# Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest

**Emia Rosta Br. Sebayang [1], Yulison Herry Chrisnanto [2], Melina [3]**
Universitas Jenderal Achmad Yani Cimahi [1.2.3]
Jl. Terusan Jend. Sudirman, Cibeber, Cimahi Selatan, Cimahi, Jawa Barat 40531
Sur-el : Emia.rosta@student.unjani.ac.id [1], yhc.if@unjani.ac.id[2], melina@lecture.unjani.ac.id[3]

***Abstract :*** *Mental health is an integral part of human well-being. Mental health disorders can affect individuals in various aspects of life. Work pressure, heavy workload, and an unhealthy lifestyle can be the main causes of mental health disorders in the workplace, such as industrial technology. Employees' mental health problems in the workplace often do not receive enough attention because they cannot be seen physically. Mental health has a significant impact on the performance that will be shown by employees in contributing to the company, it requires the company's prudence and sensitivity in observing and understanding the mental health conditions of employees. In this study, the Open Source Mental Illness (OSMI) survey data was classified using the Random Forest algorithm with the ensemble method, as well as the bootstrap tree method to improve the performance of the Random Forest algorithm in determining the accuracy of mental health data. The Random Forest algorithm is an ensemble learning method that combines several decision trees to improve prediction accuracy. Classification is carried out using a bootstrap tree which takes training data to train a model or ensemble so that it can take patterns and relations from the data to carry out classification, the Random Forest algorithm is an ensemble learning method that combines several decision trees for research with 80% training data and 20 test data %. The results of this study indicate a fairly good level of accuracy, which is 84%, so that it can make an important contribution in understanding the level of mental health disorders experienced by technology industry employees. The expected results of this research can improve the quality of life and productivity of employees at work.*

***Keywords:*** *classification, machine learning, mental health, random forest, technology industry*

***Abstrak :*** *Kesehatan mental adalah bagian yang tak terpisahkan dari kesejahteraan manusia. Gangguan kesehatan mental dapat mempengaruhi individu dalam berbagai aspek kehidupan. Tekanan kerja, beban kerja yang berat, dan gaya hidup yang tidak sehat dapat menjadi penyebab utama adanya gangguan kesehatan mental di lingkungan tempat kerja, seperti industri teknologi. Masalah kesehatan mental karyawan di tempat kerja sering kali tidak mendapat perhatian yang cukup karena tidak dapat terlihat secara fisik. Kesehatan mental memiliki dampak signifikan terhadap kinerja yang akan ditunjukkan oleh karyawan dalam memberikan kontribusi bagi perusahaan dibutuhkan ketelitian dan kepekaan perusahaan dalam mengamati dan memahami kondisi kesehatan mental karyawan. Pada penelitian ini, data survey Open Source Mental Illness (OSMI) diklasifikasikan dengan menggunakan algoritma Random Forest dengan metode ensemble, serta metode bootstrap tree untuk meningkatkan performa dari algoritma Random Forest dalam menentukan akurasi data kesehatan mental. Algoritma Random Forest merupakan metode ensemble learning yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi. Klasifikasi dilakukan menggunakan bootstrap tree yang mengambil data yang di latih untuk melatih model atau ensemble sehingga dapat mengambil pola dan relasi dari data tersebut untuk melakukan klasifikasi, algoritma Random Forest merupakan metode ensemble learning yang menggabungkan beberapa pohon keputusan untuk penelitian dengan data latih 80% dan data uji 20%. Hasil penelitian ini menunjukkan tingkat akurasi yang cukup baik, yaitu 84% sehingga dapat memberikan kontribusi penting dalam memahami tingkat gangguan kesehatan mental yang dialami oleh karyawan industri teknologi. Diharapkan hasil penelitian ini dapat meningkatkan kualitas hidup dan produktivitas karyawan dalam bekerja.*

***Kata kunci****: algoritma random forest, industri teknologi, klasifikasi, kesehatan mental, machine learning.*

## 1. PENDAHULUAN

Kemajuan teknologi informasi yang pesat di berbagai bidang kehidupan sejalan dengan peningkatan volume data yang dihasilkan, termasuk di industri, kesehatan dan bidang lainnya [1]. Perkembangan ini juga membawa tantangan baru, terutama dalam hal kesehatan mental karyawan di industri teknologi. Tekanan kerja yang tinggi, tengat waktu yang ketat, dan lingkungan kerja yang kompetitif dapat seringkali menyebabkan stres dan ketidakseimbangan mental [2]. Masalah kesehatan mental karyawan di tempat kerja sering kali tidak mendapat perhatian yang cukup karena tidak dapat terlihat secara fisik. Oleh karena itu, dibutuhkan perhatian dari perusahaan dalam mengamati dan memahami kondisi kesehatan mental karyawan. Kesehatan mental memiliki dampak signifikan terhadap kinerja yang akan ditunjukkan oleh karyawan dalam memberikan kontribusi bagi perusahaan [3].

Menurut *World Health Organization (WHO)*, kesehatan mental mengacu pada kondisi sejahtera yang disadari oleh individu, yang memiliki kemampuan dalam mengelola stres kehidupan secara tepat, dan dapat bekerja dengan produktif, serta berperan serta aktif dalam masyarakat [4]. Dalam industri teknologi itu sendiri, terdapat defisiensi pada pengertian mendasar mengenai seberapa pentingnya masalah kesehatan mental pada lingkungan kerja industri teknologi tersebut. Tidak sedikit perusahaan yang kurang memperhatikan efek yang timbul pada pekerja tentang kondisi kesehatan mental yang dimiliki oleh pekerja. Hal ini disadari oleh *Open Source Mental Illness (OSMI)*, suatu perusahaan *non-profit* yang bergerak dalam bidang kesehatan, khususnya kesehatan mental, pada industri dan komunitas teknologi menjalankan survei mengenai kesadaran kesehatan mental pada pekerja dalam industri teknologi pada tahun 2014 sampai tahun 2015 [5].

Algoritma *Random Forest* merupakan salah satu algoritma klasifikasi yang paling cocok untuk digunakan dalam diagnosis kesalahan. Algoritma ini berfungsi dengan cepat bahkan dalam kumpulan data yang besar yang memiliki kelemahan menghasilkan *overfitting* pada data yang tidak seimbang [6]. Metode dengan menggunakan *bootstrap tree* dapat meningkatkan kemampuan *Random Forest* dalam mengklasifikasi data kesehatan mental di industri teknologi. Kemampuan *bootstrap tree* dapat mengatasi overfitting pada data yang tidak seimbang untuk diklasifikasi dapat memberikan hasil yang baik untuk data yang hilang [7]. *Random Forest* telah terbukti efektif dalam banyak aplikasi, akan tetapi masih ada ruang untuk meningkatkan akurasi prediksi yang dihasilkan oleh algoritma ini. Terdapat beberapa tantangan yang dapat dihadapi, seperti pemilihan parameter yang optimasi, pengelolaan data yang tidak seimbang, dan penangan fitur yang tidak relevan [8].

