

# Teknik SMOTE untuk Mengatasi *Imbalance Class* dalam Klasifikasi *Bank Customer Churn* Menggunakan Algoritma *Naïve bayes* dan *Logistic Regression*

Chiekal Mulia<sup>1</sup>, Aliyah Kurniasih<sup>2</sup>

Program Studi Ilmu Komputer<sup>1,2</sup>

Sekolah Tinggi Ilmu Manajemen dan Ilmu Komputer ESQ<sup>1,2</sup>

JL TB Simatupang, Cilandak Timur, Kec. Ps. Minggu, Kota Jakarta Selatan 12560

[chiekal.mulia@students.esqbs.ac.id](mailto:chiekal.mulia@students.esqbs.ac.id)<sup>1</sup>, [aliyah.kurniasih@esqbs.ac.id](mailto:aliyah.kurniasih@esqbs.ac.id)<sup>2</sup>

**Abstrak.** *Churn* nasabah menjadi permasalahan serius bagi *Bank* karena kehilangan nasabah berarti kehilangan pendapatan dan potensi kerugian jangka panjang. Untuk mencegah hal tersebut, diperlukannya teknologi masa kini untuk mengklasifikasi nasabah yang akan pergi (*churn*) atau bertahan melalui pemodelan *machine learning* algoritma *Naïve Bayes Classifier* dan algoritma *Logistic Regression*. Model klasifikasi dibuat untuk mempertahankan nasabah *Bank* berdasarkan data nasabah. Teknik SMOTE digunakan untuk mengatasi masalah *imbalance class* pada data yang digunakan, dan model di evaluasi dengan nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Model terbaik yang dihasilkan ketika menggunakan teknik SMOTE ternyata mengalami penurunan nilai akurasi sebesar 13% pada algoritma *Gaussian Naïve Bayes*, model tanpa teknik SMOTE dengan akurasi sebesar 82% dan dengan teknik SMOTE sebesar 69%. *Gaussian Naïve Bayes* berhasil mengungguli jika dibandingkan dengan *Complement Naïve Bayes* dan *Logistic Regression*.

**Kata Kunci:** klasifikasi, *data mining*, SMOTE, *logistic regression*, *naïve bayes*.

## 1 Pendahuluan

Industri perbankan dan jasa keuangan telah mengalami perkembangan pesat belakangan ini karena adanya kemajuan teknologi dan kebijakan deregulasi. Deregulasi finansial telah menciptakan fenomena baru yang meningkatkan persaingan, termasuk persaingan di sektor perbankan. Persaingan ini disebabkan oleh pertumbuhan setiap *Bank*. Pertumbuhan *Bank – Bank* ini sangat dipengaruhi oleh kemampuan mereka dalam mengumpulkan dana dari masyarakat, baik dalam jumlah kecil maupun besar [1]. Dibalik itu semua ada satu keadaan, dimana nasabah *Bank* menghentikan hubungan atau layanan dengan *Bank* tersebut, istilah ini disebut *Churn*. *Churn* nasabah menjadi permasalahan serius bagi *Bank* karena kehilangan nasabah berarti kehilangan pendapatan dan potensi kerugian jangka panjang. Faktor – faktor yang dapat menyebabkan *churn* nasabah antara lain kekecewaan terhadap pelayanan atau produk, ketidakpuasan, kurangnya keuntungan atau manfaat yang dirasakan, perubahan kebutuhan atau preferensi nasabah, serta penawaran yang lebih baik dari pesaing di pasar.

