

## **BAB III**

### **METODOLOGI**

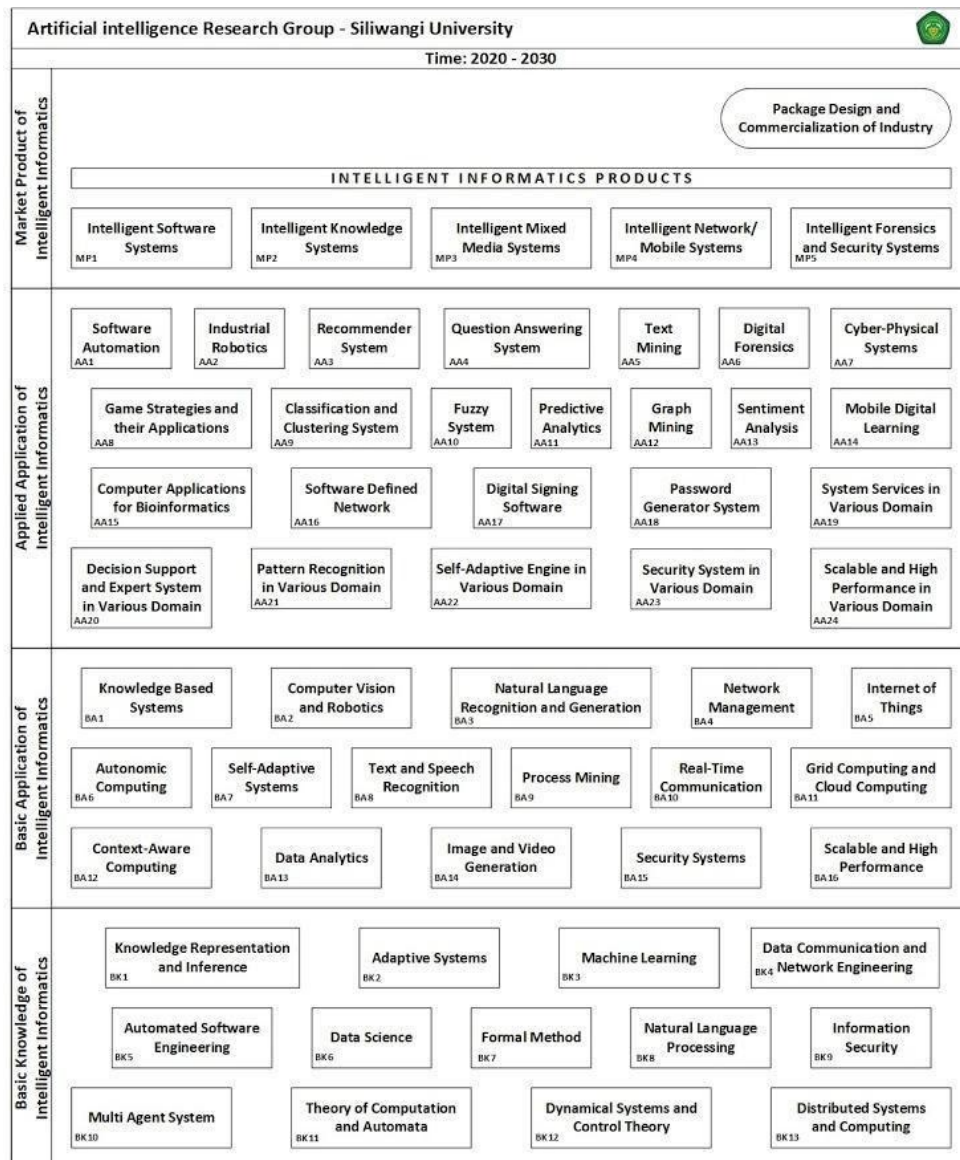
#### **3.1 Metodologi Penelitian**

Penentuan metodologi penelitian merupakan langkah yang harus digunakan karena dapat menentukan berhasil tidaknya penelitian yang dilakukan. Teknik yang disusun secara teratur, sistematis pada metodologi penelitian bertujuan untuk menemukan fakta-fakta atau prinsip-prinsip, mengembangkan dan menguji kebenaran ilmiah suatu pengetahuan (Sugiyono, 2016). Metodologi penelitian yang digunakan dalam penelitian ini adalah metode eksperimen.

