

**KLASIFIKASI KELOMPOK KEILMUAN INFORMATIKA
BERDASARKAN NILAI MENGGUNAKAN METODE
*K-NEAREST NEIGHBOR***

SKRIPSI

Oleh:
CANDRA HAFIDZ ARDANA
NIM. 200605110074



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**KLASIFIKASI KELOMPOK KEILMUAN INFORMATIKA
BERDASARKAN NILAI MENGGUNAKAN METODE
*K-NEAREST NEIGHBOR***

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam Memperoleh
Gelar Sarjana Komputer (S.Kom)

Oleh:
CANDRA HAFIDZ ARDANA
NIM. 200605110074

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN

**KLASIFIKASI KELOMPOK KEILMUAN INFORMATIKA
BERDASARKAN NILAI MENGGUNAKAN METODE
*K-NEAREST NEIGHBOR***

SKRIPSI

Oleh:
CANDRA HAFIDZ ARDANA
NIM. 200605110074

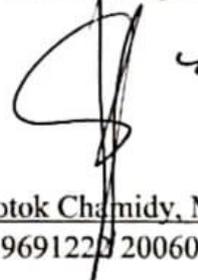
Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 6 Mei 2024

Pembimbing I,



Syahiduz Zaman, M.Kom
NIP. 19700502 200501 1 005

Pembimbing II,



Dr. Totok Chaididy, M.Kom
NIP. 19691220 200604 1 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachri Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI KELOMPOK KEILMUAN INFORMATIKA BERDASARKAN NILAI MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

SKRIPSI

Oleh:

CANDRA HAFIDZ ARDANA
NIM. 200605110074

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 10 Juni 2024

Susunan Dewan Penguji

Ketua Penguji : Prof. Dr. Suhartono, M.Kom
NIP. 19680519 200312 1 001

Anggota Penguji I : Agung Teguh Widodo Almais, M.T
NIP. 19860301 202321 1 016

Anggota Penguji II : Syahiduz Zaman, M.Kom
NIP. 19700502 200501 1 005

Anggota Penguji III : Dr. Totok Chamidy, M.Kom
NIP. 19691222 200604 1 001

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fochid Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Candra Hafidz Ardana
NIM : 200605110074
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Klasifikasi Kelompok Keilmuan Informatika
Berdasarkan Nilai Menggunakan Metode
K-Nearest Neighbor.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 6 Mei 2024

Yang membuat pernyataan,



Candra Hafidz Ardana
NIM. 200605110074

MOTTO

“Jangan Merasa Tertinggal, Setiap Orang Punya Proses dan Rezekinya Masing-Masing.”

-- Q.S Maryam : 4 --

“Kita akan Baik-Baik saja Selama Kita Tidak Membandingkan Diri Kita dengan Orang Lain, Karena Manusia Mempunyai Bagiannya Masing-Masing, Fokuslah pada Bagianmu dan Bersyukurlah atas Apa yang Kamu Miliki Saat Ini”

-- Gus Iqdam Muhammad --

HALAMAN PERSEMBAHAN

Puji syukur kepada Allah SWT, sholawat serta salam kehadiran Rasul-Nya. Penulis mempersembahkan hasil karya ini kepada kedua orang tua penulis, Bapak Eka Asta Ardana dan Ibu Nanik Yulianti yang selalu memberikan dukungan, do'a, semangat, serta motivasi yang begitu besar kepada penulis.

Dosen pembimbing penulis Bapak Syahiduz Zaman, M.Kom dan Bapak Dr. Totok Chamidy, M.Kom yang telah sabar membimbing penulis sejak penyusunan tahap awal skripsi mulai praproposal sampai tahap akhir skripsi dan selalu memberikan semangat, masukan, serta Arah dalam penyusunan skripsi ini.

Seluruh bapak/ibu dosen dan jajaran civitas akademika program studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberi ilmu dan memberi kelancaran dalam penyusunan skripsi ini.

Teman-teman seperjuangan mulai penulis masuk ke Universitas Islam Negeri Maulana Malik Ibrahim Malang khususnya Acid Blue Esport yang sudah penulis anggap sebagai keluarga yang selalu memberikan semangat dan dukungan untuk belajar bersama tanpa menjatuhkan.

Keluarga besar Integer (Teknik Informatika angkatan tahun 2020) yang telah memberikan semangat dan doa.

Orang-orang yang penulis sayangi, yang tidak bisa penulis sebutkan satu per satu yang selalu memberikan semangat dan dukungan untuk menyelesaikan skripsi ini hingga selesai.

Diri saya sendiri.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan rahmat dan karunia kepada kita semua, sehingga penulis bisa menyelesaikan skripsi ini yang berjudul “Klasifikasi Kelompok Keilmuan Informatika Berdasarkan Nilai Menggunakan Metode *K-Nearest Neighbor*”. Dalam kesempatan ini, penulis ingin mengucapkan terima kasih sebanyak-banyaknya kepada semua pihak yang telah memberikan bantuan, dukungan, dan dorongan kepada penulis sepanjang perjalanan penulisan skripsi ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Prof. Dr. M. Zainuddin MA, selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT., IPM, selaku Kepala Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Syahiduz Zaman, M.Kom, selaku Dosen Pembimbing I yang telah membimbing penulis dalam penyusunan skripsi ini hingga selesai.
5. Dr. Totok Chamidy, M.Kom, selaku Dosen Pembimbing II yang telah membimbing penulis dalam penyusunan skripsi ini hingga selesai.
6. Dr. Cahyo Crysdiyan, M.Cs, selaku Dosen Wali yang telah memberikan saran dan masukan untuk kebaikan penulis.

7. Seluruh dosen Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan ilmu selama kegiatan perkuliahan berlangsung.
8. Orang tua dan keluarga penulis yang senantiasa memberikan do'a, dukungan, semangat, serta kasih sayang yang menjadi pendorong utama penulis untuk menyelesaikan skripsi ini.
9. Adlian Aldita Alif Aisyah Ainur Khoyum, teman terdekat sekaligus teman untuk bercerita yang selalu memberikan dukungan kepada penulis
10. Keluarga besar Acid Blue Esport yang senantiasa selalu menemani dan memberikan semangat hingga dapat menempuh gelar sarjana bersama.
11. Teman-teman angkatan 2020 yang selalu memberikan semangat dan dorongan kepada penulis.
12. Semua pihak yang telah banyak membantu penulis dalam penyusunan skripsi ini yang tidak bisa penulis sebutkan semuanya.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan dan kelemahan, oleh karena itu kami sangat mengharapkan kritik, saran, dan masukan yang membangun untuk pembaca. Penulis berharap semoga skripsi ini dapat memberikan manfaat kepada pembaca dan dapat berkontribusi dalam pengembangan ilmu pengetahuan dan kehidupan di masa yang akan datang.

Malang, 5 Mei 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
ABSTRAK	xiv
ABSTRACT	xv
مستخلص البحث.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah.....	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB II KAJIAN PUSTAKA	7
2.1 Penelitian Terdahulu	7
2.2 Program Studi Teknik Informatika UIN Malang	11
2.3 <i>Data Mining</i>	13
2.3.1 Pengelompokkan <i>Data Mining</i>	15
2.3.2 Pemrosesan <i>Data Mining</i>	17
2.4 Klasifikasi	19
2.5 <i>K-Nearest Neighbor</i> (KNN).....	20
2.6 <i>Confusion Matrix</i>	22
2.7 Python	24
2.8 <i>Jupyter Notebook</i>	25
BAB III METODOLOGI PENELITIAN	27
3.1 Tahapan Penelitian	27
3.2 Analisis Kebutuhan Informasi.....	28
3.3 Metode Pengumpulan Data	28
3.4 Teknik Analisis Data.....	29
3.5 Flowchart Algoritma <i>K-Nearest Neighbor</i>	30
3.6 Skenario Pengujian.....	31
BAB IV HASIL DAN PEMBAHASAN	34
4.1 <i>Data Understanding</i>	34
4.2 <i>Data Preparation</i>	34
4.3 <i>Data Modelling</i>	35
4.3.1 Menentukan Atribut dan Label	35

4.3.2 Membagi Data <i>Training</i> dan Data <i>Testing</i>	36
4.3.3 Implementasi Algoritma KNN.....	37
4.4 <i>Evaluation</i> atau Pengujian.....	38
4.4.1 Pengujian Terhadap Variasi Data	39
4.4.2 Pengujian Mencari Nilai K Terbaik	39
4.4.3 <i>Confusion Matrix</i>	40
4.5 Pembahasan.....	54
4.5.1 Pembahasan Pengujian Variasi Data.....	54
4.5.2 Pembahasan Nilai K Terbaik	56
4.5.3 Pembahasan Pengujian Kelas.....	56
4.6 Integrasi Islam.....	61
BAB V KESIMPULAN DAN SARAN	63
5.1 Kesimpulan	63
5.2 Saran.....	64
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian	27
Gambar 3. 2 Flowchart Algoritma K-Nearest Neighbor.....	31
Gambar 4. 1 Daftar Nilai Mata Kuliah dan IPK Mahasiswa Teknik Informatika Tahun 2018 UIN Malang.....	34
Gambar 4. 2 Dataset.....	35
Gambar 4. 3 Jumlah Nilai pada Setiap Label.....	36
Gambar 4. 4 Plot Akurasi Pelatihan dan Pengujian	38
Gambar 4. 5 Confusion Matrix Skenario A	43
Gambar 4. 6 Confusion Matrix Skenario B	46
Gambar 4. 7 Confusion Matrix Skenario C	49
Gambar 4. 8 Confusion Matrix Skenario D	53
Gambar 4. 9 Grafik Nilai Akurasi Setiap Class	57
Gambar 4. 10 Grafik Presisi Setiap Class	58
Gambar 4. 11 Grafik Nilai Recall Setiap Class.....	58
Gambar 4. 12 Grafik Nilai f-1 Score Setiap Class	59

DAFTAR TABEL

Tabel 2. 1 Perbandingan penelitian.....	7
Tabel 2. 2 Confusion Matrix	23
Tabel 3. 1 Nilai Mata Kuliah.....	30
Tabel 3. 2 Skenario Pengujian Setiap Class.....	32
Tabel 4. 1 Kelompok Keilmuan Teknik Informatika Universitas Islam Maulana Malik Ibrahim Malang.....	36
Tabel 4. 2 Pembagian Dataset.....	39
Tabel 4. 3 Pengujian Nilai K.....	40
Tabel 4. 4 Confusion Matrix Class Multimedia.....	41
Tabel 4. 5 Confusion Matrix Class Information System	41
Tabel 4. 6 Confusion Matrix Class Software Engineering	41
Tabel 4. 7 Confusion Matrix Class Intelligent System	41
Tabel 4. 8 Confusion Matrix Class System & Network	42
Tabel 4. 9 Confusion Matrix Class Web & Mobile Programming	42
Tabel 4. 10 Performa Skenario A.....	43
Tabel 4. 11 Confusion Matrix Class Multimedia.....	44
Tabel 4. 12 Confusion Matrix Class Information System.....	44
Tabel 4. 13 Confusion Matrix Class Software Engineering	45
Tabel 4. 14 Confusion Matrix Class Intelligent System	45
Tabel 4. 15 Confusion Matrix Class System & Network	45
Tabel 4. 16 Confusion Matrix Class Web & Mobile Programming	46
Tabel 4. 17 Performa Skenario B	46
Tabel 4. 18 Confusion Matrix Class Multimedia.....	47
Tabel 4. 19 Confusion Matrix Class Information System.....	48
Tabel 4. 20 Confusion Matrix Class Software Engineering	48
Tabel 4. 21 Confusion Matrix Class Intelligent System	48
Tabel 4. 22 Confusion Matrix Class System & Network	49
Tabel 4. 23 Confusion Matrix Class Web & Mobile Programming	49
Tabel 4. 24 Performa Skenario C	50
Tabel 4. 25 Confusion Matrix Class Multimedia.....	51
Tabel 4. 26 Confusion Matrix Class Information System.....	51
Tabel 4. 27 Confusion Matrix Class Software Engineering	51
Tabel 4. 28 Confusion Matrix Class Intelligent System	52
Tabel 4. 29 Confusion Matrix Class System & Network	52
Tabel 4. 30 Confusion Matrix Class Web & Mobile Programming	52
Tabel 4. 31 Performa Skenario D.....	53

ABSTRAK

Ardana, Candra Hafidz. 2024. **Klasifikasi Kelompok Keilmuan Informatika Berdasarkan Nilai Menggunakan Metode *K-Nearest Neighbor***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Syahiduz Zaman, M.Kom, (II) Dr. Totok Chamidy, M.Kom.

Kata kunci: *K-Nearest Neighbor*, Kelompok Keilmuan, Klasifikasi

Kelompok keahlian pada Program Studi Teknik Informatika Universitas Islam Maulana Malik Ibrahim Malang merupakan kelompok keahlian mengacu pada distribusi atau pengelompokan bidang-bidang khusus yang mencakup berbagai aspek atau disiplin ilmu di dunia teknik informatika. Tujuan dari pembagian ini adalah untuk memberikan spesialisasi kepada mahasiswa agar mereka dapat mengembangkan keahlian dan pengetahuan yang mendalam dalam bidang tertentu sesuai dengan minat dan kebutuhan pasar. Metode yang digunakan dalam penelitian ini menggunakan metode *K-Nearest Neighbor* (KNN) untuk memecahkan masalah yang terjadi dengan harapan akan memberikan pemahaman yang lebih baik tentang peminatan kelompok keahlian informatika. Pengujian dilakukan dengan membagi dataset rasio ke dalam beberapa bagian yang telah ditentukan dan mencari nilai akurasi dari masing-masing kelas. Hasil dari pengujian rasio pembagian data sebanyak 95 data dengan nilai $K=10$ diperoleh nilai tingkat akurasi sebesar 60% dengan pembagian data sebanyak 80 data sebagai data latih dan 15 data sebagai data uji. Hasil pengujian dari masing-masing kelas diperoleh *Class Multimedia* memperoleh akurasi sebesar 57.75%, *Class Information System* memperoleh akurasi sebesar 78.75%, *Class Software Engineering* memperoleh akurasi sebesar 92.25%, *Class Intelligent System* memperoleh akurasi sebesar 57.25% , *Class System & Network* memperoleh akurasi sebesar 95.25%, dan *Class Web & Mobile Programming* memperoleh akurasi sebesar 95.25%. Maka implementasi metode *K-Nearest Neighbor* (KNN) untuk klasifikasi kelompok keilmuan teknik informatika UIN Malang dapat digunakan sebagai sarana bagi mahasiswa yang masih mengalami kebingungan untuk menentukan keminatan kelompok keilmuan yang akan diambil.

ABSTRACT

Ardana, Candra Hafidz. 2024. **Classification of Informatics Scientific Groups Based on Values Using the K-Nearest Neighbor Method**. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Supervisors: (I) Syahiduz Zaman, M.Kom, (II) Dr. Totok Chamidy, M.Kom.

The expertise group in the Informatics Engineering Study Program at the Islamic University of Maulana Malik Ibrahim Malang is an expertise group referring to the distribution or grouping of special fields that cover various aspects or disciplines in the world of informatics engineering. The aim of this division is to provide specialization to students so that they can develop in-depth skills and knowledge in certain fields according to interests and market needs. The method used in this research uses the K-Nearest Neighbor (KNN) method to solve problems that occur in the hope of providing a better understanding of the informatics expertise group specialization. Testing is carried out by dividing the ratio dataset into several predetermined parts and looking for the accuracy value of each *Class*. The results of testing the data sharing ratio of 95 data with a value of $K=10$ obtained an accuracy level of 60% with a data sharing of 80 data as training data and 15 data as test data. The test results for each *Class* obtained that the *Multimedia Class* obtained an accuracy of 57.75%, the *Information System Class* obtained an accuracy of 78.75%, the *Software Engineering Class* obtained an accuracy of 92.25%, the *Intelligent System Class* obtained an accuracy of 57.25% , the *System & Network Class* obtained an accuracy of 95.25%, and the *Web & Mobile Programming Class* obtained an accuracy of 95.25%. So the implementation of the K-Nearest Neighbor (KNN) method for *Classifying* scientific groups in informatics engineering at UIN Malang can be used as a means for students who are still confused about determining the interests of the scientific groups they will take.

Keywords: *K-Nearest Neighbor, Scientific Groups, Classification*

مستخلص البحث

أردانا، جانديرا حافظ. 2024. تصنيف المجموعات العلمية المعلوماتية على أساس القيمة باستخدام طريقة كي أقرب جار (*K-Nearest Neighbor*). البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: شاهد الزمان، الماجستير. المشرف الثاني: د. توتوك حميدي، الماجستير.

الكلمات الرئيسية: تصنيف، مجموعة علمية، كي أقرب جار.

