

Prediksi Tingkat Risiko Depresi Pada Mahasiswa Menggunakan Algoritma K-Nearest Neighbor

Predicting Depression Risk Levels in College Students Using the K-Nearest Neighbor Algorithm

Attagantari*¹, Arief Hidayat², Akhmad Pandhu Wijaya³

^{1,2,3} Program Studi Informatika, Fakultas Teknik, Universitas Wahid Hasyim, Indonesia
e-mail: *¹21103041043@unwahas.ac.id, ²rifmillenia@gmail.com, ³pandhudsn@unwahas.ac.id

Abstrak

Mahasiswa sering mengalami tekanan dari aspek akademik, sosial, dan pribadi yang dapat menyebabkan gangguan mental seperti depresi. Jika tidak ditangani, kondisi ini berisiko mengganggu pencapaian akademik, relasi sosial, hingga memicu tindakan ekstrem. Penelitian ini bertujuan untuk merancang dan mengembangkan sebuah sistem guna memprediksi tingkat risiko depresi pada mahasiswa menggunakan algoritma klasifikasi K-Nearest Neighbor (KNN). Data dikumpulkan dari 300 mahasiswa melalui kuesioner dengan skala Likert. Sistem dikembangkan menggunakan Python untuk tahap pra-pemrosesan. Proses analisis mencakup seleksi data, pembersihan, transformasi, klasifikasi dengan KNN, dan evaluasi kinerja menggunakan confusion matrix. Dari pengujian terhadap 60 data uji, diperoleh akurasi 85%, presisi 90%, dan recall 83%. Sistem ini berpotensi digunakan sebagai alat bantu deteksi dini bagi mahasiswa dan institusi pendidikan.

Kata Kunci: Depresi, Mahasiswa, K-Nearest Neighbor, Klasifikasi, Data Mining

Abstract

College students often experience academic, social, and personal stress that can lead to mental health issues such as depression. If left untreated, this condition can disrupt academic achievement and social relationships, and even trigger extreme behavior. This study aims to design and develop a system to predict the risk of depression in college students using the K-Nearest Neighbor (KNN) classification algorithm. Data were collected from 300 students through a Likert-scale questionnaire. The system was developed using Python for pre-processing. The analysis process included data selection, cleaning, transformation, classification with KNN, and performance evaluation using a confusion matrix. Testing of 60 test data sets yielded 85% accuracy, 90% precision, and 83% recall. This system has the potential to be used as an early detection tool for students and educational institutions.

Keywords : Depression, Students, K-Nearest Neighbor, Classification, Data Mining

PENDAHULUAN

Mahasiswa menghadapi masalah kesehatan mental yang semakin meningkat, terutama mereka yang menghadapi tekanan akademis, sosial, dan pribadi yang kompleks. Tekanan tersebut kerap menjadi pemicu munculnya masalah kesehatan mental seperti depresi. Jika tidak segera ditangani secara tepat, kondisi ini berpotensi berdampak negatif terhadap pencapaian akademik, kondisi fisik, serta interaksi sosial mahasiswa. Mahasiswa mengalami peningkatan depresi dalam beberapa tahun terakhir. Kondisi ini membutuhkan tindakan pencegahan dan deteksi dini yang lebih baik. Salah satu masalah kesehatan mental yang paling sering dialami oleh remaja dan dewasa muda adalah depresi. Salah satu tipe masalah kesehatan mental yang dikenal sebagai depresi adalah yang ditandai dengan hilangnya minat dan motivasi untuk beraktivitas, munculnya pikiran negatif, perasaan sedih yang berkepanjangan, dan kesulitan bersosialisasi.