Metode eksperimen adalah suatu cara untuk mencari hubungan sebab akibat antara dua faktor yang ditimbulkan oleh peneliti dengan mengeliminasi atau mengurangi faktor – faktor lain yang mengganggu(Arikunto, 2019)

#### **3.2 Roadmap Penelitian**

Roadmap penelitian merupakan peta perjalanan dari sebuah penelitian untuk menjelaskan arah penelitian yang dituju. Roadmap pada penelitian ini mengacu kepada Roadmap *Artificial Intelligence Research Group* - Universitas Siliwangi tahun 2020 – 2030. Roadmap Artificial Intelligence Research Group – Universitas Siliwangi tahun 2020 – 2030 ini merupakan kolaborasi antara Kelompok Keahlian (KK) Informatika dan Sistem Intelijen (ISI) bersama dengan Kelompok Keahlian (KK) Jaringan, Keamanan, dan Forensika Digital (JKF) jurusan Informatika, Fakultas Teknik, Universitas Siliwangi. *Roadmap Artificial Intelligence Research Group* - Universitas Siliwangi tahun 2020 – 2030 ditunjukkan pada Gambar 3.1

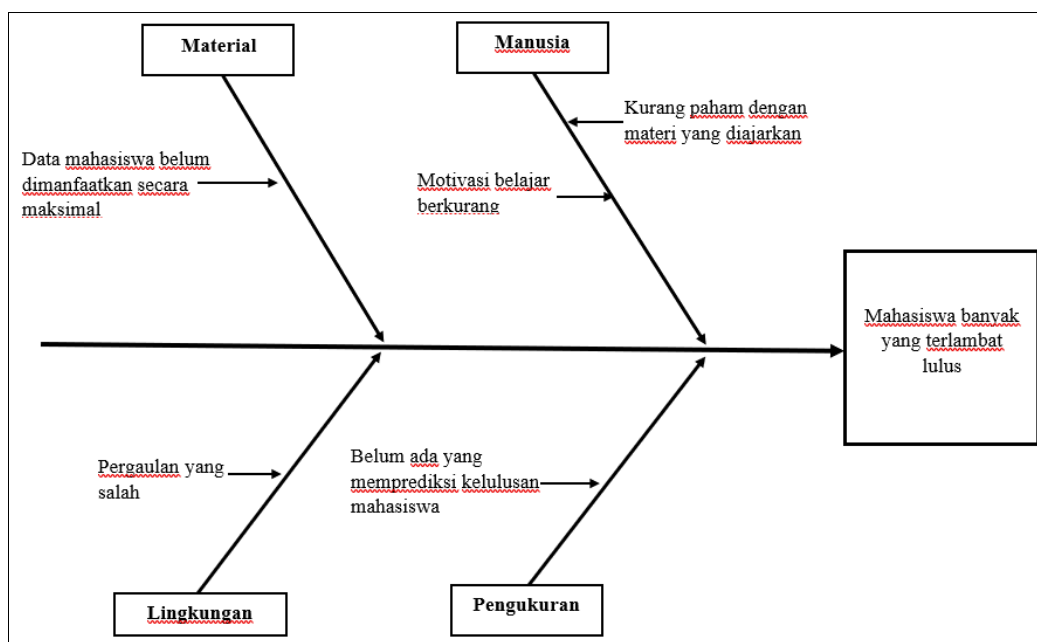


Gambar 3.1 Roadmap AI Research Group Universitas Siliwangi 2020- 2030 (Sumber AIS, 2019)

Gambar 3.1 merupakan Road Map penelitian dari AI Research Group Universitas Siliwangi 2020-2023, Basic Knowledge yang digunakan pada penelitian ini yaitu Data Science dan Machine Learning, Basic Application pada penelitian ini yaitu Data Analytics dan Process Mining, dan pada Applied Application yaitu Predictive Analytics.

