

KLASIFIKASI MENTAL HEALTH ISSUE PADA MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Faziz Ammarafi Aryamada^{1*}, Tommy Dwi Putra²

¹Informatika

(STMIK Amikom Surakarta)

^{1*}faziz.10334@mhs.amikomsolo.ac.id

²Informatika

(STMIK Amikom Surakarta)

²tommy@dosen.amikomsolo.ac.id

Abstrak— Masalah kesehatan mental yang semakin meningkat di kalangan mahasiswa menjadi semakin mengkhawatirkan. Jika tidak ditangani, masalah ini dapat berdampak buruk pada kinerja akademis, hubungan sosial, dan kualitas hidup secara umum. Penelitian ini mengklasifikasikan masalah kesehatan mental mahasiswa menggunakan algoritma K-Nearest Neighbor (KNN). Dataset yang digunakan berasal dari Kaggle, yang terdiri dari berbagai atribut seperti tingkat stres, kecemasan, depresi, dan faktor-faktor lain yang dapat mempengaruhi kesehatan mental. Dalam penelitian ini, data yang digunakan diproses dan dianalisis menggunakan perangkat lunak RapidMiner, yang memudahkan penerapan algoritma KNN. Hasil penelitian ini membuktikan bahwa algoritma KNN dapat menjadi cara yang efisien untuk mengklasifikasikan masalah kesehatan mental dengan tingkat akurasi yang tinggi dalam memprediksi hasil berdasarkan atribut yang diberikan. Temuan ini menyarankan bahwa deteksi dan klasifikasi dini masalah kesehatan mental dapat memungkinkan intervensi tepat waktu untuk meningkatkan kesejahteraan mental dan kualitas hidup mahasiswa. Hasil klasifikasi menggunakan KNN akurasi terbaik adalah 100%

Kata kunci— Kesehatan Mental, K-Nearest Neighbor, Klasifikasi, Mahasiswa, RapidMiner, Penggalian Data.

Abstract— The growing mental health problems in college students are increasingly worrying. If not attended to, these problems could result in devastating impacts on their school performance, social relations, and general life. This paper classified mental health problems of students using the K-Nearest Neighbor algorithm. The dataset used is from Kaggle, consisting of multiple attributes: stress levels, anxiety, depression, and other factors likely to affect mental health. In this research, the data being used will be processed and analyzed using the RapidMiner software, which makes it easy and simple to apply the KNN algorithm. This research results in proof that the KNN algorithm can be an efficient way for the classification of mental health issues with a high accuracy rate in predicting the outcome based on the given attributes. The findings suggest that early detection and classification of mental health problems would enable timely intervention to improve the mental well-being and quality of life for these students. The classification result using KNN with the best accuracy is 100%

Keywords— Mental Health, K-Nearest Neighbor, Clasification, College, Students, RapidMiner, Data Mining.

I. PENDAHULUAN

Kesehatan mental telah menjadi masalah penting dan dianggap serius dalam beberapa tahun terakhir, terutama di kalangan mahasiswa. Mahasiswa sering mengalami tekanan akademis, sosial, dan ekonomi, yang mungkin mudah dipengaruhi oleh kesehatan mental. Jika ditangani dan diamati dengan benar, masalah kesehatan mental dapat memengaruhi prestasi akademik dan hubungan sosial serta kualitas hidup [1]. Oleh karena itu, metode yang efektif untuk mendeteksi dan mengklasifikasikan

masalah kesehatan mental di kalangan mahasiswa harus dikembangkan [2].

Salah satu metode yang digunakan dalam mengklasifikasikan masalah kesehatan mental adalah melalui penggunaan algoritma K-Nearest Neighbor [3]. KNN adalah algoritma sederhana namun kuat yang mengklasifikasikan data menurut kedekatan data baru dengan data yang ada [4]. Keefektifan algoritma ini telah diuji pada spektrum aplikasi yang luas, termasuk bidang klasifikasi kesehatan mental dan sosialisasi terhadap masyarakat sekitar [5].

