

Penerapan Data Mining Dengan Algoritma SMOTE dan Naive Bayes untuk Klasifikasi Kelayakan Penerimaan Bantuan Sosial pada Kantor Camat Medan Timur

Application of Data Mining with SMOTE and Naive Bayes Algorithms for Classification of Eligibility for Social Assistance Receipts at the East Medan District Office

Dina L.Simorangkir^{a,1}, Rita Novita Sari^{b,2}

^aSistem Informasi, Universitas Potensi Utama, Medan 20241, Sumatera Utara, Indonesia

^bSistem Informasi, Universitas Islam Negeri Sumatera Utara, Deli Serdang 20353, Sumatera Utara, Indonesia
Email: simorangkirdina383@gmail.com¹, ritanovitasari@uinsu.ac.id²

ABSTRAK

Kantor Camat Medan Timur menghadapi tantangan dalam menentukan kelayakan penerima bantuan sosial akibat ketidakakuratan data, kriteria yang tidak konsisten, dan metode seleksi manual yang kurang efisien. Penelitian ini bertujuan untuk mengatasi permasalahan tersebut melalui penerapan algoritma SMOTE dan Naive Bayes. SMOTE digunakan untuk menangani ketidakseimbangan data, sementara Naive Bayes diterapkan untuk melakukan klasifikasi berdasarkan probabilitas. Proses penelitian dimulai dengan pengumpulan data penerima bantuan sosial, yang kemudian diolah menggunakan algoritma SMOTE untuk meningkatkan jumlah data dalam kategori minoritas. Selanjutnya, algoritma Naive Bayes diterapkan untuk melakukan klasifikasi kelayakan penerima berdasarkan fitur-fitur yang tersedia. Hasil dari perhitungan menunjukkan bahwa kombinasi kedua algoritma ini berhasil meningkatkan akurasi klasifikasi secara signifikan. Berdasarkan perhitungan nilai tertinggi yaitu sebesar 0,008 "Layak", sehingga data yang diuji dikategorikan Layak menerima Bantuan Sosial. Jika dipersentasekan, nilai tersebut setara dengan $0,008 \times 100 = 0,75$ atau 75%. Selain itu, sistem yang dikembangkan dalam penelitian ini dirancang untuk memudahkan implementasi di instansi pemerintah. Dengan transparansi dan efisiensi yang dihasilkan, sistem ini mampu membantu Kantor Camat Medan Timur dalam menyalurkan bantuan sosial secara lebih tepat sasaran.

Kata Kunci : Data Mining, Klasifikasi, Bantuan Sosial, SMOTE, Naive Bayes.

ABSTRACT

The East Medan Sub-district Office faces challenges in determining the eligibility of social assistance recipients due to data inaccuracy, inconsistent criteria, and inefficient manual selection methods. This study aims to address these issues through the application of the SMOTE and Naive Bayes algorithms. SMOTE is used to address data imbalance, while Naive Bayes is applied to classify based on probability. The research process begins with collecting data on social assistance recipients, which is then processed using the SMOTE algorithm to increase the number of data in the minority category. Next, the Naive Bayes algorithm is applied to classify recipient eligibility based on available features. The calculation results show that the combination of these two algorithms significantly improved classification accuracy. Based on the calculation, the highest value is 0.008 "Eligible", so the tested data is categorized as Eligible to receive Social Assistance. If expressed as a percentage, this value is equivalent to $0.008 \times 100 = 0.75$ or 75%. Furthermore, the system developed in this study is designed to facilitate implementation in government agencies. With the resulting transparency and efficiency, this system is able to assist the East Medan Sub-district Office in distributing social assistance more precisely.

Keywords : Data Mining, Classification, Social Assistance, SMOTE, Naive Bayes

Info Artikel :

Disubmit: 11 September 2025

Disubmit: 11 September 2025

Disubmit: 11 September 2025

Copyright © 2025 – senadimu. All rights reserved.

1. PENDAHULUAN

Perkembangan teknologi dan komunikasi yang pesat saat ini memberikan dampak yang sangat besar bagi kehidupan sehari – hari manusia. Perkembangan teknologi informasi juga memiliki sifat positif maupun negatif di kehidupan manusia [1]. Pemanfaatan teknologi informasi secara efektif sangat penting untuk mengefektifkan aktivitas ketenagakerjaan manusia, seperti pengelolaan data administrasi pegawai, guru, dan siswa berbasis database teknologi secara efektif sangat penting untuk memperlancar aktivitas tenaga kerja manusia [2]. Hal itu berdampak pada semakin canggihnya kemajuan teknologi yang menjadikan segala hal terkait dengan teknologi, informasi dan komunikasi baik itu yang positif maupun negatif dapat diperoleh secara cepat, tepat dan akurat [3]. Data mentah menjadi informasi ketika data tersebut bermakna, relevan, dan dapat digunakan untuk tujuan tertentu, seperti pengambilan keputusan [4].

Kantor Camat Medan Timur merupakan salah satu instansi pemerintah yang bertugas memberikan pelayanan administrasi publik kepada masyarakat di wilayah Kecamatan Medan Timur. Salah satu tugas pentingnya adalah memastikan program-program pemerintah, seperti penyaluran bantuan sosial, dapat diterima oleh masyarakat yang berhak. Bantuan sosial yang disalurkan berupa Bantuan Pangan Non-Tunai (BPNT), yaitu program bantuan berupa bahan pangan dengan nominal Rp200.000 per bulan yang diberikan kepada keluarga penerima manfaat (KPM) yang tergolong kurang mampu. Berdasarkan data dari Badan Pusat Statistik (BPS) tahun 2023, jumlah penduduk di Kecamatan Medan Timur mencapai 117.035 jiwa. Dengan jumlah penduduk yang padat dan beragam tersebut, Kantor Camat Medan Timur menghadapi tantangan besar dalam mengelola data penerima dan memastikan bantuan ini tepat sasaran, terutama ketika penyaluran bantuan dilakukan secara manual tanpa dukungan sistem klasifikasi yang terstruktur.

