

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Efisiensi kampanye pemasaran, khususnya *telemarketing*, menjadi prioritas strategis bagi perbankan untuk menekan biaya operasional yang tinggi [1]. Algoritma XGBoost telah menjadi standar industri dalam berbagai aplikasi klasifikasi karena akurasi yang superior [2]. Namun, evaluasi model ini seringkali hanya terpaku pada metrik label akhir tanpa memperhatikan kualitas probabilitas prediksi. Padahal, keputusan bisnis berbasis prioritas nasabah membutuhkan estimasi peluang yang presisi, bukan sekadar prediksi biner.

Permasalahan mendasar muncul karena algoritma *boosting* cenderung menghasilkan prediksi *overconfident* akibat minimasi *margin error* yang agresif [3]. Distorsi probabilitas ini diperburuk oleh karakteristik *class imbalance* pada data pemasaran bank, yang menyebabkan bias estimasi risiko secara sistematis [4]. Tanpa mekanisme kalibrasi yang tepat, model rentan memberikan skor kepercayaan tinggi secara keliru pada nasabah non-potensial [5], sehingga meningkatkan risiko kesalahan fatal dalam pengambilan keputusan bisnis berbasis prioritas.

Mitigasi bias umumnya dilakukan menggunakan metode pasca-pemrosesan standar seperti *Platt Scaling* atau *Isotonic Regression*. Namun, kedua metode ini memiliki keterbatasan fundamental; *Platt Scaling* terbatas pada asumsi distribusi error berbentuk *sigmoid* yang kaku, sedangkan *Isotonic Regression* rentan mengalami *overfitting* pada data terbatas [6]. Penggunaan metode tunggal seringkali gagal menangani pola distorsi probabilitas yang kompleks secara optimal, sehingga diperlukan pendekatan yang lebih *robust* [7].

Berdasarkan keterbatasan tersebut, penelitian ini mengajukan pendekatan *Hybrid Calibration Ensemble* yang mengintegrasikan metode parametrik dan non-parametrik untuk meminimalkan *miscalibration* melalui prinsip *ensemble learning* [8]. Metode gabungan ini bertujuan menyeimbangkan stabilitas dan fleksibilitas model agar menghasilkan estimasi risiko yang lebih adaptif terhadap variasi data.

Selanjutnya, efektivitas peningkatan reliabilitas model dievaluasi secara kuantitatif menggunakan metrik *Brier Score* dan *Estimated Calibration Error* untuk memverifikasi validitas probabilitas yang dihasilkan [9].

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan metode *Hybrid Calibration Ensemble* pada algoritma *XGBoost* dalam konteks permasalahan *miscalibration* pada dataset pemasaran bank yang tidak seimbang (*imbalanced*)?
2. Bagaimana perbandingan reliabilitas probabilitas antara model *XGBoost* tanpa kalibrasi (*baseline uncalibrated*) dan model dengan pendekatan *Hybrid Calibration Ensemble* yang diukur menggunakan *Estimated Calibration Error* dan *Brier Score*?

1.3 Batasan Masalah

1. Dataset yang digunakan merupakan dataset sekunder *Bank Marketing* yang dipublikasikan melalui Kaggle. Dataset ini digunakan sebagai representasi data klasifikasi biner yang bersifat *imbalanced*, tanpa bermaksud melakukan generalisasi pada konteks perbankan di wilayah tertentu.
2. Algoritma klasifikasi yang digunakan sebagai *base learner* dibatasi hanya pada *Extreme Gradient Boosting (XGBoost)*. Penelitian ini tidak membandingkan performa *XGBoost* dengan algoritma klasifikasi lain, melainkan fokus pada perbaikan kualitas probabilitasnya.
3. Sebagai pembandingan (*baseline*), penelitian ini menggunakan model *XGBoost* tanpa proses kalibrasi (*uncalibrated*), sehingga analisis difokuskan pada perbandingan reliabilitas probabilitas antara model *baseline* dan model dengan pendekatan *Hybrid Calibration Ensemble*.
4. Evaluasi performa model difokuskan pada metrik reliabilitas probabilitas (*reliability*), yaitu *Brier Score*, *ECE* dan visualisasi *Reliability Diagram*. Metrik klasifikasi standar seperti Akurasi, Presisi, atau *F1-Score* hanya digunakan sebagai metrik pendukung, bukan indikator keberhasilan utama.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan utama penelitian ini adalah menerapkan pendekatan *Hybrid Calibration Ensemble* pada model *XGBoost* sebagai metode kalibrasi probabilitas pasca pelatihan dan membandingkannya dengan model *XGBoost* tanpa kalibrasi sebagai *baseline*. Penelitian ini bertujuan untuk menganalisis dampak penerapan pendekatan *hybrid* terhadap kualitas probabilitas prediksi, khususnya pada dataset pemasaran bank yang bersifat *imbalanced*. Evaluasi dilakukan secara empiris menggunakan *Brier Score* untuk mengukur peningkatan reliabilitas probabilitas antara model *baseline uncalibrated* dan model dengan *Hybrid Calibration Ensemble*.

