

# Perbandingan Algoritma *Support Vector Machine (SVM)* dan *Decision Tree* Untuk Deteksi Tingkat Depresi Mahasiswa

Hermawan Syahputra<sup>1</sup>, Steven Imanuel Naibaho<sup>1</sup>, M. Ari Maulana<sup>1</sup>, Ihsan Zulfahmi<sup>1,\*</sup>, Edward Perdana Sinaga<sup>1</sup>

<sup>1</sup>Ilmu Komputer; Universitas Negeri Medan; Jl. Willem Iskandar / Pasar V, Medan, Sumatera Utara Kode Pos 20221 Telp. (061) 6613365, Fax. (061) 6614002 / 6613319; e-mail: [hsyahputra@unimed.ac.id](mailto:hsyahputra@unimed.ac.id), [stevnuel@mhs.unimed.ac.id](mailto:stevnuel@mhs.unimed.ac.id), [ihsanzulfahmi@mhs.unimed.ac.id](mailto:ihsanzulfahmi@mhs.unimed.ac.id), [m.arimaulana05@gmail.com](mailto:m.arimaulana05@gmail.com), [edwardperdanasinaga@mhs.unimed.ac.id](mailto:edwardperdanasinaga@mhs.unimed.ac.id)

\* Korespondensi: [ihsanzulfahmi@mhs.unimed.ac.id](mailto:ihsanzulfahmi@mhs.unimed.ac.id),

Diterima: 26 Mei 2023; Review: 11 Juni 2023; Disetujui: 24 Juni 2023

Cara sitasi: Hermawan Syahputra, Steven Imanuel Naibaho, M. Ari Maulana, Ihsan Zulfahmi, Edward Perdana Sinaga. 2023. A Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree Untuk Deteksi Tingkat Depresi Mahasiswa. Bina Insani ICT Journal. Vol. 10 (1): halaman 52-61.

**Abstrak:** Depresi adalah gangguan kesehatan mental umum yang dapat mempengaruhi orang-orang dari segala usia dan dari semua lapisan masyarakat. Deteksi dan intervensi dini sangat penting untuk mengelola depresi dan mencegah konsekuensi negatifnya. Dalam beberapa tahun terakhir, algoritma pembelajaran mesin telah digunakan untuk mendeteksi tingkat depresi pada individu. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Support Vector Machine (SVM) dan Decision Tree dalam mendeteksi tingkat depresi pada mahasiswa. Kumpulan data yang digunakan dalam penelitian ini dikumpulkan dari 787 mahasiswa di Universitas Lahore dan memasukkan 19 variabel yang berkaitan dengan gejala depresi, gaya hidup, dan demografi. Algoritma Chi-Square digunakan untuk pemilihan fitur dan algoritma Decision Tree dan SVM digunakan untuk pemodelan. Studi tersebut menemukan bahwa kedua algoritme tersebut sangat akurat dalam mendeteksi depresi pada mahasiswa, dengan model SVM memiliki tingkat akurasi 95% lebih tinggi daripada model pohon keputusan. Penulis telah menyajikan data dan tabel secara visual yang menunjukkan akurasi dan ukuran evaluasi model. Studi menyimpulkan bahwa model SVM sangat akurat dalam mendeteksi tingkat depresi pada mahasiswa dan dapat digunakan sebagai alat yang dapat diandalkan untuk deteksi dini dan intervensi. Hasil penelitian ini memiliki implikasi penting bagi psikolog untuk mengambil tindakan. Algoritma pembelajaran mesin dapat digunakan untuk mendeteksi tingkat depresi individu, yang pada gilirannya dapat membantu deteksi dini dan intervensi. Ini dapat mengarah pada manajemen depresi yang lebih baik dan hasil kesehatan mental yang lebih baik. Namun, penelitian lebih lanjut diperlukan untuk memvalidasi hasil penelitian ini dan untuk mengeksplorasi potensi algoritma pembelajaran mesin dalam mendeteksi tingkat depresi pada populasi lain.

**Kata kunci:** *Support Vector Machine*, Depresi, Mahasiswa, Decision Tree

**Abstract:** Depression is a common mental health disorder that can affect people of all ages and from all walks of life. Early detection and intervention are crucial in managing depression and preventing its negative consequences. In recent years, machine learning algorithms have been used to detect depression levels in individuals. This study aimed to compare the performance of Support Vector Machine (SVM) and Decision Tree algorithms in detecting depression levels in college students. The dataset used in this study was collected from 787 students at the University of Lahore and included 19 variables related to depression symptoms, lifestyle, and demographics. The Chi-Square algorithm was used for feature selection, and Decision Tree and SVM algorithms were used for modeling. The study found that both algorithms were highly accurate in detecting

depression levels in students, with the SVM model having a higher accuracy rate of 95% compared to the decision tree model. The study also discussed the process of evaluating machine learning models using relevant metrics such as Classification Report and Confusion Matrix. The author presented visualizations of the data and tables showing the accuracy and evaluation metrics of the models. The study concluded that the SVM model had good accuracy in detecting depression levels in college students and can be used as a reliable tool for early detection and intervention. The findings of this study have important implications for psychologists to take action. Machine learning algorithms can be used to detect individual depression levels, which, in turn, can aid in early detection and intervention. This can lead to better depression management and improved mental health outcomes. However, further research is needed to validate the findings of this study and to explore the potential of machine learning algorithms in detecting depression levels in other populations.

