Jurnal Impresi Indonesia (JII)
p-ISSN: 2828-1284 e-ISSN: 2810-062x
website: https: //rivierapublishing.id/JII/ index.php / jii /index

Analisis Metode Klasifikasi Nasabah Potensial dalam Membuka Deposito Jangka Panjang Melalui *Telemarketing* Menggunakan Metode *Gradient Boosting Classifier*

Michael Richard Takakobi[1], Kristoko Dwi Hartomo[2]
Universitas Kristen Satya Wacana, Indonesia[1,2]
Email: 682018151@student.uksw.edu, kristoko@uksw.edu

Abstrak:
Penelitian ini bertujuan untuk melihat klasifikasi nasabah dalam membuka sebuah deposito jangka panjang. Penelitian ini akan menggunakan algoritma Machine Learning, yaitu Gradient Boosting Classifier. Dataset yang digunakan diambil dari arsip dataset UC Irvine Machine Learning Repository yang mencangkup 41.188 sampel dengan 20 variabel. Dataset Kampanye bank di Portugal menggunakan metode penawaran dilakukan secara jarak jauh atau tidak langsung yang biasa disebut dengan Telemarketing. Nasabah dalam dataset terdiri dari berbagai latar belakang. Penelitian memberikan hasil baik dalam klasifikasi menggunakan metode Gradient Boosting Classifier, metode ini dikombinasikan dengan teknik random oversamping untuk mengatasi data imbalance. Penelitian menghasilkan nilai ROC-AUC sebesar 0.81. Hasil penelitian juga memberikan informasi yang dapat digunakan dalam pengambilkan Keputusan terkait dengan kampanye telemarketing selanjutnya. Penelitian ini merekomendasikan penerapan model ini untuk meningkatkan efisiensi telemarketing dengan menargetkan nasabah berpotensi, sekaligus mengurangi biaya operasional.

Kata kunci: Telemarketing, Deposito Jangka Panjang, Klasifikasi, Gradient Boosting Classifier.

Abstract:
This study aims to look at the classification of customers in opening a long-term deposit. This research will use a Machine Learning algorithm, namely the Gradient Boosting Classifier. The dataset used was taken from the UC Irvine Machine Learning Repository dataset archive which included 41,188 samples with 20 variables. Dataset Bank campaigns in Portugal using a remote or indirect bidding method commonly known as Telemarketing. Customers in the dataset come from a variety of backgrounds. The research gave good results in classification using the Gradient Boosting Classifier method, this method was combined with random overside techniques to overcome data imbalance. The study resulted in a ROC-AUC value of 0.81. The results of the study also provide information that can be used in making decisions related to the next telemarketing campaign. This study recommends the application of this model to improve telemarketing efficiency by targeting potential customers, while reducing operational costs.

***Keywords**: Telemarketing, Long-Term Deposits, Classification, Gradient Boosting Classifier.*

PENDAHULUAN

Industri perbankan saat ini menghadapi tantangan besar dalam mengkonversi jumlah nasabah untuk membuka tabunggan deposito jangka panjang. Deposito adalah produk investasi

dengan menyimpan uang dan penarikannya hanya bisa dilakukan dengan kurun waktu tertentu yang telah di janjikan oleh pihak bank dengan persetujuan nasabah (Adani et al., 2018). Deposito jangka panjang menjadi salah satu produk unggulan bank karena menawarkan keuntunan bunga yang lebih tinggi bagi nasabah dan menawarkan stabilitas dana bagi bank. Namun, tingkat keberhasilan penawaran deposito melalui metode konvensional masih tergolong rendah (Istyaningsih, 2015).