Informasi Artikel:

Submitted: bulan 20xx, **Accepted:** bulan 20xx, **Published:** Mei / November 20xx

ISSN: 2685-4902 (media online), **Website:** <http://jurnal.umus.ac.id/index.php/intech>

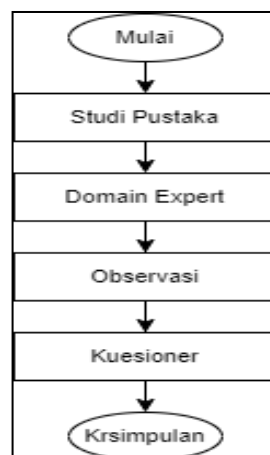
Dalam beberapa kasus, kondisi ini bahkan dapat menyebabkan bunuh diri. Sekitar 300 juta orang, atau 3,76% dari populasi global, mengalami depresi. World Health Organization (WHO) memperkirakan bahwa dalam 15 tahun mendatang, depresi akan menjadi salah satu faktor utama penyebab kematian. Depresi memiliki dampak yang signifikan terhadap kualitas hidup, yang mencakup penurunan prestasi, penurunan kreativitas, dan penurunan produktivitas di tempat kerja. Akibatnya, orang kesulitan untuk menikmati aktivitas yang biasanya menyenangkan (Purwanto, Sri, and Tamad 2022). Salah satu cara untuk mengidentifikasi risiko depresi secara cepat dan tepat adalah dengan memanfaatkan teknologi berbasis kecerdasan buatan, khususnya algoritma klasifikasi.

Dalam studi ini, pendekatan yang digunakan adalah metode kuantitatif berbasis data mining, dengan memanfaatkan algoritma K-Nearest Neighbor (KNN) sebagai teknik klasifikasi utama untuk mengidentifikasi seseorang yang memiliki kemungkinan gejala depresi (Abimanyu, Bahtiar, and Adi Sarwoko 2023). K-Nearest Neighbors (KNN) merupakan salah satu cara yang termasuk dalam metode pengelompokan pada penambahan data dan juga menjadi salah satu metode yang potensial dalam mengklasifikasi data berdasarkan kemiripan atau kedekatan nilai-nilai atribut pada dataset (Sudarsono and Lestari 2020).

Penelitian ini bertujuan untuk menciptakan sebuah sistem analisis risiko depresi pada mahasiswa menggunakan algoritma KNN, yang dapat membantu dalam deteksi dini. Diharapkan, sistem ini mampu menyajikan informasi yang akurat sebagai alat diagnose diri bagi mahasiswa, dosen, serta pihak kampus dalam mendukung upaya pencegahan dan penanganan depresi (Utami 2024). Pada penelitian ini, saya menggunakan bahasa pemrograman *python* yang merupakan bahasa pemrograman berbasis objek yang dapat diinteraksi secara interaktif. Bahasa pemrograman *Python* dirancang khusus untuk memudahkan programmer dalam membuat program dengan efisiensi waktu, kemudahan pengembangan, dan kompatibilitas dengan sistem. Python dapat digunakan untuk membuat 16 aplikasi mandiri atau pemrograman skrip (Triono, Budi, and Abdillah 2023).

METODE PENELITIAN

Penelitian ini menerapkan metode kuantitatif menggunakan pendekatan klasifikasi berbasis data mining. Proses dimulai dengan studi pustaka dan diskusi dengan *domain expert* untuk menyusun kuesioner sebagai instrumen pengumpulan data. Data dikumpulkan dari 300 responden mahasiswa menggunakan skala Likert 4 poin, kemudian dilabeli menjadi tiga kelas risiko depresi: normal, ringan, dan berat.



Gambar 2.. 1 Alur Metode Penelitian

2.1 Teknik Pengolahan Data

Tahapan pengolahan data mencakup:

- Seleksi data : Memilih data relevan dari kuesioner.
- Preprocessing : Membersihkan data duplikat dan tidak konsisten.
- Transformasi Data : Mengubah data menjadi bentuk numerik, termasuk konversi jenis kelamin.
- Klasifikasi : Data analisis memakai algoritma K-Nearest Neighbor (KNN) dengan tetangga terdekat menggunakan parameter $K=3$ dan perhitungan jarak menggunakan *Euclidean Distance*.
- Evaluasi : Modul diperiksa dengan Matriks Kebingungan dengan metrik akurasi, presisi, dan recall.

