

ANALISIS FAKTOR KEPUTUSAN NASABAH BERLANGGANAN TERM DEPOSIT: PERBANDINGAN RANDOM FOREST DAN XGBOOST

Rizky Wahyudi, Kristin Impana Manik, Muhammad Alfin,
John Bush Henrydunan, Arnita, Fanny Ramadhani

Ilmu Komputer, Universitas Negeri Medan

Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan,
Kabupaten Deli Serdang, Sumatera Utara, Indonesia
rizkyw23@mhs.unimed.ac.id

ABSTRAK

Pemasaran produk perbankan seperti term deposit memerlukan strategi yang tepat untuk meningkatkan tingkat langganan nasabah. Berbagai faktor yang memengaruhi keputusan nasabah menjadi tantangan bagi bank dalam menargetkan calon pelanggan potensial. Penelitian ini membandingkan performa Random Forest dan XGBoost dalam memprediksi keputusan nasabah menggunakan dataset Bank Marketing dari UCI Machine Learning Repository. Tahapan preprocessing mencakup pemeriksaan nilai hilang, penanganan outlier dengan IQR, transformasi fitur kategorikal menggunakan label encoding, dan oversampling dengan SMOTE. Evaluasi model dilakukan dengan accuracy, F1-Score, dan ROC-AUC. Hasil penelitian menunjukkan bahwa XGBoost lebih unggul dibandingkan Random Forest, dengan accuracy 95,30%, F1-score 95,40%, dan ROC-AUC 95,29% pada rasio 80:20. Komunikasi dan durasi interaksi dalam pemasaran produk term deposit menjadi faktor paling signifikan dalam memengaruhi keputusan nasabah. Kesimpulan umum dari penelitian ini adalah XGBoost menjadi algoritma yang lebih efektif untuk prediksi keputusan nasabah berlangganan term deposit daripada Random Forest terutama dalam menangani data tidak seimbang dan hubungan non-linear.

Kata kunci : term deposit, keputusan nasabah, Random Forest, XGBoost, ROC-AUC

1. PENDAHULUAN

Dalam beberapa tahun terakhir, industri perbankan menghadapi tantangan dalam mempertahankan dan menarik nasabah untuk berinvestasi dalam deposito berjangka. Salah satu fenomena yang mencolok adalah penurunan minat nasabah terhadap produk deposito berjangka di tengah maraknya instrumen investasi alternatif seperti reksa dana, obligasi, dan aset digital [1].

Di sisi lain, fluktuasi suku bunga yang dipengaruhi oleh kebijakan moneter serta ketidakpastian ekonomi global juga berdampak pada keputusan nasabah dalam memilih instrumen investasi yang aman dan menguntungkan. Di era digitalisasi ini, kemudahan akses informasi dan layanan keuangan berbasis teknologi semakin mempengaruhi pola keputusan nasabah dalam berlangganan deposito berjangka [2], [3].

Bank berperan penting dalam mengelola dana masyarakat, termasuk menawarkan deposito berjangka sebagai investasi aman dengan keuntungan lebih tinggi dari tabungan biasa. Namun perubahan ekonomi dan perkembangan layanan keuangan digital telah menggeser preferensi nasabah terhadap produk ini [4].

Oleh karena itu, memahami faktor-faktor yang mempengaruhi keputusan nasabah menjadi hal yang krusial bagi bank untuk meningkatkan efektivitas strategi pemasaran mereka.

Memahami faktor-faktor yang memengaruhi keputusan nasabah dalam berlangganan term deposit menjadi krusial, terutama dengan penerapan machine learning yang semakin luas di sektor keuangan. Machine learning telah terbukti meningkatkan akurasi

prediksi keputusan nasabah khususnya dengan XGBoost yang sering kali lebih unggul dalam memprediksi calon nasabah deposito [5].

Secara teknis, XGBoost lebih efektif dalam menangani data tidak seimbang, memiliki akurasi lebih tinggi, dan waktu pelatihan lebih cepat berkat teknik peningkatan gradiennya. Sementara Random Forest lebih stabil pada dataset besar dan lebih andal untuk dataset kecil meskipun berisiko overfitting jika tidak dikonfigurasi dengan baik [6].

Oleh karena itu, pemilihan algoritma yang tepat menjadi faktor penting bagi bank dalam mengoptimalkan strategi pemasaran deposito berjangka.

Pemilihan Random Forest dan XGBoost didasarkan pada keunggulan kedua algoritma dalam penelitian sebelumnya. Misalnya, penelitian yang dilakukan oleh [7] dalam studi “Perbandingan Algoritma SVM, Random Forest dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit” melaporkan bahwa XGBoost mencapai akurasi sekitar 82% dibandingkan dengan 71% pada Random Forest, selain itu penelitian yang dilakukan oleh [8] mengungkapkan hasil dari model terbaik yaitu random forest dan xgboost, dengan akurasi sebesar 91.7% dalam konteks prediksi term deposit. Temuan-temuan ini mendasari pemilihan kedua algoritma tersebut untuk dibandingkan, karena memberikan gambaran yang jelas mengenai keunggulan XGBoost dan Random Forest dalam menangani kompleksitas data dan ketidakseimbangan kelas dibandingkan dengan algoritma lain, sehingga relevan untuk diterapkan

dalam memprediksi keputusan nasabah berlangganan term deposit.

