengaruh dalam memprediksi kesehatan mental karyawan berdasarkan faktor demografi dan lingkungan kerja. Hal ini membantu untuk memberikan pemahaman fitur mana yang paling berpengaruh dalam model. Pada grafik fitur dengan *importance* tertinggi adalah "Would you feel comfortable discussing a mental health issue with your coworkers?", diikuti oleh "Does your employer provide mental health benefits as part of healthcare coverage?". Temuan ini menunjukkan bahwa faktor lingkungan kerja, terutama kebijakan perusahaan dan

budaya organisasi terkait kesehatan mental, lebih berpengaruh dibanding faktor individu. Ini dapat menjadi dasar bagi penelitian lebih lanjut tentang bagaimana intervensi kebijakan dapat meningkatkan kesejahteraan mental pekerja.



Gambar 11. Feature Importance Berdasarkan Demografi

Pada grafik faktor demografi yang paling tinggi tingkat pentingnya dalam memprediksi (Kemungkinan diagnosis gangguan Kesehatan mental) pada model *random forest* adalah *What is your age ?*, Hal ini disebabkan oleh perbedaan tingkat stres, tanggung jawab, dan pengalaman hidup yang dialami oleh karyawan dari berbagai kelompok usia. Oleh karena itu perusahaan perlu memperhatikan kebutuhan kesehatan mental karyawan dari berbagai kelompok usia. Misalnya, karyawan yang lebih muda membutuhkan lebih banyak dukungan dalam menghadapi tekanan kerja, sementara karyawan yang lebih tua membutuhkan bantuan dalam mengelola stres terkait karir atau kehidupan pribadi.



Gambar 12. Feature Importance Lingkungan Kerja

Pada Gambar 10 menunjukkan *feature importance* atau kepentingan setiap fitur yang terkait dengan faktor lingkungan kerja dalam memprediksi diagnosis gangguan kesehatan mental menggunakan model *Random Forest*. Pada grafik menunjukan bahwa fitur "*Would you feel comfortable discussing a mental health issue with your coworkers?*" memiliki *importance* tertinggi. Hal ini menunjukkan bahwa kenyamanan dalam berdiskusi tentang kesehatan mental dengan rekan kerja merupakan faktor kunci dalam memprediksi risiko kesehatan mental. Oleh karena itu perusahaan perlu menciptakan lingkungan kerja yang mendukung dan inklusif, di mana karyawan merasa aman dan nyaman untuk membicarakan masalah kesehatan mental mereka. Hal ini dapat mengurangi stigma terkait kesehatan mental dan mendorong karyawan untuk mencari bantuan lebih awal jika diperlukan. Kemudian ada juga "*Does your employer provide mental health benefits as part of healthcare coverage?*" Fitur ini juga memiliki tingkat

kepentingan yang tinggi. Ketersediaan tunjangan kesehatan mental menunjukkan bahwa perusahaan peduli terhadap kesejahteraan mental karyawan. Hal ini menunjukkan bahwa Perusahaan yang menyediakan tunjangan kesehatan mental cenderung memiliki karyawan dengan tingkat kesejahteraan mental yang lebih baik. Ini menunjukkan pentingnya kebijakan perusahaan dalam mendukung kesehatan mental karyawan melalui program-program yang konkret.

Dibandingkan dengan penelitian sebelumnya, model dalam penelitian ini menunjukkan peningkatan performa yang signifikan. Khurana et al. (2022) menggunakan kombinasi Boosting, Random Forest, dan Logistic Regression, tetapi hanya mencapai akurasi 81.7% akibat ketidakseimbangan dataset dan metode preprocessing yang kurang optimal. Sementara itu, Mohamed et al. (2023) menggunakan model hibrida (Random Forest, SVM, dan MLP) dengan akurasi 97.68%, tetapi datasetnya sangat terbatas (hanya 215 sampel), sehingga berisiko mengalami overfitting dan kurang dapat digeneralisasi. Penelitian oleh Vaishnavi et al. (2022) menunjukkan bahwa model Random Forest memiliki akurasi 90%, tetapi penelitian ini tidak mempertimbangkan faktor lingkungan kerja sebagai variabel prediktor. Selain itu, penggunaan semua variabel dalam penelitian tersebut dapat menyebabkan noise dalam model, sehingga memengaruhi akurasi prediksi.