**Keywords:** *Support Vector Machine, Depression, Students, Decision Tree*

## 1. Pendahuluan

Depresi adalah gangguan suasana hati yang ditandai oleh perasaan sedih yang persisten, kehilangan minat atau kesenangan dalam aktivitas sehari-hari, perubahan nafsu makan atau tidur, perasaan lelah yang berlebihan, penurunan energi, dan berbagai gejala fisik dan kognitif lainnya. [1] Depresi dapat bervariasi dalam tingkat keparahan, mulai dari depresi ringan hingga depresi mayor yang parah. Ini adalah gangguan yang umum, yang dapat mempengaruhi orang dari segala usia dan latar belakang. [2] Depresi juga dapat memengaruhi individu secara berulang atau kronis, atau muncul dalam episode yang terpisah.

Dari penelitian yang sudah dilakukan oleh beberapa orang didapat bahwa [3] dalam jurnal Rancang Bangun Sistem Pakar Diagnosa tingkat Depresi Pada Mahasiswa Tingkat Akhir Menggunakan Metode Fuzzy Tsukamoto (Studi Kasus : Universitas Siliwangi) tingkat akurasi yang dihasilkan dari metode Fuzzy Tsukamoto salah 96%, dan masih banyak metode lainnya yang sudah digunakan untuk mendeteksi tingkat depresi atau stres mahasiswa seperti pada jurnal Sistem Pakar Diagnosa Depresi Mahasiswa Akhir Dengan Metode Certainty Factor Berbasis Mobile Karya Apip Supiandi dkk. Yang menggunakan metode Certainty Factor Metode Certainty Factor merupakan metode yang mendefinisikan ukuran kepastian terhadap suatu fakta atau aturan, untuk menggambarkan tingkat keyakinan pakar terhadap masalah yang sedang dihadapi, dengan menggunakan Certainty Factor ini dapat menggambarkan tingkat keyakinan pakar pada penelitian mereka.

[4]. Juga melakukan penelitian dengan judul Diagnosa Tingkat Depresi Mahasiswa Selama Masa Pandemi Covid-19 Menggunakan Algoritma Random Forest didapat hasil prediksi tingkat depresi menggunakan algoritma Random Forest memiliki akurasi sebesar 85,94%. Pada setiap tingkat depresi, memiliki persentase accuracy dan specificity diatas 88%, sedangkan untuk sensitivity, pada tingkat depresi normal, gejala depresi ringan, dan depresi ringan memiliki persentase diatas 90%, pada tingkat depresi berat sebesar 74%, dan pada tingkat depresi sedang sebesar 47%.

Selanjutnya Penelitian dengan tema yang mirip yaitu Komparasi Metode Machine Learning pada Diagnosa Gangguan Kejiwaan Depresi karya [5]. Dimana penulis dalam penelitian menggunakan 4 metode diantaranya : Naïve Bayes dengan hasil Precision sebesar 73%, recall sebesar 82% dan accuracy sebesar 75%; KNN dengan hasil Precision sebesar 77%, recall sebesar 73% dan accuracy sebesar 75%; Decision Tree dengan hasil Precision sebesar 75%, recall sebesar 80% dan accuracy sebesar 76%, Random Forest dengan hasil Precision sebesar 76%, recall sebesar 90% dan accuracy sebesar 80% sehingga didapat Random Forest yang memiliki tingkat akurasi tertinggi sebesar 80%.

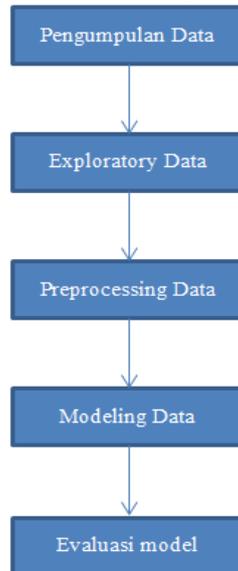
Dari hasil pembahasan diatas, maka penulis melakukan sebuah penelitian yaitu "Perbandingan performa algoritma Decision Tree dan Support Vector Machine dalam mendeteksi tingkat depresi mahasiswa". [6] Metode Decision Tree dipilih karena kemampuannya dalam menggambarkan dan memahami hubungan antara variabel input dengan variabel target. Dalam hal ini, variabel input meliputi data demografis, riwayat kesehatan mental, dan faktor-faktor lain yang berpotensi terkait dengan depresi. Decision Tree dapat memberikan pemahaman yang intuitif tentang faktor-faktor apa yang paling berpengaruh dalam menentukan tingkat depresi.

Selain itu, SVM dipilih karena kemampuannya dalam menangani masalah klasifikasi pada dataset yang kompleks. [7]SVM membangun batas keputusan yang optimal di antara kelas-kelas yang berbeda dalam ruang fitur multidimensi. Dengan demikian, SVM dapat membantu mengidentifikasi pola tersembunyi dalam data yang dapat memprediksi tingkat depresi mahasiswa.

