

Analysis of Gradient Boosted Trees Algorithm in Breast Cancer Classification

Cantika Okzen Suryaputri ^{1*}, Majid Rahardi ^{2**}

* Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
cantikaokzensuryaputri@students.amikom.ac.id¹, majid@amikom.ac.id²

Article Info

Article history:

Received 2025-11-29

Revised 2026-01-12

Accepted 2026-01-13

Keyword:

*Breast Cancer Classification,
CatBoost Gradient Boosted Trees,
Machine Learning,
SHAP Explainability.*

ABSTRACT

Early and accurate classification of breast cancer is essential to support clinical diagnostic processes and improve patient outcomes. This study proposes a comprehensive machine learning pipeline based on Gradient Boosted Tree algorithms to classify breast tumors into benign and malignant categories. The proposed framework integrates several preprocessing stages, including outlier handling using the Local Outlier Factor (LOF), feature normalization with StandardScaler, class imbalance handling using SMOTE, and feature selection through ANOVA-based SelectKBest. Five ensemble learning models—XGBoost, LightGBM, CatBoost, HistGradientBoosting, and GradientBoosting—were trained and evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The experimental results show that all models achieved strong and comparable classification performance. Among them, CatBoost obtained the highest ROC-AUC value of 0.9960, along with an accuracy of 0.9649, precision of 0.9750, recall of 0.9286, and F1-score of 0.9512. Statistical evaluation using the DeLong test indicated that the differences in ROC-AUC among the evaluated models were not statistically significant ($p > 0.05$), suggesting similar discriminative capabilities across models. To enhance model interpretability, SHAP (SHapley Additive exPlanations) was applied to the CatBoost model as a representative classifier. The results show that features related to nuclear size and shape, such as radius, area, perimeter, and concavity, contributed most significantly to malignant predictions. This study demonstrates that the integration of robust preprocessing techniques, Gradient Boosted Tree models, and explainable machine learning provides an accurate and interpretable approach for breast cancer classification. However, the evaluation was conducted on a single public dataset without external validation, and further studies using independent and real-world datasets are required before clinical deployment.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Kanker payudara adalah keganasan yang terjadi pada jaringan payudara dan dapat berkembang dari epitel duktus maupun lobulus yang menjadi salah satu jenis kanker yang paling banyak ditemukan di Indonesia [1]. Kanker payudara adalah penyebab utama kematian akibat kanker pada perempuan di Indonesia, selain itu menurut data dari GLOBOCAN (Global Burden of Cancer) yang dirilis oleh International Agency for Research on Cancer (IARC)

menunjukkan bahwa pada 2018 tercatat 18,1 juta kasus kanker baru dan 9,6 juta kematian karena kanker secara global [2], [3]. Pada tahun 2022 tercatat 66.271 kasus baru dan 22.598 kematian atau sekitar 64 perempuan meninggal setiap hari akibat kanker payudara [4]. Permasalahan kanker payudara di Indonesia semakin memprihatinkan karena lebih dari 60% pasien kanker payudara baru memeriksakan diri ke dokter ketika penyakit sudah mencapai stadium lanjut [5]. Jika terdeteksi pada stadium dini, peluang bertahan hidup hingga lima tahun dapat melampaui 97%, sedangkan pada

stadium lanjut turun menjadi kurang dari 20%; sejalan dengan itu, data BPJS Kesehatan menunjukkan angka kelangsungan hidup lima tahun kanker payudara di Indonesia sebesar 56,8% [4]. Tingkat deteksi dini yang rendah serta keterlambatan penanganan kanker payudara tidak hanya disebabkan kurangnya kesadaran pasien, tetapi juga oleh keterbatasan metode diagnosis konvensional seperti mammografi dan biopsi yang bergantung pada interpretasi manual tenaga medis sehingga rentan menghasilkan diagnosis yang kurang akurat, terutama pada kasus dengan karakteristik sel yang sulit dibedakan antara jinak dan ganas [6], [7]. Perkembangan teknologi kecerdasan buatan (Artificial Intelligence) dan pembelajaran mesin (Machine Learning) telah membawa kemajuan besar di sektor kesehatan, terutama dalam hal diagnosis dan prediksi penyakit yang didasarkan pada data medis [8]. Dengan adanya keterbatasan metode konvensional mendorong perlunya pemanfaatan teknologi machine learning yang mampu mengolah data medis dengan lebih cepat dan akurat serta mengenali pola kompleks dalam data klinis untuk meningkatkan ketepatan identifikasi kanker pada tahap awal [9].

Salah satu pendekatan machine learning yang populer adalah ensemble learning, yaitu metode yang menggabungkan beberapa model untuk meningkatkan akurasi dan kestabilan hasil [10]. Berkembangnya metode Gradient Boosted Trees (GBT) sebagai salah satu bentuk ensemble yang lebih canggih dan efektif, dengan menggabungkan banyak pohon keputusan secara berurutan, di mana setiap pohon dilatih untuk mengoreksi error pohon sebelumnya, sehingga mampu menangkap pola kompleks secara lebih efektif dibandingkan metode tradisional [11].

Penelitian terdahulu, pada prediksi diabetes mellitus, LightGBM dan XGBoost mampu mencapai akurasi di atas 97% menunjukkan superioritas boosting dalam menangkap pola klinis non-linear yang sulit ditangani model tradisional [12]. XGBoost mencapai akurasi 93,83% pada klasifikasi kanker paru-paru dan tetap stabil pada berbagai skenario pembagian data, menunjukkan konsistensinya pada dataset besar [13]. Dengan kombinasi metode seleksi fitur mRMR dengan LightGBM menghasilkan akurasi 98% untuk prediksi kanker payudara [14]. LightGBM menunjukkan performa tertinggi dengan akurasi 95,3%, recall 94,8%, precision 95,5%, dan AUC 0.987, menegaskan superioritasnya dalam klasifikasi kanker payudara berbasis boosting [15]. Penggunaan Gradient Boosting Classifier dalam sistem pendukung keputusan klinis mampu mengidentifikasi lebih dari 98% kasus high-alert drug mismatch pada data uji dan 99% pada data evaluasi [16]. HistGradientBoosting terbukti memberikan hasil terbaik dibanding model lain seperti Logistic Regression, Random Forest, AdaBoost, dan Gradient Boosting konvensional, terutama pada metrik AUC > 0.99 untuk dua dataset utama dengan akurasi 96,25% dan 96,48% [17]. Model CatBoost mencapai AUC 0.836 dan F1-Score 0.735 pada prediksi

rawat inap pasien gagal gantung lansia, sementara analisis SHAP mengidentifikasi HGB, NT-proBNP, dan riwayat merokok sebagai prediktor paling berpengaruh [18]. Pada kerangka semi-supervised learning berbasis CT-radiomics dan XGBoost yang mampu mencapai akurasi 0.90 pada prediksi keselamatan hidup kanker paru, dengan SHAP yang menunjukkan peningkatan separabilitas fitur tekstur akibat pemanfaatan data tak berlabel [19]. Pengembangan pipeline prediksi diabetes yang menggabungkan SMOTE, seleksi fitur, dan interpretasi SHAP untuk menghasilkan model yang akurat sekaligus transparan, di mana fitur seperti usia, tekanan darah, dan BMI terbukti sebagai prediktor dominan [20]. Penggunaan algoritma tree-ensemble seperti Random Forest, XGBoost, dan LightGBM mampu mencapai akurasi hingga 97,38% dalam klasifikasi risiko kanker paru, sementara integrasi SHAP dan LIME memberikan interpretasi visual yang transparan untuk mendukung tenaga medis dalam memahami faktor risiko utama pada setiap prediksi [21]. XGBoost mencapai akurasi 87,4% dengan AUC 0,949 dalam memprediksi risiko penyakit kardiovaskular pada pasien diabetes berbasis asupan antioksidan, sementara analisis SHAP mengidentifikasi Daidzein, magnesium, dan ECGG sebagai fitur paling berpengaruh dalam keputusan model [22].

Dengan demikian, penelitian ini diposisikan untuk memperluas kajian komparatif terhadap lima varian algoritma Gradient Boosted Trees dalam satu rancangan eksperimen yang konsisten dan terstruktur. Berbeda dengan sebagian besar penelitian sebelumnya yang umumnya berfokus pada optimalisasi performa satu atau dua algoritma tertentu, studi ini menyajikan perbandingan sistematis antar beberapa model boosting dalam satu pipeline yang sama, sehingga memungkinkan evaluasi yang lebih adil dan komprehensif. Selain aspek performa, penelitian ini juga mempertimbangkan interpretabilitas model melalui pendekatan Explainable Artificial Intelligence (XAI) berbasis SHAP sebagai dasar analisis transparansi keputusan model.

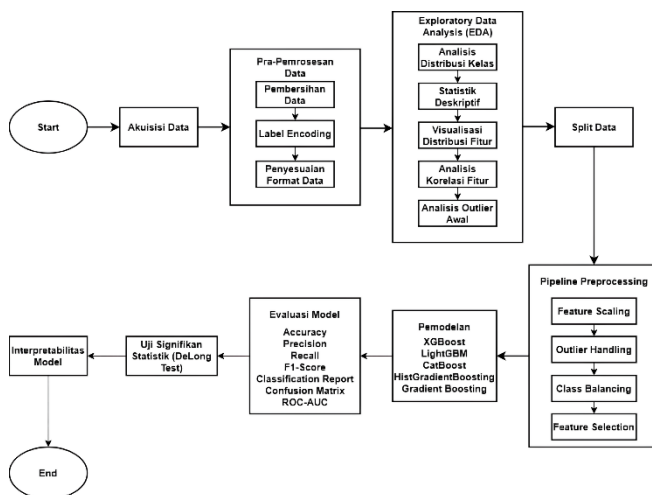
Meskipun berbagai studi terdahulu telah menunjukkan bahwa algoritma boosting mampu menghasilkan performa tinggi pada klasifikasi data medis, ruang kajian yang mengaitkan preprocessing, evaluasi multi-metrik, dan interpretabilitas model dalam satu kerangka eksperimen yang terintegrasi masih relatif terbatas. Padahal, dalam konteks kesehatan, transparansi dan keterjelasan alasan di balik keputusan model sama pentingnya dengan akurasi prediksi. Oleh karena itu, diperlukan kajian komparatif yang tidak hanya membandingkan performa numerik, tetapi juga mampu menjelaskan kontribusi fitur terhadap keputusan model secara global maupun lokal.