Beberapa penelitian terdahulu yang mengkaji tentang *Random Forest* C*lassification* yaitu penelitian oleh Wakiru, dkk. (2018), menyatakan bahwa *Random Forest* menawarkan akurasi prediksi yang tinggi dan mampu mengidentifikasi atau menilai pengaruh variabel terhadap hasil atau prediksi yang tidak memiliki nilai yang hilang. Pemilihan variabel menggunakan *Random Forest* menemukan variabel-variabel yang penting dan dapat diterima untuk klasifikasi berdasarkan kumpulan data. Pohon keputusan dimodelkan dan diatur untuk meningkatkan kekuatan prediktifnya dari 96,61% menjadi 97,53% sehingga meningkatkan skor

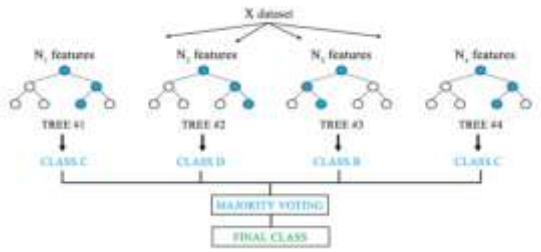
keluaran klasifikasi menuju akurasi tinggi sehingga keandalannya tinggi [9]. Kemudian, penelitian yang dilakukan oleh Widya Apriliah dkk (2020) yang menggunakan tiga algoritma klasifikasi *machine learning* yaitu *Suport Vector Machine*, *Naïve Bayes* dan *Random Forest.* Hasil penelitiannya menunjukkan kecukupan sistem yang dirancang dengan menggunakan algoritma *Random Forest* mencapai sebesar 97,88%. Ini menyatakan algoritma *Random Forest* telah bekerja dengan akurasi terbaik [10]. Penelitian yang dilakukan oleh Nugraha dkk (2020), yang dapat memberikan solusi untuk penentuan kelayakan pemberian kredit dengan hasil pengujian dengan algoritma klasifikasi *Random Forest* mampu menganalisis kredit yang bermasalah dan yang debitur yang tidak bermasalah dan yang debitur yang tidak bermasalah dengan nilai akurasi sebesar 87,88% [11]. Selanjutnya, penelitian oleh Sarifah dkk (2022) yang mengklasifikasi penyakit daun padi menggunakan *Random Forest* dan Color Histogram" yang mencapai akurasi sebesar 99,65% dari metode yang diusulkan dalam penelitian tersebut. Hasil penelitiannya menunjukkan bahwa penggunaan metode *Random Forest* dan Color Histogram efektif dalam mengklasifikasi penyakit daun pada tanaman padi dengan tingkat akurasi yang sangat tinggi [12].

Berdasarkan latar belakang, penelitian ini menggunakan data survei kesehatan mental di industri teknologi menggunakan algoritma *Random Forest* dengan data yang bersifat *imbalanced,* berbeda dengan penelitian sebelumnya yang menggunakan data kesehatan mental dari psikologi yang bersifat *balanced.* Penelitian ini menggabungkan variasi dalam pembentukan pohon keputusan, penggabungan hasil prediksi dari setiap pohon, dan pemilihan fitur yang lebih optimal dengan menggunakan teknik *ensemble bootstrap tree* untuk meningkatkan akurasi prediksi. Diharapkan penelitian ini dapat memberikan kontribusi dalam mengembangkan metode *ensemble bootstrap tree* untuk meningkatkan akurasi prediksi menggunakan algoritma *Random Forest.* Akurasi memprediksi keluaran dengan akurasi tinggi, bahkan kumpulan data besar akan mendapatkan akurasi yang baik [6].

## 2. METODOLOGI PENELITIAN

### 2.1 *Random Forest*

*Random Forest* adalah salah satu algoritma klasifikasi dan regresi yang popular dalam *data mining* dan *machine learning*. *Random forest* menggabungkan konsep dari metode *ensemble* dan *decision tree* [13]. *Random forest* adalah suatu algoritma yang digunakan pada klasifikasi data jumlah besar. Klasifikasi *Random Forest* dilakukan mealui penggabungan pohon dengan melakukan *training* yang dilakukan membangkitkan pohon klasifikasi dengan banyak versi yang kemudian mengkombinasikannya untuk memperoleh prediksi akhir, maka dalam *Random Forest* proses pengacakan untuk membentuk pohon klasifikasi tidak hanya dilakukan untuk data sampel saja melainkan juga pada pengambilan variabel prediktor [14].




Gambar 2. Pohon Random Forest

## 2.2 Proses Klasifikasi menggunakan Random Forest

Sehingga, proses ini akan menghasilkan kumpulan pohon klasifikasi dengan ukuran dan bentuk yang berbeda-beda. Hasil yang diharapkan adalah suatu kumpulan pohon klasifikasi yang memiliki korelasi kecil antar pohon. Korelasi yang kecil akan menurunkan hasil kesalahan prediksi *Random Forest* [15]. Berikut langkah-langkah algoritma yang dikerjakan dalam penggunaan algoritma *Random Forest* yaitu [16]:

1. Atribut data set yang digunakan menggunakan 15 atribut data yang telah melalui data selection, data cleaning data transformasi. Nama-nama atribut yang digunakan yaitu :
   *Age, Gender, Family History, Treatment, Work Interfere, Remote Work, Tech Company, Care Options, Wellness program, Seek Help, Anonymity, Mental Health Consequence, Supervisor, Mental health Interview, Phys Health Interview.*
2. Melakukan *bootstrap sampling* untuk mengambil sampel acak dengan penggantian dari *Dataset* pelatihan sehingga menghasilkan beberapa *Dataset* pelatihan yang berbeda untuk digunakan dalam membangun setiap *decision tree.*
3. Membangun *decision tree* pada setiap *Dataset* pelatihan yang dihasilkan dari *bootstrap tree*.
4. Menggabungkan prediksi *decision tree* yang dapat dilakukan dengan menggunakan mayoritas suara dari prediksi *decision tree*.
5. Mengevaluasi performa model *Random Forest* menggunakan metrik yang relevan, seperti *akurasi, presisi, recall*, atau *MSE (Mean Squared Error) untuk regresi*. *Dataset* validasi atau metode validasi silang (*cross-validation*) untuk menguji kinerja model.
6. Menyesuaikan *hyperparameter* yang umum seperti *decision tree* (*n_estimators*), jumlah fitur yang dipertimbangkan di setiap simpul (*max_features*), dan kedalaman maksimum decision tree (*max_depth*).
7. Memprediksi pada data baru dengan mengambil mayoritas suara (klasifikasi) dari semua *decision tree*.
8. Mengintrepretasikan hasil *Random Forest* untuk mendapatkan wawasan tentang faktor-faktor penting dalam prediksi atau hubungan antar variabel.

