

Analisis Perbandingan Performa Algoritma XGBoost dan LightGBM pada Klasifikasi Kanker Payudara

Danang Wijayanto¹, Bambang Pilu Hartato²

danangwijayanto@students.amikom.ac.id, Bambang.pilu@amikom.ac.id

Universitas Amikom Yogyakarta

Informasi Artikel

Diterima : 3 Apr 2024

Direview : 6 Apr 2024

Disetujui : 25 Apr 2024

Kata Kunci

XGBoost, LightGBM,
Klasifikasi, Kanker
Payudara, perbandingan

Abstrak

Kanker payudara adalah salah satu dari jenis kanker yang paling sering terjadi dan menyerang kalangan wanita di seluruh dunia. Dilihat dari kasus kematiannya, kanker payudara menempati posisi kedua kematian yang ditimbulkan akibat kanker. Metode *Fine needle aspiration* merupakan salah satu cara untuk mendeteksi kanker payudara secara dini, tetapi terdapat beberapa kekurangan seperti sampel yang terbatas sehingga mempengaruhi hasil akurasi diagnosis atau ketergantungan pada keahlian dan pengalaman yang melakukan metode tersebut. Machine learning dinilai dapat membantu mengatasi permasalahan di bidang kesehatan diantaranya dapat mendiagnosis seseorang terkena kanker atau tidak dengan menggunakan algoritma XGBoost dan LightGBM. XGBoost dan LightGBM merupakan algoritma yang efisien dalam melakukan pembelajaran dan memiliki perbedaan pada strategi pembelajaran yaitu *level-wise* dan *leaf-wise*. penelitian ini akan melakukan perbandingan performa akurasi, sensitivitas dan spesifisitas dari dua algoritma yaitu XBoost dan LightGBM untuk melihat algoritma mana yang dapat melakukan klasifikasi lebih baik. Dari hasil Percobaan diperoleh hasil XGBoost memiliki performa yang lebih bagus dengan memperoleh hasil rata-rata akurasi sebesar 97,03%, rata-rata sensitivitas 97,40% dan rata-rata spesifisitas 96,81% sedangkan LightGBM memperoleh rata-rata akurasi sebesar 95,59%, rata-rata sensitivitas 94,70% dan rata-rata spesifisitas 96,10%.

Keywords

XGBoost, LightGBM,
Classification, Breast
Cancer, comparison

Abstract

Breast cancer is one of the most common types of cancer and attacks women throughout the world. Judging from death cases, breast cancer is in second place in deaths caused by cancer. The fine needle aspiration method is one way to detect breast cancer early, but there are several disadvantages such as limited samples which affect the accuracy of the diagnosis or dependence on the skill and experience of the person carrying out the method. Machine learning is considered to be able to help overcome problems in the health sector, including being able to diagnose whether someone has cancer or not using the XGBoost and LightGBM algorithms. XGBoost and LightGBM are efficient algorithms for learning and have differences in learning strategies, namely level-wise and leaf-wise. This research will compare the accuracy, sensitivity and specificity performance of two algorithms, namely XBoost and LightGBM, to see which algorithm can perform better classification. From the experimental results, it was found that XGBoost had better performance by obtaining an average accuracy of 97.03%, an average sensitivity of 97.40% and an average specificity of 96.81%, while LightGBM obtained an average accuracy of 95.59%, average sensitivity 94.70% and average specificity 96.10%.

A. Pendahuluan

Kanker adalah sebuah penyakit yang muncul ketika sel tubuh yang normal menjadi bermutasi dan menyebar ke bagian tubuh lain melalui sistem peredaran darah [1]. Diperkirakan, terdapat 18,1 juta kasus kanker baru yang ada diseluruh dunia pada tahun 2020 dan diprediksi 10 juta kematian yang disebabkan kasus kanker ini. Beberapa jenis kanker yang paling umum terjadi yaitu kanker payudara, usus, prostat dan paru-paru [2]. Kanker payudara menjadi salah satu tipe kanker yang sangat sering terjadi dan paling banyak diderita oleh perempuan. Pada tahun 2020, terdapat 2,26 juta kasus baru dan 685.000 kematian yang terjadi akibat kanker payudara [3]. Kanker payudara terjadi ketika sel abnormal berkembang dan menyebar tanpa kendali ke seluruh bagian tubuh. Sel kanker muncul mula-mula pada saluran susu di payudara dan berkembang melalui jaringan payudara lainnya yang akhirnya menyebabkan tumor dan berpeluang menyebar ke bagian jaringan organ lainnya. Secara umum, terdapat dua jenis sel abnormal yang dapat dilakukan proses deteksi yaitu sel jinak (*benign*) dan sel kanker (*malignant*). Sel jinak memiliki sifat tidak menyebar dan tidak merusak jaringan-jaringan yang ada di sekitar sedangkan sel kanker dapat menyebar dengan cepat dan merusak jaringan-jaringan yang ada pada penderitanya[4].

