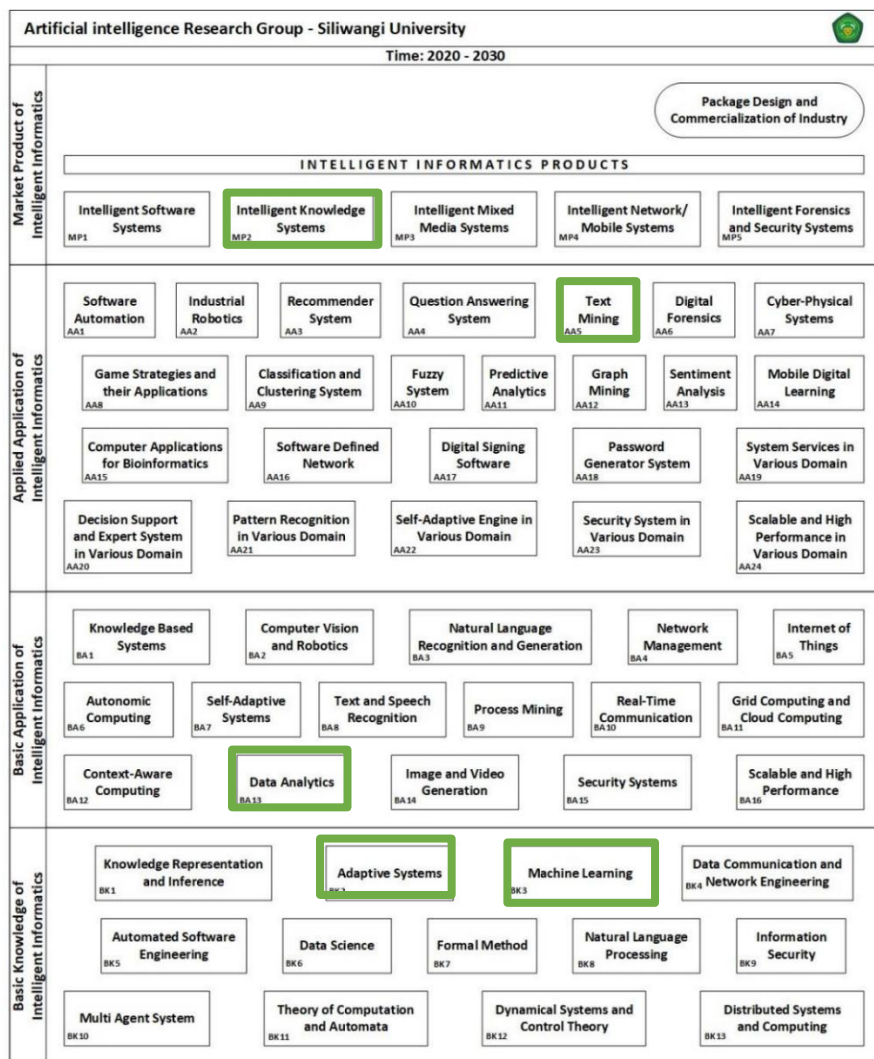


BAB III

METODOLOGI PENELITIAN

3.1 Roadmap Penelitian

Penelitian yang dijelaskan dalam penelitian ini sejalan dengan *Roadmap Artificial Intelligence Research Group – Universitas Siliwangi tahun 2020 – 2030* yang dapat dilihat pada Gambar 3.1