Untuk mencegah hal tersebut dapat memanfaatkan teknologi masa kini dengan membuat model prediksi menggunakan *Data Mining*. Seperti pada penelitian terdahulu yang dilakukan oleh Aksama, dkk [1] memprediksi *Churn* Nasabah *Bank* dengan algoritma *Naïve Bayes* dan ID3 pada 10.000 data, model yang mencapai hasil optimal yaitu *Naïve Bayes* dengan akurasi sebesar 85,17%, sementara penerapan metode ID3 hanya menghasilkan akurasi sekitar 79,17%, dimana pembagian data dengan perbandingan 6:4 data latih dan data uji. Penelitian selanjutnya yang dilakukan oleh Fitriana, dkk [2] juga memprediksi Nasabah yang berpotensi *churn* menggunakan Algoritma Klasifikasi *Naïve Bayes* mendapatkan hasil akurasi sebesar 86,17% dengan tingkat kesalahan mencapai 13,83%. Penelitian berikutnya yang dilakukan oleh Irmanda, dkk [3], yang membandingkan metode jaringan syaraf tiruan dan *decision tree* untuk prediksi *churn*, menghasilkan bahwa jaringan saraf tiruan memiliki akurasi yang lebih unggul dibandingkan dengan *decision tree* yaitu mencapai 86%, *precision* 71%, akan tetapi *recall* yang dihasilkan *decision tree* menunjukkan kinerja yang lebih baik dibandingkan jaringan saraf tiruan sebesar 58%.

Pada penelitian ini menggunakan data pelanggan pemegang rekening multinational *Bank* yang bertujuan untuk mengklasifikasikan nasabah yang akan meninggalkan *Bank* (*churn*) atau bertahan. *Churn* memungkinkan perusahaan mengembangkan program loyalitas dan kampanye retensi untuk mempertahankan pelanggan sebanyak mungkin Dimana pada data yang digunakan terjadi sebuah permasalahan *imbalance class*, yaitu data dengan kelas 0 (bertahan) mendominasi data terbanyak dan kelas 1 (*churn*) mendominasi data yang sedikit.

Algoritma yang digunakan untuk membuat model klasifikasi *machine learning* yaitu algoritma dari keluarga *Naïve Bayes Classifier* diantaranya *Gaussian Naïve Bayes*, *Complement Naïve Bayes*, dan *Logistic Regression*. Dan untuk mengatasi permasalahan pada *imbalance class* menggunakan Teknik SMOTE. Konsep yang dibangun yaitu membuat model dengan tidak menggunakan pra-proses data teknik SMOTE, kemudian dibandingkan dengan yang telah menggunakan SMOTE pada ketiga algoritma, dan yang terakhir membandingkan dari ketiga algoritma tersebut.

## 2 Dasar Teori

### 2.1 Bank customer

Dalam Undang-Undang Nomor 10 Tahun 1998 tentang Perubahan Atas Undang-Undang Nomor 7 Tahun 1992 tentang Perbankan, terdapat dua jenis nasabah, yaitu nasabah penyimpan dan nasabah debitur. Nasabah penyimpan merujuk pada nasabah yang menempatkan dananya di *Bank* dalam bentuk simpanan berdasarkan kesepakatan antara *Bank* dan nasabah tersebut. Sementara itu, nasabah debitur adalah nasabah yang mendapatkan fasilitas kredit atau pembiayaan sesuai dengan prinsip syariah atau setara dengannya, berdasarkan kesepakatan antara *Bank* dan nasabah yang bersangkutan [4].

### 2.2 Churn

*Churn* nasabah terjadi ketika nasabah memutuskan untuk tidak menggunakan sebagai pengguna *Bank* tersebut lagi. Fenomena ini dikenal dengan istilah *churn* karena nasabah memutuskan untuk mengakhiri hubungan dengan *Bank* tertentu dan mencari pelayanan di *Bank* lain. Hal ini menunjukkan bahwa nasabah merasa bahwa lebih mendapatkan keuntungan menggunakan *Bank* yang lain. Oleh karena itu, *churn* nasabah perlu ditangani dengan cepat untuk menghindari dampak yang signifikan bagi *Bank*. Semakin banyak nasabah yang melakukan *churn*, *Bank* harus mengevaluasi layanan yang mereka berikan. Jelas bahwa memperoleh nasabah baru akan memerlukan biaya yang lebih tinggi daripada mempertahankan nasabah yang sudah ada [5].

