

Penerapan Algoritma Machine Learning Untuk Memprediksi Term Deposit Nasabah Perbankan

Muhammad Rasikh Azfa Riyyasy^{1*}, Wahyu Nouval Aghniya², Henri Tantyoko³
^{1,2,3}Fakultas Informatika, Institut Teknologi Telkom Purwokerto, Indonesia

*Corresponding Author: 21110005@ittelkom-pwt.ac.id

Abstract

The banking sector holds a pivotal role in influencing a country's economic conditions, driving continuous innovation in its product offerings. Among these offerings, deposits stand out as a fundamental financial instrument. Deposits represent savings with withdrawal restrictions based on agreements between depositors and banks. As time progresses, banks introduce various forms of investments as new products, prompting shifts in customer preferences towards these alternatives. This research aims to elucidate the term deposit preferences of banking customers. Utilizing Random Forest, Logistic Regression, SVC, and XGBoost methods, the study seeks to uncover patterns in customer behavior related to deposit terms. The primary goal is to enhance our understanding of these preferences, aiding banks in tailoring their products accordingly. The research findings reveal that the random forest and xgboost models emerge as the most effective, achieving a commendable accuracy of 91.7%, demonstrating their reliability without overfitting. The objective is not only to identify deposit term preferences but also to provide actionable insights for banks to refine their strategies, making them more customer-centric and adaptive in a dynamic market.

Keywords: Customers, Deposits, Random Forest, Logistic Regression, SVC, XGBoost

Abstrak

Sektor perbankan memiliki pengaruh yang sangat signifikan terhadap kondisi ekonomi suatu negara, mendorong terus dilakukannya inovasi dalam produk-produknya. Salah satu produk yang menjadi fokus adalah deposito, sebagai instrumen keuangan yang mendasar. Deposito adalah bentuk simpanan dengan pembatasan penarikan berdasarkan kesepakatan antara nasabah penyimpan dan bank. Seiring berjalannya waktu, bank memperkenalkan berbagai bentuk investasi baru, mempengaruhi pergeseran preferensi nasabah terhadap produk-produk tersebut. Penelitian ini bertujuan untuk menggali preferensi nasabah terkait term deposit. Dengan menggunakan metode Random Forest, Logistic Regression, SVC, dan XGBoost, penelitian ini mencoba mengidentifikasi pola perilaku nasabah terkait dengan pilihan term deposit. Tujuan utamanya adalah meningkatkan pemahaman tentang preferensi tersebut, membantu bank dalam menyesuaikan produk mereka secara efektif. Hasil penelitian menunjukkan bahwa model random forest dan xgboost menjadi yang paling efektif, dengan akurasi mencapai 91,7%, tanpa menunjukkan overfitting. Tujuan penelitian bukan hanya untuk mengidentifikasi preferensi term deposit, tetapi juga memberikan wawasan yang dapat diimplementasikan bagi bank untuk menyempurnakan strategi mereka, menjadikannya lebih berorientasi pada pelanggan dan adaptif dalam pasar yang dinamis.

Kata Kunci: Nasabah, Deposito, Random Forest, Logistic Regression, SVC, XGBoost

I. INTRODUCTION

Seiring dengan perkembangan ekonomi yang pesat, kebutuhan finansial manusia juga meningkat. Sektor perbankan mempunyai peranan yang sangat strategis dalam menggerakkan roda perekonomian suatu negara sehingga perbankan harus selalu terus berinovasi dengan memperbaiki kualitas dan kuantitas nya [1]. Untuk memenuhi kebutuhan tersebut, manusia memerlukan kemudahan dan keamanan, dan inilah yang ditawarkan oleh bank melalui produk depositonya. Bank mampu membantu individu yang memenuhi kriteria tertentu untuk melakukan deposit, yang mana bank akan memberikan bunga atas deposit tersebut. Bisnis deposit ini memberikan keuntungan bagi bank melalui bunga yang dikenakan kepada nasabah.

Deposito menurut Undang-Undang Nomor 10 Tahun 1998 adalah simpanan yang penarikannya hanya dapat dilakukan pada waktu tertentu berdasarkan perjanjian nasabah penyimpan dengan bank. Deposito merupakan salah satu tempat bagi nasabah untuk melakukan investasi dalam bentuk surat-surat berharga yang kedepannya akan diberikan imbalan bunga dan bunga yang diberikan merupakan bunga yang tertinggi [2].