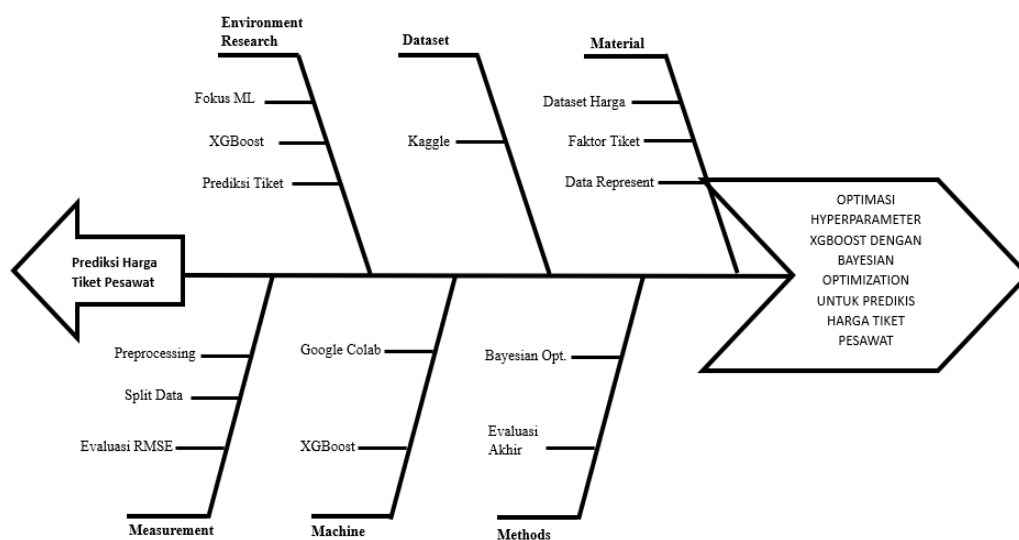


Gambar 3. 1 RoadMap Penelitian (AIS Universitas Siliwangi, 2020)

Penelitian ini juga disusun dengan mengacu pada *Roadmap Artificial Intelligence Research Group – Universitas Siliwangi 2020–2030*, sebagaimana ditunjukkan pada Gambar 3.1, sehingga memiliki keterkaitan langsung dengan arah pengembangan riset yang telah ditetapkan. yang telah dipublikasikan oleh universitas. Menyoroti beragam topik dalam *Intelligent Informatics Products*. Beberapa area yang tercakup di antaranya *Predictive Analytics*, *Machine learning*, *Data Analytics*, dan Optimasi Model (sering disebut *Adaptive systems*).

Penelitian yang dilakukan berada dalam ranah *Machine Learning* dan *Predictive Analytics* dengan tujuan untuk membangun *Intelligent Knowledge Systems*. Proses data *analytics* dalam penelitian ini mencakup tahapan akuisisi data, pembersihan data, serta pra-pemrosesan data untuk memastikan kualitas data yang digunakan dalam pemodelan. Selanjutnya, pemodelan dilakukan menggunakan algoritma *Extreme Gradient Boosting (XGBoost)* yang termasuk dalam pendekatan *ensemble learning* dan merupakan bagian dari *adaptive systems*, sehingga model mampu menyesuaikan diri terhadap pola data yang kompleks dan memberikan performa prediksi yang optimal pada berbagai kondisi. Evaluasi model dilakukan secara menyeluruh menggunakan metrik regresi untuk menilai tingkat akurasi dan kemampuan generalisasi model. Dengan demikian, seluruh rangkaian proses mulai dari pemilihan dataset, tahapan pengolahan data, pemodelan, hingga evaluasi model berkontribusi secara langsung dalam mendukung pencapaian tujuan *Intelligent Knowledge Systems* sesuai dengan roadmap penelitian yang ditunjukkan pada gambar.

Fishbone Diagram merupakan alat yang efektif untuk mengidentifikasi, mengeksplorasi, dan menganalisis berbagai faktor yang menjadi sumber inovasi serta evolusinya, termasuk teknologi yang memiliki tujuan umum (Coccia, 2020). *Fishbone Diagram* juga memfasilitasi pemetaan hubungan sebab akibat antara berbagai faktor penyebab, sehingga memudahkan identifikasi area kritis untuk perbaikan dan pengembangan strategi inovasi yang lebih terintegrasi. *Fishbone* pada penelitian diadaptasi dari peneliti sebelumnya yang dilakukan oleh (Ferreira, 2021), dengan beberapa perubahan yang menyesuaikan penelitian yang dilakukan seperti pada gambar 3.2.



