



**PENERAPAN ALGORITMA *K-NEAREST NEIGHBOR* UNTUK
MEMPREDIKSI KELULUSAN MAHASISWA DI UNIVERSITAS
DINAMIKA**



Program Studi

S1 SISTEM INFORMASI

**UNIVERSITAS
Dinamika**

Oleh:

FARHANNUR PRIAMBODO

22410100081

FAKULTAS TEKNOLOGI DAN INFORMATIKA

UNIVERSITAS DINAMIKA

2026

**PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK
MEMPREDIKSI KELULUSAN MAHASISWA DI UNIVERSITAS
DINAMIKA**

TUGAS AKHIR

**Diajukan sebagai salah syarat untuk menyelesaikan
Program Sarjana**



**UNIVERSITAS
Dinamika**

Oleh:

**Nama : Farhannur Priambodo
NIM : 22410100081
Program Studi : S1 Sistem Informasi**

**FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA**

2026

Tugas Akhir

PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI KELULUSAN MAHASISWA DI UNIVERSITAS DINAMIKA

Dipersiapkan dan disusun oleh

Farhannur Priambodo

NIM: 22410100081

Susunan Dewan Pembahas

Pembimbing

I. Vivine Nurcahyawati, M.Kom.

NIDN. 0723018101

II. Sulistiowati, S.Si., M.M.

NIDN. 0719016801

Digitally signed
by Vivine
Nurcahyawati
Date: 2026.03.03
16:06:01 +07'00'

Pembahas

Pantjawati Sudarmaningtyas, S.Kom., M.Eng.

NIDN. 0712066801

Pantjawati
Sudarmaningtyas
2026.03.04
11:53:34 +07'00'

Tugas Akhir ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana:

Fakultas Teknologi dan Informatika
UNIVERSITAS
Dinamika

Tan Amelia, S.Kom., M.MT., Ph.D.

NIDN. 0728017602

Dekan Fakultas Teknologi dan Informatika
UNIVERSITAS DINAMIKA

"Teruslah belajar, meskipun hal terkecil sekalipun."

-Farhannur Priambodo-



UNIVERSITAS
Dinamika

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Saya ingin mengucapkan terima kasih kepada kedua orang tua saya, saudara-saudara saya, dan keluarga besar yang telah memberikan dukungan dan do'a dalam proses penyelesaian Laporan Tugas Akhir ini. Terima Kasih.



UNIVERSITAS
Dinamika

**SURAT PERNYATAAN
PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH**

Sebagai mahasiswa **Universitas Dinamika**, saya:

Nama : **Farhannur Priambodo**
NIM : **22410100081**
Program Studi : **S1 Sistem Informasi**
Fakultas : **Fakultas Teknologi dan Informatika**
Jenis Karya : **Tugas Akhir**
Judul Karya : **PENERAPAN ALGORITMA K-NEAREST NEIGHBOR
UNTUK MEMPREDIKSI KELULUSAN MAHASISWA DI
UNIVERSITAS DINAMIKA**

Menyatakan dengan sesungguhnya bahwa:

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, Saya menyetujui memberikan kepada **Universitas Dinamika** Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas seluruh isi/sebagian karya ilmiah Saya tersebut di atas untuk disimpan, diahliemediakan, dan dikelola dalam bentuk pangkalan data (database) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama Saya sebagai penulis atau pencipta and sebagai pemilik Hak Cipta.
2. Karya tersebut di atas adalah hasil karya asli Saya, bukan plagiat baik sebagian maupun keseluruhan. Kutipan, karya, atau pendapat orang lain yang ada di dalam karya ilmiah ini semata-mata hanya sebagai rujukan yang dicantumkan dalam Daftar Pustaka Saya.
3. Apabila di kemudian hari ditemukan dan terbukti terdapat tindakan plagiasi pada karya ilmiah ini, maka Saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada Saya.

Demikian surat pernyataan ini Saya buat dengan sebenar-benarnya.

Surabaya, 19 Desember 2025



Farhannur Priambodo
NIM : 22410100081

ABSTRAK

Salah satu faktor penentu kualitas perguruan tinggi adalah persentase mahasiswa untuk menyelesaikan studi tepat waktu, dan masalah kelulusan mahasiswa banyak dialami oleh perguruan tinggi. Universitas Dinamika (UNDIKA) adalah sebuah Perguruan Tinggi Swasta (PTS) di Surabaya. UNDIKA memiliki 3 fakultas antara lain Fakultas Ekonomi dan Bisnis, Fakultas Teknologi dan Informatika, dan Fakultas Desain dan Industri Kreatif. Setiap Prodi memiliki cara atau proses tersendiri dalam memantau kelulusan mahasiswa. Permasalahannya adalah saat ini pihak prodi masih belum memiliki metode sistematis untuk memprediksi kelulusan mahasiswa, sehingga proses prediksi yang dilakukan masih dilakukan dengan memperkirakan berapa banyak semester mahasiswa untuk menyelesaikan masa studi berdasarkan laporan dari dosen wali. Dengan algoritma *K-Nearest Neighbor* (K-NN) dapat memprediksi ketepatan waktu kelulusan mahasiswa dengan cara melatih data train mahasiswa reguler S1 Sistem informasi yang telah lulus sejumlah 193 data semester 6 dan 8 dan memprediksi data test mahasiswa reguler aktif S1 Sistem Informasi sejumlah 84 data mahasiswa semester 6 dan 87 data mahasiswa semester 8. Hasil prediksi dengan algoritma K-NN menghasilkan visualisasi *pie chart* pada data semester 6 menunjukkan 45 mahasiswa diprediksi lulus tepat waktu dan 39 mahasiswa diprediksi lulus tidak tepat waktu. Sedangkan pada *pie chart* data semester 8 menunjukkan 13 mahasiswa diprediksi lulus tepat waktu dan 74 mahasiswa diprediksi lulus tidak tepat waktu. Kemudian visualisasi *scatter plot* digunakan untuk menampilkan titik data test semester 6 dan semester 8 pada sumbu X dan Y. Hasil evaluasi algoritma K-NN dengan menggunakan data alumni S1 Sistem informasi lulusan semester 24.1 dan 24.2 sejumlah 100 data menghasilkan nilai akurasi sebesar 69% untuk data semester 6 dan 85% untuk data semester 8. Berdasarkan hasil penelitian menunjukkan bahwa K-NN telah berfungsi dengan benar dan mampu memprediksi ketepatan waktu kelulusan mahasiswa dengan menghasilkan prediksi terhadap data mahasiswa reguler aktif semester 6 dan 8 dalam bentuk visualisasi *pie chart*, *scatter plot*, dan tabel sehingga prodi dapat memberikan arahan kepada dosen wali terhadap mahasiswa yang diprediksi lulus tidak tepat waktu.

Kata Kunci: *K-Nearest Neighbor*, *Data Mining*, *Kelulusan Mahasiswa*, *Confusion Matrix*, *Klasifikasi*

KATA PENGANTAR

Puji syukur kehadiran Allah SWT berkat Rahmat dan hidayah-Nya sehingga penulis mampu menyelesaikan Tugas Akhir dengan judul “PENERAPAN ALGORITMA *K-NEAREST NEIGHBOR* UNTUK MEMPREDIKSI KELULUSAN MAHASISWA DI UNIVERSITAS DINAMIKA”. Laporan Tugas Akhir ini disusun sebagai salah satu syarat untuk menyelesaikan program strata satu di Universitas Dinamika. Oleh karena itu penulis menyampaikan rasa terima kasih kepada:

1. Keluarga, yaitu kedua orang tua, kakak laki-laki dan adik perempuan saya, yang telah mendukung saya selalu.
2. Ibu Vivine Nurcahyawati, M.Kom. selaku Dosen Pembimbing pertama dalam Tugas Akhir saya yang selalu memberikan saya dukungan dan bimbingan penuh untuk karya saya.
3. Ibu Sulistiowati, S.Si., M.M. selaku Dosen Pembimbing kedua dalam kegiatan membantu pengerjaan Tugas Akhir saya.
4. Ibu Endra Rahmawati, M.Kom. selaku Ketua Program Studi S1 Sistem Informasi yang telah mendukung saya dengan Tugas Akhir ini.
5. Ibu Tan Amelia, S.Kom., M.MT., Ph.D. selaku Dekan Fakultas Teknologi dan Informatika.
6. Rezza Santika Putra, Arya Tegar, dan M. Ihksan selaku orang-orang yang selalu mendukung dan menemani saya ketika lelah dan tidak tahu arah.
7. Teman-teman di Universitas Dinamika, khususnya pada program studi S1 Sistem Informasi yang selalu menemani, memberikan dukungan dan membantu penulis.

Surabaya, 3 Maret 2026

Penulis

DAFTAR ISI

	Halaman
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	14
1.1 Latar Belakang.....	14
1.2 Rumusan Masalah.....	16
1.3 Batasan Masalah	16
1.4 Tujuan.....	16
1.5 Manfaat.....	16
BAB II LANDASAN TEORI	18
2.1 <i>K-Nearest Neighbor</i>	18
2.2 CRISP-DM	19
2.3 <i>Confusion Matrix</i>	22
2.4 <i>Python</i>	23
2.4.1 NumPy.....	23
2.4.2 Pandas.....	23
2.5 Korelasi <i>Spearman</i>	24
2.6 Variance Inflation Factor (VIF).....	24
BAB III METODOLOGI PENELITIAN	26
3.1 Tahap <i>Business Understanding</i>	26
3.2 Tahap <i>Data Understanding</i>	27
3.3 <i>Data Preparation</i>	29
3.4 <i>Modelling</i>	29
3.5 <i>Evaluation</i>	31
BAB IV HASIL DAN PEMBAHASAN	32
4.1 <i>Business Understanding</i>	32
4.2 <i>Data Understanding</i>	32

4.3	<i>Data Preparation</i>	36
4.4	Modelling.....	38
4.4.1	Proses Import Data	40
4.4.2	Prediksi pada Data Test.....	42
4.5	<i>Evaluation</i>	45
BAB V PENUTUP		48
5.1	Kesimpulan.....	48
5.2	Saran	48
DAFTAR PUSTAKA		49



UNIVERSITAS
Dinamika

DAFTAR TABEL

	Halaman
Tabel 2. 1 <i>Confusion Matrix</i>	22
Tabel 3. 1 Hasil Wawancara	27
Tabel 3. 2 Atribut Yang Diperoleh	28
Tabel 4. 1 Nilai Korelasi <i>Spearman</i>	33
Tabel 4. 2 Seleksi Atribut Berdasarkan Nilai Korelasi <i>Spearman</i>	35
Tabel 4. 3 Seleksi Atribut Setelah Perhitungan VIF	36
Tabel 4. 4 <i>Data Train</i> Semester 6	37
Tabel 4. 5 <i>Data Train</i> Semester 8	37
Tabel 4. 6 <i>Data Test</i> Semester 6	37
Tabel 4. 7 <i>Data Test</i> Semester 8	37
Tabel 4. 8 Hasil <i>Preprocessing Data Train</i> Semester 6.....	38
Tabel 4. 9 Hasil <i>Preprocessing Data Train</i> Semester 8.....	38
Tabel 4. 10 Tabel Nilai Akurasi Pada <i>Data Train</i> Berdasarkan Nilai K	39
Tabel L2. 1 Hasil Perhitungan Jarak <i>Euclidean</i> dan Prediksi Excel Data Alumni Semester 6	55
Tabel L2. 2 Hasil Perhitungan Jarak <i>Euclidean</i> dan Prediksi Excel Data Alumni Semester 8	57