مجموعة الخبرة في قسم الهندسة المعلوماتية بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج هي مجموعة من الخبرات التي تشير إلى توزيع أو تجميع المجالات الخاصة التي تغطي مختلف الجوانب أو التخصصات في عالم الهندسة المعلوماتية. الغرض من هذا التقسيم هو توفير التخصص للطلاب حتى يتمكنوا من تطوير الخبرة والمعرفة المتعمقة في مجال معين وفقا لاهتمامات السوق واحتياجاته. تستخدم الطريقة المستخدمة في هذه الدراسة طريقة كي أقرب جار (*K-Nearest Neighbor (KNN)*) لحل المشكلات التي تحدث على أمل أن توفر فهما أفضل لتخصص مجموعات الخبرة المعلوماتية. تم إجراء الاختبار عن طريق تقسيم مجموعة بيانات النسبة إلى عدة أجزاء محددة مسبقا والبحث عن قيمة الدقة لكل فئة. حصلت نتائج اختبار نسبة مشاركة البيانات ل 95 بيانات بقيمة كي = 10 على قيمة مستوى دقة 60%. مع تقسيم البيانات ل 80 بيانات كبيانات تدريب و 15 بيانات كبيانات اختبار. وحصلت نتائج اختبار كل صنف على دقة بلغت 57.75% لفئة الوسائط المتعددة، ودقة بلغت 78.75% لفئة نظم المعلومات، ودقة 92.25% لفئة هندسة البرمجيات، ودقة 57.25% لفئة النظام الذكي، ودقة 95.25% لفئة النظام والشبكات، ودقة 95.25% لفئة برمجة الويب والجوال. لذلك، يمكن استخدام تطبيق طريقة كي أقرب جار (*K-Nearest Neighbor (KNN)*) لتصنيف المجموعات العلمية لهندسة المعلوماتية بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج كوسيلة للطلاب الذين لا يزالون في حيرة من أمرهم لتحديد اهتمام المجموعة العلمية التي يجب اتخاذها.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kelompok keahlian yang ada dalam Program Studi Teknik Informatika Universitas Islam Maulana Malik Ibrahim Malang merupakan kelompok keahlian mengacu pada pembagian atau pengelompokan bidang-bidang khusus yang mencakup berbagai aspek atau disiplin ilmu di dalam dunia teknik informatika. Tujuan dari pembagian ini adalah untuk memberikan spesialisasi kepada mahasiswa agar mereka dapat mengembangkan keahlian dan pengetahuan yang mendalam dalam area tertentu sesuai dengan minat dan kebutuhan pasar (*SI-Teknik Informatika UIN Maulana Malik Ibrahim Malang*, diakses pada 05-06-2024 pukul 12.01).

Perubahan minat mahasiswa terhadap topik skripsi mereka merupakan permasalahan yang terjadi. Beberapa kesalahan dalam pemilihan mata kuliah peminatan di semester-semester sebelumnya menjadi salah satu penyebabnya. Penyebab lainnya yaitu adanya perbedaan antara nilai akademik *core courses* dan *elective courses* bersamaan dengan minat mahasiswa itu sendiri (Wulandari & Etikasari, 2019). Misalnya, peminatan dalam bidang *Software Engineering* memerlukan dukungan nilai yang baik pada *core courses* seperti *Algorithms & Programming*, *Object-Oriented Programming*, dan *Software Engineering*. Selain itu, *elective courses* yang relevan seperti *Software Quality*, *Software Management* dan lain-lain yang mendukung peminatan tersebut. Faktor eksternal seperti *peer*

influence dan *faculty guidance* juga dapat mempengaruhi pilihan minat mahasiswa (Prasetyo & Maksum, 2013).

Penelitian ini memiliki nilai penting dalam menghadapi dinamika perkembangan bidang teknik informatika dan kebutuhan industri yang semakin kompleks. Dengan adanya sistem klasifikasi peminatan kelompok keahlian mahasiswa teknik informatika berdasarkan nilai akademis menggunakan metode *K-Nearest Neighbor* (KNN), perguruan tinggi dapat lebih efektif merespons keberagaman keahlian mahasiswa (Anshori dkk., 2018). Hal ini tidak hanya memberikan manfaat dalam perencanaan kurikulum dan pengembangan program studi, tetapi juga memastikan bahwa lulusan teknik informatika memiliki keahlian yang sesuai dengan tuntutan industri. Tidak hanya itu, diharapkan bahwa hasil penelitian ini bisa memberikan landasan data yang kuat bagi kebijakan pengembangan SDM di bidang teknik informatika, sehingga daya saing lulusan di pasar kerja yang terus berubah dan menuntut keahlian yang beragam dapat ditingkatkan.

Dalam Islam, pendidikan dianggap sangat penting dan mencakup pengembangan bakat dan potensi setiap orang. Al-Quran menunjukkan betapa pentingnya memanfaatkan potensi yang diberikan kepada setiap orang (Hadi, 2013). Penelitian ini diilhami oleh ajaran Islam yang mendorong setiap orang untuk mengembangkan kemampuan mereka sendiri yang sesuai dengan preferensi mereka. Sebagai contoh, dalam Surat Al-Qasas ayat 68 yang menyatakan:

وَرَبُّكَ يَخْلُقُ مَا يَشَاءُ وَيَخْتَارُ مَا كَانَ لَهُمُ الْخِيَرَةُ سُبْحَانَ اللَّهِ وَتَعَالَى عَمَّا يُشْرِكُونَ

“Dan Tuhanmu menciptakan apa yang Dia kehendaki dan memilihnya. Sekali-kali tidak ada pilihan bagi mereka. Maha Suci Allah dan Maha Tinggi dari apa yang mereka persekutukan (dengan Dia)” (Q.S Al-Qasas:68).

Menurut Tafsir Jalalain bahwasanya apa yang dikehendaki oleh Allah diciptakan sesuai dengan kehendak-Nya. Allah juga memilih siapa saja yang Dia kehendaki untuk menjadi rasul-Nya. Tidak ada bagi manusia pilihan dalam masalah ini, tetapi pilihan itu mutlak milik Allah yang Maha Suci dari segala sekutu yang mereka sifatkan kepada-Nya. Ayat Al-Quran tersebut menggambarkan bahwa Allah adalah Sang Pencipta yang Maha Kuasa yang telah menciptakan beragam pilihan dan kebebasan bagi manusia. Dalam penelitian ini, ayat tersebut dapat diartikan sebagai menciptakan beragam minat dan preferensi mahasiswa dalam memilih peminatan kelompok keahlian yang sesuai dengan keinginan mereka.

Penelitian ini diharapkan akan memberikan dampak positif pada dunia pendidikan terutama universitas, yang merupakan komponen penting dari tatanan sosial masyarakat Islam. Penelitian ini berusaha untuk menciptakan lingkungan pendidikan yang lebih inklusif dan berpusat pada mahasiswa, sejalan dengan ajaran Islam tentang keadilan dan pemberian kesempatan yang adil. Dengan demikian, penelitian ini tidak hanya memiliki nilai akademik, tetapi juga mencerminkan nilai-nilai dan prinsip-prinsip Islam dalam hal pengembangan potensi setiap orang.

Sebagaimana penelitian yang telah dilakukan oleh (Zulfallah, 2022), yang berjudul "Implementasi Algoritma KNN Dalam Mengukur Ketepatan Kelulusan Mahasiswa UIN Syarif Hidayatullah Jakarta.", menghasilkan prediksi yang baik berdasarkan *Confusion Matrix* yang mengindikasikan kemampuan algoritma KNN dalam memprediksi kelulusan mahasiswa. Sementara itu, penelitian (Umaidah &

Purwanto, 2019), yang berjudul “Penerapan Algoritma K-Nearest Neighbor (KNN) Dengan Pencarian Optimal Untuk Prediksi Prestasi Siswa.”, menunjukkan bahwa penggunaan algoritma KNN pada kluster 2 berhasil memberikan prediksi yang sangat baik dengan tingkat akurasi mencapai 93,63%. Penentuan nilai k -optimal didasarkan pada hasil validasi silang *k-fold* untuk mengevaluasi performa algoritma yang digunakan.

Merujuk pada hal tersebut, kebaruan penelitian ini penulis akan memperluas fokus cakupan penelitian pada aspek-aspek yang berbeda atau metode yang berbeda dalam konteks penggunaan algoritma KNN, memberikan kontribusi penting dalam pengembangan sistem algoritma tersebut dalam konteks pendidikan. Metode KNN akan digunakan penulis dalam penelitian ini guna memecahkan masalah yang telah dibahas di atas, dengan harapan akan memberikan pemahaman yang lebih baik tentang peminatan kelompok keahlian informatika.

1.2 Rumusan Masalah

Latar belakang di atas menjadi pertimbangan untuk merumuskan masalah sebagai berikut: “Bagaimana mengklasifikasikan kelompok keilmuan berdasarkan nilai menggunakan metode *K-Nearest Neighbors* dan berapa tingkat akurasi yang diperoleh?”.

1.3 Tujuan Penelitian

Tujuan penelitian ini untuk mengklasifikasikan kelompok keilmuan informatika dan mengetahui tingkat akurasi yang diperoleh menggunakan metode *K-Nearest Neighbor* (KNN).

1.4 Batasan Masalah

Pentingnya menetapkan batasan masalah dalam sebuah penelitian adalah untuk memastikan struktur yang lebih teratur dan memudahkan pembahasan, sehingga pencapaian tujuan penelitian menjadi lebih mudah. Berikut ini adalah batasan masalah untuk penelitian ini:

1. Penelitian ini dilaksanakan di Program Studi Teknik Informatika UIN Malang.
2. Terdapat 6 kelompok keilmuan, yaitu *Multimedia*, *Information System*, *Software Engineering*, *Intelligent System*, *System & Network*, dan *Web & Mobile Programming*.
3. Data penelitian ini menggunakan data dari akademik berupa nilai mata kuliah wajib dan indeks prestasi kumulatif.

1.5 Manfaat Penelitian

Beberapa manfaat yang diperoleh dari penelitian ini antara lain:

1. Memudahkan mahasiswa dalam memahami kelompok keilmuan informatika dengan lebih baik.
2. Memberikan pemahaman yang lebih baik terhadap keberagaman keahlian mahasiswa, memastikan bahwa mereka ditempatkan pada kelompok keahlian yang sesuai dengan potensi dan minat mereka.
3. Membantu mahasiswa memilih peminatan yang sesuai dengan kemampuan dan minat mereka, mengurangi perubahan pemikiran yang disebabkan oleh faktor-faktor tertentu.

1.6 Sistematika Penulisan

Dalam mencatat tugas akhir, biasanya dilakukan berdasarkan urutan tertentu yang mencakup beberapa bagian seperti Bab Pendahuluan, kajian pustaka, metode penelitian, hasil dan pembahasan, serta penutup. Berikut adalah penjelasan rinci terkait pencatatan tugas akhir:

BAB I: PENDAHULUAN

Mencakup latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II: KAJIAN PUSTAKA

Studi pustaka yang di dalamnya dibahas mengenai teori-teori serta jurnal acuan yang akan dipakai pada penelitian.

BAB III: METODOLOGI PENELITIAN

Meliputi desain penelitian dan implementasi teori-teori yang telah diuraikan pada Bab 2.

BAB IV: HASIL DAN PEMBAHASAN

Pada bagian hasil dan pembahasan, akan dijelaskan hasil implementasi sistem berdasarkan rancangan yang telah disusun, dengan tujuan membuktikan kesesuaian hasil penelitian dengan tujuan yang telah diuraikan pada Bab 1.

BAB V: KESIMPULAN DAN SARAN

Bab terakhir ini mencakup kesimpulan dari penelitian serta rekomendasi untuk pengembangan penelitian di masa mendatang.

BAB II

KAJIAN PUSTAKA

2.1 Penelitian Terdahulu

Tabel 2.1 menunjukkan daftar penelitian-penelitian yang berkaitan dengan metode dan penelitian yang sedang dilakukan penulis.

Tabel 2.1 Perbandingan penelitian

No	Penulis / Lembaga / Tahun	Metode	Pembahasan	Hasil
1	Farhan Hidayat Zulfallah / UIN Syarif Hidayatullah Jakarta / 2022	KNN	Penelitian ini membahas mengenai prediksi kelulusan mahasiswa UIN Syarif Hidayatullah Jakarta menggunakan algoritma KNN.	<i>Confusion Matrix</i> yang menunjukkan hasil yang baik sehingga penelitiannya dapat digunakan sebagai tolak ukur ketepatan kelulusan mahasiswa UIN Syarif Hidayatullah Jakarta.
2	Clariza Adelina Rachma / UIN Maulana Malik Ibrahim Malang / 2022	KNN	Penelitian ini membahas tentang kedalaman kemiskinan di Jawa Timur tahun 2020 menggunakan algoritma KNN yang dapat membantu program pemberantasan kemiskinan pemerintah.	Pengklasifikasian tingkat kemiskinan menggunakan algoritma KNN menunjukkan akurasi tertinggi sebesar 76,67% dengan nilai parameter k terbaik k = 1 dan k = 2.
3.	Umaidah dan Purwanto / Universitas Singaperbangsa Karawang / 2019	KNN	Penelitian ini mengenai penilaian kinerja siswa dan membantu sekolah membuat keputusan lebih baik menggunakan algoritma KNN	Pemilihan nilai k-fold cross-validation digunakan untuk menghitung klaster. Algoritma KNN pada klaster 2 mencapai akurasi 93,63%. Nilai k-optimal ditentukan

No	Penulis / Lembaga / Tahun	Metode	Pembahasan	Hasil
			dengan nilai K-optimal.	dari validasi silang k-fold.
4.	Yulianto, dkk. / Universitas Nasional / 2020	KNN dan <i>Decision Tree C4.4</i>	Penelitian ini membahas mengenai penggunaan algoritma KNN dan Decision Tree C4.5 dalam menganalisis data untuk meningkatkan kinerja mahasiswa.	Pengujian algoritma KNN menghasilkan akurasi 59,32%, lebih baik daripada decision tree C4.5 yang mencapai 54,80%. Oleh karena itu, algoritma KNN digunakan untuk memprediksi kelulusan mahasiswa.
5.	Syarah Nur Latifah / Universitas Telkom / 2019	ANN	Penelitian ini membahas mengenai penggunaan metode klasifikasi ANN karena kemampuannya yang mampu menangani data yang tidak terstruktur.	Penggunaan ANN mempunyai akurasi sekitar 94,81% dalam memprediksi bidang peminatan siswa. Sehingga ANN dapat digunakan untuk memprediksi bidang peminatan mahasiswa.
6.	Aulia Wahyu Hanifah / UIN Maulana Malik Ibrahim Malang / 2023	<i>Collaborative Filtering</i>	Penelitian ini membahas mengenai adanya sistem rekomendasi susu formula dapat membantu orang tua dalam menentukan pilihan yang tepat.	Metode collaborative filtering dengan 100 responden melalui kuesioner menghasilkan sistem rekomendasi yang akurat melalui sampel yang representatif.
7.	Candra Hafidz Ardana / UIN Maulana Malik Ibrahim Malang / 2024	KNN	Penelitian ini membahas tentang klasifikasi kelompok keilmuan informatika UIN Malang yang dapat membantu mahasiswa dalam memilih kelompok keilmuan yang sesuai dengan preferensi dan minat mereka.	Penggunaan metode KNN memiliki akurasi sekitar 60% dengan nilai K=10 sehingga dapat membantu mahasiswa untuk memilih kelompok keilmuan informatika.

Penelitian terkait yang dijadikan referensi penulis berasal dari sumber-sumber yang relevan dengan penelitian yang sedang dilakukan.

Penelitian yang dilakukan oleh (Zulfallah, 2022) dengan judul “Implementasi Algoritma KNN Dalam Mengukur Ketepatan Kelulusan Mahasiswa UIN Syarif Hidayatullah Jakarta”. Hasil Penelitiannya adalah *Confusion Matrix* yang menunjukkan hasil yang baik sehingga penelitiannya dapat digunakan untuk keakuratan dalam memprediksi kelulusan mahasiswa UIN Syarif Hidayatullah Jakarta. Di mana pengukuran tersebut merupakan prediksi kelulusan menggunakan hasil algoritma KNN.

Penelitian yang dilakukan oleh (Rachma, 2022) dengan judul “Implementasi Algoritma *K-Nearest Neighbor* dalam Penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur”. Hasil Penelitiannya adalah digunakannya algoritma KNN untuk mengklasifikasi tingkat kemiskinan yang menunjukkan tingkat akurasi tertinggi sebesar 76,67% dengan nilai parameter k terbaik $k = 1$ dan $k = 2$.

Penelitian yang dilakukan oleh (Umaidah & Purwantoro, 2019) dengan judul “Penerapan Algoritma *K-Nearest Neighbor (K-nn)* Dengan Pencarian Optimal Untuk Prediksi Prestasi Siswa”. Hasil penelitiannya adalah Pemilihan nilai cross-validation k-fold yang hasilnya digunakan untuk menghitung nilai kluster. Algoritma KNN yang diterapkan pada kluster 2 berhasil memberikan prediksi yang sangat baik dengan tingkat akurasi mencapai 93,63%. Penentuan nilai k-optimal didasarkan pada hasil dari validasi silang k-fold untuk mengevaluasi performa algoritma yang digunakan.

Penelitian yang dilakukan oleh (Yulianto dkk., 2020) berjudul “*Implementation Educational Data mining For Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data mining Method and Decision tree C4.5*”. Hasil penelitiannya adalah pengujian nilai akurasi pada model algoritma KNN menghasilkan tingkat akurasi sebesar 59,32% jauh lebih baik daripada menggunakan model algoritma decision tree C4.5 yang tingkat akurasi dihasilkannya sebesar 54,80%. Sehingga untuk memprediksi kelulusan studi mahasiswa yang digunakan adalah algoritma KNN.