Gambar 3. 2 *Fishbone* Diagram

Gambar 3.2 menampilkan *fishbone* diagram yang mengilustrasikan komponen-komponen utama dalam penelitian optimasi *Hyperparameter XGBoost* dengan *Bayesian Optimization* untuk Prediksi Harga Tiket Pesawat. Diagram ini menunjukkan keterkaitan antara beberapa aspek penting, yaitu:

1. *Environment Research*

Berfokus pada bidang *Machine learning*, *Data Analytics*, dan *Predictive Analytics* yang mendasari tujuan penelitian. Penerapan *XGBoost* untuk memprediksi harga tiket pesawat menjadi bagian dari upaya pengembangan *Intelligent Knowledge Systems*.

2. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari platform *Kaggle*, yaitu *Flight Price Prediction* yang bersumber dari situs *EaseMyTrip*. *Dataset* yang telah dibersihkan atau *Clean_Dataset.csv* terdiri dari 300.261 baris dan 11 fitur, termasuk informasi maskapai, rute penerbangan, waktu keberangkatan dan kedatangan, jumlah pemberhentian, kelas penerbangan, durasi perjalanan, hari tersisa sebelum keberangkatan, serta harga tiket sebagai variabel target. Fitur-fitur tersebut merepresentasikan faktor utama yang memengaruhi harga tiket dan sangat relevan untuk membangun model prediksi. Ketersediaan data yang lengkap dan terstruktur mendukung pengembangan model *XGBoost* yang akan dioptimalkan menggunakan *Bayesian Optimization*.

3. *Material*

Mencakup dataset harga tiket pesawat beserta faktor-faktor yang memengaruhi perubahan harga seperti waktu pemesanan, musim, dan tujuan penerbangan. Kualitas serta representasi data memengaruhi kemampuan model untuk melakukan generalisasi dengan baik.

4. *Measurement*

Menitikberatkan pada proses *preprocessing* misalnya pembersihan data, normalisasi, dan penanganan data yang tidak lengkap, pembagian data (*splitting*) menjadi *training* dan *testing*, serta evaluasi awal melalui metrik seperti RMSE. Tahap ini memastikan data siap digunakan untuk pelatihan model dan penilaian kinerja.

5. *Machine*

Mengacu pada perangkat atau platform yang digunakan untuk menjalankan eksperimen, yakni *Google Colab* atau *Kaggle Notebook*. Ketersediaan sumber daya komputasi yang memadai memungkinkan proses pelatihan *XGBoost* berjalan lebih efisien.

6. *Methods*

Menunjukkan pendekatan yang diambil untuk mencapai performa model yang optimal, yaitu *Bayesian Optimization* untuk menyetel *Hyperparameter XGBoost* secara efektif. Selain itu, aspek evaluasi akurasi juga menjadi perhatian agar model dapat diandalkan.

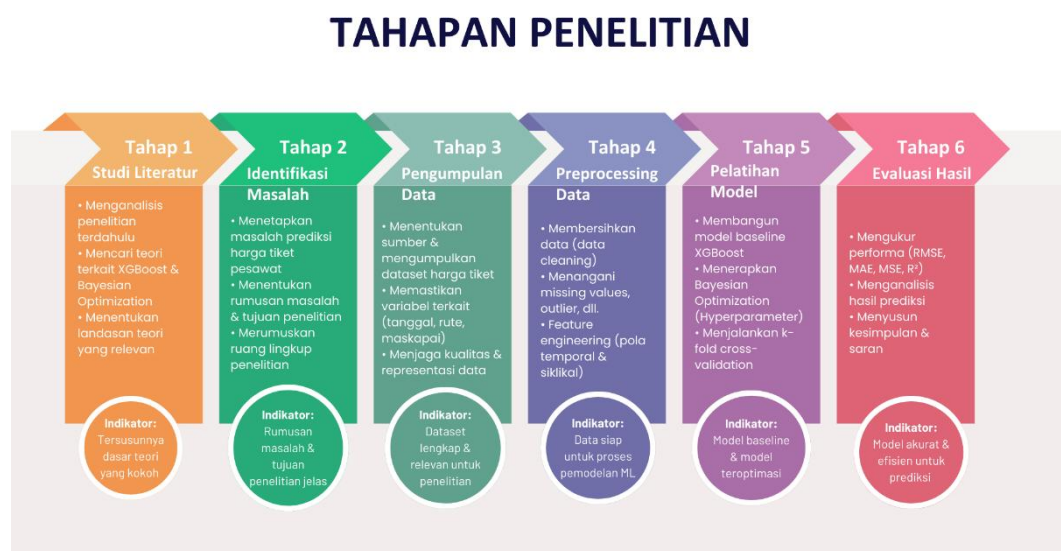
Bagian tulang ikan yang mengarah ke Optimasi *Hyperparameter XGBoost* menegaskan bahwa seluruh faktor (*Environment Research, Dataset, Material, Measurement, Machine, dan Methods*) saling berkaitan untuk meningkatkan kinerja prediksi harga tiket pesawat.

