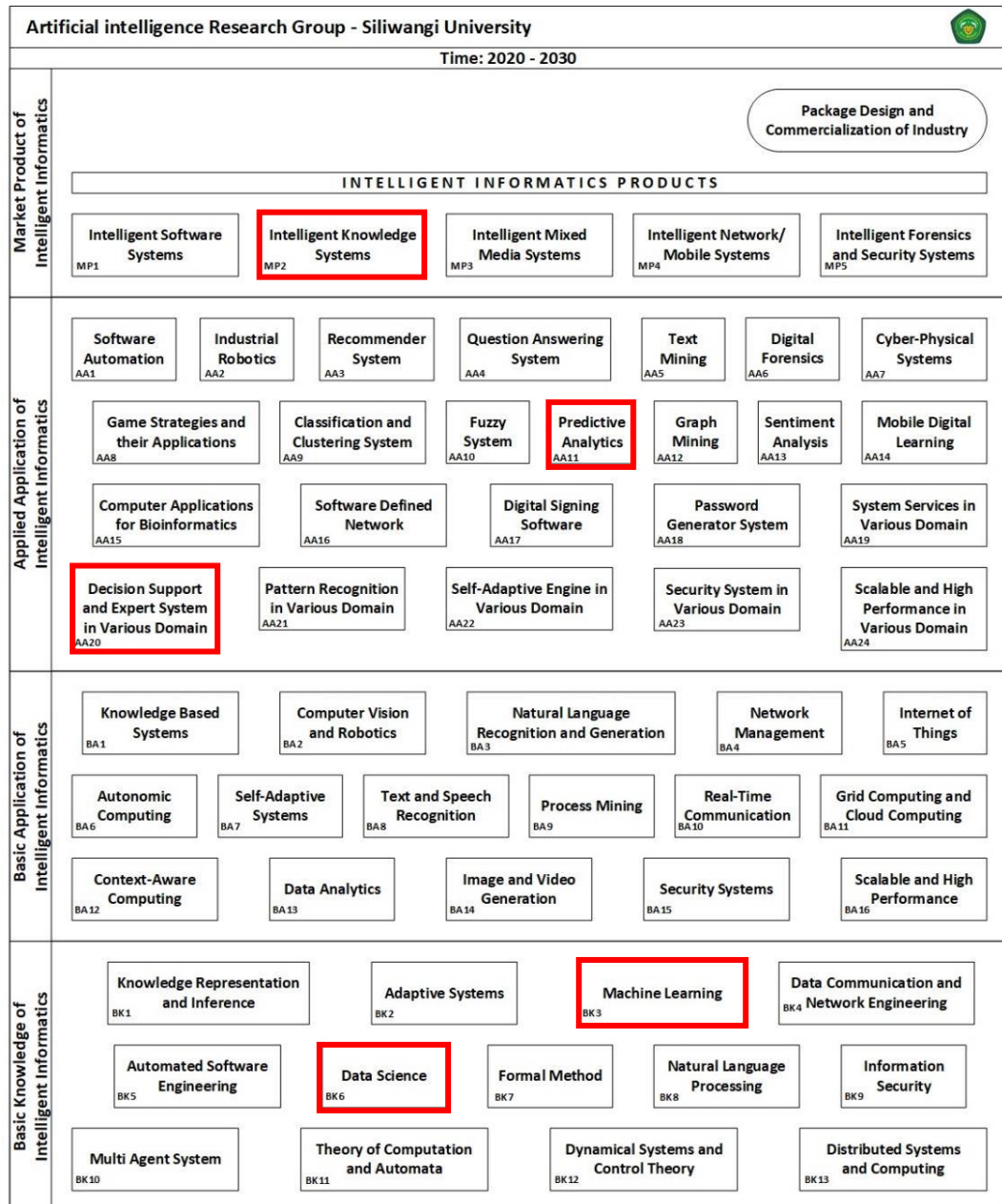


## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Roadmap Penelitian

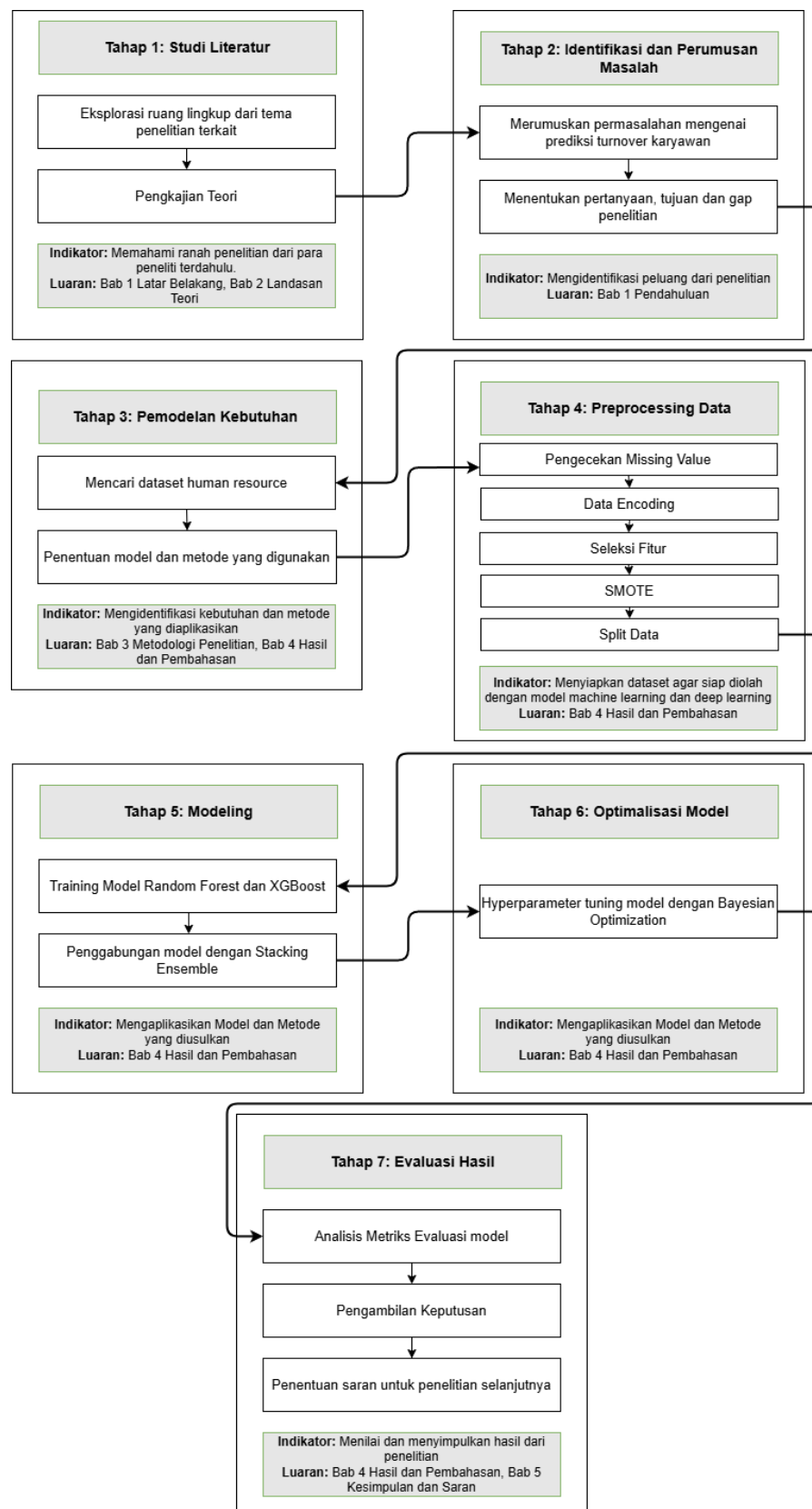


Gambar 3. 1 Roadmap Penelitian (AIS Universitas Siliwangi, 2020)

Penelitian yang dijelaskan dalam proposal ini sejalan dengan Roadmap Artificial Intelligence Research Group – Universitas Siliwangi tahun 2020 – 2030, sebagaimana terlihat pada Gambar 3.1 yang telah dipublikasikan oleh universitas. Gambar 3.1 menampilkan roadmap penelitian Artificial Intelligence Research Group – Siliwangi University periode 2020–2030, yang menyoroti beragam topik dalam *Intelligent Informatics Products*. Penelitian ini berfokus pada kajian yang ditandai kotak warna merah, yaitu mengenai *Data Science* dan *Machine Learning*, yang dikembangkan dengan tujuan untuk mendapatkan fungsi sebagai *Predictive Analytics* dan diharapkan menjadi *Decision Support and Expert System* pada domain prediksi *turnover*, sehingga akan terbentuk *Intelligent Knowledge System*.