## 2. Metode Penelitian

### a. Alur Proses Penelitian:

Untuk alur proses penelitian yang penulis lakukan dapat dilihat pada gambar dibawah ini.



Sumber: Hasil Penelitian (2023)

Gambar 1. Diagram Alur Penelitian

### b. Data Introduction

Dataset yang digunakan berasal dari Kaggle dan berisi tentang data depresi dan kecemasan. Dataset ini dikumpulkan dari mahasiswa Universitas Lahore dengan partisipasi 787 peserta. Data ini didasarkan pada inventarisasi Beck Depression dan Beck Anxiety. Gangguan kecemasan merupakan jenis gangguan mental yang umum terjadi, mempengaruhi sekitar satu dari setiap empat orang dalam hidup mereka. Terdapat 19 variabel/kelas serta 783 data dari setiap variabel atau nilai kelas nya. Kolom-kolom pada dataset ini meliputi : ID (identitas mahasiswa), school\_year (tahun menjadi mahasiswa), Age (usia mahasiswa), Gender (jenis kelamin), BMI (ukuran berat badan), Who\_BMI (tingkat berat badan), Depression\_treatment (penanganan depresi), Depression\_diagnosis (diagnosis depresi), Gad\_Score (nilai gangguan kecemasan umum), Suicidal (rencana bunuh diri), Anxiety\_severity (kecemasan parah), Anxiety\_treatment (penanganan kecemasan), Anxiousness (kecemasan dan kekhawatiran berlebih), dan Sleepness (gangguan tidur).

Gambar yang terlampir menunjukkan contoh sampel data dalam dataset tersebut.

|   | id | school_year | age | gender | bmi       | who_bmi         | phq_score | depression_severity | depressiveness | suicidal |
|---|----|-------------|-----|--------|-----------|-----------------|-----------|---------------------|----------------|----------|
| 0 | 1  | 1           | 19  | male   | 33.333333 | Class I Obesity | 9         | Mild                | False          | False    |
| 1 | 2  | 1           | 18  | male   | 19.841270 | Normal          | 8         | Mild                | False          | False    |
| 2 | 3  | 1           | 19  | male   | 25.102391 | Overweight      | 8         | Mild                | False          | False    |
| 3 | 4  | 1           | 18  | female | 23.738662 | Normal          | 19        | Moderately severe   | True           | True     |
| 4 | 5  | 1           | 18  | male   | 25.617284 | Overweight      | 6         | Mild                | False          | False    |

Sumber: Hasil Penelitian (2023)

Gambar 2. Sampel Data Depresi

### c. Data Preprocessing

Setelah memiliki dataset, selanjutnya adalah melakukan Preprocessing Data. Tujuan daripada Preprocessing data ini adalah untuk menghilangkan missing value dan juga outlier. Dari dataset

yang peneliti gunakan, terdapat 41 data yang bernilai missing value. Lalu peneliti melakukan penanganan missing value dengan menghapus data variabel yang terdapat missing value nya agar tidak terjadi kesalahan pada model yang dibangun. Selanjutnya, peneliti melakukan Label Encoder untuk mengubah tipe data keseluruhan menjadi numeric, dikarenakan model hanya bisa membaca data dalam tipe numeric.

d. Feature Selection

Selanjutnya peneliti melakukan seleksi fitur. Seleksi fitur (feature selection) adalah proses pemilihan subset dari fitur (variabel atau atribut) yang tersedia dalam dataset untuk digunakan dalam pembuatan model. Tujuan utama seleksi fitur adalah untuk meningkatkan kinerja model dengan menghilangkan fitur yang tidak relevan, redundan, atau yang memiliki dampak negatif pada prediksi. Untuk feature selection peneliti menggunakan data “depressiveness” sebagai variabel dependen atau target. Selanjutnya peneliti melakukan feature selection yang bertujuan untuk membuat proses pemodelan semakin akurat. Algoritma yang digunakan didalam seleksi fitur adalah algoritma Chi Square Algoritma Chi-square merupakan metode statistik yang digunakan dalam machine learning untuk mengevaluasi hubungan antara dua variabel kategorikal dalam dataset. Algoritma ini menggunakan uji Chi-square untuk menguji independensi atau ketergantungan antara dua variabel. [6] Didapatkan hasil seleksi fitur menggunakan Algoritma Chi Square adalah sebagai berikut :

- a) Bmi
- b) Phq\_score
- c) Suicidal
- d) Depression\_diagnosis
- e) Depression\_treatment
- f) Gad\_score
- g) Anxiety\_severity
- h) Anxiousness
- i) Anxiety\_diagnosis
- j) Anxiety\_treatment
- k) Epworth\_score
- l) sleepiness

e. Modeling Data

Penulis menggunakan Decision tree sebagai salah satu metode untuk menyelesaikan persoalan, dimana Decision tree adalah model machine learning yang memprediksi nilai target dengan membuat serangkaian keputusan. Alasan dipilihnya metode Decision tree karena kemampuannya dalam menggambarkan dan memahami hubungan antara variabel input dengan variabel target dan Decision tree biasanya meniru kemampuan berpikir manusia untuk memproses pengambilan keputusan yang kompleks menjadi lebih simple, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Sehingga mudah dimengerti. berdasarkan fitur yang diberikan. Secara umum, rumus perhitungan decision tree melibatkan beberapa tahap.

