

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

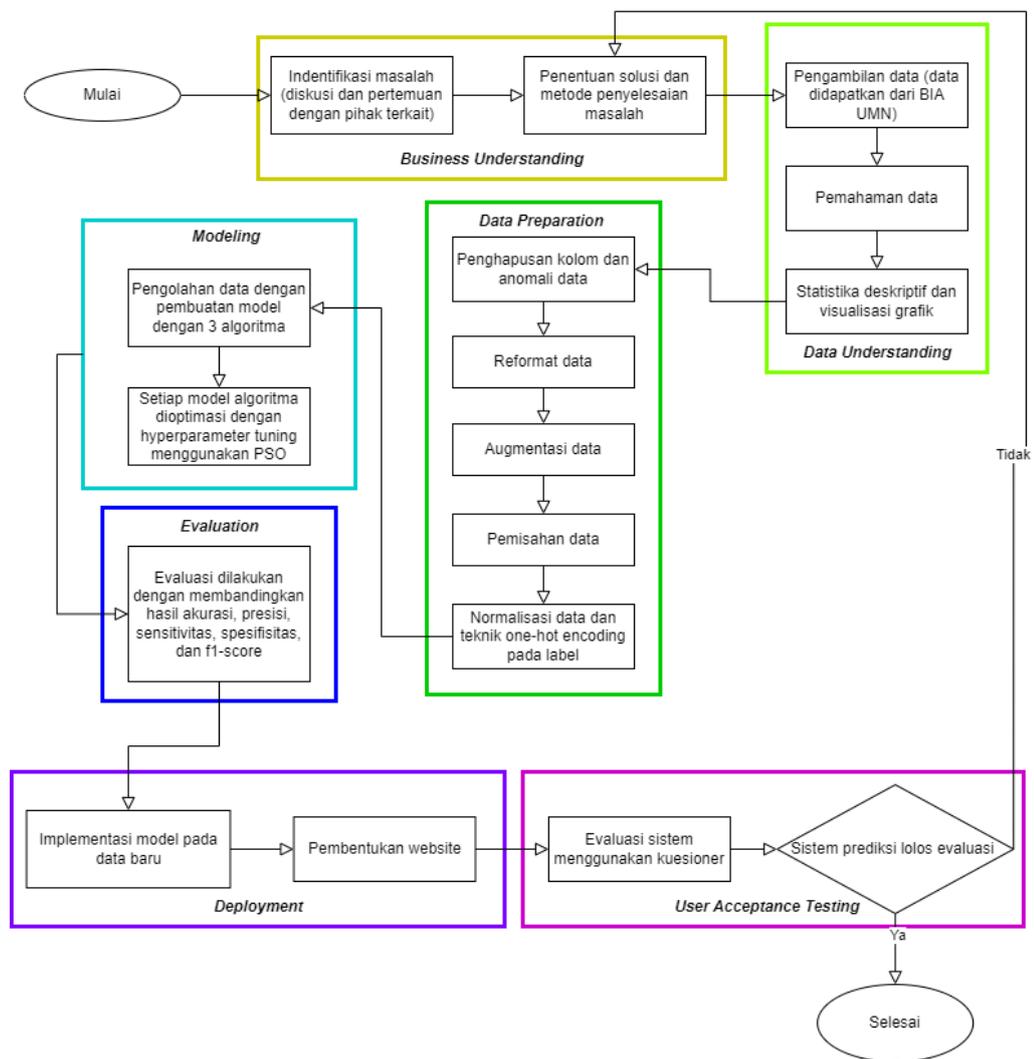
Penelitian ini dilakukan untuk melakukan prediksi waktu kelulusan mahasiswa dari Program Studi Sistem Informasi Universitas Multimedia Nusantara. Universitas Multimedia Nusantara, sebagai salah satu perguruan tinggi swasta, memiliki 16 program studi untuk program strata 1 dan program strata 2. Objek penelitian ini akan berfokus pada Program Studi Strata 1 Sistem Informasi dikarenakan terdapat perbedaan ketentuan lini masa studi perkuliahan pada program studi lainnya.

