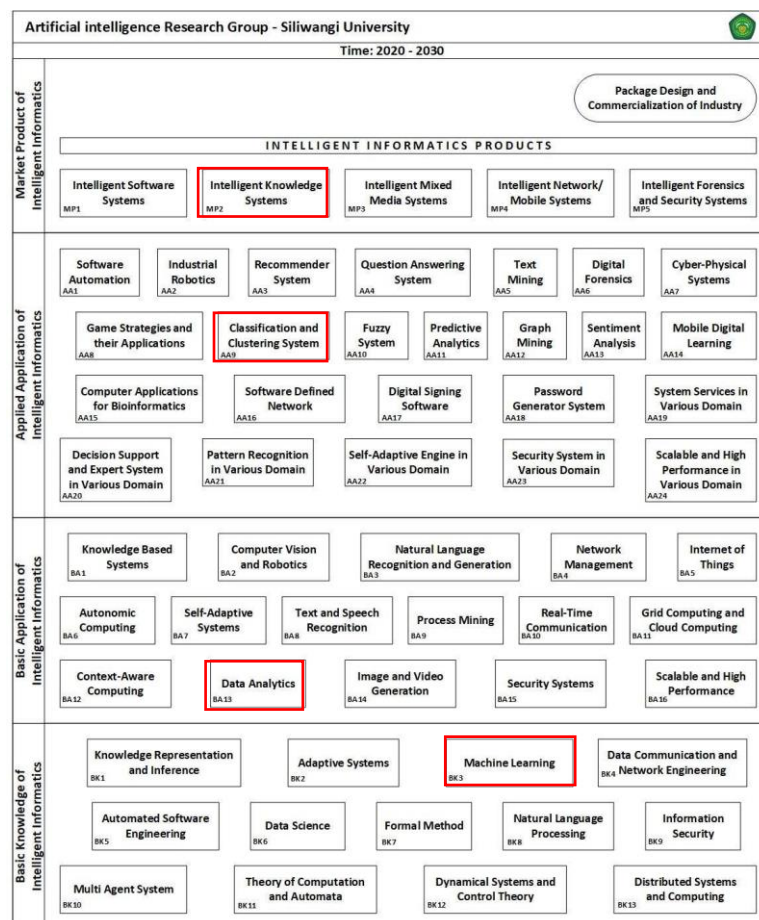


BAB III

METODOLOGI

3.1 Roadmap Penelitian

Roadmap dalam penelitian ini mengacu pada Rencana Strategis Kelompok Riset Kecerdasan Buatan Universitas Siliwangi untuk periode 2020 hingga 2030. Rencana tersebut merupakan hasil kolaborasi antara Kelompok Keahlian Informatika dan Sistem Intelijen (ISI) dengan Kelompok Keahlian Jaringan, Keamanan, dan Forensika Digital (JKF) di Jurusan Informatika, Fakultas Teknik, Universitas Siliwangi. Rencana tersebut dapat dilihat pada Gambar 3.1.