Berdasarkan kebutuhan tersebut, penelitian ini menyajikan studi komparatif terhadap lima algoritma Gradient Boosted Trees, yaitu XGBoost, LightGBM, CatBoost, HistGradientBoosting, dan GradientBoosting, untuk klasifikasi kanker payudara dalam satu rancangan eksperimen yang konsisten. Penelitian ini mengintegrasikan pipeline

preprocessing yang komprehensif, meliputi pembersihan data, deteksi dan penghapusan outlier menggunakan Local Outlier Factor (LOF), normalisasi fitur, penyeimbangan kelas menggunakan SMOTE, serta seleksi fitur berbasis ANOVA melalui SelectKBest untuk memperoleh fitur tumor yang paling relevan. Evaluasi performa dilakukan menggunakan pendekatan multi-metrik, yaitu accuracy, precision, recall, F1-score, dan ROC-AUC, serta dilengkapi dengan uji signifikansi statistik menggunakan DeLong test untuk membandingkan perbedaan ROC-AUC antar model.

Hasil penelitian menunjukkan bahwa seluruh model menghasilkan performa klasifikasi yang tinggi dan relatif sebanding, di mana CatBoost memperoleh nilai ROC-AUC tertinggi. Namun demikian, uji DeLong menunjukkan bahwa perbedaan performa antar model tidak signifikan secara statistik. Untuk meningkatkan transparansi dan kepercayaan terhadap model, analisis SHAP diterapkan pada model CatBoost sebagai model representatif guna menginterpretasikan kontribusi fitur secara global dan lokal. Analisis ini menunjukkan bahwa fitur-fitur yang berkaitan dengan ukuran dan bentuk inti sel tumor, seperti area, perimeter, dan concave points, memiliki pengaruh dominan terhadap prediksi keganasan. Dengan demikian, kontribusi utama penelitian ini terletak pada penyajian evaluasi yang mengintegrasikan aspek performa dan interpretabilitas dalam satu pipeline terstruktur, sehingga tidak hanya mengidentifikasi model dengan performa terbaik secara numerik, tetapi juga menjelaskan dasar pengambilan keputusan model secara transparan dan relevan secara klinis.

II. METODE



Gambar 1 Alur Penelitian

A. Akuisisi Data

Data yang digunakan dalam penelitian ini diperoleh melalui proses akuisisi data sekunder, yaitu pengambilan dataset yang telah tersedia secara publik dari platform Kaggle

[23]. Dataset yang digunakan berjudul Breast Cancer Dataset yang dikembangkan oleh YasserH. Dataset ini terdiri dari 569 sampel data sel payudara, yang dikategorikan menjadi dua kelas, yaitu Malignant (ganas) sebanyak 212 sampel dan Benign (jinak) sebanyak 357 sampel. Setiap sampel data direpresentasikan dalam 32 kolom atribut, yang mencakup satu atribut label diagnosis dan 31 fitur numerik hasil ekstraksi karakteristik morfologi inti sel epitel payudara dari citra digital. Proses akuisisi dilakukan dengan cara mengunduh dataset dalam format CSV dari Kaggle, kemudian memuatnya ke dalam lingkungan pemrograman Python menggunakan pustaka pandas untuk keperluan eksplorasi dan pemrosesan lebih lanjut.

B. Pra-Pemrosesan Data

Tahap pemrosesan data dilakukan untuk memastikan dataset berada dalam kondisi yang konsisten, bersih, dan siap digunakan pada tahap berikutnya. Langkah awal pada pra-pemrosesan data adalah pembersihan data dengan menyesuaikan penamaan atribut pada beberapa kolom, seperti 'concave points_mean' dan 'concave points_worst', diubah menjadi 'concave_points_mean' dan 'concave_points_worst', proses ini dilakukan untuk menghilangkan spasi yang dapat mengganggu proses pemanggilan variabel, kolom id dihapus karena tidak memiliki kontribusi terhadap proses klasifikasi. Selanjutnya, kolom diagnosis yang semula bertipe kategorikal dikodekan ke dalam bentuk numerik, dengan nilai 0 merepresentasikan kelas jinak (Benign) dan nilai 1 merepresentasikan kelas ganas (Malignant). Proses pengkodean ini bertujuan agar label kelas dapat diproses oleh algoritma pembelajaran mesin. Setelah itu, dataset dipisahkan ke dalam matriks fitur (X) yang terdiri dari 30 fitur numerik dan vektor target (y) yang berisi label kelas. Distribusi kelas menunjukkan 357 sampel kelas jinak dan 212 sampel kelas ganas. Dataset yang telah melalui tahap pra-pemrosesan ini selanjutnya digunakan pada tahap eksplorasi data lanjutan, pembagian data, serta pemodelan dan evaluasi kinerja model.

C. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah pendekatan sistematis dalam menganalisis data yang dirancang untuk mengeksplorasi dan memahami struktur serta karakteristik fundamental dari suatu dataset [24]. Pada tahap eksplorasi data, dilakukan analisis menyeluruh terhadap Breast Cancer Dataset untuk memahami struktur, karakteristik, dan distribusi data sebelum memasuki proses preprocessing dan pemodelan selanjutnya. Analisis awal dilakukan dengan meninjau struktur dataset menggunakan fungsi inspeksi data untuk mengetahui jumlah sampel, jumlah atribut, tipe data, serta keberadaan nilai hilang. Visualisasi distribusi kelas dilakukan untuk melihat proporsi data Malignant dan Benign, yang menunjukkan adanya ketidakseimbangan kelas dalam dataset. Selanjutnya, Untuk memahami perbedaan karakteristik fitur antar kelas, dilakukan analisis statistik

deskriptif dengan menghitung nilai rata-rata setiap fitur berdasarkan kelas diagnosis. Analisis hubungan antar fitur dilakukan menggunakan heatmap korelasi untuk mengidentifikasi tingkat keterkaitan antar atribut. Heatmap korelasi menunjukkan adanya korelasi tinggi antar beberapa fitur yang berkaitan dengan ukuran dan bentuk inti sel, seperti radius, perimeter, dan area. Temuan ini mengindikasikan potensi redundansi fitur dan mendukung penerapan seleksi fitur pada tahap pemodelan. Analisis hubungan antar fitur dilakukan menggunakan heatmap korelasi untuk mengidentifikasi keterkaitan antar atribut. Hasil analisis menunjukkan adanya korelasi tinggi pada beberapa fitur yang berkaitan dengan ukuran dan bentuk inti sel, sehingga mengindikasikan potensi redundansi fitur dan mendukung penerapan seleksi fitur. Selain itu, visualisasi boxplot dan Kernel Density Estimation (KDE) digunakan untuk menganalisis sebaran serta perbedaan distribusi fitur antara kelas jinak dan ganas. Hasil visualisasi menunjukkan perbedaan distribusi yang jelas pada sebagian besar fitur utama, yang mengindikasikan potensi separabilitas kelas sebelum proses pemodelan.

D. Split Data

Tahap split data dilakukan untuk membagi dataset menjadi data latih (training set) dan data uji (testing set) sebelum proses pemodelan. Pembagian ini bertujuan memastikan bahwa model dievaluasi menggunakan data yang tidak pernah dilihat sebelumnya sehingga hasil pengujian lebih objektif dan tidak bias. Pada penelitian ini digunakan fungsi `train_test_split` dengan proporsi 80% sebagai data latih dan 20% sebagai data uji.

E. Pipeline Preprocessing

Tahap preprocessing berperan penting dalam memastikan bahwa data yang dimanfaatkan adalah data yang berkualitas baik, lebih mudah dianalisis, dan mampu menghasilkan kesimpulan yang lebih tepat dan bermakna [25]. Untuk memastikan konsistensi proses serta mencegah terjadinya data leakage, tahapan preprocessing lanjutan diintegrasikan ke dalam sebuah pipeline yang diterapkan hanya pada data pelatihan. Pipeline preprocessing dirancang untuk menangani permasalahan perbedaan skala fitur, keberadaan outlier, ketidakseimbangan kelas, serta redundansi fitur secara terstruktur dan berurutan. Tahap pertama dalam pipeline adalah feature scaling menggunakan `StandardScaler`, yang menstandarisasi setiap fitur agar memiliki nilai rata-rata nol dan standar deviasi satu. Normalisasi ini diperlukan untuk memastikan bahwa seluruh fitur berada pada skala yang sebanding sehingga tidak mendominasi proses pembelajaran model. Tahap berikutnya adalah penanganan outlier menggunakan metode `Local Outlier Factor (LOF)`. Metode `Local Outlier Factor (LOF)` diterapkan untuk mendeteksi dan menghapus outlier sehingga distribusi data menjadi lebih stabil, karena metode ini menilai anomali berdasarkan perbandingan kerapatan

lokal setiap titik data terhadap tetangganya sehingga efektivitas deteksi tetap terjaga meskipun dataset memiliki ukuran besar dan fitur yang kompleks [26]. Selanjutnya, penyeimbangan kelas dilakukan menggunakan `Synthetic Minority Over-sampling Technique (SMOTE)`. Metode ini digunakan untuk mengatasi ketidakseimbangan kelas dengan menghasilkan sampel sintesis pada kelas minoritas, sehingga distribusi kelas pada data pelatihan menjadi lebih seimbang dan model dapat belajar secara lebih representatif. Tahap terakhir dalam pipeline adalah seleksi fitur menggunakan `SelectKBest` berbasis uji statistik ANOVA. Metode ini digunakan untuk memilih sejumlah fitur dengan kontribusi paling signifikan terhadap target klasifikasi. Dengan mengurangi jumlah fitur yang redundan atau kurang relevan, proses ini bertujuan untuk meningkatkan efisiensi komputasi serta kemampuan generalisasi model [27].