Hampir semua bank menawarkan layanan deposit kepada nasabah yang memenuhi kriteria yang telah ditetapkan. Namun, hal ini menciptakan persaingan antar bank. Persaingan ini mendorong bank untuk melakukan analisis terhadap nasabah yang berpotensi untuk menarik deposit mereka dan beralih ke bank lain. Keberhasilan bisnis bank sangat bergantung pada sejauh mana bank tersebut mampu memberikan pelayanan terbaik kepada setiap nasabahnya.

Nasabah merupakan pelanggan (customer) baik individu maupun perusahaan yang mendapatkan manfaat atau kegunaan dari produk dan jasa dari sebuah perusahaan perbankan, yang meliputi kegiatan pembelian, penyewaan, serta layanan jasa [7]. Nasabah perbankan merupakan seorang pelanggan atau pemakai jasa suatu produk dalam industri perbankan. Produk bisa berupa deposit dari investasi yang dipilih oleh setiap nasabah. Suatu bank juga harus berusaha memberikan pelayanan terbaiknya untuk mendapat kepercayaan dan loyalitas dari setiap nasabah guna mencapai tujuan atau target dari bank tersebut. Untuk kepentingan nasabah, bank wajib menyediakan informasi mengenai kemungkinan timbulnya resiko kerugian sehubungan dengan transaksi nasabah yang dilakukan melalui bank [8]. Jadi, nasabah perbankan merupakan sumber pendapatan dari suatu bank karena keuntungan bank didapat dari banyaknya jumlah nasabah dan transaksinya.

Pada saat ini, para nasabah perbankan menyimpan dana dalam bentuk deposito pada bank sangat berbeda dengan yang sebelum-sebelumnya. Jika sebelumnya para nasabah menjadi pihak yang sering terabaikan, sekarang ini para nasabah menjadi penentu kelangsungan hidup sebuah perusahaan tersebut [3]. Bahkan pada beberapa kasus, perusahaan terpaksa tutup dan bangkrut karena faktor kurangnya nasabah yang menguntungkan. Kehilangan nasabah menjadi masalah serius bagi bank, karena nasabah adalah sumber utama keuntungan mereka. Salah satu sumber pendapatan terbesar bank adalah dari penawaran deposit yang mereka berikan kepada nasabah. Seiring dengan berjalannya waktu, perbankan mengeluarkan berbagai macam jenis dan bentuk investasi sebagai produk baru yang ditawarkan kepada para nasabah. Dampak kehadiran dari berbagai macam produk investasi tersebut akan membuat para nasabah yang dulunya menggunakan deposito sebagai produk investasi menjadi beralih ke produk investasi yang baru dengan beberapa macam faktor yang mempengaruhinya.

Berdasarkan pada penelitian sebelumnya, mengenai perbandingan algoritma machine learning dengan kombinasi tahapan preprocessing untuk memprediksi potensi hilangnya nasabah menunjukkan bahwa tahapan data preprocessing tidak memberikan pengaruh yang signifikan pada akurasi model algoritma machine learning [4]. Disisi lain, menggunakan metode klasifikasi data mining seperti neural network untuk mengklasifikasi nasabah yang layak mendapat pinjaman dengan memperhitungkan beberapa parameter yang tersedia menunjukkan hasil akurasi sebesar 98,24% [5]. Pada penelitian yang memprediksi potensi nasabah yang membuka simpanan deposito menunjukkan bahwa algoritma Particle Swarm Optimization (PSO) mampu mengoptimalkan algoritma Naive Baiyes dengan pembobotan atribut terlebih dahulu yang mengakibatkan akurasi model meningkat [6].

Di era digital yang berkembang pesat ini, peran teknologi informasi juga semakin diperkuat, sehingga membuka peluang baru bagi berbagai sektor, termasuk sektor perbankan. Teknologi seperti algoritma

machine learning dapat membantu sektor perbankan dalam memprediksi nasabah yang memenuhi kriteria untuk melakukan deposit. Penerapan algoritma machine learning dalam analisis nasabah memberikan kemampuan prediktif yang lebih tinggi, memungkinkan bank untuk mengidentifikasi potensi nasabah deposit dengan lebih tepat dan efisien. Pemanfaatan algoritma machine learning dalam konteks perbankan tidak hanya memberikan keunggulan analitis, tetapi juga memungkinkan bank untuk mengoptimalkan strategi pemasaran, menyusun penawaran yang lebih personal, dan merancang program loyalitas yang dapat menarik nasabah untuk tetap berinvestasi.

