

Implementasi Metode *K-Nearest Neighbor* untuk Rekomendasi Keminatan Studi (Studi Kasus: Jurusan Teknik Informatika Universitas Brawijaya)

Luthfi Anshori¹, Rekyan Regasari Mardi Putri², Tibyani³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹luthfianshori.tif@gmail.com, ²rekyan.rmp@ub.ac.id, ³tibyani@ub.ac.id

Abstrak

Prodi Informatika Universitas Brawijaya mempunyai 5(lima) keminatan yang akan ditempuh mahasiswa dalam menjalani perkuliahan. Mengingat sangat pentingnya keminatan ini untuk perkuliahan mahasiswa yang nantinya juga berpengaruh pada tugas akhir (skripsi) hingga kelulusan mahasiswa, maka mahasiswa diharapkan untuk memilih keminatan yang sesuai dengan minat dan bakatnya. Berdasarkan Buku Pedoman Fakultas Ilmu Komputer Universitas Brawijaya tahun 2016 keminatan yang ada pada prodi Informatika ada 5 (lima) keminatan yaitu Rekayasa Perangkat Lunak, Komputasi Cerdas, Perangkat bergerak, Jaringan Komputer serta *Game*. Agar mahasiswa dapat memilih keminatan yang sesuai maka di perlukan sistem rekomendasi untuk keminatan mahasiswa dengan harapan mahasiswa dapat memilih keminatan yang sesuai dengan bakat, keinginan serta tentu saja nilai mata kuliah wajib yang sudah ditempuh. Dengan mengimplementasikan metode *K-Nearest Neighbor* yang bertujuan untuk memberikan rekomendasi keminatan berdasarkan atribut nilai-nilai mata kuliah wajib. Pengujian dilakukan dengan membandingkan hasil keminatan yang sudah diverifikasi oleh akademik kampus dengan hasil dari sistem. Dari pengujian data uji sebanyak 30 data mahasiswa yang sudah memilih keminatan maka didapatkan hasil akurasi sebanyak 76,66% terhadap data latih dengan nilai *K* yang paling optimal adalah *K* bernilai 10, maka dapat disimpulkan sistem rekomendasi keminatan ini dapat membantu mahasiswa memberikan rekomendasi keminatan yang akan dipilih.

Kata kunci: sistem Rekomendasi, keminatan, *k-nearest neighbor*, *KNN*

Abstract

*The Informatics Engineering Programme in Brawijaya University has 5 (five) interests that should be taken by each students in their course period. Considering the importance of this interests to students' academic process which later will affect their final project until their graduation, therefore students are expected to choose their insterests based on their attentiveness and ability. Based on the Faculty of Computer Science Handbook Universitas Brawijaya 2016, there are 5 (five) interests existing in Informatics Programme, namely Software Engineering, Smart Computing, Mobile devices, Computer Networks and Games. In order for students to choose the right interest then they need a recommendation system for students so that students can choose interest that suites their talent, desire and scores of courses that have been taken. By implementing *K-Nearest Neighbor* method which aims to give recommendation of interests based on attribute of compulsory courses scores. Testing is done by comparing the results of the interests that has been verified by campus' academic with the results of the system. From the test of data train as much as 30 data of students which have chosen their interests hence gotten the result of accuracy as much 76.66% to trained data with the most optimal *K* value, the *K* value 10, therefore can be concluded this recommendation of interest system can help student give recommendation of interest which will be chosen.*

Keyword : recommendation system, interests, *k-nearest neighbor*, *KNN*.

1. PENDAHULUAN

Keminatan pada Fakultas Ilmu Komputer (FILKOM) prodi Informatika dan Ilmu

Komputer Universitas Brawijaya merupakan pilihan yang harus di pilih oleh mahasiswa berdasarkan keahlian khusus dan potensi yang nantinya akan berpengaruh pada tugas akhir (skripsi) yang akan diambil sehingga sangat

berpengaruh pada kelulusan mahasiswa.

Permasalahan yang terjadi adalah masih dapat terjadinya perubahan pemikiran mahasiswa terhadap keminatan skripsinya yang telah dipilihnya karena beberapa disebabkan oleh kesalahan dalam pengambilan mata kuliah pilihan pada semester-semester sebelumnya, penyebab lain adalah dikarenakan oleh nilai akademik di mata kuliah dasar dan mata kuliah pilihan serta keinginan yang terkadang berbeda. Sebagai salah satu contoh ialah keminatan Rekayasa Perangkat Lunak yang harus ditunjang oleh nilai mata kuliah wajib diantaranya seperti Pemrograman Dasar, Pemrograman Lanjut, Rekayasa Perangkat Lunak sendiri serta mata kuliah pilihan yang mendukung adalah Manajemen Proyek Perangkat Lunak, Pemodelan Perangkat Lunak serta Pengujian Perangkat Lunak sedangkan mata kuliah Pemrograman dasar dan Pemrograman Lanjut juga bisa digunakan untuk menunjang keminatan lainnya seperti keminatan perangkat bergerak (*mobile*) dan keminatan game. Sehingga bisa saja keminatan tersebut dipengaruhi nilai mata kuliah wajib di keminatan yang lainnya. Beberapa faktor eksternal seperti teman dan dosen juga dapat mempengaruhi pertimbangan penentuan keminatan yang akan dipilih. Beberapa faktor yang mempengaruhi keminatan tersebut secara langsung mengakibatkan pengerjaan skripsi menjadi terhambat dan berkemungkinan untuk memulai dari awal lagi. Bahan perbandingan atau panduan rekomendasi dalam pemilihan kriteria keminatan berdasarkan mata kuliah yang telah ditempuh masih belum tersedia.