Dalam makalah ini, kita akan menggunakan algoritma KNN untuk mengklasifikasikan masalah kesehatan mental di kalangan mahasiswa. Atribut akan digunakan yang dapat memengaruhi kesehatan mental, seperti tingkat stres, kecemasan, depresi, atau fitur lainnya. Penelitian ini akan dilakukan pada perangkat lunak RapidMiner, yang menawarkan beragam alat untuk memproses dan menganalisis data dengan sukses [7].

Penelitian ini dilakukan untuk berkontribusi pada pendeteksian dini masalah kesehatan mental siswa, sehingga intervensi dini dapat dilakukan dan langkah-langkah yang tepat dapat dilakukan, diharapkan dapat berkontribusi pada peningkatan kesehatan mental dan kualitas hidup mereka diteliti.

Penelitian ini menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) mampu menghasilkan tingkat akurasi yang sangat tinggi dalam proses klasifikasi kesehatan mental mahasiswa. Meskipun demikian, hasil akurasi yang mencapai 100% perlu dianalisis secara lebih mendalam karena pada penelitian klasifikasi kondisi tersebut dapat dipengaruhi oleh karakteristik dataset maupun proses validasi model. Beberapa atribut, seperti *Mental Health History*, *Treatment*, dan *Mental Health Interview*, memiliki hubungan yang sangat erat dengan kondisi kesehatan mental sehingga berpotensi memberikan kontribusi yang sangat besar terhadap proses klasifikasi. Selain itu, penelitian ini belum menjelaskan secara rinci parameter yang digunakan, seperti nilai k , metode pembagian data pelatihan dan data pengujian, maupun teknik validasi model, sehingga aspek reproduktibilitas penelitian masih perlu ditingkatkan [8].

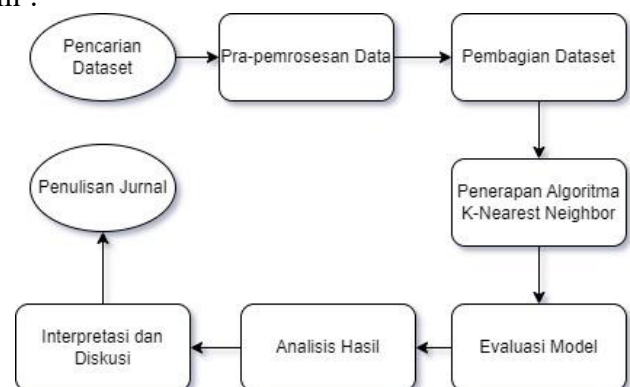
Pada tahap pengolahan data, penelitian menjelaskan bahwa dilakukan proses seleksi fitur untuk memilih atribut yang relevan berdasarkan analisis korelasi. Namun, hasil yang diperoleh menunjukkan bahwa seluruh atribut pada dataset masih digunakan pada proses klasifikasi sehingga belum terdapat pengurangan jumlah fitur yang digunakan. Bahkan beberapa atribut yang memiliki nilai korelasi sangat rendah tetap dipertahankan dalam pemodelan.

Oleh karena itu, proses seleksi fitur dapat dioptimalkan pada penelitian selanjutnya agar model yang dihasilkan menjadi lebih sederhana, efisien, dan memiliki kemampuan generalisasi yang lebih baik [9].

Evaluasi performa model pada penelitian ini masih berfokus pada nilai akurasi, waktu proses, dan penggunaan memori. Padahal, untuk penelitian klasifikasi, terutama pada bidang kesehatan mental, diperlukan metrik evaluasi yang lebih komprehensif, seperti *confusion matrix*, *precision*, *recall*, dan *F1-score*. Penggunaan metrik tersebut akan memberikan gambaran yang lebih lengkap mengenai kemampuan model dalam mengidentifikasi setiap kelas. Selain itu, perbandingan performa KNN dengan algoritma klasifikasi lainnya juga perlu dilakukan agar keunggulan maupun keterbatasan metode yang digunakan dapat dianalisis secara lebih objektif [10].