Dalam menghadapi tantangan tersebut, tujuan penelitian ini adalah untuk mengimplementasikan algoritma klasifikasi machine learning yang berupa Random Forest, XGBoost, SVC, dan Logistic Regression guna memprediksi dengan akurasi tinggi apakah seseorang akan menjadi nasabah dari bank tersebut. Pemilihan keempat metode tersebut didasarkan pada masih sedikit penelitian untuk memprediksi deposito nasabah menggunakan keempat metode tersebut. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam mendukung strategi pemasaran dan pengelolaan nasabah yang lebih efektif bagi lembaga keuangan tersebut. Berdasarkan permasalahan tersebut, tujuan penelitian ini yaitu untuk memprediksi apakah seorang akan menjadi nasabah dari bank tersebut menggunakan algoritma klasifikasi machine learning.

II. LITERATURE REVIEW

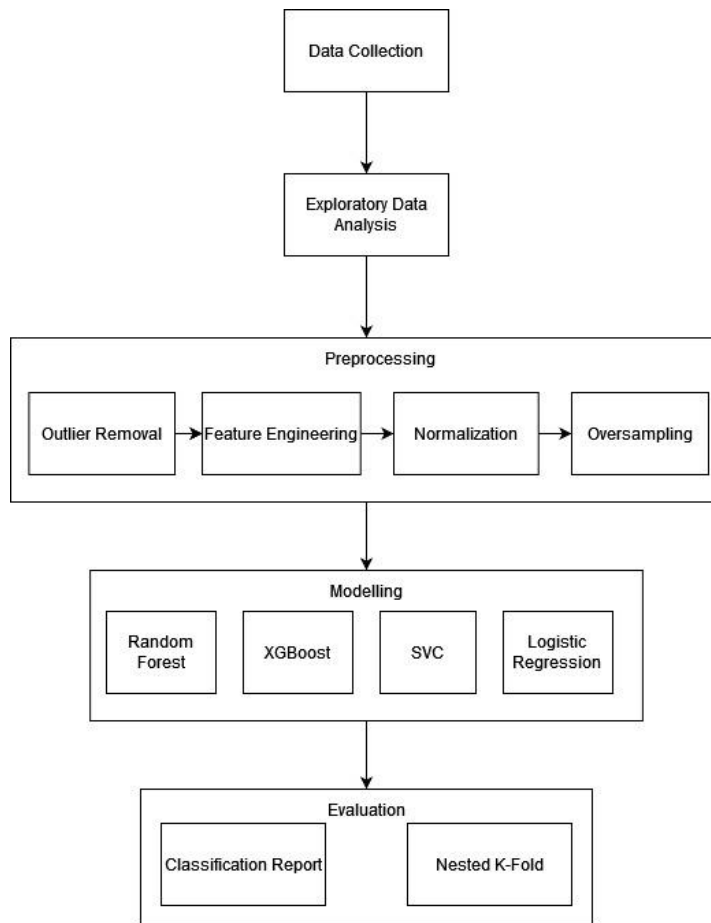
Terdapat beberapa penelitian mengenai prediksi term deposit nasabah perbankan, Prabowo et al. [6] melakukan prediksi potensi nasabah buka simpanan deposito menggunakan Naive Bayes yang berbasis PSO. PSO digunakan sebagai feature selection untuk meningkatkan hasil akurasi. Tingkat akurasi dari prediksi nasabah tersebut ketika menggunakan Naive Bayes sebesar 82,19%, sedangkan ketika menggunakan Naive Bayes yang berbasis PSO menghasilkan akurasi sebesar 89,70%.

Nia Nuraeni [9] melakukan penelitian klasifikasi data mining untuk prediksi potensi nasabah buka deposito berjangka. Algoritma klasifikasi yang digunakan yaitu Decision Tree, Naive Bayes dan KNN. Dengan membandingkan ketiga algoritma tersebut, didapat bahwa algoritma Decision Tree memiliki akurasi sebesar 91.26%, Naive Bayes sebesar 86.96%, dan KNN sebesar 90.39%.