### 2.3 Naïve Bayes

*Naive Bayes* merupakan metode klasifikasi yang mengestimasi probabilitas dengan menganalisis frekuensi dan kombinasi nilai dalam suatu set data [4]. Ilmuwan Inggris Thomas Bayes adalah tokoh yang memperkenalkan algoritma *Naive Bayes*, yang bekerja dengan memproyeksikan peluang-peluang di masa mendatang berdasarkan pengalaman sebelumnya, sehingga sering kali disebut sebagai *Teorema Bayes*. Keunggulan dalam penerapan *Naive Bayes* terletak pada kemampuannya menggunakan data latihan yang relatif sedikit untuk mengatur parameter-parameter yang diperlukan dalam proses klasifikasi. Algoritma ini mengasumsikan independensi variabel, sehingga hanya varians yang terkait dengan variabel dalam sebuah kelas yang menjadi fokus dalam menentukan klasifikasi, dan tidak perlu memperhitungkan seluruh matriks kovarians. *Teorema Bayes*, yang menghitung probabilitas suatu kejadian berdasarkan kejadian lain yang sudah terjadi, diekspresikan secara matematis dalam persamaan berikut:

$$P(B) = \frac{P(A)P(B|A)}{P(B)} \quad (1)$$

$P(A|B)$  = Probabilitas bahwa A akan terjadi ketika kejadian B terjadi.

$P(A)$  = Probabilitas terjadinya kejadian A.

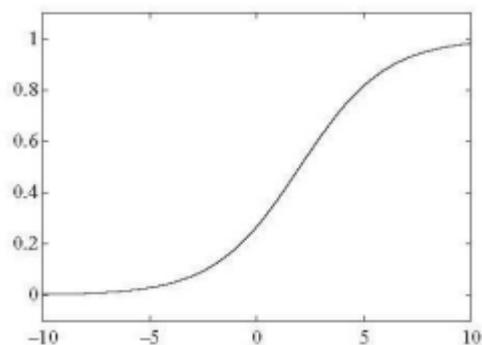
$P(B|A)$  = Probabilitas bahwa kejadian B terjadi ketika kejadian A terjadi

$P(B)$  = Probabilitas terjadinya kejadian B.

### 2.4 Logistic Regression

Algoritma *Data Mining* yang cukup terkenal adalah *logistic regression*. Secara banyak hal, *regresi linier* dan *logistic regression* memiliki kesamaan. Namun, perbedaan utama terletak pada penggunaan keduanya. Algoritma *regresi linier* digunakan untuk memprediksi atau memperkirakan nilai, sedangkan *logistic regression* digunakan untuk tugas klasifikasi [6]. Kemudian ada juga yang berpendapat Algoritma *logistic regression* adalah

salah satu algoritma *machine learning* yang populer digunakan dalam masalah klasifikasi. Algoritma ini merupakan metode analisis prediktif yang berbasis pada konsep probabilitas [7].



**Gambar 1.** Gambar diatas merupakan contoh dari algoritma *Logistic Regression*

## 2.5 SMOTE

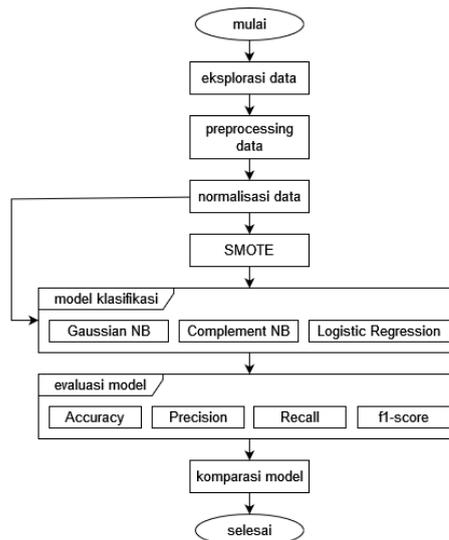
SMOTE adalah sebuah teknik yang digunakan untuk mengimbangi jumlah distribusi sampel data pada kelas minoritas dengan memilih sampel data tersebut sehingga jumlahnya seimbang dengan jumlah sampel data kelas mayoritas. Dengan menggunakan metode ini, data pada kelas minoritas dapat di-"*oversampling*" sehingga dataset menjadi lebih seimbang antara kelas mayoritas dan kelas minoritas [8]. Pada algoritma SMOTE, langkah pertama adalah mengambil selisih antara vektor fitur dari sampel pada kelas minoritas dengan nilai vektor tetangga terdekat dari kelas minoritas. Selanjutnya, selisih tersebut dikalikan dengan angka acak antara 0 dan 1. Kemudian, hasil perkalian tersebut ditambahkan dengan vektor fitur asli untuk mendapatkan nilai vektor baru yang disintesis. Dengan cara ini, SMOTE dapat menghasilkan sampel baru yang mewakili variasi dalam kelas minoritas [9].