Aspirasi jarum halus (*Fine Needle Aspiration*) adalah salah satu dari beberapa teknik medis yang digunakan oleh dokter dengan tahapan menggunakan jarum halus untuk mengambil sampel sel atau cairan dari area tertentu dalam tubuh pasien dengan tujuan untuk mengidentifikasi kondisi pasien jika terdapat penyakit seperti tumor, infeksi atau kista. Semakin berkembangnya teknologi terutama di bidang *artificial intelligence*, memberikan banyak kontribusi terhadap bidang-bidang lain salah satunya kesehatan. *Machine learning* yang merupakan cabang dari *artificial intelligence* dapat membantu dalam menyelesaikan permasalahan-permasalahan yang ada di bidang kesehatan. Metode klasifikasi yang menjadi bagian dari *supervised learning* pada *machine learning* dapat membantu dalam mengidentifikasi kanker payudara untuk melihat hasil analisis berupa sel kanker(*malignant*) atau sel jinak(*benign*). Sampel gambar digital dari massa payudara yang telah diambil menggunakan aspirasi jarum halus (*Fine Needle Aspiration*) dan dihitung karakteristik inti selnya bisa digunakan untuk melatih sebuah model *machine learning* untuk tugas klasifikasi. Model yang telah dibuat dapat membantu ahli patologi untuk mengambil keputusan yang lebih efisien dan akurat serta mengurangi *human error* yang dapat terjadi. Selain itu, teknologi *machine learning* ini juga akan memberikan inovasi baru dalam bidang kesehatan yang berpengaruh pada meningkatkannya kualitas layanan kesehatan yang diberikan.

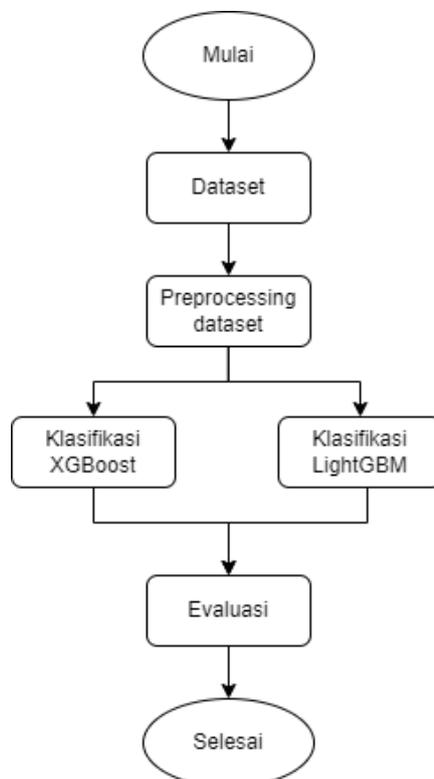
Penelitian sebelumnya yang dilakukan pada [5] menggunakan algoritma *logistic regression* serta menerapkan *hyperparameter tuning* memperoleh tingkat akurasi yang sangat baik yaitu 98%. Penelitian lainnya pada [6] melakukan komparasi algoritma *Multi Layer Perception (MLP)* dan *Support Vector machine (SVM)* menggunakan aplikasi *orange data mining* memperoleh hasil algoritma *Multi Layer Perception (MLP)* dengan aktivasi *logistic* dan optimasi *adam* lebih baik dibandingkan *Support Vector Machine (SVM)* dengan akurasi sebesar 97,7%.

Penelitian lainnya pada [7] menggunakan algoritma K-Nearest Neighbor (KNN) dan Decision Tree C4.5 pada proses klasifikasi dan menghasilkan akurasi sebesar 96,49% pada algoritma C4.5 dan 94,73% pada algoritma K-Nearest Neighbor (KNN). Pada algoritma C4.5 metode imputasi tidak berdampak signifikan pada hasil akurasi yang diperoleh.