Entropy adalah ukuran ketidakpastian atau keacakan pada data. Semakin homogen atau seragam data, maka nilai entropy semakin rendah. Rumus untuk menghitung entropy:

$$E = - \sum_{i=1}^n P_i \log_2 P_i$$

di mana  $P_i$  adalah proporsi jumlah contoh yang memiliki nilai target  $i$  dalam set data saat ini. Hitung informasi gain untuk setiap fitur:

Informasi gain adalah ukuran seberapa banyak informasi yang diperoleh dengan membagi data berdasarkan fitur tertentu. Fitur dengan informasi gain tertinggi akan dipilih sebagai fitur pemisah berikutnya.

$$IG(D, f) = E(D) - \sum_{v \in V} \left| \frac{D_v}{D} \right| E(D_v)$$

di mana  $D$  adalah set data saat ini,  $f$  adalah fitur yang akan dievaluasi,  $V$  adalah himpunan nilai yang mungkin untuk fitur  $f$ ,  $D_v$  adalah subset data yang memenuhi kondisi fitur  $f = v$ , dan  $E(D_v)$  adalah nilai entropy dari subset data  $D_v$ .

Selain menggunakan metode Decision tree peneliti juga menggunakan SVM sebagai alternatif, alasan dipilihnya SVM karena kemampuannya dalam menangani masalah klasifikasi pada dataset yang kompleks. Pemecahan masalah yang dilakukan menggunakan hyperparameter dengan cara menentukan rentang nilai untuk C dan gamma yang akan dicoba. Misalnya, C dapat diatur pada nilai [0.1, 1, 10] dan gamma pada nilai [0.001, 0.01, 0.1]. Bagi data menjadi set pelatihan dan validasi. Untuk setiap kombinasi nilai C dan gamma pada rentang yang ditentukan, melatih model SVM pada set pelatihan dan evaluasi pada set validasi. Dan juga dengan RBF dimana Hitung jarak antara data uji dan setiap vektor dukungan (support vector) pada model SVM menggunakan kernel RBF. Formula untuk menghitung jarak dengan kernel RBF adalah sebagai berikut:

$$K(x, x_l) = \exp(-\text{gama} * \|x - x_l\|^2)$$

Formula untuk menghitung nilai prediksi pada SVM RBF adalah sebagai berikut:

$$f(x) = \text{sum}(\text{alpha} * y * K(x, x_i))$$

Di mana  $f(x)$  adalah nilai prediksi untuk data uji,  $\alpha$  adalah bobot dari vektor dukungan,  $y$  adalah nilai target dari vektor dukungan, dan  $K(x, x_i)$  adalah jarak antara data uji dan vektor dukungan dengan kernel RBF.

#### f. Evaluasi Model

Evaluasi model pada machine learning adalah proses untuk mengukur kinerja suatu model prediksi atau klasifikasi dengan menggunakan metrik yang relevan. Evaluasi model sangat penting untuk menilai sejauh mana model tersebut mampu melakukan prediksi atau klasifikasi yang akurat dan dapat diandalkan. Metrik Evaluasi yang digunakan didalam penelitian ini yaitu Classification Report yang terbagi kedalam beberapa bagian yaitu: Accuracy, Precision, Recall, F1-Score dan Support, serta Confusion Matriks.

Evaluasi model kali ini menggunakan Confusion Matriks dengan menggunakan confusion matrix, penulis dapat melihat secara detail bagaimana model melakukan prediksi untuk setiap kelas dan mengukur kinerjanya dengan menggunakan metrik evaluasi yang sesuai dengan konteks masalah yang tertuang didalam penelitian ini.

Confusion matrix biasanya memiliki bentuk tabel persegi dengan empat sel utama:

1. True Positive (TP): Jumlah data yang benar diprediksi sebagai positif oleh model.
2. True Negative (TN): Jumlah data yang benar diprediksi sebagai negatif oleh model.
3. False Positive (FP) atau Type I error: Jumlah data yang sebenarnya negatif tetapi salah diprediksi sebagai positif oleh model.
4. False Negative (FN) atau Type II error: Jumlah data yang sebenarnya positif tetapi salah diprediksi sebagai negatif oleh model.

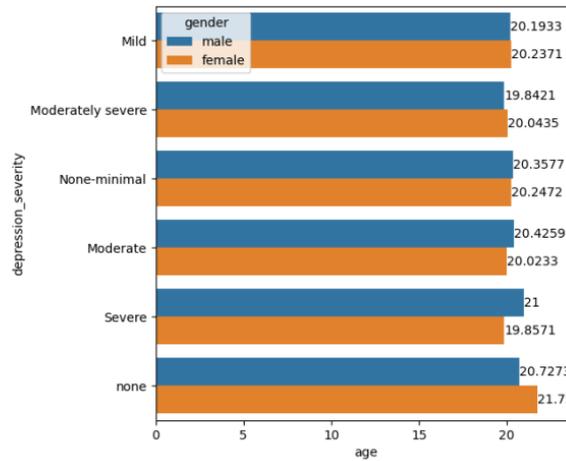
Confusion matrix memberikan informasi yang lebih rinci tentang performa model klasifikasi, yang memungkinkan kita untuk menghitung metrik evaluasi seperti akurasi, presisi, recall, dan lain-lain. Berikut adalah beberapa metrik evaluasi yang dapat dihitung menggunakan confusion matrix:

