

KLASIFIKASI KARAKTERISTIK KEPERIBADIAN SISWA BERDASARKAN THE BIG FIVE PERSONALITY DENGAN MENGGUNAKAN METODE K- NEAREST NEIGHBOR (KNN)

Maruli Tua Silaen¹

¹²Program Studi Teknik Informatika, Fakultas Teknik, Universitas Tanjungpura

Jln. Prof. Dr.H.Hadari Nawawi, Pontianak, 78124

¹ marulituasilaen76@gmail.com,

Abstract

The problem faced in this study is that many students do not know their potential, making it difficult to develop their abilities [2]. Therefore, this study aims to assist students in recognizing their potential and developing their abilities by classifying student personalities based on report cards, The Big Five Personality questionnaire, knowledge values, and skills values as research parameters [3]. The method used in this study is the K-Nearest Neighbor (KNN) algorithm which is applied in an application using The Big Five Personality method [4]. Data was taken from a questionnaire containing 50 questions representing the five personality types, namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Data on report cards and student skill scores are used as supporting attributes in the classification process, while the target objectives of the classification results are personality types which are grouped into five types of personality types based on The Big Five Personality as labeling targets [6]. The most classification result is Conscientiousness personality. Testing is done by dividing the data into training data and test data randomly with a ratio of 8:2, and produces an accuracy of 95% with the confusion matrix method. The results of this study indicate that building a classification model requires good quality in the selection of research parameters that can affect the performance and accuracy of the model.

Keywords : K-Nearest Neighbor, Klasifikasi karakteristik, The Big Five Personality, Confusion Matrix

Abstrak

Permasalahan yang dihadapi dalam penelitian ini adalah banyaknya siswa yang tidak mengetahui potensi mereka sehingga sulit untuk mengembangkan kemampuan yang dimiliki [2]. Oleh karena itu, penelitian ini bertujuan untuk membantu siswa dalam mengenali potensi mereka dan mengembangkan kemampuan yang dimiliki dengan mengklasifikasikan kepribadian siswa berdasarkan nilai raport, kuesioner The Big Five Personality, nilai pengetahuan, dan nilai keterampilan sebagai parameter penelitian [3]. Metode yang digunakan dalam penelitian ini adalah algoritma K-Nearest Neighbor (KNN) yang diterapkan dalam sebuah aplikasi menggunakan metode The Big Five Personality [4]. Data diambil dari kuesioner yang berisikan 50 pertanyaan yang mewakili kelima jenis kepribadian, yaitu Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism. Data nilai raport dan nilai keterampilan siswa digunakan sebagai atribut pendukung dalam proses klasifikasi, sementara target tujuan dari hasil klasifikasi adalah jenis kepribadian yang dikelompokkan ke dalam lima jenis tipe kepribadian berdasarkan The Big Five Personality sebagai labeling target [6]. Hasil klasifikasi terbanyak adalah kepribadian Conscientiousness. Pengujian dilakukan dengan membagi data menjadi data latih dan data uji secara acak dengan perbandingan 8:2, dan menghasilkan akurasi sebesar 95% dengan metode confusion matrix. Hasil penelitian ini menunjukkan bahwa membangun model klasifikasi memerlukan kualitas yang baik dalam pemilihan parameter penelitian yang dapat mempengaruhi kinerja dan akurasi model.

Kata kunci : K-Nearest Neighbor, Classification of characteristics, The Big Five Personality, Confusion Matrix

1. PENDAHULUAN

Kepribadian merupakan faktor yang penting yang melekat pada seorang manusia, tetapi faktanya manusia sering lupamemahami arti kepribadian itu sendiri. Cabang ilmu psikologi yang mempelajari kepribadian, yaitu psikologi kepribadian, membutuhkan model deskriptif

(taksonomi) untuk menggambarkan konsep kepribadian. Taksonomi ini membantu para peneliti dalam menganalisis sumber utama dari kepribadian dengan lebih mudah daripada harus melakukan pemeriksaan terhadap ribuan atribut yang membuat masing-masing individu unik dan berbeda [1].

Pada jenjang pendidikan, siswa diharapkan untuk dapat lebih rajin dalam aktivitas belajar. Pendidikan yang bermutu dari di sekolah maupun dari rumah melalui orang tua tidaklah cukup jika tidak memahami potensi yang dimiliki oleh seorang anak. Tanpa pengetahuan tentang potensi seseorang, sumber daya manusia akan terus menurun. Potensi adalah sebuah kemampuan yang dipunyai oleh masing-masing orang yang memiliki kemungkinan atau kecenderungan untuk dikembangkan seperti prestasi dengan kemampuan yang terpendam, namun anak-anak seringkali mengalami kesulitan dalam memahami diri mereka sendiri [2].

Salah satu persoalan cukup yang sering muncul adalah kurangnya pengetahuan guru tentang kemampuan potensi siswa berdasarkan kepribadian yang mereka miliki. Ditambah lagi siswa juga tidak menyadari kemampuan yang mereka miliki. Hal ini menyebabkan banyak masalah, seperti potensi siswa tidak terwakili dengan baik karena ketidaktauan akan potensi siswa, sehingga setelah lulus sekolah siswa hanya mengikuti arus atau tren yang sedang populer tanpa mempertimbangkan potensi yang sebenarnya. Penelitian ini berfokus pada pengembangan karakter siswa, terutama siswa SMA [3]. Kelompok siswa SMA ini dipilih karena karena pada usia ini, adalah waktu yang tepat untuk menyusun rencana karir mereka. Penelitian dilakukan di SMA Negeri 01 Tayan Hulu. Pada penelitian ini akan mengembangkan pengelompokan kepribadian berdasarkan The Big Five Personality, ini adalah suatu pendekatan dalam psikologi kepribadian yang mengelompokkan kepribadian melalui analisis faktor yang dikembangkan oleh Allport dan Cattell. The Big Five Personality adalah suatu pendekatan untuk menganalisis kepribadian manusia melalui lima domain kepribadian yang telah dibentuk dengan menggunakan analisis factor.