Penelitian ini dirancang untuk mendukung pencapaian tujuan pada bidang *Machine learning* dan *predictive analytics* dalam *roadmap* penelitian. Dataset harga tiket melalui tahapan akuisisi dan *preprocessing* pada bagian *Measurement*, lalu

diolah menggunakan pendekatan *XGBoost* untuk bagian *Methods* yang dioptimasi dengan *Bayesian Optimization*. Platform komputasi seperti *Google Colab* pada bagian *Machine* memungkinkan proses komputasi berjalan efisien. Dengan demikian, penelitian ini berkontribusi pada pengembangan *Intelligent Knowledge Systems* melalui pemodelan prediktif yang akurat dan andal dalam memproyeksikan harga tiket pesawat.

3.2 Tahapan Penelitian

Tahapan penelitian yang diterapkan dimulai dari studi literatur, lalu identifikasi masalah, pengumpulan data, *preprocessing* data, pelatihan model, evaluasi hasil. Rangkaian tahapan penelitian ini dapat diilustrasikan pada Gambar 3.3



Gambar 3. 3 Tahapan Penelitian

Gambar 3.3 menampilkan rangkaian tahapan penelitian yang dilakukan secara sistematis, mulai dari studi literatur, akuisisi data, pra-pemrosesan data,

pemodelan, evaluasi model, hingga penarikan kesimpulan. Tahapan-tahapan ini dirancang untuk memastikan penelitian berjalan terstruktur, sehingga hasil akhir yang diperoleh dapat diandalkan dan sesuai dengan tujuan penelitian. Gambar 3.3 menunjukkan enam tahapan utama yang dilalui dalam penelitian ini, yaitu:

3.2.1 Studi Literatur

Pada tahap ini, dilakukan pencarian dan penelaahan referensi dari jurnal, buku, maupun karya ilmiah lain yang relevan dengan topik penelitian. Tujuannya adalah untuk memahami konsep dasar, metode yang telah dikembangkan, serta mengidentifikasi gap penelitian.

3.2.2 Identifikasi Masalah

Tahap ini bertujuan merumuskan secara spesifik permasalahan yang ingin diselesaikan. enentukan fokus penelitian, ruang lingkup, serta tujuan yang hendak dicapai, sehingga arah penelitian menjadi jelas.

3.2.3 Pengumpulan Data

Dataset yang digunakan berasal dari platform *Kaggle* dengan judul '*Flight Price Prediction*', disediakan oleh *EaseMyTrip*. Data terdiri atas 300.261 baris dan 11 fitur, mencakup informasi maskapai, rute penerbangan, waktu keberangkatan dan kedatangan, jumlah pemberhentian, kelas penerbangan, durasi perjalanan, hari menuju keberangkatan, serta harga tiket sebagai variabel target. Sebelum digunakan dalam pelatihan model, data melalui proses pembersihan dan pra-pemrosesan untuk memastikan kualitas dan konsistensinya. Dataset ini dipilih karena strukturnya sesuai untuk membangun model prediktif harga tiket pesawat berbasis *XGBoost*,

dengan optimasi *Hyperparameter* menggunakan *Bayesian Optimization* guna mencapai hasil yang lebih akurat dan efisien.

