

## KOMPARASI MODEL MACHINE LEARNING UNTUK PREDIKSI PENERIMAAN DEPOSITO BERJANGKA PADA KAMPANYE TELEMARKETING BANK

Marcellio Aurel Christian

Universitas Pembangunan Nasional “Veteran” Jawa Timur  
[22082010019@student.upnjatim.ac.id](mailto:22082010019@student.upnjatim.ac.id)

### Abstract

*Telemarketing campaigns are widely used by banks to promote term deposit products, yet their success rate is often low because the offers are not targeted to the right customers. This study aims to compare several machine learning models for predicting customers' acceptance of term deposit offers, so that banks can conduct more effective and efficient campaigns. The dataset used is the Bank Marketing Dataset, which contains 45,211 customer records with demographic, socio-economic, and campaign-related attributes. The research stages include exploratory data analysis to understand the data characteristics and class imbalance, followed by data preprocessing such as handling “unknown” values, encoding categorical variables using one-hot encoding, transforming the target label into binary classes, and splitting the data into training and test sets using a stratified scheme. The models evaluated in this study are Decision Tree, Random Forest, and XGBoost, which are further optimized using Grid Search Cross-Validation. Model performance is measured using accuracy, precision, recall, and F1-score. The experimental results show that the tuned XGBoost model achieves the best performance with accuracy above 90% and stable results across different data subsets. This model can be utilized as a decision support tool to prioritize customers with a high probability of accepting term deposit offers and to improve the efficiency of telemarketing campaigns.*

**Keywords:** bank marketing dataset, machine learning, classification, term deposit, telemarketing.

### Abstrak

Kampanye telemarketing merupakan salah satu strategi yang umum digunakan perbankan untuk memasarkan produk deposito berjangka. Namun, tingkat keberhasilan kampanye ini sering kali rendah karena penawaran tidak tepat sasaran dan tidak sesuai dengan karakteristik calon nasabah. Penelitian ini bertujuan untuk membandingkan beberapa model machine learning dalam memprediksi penerimaan deposito berjangka sehingga dapat membantu bank menyusun strategi kampanye yang lebih efektif. Data yang digunakan adalah Bank Marketing Dataset yang berisi 45.211 data nasabah dengan atribut demografis, sosial, ekonomi, serta informasi terkait kampanye telemarketing. Tahapan penelitian meliputi exploratory data analysis untuk memahami karakteristik data dan ketidakseimbangan kelas, dilanjutkan dengan data preprocessing berupa penanganan nilai “unknown”, pengkodean variabel kategorikal menggunakan one-hot encoding, transformasi label target menjadi kelas biner, serta pemisahan data menjadi data latih dan data uji secara stratified. Selanjutnya dilakukan pemodelan menggunakan algoritma Decision Tree, Random Forest, dan XGBoost yang kemudian dioptimasi dengan Grid Search Cross-Validation. Evaluasi model dilakukan menggunakan metrik

accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa XGBoost teroptimasi memberikan performa terbaik dengan akurasi di atas 90% dan stabil pada berbagai subset data. Model ini berpotensi dimanfaatkan sebagai sistem pendukung keputusan untuk memprioritaskan calon nasabah dengan probabilitas tinggi menerima penawaran deposito berjangka dan sekaligus meningkatkan efisiensi kampanye telemarketing.