Data mahasiswa yang digunakan untuk penelitian ini didapatkan dari Biro Informasi Akademik Universitas Multimedia Nusantara yang telah diajukan pada Januari 2023 seperti yang ditunjukkan pada Lampiran B Surat Perizinan Pengambilan Data dan diperoleh pada Februari 2023, seperti yang ditunjukkan pada Lampiran C Surat Pengambilan Data. Data yang digunakan pada penelitian ini yaitu data mahasiswa tahun angkatan 2014-2023 Program Studi Sistem Informasi di Universitas Multimedia Nusantara. Data tersebut akan menjadi acuan pembelajaran model klasifikasi dimana model klasifikasi tersebut akan digunakan untuk memprediksi waktu kelulusan mahasiswa yang berstatus aktif.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Alur atau tahapan yang dilakukan pada penelitian ini ditunjukkan pada Gambar 3.1. Alur penelitian digunakan sebagai acuan untuk membantu proses pengerjaan proyek menjadi lebih terarah. Penelitian dimulai dengan mengidentifikasi masalah yang terjadi yaitu lulus dengan terlambat hingga terjadinya putus kuliah dapat memberikan berbagai dampak negatif yang tidak hanya dirasakan oleh pihak mahasiswa, namun juga oleh pihak kampus.



Gambar 3.1 Alur Penelitian

Berdasarkan permasalahan yang telah diidentifikasi, solusi yang ditawarkan adalah dengan adanya pembuatan sistem prediksi waktu kelulusan mahasiswa yang dapat digunakan sebagai acuan para dosen pembimbing untuk melakukan penyesuaian dalam memberikan arahan atau bimbingannya kepada mahasiswa, serta memberikan *awareness* bagi pihak mahasiswa terkait prediksi waktu kelulusannya. Metode yang digunakan pada penelitian ini adalah metode *data mining* dengan penggunaan *framework* CRISP-DM. Data sebagai penunjang penelitian didapatkan dari Biro Informasi Akademik Universitas Multimedia Nusantara. Data yang diberikan terbagi menjadi 2 kelompok yaitu data mahasiswa yang berstatus lulus dan aktif. Rentang periode data ini dipilih karena alasan relevansi, data mahasiswa berstatus aktif memiliki rentang tahun

angkatan 2017 hingga 2023, tahun 2017 merupakan tahun angkatan terakhir yang dapat lulus pada masa pengerjaan penelitian ini dan tahun angkatan 2023 merupakan tahun angkatan terbaru pada masa pengerjaan penelitian ini. Data mahasiswa berstatus lulus memiliki rentang tahun angkatan 2014 hingga 2023, hal ini dikarenakan model butuh mempelajari pola mahasiswa yang lulus 14 semester. Data yang digunakan pada penelitian ini akan dipersiapkan melalui 5 proses yaitu penghapusan kolom dan anomali data, *reformat* data, augmentasi data, pemisahan data, dan normalisasi data serta penggunaan teknik *one-hot encoding* pada label. Pengolahan data dilakukan dengan cara pembuatan model prediksi berdasarkan klasifikasi dengan membandingkan 3 algoritma yaitu Naïve Bayes, Support Vector Machine, dan Neural Network yang akan dioptimasi (*hyperparameter tuning*) menggunakan algoritma Particle Swarm Optimization. Hasil ketiga algoritma tersebut akan dibandingkan menggunakan metrik akurasi, presisi, sensitivitas, spesifisitas, dan *f1-score*. Hasil algoritma terbaik akan digunakan sebagai model untuk melakukan pengimplementasian prediksi waktu kelulusan mahasiswa dengan menggunakan data mahasiswa berstatus aktif. Hasil prediksi akan ditampilkan pada *website*. Sistem prediksi dan *website* akan diuji dengan *User Acceptance Testing* menggunakan penyebaran kuesioner. *User Acceptance Testing* dilakukan untuk menganalisa apakah sistem yang dibuat sudah menjawab permasalahan yang diidentifikasi. Sistem dinyatakan lolos dari evaluasi UAT jika sistem prediksi mendapatkan penilaian 4 atau baik dalam *rating* secara keseluruhan dari sisi pengguna mahasiswa maupun staf akademik atau dosen pembimbing.

