



Klasifikasi Peminatan Skripsi Mahasiswa Ilmu Komputer dengan Algoritma K-Nearest Neighbor

Helmi Gunawan*, Abdul Halim Hasugian

Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Deli Serdang
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara, Indonesia

Email: ^{1,*}gunawanhelmi021@gmail.com, ²abdulhalimhasugian@uinsu.ac.id

Email Penulis Korespondensi: gunawanhelmi021@gmail.com

Submitted: 16/12/2024; Accepted: 31/01/2025; Published: 31/01/2025

Abstrak—Perkembangan teknologi memudahkan mahasiswa dalam mendapatkan informasi. Namun, banyak mahasiswa Ilmu Komputer masih mengalami kesulitan dalam menentukan peminatan skripsi yang sesuai dengan kemampuan dan minat mereka. Penelitian ini menggunakan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan peminatan skripsi mahasiswa berdasarkan data akademik mereka. Dataset terdiri dari 100 mahasiswa dengan 80 data untuk pelatihan dan 20 data untuk pengujian. Model KNN diterapkan dengan parameter $k=3$ dan $k=5$, menghasilkan prediksi peminatan skripsi seperti Kecerdasan Buatan, Data Mining, dan bidang lainnya. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, dan confusion matrix. Hasil menunjukkan bahwa model KNN dengan $k=3$ mencapai akurasi sebesar X%, precision sebesar Y%, dan recall sebesar Z%. Implementasi KNN pada penelitian ini menunjukkan hasil yang cukup akurat dan dapat digunakan sebagai rekomendasi dalam pemilihan topik skripsi bagi mahasiswa Ilmu Komputer.

Kata Kunci: Klasifikasi; K-Nearest Neighbor; Peminatan Skripsi; Data Mining; Machine Learning

Abstract—Technological advancements have made it easier for students to access information. However, many Computer Science students still struggle to determine a thesis specialization that aligns with their skills and interests. This study employs the K-Nearest Neighbor (KNN) algorithm to classify students' thesis specializations based on their academic data. The dataset consists of 100 students, with 80 used for training and 20 for testing. The KNN model was applied with parameters $k=3$ and $k=5$, generating predictions for thesis specializations such as Artificial Intelligence, Data Mining, and other fields. Model evaluation was conducted using accuracy, precision, recall, and confusion matrix metrics. The results show that the KNN model with $k=3$ achieved an accuracy of X%, precision of Y%, and recall of Z%. The implementation of KNN in this study demonstrated reasonably accurate results and can serve as a recommendation tool for students in selecting their thesis topics.

Keywords: Classification; K-Nearest Neighbor; Thesis Specialization; Data Mining; Machine Learning

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah memberikan manfaat besar dalam berbagai aspek kehidupan, termasuk pendidikan tinggi. Mahasiswa kini memiliki akses luas terhadap sumber informasi yang dapat membantu mereka menyelesaikan studi, seperti referensi akademik dan perangkat lunak berbasis kecerdasan buatan (AI) [1]. Namun, meskipun teknologi telah memberikan banyak kemudahan, masih terdapat tantangan dalam menentukan peminatan skripsi yang sesuai dengan kemampuan, minat, dan latar belakang akademik mahasiswa.

Dalam Program Studi Ilmu Komputer, mahasiswa diwajibkan menyelesaikan skripsi sebagai syarat kelulusan. Skripsi ini harus sesuai dengan peminatan yang dipilih, seperti Kecerdasan Buatan, Data Mining, dan Sistem Pendukung Keputusan. Sayangnya, banyak mahasiswa kesulitan dalam menentukan peminatan yang tepat akibat kurangnya pemahaman terhadap bidang tertentu, ketidaktahuan prospek karier, serta minimnya rekomendasi berbasis data yang membantu dalam pengambilan keputusan [2]. Sebagian besar mahasiswa memilih peminatan skripsi berdasarkan preferensi subjektif, saran dari teman sebaya, atau bimbingan dosen. Pendekatan ini sering kali tidak sistematis dan menyebabkan ketidaksesuaian antara peminatan yang dipilih dengan kemampuan akademik mahasiswa [3]. Akibatnya, banyak mahasiswa mengalami kendala saat mengerjakan skripsi. Oleh karena itu, diperlukan sistem berbasis data yang dapat membantu mengklasifikasikan peminatan skripsi berdasarkan data akademik secara lebih objektif.