Penelitian kali ini akan melakukan percobaan klasifikasi dengan membandingkan algoritma XGBoost dan LightBGM untuk melihat algoritma mana yang memiliki performa lebih baik dalam proses klasifikasi kanker payudara. Kedua algoritma tersebut termasuk kedalam metode *ensemble learning* yang memiliki keunggulan dengan mengumpulkan model lemah akan digabung dengan bertahap hingga memperoleh model yang lebih kuat. Walaupun memiliki dasar yang sama dari metode *ensemble learning*, strategi pembelajaran XGBoost menggunakan pendekatan level-wise sedangkan LightGBM menggunakan pendekatan leaf-wise yang bisa berpengaruh terhadap hasil yang diperoleh. Oleh karena itu, pada penelitian kali ini akan dilakukan percobaan pada kedua algoritma tersebut agar dapat memberikan sudut pandang baru dari perbandingan algoritma-algoritma yang digunakan.

B. Metode Penelitian

Pada penelitian ini akan mencoba menerapkan algoritma XGBoost dan LightBGM pada dataset kanker payudara dan terdiri dari beberapa tahapan secara garis besar hingga menghasilkan output yang diharapkan. Dapat dilihat dari diagram alir pada gambar di bawah ini.



Gambar 1. Tahapan Penelitian

1. Dataset Kanker Payudara

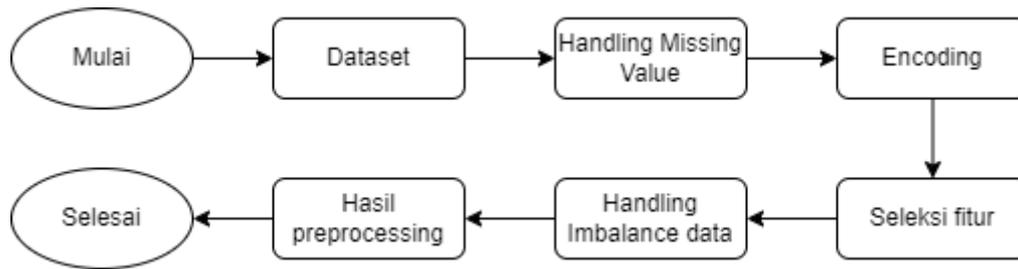
Data yang digunakan pada penilitan kali ini bersumber dari data publik kaggle.com (<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>). Terdapat 569 baris dan 31 kolom pada dataset yang akan digunakan [8]. Data berupa hasil perhitungan gambar digital massa payudara yang telah dilakukan dengan metode *Fine Needle Aspiration*.

Tabel 1. Data Kanker Payudara

No	Atribut	Type
1	diagnosis	String
2	radius_mean	float
3	texture_mean	Float
4	perimeter_mean	Float
5	area_mean	Float
6	smoothness_mean	Float
7	compactness_mean	Float
8	concavity_mean	Float
9	concave points_mean	Float
10	symmetry_mean	Float
11	fractal_dimension_mean	Float
12	radius_se	Float
13	texture_se	Float
14	perimeter_se	Float
15	area_se	Float
16	smoothness_se	Float
17	compactness_se	Float
18	compactness_se	Float
19	concave points_se	Float
20	symmetry_se	Float
21	fractal_dimension_se	Float
22	radius_worst	Float
23	texture_worst	Float
24	perimeter_worst	Float
25	area_worst	Float
26	smoothness_worst	Float
27	compactness_worst	Float
28	concavity_worst	Float
29	concave points_worst	Float
30	symmetry_worst	Float
31	fractal_dimension_worst	float

2. Preprocessing Dataset

Selanjutnya pada tahap preprocessing akan dilakukan beberapa tahapan untuk mengolah data yang akan digunakan. Tujuan pada preprocessing ini agar model *machine learning* yang digunakan semakin akurat pada saat mempelajari data. Berikut adalah diagram alir dari tahap preprocessing.



Gambar 2. Tahapan Preprocessing

2.1. Handling Missing Value

Handling missing value dilakukan untuk mengatasi nilai yang kosong pada tiap kolom. Nilai yang kosong dapat terjadi ketika kesalahan pada saat menginput data [9]. Ketika terdapat nilai yang kosong pada salah satu kolom maka algoritma tidak dapat bekerja.

2.2. Encoding

Encoding dilakukan untuk mengubah atribut yang masih memiliki tipe data selain numerik. Ketika sudah pada tahap pembuatan model, maka semua tipe data pada dataset harus bertipe numerik.

2.3. Seleksi Fitur

Agar algoritma yang digunakan semakin optimal dan memperoleh hasil performa yang bagus, seleksi fitur perlu dilakukan untuk mengurangi kompleksitas model dan waktu komputasi [10]. Fitur yang memiliki nilai korelasi kecil akan dihapus dari dataset.