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 Tahap <i>Preprocessing Data</i>	21
Gambar 3. 1 Skema CRISP-DM.....	26
Gambar 4. 1 Hasil Perhitungan VIF.....	35
Gambar 4. 2 <i>Import Data Train</i>	40
Gambar 4. 3 <i>Import Data Test</i>	41
Gambar 4. 4 Hasil Prediksi <i>Data Test</i>	42
Gambar 4. 5 <i>Pie Chart</i> Hasil Prediksi Data Semester 6.....	43
Gambar 4. 6 <i>Pie Chart</i> Hasil Prediksi Data Semester 8.....	43
Gambar 4. 9 Nilai Akurasi dan <i>Confusion Matrix</i> Perhitungan Excel Data Alumni Semester 6.....	45
Gambar 4. 10 Nilai Akurasi dan <i>Confusion Matrix</i> Perhitungan Excel Data Alumni Semester 8.....	46
Gambar 4. 11 Nilai Akurasi dan <i>Confusion Matrix</i> Perhitungan <i>Python</i> Data Alumni Semester 6.....	46
Gambar 4. 12 Nilai Akurasi dan <i>Confusion Matrix</i> Perhitungan <i>Python</i> Data Alumni Semester 8.....	47
Gambar L1. 1 <i>Source Code Python Import Data Train</i>	51
Gambar L1. 2 <i>Source Code Python Import Data Test</i>	51
Gambar L1. 3 <i>Source Code Python Display Data Train</i>	51
Gambar L1. 4 <i>Source Code Python Display Data Test</i>	52
Gambar L1. 5 <i>Source Code Python</i> Pemodelan K-NN.....	52
Gambar L1. 6 <i>Source Code Python</i> Visualisasi <i>Scatter Plot</i>	53
Gambar L1. 7 <i>Source Code Python</i> Visualisasi <i>Pie Chart</i>	53
Gambar L1. 8 <i>Source Code Python</i> Perhitungan Nilai Akurasi Dan <i>Confusion Matrix</i>	54
Gambar L3. 1 Hasil Cek Plagiarism.....	59
Gambar L4. 1 Kartu Bimbingan TA.....	60

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Source Code Aplikasi	51
Lampiran 2 Hasil Perhitungan Jarak <i>Euclidean</i> dan Prediksi dengan Excel	55
Lampiran 3 Hasil Cek Plagiat	59
Lampiran 4 Kartu Bimbingan TA	60
Lampiran 5 Biodata Penulis	61



UNIVERSITAS
Dinamika

BAB I PENDAHULUAN

1.1 Latar Belakang

Pendidikan adalah sarana yang ditujukan pada manusia untuk dapat mengembangkan potensi diri dengan proses pembelajaran yang diperoleh, menurut Undang-Undang Dasar 1945 pasal 31 Ayat 1 menyebutkan bahwa “setiap warga negara berhak mendapatkan pendidikan”(Fadia & Fitri, 2021). Jenjang pendidikan di Indonesia dibagi menjadi pendidikan umum yang terdiri atas SD (Sekolah Dasar), SMP (Sekolah Menengah Pertama), SMA (Sekolah Menengah Atas), dan Perguruan Tinggi (Rembangsupu et al., 2022).

Perguruan tinggi memiliki jenis-jenis tersendiri antara lain pendidikan profesi, pendidikan akademik, dan pendidikan vokasi. Pendidikan akademik terdiri atas program sarjana, magister dan doktor. Pendidikan akademik terdiri atas program sarjana, program magister dan program doktor. Pendidikan vokasi terdiri atas program diploma seperti D1, D2, D3, dan D4, pendidikan akademik terdiri atas sarjana terapan, sarjana terapan yang bergelar ahli, magister terapan, doktor terapan, profesi dan spesialis (Gustini et al., 2023).

Salah satu faktor penentu kualitas perguruan tinggi adalah persentase mahasiswa untuk menyelesaikan studi tepat waktu, dan masalah kelulusan mahasiswa banyak dialami oleh perguruan tinggi. Lama studi berpengaruh pada kualitas program studi karena merupakan salah satu faktor pada penilaian akreditasi yaitu penilaian terhadap lama studi mahasiswa yang telah lulus. Lama studi mahasiswa dapat dipengaruhi oleh faktor yang dapat dilihat pada Indeks Prestasi Kumulatif (IPK) yang menunjukkan keseluruhan nilai yang didapatkan oleh mahasiswa setiap semester. IPK didapatkan dengan menjumlahkan nilai semua mata kuliah yang telah ditempuh dan dibagi oleh total Satuan Kredit Semester (SKS) (Yuda et al., 2022).

Universitas Dinamika (UNDIKA) adalah sebuah Perguruan Tinggi Swasta (PTS) di Surabaya. UNDIKA memiliki 3 fakultas antara lain Fakultas Ekonomi dan Bisnis, Fakultas Teknologi dan Informatika, dan Fakultas Desain dan Industri Kreatif, dan 8 program studi yang terdiri atas S1 Sistem Informasi, D3 Sistem

Informasi, S1 Manajemen, S1 Teknik Komputer, S1 Desain Produk, S1 Akuntansi, S1 Desain Komunikasi Visual, dan D4 Produksi Film dan Televisi. Setiap Prodi memiliki cara atau proses tersendiri dalam memantau kelulusan mahasiswa. Prodi S1 Sistem Informasi memantau kelulusan mahasiswa dengan cara merekap laporan mahasiswa yang diperoleh dari masing-masing dosen wali, kemudian dianalisis dan diprediksi secara manual seberapa lama mahasiswa akan menyelesaikan masa studi kuliah. Rekap laporan mahasiswa kemudian dibuat ke dalam bentuk *dashboard* yang nantinya akan dilaporkan kepada Dekan FTI.

Permasalahannya adalah saat ini pihak prodi masih belum memiliki metode sistematis untuk memprediksi kelulusan mahasiswa, sehingga proses prediksi yang dilakukan masih dilakukan dengan memperkirakan berapa banyak semester mahasiswa untuk menyelesaikan masa studi berdasarkan laporan dari dosen wali. Berdasarkan latar belakang maka dapat ditemukan sebuah masalah yaitu prodi belum menerapkan metode dalam memprediksi ketepatan waktu kelulusan mahasiswa. Maka dari itu dibutuhkan sebuah solusi untuk menyelesaikan masalah. Salah satu solusinya adalah dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN).

Algoritma K-NN dapat digunakan untuk memprediksi, mengklasifikasi, memperkirakan dan mendapatkan informasi yang berguna berdasarkan data dengan jumlah besar (Manullang & Sianturi, 2021). Pada penelitian milik (Kartarina et al., 2021) menghasilkan nilai akurasi penerapan Algoritma K-NN sebesar 96,18% dan algoritma *Naïve Bayes* sebesar 91,94%. Pada penelitian milik (Manullang & Sianturi, 2021) menunjukkan hasil penerapan algoritma K-NN dari 215 data, mendapatkan hasil prediksi jumlah mahasiswa yang lulus untuk Program Studi Teknik Informatika sejumlah 148 mahasiswa, dan jumlah mahasiswa yang tidak lulus sejumlah 67 mahasiswa. Pada penelitian milik (Lestari & Andriansyah, 2021) menghasilkan nilai akurasi dan nilai k dengan menggunakan algoritma K-NN sebesar 93,55% pada nilai $k=3$, dan dengan menggunakan metode *Forward Selection* sebesar 99,39% pada nilai $k=3$. Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma K-NN dapat melakukan prediksi kelulusan mahasiswa dengan baik dengan menghasilkan nilai akurasi yang tinggi.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka dapat dirumuskan sebuah rumusan masalah: Bagaimana menerapkan algoritma K-NN dalam memprediksi kelulusan mahasiswa di Universitas Dinamika?

1.3 Batasan Masalah

Berdasarkan rumusan masalah yang telah dirumuskan, maka dapat ditentukan beberapa poin batasan masalah sebagai berikut:

1. Data yang digunakan sebagai *data train* adalah data mahasiswa reguler S1 Sistem Informasi yang telah lulus pada tahun 2019 dan 2020.
2. Data yang digunakan sebagai *data test* adalah data mahasiswa reguler aktif angkatan 2021 dan 2022 S1 Sistem Informasi yang telah atau sedang menempuh semester 6 sampai dengan 8.
3. Output penelitian berupa hasil dan tingkat akurasi prediksi “Tepat Waktu” dan “Tidak Tepat Waktu”.
4. Proses Klasifikasi, Evaluasi Data, dan Visualisasi data dikerjakan menggunakan Python GUI.

1.4 Tujuan

Tujuan yang akan dicapai pada penelitian ini adalah menghasilkan prediksi algoritma K-NN ke dalam bentuk *pie chart*, *scatter plot*, dan tabel sehingga dapat ditentukan ketepatan waktu kelulusan mahasiswa S1 Sistem Informasi di Universitas Dinamika.

1.5 Manfaat

Beberapa manfaat yang didapat dari penelitian ini adalah:

1. Bagi Program Studi

Dengan menggunakan program dari penelitian ini diharapkan untuk digunakan sebagai bahan pertimbangan dalam menentukan ketepatan waktu kelulusan mahasiswa secara langsung, tanpa perlu adanya laporan dari dosen wali.

2. Bagi Dosen Wali

Penelitian ini diharapkan dapat membantu dosen wali untuk mendapatkan wawasan terkait hasil prediksi ketepatan waktu kelulusan mahasiswanya, sehingga hasil prediksi dapat dijadikan sebagai salah satu acuan untuk terus mendukung mahasiswa walinya yang masih aktif untuk dapat menyelesaikan program studi tepat waktu.

3. Bagi Peneliti dan Akademisi

Penelitian ini memiliki kontribusi pada kajian akademis di bidang akademik. Dengan menggunakan analisis klasifikasi seperti K-NN penelitian ini diharapkan dapat menjadi contoh penerapan metode klasifikasi untuk memahami pola ketepatan waktu kelulusan mahasiswa.



UNIVERSITAS
Dinamika

BAB II

LANDASAN TEORI

2.1 *K-Nearest Neighbor*

Algoritma K-NN adalah metode klasifikasi objek berdasarkan *data train* yang memiliki jarak terdekat dengan objek. Teknik ini mudah digunakan dan sederhana untuk diimplementasikan. Klasifikasi K-NN mirip dengan *clustering*, yaitu mengelompokkan data berdasarkan jarak data tersebut ke beberapa data terdekat (Putra et al., 2024b). K-NN adalah salah satu teknik *lazy learning*, dan termasuk algoritma kelompok *instance-based learning*. K-NN dilakukan dengan mencari kelompok *k* objek pada *data train* terdekat dengan objek data baru atau *data test*. Sebelum mencari jarak data ke tetangga adalah menentukan nilai *k* tetangga (*neighbor*), kemudian untuk mendefinisikan jarak antara kedua titik yaitu titik pada *data train* dan titik pada *data test*, maka digunakan rumus *Euclidean* dengan persamaan:

$$euc = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Keterangan:

- euc : jarak Euclidean
- p : sampel data / *data train*
- q : data uji / *data test*
- i : variabel data
- n : dimensi data

Langkah-langkah K-NN:

1. Menentukan parameter *k* (jumlah data tetangga terdekat), parameter *k* pada testing ditentukan berdasarkan nilai *k* optimum pada saat pelatihan pada data train.
2. Menghitung kuadrat *Euclidean Distance* masing-masing objek terhadap data sampel atau *data train* yang diberikan dengan menggunakan rumus (1).
3. Mengurutkan objek-objek tersebut ke kelompok data yang memiliki *Euclidean Distance* terkecil
4. Mengumpulkan kategori *Y* (klasifikasi K-NN)

5. Mencari jumlah kelas dari tetangga terdekat dan menetapkan kelas dengan jumlah data terbanyak sebagai kelas yang akan dievaluasi.

2.2 CRISP-DM

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah kerangka kerja terstruktur yang digunakan secara luas pada data mining. CRISP-DM menyediakan tahap-tahap untuk memandu proyek data mining, membantu memastikan bahwa proyek dapat berjalan dengan efektif dan mendapatkan hasil yang bermanfaat (Murti, 2024). CRISP-DM memiliki 6 tahap antara lain:

1. Tahap *Business Understanding*

Tahap *Business Understanding* berfokus pada pemahaman tujuan penelitian, persyaratan proyek, dan cara data mining dapat memenuhi tujuan penelitian.