F. Pemodelan

Tahap pemodelan dilakukan dengan menerapkan lima algoritma `Gradient Boosted Trees` untuk melakukan klasifikasi kanker payudara, yaitu `XGBoost`, `LightGBM`, `CatBoost`, `HistGradientBoosting`, dan `GradientBoosting`. Kelima algoritma ini dipilih karena kemampuannya dalam menangani data non-linear serta performanya yang baik pada permasalahan klasifikasi berbasis data medis. Setiap model dilatih menggunakan pipeline preprocessing yang sama untuk memastikan konsistensi dan keadilan dalam proses perbandingan. Pipeline ini mencakup proses normalisasi fitur, penanganan outlier, penyeimbangan kelas, serta seleksi fitur, yang seluruhnya diterapkan hanya pada data pelatihan untuk mencegah terjadinya data leakage. Proses pelatihan model dilakukan dengan menyesuaikan parameter dasar masing-masing algoritma dan menggunakan nilai `random state` yang sama untuk memastikan reproduktibilitas hasil eksperimen. Setelah model dilatih, kinerja model dievaluasi menggunakan data pengujian yang tidak terlibat dalam proses pelatihan.

G. Evaluasi Model

Tahap evaluasi dilakukan dengan mengukur performa model pada data uji menggunakan metrik akurasi, presisi, recall, dan F1-score melalui `classification report` untuk menilai ketepatan serta keseimbangan kinerja klasifikasi. Selain itu, `confusion matrix` divisualisasikan menggunakan heatmap untuk melihat distribusi prediksi benar dan salah pada tiap kelas, yang penting dalam konteks kesehatan untuk mencegah salah diagnosis pada kasus `Malignant`. Evaluasi juga mencakup analisis kurva ROC dan nilai AUC untuk menilai kemampuan model membedakan kelas positif dan negatif secara lebih komprehensif.

H. Uji Signifikan Statistik (DeLong Test)

Uji signifikansi statistik dilakukan untuk menilai apakah perbedaan kinerja antar model klasifikasi bersifat signifikan secara statistik. Dalam penelitian ini, uji DeLong digunakan

untuk membandingkan nilai area under the receiver operating characteristic curve (ROC-AUC) antar model, karena uji ini secara khusus dirancang untuk membandingkan dua kurva ROC yang dihasilkan dari dataset pengujian yang sama. Uji DeLong diterapkan dengan menghitung nilai ROC-AUC dari dua model yang dibandingkan, kemudian memperkirakan kovarians dari perbedaan AUC menggunakan pendekatan non-parametrik. Perhitungan ini mempertimbangkan keterkaitan antar prediksi model terhadap sampel data yang sama, sehingga memberikan estimasi varians yang lebih akurat dibandingkan metode perbandingan sederhana. Dalam implementasinya, probabilitas prediksi dari masing-masing model diurutkan berdasarkan label kelas aktual, kemudian digunakan untuk menghitung nilai mid-rank dan kovarians AUC. Selanjutnya, nilai statistik z dihitung berdasarkan selisih AUC dan variansnya, dan nilai p -value diperoleh dari distribusi normal baku. Nilai p -value digunakan untuk menentukan signifikansi perbedaan kinerja antar model, dengan tingkat signifikansi yang ditetapkan sebesar 0,05. Hasil uji DeLong disajikan dalam bentuk tabel perbandingan yang memuat nilai AUC masing-masing model, selisih AUC, serta p -value. Uji ini digunakan sebagai dasar untuk menilai apakah model dengan performa numerik tertinggi benar-benar unggul secara statistik, atau apakah perbedaan performa antar model bersifat tidak signifikan.

I. Interpretabilitas Model

Interpretasi model pada penelitian ini dilakukan menggunakan pendekatan Shapley Additive Explanations (SHAP) untuk memperoleh pemahaman yang komprehensif terhadap perilaku model CatBoost sebagai model representatif dengan nilai ROC-AUC tertinggi. Analisis interpretabilitas global dilakukan menggunakan summary plot dan nilai rata-rata absolut SHAP (mean absolute SHAP value) untuk mengidentifikasi fitur-fitur yang memberikan kontribusi terbesar terhadap keputusan klasifikasi. Hasil analisis menunjukkan bahwa fitur-fitur yang berkaitan dengan ukuran dan bentuk inti sel, seperti `radius_mean`, `area_mean`, `perimeter_mean`, dan `concave_points_mean`, memiliki pengaruh dominan dalam membedakan antara kelas Malignant dan Benign, sehingga memberikan justifikasi statistik terhadap pentingnya fitur-fitur tersebut. Selain itu, interpretasi lokal dilakukan untuk menilai pengaruh individual setiap fitur terhadap prediksi pada sampel data tertentu, sehingga memungkinkan penelusuran faktor spesifik yang mendorong model menghasilkan keputusan klasifikasi tertentu. Pendekatan interpretabilitas secara global dan lokal ini meningkatkan transparansi model serta memperkuat kepercayaan terhadap hasil prediksi, sehingga model yang dihasilkan tidak hanya menunjukkan performa yang tinggi, tetapi juga dapat dipertanggungjawabkan dan relevan secara klinis dalam mendukung pengambilan keputusan diagnosis kanker payudara.

III. HASIL DAN PEMBAHASAN

A. Akuisisi Data

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dan merupakan dataset publik yang banyak digunakan dalam penelitian klasifikasi kanker payudara. Dataset ini berisi 569 sampel data tumor payudara yang diklasifikasikan ke dalam dua kelas, yaitu Benign (jinak) dan Malignant (ganas). Distribusi kelas terdiri dari 357 sampel Benign dan 212 sampel Malignant, yang menunjukkan adanya ketidakseimbangan kelas dengan dominasi sampel jinak. Setiap sampel data direpresentasikan oleh 30 fitur numerik yang menggambarkan karakteristik morfologi inti sel epitel payudara hasil ekstraksi dari citra digital. Fitur-fitur tersebut mencakup karakteristik ukuran inti sel, seperti radius, perimeter, dan area karakteristik bentuk dan ketidakrataan kontur, seperti compactness, concavity, dan concave points, serta karakteristik tekstur dan simetri, seperti texture, smoothness, symmetry, dan fractal dimension. Secara struktural, fitur-fitur dalam dataset dikelompokkan ke dalam tiga jenis pengukuran, yaitu nilai rata-rata (mean), standar deviasi/galat (standard error, se), dan nilai ekstrem (worst) dari masing-masing karakteristik inti sel. Pembagian ini memungkinkan model untuk menangkap informasi tidak hanya dari nilai tipikal suatu fitur, tetapi juga dari variasi dan nilai ekstrem yang sering berkaitan dengan tingkat keganasan sel tumor. Berdasarkan karakteristik tersebut, dataset ini menyediakan informasi yang kaya dan relevan untuk membedakan antara tumor jinak dan ganas, serta sesuai digunakan sebagai dasar pengembangan dan evaluasi model klasifikasi kanker payudara.

B. Pra-Pemrosesan Data

Tahap pemrosesan data dilakukan untuk memastikan dataset berada dalam kondisi yang konsisten, bersih, dan siap digunakan pada tahap berikutnya. Langkah awal pada pra-pemrosesan data adalah pembersihan data dengan menyesuaikan penamaan atribut pada beberapa kolom, seperti 'concave points_mean' dan 'concave points_worst', diubah menjadi 'concave_points_mean' dan 'concave_points_worst', proses ini dilakukan untuk menghilangkan spasi yang dapat mengganggu proses pemanggilan variabel, kolom id dihapus karena tidak memiliki kontribusi terhadap proses klasifikasi. Selanjutnya, kolom diagnosis yang semula bertipe kategorikal dikodekan ke dalam bentuk numerik, dengan nilai 0 merepresentasikan kelas jinak (Benign) dan nilai 1 merepresentasikan kelas ganas (Malignant). Proses pengkodean ini bertujuan agar label kelas dapat diproses oleh algoritma pembelajaran mesin. Setelah itu, dataset dipisahkan ke dalam matriks fitur (X) yang terdiri dari 30 fitur numerik dan vektor target (y) yang berisi label kelas. Distribusi kelas menunjukkan bahwa dataset bersifat tidak seimbang, dengan 357 sampel kelas jinak dan 212 sampel kelas ganas. Dataset yang telah melalui

tahap pra-pemrosesan ini selanjutnya digunakan pada tahap eksplorasi data lanjutan, pembagian data, serta pemodelan dan evaluasi kinerja model.

C. Exploratory Data Analysis (EDA)

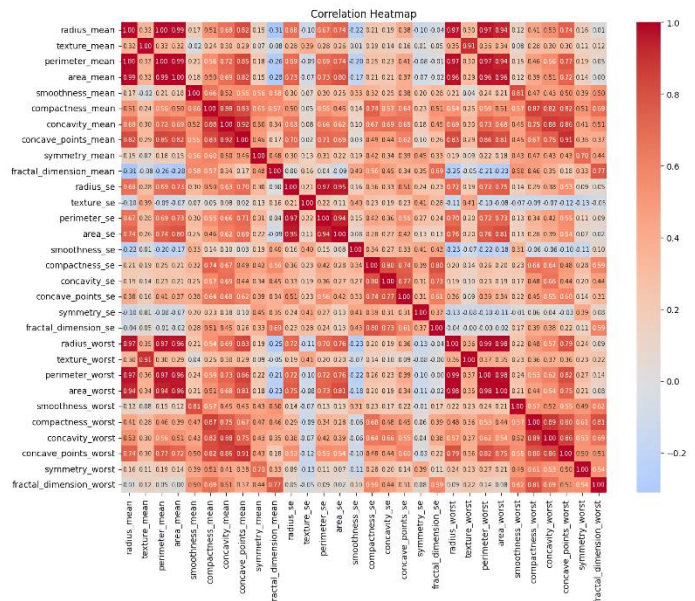
Hasil Exploratory Data Analysis (EDA) memberikan gambaran awal mengenai karakteristik dataset kanker payudara yang digunakan dalam penelitian ini. Dataset terdiri dari 569 sampel dengan 30 fitur numerik dan satu label kelas diagnosis, serta tidak ditemukan nilai hilang pada seluruh atribut. Hal ini menunjukkan bahwa dataset berada dalam kondisi yang baik dan layak untuk digunakan pada tahap analisis dan pemodelan lanjutan.