Bank dapat meningkatkan efektivitas kampanye pemasaran mereka dan memberikan penawaran yang lebih individual kepada konsumen dengan memanfaatkan teknologi pembelajaran mesin. Oleh karena itu, tujuan dari penelitian ini adalah untuk mengevaluasi dan membandingkan seberapa baik algoritma Random Forest dan XGBoost dalam memprediksi apakah nasabah akan memilih untuk membuka deposito berjangka atau tidak. Diharapkan bahwa temuan penelitian ini akan memberikan informasi penting bagi sektor perbankan untuk membuat rencana pemasaran yang lebih sukses dan menarik lebih banyak nasabah deposito jangka panjang.

Untuk menilai stabilitas kinerja model dalam situasi data yang berbeda, penelitian ini menerapkan dua skenario pembagian data, yaitu 80:20 dan 70:30, berbeda dengan penelitian sebelumnya yang hanya menggunakan rasio data latih dan data uji sebesar 80:20 saja. Selain itu, penelitian sebelumnya lebih banyak berfokus pada metrik evaluasi seperti accuracy, precision, dan F1-score, sedangkan penelitian ini menambahkan ROC-AUC sebagai salah satu metrik utama dalam mengukur performa model pada data yang tidak seimbang. Dengan pendekatan ini, penelitian ini diharapkan bisa memberikan kontribusi baru dalam pemahaman faktor keputusan nasabah dengan analisis yang lebih komprehensif serta evaluasi model yang lebih menyeluruh.

2. TINJAUAN PUSTAKA

2.1. Teori Keputusan Nasabah

Teori keputusan nasabah menjelaskan bagaimana konsumen, khususnya yang berada di industri perbankan syariah, mengambil keputusan tentang barang dan jasa. Ada beberapa langkah dalam proses ini, termasuk identifikasi masalah, pengumpulan informasi, penilaian alternatif, keputusan pembelian, dan perilaku pasca pembelian. [9], [10].

Pelanggan menimbang berbagai elemen yang dapat memengaruhi pilihan mereka pada setiap tahap. Elemen-elemen ini termasuk demografi yang dapat memengaruhi persyaratan dan preferensi pelanggan, seperti usia, pendidikan, dan status pekerjaan [11]. Selain itu, kondisi ekonomi, seperti pendapatan dan stabilitas finansial, juga berperan penting dalam menentukan keputusan nasabah untuk berlangganan produk keuangan. Dengan memahami teori ini, lembaga keuangan dapat merancang strategi pemasaran yang lebih efektif dan sesuai dengan kebutuhan nasabah.

Terdapat beberapa faktor individu dan makroekonomi yang mempengaruhi keputusan nasabah dalam membuka deposito berjangka. Faktor individu tersebut meliputi karakteristik demografi seperti jenis pekerjaan, status perkawinan, dan respon terhadap promosi [12].

2.2. Term Deposit

Deposito berjangka adalah deposito yang menurut ketentuan perjanjian deposan dengan bank hanya dapat ditarik pada waktu tertentu. Jangka waktu penyimpanan dapat diperpanjang secara otomatis sesuai dengan kontrak awal dan biasanya antara tiga, enam, dua belas, dan dua puluh empat bulan. Undang-undang Perbankan No. 10 tahun 1998 memberikan landasan hukum untuk deposito berjangka [13].

Simpanan pihak ketiga dalam bentuk rupiah dan mata uang asing dikenal sebagai deposito berjangka. Deposito ini diterbitkan atas nama nasabah bank dan hanya dapat dicairkan pada waktu tertentu yang disesuaikan dengan perjanjian antara nasabah dan bank [14]. Pinalti akan dikenakan jika deposito dicairkan sebelum jatuh tempo [15].

2.3. Random Forest

Random Forest adalah kumpulan dari pohon keputusan yang dibangun menggunakan sampel acak, namun memiliki aturan pemisahan simpul yang berbeda. Model ini beroperasi dengan memanfaatkan subset fitur pada setiap pohon dan mencoba menemukan ambang batas terbaik untuk memisahkan data. Akibatnya, terdapat banyak pohon yang dilatih dengan cara yang lebih lemah, dan setiap pohon menghasilkan prediksi yang berbeda. Interpretasi hasil tersebut dapat dilakukan dengan dua cara, yang paling umum adalah berdasarkan suara terbanyak, yang dianggap sebagai kelas yang benar [16].

Random Forest merupakan algoritma ensemble yang mengambil keputusan akhir melalui pemungutan suara mayoritas dari semua model pohon keputusan yang ada [17].

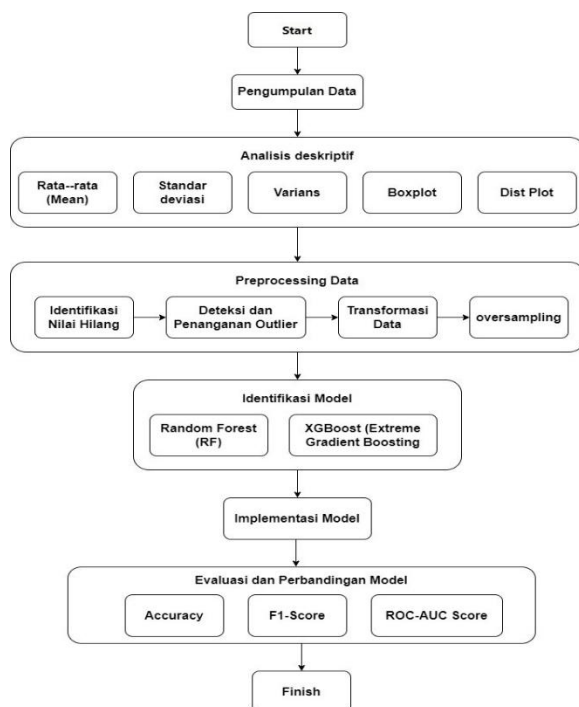
Secara umum, semakin banyak jumlah pohon dalam Random Forest, semakin stabil dan akurat prediksinya. Namun, jika jumlah pohon terlalu sedikit, model memiliki varians tinggi dan cenderung memberikan hasil yang kurang akurat [18].