2.2 Implementasi Sistem

Sistem dibangun menggunakan Python (untuk preprocessing) dan PHP CodeIgniter 3 (untuk antarmuka web). Dari 302 data, dilakukan pembagian data mencapai 80% 80% digunakan sebagai data latih dan 20% sebagai data uji. Pengujian model dilakukan dengan validasi menggunakan *confusion matrix*. Sistem mampu mencapai performa klasifikasi yang baik dengan akurasi 84%, presisi 88%, dan recall 82%.

HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset

Penelitian ini memanfaatkan data hasil kuesioner dari 300 responden yang merupakan mahasiswa aktif dengan rentang usia 19 – 24 tahun. Kuesioner terdiri dari 10 indikator risiko depresi, seperti masalah keuangan, hubungan social, perasaan putus asa, hingga tingkat kepanikan dan mudah tersinggung. Setiap indikator dinilai menggunakan skala likert 4 poin (0-3), sehingga total skor maksimum per responden adalah 30 poin.

3.2 Metode Pelabelan

Label kategori risiko depresi ditentukan berdasarkan total skor kuesioner tiap responden. Tiga kategori yang digunakan dalam klasifikasi adalah :

1. Normal : Skor rendah, tidak menunjukkan gejala depresi signifikan.
2. Ringan : Skor sedang, menunjukkan gejala awal depresi.
3. Berat : Skor tinggi, mengindikasikan gejala depresi yang cukup serius dan berkelanjutan.

Proses pelabelan ini dilakukan secara manual berdasarkan referensi dari domain expert di bidang kesehatan mental dan literatur sebelumnya.

3.3 Proses Transformasi

Penelitian ini memerlukan transformasi dikarenakan masih ada beberapa data string yang nantinya tidak bisa diproses oleh sistem. Data tersebut adalah kolom jenis kelamin. Jenis kelamin laki-laki diubah menjadi angka 1 dan jenis kelamin perempuan diubah dalam angka 2.

3.4 Proses Klasifikasi

Klasifikasi dilakukan dengan menghitung jarak *Euclidean* antara data uji dan data latih. Hasil dari 3 tetangga terdekat menentukan label klasifikasi Berdasarkan mayoritas. Proses ini dilakukan secara otomatis dalam sistem.

Berikut Rumus menghitung Jarak *Euclidean* :

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - Y_k)^2}$$

Dengan penjelasan :

- a. X : Nilai data latih
- b. Y : Nilai data uji
- c. N : Jumlah Fitur

Contoh Klasifikasi :

- Data uji dengan skor tertentu dihitung jaraknya terhadap data latih
- Tiga data terdekat dianalisis
- Mayoritas label diantara tiga tetangga menentukan prediksi akhir

3.5 Implementasi Algoritma KNN

Metode klasifikasi yang digunakan adalah K-Nearest Neighbor (KNN), dengan parameter $K = 3$. KNN bekerja dengan menghitung jarak Euclidean antara data uji dan data latih untuk menentukan kategori berdasarkan mayoritas tetangga terdekat.

Dataset dibagi menjadi :

80% (240 data) sebagai data training

20% (60 data) sebagai data testing

3.6 Hasil Evaluasi Model

Evaluasi model disini menggunakan *confusion matrix*. Berikut merupakan persamaan yang dipakai untuk menentukan nilai akurasi, presisi, dan recall.