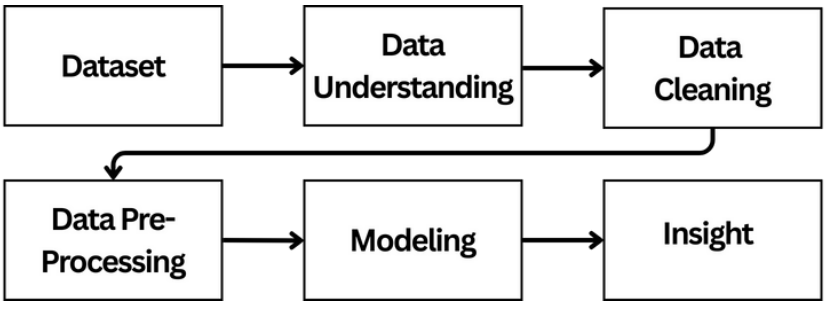
Telemarketing merupakan metode pemasaran langsung yang menggunakan telepon untuk menjangkau dan menawarkan produk atau layanan kepada calon pelanggan (Muneeb Asif, 2018). Melalui Telemarketing, bank dapat menjangkau lebih banyak orang secara efesien. *Telemarketing* juga memiliki kelemahan dimana dengan begitu besarnya database sebuah bank, maka tidak semua atribut dibutuhkan kedalam pemilihan nasabah potensial. Oleh karena itu diperlukan pendekatan berbasis teknologi yang mampu mengolah *database* yang besar untuk menemukan pola-pola yang dapat mengklasifikasikan nasabah dalam membuka deposito jangka panjang (Vongchalerm, 2022).

Dalam beberapa tahun terakhir, berbagai metode klasifikasi berbasis *data mining* dan *machine learning* telah banyak diterapkan untuk membantu bank dalam mengidentifikasi nasabah potensial. Metode-motode seperti *K-Nearest Neighbor*, *Logistic Regression*, *XGBoost*, *Decision Tree*, *Naïve Bayes*, *Neural Network* dan *Support Vector Machine* telah dibandingkan dalam berbagai penelitian untuk mengetahui tingkat akurasi dan efetkivitasnya dalam memprediksi nasabah yang berpotensi membuka deposito jangka Panjang melalui telemarketing (Adani et al., 2018; Huang, 2024; Moro et al., 2014; Muneeb Asif, 2018; Vongchalerm, 2022). Hasil penelitian menunjukkan bahwa setiap model memiliki kelebihan dan kekurangan masing-masing, serta tingkat akurasi yang bervariasi tergantung pada karakteristik data dan parameter yang digunakan.

Dalam penelitian ini, penggunaan statistika dan metode klasifikasi digunakan untuk melakukan klasifikasi terhadap *dataset telemarketing* sebuah bank di Portugal. Secara Spesifik penelitian ini akan menggunakan metode *Gradient Boosting Classifier*. Untuk melakukan evaluasi terhadap hasil klasifikasi akan digunakan *confusion matrix* dan ROC-AUC untuk menilai performa tiap metode.

METODE PENELITIAN

Tahap-tahap yang digunakan pada penelitian ini adalah (1) Data *Understanding*, (2) Data *Cleaning*, (3) Data *Pre-Processing*, (4) *Modelling*, (5) *Insight*.



Gambar 1. Diagram Metode Penelitian

A. Data Understanding

Data *Understanding* merupakan tahapan pengumpulan, eksplorasi dan evaluasi data awal yang akan digunakan dalam analisis atau permodelan. Tujuan utamanya adalah untuk memahami karateristik, struktur, kualitass dan potensi informasi yang terkandung dalam data, serta mengidentifikasi masalah yang mungkin ada sebelum data diproses lebih lanjut. Data pada penelitian ini diambil dari database UCI Repository (https://archive.ics.uci.edu/ml/datasets/Bank+Marketing) , tentang sebuah kampanye Bank di Portugal untuk pembukaan deposito jangka panjang (Budiman et al., 2014)(Hasanah et al., 2021).

B. Data Cleaning

Data *Cleaning* adalah proses mendeteksi dan memperbaiki atau menghapus data yang salah, tidak lengkap, duplikat, atau tidak konsisten dalam sebuah dataset untuk meningkatkan kualitas data. Proses ini menjadi krusial dikarenakan kualitas data secara signifikan mempengaruhi kualitas akurasi analisis, pemodelan dan akhirnya pengambilan keputusan (Guo et al., 2023).