Menurut teori kepribadian The Big Five,

manusia dikelompokkan menjadi lima tipe kepribadian, yaitu Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism [4]. Individu yang memiliki kepribadian Openness memiliki imajinasi yang aktif, sensitif terhadap perbedaan, rasa ingin tahu intelektual, dan kebebasan dalam berpendapat. Sementara itu, individu yang memiliki kepribadian Conscientiousness memiliki perencanaan yang matang, yang berkaitan dengan kontrol diri, organisasi, dan keterampilan dalam menyelesaikan tugas-tugas, yang biasanya terkait dengan prestasi dalam bidang karier dan akademik. Individu yang memiliki kepribadian Extraversion senang bersosialisasi. Sedangkan individu yang memiliki kepribadian Neuroticism cenderung merasakan perasaan-perasaan negatif seperti ketakutan, kekhawatiran, kesedihan, canggung, marah, bersalah, dan benci.

Untuk mengatasi permasalahan yang ada, diperlukan sebuah sistem untuk mengklasifikasikan kepribadian siswa berdasarkan The Big Five Personality agar potensi siswa dapat dikenali lebih awal. Langkah pertama yang harus dilakukan adalah mengetahui potensi siswa dengan cara mengelompokkan kepribadian siswa ke dalam kategori Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism dengan menggunakan teknik klasifikasi K- Nearest Neighbor (KNN) [5]. Dengan demikian, pihak terkait seperti pendidik di sekolah atau orang tua dapat mengatasi dan mengembangkan potensi yang ada. KNN memiliki karakteristik untuk melakukan klasifikasi terhadap objek yang berdasarkan pada data pembelajaran yang jaraknya paling dekat dengan objek tersebut [6]. Alat ukur kepribadian dapat dikorelasikan dengan beberapa variabel seperti nilai sikap, nilai pengetahuan, dan nilai keterampilan. Berdasarkan masalah yang ada, penyusun melakukan penelitian tentang Klasifikasi Karakteristik Kepribadian Siswa di SMA Negeri 01 Tayan Hulu Berdasarkan The Big Five Personality dengan Menggunakan Metode K-Nearest Neighbor (KNN) sebagai solusi yang baik untuk mengatasi permasalahan tersebut.

Berdasarkan perkembangan terkini dalam bidang psikologi dan teknologi informasi, penelitian terbaru telah menunjukkan minat yang signifikan dalam mengembangkan aplikasi yang dapat mengklasifikasikan kepribadian siswa dengan tingkat akurasi yang tinggi. Salah satu teknik yang digunakan dalam penelitian ini adalah metode K-Nearest Neighbor (KNN) yang merupakan metode pembelajaran mesin yang terkenal dan sering digunakan dalam pengklasifikasian data [7]. Selain itu, teori The Big Five Personality, yang merupakan salah satu teori

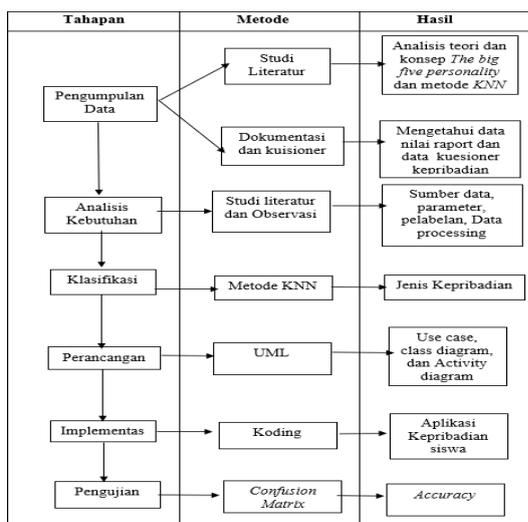
kepribadian paling terkemuka saat ini, digunakan sebagai dasar untuk mengklasifikasikan karakteristik kepribadian siswa. Penelitian ini bertujuan untuk mengembangkan sebuah aplikasi yang dapat membantu dalam menganalisis kepribadian siswa secara sistematis berdasarkan data yang telah ditentukan sebagai parameter penilaian. Dengan memadukan metode KNN dan teori The Big Five Personality, aplikasi ini dapat memberikan hasil yang akurat dan baik dalam mengklasifikasikan karakteristik kepribadian

siswa. Oleh karena itu, penelitian ini dianggap sebagai state of the art dalam pengembangan aplikasi untuk menganalisis kepribadian siswa secara sistematis menggunakan teknik-teknik terbaru dalam bidang psikologi dan teknologi informasi.

2. METODOLOGI PENELITIAN

2.1. Skema Alur Penelitian

Berikut adalah tahapan penelitian yang akan dilakukan, seperti yang ditunjukkan pada Gambar 1. Tahapan penelitian tersebut adalah sebagai berikut terdiri dari enam langkah dimana masing-masing langkah diikuti oleh metode dan hasil.



