

BAB I

PENDAHULUAN

1.1 Latar Belakang

Keberhasilan lulusan memasuki dunia kerja merupakan salah satu indikator penting kualitas pendidikan tinggi. Pusat Karir memegang peran strategis dalam mempercepat masa tunggu lulusan, karena indikator ini berkontribusi langsung pada IKU 1: Lulusan Mendapat Pekerjaan yang Layak [1]. Untuk mengukur dan memantau capaian tersebut, perguruan tinggi secara rutin melaksanakan *tracer study*. *Tracer study* adalah studi mengenai lulusan lembaga penyelenggara pendidikan tinggi [2]. Dalam penerapan di Indonesia, *tracer study* menyediakan informasi penting tentang keterkaitan pendidikan tinggi dengan dunia kerja profesional [3]. Berdasarkan data *tracer study* salah satu Universitas tahun 2020-2023, terdapat 493 lulusan yang telah bekerja dengan rata-rata masa tunggu 4,08 bulan, dan 20,7% dari lulusan yang bekerja menunggu lebih dari 6 bulan dalam mendapat pekerjaan pertama [4].

Selain itu, Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) menempatkan waktu tunggu lulusan sebagai salah satu elemen penilaian dalam instrumen akreditasi program studi maupun institusi. Namun, praktik pengolahan *tracer study* di banyak perguruan tinggi Indonesia masih dominan bersifat deskriptif, sebagaimana terlihat pada laporan pelaksanaan *tracer study* di beberapa kampus dan format tabel BAN-PT. [5] Konsekuensinya belum tersedia perkiraan masa tunggu pada tingkat individu yang siap pakai untuk dimanfaatkan oleh pusat karir untuk merancang program / kegiatan yang relevan untuk mempercepat masa tunggu. Dengan demikian, terdapat kesenjangan antara ketersediaan data *tracer study* dan kebutuhan pengambilan keputusan berbasis prediksi yang mampu menghasilkan perkiraan masa tunggu dan menjelaskan faktor yang paling berperan dalam mempercepat masa tunggu.

Dari permasalahan tersebut, penelitian ini mengusulkan model prediksi masa tunggu menggabungkan model hibrida *stacking ElasticNet* dan *LightGBM* berdasarkan beberapa variabel terpilih. Dalam penerapannya, *ElasticNet* dipilih karena model tersebut dapat berperan sebagai penimbang variabel yang stabil pada data dengan banyak variabel yang saling berkorelasi / multikolinearitas. [6]. Hal

tersebut didukung dengan penelitian yang melakukan pemodelan variabel yang saling berkorelasi dengan memberi RMSE terkecil dari metode lain yang diuji [7]. Selain itu *ElasticNet* dikenal efektif untuk menangani multikolinearitas sekaligus melakukan seleksi variabel pada data dengan banyak fitur yang berkorelasi.[8].

Sementara itu, *LightGBM* dipilih karena berbagai studi sebelumnya menunjukkan kinerja yang baik pada hubungan non-linear, dan unggul dibandingkan dengan model lain seperti *logistic regression*, *elastic net*, *random forest*, dan *XGBoost* [9]. Hasil penelitian lain juga menegaskan bahwa model berbasis pohon, termasuk *LightGBM*, secara konsisten mengungguli model regresi dalam hal akurasi prediksi [10].

Pendekatan *stacking* sendiri merupakan teknik *ensemble* yang mengombinasikan beberapa pembelajar dasar (*base learner*) melalui sebuah pembelajar-meta (*meta-learner*) untuk menggabungkan prediksi. [11] Bukti empiris menunjukkan bahwa rancangan *stacking ElasticNet* dan *LightGBM* mampu meningkatkan kinerja prediksi, di mana model *stacking* mencapai nilai $R^2 = 0,9298$ lebih tinggi dibandingkan model tunggal *ElasticNet* $R^2 = 0,8475$ dan *LightGBM* $R^2 = 0,8011$ temuan ini mengindikasikan bahwa penggabungan prediksi model *stacking ElasticNet* dan *LightGBM* dapat memperbaiki akurasi sekaligus menurunkan error dibandingkan model tunggal. [8]