Metode Klasifikasi dengan Algoritma K-Nearest Neighbor (k-NN) adalah suatu metode yang menggunakan algoritma supervised, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN yang bertujuan untuk mengklasifikasi objek baru berdasarkan atribut dan *training sample* (L.D & Larose, 2005). Penulis menerapkan metode *K-Nearest Neighbor* untuk membantu menyelesaikan permasalahan rekomendasi keminatan tersebut karena metode ini adalah sebuah metode yang efektif apabila mempunyai data latih yang besar. Algoritma ini juga merupakan salah satu teknik *lazy learning*. KNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data *testing* (Wu, 2009).

Oleh karena itu, penelitian ini akan

menganalisis penerapan teknik klasifikasi K-Nearest Neighbor untuk mengelompokkan data nilai akademik mahasiswa dengan rekomendasi keminatan.

2. LANDASAN KEPUSTAKAAN

Penulis memulai penelitian ini dengan terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Dari studi kepustakaan itu, penulis menemukan beberapa penelitian yang mendorong untuk mengangkat tema seperti diatas. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian penulis, antara lain adalah penelitian mengenai algoritma yang digunakan penulis yang akan diangkat oleh penulis.

Tabel 1. Penelitian Terkait

No.	Judul	Metode (proses)	Hasil
1.	Implementasi Algoritma <i>K-Nearest Neighbor</i> (KNN) Untuk Menentukan Tingkat Resiko Penyakit Lemak Darah (Profil Lipid) (Novia Ayu Kumalasari, 2014)	<i>K-Nearest Neighbor</i> (KNN)	<ul style="list-style-type: none"> • 140 data menghasilkan nilai akurasi sebesar 85,55% • 73,55% untuk 60 data latih.
2.	Implementasi Algoritma <i>K-Nearest Neighbor</i> (KNN) Yang Berdasarkan One Pass Clustering Untuk Katagori Teks (Andreas Daniel Arifin, 2014)	<i>K-Nearest Neighbor</i> (KNN) yang Berdasarkan <i>One Pass Clustering</i>	Menghasilkan Performa one pass KNN lebih baik daripada performa KNN konvensional di mana performa one pass KNN bernilai 88% pada pada epsilon sebesar 4 dan nilai k sebesar 30 dibanding dengan performa KNN yang bernilai hanya 85% saat nilai k = 50.

2.1 Klasifikasi

Klasifikasi adalah sebuah metode untuk mengelompokkan data secara sistem menurut aturan dan kaidah yang telah ditetapkan. Klasifikasi juga dapat diartikan pengelompokkan data atau objek baru berdasarkan variabel yang diamati dengan tujuan untuk memprediksi suatu objek dari yang masih belum diketahui kelas atau kategorinya (Tri Halomoan Simanjuntak,2014).

Klasifikasi merupakan teknik data mining yang melihat atribut dari kelompok data yang sudah didefinisikan sebelumnya. Sehingga dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang sudah diklasifikasi dan menggunakan hasilnya untuk memberikan beberapa aturan. Aturan-aturan tersebut dipakai pada data-data baru untuk selanjutnya diklasifikasi (Siti Hardiyanti, 2014).

Tujuan dari klasifikasi yaitu supaya record-record yang tidak diketahui pada kategori mana sebelumnya dapat dikelompokkan kelasnya secara akurat (Sofa Zainuddin,2014).