**Kata kunci:** bank marketing dataset, machine learning, klasifikasi, deposito berjangka, telemarketing.

## PENDAHULUAN

Industri perbankan menghadapi persaingan yang semakin ketat dalam upaya menghimpun dana pihak ketiga, salah satunya melalui produk deposito berjangka. Bank menggunakan berbagai strategi pemasaran untuk menarik minat nasabah, termasuk kampanye pemasaran langsung (direct marketing) melalui telepon. Meskipun demikian, tingkat keberhasilan kampanye telemarketing seringkali rendah karena sebagian besar calon nasabah tidak tertarik dengan penawaran yang disampaikan. Kondisi ini berdampak pada rendahnya rasio konversi, tingginya beban kerja agen telemarketing, serta pemborosan biaya operasional yang tidak sebanding dengan jumlah nasabah baru yang berhasil diperoleh. Bank Marketing Dataset yang dikumpulkan dari sebuah bank di Portugal menyediakan data historis kampanye telemarketing yang cukup lengkap. Dataset ini memuat atribut demografis, sosial, ekonomi, serta informasi terkait interaksi kampanye, seperti jumlah kontak, durasi panggilan, dan hasil kampanye sebelumnya. Variabel target pada dataset tersebut adalah keputusan nasabah untuk berlangganan produk deposito berjangka dengan dua kemungkinan label, yaitu yes dan no. Distribusi target yang tidak seimbang, di mana jumlah nasabah yang menolak penawaran jauh lebih besar dibandingkan yang menerima, menjadikan permasalahan ini menantang dan relevan untuk diselesaikan dengan pendekatan machine learning (Moro et al., 2014).

Pendekatan data-driven memungkinkan bank memanfaatkan data historis untuk membangun model prediktif yang dapat mengestimasi kecenderungan calon nasabah menerima atau menolak penawaran deposito berjangka. Berbagai algoritma klasifikasi telah digunakan pada Bank Marketing Dataset, seperti K-Nearest Neighbor (KNN), Decision Tree, Random Forest, dan XGBoost. Salah satu penelitian melaporkan bahwa XGBoost memberikan kinerja terbaik dengan akurasi mencapai 92,36%, diikuti oleh Decision Tree, KNN, dan Random Forest. Hasil tersebut menunjukkan bahwa model ensemble dan gradient boosting berpotensi memberikan performa yang unggul dalam konteks prediksi penerimaan deposito berjangka (Maulidah, 2023). Meskipun telah dilakukan beberapa penelitian sebelumnya, masih diperlukan kajian yang secara sistematis membandingkan kinerja beberapa algoritma machine learning dengan alur pemodelan yang terstruktur, mulai dari exploratory data analysis (EDA), preprocessing, pembentukan fitur, hingga optimasi hyperparameter dan evaluasi akhir. Selain itu,

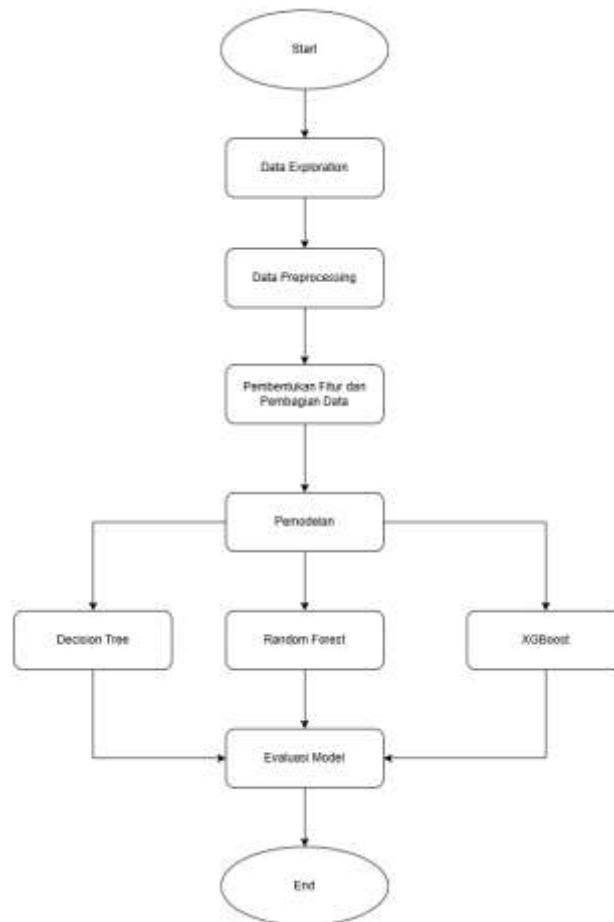
dibutuhkan pula pembahasan yang menekankan implikasi praktis model terbaik terhadap proses bisnis kampanye telemarketing, misalnya sebagai dasar penyusunan sistem lead scoring untuk memprioritaskan calon nasabah yang berpotensi tinggi menerima penawaran.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada komparasi model machine learning untuk prediksi penerimaan deposito berjangka pada kampanye telemarketing bank menggunakan Bank Marketing Dataset. Tujuan penelitian adalah (1) membangun dan membandingkan model Decision Tree, Random Forest, dan XGBoost dalam memprediksi keputusan nasabah terhadap penawaran deposito berjangka; (2) melakukan optimasi hyperparameter menggunakan Grid Search Cross-Validation untuk memperoleh konfigurasi model terbaik; serta (3) menganalisis potensi pemanfaatan model terbaik sebagai pendukung pengambilan keputusan dalam strategi kampanye telemarketing.