Permasalahan utama yang sering muncul dalam proses penyaluran bantuan sosial adalah ketidakakuratan data penerima, seperti adanya data yang tidak lengkap, tidak valid, atau tidak terverifikasi dengan baik. Selain itu, metode seleksi manual yang selama ini diterapkan cenderung memakan waktu dan kurang efisien. Metode manual ini dilakukan dengan cara petugas mendatangi langsung rumah warga untuk melakukan pendataan. Seluruh data yang dikumpulkan kemudian dicatat secara tertulis atau diinput ke dalam spreadsheet sederhana tanpa melalui proses analisis sistematis. Penilaian kelayakan penerima bantuan pun hanya didasarkan pada pertimbangan subjektif dan pengalaman petugas lapangan, tanpa menggunakan kriteria yang terstandarisasi secara konsisten. Akibatnya, proses ini rawan terhadap kesalahan dalam penilaian, membutuhkan waktu yang lama, dan sulit dijalankan secara efektif ketika jumlah calon penerima sangat banyak. Hal ini berkontribusi pada ketidaktepatan sasaran dalam penyaluran bantuan, seperti adanya penerima bantuan yang sebenarnya tidak layak atau sebaliknya, yang akhirnya berdampak pada bantuan sosial yang tidak tepat sasaran.

Data mining adalah proses menganalisis kumpulan data yang besar untuk menemukan pola, hubungan, tren, atau anomali yang sebelumnya tidak diketahui dan bermakna [5]. Inti dari data mining adalah mengubah data mentah menjadi informasi yang berguna dan selanjutnya menjadi pengetahuan (*knowledge*) yang dapat ditindaklanjuti untuk mendukung pengambilan keputusan [6]. Proses ini menerapkan berbagai teknik dari bidang statistik, kecerdasan buatan (*artificial intelligence*), dan pembelajaran mesin (*machine learning*), seperti klasifikasi, klastering, dan asosiasi [7]. Tujuannya tidak hanya mendeskripsikan apa yang telah terjadi (analisis deskriptif) tetapi juga untuk memprediksi perilaku atau tren di masa depan (analisis prediktif), sehingga memberikan wawasan strategis [8]. Data mining sering dianggap sebagai langkah inti dalam proses yang lebih besar yang disebut Knowledge Discovery in Databases (KDD), yaitu proses menyeluruh untuk menemukan pengetahuan dari data [9].

Sebagai solusi, penelitian ini mengintegrasikan algoritma SMOTE dan Naive Bayes untuk membangun sistem klasifikasi kelayakan penerima bantuan sosial. SMOTE bertugas melakukan upsampling atas imbalanced data yang timbul dalam dataset [10]. SMOTE merupakan salah satu teknik data balancing yang bekerja dengan cara menghasilkan sampel sintetis pada kelas minoritas berdasarkan interpolasi antar data yang ada [11]. Naive Bayes adalah sebuah algoritma supervised learning berdasarkan teorema Bayes yang digunakan untuk memecahkan masalah klasifikasi dengan mengikuti

pendekatan probabilistic [12]. Algoritma Naive Bayes merupakan sebuah klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes [13].

Penelitian “Klasifikasi Masyarakat Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Naïve Bayes Dan Smote”. Berdasarkan hasil model kinerja klasifikasi yang diperoleh, model yang dihasilkan dapat diimplementasikan ke dalam sistem aplikasi pendukung keputusan untuk membantu Desa dalam menentukan penerima BLT DD agar lebih cepat dan mudah [14]. Penelitian “Penerapan Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Kepribadian MbtI Menggunakan Naive Bayes Classifier”. Dengan demikian, penelitian ini memberikan pemahaman tentang penggunaan algoritma Naive Bayes Classifier dan teknik SMOTE dalam prediksi kepribadian MBTI, dengan potensi peningkatan kinerja melalui penggunaan algoritma Logistic Regression [15].

2. METODE

1. Synthetic Minority Oversampling Technique (SMOTE)

SMOTE: “*Synthetic Minority Oversampling Technique*”. SMOTE merupakan metode resampling dimana metode ini akan mengekstrak dan membentuk data baru (data sintetik) berdasarkan nilai data dari kelas minoritas sehingga distribusi data antar kelas seimbang. Sebagian besar metode *oversampling* memiliki prinsip mereproduksi data baru secara acak, sehingga metode SMOTE lebih memilih *oversampling* dengan pendekatan k-NN. Pengolahan data sintetik pada data dengan nilai fitur numerik berbeda dengan data yang memiliki nilai fitur kategori. Data *numerik* diukur kemiripannya dengan persamaan *Euclidean*, sedangkan data kategorik dilakukan dengan menggunakan rumus *Value Difference Metric* (VDM) [16], yaitu.

$$d(V_1, V_2) = \sum_{i=1}^N \left| \frac{c_{1i}}{c_1} - \frac{c_{2i}}{c_2} \right| \dots \dots \dots (1)$$