- Akurasi = $\frac{TP + TN}{TP + TN + FP + FN}$
- Presisi = $\frac{TP}{TP + FP}$
- Recall = $\frac{TP}{TP + FN}$

	Prediksi Berat	Prediksi Normal	Prediksi Ringan
Aktual Berat	6	0	2
Aktual Normal	0	25	0
Aktual Ringan	0	7	20

Evaluasi

menggunakan

confusion matrix terhadap 60 data uji menghasilkan :

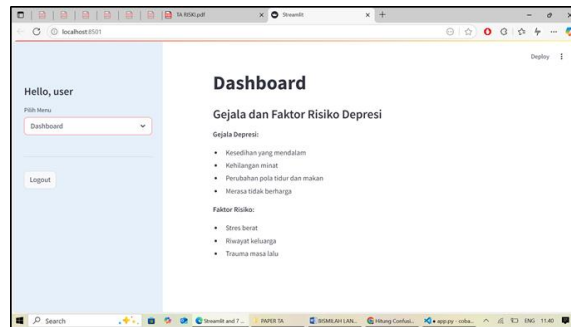
- Akurasi : 85% : Menunjukkan bahwa model mampu mengklasifikasikan data dengan benar pada sebagian besar kasus.
- Presisi : 90% : Menandakan bahwa sebagian besar prediksi kategori positif (misalnya depresi berat) adalah benar
- Recall : 83% : mengindikasikan bahwa model berhasil mendeteksi sebagian besar kasus nyata yang termasuk dalam kategori tersebut.

Hasil ini menunjukkan bahwa pendekatan KNN mempunyai performa yang cukup bagus dan efektif dalam mengklasifikasikan tingkatan risiko depresi pada mahasiswa, meskipun model sederhana dan berbasis kedekatan jarak.

3.7 Implementasi Sistem

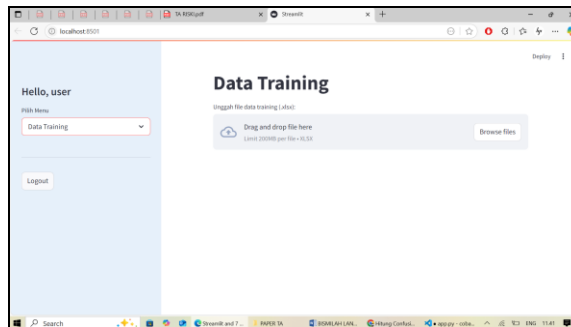
Dalam penelitian ini, implementasi alat uji metode K-Nearest Neighbor (K-NN) digunakan untuk mengklasifikasikan sebanyak 300 data. Data tersebut terbagi menjadi dua

sejumlah 240 data dipakai sebagai data *training* dan 60 data digunakan sebagai data *testing*. Berikut ini adalah tampilan dari sistem yang sudah dikembangkan oleh penulis.



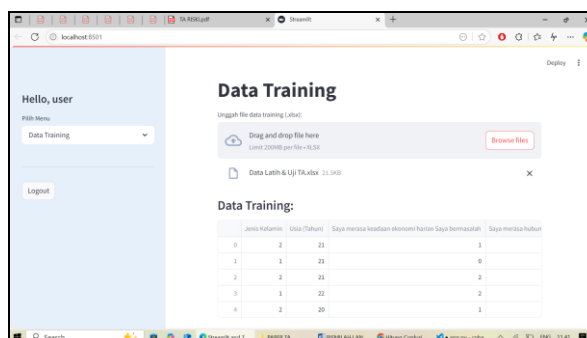
Gambar 3. 1 Tampilan Menu Dashboard

Pada gambar 3.1 merupakan halaman dashboard yang berisi tentang penjelasan gejala dan factor risiko depresi Sesuai dengan tema sistem tersebut yang memprediksi tingkat risiko depresi.



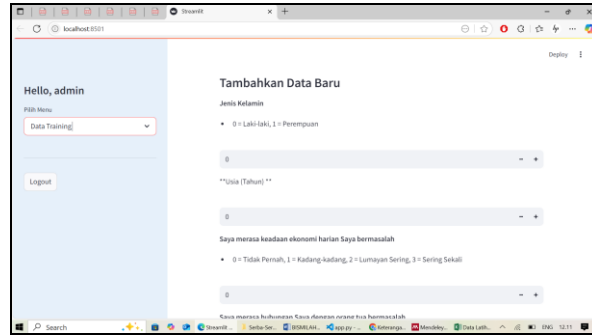
Gambar 3. 2 Tampilan Menu Data Training

Pada gambar 3.2 merupakan halaman data *training* untuk mengupload dataset yang akan diproses dan diprediksi dengan upload file berupa xlsx.