## 2.6 Data Mining

*Data Mining* merupakan menggali pengetahuan untuk mengekstrak informasi yang tersembunyi dari data kemudian menganalisis data untuk mencari pola dalam data dengan tujuan untuk mendapatkan pengetahuan dari data. *Data Mining* diperlukan ketika ukuran data yang semakin besar secara eksponensial ataupun data yang semakin heterogen dan kompleks. Bisa digunakan ketika untuk mencari pola-pola yang dapat di pahami manusia yang menjelaskan karakteristik data (secara deskripsi), juga membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi (secara prediksi).

## 3 Metode Penelitian

Metode penelitian diperlukan untuk dapat melaksanakan penelitian dengan sistematis. Gambar 1 menyajikan metode yang dilakukan untuk mendapatkan hasil model klasifikasi terbaik. Dataset dan metode – metode tersebut dijelaskan sebagai berikut:



**Gambar 2.** Alur Metode Penelitian

### 3.1 Data Penelitian

Proses pencarian data pada yaitu menggunakan data publik yang didapatkan dari situs Kaggle[10]. Dataset tersebut berisi data nasabah yang pada *Bank* multinasional yang bertujuan untuk mengetahui apa yang menyebabkan keputusan nasabah untuk keluar dari *Bank* tersebut (*Churn*). Dataset tersebut diperbarui pada bulan Maret tahun 2023. Detail dari setiap atribut pada dataset disajikan pada Tabel 1. Dibawah ini.

**Tabel 1.** Deskripsi Atribut Dataset

Atribut	Keterangan
customer_id	Nomor akun
credit_score	Nilai kredit
country	Negara tempat tinggal
gender	Jenis kelamin
age	Usia
tenure	Sejak berapa tahun dia mempunyai rekening Bank di Bank tersebut
balance	Saldo rekening
products_number	Jumlah produk dari Bank
credit_card	Apakah pelanggan ini memiliki kartu kredit?
active_number	Apakah dia anggota Bank yang aktif?
estimated_salary	Gaji pemegang rekening
churn	Status churn

### 3.2 Data Penelitian

Pada tahap ini proses yang dilakukan diantaranya yaitu mengecek jumlah data dan jumlah atribut, mengecek tipe data dan isi data, mengecek unik data dari atribut target model klasifikasi ‘churn’ dan distribusi jumlah datanya, mengecek data yang *missing values* dan duplikat.

### 3.3 Preprocessing Data

Pemrosesan data adalah langkah untuk mengatasi beberapa masalah yang mungkin terjadi pada data yang dapat mengganggu *training* model. Pemrosesan data akan membantu meningkatkan presisi dan kinerja *data mining* serta mencegah kesalahan dalam *data mining*. Tujuannya adalah untuk memudahkan pemahaman data sehingga mempermudah pemilihan teknik dan metode *data mining* yang sesuai. Pada tahap ini proses yang dilakukan yaitu *feature selection* dengan menghapus atribut yang dirasa tidak memberikan pengaruh terhadap target model klasifikasi, dan melakukan pembagian data uji 20% dan data latih 80% dari total data. Dimana data latih digunakan untuk *training* model klasifikasi dan data uji digunakan untuk melakukan evaluasi model klasifikasi.