3.2.2 Metode Data Mining

Penelitian ini terdapat dua metodologi yang akan digunakan yakni, metodologi pemahaman data dan metodologi pemecahan masalah. Berikut merupakan tabel perbandingan *framework* yang digunakan sebagai metodologi pemahaman data, yaitu CRISP-DM, SEMMA, dan KDD:

Tabel 3.1 Perbandingan CRISP-DM, SEMMA, dan KDD

Indikator	CRISP-DM [29]	SEMMA [31]	KDD [30]
Tingkat kompleksitas	Tahapan <i>framework</i> CRISP-DM merupakan adaptasi dan penyederhanaan dari tahapan <i>framework</i> KDD.	Tahapan <i>framework</i> SEMMA paling sederhana jika dibandingkan dengan kedua <i>framework</i> lainnya.	Tahapan <i>framework</i> KDD merupakan tahapan yang paling lengkap namun paling kompleks dibandingkan dengan kedua <i>framework</i> lainnya.
Langkah-langkah	<ol style="list-style-type: none"> 1. <i>Business understanding</i> 2. <i>Data understanding</i> 3. <i>Data preparation</i> 4. <i>Modeling</i> 5. <i>Evaluation</i> 6. <i>Deployment</i> 	<ol style="list-style-type: none"> 1. <i>Sample</i> 2. <i>Explore</i> 3. <i>Modify</i> 4. <i>Model</i> 5. <i>Assessment</i> 	<ol style="list-style-type: none"> 1. <i>Understanding the application area</i> 2. <i>Extracting the target data set</i> 3. <i>Data cleaning and preprocessing</i> 4. <i>Data integration</i> 5. <i>Data reduction and projection</i> 6. <i>Choosing the data mining algorithm(s)</i> 7. <i>Data mining</i> 8. <i>Interpretation</i> 9. <i>Using discovered knowledge</i>

Mengacu pada Tabel 3.1, penelitian ini akan menggunakan *framework* CRISP-DM sebagai metodologi pemahaman data, dimana pemilihan metodologi ini didasari dari penelitian-penelitian sebelumnya. *Framework* CRISP-DM merupakan bentuk adaptasi dan penyederhanaan dari *framework* KDD, namun lebih komprehensif dari *framework* SEMMA[31]. *Framework* CRISP-DM juga dipilih sebagai *framework* yang dipakai pada penelitian ini karena setiap langkah atau fase yang dilakukan merupakan langkah yang sesuai dengan penelitian ini, dengan contoh tidak adanya integrasi data seperti yang dilakukan pada tahap *framework* KDD. Di sisi lain, *framework* SEMMA juga kurang sesuai karena tidak adanya tahap identifikasi objektif penelitian seperti yang dilakukan pada tahap *business understanding* pada *framework* CRISP-DM

Berikut merupakan langkah-langkah atau tahapan-tahapan dari *framework* CRISP-DM yang akan diterapkan pada penelitian ini:

3.2.2.1 *Business Understanding*

Business Understanding merupakan tahapan pertama dalam metode penelitian. Pada tahapan ini peneliti akan menganalisis dan mengetahui

lebih dalam tentang tujuan dan latar belakang dari berjalannya proyek yang akan dilakukan [27]. Dalam hal ini, latar belakang atau tujuan dari dilakukannya proyek ini yaitu pembuatan sistem prediksi waktu kelulusan mahasiswa sebagai salah satu solusi terhadap keterlambatan waktu kelulusan hingga putus kuliah dengan upaya meningkatkan kesadaran mahasiswa serta sebagai salah satu acuan bagi dosen pembimbing maupun staf akademik untuk menyesuaikan pemberian bimbingan akademik kepada para mahasiswa bimbingan. Hasil prediksi merupakan waktu kelulusan mahasiswa dalam satuan semester. Algoritma yang digunakan adalah Naïve Bayes, Support Vector Machine, dan Neural Network karena ketiga algoritma tersebut memiliki performa yang baik pada penelitian sebelumnya.