Penelitian yang dilakukan oleh (Latifah, 2019) dengan judul “Analisis Prediksi Pemilihan Bidang Peminatan Menggunakan Metode Klasifikasi dengan Algoritma *Artificial Neural Network* (Studi Kasus: Program Studi Sistem Informasi Universitas Telkom)”. Hasil penelitiannya adalah penggunaan *Artificial Neural Network* (ANN) mempunyai akurasi sekitar 94,81% dalam memprediksi bidang peminatan siswa. Sehingga ANN dapat digunakan untuk memprediksi bidang peminatan mahasiswa.

Penelitian yang dilakukan oleh (Hanifah, 2023) dengan judul “Sistem Rekomendasi Pemilihan Susu Formula Menggunakan *User-Based Collaborative Filtering*”. Hasil penelitiannya adalah penggunaan metode *collaborative filtering* menggunakan 100 data responden melalui kuesioner dengan memberikan sampel yang lebih representatif telah menunjukkan sistem rekomendasi dengan prediksi yang cukup akurat.

2.2 Program Studi Teknik Informatika UIN Malang

Program Studi Teknik Informatika yang selanjutnya disebut Prodi TI, Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang yang selanjutnya disebut Fakultas Saintek UIN Malang, menyelenggarakan program sarjana (S-1) Teknik Informatika sesuai dengan Keputusan Menteri Pendidikan Nasional No. 05/MPN/HK/1004 tanggal 23 Januari 2004, dan Surat Keputusan Direktur Jenderal Departemen Agama No DJ.II/54/2005 tanggal 28 Maret 2005. Berdasarkan Surat Keputusan Lembaga Akreditasi Mandiri Informatika dan Komputer (LAM INFOKOM) No 040/SK/LAM-INFOKOM/Ak/S/III/2024 pada tanggal 22 Maret 2024, Jurusan Teknik Informatika menyanggah akreditasi unggul.

Tujuan Prodi TI adalah melahirkan individu dan teknologi yang unggul, mampu mengembangkan serta mengaplikasikan Teknologi Informasi dengan memperhatikan nilai-nilai Islam yang terdapat dalam Al-Qur'an dan Al-Hadits, guna menyumbang pada kemajuan Teknologi Informasi. Di sini, prinsip-prinsip Al-Qur'an dan Al-Hadits menjadi pedoman moral bagi semua ilmuwan (*Profile – Teknik Informatika – UIN Maulana Malik Ibrahim Malang*, diakses pada 17-11-2023 pukul 12.44).

Program Studi Teknik Informatika tidak hanya menjadi wadah untuk menimba ilmu secara teoritis di bidang teknologi informasi, namun juga bertujuan untuk menghasilkan sumber daya manusia unggul yang mampu beradaptasi dengan cepat terhadap perkembangan dinamis di bidang tersebut. Lulusan dari jurusan ini

diharapkan tidak hanya memiliki pemahaman mendalam tentang konsep teoritis tetapi juga keterampilan praktis yang dapat diterapkan dalam situasi dunia nyata.

Visi program studi tahun 2030 ini bercita-cita menjadi program studi pendidikan tinggi dan penelitian kelas dunia di bidang Informatika khususnya Intelijen dan Rekayasa Perangkat Lunak yang berlandaskan nilai-nilai Islam. Visi berprestasi ini diharapkan dapat mendorong kemajuan yang signifikan di masyarakat, menciptakan lulusan yang tidak hanya kompeten secara teknis tetapi juga memiliki landasan nilai yang kuat. Sedangkan, misi Program Studi TI meliputi yaitu menyelenggarakan pendidikan tinggi Informatika di bidang Intelijen dan Rekayasa Perangkat Lunak, serta menginternalisasi nilai-nilai ulul albab. Penelitian yang disebarkan bertujuan untuk memperluas cakupan pengetahuan Informatika dalam bidang Intelijen dan Rekayasa Perangkat Lunak atau yang berperan dalam memecahkan masalah global dengan berlandaskan nilai-nilai Islam. Selain itu, dilakukan juga transfer teknologi melalui pengembangan dan pendistribusian hasil penelitian untuk mendorong kemajuan masyarakat (*PDDikti - Pangkalan Data Pendidikan Tinggi* **diakses** pada 05-06-2024 pukul 13.08).

Misi Program Studi TI meliputi menyelenggarakan pendidikan tinggi bidang Intelijen dan *Software Engineering* dengan internalisasi nilai-nilai ulul albab, pengembangan penelitian yang memperluas ilmu pengetahuan atau berkontribusi dalam penyelesaian permasalahan global yang didasarkan pada nilai-nilai Islam, dan implementasi transfer teknologi dengan pengembangan dan diseminasi hasil kajian. Semua misi ini bertujuan untuk mencapai kemajuan masyarakat melalui pendidikan, penelitian, dan penerapan teknologi.

Program kompetensi ini meliputi penguasaan pengetahuan dan keterampilan dalam berbagai aplikasi, penggunaan metode dan teknik pemecahan masalah berbasis komputer, serta kemampuan untuk menyerap, mengembangkan, dan memajukan ilmu pengetahuan dan teknologi komputer dan informatika. Dengan demikian, di pasar global diharapkan lulusan memiliki kemampuan untuk bersaing memberikan peran positif terhadap perkembangan teknologi dan berperan penting dalam kemajuan masyarakat secara keseluruhan (*PDDikti - Pangkalan Data Pendidikan Tinggi*, diakses pada 17-11-2023 pukul 01.20).

Berdasarkan Surat Keputusan Dekan Fakultas Saintek UIN Malang No 309/FST/HK.00.5/01/2023 Pada tanggal 27 Januari 2023, memiliki 6 bidang kelompok keilmuan, yaitu *Multimedia, Information System, Software Engineering, Intelligent System, System & Network, Web & Mobile Programming*.

Pemilihan kelompok keilmuan adalah proses di mana mahasiswa pada semester tertentu memilih program kelompok keahlian yang diinginkan agar fokus pada bidang tertentu berdasarkan minat dan bakat mereka. Proses ini terkait erat dengan pembimbingan, yang bertujuan untuk mengarahkan mahasiswa agar dapat memilih program keahlian atau konsentrasi yang tepat.

2.3 Data Mining

Data mining adalah serangkaian tindakan yang digunakan untuk mengekstraksi nilai tambahan secara manual dari informasi yang sebelumnya tidak diketahui dari kumpulan data (Mardi, 2017). Proses data mining dilakukan pada jumlah data yang besar yang tersimpan dalam basis data untuk menemukan pola,

hubungan, dan tren baru dengan menggunakan teknik pengenalan pola seperti matematika dan statistik.

Data mining merupakan salah satu jenis ilmu yang berfokus pada pencarian, penggalian, atau penambangan pengetahuan dari data atau informasi yang dimiliki. *Data mining* juga dikenal sebagai Penemuan Pengetahuan dalam Basis Data. Proses pengumpulan, penggunaan, dan analisis data untuk menemukan hubungan, pola, atau keteraturan dalam kumpulan data yang sangat besar dikenal sebagai KDD (*Knowledge Discovery in Databases*).

Data mining didefinisikan sebagai proses berulang yang menentukan kemajuan melalui penemuan yang dilakukan secara manual atau otomatis (Kantardzic, M, 2019). *Data mining* paling efektif dalam skenario analisis eksplorasi yang hasilnya tidak jelas. *Data mining* merupakan proses mengekstraksi informasi yang baru, berharga, dan signifikan dari sekumpulan data yang besar. Proses ini melibatkan kolaborasi antara manusia dan komputer. Keberhasilan dalam proses ini dicapai melalui penyeimbangan antara pengetahuan manusia yang ahli dalam masalah dan tujuan tertentu dengan kemampuan pencarian komputer.

Prediksi dan deskripsi biasanya merupakan dua tujuan utama *data mining*. Prediksi melibatkan penggunaan bidang atau variabel tertentu dalam kumpulan data untuk memprediksi nilai masa depan dari variabel yang diminati atau nilai yang tidak diketahui. Meskipun data mining sering dianggap sebagai bagian dari proses penemuan pengetahuan, sebelumnya, perspektif mengindikasikan pentingnya langkah ini karena mengungkapkan pola tersembunyi untuk evaluasi. Namun, istilah data mining sering dipakai dalam konteks yang lebih luas untuk merujuk

pada proses keseluruhan penemuan pengetahuan dalam berbagai lingkungan industri, media, dan penelitian, mungkin karena istilahnya yang lebih singkat. Oleh karena itu, data mining juga sering didefinisikan sebagai proses menggali pola dan pengetahuan menarik dari kumpulan data yang besar. Sumber data bervariasi, mulai dari data yang dialirkan ke sistem, database, gudang data, web, dan tempat penyimpanan lainnya (Han dkk., 2012).

2.3.1 Pengelompokkan *Data Mining*

1. Deskripsi

Peneliti dan analis sering mencari cara sederhana untuk mengidentifikasi pola dan tren dalam data. Menjelaskan pola dan tren ini sering kali dapat memberikan wawasan atau kemungkinan penjelasan mengenai fenomena yang diamati.

2. Estimasi

Estimasi memiliki kesamaan dengan klasifikasi, tetapi variabel target dalam estimasi adalah numerik, bukan kategorikal. Model dikembangkan dengan menggunakan data lengkap yang mencakup nilai variabel target sebagai prediktor. Kemudian pada tahap selanjutnya, nilai variabel prediksi diestimasi sesuai dengan nilai variabel target tersebut. Misalnya algoritma yang digunakan untuk menerapkan estimasi adalah algoritma *Neural Network*..

3. Prediksi

Prediksi memiliki konsep yang hampir sama dengan klasifikasi dan estimasi, dengan perbedaan bahwa variabel prediksi dalam prediksi memiliki nilai hasil yang lebih kontinu daripada variabel kategori dalam klasifikasi. Dalam konteks prediksi, sejumlah pendekatan yang sama yang digunakan untuk klasifikasi dan estimasi

dapat diterapkan sesuai kebutuhan. Sebagai contoh, memperkirakan harga saham beberapa bulan ke depan.

Jika sesuai, metode dan teknik yang dipakai untuk klasifikasi dan estimasi juga dapat diterapkan untuk prediksi yang mencakup metode statistik tradisional seperti estimasi titik dan estimasi interval kepercayaan, regresi linear sederhana dan korelasi, dan regresi berganda. Contoh algoritma yang mengimplementasikan prediksi adalah algoritma *K-Nearest Neighbor*, *Decision Tree*, dan lain-lain.

4. Klasifikasi

Klasifikasi melibatkan penggunaan kategori untuk variabel target. Misalnya, pengelompokan pendapatan dapat dibagi menjadi tiga kategori: tinggi, sedang, dan rendah. Model *data mining* mengevaluasi sejumlah besar catatan, yang masing-masing berisi informasi tentang variabel prediktor.

5. *Clustering*

Clustering adalah proses pengelompokan catatan, observasi, atau entitas yang sebanding untuk membuat kelas objek dengan karakteristik yang sama. Kumpulan rekaman yang sangat mirip dan berbeda satu sama lain disebut cluster. *Clustering*, tidak seperti klasifikasi, tidak melibatkan estimasi, klasifikasi, atau prediksi nilai variabel target. Algoritma pengelompokan, pada sisi lain, berusaha untuk membagi data secara keseluruhan menjadi kelompok-kelompok yang sejenis. Kesamaan antara entri data dalam satu kelompok diusahakan untuk maksimum, sementara kesamaan antara entri data dalam kelompok lain diusahakan untuk minimal. *Clustering* sering menjadi langkah awal dalam proses data mining, di mana kelompok-kelompok yang terbentuk digunakan sebagai rekomendasi untuk

pendekatan lebih dalam di masa mendatang. Contoh algoritma untuk mengimplementasi *clustering* adalah algoritma *K-Means*.

6. Asosiasi

Dalam *data mining*, tugas asosiasi adalah menemukan fitur-fitur yang sering muncul bersamaan pada waktu yang sama (Jollyta dkk., 2020). Asosiasi mencoba untuk menemukan kriteria yang akan digunakan untuk mengukur hubungan antara dua atau lebih kualitas. Contoh algoritma untuk mengimplementasikan asosiasi adalah algoritma GRI dan algoritma apriori.

2.3.2 Pemrosesan *Data Mining*

Data mining dapat digambarkan sebagai serangkaian proses yang terdiri dari beberapa tahapan yang melibatkan interaksi dengan pengguna secara langsung atau melalui sumber daya pengetahuan (Marisa, 2013). Tahap-tahap *data mining* mencakup :

1. Pembersihan data (*data cleaning*)

Proses pembersihan data mengeliminasi gangguan serta penghapusan data yang tidak konsisten atau tidak relevan. Langkah-langkah seperti menghapus entri duplikat dan memperbaiki kesalahan data, termasuk kesalahan pengetikan, merupakan langkah lain dalam proses ini.

2. Integrasi data

Data yang dipakai untuk *data mining* dapat berasal dari beberapa database atau file teks, atau hanya dari satu database. Nama, jenis produk, nomor pelanggan,

dan atribut lain yang digunakan untuk membedakan entitas unik diintegrasikan dengan data.

3. Pemilihan data

Kita hanya akan mengambil data yang relevan untuk analisis dari database, sehingga tidak semua data dalam *database* akan digunakan. Sebelum memulai tahap penambangan informasi dalam KDD, harus dilakukan tahap pemilihan data dari operasional pengumpulan data.

4. Transformasi data

Data disesuaikan agar dapat diproses dalam pengolahan data. Proses ini mencakup konversi atau penyatuan data ke dalam format yang sesuai untuk analisis.

5. Proses *mining*

Adalah fase utama dari pendekatan yang dipakai untuk menemukan pengetahuan penting yang tersembunyi dalam data. Dengan menyaring sejumlah besar data yang disimpan dalam penyimpanan, proses ini memungkinkan penemuan hubungan, pola, dan tren baru yang signifikan..

6. Evaluasi pola

Proses ini bertujuan untuk mengenali pola-pola menarik dan relevan yang telah ditemukan sehingga dapat dimasukkan ke dalam basis pengetahuan.

7. Presentasi pengetahuan

Proses ini bertujuan untuk mengenali pola-pola menarik dan relevan yang telah ditemukan. Pada proses ini juga melihat dan menyampaikan informasi tentang bagaimana pengguna mengumpulkan informasi.

2.4 Klasifikasi

Klasifikasi adalah proses pengelompokan data berdasarkan target atau label kelasnya (Wahyono dkk., 2020). Proses klasifikasi ini digunakan untuk membuat keputusan berdasarkan analisis data historis dengan menggunakan algoritma perhitungan. Saat ini, klasifikasi digunakan secara luas dalam berbagai bidang seperti kesehatan, pendidikan, komputer, dan berbagai bidang lainnya.

Algoritma yang digunakan dalam klasifikasi ini termasuk dalam kategori pembelajaran terawasi, di mana data target atau label berfungsi sebagai pengawas yang membimbing proses pembelajaran untuk mencapai tingkat akurasi tertentu. Beberapa algoritma yang umum digunakan untuk menangani masalah klasifikasi meliputi *K-Nearest Neighbor*, *Decision Tree*, dan lainnya.

Klasifikasi adalah suatu metode dalam *data mining* yang menguji karakteristik atribut dari kelompok data yang telah ditetapkan. Metode ini memberikan pemahaman tentang data baru dengan menganalisis data yang telah dikelompokkan, sehingga menghasilkan wawasan atau serangkaian aturan. Aturan ini dapat diterapkan pada data baru untuk mengklasifikasikannya ke dalam kategori atau kelas tertentu.

Menurut Kustanti dkk., (2018), dalam klasifikasi, langkah-langkah dilakukan untuk menciptakan model yang dapat menjelaskan dan membedakan kelas data atau konsep, sehingga dapat digunakan untuk memprediksi kelas objek yang label kelasnya tidak diketahui.

2.5 *K-Nearest Neighbor (KNN)*

Menurut (Boiculese dkk., 2013), algoritma K-Nearest Neighbor (KNN) merupakan suatu teknik klasifikasi yang bergantung pada perbandingan tingkat kesamaan antara data pembelajaran (data latih dan data uji) yang dihitung berdasarkan jarak terdekat (distance) di antara mereka. Prinsip dasar dari KNN adalah mengidentifikasi objek terdekat dari data yang akan dievaluasi dengan menggunakan data latihan yang ada. Tetangga terdekat merujuk pada objek latihan yang memiliki kesamaan nilai terbesar atau perbedaan nilai terkecil dengan data yang sedang diuji. KNN merupakan algoritma *supervised learning* yang dipakai untuk mengklasifikasi berdasarkan data pelatihan. Setiap data pelatihan mewakili satu kelas data, dan saat data uji tidak memiliki label kelas, KNN akan mencari data yang paling mirip dengan data uji tersebut. Metode yang paling sering diterapkan untuk mengukur jarak dalam algoritma KNN salah satunya adalah menggunakan metode Euclidean distance (Aggarwal & Zhai, 2012). Metode *Euclidean distance* biasanya digunakan khususnya untuk data berjenis numerik. Rumus perhitungan *Euclidean distance* adalah sebagai berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_1 - x_2)^2} \quad (2.1)$$

Keterangan:

$d(x,y)$ = jarak kedekatan antara data x ke data y
 x_i = data *testing* (data uji) ke i
 y_i = data *training* (data latih) ke i
 n = jumlah atribut 1 sampai n

Menurut (Permana dkk., 2023) Langkah-langkah menghitung algoritma KNN adalah sebagai berikut :

1. Menentukan parameter k (seleksi nilai k secara manual)
2. Menghitung jarak antara data *training* (data latih) dan data *testing* (data uji)
3. Mengurutkan data *training* berdasarkan jarak yang dihasilkan (dari yang terkecil hingga yang terbesar)
4. Menetapkan kelas yang sesuai dengan memilih kelas yang memiliki jumlah terbanyak dari nilai k pada data *testing* (data uji).