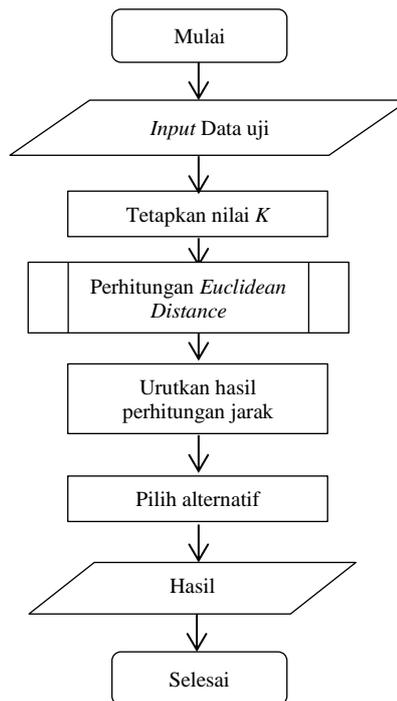
2.2 K-Nearest Neighbor

Salah satu metode klasifikasi yang sering digunakan adalah metode *K-Nearest Neighbor*. Penggunaan *K-Nearest Neighbor* bertujuan untuk mengklasifikasikan objek baru berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek baru tersebut. Teknik algoritma *K-Nearest Neighbor* mudah untuk diimplementasikan. Dalam hal ini jumlah data atau biasa disebut dengan tetangga terdekat ditentukan oleh user yang dinyatakan oleh *k*.

Adapun langkah-langkah dalam metode *K-Nearest Neighbor* adalah sebagai berikut:

1. Menghitung jarak *Euclidean*
Rumus jarak *euclidean* :
$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

Di mana:
 $d(x_i, x_j)$ = jarak *Euclidean*
 x_i = *record* ke-*i*
 x_j = *record* ke-*j*
 a_r = data ke-*r*
2. Mengurutkan berdasarkan nilai *Euclidean distance*
3. Menentukan *k record* klasifikasi terdekat
4. Target *output* merupakan kelas yang mayoritas



Gambar 1. Flowchart Metode *K-Nearest Neighbor*

2.3 Euclidean Distance

Metode *K-Nearest Neighbor* dilakukan perhitungan kuadrat jarak euclid (*query instance*) masing-masing obyek terhadap data sampel yang diberikan. Dalam perhitungan jarak ada beberapa rumus jarak contohnya seperti *Euclidian Distance*, *Square euclidean distance*, dan *Manhattan distance*. Karena secara umum untuk mendefinisikan jarak antara dua objek *x* dan *y*, maka digunakan rumus

$$X1 = [X1, X2, \dots, Xp] \text{ dan } Y1 = [Y1, Y2, \dots, Yp]$$

$$D2(x,y) = (X1 - Y1)^2 + (X2 - Y2)^2 + \dots + (Xp - Yp)^2$$

$$= (X - Y) (X - Y) = (Y - X) (Y - X) \quad (2)$$

dimana matriks *D(a,b)* adalah jarak skalar dari kedua vektor *a* dan *b* dari matriks dengan ukuran *d* dimensi

3. METODOLOGI

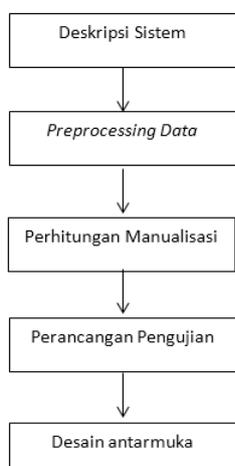
Bab ini menjelaskan metode dan langkah-langkah kajian ilmiah yang akan dilakukan dalam penyusunan skripsi. Pada bab ini akan dijelaskan tahapan susunan kegiatan penelitian untuk mengimplementasikan metode *K-Nearest Neighbor* (*KNN*) secara terstruktur serta perancangan sistem yang baik

Perancangan sistem dalam penelitian ini menggunakan metode *K-Nearest Neighbor* (*KNN*) sebagai metode dalam proses

penghitungan, dan perbandingan nilai terbesar untuk kesimpulan akhir.

4. PERANCANGAN

Pada bagian ini menjelaskan mengenai tentang analisis kebutuhan dan perancangan untuk membangun sebuah sistem dalam memberikan informasi kepada pengguna mengenai merekomendasikan keminatan laboratorium. Klasifikasi pada sistem ini menggunakan metode *K-Nearest Neighbor*. Pada gambar 4.1 menunjukkan gambaran proses yang dilakukan pada bab perancangan ini



Gambar 2. Pohon perancangan

4.1 Deskripsi Sistem

Secara garis besar sistem yang akan dikembangkan merupakan sebuah sistem yang mengimplementasikan metode *K-Nearest Neighbor* untuk merekomendasikan keminatan laboratorium pada mahasiswa Fakultas Ilmu Komputer Universitas Brawijaya. Sistem akan mengelola data dari 24 matakuliah dimana yang akan mengarah ke 5 rekomendasi keminatan laboratorium.

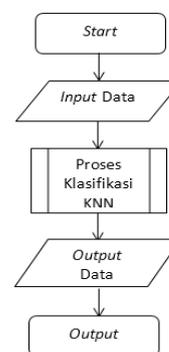
4.2 Preprocessing Data

Adapun dari setiap mata kuliah memiliki nilai yang terdiri dari nilai huruf yang nantinya setiap nilai huruf tersebut akan diubah menjadi nilai angka dengan proses *scoring*. Pada preproses data ini akan dilakukan penyeleksian terhadap data yang diseleksi dimana data tersebut telah dipisahkan berdasarkan pada kriteria yang ada. Kriteria pada preproses ini memiliki 24 data matakuliah serta 5 lab yang direkomendasikan. Nilai pembobotan telah didapatkan untuk setiap matakuliah yang bertujuan untuk mempermudah sistem dalam

tahap perhitungan.