Gambar 1. Metodologi Penelitian

2.2. Pengumpulan Data

Penelitian ini melakukan pengumpulan data untuk memperoleh informasi yang akan digunakan selama proses penelitian. Terdapat dua metode pengumpulan data yang digunakan, yaitu studi literatur dan dokumentasi [15]. Studi literatur dilakukan dengan mempelajari teori,

data, atau informasi riset melalui sumber bacaan seperti jurnal ilmiah, buku referensi, dan bahan publikasi lain yang tersedia. Dalam studi literatur penelitian ini, diantaranya akan dipelajari teori dan konsep The Big Five Personality serta metode Klasifikasi K-Nearest Neighbor. Sedangkan studi dokumentasi dan kuisisioner digunakan untuk mengumpulkan data nilai raport kelas 10 semester ganjil dan data kuisisioner kepribadian

2.3. Analisa Data

Analisis kebutuhan dilakukan untuk mengidentifikasi elemen yang dibutuhkan dan menentukan penelitian yang akan dilakukan. Hasil pengumpulan data melalui studi literatur dan dokumentasi menunjukkan bahwa poin utama penelitian ini adalah mengembangkan potensi siswa berdasarkan kepribadian

2.4. Tahapan Klasifikasi

A. Dataset

Data yang digunakan dalam penelitian ini adalah nilai raport kelas 10 semester ganjil tahun ajaran 2021/2022. Proses penilaian pembelajaran di sekolah didasarkan pada peraturan Menteri Pendidikan Nasional nomor 20 tahun 2007 tentang Standar Penilaian Pendidikan yang menggunakan teknik dan instrumen penilaian seperti pendidik melakukan penilaian terhadap hasil belajar peserta didik dengan menggunakan berbagai teknik penilaian, di antaranya tes, observasi, penugasan perseorangan atau kelompok, dan bentuk lain yang sesuai dengan karakteristik kompetensi dari masing-masing siswa dan tingkat perkembangan peserta didik [8]. Teknik tes yang dapat digunakan adalah tes tertulis, tes lisan, dan tes praktik atau tes kinerja. Observasi atau pengamatan dapat dilakukan selama proses pembelajaran atau di luar kegiatan pembelajaran [9]. Penugasan perseorangan atau kelompok dapat berupa tugas rumah atau proyek. Berdasarkan penilaian pembelajaran siswa, terdapat dua parameter utama yaitu Pengetahuan dan Keterampilan yang dijadikan sebagai bahan penelitian agar penelitian lebih spesifik dan termasuk dalam aspek-aspek pembentukan kepribadian siswa sebagai atribut pendukung. Selanjutnya, dalam proses klasifikasi diperlukan data jenis kepribadian siswa yang dijadikan sebagai label target yang didapat dari hasil kuisisioner kepribadian siswa. Untuk mengetahui jenis kepribadian siswa, digunakan metode The Big Five Personality yang akan mengelompokkan kepribadian siswa ke dalam 5 jenis kepribadian seperti Openness, Conscientiousness, Agreeableness, Extraversion, dan Neuroticism

[10]. Data kepribadian diperoleh dengan cara mengisi form kuesioner di situs openpsychometrics.org yang terdiri dari 50 pertanyaan yang dikembangkan oleh Goldberg pada tahun 1992. Data jenis kepribadian akan digunakan sebagai label data pembelajaran siswa dalam proses klasifikasi. Data yang digunakan berisi 100 dataset yang terbagi atas 80 data training dan 20 data testing. Data latih berisi data nilai pengetahuan, nilai keterampilan, dan jenis kepribadian, sedangkan data uji berisi data nilai pengetahuan dan nilai keterampilan yang belum diberi label jenis kepribadian. Jenis kepribadian diberi label dengan menggunakan metode K-Nearest Neighbor dengan menghitung nilai jarak kedekatan atau kemiripan antara data training dan data testing dengan Euclidean Distance. Untuk pengujian, digunakan split validation untuk membagi dataset dan confusion matrix untuk menentukan Akurasi, recall, dan Precision [11]. Data yang digunakan dalam penelitian ini menggunakan 100 dataset nilai raport dari kelas 10 semester ganjil

C. KETERAMPILAN
Kriteria Ketuntasan Minimal = 65

No	Mata Pelajaran	Nilai	Predikat	Keterampilan
				Deskripsi
Kelompok A				
1	Pendidikan Agama, Katolik dan Budi Pekerti	15	D	Memiliki penguasaan sebagian keterampilan, terutama dalam Melakukan aktivitas (misalnya menulis refleksi)
2	Pendidikan Pancasila dan Kewarganegaraan			
3	Bahasa Indonesia	65	C	Memiliki penguasaan keterampilan cukup baik, terutama dalam Menginterpretasi teks laporan hasil observasi
4	Matematika (Umum)	77	B	Memiliki penguasaan keterampilan cukup baik, bahkan sangat baik dalam Menyusun masalah yang berkaitan dengan pertidaksamaan rasional dan rasional suku rasional
5	Sejarah Indonesia	11	D	Memiliki penguasaan sebagian keterampilan, terutama dalam perubahan dan keterkaitan dalam menggali peristiwa sejarah
6	Bahasa Inggris			
Kelompok B				
1	Seni Budaya	65	C	Memiliki penguasaan keterampilan cukup baik, terutama dalam Mengakan ide/gagasan sesuai konsep
2	Pendidikan Jasmani, Olahraga, dan Kesehatan			
3	Prakarya dan Kewirausahaan	70	C	Memiliki penguasaan keterampilan cukup baik, bahkan terampil dalam Menyusun perencanaan usaha pengolahan makanan awetan
Kelompok C				
1	Matematika (Peminatan)	65	C	Memiliki penguasaan keterampilan cukup baik, terutama dalam Menyajikan dan menganalisis tren masalah yang berkaitan dengan fungsi eksponensial dan fungsi logaritma
2	Fisika	11	D	Memiliki penguasaan sebagian keterampilan, terutama dalam Menyajikan data dan grafik hasil percobaan GLB dan GLBB
3	Kimia			
4	Biologi	65	C	Memiliki penguasaan sebagian keterampilan, terutama dalam observasi tingkat keanekaragaman hayati di Indonesia
5	Bahasa dan Sastra Inggris	11	D	Memiliki penguasaan sebagian keterampilan, terutama dalam Menyusun teks lisan terkait kesehatan
6	Bahasa dan Sastra Indonesia			