2.4. Handling Imbalance Data

Ketika distribusi nilai pada dataset tidak tersebar dengan merata, maka dapat menyebabkan model yang dilatih akan menjadi bias terhadap nilai minoritas. Maka diperlukan handling imbalance data untuk mengimbangi antara nilai mayoritas dan minoritas. Pada penelitian kali ini akan menggunakan teknik SMOTE karena memiliki kinerja yang seimbang dan akurat [11].

3. XGBoost

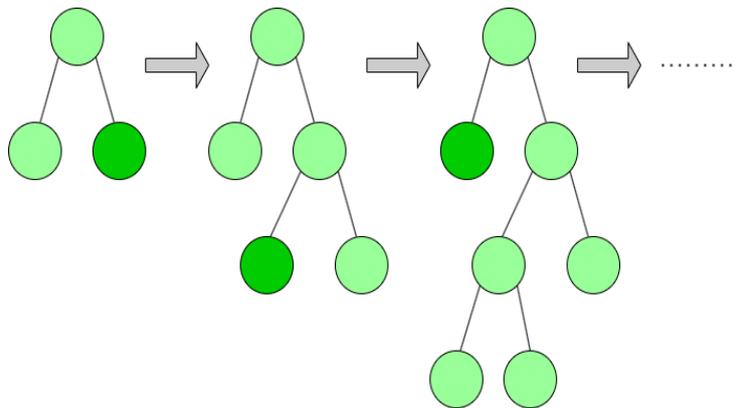
XGBoost merupakan implementasi dari gradient boosting yang memiliki cara kerja dengan menggabungkan kumpulan pohon keputusan yang diperoleh secara bertahap. Selanjutnya tiap pohon keputusan akan mengoreksi prediksi model sebelumnya hingga memperoleh hasil yang paling optimal. Algoritma akan menggunakan K fungsi tambahan untuk memprediksi nilai keluaran Y_i untuk tiap X_i pada persamaan (1) dan mengukur kesalahan prediksi yang terjadi pada model terhadap nilai target pada persamaan (2) [12].

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^K f_k(\mathbf{x}_i) \quad (1)$$

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

4. LightGBM

LightGBM adalah algoritma yang termasuk dalam bagian gradient boosting yang menerapkan teknik leaf wise. Tiap langkah-langkahnya akan memilih leaf dengan pembagian keuntungan terbesar dan terus membaginya agar mengurangi kesalahan dan memperoleh akurasi lebih baik [13].



Gambazr 3. Ilustrasi LightGBM

5. Evaluasi

Pada tahap akhir akan dilakukan evaluasi untuk melihat performa tiap model yang telah menerapkan algoritma XGBoost dan LightGBM. Confussion matrix akan digunakan sebagai alat evaluasi untuk penelitian kali ini. Bentuk confussion Matrix berupa tabel untuk membandingkan hasil klasifikasi sistem terhadap hasil sesungguhnya [14]. Dari hasil Confussion Matrix tersebut akan digunakan matrix evaluasi yaitu akurasi, sensitivitas, dan spesifisitas.

		PREDICTED	
		Positive	Negative
ACTUAL	Positive	TP	FN
	Negative	FP	TN

Gambar 4. Confussion Matrix

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Sensitivitas} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Spesifisitas} = \frac{TN}{FP + TN} \quad (5)$$

C. Hasil dan Pembahasan

1. Preprocessing Data

Pada penjelasan sebelumnya, preprocessing data akan dilakukan untuk mengolah dataset agar memiliki kualitas lebih bagus ketika dilakukan pelatihan model menggunakan algoritma. Terdiri dari beberapa tahapan sebagai berikut.

1.1. Handling Missing Value

Terdapat nilai null pada dataset kanker payudara yang bisa dilihat pada gambar 5. Atribut Unnamed:32 akan dihapus berserta dengan id karena tidak berhubungan dengan hasil yang diperoleh.

