

BAB III

METODOLOGI PENELITIAN

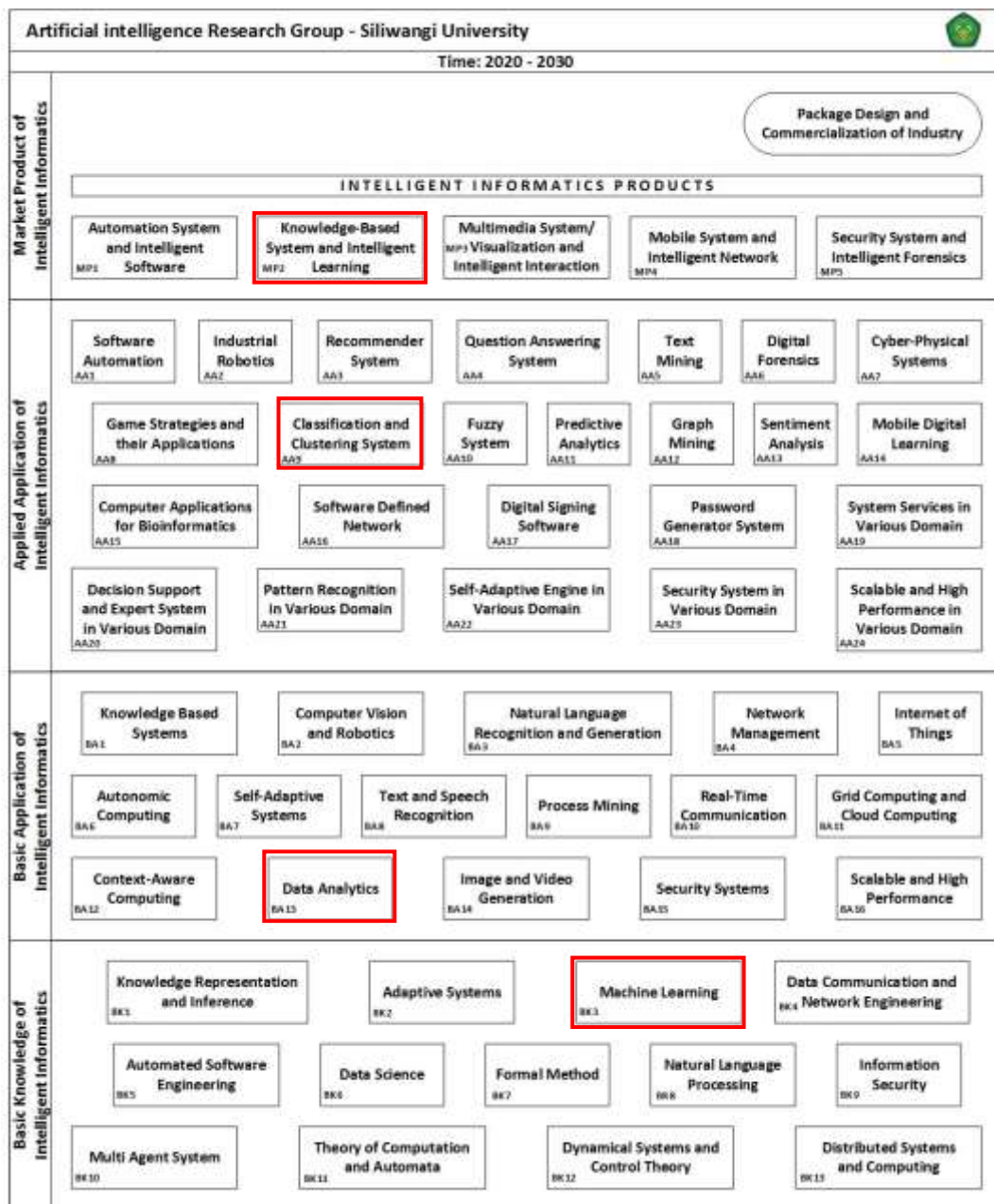
3.1 Metodologi Penelitian

Metode yang digunakan dalam penelitian ini adalah penelitian kuantitatif dengan orientasi eksperimen. Penelitian kuantitatif dilakukan karena penelitian berasal dari sumber data yang telah diuji dengan sistem komputasi dan menggunakan rumus matematika sehingga hasil akhir analisis bersifat statistik atau angka. Penelitian dengan orientasi eksperimen dilakukan karena pada penelitian ini menghasilkan nilai dari setiap variabel, sehingga perlu diukur pengaruh antara variabel independen dan dependen (Creswell et al., 2007).