Rumus perhitungan *IndexGini* dan *GiniGain* sebagai berikut :

$$IndexGini(S) = 1 - \sum_{i=1}^{k}(Pi^2) \quad (1)$$

$$GiniGain = Gini(\text{A,S}) - \sum_{k=1}^{k} \frac{|Si|}{|S|} xGini(Si) \quad (2)$$

Dimana

$S$ : himpunan gugus data latih

$K$ : Banyak kelas

$P$i : Peluang S dari kelas ke-i

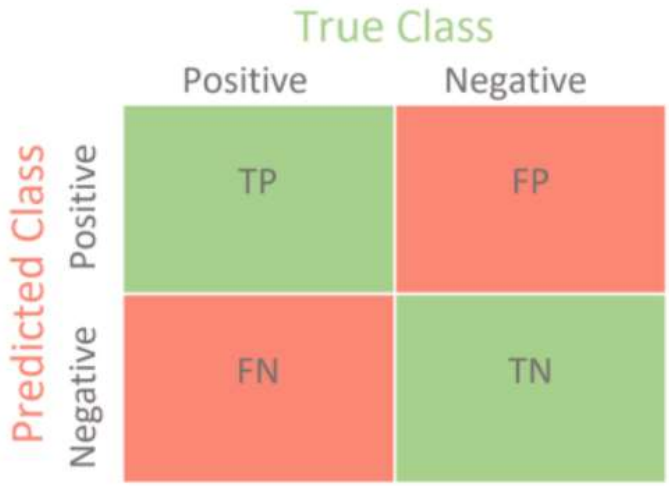
$\boldsymbol{Si}$ **:** Partisi S yang disebabkan oleh peubah AS

### 2.3 Pengujian Model

Pengujian yang dilakukan pada model dilakukan untuk mengevaluasi dan mengevaluasi kinerja algoritma yang telah dilatih. Pada penelitian ini, pengujian model menggunakan *confusion matriks*. Tabel ini, juga dikenal sebagai *confusion matriks*, digunakan untuk menganalisis kinerja model klasifikasi dalam memprediksi kelas target. Metode ini menunjukkan kemampuan model untuk mengklasifikasikan data dengan akurat serta kesalahan yang dibuat selama proses [17].

Empat sel elemen penting dalam matriks confusion membantu mengevaluasi kinerja model:

- *True Positive* (TP) adalah jumlah kasus yang diklasifikasikan dengan benar sebagai positif oleh model.
- *True Negative* (TN) adalah jumlah kasus yang diklasifikasikan dengan benar sebagai negatif oleh model.
- *False Positive* (FP) adalah jumlah kasus yang diklasifikasikan dengan salah sebagai positif oleh model, meskipun sebenarnya negatif.
- *False Negative* (FN) adalah jumlah kasus yang diklasifikasikan dengan salah sebagai negatif oleh model walaupun sebenarnya negatif.




Gambar 2. 1 *Confusion Matrix*

- *Accuracy* megukur seberapa sering model memberikan prediksi yang benar secara keseluruhan.Adapaun rumus dari *accuracy* yaitu :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

- *Precision* mengukur sejauh mana model memberikan hasil positif yang benar dari semua hasil positif yang diberikan. Adapun rumus dari *Precision* yaitu :

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

- *Recall* atau *Sensitivity* mengukur seberapa baik model dapat mengidentifikasi kasus positif yang sebenarnya. Adapun rumusan *Recall* atau *Sensitivity* yaitu:

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

- *F1-Score* menggambarkan perbandingan *precision* dan *recall*. Jika kumpulan data berisi jumlah data *False Negative* dan *False Positive* yang sangat mirip, namun jika angkanya tidak dekat, maka menggunakan *F1-Score* sebagai referensi. Adapun rumusan *F1-Score*:

$$F1 - Score = \frac{1}{\frac{1}{recall}+\frac{1}{precisiom}} \quad (6)$$

## 3. HASIL DAN PEMBAHASAN

Tabel 3. 1 Sampel Data

| ***Gender*** | ***Family history*** | ***Treatment*** | ***Anonymity*** | ***Mental Health Consequence*** | ***…*** | ***Phys health Interview*** |
|---|---|---|---|---|---|---|
| *Female* | *No* | *Yes* | *Yes* | *No* | *…* | *Maybe* |
| *Female* | *No* | *Yes* | *Yes* | *No* | *…* | *Maybe* |
| *Male* | *Yes* | *Yes* | *No* | *Yes* | *…* | *Maybe* |
| *Male* | *No* | *No* | *Don't know* | *No* | *…* | *Yes* |
| *Male* | *Yes* | *No* | *Don't know* | *No* | *…* | *Maybe* |
| *Female* | *Yes* | *Yes* | *No* | *Maybe* | *…* | *No* |
| *Male* | *No* | *No* | *Yes* | *No* | *…* | *No* |
| *…* | *…* | *…* | *…* | *…* | *…* | *…* |
| *Female* | *No* | *No* | *Don't know* | *Yes* | *…* | *No* |

Berdasarkan Tabel 3.6 Selection Atribut Survei Data Kesehatan Mental Di Industri Kesehatan Mental di atas yang digunakan berjumlah 15 (Lima Belas) atribut, yaitu :

1. *Age*
2. *Gender*
3. *Family History*
4. *Treatment*

5. *Work Interfere*
6. *Remote Work*
7. *Tech Company*
8. *Care Option*
9. *Wellness Program*
10. *Seek Help*
11. *Anonymity*
12. *Mental Health Consequence*
13. *Supervisor*
14. *Mental Health Inteview*
15. *Phys Health Interview*