C. Data Pre-Processing

Data *Pre-Processing* merupakan tahapan persiapan mencakup semua aktivitas untuk mempersiapkan *dataset* akhir dari adat awal mentah (Ristyawan et al., 2025). Tahapan ini mengimplementasikan beberapa proses tahapan. Setiap tahapan bertujuan agar data menjadi lebih optimal dan siap untuk dimasukan kedalam model akhir (Airlangga et al., 2024).

D. Modeling

Modeling merupakan tahapan yang secara langsung melibatkan *Machine Learning* untuk penentuan teknik *data mining*, dan menentukan algoritma yang akan digunakan sesuai dengan proses bisnis yang berjalan (Alsolami et al., 1441). Pada penelitian kali ini akan menggunakan algoritma *Gradient Boosting Classifier*.

Proses model evaluation penting dalam *machine learning*. Permodelan pada penelitian ini akan dievaluasi menggunakan *confusion matrix* dan juga ROC-AUC.

E. Insight

Insight merupakan tahapan yang bertujuan untuk menarik kesimpulan dari tahap-tahap yang telah dilakukan sebelumnya. *Output* pada tahapan ini akan ada pada Kesimpulan (Simarmata et al., 2022).

HASIL DAN PEMBAHASAN

A. Data Understanding

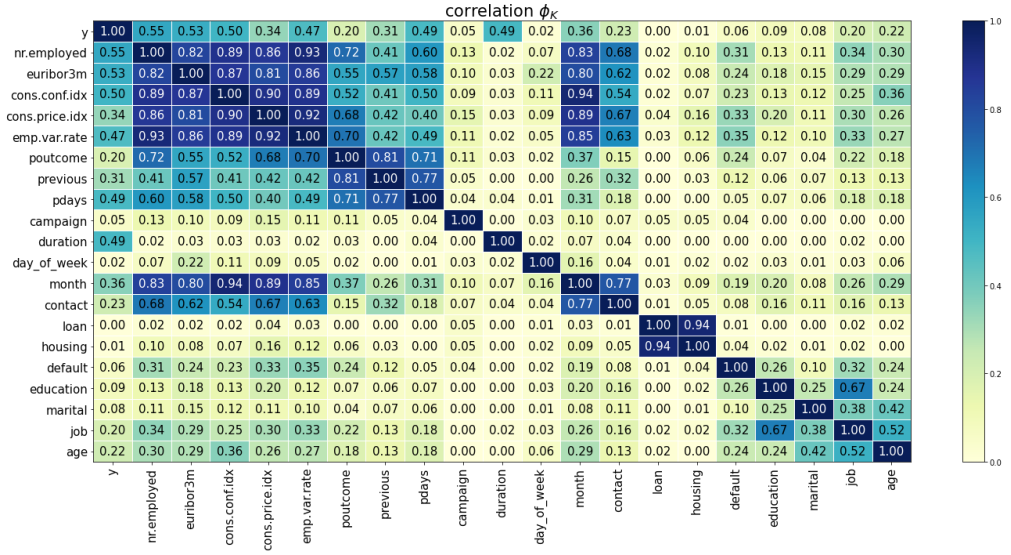
Dataset terdiri dari 41.188 record, 20 atribut dengan 10 atribut numerikal dan 10 atribut kategorikal dan juga 1 target kelas.

Tabel 1. Deskripsi Dataset

No	Fitur	Deskripsi	Tipe Data
1	Age	Umur nasabah	Int64
2	Job	Pekerjaan nasabah	*Object*
3	Marital	Status pernikahan nasabah	*Object*
4	Education	Pendidikan terakhir nasabah	*Object*
5	Default	Nasabah memiliki tunggakan	*Object*