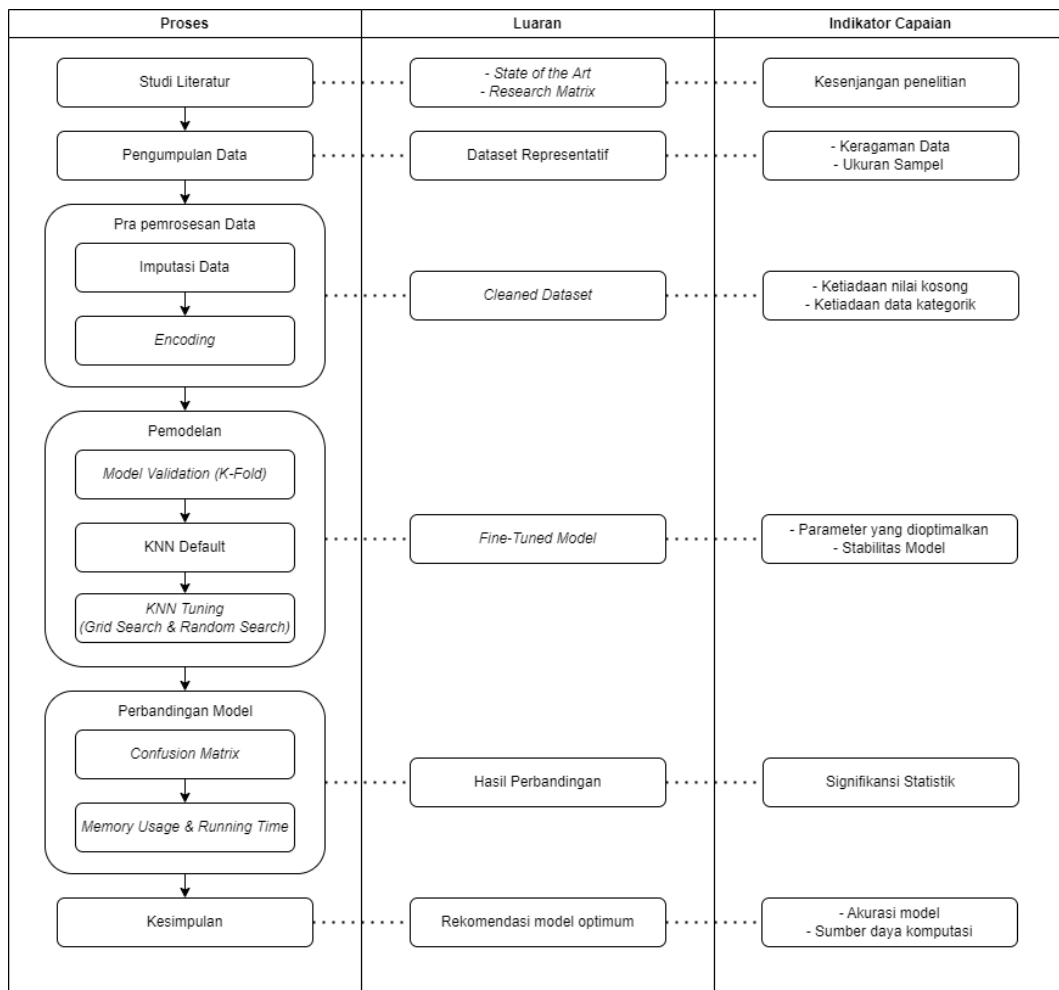


Gambar 3. 1 Roadmap penelitian

Berdasarkan Roadmap penelitian dari AI Research Group Universitas Siliwangi pada Gambar 3.1, penelitian ini didasarkan pada pengetahuan dasar dalam bidang *Machine Learning*, dengan fokus pada penerapan dasar dalam *Data Analytics*. Sistem yang diimplementasikan dalam penelitian ini adalah *Classification and Clustering system*, yang bertujuan untuk mengklasifikasikan dan mengelompokkan data dengan tepat. Sebagai solusi aplikatif, produk pasar yang digunakan adalah *Intelligent Knowledge System*, yang memanfaatkan hasil penelitian untuk menciptakan sistem yang cerdas dalam mengelola pengetahuan.

3.2 Metode Penelitian

Dalam penelitian ini digunakan berbagai kumpulan data publik yang tersedia untuk melakukan pengujian dalam menyelesaikan permasalahan klasifikasi. Penelitian ini menggunakan algoritma *K-Nearest Neighbors* (KNN) untuk melakukan klasifikasi data, kemudian membandingkan dua metode tuning parameter, yaitu *Grid Search* dan *Random Search*. Untuk menyelidiki dan memahami perbandingan kinerja antara *Grid Search* dan *Random Search* dalam konteks klasifikasi data menggunakan algoritma KNN, digunakan tiga indikator yaitu *accuracy*, *memory usage*, dan *running time*. Proses penelitian ini melibatkan beberapa tahapan, diantaranya studi literatur, pengumpulan data, pemrosesan data, pembuatan model, dan evaluasi hasil seperti yang disajikan pada gambar 3.2.



Gambar 3. 2 Metode Penelitian

3.2.1 Studi Literatur

Tahapan awal dilakukan dengan mencari *survey paper* yang relevan dengan topik penelitian yang sedang dilakukan. Kemudian melanjutkan pencarian dengan memilih *technical paper* yang terindeks, sehingga hasilnya dapat menciptakan lingkungan *state of the art* yang mendukung kelancaran penelitian ini. Proses ini membantu dalam memperoleh pemahaman yang lebih mendalam tentang perkembangan terbaru dalam domain penelitian ini dan memastikan bahwa pendekatan yang diambil sesuai dengan tren dan praktik terkini.

