

Data Mining Application for Predicting Student Graduation Rates Using the SMOTE and K-NN Methods at SMK PAB Helvetia

^{a,b} Sistem Informasi, Universitas Potensi Utama, Medan 20241, Sumatera Utara, Indonesia
Email: muhammadrizkyazhari882@gmail.com¹, helmikurniawan77@gmail.com²

yang dihadapi adalah tingkat kelulusan siswa yang bervariasi setiap tahunnya. Tingkat kelulusan siswa merupakan salah satu indikator utama keberhasilan sebuah institusi pendidikan, termasuk di Sekolah Menengah Kejuruan (SMK) PAB Helvetia. Kelulusan siswa tidak hanya menjadi bukti pencapaian siswa dalam menyelesaikan program pendidikan, tetapi juga mencerminkan kualitas pembelajaran yang diberikan oleh sekolah.

Namun, berdasarkan data kelulusan siswa SMK PAB Helvetia, terdapat fenomena yang menunjukkan bahwa meskipun tingkat kelulusan mencapai 95%, masih ada sekitar 5% siswa yang tidak berhasil lulus dan angka itu berubah-ubah setiap tahunnya. Hal ini menjadi perhatian serius karena siswa yang tidak lulus biasanya menunjukkan kesenjangan dalam beberapa aspek, seperti Nilai Ujian Sekolah (US) Teori, Nilai Ujian Sekolah (US) Peraktek, Nilai Ujian Tengah Semester (UTS), Nilai Prilaku Siswa. Ketidakmampuan sekolah untuk memprediksi tingkat kelulusan dengan akurat sering kali menyulitkan pihak manajemen dalam merancang strategi pembelajaran yang efektif. Oleh karena itu, diperlukan sistem yang dapat membantu sekolah menganalisis data siswa untuk memprediksi kemungkinan kelulusan mereka dengan lebih baik.

Data mining adalah proses menganalisis set data yang besar untuk menemukan pola, korelasi, dan anomali yang sebelumnya tidak diketahui untuk memprediksi tren atau hasil di masa depan [1]. Data mining adalah suatu teknik dalam bisnis untuk menggali informasi berharga dari data historis guna mendukung pengambilan keputusan yang lebih strategis, seperti meningkatkan penjualan, memahami perilaku pelanggan, atau mengurangi risiko [2]. Data adalah fakta mentah, angka, kata-kata, pengamatan, atau deskripsi tentang suatu entitas (seperti orang, benda, atau peristiwa) yang belum diolah dan belum memiliki makna tertentu [3]. Data adalah representasi formal dari informasi atau fakta dalam bentuk yang terstruktur (seperti tabel, database) atau tidak terstruktur (seperti teks, gambar, video), sehingga dapat disimpan, diproses, dan dikomunikasikan oleh computer [4]. Data adalah bahan baku atau bahan mentah yang, setelah melalui proses pengorganisasian, kontekstualisasi, dan pengolahan, akan menjadi informasi yang bernilai dan bermakna untuk pengambilan keputusan [5].

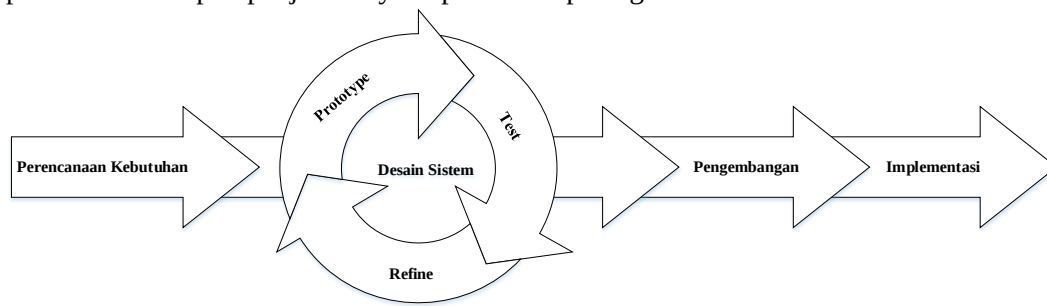
K-NN (K-Nearest Neighbors) adalah metode klasifikasi yang mengelompokkan data baru berdasarkan kedekatannya dengan data-data lain yang sudah diketahui kelasnya dalam ruang dimensi tertentu [6]. Metode ini bekerja dengan membandingkan jarak suatu data titik baru (query) ke sejumlah 'K' data tetangga terdekatnya, dimana kelas data baru tersebut akan ditentukan oleh kelas yang paling banyak muncul di antara tetangga terdekatnya [7]. Nilai 'K' (jumlah tetangga terdekat) merupakan parameter penting yang mempengaruhi hasil; nilai K yang terlalu kecil dapat membuat model sensitif terhadap noise, sedangkan nilai K yang terlalu besar dapat membuat batas keputusan menjadi terlalu halus [8]. SMOTE (Synthetic Minority Oversampling Technique) adalah teknik oversampling yang menangani data tidak seimbang dengan membuat data sintetis baru untuk kelas minoritas [9]. Tujuan utama SMOTE adalah untuk mengatasi bias model machine learning yang disebabkan oleh ketidakseimbangan jumlah data antar kelas, sehingga model dapat belajar karakteristik kelas minoritas dengan lebih baik dan tidak didominasi oleh kelas mayoritas [10].