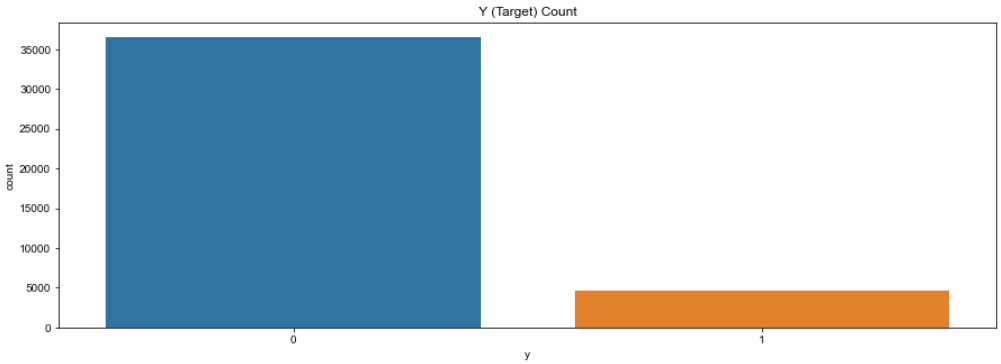
No	Fitur	Deskripsi	Tipe Data
6	Housing	Nasabah memiliki KPR	*Object*
7	Loan	Nasabah memiliki hutang ke bank	*Object*
8	Contact	Tipe kontak dengan nasabah	*Object*
9	Month	Bulan terakhir kontak dengan nasabah	*Object*
10	Day of Week	Hari terakhir kontak dengan nasabah	*Object*
11	Duration	Durasi kontak dengan klien (dalam detik)	Int64
12	Campaign	Jumlah kontak selama kampanye bank	Int64
13	Pdays	Jumlah hari berlalu dari kedatangan terakhir nasabah	Int64
14	Previous	Jumlah kontak dengan nasabah sebelum kampanye bank ini	Int64
15	Poutcome	Hasil dari kampanye pemasaran terakhir	*Object*
16	Emp.var.rate	Variasi lapangan kerja	Float64
17	Cos.Price.IDX	Index harga konsumen	Float64
18	Cos.Conf.IDX	Index kepercayaan diri konsumen	Float64
19	Euribor3m	Suku bunga Euribor 3 bulan	Float64
20	Nr.employed	Jumlah Karyawab	Float64
21	y	Status sudah belumnya nasabah membuka deposito jangka panjang	object

Sebelum dilakukan data *cleaning* dan *preprocessing* untuk modeling, dilakukan beberapa visualiasi untuk dapat lebih mengerti *dataset*. Visualisasi korelasi digunakan untuk melihat korelasi dari tiap kolom terhadap kolom lainnya, khususnya terhadapat target kelas ‘y’.



Gambar 2. Visualisasi Korelasi

Tahapan terakhir dari data *understanding* adalah melakukan pengecekan terhadapat kelas target ‘y’. Dapat dilihat dari visualiasi data di Gambar 3 bahwa data di kelas *target imbalance*. Dimana 1 kelas nilai secara signifikan lebih besar dari kelas lainnya.



Gambar 3. Visualisasi Kelas Target

B. Data Cleaning

1. Handling Duplicates

Pada proses ini diawali dengan pengecekan pada berapa jumlah baris duplikat dengan menggunakan kode dibawah.

Kode Program 1. Kode Program untuk melihat baris duplikat

```
print(df[df.duplicated(keep=False)].shape)
```

Setelah dijalankan menunjukkan terdapat 24 baris duplikat, setelah mengetahui jumlah baris yang duplikat, akan dijalankan sebuah kode, untuk hanya menyisakan kemunculkan baris duplikat terakhir.

Kode Program 2. Kode Program untuk menyisakan kemunculan baris duplikat terakhir

```
df.drop_duplicates(keep='last',inplace=True)
```

2. Handling NaN

Pada proses ini pertama harus dilakukan kovenrsi nilai "*unknown*" pada dataset menjadi NaN.

Tabel 2. Tabel Deskripsi nilai Unknown

Kolom	Jumlah baris Unknown	Presentase Unknown
job	330	0.8%
marital	80	0.19%
education	1731	4.2%
default	8597	20.87%
housing	990	2.4%
loan	990	2.4%

Untuk merubah nilai Unknown menjadi NaN, maka akan dijalankan kode dibawah.

