

BAB III

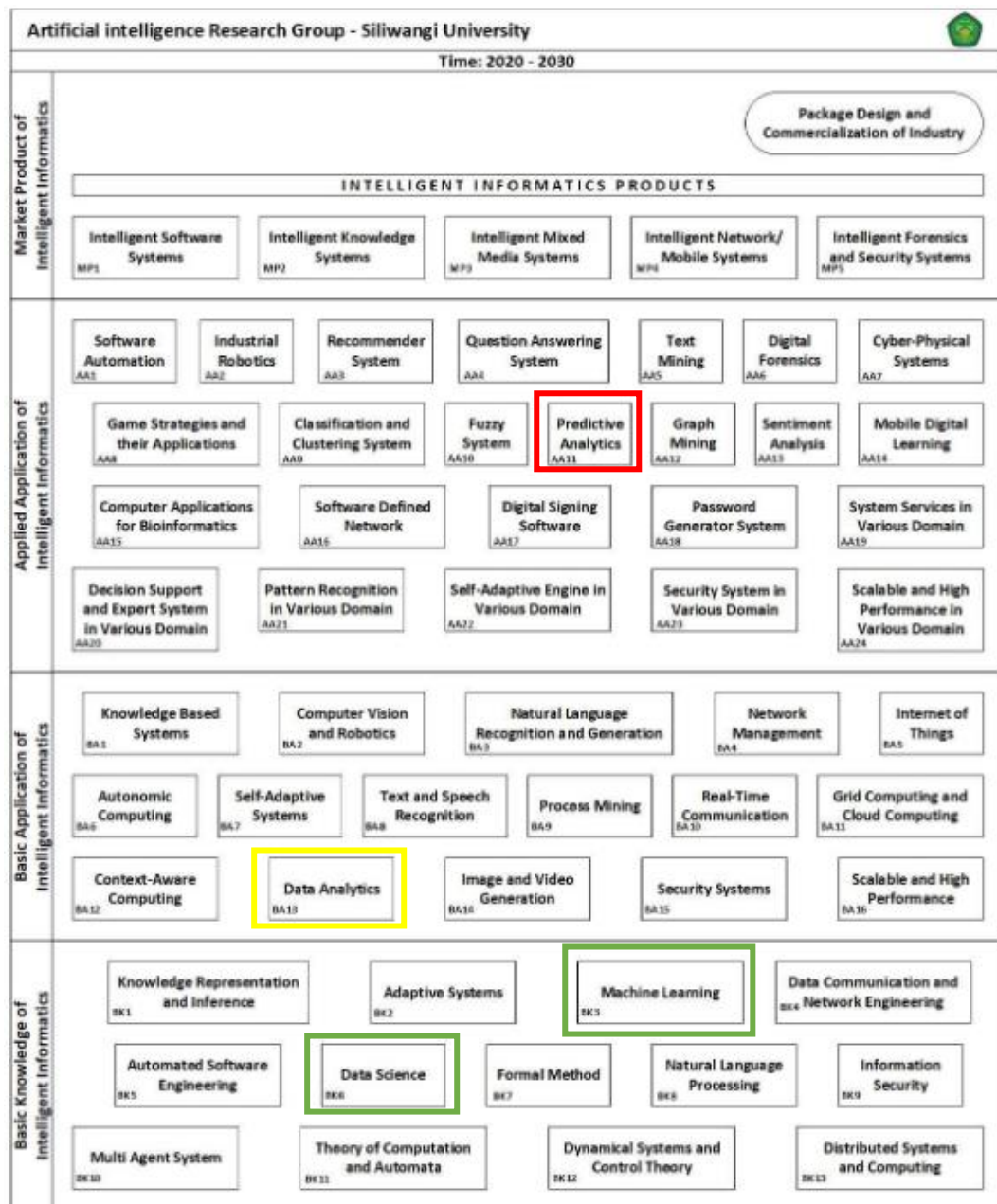
METODOLOGI PENELITIAN

3.1 Metode Penelitian

Menurut Prof. Dr Suryana (2012), Metode Penelitian atau ilmiah merupakan Langkah dalam mendapatkan pengetahuan ilmiah. Pada Penelitian ini, objek yang digunakan yaitu *temperature, humidity, pH, rainfall* dan unsur hara tanah nitrogen, pospor, dan kalium. Penelitian yang dilakukan adalah dengan mengukur performa kinerja serta keakuratan dari algoritma *Random forest* dalam memprediksi konsentrasi unsur hara tanah .