4.3 Perancangan Proses

Perancangan dalam merekomendasikan keminatan laboratorium untuk sistem ini adalah dengan memasukkan data mata kuliah wajib dan mata kuliah pilihan. Selanjutnya sistem akan melakukan perhitungan jarak *euclidean* antar data latih. Hal ini dilakukan untuk mendukung proses selanjutnya untuk menentukan apakah data tersebut masuk dalam klasifikasi. Kemudian sistem akan menghitung jarak antar data latih dan data uji. Klasifikasi perekomendasi laboratorium dilanjutkan dengan pemilihan alternatif terbanyak. Pada Gambar 4.2 akan menunjukan diagram alir dari proses perhitungan yang dilakukan



Gambar 3. Flowchart Umum Sistem

4.4 Perhitungan Manual

Dalam perhitungan manual ini data yang telah diperoleh sebelumnya akan dihitung dengan menggunakan metode K-NN untuk mengetahui kualitas data yang dimiliki apakah cukup baik untuk bisa menjadi data latih sebagai dasar pembuatan program. Pada metode K-NN ini untuk menentukan kelas dilihat dari data yang memiliki jarak terdekat dari data test cara penghitungan.

Untuk melakukan proses manualisasi data pada penelitian “Implementasi Metode *K-Nearest Neighbor* Pada Rekomendasi Keminatan Laboratorium (Studi Kasus: Fakultas Ilmu Komputer Universitas Brawijaya)” ini, maka digunakan 227 dataset.

Penilaian huruf akan diubah menjadi data *testing* untuk lebih jelasnya dapat dilihat pada tabel 2.

Tabel 2. Perubahan Nilai (Scoring)

Nilai Huruf	Nilai Angka
A	4

B+	3,5
B	3
C+	2,5
C	2
D+	1,5
D	1
E	0

Langkah selanjutnya melakukan klasifikasi pada data *testing* menentukan parameter $K = 7$ dengan cara menghitung kuadrat jarak euclid (*euclidean distance*) masing-masing obyek terhadap data training yang diberikan untuk lebih jelasnya ditampilkan pada Tabel 3 berikut ini.

Tabel 3. Hasil *scoring* Data Training

NO	PD / DPK	PTIHK / STI	KB	KJ	Kelas
1	3	3	3,5	3	RPL
2	3	3,5	4	4	PAPB
3	3,5	3,5	4	3,5	KC
4	3	3,5	4	3,5	Jaringan
5	3	4	4	3,5	KC
6	3	2,5	4	4	PAPB
7	3,5	3	3,5	3	KC
8	2,5	4	4	3,5	PAPB
...
...
211	4	3	3,5	3	RPL
212	3,5	3	3,5	3,5	RPL
213	2,5	3	0	0	RPL
214	4	3	3,5	2	RPL

Berikut merupakan contoh perhitungan untuk data training ke 1 terhadap data *testing* untuk lebih jelasnya akan ditampilkan pada Tabel 4 berikut ini

Tabel 4. Contoh hasil perhitungan Jarak

No	Data Latih	Data Distance 1
1	data latih 1	3,905124838
2	data latih 2	3,082207001
3	data latih 3	4,213074887
4	data latih 4	3,122498999
5	data latih 5	3,31662479
6	data latih 6	3,122498999
7	data latih 7	3,240370349
...
...
204	data latih 204	3,640054945

221	data latih 221	17,82554347
222	data latih 222	17,82554347

Berikut merupakan perhitungan manual pencarian nilai *euclidean distance*.

$$\begin{aligned}
 & D1, D222 \\
 & = \sqrt{(4 - 3)^2 + (3.5 - 3)^2 + \dots + (4 - 3.5)^2 + (4 - 3)^2} \\
 & = \sqrt{(1)^2 + (0.5)^2 + (1)^2 + \dots + (0.5)^2 + (1)^2} \\
 & = \sqrt{1 + 0.25 + 1 + \dots + 0.25 + 1} \\
 & = 3,905 \tag{3}
 \end{aligned}$$

Pada langkah perhitungan selanjutnya adalah mengurutkan nilai – nilai kedalam kelompok dari jarak *euclidian* terkecil

Tabel 5. Urutkan berdasarkan jarak

No	Jarak Ke Masing-Masing Data Latih	Kelas
188	2,549509757	RPL
124	2,783882181	RPL
133	2,783882181	RPL
157	2,783882181	RPL
150	3,041381265	RPL
159	3,041381265	RPL
2	3,082207001	PAPB
93	3,082207001	RPL
97	3,082207001	RPL
4	3,122498999	Jaringan
6	3,122498999	PAPB
146	3,122498999	RPL
...
...
85	17,82554347	RPL
189	17,82554347	RPL
194	17,82554347	RPL
221	17,82554347	RPL
222	17,82554347	RPL