2. Tahap *Data Understanding*

Tahap *Data Understanding* melibatkan pengumpulan data awal, eksplorasi data, dan identifikasi masalah atau potensi masalah pada data.

3. Tahap *Data Preparation*

Tahap *Data Preparation* melibatkan tahap pada *Data Preprocessing*. *Data Preprocessing* adalah salah satu tugas data mining yang memiliki persiapan dan transformasi data menjadi bentuk yang pas untuk prosedur mining. *Data Preprocessing* memiliki tujuan untuk mengurangi ukuran data, normalisasi data, mencari relasi antara data, menghapus *outlier*, dan mengekstrak fitur untuk data, termasuk beberapa teknik antara lain *data cleaning*, integrasi, transformasi, dan reduksi (Putra et al., 2024a).

- a. *Data Cleaning* (Pembersihan Data)

Data Cleaning merupakan proses penghilangan data yang tidak relevan atau konsisten. Data yang diperoleh pada umumnya memiliki isi yang tidak lengkap, seperti data yang tidak valid atau data yang hilang. Pembersihan data yang dilakukan akan mengurangi kompleksitas dan jumlah data yang akan dimodelkan, sehingga performa teknik data mining akan terpengaruh. Contoh *data cleaning* adalah dengan mengisi data yang tidak memiliki nilai dengan median, rata-rata atau nilai

terdekat, mengubah format tanggal yang berbeda menjadi format yang seragam.

b. *Data Integration* (Integrasi Data)

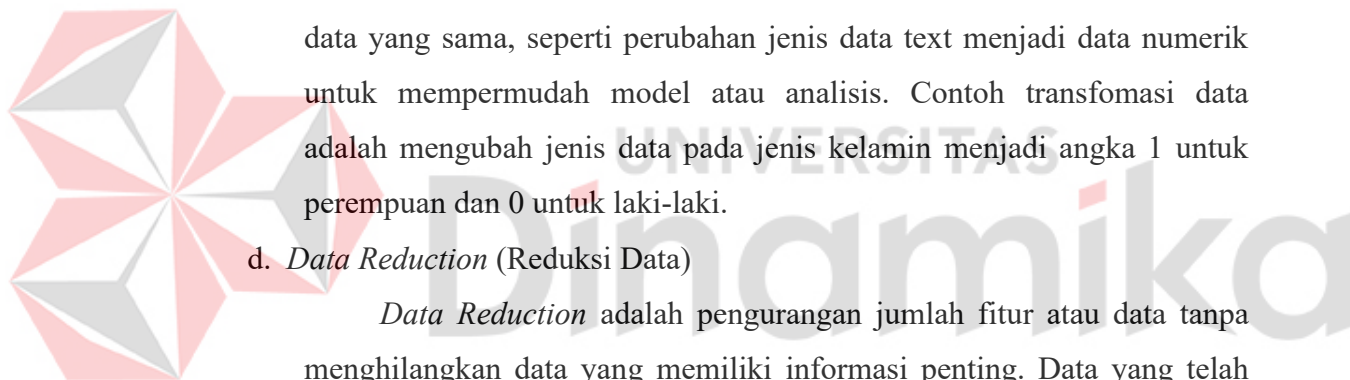
Data Integration adalah penggabungan dataset dari berbagai macam *database* ke *database* baru. Data yang diperoleh memiliki kemungkinan berasal dari *database* yang berbeda. Integrasi data harus dilakukan secara teliti dan cermat karena memiliki resiko hasil data yang meyimpang. Contoh integrasi data adalah menggabungkan data karyawan dengan data departemen dengan memastikan format atribut konsisten antar dua data.

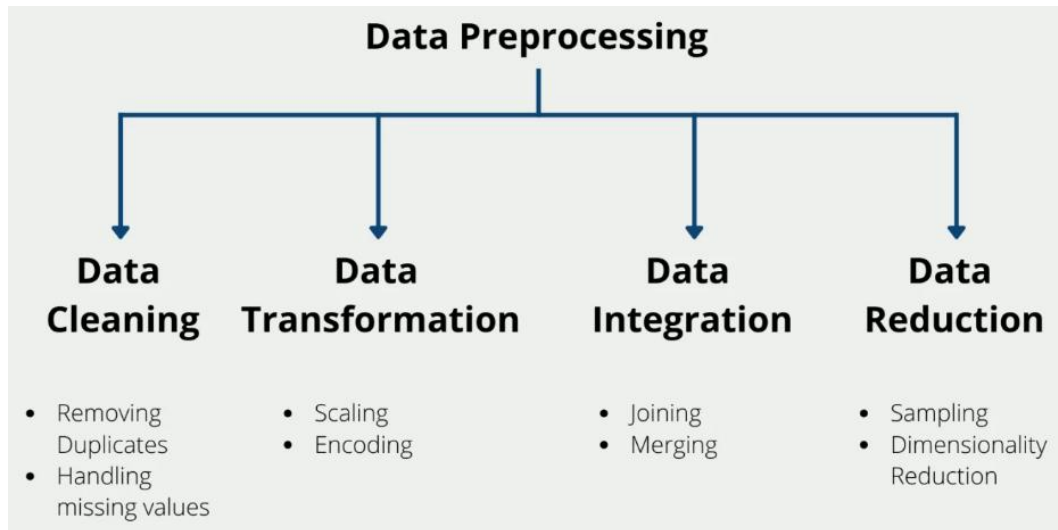
c. *Data Transformation* (Transformasi Data)

Data Transformation adalah perubahan atau normalisasi data untuk menyamakan jenis data pada semua atribut sehingga memiliki rentang data yang sama, seperti perubahan jenis data text menjadi data numerik untuk mempermudah model atau analisis. Contoh transformasi data adalah mengubah jenis data pada jenis kelamin menjadi angka 1 untuk perempuan dan 0 untuk laki-laki.

d. *Data Reduction* (Reduksi Data)

Data Reduction adalah pengurangan jumlah fitur atau data tanpa menghilangkan data yang memiliki informasi penting. Data yang telah diproses tidak selalu berisi data yang relevan, sehingga perlu dilakukan reduksi data. Contoh reduksi data adalah menghapus fitur yang tidak penting, melakukan pemilihan fitur untuk memilih fitur yang memiliki informasi penting, atau melakukan sampling data dengan menggunakan sebagian kecil data yang representatif (Qisthiano et al., 2021).





Gambar 2. 1 Tahap *Preprocessing Data*

4. Tahap *Modelling*

Tahap *Modeling* melibatkan pemilihan dan penerapan metode pemodelan yang sesuai untuk menjawab tujuan penelitian, seperti algoritma *machine learning*.

5. Tahap *Evaluation*

Tahap *Evaluation* melibatkan evaluasi model untuk memastikan bahwa model yang digunakan efektif dan sesuai dengan tujuan penelitian. *Confusion Matrix* adalah alat utama untuk mengukur kinerja klasifikasi biner maupun multi-kelas. Proses evaluasi pada *Confusion Matrix* adalah dengan mengukur nilai akurasi, presisi, dan *recall* sebuah model klasifikasi. Akurasi adalah nilai seberapa sering model mengklasifikasi dengan benar sesuai dengan label aktual, Presisi adalah berapa banyak data dari salah satu kelas yang diprediksi benar-benar memiliki kelas aktual yang sama, *Recall* adalah berapa banyak data yang berhasil dideteksi dari semua data pada salah satu kelas aktual.

6. Tahap *Deployment*

Tahap *Deployment* melibatkan penerapan model dalam lingkungan operasional dan pemantauan kinerja model dari waktu ke waktu.

2.3 Confusion Matrix

Confusion Matrix adalah tabel yang berfungsi untuk menggambarkan performa model klasifikasi pada Data Test yang nilai sebenarnya diketahui. *Confusion matrix* menunjukkan langkah-langkah ketika klasifikasi bingung dalam menentukan *class*-nya pada saat membuat prediksi. *Confusion Matrix* memberi informasi yang detil tidak hanya mengenai kesalahan yang dibuat oleh *classifier* tetapi lebih penting lagi jenis kesalahan yang dibuat (Said et al., 2022).

Confusion matrix menampilkan hasil akurasi *classifier* dengan membandingkan kelas aktual dan prediksi. *Binary classifier* memprediksi semua *instance* dari *data test* sebagai negatif atau positif. Hasil klasifikasi menampilkan empat hasil antara lain *True Negative* (TN), *False Negative* (FN), *True Positive* (TP), dan *False Positive* (FP). TP menghasilkan nilai prediksi benar diprediksi sebagai positif aktual, dan FP menghasilkan nilai prediksi salah diprediksi sebagai positif aktual. misal, nilai-nilai negatif diprediksi sebagai positif. Sedangkan FN menghasilkan nilai prediksi benar diprediksi sebagai negatif aktual dan TN menghasilkan nilai prediksi benar diprediksi sebagai negatif aktual.

Tabel 2. 1 *Confusion Matrix*

		<i>Predicted Values</i>	
		<i>False</i>	<i>True</i>
<i>Acuta Values</i>	<i>False</i>	TN (<i>True Negative</i>)	FN (<i>False Negative</i>) <i>Type II Error</i>
	<i>True</i>	FP (<i>False Positive</i>) <i>Type I Error</i>	TP (<i>True Positive</i>)

Berdasarkan *confusion matrix* dapat dibuat sebuah metrik pengukuran untuk mendapatkan nilai akurasi (*Accuracy*). *Accuracy* adalah sebuah metrik evaluasi untuk mengukur seberapa baik sebuah model membuat prediksi yang benar berdasarkan total prediksi yang telah dilakukan (Riadi et al., 2024).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

2.4 Python

Python adalah sebuah bahasa pemrograman yang biasa dipakai untuk membuat sebuah program/aplikasi, membangun situs web, mengotomatiskan tugas, dan menganalisis data. *Python* dikenalkan pada tahun 1991 oleh Guido van Rossum dan telah berkembang sebagai bahasa pemrograman yang umum digunakan di dunia, terutama bidang pengembangan web, *data science*, dan otomatisasi. *Python* mendukung berbagai macam pemrograman, termasuk pemrograman fungsional, prosedural, dan objek sehingga pemakaiannya dapat disesuaikan oleh kebutuhan pengembang. *Python* memiliki banyak pustaka yang dapat digunakan, Pandas untuk memanipulasi data, NumPy untuk pengoperasian numerik, dan Matplotlib untuk memvisualisasikan data. Maka dari itu *Python* adalah pilihan ideal bagi ilmuwan data dan analisis data (Candra, 2025).

2.4.1 NumPy

Numerical Python (NumPy) adalah *library* pada *Python* untuk mengkomputasi data numerik. Salah satu kemampuan dari NumPy adalah menangani *array* multidimensi (dua atau lebih) dibandingkan dengan *python* biasa yang hanya dapat menampung data satu dimensi. NumPy sangat berguna untuk menganalisis data, karena sering kali mengelola data terstruktur dalam bentuk matriks atau tabel (Candra, 2025).

2.4.2 Pandas

Pandas adalah *Python library* yang dirancang untuk menganalisis dan memanipulasi data. Pandas menawarkan alat analisis yang kuat dan struktur data yang fleksibel sehingga pengerjaan data lebih mudah. Pandas sangat cocok untuk menganalisis data dengan berbagai macam format, termasuk database SQL, CSV, dan Excel. Salah satu keunggulan Pandas adalah kemampuan untuk menangani *missing value* (data yang hilang) dan pemfilteran yang cepat. Dengan Pandas, user dapat memanipulasi, menganalisis, dan membersihkan data hanya dengan sedikit kode (Candra, 2025).