## **METODE PENELITIAN**

Bagian ini menjelaskan data yang digunakan serta tahapan penelitian yang dilakukan mulai dari eksplorasi data, pra pemrosesan, pembentukan fitur dan target, pemodelan, hingga evaluasi model. Seluruh proses komputasi dilakukan menggunakan bahasa pemrograman Python dengan bantuan beberapa pustaka seperti pandas, scikit-learn, dan XGBoost pada lingkungan Google Colab.

Data yang digunakan dalam penelitian ini adalah Bank Marketing Dataset yang berasal dari kampanye pemasaran langsung sebuah bank di Portugal. Dataset terdiri atas 45.211 baris dan 17 atribut yang merepresentasikan karakteristik nasabah serta informasi kampanye. Terdapat tujuh atribut numerik, yaitu age, balance, day, duration, campaign, pdays, dan previous, serta sepuluh atribut kategorikal, yaitu job, marital, education, default, housing, loan, contact, month, poutcome, dan y sebagai variabel target. Variabel y menyatakan keputusan nasabah untuk berlangganan deposito berjangka dengan dua nilai, yaitu yes (menerima) dan no (menolak), sehingga permasalahan yang dikaji merupakan tugas klasifikasi biner. Distribusi kelas pada variabel target tidak seimbang, di mana jumlah nasabah dengan label no jauh lebih banyak dibandingkan label yes, sehingga diperlukan pendekatan pemodelan dan evaluasi yang memperhatikan kondisi imbalanced data (Moro et al., 2014)



Gambar 1. Diagram Alir Penelitian (Maulidah, 2023)

Secara umum, alur penelitian terdiri atas beberapa tahapan utama, yaitu *exploratory data analysis* (EDA), *data preprocessing*, pembentukan fitur dan target serta pembagian data, pemodelan menggunakan beberapa algoritma klasifikasi, dan evaluasi dan komparasi kinerja model.

Tahap *exploratory data analysis* (EDA) dilakukan untuk memahami karakteristik awal Bank Marketing Dataset sebelum data digunakan dalam pemodelan. Pada tahap ini, peneliti mengamati struktur data, jenis atribut, serta distribusi masing-masing variabel baik numerik maupun kategorikal. Analisis dilakukan dengan melihat statistik deskriptif, distribusi nilai pada fitur numerik seperti *age*, *balance*, dan *duration*, serta frekuensi kategori pada fitur seperti *job*, *marital*, dan *education*. Selain itu, pada tahap ini juga diidentifikasi adanya nilai khusus “unknown” pada beberapa fitur kategorikal, serta keberadaan outlier pada fitur numerik yang memiliki rentang nilai sangat lebar. EDA juga digunakan untuk melihat distribusi kelas pada variabel target *y* dan mengonfirmasi bahwa data bersifat tidak seimbang, di mana jumlah nasabah yang menolak penawaran deposito berjangka jauh lebih banyak dibandingkan yang menerima. Hasil pengamatan EDA menjadi dasar dalam merancang strategi pra pemrosesan data dan pendekatan pemodelan yang akan digunakan pada tahap berikutnya.