Tabel interval predikat berdasarkan KKM

KKM	Predikat			
	D	C	B	A
65	Nilai < 70	70 ≤ Nilai < 80	80 ≤ Nilai < 90	Nilai ≥ 90

Gambar 2. Data nilai keterampilan

B. PENGETAHUAN
Kriteria Ketuntasan Minimal = 65

No	Mata Pelajaran	Nilai	Predikat	Pengetahuan
				Deskripsi
Kelompok A				
1	Pendidikan Agama, Katolik dan Budi Pekerti	65	D	Memiliki penguasaan sebagian pengetahuan cukup baik dalam Memahami konsep/konsep diri sebagai citra, perlu peningkatan dalam Memahami jati diri sebagai perempuan atau laki-laki
2	Pendidikan Pancasila dan Kewarganegaraan			
3	Bahasa Indonesia	70	C	Memiliki penguasaan pengetahuan yang cukup baik dalam Mengevaluasi teks anekdot dari aspek media sosial, bahkan pengetahuan Mengidentifikasi laporan hasil observasi baik
4	Matematika (Umum)	65	C	Memiliki penguasaan pengetahuan yang cukup baik, terutama Menyusun sistem persamaan linear tiga variabel dari masalah kontekstual
5	Sejarah Indonesia	11	D	Memiliki penguasaan sebagian pengetahuan, terutama dalam berpikir kronologis, diakronik, sinkronik, ruang, dan waktu
6	Bahasa Inggris			
Kelompok B				
1	Seni Budaya	65	C	Memiliki penguasaan pengetahuan yang cukup baik, terutama Memahami konsep, teknik dalam ragam gerak tari tradisi
2	Pendidikan Jasmani, Olahraga, dan Kesehatan			
3	Prakarya dan Kewirausahaan	70	C	Memiliki penguasaan pengetahuan yang cukup baik dalam Menganalisis evaluasi kegiatan budidaya ternak/peternakan, bahkan pengetahuan Memahami perencanaan usaha produk teknologi transportasi baik
Kelompok C				
1	Matematika (Peminatan)	65	C	Memiliki penguasaan pengetahuan yang cukup baik, terutama Mendeskripsikan dan membuat penyajian fungsi eksponensial dan fungsi logaritma menggunakan masalah kontekstual, serta keberfektifannya
2	Fisika	11	D	Memiliki penguasaan sebagian pengetahuan, terutama dalam Menjelaskan konsep ilmu Fisika dan perannya dalam kehidupan
3	Kimia			
4	Biologi	65	C	Memiliki penguasaan sebagian pengetahuan, terutama dalam Mengidentifikasi struktur, cara hidup, reproduksi bakteri
5	Bahasa dan Sastra Inggris	22	D	Memiliki penguasaan sebagian pengetahuan cukup baik dalam teks recount lisan dan tulis, dalam bentuk biografi, perlu peningkatan dalam memberi dan meminta informasi terkait hubungan sebab
6	Bahasa dan Sastra Indonesia			

Gambar 3. Data nilai Pengetahuan

Data kuesioner kepribadian yang didasarkan pada The Big Five Personality diperoleh dari hasil

tes kepribadian yang tersedia di situs openpsychometrics.org [12], yang dikembangkan oleh Goldberg pada tahun 1992. Terdapat 50 pertanyaan yang masing-masing memiliki lima pilihan jawaban, yaitu sangat setuju, setuju, netral, tidak setuju, dan sangat tidak setuju. Form kuesioner terdiri atas 50 pertanyaan yang terorganisir berdasarkan lima kepribadian. Pertanyaan nomor 1, 11, 16, 21, 26, 3, 36, 41, 46 mewakili skala domain ekstraversi, pertanyaan nomor 2, 7, 12, 17, 22, 27, 32, 37, 42, 47 mewakili skala domain sikap yang mudah bersetuju, pertanyaan nomor 3, 8, 13, 18, 23, 28, 33, 38, 43, 48 mewakili skala konsistensi, pertanyaan nomor 4, 9, 14, 19, 24, 29, 34, 39, 44, 49 mewakili neuroticisme, dan pertanyaan nomor 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 mewakili kebukaan. Form kuesioner dapat dilihat pada Tabel 1 beberapa contoh pertanyaan kuesioner.