Penelitian lain yang menggabungkan metode *ElasticNet* dan *LightGBM* juga terlihat mengungguli seluruh model tunggal, seperti penelitian tentang prediksi kemiskinan kabupaten/kota di Jawa Tengah, di mana pendekatan *stacking* dilaporkan mengungguli seluruh model tunggal dengan $R^2 = 0,942$. Sementara *ElasticNet* sebagai model tunggal mencapai $R^2 = 0,543$ dan *LightGBM* memiliki $R^2=0,851$. [12] Selain itu, pada studi prediksi risiko klinis, *stacking ensemble* memperoleh AUC tertinggi 0,795 pada *testing set* dan melampaui model individual seperti *ElasticNet* (AUC 0,768) serta *LightGBM* (AUC 0,778), dengan akurasi model *stacking* sebesar 0,728. [13] Secara keseluruhan, temuan-temuan tersebut memperkuat argumen bahwa *stacking* layak dipilih untuk permasalahan prediksi yang dipengaruhi relasi data yang kompleks dan heterogen, karena mampu memanfaatkan karakteristik antar model untuk meningkatkan performa prediksi.

Dalam Implementasinya, hasil dari penelitian ini diharapkan dapat membantu Universitas memantau capaian, misal banyaknya lulusan yang waktu tunggu < 6 Bulan. Temuan variabel yang berpengaruh dalam masa tunggu dipakai untuk menargetkan kegiatan yang akan dilaksanakan oleh Universitas untuk menekan lulusan dengan waktu tunggu ≥ 6 bulan, serta mendukung skor akreditasi karena BAN-PT memberi nilai maksimum untuk waktu tunggu < 6 bulan (BAN-PT, 2019).

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang model prediksi masa tunggu berbasis data *tracer study* menggunakan pendekatan Hibrida *stacking* yang mengintegrasikan *ElasticNet* dan *LightGBM*?
2. Bagaimana kinerja pendekatan Hibrida *stacking ElasticNet* dan *LightGBM* dibandingkan dengan model tunggal *ElasticNet* dan model tunggal *LightGBM* berdasarkan evaluasi *k-fold cross-validation* menggunakan metrik R^2 , RMSE, dan MAE?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan penelitian ini adalah sebagai berikut:

1. Mengembangkan model hibrida *stacking* yang mengintegrasikan *ElasticNet* dan *LightGBM* untuk memprediksi masa tunggu berdasarkan variabel *tracer study* terpilih.
2. Membandingkan kinerja model hibrida *stacking* terhadap model tunggal *ElasticNet* dan model tunggal *LightGBM* menggunakan skema *k-fold cross-validation*.
3. Menyajikan interpretasi faktor-faktor dominan berdasarkan keluaran model.

1.4 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Akademik : Menyajikan rancangan model hibrida *stacking ElasticNet - LightGBM* yang dapat direplikasi pada penelitian prediksi berbasis data *tracer study*.
2. Universitas : Mendukung pemantauan capaian IKU dan akreditasi melalui pemanfaatan data *tracer study* secara prediktif untuk mengidentifikasi risiko masa tunggu yang panjang.
3. Pusat Karir : Menyediakan estimasi masa tunggu pada tingkat individu serta informasi faktor dominan yang berkorelasi dengan masa tunggu sebagai dasar penyusunan program dan intervensi yang lebih tepat sasaran.

1.5 Ruang Lingkup Penelitian

Agar pembahasan tetap fokus dan tidak melebar, ruang lingkup penelitian ini ditetapkan sebagai berikut:

1. Jumlah Data : 493 Data responden lulusan 2020-2023 dengan berstatus bekerja dari *tracer study* Universitas XYZ
2. Variabel Target : Masa Tunggu
3. Variabel Prediktor : Masa_studi, Provinsi, Prodi, IPK, TOEFL, JenKel, SSKM, N_KP, N_TA, Masa_carikerja, Jml_Lamaran, Magang
4. Metode : Model tunggal *ElasticNet*, model tunggal *LightGBM*, dan model hibrida *stacking ElasticNet-LightGBM*.
5. Penilaian Hasil : Menggunakan *5-fold cross-validation* untuk mengukur stabilitas. hasil dilaporkan menggunakan metrik R^2 , RMSE, dan MAE.
6. Sumber Data : Responden *tracer study* jenjang S1 dan D4 dengan masa studi minimum 7 semester.

