Journal of Information System and Application Development
Volume 4 Number 1 March 2026 page 195-200
P-ISSN: 2988-5698 | E-ISSN: 2988-4721
https://jurnal.unmer.ac.id/index.php/jisad

Perbandingan Decision Tree dan Naive Bayes untuk Prediksi Customer Churn

Comparison of Decision Tree and Naive Bayes for Customer Churn Prediction

Firman Firdaus*, Lamsihar Pandapotan Pane
Program Studi Sistem Informasi,Fakultas Teknik dan Informatika,Universitas Bina Sarana Informatika,Jakarta,Indonesia

E-mail: * firman.frd98@gmail.com

Abstract. *Customer churn is a critical issue faced by companies as it directly affects revenue and business sustainability. Therefore, an accurate prediction method is required to identify customers who are likely to discontinue using a service. This study aims to compare the performance of Decision Tree and Naive Bayes algorithms in predicting customer churn based on service usage patterns. The dataset used in this study is the Customer Subscription Churn and Usage Patterns dataset consisting of 2,801 records with churn labels categorized as Yes and No. Data processing was conducted using RapidMiner through several stages, including data preprocessing, handling imbalanced data using the Synthetic Minority Over-sampling Technique (SMOTE), and model evaluation using cross-validation. The experimental results show that the Decision Tree algorithm achieved better overall performance compared to Naive Bayes, with an accuracy of 65.04% and an AUC value of 0.704, while Naive Bayes obtained an accuracy of 62.68% and an AUC value of 0.737. Although Naive Bayes produced a slightly higher AUC, the Decision Tree algorithm demonstrated a more balanced performance in terms of accuracy, precision, and recall for customer churn prediction. Based on these results, it can be concluded that the Decision Tree algorithm is more effective for predicting customer churn on the dataset used in this study.*

Keywords: *Customer Churn, Data Mining, Decision Tree, Naive Bayes, SMOTE*

Abstrak. *Customer churn* merupakan permasalahan penting yang dihadapi perusahaan karena dapat berdampak langsung terhadap penurunan pendapatan dan keberlangsungan bisnis. Oleh karena itu, diperlukan metode prediksi yang akurat untuk mengidentifikasi pelanggan yang berpotensi berhenti menggunakan layanan. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Decision Tree* dan *Naive Bayes* dalam memprediksi *customer churn* berdasarkan pola penggunaan layanan. Dataset yang digunakan merupakan data *Customer Subscription Churn and Usage Patterns* yang terdiri dari 2.801 data dengan label churn bernilai Yes dan No. Proses pengolahan data dilakukan menggunakan *RapidMiner* dengan tahapan *preprocessing*, penanganan data tidak seimbang menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE), serta evaluasi model menggunakan *cross validation*. Hasil penelitian menunjukkan bahwa *algoritma Decision Tree* menghasilkan kinerja yang lebih baik dibandingkan *Naive Bayes* dengan nilai akurasi sebesar 65,04% dan AUC sebesar 0,704, sedangkan *Naive Bayes* memperoleh akurasi sebesar 62,68% dan *AUC* sebesar 0,737. Meskipun *Naive Bayes* memiliki nilai *AUC* yang sedikit lebih tinggi, *Decision Tree* menunjukkan keseimbangan yang lebih baik antara akurasi, *precision*, dan *recall* dalam memprediksi *customer churn*. Berdasarkan hasil tersebut, dapat disimpulkan bahwa *Decision Tree* lebih efektif digunakan untuk prediksi *customer churn* pada dataset yang digunakan dalam penelitian ini.

Kata kunci: *Customer Churn*, Data Mining, *Decision Tree*, *Naive Bayes*, *SMOTE*

Submitted: 24-02-2026 | ***Accepted:*** 25-03-2026 | ***Published:*** 31-03-2026

How to Cite:
F. Firdaus, and L. P. Pane, "Perbandingan Decision Tree dan Naive Bayes untuk Prediksi Customer Churn" *Journal of Information System and Application Development (JISAD)*, vol. 4, no. 1, pp. 195-200, 2026, doi: 10.26905/jisad.v4i1.16837

DOI: 10.26905/jisad.v4i1.16837

PENDAHULUAN

Customer churn merupakan kondisi ketika pelanggan memutuskan untuk berhenti menggunakan suatu produk atau layanan dalam periode tertentu. Fenomena ini menjadi permasalahan serius bagi perusahaan karena dapat menyebabkan penurunan pendapatan serta meningkatnya biaya untuk memperoleh pelanggan baru. Beberapa penelitian menunjukkan bahwa kemampuan memprediksi *customer churn* secara dini dapat membantu perusahaan dalam menyusun strategi retensi pelanggan yang lebih efektif dan tepat sasaran [1], [2] .