TABEL 1. KUESIONER KEPRIKIBADIAN

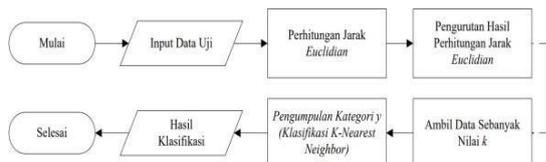
NO	Pertanyaan
1.	Saya adalah kehidupan pesta.
2.	Saya merasa sedikit perhatian terhadap orang lain.
3.	Saya selalu siap.
4.	Saya mudah stres.
5.	Saya memiliki kosakata yang kaya.
6.	Saya tidak banyak bicara.
7.	Saya tertarik pada orang.
8.	Saya meninggalkan barang-barang saya di sekitar.
9.	Saya santai sebagian besar waktu.
10.	Saya mengalami kesulitan memahami ide-ide abstrak.

Seluruh tahapan dijelaskan pada bagian 3 termasuk bagaimana melakukan pengembangan atau implementasi dan pengujian. Jika penjelasan proses penelitian menggunakan gambar dan tabel, maka gambar dan tabel harus disajikan dengan judul tabel dan gambar disertai dengan nomor urut. Contoh tabel seperti pada Tabel 1. Setiap tabel diberikan judul tabel yang diletakkan di atas tabel dengan *style* rata kiri, sedangkan judul gambar diletakkan di bawah gambar dengan *style* rata tengah. Rujukan tabel dan gambar di dalam teks diketik dengan nomer urut tabel dan gambar

dengan huruf awal kapital, seperti Tabel I.

B. Preprocessing

Preprocessing dilakukan untuk mengubah data mentah yang diperoleh dari proses pengambilan data menjadi format data yang lebih efisien untuk digunakan dalam proses klasifikasi [13]. Setelah melalui tahap preprocessing, data tersebut disimpan sebagai data baru dalam



format .csv dan terdiri dari 100 data set yang akan digunakan dalam proses pemodelan. Adapun tahapan-tahapan yang dilakukan dalam proses ini pertama yaitu Pembersihan data merupakan tahap dari preprocessing yang melibatkan analisis kualitas data. Proses ini juga dapat mengoreksi atau menghapus data yang tidak diperlukan untuk dipersiapkan sebagai data yang akan digunakan dalam pembuatan model [14]. Sebagai contoh, data nilai raport yang tidak memiliki data kepribadian harus dihapus agar dataset memiliki komposisi yang sama dari masing-masing data. Tahapan kedua penggabungan data dilakukan untuk menggabungkan data-data yang telah didapatkan seperti data nilai raport yang memiliki format data excel dan data kuesioner kepribadian memiliki format data docx. Tujuan dari dilakukannya penggabungan data agar memiliki format yang sama. Tahapan ketiga pelabelan data merupakan tahap mengidentifikasi data mentah dan menambahkan label, seperti jenis kepribadian dominan, pada data nilai raport untuk memberi makna dan informasi konteks sehingga model pembelajaran mesin dapat memahami dan menggunakannya. Tahapan keempat Split data adalah suatu teknik yang digunakan untuk membagi dataset menjadi dua bagian, yaitu data latih dan data uji, secara acak dengan perbandingan 80:20. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk mengevaluasi model tersebut. Ini dilakukan untuk menghindari plagiat dan memastikan bahwa model yang dihasilkan dapat bekerja dengan baik pada data yang belum pernah dilihat sebelumnya [15].

C. Klasifikasi

Penciptaan model dengan menggunakan metode terbaik yang sesuai dengan data dari beberapa metode pengklasifikasian yang telah digunakan oleh peneliti sebelumnya, yaitu K-Nearest Neighbor (KNN) [16]. Pemodelan dengan

menggunakan algoritma KNN dipilih karena data yang digunakan adalah data numerical. Proses klasifikasi kepribadian dengan menerapkan algoritma KNN menggunakan pembelajaran supervised learning, di mana hasil dari data yang diuji dikategorikan berdasarkan keanggotaan terdekat yang terbanyak dari data uji. Pada tahap uji, jarak antara data uji dan data latih dihitung dan sejumlah keanggotaan terdekat diambil [17]. Data uji diprediksi masuk ke dalam keanggotaan terdekat yang terbanyak dari data latih. Alur perhitungan K-Nearest Neighbor dapat dilihat pada gambar 4.

Gambar 4. Alur Perhitungan K-Nearest Neighbor

3. HASIL DAN PEMBAHASAN

Hasil klasifikasi kepribadian siswa dengan menggunakan algoritma K-Nearest Neighbor (KNN) berdasarkan The Big Five Personality telah diuji dengan perhitungan jarak Euclidean. Uji coba ini menggunakan 100 dataset yang terdiri dari 80 data latih dan 20 data uji, dengan mengambil 2 atribut yaitu nilai pengetahuan dan nilai keterampilan. Data latih yang terdiri dari 80 data. Berikut beberapa data latih pada tabel II berikut.

TABEL 2. DATA LATIH

NO	NAMA	PENGETAHUAN	KETERAMPILAN	KEPRIBADIAN
1	ANGGRAENI GHEA SAPUTRI	80,20	79,73	Openness
2	ARNOLDUS YANSEN BOTAMA LUBUR	81,40	79,40	Neuroticism
3	AURELIA SHOSELA GIZSCKA JUMIN	81,07	79,20	Extraversion
4	BENEDICTUS APRILIAN RAYARUNA	80,93	78,73	Neuroticism
5	BENEDIKTUS ALFONS SOWADA	82,40	80,80	Conscientiousness

Tabel II menunjukkan distribusi data kepribadian pada data latih yang digunakan untuk melatih model atau algoritma dalam proses pembelajaran mesin. Data latih terdiri dari 5 kategori kepribadian, yaitu Openness, Conscientiousness, Extroversion, Agreeableness, dan Neuroticism. Jumlah masing-masing kategori ditunjukkan dalam kolom "Jumlah" dan persentase masing-masing kategori ditunjukkan dalam kolom "Persentase". Openness menunjukkan kemampuan seseorang dalam memahami dan menerima ide-ide baru serta

memiliki minat terhadap hal-hal baru dan beragam.