3.2 Peta jalan (*Road Map*) Penelitian

Peta jalan (*Road Map*) pada penelitian ini mengacu pada Peta Jalan *Artificial Intelligence Research Group* Universitas Siliwangi tahun 2020 – 2030. Peta Jalan *Artificial Intelligence Research Group* Universitas Siliwangi merupakan kolaborasi antara KK (Kelompok Keahlian) Informatika dan Sistem Inteligen (ISI) bersama dengan KK (Kelompok Keahlian) Jaringan, Keamanan, dan Forensik Digital (JKF) jurusan Informatika Fakultas Teknik, Unversitas Siliwangi.



Gambar 3. 1 Peta Jalan Artificial Intelligence Research (AIS) Group Universitas Siliwangi 2020 - 2030

Pada penelitian ini, berdasarkan Peta Jalan Artificial Intelligence Research (AIS) Group Universitas Siliwangi 2020 – 2030, spesifikasi setiap lapisan yang digunakan diantaranya pertama *Basic Knowledge* yang digunakan pada penelitian

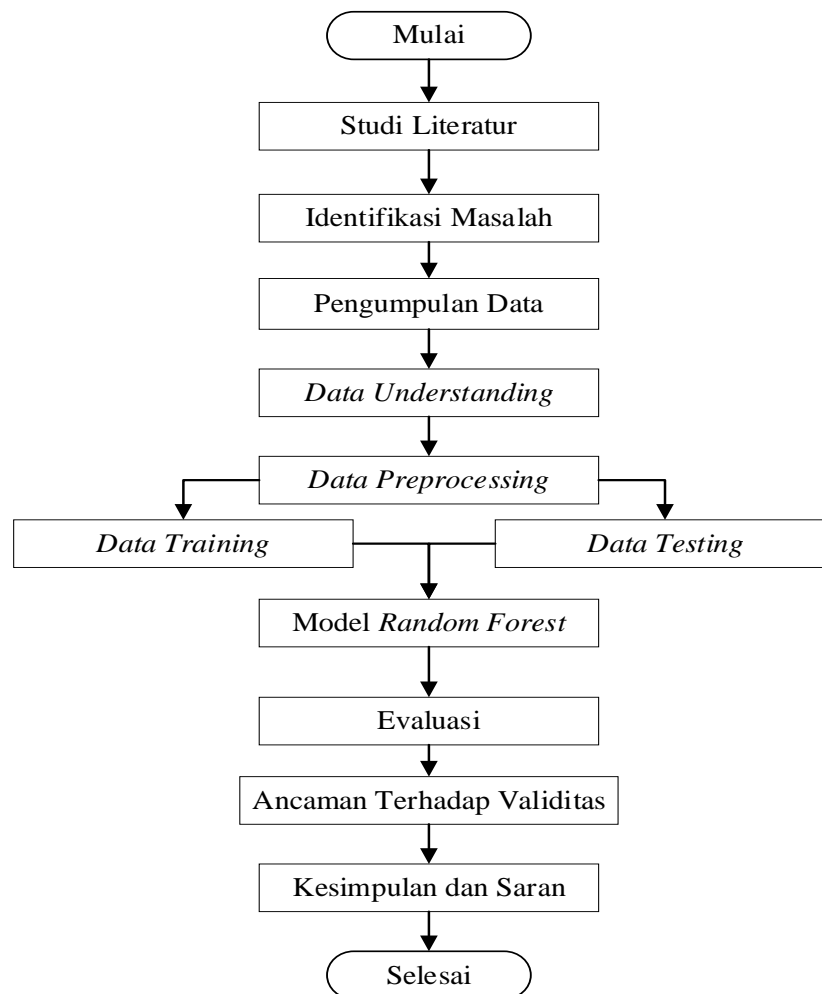
ini yaitu *Machine Learning* dan *Data Science*. *Basic Knowledge* merupakan lapisan yang paling utama untuk dilakukan, karena pada lapisan ini harus melakukan pemahaman mengenai *Basic Knowledge* yang akan digunakan dalam penelitian ini. Selain itu, harus melakukan studi literatur dilakukan untuk memahami teori yang terkait dengan penelitian serta untuk mengumpulkan informasi yang sesuai dengan topik dari penelitian yang dilakukan.