Penelitian “Optimasi K-Nearest Neighbor Menggunakan Algoritma Smote Untuk Mengatasi Imbalance Class Pada Klasifikasi Analisis Sentimen”. Penelitian ini menyoroti pentingnya mempertimbangkan metode oversampling seperti SMOTE untuk meningkatkan performa KNN dalam mengatasi ketidakseimbangan data pada klasifikasi analisis sentimen di media sosial Twitter [11]. Penelitian “Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN”. Hasil klasifikasi dievaluasi akurasi menggunakan 10fold-cross validation dengan membandingkan hasil klasifikasi yang hanya menggunakan KNN dan menggunakan KNN dan SMOTE [12]. Penelitian “Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE”. Tujuan penelitian ini adalah untuk mengetahui sentimen masyarakat terhadap implementasi MBKM serta melakukan optimasi terhadap performa metode KNN melalui metode SMOTE dalam melakukan analisis sentiment [13]. Penelitian “Klasifikasi Kelayakan Calon Kreditur menggunakan Metode Syntethic Minority Over-Sampling Technique (SMOTE) dan K-Nearest Neighbor (KNN)”. Nilai akurasi yang didapatkan ketika data diklasifikasikan dengan menggunakan metode KNN tanpa menggunakan SMOTE adalah 60% [14]. Penelitian “Penerapan Synthetic Minority

Oversampling Technique (SMOTE) untuk Imbalance Class pada Data Text Menggunakan KNN". Tujuan dari penelitian ini yaitu untuk mengetahui performansi penanganan ketidakseimbangan kelas pada data review aplikasi Tokopedia menggunakan algoritma kNN dan SMOTE [15].

2. METODE

Rapid Application Development (RAD) adalah sebuah metodologi pengembangan perangkat lunak yang menekankan pada kecepatan proses pengembangan melalui pendekatan iteratif dan kolaboratif [16]. RAD bertujuan untuk mempercepat pengembangan aplikasi dengan melibatkan pengguna akhir secara aktif dalam setiap tahap proses, sehingga kebutuhan dan perubahan dapat langsung diakomodasi. Dalam model ini, fokus utama adalah menghasilkan prototipe fungsional dalam waktu singkat untuk segera diuji dan diperbaiki berdasarkan masukan pengguna. Metode RAD terdiri dari beberapa fase utama, yaitu perencanaan kebutuhan, desain pengguna (*prototyping*), pengembangan, dan implementasi. Adapun penjelasan ya dapat dilihat pada gambar berikut :



Gambar 1 RAD (Rapid Application Development)

Berikut adalah penjelasan setiap tahap, disesuaikan dengan penelitian tentang aplikasi prediksi tingkat kelulusan siswa menggunakan metode SMOTE dan K-NN:

1. Perencanaan Kebutuhan
Tahap ini merupakan langkah awal untuk mengidentifikasi kebutuhan sistem prediksi tingkat kelulusan siswa. Peneliti dan pihak terkait, seperti guru dan kepala sekolah, berdiskusi untuk menentukan data-data yang dibutuhkan.
2. Desain Sistem
Pada tahap ini, mulai merancang prototipe sistem dengan melibatkan siklus iteratif (*Prototipe*, *Tes*, dan *Refine*):
 - a. *Prototipe*: Sebuah *prototipe* awal dari aplikasi dirancang untuk memvisualisasikan alur kerja sistem, termasuk antarmuka pengguna dan proses prediksi tingkat kelulusan siswa.
 - b. *Tes*: *Prototipe* diuji oleh pengguna, seperti guru, untuk memastikan bahwa sistem sesuai dengan kebutuhan mereka. Data uji awal digunakan untuk mengukur keakuratan metode SMOTE dan K-NN.
 - c. *Refine*: Berdasarkan masukan dari pengguna, *prototipe* diperbaiki hingga memenuhi ekspektasi dan dapat dioptimalkan atau tampilan antarmuka disesuaikan agar lebih *user-friendly*.
3. Pengembangan
Setelah *prototipe* disepakati, sistem dikembangkan secara penuh menggunakan bahasa pemrograman dan *platform* yang telah ditentukan yaitu PHP dan menggunakan *database* MySQL. Pada tahap ini, semua fitur utama, seperti preprocessing data dengan SMOTE dan implementasi algoritma K-NN untuk prediksi, diintegrasikan ke dalam aplikasi.

Implementasi

Tahap akhir adalah implementasi aplikasi di SMK PAB Helvetia. Sistem ini diuji secara langsung dengan data siswa nyata untuk memprediksi tingkat kelulusan mereka. Implementasi juga melibatkan pelatihan kepada pengguna, seperti guru, agar mereka dapat menggunakan aplikasi dengan baik.

3. HASIL DAN PEMBAHASAN

Metode SMOTE (*Synthetic Minority Oversampling Technique*) adalah teknik yang digunakan untuk mengatasi masalah ketidakseimbangan data dengan cara menghasilkan sampel sintesis dari kelas minoritas. Dengan menciptakan contoh-contoh data baru yang menyerupai data asli pada kelas yang kurang terwakili, SMOTE membantu meningkatkan representativitas data dan mengurangi bias yang dapat terjadi pada proses klasifikasi. Sedangkan *K-Nearest Neighbor* (K-NN) adalah algoritma pembelajaran mesin yang melakukan pengklasifikasian berdasarkan kedekatan data. Algoritma ini menghitung jarak antara data yang akan diklasifikasikan dengan data pelatihan dan menentukan kelas berdasarkan mayoritas dari k data terdekat. Dalam konteks penelitian ini, kombinasi SMOTE dan K-NN diharapkan dapat menghasilkan sistem prediksi tingkat kelulusan siswa yang lebih akurat, karena SMOTE mengatasi permasalahan ketidakseimbangan data sedangkan K-NN memberikan hasil klasifikasi yang sederhana namun efektif. Dalam studi kasus ini, kita akan mengimplementasikan SMOTE untuk menyeimbangkan dataset yang tidak seimbang, kemudian menggunakan algoritma K-NN untuk melakukan klasifikasi. Tujuannya adalah untuk melihat bagaimana kombinasi SMOTE dan K-NN dapat meningkatkan performa model, terutama dalam memprediksi kelas minoritas. Adapun Langkah-langkah penyelesaian ya dapat dijelaskan sebagai berikut :

1. Data Kelulusan Siswa

Data kelulusan siswa ini data yang digunakan sebagai data latih pada penelitian ini, Adapun data ya dapat dilihat sebagai berikut :

Tabel 1 Data Kelulusan Siswa

No.	NISN	Siswa	Nilai US Teori	Nilai US Praktik	Nilai UTS	Nilai Prilaku	Rata-Rata	Keterangan
1	9962241017	Ahkamul Fajri	76	71	66	82	73,75	Tidak Lulus
2	9941062021	Ahmad Pauji	71	69	71	79	72,5	Tidak Lulus
3	9962241000	Andika Rahadian Febri	89	77	64	68	74,5	Tidak Lulus
4	9962241168	Anissa Zahra Setyarini	77	70	79	73	74,75	Tidak Lulus
5	9962241142	Aulia Rahma Widyowati	72	70	83	66	72,75	Tidak Lulus
6	9962241010	Chaerul Nisa	67	74	60	62	65,75	Tidak Lulus
7	9952346185	Dendi	77	78	70	66	72,75	Tidak Lulus
8	9962241028	Desi Lidiawati	76	77	72	71	74	Tidak Lulus
9	9952346132	Dewi Fitriana Sari	79	71	78	70	74,5	Tidak Lulus
10	9941061987	Elin Febriyanti Futri	70	73	70	83	74	Tidak Lulus
...	...	...	...	...	...	...	...	...
123	9971688793	Ghina Nurzanah	82	69	85	90	81,5	Lulus

