

BUSINESS REQUIREMENT DOCUMENT (BRD)

Evaluasi Kinerja dan Interpretabilitas Model XGBoost Terkalibrasi untuk Prediksi Risiko Diabetes menggunakan Explainable AI SHAP dan LIME

Versi Dokumen: 2.0

Tanggal: 27 Oktober 2025

Disusun Oleh: Nadia Recilia Dahmi – 20220801041

1. Executive Summary

Dokumen ini menjelaskan kebutuhan dan spesifikasi untuk penelitian yang berfokus pada pengembangan serta evaluasi model *Machine Learning* berbasis XGBoost dalam memprediksi risiko diabetes. Fokus utama penelitian ini adalah memastikan model tidak hanya memiliki akurasi tinggi, tetapi juga mampu memberikan estimasi probabilitas yang terkalibrasi dan penjelasan hasil yang transparan melalui metode *Explainable AI* (XAI), yaitu SHAP (SHapley Additive exPlanations) dan LIME (Local Interpretable Model-Agnostic Explanations).

Penelitian ini dilatarbelakangi oleh fakta bahwa banyak model prediktif medis yang berfungsi seperti “kotak hitam” (*black box*) — meskipun akurat, namun sulit dijelaskan kepada tenaga medis dan tidak dapat memberikan tingkat kepercayaan probabilitas yang realistis. Oleh karena itu, penelitian ini berupaya membangun framework model XGBoost yang terkalibrasi dan terinterpretasi, dengan hasil yang dapat dipercaya, mudah dipahami, serta siap diintegrasikan dalam sistem pendukung keputusan klinis.

Output dari penelitian ini meliputi model terkalibrasi, pipeline evaluasi, dan laporan teknis komprehensif yang mencakup kinerja model, hasil kalibrasi probabilitas, serta analisis interpretabilitas menggunakan SHAP dan LIME. Hasil akhir diharapkan dapat menjadi kontribusi nyata dalam pengembangan sistem kecerdasan buatan yang transparan, akuntabel, dan sesuai untuk penerapan di bidang kesehatan.

2. Latar Belakang dan Permasalahan (Business Case)

Latar Belakang

Prediksi dini terhadap risiko diabetes merupakan komponen penting dalam strategi pencegahan penyakit kronis. Model berbasis *tree ensemble* seperti XGBoost telah terbukti unggul dalam hal akurasi dan efisiensi, namun masih menghadapi dua kelemahan mendasar: (1) probabilitas prediksi yang tidak selalu mencerminkan risiko sebenarnya (tidak terkalibrasi), dan

(2) kurangnya transparansi dalam proses pengambilan keputusan model.

Keterbatasan ini menyebabkan kesulitan dalam penerapan hasil model pada konteks medis, di mana penjelasan dan keandalan prediksi memiliki nilai yang sama pentingnya dengan akurasi.

Permasalahan Utama

1. Probabilitas Tidak Terkalibrasi — Model cenderung menghasilkan estimasi risiko yang terlalu tinggi (*overconfident*) atau terlalu rendah (*underconfident*), yang berpotensi menyesatkan analisis klinis.
2. Interpretabilitas Rendah — Model XGBoost sulit dijelaskan kepada non-ahli karena bersifat kompleks dan tidak memberikan alasan eksplisit di balik prediksinya.
3. Evaluasi Tidak Komprehensif — Sebagian besar penelitian berfokus pada metrik akurasi (AUC) tanpa mempertimbangkan reliabilitas kalibrasi dan konsistensi penjelasan model.

3. Tujuan Bisnis dan Manfaat (Business Objectives & Benefits)

Tujuan Bisnis	Manfaat yang Diharapkan
Mengevaluasi pengaruh teknik kalibrasi terhadap keandalan probabilitas model XGBoost.	Bagi Peneliti: <ul style="list-style-type: none">• Memberikan insight mengenai metode kalibrasi terbaik untuk XGBoost dalam konteks diabetes.• Menghasilkan model dengan probabilitas yang lebih dapat diandalkan untuk penelitian lanjutan.
Membandingkan efektivitas SHAP dan LIME dalam meningkatkan interpretabilitas model.	Bagi Pengembang Model: <ul style="list-style-type: none">• Panduan dalam memilih teknik XAI yang sesuai untuk kebutuhan interpretabilitas global (SHAP) maupun lokal (LIME).• Meningkatkan transparansi dan kepercayaan terhadap model.