Langkah selanjutnya merupakan pengumpulan kategori Y (Klasifikasi *nearest neighbor*), pada proses ini pengambilan data sesuai dengan jumlah K yang telah ditentukan pada *step* 1 dimana jumlah K yang kita tentukan adalah $K = 7$, sehingga pada proses ini dipilih 7 data terbaik, yang hasilnya akan ditampilkan pada Tabel 6 sebagai berikut

Tabel 6. Urutan berdasarkan Jarak

No	Jarak Ke Masing-Masing Data Latih	Kelas
188	2,54951	RPL
124	2,783882	RPL

133	2,783882	RPL
157	2,783882	RPL
150	3,041381	RPL
159	3,041381	RPL
2	3,082207	Mobile

Kemudian dipilih dari data yang paling banyak tersebut sebagai kelas data yang akan dievaluasi, berikut merupakan munculan data terbanyak yang akan dijelaskan pada Tabel 7.

Tabel 7. Data terbanyak

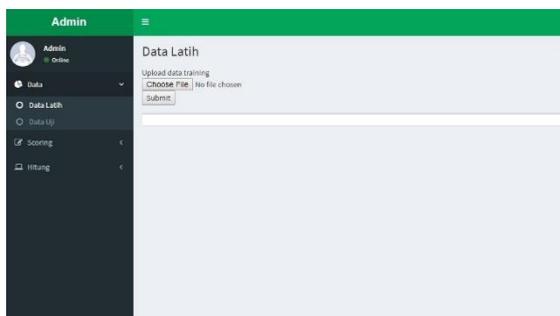
Kelas	Jumlah Data
Mobile	1
RPL	6
Jaringan	0
KC	0
Game	0

5. IMPLEMENTASI

Sistem ini memiliki dua algoritma yang diimplementasikan ke dalam bentuk *source code*, yaitu implementasi algoritma perhitungan *Euclidean Distance* dan algoritma *K-Nearest Neighbor* kemudian setelah implementasi algoritma maka akan ditampilkan dalam antarmuka sistem.

5.1 Antarmuka Halaman Utama

Pada halaman utama pengguna sistem dapat melakukan *submit* file dari *Microsoft Excel* yang sudah diformat menjadi ekstensi *.xlsx*. pada halaman ini terdapat tombol untuk memilih dan melakukan *submit* file. Halaman utama dapat dilihat pada gambar 4.

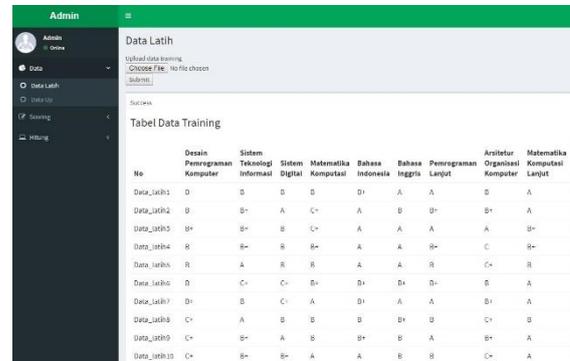


Gambar 4. Antarmuka Halaman Utama

5.2 Antarmuka halaman Upload Data Latih

Pada halaman upload data latih adalah sama dengan halaman utama karena halaman ini menjadi satu kesatuan dengan halaman utama yang dimana pengguna sistem dapat memilih file data latih dan melakukan *submit* file tersebut.

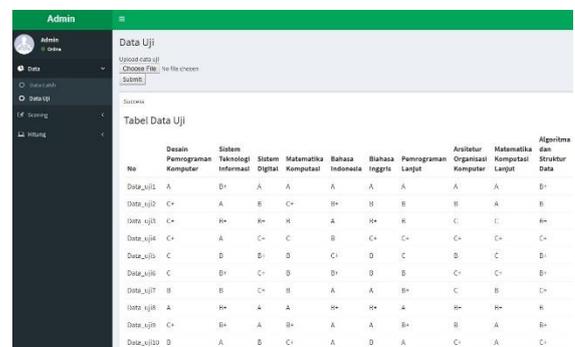
Kemudian setelah kita melakukan *submit* maka sistem akan menampilkan data latih berbentuk tabel. Halaman upload data latih dapat dilihat pada gambar 5.



Gambar 5. Antarmuka Halaman Upload Data Latih

5.3 Antarmuka halaman Upload Data Uji

Pada halaman upload data uji sama halnya dengan halaman upload data latih perbedaannya ialah file yang akan diupload adalah file data uji dimana pengguna system dapat memilih file data uji kemudian melakukan *submit* file. Setelah melakukan *submit* maka akan ditampilkan tabel yang berisikan data uji yang akan digunakan. Halaman upload data uji dapat dilihat pada gambar 6.