Soraya Dachi et al [10] melakukan penelitian dengan membandingkan XGBoost dan Random Forest untuk klasifikasi keputusan kredit. Penelitian tersebut menghasilkan kinerja model algoritma XGBoost memiliki akurasi sebesar 100% untuk 10.000 dan 100.000 data, sedangkan Random Forest memiliki akurasi sebesar 99% untuk 10.000 dan 100.000 data.

Penelitian selanjutnya dilakukan oleh Audenza Maulana et al [11] dalam membandingkan metode Logistic Regression, MultinomialNB, SVM, dan KNN pada ulasan aplikasi Gojek. Untuk evaluasi menggunakan classification report dan confusion matrix. Metode KNN memiliki skor recall paling tinggi sebesar 93,52%, namun skor lainnya sangat rendah. Untuk metode yang memiliki performa cukup baik pada keseluruhan skornya yaitu metode MultinomialNB dan SVM yaitu memiliki rentang skor sekitar 76%.

III. RESEARCH METHOD



Gambar 1. Research Method

3.1 Data Collection

Dataset yang digunakan untuk proses klasifikasi ini merupakan dataset yang kita ambil dari Hugging Face dan mempunyai variabel independent sebanyak 16 kolom yang terdiri age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, dan poutcome. Dan targetnya adalah y yang merupakan apakah nasabah melakukan term deposit atau tidak.

3.2 Exploratory Data Analysis

EDA (exploratory data analysis) merupakan tahapan yang dilakukan untuk mengetahui karakteristik dan mendapatkan insight dari data melalui visualisasi, sehingga kita dapat melakukan treatment yang sesuai untuk mencapai hasil yang terbaik [12]. Exploratory Data Analysis (EDA) adalah pendekatan analisis data di mana fitur dan karakteristik data ditinjau tanpa harus menerapkan model tertentu pada data [13]. EDA membantu menganalisis kumpulan data untuk merangkum fokus karakteristik statistiknya pada aspek ukuran tendensi sentral (terdiri dari mean, modus dan median), ukuran spread (terdiri dari standar deviasi dan varians), bentuk distribusi dan keberadaan outlier [14]. Biasanya, EDA digunakan pada tahap awal ketika data masih mentah atau sebelum model apa pun dipilih untuk komponen struktural atau stokastik. Tahap EDA pada penelitian ini terdiri dari visualisasi grafik Box Plot, Count Plot dan Dist Plot. Pada Box Plot dapat menggambarkan beberapa ukuran statistik seperti nilai minimum, kuartil pertama, median,

kuartil ketiga, nilai maksimum dan ada tidaknya nilai outlier. Pada Count Plot digunakan untuk merepresentasikan jumlah observasi variabel kategori untuk setiap grup dengan menggunakan grafik diagram batang. Pada Dist Plot dapat memvisualisasikan distribusi dari data numerik atau menggambarkan variasi sebaran data.

3.3 Preprocessing

Preprocessing merupakan tahap yang sangat penting dan membutuhkan waktu, dimana data akan diolah sebelum diterapkan di tahap modelling [15]. Preprocessing yang dilakukan pada penelitian ini berupa penanganan menghapus outlier, Feature engineering, normalisasi, dan oversampling. Hal tersebut dilakukan berdasarkan hasil EDA sebelumnya.

3.4 Modelling

Pada tahap ini, data yang telah melewati tahap preprocessing akan melalui proses pelatihan menggunakan model Random Forest, Logistic Regression, SVC, dan XGBoost yang telah dibangun. Berikut adalah penjelasan singkat mengenai model yang digunakan:

a) Random Forest

Random Forest biasa digunakan untuk menyelesaikan masalah yang berhubungan dengan klasifikasi, regresi, dan sebagainya. Algoritma ini merupakan gabungan dari beberapa decision trees dimana setiap tree bergantung pada nilai random vector. Untuk rumus persamaan Random Forest yang terdiri dari N trees yaitu sebagai berikut.

$$l(y) = \operatorname{argmax}_c \left(\sum_{n=1}^N l_{h_n(y)=c} \right)$$

b) XGBoost

Metode XGBoost merupakan pengembangan dari gradient boosting. XGBoost menggunakan model yang lebih teratur untuk membangun struktur pohon regresi, sehingga dapat memberikan kinerja yang lebih baik dan mampu mengurangi kompleksitas model untuk menghindari overfitting [16].

c) Support Vector Classifier (SVC)