Analisis distribusi kelas menunjukkan bahwa dataset terdiri dari 357 sampel kelas jinak (Benign) dan 212 sampel kelas ganas (Malignant). Distribusi ini mengindikasikan adanya ketidakseimbangan kelas, di mana proporsi sampel jinak lebih besar dibandingkan sampel ganas. Kondisi ketidakseimbangan ini berpotensi memengaruhi proses pembelajaran model, khususnya dalam mengenali kelas minoritas, sehingga diperlukan penerapan teknik penyeimbangan kelas pada tahap preprocessing untuk meningkatkan kinerja klasifikasi.

TABEL 1
CONTOH STATISTIK DESKRIPTIF 5 FITUR

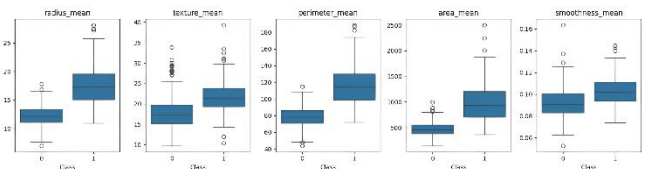
Fitur	Diagnosis	
	0	1
area_worst	558.899440	1422.286321
area_mean	462.790196	978.376415
perimeter_worst	87.005938	141.370330
perimeter_mean	78.075406	115.365377
area_se	21.135148	72.672406

Pada tabel 1 analisis statistik deskriptif berdasarkan nilai rata-rata fitur pada masing-masing kelas memperlihatkan perbedaan karakteristik yang cukup jelas antara kelas jinak (Benign) dan ganas (Malignant). Sebagian besar fitur menunjukkan nilai rata-rata yang lebih tinggi pada kelas ganas, khususnya fitur-fitur yang berkaitan dengan ukuran inti sel seperti radius, perimeter, dan area. Hal ini mengindikasikan bahwa sel ganas cenderung memiliki ukuran inti yang lebih besar dan struktur yang lebih kompleks dibandingkan sel jinak, yang sejalan dengan karakteristik biologis kanker payudara.



Gambar 2 Heatmap Korelasi

Dari gambar 3, Hasil visualisasi heatmap korelasi memperlihatkan adanya korelasi yang sangat kuat antar beberapa fitur, terutama fitur-fitur yang merepresentasikan ukuran inti sel, seperti radius_mean, perimeter_mean, dan area_mean, serta padanan fitur worst-nya. Tingginya korelasi ini menunjukkan adanya redundansi informasi antar fitur, yang berpotensi meningkatkan kompleksitas model tanpa memberikan informasi tambahan yang signifikan. Oleh karena itu, temuan ini mendukung penerapan seleksi fitur pada tahap pemodelan untuk meningkatkan efisiensi dan kemampuan generalisasi model.



Gambar 3 Contoh Visualisasi Boxplot Analisis Outlier Awal 5 Fitur

Gambar 4 visualisasi boxplot digunakan untuk menganalisis sebaran nilai fitur pada masing-masing kelas serta mengidentifikasi potensi outlier. Hasil boxplot menunjukkan bahwa pada sebagian besar fitur utama, kelas ganas memiliki median nilai yang lebih tinggi serta rentang distribusi yang lebih lebar dibandingkan kelas jinak. Selain itu, keberadaan nilai ekstrem pada beberapa fitur mengindikasikan adanya sampel dengan karakteristik yang menyimpang, yang berpotensi memengaruhi proses pembelajaran model apabila tidak ditangani dengan baik.

D. Split Data

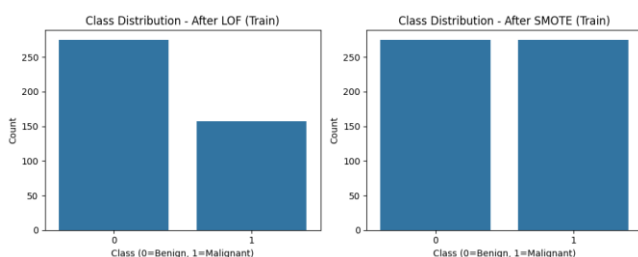
Hasil pembagian data menunjukkan bahwa dataset dibagi menjadi data latih (training set) dan data uji (testing set)

dengan proporsi 80% : 20% menggunakan metode stratified split. Pendekatan ini digunakan untuk memastikan bahwa proporsi kelas jinak (Benign) dan ganas (Malignant) tetap terjaga secara konsisten pada kedua subset data. Dari total 569 sampel, diperoleh 455 sampel sebagai data latih dan 114 sampel sebagai data uji. Distribusi kelas pada data latih terdiri dari 285 sampel Benign dan 170 sampel Malignant, sedangkan data uji terdiri dari 72 sampel Benign dan 42 sampel Malignant. Konsistensi distribusi kelas antara data latih dan data uji menunjukkan bahwa proses pembagian data dilakukan secara representatif. Pembagian data ini memastikan bahwa model dilatih menggunakan data yang cukup besar dan beragam, sekaligus diuji menggunakan data yang benar-benar terpisah dari proses pelatihan.

E. Pipeline Preprocessing

Pada tahap preprocessing, data latih terlebih dahulu dinormalisasi menggunakan StandardScaler untuk memastikan seluruh fitur numerik berada pada skala yang sebanding. Normalisasi ini penting mengingat fitur dalam dataset memiliki rentang nilai yang bervariasi, sehingga dapat memengaruhi kinerja algoritma deteksi outlier dan seleksi fitur apabila tidak distandarisasi.

Deteksi outlier dilakukan menggunakan Local Outlier Factor (LOF) pada data latih yang telah diskalakan. Hasil deteksi menunjukkan bahwa dari total 455 sampel data latih, teridentifikasi 23 sampel sebagai outlier dan 432 sampel sebagai inlier, dengan persentase outlier sebesar 5,05%. Jumlah ini relatif kecil dan menunjukkan bahwa mayoritas data memiliki pola yang konsisten. Penghapusan outlier dilakukan hanya pada data latih untuk mengurangi pengaruh sampel ekstrem terhadap proses pembelajaran model, sekaligus menjaga objektivitas evaluasi karena data uji tidak dimodifikasi.



Gambar 4 Distribusi Kelas Sebelum dan Sesudah SMOTE

Setelah penghapusan outlier, distribusi kelas pada data latih yang terlihat pada gambar 4 masih menunjukkan ketidakseimbangan, di mana jumlah sampel kelas Benign lebih besar dibandingkan kelas Malignant. Untuk mengatasi permasalahan tersebut, diterapkan teknik Synthetic Minority Over-sampling Technique (SMOTE) pada data latih setelah LOF. Hasil visualisasi distribusi kelas menunjukkan bahwa setelah penerapan SMOTE, jumlah sampel pada kedua kelas menjadi seimbang. Penyeimbangan ini bertujuan untuk mengurangi bias model terhadap kelas mayoritas dan

meningkatkan kemampuan model dalam mengenali kasus kanker ganas sebagai kelas minoritas.

TABEL 2
HASIL PERINGKAT FITUR (10 FITUR)

Fitur	F-Score	P-Value
concave_points_worst	1087.409091	3.256932e-132
perimeter_worst	910.318181	1.460672e-118
concave_points_mean	880.128564	4.535272e-116
radius_worst	872.552110	1.951186e-115
perimeter_mean	720.615375	5.861753e-102
radius_mean	669.607332	4.554457e-97
area_worst	648.429848	5.616536e-95
concavity_mean	602.157880	2.821556e-90
area_mean	583.902059	2.278223e-88
concavity_worst	536.351327	2.976444e-83

Tahap selanjutnya adalah seleksi fitur menggunakan metode ANOVA (f_classif) dengan pendekatan SelectKBest. Uji ANOVA dilakukan pada data latih yang telah melalui LOF dan SMOTE, sehingga perhitungan signifikansi fitur tidak dipengaruhi oleh outlier maupun ketidakseimbangan kelas. Hasil ANOVA pada tabel 2 menunjukkan bahwa sejumlah fitur memiliki nilai F-score yang sangat tinggi disertai p-value yang sangat kecil, yang mengindikasikan adanya perbedaan yang signifikan secara statistik antara kelas Benign dan Malignant. Berdasarkan hasil SelectKBest dengan nilai K = 10, diperoleh sepuluh fitur paling relevan, yaitu radius_mean, perimeter_mean, area_mean, concavity_mean, concave_points_mean, radius_worst, perimeter_worst, area_worst, concavity_worst, dan concave_points_worst. Fitur-fitur ini didominasi oleh karakteristik yang berkaitan dengan ukuran inti sel (radius, perimeter, area) serta ketidakrataannya kontur dan kedalaman lekukan inti sel (concavity dan concave points). Dominasi fitur-fitur tersebut sejalan dengan karakteristik klinis kanker payudara, di mana sel ganas cenderung memiliki ukuran inti yang lebih besar, bentuk yang tidak teratur, serta kontur yang lebih kompleks dibandingkan sel jinak.

Dengan demikian, hasil seleksi fitur ini tidak hanya signifikan secara statistik, tetapi juga relevan secara biologis dan klinis. Fitur-fitur terpilih selanjutnya digunakan sebagai input utama pada tahap pemodelan menggunakan algoritma Gradient Boosted Trees, sehingga diharapkan dapat meningkatkan kinerja model sekaligus menjaga efisiensi dan interpretabilitas.