	Nama Kolom	Tipe Data	Null
id	id	int64	0
diagnosis	diagnosis	object	0
radius_mean	radius_mean	float64	0
texture_mean	texture_mean	float64	0
perimeter_mean	perimeter_mean	float64	0
area_mean	area_mean	float64	0
smoothness_mean	smoothness_mean	float64	0
compactness_mean	compactness_mean	float64	0
concavity_mean	concavity_mean	float64	0
concave points_mean	concave points_mean	float64	0
symmetry_mean	symmetry_mean	float64	0
fractal_dimension_mean	fractal_dimension_mean	float64	0
radius_se	radius_se	float64	0
texture_se	texture_se	float64	0
perimeter_se	perimeter_se	float64	0
area_se	area_se	float64	0
smoothness_se	smoothness_se	float64	0
compactness_se	compactness_se	float64	0
concavity_se	concavity_se	float64	0
concave points_se	concave points_se	float64	0
symmetry_se	symmetry_se	float64	0
fractal_dimension_se	fractal_dimension_se	float64	0
radius_worst	radius_worst	float64	0
texture_worst	texture_worst	float64	0
perimeter_worst	perimeter_worst	float64	0
area_worst	area_worst	float64	0
smoothness_worst	smoothness_worst	float64	0
compactness_worst	compactness_worst	float64	0
concavity_worst	concavity_worst	float64	0
concave points_worst	concave points_worst	float64	0
symmetry_worst	symmetry_worst	float64	0
fractal_dimension_worst	fractal_dimension_worst	float64	0
Unnamed: 32	Unnamed: 32	float64	569

Gambar 5. Keterangan Data Kanker Payudara

1.2. Encoding

Encoding akan dilakukan pada atribut diagnosis yang masih bertipe object dan diubah menjadi tipe data numerik.

	Nama Kolom	Tipe Data	Total
0	diagnosis	object	569

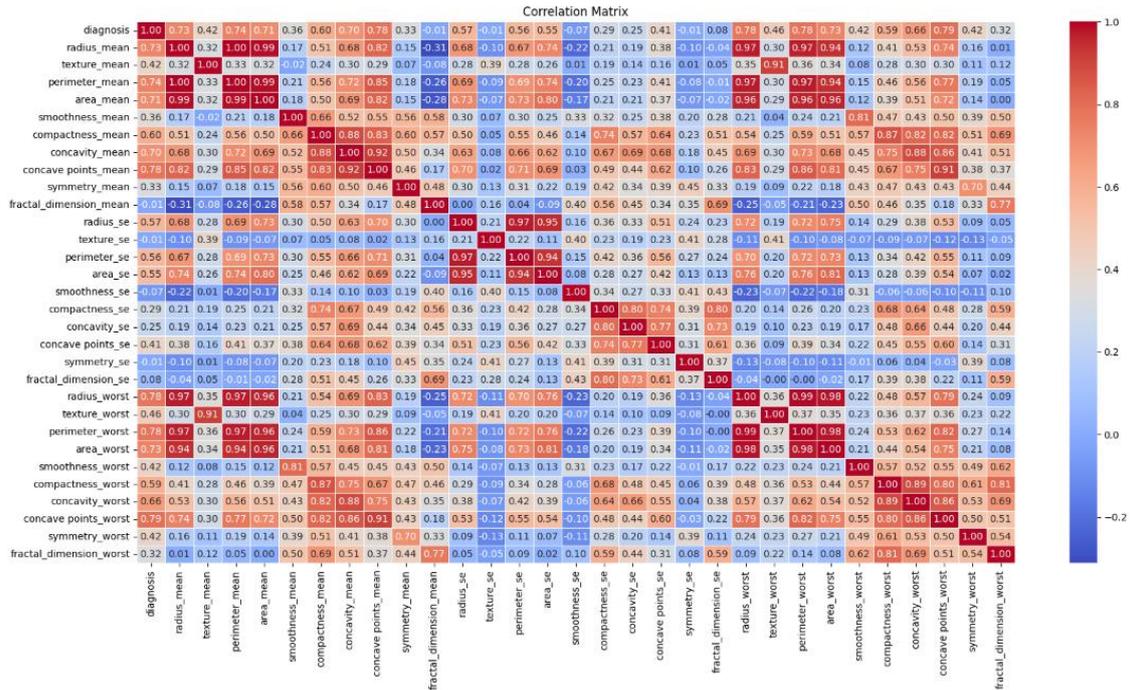
Gambar 6. Sebelum encoding

	Nama Kolom	Tipe Data	Total
0	diagnosis	float64	569

Gambar 7. Setelah encoding

1.3. Seleksi Fitur

Seleksi fitur diterapkan untuk mengurangi kompleksitas dan dimensi pada dataset. Dapat dilihat pada gambar 8, korelasi matrix antar atribut. Atribut yang memiliki nilai korelasi dibawah 0,2 terhadap variabel terikat (diagnosis) akan dihapus pada dataset. Maka terdapat 5 atribut yang memiliki korelasi dibawah 0,2 sehingga jumlah artibut yang semula 31 menjadi 26.