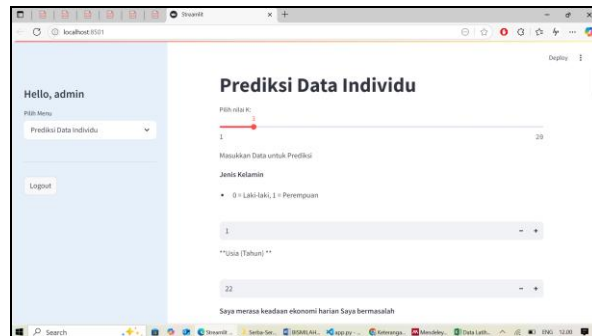
Gambar 3. 3 Tampilan Menu Menampilkan Data Training

Pada gambar 3.3 merupakan halaman isi dari data *training*, isinya menampilkan dataset yang akan diklasifikasikan didalam sistem tersebut.



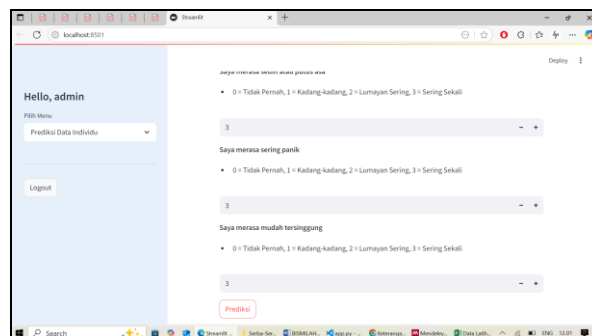
Gambar 3. 4 Tampilan Menu Tambah Data Baru

Pada gambar 3.4 merupakan menu data training yang berisi halaman tambah data baru. Jika ada data baru yang akan ikut di prediksi maka bisa mengisi terlebih dahulu pada halaman tambah data baru.



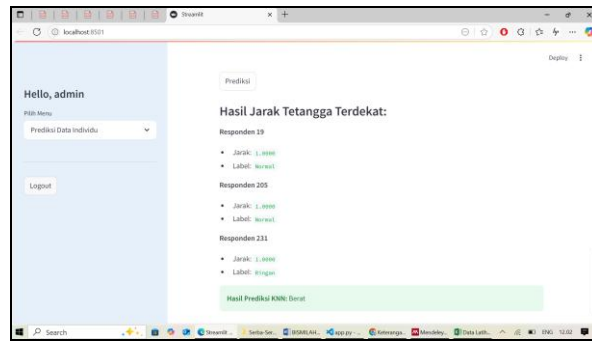
Gambar 3. 5 Tampilan Menu Prediksi Data Individu

Pada gambar 3.5 merupakan halaman prediksi data individu yang digunakan untuk memprediksi risiko depresi Berdasarkan tetangga terdekat (nilai k).



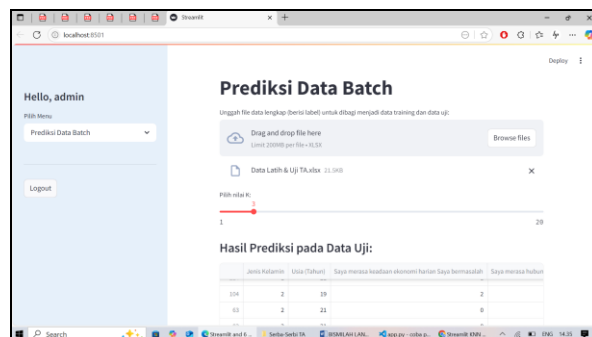
Gambar 3. 6 Tampilan Halaman Isi Prediksi Data Individu

Pada gambar 3.6 merupakan halaman isi prediksi data individu yang berisi kolom untuk diisi, kemudian jika sudah terisi semua maka klik tombol prediksi.