F. Pemodelan

Tahap modeling dilakukan dengan membangun lima algoritma Gradient Boosted Trees, yaitu XGBoost, LightGBM, CatBoost, HistGradientBoosting, dan GradientBoosting, yang seluruhnya diintegrasikan dalam satu pipeline preprocessing yang konsisten. Pipeline ini mencakup normalisasi fitur, penghapusan outlier menggunakan LOF, penyeimbangan kelas dengan SMOTE, serta seleksi fitur berbasis ANOVA, yang seluruhnya diterapkan hanya pada

data latih untuk mencegah terjadinya data leakage. Setiap model kemudian dilatih menggunakan data latih dan dievaluasi menggunakan data uji yang tidak mengalami transformasi apapun selain scaling dan seleksi fitur yang telah dipelajari dari data latih.

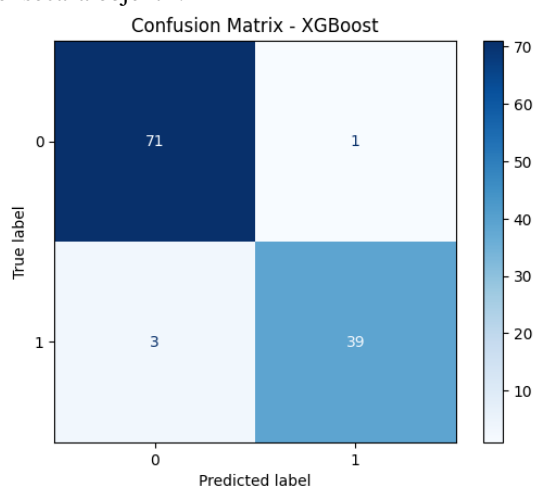
XGBoost digunakan dengan fungsi objektif logloss untuk klasifikasi biner dan dikenal mampu menangani hubungan non-linear secara efektif. LightGBM mengadopsi strategi leaf-wise growth yang meningkatkan efisiensi pelatihan pada data berdimensi tinggi. CatBoost menerapkan ordered boosting yang meningkatkan stabilitas model dan mengurangi overfitting tanpa banyak penyetelan parameter. HistGradientBoosting memanfaatkan pendekatan histogram untuk mempercepat komputasi dan mengurangi kompleksitas model, sedangkan GradientBoosting konvensional digunakan sebagai model pembandingan dasar.

G. Evaluasi Model

	Model	Accuracy	Precision	Recall	F1	ROC-AUC
2	CatBoost	0.9649	0.9750	0.9286	0.9512	0.9960
3	HistGradientBoosting	0.9561	0.9744	0.9048	0.9383	0.9957
1	LightGBM	0.9561	0.9744	0.9048	0.9383	0.9944
4	GradientBoosting	0.9561	0.9744	0.9048	0.9383	0.9940
0	XGBoost	0.9649	0.9750	0.9286	0.9512	0.9904

Gambar 5 Evaluasi Model

. Pada gambar 5 evaluasi kinerja model dilakukan menggunakan beberapa metrik klasifikasi yang umum digunakan pada penelitian medis, yaitu accuracy, precision, recall, F1-score, dan ROC-AUC, serta dianalisis lebih lanjut melalui confusion matrix dan kurva ROC. Seluruh evaluasi dilakukan pada data uji yang tidak terlibat dalam proses pelatihan maupun preprocessing lanjutan, sehingga hasil yang diperoleh merepresentasikan kemampuan generalisasi model secara objektif.



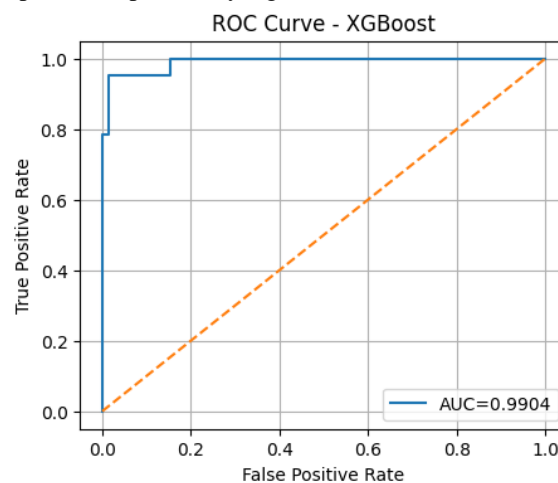
Gambar 6 Confusion Matrix XGBoost

Pada gambar 6 model XGBoost berhasil mengklasifikasikan 71 dari 72 sampel Benign dengan benar, dengan hanya 1 kesalahan false positive. Pada kelas Malignant, model mampu mengidentifikasi 39 dari 42 sampel secara benar, dengan 3 kasus false negative. Pola ini menunjukkan bahwa XGBoost memiliki kemampuan yang baik dalam membedakan kedua kelas, meskipun masih terdapat sejumlah kecil kasus kanker ganas yang salah diklasifikasikan sebagai jinak.

Classification Report - XGBoost				
	precision	recall	f1-score	support
0	0.9595	0.9861	0.9726	72
1	0.9750	0.9286	0.9512	42
accuracy			0.9649	114
macro avg	0.9672	0.9573	0.9619	114
weighted avg	0.9652	0.9649	0.9647	114

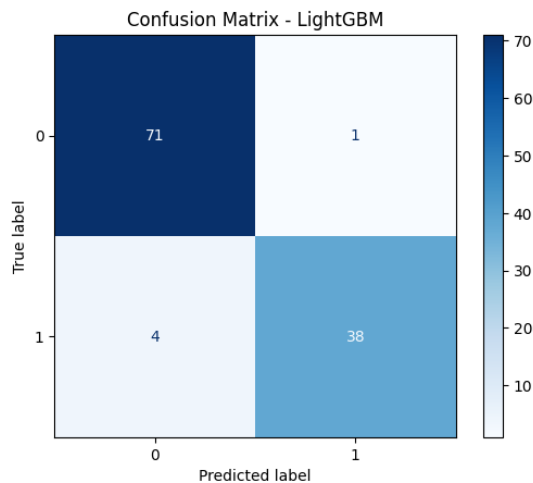
Gambar 7 Classification Report XGBoost

Pada gambar 7 XGBoost mencapai akurasi 0.9649. Untuk kelas Malignant, model menghasilkan precision 0.975, recall 0.9286, dan F1-score 0.9512. Nilai precision yang tinggi menunjukkan bahwa prediksi Malignant sangat akurat, sementara recall yang tinggi menandakan bahwa sebagian besar kasus kanker ganas berhasil terdeteksi, meskipun masih terdapat beberapa kasus yang terlewat.

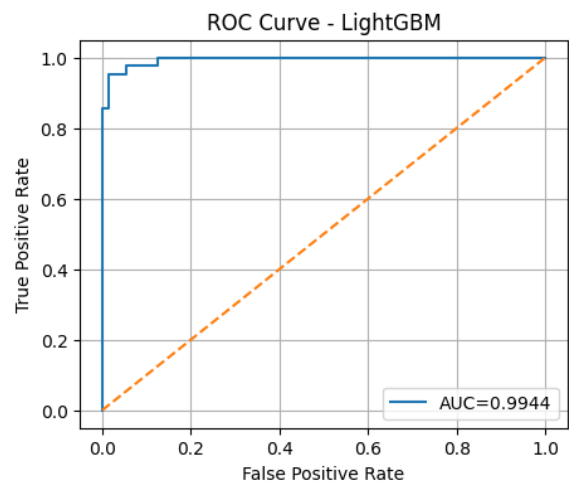


Gambar 8 ROC Curve XGBoost

Nilai ROC-AUC pada gambar 8 sebesar 0.9904 menunjukkan bahwa XGBoost memiliki kemampuan diskriminatif yang sangat baik. Kurva ROC yang mendekati sudut kiri atas menandakan keseimbangan yang baik antara sensitivitas dan spesifisitas pada berbagai ambang klasifikasi.



Gambar 9 Confusion Matrix LightGBM



Gambar 11 ROC Curve LightGBM

Pada gambar 9 LightGBM mampu mengklasifikasikan 71 dari 72 sampel Benign dengan benar, dengan 1 false positive. Namun, pada kelas Malignant, terdapat 4 false negative, sehingga 38 dari 42 sampel terklasifikasi dengan benar. Hal ini menunjukkan bahwa LightGBM sedikit lebih konservatif dalam mendeteksi kelas Malignant dibandingkan XGBoost dan CatBoost.

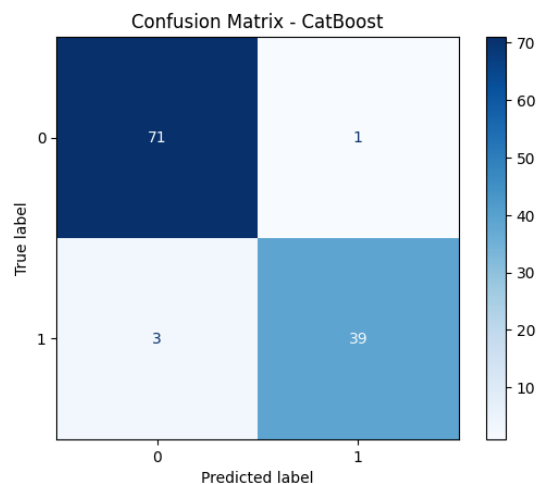
Classification Report - LightGBM

	precision	recall	f1-score	support
0	0.9467	0.9861	0.9660	72
1	0.9744	0.9048	0.9383	42
accuracy			0.9561	114
macro avg	0.9605	0.9454	0.9521	114
weighted avg	0.9569	0.9561	0.9558	114

Gambar 10 Classification Report LightGBM

Pada gambar 10 model LightGBM mencapai akurasi 0.9561. Untuk kelas Malignant, LightGBM menghasilkan precision 0.9744, recall 0.9048, dan F1-score 0.9383. Nilai recall yang sedikit lebih rendah mengindikasikan bahwa sebagian kecil kasus kanker ganas masih belum terdeteksi, meskipun secara keseluruhan performa model tetap sangat baik.