2. Naive Bayes

Naive Bayes menerapkan konsep probabilitas dari teori ilmu statistik untuk menangani kasus *Supervised Learning*, di mana terdapat label, kelas, atau target yang menjadi acuan dalam himpunan data. Model klasifikasi *Naive Bayes* melibatkan beberapa tahap penyelesaian, dimulai dari pelatihan data (*training*) dan diakhiri dengan pengujian (*testing*), dengan tujuan menghasilkan keputusan yang akurat. Metode *Naive Bayes* merupakan salah satu pendekatan yang mengandalkan *teorema Bayes*, sebuah algoritma klasifikasi statistik yang dapat digunakan untuk memperkirakan probabilitas suatu kejadian. *Naive Bayes* didasarkan pada asumsi bahwa nilai-nilai atribut secara independen terhadap satu sama lain ketika nilai outputnya diberikan. Oleh karena itu, kelebihan dalam proses data mining menggunakan *Naive Bayes* adalah bahwa model algoritma ini hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter yang diperlukan.

$$P(H|X) = \frac{P(H|X) \times P(H)}{P(X)} \dots \dots \dots (1)$$

X = Sampel data yang memiliki kelas (label) yang tidak diketahui

H = Hipotesa bahwa X adalah data kelas (label)

P(H) = Peluang dari hipotesa H

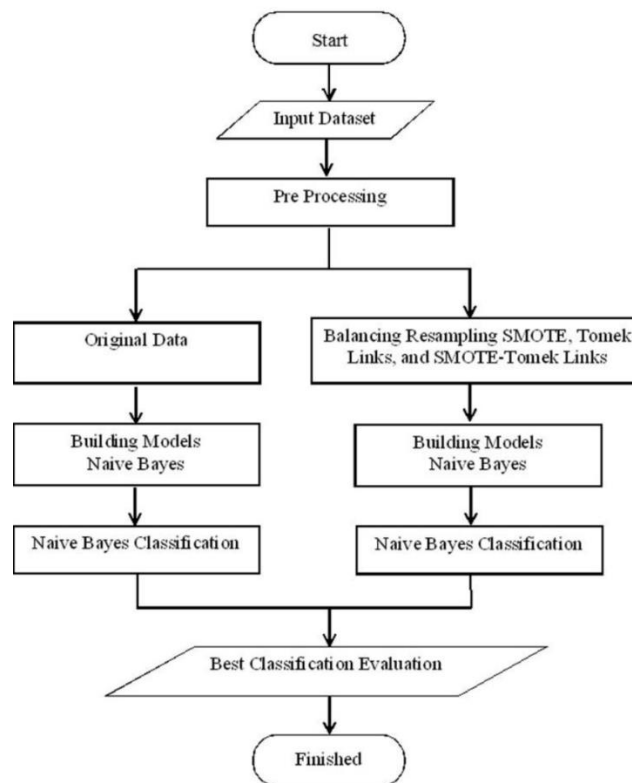
P(X) = Peluang dari data sampel yang di amati

P(H|X) = Peluang data sampel X bila diamsusikan bahwa hipotesa benar [17].

3. Flowchart Metode SMOTE dan Naïve Bayes

Pada tahapan penelitian yang menerapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) dan algoritma klasifikasi *Naive Bayes*, diperlukan suatu alur kerja yang sistematis agar setiap proses dapat berjalan secara terstruktur dan menghasilkan model klasifikasi yang optimal. Flowchart berikut disusun untuk memberikan gambaran menyeluruh mengenai proses pengolahan data mulai dari tahap awal hingga evaluasi akhir, sehingga membantu dalam memahami hubungan antara proses *pre-*

processing, penyeimbangan kelas, pembangunan model, serta penilaian performa klasifikasi. Dengan adanya diagram alir ini, seluruh rangkaian metode dapat diuji, dibandingkan, dan dianalisis secara lebih objektif sesuai dengan prinsip penelitian berbasis data.



Gambar 1. Flowchart SMOTE dan Naïve Bayes

Flowchart menunjukkan langkah-langkah sistematis dimulai dari tahap *Start*, kemudian dilanjutkan dengan proses memasukkan dataset (*Input Dataset*). Dataset tersebut menjalani tahapan *Pre Processing* untuk memastikan kualitas data sebelum digunakan dalam proses klasifikasi. Setelah pra-pemrosesan selesai, alur terbagi menjadi dua jalur utama, yaitu penggunaan data asli (*Original Data*) dan penggunaan data hasil penyeimbangan melalui teknik *resampling*, yang meliputi SMOTE, Tomek Links, dan SMOTE–Tomek Links. Pada masing-masing jalur, dilakukan pembangunan model menggunakan algoritma Naïve Bayes, yang kemudian menghasilkan proses Naïve Bayes Classification. Kedua hasil klasifikasi tersebut selanjutnya dibandingkan melalui tahap *Best Classification Evaluation* guna menentukan metode mana yang memberikan performa paling optimal dalam menangani ketidakseimbangan kelas. Proses berakhir pada tahap *Finished* setelah model terbaik berhasil diidentifikasi.

3. HASIL DAN PEMBAHASAN

Metode SMOTE (*Synthetic Minority Over-sampling Technique*) dan *Naive Bayes* sering digunakan bersama untuk menangani masalah ketidakseimbangan kelas (*class imbalance*) dalam klasifikasi data. SMOTE mengatasi ketidakseimbangan dengan membuat sampel sintesis untuk kelas minoritas berdasarkan *k-nearest neighbors*, sehingga distribusi kelas lebih seimbang. Sementara itu, *Naive Bayes* adalah algoritma klasifikasi probabilistik yang sederhana dan efisien, meskipun mengasumsikan kemandirian antar fitur. Dengan menggabungkan SMOTE dan *Naive Bayes*, performa model klasifikasi dapat meningkat karena SMOTE mengurangi bias terhadap kelas mayoritas, sementara *Naive Bayes*

memprediksi kelas berdasarkan probabilitas. Kombinasi ini efektif untuk dataset dengan ketidakseimbangan kelas yang signifikan.

1. Identifikasi Atribut

Atribut yang digunakan untuk memastikan penerimaan bantuan sosial (bansos) data yang digunakan sebagai berikut.