SVC merupakan ekstensi dari maximal margin classifier yang mengasumsikan bahwa memperbolehkan beberapa data diklasifikasikan salah dengan parameter kontrol ϵ yang melambangkan apakah suatu observasi boleh berada pada ruang yang tidak tepat. Berikut merupakan persamaan SVC.

$$y_i(x_i \cdot w + b) \geq 1(1 - \epsilon_i)$$

Untuk membatasi banyaknya kesalahan yang dibuat dapat menggunakan constraint baru seperti berikut.

$$\epsilon \geq 0; \sum \epsilon_i \leq C$$

d) Logistic Regression

Regresi logistik merupakan bentuk regresi yang digunakan ketika outputnya berupa nilai biner dan juga memungkinkan nilai target prediksinya dalam interval suatu nilai peluang, yaitu $[0;1]$, dengan inputnya berupa data yang bisa sangat variatif (nominal, ordinal, interval atau tingkat rasio atau lainnya) dan dengan dimensi berapapun [17]. Berikut merupakan persamaan dari model Logistic Regression.

$$y = b_0 - b_1x$$

Untuk fungsi sigmoid yaitu:

$$p = \frac{e^y}{e^y + 1}$$

Sehingga didapat persamaan:

$$\ln\left(\frac{p}{1-p}\right) = b_0 - b_1x$$

3.5 Evaluation

Evaluation adalah tahap terakhir dari penelitian ini, tahap ini bertujuan untuk mengukur seberapa baik model yang telah kita buat dalam mengklasifikasikan data yang diinputkan. Ukuran yang diterapkan dalam penelitian ini berdasarkan classification report dan nested k-fold.

a. Classification Report

Untuk evaluasi menggunakan classification report dihitung berdasarkan besaran dari precision, recall, akurasi, dan f1-score. Untuk rumus dari keempat besaran tersebut dapat menggunakan persamaan berikut.

$$Precision = \frac{(TP)}{(TP + FP)}$$

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

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

$$F1\ Score = \frac{2(Recall\ Precision)}{(Recall + Precision)}$$

Keterangan :

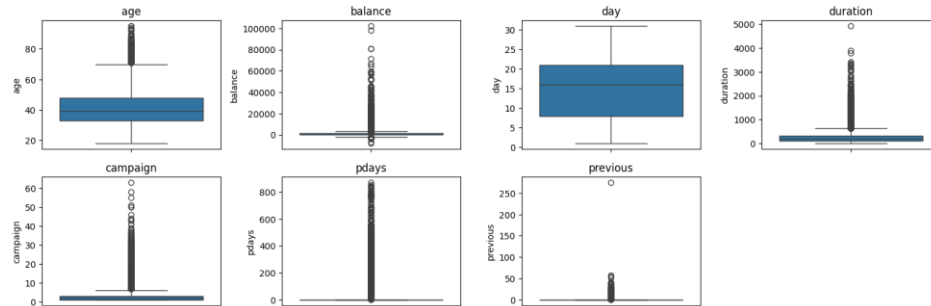
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

IV. RESULTS AND DISCUSSION

4.1 Exploratory Data Analysis

Pada tahap ini telah dilakukan analisis visualisasi seperti berikut:

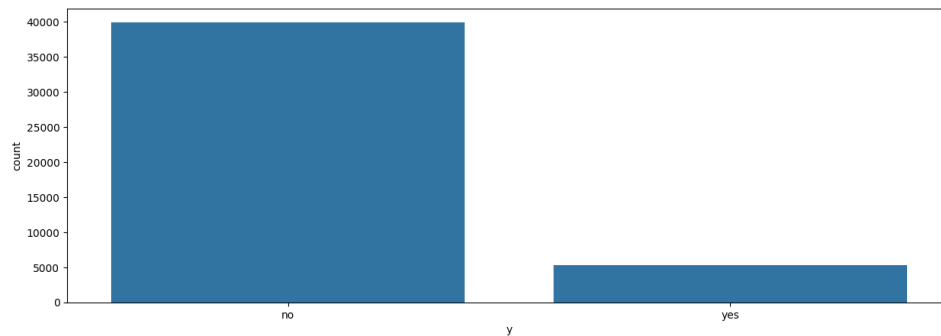
a. *Box Plot*



Gambar 2. Box Plot

Dalam hasil visualisasi box plot yang disajikan di atas, teramati bahwa sejumlah outlier yang signifikan terdeteksi pada variabel *age*, *balance*, *duration*, *campaign*, *pdays*, dan *previous*. Oleh sebab itu, selanjutnya pasti akan dilakukan *preprocessing* berupa *outlier removal*.