Tujuan Bisnis	Manfaat yang Diharapkan
Menghasilkan rekomendasi praktis untuk pengembangan model prediksi diabetes yang robust dan dapat diinterpretasi.	Bagi Stakeholder Klinis: <ul style="list-style-type: none">• Dasar evidence-based untuk mengadopsi model yang tidak hanya akurat tetapi juga dapat dijelaskan dan dipercaya.

4. Ruang Lingkup (Scope)

Dalam Cakupan (In-Scope):

- 1) Implementasi dan evaluasi model XGBoost untuk prediksi diabetes.
- 2) Penerapan teknik kalibrasi (Platt Scaling, Isotonic Regression) dan evaluasi dampaknya terhadap metrik kalibrasi (Brier Score, ECE).
- 3) Implementasi SHAP untuk interpretabilitas global dan LIME untuk interpretabilitas lokal.
- 4) Perbandingan kinerja model sebelum dan setelah kalibrasi menggunakan metrik diskriminasi (AUC, F1-Score) dan kalibrasi (Brier, ECE).

Di Luar Cakupan (Out-of-Scope):

- 1) Pengembangan dashboard atau aplikasi end-user.
- 2) Integrasi dengan sistem klinis atau EHR.
- 3) Deployment model dalam environment produksi.
- 4) Pengumpulan data baru atau validasi eksternal berskala besar.

5. Functional Requirements (Persyaratan Fungsional)

ID	Persyaratan Fungsional	Prioritas
FR-01	Pelatihan Model XGBoost: Sistem harus mampu melatih model XGBoost dengan hyperparameter yang dioptimasi.	Tinggi
FR-02	Kalibrasi Probabilitas: Sistem harus mengimplementasikan metode kalibrasi (Isotonic Regression).	Tinggi

ID	Persyaratan Fungsional	Prioritas
FR-03	Interpretabilitas dengan SHAP: Sistem harus menghasilkan SHAP summary plot dan SHAP values untuk setiap instansi.	Tinggi
FR-04	Interpretabilitas dengan LIME: Sistem harus mampu menghasilkan penjelasan lokal untuk prediksi individu menggunakan LIME.	Tinggi
FR-05	Evaluasi Kinerja: Sistem harus menghitung metrik diskriminasi (AUC, Accuracy, Precision, Recall, F1-Score) dan metrik kalibrasi (Brier Score, ECE).	Tinggi
FR-06	Pelaporan Komparatif: Sistem harus menghasilkan laporan yang membandingkan kinerja dan interpretabilitas model sebelum dan setelah kalibrasi.	Sedang

6. Non-Functional Requirements (Persyaratan Non-Fungsional)

ID	Persyaratan Non-Fungsional
NFR-01	Reproducibility: Seluruh eksperimen harus dapat direproduksi dengan deviasi $< 1\%$ pada metrik utama
NFR-02	Konsistensi Interpretabilitas: Penjelasan SHAP dan LIME harus konsisten secara kualitatif untuk sampel yang sama di berbagai model
NFR-03	Standar Performa: Minimal satu model harus mencapai $AUC > 0.85$ dan Brier Score < 0.15 setelah kalibrasi

ID	Persyaratan Non-Fungsional
NFR-04	Efisiensi Komputasi: Pipeline harus dapat dijalankan dalam waktu < 24 jam pada hardware standar penelitian
NFR-05	Kualitas Dokumentasi: Dokumentasi harus memungkinkan peneliti lain mereproduksi seluruh hasil dalam waktu < 1 minggu