Algoritma KNN merupakan salah satu metode klasifikasi yang dipakai dalam dunia *machine learning* yang memiliki kelebihan dalam beberapa aspek. Namun demikian, seperti halnya setiap algoritma, KNN juga memiliki kekurangan dalam konteks pemodelan dan klasifikasi data. Menurut (Kumar dkk., 2018), kelebihan algoritma KNN antara lain :

1. KNN tidak memerlukan proses pelatihan sebelum melakukan prediksi, yang memungkinkan penambahan data baru dengan kemudahan tanpa mengurangi tingkat akurasi.
2. KNN tergolong sebagai *Learner* yang bersifat *Lazy*, sehingga algoritma ini memiliki kecepatan yang lebih tinggi jika dibandingkan dengan beberapa algoritma lainnya.
3. Implementasi KNN sangat sederhana karena hanya melibatkan dua parameter, yaitu nilai K dan matrik jarak.

Sedangkan, kekurangan algoritma KNN antara lain :

1. Kesulitan mengatasi volume data yang besar terjadi karena kinerja algoritma menurun saat harus menghitung jarak antara titik baru dan yang sudah ada dalam dataset yang luas.

2. Sebelum menerapkan algoritma KNN pada dataset, penting untuk melakukan standarisasi dan normalisasi. Jika tidak, algoritma KNN dapat menghasilkan prediksi yang tidak akurat.

2.6 *Confusion Matrix*

Confusion Matrix adalah dasar penting untuk menilai kinerja algoritma klasifikasi karena dianggap sebagai salah satu metode pengukuran keputusan paling umum dalam pembelajaran mesin yang diawasi. *Confusion Matrix* merupakan alat visual yang berguna untuk mempelajari tingkat kebingungan atau kesalahan model di berbagai kelas. Dalam pembelajaran mesin, *Confusion Matrix* tidak terbatas pada jenis algoritma klasifikasi tertentu. Hal ini memberikan gambaran komprehensif yang tidak tergantung pada algoritma tertentu yang digunakan (Xu dkk., 2020).

Confusion Matrix adalah dasar untuk memahami lebih lanjut ketidakpastian pembelajaran mesin dan alat yang umum digunakan untuk menilai kinerja. Dalam konteks klasifikasi, *Confusion Matrix* merupakan alat evaluasi yang penting, terutama untuk mengevaluasi seberapa baik model klasifikasi dapat memprediksi kelas yang tepat. *Confusion Matrix* memberikan pemahaman yang lebih mendalam tentang performa model dibandingkan metrik evaluasi sederhana seperti akurasi, karena matriks ini berisi informasi yang membandingkan hasil klasifikasi sistem dengan keadaan sebenarnya.

Pada dasarnya, empat kategori utama hasil prediksi dibagi berdasarkan matriks konfusi. *Benar Positif (TP)* adalah ketika model memprediksi kelas positif dengan benar; *Benar Negatif (TN)* adalah ketika model memprediksi kelas negatif dengan benar; *Salah Positif (FP)* adalah ketika model memprediksi kelas positif

padahal seharusnya kelas negatif; dan *Salah Negatif* (FN) adalah ketika model memprediksi kelas negatif padahal seharusnya negatif. Konsep *Confusion Matrix* dapat dilihat pada tabel 2.2 berikut.

Tabel 2. 2 *Confusion Matrix*

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan:

- TP : Jumlah data yang bernilai positif dan diprediksi benar positif
 TN : Jumlah data yang bernilai negatif dan diprediksi benar negatif
 FP : Jumlah data yang bernilai negatif tetapi diprediksi sebagai positif
 FN : Jumlah data yang bernilai positif tetapi diprediksi sebagai negative

Berdasarkan *Confusion Matrix*, akan diperoleh nilai seperti akurasi, presisi, dan recall yang dapat dihitung menggunakan informasi ini. Akurasi mengukur keakuratan keseluruhan prediksi model, sedangkan presisi mengukur keakuratan prediksi positif. Recall menunjukkan kemampuan model untuk menemukan semua kasus positif yang seharusnya.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (2.3)$$

$$Precision = \frac{TP+TN}{TP+FP} \times 100\% \quad (2.4)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.5)$$

Penelitian (Rahman dkk., 2017) dapat menawarkan perspektif yang lebih spesifik atau relevan mengenai penggunaan *Confusion Matrix* dalam konteks tertentu. Namun perlu diingat bahwa pemahaman dan penerapan matriks konfusi

dapat berbeda-beda bergantung pada karakteristik kumpulan data dan jenis analisis yang diperlukan.

2.7 Python

Python merupakan bahasa pemrograman yang biasanya dipakai untuk membuat aplikasi berbasis website karena memiliki filosofi desain yang mengacu pada tingkat keterbacaan kode yang mudan dan bersifat interpretatif dengan banyak fungsi yang dapat dijalankan (Syahrudin & Kurniawan, 2018). Selain itu, Phyton juga mempunyai kemampuan sintaks yang mudah dipahami dengan beragam pustaka komprehensif yang mendukung.

Python awalnya dikembangkan sebagai bahasa skrip, namun kini telah merambah berbagai bidang pengembangan perangkat lunak dan dapat beroperasi di berbagai sistem operasi, termasuk Linux/Unix, Windows, Mac OS X, Palm, Java Virtual Machine, Amiga, dan Symbian (produk Nokia). Distribusi Python saat ini tersedia dalam berbagai lisensi dan versi yang berbeda, dengan prinsipnya dapat diperoleh secara gratis untuk keperluan perdagangan.

Kemampuan Python meliputi penggunaannya di server untuk membangun aplikasi web, integrasi dengan perangkat lunak untuk mengatur alur kerja, koneksi dengan sistem basis data, manipulasi file, pengolahan data berskala besar, serta melakukan perhitungan matematika yang kompleks. Pengembangan prototipe atau perangkat lunak siap produksi juga dapat dilakukan dengan Python. Selain itu, Python memiliki banyak perpustakaan yang tersedia dan dapat digunakan secara fungsional, prosedural, atau berorientasi pada objek.

Meskipun Python versi 2 tidak diperbarui lagi kecuali untuk pembaruan keamanan, Python 3 adalah versi terbaik. Python memiliki banyak keuntungan dibandingkan dengan bahasa pemrograman lainnya. Python memiliki desain yang mudah dibaca dan mirip dengan bahasa Inggris serta matematika. Tidak seperti banyak bahasa lain yang menggunakan titik koma atau tanda kurung untuk mengakhiri perintah, Python menggunakan baris baru. Selain itu, Python juga menggunakan indentasi, yaitu penggunaan spasi, untuk menentukan cakupan seperti dalam loop, fungsi, dan kelas, sedangkan bahasa lain cenderung memakai kurung kurawal (Ma'arif, 2020).

2.8 *Jupyter Notebook*

Jupyter notebook juga disebut sebagai notebook komputasi adalah aplikasi web sumber terbuka (*open source*) yang tersedia secara gratis. Pengguna dapat membuat dan berbagi dokumen yang berisi teks naratif, visualisasi, persamaan, dan kode langsung di platform ini. Untuk berbagai tugas, seperti penerapan pembelajaran mesin, pembersihan dan transformasi data, dan sebagainya, digunakan.

Meskipun notebook komputasi telah ada sejak lama, *Jupyter* menjadi sangat populer akhir-akhir ini. Pertumbuhan pesat platform ini didorong oleh komunitas pengguna dan pengembang yang antusias serta arsitektur yang didesain ulang yang memungkinkannya berinteraksi dengan berbagai bahasa pemrograman. Meskipun *Jupyter notebook* semakin populer, namun tetap saja ada kelebihan dan kekurangannya (Perkel, 2018), beberapa di antaranya adalah :

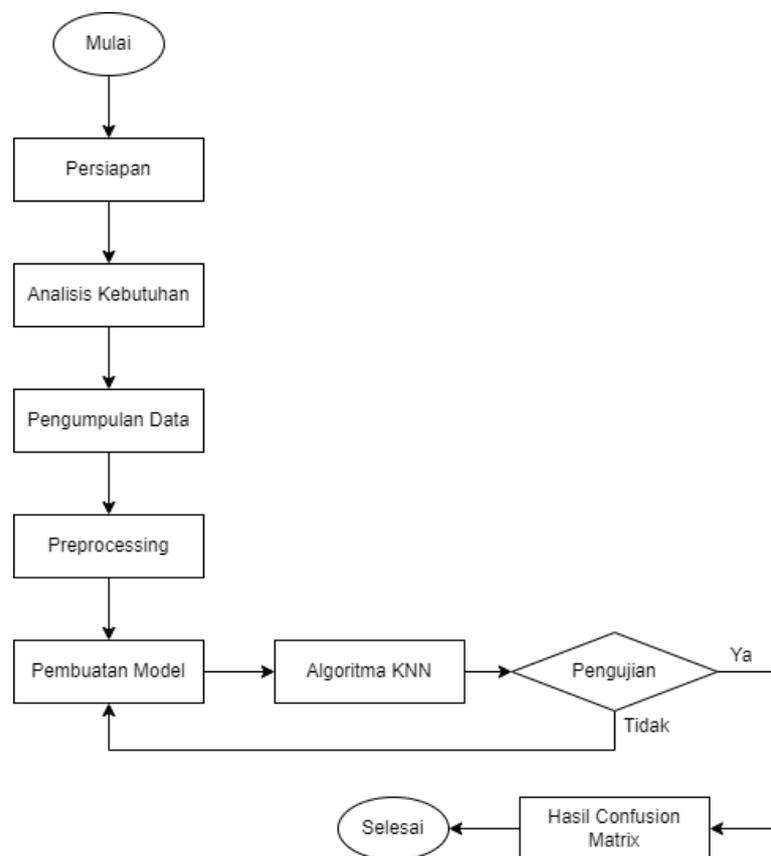
Kelebihan *jupyter notebook* :

1. Merupakan platform terbaik untuk memulai pekerjaan dalam ilmu data atau analisis data.
2. Mudah untuk berbagi gambar dan buku catatan dengan orang lain.
3. Terdapat fungsi tambahan serta fitur markdown, yang memungkinkan penambahan catatan untuk membuat kode lebih teratur dan mudah dipahami pembaca.
4. Sedangkan, kekurangan *jupyter notebook* yaitu beberapa fitur yang biasanya tersedia dalam lingkungan pengembangan terpadu (IDE) masih belum tersedia.

BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini akan dilakukan beberapa tahapan penelitian, berikut merupakan diagram alur penelitian yang akan dilakukan:



Gambar 3. 1. Tahapan Penelitian

Dalam melakukan penelitian ini, terdapat beberapa tahapan, yaitu:

a. Persiapan

Pada tahap persiapan, dilakukan dengan penyusunan proposal penelitian dan mengurus seluruh perizinan dalam penelitian.

b. Analisis Kebutuhan

Dilakukan melalui studi literatur dan observasi yang menghasilkan sumber data, kebutuhan sistem, dan beberapa jenis kelompok keilmuan informatika.

c. Pengumpulan Data

Data diperoleh dari pihak akademik program studi teknik informatika UIN Malang yang akan menghasilkan informasi dari data tersebut.

d. Preprocessing

Dilakukan normalisasi data yang akan menghasilkan pembersihan data dan mengubah ke format yang sesuai.

e. Pembuatan Model Klasifikasi

Pembuatan model menggunakan metode KNN yang akan menghasilkan klasifikasi kelompok keilmuan informatika.

f. Pengujian

Dilakukan dengan metode *Confusion Matrix* yang akan menghasilkan tingkat akurasi dari sistem yang telah dibangun.

3.2 Analisis Kebutuhan Informasi

Pada kebutuhan informasi peneliti membutuhkan beberapa informasi tentang data dari mahasiswa untuk menentukan jenis kelompok keilmuan informatika. Informasi tahapan ini diperoleh dari survey dan studi literatur.

3.3 Metode Pengumpulan Data

Dalam menyusun penelitian ini, dibutuhkan data informasi yang cukup lengkap sebagai dasar untuk mendukung keakuratan bahan pembahasan. Sehingga,

pengumpulan data dilakukan untuk mendapatkan informasi yang diperlukan, dimana data yang dikumpulkan adalah data primer yang didapatkan langsung dari bagian akademik UIN Malang. Salah satu sumber data yang akan dimanfaatkan adalah data akademik mahasiswa Prodi TI di UIN Malang. Data tersebut meliputi berbagai jenis informasi, seperti nilai mata kuliah wajib mahasiswa dan indeks prestasi kumulatif (IPK).

3.4 Teknik Analisis Data

Analisis data menurut (Sugiyono, 2018) adalah proses pencarian dan penyusunan data secara sistematis yang telah diperoleh dari berbagai metode pengumpulan data, seperti studi literatur dan wawancara. Proses ini melibatkan pengorganisasian data ke dalam beberapa kategori agar dapat dipahami dengan mudah oleh peneliti maupun orang lain.

Kriteria yang terdapat pada penelitian kali ini merupakan variabel yang akan diteliti adalah sebagai berikut:

a. Nilai Mata Kuliah Wajib

Nilai yang diambil dari mata kuliah wajib adalah nilai yang telah diperoleh dari mahasiswa aktif pada perguruan tinggi. Nilai mata kuliah wajib ini yang akan digunakan hanya nilai dari matakuliah yang berhubungan dengan prodi TI. Nilai yang terdapat di UIN Malang dapat dilihat pada tabel berikut.

Tabel 3. 1 Nilai Mata Kuliah

Nilai Dalam Huruf	Nilai Dalam Angka
Nilai A	4
Nilai B+	3,5
Nilai B	3
Nilai C+	2,5
Nilai C	2
Nilai D	1
Nilai E	0

(Buku Pedoman Pendidikan UIN Malang, 2022) Berdasarkan Pasal 3 tentang Penetapan Nilai Akhir Mata kuliah, Nilai akhir suatu mata kuliah ditentukan berdasarkan nilai dari komponen penilaian yang telah ditentukan oleh dosen dan disampaikan kepada mahasiswa pada awal perkuliahan. Nilai akhir ditentukan dengan mengonversi nilai berdasarkan pedoman konversi. Pada program sarjana, konversi nilai dinyatakan dengan huruf A, B+, B, C+, C, D, dan E yang masing-masing memiliki rentang skor sebagai berikut: A (81-100), B+ (76-80), B (71-75), C+ (66-70), C (61-65), D (51-60), dan E (<50).

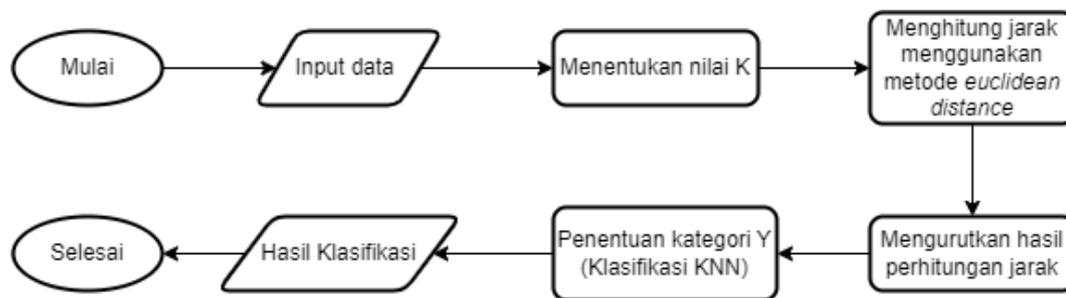
b. Indeks Prestasi Kumulatif

Indeks Prestasi Kumulatif (IPK) merupakan nilai yang mencerminkan prestasi akademis seorang mahasiswa selama masa studi di perguruan tinggi. IPK dihitung berdasarkan rata-rata nilai semua mata kuliah yang diambil oleh mahasiswa dari awal hingga akhir masa perkuliahan. Skala nilai IPK memiliki skala dari 0-4.

3.5 Flowchart Algoritma *K-Nearest Neighbor*

Proses pengidentifikasian bidang kelompok keilmuan informatika dengan menggunakan metode KNN akan melalui beberapa langkah. Pertama, data akan

dimasukkan, kemudian nilai K akan ditentukan (jumlah tetangga terdekat). Selanjutnya, dilakukan normalisasi menggunakan metode min-max untuk setiap atribut. Langkah berikutnya adalah menghitung jarak antara data pelatihan dan data pengujian dengan menggunakan rumus *euclidean*. Kemudian, jarak terdekat akan diurutkan hingga mencapai nilai K , dan kategori Y akan dikumpulkan dari baris tetangga terdekat. Setelah urutan tersebut diperoleh, langkah terakhir adalah menentukan kelompok keilmuan berdasarkan jarak dari tetangga terdekat.



Gambar 3. 2 Flowchart Algoritma K-Nearest Neighbor

3.6 Skenario Pengujian

Setiap model akan dilakukan uji coba dengan menggunakan data latih klasifikasi menggunakan *K-Nearest Neighbor*. Pada data testing akan digunakan untuk pengujian akurasi dan performa dari model yang telah dibangun. Pengujian akan dilakukan beberapa kali agar mendapatkan hasil yang terbaik. Berikut skenario pengujian dapat dilihat pada tabel 3.2.