3.2.2.2 Data Understanding

Data Understanding merupakan tahapan untuk memahami dan mengenal data lebih dalam dan detail [29]. Penelitian ini akan menggunakan data mahasiswa Universitas Multimedia Nusantara dimana periode data yang digunakan yaitu data dari tahun 2014-2023 yang berjumlah 9034 baris dan 16 kolom. Data yang digunakan merupakan data tabular. Variabel-variabel yang akan digunakan yaitu indeks prestasi kumulatif (IPK) per semester, indeks prestasi semester (IPS) per semester, semester yang ditempuh saat ini, jumlah SKS yang diambil per semester, jumlah SKS yang lulus per semester, jumlah SKS yang tidak lulus per semester, yotal satuan kredit semester (SKS) yang telah diambil, total satuan kredit semester (SKS) yang telah lulus, dan keikutsertaan semester antara. Jumlah populasi data pada penelitian ini adalah mahasiswa Universitas Multimedia Nusantara yang berjumlah 11012 mahasiswa, sedangkan sampel data yang digunakan merupakan mahasiswa Jurusan Sistem Informasi Universitas Multimedia Nusantara yang berjumlah 1013 mahasiswa. Validasi sistem prediksi konsep *User Acceptance Testing* dilakukan dengan menggunakan teknik *simple random sampling* dengan populasi berjumlah 844 mahasiswa dan sampel berjumlah 72 mahasiswa. Pada tahap *data understanding* juga

dilakukan EDA atau *exploratory data analysis* untuk memahami dan mengidentifikasi data lebih lanjut sebelum memasuki tahap *data preparation* yang dengan 2 cara yaitu statistika deskriptif untuk variabel numerikal dan visualisasi dengan grafik untuk variabel kategorikal.

3.2.2.3 *Data Preparation*

Data preparation merupakan tahap dimana data akan dipersiapkan secara terstruktur dan terorganisir agar dapat digunakan dengan tepat dan sesuai di tahap selanjutnya [56]. Pada penelitian ini, data akan dipersiapkan melalui berbagai proses yaitu penghapusan kolom dan anomali data, *reformat* data, augmentasi data, pemisahan data, dan normalisasi data serta penggunaan teknik *one-hot encoding* pada label. *Data cleansing* dilakukan untuk merapihkan struktur data agar data layak pakai [57] dimana metode *data cleansing* berlaku pada proses penghapusan kolom dan anomali data, serta *reformat* data pada penelitian ini. Beberapa hal yang dilakukan dalam proses pertama, yakni penghapusan kolom-kolom data yang tidak memiliki informasi yang relevan pada pelatihan model, serta penanganan anomali data dimana terdapat dalam data mahasiswa yang telah lulus terdapat 1 tambahan semester yang tercatat dari total semester yang seharusnya. *Reformat* data dilakukan karena data yang diperoleh menampilkan setiap baris mewakili hasil studi mahasiswa pada semester tertentu. Data diolah sehingga format yang digunakan menjadi data dengan setiap baris mewakili setiap individu mahasiswa yang berbeda dengan menggabungkan semua data performa mahasiswa dari semester pertama hingga semester lulus. Tahap selanjutnya merupakan augmentasi data mengadopsi konsep *expanding window*. *Expanding window* merupakan salah satu teknik augmentasi yang umumnya digunakan untuk permasalahan *time series* dimana data akan dipisah dan akan diakumulasi untuk terus bertambah seiring dengan berjalannya waktu [58]. Augmentasi data yang dilakukan adalah menduplikasi data sebanyak 20 kali dimana setiap baris merupakan representasi performa mahasiswa yang diakumulasi dari semester 1 hingga 14, hal ini dilakukan untuk menambahkan entitas baru agar model dapat

menemukan pola tertentu dalam melakukan prediksi. Langkah selanjutnya yang dilakukan merupakan penanganan nilai *null*, hal ini dilakukan dengan menggunakan teknik *data imputation* dan kolom indikator. Setelah itu akan dilakukan *split data* menjadi data *training* dan *testing* dimana *split data* ini akan dilakukan untuk memvalidasi dan mendukung tahap pemodelan. Proses dilanjutkan dengan pemisahan data menggunakan metode *k-fold cross validation* yang terstratifikasi, hal ini dikarenakan *stratified k-fold cross validation* dapat mengatasi permasalahan data yang tidak seimbang dengan mendistribusikan setiap kelas secara merata menggunakan rasio ke setiap lipatan [59]. Langkah terakhir yang dilakukan adalah melakukan normalisasi data serta penggunaan teknik *one-hot encoding* pada label. Normalisasi data dilakukan agar nilai data memiliki rata rata sebesar 0 dan varian sebesar 1 dengan menggunakan fungsi *StandarScaler()*, serta teknik *one-hot encoding* dilakukan agar label tidak memiliki nilai 7 hingga 14 yang mewakili semester lulus, melainkan nilai antara 1 atau 0 [60].