Tabel 1 Atribut

No	Atribut	Subatribut	Nilai
A1	Mata Pencarian	Wiraswasta	4
		Pegawai Swasta/Buruh	3
		Petani	2
		Pegawai Negeri Sipil	1
A2	Penghasilan	< Rp. 1.000.000	4
		< Rp. 2.000.000	3
		< RP. 3. 000.000	2
		< Rp. 4.000.000	1
A3	Tanggungan	4	4
		3	3
		2	2
		1	1
A4	Kondisi Rumah	Tidak Layak	4
		Kurang Layak	3
		Cukup Layak	2
		Layak	1

2. Identifikasi Dataset

Dataset yang digunakan untuk memastikan penerimaan bantuan sosial (bansos) data yang digunakan sebagai berikut.

Tabel 2 Dataset

No.	Warga	A1	A2	A3	A4	Kelas
1	Hiras Manurung	Pegawai Swasta/Buruh	2000000	2	Kurang Layak	Layak
2	Milakarmelia	Pegawai Negeri Sipil	4000000	3	Layak	Tidak Layak
3	Hotmaidah Fitri Ritonga	Pegawai Negeri Sipil	4000000	3	Tidak Layak	Tidak Layak
4	Gusriati Dewi	Pegawai Negeri Sipil	4000000	1	Tidak Layak	Tidak Layak
5	Eri Erianto	Wiraswasta	2000000	2	Kurang Layak	Layak
6	Masyitoh Ritonga	Pegawai Swasta/Buruh	3000000	3	Cukup Layak	Layak
7	Rendi Nurindah Sari	Pegawai Swasta/Buruh	2000000	4	Tidak Layak	Layak
8	Huta Darawati	Wiraswasta	2000000	2	Kurang Layak	Layak
9	Eko Tri Kurniawan	Petani	3000000	4	Layak	Layak
10	Elisabet Elvira	Wiraswasta	1000000	2	Kurang Layak	Layak
11	Yunita	Pegawai Negeri Sipil	4000000	1	Cukup Layak	Tidak Layak
12	Tini	Pegawai Swasta/Buruh	2000000	4	Tidak Layak	Layak
13	Nurbaya	Wiraswasta	3000000	1	Layak	Layak
14	Fatmawati	Petani	3000000	3	Cukup Layak	Layak
15	Miswati	Wiraswasta	1000000	3	Cukup Layak	Layak
16	Bincar Siregar	Petani	3000000	1	Cukup Layak	Layak
17	Ramadhan	Wiraswasta	3000000	4	Tidak Layak	Layak
18	Suyanti	Pegawai Negeri Sipil	4000000	2	Kurang Layak	Tidak Layak
19	Lindawati Tambunan	Pegawai Negeri Sipil	4000000	2	Cukup Layak	Tidak Layak
20	Sutaryo	Pegawai Negeri Sipil	4000000	1	Kurang Layak	Tidak Layak
21	Yusliana	Petani	1000000	2	Layak	Layak
22	Eka Sutiana	Wiraswasta	2000000	3	Kurang Layak	Layak
23	Wita Wardani	Petani	1000000	1	Cukup Layak	Layak
24	Valentino	Wiraswasta	2000000	2	Layak	Layak

No.	Warga	A1	A2	A3	A4	Kelas
25	Setiawarman	Pegawai Negeri Sipil	4000000	3	Kurang Layak	Tidak Layak
26	Mariani	Pegawai Swasta/Buruh	2000000	4	Cukup Layak	Layak
27	Sri Dewi	Petani	3000000	4	Tidak Layak	Layak
28	Fajar Ajie Maulana Dika	Wiraswasta	1000000	4	Cukup Layak	Layak
29	Syahrizal	Wiraswasta	2000000	3	Kurang Layak	Layak
30	Tri Bayu Siswanto	Petani	2000000	1	Tidak Layak	Layak
31	Romla Siregar	Wiraswasta	3000000	1	Tidak Layak	Layak
32	Eva Khairunnisa Pane	Pegawai Negeri Sipil	4000000	2	Tidak Layak	Tidak Layak
33	Ernawati	Wiraswasta	2000000	3	Layak	Layak
34	Subagiok	Pegawai Negeri Sipil	4000000	3	Tidak Layak	Tidak Layak
35	Irma Indriani	Pegawai Negeri Sipil	4000000	3	Cukup Layak	Tidak Layak
36	Elfrida Boru Sianturi	Petani	2000000	1	Kurang Layak	Layak
37	Supriadi	Pegawai Swasta/Buruh	3000000	3	Tidak Layak	Layak
38	Indra Rinolta Sibagariang	Petani	3000000	3	Kurang Layak	Layak
39	Irma Suryani	Pegawai Negeri Sipil	4000000	2	Kurang Layak	Tidak Layak
40	Tria Nopita Sari	Pegawai Negeri Sipil	4000000	4	Layak	Tidak Layak
41	Susiyanti	Petani	1000000	1	Cukup Layak	Layak
42	Putri Ramadani	Pegawai Swasta/Buruh	2000000	3	Layak	Layak
43	Bahtiar Siregar	Pegawai Negeri Sipil	4000000	3	Cukup Layak	Tidak Layak
44	Ira Maya Sopha	Pegawai Swasta/Buruh	3000000	1	Cukup Layak	Layak
45	Ananda Hadijah	Pegawai Negeri Sipil	4000000	2	Kurang Layak	Tidak Layak
46	Fanotona Zega	Petani	3000000	1	Cukup Layak	Layak
47	Rusmini	Pegawai Swasta/Buruh	1000000	1	Kurang Layak	Layak
48	Angga Prasetya	Pegawai Swasta/Buruh	2000000	4	Tidak Layak	Layak
49	Ahmad Renaldi Lbs	Wiraswasta	2000000	3	Tidak Layak	Layak
50	Syahputra	Petani	3000000	2	Cukup Layak	Layak
...	...	...	...	...	...	...
185	Lenny Triana	Petani	3000000	3	Layak	Layak