Dari tabel tersebut, terdapat 1 data kepribadian dengan Openness, atau 1.25% dari total data latih. Conscientiousness menunjukkan kemampuan seseorang dalam mengatur dan mengelola waktu dengan baik, serta memiliki sikap disiplin dan tanggung jawab. Dari tabel tersebut, terdapat 37 data kepribadian dengan Conscientiousness, atau 46.25% dari total data latih. Extroversion menunjukkan kemampuan seseorang dalam bersosialisasi dan berkomunikasi dengan orang lain, serta memiliki kepercayaan diri yang tinggi. Dari tabel tersebut terdapat 24 data kepribadian dengan *Agreeableness* menunjukkan kemampuan seseorang dalam bekerja sama dengan orang lain, serta memiliki sikap empati dan *cooperatif*. Dari tabel tersebut, terdapat 5 data kepribadian dengan *Agreeableness*, atau 6.25% dari total data latih. *Neuroticism* menunjukkan kemampuan seseorang dalam mengendalikan emosi dan menghadapi stress, serta memiliki tingkat stabilitas emosi yang rendah. Dari tabel tersebut, terdapat 13 data kepribadian dengan *Neuroticism*, atau 16.25% dari total data latih. Total data latih yang

ada adalah 80, seperti ditunjukkan oleh jumlah total dari masing- masing kategori kepribadian.

TABEL 3. DISTRIBUSI DATA LATIH

KEPRIBADIAN	JUMLAH	PERSENTASE
OPENNESS	1	1.25%
CONSCIENTIOUSNESS	37	46.25%
EXTROVERSION	24	30%
AGREEABLENESS	5	6.25%
NEUROTICISM	13	16.25%

Tabel III menunjukkan data yang digunakan untuk mengevaluasi kemampuan model atau algoritma dalam pembelajaran mesin. Data ini digunakan untuk mengukur kemampuan model atau algoritma dalam menyelesaikan masalah yang belum pernah dilihat sebelumnya. Jumlah total data uji adalah 20, yang ditunjukkan oleh jumlah baris didalam tabel. Nilai pengetahuan dan keterampilan siswa ditunjukkan dalam persentase, dengan skala 0-100%.

TABEL 4. DATA UJI.

NO	NAMA	PENGETAHUAN	KETERAMPILAN
1	MELIANA EVA	79.20	78.60
2	MARKO DIMUS DINATA	74.53	74.00
3	MARIANUS CIE PHEN	77.60	76.07
4	PUTRI HANDAYANI	81.67	81.33
5	NATAN ARDIAN JULIANTO	77.40	76.87

Perhitungan yang dilakukan adalah data uji dengan nama Meliana Eva terhadap seluruh data latih.

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

$$d = \sqrt{(80,20 - 77,40)^2 + (79,73 - 76,87)^2}$$

$$= 4.0024492$$

data lain. Jarak Euclidean merupakan salah satu metode untuk menghitung jarak antara dua titik dalam ruang n-dimensi. Hasil perhitungan ini dapat digunakan untuk mengidentifikasi titik yang paling dekat atau paling jauh dari titik acuan yang ditentukan.

TABEL 5. DATA UJI DATA JARAK EUCLIDEAN MELIANA EVA.

NO	NAMA SISWA	PENGETAHUAN	KETERAMPILAN	KEPRIBADIAN	JARAK	RANK
1	ERIN	77.53	76.80	NEUROTICISM	0.14	1
2	PETRODIKA MAHARANI	77.20	76.87	CONSCIENTIOUSNESS	0.2	2
3	VALENTINO FERDY P.H	77.00	76.93	CONSCIENTIOUSNESS	0.40	3
4	ELSYA VIVI CHRISTIN BARAHEN PUTRI	77.27	76.47	NEUROTICISM	0.42	4
5	CRISTINA AE. SIMARMATA	77.73	77.40	CONSCIENTIOUSNESS	0.62	5
6	JESI TASIANA	76.73	77.07	CONSCIENTIOUSNESS	0.69	6
7	CHARLES WINSSEN	77.60	76.07	NEUROTICISM	0.82	7
8	GLADIS GRACE	77.80	76.13	EXTROVERSION	0.84	8
9	CLEARESTA MARCHYA	77.33	77.80	EXTROVERSION	0.93	9
10	MERVA	78.27	76.53	CONSCIENTIOUSNESS	0.93	10
11	OKTAVIA MEISA	76.33	76.93	EXTROVERSION	1.07	11
12	DANIEL SAPUTRA	76.67	75.93	CONSCIENTIOUSNESS	1.19	12

Setelah mengurutkan nilai jarak Euclidean, pencarian nilai k dilakukan dengan menggunakan nilai k = 3. Sebanyak 3 tetangga mayoritas teratas yang terpilih sebagai hasil keputusan yang akan digunakan, yaitu sesuai dengan perhitungan yang terlihat pada tabel VI, yang menunjukkan bahwa kepribadian *Conscientiousness* merupakan kepribadian mayoritas dari 3 tetangga terdekat. Dengan demikian, dapat disimpulkan bahwa Meliana Eva memiliki kepribadian *Conscientiousness* berdasarkan proses pengklasifikasian.