Gambar 8. Korelasi matrix

1.4. Handling Imbalance Data

Teknik SMOTE akan digunakan pada tahap handling imbalance data untuk mengatasi ketidakseimbangan nilai pada data *train*.

```
smote = SMOTE()
X_resampled, y_resampled = smote.fit_resample(X_train,y_train)
```

Gambar 9. Penerapan teknik SMOTE pada data train

2. Klasifikasi Model Menggunakan XGBoost dan LightGBM

Setelah melalui tahap preprocessing, data akan di bagi menjadi data *train* dan data *test* dengan komposisi 70% : 30% sebelum melakukan pelatihan model. Pada data *train* akan diterapkan teknik SMOTE agar pelatihan model lebih akurat seperti pada penjelasan sebelumnya. penelitian kali ini akan melakukan 30 kali percobaan pelatihan pada tiap model XGBoost ataupun LightBGM dengan tujuan mendapatkan nilai rata-rata dari hasil evaluasi serta memberikan hasil yang lebih realistis.

Tabel 2. Hasil perhitungan model dengan algoritma LightGBM

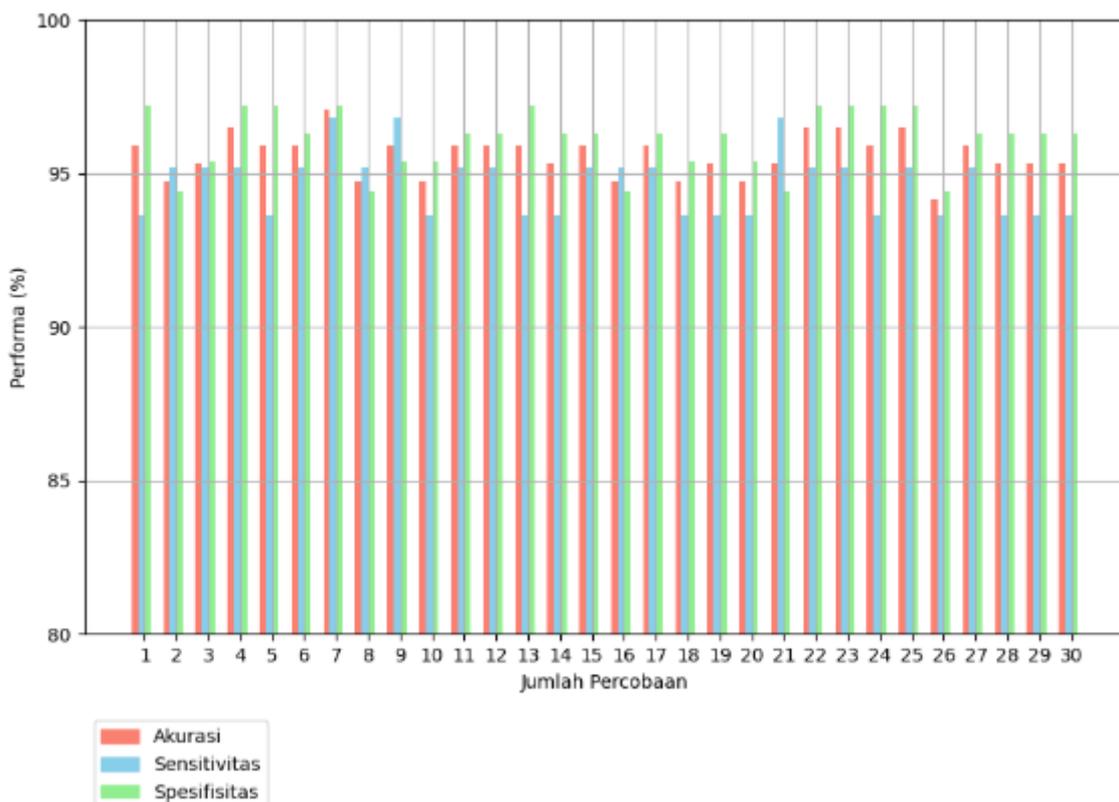
Percobaan	Akurasi	Sensitivitas	Spesifisitas
1	95,90%	93,65%	97,22%
2	94,73%	95,23%	94,44%
3	95,32%	95,23%	95,37%
4	96,49%	95,23%	97,22%
5	95,90%	93,65%	97,22%
.....
30	95,32%	93,65%	96,29%
Rata-rata	95,59%	94,70%	96,10%

Tabel 2 menampilkan hasil percobaan model menggunakan algoritma LightGBM sebanyak 30 kali dan menunjukkan bahwa model memperoleh rata-rata akurasi 95,59%, rata-rata sensitivitas 94,70%, dan rata-rata spesifisitas 96,10%.