3. Konversi Dataset

Kemudian dari 185 sampel tersebut dilakukan pengkonversian data yang telah disesuaikan dengan nilai yang dimiliki setiap atribut dan subatribut, tujuan pengkonversian data tersebut kedalam bentuk nilai dengan skala 1 s/d 4 agar bisa dilakukan perhitungan metode SMOTE, adapun hasil ya dapat dijelaskan pada tabel berikut :

Tabel 3 Konversi Dataset

No.	Warga	A1	A2	A3	A4	Kelas
1	Hiras Manurung	3	3	2	3	Layak
2	Milakarmelia	1	1	3	1	Tidak Layak
3	Hotmaidah Fitri Ritonga	1	1	3	4	Tidak Layak
4	Gusriati Dewi	1	1	1	4	Tidak Layak
5	Eri Erianto	4	3	2	3	Layak
6	Masyitoh Ritonga	3	2	3	2	Layak
7	Rendi Nurindah Sari	3	3	4	4	Layak
8	Huta Darawati	4	3	2	3	Layak
9	Eko Tri Kurniawan	2	2	4	1	Layak
10	Elisabet Elvira	4	4	2	3	Layak
11	Yunita	1	1	1	2	Tidak Layak
12	Tini	3	3	4	4	Layak
13	Nurbaya	4	2	1	1	Layak
14	Fatmawati	2	2	3	2	Layak
15	Miswati	4	4	3	2	Layak
16	Bincar Siregar	2	2	1	2	Layak

No.	Warga	A1	A2	A3	A4	Kelas
17	Ramadhan	4	2	4	4	Layak
18	Suyanti	1	1	2	3	Tidak Layak
19	Lindawati Tambunan	1	1	2	2	Tidak Layak
20	Sutaryo	1	1	1	3	Tidak Layak
21	Yusliana	2	4	2	1	Layak
22	Eka Sutiana	4	3	3	3	Layak
23	Wita Wardani	2	4	1	2	Layak
24	Valentino	4	3	2	1	Layak
25	Setiawarman	1	1	3	3	Tidak Layak
26	Mariani	3	3	4	2	Layak
27	Sri Dewi	2	2	4	4	Layak
28	Fajar Ajie Maulana Dika	4	4	4	2	Layak
29	Syahrizal	4	3	3	3	Layak
30	Tri Bayu Siswanto	2	3	1	4	Layak
31	Romla Siregar	4	2	1	4	Layak
32	Eva Khairunnisa Pane	1	1	2	4	Tidak Layak
33	Ernawati	4	3	3	1	Layak
34	Subagiok	1	1	3	4	Tidak Layak
35	Irma Indriani	1	1	3	2	Tidak Layak
36	Elfrida Boru Sianturi	2	3	1	3	Layak
37	Supriadi	3	2	3	4	Layak
38	Indra Rinolta Sibagariang	2	2	3	3	Layak
39	Irma Suryani	1	1	2	3	Tidak Layak
40	Tria Nopita Sari	1	1	4	1	Tidak Layak
41	Susiyanti	2	4	1	2	Layak
42	Putri Ramadani	3	3	3	1	Layak
43	Bahtiar Siregar	1	1	3	2	Tidak Layak
44	Ira Maya Sopha	3	2	1	2	Layak
45	Ananda Hadijah	1	1	2	3	Tidak Layak
46	Fanotona Zega	2	2	1	2	Layak
47	Rusmini	3	4	1	3	Layak
48	Angga Prasetya	3	3	4	4	Layak
49	Ahmad Renaldi Lbs	4	3	3	4	Layak
50	Syahputra	2	2	2	2	Layak
..	...	...	...	...	...	...
185	Lenny Triana	2	2	3	1	Layak

4. SMOTE

Setelah selesai dilakukan konversi data, selanjutnya dilakukan penghitungan jumlah kelas yang dapat dijelaskan pada tabel berikut :

Tabel 4 Jumlah Kelas Minoritas dan Mayoritas

No	Kelas	Jumlah
1	Layak	137
2	Tidak Layak	48

Dari tabel diatas dapat dijelaskan bawasanya yang termasuk kelas minoritas adalah kelas “Tidak Layak” karen jumlah kelas “Tidak Layak” lebih sedikit dibandingkan jumlah kelas “Layak”. Sehingga dapat dilakukan pengelompokan pada kelas minoritas yang dapat dilihat pada tabel berikut :

Tabel 5 Kelas Minoritas

No.	Warga	A1	A2	A3	A4	Kelas
2	Milakarmelia	1	1	3	1	Tidak Layak
3	Hotmaidah Fitri Ritonga	1	1	3	4	Tidak Layak