Tahap data preprocessing bertujuan menyiapkan data agar berada dalam kondisi yang bersih, konsisten, dan sesuai dengan kebutuhan algoritma machine learning. Langkah pertama yang dilakukan adalah menangani nilai “unknown” pada fitur kategorikal, seperti job, education, contact, dan poutcome. Nilai “unknown” tidak dihapus, tetapi digantikan dengan modus dari masing-masing fitur sehingga informasi data tetap terjaga tanpa mengurangi jumlah baris secara signifikan. Selanjutnya, seluruh fitur kategorikal selain y dikonversi menjadi bentuk numerik menggunakan teknik One-Hot Encoding, dengan pengaturan penghapusan satu kategori dasar untuk menghindari dummy variable trap. Tahap ini menghasilkan sekumpulan fitur biner yang merepresentasikan keberadaan suatu kategori (Developers, n.d.-a). Setelah itu, variabel target y yang semula bertipe teks dengan nilai yes dan no dipetakan menjadi label biner, yaitu 1 untuk nasabah yang menerima penawaran deposito berjangka dan 0 untuk nasabah yang menolak. Melalui rangkaian proses ini, diperoleh dataframe akhir yang sepenuhnya numerik dan siap digunakan pada tahap pembentukan fitur dan pemodelan.

Setelah proses preprocessing selesai, dilakukan pembentukan fitur dan target untuk kebutuhan pelatihan model. Variabel fitur (X) didefinisikan sebagai seluruh kolom pada data frame kecuali kolom y, sedangkan variabel target (y) berisi label biner hasil transformasi sebelumnya yang merepresentasikan keputusan nasabah terhadap penawaran deposito berjangka. Pada tahap ini dilakukan pemeriksaan ulang untuk memastikan bahwa seluruh fitur bertipe numerik dan tidak terdapat nilai hilang (missing values) pada X maupun y, sehingga tidak menimbulkan error ketika proses pelatihan model dijalankan. Selanjutnya, data dibagi menjadi dua bagian menggunakan metode train-test split dengan proporsi 80% sebagai data latih dan 20% sebagai data uji. Pembagian ini menggunakan skema stratified split terhadap variabel target, sehingga distribusi kelas yes dan no pada data latih dan data uji tetap sebanding dengan distribusi awal dataset (Developers, n.d.-b). Dengan demikian, evaluasi model pada data uji diharapkan dapat menggambarkan kinerja model secara lebih representatif terhadap data baru.

Tahap pemodelan dilakukan untuk membangun model klasifikasi yang mampu memprediksi penerimaan deposito berjangka berdasarkan fitur-fitur yang telah disiapkan. Tiga algoritma machine learning yang digunakan dalam penelitian ini adalah Decision Tree, Random Forest, dan XGBoost. Proses pemodelan diawali dengan baseline modelling, yaitu pelatihan ketiga model menggunakan parameter standar pada data latih dan pengujian pada data uji untuk memperoleh gambaran awal kinerja masing-masing algoritma. Nilai akurasi, precision, recall, dan F1-score dihitung untuk menilai kemampuan model dalam melakukan klasifikasi. Setelah performa awal diperoleh, dilakukan tahap optimasi menggunakan Grid Search Cross-Validation untuk mencari kombinasi hyperparameter terbaik bagi setiap algoritma. Proses grid search ini menggunakan skema 5-fold Stratified Cross-Validation sehingga setiap kombinasi

parameter dievaluasi pada lima lipatan data dengan distribusi kelas yang seimbang. Model dengan rata-rata akurasi terbaik pada proses cross-validation kemudian dipilih sebagai model teroptimasi untuk Decision Tree, Random Forest, dan XGBoost.

Tahap evaluasi model bertujuan menilai dan membandingkan kinerja ketiga model teroptimasi yang telah dibangun. Penilaian kinerja dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score untuk menggambarkan kemampuan model dalam mengklasifikasikan nasabah yang menerima maupun menolak penawaran deposito berjangka. Selain itu, confusion matrix digunakan untuk menganalisis distribusi prediksi benar dan salah pada masing-masing kelas sehingga dapat diketahui pola kesalahan model, seperti false positive dan false negative. Hasil evaluasi dari setiap algoritma kemudian dibandingkan untuk menentukan model dengan performa terbaik yang direkomendasikan sebagai kandidat utama dalam mendukung pengambilan keputusan pada kampanye telemarketing deposito berjangka.