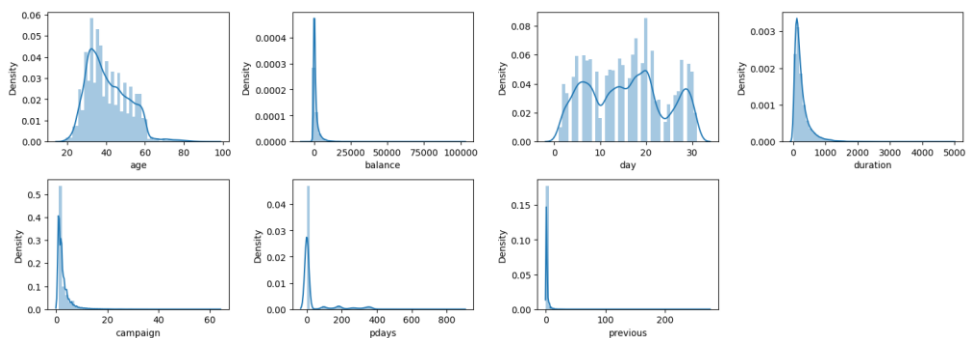
b. *Count Plot*



Gambar 3. Count Plot

Dalam hasil visualisasi count plot yang disajikan di atas, telah ditemukan ketidakseimbangan pada distribusi kelas variabel target. Oleh sebab itu, selanjutnya akan dilakukan metode *oversampling* untuk mengatasi ketidakseimbangan tersebut.

c. *Dist Plot*



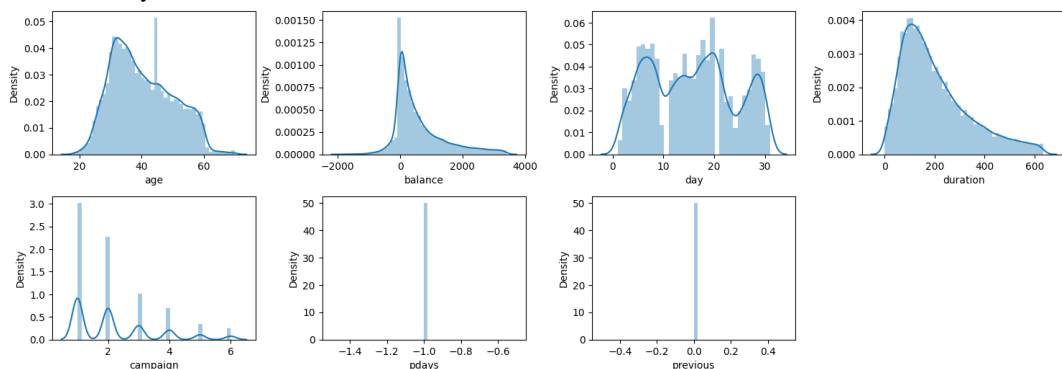
Gambar 4. Dist Plot

Dalam hasil visualisasi dist plot yang disajikan di atas, telah ditemukan indikasi *skewness* dan *outlier*. Oleh sebab itu, kita akan melakukan menghilangkan outlier seperti yang telah disebutkan sebelumnya dan melakukan transformasi data jika hasil penghapusan outlier dan model tidak memuaskan.

4.2 Preprocessing

a. Outlier Removal

Setelah kita lakukan feature engineering, tahap selanjutnya adalah outlier removal. Metode yang peneliti gunakan adalah inter-quartile range. Berikut adalah persebaran data setelah dilakukannya outlier removal