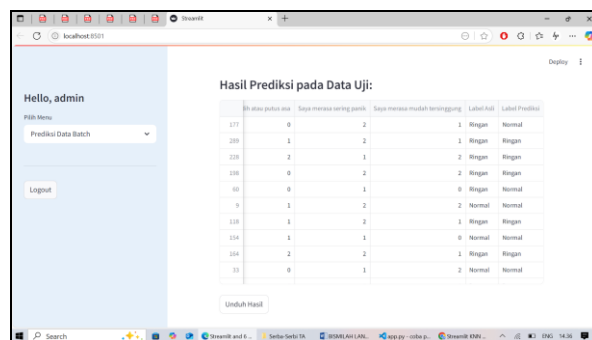
Gambar 3. 7 Tampilan Halaman Hasil Prediksi Data Individu

Pada gambar 3.7 merupakan halaman hasil prediksi data individu yang berisi prediksi hasil jarak tetangga terdekat, dengan Berdasarkan nilai K.



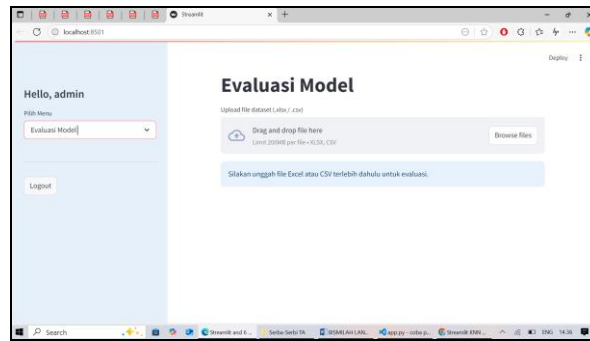
Gambar 3. 8 Tampilan Menu Prediksi Data Batch

Pada gambar 3.8 merupakan halaman prediksi data batch yang berfungsi untuk memproses label prediksi, dengan mengupload file dataset terlebih dahulu.



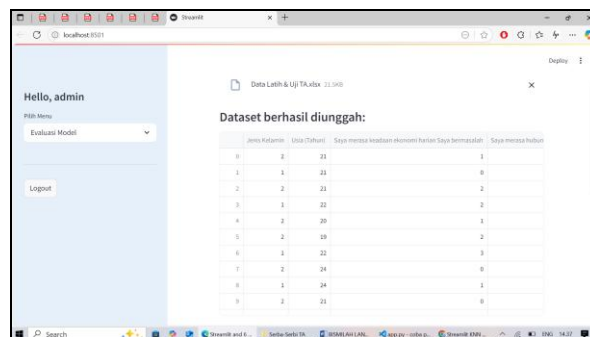
Gambar 3. 9 Tampilan Halaman Hasil Prediksi Data Uji

Pada gambar 3.9 merupakan halaman prediksi data uji yang menampilkan label baru/label prediksi.



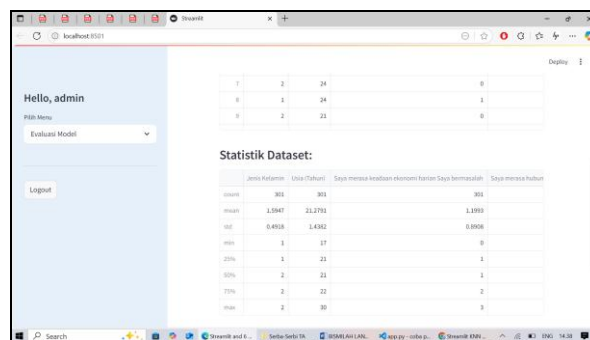
Gambar 3. 10 Tampilan Menu Evaluasi Model

Pada gambar 3.10 merupakan halaman Evaluasi Model, yang berfungsi untuk menghitung confusion matrix yaitu akurasi model, presisi, recall.