3.2.2.4 Modeling

Pada tahapan ini, akan dilakukan pemodelan dengan menggunakan data yang telah diolah sebelumnya dengan berbagai algoritma *data mining* [61]. Pemodelan yang dilakukan pada penelitian ini yaitu pembuatan model prediksi dengan jenis klasifikasi dengan membandingkan 3 algoritma yaitu Neural Network, Naïve Bayes, dan Support Vector Machine. Pelatihan model dilakukan pada *software* Visual Studio Code dengan bahasa pemrograman Python. Ketiga algoritma tersebut juga akan di optimasi menggunakan algoritma PSO (*Particle Swarm Optimization*) agar perbandingan yang dilakukan dari performa algoritma yang optimal, bukan dari parameter secara *default* atau parameter ditentukan secara subjektif. Hasil performa dari ketiga algoritma tersebut akan dibandingkan dan dipilih pada tahap berikutnya yaitu *evaluation*. Model prediksi dapat terus dilakukan pembaruan dengan data baru mahasiswa agar pola dari data dapat lebih relevan.

3.2.2.5 Evaluation

Tahapan ini merupakan tahapan saat pengolahan data dapat menguji dan mengetahui performa model yang telah dibuat. Di tahap ini, peneliti dapat menentukan apakah model yang telah dibentuk dapat diteruskan dan diimplementasikan pada tahap *deployment* [29]. Hasil performa model pada tahap sebelumnya akan dibandingkan dengan beberapa metrik yang telah ditetapkan yaitu akurasi, presisi, sensitivitas, spesifisitas, *f1-score*, dan *confusion matrix*. Model dengan performa terbaik akan dipilih dan digunakan langsung dengan mengimplementasikannya pada data mahasiswa yang masih aktif.

Evaluasi juga dilakukan pada sistem prediksi yang telah diimplementasikan di *website*. Evaluasi dilakukan konsep *User Acceptance Testing* menggunakan metode penyebaran kuesioner dengan skala Likert. Pertanyaan kuesioner akan disebarakan pada 2 pihak yaitu pihak mahasiswa dan pihak dosen pembimbing atau staf akademik. Berikut merupakan pertanyaan kuesioner yang disebarakan untuk pihak mahasiswa:

- 1) Hasil prediksi dapat memberikan informasi atau *insight* terkait prediksi waktu kelulusan mahasiswa bagi mahasiswa
- 2) Seberapa bermanfaat sistem prediksi bagi mahasiswa
- 3) Hasil prediksi dapat meningkatkan *awareness* terkait lulus dengan tepat waktu bagi mahasiswa
- 4) Rating secara keseluruhan terkait sistem prediksi waktu kelulusan mahasiswa

Terdapat sedikit perbedaan Berikut merupakan pertanyaan kuesioner yang disebarakan untuk pihak mahasiswa:

- 1) Hasil prediksi dapat memberikan informasi atau *insight* terkait prediksi waktu kelulusan mahasiswa bagi dosen pembimbing atau staf akademik

- 2) Hasil prediksi dapat membantu dosen pembimbing untuk menyesuaikan bimbingan yang diberikan kepada mahasiswa bimbingan
- 3) Seberapa bermanfaat sistem prediksi bagi dosen pembimbing dan staf akademik
- 4) Rating secara keseluruhan terkait sistem prediksi waktu kelulusan mahasiswa

Pernyataan-pernyataan tersebut dapat dinilai atau ditanggapi dari rentang 1-5 berdasarkan konsep skala Likert. Nilai 1 menggambarkan pendapat bahwa responden sangat tidak setuju, dan nilai 5 merepresentasikan pendapat bahwa responden sangat setuju.