Tabel 3. Hasil perhitungan model dengan algoritma XGBoost

Percobaan	Akurasi	Sensitivitas	Spesifisitas
1	97,66%	96,82%	98,14%
2	97,07%	98,41%	96,29%
3	97,07%	98,41%	96,29%
4	97,66%	98,41%	97,22%
5	97,66%	96,82%	98,14%
.....
30	97,66%	98,41%	97,22%
Rata-rata	97,03%	97,40%	96,81%

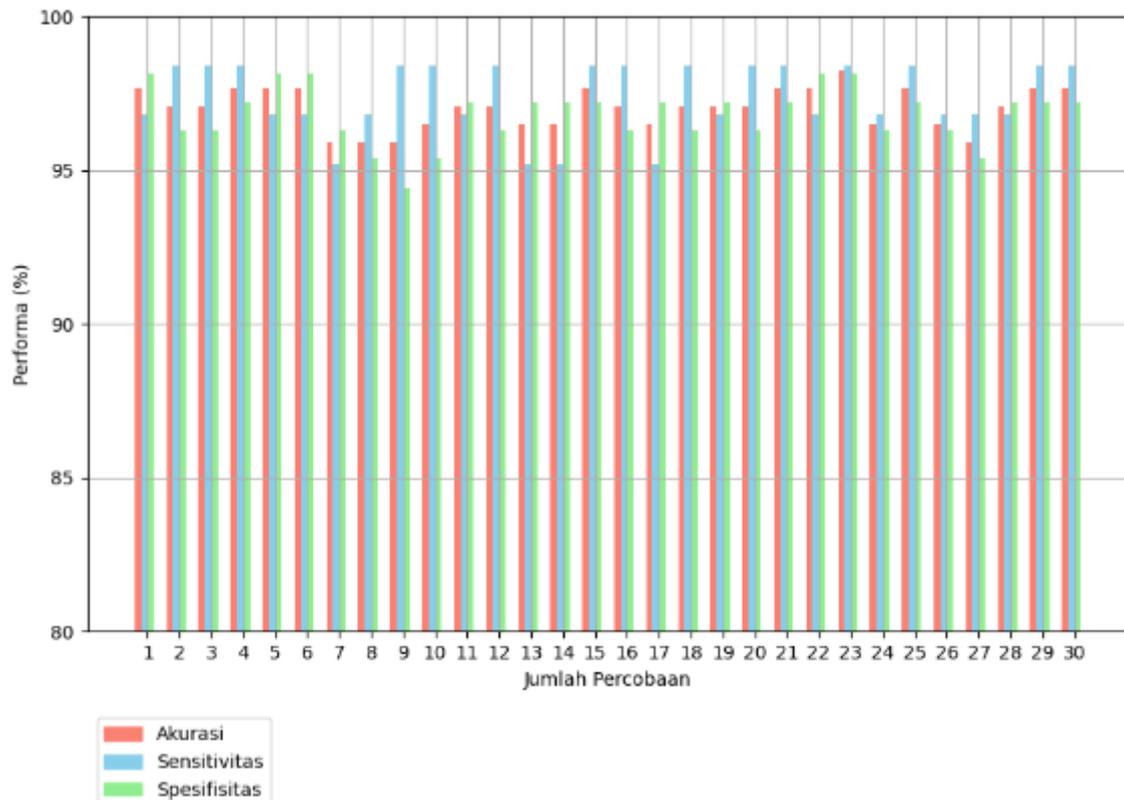
Tabel 3 menampilkan hasil percobaan model menggunakan algoritma XGBoost sebanyak 30 kali dan menghasilkan rata-rata akurasi sebesar 97,03%, rata-rata sensitivitas 97,40%, dan rata-rata spesifisitas 96,81%.



Gambar 9. Performa Model dengan Algoritma LightGBM

Gambar 9 menampilkan performa akurasi, sensitivitas, dan spesifisitas yang dihasilkan dari percobaan model sebanyak 30 kali menggunakan algoritma LightGBM. Dapat dilihat akurasi memiliki nilai relatif diatas 95% walaupun terdapat beberapa kali mengalami penurunan pada saat pelatihan model sedangkan sensitivitas cenderung memiliki nilai di antara 90- 95% dan beberapa

kali mengalami kenaikan. Spesifisitas menunjukkan kondisi yang kurang lebih sama dengan nilai akurasi yang relatif diatas 95%.