TABEL 6. HASIL KEPUTUSAN TETANGGA TERDEKAT MAYORITAS.

NO	NAMA	PENGETAHUAN	KETERAMPILAN	KEPRIBADIAN	JARAK	RANK
1	ERIN	77.53	76.80	NEUROTICISM	0.147	1
2	PETRODIKA MAHARANI	77.20	76.87	CONSCIENTIOUSNESS	0.2	2
3	VALENTINO FERDY P.H	77.00	76.93	CONSCIENTIOUSNESS	0.404	3

Perhitungan yang dilakukan menggunakan data uji Meliana Eva terhadap data latih Anggraeni Ghea Saputri menghasilkan nilai jarak Euclidean sebesar 4.0024492. Setelah selesai melakukan perhitungan dengan rumus yang sama pada semua data uji terhadap data latih, hasilnya diurutkan berdasarkan nilai jarak Euclidean terkecil hingga terbesar yang ditampilkan dalam Tabel IV. Langkah-langkah yang sama diulangi untuk Pada Tabel V menunjukkan hasil perhitungan jarak Euclidean dari data Meliana Eva dengan data semua data pengujian, di mana setiap data pengujian akan mengalami proses perhitungan jarak terhadap delapan puluh data latih. Setelah semua proses selesai, hasil akhir pengklasifikasi data pengujian akan ditampilkan pada tabel VII.

TABEL 7. HASIL KLASIFIKASI DATA UJI

NO	NAMA	PENGETAHUAN	KETERAMPILAN	KEPRIBADIAN
1	MARIANUS CIE PHEN	77.60	76.07	NEUROTICISM
2	MARKO DIMUS DINATA	74.53	74.00	NEUROTICISM
3	MELIANA EVA	79.20	78.60	AGREEABLENESS
4	PUTRI HANDAYANI	81.67	81.33	CONSCIENTIOUSNESS
5	NATAN ARDIAN JULIANTO	77.40	76.87	CONSCIENTIOUSNESS

6	NOVITA DEWI	78.13	77.53	CONSCIENTIOUSNESS
7	PUSMA ELIK	74.13	74.93	CONSCIENTIOUSNESS
8	REKONELIS RIKI	75.20	75.13	EXTROVERSION
9	TUSIMAN	72.93	73.00	EXTROVERSION
10	VIKTORIUS	76.67	75.87	CONSCIENTIOUSNESS

Tabel VIII menunjukkan bagaimana data uji yang digunakan dalam proses pelatihan dan pengujian model atau algoritma terdistribusi. Distribusi data uji merupakan jumlah dan persentase dari setiap kelas data yang terdapat dalam data uji. Hal ini penting untuk diketahui karena dapat mempengaruhi keakuratan hasil dari pelatihan dan pengujian model atau algoritma.

TABEL 8. DISTRIBUSI DATA UJI

Kepribadian	Jumlah	Persentase
Openness	0	0%
Conscientiousness	11	5%
Extroversion	4	20%
Agreeableness	1	5%
Neuroticism	4	20%

4. Kesimpulan dan Saran

Penelitian yang dilakukan untuk membangun model klasifikasi kepribadian siswa menggunakan metode *K-Nearest Neighbor* dan *The Big Five Personality* telah menghasilkan beberapa kesimpulan. Pertama, penelitian ini fokus pada implementasi algoritma *K-Nearest Neighbor* untuk mengetahui potensi siswa berdasarkan kepribadian mereka yang terdiri dari nilai pengetahuan dan nilai keterampilan yang diperoleh di sekolah. Hasil klasifikasi terbanyak adalah menghasilkan kepribadian *Conscientiousness*. Kedua, data hasil pembelajaran siswa yang terdiri atas nilai pengetahuan dan nilai keterampilan digunakan sebagai parameter penelitian karena kedua parameter tersebut merupakan bagian dari aspek-aspek yang membentuk kepribadian siswa. Ketiga, data nilai raport dan data kuesioner *The Big Five Personality* yang diambil dari bulan September 2022 terdiri dari 100 dataset yang kemudian dibagi menjadi data latih dan data uji. Data latih memiliki labeling data berupa jenis kepribadian dominan, sedangkan data uji tidak memiliki labeling data

yang akan digunakan untuk menguji sistem yang dibangun. Keempat, data uji yang digunakan terdiri atas 80 dataset yang terdiri dari 1 kepribadian *openness*, 37 kepribadian *conscientiousness*, 24 kepribadian *extraversion*, 5 kepribadian *agreeableness*, dan 13 kepribadian *neuroticism*. Sedangkan data uji sebanyak 20 dataset yang terdiri atas 0 kepribadian *openness*, 11 kepribadian *conscientiousness*, 4 kepribadian *extraversion*, 1 kepribadian *Agreeableness* dan 4 kepribadian *neuroticism*. Kelima, pengujian yang dilakukan menggunakan metode *confusion matrix* untuk mengukur *Accuracy*. Hasil pengujian dengan membagi data menjadi data latih dan data uji secara acak dengan perbandingan 8:2 menunjukkan *Accuracy* sebesar 95%. Dari hasil penelitian ini dapat diketahui bahwa membangun model klasifikasi memerlukan kualitas yang baik terutama dalam pemilihan parameter penelitian yang digunakan dalam proses klasifikasi yang mempengaruhi kinerja model serta nilai akurasi model.