1.6 Batasan Permasalahan

Untuk memperjelas batas pembahasan, batasan permasalahan dalam penelitian ini ditetapkan sebagai berikut:

1. Penelitian ini hanya menggunakan variabel prediktor yang tersedia pada data akademik dan *tracer study* Universitas XYZ, tanpa menambahkan variabel dari sumber data eksternal.
2. Penelitian tidak bertujuan menjelaskan hubungan sebab-akibat. Interpretasi yang disajikan terbatas pada kontribusi prediktor terhadap kinerja prediksi seperti *koefisien/importance* dalam konteks pemodelan.
3. Penelitian difokuskan pada pemodelan regresi untuk memprediksi masa tunggu dan tidak mencakup perumusan masalah klasifikasi.

1.7 Kontribusi Penelitian

Berdasarkan Kajian pustaka menunjukkan bahwa *tracer study* di Indonesia masih dominan dimanfaatkan sebagai laporan deskriptif, sehingga pengembangan model prediksi masa tunggu pada level individu masih terbatas. Di sisi lain, studi internasional mulai memanfaatkan *machine learning* untuk *outcome* lulusan, namun umumnya berfokus pada status kerja atau pendapatan, bukan durasi masa tunggu. Pada aspek metodologis, literatur juga menunjukkan pendekatan hibrida termasuk penggabungan model regularisasi dan *boosting* telah banyak diuji pada domain lain, tetapi penerapannya pada konteks pendidikan tinggi dan data *tracer study* masih jarang. Berangkat dari celah tersebut, kontribusi penelitian ini dirumuskan dalam tiga aspek, yaitu kontribusi teoritis, praktis, dan metodologis.

1.7.1 Kontribusi Teoritis

Kontribusi teoritis penelitian ini adalah memperkuat posisi *tracer study* sebagai instrumen prediktif. Dengan menjadikan masa tunggu sebagai target regresi dan membangun model prediksi pada level individu, penelitian ini memperkuat pemahaman bahwa masa tunggu merupakan salah satu indikator langsung efektivitas perguruan tinggi dalam menyiapkan lulusan siap kerja. Arah ini konsisten dengan gap *tracer study* yang menegaskan minimnya kajian prediktif masa tunggu individual pada konteks Indonesia.

Selain itu, penelitian ini memberi bukti empiris berbasis model bahwa sinyal prediktif dalam *tracer study* dapat didominasi oleh faktor tertentu, bukan merata pada semua variabel. Hasil interpretasi menunjukkan model terutama bertumpu pada Masa_Studi dan Magang sebagai prediktor dominan, sementara fitur lain berkontribusi lebih kecil sebagai pelengkap. Temuan ini mempertegas fokus teoritis pada faktor transisi studi-kerja dalam konteks data penelitian.

1.7.2 Kontribusi Praktis

Kontribusi praktis penelitian ini terletak pada tersedianya model dan luaran yang dapat dipakai sebagai alat bantu pengambilan keputusan bagi universitas. Penelitian mengaitkan masa tunggu dengan kebutuhan pemantauan akreditasi dan IKU, serta menekankan kebutuhan perkiraan masa tunggu di level individu untuk mendukung perencanaan program yang mendukung percepatan masa tunggu.

Luaran penelitian ini tidak berhenti pada ringkasan metrik, tetapi juga menyediakan bentuk lain yang dapat dibaca pada level individu sehingga universitas dapat memanfaatkan hasil prediksi untuk pemetaan risiko masa tunggu dan penentuan prioritas pendampingan.

1.7.3 Kontribusi Metodologis

Kontribusi metodologis penelitian ini adalah menghadirkan *pipeline* pemodelan regresi untuk data *tracer study* dimana seluruh proses dibangun dalam *pipeline* untuk mencegah data *leakage*, mencakup *imputasi* nilai hilang, standarisasi fitur numerik, *ordinal encoding* untuk variabel berjenjang, serta *one-hot encoding* untuk variabel nominal.

Kontribusi metodologis yang paling spesifik adalah implementasi *stacking* berbasis *out-of-fold* (OOF) dimana *ElasticNet* dan *LightGBM* digunakan sebagai base learner, sedangkan *ElasticNet* sebagai meta-learner yang dilatih dari prediksi OOF melalui *StackingRegressor*, sehingga *meta-learner* belajar dari prediksi yang dibentuk melalui validasi silang. Selanjutnya, penelitian menambahkan validasi eksternal cohort 2024 sebagai uji generalisasi pada data berbeda. Hal tersebut memperkuat kewajaran metodologi dan memisahkan klaim performa internal dari kemampuan generalisasi.