Perkembangan teknologi informasi mendorong perusahaan untuk memanfaatkan data pelanggan dalam jumlah besar sebagai dasar pengambilan keputusan. Dalam konteks ini, data *mining* dan *machine learning* banyak digunakan untuk menganalisis pola perilaku pelanggan serta memprediksi kemungkinan terjadinya *churn*. Metode klasifikasi merupakan salah satu teknik data mining yang paling umum digunakan dalam prediksi *customer churn* karena mampu mengelompokkan pelanggan ke dalam kelas *churn* dan *non-churn* berdasarkan *attribut* tertentu [3], [4].

Berbagai algoritma klasifikasi telah diterapkan dalam penelitian *customer churn*, diantaranya *Decision Tree* dan *Naive Bayes*. *Decision Tree* banyak digunakan karena menghasilkan model yang mudah dipahami dan mampu menunjukkan atribut yang berpengaruh terhadap keputusan *churn*. *Decision Tree* mampu memberikan performa yang cukup baik dalam memprediksi *churn* pada sektor perbankan dan layanan telekomunikasi [1], [5]. Sementara itu, *Naive Bayes* merupakan algoritma *probabilistik* yang menggunakan Teorema Bayes dengan asumsi independensi antar atribut. Meskipun memiliki asumsi yang sederhana, *Naive Bayes* telah terbukti efektif dalam berbagai studi prediksi *churn* pelanggan [6], [7], [8].

Namun, salah satu tantangan utama dalam penelitian *customer churn* adalah kondisi data yang tidak seimbang (*imbalanced* data), dimana jumlah pelanggan *non-churn* umumnya lebih dominan dibandingkan pelanggan *churn*. Ketidakseimbangan data ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas dan menurunkan kemampuan model dalam mendeteksi pelanggan *churn* sebagai kelas minoritas [9], [10]. Oleh karena itu, diperlukan teknik khusus untuk menangani permasalahan data tidak seimbang agar model klasifikasi dapat bekerja secara optimal.

Salah satu teknik yang banyak digunakan untuk mengatasi ketidakseimbangan data adalah *SMOTE*. Teknik ini bekerja dengan menghasilkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang. Beberapa penelitian menunjukkan bahwa penerapan *SMOTE* dapat meningkatkan kinerja algoritma klasifikasi dalam prediksi *churn* maupun permasalahan klasifikasi lainnya [11], [12]. Selain itu, *SMOTE* efektif digunakan untuk meningkatkan performa *Naive Bayes* pada data tidak seimbang meskipun diterapkan pada domain yang berbeda [13].