1.5 Manfaat Penelitian

Secara teoritis, hasil penelitian ini diharapkan dapat memperkaya literatur ilmu data (*Data Science*), khususnya mengenai efektivitas metode *ensemble* dalam menangani masalah kalibrasi probabilitas pada dataset dengan karakteristik tidak seimbang (*imbalanced data*). Selain itu, penelitian ini memberikan bukti empiris mengenai keunggulan pendekatan *Hybrid Calibration Ensemble* dibandingkan metode kalibrasi tunggal dalam menghasilkan estimasi risiko yang lebih reliabel.

Secara praktis, manfaat penelitian ini ditujukan bagi industri perbankan dan peneliti selanjutnya. Bagi industri perbankan, model kalibrasi yang dihasilkan dapat dimanfaatkan sebagai sistem pendukung keputusan (*Decision Support System*) untuk menentukan prioritas nasabah secara lebih presisi, yang berdampak langsung pada peningkatan efisiensi biaya operasional kampanye pemasaran (telemarketing) dan optimalisasi sumber daya manusia. Sementara itu, bagi peneliti selanjutnya, kajian ini dapat menjadi referensi metodologis dan rujukan ilmiah bagi penelitian lanjutan yang berfokus pada perbaikan kualitas probabilitas algoritma *Machine Learning*, khususnya pada kasus klasifikasi risiko keuangan atau domain lain yang memiliki distribusi data tidak seimbang.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran menyeluruh mengenai isi dan alur pembahasan dalam penelitian. Uraian singkat mengenai isi masing-masing bab adalah sebagai berikut:

BAB I PENDAHULUAN, berisi gambaran umum penelitian yang meliputi latar belakang masalah mengenai urgensi kalibrasi probabilitas pada data perbankan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian bagi aspek teoritis maupun praktis, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA, memaparkan teori-teori dasar yang mendukung penelitian, meliputi konsep *Imbalanced Data*, algoritma *Extreme Gradient Boosting (XGBoost)*, teori kalibrasi probabilitas (*Probability Calibration*), serta tinjauan pustaka mengenai metode *Platt Scaling*, *Isotonic Regression*, dan metrik evaluasi *Brier Score*. Bab ini juga memuat penelitian terkait (*related works*) yang menjadi landasan pengembangan metode *Hybrid*.

BAB III METODE PENELITIAN, menjelaskan tahapan dan metodologi yang digunakan dalam menyelesaikan masalah penelitian. Pembahasan meliputi deskripsi dataset *Bank Marketing*, alur pra-pemrosesan data (*preprocessing*), skenario penanganan data tidak seimbang, perancangan metode usulan *Hybrid Calibration Ensemble*, serta prosedur pengujian dan evaluasi model.

BAB IV HASIL DAN PEMBAHASAN, menguraikan hasil eksperimen komputasi yang telah dilakukan. Penulis menyajikan analisis perbandingan performa antara model tanpa kalibrasi, model dengan kalibrasi standar, dan model dengan metode usulan *Hybrid*. Pembahasan difokuskan pada analisis grafik *Reliability Diagram* dan nilai *Brier Score* untuk memvalidasi efektivitas solusi yang ditawarkan.

BAB V PENUTUP, kesimpulan yang merangkum hasil temuan penelitian berdasarkan tujuan yang telah ditetapkan, serta saran-saran konstruktif bagi pengembangan penelitian selanjutnya di masa mendatang.