2.5 Korelasi *Spearman*

Korelasi *Spearman* adalah uji statistik untuk mengetahui hubungan antar dua variabel dengan skala Ordinal. Korelasi *Spearman* memiliki koefisien statistik-non parametik, dikarenakan data yang diperoleh tidak berdistribusi normal. Korelasi *Spearman* digunakan untuk mengevaluasi performa fungsi monotonik dari fungsi yang sesuai dengan perintah, sehingga memvisualisasikan hubungan antara dua variabel tanpa asumsi distribusi frekuensi variabel yang diteliti.

Nilai koefisien antara Korelasi *Pearson* dengan *Spearman* sama, namun *Spearman* memiliki keuntungan dimana analisis dengan basis peringkat lebih mudah dihitung daripada analisis dengan basis numerik dan dapat digunakan untuk korelasi non linear. Metode yang digunakan korelasi ini adalah menghitung peringkat korelasi mulai dari +1, dimana nilai tersebut berarti korelasi sempurna dalam derajat kemiringan positif dan -1 untuk nilai korelasi sempurna dalam derajat kemiringan negatif. Jika nilai korelasi tersebut bernilai di atas 0,5 atau di bawah -0,5 maka korelasi tersebut dinamakan hubungan yang kuat (Pamungkasih, 2023a).

Korelasi *Spearman* memiliki rumus perhitungan sebagai berikut.

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)} \quad (3)$$

Keterangan:

r_s = Nilai *Korelasi Spearman*

d = Selisih antara X dan Y

n = Jumlah pasangan (data)

2.6 Variance Inflation Factor (VIF)

VIF berfungsi untuk mengukur besar varians dari koefisien regresi yang diestimasi meningkat diakibatkan adanya korelasi antar variabel bebas. Pada model regresi, VIF dihitung pada setiap variabel bebas dengan menggunakan rumus sebagai berikut.

$$VIF = 1/(1 - R^2) \quad (4)$$

Keterangan:

R^2 = Koefisien determinasi variabel bebas

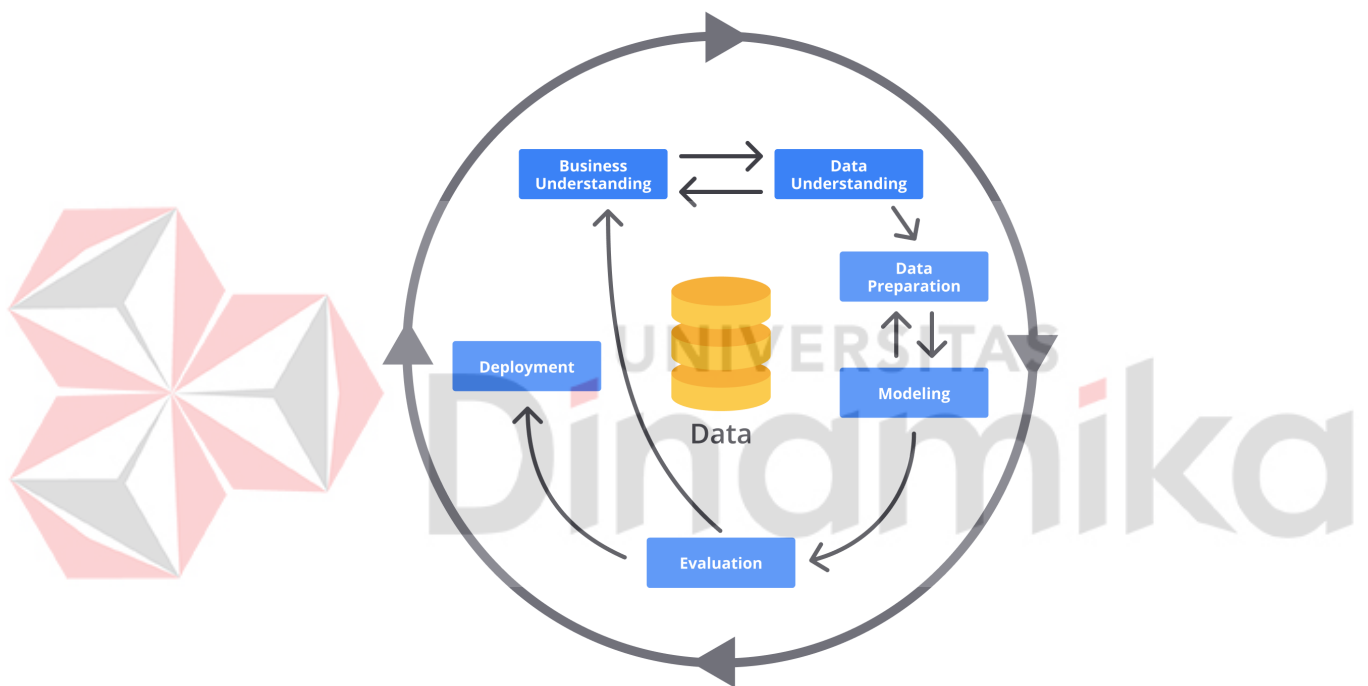
Nilai VIF sebesar 1 menyatakan bahwa variabel bebas tidak memiliki multikolinieritas, dan nilai di atas 1 menyatakan bahwa terdapat peningkatan multikolinieritas. Nilai VIF di atas 10 dianggap tinggi dan mengindikasikan multikolinieritas parah (Kyriazos & Poga, 2023).



UNIVERSITAS
Dinamika

BAB III METODOLOGI PENELITIAN

Metodologi penelitian adalah serangkaian prosedur, alur sistematis, dan teknik yang digunakan pada penelitian untuk mengumpulkan, menganalisis, dan menginterpretasikan data dengan tujuan untuk memecahkan atau memenuhi tujuan penelitian. Pada penelitian ini menggunakan CRISP-DM sebagai metodologi penelitian.



Gambar 3. 1 Skema CRISP-DM

3.1 Tahap *Business Understanding*

Tahap *Business Understanding* memahami masalah pada Program Studi S1 Sistem Informasi di Universitas Dinamika terkait ketelambatan kelulusan mahasiswa secara mendalam dan menentukan tujuan penelitian berdasarkan rumusan masalah yang telah dirumuskan. Hal pertama yang dilakukan adalah wawancara dengan Prodi terkait metode prediksi kelulusan mahasiswa saat ini sehingga menghasilkan hasil wawancara pada Tabel 3. 1.

Tabel 3. 1 Hasil Wawancara

NO	PERTANYAAN	JAWABAN NARASUMBER
1	Bagaimana proses memprediksi kelulusan mahasiswa di Prodi S1 Sistem Informasi saat ini?	Proses memprediksinya dengan cara melihat semester aktif mahasiswa saat ini, kemudian memperkirakan seberapa lama mahasiswa bakal menyelesaikan program studinya.
2	Bagaimanakah cara mendapatkan data mahasiswa tersebut?	Berdasarkan laporan dari dosen wali yang dilaporkan sebanyak 3 kali dalam satu semester
3	Apa isi laporan dari dosen wali?	Biasanya berupa keterangan presensi dan keluhan atau kendala dari mahasiswa walinya. Dari situ bisa diperkirakan berapa lama mahasiswa akan menyelesaikan program studinya. Dari laporan tersebut saya rekap untuk saya laporkan ke Dekan.
4	Apakah ada metode yang lebih efektif untuk memprediksi ketepatan waktu kelulusan mahasiswa?	Sampai saat ini masih belum ada metode yang sistematis untuk memprediksinya, karena masih mengandalkan laporan dari dosen wali.

3.2 Tahap *Data Understanding*

Tahap *Data Understanding* mengumpulkan dan memahami data yang diperoleh. Data yang akan digunakan adalah data mahasiswa reguler S1 Sistem Informasi yang diperoleh dari *database* mahasiswa di bagian Pengembangan dan Penerapan Teknologi Informasi (PPTI) Universitas Dinamika. Data yang diperoleh berupa tiga file excel berisi data mahasiswa yang telah lulus pada tahun 2019 sampai dengan 2020 yang akan berfungsi sebagai *data train*, data mahasiswa reguler aktif angkatan 2021 dan 2022 sebagai *data test*, dan data alumni lulusan semester 24.1 dan 24.1 sebagai *data test* untuk pengukuran nilai akurasi.

Data yang diperoleh memiliki atribut yang kemudian dihitung nilai korelasi dengan metode Korelasi *Spearman* pada IBM SPSS Statistic. Tujuan penggunaan Korelasi *Spearman* adalah untuk melihat tingkat keeratan antara dua variabel, melihat arah hubungan dua variabel, dan melihat apakah hubungan variabel signifikan atau tidak (Pamungkasih, 2023b). Pada penelitian ini Korelasi *Spearman* digunakan untuk melihat ada tidaknya hubungan antara variabel bebas dengan variabel terikat (status kelulusan). Nilai korelasi setiap atribut diurutkan berdasarkan nilai korelasi tertinggi dan diseleksi atributnya dengan melihat pedoman derajat hubungan. Atribut data yang diperoleh dapat dilihat pada Tabel 3. 2.

Tabel 3. 2 Atribut Yang Diperoleh

NO.	NAMA ATRIBUT	KETERANGAN
1.	NIM	Nomor Induk Mahasiswa
2.	NAMA	Nama mahasiswa
3.	JENIS KELAMIN	Jenis kelamin mahasiswa
4.	UMUR	Usia mahasiswa
5.	JUMLAH AKTIF	Jumlah semester berstatus registrasi "Aktif Tidak Ambil TA"
6.	JUMLAH CUTI	Jumlah semester berstatus registrasi "Cuti"
7.	JUMLAH AMBIL TA	Jumlah semester berstatus registrasi "Aktif Ambil TA"
8.	JUMLAH MK MENGULANG	Jumlah mata kuliah berstatus Ulang dengan nilai di bawah C
9.	IPP1	Indeks Prestasi Persiapan semester 1
10.	IPP2	Indeks Prestasi Persiapan semester 1 dan 2
11.	IPP3	Indeks Prestasi Persiapan semester 1, 2, dan 3
12.	IPP4	Indeks Prestasi Persiapan semester 1 sampai dengan 4
13.	IPS1	Indeks Prestasi Sementara semester 1
14.	IPS2	Indeks Prestasi Sementara semester 2
15.	IPS3	Indeks Prestasi Sementara semester 3
16.	IPS4	Indeks Prestasi Sementara semester 4
17.	IPS5	Indeks Prestasi Sementara semester 5
18.	IPS6	Indeks Prestasi Sementara semester 6
19.	IPK	Indeks Prestasi Kumulatif
20.	STATUS KELULUSAN	Status lulus mahasiswa

Berdasarkan data beserta atribut yang diperoleh kemudian dihitung dengan menggunakan Korelasi Spearman untuk menghitung nilai korelasi setiap atribut data kecuali NIM dan nama terhadap STATUS KELULUSAN untuk menentukan atribut tersebut memiliki korelasi terhadap kelulusan atau tidak berdasarkan pedoman derajat hubungan pada Gambar 3. 2 (Pamungkasih, 2023a).

Nilai koefisien korelasi	Kriteria korelasi
(1)	(2)
0,00 – 0,25	hubungan sangat lemah
0,26 – 0,50	hubungan cukup
0,51 – 0,75	hubungan kuat
0,76 – 0,99	hubungan sangat kuat
1,00	hubungan sempurna

Gambar 3. 2 Pedoman Derajat Hubungan

3.3 Data Preparation

Tahap *Data Preparation* memperbaiki masalah pada data yang telah diidentifikasi pada tahap *Data Understanding*. Data mahasiswa yang atributnya telah dihitung nilai korelasinya kemudian dinormalisasikan dengan menggunakan preprocessing data. Salah satu proses yang dilakukan adalah dengan menggunakan transformasi data, dimana pada data train terdapat data berisi teks pada kolom STATUS KELULUSAN dengan nilai “Tepat Waktu” dan “Tidak Tepat Waktu diubah menjadi data numerik.