Kode Program 3. Kode Program untuk merubah nilai 'unknown' menjadi Nan

```
df.repalce({'unknown' : pd.np.nan})
```

Tabel 3. Deskripsi nilai NaN

Kolom	Jumlah Baris	Presentase
job	330	0.8 %
marital	80	0.19 %
education	1731	4.2 %
default	8597	20.87 %
housing	990	2.4%
loan	990	2.4 %

Dengan melihat bahwa kolom default memiliki lebih dari 20% baris dengan nilai NaN, sehingga diputuskan untuk drop kolom default dari dataset akhir untuk permodelan. Selain kolom default, semua baris yang memiliki nilai NaN juga akan di drop.

Kode Program 3. Kode Program untuk melakukan drop kolom default dan baris dengan nilai Nan

```
df.drop(columns = 'default', inplace = True) #Kode untuk drop kolom default
df.dropna(inplace=True)      #Kode untuk drop baris dengan Nan
```

F. Data Pre-Processing

Proses pertama pada tahapan Data *Pre-Processing* adalah melakukan *drop* pada kolom duration, karena tidak relevan. Dikarenakan kita pihak bank baru akan melakukan kontak untuk kampanye, belum diketahui durasi kontak yang akan terjadi, untuk itu kolom duration akan di *drop*.

Proses selanjutnya akan dilakukan pengelompokkan terhadap beberapa nilai pada sebuah kolom, untuk mempercepat proses komputasi dan meningkatkan efisiensi. Kolom pertama adalah kolom Education dimana ada 3 kolom *basic education* untuk 4y, 6y dan 9y. ketiga nilai akan dilakukan pengelompokan menjadi 1 nilai basic education.

Tabel 4. Distribusi education

Education	Jumlah Baris
Basic.education	12.061
High.school	9241
Professional.course	5.097
University.degree	11.817
illiterate	18

Setelah kolom education, akan dilakukan pengelompokan nilai pada kolom age. Kolom age sebelumnya memiliki 77 nilai unik, akan dikelompok menjadi 7 nilai. Proses pengelompokan dilakukan menjadi rentang usia.

Tabel 5. Distribusi age

Rentang Umur	age	Jumlah Baris
< 20	1	75
20 - 29	2	5.290
30 - 39	3	15.946
40 - 49	4	9.892
50 - 59	5	6.415

Rentang Umur	age	Jumlah Baris
60 - 69	6	698
> 70	7	454

Proses selanjutnya adalah dengan mengubah nilai 999 pada kolom Pdays menjadi, Not Contacted dan nilai lainnya menjadi Contacted.

Tabel 6. Distribusi pdays

pdays	Jumlah Baris
Not Contacted	36.868
Contacted	1.366

Setelah merubah nilai pada kolom Pdays, akan dilakukan *outlier handling* untuk kolom Campaign. *Outlier handling* dilakukan dengan menghilangkan nilai *outlier* pada *dataset*. Nilai *outlier* dihilangkan agar dapat meminimalkan jumlah kontak dalam penawaran sampai nasabah membuka deposito jangka panjang.

Tahapan selanjutnya adalah melakukan encoding untuk merubah nilai kategorikal sehingga dapat direpsentasikan secara numerik. Untuk kolom housing dan loan, nilai no akan direpresentasikan oleh 0 dan nilai yes akan dipresentasikan oleh nilai 1. Pada kolom pdays nilai Not Contacted akan direpresentasikan oleh nilai 0 dan nilai Contacted akan di representasikan oleh nilai 1. Penambahan fitur pada . Pada kolom education nilai illiterate akan direpresentasikan oleh nilai 0, nilai basic.education akan direpresentasikan oleh nilai 1, nilai high.school akan direpsentasikan oleh nilai 2, nilai university.degree akan direpresentasikan oleh nilai 3, dan nilai professional.course akan direpresentasikan oleh nilai 4. Untuk kolom month dan job nilai akan direpresentasikan oleh jumlah frekuensi nilai pada kolom. Pada kolom age, marital, contact, day_of_week, campaign, dan poutcome dilakukan one-hot encoding. Penambahan fitur pada age menjadi age_1, age_2, age_3, age_4, age_5, age_6, age_7. Fitur marital menjadi marital_divorced, marital_married dan marital_single. Fitur contact menjadi contact_cellular dan contact_telephone. Fitur day_of_week menjadi day_of_week_fri, day_of_week_mon, day_of_week_thu, day_of_week_tue, day_of_week_wed. Fitur campaign menjadi campaign_1, campaign_2, campaign_3, campaign_4, campaign_5, campaign_6. Fitur poutcome menjadi poutcome_failure, poutcome_nonexistent, poutcome_success.