### 3.4 Normalisasi Data

Untuk menghindari bias dalam proses *data mining*, perlu melakukan normalisasi atau distandarisasi terhadap nilai-nilai atribut data yang memiliki rentang nilai yang berbeda-beda. Normalisasi dilakukan dengan mengubah rentang nilai menjadi lebih kecil, misalnya [0, 1] atau [-1, 1]. Hal ini dilakukan agar semua atribut memiliki bobot yang sama. Normalisasi yang digunakan pada penelitian ini menggunakan MinMaxScaler mengubah variabel numerik ke dalam rentang tertentu antara 0 dan 1.

### 3.5 SMOTE

SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan metode yang digunakan untuk mensintesis *instance* minoritas baru di antara *instance* minoritas yang sudah ada. Metode ini menghasilkan catatan pelatihan virtual dengan menggunakan interpolasi linear untuk kelas minoritas. Penerapan SMOTE yang dilakukan pada penelitian ini hanya pada data latih.

### 3.6 Klasifikasi Model

Algoritma yang digunakan pada yaitu *Gaussian Naïve Bayes*, *Complement Naïve Bayes*, dan *Logistic Regression*. Dimana algoritma ini digunakan karena pada dasarnya algoritma tersebut bekerja untuk klasifikasi *biner* atau dua kelas, hal ini sesuai dengan dataset yang digunakan yaitu memprediksi nasabah yang pergi atau tidak.

### 3.7 Evaluasi Model

Evaluasi model dilakukan untuk tujuan mengukur kinerja model klasifikasi secara keseluruhan, dengan menggunakan *metrix evaluation* yaitu *accuracy*, serta *precision*, *recall*, *f1-score* pada *average macro*.

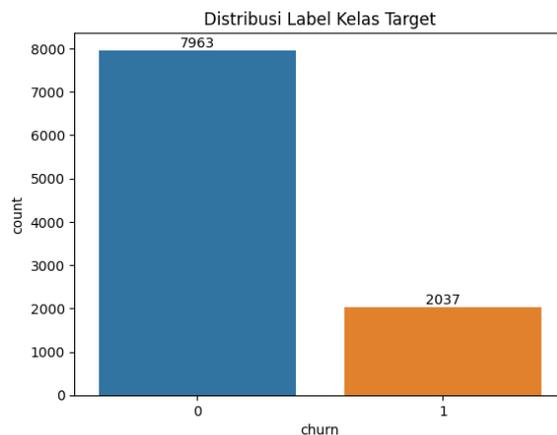
## 4 Hasil Penelitian

### 4.1 Hasil Data Eksplorasi

Hasil dari tahap eksplorasi data bahwa terdapat jumlah data sebanyak 10.000 *record* dan 12 atribut. Tabel 2 merupakan tipe data dari setiap atribut dan adanya isi data yang sesuai dengan tipe data. Unik data dari atribut target model klasifikasi 'churn' yaitu label 0 (tidak pergi atau bertahan) dan label 1 (pergi atau *churn*). Gambar 2 distribusi jumlah berdasarkan target, dari hasil tersebut telah terdapat adanya data *imbalance class*. kemudian tidak terdapat *missing values* dan duplikat data.

**Tabel 2.** Tipe Data Atribut

Tipe Data	Atribut
integer	customer_id, credit_score, age, tenure, product_number, credit_card, active_member, churn
float	balance, estimated_salary
object (string kategorikal)	country, gender



**Gambar 3.** Distribusi Target

## 4.2 Hasil *Preprocessing* Data

Hasil dari tahap *preprocessing* data dengan *feature selection* yaitu menghapus 4 atribut yang merupakan atribut demografis pelanggan diantaranya atribut 'customer\_id', 'age', 'country', dan 'gender'. Sehingga atribut yang tersisa tinggal 8, yaitu 7 atribut untuk *input* model diantaranya 'credit\_score', 'tenure', 'product\_number', 'credit\_card', 'active\_number', 'balance', 'estimated\_salary', dan atribut target model yaitu 'churn'. Tabel 3 data latih 80% yang dihasilkan sebanyak 8.000 data, dan data uji 20% yang dihasilkan sebanyak 2.000 data. Dimana komponen X merupakan input dari atribut model (7 atribut), dan komponen y merupakan output dari atribut target model (1 atribut).