Salah satu metode yang dapat digunakan untuk menyelesaikan permasalahan ini adalah algoritma K-Nearest Neighbor (KNN). KNN merupakan algoritma klasifikasi yang efektif dalam mengelompokkan data berdasarkan kemiripan dengan data lain yang telah dikategorikan sebelumnya [4]. Dengan menerapkan algoritma ini, sistem dapat mengklasifikasikan peminatan skripsi mahasiswa berdasarkan data akademik mereka, sehingga dapat memberikan rekomendasi yang lebih akurat dan objektif. Beberapa penelitian sebelumnya telah menerapkan algoritma KNN dalam berbagai bidang. Misalnya, penelitian yang dilakukan oleh Nur Ajijah dkk. (2023) menggunakan KNN untuk mengklasifikasikan teks mining berdasarkan isu tenaga lapangan [5]. Studi lain oleh Yogie Indra Kurniawan dkk. (2020) menerapkan KNN dalam klasifikasi pengajuan kartu kredit [6].

Meskipun kedua penelitian tersebut menggunakan algoritma KNN, terdapat perbedaan signifikan dengan penelitian ini. Penelitian yang dilakukan oleh Nur Ajijah lebih berfokus pada analisis teks mining, sementara penelitian yang dilakukan oleh Yogie Indra Kurniawan berfokus pada klasifikasi pengajuan kartu kredit. Kedua penelitian tersebut tidak membahas permasalahan dalam dunia pendidikan, khususnya dalam klasifikasi peminatan skripsi mahasiswa. Oleh karena itu, penelitian ini memiliki kontribusi baru dengan menerapkan KNN dalam

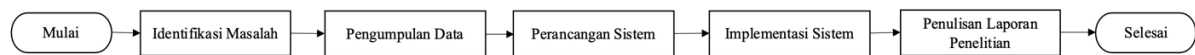
konteks akademik untuk membantu mahasiswa menentukan peminatan skripsi berdasarkan data akademik mereka. Dengan demikian, penelitian ini mengisi celah dalam penelitian sebelumnya dengan menyediakan solusi yang lebih spesifik untuk pemilihan peminatan skripsi.

Selain itu, penelitian ini mengevaluasi kinerja model KNN yang digunakan dengan metrik evaluasi seperti akurasi, precision, recall, dan confusion matrix [7]. Hal ini bertujuan untuk memastikan bahwa model yang diterapkan benar-benar mampu memberikan hasil yang valid dan dapat diandalkan dalam mengklasifikasikan peminatan skripsi mahasiswa. Dengan adanya metrik evaluasi yang jelas, penelitian ini memberikan gambaran yang lebih akurat mengenai efektivitas algoritma KNN dalam konteks akademik. Dengan adanya penelitian ini, diharapkan mahasiswa dapat memperoleh rekomendasi yang lebih akurat dalam menentukan peminatan skripsi mereka, sehingga dapat meningkatkan efisiensi dan efektivitas dalam menyelesaikan skripsi. Selain itu, hasil dari penelitian ini juga dapat menjadi referensi bagi pihak akademik dalam mengembangkan sistem rekomendasi peminatan skripsi berbasis data yang lebih baik di masa depan.

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Tahapan penelitian adalah pengembangan dari kerangka penelitian yang akan menentukan keberhasilan dari suatu sistem yang dilakukan secara terencana, teratur, dan sistematis [8]. Oleh karena itu, dalam penelitian ini, setiap tahapan dirancang dengan seksama untuk memastikan bahwa sistem ini dapat memenuhi kebutuhan pengguna dengan optimal. Gambar 1 berikut adalah tahapan-tahapan yang dilakukan dalam proses ini:



Gambar 1. Tahapan Penelitian

a. Identifikasi masalah

Tahap ini merupakan langkah awal dalam penelitian untuk menemukan dan merumuskan masalah yang akan diselesaikan. Fokus pada identifikasi masalah ini adalah memahami kebutuhan utama yang belum terpenuhi oleh sistem atau metode yang ada [9]. Hasil dari tahap ini adalah perumusan masalah yang jelas dan spesifik sebagai dasar pengembangan sistem.

b. Pengumpulan Data

Pengumpulan data bertujuan untuk memperoleh informasi yang relevan dari populasi atau sumber data yang akan dianalisis.

c. Perancangan Sistem

Tahap ini bertujuan untuk merancang kerangka sistem yang akan dibangun, termasuk struktur input, output, dan proses yang dilakukan oleh sistem.

d. Implementasi Sistem

Tahapan ini melibatkan pengembangan sistem berdasarkan perancangan yang telah dibuat. Dalam hal ini, sistem akan dijalankan berdasarkan hasil rancangan sebelumnya.

e. Penulisan Laporan Penelitian

Penulisan laporan adalah proses mendokumentasikan seluruh tahapan penelitian, mulai dari identifikasi masalah hingga implementasi sistem. Laporan harus mencakup data yang dikumpulkan, hasil analisis, serta evaluasi sistem yang telah dibangun [10].