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode klasifikasi KNN (K- NEAREST NEIGHBOR). Untuk mendapatkan hasil dari klasifikasi diperlukan beberapa tahapan antara lain pengumpulan, penyeleksian dan pra-pemrosesan. Data yang telah diproses sebelumnya akan diklasifikasikan menggunakan algoritma KNN (K- NEAREST NEIGHBOR). Berikut merupakan langkah-langkah dalam penelitian ini :



Gambar 1. Metode Penelitian

A. Pengumpulan Data

Data ini diperoleh dari website kaggle yang memiliki 292.265 dataset dan 17 atribut. Dataset yang digunakan dapat diakses melalui link

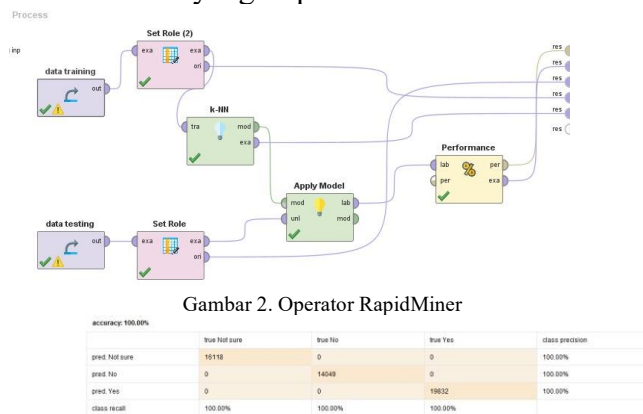
<https://www.kaggle.com/datasets/divaniazzahra/mental-health-dataset>. Tabel 1 merupakan atribut keseluruhan dari dataset yang diperoleh dari Kaggle.

Tabel 1. Atribut Dataset

Atribut	Keterangan
Atribut	Date
Timestamp	Male/Female
Gender	U.S (59%) dll
Country	Corporate, Student, Business &
Occupation	Housewife
Self employed,	Yes/No
Family history	Yes/No
Treatment	Yes/No
Days Indoors	1-14 Days
Growing Stress	Yes/No
Changes Habits	Yes/No
Mental Health History	Yes/Maybe/No
Mood Swings	Low/Medium
Coping Struggles	Yes/No
Work Interest	Yes/Maybe/No
Social Weakness	Yes/Maybe/No
Mental health interview	Yes/Maybe/No
Care options	Yes/No/Not Sure

B. Pengolahan Data

Tahap ini dilakukan karena ada atribut yang tidak relevan untuk digunakan, proses ini dilakukan untuk memilih atribut yang relevan pada *Mental Health History*. Gambar 2 merupakan operator RapidMiner untuk memilih atribut yang relevan, pada Gambar 3 dan Tabel 2 adalah atribut yang terpilih.



Gambar 2. Operator RapidMiner

Gambar 3. Hasil akurasi dari algoritma K-Nearest Neighbor



Gambar 4. Operator RapidMiner untuk mencari korelasi atribut

attribute	weight
Gender	0
Coping_Struggles	0.000
Growing_Stress	0.000
Mental_Health_History	0.000
Changes_Habits	0.000
Mood_Swings	0.000
Social_Weakness	0.000
Work_Interest	0.000
Occupation	0.000
Days_Indoors	0.000
self_employed	0.009
mental_health_interview	0.012
family_history	0.017
Timestamp	0.047
treatment	0.049
Country	0.171

Gambar 5. Hasil dari operator RapidMiner untuk mencari korelasi atribut

Tabel 2. Atribut yang terpilih

Atribut Terpilih	Keterangan
Timestamp	Date
Gender	Male/Female
Country	U.S (59%) dll
Occupation	Corporate, Student, Business & Housewife
Self employed,	Yes/No
Family history	Yes/No
Treatment	Yes/No
Days Indoors	1-14 Days
Growing Stress	Yes/No
Changes Habits	Yes/No
Mental Health History	Yes/Maybe/No
Mood Swings	Low/Medium
Coping Struggles	Yes/No
Work Interest	Yes/Maybe/No
Social Weakness	Yes/Maybe/No
Mental health interview	Yes/Maybe/No
Care options	Yes/No/Not Sure