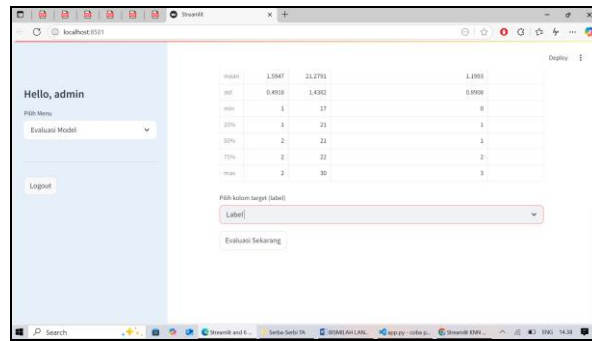
Gambar 3. 11 Tampilan Halaman Unggah Dataset

Pada gambar 3.11 merupakan halaman unggah dataset pada menu evaluasi model yang nantinya akan diproses *confusion matrix*.



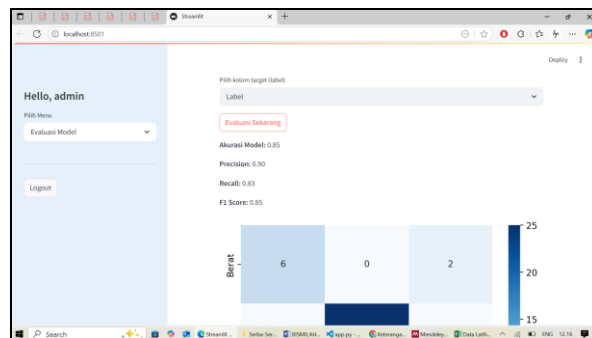
Gambar 3. 12 Tampilan halaman statistic dataset

Pada gambar 3.12 merupakan halaman statistic dataset yang berada pada menu evaluasi model, gunanya untuk melihat statistic dataset yang akan diproses.



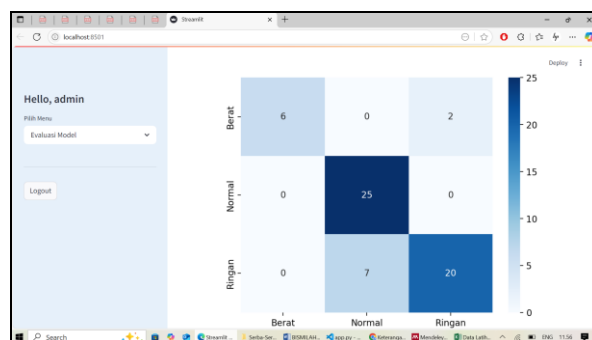
Gambar 3. 13 Tampilan Halaman Pemilihan Kolom Target

Pada gambar 3.13 merupakan halaman pemilihan kolom target pada menu evaluasi model, pada menu ini anda diwajibkan untuk memilih label untuk di hitung confusion matrixnya.



Gambar 3. 14 Tampilan Halaman Hasil Confusion Matrix

Pada gambar 3.14 merupakan halaman hasil confusion matrixnya pada menu evaluasi model, yang menampilkan akurasi model, presisi, recall.



Gambar 3. 15 Tampilan Halaman Grafik Confusion Matrix

Pada gambar 3.15 merupakan halaman grafik confusion matrix pada menu evaluasi model, yang gunanya untuk melihat hasil jumlah prediksi berat, prediksi normal, prediksi ringan.

3.8 Pembahasan

Nilai akurasi yang tinggi (85%) menandakan bahwa sistem dapat menjadi alat bantu yang andal dalam mendeteksi tingkat risiko depresi. Presisi sebesar 90% penting dalam konteks ini karena kesalahan dalam mengklasifikasikan seseorang sebagai “tidak depresi” padahal

sebenarnya berisiko, dapat berdampak serius. Nilai recall 83% juga menunjukkan sensitivitas yang cukup baik, namun masih ada ruang untuk perbaikan agar kasus yang benar-benar berisiko tidak terlewat.