2.4. XGBoost

XGBoost (Extreme Gradient Boosting) merupakan evolusi lanjutan dari algoritma Gradient Tree Boosting berbasis ensemble, yang mampu menangani masalah machine learning dalam skala besar secara efektif. XGBoost menawarkan keunggulan dalam hal performa, efisiensi waktu, dan penggunaan memori, sehingga telah diaplikasikan di berbagai bidang penelitian, mulai dari kesehatan, penilaian risiko kredit, hingga metagenomic [19].

Algoritma ini mengoptimalkan proses boosting dengan teknik regulasi dan peningkatan efisiensi komputasi. XGBoost awalnya dikembangkan oleh Tianqi Chen sebagai proyek penelitian, tetapi segera menjadi standar industri dalam kompetisi data science seperti Kaggle karena kemampuannya dalam menangani masalah data tabular dengan efisiensi tinggi [20].

3. METODE PENELITIAN



Gambar 1. Diagram alir penelitian

Penelitian ini membandingkan performa algoritma Random Forest dan XGBoost dalam memprediksi keputusan nasabah terhadap term deposit menggunakan dataset Bank Marketing dari UCI Machine Learning Repository. Diagram alir pada Gambar 1 memberikan gambaran umum tentang keseluruhan alur kerja penelitian.

3.1. Pengumpulan Data

Penelitian ini menggunakan dataset Bank Marketing yaitu "bank.csv" dari UCI Machine Learning Repository yang mencakup 4521 data nasabah. Dataset yang digunakan ini adalah dataset perbankan yang berisi informasi mengenai keputusan nasabah untuk berlangganan term deposit. Karakteristik dataset ini adalah multivariat dengan tipe data kategorikal dan integer (numerik). Dataset ini terdiri dari 16 fitur dan 1 kelas target sebagaimana dijelaskan pada Tabel 1 berikut.

Tabel 1. Deskripsi fitur dataset

No	Fitur	Jenis data	Skala data	Keterangan
1	age	Numerik	Rasio	Usia nasabah
2	job	Kategorik	Nominal	Pekerjaan nasabah
3	marital	Kategorik	Nominal	Status pernikahan
4	education	Kategorik	Ordinal	Tingkat pendidikan (SD < SMP < SMA < Universitas)
5	balance	Numerik	Rasio	Saldo rekening (Euro)
6	housing	Kategorik	Nominal	Status kepemilikan rumah
7	loan	Kategorik	Nominal	Status pinjaman pribadi
8	contact	Kategorik	Nominal	Jenis komunikasi terakhir
9	day	Numerik	Ordinal	Hari terakhir kontak dalam sebulan (1-31)
10	month	Kategorik	Nominal	Bulan terakhir kontak
11	duration	Numerik	Rasio	Durasi terakhir kontak (detik)
12	campaign	Numerik	Rasio	Jumlah kontak selama kampanye ini
13	pdays	Numerik	Rasio	Hari sejak terakhir dihubungi (-1 berarti belum pernah)
14	previous	Numerik	Rasio	Jumlah kontak sebelum kampanye ini
15	poutcome	Kategorik	Nominal	Hasil kampanye pemasaran sebelumnya
16	default	Kategorik	Nominal	Status kredit macet
17	y	Kategorik	Biner	Keputusan nasabah untuk berlangganan term deposit

3.2. Analisis Deskriptif

Analisis deskriptif dilakukan untuk mendapatkan pemahaman tentang karakteristik dataset. Ini termasuk mengevaluasi apakah skala data tidak seragam. Analisis ini menggunakan parameter rata-rata (mean), standar deviasi, varians, dan boxplot untuk melihat distribusi data dan menemukan pencilan (outlier). Selain itu, dist plot digunakan untuk melihat distribusi distribusi data.

3.3. Preprocessing Data

Tahapan preprocessing meliputi beberapa langkah berikut. Dataset diperiksa untuk mengetahui apakah terdapat nilai yang hilang. Selanjutnya, outlier

diidentifikasi menggunakan boxplot dan ditangani dengan metode Interquartile Range (IQR) untuk menghapus pencilan yang ekstrem. Pada tahap transformasi data, fitur kategorikal dikonversi menjadi numerik menggunakan teknik label encoding agar model dapat memahami tipe data dalam numerik. Terakhir, untuk mengatasi perbedaan skala, dilakukan oversampling menggunakan SMOTE.