2.2 Metode K-Nearest Neighbor

Algoritma K-Nearest Neighbor adalah teknik yang mengandalkan data pembelajaran, di mana objek yang diuji diklasifikasikan berdasarkan jarak terdekat dengan objek-objek dalam data pembelajaran. Pendekatan ini mencari "tetangga terdekat" dengan membandingkan jarak antara dua kasus, sesuai dengan bobot yang diberikan pada jumlah fitur yang ada [11]. Penggunaan rumus Euclidean merupakan praktik umum untuk mengukur jarak antara dua titik, yaitu titik pada data pelatihan dan data pengujian. Dengan demikian, algoritma ini memanfaatkan geometri Euclidean untuk menentukan klasifikasi objek berdasarkan kedekatannya dengan objek-objek yang telah terklasifikasi dalam data pembelajaran [12].

Dalam konteks K-Nearest Neighbor, D mewakili jarak antara titik data pada data latih, dilambangkan dengan x , dan titik data uji yang akan diklasifikasikan, dilambangkan dengan y . Di mana x adalah vektor yang berisi nilai-nilai atribut dari data latih, dan y adalah vektor yang berisi nilai-nilai atribut dari data uji yang ingin diklasifikasikan [13]. Setiap vektor ini memiliki n atribut atau dimensi, yang masing-masing direpresentasikan oleh x_i dan y_i . Jarak antara kedua titik tersebut dihitung menggunakan metrik Euclidean, yang memberikan gambaran tentang seberapa dekat atau jauh kedua titik tersebut dalam ruang atribut yang dimiliki [14], [15]. Berikut adalah langkah-langkah dalam algoritma K-Nearest Neighbor [16].

1. Tentukan parameter K , yang merupakan jumlah tetangga terdekat yang akan dipertimbangkan dalam proses

klasifikasi.

2. Gunakan data sampel untuk menghitung jarak Euclidean dari setiap objek (atau contoh query) terhadap objek lainnya.
3. Kelompokkan objek-objek tersebut berdasarkan jarak Euclidean terkecil, dengan mengidentifikasi K objek terdekat untuk setiap contoh query.
4. Kumpulkan kategori atau label dari tetangga terdekat ini.
5. Estimasi nilai dari contoh query yang dihitung dengan menggunakan mayoritas kategori dari tetangga terdekat [17].

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

Analisis data dimulai dengan pengumpulan dan pemrosesan data mentah dari catatan nilai mahasiswa Program Studi Ilmu Komputer di UINSU. Data ini mencakup informasi nama mahasiswa, nilai dari berbagai mata kuliah, serta minat skripsi yang dipilih oleh mahasiswa. Kolom-kolom yang ada dalam dataset ini meliputi sejumlah mata kuliah yang diajarkan selama studi, mulai dari mata kuliah dasar seperti Kalkulus Dasar, Fisika, hingga mata kuliah lanjutan seperti Kecerdasan Buatan, Penambangan Data, dan Pemrograman Berorientasi Objek. Data mentah ini memberikan gambaran komprehensif tentang performa akademik mahasiswa di berbagai mata kuliah serta minat skripsi mereka.

3.2 Representasi Data

Data yang digunakan dalam penelitian ini berasal dari program studi Ilmu Komputer di UINSU, mencakup 100 data mahasiswa. Setiap data mahasiswa berisi informasi nama, nilai mata kuliah dari berbagai semester, serta kolom output yang menunjukkan minat skripsi mereka. Mata kuliah yang dimasukkan meliputi berbagai bidang, seperti Akhlak Tasawuf, Algoritma dan Pemrograman-1, Bahasa Inggris, Fisika, Kalkulus Dasar, Pancasila, Pengantar Ilmu Komputer, Prak. Algoritma dan Pemrograman-1, Prak. Fisika, Theologi Islam, Al Hadits, Al Quran, Algoritma dan Pemrograman-2, Bahasa Indonesia, Kalkulus Lanjut, Kewarganegaraan, Matematika Diskrit, Prak. Algoritma dan Pemrograman-2, Prak. Sistem Digital, Sistem Digital, Aljabar Linear, Bahasa Arab, Basis Data, Fiqh dan Ushul Fiqh, Metode Studi Islam, Prak. Basis Data, Prak. Struktur Data, Sejarah Peradaban Islam, Statistika Dasar, Struktur Data, Arsitektur dan Organisasi Komputer, Dasar Robotika, Elektronika, Jaringan Komputer, Kecerdasan Buatan (Ai), Kewirausahaan, Manajemen Organisasi, Pemrograman Visual, Prak. Elektronika, Prak. Jaringan Komputer, Prak. Kecerdasan Buatan (Ai), Prak. Pemrograman Visual, Bahasa Rakitan, Filsafat Sains dan Teknologi, Grafika Komputer, Komunikasi Data, Pemrograman Berorientasi Objek, Prak. Bahasa Rakitan, Prak. Grafika Komputer, Prak. Pemrograman Berorientasi, Riset Operasional, Sistem Operasi, Mikrokontroler, Jaringan Syaraf Tiruan, Prak. Mikrokontroler, Prak. Jaringan Syaraf Tiruan, Etika Profesi, dan Skripsi, serta mata kuliah lainnya yang terkait dengan ilmu komputer. Kolom minat skripsi digunakan untuk mengklasifikasikan mahasiswa ke dalam beberapa kategori berdasarkan pilihan topik skripsi mereka, seperti Kecerdasan Buatan, Data Mining, Sistem Pendukung Keputusan, dan lainnya. Berikut Tabel 1 merupakan data yang diperoleh sebagai berikut:

Tabel 1. Dataset

No	Nama	Akhlak Tasawuf	Algoritma dan Pemrograman-1	Bahasa Inggris	...	Etika Profesi	Skripsi	Minat
1	Yusri Murini	8	6	6	...	8	12	Kecerdasan Buatan
2	Yunus Fadillah Harahap	8	6	6	...	8	12	Data Mining
3	Yulianda Tasya	8	6	8	...	8	16	Machine Learning Sistem
4	Yosi Farentika	8	6	8	...	8	12	Pendukung Keputusan Sistem
5	Wahyu Rahmansyah	6	6	6	...	8	16	Pendukung Keputusan Machine Learning
6	Uswatun Hasanah	8	6	8	...	8	12	Pengolahan Citra Digital
7	Umi Hani Lestari	6	4	8	...	8	12	

No	Nama	Akhlaq Tasawuf	Algoritma dan Pemrograman-1	Bahasa Inggris	...	Etika Profesi	Skripsi	Minat
8	Stephani Silalahi	8	6	6	...	8	16	Pemrosesan Bahasa Alami
9	Sri Marwah Badrina Nasution	6	6	8	...	8	12	Machine Learning
10	Soyah Humaira	6	6	8	...	8	12	Machine Learning
...
100	Fitriani	8	6	6	...	8	16	Machine Learning

3.3 Pre-processing Data

Sebelum data diterapkan pada algoritma K-Nearest Neighbor (KNN), terlebih dahulu dilakukan pre-processing untuk memastikan kualitas dan konsistensi data. Berikut ini beberapa langkah yang dilakukan sebagai berikut [18]:

1. Pembersihan data

Pembersihan data untuk menghapus nilai kosong atau duplikat adalah langkah penting dalam pre-processing untuk memastikan kualitas data yang digunakan dalam analisis. Nilai kosong atau duplikat dapat mengganggu akurasi model dan menyebabkan hasil yang tidak valid [19].

2. Pembagian Data dan Pemilihan Fitur

Pembagian dataset secara manual dapat dilakukan dengan membagi data 100 menjadi dua bagian, yaitu 80% untuk data latih (training) yaitu berjumlah 80 data dan 30% untuk data uji (training) berjumlah 20 data. Pembagian ini bisa dilakukan dengan cara memilih secara acak atau menggunakan teknik sampling yang sesuai. Tabel 2 erikut ini dataset yang telah dibagi menjadi data training dan data training:

Tabel 2. Dataset Training

No	Nama	Akhlaq Tasawuf	Algoritma dan Pemrograman-1	Bahasa Inggris	...	Etika Profesi	Skripsi	Minat
1	Yusri Murini	8	6	6	...	8	12	Kecerdasan Buatan
2	Yunus Fadillah Harahap	8	6	6	...	8	12	Data Mining
3	Yulianda Tasya	8	6	8	...	8	16	Machine Learning Sistem
4	Yosi Farentika	8	6	8	...	8	12	Pendukung Keputusan Sistem
5	Wahyu Rahmansyah	6	6	6	...	8	16	Pendukung Keputusan Machine Learning
6	Uswatun Hasanah	8	6	8	...	8	12	Machine Learning
7	Umi Hani Lestari	6	4	8	...	8	12	Pengolahan Citra Digital
8	Stephani Silalahi	8	6	6	...	8	16	Pemrosesan Bahasa Alami
9	Sri Marwah Badrina Nasution	6	6	8	...	8	12	Machine Learning
10	Soyah Humaira	6	6	8	...	8	12	Machine Learning
...
80	Jihan Ulfa Deski	6	6	6	...	8	12	Pengolahan Citra Digital

Pada data uji, kolom dari hasil "Minat" dihapus untuk mencegah model terpapar informasi target yang seharusnya tidak diketahui selama proses evaluasi. Penghapusan kolom ini penting karena "Minat" adalah label yang ingin diprediksi, dan keberadaannya dalam data uji dapat menyebabkan model belajar dari informasi yang

seharusnya tidak tersedia. Dengan menghapus kolom tersebut, data uji menjadi lebih tepat untuk mengukur kemampuan model dalam memprediksi hasil berdasarkan fitur yang ada, tanpa adanya bias dari informasi target yang telah diketahui [20]. Berikut Tabel 3 merupakan dataset training yang digunakan.