Setelah melakukan proses *encoding*, akan dilakukan proses scaling menggunakan *min-max scaler* agar nilai secara keseluruh berada pada rentang lebih besar sama dengan 0 sampai lebih kecil sama dengan 1.

Tahapan terakhir adalah akan dilakukan *random oversampling* terhadap target kelas terdapat pada Gambar 3, dikarenakan terjadinya ketimpangan pada jumlah kelas 0 dan 1. Teknik yang digunakan adalah teknik random oversampling dengan *sampling strategy* 0.6.

Tabel 7. Distribusi Data Tujuan

Kelas	Sebelum Oversampling	Sesudah Oversampling
0	22.937	22.937
1	2.993	13.762

G. Modeling

Pengolahan data menggunakan *Visual Studio Code* dan *Python.* Tahapan pertama adalah dengan membagi data yang telah melalui proses data *cleaning* dan data *pre-processing* sebesar 80% untuk data *training* dan 20% untuk data *testing*. Algoritma *Gradient Boosting Classifier* yang digunakan berasal dari *library scikit-learn*. Proses algoritma *Gradient Boosting Classifier* dapat dilihat pada *pseudocode* berikut.

Algorithm 1 Gradient Boosting Classifier Pseudocode

Input: Training dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, a differentiable loss function $L(y, F(x))$, number of iterations or regression trees M, learning rate ν.

Initialize: Model $F_0(x)$ with a constant value:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{N} L(y_i, \gamma).$$

for $m = 1$ to M **do**

A. Compute pseudo-residuals:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}, \text{ for } i = 1...N$$

B. Fit a weak learner $F_m(x)$ to pseudo-residuals r_{im}, and create a terminal region R_{jm}, for $j = 1...J_m$

C. For $j = 1...J_m$ compute:

$$\gamma_{jm} = \arg\min_{\gamma} \sum_{x \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$$

end for

Output: Final model $F_0(x)$ and $F_m(x)$ for $m = 1...M$.

Gambar 3. Pseudocode GBC

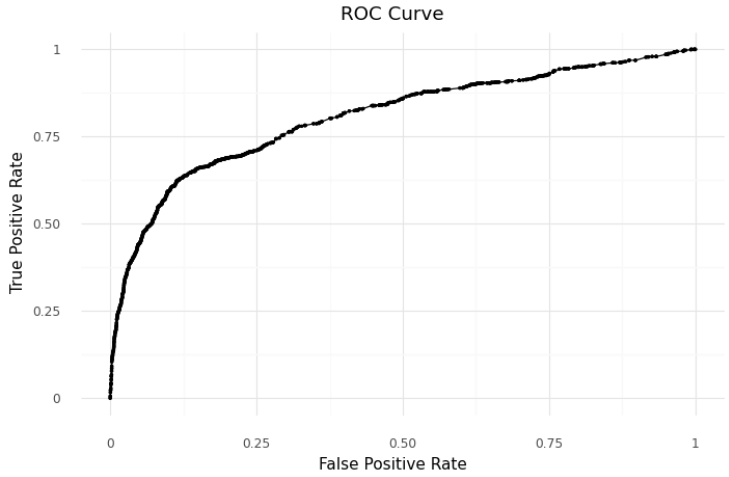
Tahap awal yang dilakukan adalah melakukan input data dan parameter. Data yang dimasukan adalah data *training* dan parameter L(y, F(x)) merupakan *loss function* yang dapat dibedakan yang adalah log peluang negatif. M merupakan banyak iterasi yang ingin dilakukan. V merupakan *learning rate*. Tahap ke-2 adalah menginisiasi sebuah fungsi dengan nilai konstan. Untuk menemukan nilai optimum dari γ (log peluang) yang meminimalkan nilai dari *loss function*. Tahap ke-3 merupakan sebuah perulangan untuk mencari parameter terbaik dalam data training. Tahap ke-4 menyimpan parameter terbaik berdasarkan perulangan kedalam model akhir. Setelah ada model terbaik, akan dilakukan pengetesan melakukan data *testing*.