C. Pemrosesan Data

Tahap ini dilakukan untuk memastikan pada atribut dataset yang dipakai tidak ada Missing

Value, pada Gambar 4 pengecekan Missing Value dilakukan menggunakan operator RapidMiner dan pada Tabel 3 merupakan hasil dari operator RapidMiner.



Gambar 6. Operator RapidMiner untuk mengecek Missing Value

Tabel 3. Cek Missing Value

Attribute	Missing Value
self_employed	2.48%
Timestamp	0.00%
care_options	0.00%
mental_health_interview	0.00%
treatment	0.00%
Country	0.00%
family_history	0.00%
Growing_Stress	0.00%
Gender	0.00%
Coping_Struggles	0.00%
Social_Weakness	0.00%
Changes_Habits	0.00%
Mood_Swings	0.00%
Work_Interest	0.00%
Occupation	0.00%
Days Indoors	0.00%

D. Penggunaan Data Mining

1. Data Mining

Data mining adalah proses mengekstraksi atau menggali data dan informasi dalam jumlah besar, yang sebelumnya tidak diketahui, namun memiliki wawasan mendalam dan bernilai, dari basis data besar yang kemudian digunakan untuk membuat keputusan bisnis yang lebih strategis [9].

2. RapidMiner

Rapid Miner adalah perangkat lunak yang dikembangkan oleh Dr. Markus Hofmann dari Blanchardstown Institute of Technology dan Ralf Klinkenberg dari rapid-i.com [12]. Perangkat ini memiliki antarmuka GUI (*graphical user interface*) yang memudahkan pengguna dalam mengoperasikan perangkat lunak, bersifat *open source*, dan dirancang khusus untuk keperluan data mining.

3. Klasifikasi

Klasifikasi adalah proses pencarian serangkaian model (fungsi) yang dapat digunakan untuk mendefinisikan dan mengelompokkan data atau konsep ke dalam kelas-kelas tertentu, dengan tujuan

menggunakan model tersebut untuk memprediksi kelas dari objek yang belum diketahui [4].

4. K-Nearest Neighbors (KNN)

K-nearest neighbors atau knn adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (train data sets), yang diambil dari k tetangga terdekatnya (*nearest neighbors*) [9]. Dengan k merupakan banyaknya tetangga terdekat. Teknik pencarian tetangga terdekat yang umum dilakukan dengan menggunakan formula jarak euclidean. Berikut beberapa formula yang digunakan dalam algoritma knn: *Euclidean Distance*, Jarak *Euclidean* adalah formula untuk mencari jarak antara 2 titik dalam ruang dua dimensi.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

Dimana :

$d(x,y)$: ini adalah jarak antara dua titik x dan y. Simbol d biasanya digunakan untuk merepresentasikan jarak (dari kata «distance»).

x ; ini adalah vektor yang merepresentasikan koordinat titik pertama dalam ruang fitur.

y : ini adalah vektor yang merepresentasikan koordinat titik kedua dalam ruang fitur.

$x_1, x_2, \dots, x_n$: komponen-komponen dari vektor x.

Jika ada n fitur dalam dataset, maka x memiliki n komponen.

$y_1, y_2, \dots, y_n$: komponen-komponen dari vektor y. Jika ada n fitur dalam dataset, maka y juga memiliki n komponen.

III. HASIL DAN PEMBAHASAN

Pada penelitian ini, algoritma K-Nearest Neighbor digunakan sebagai klasifikasi masalah kesehatan mental di kalangan mahasiswa. Dalam penelitian ini, data yang digunakan terdiri dari berbagai atribut relevan seperti tingkat stres, kecemasan, depresi, serta faktor-faktor lainnya yang dapat mempengaruhi kesehatan mental.