3.4 Modelling

Tahap *Modelling* pada metode CRISP-DM melakukan analisis menggunakan algoritma K-NN terhadap data yang diperoleh dan telah melalui *preprocess*. Pada penelitian ini data akan dimasukkan ke program dengan menggunakan scikit-learn pada *Python Library* untuk melakukan pemodelan K-NN. K-NN menyimpan data train tanpa mempelajari pola terlebih dahulu kemudian pada saat data test dimasukkan sebagai data baru, K-NN menentukan jumlah titik data terdekat milik *data train* (*neighbor*) dan menghitung jarak ke semua titik data tersebut (*Euclidean Distance*) dari titik data test. Kemudian label pada *data test* diprediksi dengan menggunakan pemilihan mayoritas pada label *data train* yang jaraknya telah dihitung. Berdasarkan rumus (1) maka dapat dilakukan simulasi perhitungan jarak Euclidean sebagai berikut.

Data test: JUMLAH AMBIL TA=0, JUMLAH MK MENGULANG=2, IPK=3,25

Neighbor 1: JUMLAH AMBIL TA=0, JUMLAH MK MENGULANG=1, IPK=3,33, STATUS KELULUSAN=1

$$euc = \sqrt{(0 - 0)^2 + (2 - 1)^2 + (3,25 - 3,33)^2}$$

$$euc = \sqrt{0 + 1 + 0,0064}$$

$$euc = 1,003$$

Neighbor 2: JUMLAH AMBIL TA=1, JUMLAH MK MENGULANG=4, IPK=2,85, STATUS KELULUSAN=0

$$euc = \sqrt{(0 - 1)^2 + (2 - 4)^2 + (3,25 - 2,85)^2}$$

$$euc = \sqrt{1 + 4 + 0,16}$$

$$euc = 2,271$$

Neighbor 3: JUMLAH AMBIL TA=0, JUMLAH MK MENGULANG=0, IPK=3,77, STATUS KELULUSAN=1

$$euc = \sqrt{(0 - 0)^2 + (2 - 0)^2 + (3,25 - 3,77)^2}$$

$$euc = \sqrt{0 + 4 + 0,274}$$

$$euc = 2,006$$

Berdasarkan ketiga *neighbor* yang jaraknya telah dihitung, kemudian *data test* diprediksi dengan cara memilih mayoritas dari nilai label pada *neighbor*. Pada simulasi ini terdapat 2 *neighbor* yang memiliki STATUS KELULUSAN dengan nilai 1 dan 1 *neighbor* yang memiliki STATUS KELULUSAN dengan nilai 0. Maka dari itu STATUS KELULUSAN pada *data test* diisi nilai 1 sehingga diprediksi lulus tepat waktu.

Prediksi yang telah dilakukan divisualisasikan ke dalam bentuk *pie chart* dan *scatter plot* untuk menampilkan distribusi label prediksi pada *pie chart* dan titik data test pada *scatter plot*. *Pie chart* bertujuan untuk menampilkan jumlah mahasiswa yang telah diprediksi dalam bentuk lingkaran dengan masing-masing irisan menunjukkan jumlah mahasiswa dengan label prediksi “Tepat Waktu” dan “Tidak Tepat Waktu”. *Scatter Plot* bertujuan untuk menampilkan data test dalam ruang 2D sehingga memudahkan pengamatan bagaimana titik-titik data terkelompok berdasarkan kelasnya.

3.5 *Evaluation*

Tahap *Evaluation* memvisualisasikan dan mengevaluasi hasil pemodelan K-NN yang telah dilakukan. Proses evaluasi dilakukan dengan menampilkan *confusion matrix* sehingga dapat dihitung nilai akurasi berdasarkan jumlah data TP ditambah TN dibagi dengan jumlah total data. Hasil akurasi dapat menunjukkan keakuratan algoritma K-NN dalam memprediksi data test berdasarkan atribut data yang telah diseleksi dan di-*preprocess*. Kemudian kesimpulan dapat ditarik berdasarkan hasil prediksi ketepatan waktu kelulusan mahasiswa dengan K-NN dan nilai akurasi dari *confusion matrix*.



UNIVERSITAS
Dinamika

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Business Understanding*

Permasalahan yang menjadi pokok penelitian ini adalah banyaknya mahasiswa S1 Sistem Informasi di Universitas Dinamika yang lulus melebihi batas normal masa studi yaitu 8 semester. Kasus tersebut memiliki kemungkinan berdampak pada kualitas perguruan tinggi. Hasil laporan dari setiap dosen wali hanya dapat diperoleh per dua bulan sejumlah tiga kali persemester, sehingga proses pemantauan kelulusan yang dilakukan oleh prodi tidak dapat dilakukan kapanpun.

Ketidakmampuan untuk memantau kelulusan mahasiswa secara langsung akan menyulitkan prodi untuk menyiasati tindakan mahasiswa yang akan dan sedang menempuh Tugas Akhir (TA), termasuk mahasiswa yang masih menempuh beberapa semester. Pihak prodi tidak memiliki metode sistematis untuk memprediksi kelulusan mahasiswa, dimana metode ini dapat digunakan untuk memprediksi apakah mahasiswa dapat lulus dengan tepat waktu atau tidak.

Berasarkan permasalahan yang ada, maka pada penelitian dapat diidentifikasi tujuan utama penelitian yaitu menerapkan metode K-NN sebagai alat bantu prodi untuk memprediksi ketepatan waktu kelulusan mahasiswa sehingga pihak prodi dapat mengantisipasi keterlambatan kelulusan mahasiswa yang masih aktif.

4.2 *Data Understanding*

Pada penelitian ini, data yang didapat berasal dari *database* mahasiswa di PPTI Universitas Dinamika. Data berisi mahasiswa reguler S1 Sistem Informasi yang lulus pada tahun 2019 dan 2020 sebagai *data train* dan mahasiswa reguler aktif S1 Sistem Informasi angkatan 2021 dan 2022 yang telah atau sedang menempuh semester 6 sampai dengan 8. Data dengan atribut yang diperoleh dihitung nilai korelasinya dengan Korelasi Spearman sehingga nilai korelasinya dapat dilihat pada Tabel 4. 1.

Tabel 4. 1 Nilai Korelasi *Spearman*

ATRIBUT	SPEARMAN	SIG. (2-TAILED)
JENIS KELAMIN	0,320	0,000
UMUR	-0,063	0,383
JUMLAH AKTIF	-0,112	0,122
JUMLAH CUTI	-0,040	0,578
JUMLAH AMBIL TA	0,636	0,000
JUMLAH MK MENGULANG	-0,401	0,000
IPP1	0,499	0,000
IPP2	0,541	0,000
IPP3	0,528	0,000
IPP4	0,522	0,000
IPS1	0,497	0,000
IPS2	0,504	0,000
IPS3	0,561	0,000
IPS4	0,520	0,000
IPS5	0,463	0,000
IPS6	0,580	0,000
IPK	0,687	0,000

Berdasarkan hasil perhitungan korelasi, maka dapat diketahui korelasi setiap atribut data terhadap status kelulusan dengan melihat nilai r (*Spearman*) sehingga korelasinya dapat dijelaskan sebagai berikut.

1. Jenis Kelamin

Jenis Kelamin menghasilkan nilai r sebesar 0,322 (korelasi positif), menunjukkan bahwa memiliki korelasi yang cukup terhadap status kelulusan. Mahasiswa perempuan cenderung lulus tepat waktu daripada mahasiswa laki-laki.

2. Umur

Umur menghasilkan nilai r sebesar -0,063 (korelasi negatif), menunjukkan bahwa umur memiliki hubungan terbalik dengan yang sangat lemah terhadap status kelulusan, namun tingkat korelasinya sangat rendah sehingga tidak memengaruhi ketepatan waktu kelulusan.

3. Jumlah Aktif

Jumlah aktif menghasilkan nilai r sebesar -0,112 (korelasi negatif), menunjukkan bahwa terdapat hubungan terbalik antara jumlah semester aktif

dan status kelulusan, namun tingkat korelasinya sangat lemah sehingga tidak memengaruhi ketepatan waktu kelulusan

4. Jumlah Cuti

Jumlah cuti menghasilkan nilai r sebesar $-0,040$ (korelasi negatif) menunjukkan bahwa jumlah cuti memiliki hubungan terbalik terhadap status kelulusan, namun tingkat korelasinya sangat rendah sehingga tidak memengaruhi ketepatan waktu kelulusan.

5. Jumlah Ambil TA

Jumlah ambil TA menghasilkan nilai r sebesar $0,636$ (korelasi positif) menandakan bahwa jumlah ambil TA memiliki korelasi yang kuat terhadap status kelulusan.

6. Jumlah MK Mengulang

Jumlah MK Mengulang menghasilkan nilai r sebesar $-0,401$ (korelasi negatif), menunjukkan bahwa jumlah MK mengulang memiliki hubungan terbalik terhadap status kelulusan dan nilai korelasinya termasuk kategori cukup. Mahasiswa yang jarang mengulang mata kuliah atau tidak mengulang sama sekali cenderung lulus tepat waktu.

7. IPP1 sampai IPP4

IPP1 hingga IPP4 menghasilkan nilai r dengan skala nilai $0,499-0,541$ (korelasi positif), menunjukkan bahwa nilai IPP memiliki korelasi cukup hingga kuat terhadap status kelulusan. Mahasiswa dengan nilai IPP yang tinggi cenderung lulus tepat waktu.

8. IPS1 sampai IPS6

IPS1 hingga IPS6 menghasilkan nilai r dengan skala nilai $0,463-0,580$ (korelasi positif), menandakan bahwa nilai IPS memiliki korelasi cukup hingga kuat terhadap status kelulusan. Mahasiswa dengan nilai IPS tinggi cenderung lulus tepat waktu.

9. IPK

IPK menghasilkan nilai r sebesar $0,687$ (korelasi positif), menandakan bahwa IPK memiliki korelasi kuat terhadap status kelulusan. Mahasiswa dengan IPK tinggi cenderung lulus tepat waktu.

Berdasarkan hasil perhitungan di atas maka dapat dilakukan proses seleksi atribut data yang akan digunakan untuk pemodelan dengan K-NN dengan memilih fitur yang memiliki nilai korelasi kuat sehingga seleksi atribut dapat dilihat pada Tabel 4. 2.

Tabel 4. 2 Seleksi Atribut Berdasarkan Nilai Korelasi Spearman

ATRIBUT	SPEARMAN
JUMLAH AMBIL TA	0,636
IPP2	0,541
IPP3	0,528
IPP4	0,522
IPS3	0,561
IPS4	0,520
IPS6	0,580
IPK	0,687

Pada atribut dengan nilai korelasi kuat terdapat atribut IPP2, IPP3, IPP4, IPS3, IPS4, dan IPS6, di mana atribut-atribut tersebut memiliki fungsi yang sama dengan IPK sama yaitu sebagai indeks prestasi dan memiliki nilai korelasi yang kuat sehingga terjadi multikolinearitas yang akan memengaruhi hasil prediksi pada saat pemodelan. Cara untuk mendeteksinya adalah dengan menggunakan metode (VIF) dalam model regresi linear pada IBM SPSS *Statistics* untuk menghitung nilai VIF sehingga didapat hasil perhitungan VIF pada Gambar 4. 1.