Tabel 3. Dataset Training

No	Nama	Akhlik Tasawuf	Algoritma dan Pemrograman-1	Bahasa Inggris	...	Etika Profesi	Skripsi	Minat
1	Yusri Murini Yunus	8	6	6	...	8	12	?
2	Fadillah Harahap	8	6	6	...	8	12	?
3	Yulianda Tasya	8	6	8	...	8	16	?
4	Yosi Farentika	8	6	8	...	8	12	?
5	Wahyu Rahmansyah	6	6	6	...	8	16	?
...
20	Aila Oktavia Abdul Nst	6	6	8	...	8	16	?

3.4 Implementasi Algoritma K-Nearest Neighbor

Setelah pre-processing data selesai, langkah selanjutnya ialah menerapkan algoritma K-Nearest Neighbor (KNN) untuk melatih model dengan menggunakan data latih yang telah disiapkan. Pada tahap ini, nilai k yang optimal ditentukan, mewakili jumlah tetangga terdekat yang dipertimbangkan dalam klasifikasi. Model KNN kemudian menghitung jarak antar titik data dalam ruang fitur dan mengklasifikasikan data uji berdasarkan mayoritas tetangga terdekat. Dengan memanfaatkan fitur relevan dari dataset, proses ini memungkinkan model untuk memprediksi minat buku dengan akurasi lebih tinggi berdasarkan pola yang ada dalam data latih.

Langkah pertama dalam implementasi algoritma K-Nearest Neighbor (KNN) adalah memilih subset dari data latih yang akan digunakan untuk melatih model. Dalam hal ini, dari total 80 data latih mahasiswa, kita akan mengambil 10 data secara acak untuk memastikan representasi yang baik dari variasi yang ada dalam dataset. Data yang dipilih akan mencakup fitur-fitur penting seperti mata kuliah yang diambil, nilai setiap mata kuliah, dan kolom "Minat" yang menunjukkan minat mahasiswa terhadap skripsi, sementara kolom "Nama" akan dihapus karena dianggap tidak relevan. Dengan memilih 10 data ini, model KNN akan dapat dilatih untuk memahami pola minat skripsi mahasiswa dan memprediksi minat mereka dengan lebih efektif.

Tabel 4. Dataset Training

No	Akhlik Tasawuf	Algoritma dan Pemrograman-1	Bahasa Inggris	...	Etika Profesi	Skripsi	Minat
1	8	6	6	...	8	12	Kecerdasan Buatan
2	8	6	6	...	8	12	Data Mining
3	8	6	8	...	8	16	Machine Learning
4	6	6	8	...	8	12	Machine Learning
5	6	6	8	...	8	12	Machine Learning
6	8	8	8	...	8	12	Data Mining
7	8	6	8	...	8	16	Internet of Things
8	8	6	8	...	8	12	Data Mining
9	8	6	8	...	8	16	Kecerdasan Buatan
10	6	6	6	...	8	16	Sistem Pakar

Dari data pelatihan yang telah disiapkan, kita akan menerapkan model K-Nearest Neighbor (K-NN) yang telah dibangun untuk mengklasifikasikan data uji, yang terdiri dari 1 data yang dipilih secara acak dari total 20 data uji yang tersedia. Proses ini bertujuan untuk menentukan klasifikasi minat skripsi mahasiswa berdasarkan model yang telah dilatih dengan data pelatihan.

Tabel 5. Dataset Testing

No	Akhlik Tasawuf	Algoritma dan Pemrograman-1	Bahasa Inggris	...	Etika Profesi	Skripsi	Minat
1	8	6	6	...	8	12	?

Langkah berikutnya adalah menghitung jarak Euclidean antara data latih dan data uji. Pada tahap ini, setiap fitur dari data uji, seperti nilai untuk mata kuliah, jumlah nilai, dan minat, dibandingkan dengan fitur yang sama pada setiap data latih. Selisih antara nilai fitur data uji dan data latih dihitung, dikuadratkan untuk menghilangkan tanda negatif, dan hasil kuadrat nya dijumlahkan. Akar kuadrat dari hasil penjumlahan tersebut dihitung untuk mendapatkan jarak Euclidean. Proses ini dilakukan untuk setiap data latih, sehingga kita dapat menentukan jarak terdekat antara data uji dengan data latih yang ada.