No.	Warga	A1	A2	A3	A4	Kelas
4	Gusriati Dewi	1	1	1	4	Tidak Layak
11	Yunita	1	1	1	2	Tidak Layak
18	Suyanti	1	1	2	3	Tidak Layak
19	Lindawati Tambunan	1	1	2	2	Tidak Layak
20	Sutaryo	1	1	1	3	Tidak Layak
25	Setiawarman	1	1	3	3	Tidak Layak
32	Eva Khairunnisa Pane	1	1	2	4	Tidak Layak
34	Subagiok	1	1	3	4	Tidak Layak
35	Irma Indriani	1	1	3	2	Tidak Layak
39	Irma Suryani	1	1	2	3	Tidak Layak
40	Tria Nopita Sari	1	1	4	1	Tidak Layak
43	Bahtiar Siregar	1	1	3	2	Tidak Layak
45	Ananda Hadijah	1	1	2	3	Tidak Layak
54	Hardian Singgih Prayogi	1	1	2	3	Tidak Layak
56	Eni Purwati	1	1	4	2	Tidak Layak
57	Harianto	1	1	2	3	Tidak Layak
61	Rika Armayani	1	1	2	2	Tidak Layak
62	Nur Aisyah Putri	1	1	2	3	Tidak Layak
64	Warta Kesuma	1	1	1	3	Tidak Layak
66	Zainuddin	1	1	2	4	Tidak Layak
68	Endang Rahayu Lubis	1	1	1	1	Tidak Layak
72	Marwahni Nandra Utari	1	1	3	1	Tidak Layak
74	Nurlita	1	1	2	2	Tidak Layak
78	Emilia Puspita	1	1	2	2	Tidak Layak
85	Chairia Ananda Tandias	1	1	4	3	Tidak Layak
96	Handoyo	1	1	1	4	Tidak Layak
97	Supriyadi	1	1	2	1	Tidak Layak
100	Nur Ainun	1	1	1	3	Tidak Layak
101	Nurna Dewi	1	1	1	3	Tidak Layak
102	Alfiatika	1	1	4	1	Tidak Layak
106	Sri Madonna Lubis	1	1	2	1	Tidak Layak
107	Mulya Agustini	1	1	1	3	Tidak Layak
112	Hakim Efendi	1	1	4	4	Tidak Layak
113	Ganda Ardian	1	1	4	2	Tidak Layak
116	Susilawati	1	1	1	1	Tidak Layak
119	Rizky Ramadhan	1	1	1	2	Tidak Layak
126	Supiana	1	1	1	3	Tidak Layak
130	Sondang Maduma Sitorus	1	1	2	4	Tidak Layak
132	Mia Bramyanti	1	1	1	4	Tidak Layak
149	Yusrizal	1	1	2	4	Tidak Layak
161	Rumondang Sihombing	1	1	3	1	Tidak Layak
162	Jamaludin	1	1	4	4	Tidak Layak
172	Freddy Pradinata Siadari	1	1	3	4	Tidak Layak
174	Srikandi Vitri	1	1	3	4	Tidak Layak
175	Sri Astuti Handayani	1	1	1	1	Tidak Layak
181	Junita Wati	1	1	2	2	Tidak Layak

Pilih secara acak satu sampel dari kelas minoritas yang dapat dijelaskan pada tabel berikut :

Tabel 6 Sampel Acak Kelas Minoritas

No.	Warga	A1	A2	A3	A4	Kelas
2	Milakarmelia	1	1	3	1	Tidak Layak

Hitung sampel acak kelas minoritas dengan kelas minoritas menggunakan Jarak *Euclidean* :

$$S_{3.2} = \sqrt{(1-1)^2 + (1-1)^2 + (3-3)^2 + (4-1)^1} = 3,000$$

$$S_{4.2} = \sqrt{(1-1)^2 + (1-1)^2 + (1-3)^2 + (4-1)^1} = 3,606$$

$$S_{11.2} = \sqrt{(1-1)^2 + (1-1)^2 + (1-3)^2 + (2-1)^1} = 2,236$$

$$S_{18.2} = \sqrt{(1-1)^2 + (1-1)^2 + (2-3)^2 + (3-1)^1} = 2,236$$

$$S_{19.2} = \sqrt{(1-1)^2 + (1-1)^2 + (2-3)^2 + (2-1)^1} = 1,414$$

Dan seterusnya. Sehingga hasil Jarak *Euclidean* diambil menggunakan k=3 (3 tetangga terdekat) dapat dilihat pada tabel dibawah ini:

Tabel 7 Jarak *Euclidean*

No.	Warga	Euclidean	K=3
3	Hotmaidah Fitri Ritonga	3,000	35
4	Gusriati Dewi	3,606	45
11	Yunita	2,236	20
18	Suyanti	2,236	21
19	Lindawati Tambunan	1,414	9
20	Sutaryo	2,828	29
25	Setiawarman	2,000	16
32	Eva Khairunnisa Pane	3,162	39
34	Subagiok	3,000	36
35	Irma Indriani	1,000	3
39	Irma Suryani	2,236	22
40	Tria Nopita Sari	1,000	4
43	Bahtiar Siregar	1,000	5
45	Ananda Hadijah	2,236	23
54	Hardian Singgih Prayogi	2,236	24
56	Eni Purwati	1,414	10
57	Hariato	2,236	25
61	Rika Armayani	1,414	11
62	Nur Aisyah Putri	2,236	26
64	Warta Kesuma	2,828	30
66	Zainuddin	3,162	40
68	Endang Rahayu Lubis	2,000	17
72	Marwahni Nandra Utari	0,000	1
74	Nurlita	1,414	12
78	Emilia Puspita	1,414	13
85	Chairia Ananda Tandias	2,236	27
96	Handoyo	3,606	46
97	Supriyadi	1,000	6
100	Nur Ainun	2,828	31
101	Nurna Dewi	2,828	32
102	Alfiatika	1,000	7
106	Sri Madonna Lubis	1,000	8
107	Mulya Agustini	2,828	33
112	Hakim Efendi	3,162	41
113	Ganda Ardian	1,414	14
116	Susilawati	2,000	18
119	Rizky Ramadhan	2,236	28
126	Supiana	2,828	34
130	Sondang Maduma Sitorus	3,162	42
132	Mia Bramyanti	3,606	47
149	Yusrizal	3,162	43
161	Rumondang Sihombing	0,000	2