## **HASIL DAN PEMBAHASAN**

Bagian ini menyajikan hasil eksperimen yang dilakukan pada Bank Marketing Dataset menggunakan tiga algoritma klasifikasi, yaitu Decision Tree, Random Forest, dan XGBoost. Pembahasan meliputi hasil baseline modelling, pengaruh proses hyperparameter tuning melalui Grid Search Cross-Validation, serta komparasi kinerja akhir model teroptimasi berdasarkan metrik evaluasi yang telah ditetapkan.

### **1. Hasil Baseline Modelling**

Pada tahap baseline modelling, masing-masing algoritma dilatih menggunakan parameter standar dan dievaluasi pada data uji untuk memperoleh gambaran awal kinerja masing-masing algoritma. Hasilnya menunjukkan bahwa Decision Tree menghasilkan akurasi sebesar 86,5%. Meskipun akurasi ini tergolong cukup baik, performa model pada kelas minoritas (nasabah yang menerima penawaran, label 1) masih terbatas, dengan nilai precision sekitar 0,4272, recall 0,4518, dan F1-score 0,4391. Hal ini mengindikasikan bahwa Decision Tree cenderung lebih banyak membuat kesalahan dalam mengidentifikasi nasabah yang benar-benar berpotensi menerima penawaran deposito berjangka.

Random Forest pada baseline memberikan akurasi yang lebih tinggi, yaitu 90,67%. Model ini sangat baik dalam mengenali kelas mayoritas (label 0) dengan recall sebesar 0,9750 dan F1-score yang tinggi untuk kelas tersebut. Namun, performa pada kelas minoritas masih belum optimal, karena recall untuk label 1 hanya mencapai 0,3913 dengan F1-score 0,4952. Kondisi ini menunjukkan bahwa, meskipun akurasi keseluruhan tinggi, model Random Forest cenderung kurang sensitif dalam mendeteksi nasabah yang benar-benar menerima penawaran.

Model XGBoost pada baseline menghasilkan akurasi sebesar 90,58%, dengan kinerja yang relatif lebih seimbang untuk kelas minoritas dibanding Random Forest. Untuk label 1, XGBoost mencatat precision sekitar 0,6407, recall 0,4433, dan F1-score

0,5240. Nilai precision yang lebih tinggi ini menunjukkan bahwa prediksi positif yang dibuat XGBoost cenderung lebih tepat sasaran, meskipun masih terdapat ruang perbaikan terkait kemampuan model dalam menangkap seluruh nasabah yang sebenarnya menerima penawaran. Secara umum, hasil baseline memperlihatkan bahwa dua model ensemble (Random Forest dan XGBoost) unggul dari sisi akurasi dibandingkan Decision Tree, tetapi penanganan kelas minoritas masih menjadi tantangan pada ketiga model.

## 2. Hasil Hyperparameter Tuning

Setelah diperoleh gambaran awal kinerja model pada tahap baseline, dilakukan proses hyperparameter tuning menggunakan Grid Search Cross-Validation untuk masing-masing algoritma. Tujuan tahap ini adalah mencari kombinasi parameter yang mampu meningkatkan kinerja model sekaligus menjaga stabilitas hasil pada berbagai subset data. Untuk Decision Tree, Grid Search dilakukan terhadap parameter `max_depth`, `min_samples_split`, dan `min_samples_leaf`. Hasilnya menunjukkan bahwa konfigurasi terbaik diperoleh pada `max_depth = 5`, `min_samples_split = 2`, dan `min_samples_leaf = 1`, dengan akurasi rata-rata terbaik pada proses Grid Search sebesar 90,15%. Peningkatan akurasi ini menunjukkan bahwa pengaturan kedalaman pohon yang lebih terkontrol dapat mengurangi risiko overfitting yang biasanya muncul pada Decision Tree dengan kedalaman tak terbatas.