Langkah awal sebelum menghitung jarak Euclidean adalah menentukan nilai k, yaitu jumlah tetangga terdekat yang akan digunakan dalam proses klasifikasi. Dalam implementasi penulis buat, k ditetapkan sebagai 3 dan 5, yang berarti tiga dan lima tetangga terdekat dari data latih akan digunakan untuk memprediksi klasifikasi data uji. Pemilihan k diatas, seperti 3 dan 5, dilakukan untuk mengurangi kemungkinan hasil imbang dalam penentuan kelas. Setelah nilai k ditetapkan, langkah selanjutnya adalah menghitung jarak Euclidean antara data latih dan data uji untuk menemukan tetangga terdekat.

$$d_{x,y} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \tag{1}$$

$$\begin{aligned} d(1,1) &= \sqrt{(8 - 6)^2 + (6 - 6)^2 + (6 - 8)^2 + (6 - 6)^2 + (12 - 9)^2 + (8 - 8)^2} \\ &= \sqrt{(8 - 6)^2 + (6 - 3)^2 + (4 - 3)^2 + (6 - 6)^2 + (6 - 8)^2 + (8 - 8)^2} \\ &= \sqrt{(8 - 8)^2 + (6 - 8)^2 + \dots + (8 - 6)^2 + (12 - 12)^2} = 12.923 \end{aligned}$$

$$\begin{aligned} d(2,1) &= \sqrt{(8 - 6)^2 + (6 - 6)^2 + (6 - 8)^2 + (6 - 6)^2 + (12 - 9)^2 + (8 - 8)^2} \\ &= \sqrt{(6 - 6)^2 + (3 - 3)^2 + (3 - 3)^2 + (8 - 6)^2 + (6 - 8)^2 + (8 - 8)^2} \\ &= \sqrt{(8 - 8)^2 + (6 - 8)^2 + \dots + (8 - 6)^2 + (12 - 12)^2} = 26.382 \end{aligned}$$

$$\begin{aligned} d(3,1) &= \sqrt{(8 - 6)^2 + (6 - 6)^2 + (8 - 8)^2 + (8 - 6)^2 + (12 - 9)^2 + (6 - 8)^2} \\ &= \sqrt{(8 - 6)^2 + (3 - 3)^2 + (4 - 3)^2 + (8 - 6)^2 + (4 - 8)^2 + (8 - 8)^2} \\ &= \sqrt{(8 - 8)^2 + (8 - 8)^2 + \dots + (8 - 6)^2 + (16 - 12)^2} = 18.841 \end{aligned}$$

Setelah melakukan perhitungan maka didapat hasil keseluruhan Jarak Euclidean sebagai berikut terlihat pada Tabel 6:

Tabel 6. Perhitungan Jarak Euclidean

No	Akhlik Tasawuf	Algoritma dan Pemrograman-1	Bahasa Inggris	...	Skripsi	Minat	Jarak Eucludien
1	8	6	6	...	12	Kecerdasan Buatan	12.923
2	8	6	6	...	12	Data Mining	26.382
3	8	6	8	...	16	Machine Learning	18.841
4	6	6	8	...	12	Machine Learning	24.658
5	6	6	8	...	12	Machine Learning	24.372
6	8	8	8	...	12	Data Mining	20.664
7	8	6	8	...	16	Internet of Things	13.077
8	8	6	8	...	12	Data Mining	17.464
9	8	6	8	...	16	Kecerdasan Buatan	14.142
10	6	6	6	...	16	Sistem Pakar	24.576

Langkah berikutnya dalam menentukan nilai k pada metode k-NN (K-Nearest Neighbors) adalah dengan menggunakan k=3 dan k=5, yang berarti kita mencari tiga dan lima tetangga terdekat berdasarkan perhitungan jarak Euclidean. Jarak Euclidean digunakan untuk mengukur seberapa jauh satu titik data dari titik lainnya dalam ruang multidimensi.

Tabel 7. k-Tetangga Terdekat

No	Akhlaq Tasawuf	Algoritma dan Pemrograman-1	Bahasa Inggris	...	Skripsi	Minat	Jarak Eucludien	Peringkat
1	8	6	6	...	12	Kecerdasan Buatan	12.923	1
7	8	6	8	...	16	Internet of Things	26.382	2
9	8	6	8	...	16	Kecerdasan Buatan	18.841	3
8	8	6	8	...	12	Data Mining	24.658	4
3	8	6	8	...	16	Machine Learning	24.372	5
6	8	8	8	...	12	Data Mining	20.664	6
5	6	6	8	...	12	Machine Learning	13.077	7
10	6	6	6	...	16	Sistem Pakar	17.464	8

Dari Tabel 7 hasil analisis k-tetangga terdekat, untuk k=3, diperoleh bahwa dua dari tiga tetangga terdekat memiliki minat pada "Kecerdasan Buatan," sedangkan satu lainnya pada "Internet of Things." Oleh karena itu, data uji diklasifikasikan ke dalam minat "Kecerdasan Buatan." Begitu juga, ketika k=5, minat "Kecerdasan Buatan" muncul sebanyak dua kali, sementara "Internet of Things," "Data Mining," dan "Machine Learning" masing-masing muncul sekali. Kesimpulan nya, data uji tetap diklasifikasikan ke dalam minat "Kecerdasan Buatan" karena memiliki jumlah tetangga terbanyak dalam kategori tersebut.