3.2.2 Pengumpulan Data

Pada tahap ini, data yang dikumpulkan bersifat publik dan dapat diakses melalui internet. Semua data yang dipilih merupakan data yang terkait dengan masalah klasifikasi. Data ini diperoleh dari *University of California Irvine (UCI) repository* dan meliputi sejumlah dataset seperti *Water Quality*, *Wine Quality*, *Diabetes*, *Wisconsin Breast Cancer*, *Heart Disease*, dan *Iris*. Proses pengumpulan data dilakukan secara cermat untuk memastikan bahwa dataset yang digunakan *representatif* dan dapat memberikan hasil evaluasi yang dapat dipercaya terhadap kedua metode pencarian parameter yang akan dibandingkan.

3.2.3 Pra Pemrosesan Data

Pada tahap ini, data yang telah dikumpulkan akan diolah agar dapat diterima dengan baik oleh model yang akan digunakan. Salah satu langkah yang dilakukan adalah pengecekan *missing value* untuk memastikan tidak ada data yang hilang. Selain itu, juga diterapkan teknik *imputasi* untuk mengisi *missing value* yang ditemukan dalam dataset. Selanjutnya, dilakukan proses *encoding label* dari variabel kategorikal menjadi kelas dengan label numerik, misalnya menggunakan label 0 dan 1 jika kelas hanya terdiri dari dua kategori. Proses ini penting untuk memastikan bahwa data siap digunakan dalam proses pelatihan model.

3.2.4 Pemodelan

Pada tahap ini, dilakukan proses pemodelan algoritma *Machine Learning (ML)* menggunakan *K-Nearest Neighbors (KNN)*. Selain itu, tahap ini juga melibatkan penentuan parameter yang akan digunakan oleh *Grid Search* dan *Random Search*.

Parameter-parameter tersebut akan digunakan pada proses *Hyperparameter Tuning* untuk mencari semua kombinasi parameter-parameter yang ada pada *search space* yang telah ditentukan, kemudian mencari pasangan parameter dengan hasil model yang optimal berdasarkan nilai rata-rata dari hasil validasi terbaik. Tahap *cross validation* dilakukan dengan menggunakan metode *K-Fold Cross Validation*, sebuah teknik yang bertujuan untuk menghasilkan akurasi yang lebih baik daripada metode pengujian biasa. Teknik ini membagi dataset menjadi K bagian, di mana setiap bagian akan digunakan sebagai set pengujian (Peryanto et al., 2020b).

3.2.5 Perbandingan Evaluasi Model

Setelah model-model disesuaikan dengan parameter terbaik, langkah yang terakhir adalah melakukan evaluasi kinerjanya. Evaluasi dilakukan menggunakan *confusion matrix*, sebuah metode yang memungkinkan penilaian seberapa baik model dalam memprediksi kelas target. Dari *confusion matrix*, dapat dihitung berbagai metrik evaluasi seperti akurasi, *recall*, *precision*, dan *f-measure*, yang memberikan gambaran lebih lengkap tentang kinerja model. Perhitungan *recall*, *precision*, dan *f-measure* penting untuk memastikan bahwa model tidak hanya efektif secara keseluruhan tetapi juga mampu memberikan prediksi yang konsisten dan akurat untuk kondisi spesifik yang dihadapinya (Heydarian et al., 2022). Oleh karena itu, meskipun fokus utama pada akurasi untuk menentukan nilai K terbaik, pengukuran nilai lain tetap diperlukan untuk memahami secara menyeluruh kemampuan model dalam melakukan klasifikasi. Namun, perlu diperhatikan dalam beberapa situasi, ada *trade-off* antara metrik-metrik tersebut. *Trade-off* bisa terjadi

karena model klasifikasi sering kali harus membuat keputusan yang berimbang pada dua metrik yang bertolak belakang. Misalnya, meningkatkan *recall* bisa menurunkan presisi dan sebaliknya (Ohsaki et al., 2017). Oleh karena itu, penting untuk memahami konteks aplikasi model dan memilih metrik yang paling relevan dengan tujuan yang diharapkan. Selain itu, pada perbandingan model juga digunakan fungsi dari pustaka standar *Python* untuk mengukur *memory usage* dan *running time*.