7. Stackholders dan Pengguna

Stakeholder/Pengguna	Peran dan Kepentingan
Peneliti ML	Pengguna Utama - Melakukan eksperimen dan analisis
Data Scientist	Pengguna Sekunder - Menggunakan insights untuk proyek terapan
Penasihat Klinis	Validator - Memvalidasi interpretasi model dari perspektif klinis
Lead Penelitian	Pengambil Keputusan - Menggunakan hasil untuk perencanaan penelitian strategis

8. Deliverables

- Kode Sumber: Script Python modular untuk pelatihan, kalibrasi, evaluasi, dan interpretabilitas
- Model Terkalibrasi: File model XGBoost (.pkl/.joblib) siap digunakan
- Laporan Teknis: Hasil eksperimen, analisis perbandingan kalibrasi, SHAP & LIME
- Dokumentasi Teknis: Panduan eksekusi pipeline dan konfigurasi lingkungan
- Artefak Visualisasi: ROC Curve, Calibration Plot, Brier Diagram, SHAP Summary, dan LIME Explanation

9. Asumsi dan Kendala

Asumsi:

- Dataset yang digunakan representatif terhadap populasi klinis nyata
- Metrik teknis berkorelasi dengan utilitas klinis
- SHAP dan LIME memberikan penjelasan yang bermakna secara klini

Kendala:

- Ketersediaan sumber daya komputasi untuk SHAP pada dataset besar
- Keterbatasan data medis yang tersedia secara publik
- Keterbatasan waktu untuk hyperparameter tuning yang ekshaustif

10. Risiko dan Mitigasi

Risiko	Dampak	Probabilitas	Mitigasi
Model tidak memenuhi target performa	Tinggi	Sedang	Gunakan ensemble methods, feature engineering, advanced tuning
Kalibrasi tidak improve reliability	Sedang	Tinggi	Eksplorasi multiple calibration methods, Bayesian approaches
SHAP/LIME inconsistent across models	Sedang	Tinggi	Validasi dengan domain expert, statistical consistency tests
Computational resources insufficient	Sedang	Sedang	Sampling strategies, cloud computing, optimized implementations

11. Metrik Keberhasilan

Technical Success Metrics:

- Diskriminasi: $AUC > 0.85$ untuk minimal satu model
- Kalibrasi: $Brier\ Score < 0.15$ dan $ECE < 0.05$ untuk model terkalibrasi
- Interpretabilitas: Konsistensi $> 80\%$ antara SHAP dan LIME
- Improvement: Peningkatan $Brier\ Score > 25\%$ pasca-kalibrasi.

Research Success Metrics:

- Comparative Insights: Ranking model yang jelas berdasarkan use case
- Calibration Guidelines: Rekomendasi teknik kalibrasi per tipe model
- XAI Consistency: Pemahaman konsistensi SHAP dan LIME across models
- Reproducibility: Hasil dapat direproduksi dalam 1 minggu

12. Timeline

Tahap	Durasi	Hasil Utama
Persiapan Data	2 minggu	Dataset bersih, laporan EDA
Implementasi Model	3 minggu	Model dasar dengan hyperparameter teroptimasi
Kalibrasi & XAI	2 minggu	Model terkalibrasi, penjelasan SHAP/LIME
Evaluasi & Analisis	2 minggu	Hasil komprehensif, analisis komparatif
Pelaporan	1 minggu	Laporan akhir, dokumentasi, repository kode

13. Kesimpulan

BRD ini menegaskan ruang lingkup penelitian pada evaluasi kinerja, kalibrasi probabilitas, dan interpretabilitas model XGBoost untuk prediksi risiko diabetes menggunakan SHAP dan LIME. Penelitian ini diharapkan menghasilkan model yang akurat, terkalibrasi, dan dapat dijelaskan, sekaligus menjadi dasar pengembangan sistem kecerdasan buatan di bidang medis yang transparan, terpercaya, dan siap diintegrasikan dalam lingkungan klinis atau riset lanjutan.