3.4. Identifikasi Model

Model yang digunakan dalam penelitian ini adalah algoritma Random Forest (RF) dan XGBoost (XGB). Biasanya, Random Forest digunakan untuk menangani klasifikasi, regresi, dan masalah terkait

lainnya. Pendekatan ini menggabungkan banyak pohon keputusan, yang masing-masing bergantung pada nilai vektor acak. Nilai vektor acak menentukan nilai setiap pohon. Berikut ini adalah rumus persamaan Random Forest dengan N trees:

$$ly = \operatorname{argmax}_c \left(\sum_{n=1}^n l_{n_n(y)=c} \right) \quad (1)$$

Metode XGBoost adalah metode pengembangan dari algoritma GBDT (Gradient Boosting Decision Tree). Salah satu pustaka machine learning yang dapat digunakan untuk klasifikasi atau prediksi berdasarkan pohon keputusan disebut XGBoost. Dibandingkan dengan GBM yang ada, teknik ini memungkinkan pengoptimalan sepuluh kali lebih cepat. Parameter yang digunakan menentukan seberapa akurat temuan klasifikasi. Algoritma XGBoost dan Random Forest terdiri dari banyak pohon keputusan. Berbeda dengan random forest, yang menggunakan bagging, XGBoost membangun algoritma dengan menggunakan teknik boosting.

3.5. Implementasi Model

Random Forest digunakan sebagai model ensemble berbasis pohon keputusan yang membangun beberapa decision trees dari subset data acak dan menggabungkan hasilnya untuk meningkatkan akurasi. Model ini dilatih menggunakan parameter $n_estimators$ sebanyak 200 pohon, max_depth sebesar 20, dan $min_samples_split$ sebesar 5 untuk mengontrol kompleksitas pohon dan mengurangi overfitting. Model dilatih menggunakan dua skenario pembagian data, yaitu 70:30 dan 80:20, setelah melalui tahap preprocessing yang mencakup penanganan outlier dengan winsorizing, normalisasi menggunakan MinMaxScaler, serta oversampling menggunakan SMOTE untuk menangani ketidakseimbangan kelas.

XGBoost diterapkan sebagai model boosting berbasis pohon keputusan yang memperbaiki kesalahan prediksi melalui proses iteratif. Model ini dikonfigurasi dengan $n_estimators$ sebanyak 200, max_depth sebesar 10, dan $learning_rate$ sebesar 0.1 untuk mengontrol laju pembelajaran agar mencapai konvergensi optimal. Seperti halnya Random Forest, XGBoost juga diuji dengan pembagian data 70:30 dan 80:20 untuk mengevaluasi kestabilannya.

Implementasi kedua model dilakukan menggunakan pustaka scikit-learn dan XGBoost dalam Python. Setelah model dilatih, dilakukan evaluasi menggunakan metrik accuracy, F1-score, dan ROC AUC. Selain itu, analisis feature importance dilakukan untuk mengidentifikasi variabel yang paling berpengaruh dalam menentukan keputusan nasabah..

3.6. Evaluasi dan Perbandingan Model

Evaluasi model dilakukan menggunakan beberapa metrik utama yaitu akurasi (Accuracy), F1-Score, dan ROC-AUC Score. Hasil evaluasi dari Random Forest dan XGBoost dibandingkan berdasarkan nilai metrik tersebut untuk menentukan

metode yang lebih unggul. Perbandingan dilakukan dengan menganalisis perbedaan skor pada setiap metrik guna memperoleh model dengan performa terbaik untuk dataset ini. Evaluasi model dilakukan menggunakan metrik berikut:

$$\text{Akurasi} = \frac{TN + TP}{TN + TP + FP + FN} \quad (2)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (3)$$

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

$$\text{F1 Score} = 2 \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (5)$$

Keterangan:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

ROC-AUC dihitung berdasarkan area di bawah kurva ROC, yang menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR).

$$\text{AUC} = \frac{1}{2} \sum_{i=0}^n (x_{i+1} - x_i)(y_{i+1} - y_i) \quad (6)$$

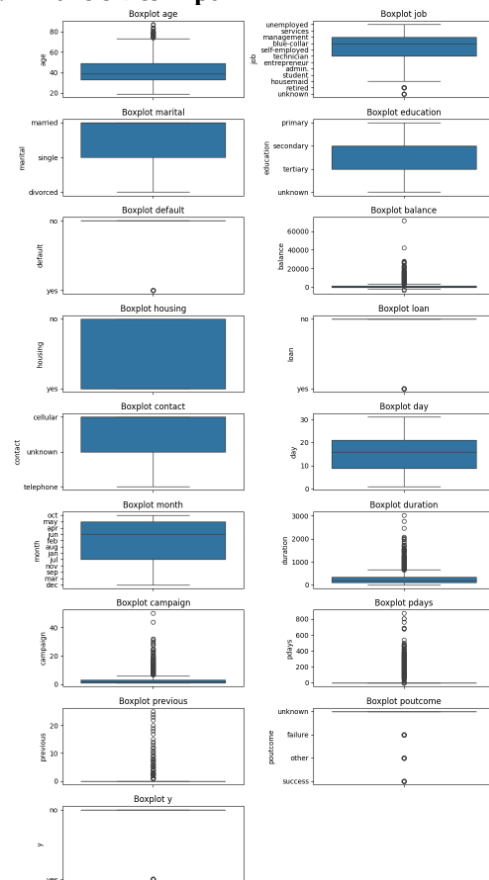
Dimana:

x_i dan y_i adalah titik-titik pada kurva ROC.

x_i adalah False Positive Rate (FPR) dan y_i adalah True Positive Rate (TPR) pada threshold tertentu.