2. SMOTE

SMOTE bekerja dengan cara membuat sampel sintetis untuk kelas minoritas dengan menggabungkan fitur-fitur dari sampel yang sudah ada. Dengan demikian, jumlah sampel kelas minoritas dapat ditingkatkan sehingga distribusi kelas menjadi lebih seimbang. Penjelasan penggunaan atribut mencakup “A1 = Nilai Ujian Sekolah (US) Teori”, “A2 = Nilai Ujian Sekolah (US) Praktik”, “A3 = Nilai Ujian Tengah Semester (UTS)”, dan “A4 = Nilai Perilaku Siswa”. Adapun data yang dilakukan proses dengan SMOTE dapat dijelaskan pada tabel dibawah ini :

Tabel 2 Dataset

No.	Siswa	A1	A2	A3	A4	Kelulusan
1	Ahkamul Fajri	76	71	66	82	Tidak Lulus
2	Ahmad Pauji	71	69	71	79	Tidak Lulus
3	Andika Rahadian Febri	89	77	64	68	Tidak Lulus
4	Anissa Zahra Setyarini	77	70	79	73	Tidak Lulus
5	Aulia Rahma Widyowati	72	70	83	66	Tidak Lulus
6	Chaerul Nisa	67	74	60	62	Tidak Lulus
7	Dendi	77	78	70	66	Tidak Lulus
8	Desi Lidiawati	76	77	72	71	Tidak Lulus
9	Dewi Fitriana Sari	79	71	78	70	Tidak Lulus
10	Elin Febriyanti Putri	70	73	70	83	Tidak Lulus
...	...	...	...	...	...	...
123	Ghina Nurzanah	82	69	85	90	Lulus

a. Identifikasi Ketidakseimbangan Data

Tentukan kelas mayoritas (kelas dengan sampel lebih banyak) dan kelas minoritas (kelas dengan sampel lebih sedikit).

Tabel 3 Kelas Mayoritas dan Minoritas

Lulus	Tidak Lulus
78	45

Dari tabel diatas dapat disimpulkan bawasanya yang menjadi kelas minoritas adalah kelas “Tidak Lulus”. Sehingga selanjutnya akan dilakukan pengelompokan kelas minoritas yang dapat dijelaskan pada tabel berikut:

Tabel 4 Kelas Minoritas

No.	Siswa	A1	A2	A3	A4	Kelas
1	Ahkamul Fajri	76	71	66	82	Tidak Lulus
2	Ahmad Pauji	71	69	71	79	Tidak Lulus
3	Andika Rahadian Febri	89	77	64	68	Tidak Lulus
4	Anissa Zahra Setyarini	77	70	79	73	Tidak Lulus
5	Aulia Rahma Widyowati	72	70	83	66	Tidak Lulus
6	Chaerul Nisa	67	74	60	62	Tidak Lulus
7	Dendi	77	78	70	66	Tidak Lulus
8	Desi Lidiawati	76	77	72	71	Tidak Lulus
9	Dewi Fitriana Sari	79	71	78	70	Tidak Lulus
10	Elin Febriyanti Putri	70	73	70	83	Tidak Lulus
11	Lola Lolita Nur Hafidah	67	83	64	72	Tidak Lulus
13	Mohammad Zeva	67	66	74	64	Tidak Lulus
14	Muhamad Adin Hapipi	71	78	68	77	Tidak Lulus
20	Muhammad Nasal Ekaputra	75	68	68	88	Tidak Lulus
21	Muhammad Yoga	79	68	60	67	Tidak Lulus
22	Nazwa Kurnia Yunita	69	72	69	81	Tidak Lulus
25	Ridho Ramadhan Ikhlas	84	82	72	60	Tidak Lulus
26	Rioni Aghanadi Phoenna	82	65	64	80	Tidak Lulus