### 3.2 Tahapan Penelitian

Penelitian ini memiliki kebaruan pada penerapan metode *Stacking Ensemble* yang mengombinasikan *Random Forest* dan *XGBoost* sebagai *base learner*, serta penggunaan *Multi-Layer Perceptron* (MLP) sebagai *meta learner* untuk meningkatkan kemampuan model dalam mempelajari pola kompleks terkait *turnover* karyawan. Selain itu, penelitian ini juga menerapkan *Bayesian Optimization* untuk memperoleh konfigurasi *hyperparameter* terbaik sehingga performa model dapat ditingkatkan secara signifikan dibandingkan pendekatan tanpa *tuning*. Pendekatan ini tidak hanya menghasilkan model dengan akurasi lebih tinggi, tetapi juga memberikan pemahaman yang lebih mendalam mengenai faktor-faktor yang paling berpengaruh terhadap keputusan karyawan meninggalkan perusahaan. Tahapan penelitian secara keseluruhan ditunjukkan pada Gambar 3.2, Gambar 3.3 dan Gambar 3.4.



Gambar 3. 2 Tahapan Penelitian

Tahap 1 Studi Literatur, tahap ini bertujuan untuk menggali ruang lingkup penelitian dan memahami konsep-konsep dasar yang berkaitan dengan prediksi *turnover* karyawan. Tinjauan literatur dilakukan dengan mempelajari berbagai penelitian sebelumnya yang relevan, guna memperoleh pemahaman yang lebih komprehensif. Hasil dari proses ini akan menjadi dasar teoritis yang mendukung penelitian, yang kemudian dituangkan dalam Bab I (Latar Belakang) dan Bab II (Tinjauan Pustaka).

Tahap 2 Identifikasi dan Perumusan Masalah Penelitian, pada tahap ini dilakukan identifikasi kesenjangan penelitian (*research gap*) berdasarkan kajian literatur yang telah dilakukan. Selanjutnya, dirumuskan permasalahan penelitian yang berkaitan dengan prediksi *turnover* karyawan. Tahap ini juga mencakup penetapan rumusan pertanyaan penelitian (*Research Question*), tujuan penelitian (*Research Objective*), serta penentuan batasan ruang lingkup penelitian. Semua hasil dari tahap ini akan disajikan dalam Bab I (Pendahuluan).

Tahap 3 Pemodelan Kebutuhan, pada tahap ini dilakukan pencarian dataset yang relevan dengan *turnover* karyawan, serta pemilihan model dan metode yang akan diterapkan dalam penelitian. Tujuan dari pemodelan kebutuhan ini adalah untuk memastikan bahwa data yang digunakan selaras dengan tujuan penelitian. Hasil dari proses ini akan dicatat dalam Bab III (Metodologi Penelitian) dan Bab IV (Hasil dan Pembahasan).

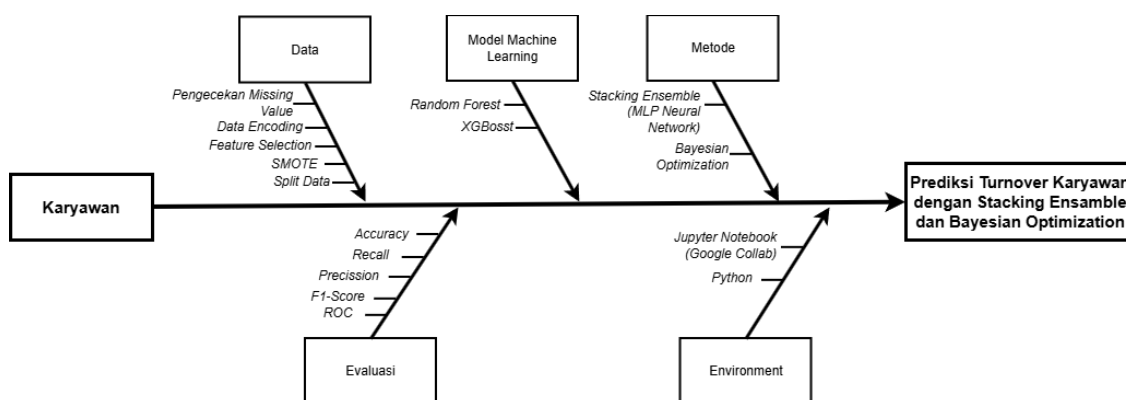
Tahap 4 Pra-pemrosesan Data, dataset diproses melalui beberapa langkah utama, dimulai dengan pengecekan missing value untuk memastikan tidak ada data