Adapun beberapa skenario pengujian dengan variasi data latih dan uji berbeda sebagai berikut :

Accuracy: 0.9362950544844929

Classification Report:

	precision	recall	f1-score	support
0	0.89	0.99	0.94	578
1	0.99	0.88	0.93	615
accuracy			0.94	1193
macro avg	0.94	0.94	0.94	1193
weighted avg	0.94	0.94	0.94	1193

Confusion Matrix:

```
[[575  3]
 [ 73 542]]
```

Gambar 13. Data latih 70% Data Uji 30%

Accuracy: 0.9352201257861635

Classification Report:

	precision	recall	f1-score	support
0	0.89	0.99	0.94	778
1	0.99	0.88	0.93	812
accuracy			0.94	1590
macro avg	0.94	0.94	0.94	1590
weighted avg	0.94	0.94	0.94	1590

Confusion Matrix:

```
[[774  4]
 [ 99 713]]
```

Gambar 14. Data Latih 60% Data Uji 40%

Secara keseluruhan, performa model cukup stabil dalam berbagai skenario, tetapi perlu peningkatan terutama dalam recall kelas 1 jika prediksi untuk kelas tersebut lebih kritis. Berdasarkan hasil prediksi yang ditampilkan, model menunjukkan akurasi yang cukup tinggi, berkisar antara 93.5% hingga 93.6%, yang

menandakan bahwa model tetap stabil meskipun dilakukan dengan scenario data latih dan data uji berbeda.

5. KESIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa kesehatan mental karyawan dipengaruhi oleh faktor demografi dan lingkungan kerja, Dimana faktor usia dan kenyamanan dalam membahas isu kesehatan mental dengan rekan kerja, serta ketersediaan tunjangan kesehatan mental dari perusahaan merupakan faktor yang paling signifikan. Model Random Forest yang digunakan dalam penelitian ini menunjukkan performa yang baik, dengan akurasi sebesar 93% dan AUC 99%, yang menunjukkan bahwa model mampu membedakan individu yang mengalami gangguan kesehatan mental dengan tingkat kesalahan yang sangat rendah. Namun, penelitian ini memiliki keterbatasan, seperti penggunaan dataset yang hanya berfokus pada industri teknologi, sifat data yang self-reported, serta kurangnya perbandingan dengan model *machine learning* lainnya seperti Gradient Boosting atau Deep Learning.

Oleh karena itu, untuk penelitian selanjutnya disarankan menggunakan dataset dari berbagai sektor industri guna meningkatkan generalisasi model, serta membandingkan model ini dengan teknik *machine learning* lainnya. Selain itu, perusahaan disarankan untuk meningkatkan kesadaran kesehatan mental melalui seminar, pelatihan mindfulness, serta program employee assistance guna mengurangi stigma dan mendorong karyawan untuk mencari bantuan lebih awal. Penyediaan layanan kesehatan mental di tempat kerja, seperti konsultasi psikologis gratis atau subsidi pengobatan, juga menjadi langkah penting dalam mengurangi risiko gangguan kesehatan mental di lingkungan kerja.