Pada Random Forest, kombinasi parameter yang diuji mencakup `n_estimators`, `max_depth`, `min_samples_split`, dan `min_samples_leaf`. Konfigurasi terbaik yang diperoleh adalah `n_estimators = 200`, `max_depth = None`, `min_samples_split = 5`, dan `min_samples_leaf = 2`, dengan akurasi rata-rata terbaik pada proses Grid Search sebesar 90,56%. Peningkatan jumlah pohon (`n_estimators`) serta pengaturan minimal sampel pada node internal dan node daun membantu model mencapai keseimbangan antara kompleksitas dan kemampuan generalisasi, sehingga meminimalkan risiko overfitting dan underfitting secara bersamaan. Untuk XGBoost, hyperparameter yang dioptimasi meliputi `n_estimators`, `max_depth`, `learning_rate`, `subsample`, dan `colsample_bytree`. Hasil Grid Search menunjukkan bahwa konfigurasi terbaik diperoleh pada `n_estimators = 200`, `max_depth = 7`, `learning_rate = 0,05`, `subsample = 0,8`, dan `colsample_bytree = 0,8`, dengan akurasi rata-rata terbaik pada proses Grid Search sebesar 90,77%. Kombinasi `learning_rate` yang relatif kecil dengan kedalaman pohon yang cukup besar dan penggunaan `subsampling` baik pada sampel maupun fitur memungkinkan XGBoost mempelajari pola kompleks secara bertahap, namun tetap menjaga kemampuan generalisasi terhadap data baru. Secara keseluruhan, proses hyperparameter tuning meningkatkan kinerja ketiga algoritma yang ditunjukkan oleh kenaikan akurasi rata-rata pada proses Grid Search dibandingkan tahap baseline.

### 3. Evaluasi Akhir dan Komparasi Model

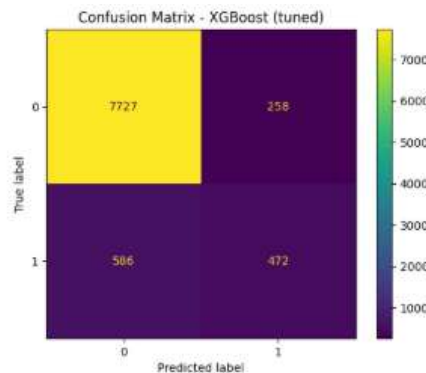
Untuk menilai kinerja ketiga model teroptimasi, evaluasi akhir dilakukan menggunakan data uji. Ringkasan hasil evaluasi berdasarkan metrik accuracy, precision, recall, dan F1-score ditampilkan pada Tabel 1.

**Tabel 1. Tabel Perbandingan Kinerja Model**

Algoritma	Akurasi	Precision	Recall	F1-Score
Decision Tree	89,95%	0.6530	0.3006	0.4117
Random Forest	90,45%	0.6633	0.3724	0.4770
XGBoost	90,67%	0.6466	0.4461	0.5280

Berdasarkan Tabel 1, Decision Tree menghasilkan akurasi sebesar 89,95% dengan precision 0,6530, recall 0,3006, dan F1-score 0,4117. Meskipun terjadi peningkatan dibandingkan model baseline, performa Decision Tree masih berada pada posisi terendah terutama karena recall kelas minoritas yang rendah, sehingga banyak kasus penerimaan deposito berjangka yang belum berhasil terdeteksi. Random Forest menunjukkan kinerja yang lebih baik dengan akurasi 90,45%, precision 0,6633, recall 0,3724, dan F1-score 0,4770, yang mengindikasikan peningkatan kemampuan model dalam mengenali nasabah yang menerima penawaran dibandingkan Decision Tree. XGBoost memberikan performa terbaik secara keseluruhan dengan akurasi 90,67% dan F1-score tertinggi sebesar 0,5280, serta recall tertinggi sebesar 0,4461. Hasil ini menunjukkan bahwa XGBoost lebih efektif dalam menangkap nasabah yang benar-benar berpotensi menerima deposito berjangka, sehingga lebih sesuai digunakan sebagai dasar lead scoring untuk memprioritaskan calon nasabah dengan probabilitas penerimaan yang lebih tinggi dalam kampanye telemarketing.

Selain metrik pada Tabel 1, analisis kesalahan klasifikasi juga dilakukan menggunakan confusion matrix untuk melihat distribusi prediksi benar dan salah pada masing-masing kelas. Confusion matrix untuk model dengan performa terbaik, yaitu XGBoost, ditunjukkan pada Gambar 2.