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-2.619	.346		-7.573	.000		
	IPP2	.019	.215	.022	.090	.928	.055	18.164
	IPP3	-.030	.223	-.032	-.132	.895	.054	18.592
	IPP4	-.273	.213	-.263	-1.280	.202	.076	13.095
	IPS3	.080	.079	.098	1.014	.312	.346	2.891
	IPS4	.033	.083	.041	.393	.695	.303	3.301
	IPS6	.022	.072	.027	.301	.764	.386	2.589
	IPK	1.054	.187	.718	5.631	.000	.198	5.049

a. Dependent Variable: STATUS KELULUSAN

Gambar 4. 1 Hasil Perhitungan VIF

Berdasarkan hasil perhitungan VIF, atribut IPP2, IPP3, dan IPP4 menghasilkan nilai $VIF > 10$ dan $Sig. > 0,05$ yang menunjukkan bahwa atribut tersebut dinyatakan multikolinearitas tinggi dan bersifat tidak signifikan. Atribut IPS3, IPS4, dan IPS6 memiliki nilai $VIF < 10$ dan $Sig. > 0,05$ yang menunjukkan bahwa atribut tersebut dinyatakan multikolinearitas rendah dan bersifat tidak signifikan. Atribut IPK memiliki nilai $VIF < 10$ dan $Sig. < 0,05$ yang menunjukkan bahwa atribut tersebut dinyatakan bebas multikolinearitas dan bersifat signifikan. Atribut dengan nilai $VIF < 10$ dan $Sig. < 0,05$ dipilih untuk menjadi variabel bebas pada saat pemodelan, maka dari itu IPK dipilih sebagai atribut untuk pemodelan dan didukung oleh nilai koefisien pada Korelasi *Spearman* tertinggi yaitu 0,687.

Setelah perhitungan VIF maka didapat hasil seleksi atribut untuk pemodelan K-NN berdasarkan nilai korelasi kuat dan nilai $VIF < 10$ sehingga hasilnya dapat dilihat pada Tabel 4. 3.

Tabel 4. 3 Seleksi Atribut Setelah Perhitungan VIF

ATRIBUT	SPEARMAN
JUMLAH AMBIL TA	0,636
IPK	0,687

Pada penelitian ini data yang diperoleh adalah data mahasiswa reguler S1 Sistem Informasi yang telah lulus pada tahun 2019 dan 2020. Mahasiswa yang telah lulus pada tahun tersebut merupakan mahasiswa angkatan 2010 sampai dengan 2016 sejumlah 193 data, dan data tersebut akan digunakan sebagai *data train* untuk melatih model K-NN. Sedangkan untuk *data test* adalah data mahasiswa reguler aktif S1 Sistem Informasi angkatan 2021 dan 2022 yang telah atau sedang menempuh semester 6 sampai dengan 8 sejumlah 84 data untuk semester 6 sejumlah 87 data untuk semester 8.

4.3 Data Preparation

Data mahasiswa yang diperoleh dan diseleksi atributnya di-*preprocess* kemudian dimasukkan ke dalam program untuk dianalisis. Data yang atributnya telah diseleksi pada tahap *Data Understanding* menghasilkan Tabel 4. 4 sebagai

data train semester 6 sejumlah 193 data, Tabel 4. 5 sebagai *data train* semester 8 sejumlah 193 data, Tabel 4. 6 sebagai *data test* semester 6 sejumlah 84 data, dan Tabel 4. 7 sebagai *data test* semester 8 sejumlah 87 data. *Data train* telah dilakukan proses pelabelan data oleh bagian PPTI dengan menggunakan Permendikbudristek No. 53 Tahun 2023 yang menetapkan masa tempuh kurikulum untuk mahasiswa sarjana selama 8 semester sebagai pedoman pelabelan.

Tabel 4. 4 *Data Train* Semester 6

NIM	Nama	Jumlah Ambil TA	IPK	Status Kelulusan
15410100147	MHS_107	0	3,28	Tepat Waktu
15410100115	MHS_108	0	2,70	Tidak Tepat Waktu
15410100184	MHS_109	0	3,40	Tepat Waktu
15410100264	MHS_110	0	3,35	Tidak Tepat Waktu
15410100266	MHS_111	0	3,47	Tepat Waktu

Tabel 4. 5 *Data Train* Semester 8

NIM	Nama	Jumlah Ambil TA	IPK	Status Kelulusan
15410100215	MHS_152	0	2,53	Tidak Tepat Waktu
15410100166	MHS_153	2	3,33	Tepat Waktu
15410100116	MHS_154	1	3,05	Tidak Tepat Waktu
15410100238	MHS_155	2	3,72	Tepat Waktu
15410100297	MHS_156	0	2,62	Tidak Tepat Waktu

Tabel 4. 6 *Data Test* Semester 6

NIM	Nama	Jumlah Ambil TA	IPK
22410100362	MHS_43	0	3,89
22410100195	MHS_44	0	3,48
22410100475	MHS_45	0	3,52
22410100421	MHS_46	0	2,53
22410100435	MHS_47	0	3,49

Tabel 4. 7 *Data Test* Semester 8

NIM	Nama	Jumlah Ambil TA	IPK
22410100379	MHS_50	0	2,98

22410100215	MHS_51	0	3,67
22410100480	MHS_52	0	3,00
22410100243	MHS_53	0	2,59
22410100324	MHS_54	0	2,90

Dataset diproses dengan cara mengecualikan NIM dan Nama, mentransformasi tipe data teks menjadi data numerik pada atribut STATUS KELULUSAN sehingga hasil *preprocess* menjadi Tabel 4. 8, dan Tabel 4. 9.

Tabel 4. 8 Hasil *Preprocessing Data Train Semester 6*

NIM	Nama	Jumlah Ambil TA	IPK	Status Kelulusan
15410100147	MHS_107	0	3,28	1
15410100115	MHS_108	0	2,70	0
15410100184	MHS_109	0	3,40	1
15410100264	MHS_110	0	3,35	0
15410100266	MHS_111	0	3,47	1

Tabel 4. 9 Hasil *Preprocessing Data Train Semester 8*

NIM	Nama	Jumlah Ambil TA	IPK	Status Kelulusan
15410100215	MHS_152	0	2,53	0
15410100166	MHS_153	2	3,33	1
15410100116	MHS_154	1	3,05	0
15410100238	MHS_155	2	3,72	1
15410100297	MHS_156	0	2,62	0

4.4 Modelling

Pada penelitian ini proses *modelling* menggunakan K-NN sebagai metode klasifikasi. Pertama nilai k ditentukan untuk menentukan jumlah data pada *data train* yang akan digunakan sebagai *neighbor* untuk *data test*, pada penelitian ini pemilihan nilai k dilakukan dengan menjalankan algoritma K-NN terhadap *data train* untuk membandingkan nilai akurasi masing-masing nilai k. Nilai k yang digunakan untuk perbandingan adalah k=3, 5, 7, 9, dan 11, kemudian nilai k dengan akurasi tertinggi akan dipilih untuk menjadi nilai k pada saat pemodelan. Hasil perhitungan nilai akurasi pada *data train* dengan masing-masing nilai k dapat dilihat pada Tabel 4. 10.

Tabel 4. 10 Tabel Nilai Akurasi Pada *Data Train* Berdasarkan Nilai K

Data	Neighbor				
	3	5	7	9	11
Semester 6	92,11%	86,84%	86,84%	84,21%	81,58%
Semester 8	94,74%	94,74%	94,74%	94,74%	94,74%

Berdasarkan tabel nilai akurasi, nilai $k=3$ memiliki nilai akurasi tertinggi sehingga dapat digunakan sebagai nilai k pada saat pemodelan dengan *data test*. *Data test* mahasiswa reguler aktif yang masih belum memiliki nilai pada kolom status kelulusan akan dihitung jaraknya dengan masing-masing data pada *data train* dengan rumus *Euclidean Distance*. Setelah jaraknya dihitung kemudian diurutkan dari jarak terkecil dan dipilih 3 data sebagai *neighbor*. Dari ketiga data tersebut dihitung berapa data dengan status kelulusan “Tepat Waktu” atau “Tidak Tepat Waktu”, kemudian dihitung mayoritas kelasnya sehingga status kelulusan pada *data test* dapat diisi berdasarkan kelas yang lebih banyak. Jika dari ketiga *neighbor* memiliki status kelulusan “Tepat Waktu” lebih banyak daripada “Tidak Tepat Waktu”, maka status kelulusan pada *data test* akan diisi “Tepat Waktu”, begitu juga sebaliknya. Contoh klasifikasi K-NN secara manual dijelaskan sebagai berikut.

Data Test

NIM: 21410100136, NAMA: MHS_3, JUMLAH AMBIL TA: 1, JUMLAH MK
ULANG: 0, IPK: 3,50

Neighbor 1

NIM: 15410100131, NAMA: MHS_96, JUMLAH AMBIL TA: 1, JUMLAH MK
ULANG: 0, IPK: 3,89, STATUS KELULUSAN: 1

$$euc = \sqrt{(1 - 1)^2 + (0 - 0)^2 + (3,50 - 3,89)^2}$$

$$euc = \sqrt{0 + 0 + 0,1521}$$

$$euc = 0,39$$

Neighbor 2

NIM: 15410100232, NAMA: MHS_104, JUMLAH AMBIL TA: 1, JUMLAH MK ULANG: 0, IPK: 3,77, STATUS KELULUSAN: 1

$$euc = \sqrt{(1 - 1)^2 + (0 - 0)^2 + (3,50 - 3,77)^2}$$

$$euc = \sqrt{0 + 0 + 0,0729}$$

$$euc = 0,27$$

Neighbor 3

NIM: 15410100268, NAMA: MHS_140, JUMLAH AMBIL TA: 1, JUMLAH MK ULANG: 0, IPK: 3,51, STATUS KELULUSAN: 1

$$euc = \sqrt{(1 - 1)^2 + (0 - 0)^2 + (3,50 - 3,51)^2}$$

$$euc = \sqrt{0 + 0 + 0,0001}$$

$$euc = 0,01$$

Dari ketiga *neighbor* yang telah dihitung jaraknya semua memiliki status kelulusan bernilai 1 yang berarti “Tepat Waktu” sehingga dengan pemilihan mayoritas dapat ditentukan kelas pada *data test* adalah 1 atau “Tepat Waktu”.

Pemodelan sistematis dijalankan dengan menggunakan *Python* GUI sebagai alat bantu untuk mengimpor data, pemodelan data, evaluasi data, dan visualisasi data, dan setiap tahapnya diuraikan pada sub-bab berikut.

4.4.1 Proses *Import* Data

Data Train					
NIM	NAMA	JUMLAH AMBIL TA	IPK	STATUS KELULUSAN	
10410100155	MHS_1	0	2,49	0	
11410100172	MHS_2	0	2,41	0	
11410100116	MHS_3	0	3,38	0	
11410100156	MHS_4	0	3,13	0	
11410100146	MHS_5	0	2,21	0	
11410100186	MHS_6	0	2,92	0	
12410100104	MHS_7	0	2,69	0	
12410100272	MHS_8	0	3,37	0	
12410100250	MHS_9	0	2,87	0	
12410100125	MHS_10	0	2,59	0	

Gambar 4. 2 *Import* Data Train

Pada program yang telah dibuat dengan *Python* GUI terdapat fitur untuk mengimpor data dengan format file excel dengan menggunakan library Pandas, dimana Pandas memunculkan dialog untuk memilih file excel, kemudian file tersebut dimasukkan ke dalam program dengan menggunakan *source code* pada Gambar L1. 1 dan ditampilkan ke dalam bentuk tabel dengan menggunakan Treeview dengan *source code* pada Gambar L1. 3.

Data yang ditampilkan dari Data Train adalah NIM, NAMA, JUMLAH AMBIL TA, IPK, dan STATUS KELULUSAN dengan penjelasan kamus data sebagai berikut:

1. JUMLAH AMBIL TA: tipe data integer dengan nilai minimal 0.
2. IPK: tipe data float dengan nilai minimal 0.
3. STATUS KELULUSAN: Status kelulusan mahasiswa dengan nilai 1 untuk “Tepat Waktu” dan 0 untuk “Tidak Tepat Waktu”.

