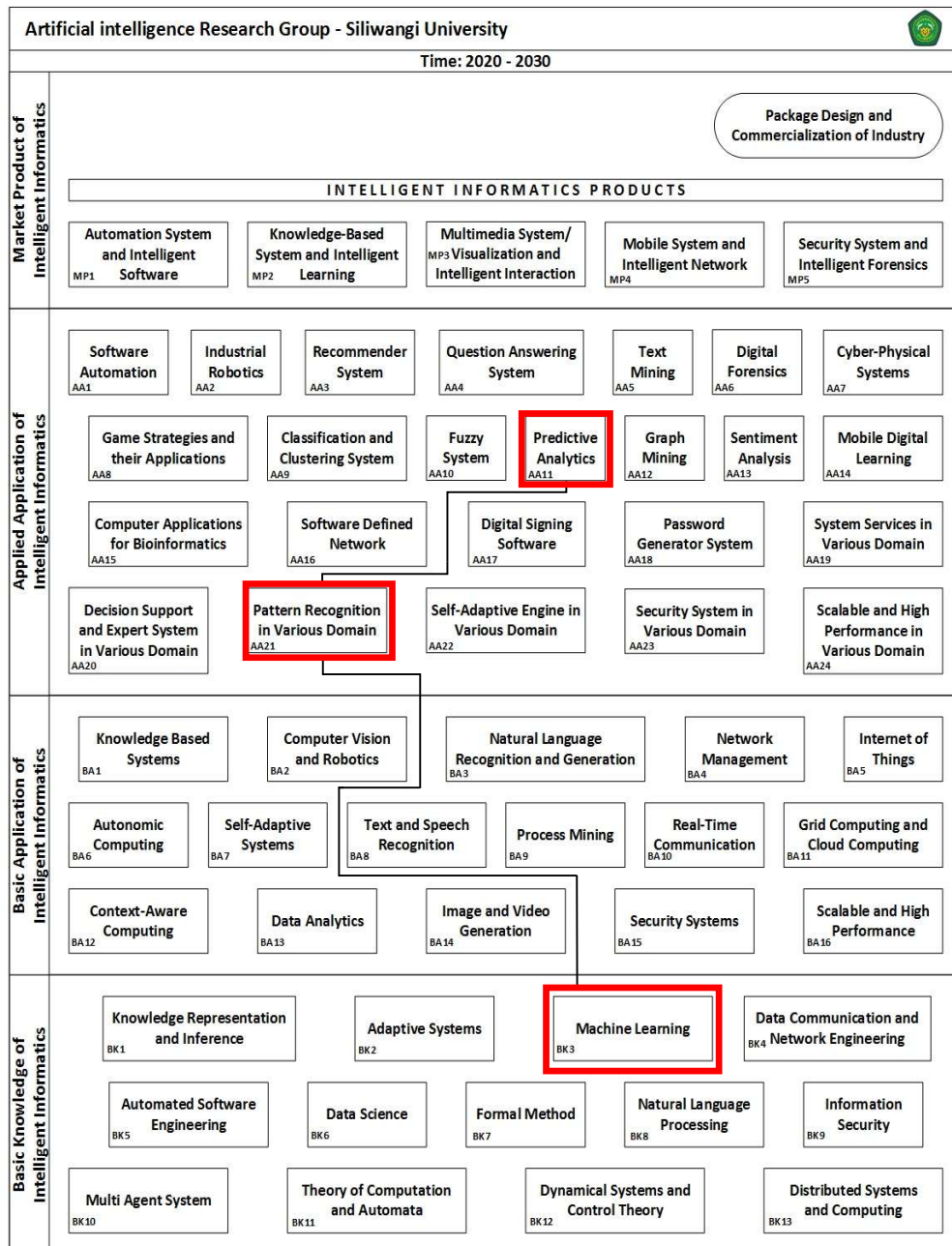


BAB III

METODOLOGI PENELITIAN

3.1 Road Map Penelitian



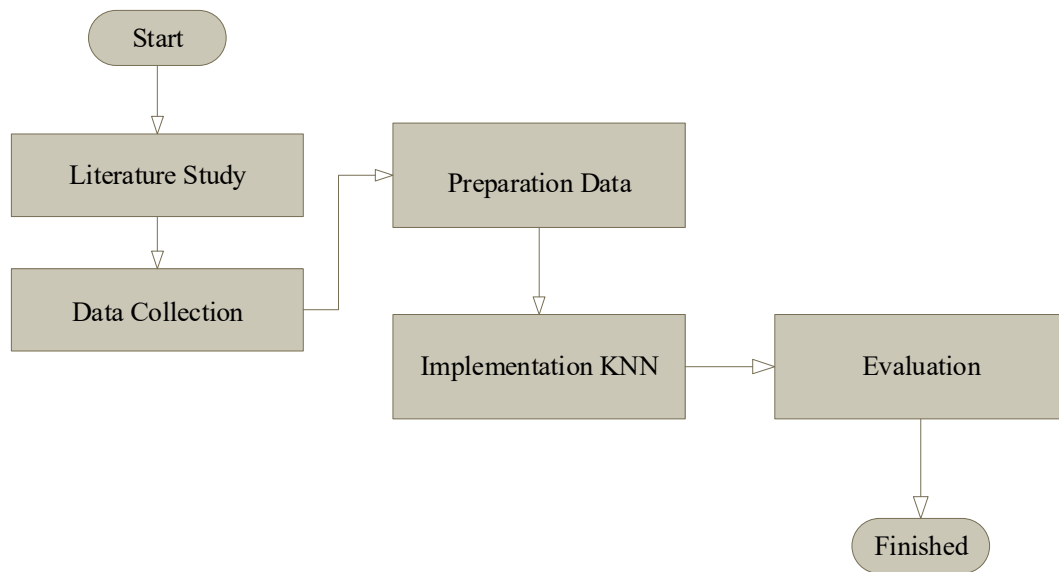
Gambar 3. 1 Roadmap AI Research Group Universitas Siliwangi 2020-2030
(Sumber AIS, 2020)

Pada penelitian ini, Kelompok Keahlian Informatika dan Sistem *Intelligen* (ISI) bersama dengan Kelompok Keahlian Jaringan, Keamanan, dan Digital Forensik (JKF) dari Jurusan Informatika, Fakultas Teknik, Universitas Siliwangi bekerja sama dalam menyusun *Roadmap Artificial Intelligence Research Group* Universitas Siliwangi untuk periode 2020–2030. Mengacu pada Gambar 3.1, bidang keilmuan yang mendasari penelitian ini meliputi *Machine Learning*, *Predictive Analytic* serta *Pattern Recognition*.

Pada proses memprediksi harga jual sapi. *Machine Learning* memberikan kerangka kerja bagi model untuk mempelajari data historis tanpa pemrograman eksplisit, di mana KNN sebagai metode pembelajaran yang diawasi mengukur kedekatan antar data menggunakan jarak *Euclidean*. *Predictive Analytics* berperan dalam memanfaatkan pola historis atribut sapi—seperti jenis, berat, usia, dan musim—untuk menghasilkan prediksi harga yang akurat, mendukung pengambilan keputusan berdasarkan