Gambar 5. Distribusi Data Setelah IQR

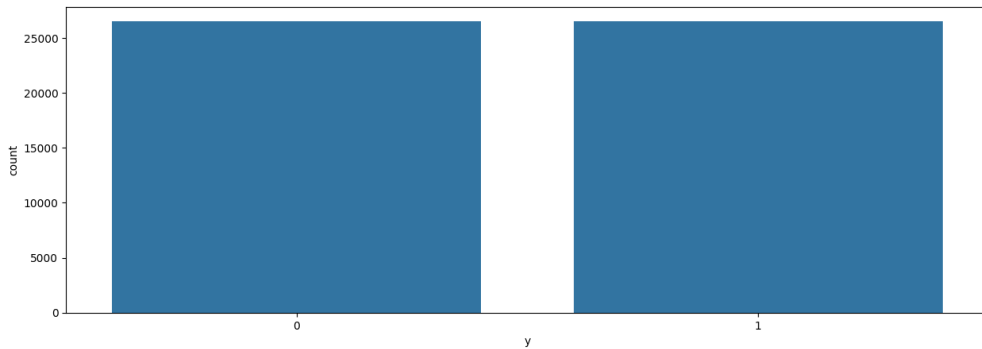
Dalam hasil visualisasi dist plot setelah dilakukannya *outlier removal* yang disajikan di atas, dapat terlihat bahwa distribusi telah mendekati normal dan outlier telah berkurang secara signifikan. Oleh sebab itu, peneliti tidak akan melakukan transformasi data.

b. Feature Engineering

Pada tahap feature engineering, peneliti membuat kolom-kolom baru berdasarkan hasil segmentasi menggunakan algoritma clustering yaitu K-Means, dan didapatkan kolom yang berkorelasi antara lain, balance by duration, duration segmentation, duration by age, job by balance, job by duration, dan age by job.

c. Oversampling

Pada tahap ini, peneliti melakukan oversampling menggunakan metode SMOTE untuk mengatasi imbalance pada data, dan didapatkan hasil seperti berikut:



Gambar 6. Count Plot Setelah *Oversampling*

Dalam hasil visualisasi count plot setelah dilakukannya *oversampling* yang disajikan di atas, dapat terlihat bahwa kelas 0 dan 1 telah seimbang. Oleh sebab itu, peneliti akan melanjutkan pada tahap selanjutnya yaitu *modelling*.

4.3 Modelling

Pada tahap modelling kami memilih kolom housing, marital, education, job, contact, loan, balance, dan kolom hasil feature engineering sebelumnya. Selanjutnya kami lakukan splitting dengan data uji sebesar 20%. Hingga pada akhirnya kami lakukan modelling menggunakan algoritma Random Forest, Logistic Regression, XGBoost Classifier, dan Support Vector Classification.

4.4 Evaluation

Tahap terakhir dari penelitian ini adalah evaluasi menggunakan classification report dan Nested K-Fold untuk menguji apakah model mengindikasikan overfitting atau tidak. Dan didapatkan hasilnya sebagai berikut:

a. Classification Report

Table 1. *Classification Report*

Method	Precision	Recall	Accuracy	F1 Score
Random Forest	0,917	0,917	0,917	0,917
XGBoost	0,917	0,913	0,913	0,913
Logistic Regression	0,723	0,723	0,723	0,723
SVM	0,797	0,796	0,796	0,795

Berdasarkan hasil evaluasi classification report yang telah dilakukan, didapatkan nilai presisi, recall, akurasi, dan f1-score terbaik adalah model random forest. Dengan nilai stabil sebesar 91.7%.

b. Nested K-Fold

Table 2. *Nested K-Fold*

Model	Nested K-Fold
RF	0.913749
LR	0.715534
XGB	0.910655
SVC	0.790579

Selanjutnya peneliti melakukan uji overfitting terhadap model yang telah dilatih menggunakan *Nested K-Fold*. Berdasarkan hasil uji tersebut didapatkan bahwa model tidak terjadinya indikasi *overfitting*, Dimana hasil *Nested K-Fold* tidak jauh berbeda dengan hasil akurasi pada *classification report* sebelumnya. Oleh sebab itu, peneliti akan menggunakan model terbaik yaitu *random forest*.

V. Conclusion

Berdasarkan hasil klasifikasi term deposit nasabah yang telah dilakukan preprocessing seperti outlier removal, oversampling normalisasi, dan feature engineering, didapatkan hasil dari model terbaik yaitu random forest dan xgboost, dengan akurasi sebesar 91.7%. Hal tersebut tidak menunjukkan adanya overfitting dikarenakan tidak banyaknya perbedaan nilai pada classification report dan hasil nested k-fold. Dengan memahami pola perilaku nasabah terkait keputusan untuk berlangganan term deposit, bank dapat mengoptimalkan strategi pemasaran mereka. Dengan menggunakan model ini, bank dapat secara efektif menargetkan segmen nasabah yang lebih cenderung untuk melakukan investasi dalam term deposit, meningkatkan efisiensi dan mengurangi biaya pemasaran yang tidak perlu.