Tabel 3. 2 Skenario Pengujian Setiap Class

Skenario Pengujian	Persentase Data Latih	Persentase Data Uji	Performa Setiap Class
A	70%	30%	<i>Multimedia</i>
			<i>Information System</i>
			<i>Software Engineering</i>
			<i>Intelligent System</i>
			<i>System & Network</i>
			<i>Web & Mobile Programming</i>
B	75%	25%	<i>Multimedia</i>
			<i>Information System</i>
			<i>Software Engineering</i>
			<i>Intelligent System</i>
			<i>System & Network</i>
			<i>Web & Mobile Programming</i>
C	80%	20%	<i>Multimedia</i>
			<i>Information System</i>
			<i>Software Engineering</i>
			<i>Intelligent System</i>
			<i>System & Network</i>
			<i>Web & Mobile Programming</i>
D	85%	15%	<i>Multimedia</i>
			<i>Information System</i>
			<i>Software Engineering</i>
			<i>Intelligent System</i>
			<i>System & Network</i>
			<i>Web & Mobile Programming</i>

Pengujian pertama adalah melakukan pembagian rasio dari data yang telah dikumpulkan. Pengujian ini dilakukan untuk menemukan rasio pembagian data training dan data testing yang paling optimal. Pembagian yang akan dilakukan sebesar 70:30 sehingga dalam dataset akan dibagi menjadi 2 bagian yaitu 70% untuk data training dan 30% untuk data testing. Pembagian ini juga dilakukan

dengan rasio 75:25, 80:20, dan 85:15 untuk mendapatkan rasio mana yang memiliki tingkat akurasi yang terbaik.

Setelah melakukan pembagian rasio yang terbaik, dilakukan pengujian dataset untuk menemukan nilai K terbaik. Penentuan nilai K yang digunakan sebanyak 10 kali yaitu dari k-1 sampai k-10. Hasil dari penentuan nilai K akan menghasilkan nilai K mana yang memiliki akurasi terbaik.

Setelah semua pengujian selesai, selanjutnya akan dilakukan proses evaluasi dari klasifikasi. Pada proses ini, sistem akan dievaluasi menggunakan teknik *Confusion Matrix* yang nantinya akan menghasilkan akurasi, presisi, recall, dan f-1 score.