data. Sementara itu, *Pattern Recognition* memungkinkan model identifikasi kemiripan antar objek berdasarkan atribut yang telah diproses, sehingga prediksi yang dihasilkan berasal dari pola kemiripan data sebelumnya. dengan demikian, bidang ketiga ini saling terintegrasi dalam membangun sistem prediksi harga sapi yang adaptif dan Andal.

3.2 Tahapan Penelitian

Tahapan penelitian yang dilakukan pada penelitian ini menggunakan kerangka penelitian yang di ambil dari (Muhamad Malik Matin, 2023) seperti pada Gambar 3.2 serangkaian langkah atau prosedur untuk melakukan penelitian berikut:



Gambar 3. 2 Alur Penelitian

Penelitian ini mencakup tahapan-tahapan yaitu :

1. Study Literatur
2. Data Collection
3. Preparation Data
4. Implementantation KNN
5. Evaluation

3.2.1 Studi Literatur

Studi literatur dilakukan dengan mengumpulkan sejumlah buku – buku, makalah dengan masalah dan tujuan peneliti yang sama (Takanwiedi & Painem, 2018). pada penelitian ini tak lepas dari referensi penelitian yang serupa sebelumnya sebagai bahan acuan diantaranya, Artikel/paper dari Jurnal Ilmiah nasional maupun internasional, Artikel/paper dari *Book Chapter*, Artikel/paper dari *Conference (Proceedings)* dan Tugas Akhir.