Dengan hasil tersebut, sistem ini berpotensi digunakan sebagai alat prediksi individu atau alat skrining awal oleh pihak kampus, konselor, atau pusat layanan mahasiswa. Penggunaan sistem berbasis KNN ini memungkinkan deteksi dini depresi secara efisien, terutama untuk populasi besar, dan dapat diarahkan ke tindakan lanjutan seperti konseling atau pemeriksaan profesional.

KESIMPULAN

Berdasarkan hasil penelitian dan analisis dari yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Nearest Neighbor (KNN) mampu digunakan secara efektif untuk memprediksi tingkat risiko depresi pada mahasiswa berdasarkan data kuesioner. Dengan pembagian data latih dan data uji sebesar 80:20, model yang dibangun menghasilkan tingkat akurasi sebesar 85%, yang menunjukkan bahwa jumlah dan kualitas data latih memberikan kontribusi signifikan terhadap performa klasifikasi. Hasil ini juga memperlihatkan bahwa algoritma KNN memiliki potensi sebagai alat bantu dalam proses skrining awal kondisi psikologis mahasiswa. Implementasi sistem ini di lingkungan kampus diharapkan dapat membantu dalam mendeteksi gejala depresi sejak dini, sehingga mahasiswa yang berisiko dapat segera memperoleh perhatian dan penanganan yang lebih lanjut.

SARAN (OPTIONAL)

Adapun saran yang disampaikan pada penelitian ini sebagai berikut :

1. Penelitian selanjutnya disarankan untuk membandingkan algoritma K-Nearest Neighbor dengan metode klasifikasi lainnya berbasis probabilitas seperti Decision Tree, Support Vector Machine (SVM), atau Random Forest untuk memperoleh hasil akurasi yang lebih optimal.
2. Sistem klasifikasi yang dikembangkan dapat ditingkatkan menjadi aplikasi berbasis web atau mobile yang lebih ramah pengguna dan efisien dalam penggunaan memori serta penyimpanan, sehingga dapat digunakan secara lebih luas oleh institusi pendidikan.

DAFTAR PUSTAKA

- Abimanyu, Satria, Nurdin Bahtiar, and Eko Adi Sarwoko. 2023. "Implementasi Metode Support Vector Machine (SVM) Dan t-Distributed Stochastic Neighbor Embedding (t-SNE) Untuk Klasifikasi Depresi." *Jurnal Masyarakat Informatika* 14(2): 146–58.
- Purwanto, Dwi, Fatiha Sri, and Utami Tamad. 2022. "Gambaran Faktor Risiko Gejala Depresi Pada Mahasiswa Teknik Mesin Universitas Muhammadiyah Purwokerto." *Jurnal Pendidikan Teknik Mesin* 22(1): 25–28.
- Sudarsono, Bernadus Gunawan, and Sri Poedji Lestari. 2020. "Diagnosa Tingkat Depresi Mahasiswa Akhir Terhadap Penelitian Ilmiah Menggunakan Algoritma K-Nearest Neighbor." *Jurnal Media Informatika Budidarma* 4(4): 1094–99. <http://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/2448>.
- Triono, Agus, Apri Setia Budi, and Rafi Abdillah. 2023. "Implementasi Peretasan Sandi Vigenere Chipper Menggunakan Bahasa Pemrograman Python." *Jurnal JOCOTIS - Journal Science Informatika and Robotics* 1(1): 1–9. <https://jurnal.ittc.web.id/index.php/jumri>.
- Utami, Sanilisa Putri. 2024. *KLASIFIKASI KESEHATAN MENTAL USIA REMAJA MENGGUNAKAN ALGORITMA DECISION*. Jakarta.