Algorithm 2 Predict with Gradient Boosting Classifier

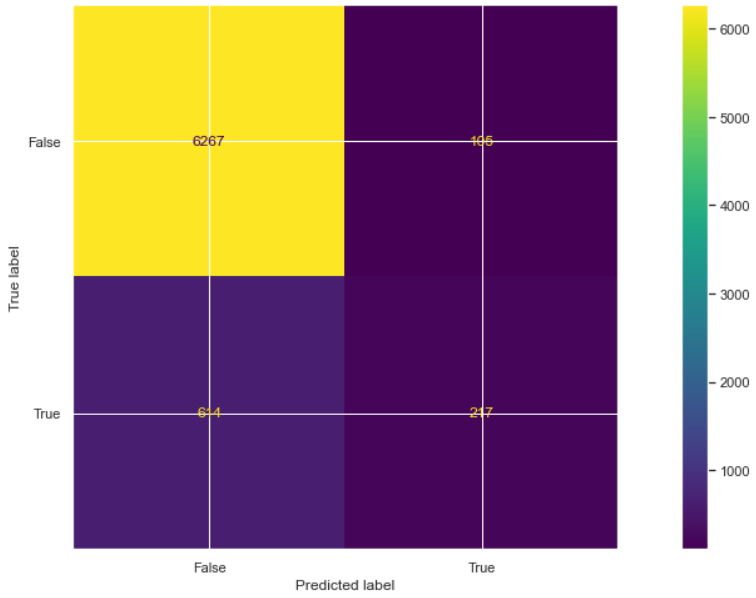
Input x: input features, $F_m(x)$: list of trained decision trees,
ν: learning rate used during training,
$total\_log\_odds \leftarrow 0$
for $m = 0$ to M **do**
  if $m = 0$ **then**
    $log\_odds \leftarrow F_m(x)$
  else
    $log\_odds \leftarrow \nu \cdot F_m(x)$
  end if
  $total\_log\_odds \leftarrow total\_log\_odds + log\_odds$
end for
$probability \leftarrow \exp(total\_log\_odds)/(1 + \exp(total\_log\_odds))$
if $probability \geq 0.5$ **then**
  $prediction \leftarrow 1$
else
  $prediction \leftarrow 0$
end if
return $prediction$

Gambar 4. Pseudocode Prediksi dengan GBC

Tahap ke-1 adalah x yang merupakan *input* data *test*, F_m(x) yang merupakan *regression tree* dan v yang merupakan *learning rate*. Tahap ke-2 adalah menginisiasi sebuah variabel Bernama *total_log_odss* untuk menyimpan total nilai dari *log_odds* tiap *regression tree*. Tahap ke-3 adalah melakukan perulangan untuk tiap F_m(x) agar dimasukan nilai x sebagai log peluang prediksi dan dikalikan dengan v, kecuali untuk iterasi 1. Setelah itu hasilnya akan ditambahkan ke dalam *total_log_odds*. Tahap ke-4 untuk melakukan perhitungan probabilitas dari *total_log_odds*. Tahap ke-5 adalah untuk membandingkan data apabila lebih besar sama dengan 1 maka hasil kelasnya adalah 1, dan apabila lebih kecil dari 0.5 maka hasil kelasnya adalah 0. Setelah dimasukkan kedalam algoritma *Gradient Boosting Classifier* data *testing* mendapatkan nilai ROC-AUC yang didapatkan adalah 0.81 yang masuk ke dalam klasifikasi baik.