Objek yang digunakan pada penelitian ini berfokus pada malware Android yang terkumpul pada dataset CCCS-CIC-AndMal2020 yang diterbitkan oleh *University Of New Brunswick, Canada*. Penelitian dilakukan dengan melakukan pengumpulan data untuk selanjutnya data akan melalui tahap *preprocessing*. Tahapan *preprocessing* dilakukan untuk normalisasi, pembersihan, *transformasi*, *random oversampling* serta *split data* agar dataset dapat dilakukan proses komputasi. Tahap selanjutnya dilakukan pemodelan dan pengujian dengan algoritma RF untuk proses klasifikasi. Performa dari algoritma RF dalam melakukan klasifikasi pada dataset CCCS-CIC-AndMal2020 akan didapatkan sebagai hasil akhir melalui evaluasi *confusion matrix*.

3.2 Peta Jalan (Road Map) Penelitian

Penelitian ini merupakan bagian dari peta jalan riset Kelompok Keahlian Informatika dan Sistem *Intelligence* (KK ISI) yang berkolaborasi dengan komunitas *AI Research Group Siliwangi University*. Kajian sebelumnya akan dijadikan dasar pengetahuan dalam penyempurnaan untuk capaian akhir dari peta jalan penelitian secara keseluruhan.



Gambar 3. 1 Roadmap/Peta Jalan Penelitian

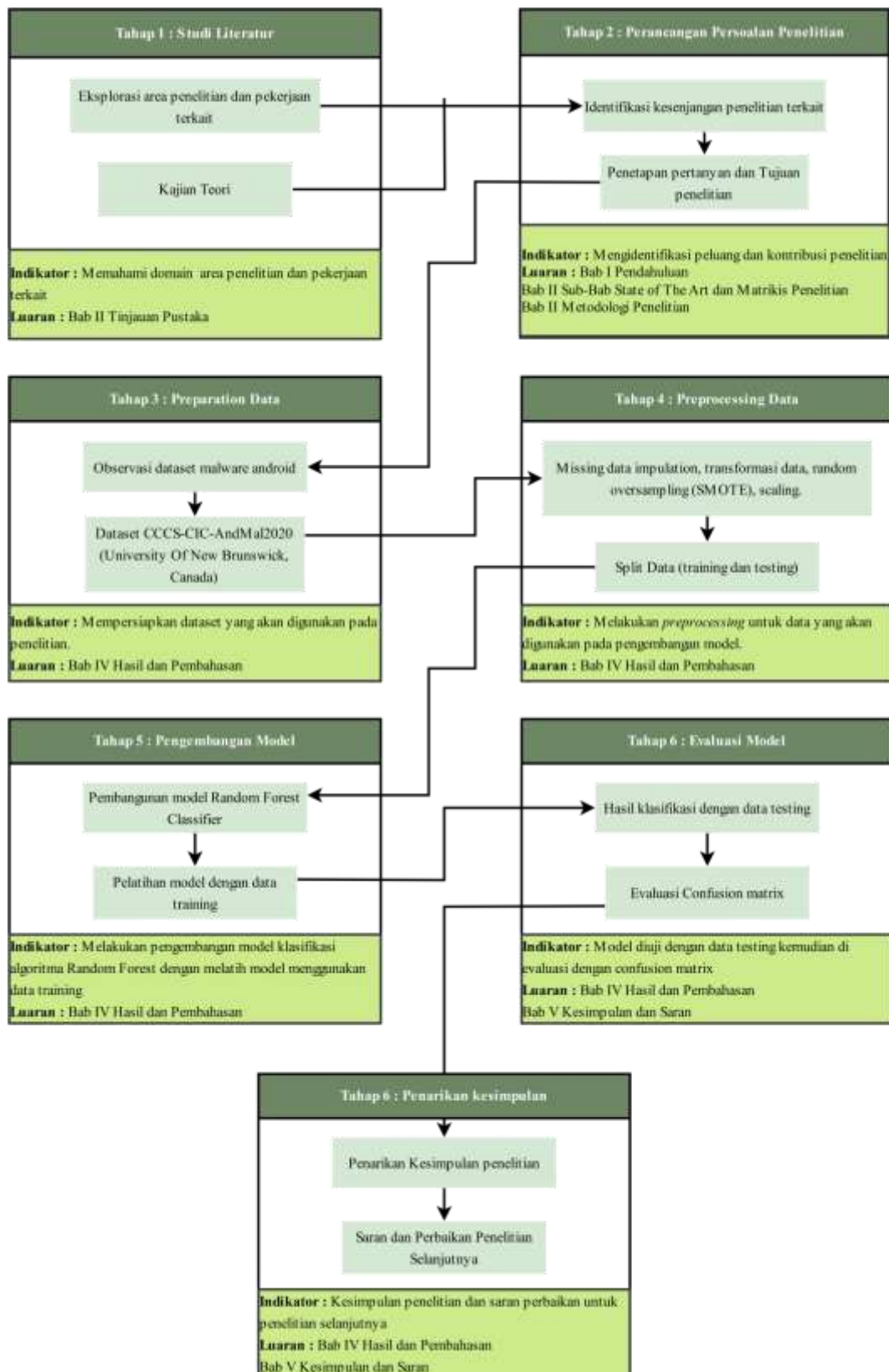
Pada gambar 3.1 terdapat beberapa kajian penelitian yang terdapat pada riset Kelompok Keahlian Informatika dan Sistem Intelligence (KK ISI) berkolaborasi dengan komunitas *AI Research Group Siliwangi University*. Kajian yang akan digunakan pada penelitian diberi tanda dengan persegi berwarna merah yang meliputi *machine learning*, *data analytics*, *classification and clustering system* untuk mencapai target berupa *knowledge-based system and intelligent learning*