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Data Understanding*

Dalam penelitian kali ini, digunakan dataset yang berasal dari Pusat Bagian Akademik UIN Malang. Dataset ini berisi kolom-kolom yang berisi kode nilai matakuliah teknik informatika dan Indeks Prestasi Kumulatif (IPK) dari masing-masing mahasiswa. Berikut merupakan gambaran tabel dari dataset tersebut ditunjukkan pada gambar 4.1.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 95 entries, 0 to 94
Data columns (total 48 columns):
#   Column      Non-Null Count  Dtype
---  -
0   1565001     95 non-null    float64
1   1565002     95 non-null    float64
2   1565003     95 non-null    float64
3   1565004     95 non-null    float64
4   1565005     95 non-null    float64
5   1565006     95 non-null    float64
6   1565007     95 non-null    float64
7   1565008     95 non-null    float64
8   1565009     95 non-null    float64
9   1565010     95 non-null    float64
10  1565011     95 non-null    float64
11  1565012     95 non-null    float64
12  1565013     95 non-null    float64
13  1565014     95 non-null    float64
14  1565015     95 non-null    float64
15  1565016     95 non-null    int64
16  1565017     95 non-null    float64
17  1565018     95 non-null    float64
18  1565019     95 non-null    float64
19  1565020     95 non-null    float64
20  1565021     95 non-null    float64
21  1565022     95 non-null    float64
22  1565023     95 non-null    float64
23  1565024     95 non-null    float64
24  1565025     95 non-null    float64
25  1565026     95 non-null    float64
26  1565027     95 non-null    float64
27  1565028     95 non-null    float64
28  1565029     95 non-null    float64
29  1565030     95 non-null    float64
30  1565031     95 non-null    float64
31  1565032     95 non-null    float64
32  1565033     95 non-null    float64
33  1565034     95 non-null    float64
34  1565035     95 non-null    float64
35  1565036     95 non-null    float64
36  1565037     95 non-null    float64
37  1565038     95 non-null    float64
38  1565039     95 non-null    float64
39  1565040     95 non-null    float64
40  1565041     95 non-null    float64
41  1565042     95 non-null    float64
42  1565043     95 non-null    float64
43  1565044     95 non-null    float64
44  1565045     95 non-null    float64
45  1565046     95 non-null    float64
46  ipk         95 non-null    float64
47  output     95 non-null    int64
dtypes: float64(46), int64(2)
memory usage: 35.8 KB

```

Gambar 4. 1 Daftar Nilai Mata Kuliah dan IPK Mahasiswa Teknik Informatika Tahun 2018 UIN Malang

4.2 *Data Preparation*

Tahapan ini dilakukan untuk memperbaiki dan menyempurnakan terhadap data yang akan dilakukan ke dalam proses *modelling*. Proses ini dilakukan agar

menghasilkan data yang optimal dan peneliti mudah dalam mengolah data. Data tersebut ditunjukkan pada gambar 4.2.

	1565001	1565002	1565003	1565004	1565005	1565006	1565007	1565008	1565009	1565010	...	1565039	1565040	1565041	1565042	1565043	1565044
0	3.5	3.0	2.5	2.0	3.5	3.0	3.5	3.5	4.0	3.5	...	4.0	4.0	3.5	3.5	4.0	4.0
1	4.0	3.5	2.5	2.0	3.0	3.0	3.5	4.0	3.5	3.5	...	3.5	3.5	3.5	3.5	4.0	4.0
2	3.0	3.0	3.0	2.0	2.5	3.0	4.0	4.0	4.0	4.0	...	3.5	3.5	4.0	4.0	4.0	4.0
3	4.0	3.5	2.0	2.5	3.5	3.5	2.0	3.5	3.5	3.0	...	4.0	3.5	4.0	4.0	4.0	4.0
4	3.5	3.0	2.0	2.5	3.0	3.5	2.0	4.0	3.5	3.5	...	4.0	4.0	3.0	4.0	4.0	4.0
...
90	4.0	3.0	4.0	2.5	2.5	3.5	3.5	3.0	3.0	3.0	...	4.0	3.0	3.5	3.5	4.0	4.0
91	3.5	3.0	3.0	2.5	3.0	3.5	2.0	4.0	3.5	2.0	...	3.5	3.5	3.5	3.5	4.0	4.0
92	3.0	2.5	3.5	4.0	2.5	3.0	4.0	3.5	4.0	3.5	...	3.5	3.5	3.5	3.5	4.0	4.0
93	3.5	3.0	2.0	2.0	2.5	3.5	2.0	3.5	4.0	3.0	...	3.5	3.0	3.5	4.0	4.0	4.0
94	4.0	3.5	3.5	4.0	3.5	3.0	3.5	3.5	4.0	4.0	...	4.0	4.0	3.5	4.0	4.0	4.0

Gambar 4. 2 Dataset

4.3 Data Modelling

Proses data mining dilakukan pada tahap ini menggunakan KNN sebagai perhitungannya. Proses ini akan menggunakan sebuah *software* pengolahan data yaitu *Jupyter Notebook*.

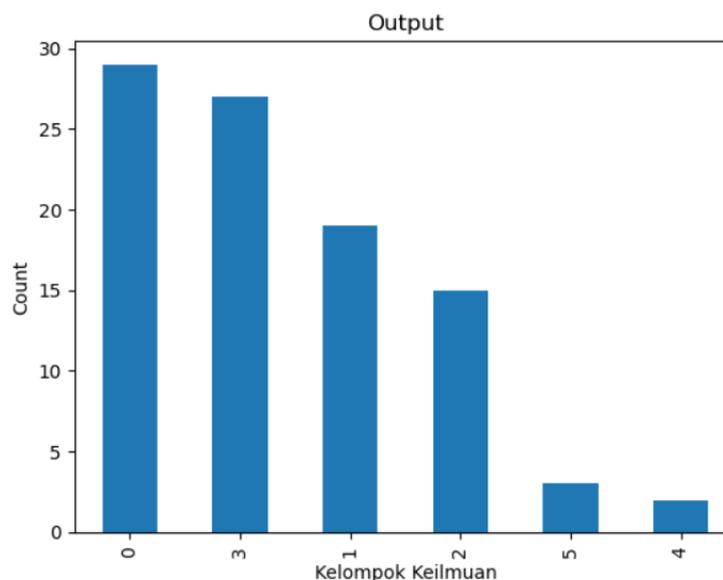
4.3.1 Menentukan Atribut dan Label

Dataset yang telah disiapkan sebagaimana yang tertera pada gambar 4.2 akan dibagi menjadi atribut dan target yang mana kolom kode mata kuliah akan bertindak sebagai atribut dan kolom output akan bertindak sebagai label. Berikut merupakan label dari kelompok keilmuan TI UIN Malang.

Tabel 4. 1 Kelompok Keilmuan TI UIN Malang

Nama Kelompok Keilmuan	Nomor
<i>Multimedia</i>	0
<i>Information System</i>	1
<i>Software Engineering</i>	2
<i>Intelligent System</i>	3
<i>System & Network</i>	4
<i>Web & Mobile Programming</i>	5

Setelah mengetahui nilai dari masing-masing label, akan diperoleh informasi jumlah data yang telah diberi label. Jumlah data tersebut ditunjukkan pada gambar 4.3 berikut.



Gambar 4. 3 Jumlah Nilai pada Setiap Label

4.3.2 Membagi Data *Training* dan Data *Testing*

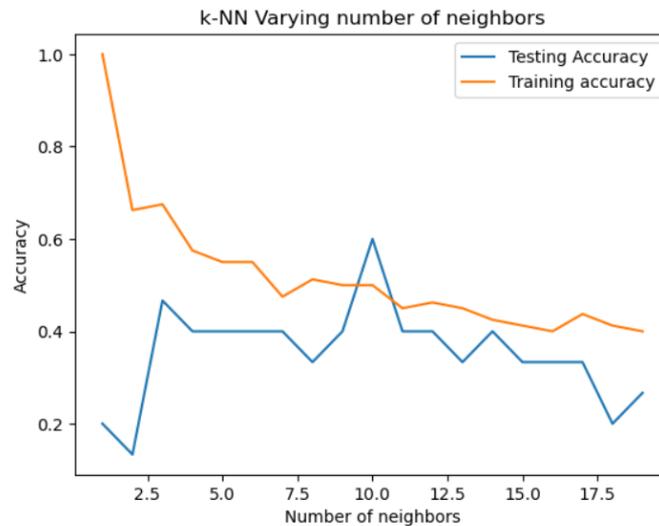
Dataset akan dibagi menjadi data *training* dan data *testing* dimana data *training* dipakai untuk melatih model *machine learning*, yaitu proses untuk mengajarkan model untuk mengenali pola atau hubungan antar target dan fitur dalam data. Data *testing* dipakai untuk menguji kinerja model *machine learning*,

yaitu untuk mengukur seberapa baik klasifikasi atau prediksi dapat dilakukan oleh model pada data yang belum pernah dilihat sebelumnya.

Satu dari beberapa metode yang paling umum dipakai untuk *splitting data* adalah menggunakan fungsi `train_test_split()` dari *library scikit-learn*. *Library scikit-learn* merupakan *library* yang menyediakan berbagai fungsi untuk *machine learning*. Fungsi `train_test_split()` dapat memisahkan antara data *training* dan data *testing* secara yang acak dan proporsional. Peneliti dapat menentukan sebanyak apa persentase data *training* dan data *testing* yang diinginkan. Fungsi `train_test_split()` ini akan mengembalikan empat objek antara lain adalah `x_train`, `y_train`, `x_test` dan `y_test` yang mana merupakan fitur data *training* dan data *testing*.

4.3.3 Implementasi Algoritma KNN

Pembuatan model *K-Nearest Neighbor* menggunakan *library* yang bernama *KNeighborsClassifier* dari *sklearn.neighbors*. Model tersebut kemudian dilatih untuk menguji akurasi dari berbagai nilai tetangga, lalu disimpan hasil akurasi pelatihan dan pengujian dalam array. Hasil dari pengukuran direpresentasikan dengan grafik jarak akurasi pelatihan dan pengujian sebagai fungsi dari nilai tetangga terdekat (k). Pemilihan jumlah tetangga terdekat atau nilai k harus dilakukan secara optimal untuk mencapai keseimbangan antara akurasi pelatihan dan pengujian.



Gambar 4. 4 Plot Akurasi Pelatihan dan Pengujian

Grafik jarak akurasi pelatihan dan pengujian tersebut memberikan panduan dalam menentukan jumlah tetangga terdekat yang optimal dari model KNN. Dari gambar 4.4 tersebut didapatkan hasil nilai K terbaik adalah $K=10$ yang mana memiliki akurasi paling tinggi dari yang lain. Kemudian, dilakukan pelatihan model KNN dengan nilai tetangga terbaik yaitu $K=10$ menggunakan *method fit()* dan menghitung akurasi model pada set pengujian dengan menggunakan *method score()* yang mana hasil akurasinya adalah 60%.

4.4 Evaluation atau Pengujian

Pengujian sistem dilakukan guna mengetahui tingkat akurasi yang didapatkan dari sistem. Hasil yang didapatkan dari sistem kemudian diuji dengan nilai K yang sudah dipilih. Hasilnya adalah sebagai berikut.

4.4.1 Pengujian Terhadap Variasi Data

Pengujian pembagian variasi antara data latih dan data uji dilakukan di tahap ini. Pengujian ini dilakukan secara acak dalam mengambil data latih dan data uji dengan jumlah perbandingan 70:30, 75:25, 80:20, dan 85:15 dari dataset yang tersedia. Pada pengujian ini nilai K yang digunakan adalah K=10. Hasil dari pengujian ditunjukkan pada tabel 4.2

Tabel 4. 2 Pembagian Dataset

No	Jumlah Data Latih (%)	Jumlah Data Uji	Akurasi
1	70	30	31%
2	75	25	25%
3	80	20	37%
4	85	15	60%

Hasil pengujian di atas menunjukkan bahwa data latih dan data uji yang memiliki tingkat akurasi yang tinggi adalah 85:15 yaitu sebesar 60% yang mana 85 mewakili jumlah persentase data latih dan 15 mewakili jumlah persentase data uji. Dari keempat skenario pembagian dataset. Sedangkan akurasi yang terendah adalah 75:25 yaitu sebesar 25%.

4.4.2 Pengujian Mencari Nilai K Terbaik

Pada tahap ini, dilakukan pengujian terhadap berbagai nilai K yang akan digunakan. Variasi data yang digunakan adalah 85:15, karena pada pengujian sebelumnya, variasi ini menunjukkan akurasi tertinggi yaitu 60%. Nilai K yang dipakai adalah dari 2 - 10. Hasil pengujian ditunjukkan pada tabel 4.2 berikut.

Tabel 4. 3 Pengujian Nilai K

No	Nilai K	Akurasi
1	K=2	13%
2	K=3	47%
3	K=4	40%
4	K=5	40%
5	K=6	40%
6	K=7	40%
7	K=8	33%
8	K=9	40%
9	K=10	60%

Hasil pengujian di atas menunjukkan bahwa nilai K yang paling optimal adalah K=10 yaitu sebesar 60%. Sedangkan, untuk nilai K yang paling rendah adalah K=2 yaitu sebesar 13%

4.4.3 *Confusion Matrix*

Model yang nantinya model tersebut akan dievaluasi untuk mengukur kinerja dari algoritma KNN diperoleh dari hasil pengujian yang telah dilakukan sebelumnya. Secara umum, untuk menggambarkan kinerja dari hasil klasifikasi adalah menggunakan *Confusion Matrix*. Matrix ini merupakan tabulasi silang yang sederhana dari kelas yang diuji dan diprediksi, yang mana sel diagonal mewakili kelas yang menunjukkan kelas yang diprediksi benar.

4.4.3.1 Hasil Uji Coba Skenario A

Sebanyak 95 data telah disiapkan dan akan dibagi ke dalam dua kategori, yaitu data latih dan data uji dengan rasio pembagian 70:30. Data latih terdiri dari 66 data, sedangkan data uji sebanyak 29 data. Hasil pengujian yang diperoleh ditunjukkan pada tabel berikut:

Tabel 4. 4 *Confusion Matrix Class Multimedia*

	Predicted True	Predicted False
Actual True	3	8
Actual False	5	13

Tabel 4.4 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 3, *Salah Positif* (FP) sejumlah 5, *Salah Negatif* (FN) sejumlah 8, dan *Benar Negatif* (TN) sejumlah 13.

Tabel 4. 5 *Confusion Matrix Class Information System*

	Predicted True	Predicted False
Actual True	1	5
Actual False	2	21

Tabel 4.5 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 1, *Salah Positif* (FP) sejumlah 2, *Salah Negatif* (FN) sejumlah 2, dan *Benar Negatif* (TN) sejumlah 21.

Tabel 4. 6 *Confusion Matrix Class Software Engineering*

	Predicted True	Predicted False
Actual True	1	3
Actual False	1	24

Tabel 4.6 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 1, *Salah Positif* (FP) sejumlah 1, *Salah Negatif* (FN) sejumlah 3, dan *Benar Negatif* (TN) sejumlah 24.

Tabel 4. 7 *Confusion Matrix Class Intelligent System*

	Predicted True	Predicted False
Actual True	4	2
Actual False	12	11

Tabel 4.7 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 4, *Salah Positif* (FP) sejumlah 12, *Salah Negatif* (FN) sejumlah 2, dan *Benar Negatif* (TN) sejumlah 11.

Tabel 4. 8 *Confusion Matrix Class System & Network*

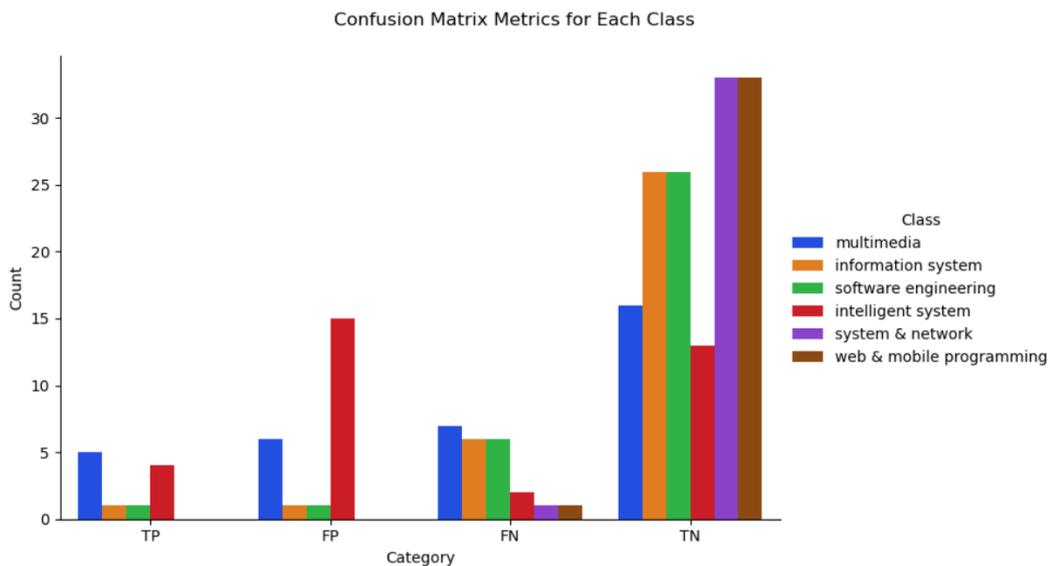
	Predicted True	Predicted False
Actual True	0	1
Actual False	0	28

Tabel 4.8 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 1, dan *Benar Negatif* (TN) sejumlah 28.

Tabel 4. 9 *Confusion Matrix Class Web & Mobile Programming*

	Predicted True	Predicted False
Actual True	0	1
Actual False	0	28

Tabel 4.9 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 1, dan *Benar Negatif* (TN) sejumlah 28.



Gambar 4. 5 *Confusion Matrix* Skenario A

Setelah mengetahui jumlah nilai *Confusion Matrix* dari masing-masing *Class*, akan dihitung performa dari masing-masing *Class*. Hasil dari performa pada skenario A ditunjukkan pada tabel 4.10 berikut.

Tabel 4. 10 Performa Skenario A

Nama Class atau Label	Akurasi	Presisi	Recall	F1-Score
<i>Multimedia</i>	55%	53%	55%	53%
<i>Information System</i>	76%	71%	76%	73%
<i>Software Engineering</i>	86%	84%	86%	84%
<i>Intelligent Systems</i>	52%	72%	52%	56%
<i>System & Network</i>	97%	93%	97%	95%
<i>Web & Mobile Programming</i>	97%	93%	97%	95%

Tabel 4.10 menunjukkan performa sistem dari masing-masing *Class*. *Class Multimedia* memiliki tingkat akurasi sebesar 55%, presisi sebesar 53%, recall sebesar 55%, dan f-1 score sebesar 53%. *Class Information System* memiliki tingkat akurasi sebesar 76%, presisi sebesar 71%, recall sebesar 76%, dan f-1 score sebesar 73%. *Class Software Engineering* memiliki tingkat akurasi sebesar 86%, presisi

sebesar 84%, recall sebesar 86%, dan f-1 score sebesar 84%. *Class Intelligent System* memiliki tingkat akurasi sebesar 52%, presisi sebesar 72%, recall sebesar 52%, dan f-1 score sebesar 56%. *Class System & Network* memiliki tingkat akurasi sebesar 97%, presisi sebesar 93%, recall sebesar 97%, dan f-1 score sebesar 95%. *Class Web & Mobile Programming* memiliki tingkat akurasi sebesar 97%, presisi sebesar 93%, recall sebesar 97%, dan f-1 score sebesar 95%.

4.4.3.2 Hasil Uji Coba Skenario B

Sebanyak data telah disiapkan dan akan dibagi menjadi dua kategori, yaitu data latih dan data uji yang memiliki pembagian rasio sebesar 75:25. Data latih memiliki jumlah 71 data dan data uji memiliki jumlah 24 data. Hasil pengujian yang diperoleh dapat dilihat sebagai berikut.

Tabel 4. 11 *Confusion Matrix Class Multimedia*

	Predicted True	Predicted False
Actual True	2	7
Actual False	5	10

Tabel 4.11 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 2, *Salah Positif* (FP) sejumlah 5, *Salah Negatif* (FN) sejumlah 7, dan *Benar Negatif* (TN) sejumlah 10.

Tabel 4. 12 *Confusion Matrix Class Information System*

	Predicted True	Predicted False
Actual True	1	4
Actual False	2	17

Tabel 4.12 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 1, *Salah Positif* (FP) sejumlah 2, *Salah Negatif* (FN) sejumlah 4, dan *Benar Negatif* (TN) sejumlah 17.

Tabel 4. 13 *Confusion Matrix Class Software Engineering*

	Predicted True	Predicted False
Actual True	1	2
Actual False	2	19

Tabel 4.13 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 1, *Salah Positif* (FP) sejumlah 2, *Salah Negatif* (FN) sejumlah 2, dan *Benar Negatif* (TN) sejumlah 19.

Tabel 4. 14 *Confusion Matrix Class Intelligent System*

	Predicted True	Predicted False
Actual True	2	3
Actual False	9	10

Tabel 4.14 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 2, *Salah Positif* (FP) sejumlah 9, *Salah Negatif* (FN) sejumlah 3, dan *Benar Negatif* (TN) sejumlah 10.

Tabel 4. 15 *Confusion Matrix Class System & Network*

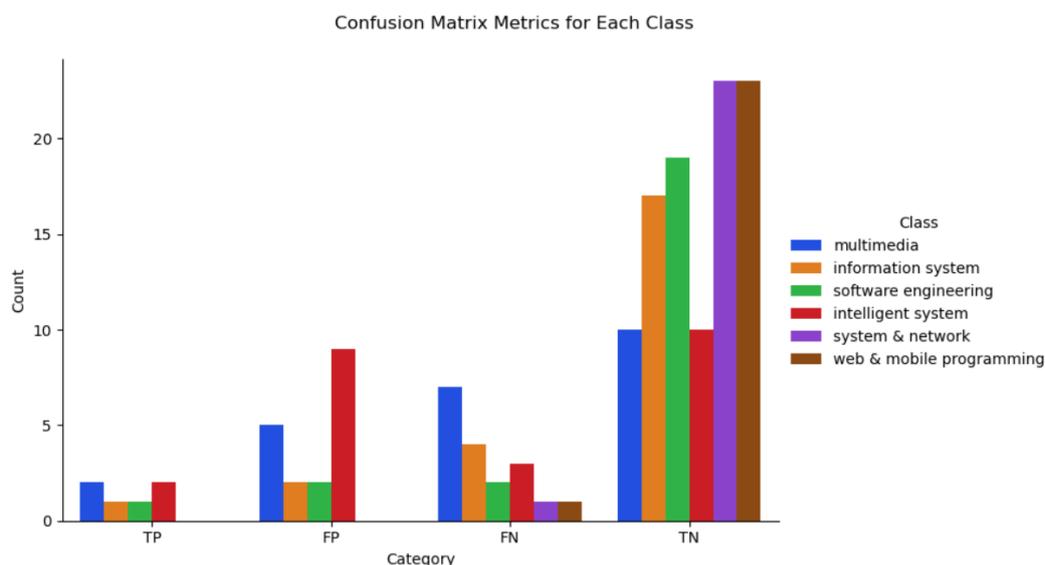