Kedua, *Basic Application* yang digunakan pada penelitian ini yaitu *Data Analytics*. Pada lapisan ini menentukan aplikasi dasar yang akan digunakan dalam penelitian ini yaitu melakukan analisis data untuk dapat memprediksi unsur hara tanah. Selain dari menentukan aplikasi dasar yang akan dilakukan dalam penelitian ini, sebelumnya harus dilakukan pengumpulan data, analisis dan pengolahan data, pengujian algoritma *Random forest*.

Terakhir, *Applied Application* yang digunakan pada penelitian ini yaitu *Predictive Analytics*. Pada penelitian ini akan dilakukan suatu analisis prediksi untuk dapat menentukan atau memprediksi unsur hara tanah dengan menggunakan algoritma *Random forest*. Pada tahapan ini, tidak hanya menentukan *Applied Application* yang digunakan dalam penelitian tetapi, melakukan pengembangan yang lebih mendalam terhadap model yang telah ditentukan untuk mendapatkan hasil yang lebih baik.

3.3 Tahapan Penelitian

Tahapan penelitian ini berdasarkan dari beberapa literatur yang disesuaikan dengan kebutuhan dengan kebutuhan pada penelitian ini (Nahya Nur F. W., 2023). Digambarkan pada gambar 3.2 penelitian dari penelitian ini.



Gambar 3. 2 Tahapan Penelitian

3.1.1 Studi Literatur

Studi literatur dilakukan untuk memahami teori yang terkait dengan penelitian serta untuk mengumpulkan informasi yang sesuai dengan topik dari penelitian yang dilakukan (Nahya Nur F. W., 2023). Informasi yang dikumpulkan berkaitan dengan prediksi unsur hara tanah, *machine learning*, algoritma *random forest*, dan lain-lain. Pencarian informasi tersebut menggunakan beberapa sumber seperti dari web, jurnal, buku, artikel dan lainnya yang berkaitan dengan penelitian yang dilakukan yaitu prediksi unsur hara tanah menggunakan algoritma *random forest*.

3.1.2 Identifikasi Masalah

Identifikasi masalah dilakukan untuk menentukan permasalahan dalam penelitian yang hasil akhir dari permasalahan tersebut akan menjadi tujuan dari penelitian. Sehingga, dapat dilakukan perbaikan mengenai permasalahan dari penelitian terdahulu.

Dalam proses penerapan algoritma *random forest* untuk kasus prediksi diperlukan data yang berkaitan dengan unsur hara tanah.

3.1.3 Pengumpulan Data

Pada penelitian ini, data yang digunakan adalah data sekunder. Data sekunder adalah data yang tidak diperoleh secara langsung ke lapangan. Data diperoleh dari kaggle yaitu menggunakan dataset yang telah dipublikasikan oleh chitrakumari pada tahun 2022. Tetapi, data yang diperoleh dari kaggle tersebut hanya diambil data yang akan digunakan dan dibutuhkan dalam penelitian ini. Data yang dibutuhkan dalam penelitian ini terbagi menjadi dua bagian yaitu data

dependent (Variabel terikat) dan data *independent* (variabel bebas). Data *dependent* terdiri dari nitrogen, pospor, dan kalium. Sedangkan untuk data *independent* terdiri dari *temperature*, *humidity*, *pH*, *rainfall*. Indikator yang diharapkan pada penelitian ini adalah memperoleh data pada *kaggle* yang dapat digunakan untuk data *training* dan data *testing* serta memperoleh sampel data yang dapat digunakan untuk proses prediksi.