kosong. Selanjutnya dilakukan data *encoding* pada fitur kategorikal agar dapat diproses oleh algoritma. Proses dilanjutkan dengan seleksi fitur berdasarkan korelasi dan relevansi untuk mengurangi redundansi. Setelah itu diterapkan SMOTE guna menangani ketidakseimbangan kelas pada variabel target. Tahap akhir adalah pembagian data menjadi data latih dan data uji dengan rasio 70:30 agar pemodelan dapat dilakukan secara objektif, di mana 70% data digunakan sebagai data pelatihan dan 30% sebagai data pengujian. Seluruh proses pra-pemrosesan ini dijelaskan lebih lanjut pada Bab IV (Hasil dan Pembahasan).

Tahap 5 Pemodelan, setelah data siap, tahap selanjutnya adalah pembangunan model dengan menggunakan algoritma *Random Forest* dan *XGBoost*. Selain itu, dilakukan eksplorasi penggabungan model menggunakan pendekatan *Stacking Ensemble* untuk meningkatkan performa prediksi, di mana lapisan *meta-learner* yang digunakan adalah *Multi-Layer Perceptron* (MLP). Hasil dari tahap ini berupa implementasi model yang terdokumentasi dalam Bab IV (Hasil dan Pembahasan).

Tahap 6 Optimasi Model, proses optimasi model ini dilakukan melalui *Hyperparameter Tuning* menggunakan *Bayesian Optimization*. Optimasi difokuskan hanya pada model *Stacking Ensemble*, khususnya pada konfigurasi *meta-learner* MLP, dengan tujuan memperoleh kombinasi parameter yang memberikan performa prediksi terbaik. Hasil optimasi tersebut kemudian dibandingkan dengan ketiga model sebelumnya untuk melihat perbedaan kinerja. Seluruh hasil proses ini disajikan pada Bab IV (Hasil dan Pembahasan).

Tahap 7 Evaluasi Hasil, tahap terakhir dari penelitian ini berfokus pada evaluasi kinerja model klasifikasi menggunakan metrik evaluasi yang relevan, yaitu *accuracy*, *precision*, *recall*, *F1-score* dan *ROC-AUC*. Selain itu, disusun rekomendasi serta perbaikan untuk penelitian selanjutnya. Luaran dari tahap ini disajikan dalam Bab IV (Hasil dan Pembahasan) dan Bab V (Kesimpulan dan Saran).



Gambar 3.3 Fishbone Diagram

Bagian ekor diagram menggambarkan fokus utama penelitian, yaitu fenomena *turnover* karyawan yang menjadi dasar pembangunan model prediksi. Bagian kepala diagram menunjukkan target akhir yang ingin dicapai, yakni menghasilkan model prediksi *turnover* karyawan yang lebih akurat dan stabil melalui pendekatan *Stacking Ensemble* yang diperkuat dengan *Bayesian Optimization*. Alur dari ekor menuju kepala memperlihatkan bagaimana berbagai faktor mulai dari data, pemodelan, metode, evaluasi, hingga lingkungan pemrograman berkontribusi langsung terhadap pencapaian tujuan penelitian.