Gambar 2. Confusion Matrix XGBoost

Berdasarkan confusion matrix pada Gambar 2, model XGBoost menghasilkan True Positive (TP) sebesar 472 dan True Negative (TN) sebesar 7727, yang menunjukkan sebagian besar prediksi pada kelas mayoritas dapat diklasifikasikan dengan baik.

Namun, masih terdapat False Negative (FN) sebesar 586, yang berarti sejumlah nasabah yang sebenarnya menerima penawaran deposito berjangka diprediksi menolak sehingga berpotensi menyebabkan peluang konversi terlewat. Di sisi lain, False Positive (FP) sebesar 258 menunjukkan adanya prediksi menerima pada nasabah yang sebenarnya menolak, yang dapat meningkatkan jumlah panggilan telemarketing yang kurang efektif. Meskipun demikian, nilai TP yang paling tinggi dibandingkan model lain sejalan dengan recall dan F1-score terbaik, sehingga XGBoost tetap menjadi kandidat utama untuk implementasi.

## **KESIMPULAN**

Penelitian ini membahas komparasi beberapa model machine learning untuk memprediksi penerimaan produk deposito berjangka pada kampanye telemarketing bank menggunakan Bank Marketing Dataset. Permasalahan utama yang dihadapi adalah ketidakseimbangan kelas pada variabel target, di mana jumlah nasabah yang menolak penawaran lebih banyak dibandingkan yang menerima, sehingga diperlukan pendekatan berbasis data untuk meningkatkan ketepatan sasaran kampanye. Tahapan penelitian meliputi exploratory data analysis, preprocessing data melalui penanganan nilai “unknown”, one-hot encoding pada fitur kategorikal, transformasi label target menjadi kelas biner, serta pembagian data secara stratified. Tiga algoritma yang diuji adalah Decision Tree, Random Forest, dan XGBoost, yang kemudian dioptimasi menggunakan Grid Search untuk memperoleh konfigurasi hyperparameter terbaik. Berdasarkan hasil evaluasi pada data uji, model XGBoost teroptimasi memberikan performa terbaik dibandingkan Decision Tree dan Random Forest. Model ini menghasilkan akurasi sebesar 90,67%, serta menunjukkan keseimbangan kinerja yang lebih baik pada kelas minoritas dengan recall 0,4461 dan F1-score 0,5280. Analisis confusion matrix juga menunjukkan bahwa XGBoost memiliki jumlah prediksi benar pada kelas positif (TP) yang lebih tinggi, sehingga lebih efektif untuk mengidentifikasi calon nasabah yang berpotensi menerima penawaran deposito berjangka. Dengan demikian, model XGBoost teroptimasi berpotensi dimanfaatkan sebagai sistem pendukung keputusan atau lead scoring untuk memprioritaskan calon nasabah, meningkatkan rasio konversi, dan menekan biaya operasional kampanye telemarketing.

Untuk pengembangan penelitian selanjutnya, peningkatan performa terutama pada kelas minoritas dapat dilakukan melalui penerapan teknik penanganan imbalanced data (misalnya class weighting atau resampling), penyesuaian threshold keputusan, serta penambahan fitur yang lebih kaya agar model mampu menangkap pola penerimaan nasabah secara lebih akurat.

## DAFTAR PUSTAKA

Developers, S. (n.d.-a). *OneHotEncoder*. Scikit-Learn Documentation.

Developers, S. (n.d.-b). *train\_test\_split*. Scikit-Learn Documentation.

Maulidah, N. (2023). Prediksi Peningkatan Jumlah Nasabah Deposito Berjangka Menggunakan Algoritma KNN, Decision Tree, Random Forest dan XGBoost. *InComTech: Jurnal Telekomunikasi Dan Komputer*, 13(2).  
<https://doi.org/10.22441/incomtech.v13i2.16921>

Moro, S., Rita, P., & Cortez, P. (2014). Bank Marketing [Dataset]. *UCI Machine Learning Repository*. <https://doi.org/10.24432/C5K306>