3.2.2 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data historis penjualan sapi dari perusahaan yang diperoleh secara manual, meliputi periode 2021 hingga 2025 dengan total 605 data setelah dilakukan proses penyaringan untuk memastikan kualitasnya. Pemilihan dataset ini dibandingkan dengan dataset lain didasarkan pada beberapa pertimbangan, antara lain karena data tersebut bersifat real, relevan langsung dengan konteks penelitian, serta mencerminkan kondisi aktual harga sapi pada perusahaan yang menjadi objek kajian. Selain itu, penggunaan data internal perusahaan memberikan jaminan konsistensi dan akurasi yang lebih baik dibandingkan data eksternal yang mungkin tidak sesuai dengan kebutuhan penelitian. Dengan demikian, dataset ini dipandang paling tepat untuk membangun model prediksi harga jual sapi menggunakan algoritma *K-Nearest Neighbors* (KNN).

3.2.3 Data Preparation

Tahap Data *Preparation* ini dilakukan beberapa hal diantaranya:

1. Pembersihan Nilai Kosong (*Missing Values*)

Pada data ada kemungkinan memiliki nilai kosong, yang menyebabkan munculnya nilai NaN (Geadalfa & Saidah, 2021). Data yang mengandung NaN pada kolom numerik kemudian dihapus menggunakan *dropna()* untuk memastikan bahwa data yang digunakan bersih dan tidak menyebabkan error saat pelatihan model (Rosydah & Widiyaningtyas, 2024).

2. Encoding Variabel Kategorikal

Variabel kategorikal yang berformat string diubah menjadi bentuk numerik (Azmi, 2024) agar dapat diproses oleh algoritma KNN. Encoding dilakukan dengan dua pendekatan:

- a. Explorasi awal menggunakan *LabelEncoder* untuk mengonversi nilai kategorikal menjadi integer (Hakim dkk., 2025).
- b. Dalam *pipeline* model digunakan encoder khusus *MultiLabelEncoder* (Hakim dkk., 2025) yang mendukung penanganan nilai NaN dan kategori yang belum pernah muncul (*unseen categories*), untuk memastikan konsistensi pada proses pelatihan dan prediksi .

3. Normalisasi Data

Algoritma KNN sangat bergantung pada perhitungan jarak antar data, semua fitur (baik numerik maupun kategorikal) perlu dinormalisasi agar berada pada skala yang seragam (Alloerung dkk., 2024). Normalisasi dilakukan menggunakan *MinMaxScaler*, yang mengubah nilai setiap fitur ke dalam rentang [0, 1] (Hazizah & Widiyaningtyas, 2024). Selain fitur input, target variabel Harga_Sapi_Rp juga dinormalisasi agar perhitungan error (seperti MSE dan RMSE) lebih stabil saat evaluasi.

4. Pembagian Data Latih dan Uji

Dataset kemudian dibagi menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*test set*) menggunakan fungsi `train_test_split` dengan perbandingan 80% data latih dan 20% data uji (Rahmani dkk., 2024).

3.2.4 Implementasi KNN

Implementasi model prediksi dilakukan menggunakan algoritma *K-Nearest Neighbors* (KNN). Algoritma ini bekerja dengan cara mencari sejumlah tetangga terdekat dari suatu data baru berdasarkan jarak tertentu, kemudian menentukan hasil prediksi berdasarkan nilai-nilai target dari tetangga tersebut. Nilai tetangga yang digunakan sangat memengaruhi tingkat akurasi model, sehingga diperlukan proses optimasi untuk menentukan nilai K yang paling sesuai (Nurfauzan & Fatimah, 2022).

Optimasi nilai K pada penelitian ini dilakukan dengan menggunakan *GridSearchCV* yang tersedia pada pustaka *scikit-learn*. Teknik ini secara sistematis mengevaluasi berbagai kombinasi *hyperparameter* pada model KNN untuk mencari konfigurasi terbaik. Parameter yang diuji seperti $n_neighbors$ jumlah tetangga yang digunakan, *weight* tipe pembobotan jarak (*uniform* atau *distance*), *metric* metode pengukuran jarak (*euclidean* atau *manhattan*).

Setiap kombinasi parameter diuji menggunakan metode *K-Fold Cross Validation* dengan nilai $k = n$, di mana data latih dibagi menjadi n bagian untuk memastikan hasil evaluasi yang lebih stabil dan mencegah terjadinya *overfitting*. Proses pencarian ini dilakukan dengan mengevaluasi performa model berdasarkan metrik *Root Mean Square Error (RMSE)* dan *Mean Absolute Error (MAE)*.