4. HASIL DAN PEMBAHASAN

4.1. Analisis Deskriptif



Gambar 2. Boxplot sebelum preprocessing

Berdasarkan Gambar 2, boxplot dari fitur-fitur dalam dataset perbankan menunjukkan distribusi data dan pencilan signifikan pada fitur numerik seperti "age", "balance", "duration", "campaign", "pdays", dan "previous". Fitur "duration" memiliki variasi besar, sedangkan nilai -1 pada "pdays" mengindikasikan nasabah yang belum dihubungi. Fitur kategorikal seperti "job", "marital", "education", "default", "housing", "loan", "contact", "month", dan "poutcome" merepresentasikan distribusi frekuensi kategori yang beragam. Informasi ini penting untuk preprocessing data, seperti penanganan pencilan dan normalisasi, demi memastikan algoritma machine learning seperti Random Forest atau XGBoost bekerja optimal pada analisis prediktif.

age	int64
job	object
marital	object
education	object
default	object
balance	int64
housing	object
loan	object
contact	object
day	int64
month	object
duration	int64
campaign	int64
pdays	int64
previous	int64
poutcome	object
y	object
dtype:	object

Gambar 3. Tipe data sebelum preprocessing

Berdasarkan Gambar 3, masih ada terdapat tipe data kategorikal (object). Untuk itu diperlukan tahapan transformasi data kategorik menjadi numerik yang dalam hal ini peneliti menggunakan metode label encoding.

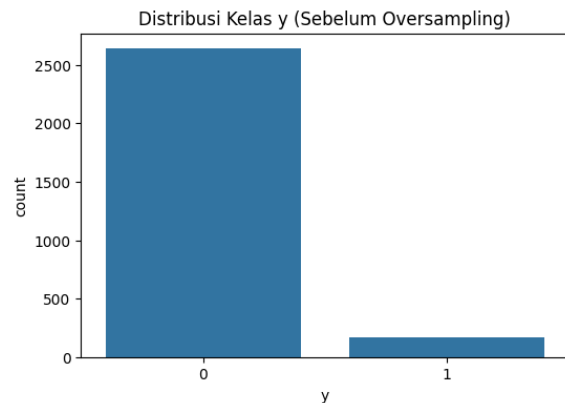
	age	balance	day	duration	campaign
count	4521.000000	4521.000000	4521.000000	4521.000000	4521.000000
mean	41.170095	1422.657819	15.915284	263.961292	2.793630
std	10.576211	3009.638142	8.247667	259.856633	3.109807
min	19.000000	-3313.000000	1.000000	4.000000	1.000000
25%	33.000000	69.000000	9.000000	104.000000	1.000000
50%	39.000000	444.000000	16.000000	185.000000	2.000000
75%	49.000000	1480.000000	21.000000	329.000000	3.000000
max	87.000000	71188.000000	31.000000	3025.000000	50.000000

	pdays	previous
count	4521.000000	4521.000000
mean	39.766645	0.542579
std	100.121124	1.693562
min	-1.000000	0.000000
25%	-1.000000	0.000000
50%	-1.000000	0.000000
75%	-1.000000	0.000000
max	871.000000	25.000000

Gambar 4. Analisis deskriptif sebelum preprocessing data

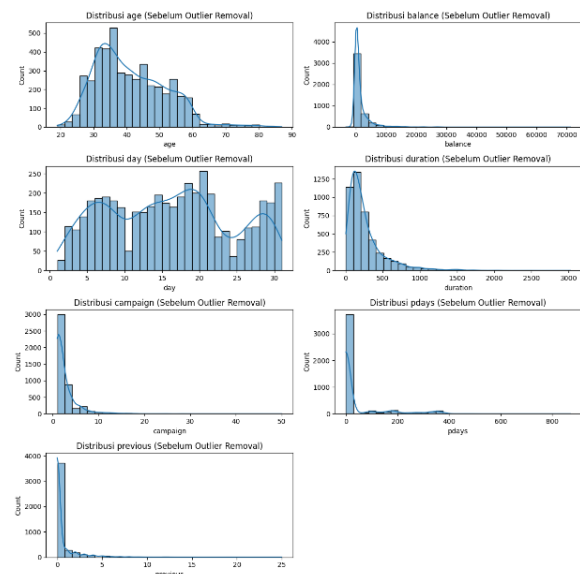
Berdasarkan Gambar 4, analisis deskriptif dataset menunjukkan bahwa fitur "age" memiliki rata-rata 41.17 tahun (std = 10.58), mengindikasikan konsentrasi usia di sekitar dewasa menengah; fitur "balance" memiliki rata-rata saldo 1.422 (std = 3.009), mengindikasikan variasi saldo yang luas; fitur "duration" memiliki rata-rata durasi percakapan 263.96 detik (std = 259.95), menunjukkan variabilitas

tinggi; fitur "campaign" memiliki rata-rata jumlah kampanye 2.79 (std = 3.10), menunjukkan sebagian besar nasabah dihubungi beberapa kali; dan fitur "pdays" memiliki rata-rata -1 (std = 100.12), menunjukkan mayoritas nasabah belum dihubungi sebelumnya. Perbedaan rata-rata dan standar deviasi ini menunjukkan adanya pencilan dan variabilitas signifikan, yang memerlukan normalisasi dan penanganan pencilan pada tahap *preprocessing* data.