- a) Akurasi (Accuracy): Akurasi mengukur sejauh mana model dapat melakukan prediksi yang benar secara keseluruhan. Rumusnya adalah:  $\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
- b) Presisi (Precision): Presisi mengukur sejauh mana prediksi positif yang dilakukan oleh model benar. Rumusnya adalah:  $\text{Presisi} = \frac{\text{TP}}{\text{TP} + \text{FP}}$
- c) Recall (Sensitivitas atau True Positive Rate): Recall mengukur sejauh mana model dapat mengidentifikasi kelas positif yang sebenarnya. Rumusnya adalah:  $\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$
- d) F1-Score: F1-score adalah ukuran komposit yang menggabungkan presisi dan recall. F1-score berguna ketika ingin mencari keseimbangan antara presisi dan recall. Rumusnya adalah:  $\text{F1-Score} = 2 * \frac{\text{Presisi} * \text{Recall}}{\text{Presisi} + \text{Recall}}$

### 3. Hasil dan Pembahasan

#### a. EDA (Exploratory Data)

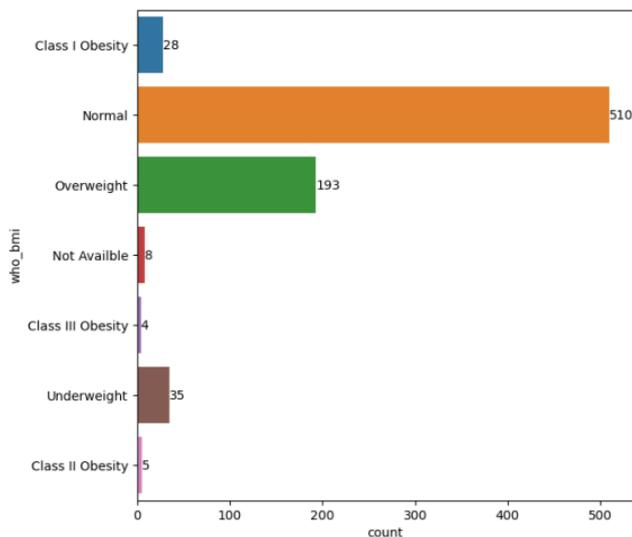
Tahapan EDA dilakukan untuk memahami dan mengenali data[7], sehingga kita dapat melakukan tahapan Preprocessing Data dan membangun Model pada data tersebut. Berikut merupakan visualisasi data dari hasil Eksplorasi Data yang telah dilakukan.



Sumber: Hasil Penelitian (2023)

Gambar 3. Visualisasi Data Tingkat Keparahan Depresi berdasarkan Gender dan umur

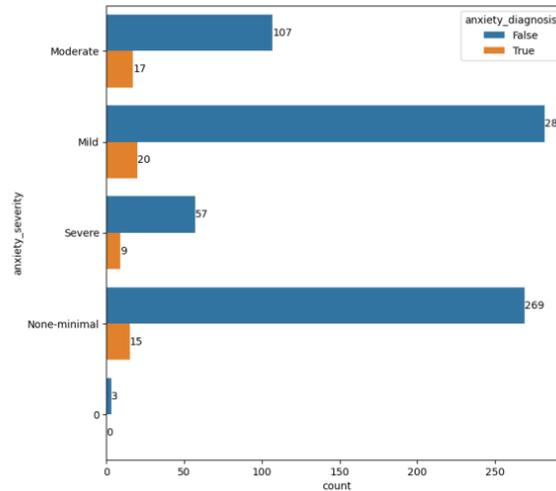
Dari grafik diagram batang diatas menunjukkan data dari kelas depression\_severity beserta variabel dari kelas nya. Dari gambar diatas dilihat bahwa orang yang memiliki depresi yang cukup parah yaitu Severe kebanyakan didominasi oleh pria berusia 21 tahun, dan untuk yang tidak memiliki depresi didominasi oleh wanita pada rentang usia 21 tahun.



Sumber: Hasil Penelitian (2023)

Gambar 4. Visualisasi data Tingkat Berat Badan

Untuk melihat tingkat depresi mahasiswa[8], juga diperlukan data berat badan mahasiswa tersebut. Dari visualisasi data kelas who\_bmi diatas menunjukkan bahwa kebanyakan mahasiswa yang diuji memiliki berat badan normal dan ada juga yang memiliki berat badan lebih sebanyak 193 mahasiswa dan yang kurus sebanyak 35 mahasiswa



Sumber: Hasil Penelitian (2023)