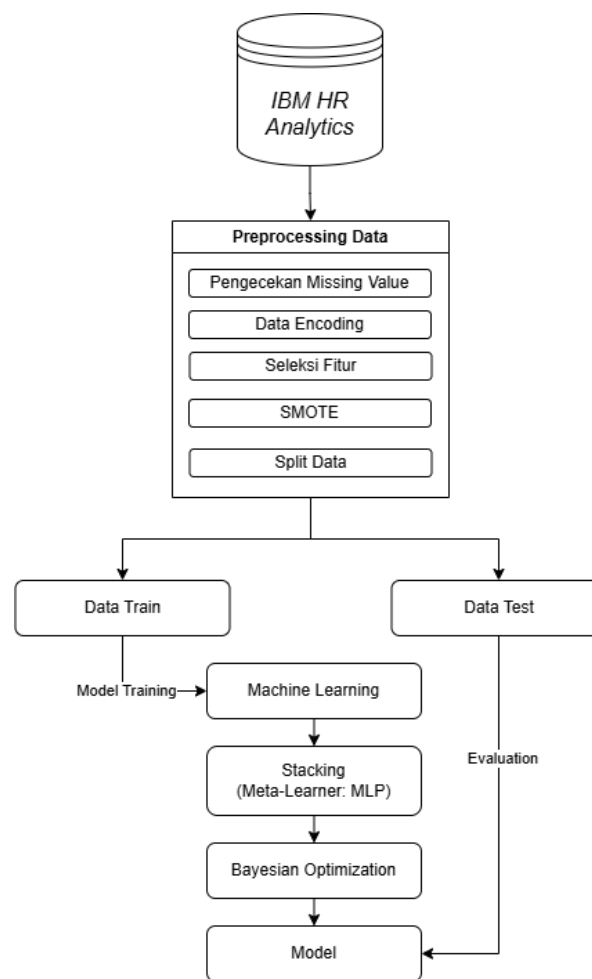
Tahapan *preprocessing* pada diagram mencakup proses penting seperti pengecekan *missing value*, *encoding* variabel kategorikal, pemilihan fitur,

penerapan SMOTE untuk menyeimbangkan kelas, serta pemisahan data menjadi set pelatihan dan pengujian. Seluruh langkah tersebut bertujuan memastikan kualitas data berada pada kondisi optimal sehingga model dapat belajar pola *turnover* secara lebih efektif.

Bagian pemodelan menampilkan penggunaan algoritma dasar *Random Forest* dan *XGBoost* yang kemudian dikombinasikan melalui *Stacking Ensemble* dengan *meta-learner MLPClassifier*. Setelah model dasar dibentuk, proses optimasi dilakukan menggunakan *Bayesian Optimization* guna memperoleh konfigurasi *hyperparameter* terbaik dan meningkatkan performa model secara keseluruhan.

Pada tahap Evaluasi, kinerja model diukur menggunakan metrik akurasi, presisi, *recall*, *F1-score*, dan AUC-ROC untuk memberikan gambaran yang menyeluruh terkait kemampuan model dalam membedakan karyawan yang bertahan dan yang berpotensi keluar. Seluruh proses penelitian diimplementasikan pada lingkungan *Google Colab* menggunakan *Python*, dengan dukungan pustaka seperti *Pandas* dan *NumPy* untuk pengolahan data, *Scikit-learn* untuk pemodelan dan evaluasi, serta *Imbalanced-Learn* untuk penerapan SMOTE. Visualisasi hasil meliputi *confusion matrix*, *ROC curve*, dan *feature importance* dibuat menggunakan *Matplotlib* dan *Seaborn*. Penggunaan *tools* tersebut memastikan seluruh alur penelitian dapat dijalankan secara sistematis, akurat, dan mudah direplikasi.

Untuk memperjelas alur pemrosesan data dan tahapan pemodelan yang digunakan dalam penelitian ini, dirancang arsitektur sistem penelitian sebagaimana ditunjukkan pada Gambar 3.4.



Gambar 3. 4 Arsitektur Sistem

Gambar 3.4 menunjukkan arsitektur sistem penelitian yang menggambarkan alur pemrosesan data mulai dari dataset *IBM HR Analytics Employee Attrition & Performance* hingga evaluasi model. Data diproses melalui tahapan pra-pemrosesan, kemudian digunakan untuk membangun model *Random Forest* dan *XGBoost* yang dikombinasikan melalui *Stacking Ensemble* dengan MLP

sebagai *meta-learner*, serta dioptimasi menggunakan *Bayesian Optimization* sebelum dilakukan evaluasi kinerja model.