No.	Warga	Euclidean	K=3
162	Jamaludin	3,162	44
172	Freddy Pradinata Siadari	3,000	37
174	Srikandi Vitri	3,000	38
175	Sri Astuti Handayani	2,000	19
181	Junita Wati	1,414	15

Tabel 8 Jarak *Euclidean* K=3

No.	Warga	Euclidean	K=3
72	Marwahni Nandra Utari	0,000	1
161	Rumondang Sihombing	0,000	2
35	Irma Indriani	1,000	3

Selanjutnya pembuatan Sampel Sintetis yaitu dengan pilih angka acak antar 0 dan 1, disini kita menggunakan angka acak 0,5 dan hitung perbedaan antara setiap sampel-sampel minoritas yang dapat dijelaskan sebagai berikut :

$$\begin{aligned}
 X_{72,2} &= ((1-1) 0,5)+1,((1-1) 0,5)+1,((3-3) 0,5)+3,((1-1) 0,5)+1 \\
 &= (1),(1),(3),(1) \\
 X_{161,2} &= ((1-1) 0,5)+1,((1-1) 0,5)+1,((3-3) 0,5)+3,((1-1) 0,5)+1 \\
 &= (1),(1),(3),(1) \\
 X_{35,2} &= ((1-1) 0,5)+1,((1-1) 0,5)+1,((3-3) 0,5)+3,((2-1) 0,5)+1 \\
 &= (1),(1),(3),(1,5)
 \end{aligned}$$

Untuk hasil ya dapat dilihat pada tabel berikut :

Tabel 9 Sampel Sintetis Baru

No.	Sampel	A1	A2	A3	A4	Kelas
72	Sampel Sintetis Baru	1	1	3	1	Tidak Layak
161	Sampel Sintetis Baru	1	1	3	1	Tidak Layak
35	Sampel Sintetis Baru	1	1	3	1,5	Tidak Layak

Sehingga dapat disimpulkan untuk Sampel Sintetis Baru hasil perhitungan Metode SMOTE dapat dilihat pada tabel diatas, setelah selesai melakukan perhitungan SMOTE selanjutnya dilakukan perhitungan Naive Bayes.

5. Naive Bayes

Hitung probabilitas setiap kelas yang dimiliki pada setiap sampel yang dapat dijelaskann pada tabel berikut :

Tabel 10 Probabilitas Kelas

Kelas		Probabilitas	
Layak	Tidak Layak	Layak	Tidak Layak
137	51	0,729	0,271
Jumlah	188		

Hitung probabilitas setiap atribut yang memiliki kelas yang sama pada setiap sampel yang dapat dijelaskann pada tabel berikut :

a. Mata Pencarian

Tabel 11 Probabilitas Mata Pencarian

A1	Kelas		Probabilitas	
	Layak	Tidak Layak	Layak	Tidak Layak
A1.a	45	0	0,328	0,000
A1.b	41	0	0,299	0,000
A1.c	51	0	0,372	0,000

A1	Kelas		Probabilitas	
	Layak	Tidak Layak	Layak	Tidak Layak
A1.d	0	51	0,000	1,000
Jumlah	137	51		

b. Penghasilan

Tabel 12 Probabilitas Penghasilan

A2	Kelas		Probabilitas	
	Layak	Tidak Layak	Layak	Tidak Layak
A2.a	34	0	0,248	0,000
A2.b	53	0	0,387	0,000
A2.c	50	0	0,365	0,000
A2.d	0	51	0,000	1,000
Jumlah	137	51		

c. Tanggungan

Tabel 13 Probabilitas Tanggungan

A3	Kelas		Probabilitas	
	Layak	Tidak Layak	Layak	Tidak Layak
A3.a	27	7	0,197	0,137
A3.b	32	13	0,234	0,255
A3.c	44	17	0,321	0,333
A3.d	34	14	0,248	0,275
Jumlah	137	51		

d. Kondisi Rumah

Tabel 14 Probabilitas Kondisi Rumah

A4	Kelas		Probabilitas	
	Layak	Tidak Layak	Layak	Tidak Layak
A4.a	38	13	0,277	0,255
A4.b	24	14	0,175	0,275
A4.c	38	12	0,277	0,235
A4.d	37	12	0,270	0,235
Jumlah	137	51		

Menghitung jumlah kasus sama dengan kelas yang sama berdasarkan data yang di uji.

Tabel 15 Data Uji

No.	Sampel	A1	A2	A3	A4	Kelas
1	Data Uji	3	3	2	2	?

Kelompokan setiap kelas yang memiliki kesamaan pada data yang diuji yang dapat dilihat pada tabel berikut :

Tabel 16 Pengelompokan Kelas

Atribut	Kelas		Probabilitas	
A1.b	41,000	0,000	0,299	0,000
A2.b	53,000	0,000	0,387	0,000
A3.c	44,000	17,000	0,321	0,333
A4.c	38,000	12,000	0,277	0,235
			0,010	0,000

Hitung probabilitas setiap kelas yang hasilnya dapat dilihat pada tabel berikut :

Tabel 17 Hasil Akhir

Kelas	Probabilitas	Normalisasi
Layak	0,008	1
Tidak Layak	0,000	0
Jumlah	0,008	

Maka hasil dari perhitungan menggunakan metode *naïve bayes* berdasarkan nilai tertinggi yaitu sebesar 0,008 “Layak”. Sehingga data yang diuji “Layak” mendapatkan Bantuan Sosial, dan untuk tingkat persentase ya $0.008 \times 100 = 0,75$ jika di persentasekan sebesar 75%.