	Predicted True	Predicted False
Actual True	0	1
Actual False	0	23

Tabel 4.15 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 1, dan *Benar Negatif* (TN) sejumlah 23.

Tabel 4. 16 *Confusion Matrix Class Web & Mobile Programming*

	Predicted True	Predicted False
Actual True	0	1
Actual False	0	23

Tabel 4.16 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 1, dan *Benar Negatif* (TN) sejumlah 23.

Gambar 4. 6 *Confusion Matrix Skenario B*

Setelah mengetahui jumlah nilai *Confusion Matrix* dari masing-masing *Class*, akan dihitung performa dari masing-masing *Class*. Hasil dari performa pada skenario B dapat dilihat pada tabel 4.17 berikut.

Tabel 4. 17 Performa Skenario B

Nama Class atau Label	Akurasi	Presisi	Recall	F1-Score
<i>Multimedia</i>	50%	47%	50%	48%
<i>Information System</i>	75%	71%	75%	73%
<i>Software Engineering</i>	83%	83%	83%	83%
<i>Intelligent Systems</i>	50%	65%	50%	55%
<i>System & Network</i>	96%	92%	96%	94%
<i>Web & Mobile Programming</i>	96%	92%	96%	94%

Dilihat dari data yang terdapat pada tabel 4.17 menunjukkan performa sistem dari masing-masing *Class*. *Class Multimedia* memiliki tingkat akurasi sebesar 50%, presisi sebesar 47%, recall sebesar 50%, dan f-1 score sebesar 48%. *Class Information System* memiliki tingkat akurasi sebesar 75%, presisi sebesar 71%, recall sebesar 75%, dan f-1 score sebesar 73%. *Class Software Engineering* memiliki tingkat akurasi sebesar 83%, presisi sebesar 83%, recall sebesar 83%, dan f-1 score sebesar 83%. *Class Intelligent System* memiliki tingkat akurasi sebesar 50%, presisi sebesar 65%, recall sebesar 50%, dan f-1 score sebesar 55%. *Class System & Network* memiliki tingkat akurasi sebesar 96%, presisi sebesar 92%, recall sebesar 96%, dan f-1 score sebesar 94%. *Class Web & Mobile Programming* memiliki tingkat akurasi sebesar 96%, presisi sebesar 92%, recall sebesar 96%, dan f-1 score sebesar 94%.

4.4.3.3 Hasil Uji Coba Skenario C

Sebanyak data telah disiapkan dan akan dibagi menjadi dua kategori, yaitu data latih dan data uji yang memiliki pembagian rasio sebesar 80:20. Data latih memiliki jumlah 76 data dan data uji memiliki jumlah 19 data. Hasil pengujian yang diperoleh dapat dilihat sebagai berikut.

Tabel 4. 18 *Confusion Matrix Class Multimedia*

	Predicted True	Predicted False
Actual True	0	8
Actual False	1	10

Tabel 4.18 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 1, *Salah Negatif* (FN) sejumlah 8, dan *Benar Negatif* (TN) sejumlah 10.

Tabel 4. 19 *Confusion Matrix Class Information System*

	Predicted True	Predicted False
Actual True	2	2
Actual False	1	14

Tabel 4.19 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 2, *Salah Positif* (FP) sejumlah 1, *Salah Negatif* (FN) sejumlah 2, dan *Benar Negatif* (TN) sejumlah 14.

Tabel 4. 20 *Confusion Matrix Class Software Engineering*

	Predicted True	Predicted False
Actual True	2	0
Actual False	0	17

Tabel 4.20 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 2, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 0, dan *Benar Negatif* (TN) sejumlah 17.

Tabel 4. 21 *Confusion Matrix Class Intelligent System*

	Predicted True	Predicted False
Actual True	3	0
Actual False	10	6

Tabel 4.21 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 3, *Salah Positif* (FP) sejumlah 10, *Salah Negatif* (FN) sejumlah 0, dan *Benar Negatif* (TN) sejumlah 6.

Tabel 4. 22 *Confusion Matrix Class System & Network*

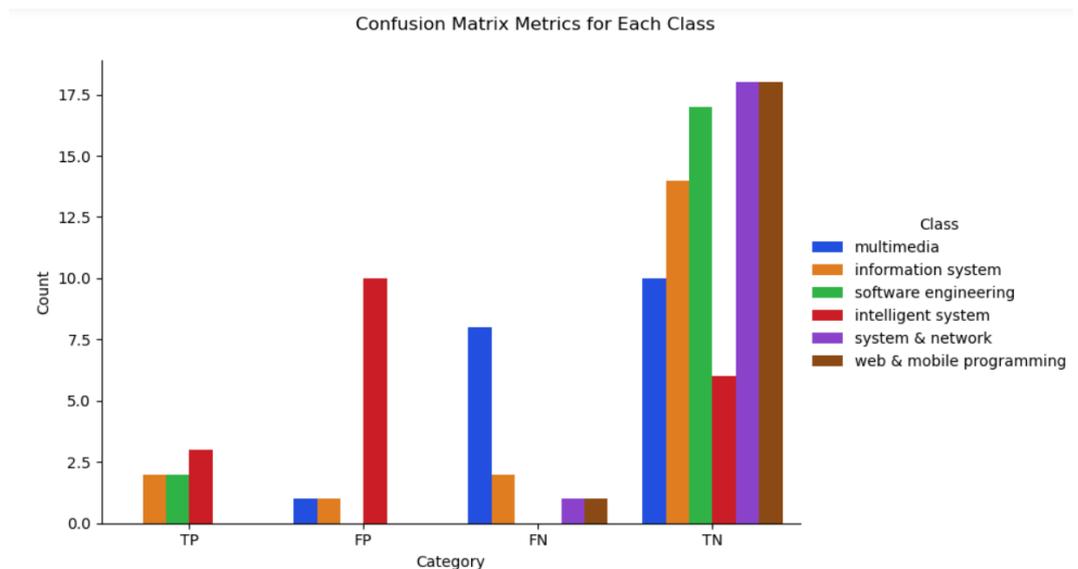
	Predicted True	Predicted False
Actual True	0	1
Actual False	0	18

Tabel 4.22 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 1, dan *Benar Negatif* (TN) sejumlah 18.

Tabel 4. 23 *Confusion Matrix Class Web & Mobile Programming*

	Predicted True	Predicted False
Actual True	0	1
Actual False	0	18

Tabel 4.23 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 1, dan *Benar Negatif* (TN) sejumlah 18.

Gambar 4. 7 *Confusion Matrix Skenario C*

Setelah mengetahui jumlah nilai *Confusion Matrix* dari masing-masing *Class*, akan dihitung performa dari masing-masing *Class*. Hasil dari performa pada skenario C dapat dilihat pada tabel 4.24 berikut.

Tabel 4. 24 Performa Skenario C

Nama Class atau Label	Akurasi	Presisi	Recall	F1-Score
<i>Multimedia</i>	53%	32%	53%	40%
<i>Information System</i>	84%	83%	84%	83%
<i>Software Engineering</i>	100%	100%	100%	100%
<i>Intelligent Systems</i>	47%	88%	47%	52%
<i>System & Network</i>	95%	90%	95%	92%
<i>Web & Mobile Programming</i>	95%	90%	95%	92%

Dilihat dari data yang terdapat pada tabel 4.24 menunjukkan performa sistem dari masing-masing *Class*. *Class Multimedia* memiliki tingkat akurasi sebesar 53%, presisi sebesar 32%, recall sebesar 53%, dan f-1 score sebesar 40%. *Class Information System* memiliki tingkat akurasi sebesar 84%, presisi sebesar 83%, recall sebesar 84%, dan f-1 score sebesar 83%. *Class Software Engineering* memiliki tingkat akurasi sebesar 100%, presisi sebesar 100%, recall sebesar 100%, dan f-1 score sebesar 100%. *Class Intelligent System* memiliki tingkat akurasi sebesar 47%, presisi sebesar 88%, recall sebesar 47%, dan f-1 score sebesar 52%. *Class System & Network* memiliki tingkat akurasi sebesar 95%, presisi sebesar 90%, recall sebesar 95%, dan f-1 score sebesar 92%. *Class Web & Mobile Programming* memiliki tingkat akurasi sebesar 95%, presisi sebesar 90%, recall sebesar 95%, dan f-1 score sebesar 92%.

4.4.3.4 Hasil Uji Coba Skenario D

Sebanyak data telah disiapkan dan akan dibagi menjadi dua kategori, yaitu data latih dan data uji yang memiliki pembagian rasio sebesar 85:15. Data latih memiliki jumlah 80 data dan data uji memiliki jumlah 14 data. Hasil pengujian yang diperoleh dapat dilihat sebagai berikut.

Tabel 4. 25 *Confusion Matrix Class Multimedia*

	Predicted True	Predicted False
Actual True	3	2
Actual False	2	8

Tabel 4.25 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 3, *Salah Positif* (FP) sejumlah 2, *Salah Negatif* (FN) sejumlah 2, dan *Benar Negatif* (TN) sejumlah 8.

Tabel 4. 26 *Confusion Matrix Class Information System*

	Predicted True	Predicted False
Actual True	2	2
Actual False	1	10

Tabel 4.26 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 2, *Salah Positif* (FP) sejumlah 1, *Salah Negatif* (FN) sejumlah 2, dan *Benar Negatif* (TN) sejumlah 10.

Tabel 4. 27 *Confusion Matrix Class Software Engineering*

	Predicted True	Predicted False
Actual True	2	0
Actual False	0	13

Tabel 4.27 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 2, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 0, dan *Benar Negatif* (TN) sejumlah 10.

Tabel 4. 28 *Confusion Matrix Class Intelligent System*

	Predicted True	Predicted False
Actual True	2	0
Actual False	3	10

Tabel 4.28 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 2, *Salah Positif* (FP) sejumlah 3, *Salah Negatif* (FN) sejumlah 0, dan *Benar Negatif* (TN) sejumlah 10.

Tabel 4. 29 *Confusion Matrix Class System & Network*

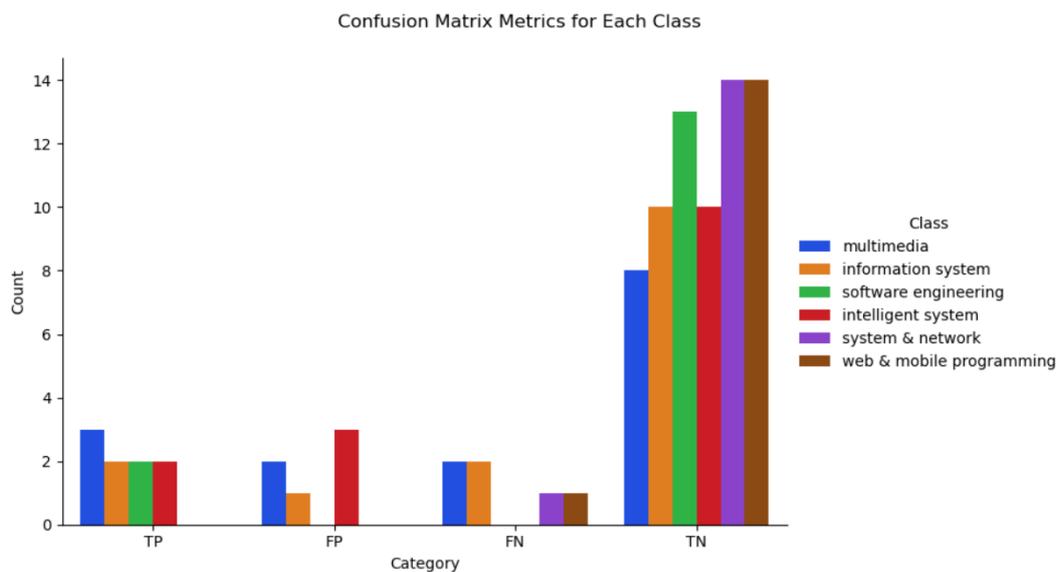
	Predicted True	Predicted False
Actual True	0	1
Actual False	0	14

Tabel 4.29 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 1, dan *Benar Negatif* (TN) sejumlah 14.

Tabel 4. 30 *Confusion Matrix Class Web & Mobile Programming*

	Predicted True	Predicted False
Actual True	0	1
Actual False	0	14

Tabel 4.30 menunjukkan jumlah nilai *Confusion Matrix* dari *Class Multimedia* yaitu *Benar Positif* (TP) sejumlah 0, *Salah Positif* (FP) sejumlah 0, *Salah Negatif* (FN) sejumlah 1, dan *Benar Negatif* (TN) sejumlah 14.



Gambar 4. 8 Confusion Matrix Skenario D

Setelah mengetahui jumlah nilai *Confusion Matrix* dari masing-masing *Class*, akan dihitung performa dari masing-masing *Class*. Hasil dari performa pada skenario D dapat dilihat pada tabel 4.31 berikut.

Tabel 4. 31 Performa Skenario D

Nama Class atau Label	Akurasi	Presisi	Recall	F1-Score
<i>Multimedia</i>	73%	73%	73%	73%
<i>Information System</i>	80%	79%	80%	79%
<i>Software Engineering</i>	100%	100%	100%	100%
<i>Intelligent Systems</i>	80%	92%	80%	83%
<i>System & Network</i>	93%	87%	93%	90%
<i>Web & Mobile Programming</i>	93%	87%	93%	90%

Dilihat dari data yang terdapat pada tabel 4.31 menunjukkan performa sistem dari masing-masing *Class*. *Class Multimedia* memiliki tingkat akurasi sebesar 73%, presisi sebesar 73%, recall sebesar 73%, dan f-1 score sebesar 73%. *Class Information System* memiliki tingkat akurasi sebesar 80%, presisi sebesar 79%, recall sebesar 80%, dan f-1 score sebesar 79%. *Class Software Engineering*

memiliki tingkat akurasi sebesar 100%, presisi sebesar 100%, recall sebesar 100%, dan f-1 score sebesar 100%. *Class Intelligent System* memiliki tingkat akurasi sebesar 80%, presisi sebesar 92%, recall sebesar 80%, dan f-1 score sebesar 83%. *Class System & Network* memiliki tingkat akurasi sebesar 93%, presisi sebesar 87%, recall sebesar 93%, dan f-1 score sebesar 90%. *Class Web & Mobile Programming* memiliki tingkat akurasi sebesar 93%, presisi sebesar 87%, recall sebesar 93%, dan f-1 score sebesar 90%.

4.5 Pembahasan

Peneliti telah melakukan beberapa skenario pengujian dengan rasio pembagian data latih dan data uji yang berbeda untuk menentukan skenario mana yang paling optimal dalam melatih model *K-Nearest Neighbors* (KNN) untuk menentukan kelompok keilmuan informatika UIN Malang.

4.5.1 Pembahasan Pengujian Variasi Data

Pada skenario A adalah pembagian data latih dan data uji sebesar 70:30 yang mana 70 mewakili persentase data latih dan 30 mewakili persentase data uji. Pada skenario pertama memiliki tingkat akurasi sebesar 31% yang mana jumlah indeks dari data latih sebesar 66 data dan jumlah indeks data uji sebesar 29 data. Skenario A juga memiliki nilai presisi sebesar 33%, nilai *recall* sebesar 31%, dan nilai *f-1 score* sebesar 29%.

Pada skenario B adalah pembagian data latih dan data uji sebesar 75:25 yang mana 75 mewakili persentase data latih dan 25 mewakili persentase data uji. Pada skenario pertama memiliki tingkat akurasi sebesar 25% yang mana jumlah indeks

dari data latih sebesar 71 data dan jumlah indeks data uji sebesar 24 data. Skenario B juga memiliki nilai presisi sebesar 26%, nilai *recall* sebesar 25%, dan nilai *f-1 score* sebesar 24%.

Pada skenario C adalah pembagian data latih dan data uji sebesar 80:20 yang mana 80 mewakili persentase data latih dan 20 mewakili persentase data uji. Pada skenario pertama memiliki tingkat akurasi sebesar 37% yang mana jumlah indeks dari data latih sebesar 76 data dan jumlah indeks data uji sebesar 19 data. Skenario C juga memiliki nilai presisi sebesar 28%, nilai *recall* sebesar 37%, dan nilai *f-1 score* sebesar 28%.

Pada skenario D adalah pembagian data latih dan data uji sebesar 85:15 yang mana 85 mewakili persentase data latih dan 15 mewakili persentase data uji. Pada skenario pertama memiliki tingkat akurasi sebesar 60% yang mana jumlah indeks dari data latih sebesar 80 data dan jumlah indeks data uji sebesar 15 data. Skenario D juga memiliki nilai presisi sebesar 56%, nilai *recall* sebesar 60%, dan nilai *f-1 score* sebesar 56%.

Secara keseluruhan dari skenario uji coba yang telah dilakukan oleh peneliti, dapat disimpulkan bahwa model KNN terbaik untuk melakukan klasifikasi kelompok keilmuan informatika UIN Malang pada dataset yang telah ada adalah menggunakan skenario pengujian yang keempat yaitu skenario D dengan pembagian rasio data sebesar 85:15. Skenario D ini menunjukkan performa yang tinggi dan juga memiliki tingkat akurasi, presisi, *recall*, dan *f-1 score* dibandingkan ketiga skenario yang lain. Jadi rasio pembagian data sebesar 85:15 dengan indeks 80 data sebagai data latih dan 15 data sebagai data uji dengan nilai $K=10$ adalah

model yang paling baik dari yang lain, karena menghasilkan performa yang tinggi pada data uji dengan nilai tingkat akurasi sebesar 60%.

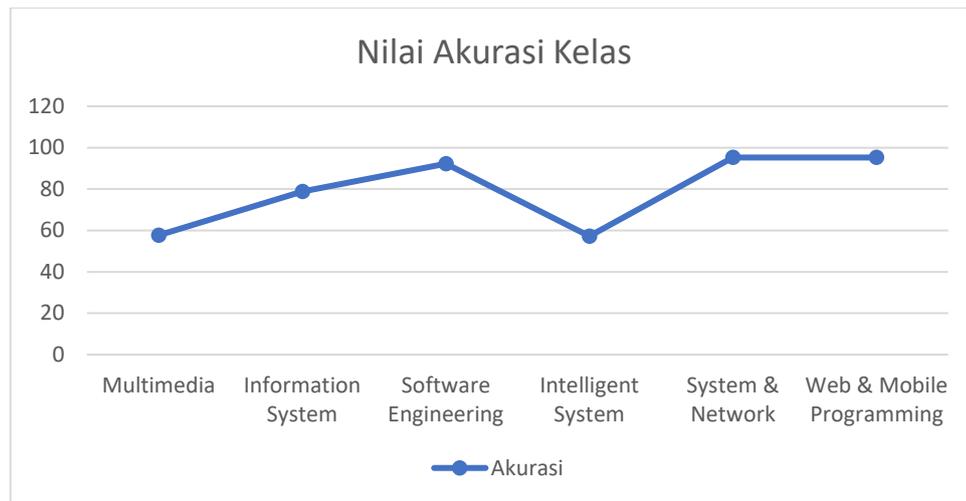
4.5.2 Pembahasan Nilai K Terbaik

Setelah melakukan pengujian terhadap pembagian rasio data, selanjutnya adalah dengan melakukan pencarian untuk menemukan nilai K terbaik menggunakan metode *K-Fold Cross Validation*. Persentase data yang digunakan adalah 85:15 karena memiliki tingkat akurasi yang paling besar dibandingkan yang lainnya. Nilai K yang diuji adalah 2 sampai 10. Dari hasil yang diperoleh dari metode *K-Fold Cross Validation* adalah nilai K=2 memiliki nilai terendah yaitu sebesar 13% dan nilai K=10 memiliki nilai tertinggi yaitu 60%

4.5.3 Pembahasan Pengujian Kelas

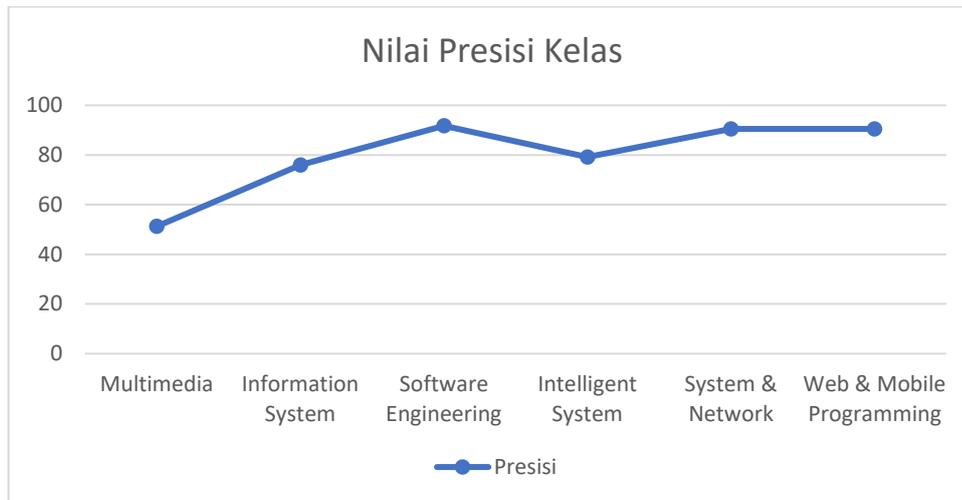
Setelah melakukan pengujian terhadap pembagian rasio data dan pencarian nilai K terbaik, selanjutnya akan dilakukan pengujian dari masing-masing *Class*. *Class Multimedia* memiliki rata-rata tingkat akurasi sebesar 57.75%, presisi sebesar 51.25%, recall sebesar 57.75%, dan f-1 score sebesar 53.5%. *Class Information System* memiliki rata-rata tingkat akurasi sebesar 78.75%, presisi sebesar 76%, recall sebesar 78.75%, dan f-1 score sebesar 77%. *Class Software Engineering* memiliki rata-rata tingkat akurasi sebesar 92.25%, presisi sebesar 91.75%, recall sebesar 92.25%, dan f-1 score sebesar 91.75%. *Class Intelligent System* memiliki rata-rata tingkat akurasi sebesar 57.25%, presisi sebesar 79.25%, recall sebesar 57.25%, dan f-1 score sebesar 61.5%. *Class System & Network* memiliki rata-rata tingkat akurasi sebesar 95.25%, presisi sebesar 90.5%, recall sebesar 95.25%, dan

f-1 score sebesar 92.75%. *Class Web & Mobile Programming* memiliki rata-rata tingkat akurasi sebesar 95.25%, presisi sebesar 90.5%, recall sebesar 95.25%, dan f-1 score sebesar 92.75%.



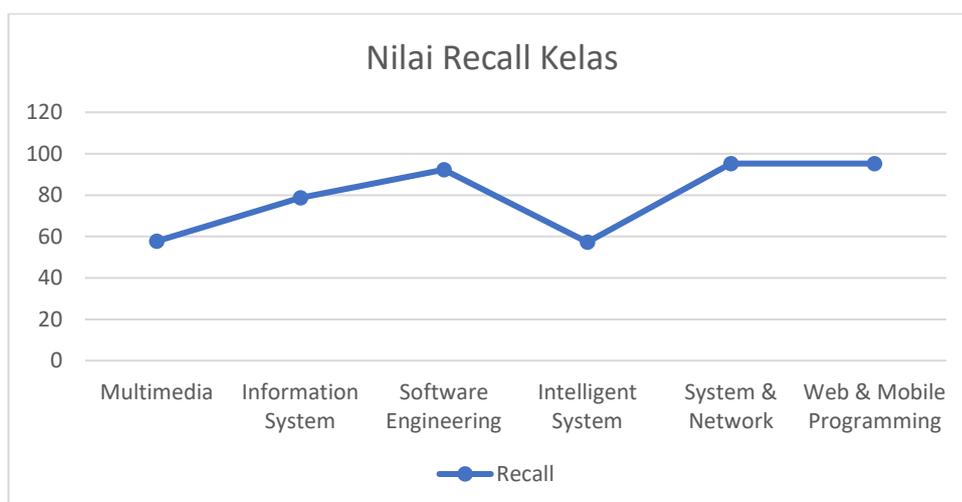
Gambar 4. 9 Grafik Nilai Akurasi Setiap Class

Berdasarkan hasil uji coba yang telah dilakukan, *Class Multimedia* memiliki rata-rata nilai tingkat akurasi sebesar 57.75%, *Class Information System* memiliki rata-rata nilai tingkat akurasi 78.75%, *Class Software Engineering* memiliki rata-rata nilai 92.25%, *Class Intelligent System* memiliki rata-rata nilai 57.25%, *Class System & Network* memiliki rata-rata nilai 95.25%, *Class Web & Mobile Programming* memiliki rata-rata nilai 95.25%.



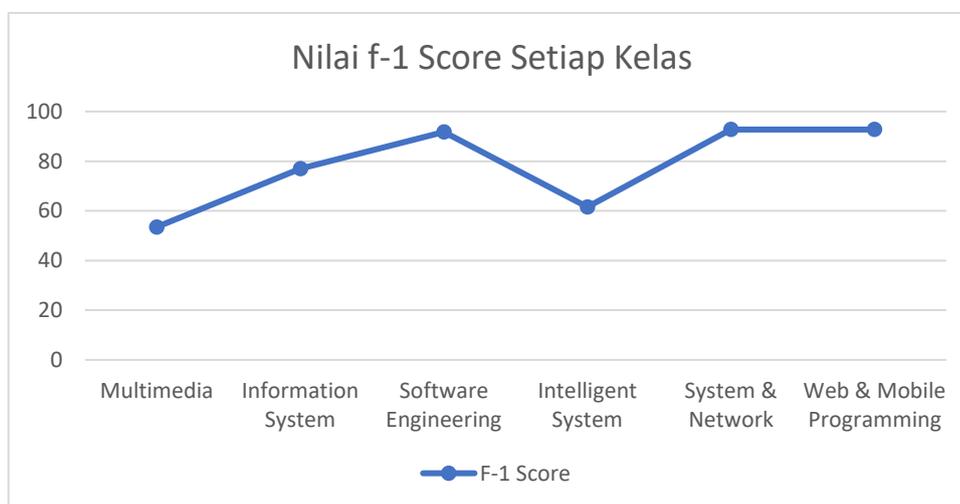
Gambar 4. 10 Grafik Presisi Setiap Class