Pada gambar 11 dengan nilai ROC-AUC sebesar 0.9944, LightGBM menunjukkan kemampuan pemisahan kelas yang sangat kuat. Kurva ROC yang tinggi mencerminkan bahwa meskipun recall sedikit lebih rendah, model tetap konsisten dalam membedakan kelas Benign dan Malignant secara probabilistik.



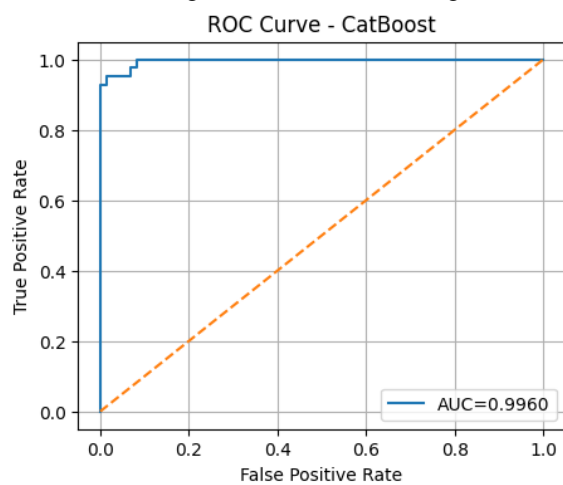
Gambar 12 Confusion Matrix CatBoost

Pada gambar 12 confusion matrix CatBoost menunjukkan hasil yang sangat seimbang, dengan 71 sampel Benign dan 39 sampel Malignant yang diklasifikasikan dengan benar. Jumlah kesalahan relatif kecil, yaitu 1 false positive dan 3 false negative, yang identik dengan pola kesalahan pada XGBoost.

Classification Report - CatBoost				
	precision	recall	f1-score	support
0	0.9595	0.9861	0.9726	72
1	0.9750	0.9286	0.9512	42
accuracy			0.9649	114
macro avg	0.9672	0.9573	0.9619	114
weighted avg	0.9652	0.9649	0.9647	114

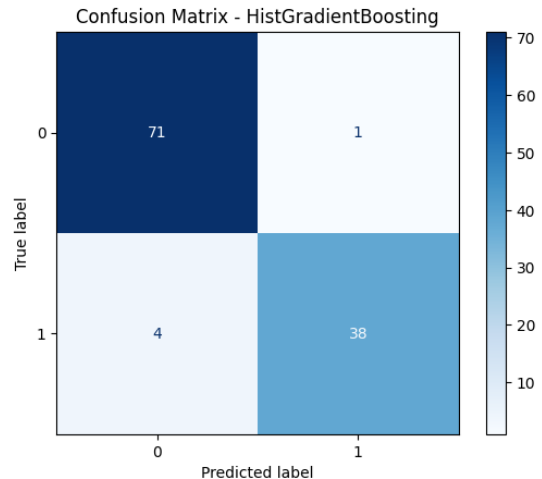
Gambar 13 Classification Report CatBoost

Model CatBoost mencapai akurasi 0.9649 pada gambar 13. Untuk kelas Malignant, diperoleh precision 0.9750, recall 0.9286, dan F1-score 0.9512. Kombinasi precision dan recall yang tinggi menunjukkan bahwa CatBoost mampu menjaga keseimbangan antara meminimalkan kesalahan positif palsu dan mendeteksi sebagian besar kasus kanker ganas.



Gambar 14 ROC Curve CatBoost

Pada gambar 14 CatBoost memperoleh ROC-AUC tertinggi sebesar 0.9960. Kurva ROC yang sangat mendekati sudut kiri atas menandakan bahwa model ini memiliki kemampuan diskriminatif terbaik di antara seluruh model yang diuji, sehingga dipilih sebagai model utama pada tahap analisis lanjutan.



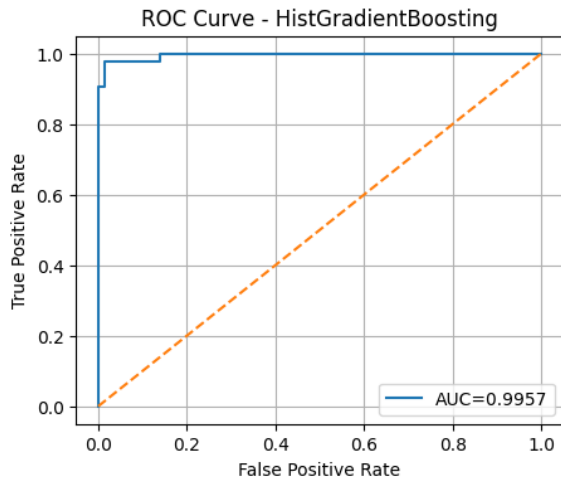
Gambar 15 Confusion Matrix HistGradientBoosting

Selanjutnya, HistGradientBoosting mengklasifikasikan 71 dari 72 sampel Benign dengan benar dan menghasilkan 1 false positive pada gambar 15. Pada kelas Malignant, terdapat 4 false negative, sehingga 38 sampel berhasil terdeteksi. Pola kesalahan ini serupa dengan LightGBM.

Classification Report - HistGradientBoosting				
	precision	recall	f1-score	support
0	0.9467	0.9861	0.9660	72
1	0.9744	0.9048	0.9383	42
accuracy			0.9561	114
macro avg	0.9605	0.9454	0.9521	114
weighted avg	0.9569	0.9561	0.9558	114

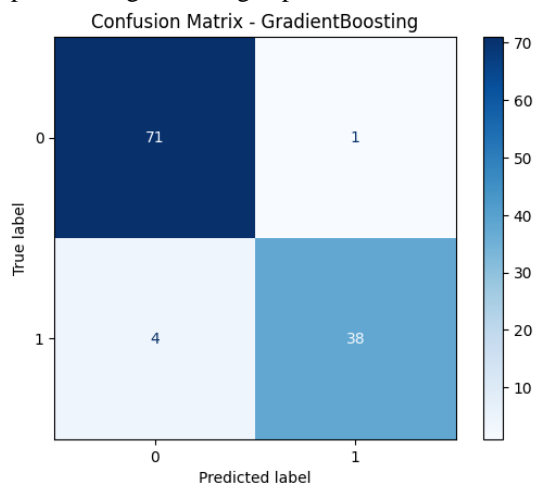
Gambar 16 Classification Report HistGradient Boosting

Model ini menghasilkan akurasi 0.9561, dengan precision 0.9744, recall 0.9048, dan F1-score 0.9383 pada kelas Malignant yang bisa dilihat pada gambar 16. Nilai ini menunjukkan performa yang stabil dan kompetitif, meskipun sensitivitas terhadap kelas Malignant sedikit lebih rendah dibandingkan XGBoost dan CatBoost.



Gambar 17 ROC Curve HistGradientBoosting

Pada gambar 17 nilai ROC-AUC sebesar 0.9957 menempatkan HistGradientBoosting sebagai salah satu model dengan kemampuan pemisahan kelas terbaik. Kurva ROC menunjukkan konsistensi model dalam membedakan kelas pada berbagai ambang keputusan.



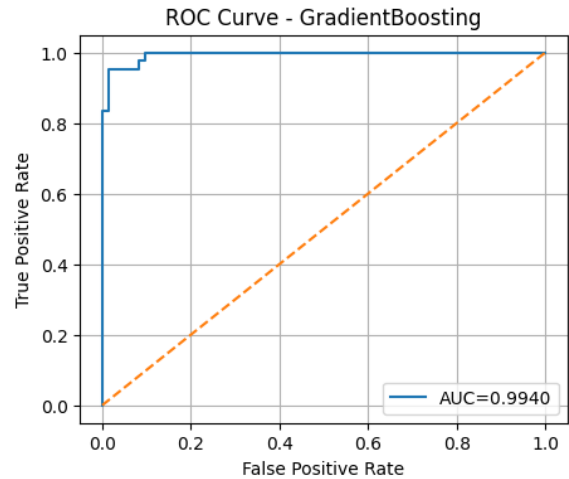
Gambar 18 Confusion Matrix GradientBoosting

Terakhir, model GradientBoosting menunjukkan hasil yang identik dengan HistGradientBoosting dan LightGBM, dengan 1 false positive dan 4 false negative yang terlihat pada gambar 18. Sebanyak 71 sampel Benign dan 38 sampel Malignant berhasil diklasifikasikan dengan benar.

Classification Report - GradientBoosting				
	precision	recall	f1-score	support
0	0.9467	0.9861	0.9660	72
1	0.9744	0.9048	0.9383	42
accuracy			0.9561	114
macro avg	0.9605	0.9454	0.9521	114
weighted avg	0.9569	0.9561	0.9558	114

Gambar 19 Classification Report GradientBoosting

Pada gambar 19 model ini mencapai akurasi 0.9561. Untuk kelas Malignant, diperoleh precision 0.9744, recall 0.9048, dan F1-score 0.9383. Performa ini menunjukkan bahwa meskipun model lebih sederhana dibandingkan varian boosting modern, GradientBoosting tetap mampu memberikan hasil yang sangat kompetitif.



Gambar 20 ROC Curve GradientBoosting

Dengan nilai ROC-AUC yang ada pada gambar 20 sebesar 0.9940, GradientBoosting masih menunjukkan kemampuan diskriminatif yang sangat baik. Kurva ROC yang tinggi mengindikasikan bahwa model ini efektif dalam membedakan antara kelas jinak dan ganas, meskipun sedikit berada di bawah CatBoost dan HistGradientBoosting.