**Tabel 3.** Hasil Pembagian Data

Keterangan	Komponen X	Komponen y
Data latih	(8.000, 7)	(8.000, 1)
Data uji	(2.000, 7)	(8.000, 1)

## 4.3 Hasil SMOTE

Tabel 4 merupakan hasil dari pada proses melakukan teknik SMOTE pada data latih, yaitu data dengan jumlah kelas minoritas label 1 (*churn*) dari jumlah awal sebanyak 1.632 data setelah di SMOTE menghasilkan data sintesis sebanyak 6.268 data mengikuti banyaknya jumlah data dengan label kelas mayoritas yaitu 0 (bertahan) sebanyak 6.368. Sehingga data input model dan output model yang dihasilkan menjadi sebanyak 12.736 data.

**Tabel 4.** Hasil SMOTE Data Latih

Label	Sebelum	Sesudah
0 (bertahan)	6.368	6.368
1 (churn)	1.632	6.368

## 4.4 Hasil Evaluasi Model

Tabel 5 merupakan hasil dari model klasifikasi yang dibuat ketika tidak menggunakan pemrosesan data dengan SMOTE, dihasilkan model terbaik pada algoritma *Gaussian Naïve Bayes* dengan tingkat akurasi sebesar 82%.

**Tabel 5.** Hasil Evaluasi Model Tanpa Smote

Model	Accuracy	Precision	Recall	F1-Score
GaussianNB	82%	84%	56%	56%

ComplementNB	58%	57%	61%	53%
Logistic Regression	80%	90%	50%	44%

Tabel 6 merupakan hasil dari model klasifikasi yang dibuat ketika menggunakan pemrosesan data teknik SMOTE, dihasilkan model terbaik pada algoritma *Gaussian Naïve Bayes* dengan tingkat akurasi sebesar 69%.

**Tabel 6.** Hasil Evaluasi Model Dengan SMOTE

Model	Accuracy	Precision	Recall	F1-Score
GaussianNB	69%	61%	66%	61%
ComplementNB	58%	57%	61%	54%
Logistic Regression	64%	58%	62%	57%

Berdasarkan dari hasil tersebut yaitu ketika model dibangun dengan menggunakan teknik SMOTE menghasilkan tingkat penurunan akurasi pada algoritma *Gaussian Naïve Bayes* sebesar 13%, dan algoritma *Logistic Regression* sebesar 16% dibandingkan ketika tidak dengan menggunakan teknik SMOTE. Penurunan ini terjadi karena sampel sintetis pada kelas minoritas yang dihasilkan berada di dekat atau bahkan terdekat dengan sampel – sampel dari kelas mayoritas dan hal ini dapat menyebabkan peningkatan *noise* dalam data. Kemudian sampel sintetis yang dihasilkan juga bisa menjadi lebih kompleks dan *overfit* ke data latih dan mengurangi kemampuan model untuk melakukan generalisasi pada data yang belum pernah dilihat sebelumnya yaitu data uji. Sehingga hal ini yang dapat menyebabkan penurunan tingkat akurasi ketika menggunakan teknik SMOTE. Akan tetapi pada algoritma Complement Naïve Bayes nilai akurasi yang dihasilkan adalah tetap atau sama yaitu sebesar 58% baik itu ketika menggunakan teknik SMOTE dan ketika tidak menggunakan teknik SMOTE. Hal ini terjadi karena algoritma Complement Naïve Bayes merupakan model yang relatif sederhana dan memiliki sifat yang membuatnya kurang sensitif terhadap perubahan dalam distribusi data sintesis yang dihasilkan.