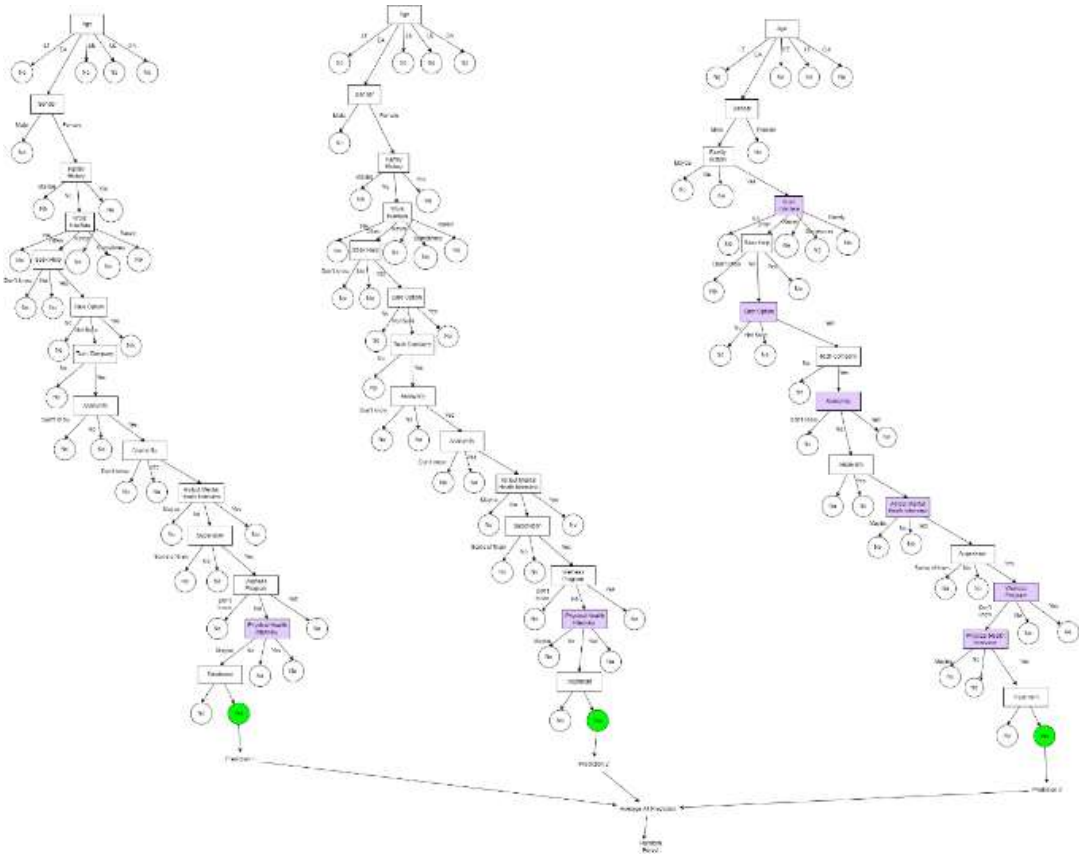
Pemilihan data kolom berdasarkan studi Pustaka yang telah peneliti lakukan sebelumnya menggunakan 10 atribut kolom [18], sedangkan pada Penelitan ini menggunakan 15 Atribut. Alasan peneliti memilih lima belas atribut yang sangat relevan pada penelitian ini yaitu kolom *Age* digunakan sebagai variabel karakteristik karena telah dilakukan penelitian tentang rentang umur dan kesehatan mental, *Gender* dimasukkan sebagai variabel karakteristik karena telah ditemukan hubungan antara jenis kelamin dan kesadaran kesehatan mental, *Family History* (Riwayat Keluarga) Riwayat keluarga tentang kesehatan mental dapat berfungsi sebagai faktor risiko atau prediktor masalah kesehatan mental pada seseorang, *Treatment* (Pengobatan) informasi apakah seseorang telah menerima perawatan kesehatan mental atau tidak, dan Jenis Kelamin dimasukkan sebagai variabel *Tech Company* Teknologi Perusahaan (Perusahaan Teknologi) Industri perusahaan dapat memiliki budaya dan lingkungan yang berbeda, yang dapat mempengaruhi kesehatan mental karyawan. *Care Option* (Opsi Perawatan) Informasi tentang opsi perawatan yang tersedia bagi individu dapat memberi tahu mereka tentang kemungkinan langkah pengobatan, *Wellness Program* (Program Kesejahteraan) Kehadiran program kesejahteraan dalam perusahaan dapat mempengaruhi kesehatan mental karyawan, *Seek Help* (mencari bantuan), *Mental Health Consequence* (Konsekuensi Kesehatan Mental) Informasi tentang akibat yang dialami seseorang karena masalah kesehatan mental dapat memberikan gambaran tentang seberapa parah masalah tersebut. *Supervisor* (Atasan) Hubungan dengan atasan dapat mempengaruhi dukungan dan lingkungan kerja yang mempengaruhi kesehatan mental. *Mental Health Interview* (Wawancara Kesehatan Mental) Faktor-faktor yang muncul dalam wawancara kesehatan mental dapat mengungkapkan lebih banyak tentang masalah kesehatan mental yang dihadapi oleh seseorang. Wawancara kesehatan fisik dapat menunjukkan kesehatan mental juga [19], [20]. Variabel label dari penelitian ini adalah *Treatment* karena untuk mencerminkan kelompok data yang

memprediksi pemahaman tingkat gangguan kesehatan mental yang dialami oleh karyawan industri teknologi apakah sudahkah pernah mencari pengobatan untuk kondisi kesehatan mental ("*Yes*" atau "*No*") [18], karena menurut *Open Source Mental Illness* (OSMI) di industri teknologi sangat mengabaikan kesehatan mental karyawan.

### 3.1 Menghitung *Gini Index* Semua Atribut

Untuk menghitung *Index Gini* (*Gini Index*) dari seluruh atribut dalam *Dataset*, perlu adanayan mengikuti rumus yang sesuai. Rumus *Index Gini* dihitung untuk setiap atribut dan digunakan sebagai metrik untuk memilih atribut terbaik yang akan menjadi pemisah dalam pembentukan pohon keputusan. Adapun rumusannya sebagai berikut:

$$IndexGini(S) = 1 - \sum_{i=1}^{k} pi^2 \quad (7)$$



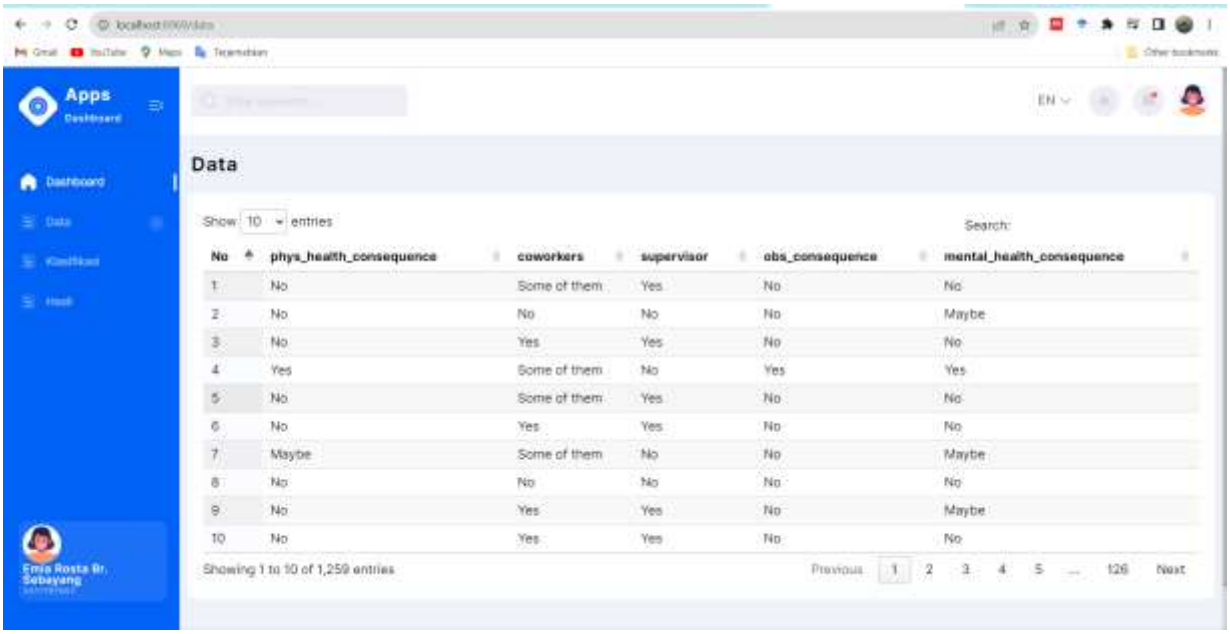
Gambar 3. 1 Pohon Keputusan

### 3.2 Hasil Evaluasi dan Analisa

1. Halaman Memasukkan Dataset

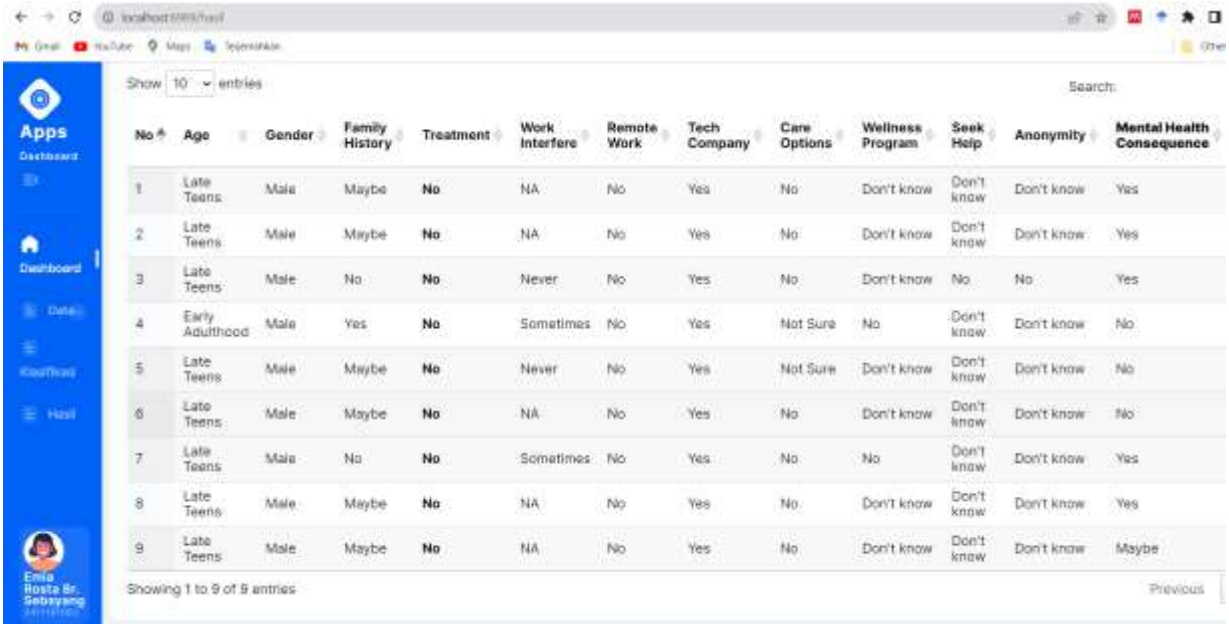

Gambar 3.1 Halaman Memasukkan Dataset

2. Halaman Menampilkan Dataset




Gambar 3.2 Halaman Menampilkan Dataset

3. Halaman Klasifikasi Data Kesehatan Mental

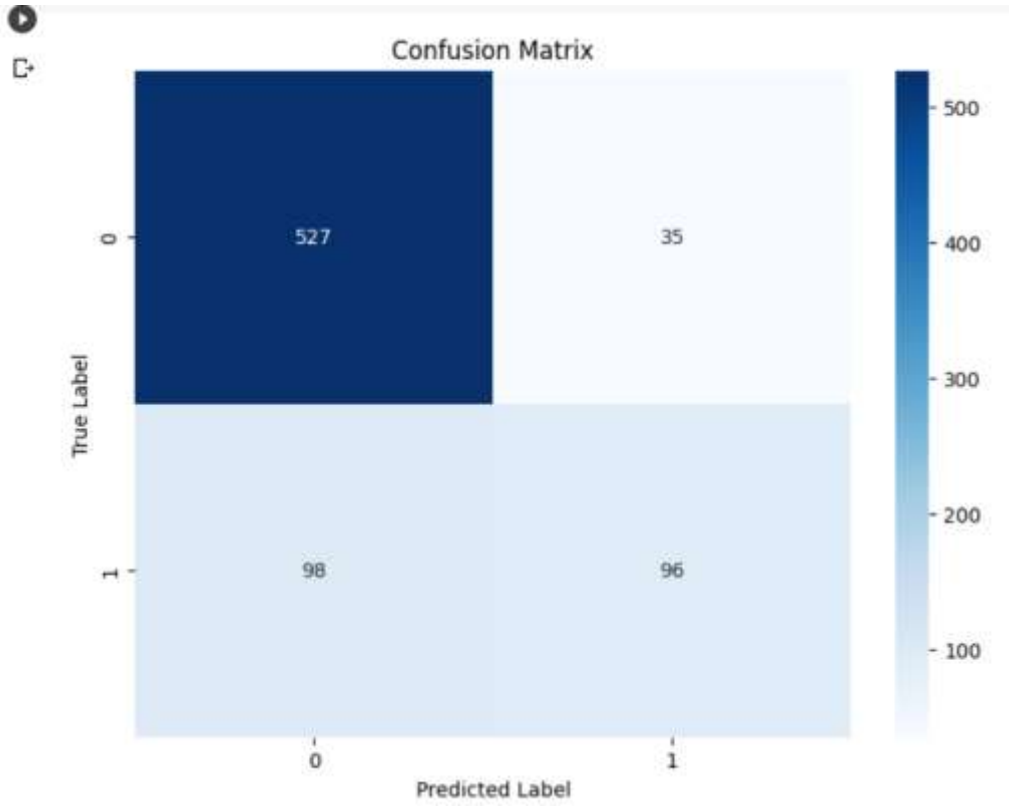



Gambar 3.3. Halaman Klasifikasi Data Kesehatan Mental

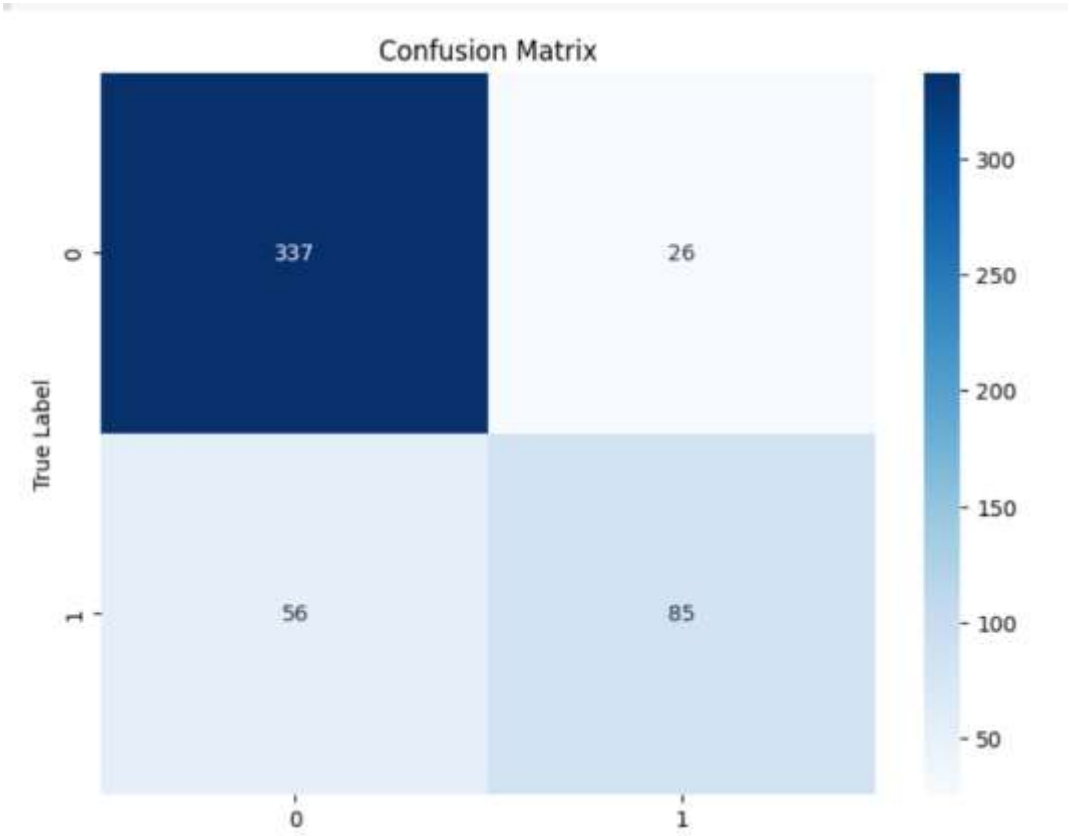
## 4. Pengujian Model

### *4.1 Skenario Dataset Rasio* 60% dan 40%

Skenario pengujian model pertama yaitu *split* data untuk data latih 60% dan data uji 40%. Hasil *accuracy* yang didapatkan 0,8293 Hasil *confusion matrix* dapat dilihat pada Gambar 4.5 dan Gambar 4.6




Gambar 4. 1 *Confusion Matrix data latih*




Gambar 4. 2 *Confusion Matrix data uji*