H. Uji Signifikan Statistik (DeLong Test)

	Model_A	Model_B	AUC_A	AUC_B	AUC_Diff (A-B)	p_value
0	CatBoost	XGBoost	0.996	0.9904	0.0056	0.1260
1	CatBoost	LightGBM	0.996	0.9944	0.0017	0.3225
3	CatBoost	GradientBoosting	0.996	0.9940	0.0020	0.3504
2	CatBoost	HistGradientBoosting	0.996	0.9957	0.0003	0.8786

Gambar 21 Hasil DeLong Test

Untuk menilai apakah perbedaan performa antar model secara statistik benar-benar signifikan, dilakukan uji DeLong terhadap nilai ROC-AUC yang dapat dilihat pada gambar 21. Uji DeLong merupakan metode non-parametrik yang umum digunakan untuk membandingkan dua kurva ROC yang dihasilkan dari model yang dievaluasi pada dataset yang sama. Pendekatan ini dipilih karena mampu mempertimbangkan korelasi antar prediksi model dan memberikan pengujian yang lebih reliabel dibandingkan perbandingan ROC-AUC secara deskriptif semata.

Berdasarkan hasil evaluasi sebelumnya, CatBoost memperoleh nilai ROC-AUC tertinggi dan oleh karena itu digunakan sebagai model acuan (Model_A) dalam pengujian DeLong. Selanjutnya, ROC-AUC CatBoost dibandingkan secara berpasangan dengan model lain, yaitu XGBoost, LightGBM, HistGradientBoosting, dan GradientBoosting.

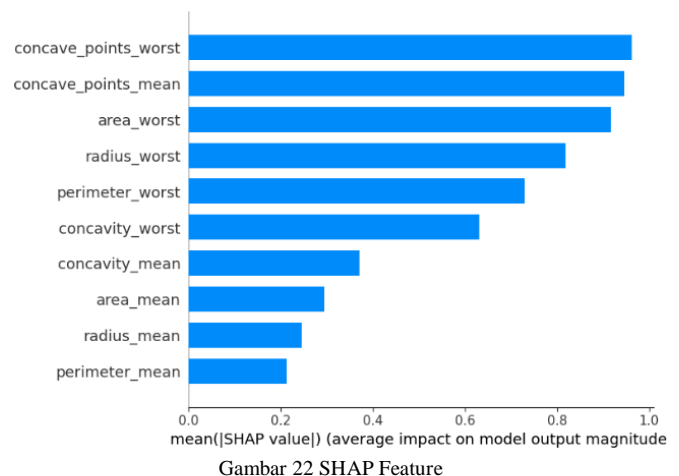
Hasil uji DeLong menunjukkan bahwa perbedaan nilai ROC-AUC antara CatBoost dan XGBoost sebesar 0.0056, dengan nilai p-value 0.1260. Meskipun CatBoost memiliki nilai ROC-AUC yang lebih tinggi, nilai p-value tersebut lebih besar dari ambang signifikansi 0,05, sehingga perbedaan performa antara kedua model tidak signifikan secara statistik. Hal serupa juga ditemukan pada perbandingan CatBoost dengan LightGBM, di mana selisih ROC-AUC sebesar 0.0017 menghasilkan p-value 0.3225, serta pada perbandingan dengan GradientBoosting yang memiliki selisih ROC-AUC 0.0020 dan p-value 0.3504.

Perbandingan antara CatBoost dan HistGradientBoosting menunjukkan selisih ROC-AUC yang sangat kecil, yaitu 0.0003, dengan p-value 0.8786, yang semakin menegaskan bahwa kedua model memiliki kemampuan diskriminatif yang hampir identik. Secara keseluruhan, seluruh nilai p-value yang diperoleh berada di atas batas signifikansi statistik ($p > 0,05$), sehingga tidak terdapat perbedaan ROC-AUC yang signifikan secara statistik antara CatBoost dan model boosting lainnya.

Temuan ini mengindikasikan bahwa meskipun CatBoost menunjukkan nilai ROC-AUC tertinggi secara numerik, keunggulan tersebut belum dapat diklaim signifikan secara statistik. Dengan demikian, performa seluruh algoritma Gradient Boosted Trees yang diuji dapat dikatakan relatif sebanding dalam hal kemampuan diskriminatif. Pemilihan CatBoost sebagai model utama dalam penelitian ini didasarkan pada kombinasi performa numerik terbaik, stabilitas prediksi, serta nilai ROC-AUC tertinggi, bukan semata-mata pada perbedaan signifikan secara statistik. Pendekatan ini relevan dalam konteks aplikasi medis, di mana model dengan performa tinggi dan stabil tetap bernilai penting meskipun perbedaannya dengan model lain tidak signifikan secara statistik.

1. Interpretabilitas Model

Interpretabilitas model pada penelitian ini dianalisis menggunakan pendekatan SHapley Additive exPlanations (SHAP) untuk memahami kontribusi setiap fitur terhadap keputusan klasifikasi kanker payudara. Analisis ini difokuskan pada model CatBoost, yang dipilih sebagai model terbaik karena menghasilkan nilai ROC-AUC tertinggi, meskipun perbedaannya dengan model lain tidak signifikan secara statistik berdasarkan uji DeLong.

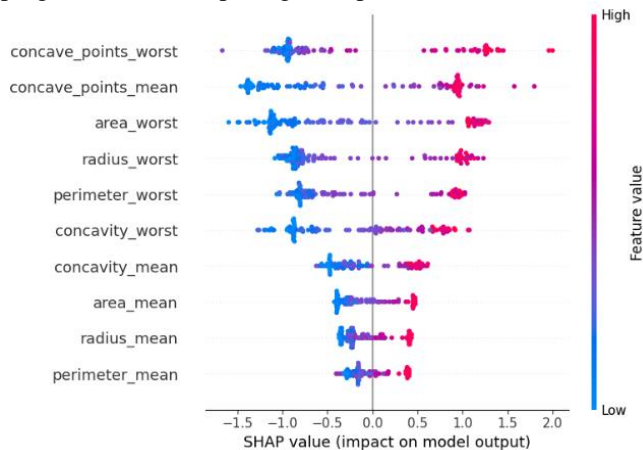


Gambar 22 SHAP Feature

Analisis global interpretability dilakukan melalui SHAP summary plot (bar) pada gambar 22 menunjukkan peringkat fitur berdasarkan nilai mean absolute SHAP. Hasil analisis menunjukkan bahwa fitur concave_points_worst, concave_points_mean, area_worst, radius_worst, dan perimeter_worst merupakan fitur dengan kontribusi terbesar terhadap prediksi kelas ganas. Dominasi fitur-fitur ini mengindikasikan bahwa karakteristik morfologi inti sel yang berkaitan dengan ketidakaturan bentuk, ukuran, dan kedalaman cekungan (concavity) memiliki peran penting dalam membedakan tumor ganas dan jinak.

Secara klinis, temuan ini sejalan dengan literatur medis yang menyatakan bahwa kanker payudara ganas umumnya menunjukkan inti sel yang lebih besar, variasi ukuran inti yang tinggi, bentuk yang tidak beraturan, serta tingkat concavity dan concave points yang lebih menonjol. Studi oleh Li et al. (2021) dalam *Analytical Cellular Pathology* melaporkan bahwa perbedaan ukuran inti (nuclear size variation), pleomorfisme inti, serta karakteristik nukleoli memiliki hubungan signifikan dengan sub tipe molekuler kanker payudara dan tingkat keganasan tumor. Sel kanker dengan karakteristik tersebut cenderung memiliki perilaku biologis yang lebih agresif dan tingkat proliferasi yang lebih tinggi. Fitur radius, area, dan perimeter, baik pada nilai mean maupun worst merepresentasikan ukuran dan ekspansi inti sel. Dalam konteks histopatologi, ukuran inti yang lebih besar dan tidak seragam merupakan indikator penting keganasan jaringan payudara, karena mencerminkan peningkatan aktivitas mitosis dan ketidakaturan siklus sel. Hal ini diperkuat oleh temuan klinis bahwa sub tipe kanker payudara dengan agresivitas tinggi, seperti HER2-positive, memiliki inti sel yang lebih besar dan pleomorfik dibandingkan sub tipe dengan prognosis lebih baik. Selain itu, fitur concavity dan concave points menggambarkan tingkat ketidakrataan dan invaginasi pada batas inti sel. Nilai concavity yang tinggi mengindikasikan bentuk inti yang tidak simetris dan tidak teratur, yang secara patologis sering dikaitkan dengan keganasan dan invasi tumor. Literatur medis menyebutkan bahwa ketidakaturan bentuk inti merupakan manifestasi

morfologis dari perubahan molekuler yang mendasari progresi kanker dan peningkatan potensi metastasis [28].



Gambar 23 SHAP Beeswarm Plot

Analisis lokal melalui beeswarm plot SHAP pada gambar 23 menunjukkan bahwa nilai fitur yang tinggi (ditandai warna merah) pada fitur-fitur utama tersebut cenderung mendorong prediksi ke arah kelas Malignant, sedangkan nilai rendah (warna biru) berkontribusi pada prediksi Benign. Pola ini konsisten dengan pemahaman klinis bahwa peningkatan ukuran dan ketidakteraturan inti sel merupakan ciri khas keganasan.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan pipeline machine learning berbasis Gradient Boosted Trees yang terintegrasi dengan tahapan preprocessing komprehensif mampu menghasilkan performa klasifikasi kanker payudara yang sangat baik. Seluruh model yang diuji, yaitu XGBoost, LightGBM, CatBoost, HistGradientBoosting, dan GradientBoosting, menunjukkan kinerja yang tinggi dan relatif sebanding pada berbagai metrik evaluasi, meliputi accuracy, precision, recall, F1-score, dan ROC-AUC.