mengenai analisis hasil classification pada *malware Android*. Fokus kajian pada peta jalan penelitian ini meliputi :

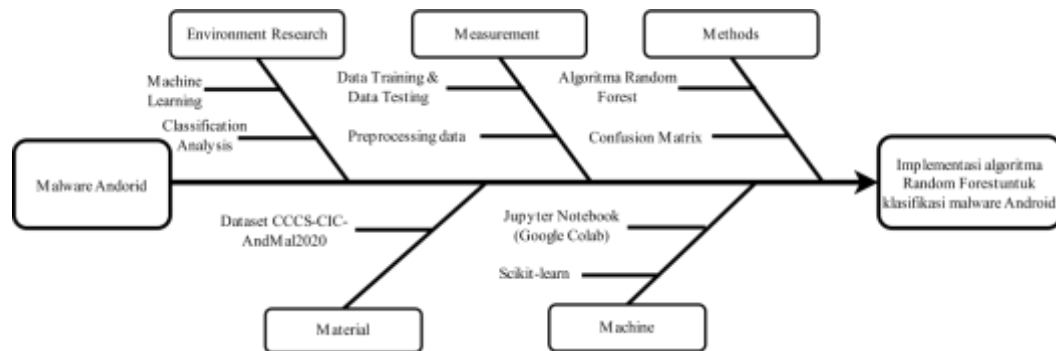
1. *Machine Learning*, yang merupakan cabang ilmu dari kecerdasan buatan (*Artificial Intelligence*) dengan algoritma yang dikembangkan untuk menganalisis dan mengenali pola dalam data.
2. *Data Analytics*, melibatkan proses pengumpulan, pembersihan, transformasi, pemrosesan dan analisis data untuk menarik kesimpulan, mendukung pengambilan keputusan serta mengeksplorasi pola dalam data.
3. *Classification and Clustering System*, merupakan sistem dalam *machine learning* yang digunakan untuk proses pengelompokan data. *Classification System* akan menjadi fokus utama dalam penelitian ini untuk mengelompokkan entitas atau objek ke dalam kategori atau kelas berdasarkan atribut-atribut tertentu. *Classification System* merupakan *supervised learning*, dimana model atau algoritma dilatih menggunakan data yang telah diberi label
4. *Knowledge-Based System and Intelligent*, merupakan capaian akhir dari peta penelitian ini yaitu untuk merepresentasikan, menyimpan, dan menggunakan pengetahuan eksplisit untuk memecahkan masalah atau membuat keputusan. Mengintegrasikan *machine learning*, *data analytics*, dan *classification system* dalam *Knowledge-Based System and Intelligent* dapat meningkatkan kecerdasan dan kemampuan adaptasi sistem.

3.3 Tahapan Penelitian

Bagian ini menjelaskan tahapan-tahapan sistematis yang akan dijadikan sebagai acuan kerangka penelitian. Tahapan penelitian ini terdiri dari studi literatur, perancangan persoalan penelitian, *preparation data*, *preprocessing data*, proses pengembangan model, evaluasi model kemudian terakhir dilakukan penarikan kesimpulan dari hasil evaluasi model. Tahapan dari penelitian ini disajikan menggunakan diagram alir seperti pada gambar 3.2, kemudian tujuan penelitian disajikan secara detail dengan menggunakan fishbone diagram pada gambar 3.3.