Gambar 5. Visualisasi Data Tingkat Kecemasan

Selanjutnya data yang diukur adalah tingkat kecemasan mahasiswa. Kelas yang digunakan yaitu kelas anxiety\_severity. Dari grafik diatas menunjukkan bahwa kebanyakan mahasiswa memiliki tingkat kecemasan yang rendah yaitu berjumlah 282 mahasiswa, dan untuk yang tidak memiliki kecemasan sebanyak 269 mahasiswa, dan yang sedang sebanyak 107 mahasiswa, sementara untuk tingkat kecemasan parah atau severe sebanyak 57 mahasiswa.

b. Modeling Data

Tahapan selanjutnya adalah peneliti melakukan pemilihan untuk variabel independen dan dependen. Dari hasil preprocessing yang telah dilakukan sebelumnya maka peneliti memutuskan bahwa data “depression\_diagnosis” akan dijadikan sebagai variabel dependen/target dari penelitian tersebut. Setelah itu, peneliti melakukan pembagian data latih dan data uji dengan membagi data dengan 33% dari keseluruhan data digunakan untuk data testing dengan random test sebanyak 42 data. Untuk hasil data yang telah dibagi kedalam data training dan testing dapat dilihat pada tabel berikut :

Tabel 1. Tabel Pembagian Data Training dan Testing

| Variabel | Train | Test |
|----------|-------|------|
| X        | 512   | 253  |
| Y        | 512   | 253  |

Sumber: Hasil Penelitian (2023)

X = Independen

Y = Dependen

Sumber: Hasil Penelitian (2023)

Untuk parameter yang akan dibandingkan dan menjadi target didalam model yang akan dibangun dapat dilihat pada tabel dibawah ini :

Tabel 2. Tabel Data Variabel Dependen dan Independen dalam model

| Variabel Dependen   | Variabel Independen(Target) |
|---|-----------------------------|
| Bmi, Phq_score, Suicidal, Depression_treatment, Gad_score, Anxiety_severity, anxiousness, Anxiety_diagnosis, Anxiety_treatment, Epworth_score, sleepiness | Depression_diagnosis        |

Sumber: Hasil Penelitian (2023)

c. Klasifikasi dengan Model Decision Tree

Untuk model Decision tree, penulis menggunakan Kriteria Entropy. Kriteria entropy digunakan untuk memilih atribut yang paling baik dalam membagi atau memisahkan himpunan data menjadi subset yang lebih homogen[9]. Saat membangun decision tree, atribut dengan perubahan

entropy yang paling signifikan atau informasi gain yang maksimum akan dipilih sebagai kriteria pemisahan pada setiap langkah. Pada penelitian ini, decision tree yang dibangun akan memiliki tingkat maksimum sebanyak 3. Ini berarti pohon keputusan akan terdiri dari 3 tingkat atau layer, di mana setiap tingkat merepresentasikan pemisahan atribut yang berbeda. Dari hasil pemodelan Decision Tree untuk kategori Entropy didapatkan Model Akurasi sebesar 0.9486 atau 94%. Untuk hasil training akurasi scorenya didapat sebesar 0.96 atau 96%, dan untuk test score nilainya sebesar 0.94 atau 94%. Di bawah ini merupakan hasil perhitungan confusion matrix dari model Decision Tree yang dibangun :

Tabel 3. hasil perhitungan confusion matrix

| Aktual (N) | Positif | Negatif |
|------------|---------|---------|
| TRUE       | 235     | 1       |
| FALSE      | 12      | 5       |

Sumber: Hasil Penelitian (2023)

Dari tabel confusion matrix menunjukkan bahwa hasil model yang dibangun menggunakan model Decision Tree menunjukkan bahwa nilai TF yaitu 235 berarti ada 235 orang diprediksi tidak terkena depresi, lalu nilai TN ada 5 orang yang diprediksi terkena depresi. Lalu ada nilai FP yaitu 1 artinya ada 1 orang yang diprediksi tidak terkena depresi, namun ternyata mereka terkena depresi, dan untuk nilai FN ada 12 berarti ada 12 orang yang diprediksi terkena depresi, namun mereka ternyata tidak terkena depresi. Selanjutnya untuk evaluasi model menggunakan classification report dilihat pada tabel dibawah ini :

Tabel 4. hasil evaluasi model decision tree

|              | Precision | Recall | F1- score | support |
|--------------|-----------|--------|-----------|---------|
| 0            | 0.95      | 1      | 0.097     | 235     |
| 1            | 0,83      | 0,29   | 0,43      | 17      |
| Accuracy     |           |        | 0.95      | 253     |
| Macro avg    | 0.89      | 0.64   | 0.7       | 253     |
| Weighted avg | 0.94      | 0.95   | 0.94      | 253     |

Sumber: Hasil Penelitian (2023)

Data yang bernilai 0 berarti tidak terkena depresi, dan 1 berarti terkena depresi. Hasil model yang telah dibangun menunjukkan bahwa 97% dari data yang telah diuji adalah tidak terkena depresi, dan untuk 43% terkena depresi. Untuk akurasi model yang dibangun sebesar 95%.

d. Klasifikasi dengan model SVM (Support Vector Machine)

Selanjutnya peneliti menggunakan Model Support Vector Machine (SVM) untuk mendeteksi tingkat depresi mahasiswa. Kernal SVM yang digunakan yaitu Default Parameter, Linear Kernel, dan Radial Kernel. Karena data yang kami gunakan bersifat kategorikal, maka yang akan menjadi acuan dalam SVM ini adalah Radial Basis Function (RBF) Kernel. Kernel ini memetakan data ke dalam ruang fitur yang tak terbatas dengan menggunakan fungsi basis Gaussian. RBF kernel memiliki parameter gamma yang mengatur seberapa jauh pengaruh setiap titik data terhadap pemisahan hyperplane. Untuk hasil Evaluasi Model menggunakan Klasifikasi Report dapat dilihat pada tabel dibawah ini.