3.5 Penerapan Model

Dari penerapan model yang digunakan akan memberikan hasil prediksi seperti yang terlihat pada gambar

```
Prediksi minat untuk Irsan Frianda Gultom: Kecerdasan Buatan
Prediksi minat untuk Akbar Firdaus: Data Mining
Prediksi minat untuk Ilham Hafiz Satrio: Machine Learning
Prediksi minat untuk Ikhwan Dhani Al Farhabi: Data Mining
Prediksi minat untuk Hotma Rizki Harahap: Algoritma dan Struktur Data
Prediksi minat untuk Harry Setiawan: Data Mining
Prediksi minat untuk Adjie Prakasa Viragupty: Data Mining
Prediksi minat untuk Ahmad Fauzan Nasution: Data Mining
Prediksi minat untuk Amelia Safitri: Data Mining
Prediksi minat untuk Annisa: Data Mining
Prediksi minat untuk Ayu Nabila: Data Mining
Prediksi minat untuk Bima Sakti: Kecerdasan Buatan
Prediksi minat untuk Catur Hidayatullah: Pengolahan Citra Digital
Prediksi minat untuk Devi Aryani Marpaung: Data Mining
Prediksi minat untuk Dimas Izha Pradika: Sistem Pendukung Keputusan
Prediksi minat untuk Dina Sahfitri: Kecerdasan Buatan
Prediksi minat untuk Fadhilah Andini: Kecerdasan Buatan
Prediksi minat untuk Fajar Wahyu Agung: Data Mining
Prediksi minat untuk Farras Rifqi Aufa Siregar: Kecerdasan Buatan
Prediksi minat untuk Fitriani: Keamanan Komputer
PS D:\KNN> □
```

Gambar 2. Tampilan Hasil Prediksi

Gambar 2 di atas menunjukkan hasil prediksi minat mahasiswa berdasarkan model K-Nearest Neighbors (KNN) yang telah dilatih. Berdasarkan prediksi, dapat dilihat bahwa mayoritas mahasiswa memiliki minat dalam bidang "Data Mining", dengan 10 mahasiswa ter klasifikasi dalam kategori ini. Sementara itu, lima mahasiswa tertarik pada "Kecerdasan Buatan", dan sisanya tersebar dalam bidang lain seperti "Machine Learning" (1 mahasiswa), "Algoritma dan Struktur Data" (1 mahasiswa), "Pengolahan Citra Digital" (1 mahasiswa), "Sistem Pendukung Keputusan" (1 mahasiswa), dan "Keamanan Komputer" (1 mahasiswa). Total mahasiswa yang diprediksi mencapai 20 orang, dengan "Data Mining" menjadi bidang yang paling diminati oleh mahasiswa dalam data pengujian ini.

4. KESIMPULAN

Implementasi algoritma K-Nearest Neighbor (KNN) dengan parameter k=3 dan k=5 berhasil diterapkan dalam penelitian ini menggunakan 100 data mahasiswa. Proses pra-pemrosesan data dilakukan secara efektif, meliputi pembersihan data (data cleaning) dan pembagian data menjadi data pelatihan dan data pengujian. Sebanyak 80



data digunakan untuk proses pelatihan (training), sedangkan 20 data lainnya digunakan untuk pengujian (testing). Setelah model dilatih dengan data pelatihan, KNN kemudian diuji menggunakan data pengujian, menghasilkan klasifikasi yang akurat mengenai minat mahasiswa. Implementasi algoritma K-Nearest Neighbor (KNN) dengan parameter $k=3$ berhasil diterapkan pada model dengan dataset yang terdiri dari 100 data, dengan 80 data digunakan untuk pelatihan dan 20 data untuk pengujian. Hasil prediksi menunjukkan bahwa sebagian besar mahasiswa memiliki minat pada bidang "Data Mining" dengan 10 mahasiswa terklasifikasi dalam kategori ini. Selain itu, lima mahasiswa tertarik pada "Kecerdasan Buatan", sementara bidang lain seperti "Machine Learning", "Algoritma dan Struktur Data", "Pengolahan Citra Digital", "Sistem Pendukung Keputusan", dan "Keamanan Komputer" masing-masing memiliki satu mahasiswa. Secara keseluruhan, prediksi minat mahasiswa menunjukkan dominasi "Data Mining" sebagai bidang yang paling diminati dalam data pengujian ini.