DAFTAR PUSTAKA

- [1] K. Vaishnavi, U. N. Kamath, B. A. Rao, dan N. V. S. Reddy, "Predicting Mental Health Illness using Machine Learning Algorithms," *J. Phys. Conf. Ser.*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012021.
- [2] F. Firmansyah dan A. Yulianto, "Pemodelan Pembelajaran Mesin untuk Prediksi Kesehatan Mental di Tempat Kerja," *J. Minfo Polgan*, vol. 13, no. 1, hal. 397–407, 2024, doi: 10.33395/jmp.v13i1.13674.
- [3] O. Nurdian, D. Sudrajat, dan A. R. Rinaldi, "Bibliometrik Analysis : Signal Preprocessing Techniques for Kualitas Sinyal Electrogram," hal. 538–545.
- [4] J. Chung dan J. Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, 2022, doi: 10.1155/2022/9970363.
- [5] S. A. Dicky Miftakhul Rizki, Odi Nurdian, "ANALISA PEMBELIAN SEPEDA MENGGUNAKAN ALGORITMA APRIORI PADA TOKO SEPEDA BRADEN BIKE."
- [6] E. S. Mohamed, T. A. Naqishbandi, S. A. C. Bukhari, I. Rauf, V. Sawrikar, dan A. Hussain, "A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms," *Healthc. Anal.*, vol. 3, no. March, hal. 100185, 2023, doi: 10.1016/j.health.2023.100185.
- [7] H. Alkahtani, T. H. H. Aldhyani, dan A. A. Alqarni, "Artificial Intelligence Models to Predict Disability for Mental Health Disorders," *J. Disabil. Res.*, vol. 3, no. 3, hal. 1–12, 2024, doi: 10.57197/jdr-2024-0022.
- [8] D. Solihudin, O. Nurdian, R. Kurniawan, dan C. L. Rohmat, "OPTIMIZATION IOT TECHNOLOGY IN WEATHER STATIONS FOR IMPROVE AGRICULTURAL SUCCESS DURING EL NIÑO ERA," vol. 10, no. 2, hal. 315–323, 2024, doi: 10.33480/jitk.v10i2.5851.OPTIMIZATION.
- [9] O. D. Adeniji, S. O. Adeyemi, dan S. A. Ajagbe, "An Improved Bagging Ensemble in Predicting Mental Disorder Using Hybridized Random Forest - Artificial Neural Network Model," *Inform.*, vol. 46, no. 4, hal. 543–550, 2022, doi: 10.31449/inf.v46i4.3916.
- [10] Y. Khurana, S. Jindal, H. Gunwant, dan D. Gupta, "Mental Health Prognosis Using Machine Learning," *SSRN Electron. J.*, no. 1, 2022, doi: 10.2139/ssrn.4060009.
- [11] N. S. Alghamdi, H. A. Hosni Mahmoud, A. Abraham, S. A. Alanazi, dan L. García-Hernández, "Predicting depression symptoms in an Arabic psychological forum," *IEEE Access*, vol. 8, hal. 57317–57334, 2020, doi: 10.1109/ACCESS.2020.2981834.
- [12] K. Saha, A. Yousuf, R. L. Boyd, J. W. Pennebaker, dan M. De Choudhury, "Social Media Discussions Predict Mental Health Consultations on College Campuses," *Sci. Rep.*, vol. 12, no. 1, hal. 1–11, 2022, doi: 10.1038/s41598-021-03423-4.
- [13] A. Meylani, E. S. Negara, U. B. Darma, dan S. Selatan, "Aplikasi Prediksi Kesehatan Menggunakan," vol. 14, no. 2, hal. 208–215, 2022.
- [14] D. A. Adler, F. Wang, D. C. Mohr, dan T. Choudhury, "Machine learning for passive mental health symptom prediction: Generalization across different longitudinal mobile sensing studies," *PLoS One*, vol. 17, no. 4 April, hal. 1–20, 2022, doi: 10.1371/journal.pone.0266516.
- [15] M. Pandey, D. Parmar, S. Mishra, dan V. Pinjarkar, "Mental Health Prediction for Juvenile Using Machine Learning Techniques," *SSRN Electron. J.*, hal. 1–10, 2021, doi: 10.2139/ssrn.3867291.