### 3.3 Diagram Fishbone

Diagram fishbone merupakan diagram yang dapat menunjukkan sebab akibat dari suatu permasalahan. Diagram *Fishbone* pada penelitian ini ditunjukkan pada gambar 3.2. Diagram ini menggambarkan hal-hal yang terkait dengan penelitian yang akan dilakukan. Objek yang akan diteliti yaitu data kelulusan mahasiswa informatika.



Gambar 3.2 Diagram *Fishbone*

Gambar 3.2 merupakan *fishbone* diagram berdasarkan dari Dr. Kaoru Ishikawa seorang ilmuwan asal Jepang. Analisis Fishbone (atau Ishikawa) adalah suatu pendekatan terstruktur yang memungkinkan dilakukan suatu analisis lebih terperinci dalam menemukan penyebab-penyebab suatu masalah, ketidaksesuaian, dan kesenjangan yang ada (Gaspers V, 2002).

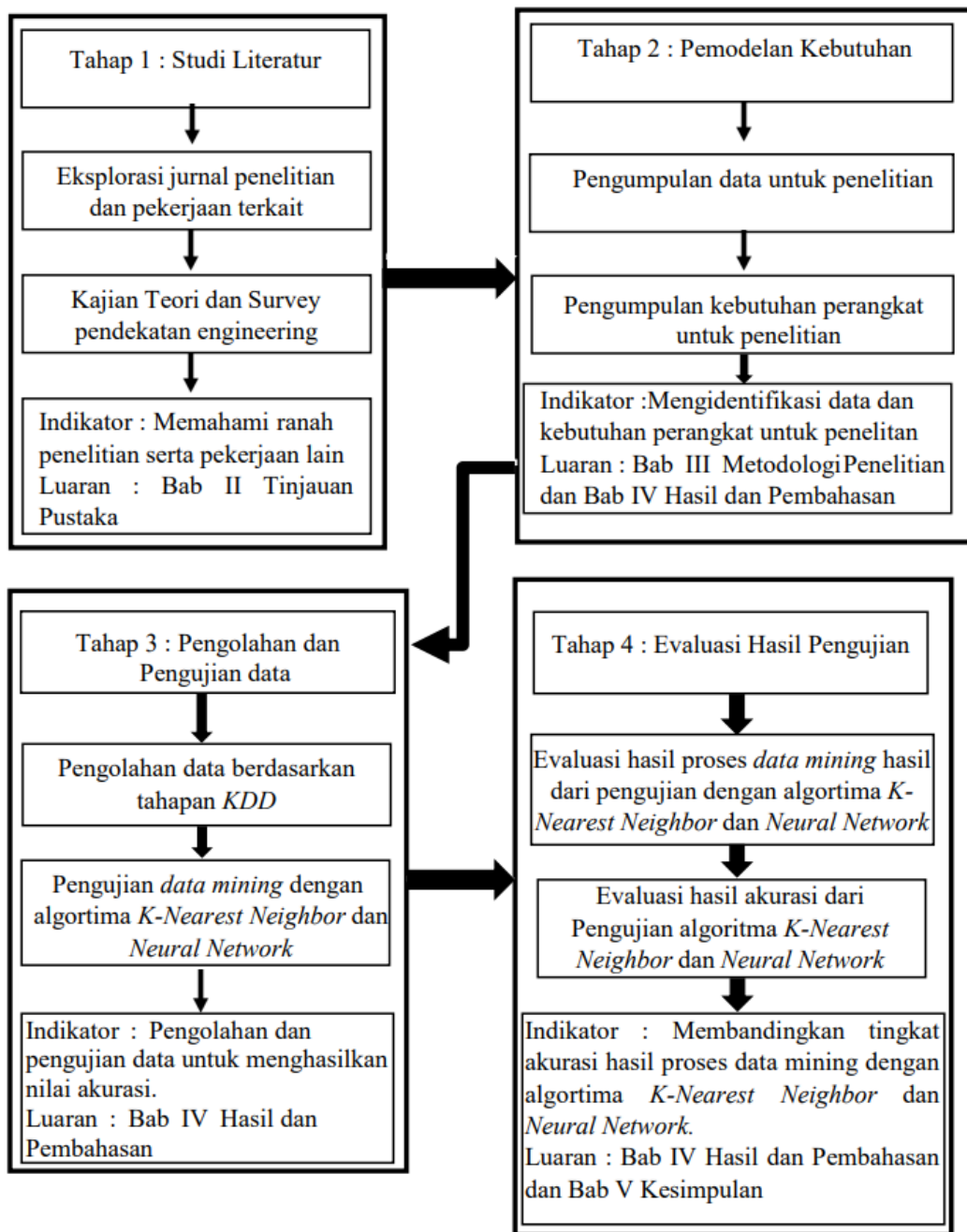
Ada 4 faktor yang menyebabkan akan dilakukannya prediksi terhadap kelulusan mahasiswa, diantaranya : Manusia, Pengukuran, Lingkungan dan Material. Manusia disini ialah mahasiswa. Mahasiswa juga termasuk kedalam