Saran untuk pengembangan penelitian selanjutnya agar tahapan penelitian menjadi lebih efektif dan model yang dihasilkan lebih baik dapat dipertimbangkan, antara lain menggunakan data siswa yang lebih banyak, menambah parameter jenis kepribadian berdasarkan aspek-aspek yang mempengaruhi terbentuknya kepribadian siswa, menggunakan parameter data latih dengan komposisi yang seimbang pada setiap kelas, dan mencoba klasifikasi dengan algoritma pengklasifikasian lain.

Daftar Pustaka:

- [1] A. D. Lestari and A. W. Utomo, "The Influence of Personality and Emotional Intelligence on Employee Performance in PT. ABC," *Jurnal Bisnis dan Kewirausahaan*, vol. 5, no. 2, pp. 106-115, 2019.
- [2] S. P. Siregar and E. Nursidah, "Kajian Pemahaman Potensi Diri pada Siswa Sekolah Menengah Pertama," *Jurnal Konseling Dan Pendidikan*, vol. 6, no. 1, pp. 22-29, 2018.
- [3] S. D. A. Safitri and A. A. Marhaeni, "Efektivitas pembelajaran berbasis karakter untuk meningkatkan karakter siswa SMA," *Jurnal Pendidikan Sains dan Teknologi*, vol. 3, no. 1, pp. 22-29, 2019.
- [4] Nurbaity, N., & Wulandari, R. D. (2018). Hubungan antara kepribadian Big Five dengan prestasi akademik pada mahasiswa Fakultas Kedokteran Universitas Riau. *Jurnal Kedokteran Riau*, 3(3), 223-230.
- [5] W. Wardana, I. Ardiyanto, and R. Hidayatulloh, "Pengklasifikasian Kepribadian Mahasiswa Menggunakan Algoritma K-Nearest Neighbor Berbasis The Big Five Personality Traits," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 4, no. 2, pp. 76-84, 2019.
- [6] Krisandi, N. and Helmi, B.P., "Penggunaan Algoritma k-Nearest Neighbor dalam Menentukan Kelas Data Hasil Produksi Kelapa Sawit di PT. Minamas Kecamatan Parindu," *Publikasi Bimaster: Ilmiah Matematika, Statistika dan Penerapannya*, vol. 2, no. 1, pp. 1-7, Jan. 2013.
- [7] Windana, Fredy, Anitarakhmi Handaratri, and Mohammad Taufan Asri Zaen, "Implementasi Metode K-Nearest Neighbor untuk Pengenalan Buah Murbei dengan Korelasi Kandungan Antosianin," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 4, no. 1, pp. 78-86, 2021.
- [8] D. Destiana, Y. Suchyadi, and F. Anjaswuri, "Pengembangan instrumen penilaian untuk meningkatkan kualitas pembelajaran produktif di sekolah dasar," *J. Pendidikan Dan Pengajaran Guru Sekolah Dasar (JPPGuseda)*, vol. 3, no. 2, pp. 119-123, 2020.
- [9] H. Hidayat, "Tes dan Pengukuran dalam Pendidikan," *Jurnal Penjaminan Mutu: Majalah Ilmiah Bidang Penjaminan Mutu Pendidikan*, vol. 3, no. 2, pp. 87-96, 2017.
- [10] Mahfud, Y. R., and Barlian, E. "Hubungan antara Tipe Kepribadian Big Five dengan Prestasi Akademik Mahasiswa," *Jurnal Pendidikan Humaniora*, vol. 5, no. 2, pp. 100-110, 2017.
- [11] Wicaksono, D. A., and Nuryani, N. "Klasifikasi citra wajah dengan metode Naïve Bayes dan k-Nearest Neighbor menggunakan confusion matrix dan split validation." *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* 3, no. 1 (2019): 114-119.
- [12] Pratama, B.Y., "Klasifikasi Kepribadian Berdasarkan Tulisan dari Twitter Menggunakan Metode Naive Bayes, KNN dan SVM," PhD dissertation, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2015.
- [13] Mutawalli, Lalu, Mohammad Taufan Asri Zaen, and Wire Bagye. "Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto)." *Jurnal Informatika dan Rekayasa Elektronik*, vol. 2, no. 2, 2019, pp. 43-51.
- [14] Windana, Fredy, Anitarakhmi Handaratri, and Mohammad Taufan Asri Zaen, "Implementasi

- Metode K-Nearest Neighbor untuk Pengenalan Buah Murbei dengan Korelasi Kandungan Antosianin," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 4, no. 1, pp. 78-86, 2021.
- [15] A. Suciati, S. Sujatmika, and S. Fatichah. "Implementasi Algoritma K-Nearest Neighbor untuk Klasifikasi Citra Tanaman Tebu." *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 2018.
- [16] Nasution, Darnisa Azzahra, Hidayah Husnul Khotimah, and Nurul Chamidah, "Metode Penelitian Kualitatif Studi Pustaka dan Studi Lapangan" in Darmalaksana, Wahyudin (ed.), *Pre-Print Digital Library UIN Sunan Gunung Djati Bandung*, vol. 5, 2020.
- [17] Al Amin, Muchammad. "Klasifikasi Kelompok Umur Manusia Berdasarkan Analisis Dimensi Fraktal Box Counting dari Citra Wajah dengan Deteksi Tepi Canny." *MATHunesa: Jurnal Ilmiah Matematika*, vol. 5, no. 2, 2017.