28	Salsa Abni Syakilla	72	67	66	87	Tidak Lulus
29	Samuel Adinugroho	66	90	65	67	Tidak Lulus
30	Siti Hoirunisa	70	68	65	86	Tidak Lulus
31	Syahrul Muzaki	82	84	66	64	Tidak Lulus
45	Elisa Lisdiana	65	72	64	85	Tidak Lulus
53	Maulana Sahril	70	66	80	67	Tidak Lulus
57	Muhamad Asep Saepullah	70	72	62	68	Tidak Lulus
59	Muhamad Jujun Alamsyah	76	69	76	63	Tidak Lulus
60	Muhamad Nur Kosasih	78	89	70	61	Tidak Lulus
67	Retno Komalasari	65	69	61	72	Tidak Lulus
74	Adi Setiawan	75	74	69	70	Tidak Lulus
78	Ananda Anwar Sadat	66	68	73	66	Tidak Lulus
81	Balqis Agnia	66	90	66	70	Tidak Lulus
84	Dita Deviana	75	73	80	65	Tidak Lulus
91	Muhamad Irgi Fahrizi	76	80	64	78	Tidak Lulus
92	Muhammad Farhan	66	71	76	60	Tidak Lulus
96	Puri Pebriani	67	77	78	75	Tidak Lulus
97	Rena Herliana Nurfalah	69	66	78	77	Tidak Lulus
100	Rio Fauzi	68	66	74	61	Tidak Lulus
101	Rivaldy Rahmiadika	75	73	68	68	Tidak Lulus
102	Rizky Ade Maulana	86	69	76	62	Tidak Lulus
103	Rusmiyanti	74	65	62	68	Tidak Lulus
109	Yudithia Andriani Hidayat	68	69	76	76	Tidak Lulus
112	Afrizal Romansyah	71	65	67	87	Tidak Lulus
116	Anisa Nur Cahya	77	69	61	71	Tidak Lulus
118	Candra Agustina Ningsih	74	72	69	84	Tidak Lulus
121	Febi Roenita Bayan	75	72	69	81	Tidak Lulus

b. Sampel Sintetis

Sampel Sintetis (data buatan) untuk kelas minoritas dengan cara menggabungkan fitur-fitur dari sampel yang sudah ada. Teknik ini melakukan duplikasi data (*oversampling*) menciptakan data baru yang lebih beragam dan representatif. Pilih sampel secara acak pada data kelas minoritas, yaitu pada sampel ke-121 yang memiliki nilai “75”, “72”, “69”, “81”. Setelah ditentukan sampel secara acak selanjutnya lakukan hitung *Euclidean* yang menggunakan jarak 3 terdekat, Adapun proses hitungnya dapat dijelaskan sebagai berikut :

$$\begin{aligned}
 S_{1.121} &= \sqrt{(76 - 75)^2 + (71 - 72)^2 + (66 - 69)^2 + (82 - 81)^2} = 3,464 \\
 S_{2.121} &= \sqrt{(71 - 75)^2 + (69 - 72)^2 + (71 - 69)^2 + (79 - 81)^2} = 5,745 \\
 S_{3.121} &= \sqrt{(89 - 75)^2 + (77 - 72)^2 + (64 - 69)^2 + (68 - 81)^2} = 20,372 \\
 S_{4.121} &= \sqrt{(77 - 75)^2 + (70 - 72)^2 + (79 - 69)^2 + (73 - 81)^2} = 13,115 \\
 S_{5.121} &= \sqrt{(72 - 75)^2 + (70 - 72)^2 + (83 - 69)^2 + (66 - 81)^2} = 20,833 \\
 S_{6.121} &= \sqrt{(67 - 75)^2 + (74 - 72)^2 + (60 - 69)^2 + (62 - 81)^2} = 22,583 \\
 S_{7.121} &= \sqrt{(77 - 75)^2 + (78 - 72)^2 + (70 - 69)^2 + (66 - 81)^2} = 16,310 \\
 S_{8.121} &= \sqrt{(76 - 75)^2 + (77 - 72)^2 + (72 - 69)^2 + (71 - 81)^2} = 11,619 \\
 S_{9.121} &= \sqrt{(79 - 75)^2 + (71 - 72)^2 + (78 - 69)^2 + (70 - 81)^2} = 14,799 \\
 S_{10.121} &= \sqrt{(70 - 75)^2 + (73 - 72)^2 + (70 - 69)^2 + (83 - 81)^2} = 5,568 \\
 S_{11.121} &= \sqrt{(67 - 75)^2 + (83 - 72)^2 + (64 - 69)^2 + (72 - 81)^2} = 17,059 \\
 S_{13.121} &= \sqrt{(67 - 75)^2 + (66 - 72)^2 + (74 - 69)^2 + (64 - 81)^2} = 20,347 \\
 S_{14.121} &= \sqrt{(71 - 75)^2 + (78 - 72)^2 + (68 - 69)^2 + (77 - 81)^2} = 8,307
 \end{aligned}$$