Setelah mengimpor *data train* kemudian mengimpor *data test* dengan format file excel ke dalam program dengan *source code* Gambar L1. 2. Isi *data test* yang telah diimpor ditampilkan ke dalam tabel dengan menggunakan Treeview pada *source code* Gambar L1. 4. Isi tabel pada *data test* antara lain NIM, NAMA, JUMLAH AMBIL TA, dan IPK. STATUS KELULUSAN tidak ditampilkan terlebih dahulu karena masih belum memiliki nilai dan akan diprediksi menggunakan K-NN.

Data Test				
NIM	NAMA	JUMLAH AMBIL TA	IPK	
21410100108	MHS_1	0	2.12	
21410100223	MHS_2	0	3.5	
21410100136	MHS_3	0	3.44	
21410100309	MHS_4	0	3.66	
21410100331	MHS_5	0	2.51	
21410100101	MHS_6	0	2.63	
21410100295	MHS_7	0	3.27	
21410100253	MHS_8	0	3.29	
21410100399	MHS_9	0	3.45	
21410100397	MHS_10	0	3.45	

Gambar 4. 3 *Import Data Test*

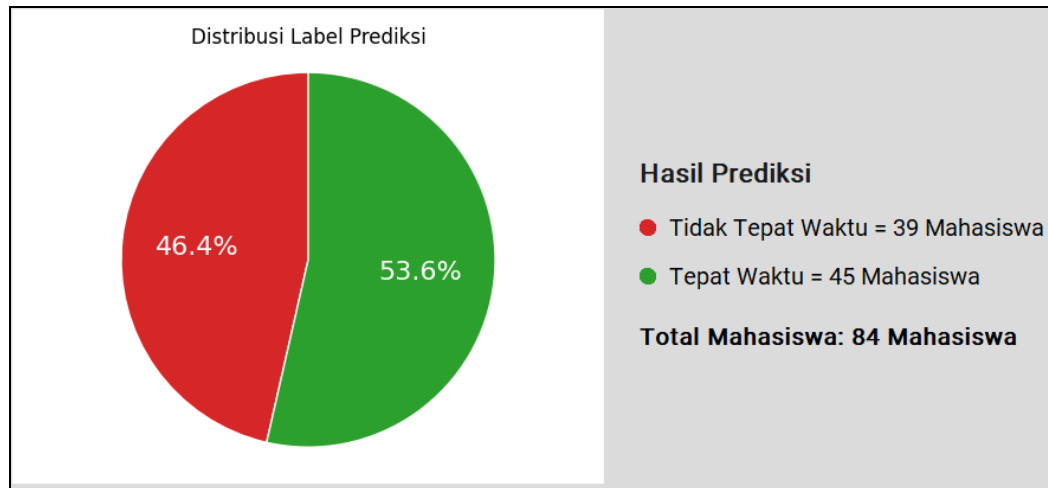
4.4.2 Prediksi pada Data Test

Data test yang telah diprediksi dengan menggunakan model K-NN akan mengisi atribut STATUS KELULUSAN yang kosong dengan label yang telah diprediksi dengan menampilkan nilai “Tepat Waktu” dan “Tidak Tepat Waktu, dan tampilan pada tabel *data test* akan ter-*update* dengan tampilan baru dengan kolom STATUS KELULUSAN sebagai kolom label. *Source code* pemodelan K-NN dapat dilihat pada Gambar L1. 5.

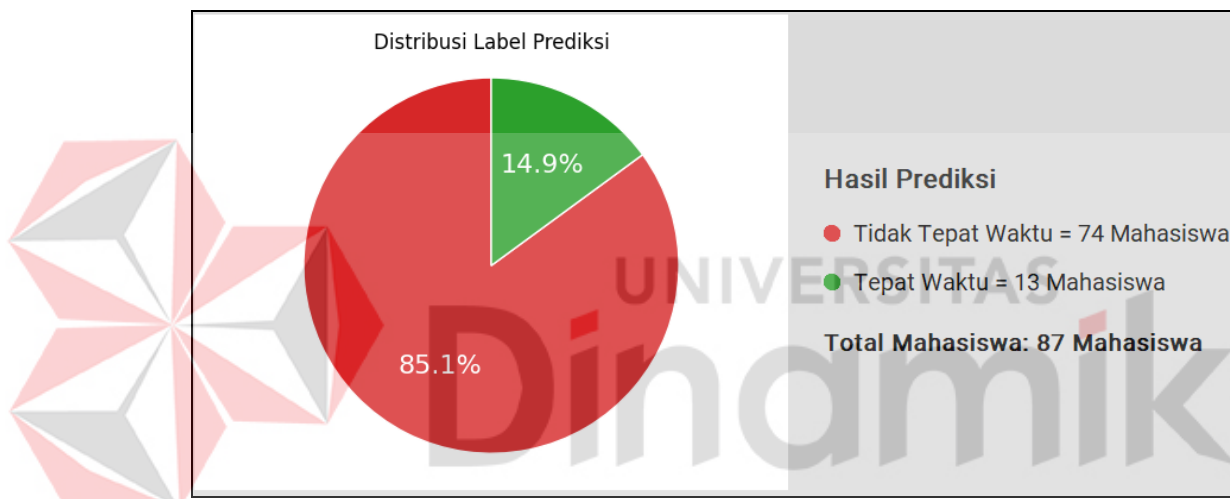
Data Test (Prediksi)						
NIM	NAMA	JUMLAH AMBIL TA	IPK	STATUS KELULUSAN		
21410100376	MHS_19	0	2.96	Tidak Tepat Waktu		
21410100494	MHS_20	0	3.25	Tidak Tepat Waktu		
21410100308	MHS_21	0	1.73	Tidak Tepat Waktu		
21410100416	MHS_22	0	2.7	Tidak Tepat Waktu		
21410100335	MHS_23	0	3.75	Tepat Waktu		
21410100471	MHS_24	0	3.63	Tepat Waktu		
21410100359	MHS_25	0	3.66	Tidak Tepat Waktu		
21410100372	MHS_26	0	3.7	Tepat Waktu		
21410100390	MHS_27	0	3.53	Tepat Waktu		
21410100408	MHS_28	0	3.84	Tepat Waktu		

Gambar 4. 4 Hasil Prediksi *Data Test*

Hasil prediksi data test dapat ditampilkan ke bentuk *pie chart* untuk menunjukkan persentase distribusi label prediksi “Tepat Waktu” dan “Tidak Tepat Waktu”. Berdasarkan *pie chart*, hasil prediksi terhadap 84 mahasiswa reguler aktif S1 Sistem Informasi pada data semester 6 menunjukkan bahwa sejumlah 45 mahasiswa (53,6%) diprediksi lulus tepat waktu dan 39 mahasiswa (46,4%) diprediksi lulus tidak tepat waktu. Sedangkan untuk data semester 8 menunjukkan bahwa dari 87 mahasiswa terdapat 13 mahasiswa (14,9%) diprediksi lulus tepat waktu dan 74 mahasiswa (85,1%) diprediksi lulus tidak tepat waktu. Tampilan *pie chart* data semester 6 dan 8 dapat dilihat pada Gambar 4. 5 dan Gambar 4. 6.

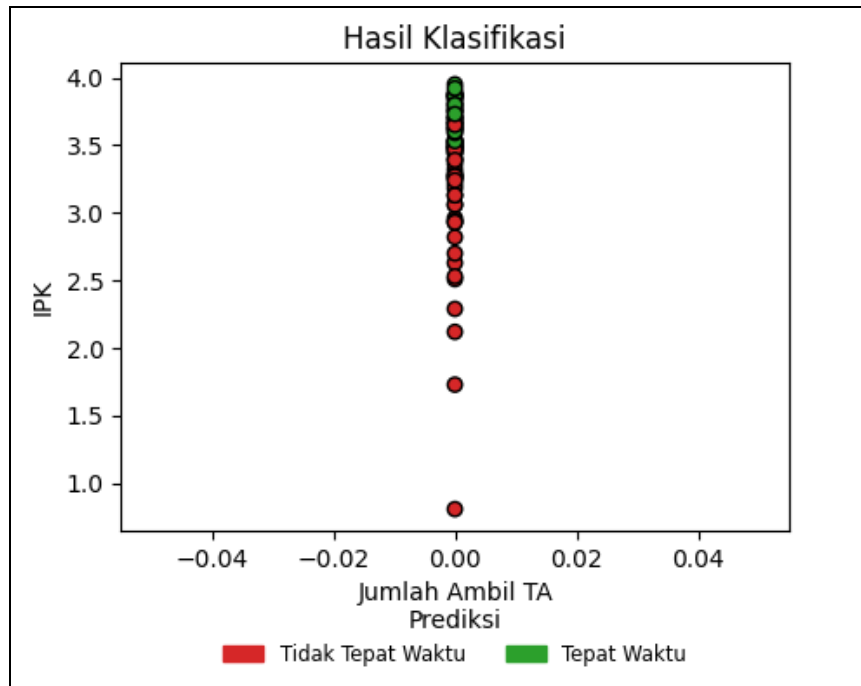


Gambar 4. 5 *Pie Chart* Hasil Prediksi Data Semester 6

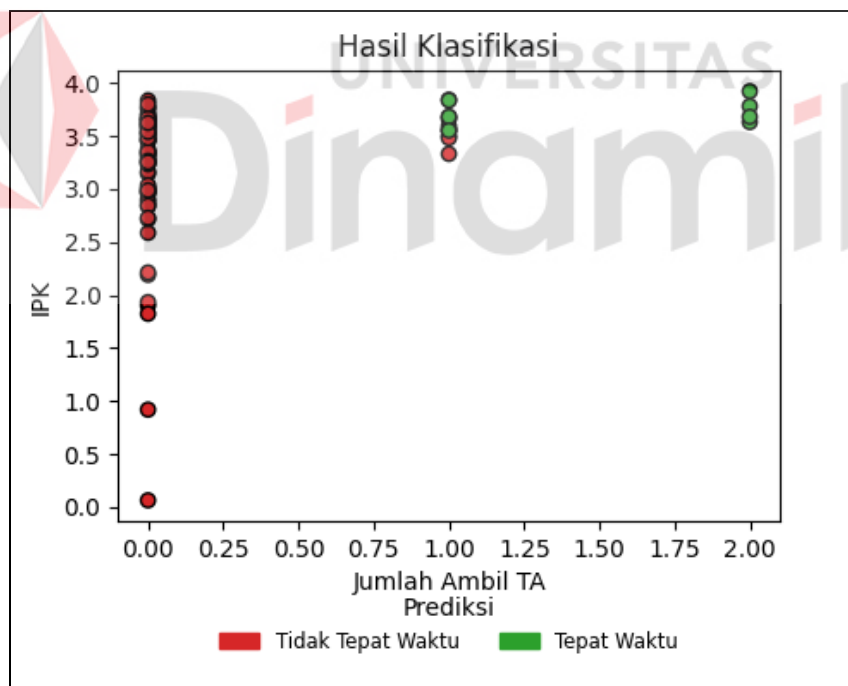


Gambar 4. 6 *Pie Chart* Hasil Prediksi Data Semester 8

Hasil prediksi juga divisualisasikan ke bentuk *scatter plot* untuk menampilkan titik data test pada sumbu X dan Y dengan indikator warna hijau untuk data dengan label “Tepat Waktu” dan merah untuk label “Tidak Tepat Waktu” sehingga tampilan mudah dibaca. *Scatter plot* data semester 6 dan 8 dapat dilihat pada Gambar 4. 7.



Gambar 4. 7 Scatter Plot Data Semester 6



Gambar 4. 8 Scatter Plot Data Semester 8

4.5 Evaluation

Proses *Modelling* yang telah dilakukan menghasilkan *confusion matrix* yang menampilkan jumlah data yang diprediksi sebagai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Nilai akurasi dihitung dengan menghitung jumlah data TP dan TN dibagi total data dengan menggunakan perhitungan excel dan source code pada Gambar L1. 8. Oleh karena itu dibutuhkan data mahasiswa dengan label aktual untuk dibandingkan dengan label prediksi. Salah satu caranya adalah dengan menggunakan data alumni yang telah lulus pada semester 24.1 dan 24.2 sejumlah 100 data sehingga nilai akurasi dapat dihitung.