Berdasarkan hasil uji coba yang telah dilakukan, *Class Multimedia* memiliki rata-rata nilai presisi sebesar 51.25%, *Class Information System* memiliki rata-rata nilai presisi 76%, *Class Software Engineering* memiliki rata-rata nilai presisi 91.75%, *Class Intelligent System* memiliki rata-rata nilai presisi 79.25%, *Class System & Network* memiliki rata-rata nilai presisi 90.5%, *Class Web & Mobile Programming* memiliki rata-rata nilai presisi 90.5%.



Gambar 4. 11 Grafik Nilai Recall Setiap Class

Berdasarkan hasil uji coba yang telah dilakukan, *Class Multimedia* memiliki rata-rata nilai recall sebesar 57.75%, *Class Information System* memiliki rata-rata nilai recall 78.75%, *Class Software Engineering* memiliki rata-rata nilai recall 92.25%, *Class Intelligent System* memiliki rata-rata nilai recall 57.25%, *Class System & Network* memiliki rata-rata nilai recall 95.25%, *Class Web & Mobile Programming* memiliki rata-rata nilai recall 95.25%.



Gambar 4. 12 Grafik Nilai f-1 Score Setiap Class

Berdasarkan hasil uji coba yang telah dilakukan, *Class Multimedia* memiliki rata-rata nilai f-1 score sebesar 53.5%, *Class Information System* memiliki rata-rata nilai f-1 score 77%, *Class Software Engineering* memiliki rata-rata nilai f-1 score 91.75%, *Class Intelligent System* memiliki rata-rata nilai f-1 score 61.5%, *Class System & Network* memiliki rata-rata nilai f-1 score 92.75%, *Class Web & Mobile Programming* memiliki rata-rata nilai f-1 score 92.75%.

Hasil dari performa dari setiap kelas, tingkat akurasi model dalam klasifikasi dapat dipengaruhi oleh berbagai faktor, termasuk ukuran dataset, distribusi data, kompleksitas kelas, dan kualitas fitur. Dalam kasus ini, jumlah data untuk kelas

Multimedia dan *Intelligent System* yang lebih banyak dibandingkan *System & Network* dan *Web & Mobile Programming* menunjukkan bahwa model mungkin menghadapi tantangan dalam generalisasi karena variabilitas yang lebih tinggi dalam data yang lebih besar, serta kemungkinan adanya lebih banyak kasus ambiguitas atau outlier. Sebaliknya, jumlah data yang sangat sedikit untuk kelas *System & Network* dan *Web & Mobile Programming* dapat membuat model lebih mudah untuk mencapai akurasi tinggi karena kesederhanaan pola dan rendahnya variabilitas dalam contoh-contoh tersebut. Namun, hal ini juga mengindikasikan risiko *overfitting*, di mana model mungkin tidak akan berkinerja baik pada data baru yang belum pernah dilihat sebelumnya. Selain itu, kelas dengan data lebih banyak sering kali lebih kompleks dan memiliki fitur yang lebih sulit dipisahkan, sementara kelas dengan data lebih sedikit mungkin memiliki fitur yang lebih distingtif dan mudah dikenali oleh model.

Proses analisis yang dilakukan oleh peneliti pada hasil pengujian implementasi metode KNN pada pemilihan kelompok keilmuan informatika di UIN Malang menunjukkan adanya kesesuaian, dilihat dari persentase keakuratan dan ketidakakuratan sistem dari pengujian tersebut. Dari sisi metode, implementasi algoritma KNN untuk pemilihan kelompok keilmuan informatika UIN Malang dapat dipakai sebagai sarana bagi mahasiswa yang masih bingung dalam menentukan kelompok keilmuan mana yang akan diambil. Dari sisi data, ketidaksesuaian hasil dari sistem diakibatkan oleh kelemahan metode dan faktor-faktor eksternal yang mempengaruhi hasil dan tidak dapat diakomodasi oleh perhitungan sistem, sehingga mengurangi akurasi sistem.

4.6 Integrasi Islam

Penelitian ini membahas tentang metode klasifikasi atau membagi sesuatu menjadi beberapa kelas. Salah satu dari banyak tujuan klasifikasi adalah untuk membuat literatur lebih mudah diidentifikasi dan dihitung sesuai kelasnya. Dalam kehidupan manusia, klasifikasi sudah sangat umum digunakan. Misalnya, ketika mahasiswa dihadapkan pada pilihan untuk memilih peminatan program studi atau jurusan, mereka berhak untuk memilih kelas mereka sendiri sesuai dengan kemampuan yang mereka miliki.

Al-Qur'an, yang dianggap sebagai kitab suci agama Islam, memberikan penjelasan yang luas tentang konsep klasifikasi. Perlu diingat bahwa Al-Qur'an telah menjadi dasar dari semua ilmu pengetahuan sejak Nabi Muhammad SAW diutus sebagai nabi terakhir. Dalam konsep klasifikasi dalam islam terdapat pada golongan yang memiliki hak menerima zakat, yang telah diklasifikasikan sesuai dengan firman Allah SWT Q.S At Taubah ayat 60 yang berbunyi :

إِنَّمَا الصَّدَقَتُ لِلْفُقَرَاءِ وَالْمَسْكِينِ وَالْعَمِلِينَ عَلَيْهَا وَالْمُؤَلَّفَةِ قُلُوبُهُمْ وَفِي الرِّقَابِ وَالْغُرَمِينَ وَفِي سَبِيلِ اللَّهِ وَأَبْنِ السَّبِيلِ
فَرِيضَةً مِّنَ اللَّهِ وَاللَّهُ عَلِيمٌ حَكِيمٌ

“Sesungguhnya zakat-zakat itu, hanyalah untuk orang-orang fakir, orang-orang miskin, pengurus-pengurus zakat, para mu'allaf yang dibujuk hatinya, untuk (memerdekakan) budak, orang-orang yang berhutang, untuk jalan Allah dan untuk mereka yang sedang dalam perjalanan, sebagai suatu ketetapan yang diwajibkan Allah, dan Allah Maha Mengetahui lagi Maha Bijaksana” (Q.S At-Taubah:60) .

Berdasarkan Tafsir Jalalain bahwasannya golongan orang yang berhak menerima zakat diklasifikasikan dalam 8 golongan, yaitu:

1. Golongan orang fakir, yaitu orang-orang tidak mampu mencukupi kebutuhannya.

2. Golongan orang miskin, yaitu orang yang sama sekali tidak memiliki apa-apa untuk mencukupi kebutuhannya.
3. Golongan amil, yaitu orang yang bertugas menarik zakat, yang membagikan zakat, juru tulis zakat, dan mengumpulkan zakat.
4. Golongan para muallaf yang dibujuk hatinya untuk masuk islam atau untuk memperkuat keislamannya.
5. Golongan Riqab atau memerdekakan hamba sahaya/budak yang berstatus mukatab.
6. Golongan Gharim atau orang-orang yang memiliki hutang dengan syarat bahwa hutang mereka bukan untuk maksiat.
7. Golongan Fi Sabilillah atau orang-orang yang berjuang di jalan hutang dengan syarat bahwa hutang mereka bukan untuk maksiat.
8. Golongan Ibnu Sabil atau orang musafir yang sedang dalam perjalanan yang kehabisan bekalnya.

Ayat tersebut menegaskan bahwa zakat harus diberikan kepada golongan-golongan yang telah ditentukan, dan tidak boleh menahan zakat dari golongan-golongan tersebut jika mereka memang ada dalam kondisi yang sesuai.

Berdasarkan penjelasan dari ayat yang disebutkan, menunjukkan bahwa konsep klasifikasi didasarkan pada ajaran islam. Al-Qur'an sangat tepat dalam mencontohkan konsep klasifikasi, seperti menentukan golongan mana yang berhak untuk menerima zakat.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian tentang klasifikasi kelompok keilmuan TI UIN Malang dengan pengambilan data sampel dari mahasiswa teknik informatika angkatan tahun 2018 yang bersumber dari bagian akademik UIN Malang, didapatkan hasil bahwa dari hasil pengujian implementasi metode KNN untuk klasifikasi kelompok keilmuan teknik informatika UIN Malang dapat dipakai sebagai panduan bagi mahasiswa yang masih bingung menentukan kelompok keilmuan mana yang sesuai dengan minat mereka. Sesuai dengan pengujian yang telah dilakukan dari 95 data, diperoleh Hasil dari pengujian rasio pembagian data sebanyak 95 data dan dengan nilai *K-Fold Cross Validation* sebesar $K=10$ diperoleh nilai tingkat akurasi sebesar 60% dengan pembagian data sebanyak 80 data sebagai data latih dan 15 data sebagai data uji. Pengujian yang telah dilakukan pada masing-masing *Class* yaitu *Multimedia* memperoleh tingkat akurasi sebesar 57.75%, *Information System* memperoleh tingkat akurasi sebesar 78.75%, *Software Engineering* memperoleh tingkat akurasi sebesar 92.25%, *Intelligent System* memperoleh tingkat akurasi sebesar 57.25%, *System & Network* memperoleh tingkat akurasi sebesar 95.25%, dan *Web & Mobile Programming* memperoleh tingkat akurasi sebesar 95.25%.

Faktor-faktor yang dapat menyebabkan terjadinya ketidaksesuaian hasil diakibatkan oleh kelemahan metode KNN dan faktor-faktor eksternal seperti

lingkungan dan pengaruh teman terhadap pilihan kelompok keilmuan mahasiswa, yang tidak selalu tercakup dalam perhitungan system.

5.2 Saran

Berdasarkan hasil penelitian tentang klasifikasi kelompok keilmuan TI UIN Malang menunjukkan bahwa penelitian tersebut masih mempunyai banyak kekurangan. Sehingga, harapan peneliti adalah adanya perbaikan untuk memajukan kinerja sistem. Berikut merupakan saran untuk penelitian di masa mendatang yang dapat dipertimbangkan antara lain:

1. Disarankan bagi peneliti selanjutnya untuk melakukan perbandingan dengan algoritma-algoritma lain yang telah digunakan dalam penelitian sebelumnya dengan melakukan eksplorasi dengan algoritma lain, seperti *Naïve Bayes* atau *Decision Tree* untuk mendapatkan wawasan yang lebih mendalam tentang keunggulan dan kelemahan dari masing-masing algoritma.
2. Menambahkan jumlah data yang terdapat pada dataset. Meskipun penelitian ini telah menggunakan data sejumlah 95 data, tetap disarankan untuk menambah jumlah data dalam dataset dengan menambahkan tahun angkatan 2020, 2021, dan seterusnya.
3. Analisis lebih komprehensif terkait faktor-faktor yang mempengaruhi. Penelitian yang akan datang diharapkan dapat menganalisa faktor penyebab yang dapat mempengaruhi ketidaksesuaian dari sistem, seperti faktor lingkungan, faktor teman, faktor orang tua, atau faktor lain yang dapat mempengaruhi hasil performa dari sistem.

DAFTAR PUSTAKA

- Aggarwal, C. C., & Zhai, C. (Ed.). (2012). *Mining Text Data*. Springer Us. <https://doi.org/10.1007/978-1-4614-3223-4>
- Anshori, L., Mp, R., Putri, R., & Tibyani, T. (2018). Implementasi Metode K-Nearest Neighbor Untuk Rekomendasi Keminatan Studi (Studi Kasus: Jurusan Teknik Informatika Universitas Brawijaya). *Insan Jurnal Psikologi Dan Kesehatan Mental*.
- Boiculescu, V. L., Dimitriu, G., & Moscalu, M. (2013). Improving Recall Of K-Nearest Neighbor Algorithm For Classes Of Uneven Size. *2013 E-Health And Bioengineering Conference (Ehb)*, 1–4. <https://doi.org/10.1109/Ehb.2013.6707403>
- Hadi, M. (2013). *Hakikat Sistem Pendidikan Islam*. 10.
- Han, J., Pei, J., & Kamber, M. (2012). *Data Mining: Concepts And Techniques*. Elsevier.
- Hanifah, A. W. (2023). *Sistem Rekomendasi Pemilihan Susu Formula Menggunakan Use Based Collaborative Filtering*. Uin Maulana Malik Ibrahim Malang.
- Jollyta, D., Ramdhan, W., & Zarlis, M. (2020). *Konsep Data Mining Dan Penerapan*. Deepublish.
- Kantardzic, M, M. (2019). *Data Mining Concepts, Models, Methods, And Algorithms 3rd Edition*. John Wiley & Sons.
- Kumar, M., Chhabra, P., & Garg, N. K. (2018). An Efficient Content Based Image Retrieval System Using Bayesnet And K-Nn. *Multimedia Tools And Applications*, 77(16), 21557–21570. <https://doi.org/10.1007/S11042-017-5587-8>
- Kustanti, E., Ba'diah, U., Sinawati, & Fadlan, M. (2018). Klasifikasi Siswa Berdasarkan Nilai Pada Bidang Ekstrakurikuler Menggunakan Metode K-Nearest Neighbor. *Journal Of Big Data Analyctic And Artificial Intelligence*, 4, 1–8.
- Latifah, S. N. (2019). *Analisis Prediksi Pemilihan Bidang Peminatan Menggunakan Metode Klasifikasi Dengan Algoritma Artificial Neural Network (Studi Kasus: Program Studi Sistem Informasi Universitas Telkom)*. Universitas Telkom.
- Ma'arif, A. (2020). *Buku Ajar Pemrograman Lanjut Bahasa Pemrograman Phyton*. Fakultas Teknologi Industri Universitas Ahmad Dahlan.

- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/Ei.2016.V2i2.1465>
- Marisa, F. (2013). Educational Data Mining (Konsep Dan Penerapan). *Jurnal Teknologi Informasi*, 90–97. <https://doi.org/10.36382/Jti-Tki.V4i2.108>
- Pddikti—Pangkalan Data Pendidikan Tinggi*. Diambil 5 Juni 2024, Dari https://pddikti.kemdikbud.go.id/Data_Prodi/Qtnfrduzmemtm0myrs00rjje luixnzetnda4njq4rem0rtc2/20201
- Perkel, J. M. (2018). Why Jupyter Is Data Scientists' Computational Notebook Of Choice. *Nature*, 563(7729), 145–146. <https://doi.org/10.1038/D41586-018-07196-1>
- Permana, A. A., Wahyuddin, S., Santoso, L. W., Wibowo, G. W. N., Wardhani, A. K., Rahmadden, Wahidin, A. J., Yuliasuti, G. E., Elisawati, Wijayanti, R. R., & Abdurrasyid. (2023). *Machine Learning*. Pt Global Eksekutif Teknologi.
- Prasetyo, Y., & Maksum, A. (2013). *Faktor-Faktor Penyebab Rendahnya Minat Siswa Terhadap Ekstrakurikuler Olahraga Di Smk Negeri 1 Slahung Ponorogo. 01*.
- Profile – Teknik Informatika – UIN Maulana Malik Ibrahim Malang*. Diambil 17 November 2023, Dari <https://informatika.uin-malang.ac.id/profile/>
- Rachma, C. A. (2022). *Implementasi Algoritma K-Nearest Neighbor Dalam Penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur*. Uin Maulana Malik Ibrahim Malang.
- Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (Rbnn). *Jurnal Informatika*, 11(1), 36. <https://doi.org/10.26555/Jifo.V11i1.A5452>
- S1-Teknik Informatika Uin Maulana Malik Ibrahim Malang*. S1-Teknik Informatika Uin Maulana Malik Ibrahim Malang. Diambil 5 Juni 2024, Dari <http://10.10.7.125:8069/profile>
- Sugiyono, P. (2018). *Metode Penelitian*.
- Syahrudin, A. N., & Kurniawan, T. (2018). Input Dan Output Pada Bahasa. *Jurnal Dasar Pemrograman Phyton Stmik*, 1–7.
- Umaidah, Y., & Purwantoro, P. (2019). Penerapan algoritma K-Nearest Neighbor(K-Nn) Dengan Pencarian Optimal untuk Prediksi Prestasi Siswa. *Jisicom (Journal Of Information System, Informatics And Computing)*, 3(2), Article 2.

- Wahyono, W., Trisna, I. N. P., Sariwening, S. L., Fajar, M., & Wijayanto, D. (2020). Comparison Of Distance Measurement On K-Nearest Neighbour In Textual Data Classification. *Jurnal Teknologi Dan Sistem Komputer*, 8(1), 54–58. <https://doi.org/10.14710/jtsiskom.8.1.2020.54-58>
- Wulandari, N., & Etikasari, P. (2019). Analisis Minat Belajar Siswa Pada Lembaga Pendidikan Pendidikan Indonesia Amerika Perumnas 3 Bekasi Dengan Metode C4.5. *Jurnal Rekayasa Informasi*, 8(1), Article 1.
- Xu, J., Zhang, Y., & Miao, D. (2020). Three-Way Confusion Matrix For Classification: A Measure Driven View. *Information Sciences*, 507, 772–794. <https://doi.org/10.1016/j.ins.2019.06.064>
- Yulianto, L. D., Triayudi, A., & Sholihati, I. D. (2020). Implementation Educational Data Mining For Analysis Of Student Performance Prediction With Comparison Of K-Nearest Neighbor Data Mining Method And Decision Tree C4.5: Implementation Educational Data Mining For Analysis Of Student Performance Prediction With Comparison Of K-Nearest Neighbor Data Mining Method And Decision Tree C4.5. *Jurnal Mantik*, 4(1), Article 1.
- Zainuddin, M. (2022). *Buku Pedoman Pendidikan 2022 UIN Malang*. Universitas Islam Maulana Malik Ibrahim Malang.
- Zulfallah, F. H. (2022). *Implementasi Algoritma KNN Dalam Mengukur Ketepatan Kelulusan Mahasiswa Uin Syarif Hidayatullah Jakarta*. Uin Syarif Hidayatullah Jakarta.