faktor yang mempengaruhi tingkat kelulusan. Hilangnya motivasi belajar dan tidak mengertinya materi yang diajarkan merupakan hal yang dapat mempengaruhi tingkat kelulusan mahasiswa. Kemudian untuk pengukuran yaitu belum adanya suatu proses prediksi terhadap kelulusan mahasiswa yang bertujuan sebagai langkah preventif supaya mahasiswa tersebut bisa lulus tepat waktu.

Lingkungan menjadi faktor selanjutnya dikarenakan banyak mahasiswa yang salah dalam memilih lingkungan pergaulan. Mahasiswa banyak yang hanya bersantai tanpa memperdulikan kuliahnya. Akibatnya banyak mahasiswa yang terlambat lulus. Selanjutnya yaitu faktor material. Faktor ini merupakan faktor terakhir yang mana maksud dari faktor ini adalah terdapat banyak data mahasiswa yang belum dimanfaatkan secara maksimal oleh instansi terkait untuk digunakan sebagai acuan dalam membuat kebijakan.

### **3.4 Tahapan Penelitian**

Tahapan penelitian merupakan salah satu hal yang penting, karena tahapan penelitian yang baik dan benar akan berpengaruh pada hasil penelitian. Tahapan penelitian harus disusun secara sistematis. Tahapan penelitian yang dilakukan pada penelitian ini yaitu dimulai dari studi literatur, pengumpulan data, pengolahan data mining, hasil serta pembahasan dan kesimpulan dan saran. Diagram tahapan penelitian dapat dilihat pada Gambar 3.3



Gambar 3.3 Diagram Tahapan Penelitian

Penelitian ini dibagi menjadi 4 tahapan utama yaitu :

1. Studi literatur. Mencari informasi mengenai prediksi kelulusan mahasiswa
2. Pemodelan kebutuhan. Mengumpulkan data dan kebutuhan perangkat

3. Pengolahan dan pengujian. Data yang telah terkumpul diolah sesuai dengan tahapan *Knowledge Discovery in Database* (KDD).
4. Evaluasi hasil pengujian. Menjelaskan hasil dari proses data mining yang dilakukan dengan menggunakan metode *K-Nearest Neighbor* dan *Neural Network*.