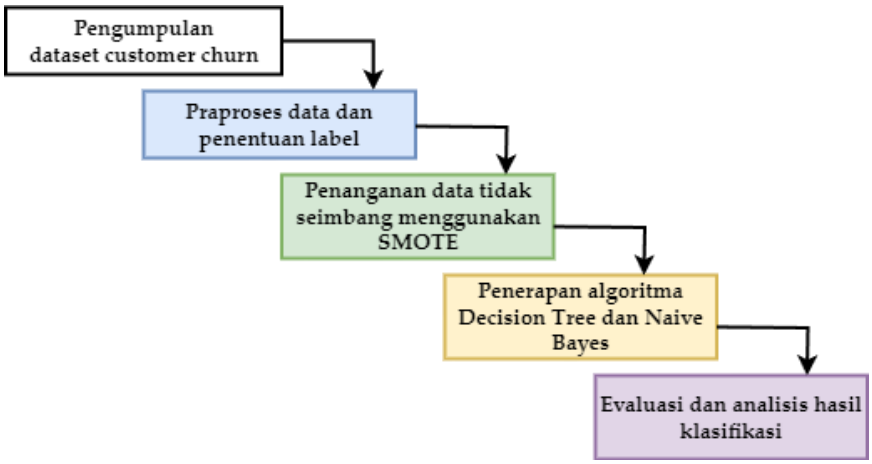
Meskipun telah banyak penelitian yang membahas prediksi *customer churn* dengan berbagai algoritma dan teknik penanganan data tidak seimbang. Hasil perbandingan kinerja antara *Decision Tree* dan *Naive Bayes* masih bervariasi tergantung pada karakteristik *dataset*, tahapan *preprocessing*, serta metode evaluasi yang digunakan [4], [14]. Oleh karena itu, diperlukan penelitian lebih lanjut untuk membandingkan kedua algoritma tersebut pada *dataset* dan skenario yang berbeda.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan kinerja algoritma *Decision Tree* dan *Naive Bayes* dalam memprediksi *customer churn* berdasarkan pola penggunaan layanan. Dataset yang digunakan merupakan *Customer Subscription Churn and Usage Patterns* dengan jumlah 2.801 data, yang diolah menggunakan perangkat lunak *RapidMiner*. Untuk mengatasi permasalahan data tidak seimbang, penelitian ini menerapkan metode *SMOTE* sebelum proses pemodelan. Evaluasi kinerja model dilakukan menggunakan teknik *cross validation* dengan *metrik* akurasi, *precision*, *recall*, *F-measure*, dan *Area Under Curve* (AUC). Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pemilihan algoritma klasifikasi yang tepat untuk prediksi *customer churn* serta menjadi referensi bagi penelitian selanjutnya di bidang sistem informasi dan data *mining*.

METODE

1. Tahap Penelitian

   Penelitian ini dilakukan melalui tahapan yang disusun secara sistematis agar proses penelitian berjalan terstruktur dan hasil yang diperoleh dapat dipertanggungjawabkan secara ilmiah. Pendekatan bertahap dalam penelitian data mining dan *prediksi customer churn* telah banyak digunakan pada penelitian sebelumnya karena mampu meningkatkan kejelasan alur penelitian dan reprodusibilitas hasil [2], [4].



Gambar 1. Alur Tahapan Penelitian

2. Dataset Penelitian

   Dataset yang digunakan dalam penelitian ini adalah *Customer Subscription Churn and Usage Patterns*, yang berisi data pelanggan beserta status *churn*. Dataset terdiri dari 2.801 data pelanggan dengan dua kelas label, yaitu *Yes (churn)* dan *No (non-churn)*. Atribut dalam dataset merepresentasikan pola penggunaan layanan pelanggan yang umum digunakan dalam penelitian prediksi churn [1], [3]. https://www.kaggle.com/datasets/jayjoshi37/customer-subscription-churn-and-usage-patterns .

3. Karakteristik dan Distribusi Dataset

   Analisis karakteristik dataset dilakukan untuk memahami struktur data dan atribut yang digunakan dalam penelitian. Informasi ini penting untuk memastikan bahwa atribut yang digunakan relevan terhadap proses prediksi *customer churn* [15].

Tabel 1. Karakteristik Dataset *Customer Churn*

No.	Nama Atribut	Tipe Data
1.	Customer_ID	Nominal
2.	Subscription_Type	Nominal
3.	Tenure	Numerik
4.	Usage_Duration	Numerik
5.	Monthly_Charge	Numerik
6.	Total_Charge	Numerik
7.	Contract_Type	Nominal
8.	Payment_Method	Nominal
9.	Customer_Service_Interaction	Numerik
10.	Churn	Nominal (Label)

Hasil analisis distribusi kelas menunjukkan bahwa dataset memiliki ketidakseimbangan jumlah data antara kelas churn dan non-churn. Kondisi ini merupakan permasalahan umum dalam

kasus *customer churn* dan dapat mempengaruhi performa model klasifikasi [9], [10].