*Tabel 4.1* Eksperimen Model 1 *Confuxion Matrix* Data latih (60%)

| | **Aktual Positif (1)** | **Aktual Negatif (0)** |
|---|---|---|
| **Prediksi Positif (1)** | TP : 527 | FP : 35 |
| **Prediksi Negatif (0)** | FN : 98 | TN : 96 |

*Tabel 4.2* Eksperimen Model 1 *Confuxion Matrix* Data uji (40%)

| | **Aktual Positif (1)** | **Aktual Negatif (0)** |
|---|---|---|
| **Prediksi Positif (1)** | TP : 337 | FP : 26 |
| **Prediksi Negatif (0)** | FN : 56 | TN : 85 |

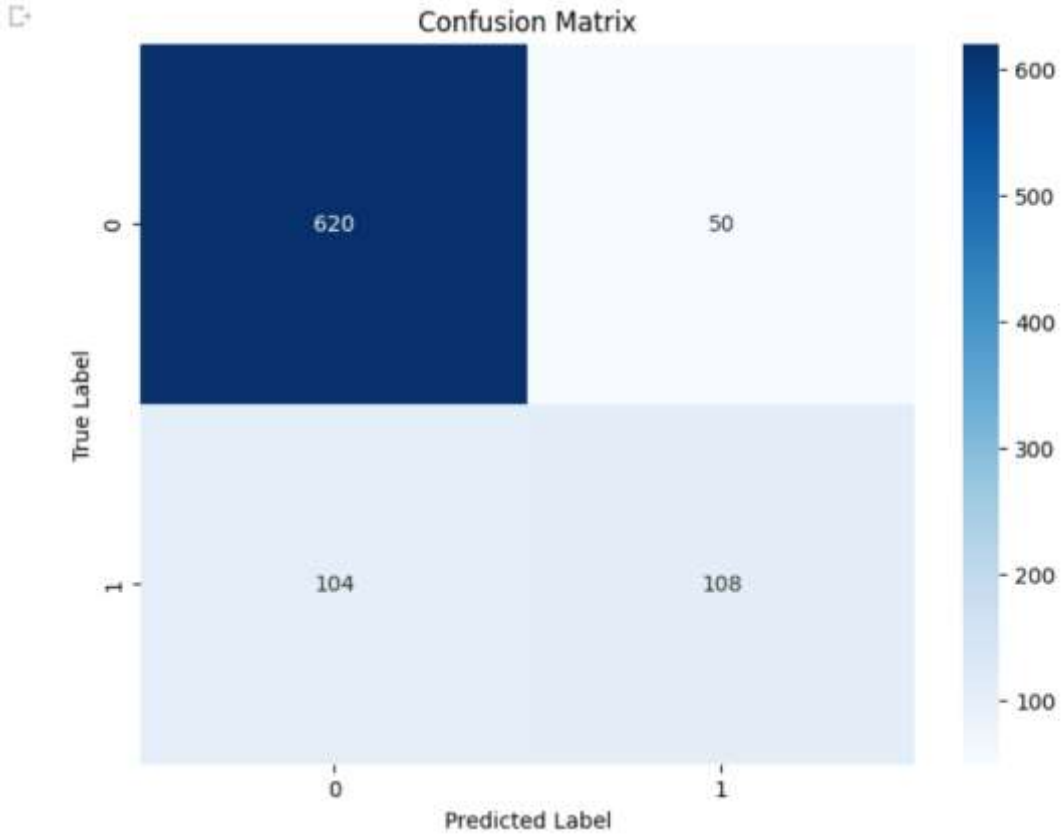
Tabel 4.3 Skenario Pengujian Model 1

| **Skenario Pembagian Data Latih 60% dan Data Uji 40%** | | |
|---|---|---|
| **Nilai** | **Data Latih** | **Data Uji** |
| **Accuracy** | 82,40% | 72,02% |
| **Precision** | 93,77% | 92,83% |
| **Recall** | 83,91% | 85,75% |
| **F1-Score** | 88,56% | 89,14% |

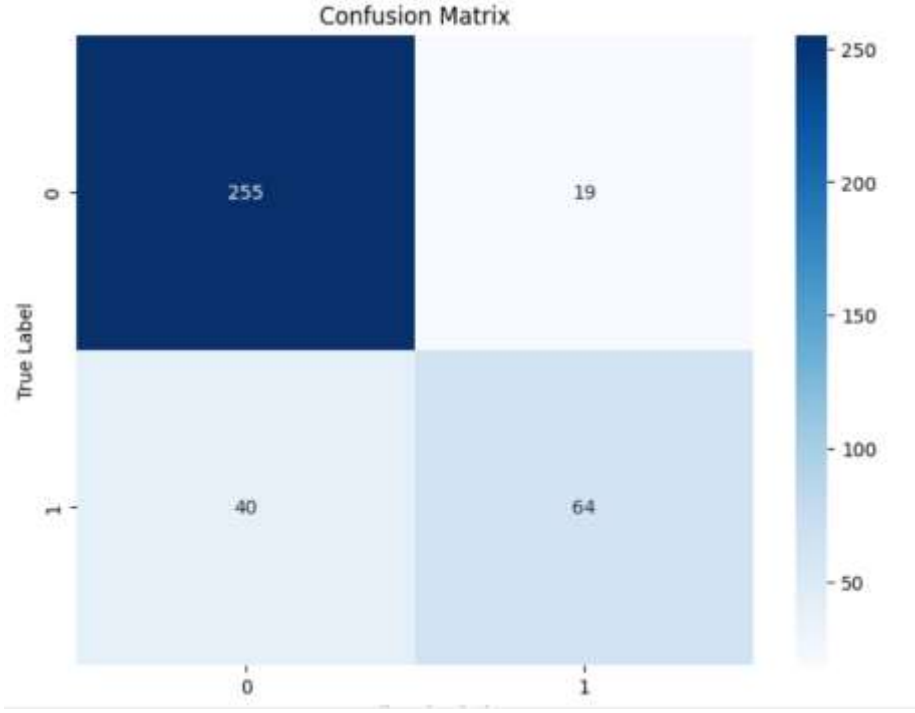
### *4.2 Skenario Dataset Rasio* 70% dan 30%