Gambar 3. 2 Tahapan Penelitian



Gambar 3. 3 *Fishbone* Diagram (Coccia, 2018)

Gambar 3.3 merupakan gambaran tujuan penelitian yang di representasikan ke dalam *fishbone* diagram. Diagram *fishbone* (disebut juga diagram Ishikawa atau diagram sebab-akibat) adalah teknik grafis untuk menunjukkan beberapa penyebab suatu peristiwa atau fenomena tertentu. *Fishbone* diagram (Ishikawa diagram), dibuat oleh Kaoru Ishikawa (1990) dalam penelitian bidang manajemen (Coccia, 2018). *Malware Android* berada dibagian ekor yang menjadi “sebab” atau objek dari pelenelitian, kemudian yang menjadi “akibatnya” pada bagian kepala dilakukan implementasi algoritma *Random Forest* untuk klasifikasi *malware Android*. Bagian tulang memiliki keterangan sebagai berikut:

1. *Environment Research*, merupakan area penelitian atau topik yang akan dijadikan fokus utama pada penelitian. Fokus utama penelitian ini adalah pada bidang *machine learning* untuk *clasification analysis*.
2. *Material*, merupakan sumber bahan atau data yang akan digunakan pada penelitian. Dataset CCCS-CIC-AndMal2020 akan menjadi bahan pada penlitian ini.
3. *Measurement*, merupakan tahapan dalam mempersiapkan data sebelum diuji. Terdapat proses split data menjadi data *training* dan data *testing*, kemudian *preprocessing data*.
4. *Machine*, merupakan mesin atau alat yang digunakan dalam penelitian. *Jupyter Notebook (Google Colab)* dan *Scikit-learn* akan digunakan pada penelitian ini.

5. *Methods*, merupakan metode atau algoritma yang akan digunakan dalam penelitian. Penelitian ini menggunakan algoritma RF untuk klasifikasi dan evaluasi model menggunakan *Confusion Matrix*.

3.3.1 Studi Literatur

Studi literatur adalah langkah awal dalam mengumpulkan konsep, teori, dan data dari berbagai sumber yang terkait dengan penelitian. Studi literatur dilakukan dengan membaca jurnal atau buku terkait yang menjelaskan tentang topik penelitian, metode, algoritma, dan tahapan eksperimen yang pernah dilakukan. Topik utama pada penelitian ini penerapan metode algoritma *machine learning* pada klasifikasi *malware Android*. Studi literatur dengan meninjau penelitian sebelumnya, akan memberi wawasan dalam pengembangan topik penelitian terkait.

3.3.2 Perancangan Persoalan Penelitian

Perancangan persoalan penelitian merupakan tahapan yang dilakukan berdasarkan analisis penelitian terdahulu. Persoalan penelitian ditentukan dengan mengidentifikasi kesenjangan, kekurangan, dan saran dari penelitian-penelitian sebelumnya sehingga dapat dilakukan perbaikan. Setelah itu, dilakukan perancangan dengan tujuan yang jelas dengan menetapkan pertanyaan penelitian, batasan penelitian, dan manfaat penelitian.

3.3.3 Preparation Data

Data yang digunakan pada penelitian ini merupakan dataset CCCS-CIC-AndMal2020 yang diterbitkan oleh *University Of New Brunswick, Canada* (<https://www.unb.ca/cic/datasets/andmal2020.html>)

3.3.4 Preprocessing Data

Preprocessing data merupakan tahap penting dalam studi klasifikasi berbasis *machine learning*. Hal ini disebabkan karena kumpulan data biasanya

memiliki *missing value*, *outlier*, atau fitur yang tidak diperlukan yang dapat mempengaruhi keakuratan model *machine learning*. Tahap persiapan data ini meliputi, *missing data imputation*, *transformasi data*, *random oversampling (SMOTE)*, *scaling* dan *split data (data training dan data testing)* akan dilakukan pada tahap *preprocessing data*.

3.3.5 Pengembangan Model

Tahap selanjutnya, data yang sudah melalui tahap *preprocessing* akan melalui tahap pengujian (klasifikasi) dengan menggunakan algoritma RF, model ini akan dijalankan dengan Google Colab dan Scikit-learn menggunakan bahasa *Python*. Model akan dilatih dengan menggunakan data training.

3.3.6 Evaluasi Model

Model akan diuji menggunakan data testing dan kemudian dilakukan evaluasi model dengan *confusion matrix* untuk melihat performa model dalam melakukan proses klasifikasi. Performa yang diukur yaitu *accuracy*, *precision*, dan *recall*

3.3.7 Penarikan Kesimpulan

Tahapan terakhir dilakukan penarikan kesimpulan berdasarkan hasil evaluasi dan performa klasifikasi dari algoritma RF dalam proses klasifikasi *malware Android*. Penarikan kesimpulan bertujuan untuk memberikan gambaran umum terhadap analisis data dan hasil evaluasi model yang mencakup keseluruhan penelitian.