Tabel 2. Distribusi Kelas Dataset *Customer Churn* (Sebelum SMOTE)

Kelas Churn	Jumlah Data	Persentase
Yes (Churn)	1.195	42,7%
No (Non-Churn)	1.606	57,3%
Total	**2.801**	**100%)**

4. Praproses Data
   Tahap praproses data dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses klasifikasi. Pada tahap ini, atribut *churn* ditetapkan sebagai label, sedangkan atribut lainnya digunakan sebagai *variabel prediktor*. Penyesuaian tipe data dilakukan agar sesuai dengan kebutuhan algoritma klasifikasi yang digunakan [8].
5. Penanganan Data Tidak Seimbang Menggunakan *SMOTE*
   Dataset *customer churn* umumnya memiliki permasalahan ketidakseimbangan kelas, dimana jumlah data *churn* lebih sedikit dibandingkan data *non-churn*. Kondisi ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas [9], [10].
   Untuk mengatasi permasalahan tersebut, penelitian ini *SMOTE*. Beberapa penelitian menunjukkan bahwa *SMOTE* mampu meningkatkan performa algoritma klasifikasi pada kasus *customer churn* dan data tidak seimbang [11], [14]. Oleh karena itu, *SMOTE* diterapkan sebelum proses pelatihan model klasifikasi pada
6. Algoritma Klasifikasi
   a. *Decision Tree*
      Decision Tree merupakan algoritma klasifikasi yang membentuk model prediksi dalam bentuk struktur pohon keputusan. Algoritma ini dipilih karena kemampuannya dalam menghasilkan model yang mudah dipahami dan diinterpretasikan. Penelitian sebelumnya menunjukkan bahwa *Decision Tree* efektif digunakan dalam prediksi *customer churn* pada berbagai sektor layanan [1], [5].
   b. Naive Bayes
      Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang menggunakan *Teorema Bayes* dengan asumsi independensi antar atribut. Algoritma ini memiliki kompleksitas komputasi yang rendah dan mampu memberikan hasil prediksi yang kompetitif. Beberapa penelitian membuktikan bahwa *Naive Bayes* efektif digunakan dalam prediksi *customer churn*, khususnya ketika dikombinasikan dengan teknik penanganan data tidak seimbang [6], [7], [8].
   c. Evaluasi Model
      Evaluasi performa model dilakukan menggunakan *metrik accuracy, precision, recall, F-measure,* dan AUC. Penggunaan berbagai metrik evaluasi diperlukan karena *accuracy* saja tidak cukup untuk merepresentasikan performa model pada dataset tidak seimbang [10], [4].
      Proses evaluasi dilakukan menggunakan teknik *cross-validation* untuk menghasilkan evaluasi model yang lebih stabil dan tidak bergantung pada satu pembagian data latih dan data uji.

HASIL DAN PEMBAHASAN

Penelitian ini membandingkan performa algoritma *Naive Bayes* dan *Decision Tree* dalam memprediksi *customer churn* berdasarkan pola penggunaan layanan pelanggan. Evaluasi dilakukan

menggunakan *metrik accuracy*, *precision*, *recall*, *F-measure*, dan *AUC* untuk memperoleh gambaran performa model secara menyeluruh, khususnya pada dataset dengan distribusi kelas yang tidak seimbang.

Tabel 3. Perbandingan Performa Algoritma Naive Bayes dan Decision Tree

Algoritma	Accuracy (%)	Precision(%)	Recall(%)	F-Measure(%)	AUC
Naïve Bayes	65,04	58,21	64,10	61,01	0,704
Decision Tree	62,68	55,01	68,87	61,17	**0,737**

Berdasarkan hasil evaluasi yang ditampilkan pada Tabel 3, algoritma *Naive Bayes* menghasilkan nilai ***accuracy* sebesar 65,04%,** sedangkan algoritma *Decision Tree* memperoleh nilai ***accuracy* sebesar 62,68%**. Hasil ini menunjukkan bahwa *Naive Bayes* memiliki kemampuan klasifikasi yang sedikit lebih baik dalam memprediksi data secara keseluruhan. Pada ***metrik precision***, *Naive Bayes* memperoleh **nilai 58,21%**, sementara *Decision Tree* menghasilkan **nilai 55,01%**. Nilai ini menunjukkan bahwa *Naive Bayes* lebih konsisten dalam memprediksi pelanggan yang benar-benar termasuk ke dalam kelas *churn* maupun *non-churn*.