3.2.2.6 Deployment

Deployment merupakan tahapan paling akhir dari *framework* CRISP-DM dimana pada tahapan ini proyek akan diimplementasikan oleh *user* [29]. Setelah penentuan algoritma terbaik berdasarkan metrik evaluasi, model akan dipilih untuk melakukan prediksi data mahasiswa aktif dan mendapatkan hasil prediksi berupa semester lulus. Hasil prediksi akan ditampilkan pada *website*. Proses prediksi dilakukan pada *back-end* tidak secara langsung pada *website* untuk meningkatkan efektivitas dan efisiensi.

Penelitian ini membandingkan 3 algoritma yang akan dipilih berdasarkan hasil performa terbaik. Berikut merupakan perbandingan algoritma klasifikasi yang digunakan untuk penelitian ini:

Tabel 3.2 Perbandingan Algoritma Klasifikasi

Indikator	Neural Network [31][33]	Naïve Bayes [35]	SVM [36]	k-NN [16]
Konsep	NN terdiri dari jaringan neuron terstruktur dengan layer dan bobot. Bobot ini dipelajari melalui proses pelatihan untuk memetakan input data ke output prediksi.	Menggunakan teorema Bayes, NB mengasumsikan kemandirian fitur dan menghitung probabilitas kelas berdasarkan fitur data	Mencari <i>hyperplane</i> pemisah optimal yang memaksimalkan margin antara data dari kelas yang berbeda	k-NN bekerja dengan mencari dengan konsep mencari kesamaan dengan menggunakan jarak. Jarak yang umumnya digunakan

Indikator	Neural Network [31][33]	Naïve Bayes [35]	SVM [36]	k-NN [16]
				adalah jarak Euclidean
Kelebihan	<ul style="list-style-type: none"> • Mampu mempelajari pola kompleks dalam data. • Memiliki akurasi prediksi yang tinggi. • Dapat menangani data <i>nonlinear</i>. 	<ul style="list-style-type: none"> • Sederhana dan mudah diimplementasikan • Cepat dalam pelatihan • Membutuhkan komputasi yang kecil. 	<ul style="list-style-type: none"> • Mampu menangani data <i>nonlinear</i>. • Memiliki akurasi prediksi yang tinggi • <i>Robust</i> terhadap <i>overfitting</i> 	<ul style="list-style-type: none"> • Mudah diimplementasikan • Dapat <i>handle</i> data yang bersifat <i>non-linear</i> yang baik
Kekurangan	<ul style="list-style-type: none"> • Membutuhkan data yang besar untuk dilatih • Membutuhkan waktu yang lama untuk dilatih • Membutuhkan komputasi yang besar. 	<ul style="list-style-type: none"> • Membutuhkan data yang independen. • Tidak dapat menangani data <i>nonlinear</i>. • Akurasi prediksi tidak setinggi NN 	<ul style="list-style-type: none"> • Membutuhkan waktu yang lama untuk dilatih • Membutuhkan komputasi yang besar • Sulit untuk memilih <i>hyperparameter</i> yang optimal 	<ul style="list-style-type: none"> • Sulit bekerja dengan data yang kompleks • Sensitif terhadap data yang bersifat <i>noisy</i> • Membutuhkan <i>resources</i> dan waktu yang lebih banyak

Berdasarkan Tabel 3.2, algoritma Neural Network, Naïve Bayes, dan Support Vector Machine dipilih dari algoritma k-Nearest Neighbors (k-NN), dimana hal ini dikarenakan algoritma k-NN sangat sensitif terhadap kualitas data, khususnya data yang bersifat *noisy* dan berdimensi kompleks [16].

Terdapat perbedaan cara algoritma Neural Network (NN), Naïve Bayes, dan SVM dalam melakukan proses klasifikasi. NN bekerja berdasarkan *layer* dan bobot. Bobot ini dipelajari melalui proses pelatihan untuk memetakan *input* data ke *output* prediksi. NN mampu mempelajari pola yang kompleks dengan menghasilkan akurasi yang tinggi, namun memiliki *cost* yang lebih besar. Berbeda dengan NN, NB dapat dijalankan dengan komputasi yang lebih kecil namun akurasi prediksi yang dihasilkan cenderung lebih kecil dari NN. Di sisi lain, algoritma SVM memiliki fungsi yang berbeda yaitu dapat menjaga agar model tidak mengalami *overfitting*.