Gambar 6. Antarmuka Halaman Upload Data Uji

5.4 Antarmuka halaman Scoring

Pada halaman *scoring*, pengguna sistem akan melakukan *scoring* dari data latih dan data uji, yang pada awalnya isi dari data tersebut ialah huruf kemudian akan dirubah menjadi angka. Halaman *scoring* dapat dilihat pada gambar 7.

No	Desain Pemrograman Komputer	Sistem Teknologi Informasi	Sistem Digital	Matematika Komputer	Bahasa Indonesia	Bahasa Inggris	Pemrograman Lanjutan	Analisis Organisasi Komputer	Matematika Komputer Lanjutan	Algoritma dan Struktur Data
Data_train1	3	3	3	3	3,5	4	5	3	4	4
Data_train2	3	3,5	4	2,5	4	3	3,5	3,5	4	4
Data_train3	3,5	3,5	3	2,5	4	4	4	4	3,5	3,5
Data_train4	3	2,5	3	3,5	1	1	2,5	2	3,5	1
Data_train5	3	4	3	3	4	4	3	2,5	3	3,5
Data_train6	3	2,5	2,9	3,5	3,5	3,5	3,5	3	4	3
Data_train7	3,5	3	2,5	4	3,5	4	4	3,5	4	3
Data_train8	2,5	4	3	3	3	3,5	3	2,5	3	3,5
Data_train9	2,5	2,5	4	3	3,5	3	4	3,5	4	4
Data_train10	2,5	3,5	3,5	4	4	3	3	2,5	4	3

No	Desain Pemrograman Komputer	Sistem Teknologi Informasi	Sistem Digital	Matematika Komputer	Bahasa Indonesia	Bahasa Inggris	Pemrograman Lanjutan	Analisis Organisasi Komputer	Matematika Komputer Lanjutan	Algoritma dan Struktur Data
Data_uji1	4	2,5	4	4	4	4	4	4	4	3,5
Data_uji2	2,5	4	3	2,5	3,5	3	3	3	4	3
Data_uji3	2,5	2,5	2,5	3	4	3,5	3	2	2	2,5
Data_uji4	2,5	4	2,5	2	3	2,5	2,5	2,5	2,5	2,5
Data_uji5	2	3	3,5	3	2,5	3	2	3	2	3,5
Data_uji6	2	3,5	2,5	3	3,5	3	3	3,5	2,5	3,5
Data_uji7	3	3	2,5	3	4	4	3,5	2	3	2,5
Data_uji8	4	2,5	4	4	3,5	3,5	4	3,5	3,5	3
Data_uji9	2,5	3,5	4	3,5	4	4	3,5	3	4	3,5
Data_uji10	2	4	3	2,5	4	3	4	2,5	4	2,5

Gambar 7. Antarmuka Halaman Scoring

5.5 Antarmuka halaman Hitung Jarak Euclidean

Pada halaman hitung jarak, data latih dan data uji yang sudah dirubah dari huruf menjadi angka kemudian akan dihitung jarak *euclidean* dari data latih dan data uji tersebut. Halaman perhitungan jarak *euclidean* dapat dilihat pada gambar 8.

	data_train1	data_train2	data_train3	data_train4	data_train5	data_train6	data_train7
data_uji1	2,5	3,5	4,2	3,5	3,5	3,5	3,5
data_uji2	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji3	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji4	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji5	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji6	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji7	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji8	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji9	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji10	4,0	3,5	3,5	3,5	3,5	3,5	3,5

Gambar 8. Antarmuka Halaman Hitung Jarak Euclidean

Setelah melakukan perhitungan *euclidean* maka pengguna dapat memasukkan nilai K yang akan digunakan pada proses selanjutnya. Halaman *input* nilai K dapat dilihat pada gambar 9.

Gambar 9. Antarmuka Halaman Input nilai K

5.6 Antarmuka halaman Hitung Nilai K-Nearest Neighbor

Pada halaman hitung nilai KNN, hasil dari

perhitungan jarak Euclidean kemudian akan dilakukan proses pengambilan hasil berdasarkan K yang sudah *diinputkan*. Halaman Hitung nilai KNN dapat dilihat pada gambar 10

	Urutan ke-1	Urutan ke-2	Urutan ke-3	Urutan ke-4	Urutan ke-5	Urutan ke-6	Urutan ke-7
data_uji1	2,5	3,5	4,2	3,5	3,5	3,5	3,5
data_uji2	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji3	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji4	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji5	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji6	4,0	3,5	3,5	3,5	3,5	3,5	3,5
data_uji7	4,0	3,5	3,5	3,5	3,5	3,5	3,5