Gambar 5. Count plot sebelum oversampling

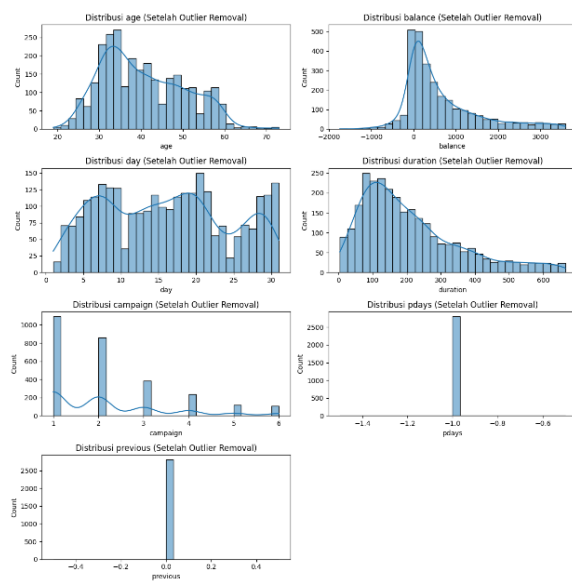
Berdasarkan Gambar 5, terjadi ketidakseimbangan pada distribusi kelas variabel target. Untuk memperbaiki ketidakseimbangan tersebut, metode oversampling dengan SMOTE akan digunakan selanjutnya..



Gambar 6. Distribusi data sebelum ICR

Hasil visualisasi distribusi data yang ditunjukkan pada Gambar 6 di atas menunjukkan indikasi skewness dan outlier. Oleh karena itu, kita akan menghilangkan outlier seperti yang disebutkan sebelumnya dan mengubah data jika hasil penghapusan outlier dan model kurang baik.

4.2. Preprocessing Data



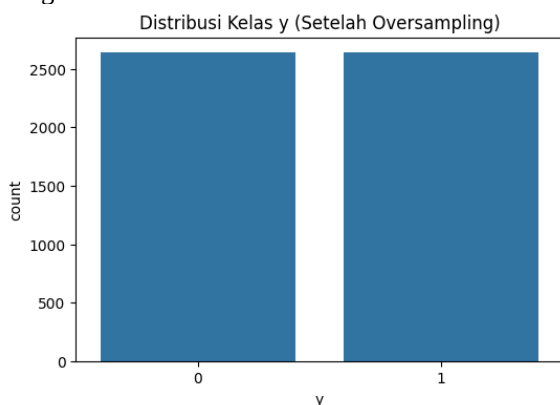
Gambar 7. Distribusi data setelah ICR

Setelah menghilangkan outlier, distribusi data telah mendekati normal dan outlier telah berkurang secara signifikan, seperti yang ditunjukkan dalam hasil visualisasi distribusi data pada Gambar 7 di atas.

```
age          int64
job          int64
marital      int64
education    int64
default      int64
balance      int64
housing      int64
loan         int64
contact      int64
day          int64
month        int64
duration     int64
campaign     int64
pdays       int64
previous     int64
poutcome     int64
y            int64
dtype: object
```

Gambar 8. Tipe data fitur setelah preprocessing

Berdasarkan Gambar 8 terlihat tipe data sudah ditransformasi dari data kategorik menjadi numerik. Dengan transformasi ini, model machine learning dapat lebih mudah memproses data karena bekerja lebih optimal dengan nilai numerik dibandingkan kategori teks.



Gambar 9. Distribusi data setelah ICR

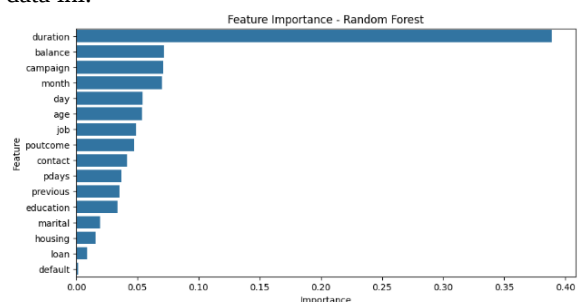
Dalam hasil visualisasi count plot setelah dilakukannya oversampling yang disajikan seperti pada Gambar 9 di atas, dapat terlihat bahwa kelas 0 dan 1 telah seimbang. Oleh karena itu, peneliti akan melanjutkan pada tahap selanjutnya yaitu evaluasi dan perbandingan model.

4.3. Evaluasi dan Perbandingan Model

Tabel 2. Hasil evaluasi model

Rasio Data	Model	Accuracy	F1-Score	ROC AUC
70:30	Random Forest	0.9302	0.9320	0.9301
70:30	XGBoost	0.9503	0.9509	0.9502
80:20	Random Forest	0.9379	0.9402	0.9376
80:20	XGBoost	0.9530	0.9540	0.9529

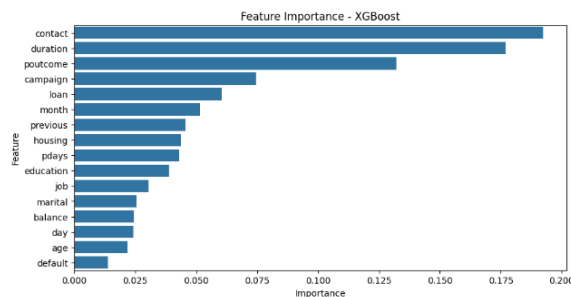
Berdasarkan Tabel 2, Hasil evaluasi menunjukkan bahwa XGBoost secara konsisten mengungguli Random Forest dalam semua metrik, baik pada rasio data 70:30 maupun 80:20. Nilai evaluasi model tertinggi didapat oleh XGBoost dengan rasio data 80:20 yang menghasilkan accuracy 95,30%, F1-Score 95,40% dan ROC AUC 95,29%. Hal ini disebabkan oleh kemampuan XGBoost dalam menangani hubungan non-linear yang kompleks serta keunggulannya dalam menangani data imbalance melalui boosting iteratif. Namun, jika diperhatikan lebih dalam, peningkatan rasio data training dari 70:30 ke 80:20 berdampak lebih signifikan pada Random Forest dibandingkan XGBoost, menunjukkan bahwa Random Forest lebih sensitif terhadap jumlah data pelatihan yang lebih besar. Meskipun demikian, XGBoost tetap lebih unggul secara keseluruhan, karena model ini mampu mempelajari pola dengan lebih baik dan mengurangi kemungkinan overfitting dibandingkan Random Forest. Dengan skor evaluasi yang lebih tinggi, XGBoost menjadi pilihan yang lebih unggul untuk prediksi keputusan nasabah berlangganan term deposit dalam skenario pembagian data ini.