### **3.4.1 Studi Literatur**

Studi literatur dilakukan dengan mempelajari dan mengkaji teori-teori melalui buku, jurnal ilmiah, laporan penelitian serta dokumen-dokumen yang berkaitan dengan penelitian. Jurnal yang terkait dengan penelitian *data mining* dikaji secara menyeluruh dan membandingkan literatur dengan hasil penelitian sebelumnya, mencari faktor-faktor yang menyebabkan perbedaan hasil dari penelitian tersebut.

### **3.4.2 Pemodelan Kebutuhan**

- a. Pengumpulan data untuk penelitian

Data yang digunakan dalam penelitian ini merupakan data mahasiswa informatika angkatan 2014 yang didapatkan dari Unit Pelaksana Teknis Teknologi Informasi dan Komunikasi (UPT TIK) Universitas Siliwangi dengan melalui tahap proses perizinan. Data yang diperoleh dari UPT TIK Universitas Siliwangi yaitu nama, nomor pokok mahasiswa, jenis kelamin, tanggal lulus, nilai indek prestasi semester 1 sampai 7 dan status kelulusan.

- b. Pengumpulan kebutuhan perangkat untuk penelitian

Perangkat keras dan perangkat lunak yang digunakan pada penelitian ini dapat dilihat pada Tabel 3.1 dan Tabel 3.2

Tabel 3.1 Perangkat keras

<i>Hardware</i>	<b>Spesifikasi</b>
Prosesor	AMD® Ryzen™ 5 2500U Processor, 2.0 GHz (2 M Cache, up to 3.6 GHz)
RAM	8192MB RAM DDR4
Grafis	AMD Radeon Vega 8 Graphics

Tabel 3.2 Perangkat lunak

<i>Software</i>	<b>Fungsi</b>
<i>Microsoft Edge</i>	Digunakan untuk mencari jurnal-jurnal terkait penelitian dan untuk mengakses <i>Google Colaboratory</i> .
<i>Google Colab</i>	Untuk menulis dan menjalankan kode <i>Python</i> , serta melakukan analisis data dan pelatihan model <i>machine learning</i> .
<i>Google Drive</i>	Untuk menyimpan dataset yang akan diakses pada saat menggunakan <i>Google Colab</i> .
<i>Rapidminer</i>	Untuk melakukan analisis proses <i>data mining</i> , serta melakukan klasifikasi data dan prediksi.

### 3.4.3 Pengolahan dan Pengujian Data

Pengolahan dan pengujian data yang dilakukan dalam penelitian ini mengikuti tahapan dalam *Knowledge Discovery in Databases* (KDD), berikut ini tahapan-tahapannya :

#### 1. Data Selection

Pemilihan data perlu dilakukan sebelum tahapan penggalian informasi dalam *Knowledge Discovery in Database* (KDD) dimulai. Data yang digunakan dalam penelitian ini adalah data kelulusan mahasiswa jurusan informatika angkatan 2014.

Adapun atribut yang digunakan yaitu NPM, Nama, Jenis kelamin, Indeks Prestasi semester 1 sampai dengan semester 7 dan status kelulusan.

## 2. Preprocessing

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus *Knowledge Discovery in Database* KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Pada tahap ini membuang data yang nilai atributnya tidak lengkap atau tidak memiliki nilai atribut sama sekali. Contoh seperti pada Tabel 3.3

Tabel 3.3 *Data Cleaning*

No	1	2	3	4	5
NPM	147006001	147006002	147006003	147006004	147006005
Nama	Ridhwan GS	Rani D	Sri Nur E	Aditya YM	Raaitha A
JK	Laki-Laki	Perempuan	Perempuan	Laki-laki	Perempuan
IP 1	3.84	3.63	3.47	-	-
IP 2	3.36	3.30	3.21	-	-
..	..	..	..	..	..
Status kelulusan	Tepat	Tepat	Tepat	-	-

Pada tabel diatas data mahasiswa yang bernama Aditya dan Raaitha akan dilakukan pembuangan data, karena nilai atributnya tidak lengkap.