Gambar 10. Antarmuka Halaman Hitung Nilai K-NN

6. PENGUJIAN DAN ANALISIS

6.1 Pengujian Akurasi

Pengujian sistem dilakukan untuk mengetahui tingkat akurasi yang dihasilkan dari system. Data yang akan diuji berjumlah 30 data sampel nilai akademik mahasiswa matematika. Hasil yang diperoleh dari sistem, akan dicocokkan dengan data yang diperoleh dari pakar dengan nilai K yang akan dipilih. Hasil yang diperoleh adalah sebagai berikut

6.1.1 Pengujian terhadap jumlah data latih

Pada tahap ini pengujian dilakukan terhadap jumlah data latih yang berbeda, dalam pengujian ini diambil 4 set data latih yang masing-masing berjumlah 40, 80, 120, 185 dengan nilai K yang digunakan adalah K=10. Hasil pengujian ini dapat dilihat pada tabel 8.

Tabel 8 Hasil pengujian terhadap jumlah data latih.

No	Jumlah Data latih	Jumlah Indeks Yang sesuai	Akurasi
1	40	6	20%
2	80	11	36,66%
3	120	12	40%
4	185	23	76,66%

Dari hasil pengujian diatas maka dapat dilihat data latih yang berjumlah 185 memiliki akurasi sebanyak 76,66% yaitu akurasi tertinggi dari keempat jumlah data latih yang berbeda. Dan yang terkecil ialah data yng berjumlah 40 yaitu 20%.

6.1.2. Pengujian mencari nilai K optimal

Pada tahap ini pengujian dilakukan terhadap nilai K yang dimana ada beberapa nilai K yang akan digunakan. Pada pengujian ini data

latih yang digunakan sebanyak 185 karena memiliki akurasi yang tinggi pada pengujian jumlah data latih data latih yaitu 76,66%. Nilai K yang digunakan adalah dari 2 sampai dengan 10. Hasil pengujian ini dapat dilihat pada tabel 9.

Tabel 9. Hasil pengujian terhadap nilai K

Nilai K	Jumlah Indeks	Akurasi
2	11	36,66%
3	11	36,66%
4	15	50%
5	20	66,66%
6	16	53,33%
7	22	73,33%
8	19	63,3%
9	21	70%
10	23	76,66%

Dari hasil pengujian diatas maka didapatkan K yang paling optimnal adalah K=10 yaitu dengan hasil akurasi sebanyak 23.

6.1.3 Pengujian Keseimbangan Data

Pada tahap ini akan dilakukan pengujian keseimbangan data yang mana data yang akan menjadi data latih adalah jumlah data yang sama pada masing-masing kelas keminatan yang diambil dan yang menjadi acuan untuk menentukan jumlah data tersebut adalah kelas yang paling sedikit memiliki jumlah data. Sedangkan yang menjadi data uji adalah sisa data dari data latih yang telah ditentukan sebelumnya. Hasil pengujian ini dapat dilihat pada tabel 11.

Tabel 10. Data yang digunakan untuk pengujian keseimbangan

Kelas	Data Latih	Data Uji
Game	20	0
Mobile	20	12
Jaringan	20	16
RPL	20	43
KC	20	56
Jumlah data	100	127

Tabel 11. Hasil dari pengujian keseimbangan data

No	Jumlah Data latih	Jumlah data uji	Jumlah Indeks	Akurasi
1	100	127	23	18,11%

Dari hasil yang diperoleh dari pengujian keseimbangan data maka didapatkan banyaknya

indeks yang sesuai dengan data uji adalah sebanyak 20 kemudian setelah diakurasi maka didapatkan hasil 18,11%.

6.1.4 Pengujian Variasi Data

Pada pengujian ini akan dilakukan pengambilan data latih dan data uji secara acak dengan jumlah data berdasarkan perbandingan 40%:60%, 50%:50%, 30%:70% dari dataset yang ada. Pada pengujian variasi data ini K yang akan digunakan adalah K=10. Hasil pengujian ini dapat dilihat pada tabel 12

Tabel 12. Hasil dari pengujian variasi data

No	Jumlah data uji (%)	Jumlah data latih (%)	Hasil Indeks	Akurasi
1	70	30	38	23,89%
2	50	50	28	24,77 %
3	40	60	25	27,47%

Berdasarkan hasil pengujian variasi data maka didapatkan hasil paling akurat adalah pada pengambilan data uji dan data latih sebanyak 60:40 yang mana 60% mewakili data uji dan 40% untuk mewakili data latih yaitu dengan prosentase sebesar 27,47%.

6.2 Analisis Pegujian Akurasi

Setelah dilakukan 4 pengujian maka dari hasil yang didapatkan dari sistem dapat dianalisis pada masing masing pengujian.