Skenario pengujian model pertama yaitu *split* data untuk data latih 70% dan data uji 30%. Hasil *accuracy* yang didapatkan 0,8412. Hasil *confusion matrix* dapat dilihat pada Gambar 4.7 dan Gambar 4.8

*Gambar 4. 1 Confusion matrix data latih*




*Gambar 4.3 Confusion matrix data uji*



*Tabel 4.4* Eksperimen Model 2 *Confuxion Matrix* Data latih (70%)

| | **Aktual Positif (1)** | **Aktual Negatif (0)** |
|---|---|---|
| **Prediksi Positif (1)** | TP : 620 | FP : 50 |
| **Prediksi Negatif (0)** | FN : 104 | TN : 108 |

Tabel 4.5 Eksperimen Model 2 *Confuxion Matrix* Data Uji (30%)

| | **Aktual Positif (1)** | **Aktual Negatif (0)** |
|---|---|---|
| **Prediksi Positif (1)** | TP : 225 | FP : 19 |
| **Prediksi Negatif (0)** | FN : 40 | TN : 64 |

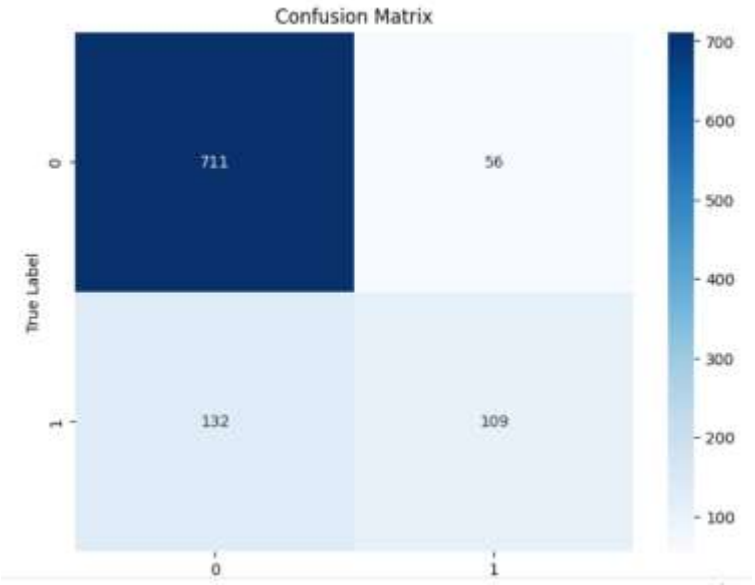
Tabel 4.6 Skenario Pengujian Model 1

| **Skenario Pembagian Data Latih 70% dan Data Uji 30%** | | |
|---|---|---|
| **Nilai** | **Data Latih** | **Data Uji** |
| **Accuracy** | 82,72% | 83,04% |
| **Precision** | 92,53% | 92,21% |
| **Recall** | 85,63% | 84,90% |
| **F1-Score** | 88,94% | 88,40% |

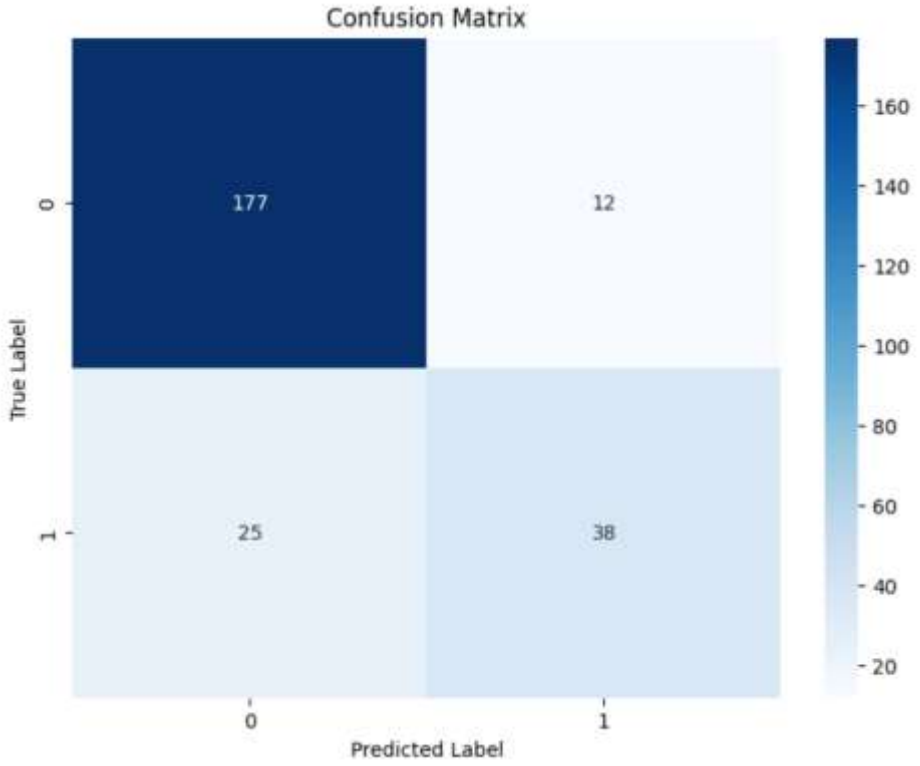
### *4.3. Skenario Dataset* 80% dan 20%

Skenario pengujian model pertama yaitu *split* data untuk data latih 80% dan data uji 20%. Hasil *accuracy* yang didapatkan 0,8412. Hasil *confusion matrix* dapat dilihat pada Gambar 4.9 dan Gambar 4.10

Gambar 4.5 *Confusion matrix data latih*



Gambar 4.6 *Confusion matrix data uji*




Tabel 4.7 Eksperimen Model 3 *Confuxion Matrix* Data Uji (80%)

| | **Aktual Positif (1)** | **Aktual Negatif (0)** |
|---|---|---|
| **Prediksi Positif (1)** | TP : 711 | FP : 56 |
| **Prediksi Negatif (0)** | FN : 132 | TN : 109 |

Tabel 4.8 Eksperimen Model 3 *Confuxion Matrix* Data Uji (20%)

| | **Aktual Positif (1)** | **Aktual Negatif (0)** |
|---|---|---|
| **Prediksi Positif (1)** | TP : 177 | FP : 12 |

Tabel 4.9 *Skenario* Pengujian Model 3

| ***Skenario* Pembagian Data Latih 80% dan Data Uji 20%** | | |
|---|---|---|
| **Nilai** | **Data Latih** | **Data Uji** |
| ***Accuracy*** | 81,34% | 7,5% |
| ***Precision*** | 92,69% | 93,65% |
| ***Recall*** | 84,34% | 87,62% |
| ***F1-Score*** | 88,31% | 90,53% |

## 5. KESIMPULAN

Metode *Random Forest* telah berhasil diimplementasikan untuk mengklasifikasikan data kesehatan mental di industri teknologi yang didapatkan dari *Kaggle.com* dengan menggunakan label/kelas "*Yes* atau "*No*" dari atribut *"Treatment*" dan menggunakan lima belas atribut yaitu atribut data yaitu *Age, Gender, Family History, Treatment, Work Interfere, Remote Work, Tech Company, Care Options, Wellness program, Seek Help, Anonymity, Mental Health Consequence, Supervisor, Mental health Interview, Phys Health Interview* mulai dari Agustus 2014 sampai Februari 2015 dengan tahap preprocess terdapat dua tahapan yaitu selection data dan cleaning data, pembagian data menjadi dua tahap yaitu data latih dan data uji, evaluasi model.