3.3 Teknik Pengumpulan Data

Penelitian ini menggunakan metode penelitian kuantitatif karena data yang digunakan merupakan data numerikal yang objektif bukan merupakan persepsi ataupun perasaan, konkrit, dapat diteliti oleh panca indra, dapat dilihat hubungan antara variabel, dan dapat diverifikasi [62]. Data yang dikumpulkan dan diperoleh adalah tipe data primer. Data primer adalah data yang dikumpulkan atau diberikan yang sifatnya langsung pada pengumpul data, sedangkan data sekunder adalah data yang telah dikumpulkan oleh pihak ketiga dimana sumber data tidak langsung dikumpulkan atau diberikan pada pengumpul data [63]. Data yang digunakan merupakan mahasiswa UMN periode tahun 2014-2022 akan diambil dan dikumpulkan dari Biro Informasi Akademik (BIA) Universitas Multimedia Nusantara yang berjumlah 9034 data mahasiswa. Data yang diperoleh terbagi menjadi 2 kelompok yaitu data mahasiswa yang berstatus lulus dan aktif. Rentang periode data ini dipilih karena alasan relevansi, data mahasiswa berstatus aktif memiliki rentang tahun angkatan 2017 hingga 2023, tahun 2017 merupakan tahun angkatan terakhir yang dapat lulus pada masa pengerjaan penelitian ini dan tahun angkatan 2023 merupakan tahun angkatan terbaru pada masa pengerjaan penelitian ini. Data mahasiswa berstatus lulus memiliki rentang tahun angkatan 2014 hingga 2023, hal ini dikarenakan model butuh mempelajari pola mahasiswa yang lulus 14 semester.

3.3.1 Populasi dan Sampel

Sampling merupakan teknik pemilihan suatu subset dari suatu populasi untuk menjadi representasi karakter dari populasi tersebut [64]. Penelitian ini menggunakan teknik *purposive sampling* sebagai salah jenis dari *non-probability sampling*. *Purposive sampling* adalah teknik pengambilan sampel yang tipe karakteristiknya telah ditentukan untuk tujuan yang relevan dengan penelitian [65]. *Purposive sampling* digunakan pada penelitian ini disebabkan oleh batasan masalah yang telah ditetapkan, relevansi dengan program studi dari peneliti, serta terdapat perbedaan rangkaian atau urutan aktivitas perkuliahan pada program studi lain yang dapat menyebabkan perbedaan pola data mahasiswa. Populasi data pada penelitian ini adalah mahasiswa Universitas Multimedia Nusantara yang berjumlah 11012 mahasiswa, sedangkan sampel data yang digunakan merupakan mahasiswa Jurusan Sistem Informasi Universitas Multimedia Nusantara yang berjumlah 1013 mahasiswa. Validasi sistem prediksi dalam bentuk penyebaran kuesioner sebagai tahap akhir validasi menggunakan konsep *User Acceptance Testing* dilakukan dengan juga menggunakan teknik *simple random sampling* terhadap kelompok mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara yang berstatus mahasiswa aktif dengan populasi berjumlah 844 mahasiswa dan sampel berjumlah 72 mahasiswa.

3.3.2 Periode Pengambilan Data

Proses pendataan data mahasiswa berlangsung pada Bulan Februari hingga Bulan Maret 2024, dimulai dari pengisian formulir sebagai perizinan dengan dokumentasi surat yang ditunjukkan pada Lampiran B Surat Perizinan Pengambilan Data hingga pemberian data dengan dokumentasi surat yang ditunjukkan pada Lampiran C Surat Pengambilan Data. Data mahasiswa diberikan oleh BIA UMN melalui *e-mail*. Data kuesioner mulai didapatkan setelah proses implementasi model pada *website* yaitu pada Bulan April 2024.

3.4 Variabel Penelitian

Penelitian ini akan menggunakan 2 tipe variabel yaitu variabel dependen dan variabel independen.