$$S_{20.121} = \sqrt{(75 - 75)^2 + (68 - 72)^2 + (68 - 69)^2 + (88 - 81)^2} = 8,124$$

$$S_{21.121} = \sqrt{(79 - 75)^2 + (68 - 72)^2 + (60 - 69)^2 + (67 - 81)^2} = 17,578$$

$$S_{22.121} = \sqrt{(69 - 75)^2 + (72 - 72)^2 + (69 - 69)^2 + (81 - 81)^2} = 6,000$$

$$S_{25.121} = \sqrt{(84 - 75)^2 + (82 - 72)^2 + (72 - 69)^2 + (60 - 81)^2} = 25,120$$

$$S_{26.121} = \sqrt{(82 - 75)^2 + (65 - 72)^2 + (64 - 69)^2 + (80 - 81)^2} = 11,136$$

$$S_{28.121} = \sqrt{(72 - 75)^2 + (67 - 72)^2 + (66 - 69)^2 + (87 - 81)^2} = 8,888$$

$$S_{29.121} = \sqrt{(66 - 75)^2 + (90 - 72)^2 + (65 - 69)^2 + (67 - 81)^2} = 24,839$$

Dan seterusnya sehingga hasilnya dapat dijelaskan pada tabel berikut :

Tabel I5 Nilai Data Baru

No.	Siswa	Nilai Sintesis	Rangking
1	Ahkamul Fajri	3,464	2
2	Ahmad Pauji	5,745	4
3	Andika Rahadian Febri	20,372	34
4	Anissa Zahra Setyarini	13,115	19
5	Aulia Rahma Widyowati	20,833	35
6	Chaerul Nisa	22,583	37
7	Dendi	16,310	25
8	Desi Lidiawati	11,619	15
9	Dewi Fitriana Sari	14,799	22
10	Elin Febriyanti Putri	5,568	3
11	Lola Lolita Nur Hafidah	17,059	27
13	Mohammad Zeva	20,347	33
14	Muhamad Adin Hapiipi	8,307	7
20	Muhammad Nasal Ekaputra	8,124	6
21	Muhammad Yoga	17,578	28
22	Nazwa Kurnia Yunita	6,000	5
25	Ridho Ramadhan Ikhlas	25,120	43
26	Rioni Aghanadi Phoenaa	11,136	12
28	Salsa Abni Syakilla	8,888	8
29	Samuel Adinugroho	24,839	42
30	Siti Hoirunisa	9,055	9
31	Syahrul Muzaki	22,159	36
45	Elisa Lisdiana	11,874	16
53	Maulana Sahril	19,442	30
57	Muhamad Asep Saepullah	15,588	23
59	Muhamad Jujun Alamsyah	19,570	32
60	Muhamad Nur Kosasih	26,439	44
67	Retno Komalasari	15,937	24
74	Adi Setiawan	11,180	13
78	Ananda Anwar Sadat	18,385	29
81	Balqis Agnia	23,130	39
84	Dita Deviana	19,442	31
91	Muhamad Irgi Fahrizi	9,950	10
92	Muhammad Farhan	23,917	41
96	Puri Pebriani	14,353	21
97	Rena Herliana Nurfalah	13,000	17
100	Rio Fauzi	22,583	38
101	Rivaldy Rahmiadika	13,077	18
102	Rizky Ade Maulana	23,238	40
103	Rusmiyanti	16,371	26
109	Yudithia Andriani Hidayat	11,489	14

112	Afrizal Romansyah	10,247	11
116	Anisa Nur Cahya	13,304	20
118	Candra Agustina Ningsih	3,162	1

Setelah selesai melakukan proses hitung *Euclidean* selanjut ambil jarak terdekat menggunakan $K=3$, yang artinya mengambil nilai sintesis yang memiliki nilai rangking 1 sampai 3 yang dapat dilihat pada tabel berikut :