Sementara itu, algoritma *Decision Tree* menunjukkan nilai ***recall*** yang lebih tinggi, **yaitu 68,87%**, dibandingkan *Naive Bayes* yang memperoleh **nilai 64,10%**. Hal ini mengindikasikan bahwa *Decision Tree* lebih mampu mengidentifikasi pelanggan yang benar-benar mengalami *churn*. Pada ***metrik F-measure***, kedua algoritma menunjukkan performa yang relatif seimbang, dengan **nilai 61,01%** untuk *Naive Bayes* dan **61,17%** untuk *Decision Tree*. Perbedaan yang sangat kecil ini menunjukkan bahwa kedua algoritma memiliki keseimbangan yang hampir sama antara ***precision*** dan ***recall***. Dari sisi *AUC*, algoritma *Decision Tree* menghasilkan **nilai 0,737**, lebih tinggi dibandingkan *Naive Bayes* dengan **nilai 0,704**. Nilai *AUC* yang lebih tinggi menunjukkan bahwa *Decision Tree* memiliki kemampuan pemisahan kelas yang lebih baik antara pelanggan *churn* dan *non-churn*.

Hasil penelitian menunjukkan bahwa kedua algoritma yang digunakan memiliki keunggulan masing-masing dalam memprediksi *customer churn*. *Naive Bayes* menunjukkan performa yang lebih baik pada *matrix accuracy* dan *precision*, sehingga cocok digunakan apabila tujuan utama adalah memperoleh hasil klasifikasi yang stabil secara keseluruhan.

Di sisi lain, *Decision Tree* menunjukkan performa yang lebih unggul pada *metrik recall* dan *AUC*. Hal ini menunjukkan bahwa algoritma ini lebih efektif dalam mendeteksi pelanggan yang berpotensi mengalami *churn*, yang sangat penting dalam konteks strategi retensi pelanggan. Kemampuan *Decision Tree* dalam membentuk pola keputusan berbasis atribut memungkinkan model untuk menangkap hubungan kompleks antara *variabe*l yang mempengaruhi *churn*.

Penerapan teknik *SMOTE* pada penelitian ini juga berkontribusi dalam meningkatkan performa kedua algoritma, khususnya pada *metrik recall* dan *AUC*. Dengan distribusi kelas yang lebih seimbang, model klasifikasi dapat mempelajari pola data minoritas dengan lebih baik, sehingga mengurangi bias terhadap kelas mayoritas.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pemilihan algoritma klasifikasi untuk prediksi *customer churn* perlu disesuaikan dengan tujuan analisis. Apabila fokus utama adalah akurasi prediksi secara umum, maka *Naive Bayes* dapat menjadi pilihan yang tepat. Namun, apabila tujuan utama adalah mendeteksi pelanggan *churn* sebanyak mungkin, *Decision Tree* lebih direkomendasikan karena memiliki nilai *recall* dan *AUC* yang lebih tinggi.

SIMPULAN DAN SARAN

Penelitian ini membandingkan kinerja algoritma *Naive Bayes* dan *Decision Tree* dalam memprediksi *customer churn* menggunakan dataset yang dianalisis melalui *RapidMiner*. Hasil penelitian menunjukkan bahwa *Naive Bayes* menghasilkan nilai *accuracy* yang sedikit lebih tinggi Sedangkan *Decision Tree* memiliki nilai *recall* yang lebih baik dalam mendeteksi pelanggan yang berpotensi *churn*, sehingga pemilihan algoritma perlu disesuaikan dengan tujuan analisis. Secara umum, kedua algoritma mampu melakukan prediksi dengan cukup baik, namun masih terdapat ruang untuk peningkatan

performa. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar, menerapkan teknik penanganan data tidak seimbang, serta membandingkan dengan algoritma lain atau metode *ensemble* agar diperoleh hasil prediksi yang lebih optimal dan aplikatif.

DAFTAR PUSTAKA

[1] A. F. Azmi And A. Voutama, "Komputa : Jurnal Ilmiah Komputer Dan Informatika Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest Dan Decision Tree Dengan Evaluasi Confusion Matrix Komputa : Jurnal Ilmiah Komputer Dan Informatika," Vol. 13, No. 1, 2024, [Online]. Available: Https://Doi.Org/10.34010/Komputa.V13i1.12639

[2] A. Budiyono, I. Nendi, M. Kuningan, P. Siber, And C. Internasional, "Customer Churn Prediction Uses Machine Learning To Improve Retention On Digital Platforms," Vol. 2, No. 2, Pp. 55–75, 2025, [Online]. Available: Https://Doi.Org/10.59261/Jdbs.V2i2.23

[3] S. E. Maghfira *Et Al.*, "Klasifikasi Customer Churn Pada Industri Bank," Vol. 06, No. 03, Pp. 648–655, 2025.