A. Variabel Independen

Variabel Independen merupakan variabel yang mempengaruhi atau menyebabkan adanya perubahan yang sesuai terhadap variabel dependen [66]. Mengacu pada data mahasiswa di Universitas Multimedia Nusantara yang akan diteliti, Variabel Independen pada penelitian ini yaitu:

- 1) Indeks Prestasi Kumulatif (IPK) per semester
- 2) Indeks Prestasi Semester (IPS) per semester
- 3) Semester yang ditempuh saat ini
- 4) Jumlah SKS yang diambil per semester
- 5) Jumlah SKS yang lulus per semester
- 6) Jumlah SKS yang tidak lulus per semester
- 7) Total Satuan Kredit Semester (SKS) yang telah diambil
- 8) Total Satuan Kredit Semester (SKS) yang telah lulus
- 9) Keikutsertaan semester antara.

Variabel NIM atau nomor induk mahasiswa pada penelitian ini akan disamarkan dan diganti menjadi data *dummy* dengan alasan konfidensial.

B. Variabel Dependen

Variabel Dependen merupakan variabel hasil yang terikat dan perubahannya dipengaruhi variabel independen [66]. Variabel dependen pada penelitian ini yaitu masa studi mahasiswa yang dipengaruhi dari variabel-variabel independen.

3.5 Teknik Analisis Data

Pada penelitian ini diperlukan *tools* untuk mengimplementasikan *data mining* untuk mendukung metode penelitian yang dipakai yaitu *framework* CRISP-DM. Berikut merupakan perbandingan dari bahasa pemrograman R dan Python yang dapat digunakan sebagai *tools* untuk melakukan analisa dalam proses *educational data mining*:

Tabel 3.3 Perbandingan R dan Python [67],[68],[69]

Parameter	R	Python
Konsep	R merupakan bahasa pemrograman yang digunakan untuk komputasi statistik, grafik, dan data analisis	Python merupakan bahasa pemrograman yang digunakan untuk pembuatan proyek seperti pengembangan web, otomatisasi, dan <i>data science</i> /analisis data
Syntax	Syntax pada bahasa pemrograman R lebih rumit dan kurva pembelajarannya tidak langsung	Syntax pada bahasa pemrograman Python lebih simpel dan mudah untuk diimplementasikan
IDE	Terdapat beberapa IDE seperti Rstudio, StatET, dan lainnya	Terdapat banyak IDE seperti Jupyter Notebook, Spyder, Pycharm, dan lainnya
<i>Popularity</i>	Bahasa pemrograman R kurang populer dibanding bahasa pemrograman Python, umumnya digunakan oleh para statistika	Bahasa pemrograman Python lebih populer dibanding bahasa pemrograman R, umumnya digunakan oleh para <i>programmer</i> dan <i>developer</i>
Visualisasi	R memiliki fitur visualisasi yang kompleks dan variatif	Python memiliki <i>tools</i> seperti Matplotlib dan Seaborn namun tidak sebaik fitur/ <i>tools</i> visualisasi pada R
Kecepatan dan Kompleksitas	Dalam memproses data, R cenderung lebih lambat dan lebih kompleks	Dalam memproses data, Python cenderung lebih cepat dan sederhana
Ukuran data	Mempunyai keterbatasan dalam memproses data yang besar karena bergantung pada ukuran RAM	Mudah dalam memproses data yang volumenya besar
<i>Packages</i> dan <i>library</i>	R memiliki <i>packages</i> dan <i>library</i> statistik yang lebih banyak dan kuat untuk analisis data	Python memiliki <i>packages</i> dan <i>library</i> untuk objektif yang lebih luas seperti <i>machine-learning</i> , analisis data, dan pengembangan web.

Berdasarkan Tabel 3.4, terlihat bahwa bahasa pemrograman Python memiliki objektif yang lebih luas dimana bahasa ini dapat digunakan sebagai alat yang mendukung dalam pembuatan sebuah proyek dimulai dari tahap awal pemrosesan data hingga tahap pengimplementasian hasil akhir dari pengolahan data [67]. Oleh karena itu, penelitian ini akan menggunakan bahasa pemrograman Python sebagai bahasa pemrograman untuk menganalisis data.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A