## 3. Transformation

Tahapan *transformation* merupakan tahap untuk merubah data menjadi bentuk yang sesuai untuk *proses data mining*. Pada penelitian ini menggunakan metode *Min-Max Normalization*, yaitu suatu metode normalisasi dengan melakukan transformasi *linier* pada data asli dan nilai yang dinormalisasi berada



dalam kisaran atau *range* tertentu. Hasil dari suatu normalisasi adalah [-1, 1] atau [0.0, 1.0]. Rumus *Min-Max Normalization* yaitu :

$$z = \frac{x - \min(x)}{[\max(x) - \min(x)]} \quad (3.1)$$

#### 4. *Data Mining*

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Metode yang digunakan pada penelitian ini adalah metode *K-Nearest Neighbor* dan *Neural Network*. Metode *K-Nearest Neighbor* memiliki atribut yang disimbolkan dengan huruf *k*, yaitu jumlah tetangga yang akan dijadikan acuan pada *K-Nearest Neighbor*, nilai *k* yaitu bilangan bulat positif yang jumlahnya ganjil ataupun genap. Rumus perhitungan jarak yang digunakan pada metode *K-Nearest Neighbor* yaitu rumus *euclidean distance*, rumusnya sebagai berikut :

$$d = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \\ = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3.2)$$

Sedangkan untuk metode *Neural Network* inialisasikan bobot jaringan antara (-0.0 sampai 1.0). Hitung input untuk simbol berdasarkan nilai input dan bobot jaringan menggunakan rumus :

$$input_j = \sum_{i=1}^n o_i w_{ij} + \theta_j \quad (3.3)$$

#### 5. *Interpretation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini

juga mencakup pemeriksaan pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

#### **3.4.4 Evaluasi hasil pengujian**

Evaluasi hasil pengujian algoritma sangat penting untuk menilai kinerja dan efektivitas suatu algoritma yang digunakan. Evaluasi ini membantu mengukur sejauh mana algoritma mampu menghasilkan hasil yang akurat dan relevan dalam konteks masalah yang sedang diselesaikan. Berikut ini beberapa evaluasi hasil pengujian algoritma :

1. Akurasi. Metrik akurasi digunakan untuk mengukur sejauh mana algoritma dapat mengklasifikasikan data dengan benar. Ini dinyatakan sebagai persentase dari jumlah prediksi yang benar dibandingkan dengan total jumlah prediksi.
2. Presisi dan Recall. Presisi dan recall adalah metrik yang umum digunakan dalam tugas klasifikasi. Presisi mengukur sejauh mana hasil positif yang diprediksi oleh algoritma adalah benar, sedangkan recall mengukur sejauh mana hasil positif yang sebenarnya diidentifikasi oleh algoritma. Metrik ini membantu dalam memahami kemampuan algoritma dalam mengklasifikasikan dengan benar kelas yang diinginkan dan menghindari kesalahan.
3. F1-Score. F1-score adalah ukuran yang menggabungkan presisi dan recall menjadi satu angka tunggal. Ini memberikan gambaran keseluruhan tentang kinerja algoritma dalam mengklasifikasikan dengan benar kelas yang diinginkan dan menghindari kesalahan.
4. Kurva ROC (Receiver Operating Characteristic). Kurva ROC digunakan dalam tugas klasifikasi untuk memvisualisasikan dan mengukur kinerja algoritma

pada berbagai ambang batas. Ini menggambarkan perbandingan antara tingkat kesalahan positif palsu (False Positive Rate, FPR) dengan tingkat kebenaran positif (True Positive Rate, TPR). Area di bawah kurva (Area Under the Curve, AUC) digunakan sebagai metrik evaluasi, di mana nilai AUC yang lebih tinggi menunjukkan kinerja yang lebih baik.

5. Confusion Matrix. Confusion matrix digunakan untuk menampilkan hasil prediksi algoritma secara terperinci. Matriks ini berisi jumlah prediksi yang benar dan salah untuk setiap kelas. Dengan menggunakan matriks ini, dapat menghitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan lainnya.