6.2.1. Pengujian terhadap jumlah Data

Pada pengujian terhadap data latih yang berbeda-beda jumlahnya, K yang digunakan adalah K=10 dan jumlah data latih yang digunakan adalah 40, 80, 120, 185 data latih. Setelah dilakukan pengujian maka didapatkan hasil yang memiliki akurasi tertinggi adalah data yang berjumlah 185 data yaitu akurasinya mencapai 76,66% sedangkan yang terendah adalah 40 data yaitu akurasinya hanya 20%. Disini terlihat semakin banyak data latih yang digunakan maka akurasi dari metode *K-Nearest Neighbor* akan semakin bagus.

6.2.2. Pengujian mencari nilai K optimal

Setelah melakukan pengujian akurasi terhadap jumlah data latih maka sudah didapatkan data latih untuk mencari K optimal yaitu dengan menggunakan 185 data latih. Nilai K yang di uji adalah 2-10 setelah kemudian setelah melakukan pengujian maka didapatkan nilai K optimal yaitu K=10 dengan akurasi

sebanyak 76,66%. Dan nilai akurasi terendah dari nilai K adalah K=2 dan K=3 akurasi hanya 36,66%.

6.2.3. Pengujian keseimbangan Data

Pada pengujian ini akan didapatkan hasil nilai kebenarannya sebesar 18,11%, hasil itu berarti data yang sudah ada memiliki komposisi yang cukup baik walaupun tidak seimbang persis namun proporsinya sudah memenuhi untuk dijadikan data latih

6.2.4. Pengujian Variasi Data

Dari 3 variasi data yang diuji, dapat dilihat bahwa variasi yang memberikan akurasi yang terbesar adalah pada variasi data ke-3 yaitu dengan membagi *dataset* menjadi 2 dengan perbandingan 60% untuk data uji dan 40% untuk data latih. Prosentasi yang dihasilkan adalah sebesar 27,47%.

6.2.5. Analisis Hasil

Proses analisis pada hasil pengujian akurasi implementasi metode *K-Nearest Neighbor* pada Keminatan studi, dengan melihat hasil persentase keakurasian dan ketidakakurasian sistem dari pengujian menunjukkan adanya ketidaksesuaian antara analisa sistem dengan analisa pakar. Ketidaksesuaian hasil yang diperoleh sistem dengan yang didapat dari hasil yang sudah terverifikasi disebabkan beberapa hal yaitu:

- **Dari sisi metode :**
Sebelum menjalankan algoritma maka K yang dinisialisasikan adalah random sehingga pengelompokan data yang dihasilkan dapat berbeda-beda.
- **Dari sisi data :**
Ketidaksesuaian data nilai dengan keminatan yang yang dipilih dikarenakan adanya factor-faktor yang tidak bisa di *cover* dengan perhitungan membuat akurasi system menjadi berkurang.

7. PENUTUP

Berdasarkan hasil perancangan, implementasi dan pengujian yang dilakukan maka dapat ditarik kesimpulan :

- Implementasi metode *K-Nearest Neighbor* untuk rekomendasi keminatan pada studi mahasiswa teknik informatika dapat digunakan sebagai sarana bagi mahasiswa yang masih bingung untuk menentukan keminatan yang akan diambil.

- Berdasarkan pengujian yang telah dilakukan dari 30 data uji dengan membandingkan antara hasil sistem dengan data yang diperoleh dari pakar maka diperoleh hasil akurasi sebesar 76,66%.
- Terjadinya ketidaksesuaian hasil dari sistem dengan hasil sudah diverifikasi disebabkan oleh kelemahan dari metode dan adanya faktor-faktor luar yang mempengaruhi hasil dan tidak dapat tercover oleh perhitungan sistem.
- Faktor lingkungan dan teman sangat mempengaruhi mahasiswa dalam mengambil keminatan yang akan ditempuh sehingga ada ketidaksesuaian antara hasil yang diperoleh dari sistem.

DAFTAR PUSTAKA

- Parvin, Hamid. 2008. *MKNN: Modified K-Nearest Neighbor*. *World Congress on Engineering and Computer Science*. USA. San Francisco.
- Ayu K, Noviana. 2014. Implementasi Algoritma *Modified K-Nearest Neighbor* (MKNN) Untuk Menentukan Tingkat Resiko Penyakit Lemak Darah (Profil Lipid). Universitas Brawijawa. Malang
- Tri Halomoan Simanjuntak, Wayan Firdaus Mahmudy dan Sutrisno. 2014. Implementasi *Modified K-Nearest Neighbor* dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai.
- Hardiyanti, Siti. 2014. Implementasi Metode *Modified K-Nearest Neighbor* (MKNN) Pada Penentuan Keminatan Sekolah Menengah Atas (SMA) (Studi Kasus: SMA Negeri 1 Seririt). Universitas Brawijaya. Malang.
- Shouman, Mai. 2012. *Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients*. University of New South Wales at the Australian Defence Force Academy Northcott Drive. Canberra ACT2600
- Kamber, H. &. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.