A. Data Processing

Tahap awal melibatkan pengumpulan data dan pengolahan data berasal dari dataset yang diambil dari situs Kaggle. Setelah melalui proses pra-pemrosesan, data tersebut kemudian diklasifikasikan menggunakan algoritma KNN. Pada tahap ini, atribut yang tidak relevan dihilangkan untuk memastikan bahwa hanya atribut yang paling signifikan yang digunakan dalam analisis [11].

B. Implementasi Algoritma KNN

Algoritma KNN diaplikasikan dengan mempertimbangkan jarak terdekat antara data baru dan data yang sudah ada. Dalam kasus ini, jarak Euclidean dipakai sebagai metrik utama untuk mengukur kedekatan antara titik-titik data. KNN dipilih karena kemampuannya yang sederhana namun efisien dalam menangani data dengan banyak fitur dan dapat memberikan hasil yang akurat dalam klasifikasi.

C. Akurasi Evaluasi

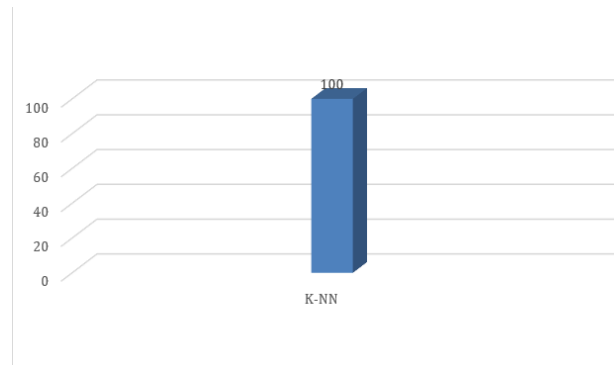
Kemudian, pengklasifikasian dilakukan dan hasil penerapan KNN menunjukkan algoritma ini dapat memberikan akurasi yang cukup tinggi. Akurasi diukur dengan perbandingan prediksi yang dihasilkan oleh model dengan data actual yang telah diketahui. Oleh karena itu, hasil ini mengindikasikan bahwa algoritma KNN dapat efektif untuk mendeteksi masalah kesehatan mental di kalangan mahasiswa.

D. Discussion

Penggunaan algoritma KNN dalam penelitian ini menunjukkan bahwa teknik ini dapat diandalkan untuk klasifikasi masalah kesehatan mental. Namun, penting untuk dicatat bahwa akurasi dari algoritma KNN dapat dipengaruhi oleh pemilihan nilai k (jumlah tetangga terdekat) dan kualitas data yang digunakan. Dalam konteks ini, pemilihan k yang optimal serta penanganan data yang bersih dan relevan sangat penting untuk mendapatkan hasil yang maksimal.

Kemungkinan pengembangan lebih lanjut juga dibuka oleh penelitian ini, seperti pengujian algoritma lain atau penggabungan KNN dengan teknik data mining lain untuk meningkatkan

performa dalam klasifikasi masalah kesehatan mental.