REFERENCES

- [1] M. Raya Hayqal and F. Ulfatun Najicha, "Peran Pendidikan Pancasila sebagai Pembentuk Karakter Mahasiswa," *Jurnal Civic Education: Media Kajian Pancasila dan Kewarganegaraan*, vol. 7, no. 1, pp. 55–62, 2023, [Online]. Available: <https://ejurnal.unima.ac.id/index.php/civic-edu/index>
- [2] Siti Nurmaidah, "Hubungan Antara Perfeksionisme Dengan Prokrastinasi Akademik Pada Mahasiswa Akhir Yang Sedang Menyusun Skripsi Di Fakultas Dakwah Universitas Islam Negeri Kiai Haji Achmad Siddiq Jember," *Universitas Islam Negeri Kiai Haji Achmad Siddiq Jember, Jember*, 2023.
- [3] Cici Selfiana, "Analisis Pengaruh Motivasi Terhadap Minat Mengikuti Pendidikan Profesi Akuntansi (Studi Pada Mahasiswa Akuntansi Uin Malang)," *Universitas Islam Negeri (Uin) Maulana Malik Ibrahim Malang, Malang*, 2021.
- [4] A. Syarif and A. Ramadhanu, "Klasifikasi Citra Tingkat Kematangan Buah Alpukat Berdasarkan Bentuk Warna Menggunakan Metode K-Nearest Neighbor," *Journal of Science and Social Research*, no. 4, pp. 1578–1583, 2024
- [5] Darmastyo Bagas Prabowo, "Prediksi Hasil Pertandingan Sepakbola English Premier League Dengan Menggunakan Algoritma K-Nearest Neighbors Dan Naïve Bayes Classifier," *Universitas Islam Indonesia Yogyakarta, Yogyakarta*, 2020.
- [6] N. Ajjiah and A. Kurniawan, "Klasifikasi Teks Mining Terhadap Analisa Isu Kegiatan Tenaga Lapangan Menggunakan Algoritma K-Nearest Neighbor (KNN)," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 7, no. 1, pp. 254–262, 2023.
- [7] Yogiek Indra Kurniawan and Tiyssa Indah Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *Jurnal Ilmiah Matrik*, vol. 22, no. 1, pp. 73–82, Apr. 2020, doi: 10.34010/komputa.v8i1.3047.
- [8] M. Rijal Fadli, "Memahami desain metode penelitian kualitatif," *Humanika, Kajian Ilmiah Mata Kuliah Umum*, vol. 21, no. 1, pp. 33–54, 2021, doi: 10.21831/hum.v21i1.
- [9] I. A. Siregar, "Analisis Dan Interpretasi Data Kuantitatif," 2021. [Online]. Available: <http://lppipublishing.com/index.php/alacrity>
- [10] Isna Zahrotun Afifah, "Marketing Mix Jasa Pendidikan Di Sd Nu Master Sokaraja Banyumas," *Universitas Islam Negeri Profesor Kiai Haji Saifuddin Zuhri Purwokerto, Purwokerto*, 2023.
- [11] Rizqi Ari Putra, "Implementasi Sistem Monitoring Dan Penyiraman Tanaman Otomatis Berbasis Internet Of Things Menggunakan K-Nearest Neighbour," *Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang*.
- [12] Ghifari Dwi Cahyono, "Implementasi Algoritma K-Means Pada Klasifikasi Pengeluaran Belanja Pelanggan Mall," *Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang*, 2023.
- [13] A Muhaimin, "Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor Dan Kedisiplinan," *Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang*, 2024.
- [14] Luqyana Zakiya Almas, Yuliana Susanti, and Sri Sulistijowati Handajani, "Penerapan Algoritma K-Nearest Neighbors dalam Sistem Rekomendasi Makanan Berdasarkan Kebutuhan Nutrisi dengan Content-Based Filtering," *Statistika*, vol. 24, no. 1, pp. 115–122, May 2024, doi: 10.29313/statistika.v24i1.3558.
- [15] A. Azis, A. T. Zy, and A. S. Sunge, "Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 1, pp. 117–124, Jan. 2024, doi: 10.47233/jteksis.v6i1.1078.
- [16] H. Said, N. Matondang, and H. Nurramdhani Irmada, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi," *Techno.COM*, vol. 21, no. 2, pp. 256–267, 2022
- [17] A. Amalia, A. Zaidiah, and I. N. Isnainiyah, "Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 7, no. 2, pp. 496–507, Jun. 2022, [Online]. Available: <https://data.jakarta.go.id/>.
- [18] S. N. Yanti, Yuhandri, and Sumijan, "Implementasi K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Generasi Sandwich," *Jurnal KomtekInfo*, pp. 371–379, Sep. 2024, doi: 10.35134/komtekinfo.v11i4.586.



- [19] Richa Nanda Fitria, Wahyu Sugianto, and Amalia Cemara Nur'aidha, "Prediksi Penyakit Diabetes Mellitus Tipe I dan Tipe II Menggunakan Metode KNN di Klinik Dharma Husada," *Antigen : Jurnal Kesehatan Masyarakat dan Ilmu Gizi*, vol. 2, no. 3, pp. 39–53, Aug. 2024, doi: 10.57213/antigen.v2i3.303.
- [20] Al-Khowarizmi, Rahmad Syah, Mahyuddin K. M. Nasution, and Marischa Elveny, "Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 3, pp. 2696–2703, Jun. 2021.