Selanjutnya diantara ketiga algoritma tersebut bahwa membuktikan hasil model terbaik pada algoritma Gaussian Naïve Bayes sebesar 82% tanpa SMOTE jika dibandingkan dengan algoritma Complement Naïve Bayes dan Logistic Regression baik itu dengan SMOTE atau tanpa SMOTE. Hal ini terjadi karena Gaussian Naïve Bayes memiliki asumsi bahwa fitur-fitur adalah independen secara kondisional yang berarti bahwa algoritma ini dapat lebih tahan terhadap fitur-fitur yang mungkin tidak relevan dengan label.

## 5 Kesimpulan

Model terbaik yang dihasilkan untuk klasifikasi data nasabah yang akan bertahan atau pergi (*churn*) walaupun mengandung data yang *imbalance class* yaitu pada algoritma *Gaussian Naïve Bayes* ketika tidak menggunakan teknik SMOTE dalam mengatasi ketidak seimbangan kelas dengan akurasi 82%. Berdasarkan hasil ini, teknik SMOTE dengan algoritma keluarga *Naïve Bayes Classifier* tidak bagus untuk digunakan pada data yang memiliki distribusi data minoritas yang terdekat dengan sampel – sampel pada data mayoritas.

Untuk penelitian kedepannya dengan dataset yang sama, beberapa hal yang dapat dilakukan yaitu menggunakan algoritma model klasifikasi lainnya seperti SVM, MLP, *Random Forest* dan lain sebagainya. Serta dilakukan pengecekan data *outlier* dan penanganannya, kemudian uji dengan melakukan teknik normalisasi data dan teknik penanganan *imbalance class* yang lainnya seperti Near Miss.

## Referensi

- [1] Miryam Clementine and Arum. (2022). Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes dan ID3. *Jurnal Processor*, vol. 17, no. 1. doi: 10.33998/processor.2022.17.1.1170.
- [2] D. Fitriana, S. Dwiasnati, H. H. H, and K. A. Baihaqi. (2021). Penerapan Metode Machine Learning untuk Prediksi Nasabah Potensial menggunakan Algoritma Klasifikasi Naïve Bayes. *Faktor Exacta*, vol. 14, no. 2, p. 92. doi: 10.30998/faktorexacta.v14i2.9297.
- [3] H. N. Irmada, R. Astriratma, and S. Afrizal. (2019). Perbandingan Metode Jaringan Syaraf Tiruan Dan Pohon Keputusan Untuk Prediksi Churn Universitas Pembangunan Nasional Veteran Jakarta. *JSI : Jurnal Sistem Informasi (E-Journal)*, vol. 11, no. 2, [Online]. Available:

- <http://ejournal.unsri.ac.id/index.php/jsi/index>
- [4] M. R. Romadhon and F. Kurniawan. (2021). A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia,” in *3rd 2021 East Indonesia Conference on Computer and Information Technology. EIconCIT 2021*, Institute of Electrical and Electronics Engineers Inc., pp. 41–44. doi: 10.1109/EIconCIT50028.2021.9431845.
- [5] K. Polat. (2019). Freezing of Gait (FoG) Detection Using Logistic Regression in Parkinson’s Disease from Acceleration Signals. *Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, IEEE.
- [6] O. Febri, R. #1, A. #2, and A. Aditsania. Handling Imbalanced Data pada Prediksi Churn menggunakan metode SMOTE dan KNN Based on Kernel Handling Imbalanced Data on Churn Prediction using SMOTE and KNN Based on Kernel Methods.
- [7] Y. Nora Marlim, L. Suryati, and N. Agustina. (2022). Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression.
- [8] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, A. Nikmatul Kasanah, U. Pujianto, T. Elektro, F. Teknik, and U. Negeri Malang. (2017). Terakreditasi SINTA Peringkat 2 Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. vol. 1, no. 3, pp. 196–201.
- [9] E. Sutoyo, M. Asri Fadlurrahman, J. Telekomunikasi Jl Terusan Buah Batu, K. Dayeuhkolot, K. Bandung, and J. Barat. JEPIN (Jurnal Edukasi dan Penelitian Informatika) Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network.
- [10] “Bank Customer Churn Dataset | Kaggle.” <https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset> (accessed Aug. 23, 2023).