REFERENCES

- [1] Heny Agustina, “Jurnal Manajemen KINERJA,” *J. Manaj. Kinerja*, vol. Vol. 3 Nom, pp. 57–63, 2017.
- [2] S. V. Wahyuni and A. Afriyeni, “Aktivitas Penghimpunan Dana Deposito Pada Pt. Bank Pembangunan Daerah (Bpd) Sumatera Barat Cabang Lintau,” *Management*, vol. 2, no. 1, pp. 25–26, 2019.
- [3] G. Suhardi, “Pengaruh Atribut Produk Yang Dipertimbangkan Dalam Pemilihan Produk Deposito Perbankan Terhadap Kepercayaan Dan Dampaknya Pada Loyalitas Nasabah,” *Kinerja*, vol. 12, no. 1, pp. 100–109, 2017, doi: 10.24002/kinerja.v12i1.1393.
- [4] Mohammad Farid Naufal, Subrata, Alvin Fernando Susanto, Christian Nathaneil Kansil, and Solichul Huda5, “Application of Machine Learning to Predict Potential Loss of Bank Customer,” *Februari*, vol. 22, no. 1, pp. 1–11, 2023.
- [5] N. Hadiano, H. B. Novitasari, and A. Rahmawati, “Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163–170, 2019, doi: 10.33480/pilar.v15i2.658.
- [6] A. D. R. Prabowo and M. Muljono, “Prediksi Nasabah Yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization,” *Techno.Com*, vol. 17, no. 2, pp. 208–219, 2018, doi: 10.33633/tc.v17i2.1648.
- [7] M. H. Nasution and S. Sutisna, “Faktor-Faktor Yang Mempengaruhi Minat Nasabah Terhadap Internet Banking,” *Nisbah J. Perbank. Syariah*, vol. 1, no. 1, p. 62, 2015, doi: 10.30997/jn.v1i1.241.
- [8] T. Abdullah and S. Wahjusaputri, *(CEK) Bank & Lembaga Keuangan*. 2018.
- [9] N. Nuraeni, “Klasifikasi Data Mining untuk Prediksi Potensi Nasabah dalam Membuat Deposito Berjangka Data Mining Classification for Predicting Customer Potential in Making Term Deposits,” *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 3, no. 01, pp. 65–75, 2021.
- [10] J. Melvin and A. Soraya, “Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit,” *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam (JURRIMIPA)*, vol. 2, no. 2, pp. 87–103, 2023.
- [11] A. Maulana, I. K. Afifah, A. Mubarrak, and K. R. Fauzan, “Comparison of Logistic Regression , Multinomialnb , Svm , and K-Nn Methods on Sentiment Analysis of Gojek App Reviews on the Google Play Store Perbandingan Metode Logistic Regression , Multinomialnb , Svm , Dan K-Nn Pada Analisis Sentimen Ulasan Aplikasi ,” vol. 4, no. 6, pp. 1487–1494, 2023.
- [12] I. Cholissodin and A. A. Soebroto, “AI , MACHINE LEARNING & DEEP LEARNING (Teori & Implementasi),” no. January, 2021.
- [13] “Exploratory Data Analysis,” in *The Concise Encyclopedia of Statistics*, New York, NY: Springer New York, 2008, pp. 192–194. doi: 10.1007/978-0-387-32833-1_136.
- [14] K. Sahoo, A. K. Samal, J. Pramanik, and S. K. Pani, “Exploratory data analysis using python,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 12, pp. 4727–4735, 2019, doi: 10.35940/ijitee.L3591.1081219.

- [15] D. Sartika and I. Saluza, “Penerapan Metode Principal Component Analysis (PCA) Pada Klasifikasi Status Kredit Nasabah Bank Sumsel Babel Cabang KM 12 Palembang Menggunakan Metode Decision Tree,” *GENERIC J. Ilmu Komput. dan Teknol. Inf.*, vol. 14, no. 2, pp. 45–49, 2022.
- [16] H. Sunata, “Komparasi Tujuh Algoritma Identifikasi Fraud ATM Pada PT. Bank Central Asia Tbk,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 3, pp. 441–450, 2020, doi: 10.35957/jatisi.v7i3.471.
- [17] N. B. Muliawan and I. A. Sulistijono, “Indonesian Journal of Computer Science,” *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023, doi: <https://doi.org/10.33022/ijcs.v12i1.3135>.