Gambar 7. Grafik Algoritma K-NN

Tabel 4. Penggunaan algoritma KNN Menggunakan RapidMiner

Algoritma	Waktu Proses	Ketelitian	Memori
K-NN	150 ms	100 %	200 KB

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa kesehatan mental merupakan permasalahan yang perlu mendapat perhatian, terutama di kalangan mahasiswa yang sering menghadapi berbagai tekanan, baik dari sisi akademik, sosial, maupun ekonomi. Oleh karena itu, diperlukan upaya deteksi dini agar masalah kesehatan mental dapat diketahui dan ditangani lebih cepat. Dalam penelitian ini, algoritma K-Nearest Neighbor (KNN) digunakan untuk mengklasifikasikan kondisi kesehatan mental mahasiswa berdasarkan atribut-atribut yang telah ditentukan. Proses pengolahan dan analisis data dibantu dengan menggunakan perangkat lunak RapidMiner sehingga penerapan algoritma KNN dapat dilakukan dengan lebih mudah dan terstruktur. Hasil penelitian menunjukkan bahwa metode KNN mampu memberikan tingkat akurasi yang cukup baik dalam mengklasifikasikan masalah kesehatan mental mahasiswa. Dengan demikian, penelitian ini diharapkan dapat menjadi salah satu referensi dalam mendukung upaya deteksi dini dan penanganan masalah kesehatan mental sehingga dapat membantu meningkatkan kualitas hidup mahasiswa.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Duta Bangsa Surakarta yang telah menyelenggarakan Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB) 2026 sebagai wadah bagi penulis untuk mempublikasikan hasil penelitian ini. Penulis juga menyampaikan terima kasih kepada STMIK Amikom Surakarta, khususnya kepada Ibu Ina Sholihah yang telah memberikan informasi kepada mahasiswa mengenai kesempatan publikasi pada kegiatan ini. Selain itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Tommy Dwi Putra, S.Kom., M.Kom. selaku dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan sehingga penulis dapat menyelesaikan penelitian dan penulisan jurnal ini dengan baik.

REFERENSI

- [1] Farhan Pratama, nst, Z. H., Zuriatul Khairi, Rahmaddeni, & Efrizoni, L. (2024). PERBANDINGAN ALGORITMA RANDOM FOREST DAN K-NEAREST NEIGHBOR DALAM KLASIFIKASI KESEHATAN MENTAL MAHASISWA. JURNAL ILMIAH BETRIK, 15(01 APRIL), 31–37.
- [2] Vera Artanti (2024). Classification of Cardiovascular Diseases Using the K-Nearest Neighbors (KNN) Algorithm Vol. 13 No. 2, June 2024, 12-27.
- [3] Septi Kenia Pita Loka, Arif Marsal (2023). Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Pada Balita. Journal Homepage: <https://journal.irpi.or.id/index.php/malcom> Vol. 3 Iss. 1 April.
- [4] Andi Maulida Argina (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. Indonesian Journal of Data and Science ISSN: 2715-9930 Vol 1, No 2, Juli.
- [5] Naja Maskuri, M., Harliana, H., Sukerti, K., & Herdian Bhakti, R. M. (2022). Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Prediksi Penyakit Stroke. Jurnal Ilmiah Intech : Information Technology Journal of UMUS, 4(01), 130–140. <https://doi.org/10.46772/intech.v4i01.751>
- [6] Fatoni (2017). Case Based Reasoning Diagnosis Penyakit Difteri dengan Algoritma K-Nearest Neighbor. Vol 4, No 3 Juli.
- [7] Dede Andri Muhammad Reza, Amril Mutoi Siregar, Rahmat Rahmat (2020). Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Kematian Akibat Penyakit Gagal Jantung. Vol. 3 No. 1 (2022): Scientific Student Journal for Information, Technology and Science.
- [8] Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. Indonesian Journal of Data and Science, 1(2), 39-43. <https://doi.org/10.33096/ijodas.v1i2.13>.
- [9] Intan Permata (2024). ANALISIS DATA MEDIS MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS. Duniadata.org Volume 1 (5).
- [10] Ahmad, I., Sulistiani, H., & Saputra, H. (2018). The Application Of Fuzzy K-Nearest Neighbour Methods For A Student Graduation Rate. Indonesian Journal of Artificial Intelligence and Data Mining, 1(1), 47–52.
- [11] Blanco C et al. Mental health of college students and their non-college-attending peers: Results from the National Epidemiologic Study on Alcohol and Related Conditions. Arch Gen Psychiatry. 2008;65(12):1429–37.
- [12] Tsania Khairani Aritonang (2020) PENGENALAN ALGORITMA PADA PEMBELAJARAN PEMROGRAMAN KOMPUTER. <https://doi.org/10.31219/osf.io/agnzm>