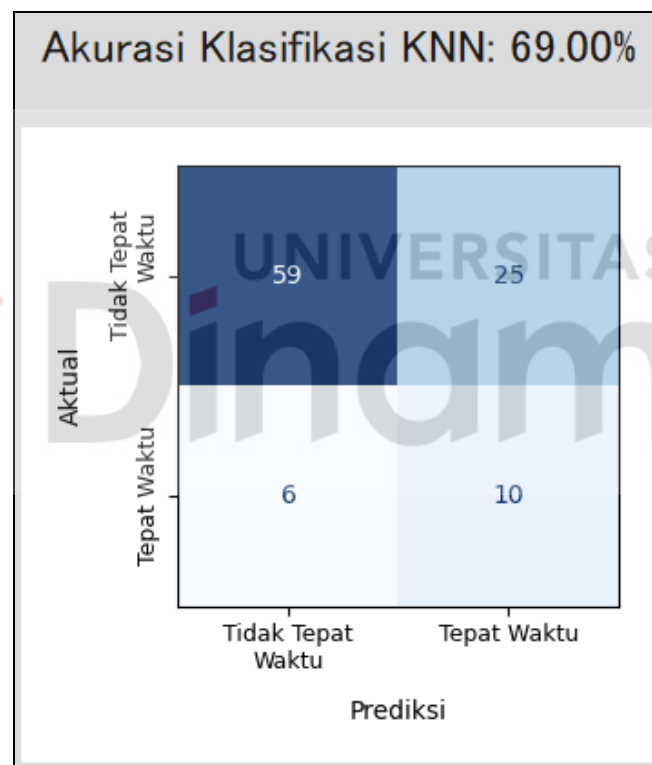
Berdasarkan perhitungan excel menghasilkan nilai akurasi sebesar 72% pada data alumni semester 6 ditampilkan pada Gambar 4. 9 dan 85% pada data alumni semester 8 ditampilkan pada Gambar 4. 10. Sedangkan pada perhitungan *Python* menghasilkan nilai akurasi pada data alumni semester 6 sebesar 69% ditampilkan pada Gambar 4. 11 dan nilai akurasi pada data alumni semester 8 sebesar 85% ditampilkan pada Gambar 4. 12.

		Prediksi	
		Tidak Tepat Waktu	Tepat Waktu
Aktual	Tidak Tepat Waktu	64	20
	Tepat Waktu	8	8
Akurasi		72,00%	

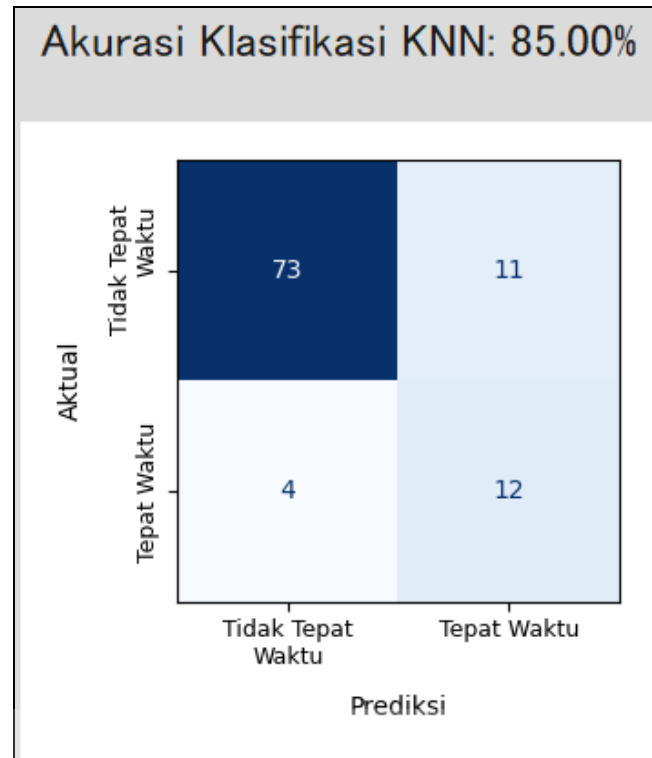
Gambar 4. 9 Nilai Akurasi dan *Confusion Matrix* Perhitungan Excel Data Alumni Semester 6

		Prediksi	
		Tidak Tepat Waktu	Tepat Waktu
Aktual	Tidak Tepat Waktu	73	11
	Tepat Waktu	4	12
Akurasi		85,00%	

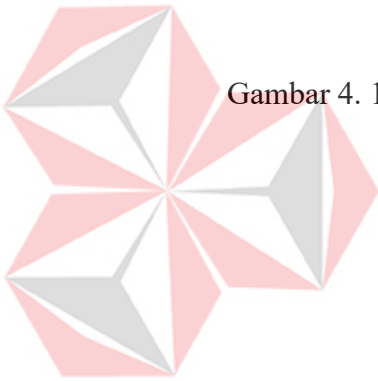
Gambar 4. 10 Nilai Akurasi dan *Confusion Matrix* Perhitungan Excel Data Alumni Semester 8



Gambar 4. 11 Nilai Akurasi dan *Confusion Matrix* Perhitungan Python Data Alumni Semester 6



Gambar 4. 12 Nilai Akurasi dan *Confusion Matrix* Perhitungan *Python Data* Alumni Semester 8



UNIVERSITAS
Dinamika

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dari penerapan algoritma K-NN untuk memprediksi ketepatan waktu kelulusan mahasiswa S1 Sistem Informasi, maka dapat dibuat kesimpulan bahwa metode yang digunakan untuk memprediksi ketepatan waktu kelulusan mahasiswa yaitu K-NN telah berfungsi dengan benar dan mampu memprediksi ketepatan waktu kelulusan mahasiswa serta menampilkan label prediksi pada tabel *data test*, menampilkan distribusi label prediksi dengan menggunakan *pie chart*, dan menampilkan titik *data test* dengan menggunakan *scatter plot*.

Visualisasi *pie chart* Pada *data test* semester 6 menghasilkan prediksi sejumlah 45 mahasiswa diprediksi lulus tepat waktu dan 39 mahasiswa diprediksi lulus tidak tepat waktu. Sedangkan pada data semester 8 menghasilkan prediksi sejumlah 13 mahasiswa diprediksi lulus tepat waktu dan 74 mahasiswa diprediksi lulus tidak tepat waktu. Berdasarkan hasil prediksi yang ada, maka sekretaris prodi dapat memberikan arahan kepada dosen wali terhadap mahasiswa yang diprediksi lulus tidak tepat waktu. Berdasarkan hasil pemodelan K-NN pada data alumni lulusan semester 24.1 dan 24.2 dengan total data 100 mahasiswa, menghasilkan nilai akurasi dengan perhitungan excel sebesar 72 % untuk data semester 6 dan 85% untuk data semester 8. Sedangkan nilai akurasi dengan perhitungan *Python* menghasilkan nilai 69% untuk data semester 6 dan 85% untuk data semester 8.

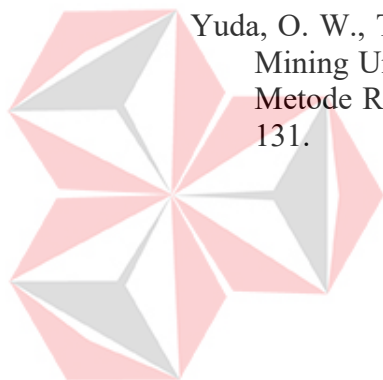
5.2 Saran

Berdasarkan hasil penelitian, saran yang dari penulis dapat dilakukan pada penelitian selanjutnya adalah menggunakan metode korelasi lain untuk proses seleksi atribut sehingga dapat menghasilkan pemilihan atribut dan hasil prediksi yang berbeda.

DAFTAR PUSTAKA

- Candra, A. P. (2025). Analisis Data Menggunakan Python: Memperkenalkan Pandas dan NumPy. *Journal of Education System and Education Development (JISED)*, 3(1), 11–16.
- Fadia, S., & Fitri, N. (2021). Problematika Kualitas Pendidikan di Indonesia. *Jurnal Pendidikan Tambusai*, 5(1), 1617.
- Gustini, S., Umardani, D., Emelda, & Calista, A. (2023). Membangun Motivasi Pelajar Untuk Melanjutkan Pendidikan Ke Jenjang Perguruan Tinggi. *NUSANTARA: Jurnal Pengabdian Kepada Masyarakat*, 3(2), 117–125.
- Kartarina, Sriwinarti, N. K., & Juniarti, N. luh P. (2021). Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 3(2), 106–112.
- Lestari, P. I., & Andriansyah, M. (2021). Analisis K-Nearest Neighbor Berdasarkan Forward Selection Untuk Prediksi Status Mahasiswa Non Aktif Pada STMIK Bani Saleh. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 6(3), 181–186.
- Manullang, R. A., & Sianturi, F. A. (2021). Penerapan Algoritma K-Nearest Neighbour Untuk Memprediksi Kelulusan Mahasiswa. *JIKOMSI [Jurnal Ilmu Komputer dan Sistem Informasi]*, 4(2), 42–50.
- Murti, E. W. (2024). Analisis Dan Perbandingan Algoritma Prediksi Dalam Mengetahui Perkiraan Peningkatan Jumlah Kasus COVID-19 Di Kabupaten Boyolali Dengan Metodologi CRISP-DM. *JIKES: Jurnal Ilmu Kesehatan*, 3(1), 24–34.
- Pamungkasih, P. (2023a). Penerapan Uji Korelasi Rank Spearman Untuk Mengetahui Hubungan Pengeluaran Rumah Tangga Untuk Makanan Dan Tingkat Kemiskinan Di Nusa Tenggara Timur Selama Pandemi COVID-19. *JSTAR*, 3(2), 1–2. <https://doi.org/10.5300/JSTAR.V3I02.49>
- Pamungkasih, P. (2023b). Penyedia Data Statistik Berkualitas untuk Indonesia Maju REFORMASI BIROKRASI. *JSTAR*, 3(2), 1–2. <https://doi.org/10.5300/JSTAR.V3I02.49>
- Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmaddeni, & Efrizoni, L. (2024a). Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia. *Institut Riset dan Publikasi Indonesia (IRPI)*, 4(1), 273–281.
- Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmaddeni, R., & Efrizoni, L. (2024b). Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 273–281. <https://doi.org/10.57152/malcom.v4i1.1085>

- Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., & Akbar, M. (2021). Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(3), 987–994.
- Rembangsupu, A., Budiman, K., Bidin, Puspita, & Rangkuti, M. Y. (2022). Studi Yuridis Tentang Jenis Dan Jalur Pendidikan Di Indonesia. *AL-AFKAR: Journal For Islamic Studies*, 5(4), 91–100.
- Riadi, I., Umar, R., & Anggara, R. (2024). Prediksi Kelulusan Tepat Waktu Berdasarkan Riwayat Akademik Menggunakan Metode K-Nearest Neighbor. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 11(2), 249–256.
- Said, H., Matondang, N., Nurramdhani Irmanda, H., & Informasi, S. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi. *Techno.COM*, 21(2), 256–267.
- Sundus, K. I., Hammo, B. H., Al-Zoubi, M. B., & Al-Omari, A. (2022). Solving the multicollinearity problem to improve the stability of machine learning algorithms applied to a fully annotated breast cancer dataset. *Informatics in Medicine Unlocked*, 33. <https://doi.org/10.1016/j.imu.2022.101088>
- Yuda, O. W., Tuti, D., Yee, L. S., & Susanti. (2022). Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest. *SATIN - Sains dan Teknologi Informasi*, 8(2), 122–131.



UNIVERSITAS
Dinamika