Gambar 10. Feature importance random forest

Berdasarkan Gambar 10, fitur “duration” memiliki pengaruh terbesar dalam keputusan nasabah berlangganan term deposit, menekankan pentingnya durasi panggilan dalam pemasaran. Fitur “balance” berada di posisi kedua, menunjukkan bahwa kondisi keuangan nasabah berperan penting, diikuti oleh

“campaign”, yang mencerminkan dampak jumlah kontak pemasaran. Fitur seperti “day”, “age”, dan “job” memiliki tingkat kepentingan sedang, tetap relevan meskipun tidak sebesar tiga fitur utama. Sementara itu, “default”, “loan”, dan “housing” memiliki pengaruh paling kecil, menunjukkan bahwa faktor pinjaman dan kepemilikan rumah kurang menentukan. Secara keseluruhan, Random Forest lebih memprioritaskan variabel terkait interaksi langsung dan kondisi finansial dibandingkan faktor demografi dalam memprediksi keputusan nasabah.



Gambar 11. Feature importance XGBoost

Berdasarkan grafik pada Gambar 11, model XGBoost menempatkan “contact” sebagai fitur paling berpengaruh, menunjukkan bahwa jenis kontak yang digunakan sangat menentukan keputusan nasabah. Fitur “duration” tetap penting di posisi kedua, menandakan bahwa lama komunikasi juga berperan besar. Fitur “poutcome” dan “campaign” memiliki dampak signifikan, mengindikasikan bahwa hasil kampanye sebelumnya serta frekuensi kontak pemasaran berkontribusi dalam prediksi. Sementara itu, fitur seperti “loan”, “month”, dan “previous” memiliki pengaruh sedang, sedangkan “default”, “age”, dan “day” memiliki dampak paling kecil. Berbeda dengan Random Forest yang lebih menitikberatkan pada durasi panggilan dan saldo nasabah, XGBoost lebih fokus pada jenis kontak dan riwayat kampanye, menandakan kemampuannya dalam menangkap pola pemasaran yang lebih kompleks.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan yang sudah dilakukan dapat disimpulkan bahwa algoritma XGBoost memiliki performa yang lebih unggul dibandingkan Random Forest dalam memprediksi keputusan nasabah untuk berlangganan term deposit yang menghasilkan accuracy 95,30%, F1-Score 95,40% dan ROC AUC 95,29% pada rasio 80:20. Fitur contact dan duration paling berpengaruh dalam keputusan nasabah, menekankan pentingnya metode komunikasi dan durasi interaksi. Selain itu, poutcome dan campaign juga berperan, menunjukkan bahwa riwayat keberhasilan kampanye serta frekuensi kontak memengaruhi keputusan. Penelitian selanjutnya dapat mengeksplorasi algoritma lain seperti LightGBM atau CatBoost serta menerapkan ensemble learning untuk meningkatkan akurasi. Selain itu, analisis dapat

diperluas dengan mempertimbangkan perilaku digital nasabah dan faktor eksternal seperti kondisi makroekonomi guna memperkaya strategi pemasaran berbasis data. Dengan memperbaiki kekurangan ini, diharapkan penelitian mendatang dapat memberikan solusi pemasaran yang lebih efektif dan relevan bagi industri perbankan.

DAFTAR PUSTAKA

- [1] D. A. Nianty and A. Marlinah, “Pengaruh Tingkat Suku Bunga Deposito terhadap Jumlah Dana Deposito,” *J. Manage.*, vol. 5, no. 3, pp. 553–566, 2022, doi: <https://doi.org/10.37531/yum.v5i3.3389>.
- [2] A. Laelasari, M. I. Fasa, and I. Susanto, “Peran Digitalisasi Dalam Strategi Pengambilan Keputusan Menjadi Nasabah Bank Syariah Indonesia,” *JICN J. Intelek dan Cendekiawan Nusantara*, vol. 1, no. 5, pp. 7188–7201, 2024, [Online]. Available: <https://jicnusantara.com/index.php/jicn/article/download/1242/1490>
- [3] R. R. Siswoyo, Supaino, and R. Agus, “Pengaruh Digitalisasi, Motivasi dan Literasi Keuangan Syariah Terhadap Keputusan Berinvestasi Tabungan Emas di Pegadaian,” *J. Keuang. Dan Perbank. Syariah*, vol. 2, no. 1, pp. 653–662, 2022, doi: <https://doi.org/10.51510/konsep.v3i1.806>.
- [4] N. Maulidah, “Prediksi Peningkatan Jumlah Nasabah Deposito Berjangka Menggunakan Algoritma KNN, Decision Tree, Random Forest Dan Xgboost,” *InComTech J. Telekomun. dan Komput.*, vol. 13, no. 2, p. 90, 2023, doi: 10.22441/incomtech.v13i2.16921.
- [5] A. N. Puteri, A. Arizal, and A. D. Achmad, “Feature Selection Correlation-Based pada Prediksi Nasabah Bank Telemarketing untuk Deposito,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 335–342, May 2021, doi: 10.30812/matrik.v20i2.1183.
- [6] J. M. A. S. Dachi and P. Sitompul, “Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit,” *J. Ris. Rumpun Mat. Dan Ilmu Pengetah. Alam*, vol. 2, no. 2, pp. 87–103, 2023, doi: 10.55606/jurrimipa.v2i2.1470.
- [7] M. R. Givari, M. R. Sulaeman, and Y. Umaidah, “Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit,” *Nuansa Inform.*, vol. 16, no. 1, pp. 141–149, 2022, doi: 10.25134/nuansa.v16i1.5406.
- [8] R. A. R. Muhammad, W. N. Aghniya, and H. Tantyoko, “Penerapan Algoritma Machine Learning Untuk Memprediksi Term Deposit Nasabah Perbankan,” *LEDGER J. Inform. Inf. Technol.*, vol. 2, no. 3, pp. 145–156, 2023,