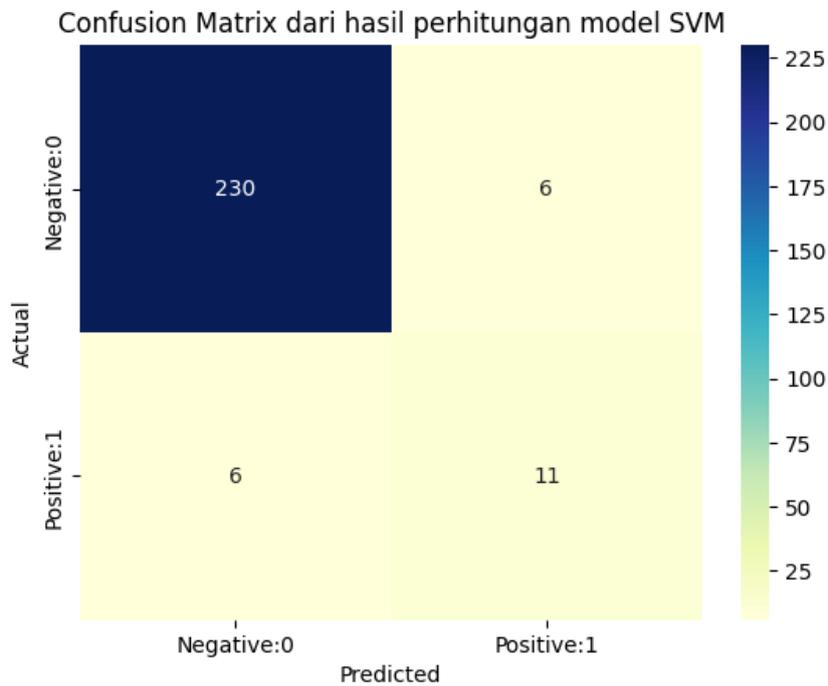
Tabel 5. hasil evaluasi model svm

|              | Precision | Recall | F1- score | support |
|--------------|-----------|--------|-----------|---------|
| 0            | 0.097     | 0.097  | 0.097     | 236     |
| 1            | 0,65      | 0,65   | 0,65      | 17      |
| Accuracy     |           |        | 0.95      | 253     |
| Macro avg    | 0.81      | 0.81   | 0.81      | 253     |
| Weighted avg | 0.95      | 0.95   | 0.95      | 253     |

Sumber: Hasil Penelitian (2023)

Dengan menggunakan SVM, didapatkan akurasi sebesar 93% dengan menggunakan default Hyperparameter untuk nilai 0 yaitu negative (prediksi mahasiswa yang tidak terkena depresi) dan akurasi untuk nilai 1 (prediksi mahasiswa yang positive terkena depresi). Selanjutnya peneliti menggunakan SVM dengan Linier Kernel. Kernel linier adalah kernel yang digunakan ketika data yang diklasifikasikan dipisahkan oleh sebuah garis atau hyperline. Untuk akurasi model SVM dengan menggunakan Linier Kernal bernilai 0.95 atau 95%.

Model SVM dengan Kernel Radial memiliki nilai akurasi yaitu 0.93 atau 93%. Tahapan selanjutnya adalah melakukan Evaluasi Model dari model yang digunakan yaitu SVM. Untuk detailnya dapat dilihat pada gambar dibawah ini.



Sumber: Hasil Penelitian (2023)

Gambar 6. hasil model svm dengan kernel radial

Dari hasil model yang telah dibangun, didapatkan bahwa pada kolom (0,0) atau True Positive adalah 230, yang artinya dari model yang diuji, terdapat 230 mahasiswa yang tidak mengalami depresi. Sementara untuk True Negative pada kolom (1,1) bernilai 11, artinya 11 orang mahasiswa yang diprediksi memiliki depresi. Untuk nilai False Positive pada kolom (0,1) bernilai 6 artinya 6 orang mahasiswa yang terdeteksi memiliki depresi namun pada kenyataannya mereka tidak mengalami depresi, dan False Negative pada kolom (1,0) bernilai 6 artinya ada 6 orang mahasiswa yang tidak terdeteksi mengalami depresi namun mereka mengalami depresi. Untuk klasifikasi report dari model SVM yang dibuat mendapat akurasi 97% untuk kategori 0 yaitu negative (prediksi mahasiswa yang tidak terkena depresi) dan akurasi untuk kategori 1 (positive) yaitu 65% untuk mahasiswa yang diprediksi mengalami depresi. Dan secara keseluruhan classification model dari model SVM yang dibangun memiliki nilai akurasi sebesar 0.95 atau 95%. Setelah itu diperoleh Classification Accuracy sebesar 0.9526 atau sekitar 95% dan Classification Errornya adalah 0.0474 atau berkisar 0.4 % kemudian Precisionnya bernilai 0.9746 atau bernilai 97% dan didapat hasil Specificity bernilai 0.6471.