[4] M. Imani, M. Joudaki, And A. Beikmohammadi, "Customer Churn Prediction : A Systematic Review Of Recent Advances , Trends , And Challenges In Machine Learning And Deep Learning," Pp. 1–38, 2025, [Online]. Available: Https://Www.Mdpi.Com/2504-4990/7/3/105

[5] A. Rizky, Y. Siregar, M. Iqbal, U. Pembangunan, And P. Budi, "Prediksi Customer Churn Pada Layanan Indihome Menggunakan Algoritma Decision Tree (Studi Kasus Pt. Telkom Akses)," Vol. 4307, No. 1, Pp. 204–211, 2025, [Online]. Available: Https://Jurnal.Goretanpena.Com/Index.Php/Jssr/Article/View/2698

[6] R. Alfarez And V. Purwayoga, "Penerapan Naïve Bayes Untuk Prediksi Customer Churn (Studi Kasus : Pt Hutchison 3 Indonesia)," Vol. 05, No. 02, Pp. 301–307, 2024, [Online]. Available: Https://Doi.Org/10.30998/Jrami.V5i2.8556

[7] M. Clementine And A. Wahyuniati, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes Dan," Vol. 17, No. 1, Pp. 9–18, 2022, [Online]. Available: Https://Doi.Org/10.33998/Processor.2022.17.1.1170

[8] I. I. Daipah, R. Astuti, And W. Prihartono, "Prediksi Churn Pelanggan Pada Layanan Desain Grafis Home Desain Menggunakan Algoritma Naïve Bayes," Vol. 13, No. 1, Pp. 1022–1028, 2025, [Online]. Available: Https://Doi.Org/10.23960/Jitet.V13i1.5811

[9] T. Kimura, "Customer Churn Prediction With Hybrid Resampling And Ensemble Learning," Vol. 25, No. 1, Pp. 1–23, 2022.

[10] H. Kaur, "A Systematic Review On Imbalanced Data Challenges In Machine Learning : Applications And Solutions A Systematic Review On Imbalanced Data Challenges In Machine Learning : Applications And Solutions," Vol. 52, No. 4, 2026, Doi: 10.1145/3343440.

[11] C. Mulia, A. Kurniasih, P. Studi, I. Komputer, And C. Timur, "Teknik Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Bank Customer Churn Menggunakan Algoritma Naïve Bayes Dan Logistic Regression," Vol. 0, Pp. 552–559, 2023, [Online]. Available: Https://Conference.Upnvj.Ac.Id/Index.Php/Senamika/Article/View/2590

[12] F. S. Pratiwi *Et Al.*, "Implementasi Metode Smote Dan Random Over- Sampling Pada Algoritma Machine Learning Untuk," Vol. 8, No. 1, Pp. 87–98, 2025, [Online]. Available: Https://Doi.Org/10.47080/Simika.V8i1.3678

[13] M. P. Pulungan, A. Purnomo, A. Kurniasih, P. Korespondensi, I. Class, And S. M. O. Technique, "Penerapan Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Kepribadian Mbti Menggunakan Naive Bayes Application Of Smote To Overcome Class Imbalance In The Mbti Personality Classification Using The Naïve Bayes Classifier," Vol. 11, No. 5, Pp. 1033–1042, 2024, Doi: 10.25126/Jtiik.2024117989.

[14] R. Suguna, J. S. Prakash, H. A. Pai, T. R. Mahesh, And V. V. Kumar, "Mitigating Class Imbalance In Churn Prediction With Ensemble Methods And Smote," Pp. 1–20, 2025, [Online]. Available: Https://Www.Nature.Com/Articles/S41598-025-01031-0

[15] Z. Nur Saputra And Z. Fatah, "Pengunaan Data Minig Untuk Mengidentifikasi Pelanggan Beresiko Tinggi," Vol. 2, No. 1, Pp. 46–51, 2025, [Online]. Available: Https://Doi.Org/10.69714/S91z1k09