Gambar 5. Visualisasi ROC Curve



Gambar 6. Visualiasi Confusion Matrix

KESIMPULAN

Penelitian ini telah melalui seluruh tahapan mulai dari *data understanding*, *data cleaning*, *data pre-processing*, *modeling*, hingga menghasilkan *insight*. Kesimpulan pertama menunjukkan bahwa algoritma Gradient Boosting Classifier memberikan performa yang baik dengan nilai ROC-AUC sebesar 0,81. Kesimpulan kedua adalah bahwa berdasarkan *confusion matrix*, pihak bank dapat mengeliminasi nasabah yang termasuk dalam klasifikasi True Negative dan lebih memfokuskan perhatian pada nasabah yang masuk dalam klasifikasi True Positive. Untuk penelitian selanjutnya, disarankan agar digunakan algoritma pemodelan lain guna membandingkan performa yang lebih optimal. Selain itu, penelitian lanjutan juga dapat menerapkan berbagai teknik lain dalam menangani masalah *imbalance class* pada variabel target.

DAFTAR PUSTAKA

Adani, Y., Studi, P., Informasi, S., & Bayes, A. N. (2018). Penerapan Algoritma Naive Bayes Untuk Memprediksi. *Jurnal Instrumentasi Dan Teknologi Informasi (JITI)*, *8*(1), 13–24.

Airlangga, G., Katolik, U., & Atma, I. (2024). *Enhancing Medical Diagnostics with Ensemble Machine Learning : A Comparative Study of Gradient Boosting , XGBoost , LightGBM , and Blended Models*. *9*, 1118–1132.

Alsolami, F. J., Saleem, F., & Al-Malaise Al-Ghamdi, A. (1441). Predicting the Accuracy for Telemarketing Process in Banks Using Data Mining. *JKAU: Comp. IT. Sci*, *9*(2), 69–83. https://doi.org/10.4197/Comp.

Budiman, I., Prahasto, T., & Christyono, Y. (2014). Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma. *Jurnal Sistem Informasi Bisnis*, *1*(3), 129–134. https://doi.org/10.21456/vol1iss3pp129-134

Guo, M., Wang, Y., Yang, Q., Li, R., Zhao, Y., Li, C., Zhu, M., Cui, Y., Jiang, X., Sheng, S., Li, Q., & Gao, R. (2023). Normal Workflow and Key Strategies for Data Cleaning Toward Real-World Data: Viewpoint. *Interactive Journal of Medical Research*, *12*, e44310. https://doi.org/10.2196/44310

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model

Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, *5*(2), 103–108. https://doi.org/10.30871/jaic.v5i2.3200

Huang, S. (2024). A Comparative Analysis of Machine Learning Algorithms for Predicting the Telemarketing Campaigns of Portuguese Banking Institutions. *Advances in Economics, Management and Political Sciences*, *94*(1), 44–53. https://doi.org/10.54254/2754-1169/94/2024ox0132

Istyaningsih, R. (2015). Studi Perilaku tentang Pengaruh Karakteristik Nasabah Bank dalam Memilih Deposito Berjangka. *Jurnal Ilmiah CIVIS*, *5*(1), 752–759.

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*(351), 22–31. https://doi.org/10.1016/j.dss.2014.03.001

Muneeb Asif. (2018). *Predicting the Success of Bank Telemarketing using various Classification Algorithms*. 14–18.

Ristyawan, A., Nugroho, A., & Amarya, T. K. (2025). *Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke*. *12*(1).

Simarmata, K. B., Hartomo, K. D., & Hartomo, K. D. (2022). Analisa Rekomendasi Fitur Persetujuan Pinjaman Perusahaan Financial Technology Menggunakan Metode Random Forest. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(3), 2055–2070. https://doi.org/10.35957/jatisi.v9i3.2258

Vongchalerm, L. (2022). *Analysis of predicting the success of the banking telemarketing campaigns by using machine learning techniques*.