3.2.4 *Preprocessing Data*

Pada tahap ini, data dipersiapkan untuk digunakan dalam proses pelatihan model. Data mentah yang diperoleh dari berbagai sumber umumnya masih mengandung banyak masalah, seperti nilai yang hilang, anomali atau *outlier*, atau format yang tidak seragam. Oleh karena itu, proses *preprocessing* menjadi langkah penting agar model pembelajaran mesin dapat memahami pola data dengan lebih baik. Tahapan *preprocessing* yang diterapkan pada penelitian ini terdiri dari beberapa langkah yakni sebagai berikut:

1. *Data Cleaning*

Langkah awal ini bertujuan untuk memastikan data yang digunakan bebas dari kesalahan atau ketidaksesuaian, seperti baris ganda, format yang salah, atau nilai kosong yang tidak berarti. Dengan membersihkan data terlebih dahulu, risiko model belajar dari informasi yang keliru dapat diminimalkan.

2. *Handling Missing Values*

Kehilangan data bisa berdampak besar terhadap hasil pelatihan. Untuk mengatasinya, digunakan teknik seperti imputasi, yaitu menggantikan nilai kosong dengan nilai rata-rata, median, atau metode lain yang sesuai dengan tipe data tersebut. Tujuannya adalah agar data tetap lengkap tanpa mengorbankan kualitas informasi.

3. *Feature Engineering*

Tahap ini bertujuan untuk mengubah data mentah menjadi fitur-fitur yang lebih bermakna bagi model. Misalnya, mengubah data kategori menjadi bentuk numerik, menambahkan fitur waktu seperti hari libur, akhir pekan, atau musim, serta mengenali pola berulang dalam data waktu. Proses ini membantu model mengenali hubungan yang berpengaruh terhadap fluktuasi harga tiket pesawat.

3.2.5 Pelatihan Model

Pada tahap pelatihan model, data yang sudah melalui tahap pra pemrosesan dilatih menggunakan algoritma *XGBoost*, dioptimasi menggunakan *Bayesian Optimization*. Tahap ini juga mencakup proses validasi, seperti *cross-validation*, guna memastikan model tidak *overfitting*. Berikut pada tabel 3.1 Merupakan rentang uji dari *Bayesian Optimization* yang di pakai.

Tabel 3. 1 *Bayesian Optimization* pada *XGBoost*

<i>Hyperparameter</i>	<i>Range Explored</i>
<i>n_estimators</i>	80, 200
<i>max_depth</i>	5, 12
<i>learning_rate</i>	0.01, 0.2, prior='log-uniform'
<i>subsample</i>	0.7, 1.0
<i>colsample_bytree</i>	0.7, 1.0

Kisaran parameter yang dieksplorasi dalam *Hyperparameter Tuning* yang ditampilkan pada Tabel 3.1 merupakan: *n_estimators* berkisar antara 80 hingga 200, *max_depth* berada dalam rentang 5 hingga 12, *learning_rate* dieksplorasi pada 0.01

hingga 0.2 dengan distribusi prior 'log-uniform', *subsample* berada dalam kisaran 0.7 hingga 1.0, serta *colsample_bytree* juga berada dalam rentang 0.7 hingga 1.0.

3.2.6 Evaluasi Hasil

Model yang telah dilatih dievaluasi menggunakan metrik evaluasi seperti RMSE, MSE, MAE, atau R^2 . Hasil evaluasi kemudian dianalisis untuk mengetahui keunggulan dan kelemahan model, sehingga dapat menyimpulkan keberhasilan penelitian dan merumuskan rekomendasi untuk pengembangan selanjutnya.

Keseluruhan tahapan ini dirancang agar penelitian berjalan secara sistematis, mulai dari dasar teori hingga evaluasi akhir, sehingga diharapkan dapat menghasilkan model yang akurat, efisien, dan sesuai dengan tujuan penelitian.