Di antara seluruh model, CatBoost menunjukkan performa numerik terbaik dengan akurasi sebesar 0,9649, precision 0,9750, recall 0,9286, F1-score 0,9512, serta ROC-AUC tertinggi sebesar 0,9960. Nilai precision yang tinggi mengindikasikan rendahnya tingkat kesalahan positif palsu pada prediksi kasus ganas, sedangkan nilai recall yang tinggi menunjukkan kemampuan model dalam mendeteksi sebagian besar kasus kanker payudara ganas. Kombinasi kedua metrik tersebut tercermin pada nilai F1-score yang tinggi, yang menunjukkan keseimbangan kinerja model dalam konteks klasifikasi medis. Meskipun CatBoost memiliki nilai ROC-AUC tertinggi secara numerik, hasil uji DeLong menunjukkan bahwa perbedaan performa antar model tidak signifikan secara statistik, sehingga seluruh algoritma Gradient Boosted Trees yang diuji dapat dikatakan memiliki kemampuan diskriminatif yang sebanding.

Selain performa, penelitian ini juga menekankan aspek interpretabilitas model melalui penerapan SHAP pada model CatBoost. Hasil analisis SHAP menunjukkan bahwa fitur-fitur yang berkaitan dengan ukuran dan bentuk inti sel, seperti radius, area, perimeter, concavity, dan concave points, memberikan kontribusi dominan terhadap prediksi keganasan. Temuan ini sejalan dengan literatur histopatologi kanker payudara, sehingga memperkuat bahwa model yang dihasilkan tidak hanya akurat secara kuantitatif, tetapi juga interpretable dan relevan secara klinis.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, proses seleksi fitur yang digunakan berbasis ANOVA bersifat univariat, sehingga hanya mempertimbangkan hubungan individual antara setiap fitur dan label kelas tanpa memperhitungkan interaksi antar fitur. Pada data medis yang kompleks, interaksi antar karakteristik sel berpotensi mengandung informasi penting yang belum sepenuhnya tertangkap oleh pendekatan ini. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penggunaan metode seleksi fitur multivariat atau embedded feature selection untuk memperoleh representasi fitur yang lebih komprehensif.

Kedua, evaluasi model dalam penelitian ini masih terbatas pada satu dataset publik, dengan pembagian data latih dan data uji yang berasal dari sumber yang sama, tanpa dilakukan validasi eksternal menggunakan dataset independen. Kondisi ini membatasi klaim keandalan model untuk penerapan klinis yang lebih luas. Oleh karena itu, diperlukan validasi lanjutan pada data dunia nyata (real-world data) atau dataset dari institusi dan populasi yang berbeda guna memastikan kemampuan generalisasi serta kestabilan model dalam konteks klinis yang sesungguhnya.

Penelitian ini menunjukkan bahwa kombinasi pipeline preprocessing yang robust, algoritma Gradient Boosted Trees, serta pendekatan Explainable Artificial Intelligence mampu menghasilkan model klasifikasi kanker payudara yang akurat, reliable secara internal, dan interpretable. Namun, pengembangan lanjutan melalui validasi eksternal, eksplorasi interaksi fitur yang lebih kompleks, serta pertimbangan aspek teknis dan etis tetap diperlukan sebelum model dapat diadopsi sebagai sistem pendukung keputusan klinis.

DAFTAR PUSTAKA

- [1] S. M. W and W. T. U, "Edukasi Kanker Payudara Pada Remaja Putri Melalui Media Daring Di SMP Negeri 1 Metro," pp. 131–134, 2020.
- [2] S. Susanto, S. A. Nugroho, and Y. T. Handoko, "Pengetahuan Ibu tentang Penyakit Kanker Payudara Berhubungan dengan Tingkat Kecemasan pada Pasien Pre Operasi Kanker Payudara," *J. Penelit. Perawat Prof.*, vol. 4, no. 2, pp. 589–598, 2022.
- [3] T. Sofa, A. Wardiyah, and Rilyani, "Faktor Risiko Kanker Payudara Pada Wanita," *J. Penelit. Perawat Prof.*, vol. 2, no. 5474, pp. 1333–1336, 2024.
- [4] <https://dinkes.kamparkab.go.id/>, "Fakta dan Permasalahan Kanker Payudara : Deteksi Dini itu Penting."
- [5] C. Hospital *et al.*, "Kualitas Hidup Pasien Kanker Payudara

- Menggunakan Kuesioner QLQ BR-23 di Rumah Sakit Kanker Dharmais Jakarta Barat,” vol. 9, no. 2, pp. 186–199, 2024.
- [6] R. R. Kadhimi and M. Y. Kamil, “Comparison of machine learning models for breast cancer diagnosis,” *IAES Int. J. Artif. Intell.*, vol. 12, no. 1, pp. 415–421, 2023.
- [7] M. A. Elsadig, A. Altigani, and H. T. Elshoush, “Breast cancer detection using machine learning approaches: a comparative study,” *Int. J. Electr. Comput. Eng.*, vol. 13, no. 1, pp. 736–745, 2023.
- [8] N. C. Ramadhan, H. H. H. T. Rohana, and A. M. Siregar, “Optimasi Algoritma Machine Learning Menggunakan Seleksi Fitur Xgboost Untuk Klasifikasi Kanker Payudara,” *TIN Terap. Inform. Nusantara*, vol. 5, no. 2, pp. 162–171, 2024.
- [9] V. Sangeetha *et al.*, “Breast cancer prediction using genetic algorithm and sand cat swarm optimization algorithm,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 37, no. 2, pp. 849–858, 2025.
- [10] P. Septiana Rizky, R. Haiban Hirzi, and U. Hidayaturohman, “Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang,” *J. Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, pp. 228–236, 2022.
- [11] L. W. Rizkallah, “Enhancing the performance of gradient boosting trees on regression problems,” *J. Big Data*, vol. 12, no. 1, 2025.
- [12] N. A. Pratama and D. W. Utomo, “Deteksi Diabetes Mellitus dengan Menggunakan Teknik Ensemble XGBoost dan LightGBM,” vol. XX, no. Xx, pp. 1–10, 2025.
- [13] R. Zizilia *et al.*, “Klasifikasi Penyakit Kanker Paru-Paru dengan Algoritma Extreme Gradient Boosting (XGBoost) dan Mutual Information sebagai Metode Feature Selection,” *Sist. J. Sist. Inf.*, vol. 14, no. 5, pp. 2540–9719, 2025.
- [14] A. Al Tawil, L. Almazaydeh, B. Alqudah, A. Z. Abualkishik, and A. A. Alwan, “Predictive modeling for breast cancer based on machine learning algorithms and features selection methods,” *Int. J. Electr. Comput. Eng.*, vol. 14, no. 2, pp. 1937–1947, 2024.
- [15] S. Akbulut, I. B. Cicek, and C. Colak, “Classification of Breast Cancer on the Strength of Potential Risk Factors with Boosting Models: A Public Health Informatics Application,” *Haseki Tip Bul.*, vol. 60, no. 3, pp. 196–203, 2022.
- [16] P. Wongyikul, N. Thongyot, P. Tantrakoolcharoen, P. Seephueng, and P. Khumrin, “High alert drugs screening using gradient boosting classifier,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–24, 2021.
- [17] N. L. Fitriyani *et al.*, “A Novel Approach Utilizing Bagging, Histogram Gradient Boosting, and Advanced Feature Selection for Predicting the Onset of Cardiovascular Diseases,” *Mathematics*, vol. 13, no. 13, pp. 1–26, 2025.
- [18] H. Luo *et al.*, “SHAP based predictive modeling for 1 year all-cause readmission risk in elderly heart failure patients: feature selection and model interpretation,” *Sci. Rep.*, vol. 14, no. 1, pp. 1–15, 2024.
- [19] M. R. Salmanpour, A. H. Pouria, S. Falahati, S. Taeb, and S. S. Mehrnia, “Robust Semi-Supervised CT Radiomics for Lung Cancer Prognosis: Cost-Effective Learning with Limited Labels and SHAP Interpretation,” no. M1, pp. 1–12.
- [20] D. Madhani, “Towards Explainable and Deployable AI for Personalized Diabetes Prediction: A Pipeline with SMOTE, Feature Selection, and SHAP-Based Model Interpretation,” *Rev. Electron. Vet.*, vol. 25, no. 1, pp. 3967–3976, 2025.
- [21] I. F. Rosyid and H. Pramaditya, “Visual Interpretation of Machine Learning Models (Random Forest) for Lung Cancer Risk Classification Using Explainable Artificial Intelligence (SHAP & LIME),” *J. Tek. Inform.*, vol. 6, no. 4, pp. 2187–2206, 2025.
- [22] X. Zhang, S. Lin, Q. Zeng, L. Peng, and C. Yan, “Machine learning and SHAP value interpretation for predicting cardiovascular disease risk in patients with diabetes using dietary antioxidants,” *Front. Nutr.*, vol. 12, no. July, pp. 1–15, 2025.
- [23] M. Yasser H, “Breast Cancer Dataset,” 2021.
- [24] I. N. Rizki, D. Prayoga, M. L. Puspita, and M. Q. Huda, “Implementasi Exploratory Data Analysis Untuk Analisis Dan Visualisasi Data Penderita Stroke Kalimantan Selatan Menggunakan Platform Tableau,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, 2024.
- [25] M. P. Pulungan, A. Purnomo, and A. Kurniasih, “Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 1033–1042, 2024.
- [26] M. Mutmainah and W. Yustanti, “Studi Komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada Analisis Anomali Kinerja Dosen,” *J. Informatics Comput. Sci.*, vol. 6, no. 02, pp. 532–540, 2024.
- [27] M. F. Thoriq, W. J. Pranoto, and F. Faldi, “Penerapan Seleksi Fitur Analysis of Variance Pada Algoritma Random Forest Classifier Dalam Klasifikasi Nilai Mahasiswa,” *Explor. J. Sist. Inf. dan Telemat.*, vol. 14, no. 2, p. 185, 2023.
- [28] J. Li, Y. Zhou, Y. Li, and Y. Liu, “Nuclear Morphological Characteristics in Breast Cancer: Correlation with Hormone Receptor and Human Epidermal Growth Factor Receptor 2,” vol. 2021, 2021.