#### 4. Kesimpulan

Dari hasil penelitian yang telah peneliti lakukan maka dapat diambil kesimpulan dari tabel hasil algoritma Decision tree yang tertera pada tabel 4 dan support vector machine yang terdapat pada tabel 5 maka peneliti dapat menarik kesimpulan seperti pada table dibawah ini.

Tabel untuk model bernilai 0 atau negative (tidak terkena depresi)

Tabel 6. tidak terkena depresi

| Model         | Precision | Recall | F-1 Score | Accuracy |
|---------------|-----------|--------|-----------|----------|
| SVM           | 0.97      | 0.97   | 0.97      | 0.95     |
| Decision Tree | 0.95      | 1      | 0.43      | 0.95     |

Sumber: Hasil Penelitian (2023)

Tabel untuk model bernilai 1 atau positif (terkena depresi).

Tabel 7.terkena depresi

| Model         | Precision | Recall | F-1 Score | Accuracy |
|---------------|-----------|--------|-----------|----------|
| SVM           | 0.65      | 0.65   | 0.65      | 0.95     |
| Decision Tree | 0.83      | 0.29   | 0.43      | 0.95     |

Sumber: Hasil Penelitian (2023)

Dapat ditarik kesimpulan bahwa deteksi yang dilakukan oleh model Decision tree dan SVM memiliki akurasi yang tinggi tetapi dalam segi akurasi yang dihasilkan svm lebih tinggi sekitar 0.02% daripada yang dihasilkan oleh decision tree, maka dapat dikatakan metode SVM memiliki keakuratan yang baik dalam mendeteksi tingkat depresi pada mahasiswa

### Referensi

- [1] N. I. Kurniati, H. Mubarak, and A. Reinaldi, "Rancang Bangun Sistem Pakar Diagnosa tingkat Depresi Pada Mahasiswa Tingkat Akhir Menggunakan Metode Fuzzy Tsukamoto(Studi Kasus : Universitas Siliwangi)," vol. 2, no. 1, 2017.
- [2] D. Septiani, U. Enri, N. Sulistiyowati, T. Informatika, and U. S. Karawang, "STRING (Satuan Tulisan Riset dan Inovasi Teknologi) DIAGNOSA TINGKAT DEPRESI MAHASISWA SELAMA MASA PANDEMI COVID-19 MENGGUNAKAN ALGORITMA RANDOM FOREST."
- [3] R. Wajhillah, S. Bahri, and A. Wibowo, "Komparasi Metode Machine Learning pada Diagnosa Gangguan Kejiwaan Depresi," 2020. [Online]. Available: <https://www.kaggle.com/everseek/depression>
- [4] M. Jordan, J. Kleinberg, and B. Schölkopf, "Pattern Recognition and Machine Learning."
- [5] A. W. Putri, B. Wibhawa, and A. S. Gutama, "Kesehatan Mental Masyarakat Indonesia (Pengetahuan, Dan Keterbukaan Masyarakat Terhadap Gangguan Kesehatan Mental)," *Prosiding Penelitian dan Pengabdian kepada Masyarakat*, vol. 2, no. 2, pp. 252–258, 2015, doi: 10.24198/jppm.v2i2.13535.
- [6] I. C. Negara and A. Prabowo, "Penggunaan Uji Chi–Square untuk Mengetahui Pengaruh Tingkat Pendidikan dan Umur terhadap Pengetahuan Penasun Mengenai HIV–AIDS di Provinsi DKI Jakarta," *Prosiding Seminar Nasional Matematika dan Terapannya 2018*, vol. 1, no. 1, pp. 1–8, 2018.
- [7] A. T. Setyanto, "Deteksi Dini Prevalensi Gangguan Kesehatan Mental Mahasiswa di Perguruan Tinggi," *Wacana*, vol. 15, no. 1, p. 66, 2023, doi: 10.20961/wacana.v15i1.69548.
- [8] D. Avianto, "Pengenalan Pola Karakter Plat Nomor Kendaraan Menggunakan Algoritma Momentum Backpropagation Neural Network," *Jurnal Informatika*, vol. 10, no. 1, pp. 1199–1209, 2016, doi: 10.26555/jifo.v10i1.a3352.
- [9] E. Deliviana, Maria Helena Erni, Putri Melina Hilery, and Novi Melly Naomi, "Jurnal Selaras : Kajian Bimbingan dan Konseling serta Psikologi Pendidikan," *Jurnal Selaras : Kajian Bimbingan dan Konseling serta Psikologi Pendidikan*, vol. 3, no. 2, pp. 129–138, 2021.
- [10] A. Suryanto and S. Nada, "Analisis Kesehatan Mental Mahasiswa Perguruan Tinggi Pada Awal Terjangkitnya Covid-19 di Indonesia," *Jurnal Citizenship Virtues*, vol. 1, no. 2, pp. 83–97, 2021.