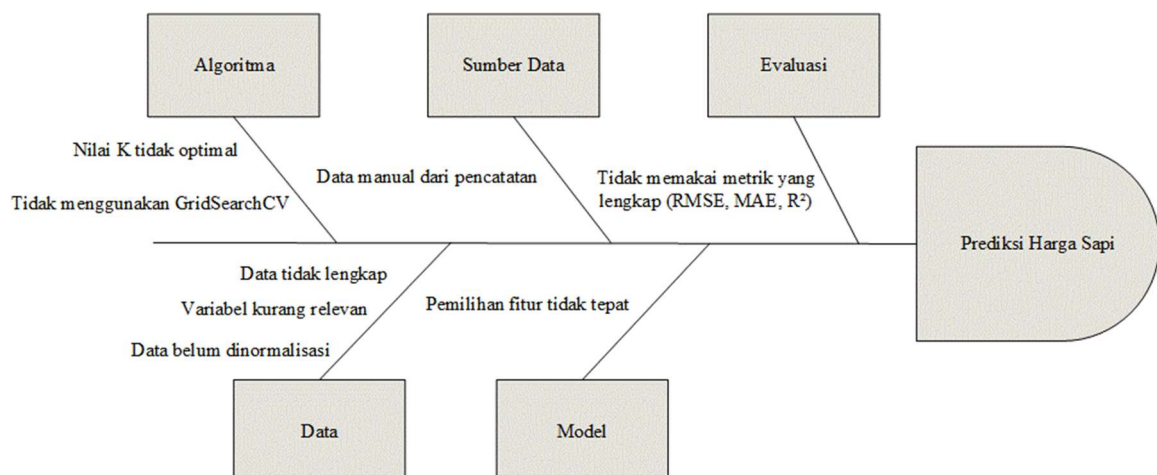
Setelah proses *GridSearchCV* selesai, diperoleh konfigurasi parameter terbaik yang menghasilkan nilai kesalahan prediksi terendah. Parameter optimal tersebut kemudian digunakan untuk melatih model akhir KNN yang digunakan dalam tahap pengujian dan evaluasi performa.

3.2.5 Evaluasi Model

Tahap terakhir dilakukan analisis dan evaluasi hasil penerapan Algoritma *K-Nearest Neighbor* (KNN) dengan menggunakan metrik evaluasi seperti RMSE, MAE, MSE, R^2 , MFE, dan SMAPE (Rahmani dkk., 2024). Kemudian hasil penelitian tersebut dibandingkan dengan hasil penelitian-penelitian sebelumnya, lalu ditarik kesimpulan dari keseluruhan proses penelitian ini.

3.3 Fishbone diagram

Fishbone diagram mengidentifikasi dan menganalisis potensi penyebab ketidakakuratan dalam proses prediksi harga jual sapi menggunakan algoritma *K-Nearest Neighbors* (KNN), penelitian ini menggunakan *fishbone diagram* (diagram tulang ikan). Diagram ini membantu dalam mengkategorikan faktor-faktor penyebab utama dari kemungkinan turunnya performa model (Aristriyana & Ahmad Fauzi, 2023).



Gambar 3. 3 Fishbone diagram

Fishbone diagram pada Gambar 3.3 digunakan untuk mengidentifikasi penyebab utama dari ketidakakuratan dalam prediksi harga jual sapi menggunakan algoritma *K-Nearest Neighbors* (KNN). Diagram ini memetakan lima kategori utama penyebab, yaitu Algoritma, Data, Sumber Data, Model, dan Evaluasi. pada aspek algoritma, permasalahan seperti pemilihan nilai K yang tidak optimal dan tidak digunakannya *GridSearchCV* dapat menurunkan performa model. dari sisi data, ketidaklengkapan, ketidaksesuaian variabel, dan data yang belum dinormalisasi menjadi kendala utama. Sumber data yang masih bersifat manual dan kurang terstruktur juga turut memengaruhi kualitas input. Selain itu, kesalahan dalam pemilihan fitur pada tahap modeling serta penggunaan metrik evaluasi yang tidak lengkap seperti RMSE, MAE, dan R^2 dapat menyebabkan hasil prediksi yang menyesatkan. dengan menggunakan diagram ini, penelitian dapat lebih fokus dalam mengantisipasi dan meminimalisasi potensi kesalahan yang berdampak pada performa model prediktif.