Gambar 10. Performa Model dengan Algoritma XGBoost

Gambar 10 menunjukkan hasil performa model menggunakan algoritma XGBoost pada pelatihan model sebanyak 30 kali percobaan. Akurasi yang didapat dari percobaan tersebut memiliki nilai selalu di atas 95% dengan performa terbaik di angka 98%. Sensitivitas menunjukkan nilai fluktuatif, namun tetap berada di rentan 95% keatas sedangkan spesifisitas hanya mengalami satu kali penurunan dibawah 95% dari 30 kali percobaan yang telah dilakukan.

D. Simpulan

Dari hasil penelitian yang telah dilakukan menggunakan algoritma LightGBM dan XGBoost dengan percobaan pelatihan model sebanyak 30 kali dan menggunakan dataset kanker payudara dengan jumlah 569 baris data dan 26 atribut dapat disimpulkan algoritma XGBoost memiliki performa lebih bagus dibandingkan LightGBM dengan unggul di ketiga matrix evaluasi yaitu rata-rata akurasi 97,03%, rata-rata sensitivitas 97,40%, dan rata-rata spesifisitas 96,81%. Walaupun XGboost lebih unggul dibandingkan LightGBM, namun performa LightGBM masih terbilang bagus dengan selisih nilai rata-rata pada akurasi 1,44%, sensitivitas 2,7% dan spesifisitas 0,71% terhadap XGBoost

E. Ucapan Terima Kasih

Terima kasih kepada Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah membantu dalam memberikan arahan pada penelitian kali ini.

F. Referensi

- [1] National Cancer Institute, <https://www.cancer.gov/about-cancer/understanding/what-is-cancer> (accessed March. 24, 2023).
- [2] Cancer Research UK, <https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer> (accessed March. 24, 2023)
- [3] World Health Organization, <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed March. 24, 2023)
- [4] Farahdiba, Balqis Aisyah, and Yusuf Sulisty Nugroho. "Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio." *Jurnal Teknik Elektro* 8.2 (2016): 43-46.
- [5] A'yunan, Yahya Anugerah Dwi Khurrota, Uce Indahyanti, and Suhendro Busono. "IMPLEMENTASI DATA MINING DALAM KLASIFIKASI DIAGNOSA KANKER PAYUDARA MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION." *Jurnal Tekinkom (Teknik Informasi dan Komputer)* 6.2 (2023): 400-407.
- [6] KUSUMA, JAKA, et al. "Komparasi Metode Multi Layer Perceptron (MLP) dan Support Vector Machine (SVM) untuk Klasifikasi Kanker Payudara." *MIND (Multimedia Artificial Intelligent Networking Database) Journal* 7.1 (2022): 51-60.
- [7] Ghalib, Fahrurrozi, and Wasilah Wasilah. "Deteksi Dini Kanker Payudara Menggunakan Algoritma K-Nearest Neighbour (KNN) dan Decision Tree C-45." *TEKNIKA: Jurnal Ilmiah Bidang Ilmu Rekayasa* 17.2 (2023): 427-434.
- [8] (UCI MACHINE LEARNING. 2017). Breast Cancer Wisconsin (Diagnostic) Data Set. <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>.
- [9] Aji, Priyo Wahyu Setiyo, S. Suprianto, and Rohman Dijaya. "Prediksi Penyakit Stroke Menggunakan Metode Random Forest." *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)* 4.4 (2023): 916-924.
- [10] G. Koneti, M. Shaik, S. Das and R. Narayanan, "Parallelized Population Based Exhaustive Replacement Method for Reproducible Features Selection and Its Applications in Drug Discovery," in 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, 2022 pp. 3837-3839.
- [11] Maldonado, Sebastián, Julio López, and Carla Vairetti. "An alternative SMOTE oversampling strategy for high-dimensional datasets." *Applied Soft Computing* 76 (2019): 380-389.
- [12] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.
- [13] Z. Qiao, "Walmart Sale Forecasting Model Based On LightGBM," in 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 2020 pp. 76-79.
- [14] Oktafiani, Rian, Arief Hermawan, and Donny Avianto. "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning." *Jurnal Sains dan Informatika* (2023): 19-28.