3.1.4 Data Understanding

Pada tahap data *understanding* ini dilakukan untuk memeriksa data, sehingga masalah yang terdapat dalam data dapat diidentifikasi. Tujuan dari tahap data *understanding* ini untuk mengumpulkan, mengidentifikasi dan memahami data dengan teliti. Dengan demikian, pada tahapan data *understanding* ini dapat mengevaluasi keandalan dari data yang digunakan dalam penelitian, dan menentukan apakah data yang digunakan sesuai dengan kebutuhan dari penelitian yang sedang dilakukan. Selain itu, tahap ini dilakukan dalam mengidentifikasi kekurangan dari data yang digunakan, yang kemudian kekurangan dari data yang digunakan dapat diatasi dalam tahap selanjutnya. Maka pada tahap *understanding* ini dengan melakukan pemahaman data, dapat mengambil keputusan yang lebih baik dan menghasilkan hasil analisis yang lebih akurat.

3.1.5 Data Preprocessing

Preprocessing data merupakan proses pengolahan data mentah menjadi data dengan bentuk yang dapat dipahami dan sesuai dengan kebutuhan penelitian yang akan digunakan sebelum dilakukan proses *training*. Pada tahap ini, yang akan dilakukan seperti *cleaning* data, dan *split data* atau pembagian dataset. Proses

cleaning data dilakukan untuk memperbaiki serta memperoleh data yang dapat mendukung dalam proses prediksi. Sedangkan untuk *split data* dilakukan untuk membagi data set menjadi dua bagian yaitu data latih (*data train*) dan data uji (*data test*).

Pada penelitian ini akan dilakukan dengan beberapa rasio untuk pembagian data *training* dan data *testing*-nya yaitu 80 : 20, 70 : 30, dan 60 : 40. Pembagian data *testing* (pengujian) dan data *training* (pelatihan) dengan ketiga rasio pembagian ini telah disesuaikan dengan jumlah dataset yang digunakan dalam penelitian ini yaitu 300 data.

3.1.6 Modelling

Setelah melakukan tahap *preprocessing*, tahap selanjutnya dilakukan pemodelan algoritma *random forest*. Dimana pada tahap ini dataset akan dipecah menjadi data *training* dan data *testing*. Data *training* digunakan untuk melatih algoritma sedangkan data *testing* digunakan untuk menentukan kinerja atau keakuratan algoritma yang telah dilatih sebelumnya. Data *training* digunakan lebih besar dibandingkan dengan data *testing*. Indikator yang diharapkan dapat tahapan ini adalah menerapkan *random forest* untuk mendapatkan suatu nilai akurasi terbaik atau performa kinerja dari *random forest*.

3.1.7 Evaluasi

Pada tahap evaluasi ini dilakukan untuk mengevaluasi akurasi yang didapatkan pada penelitian. Proses evaluasi dilakukan dengan tujuan untuk menguji

performa kinerja dari algoritma *Random forest* dalam memprediksi unsur hara tanah.

3.1.8 Ancaman Terhadap Validitas

Pada tahap ini, terdapat dua validitas yang ditentukan yaitu validitas internal dan validitas eksternal, dimana kedua validitas tersebut membantu menentukan seberapa berartinya hasil penelitian yang dilakukan. Berdasarkan penelitian yang dilakukan oleh Imam Arifin (2019), Validitas internal artinya untuk menilai seberapa baik penelitian dapat menunjukkan sebuah hubungan sebab akibat antara variabel *independent* (*temperatur, humidity, pH, rainfall*) dan variabel *dependent* (nitrogen, posfor, kalium) dengan adanya pengaruh dari variabel lain. Ancaman terhadap validitas internal dalam penelitian ini berupa *overfitting*, kurangnya representasi data, data tidak akurat, kondisi variabel berubah, dan ketidakpastian prediksi.

validitas eksternal adalah mengacu pada kemampuan hasil penelitian untuk digeneralisasikan dari sampel terhadap populasi diluar lingkup penelitian. Validitas eksternal menunjukkan seberapa baik hasil dari penelitian dapat diterapkan pada populasi yang berbeda dari sampel yang digunakan dalam penelitian (Mulyana, 2022).

3.1.9 Kesimpulan dan Saran

Kesimpulan dan Saran merupakan tahap terakhir dari penelitian ini. Pada tahap ini yang akan dilakukan yaitu menetapkan kesimpulan dari proses penelitian yang telah dilakukan serta menetapkan saran untuk penelitian selanjutnya yang akan dilakukan peneliti lain.