Hasil eksperimen yang dilakukan terhadap data uji kesehatan mental di industri teknologi tahun Agustus 2014 sampai Februari 2015 dengan acak menghasilkan jumlah klasifikasi label "*Yes*" ataupun "*No"* mengeluarkan hasil klasifikasi yang berbeda di setiap penginputan data pada masing-masing atribut yang berbeda sehingga dapat dikatakan data bersifat tidak seimbang *(imbalanced)* dan dilakukan pengujian model *Random Forest* menggunakan tiga model pembagian data latih dan data uji yang berbeda yaitu 60% data latih dan 40% data uji, 70% data latih dan 30% data uji, dan 80% data latih dan 20% data uji sehingga menghasilkan akurasi penggunaan 80% data latih dan 20% data uji menghasilkan akurasi klasifikasi tertinggi, yaitu 0,8412. Pembagian 70% dan 30% data latih juga menghasilkan akurasi yang sama, yaitu 0,8412. Hal ini menunjukkan bahwa semakin banyak data yang digunakan untuk pelatihan, atau data latih, model *Random Forest* cenderung melakukan klasifikasi data uji dengan lebih baik.

## DAFTAR PUSTAKA


[1] A. Y. Perdana, R. Latuconsina, and A. Dinimaharawati, "PREDIKSI STUNTING PADA BALITA DENGAN ALGORITMA RANDOM FOREST."

[2] U. Srinivasulu, A. Vivek, T. B. Tech, A. Dharun, and B. Tech, "Machine Learning Techniques for Stress Prediction in Working Employees."

[3] D. Winda Meidina and dan S. Netty Laura, "The Effect of Employees' Mental Health on Performance Mediated by Welfare in the Workplace (Empirical Study on Information Technology Division Employees During Work From Home," *Business Management Journal*, vol. 18, no. 1, p. p-ISSN, doi: 10.30813/bmj.

[4] Z. N. Rudianto, "PENGETAHUAN GENERASI Z TENTANG LITERASI KESEHATAN DAN KESADARAN MENTAL DI MASA PANDEMI The Impact Of Health Literacy On The Mental Health Conscioues ness Of The Z Generation In The Pandemic," 2022.

[5] Chaerudin Reza Alfaresy, "IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK ANALISIS KLASIFIKASI SURVEI KESEHATAN MENTAL INDUSTRI TEKNOLOGI (STUDI KASUS: OPEN SOURCING MENTAL ILLNESS) - Repository UPN Veteran Jakarta," 2022. https://repository.upnvj.ac.id/19626/ (accessed Nov. 30, 2022).

[6] L. Fadilah, "KLASIFIKASI RANDOM FOREST PADA DATA IMBALANCED."

[7] Y. Santur, M. Karaköse, and E. Akın, "Random Forest Based Diagnosis Approach for Rail Fault Inspection in Railways Fetal Health Pattern Classification Using Ensemble Learning View project Data Mining in Finance: Concepts, Trend and Applications View project Random Forest Based Diagnosis Approach for Rail Fault Inspection in Railways," 2016. [Online]. Available: https://www.researchgate.net/publication/314230630

[8] R. Supriyadi, W. Gata, N. Maulidah, A. Fauzi, I. Komputer, and S. Nusa Mandiri Jalan Margonda Raya No, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," vol. 13, no. 2, pp. 67–75, 2020, [Online]. Available: http://journal.stekom.ac.id/index.php/E-Bisnis■page67

[9] J. Wakiru, "00 WAKIRU et al: A decision tree-based classification framework for used oil analysis applying random forest feature selection A DECISION TREE-BASED CLASSIFICATION FRAMEWORK FOR USED OIL ANALYSIS APPLYING RANDOM FOREST FEATURE SELECTION100 WAKIRU et al: A decision tree-based classification framework for used oil analysis applying random forest feature selection," *PP 90-Journal of Applied Sciences, Engineering and Technology for Development JASETD*, vol. 3, no. 1, p. 90, 2018.

[10] W. Apriliah *et al.*, "SISTEMASI: Jurnal Sistem Informasi Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," 2021. [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[11] A. U. Zailani and N. L. Hanun, "PENERAPAN ALGORITMA KLASIFIKASI RANDOM FOREST UNTUK PENENTUAN KELAYAKAN PEMBERIAN KREDIT DI KOPERASI MITRA SEJAHTERA," *Infotech: Journal of Technology Information*, vol. 6, no. 1, pp. 7–14, Jun. 2020, doi: 10.37365/jti.v6i1.61.

[12] N. Wuryani, S. Agustiani, I. Komputer, and N. Mandiri, "Random Forest Classifier untuk Deteksi Penderita COVID-19 berbasis Citra CT Scan," *Jurnal Teknik Komputer AMIK BSI*, vol. 7, no. 2, 2021, doi: 10.31294/jtk.v4i2.

[13] S. J. Rigatti, "Random Forest," *J Insur Med*, vol. 47, no. 1, pp. 31–39, Jan. 2017, doi: 10.17849/insm-47-01-31-39.1.

[14] "https://algorit.ma/blog/random-forest-adalah-2022/."

[15] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, pp. 11–16, Jun. 2019, doi: 10.34148/teknika.v8i1.95.

[16] A. Syukron and A. Subekti, "Penerapan Metode Random Over-Under Sampling dan Random Forest untuk Klasifikasi Penilaian Kredit," *JURNAL INFORMATIKA*, vol. 5, no. 2, 2018.

[17] Z. Karimi, "Confusion Matrix Some of the authors of this publication are also working on these related projects: Data Cleaning Process View project." [Online]. Available: https://www.researchgate.net/publication/355096788

[18] Alfaresy Chaerudin Reza, “Problem Solving,” *Klasifikasi Kesehatan mental menggunakan metode naive bayes*, 2022.
[19] S. Yu, “Uncovering the hidden impacts of inequality on mental health: A global study,” *Transl Psychiatry*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/s41398-018-0148-0.
[20] E. L. Yearwood and V. P. Hines-Martin, “Editorial: Impact of social determinants of health on mental health,” *Archives of Psychiatric Nursing*, vol. 35, no. 1. W.B. Saunders, pp. A1–A2, Feb. 01, 2021. doi: 10.1016/j.apnu.2020.12.001.