4. KESIMPULAN

Adapun kesimpulan yang dapat diambil oleh penulis adalah sebagai berikut :

1. Penggabungan algoritma SMOTE dan *Naive Bayes* terbukti mampu meningkatkan performa klasifikasi kelayakan penerima bantuan sosial, khususnya dalam mengatasi masalah ketidakseimbangan data.
2. Sistem klasifikasi yang dibangun membantu pihak Kantor Camat Medan Timur dalam menyaring calon penerima bantuan secara lebih objektif, efisien, dan transparan.
3. Penggunaan data mining dalam penyaluran bantuan sosial menunjukkan bahwa teknologi informasi dapat menjadi solusi nyata dalam mendukung pemerintahan yang lebih baik dan akuntabel.

UCAPAN TERIMA KASIH

Sebagai penutup, penulis mengucapkan terima kasih kepada Universitas Potensi Utama yang telah memberikan kesempatan dan dukungan dalam penyelesaian Karya Tulis Ilmiah (KTI) ini.

REFERENSI

- [1] Fahruliyansyah, Y., & Sari, R. N. (2023). Perancangan Aplikasi Belajar Matematika Kelas IV SD Berbasis Android. *Jurnal Info Digit (JID)*, 1(3), 901-917.
- [2] Sari, R. N., Rahayu, S. L., Hayati, R. S., & Lubis, H. N. (2023). Perancangan Aplikasi Absensi Berbasis Web Responsive Design Untuk Siswa SMK. *Publikasi Pengabdian Masyarakat*, 3(2), 386-393.
- [3] Sari, R. N., & Rahayu, S. L. (2022). PELATIHAN PENGENALAN BAHASA PEMODELAN UML (UNIFIED MODELLING LANGUAGE) PADA SMA PAB 8 SAENTIS. *CORAL (Community Service Journal)*, 1(1), 29-35.
- [4] Pulungan, R. D., Sari, R. N., & Rahman, I. (2023). Rancang Bangun Aplikasi Penjadwalan Seminar Dan Sidang Tugas Akhir/Skripsi Berbasis Android Di Universitas Potensi Utama. *Jurnal Rekayasa Sistem (JUREKSI)*, 1(1), 408-423.
- [5] Setiawan, A., Kurniawan, V., & Novita, R. (2024). Penerapan Algoritma Eclat Untuk Mencari Pola Hubungan Antar Barang Pada Data transaksi Penjualan: Application of the Eclat Algorithm to Find Relationship Patterns Between Items in Sales Transaction Data. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 4(1), 9-16.
- [6] Putera, D. Y., & Rahayu, S. L. (2023). Penerapan Data Mining Dalam Memprediksi Tingkat Perceraian Pasangan Muda Di Kota Medan Menggunakan Metode NAIVE BAYES. *Jurnal Rekayasa Sistem (JUREKSI)*, 1(1), 18-29.

-
- [7] Gunawan, D. N., & Harahap, A. Y. N. (2024). Penerapan Data Mining Menggunakan Metode Clustering Dan Hierarki Divisive Clustering Dalam Evaluasi Data Penjualan. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(3 A), 1568-1579.
- [8] Nurisyifa, A. F., & Agustin, F. (2024). Penerapan Data Mining Menggunakan Metode C. 45 Dalam Cari Pola Minat Prangko Pada PT. Pos Indonesia Berbasis Web. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(2), 474-486.
- [9] Pratiwi, D. A., & Hts, D. I. G. (2024). Analisis Klasifikasi Kelulusan Siswa Baru SMKS Bina Satria Medan Memakai Data Mining Serta Metode Naive Bayes. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(3 A), 1492-1503.
- [10] Kurnia, I. P., & Saputra, D. D. (2023). Analisis sentimen dengan metode naïve bayes, SMOTE dan adaboost pada Twitter Bank BTN. *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, 7(2), 235-242.
- [11] Gultom, M. D., Destianti, R. D., & Noviamanti, G. (2025). PENERAPAN NAÏVE BAYES DENGAN BALANCING DATA SMOTE UNTUK PENINGKATAN PREDIKSI NASABAH DEPOSITO BANK. *Journal of Computer Science and Artificial Intelligence (JCSAI)*, 2(1).
- [12] Izzi, H., Setyanto, A., & Hartanto, A. D. (2025). Optimalisasi Akurasi Algoritma Naïve Bayes Dengan Metode Syntetic Minority Oversampling Technique (Smote) Pada Data Numerik. *Infotek: Jurnal Informatika dan Teknologi*, 8(1), 217-227.
- [13] Rachmat, B. K., Suwarisman, A., Afriyanti, I., Wahyudi, A., & Saputra, D. D. (2023). Analisis Sentimen Complain dan Bukan Complain pada Twitter Telkomsel dengan SMOTE dan Naïve Bayes. *J. Teknol. Inf. dan Komunikasi*, 7(1), 107-113.
- [14] Kurniadi, D., Nuraeni, F., & Firmansyah, M. (2023). Klasifikasi masyarakat penerima bantuan langsung tunai dana desa menggunakan naïve Bayes dan SMOTE. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 10(2).
- [15] Pulungan, M. P., Purnomo, A., & Kurniasih, A. (2024). Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(5), 1033-1042.
- [16] Nababan, A. A., Jannah, M., Aulina, M., & Andrian, D. (2023). Prediksi Kualitas Udara Menggunakan Xgboost Dengan Synthetic Minority Oversampling Technique (Smote) Berdasarkan Indeks Standar Pencemaran Udara (Ispu). *JTIK (Jurnal Teknik Informatika Kaputama)*, 7(1), 214-219.
- [17] Saputra, S. G. A., & Sidik, C. L. R. F. (2023). Implementasi Metode Naïve Bayes Pada Prediksi Penyakit Seliak. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika dan Komputer VOL.07, NO.01, Bulan Februari 2023* pp.8-13.