- [Online]. Available: <https://journal.ittelkom-pwt.ac.id/index.php/ledger/article/view/1416>
- [9] B. D. Putra and K. Hermanto, "Analisis Faktor-Faktor Yang Mempengaruhi Keputusan Nasabah Dalam Memilih Jasa Perbankan Syariah (Studi Kasus PT. Bank NTB Syariah KCP Lunyuk)," *J. Innov. Res. Knowl.*, vol. 1, no. 9, pp. 1003–1014, 2022, doi: <https://doi.org/10.53625/jirk.v1i9.1436>.
- [10] S. Sumantri, D. Diyah, and A. Albetris, "Analisa Faktor-Faktor yang Mempengaruhi Keputusan Nasabah Menggunakan Produk Perbankan (Studi Kasus Nasabah Bank Syariah Indonesia Kantor Cabang Pembantu Muara Bulian)," *J. Ilm. Univ. Batanghari Jambi*, vol. 21, no. 2, p. 707, 2021, doi: [10.33087/jiubj.v21i2.1533](https://doi.org/10.33087/jiubj.v21i2.1533).
- [11] M. Nazori, M. Orinaldi, and R. Saputra, "ANALISIS KEPUTUSAN NASABAH DALAM PEMBIAYAAN SYARI'AH PADA PEDAGANG ANGSO DUO KOTA JAMBI," *EKSYA J. Ekon. Syariah*, vol. 4, no. 2, pp. 125–135, 2023, doi: <https://doi.org/10.56874/eksya.v4i2.1229>.
- [12] I. Indrastata, "Strategi Promosi Perbankan Dengan Analisis Dan Implementasi Algoritma Random Forest," *AnoaTIK J. Teknol. Inf. dan Komput.*, vol. 1, no. 2, pp. 54–58, 2023, doi: [10.33772/anoatik.v1i2.8](https://doi.org/10.33772/anoatik.v1i2.8).
- [13] R. Lestari, I. Fadhillah, A. Nurasikin, and T. Handayani, "Tinjauan Hukum Islam: Akad Mudharabah dalam Simpanan Ijabah di KSPPS Tamzis Bina Utama Cabang Purbalingga," *J. Huk. DAN Din. Masy.*, vol. 22, no. 2, pp. 237–252, 2024, doi: [10.56444/hdm.v21i1](https://doi.org/10.56444/hdm.v21i1).
- [14] K. Husnaini and A. Rohman, "Implementasi Akad Mudharabah Dalam Produk Simpanan Tabungan Berjangka Sejahtera Pada Koperasi Bina Syariah Ummah Cabang Dukun Gresik," *Jesya*, vol. 6, no. 2, pp. 2190–2200, 2023, doi: [10.36778/jesya.v6i2.1211](https://doi.org/10.36778/jesya.v6i2.1211).
- [15] R. A. U. Rohaedi, "Tanggung Jawab Bank terhadap Simpanan Deposito Berjangka yang Tidak Tercatat dihubungkan dengan Perlindungan Hukum Nasabah menurut Undang-Undang Nomor 10 Tahun 1998 tentang Perbankan," *J. Ris. Ilmu Huk.*, vol. 1, no. 1, pp. 44–51, 2021, doi: [10.29313/jrih.v1i1.179](https://doi.org/10.29313/jrih.v1i1.179).
- [16] D. P. Sinambela, H. Naparin, M. Zulfadhilah, and N. Hidayah, "Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin," *J. Inf. dan Teknol.*, vol. 5, no. 3, pp. 58–64, 2023, doi: [10.60083/jidt.v5i3.393](https://doi.org/10.60083/jidt.v5i3.393).
- [17] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562>
- [18] B. Giuseppe, *Machine Learning Algorithms*, Third. Birmingham: Packt Publishing Ltd., 2017.
- [19] A. A. Saputra, B. N. Sari, and C. Rozikin, "Penerapan Algoritma Extreme Gradient Boosting (Xgboost) Untuk Analisis Risiko Kredit," *J. Ilm. Wahana Pendidik.*, vol. 10, no. 7, pp. 14–24, 2024, doi: <https://doi.org/10.5281/zenodo.10960080>.
- [20] M. Harrison, *Effective XGBoost: Tuning, Understanding, and Deploying Classification Models*. MetaSnake, 2023.