Tabel 6 Jarak Terdekat

No.	Siswa	Nilai Sintesis	Rangking
118	Candra Agustina Ningsih	3,162	1
1	Ahkamul Fajri	3,464	2
10	Elin Febriyanti Putri	5,568	3

c. Sampel Sintetis Baru

Tahap melakukan pembuatan sampel baru dengan menggunakan jarak terdekat yang didapat pada proses hitung *eucliden*, pada tahap proses pembuatan sampel baru menggunakan nilai acak (λ) 0 s/d 1. Misalnya acak (λ) menggunakan 0,5. Adapun proses hitungnya dapat dijelaskan sebagai berikut :

$$\begin{aligned}
 X_{118.121} &= (74 - 75) | (72 - 72) | (69 - 69) | (84 - 81) \\
 &= (-1) | (0) | (0) | (3) \times 0,5 \\
 &= (-0,5) | (0) | (0) | (1,5) \\
 &= (-0,5 + 75) | (0 + 72) | (0 + 69) | (1,5 + 81) \\
 &= (74,5) | (72) | (69) | (82,5) \\
 X_{1.121} &= (76 - 75) | (71 - 72) | (66 - 69) | (82 - 81) \\
 &= (1) | (-1) | (-3) | (1) \times 0,5 \\
 &= (0,5) | (-0,5) | (-1,5) | (0,5) \\
 &= (0,5 + 75) | (-0,5 + 72) | (-1,5 + 69) | (0,5 + 81) \\
 &= (75,5) | (71,5) | (67,5) | (81,5) \\
 X_{10.121} &= (70 - 75) | (73 - 72) | (70 - 69) | (83 - 81) \\
 &= (-5) | (1) | (1) | (2) \times 0,5 \\
 &= (-2,5) | (0,5) | (0,5) | (1) \\
 &= (-2,5 + 75) | (0,5 + 72) | (0,5 + 69) | (1 + 81) \\
 &= (72,5) | (72,5) | (69,5) | (82)
 \end{aligned}$$

Sehingga untuk hasil sampel baru berjumlah 3 sampel baru sesuai dengan penggunaan nilai $K=3$ dengan tahapan metode SMOTE dapat dilihat pada tabel berikut :

Tabel 7 Sampel Baru

No.	Siswa	A1	A2	A3	A4	Kelas
124	Sampel Baru	74,5	72	69	82,5	Tidak Lulus
125	Sampel Baru	75,5	71,5	67,5	81,5	Tidak Lulus
126	Sampel Baru	72,5	72,5	69,5	82	Tidak Lulus

3. K-NN

Selanjutnya hasil dari metode SMOTE dilakukan klasifikasi menggunakan K-NN. Dari tabel 1 dan tabel 7 diatas diambil untuk menjadi tambahan data training/latih untuk dilakukan pengujian data sehingga data yang digunakan berjumlah 126 sampel. Dimana data training/latih memiliki 126 *record* akan diklasifikasi dengan data testing/uji yang memiliki 10 *record* data akan diklasifikasi, adapun bentuk tabel data testing/uji dapat dilihat pada tabel dibawah ini :

Tabel 8 Data Testing/Uji

No.	Siswa	A1	A2	A3	A3	Kelas
1	Faisal Viqri	87	67	90	73	?
2	Fajar Andhika	66	71	67	78	?
3	Faras Kharisma	73	87	72	78	?

4	Fikri Islami	79	68	77	72	?
5	Muhamad Fadil	82	87	78	83	?
6	Muhammad Fahrian	75	73	82	90	?
7	Muhammad Rizki	75	90	87	73	?
8	Ndaru Tatawi Guna	80	90	82	66	?
9	Nia Apriliani	76	88	79	81	?
10	Nur Laeli Sofiyah	90	74	86	68	?

Tentukan parameter K dan hitung jarak *euclid* (*euclidean distance*) masing-masing obyek terdapat data sampel yang diberikan. Contoh perhitungan untuk data 126 (data trening/latih) terhadap data 10 (data testing/uji), akan diklasifikasikan, adapun contoh perhitungannya dapat dilihat sebagai berikut :

$$K1, K1 = \sqrt{(86 - 87)^2 + (81 - 67)^2 + (66 - 90)^2 + (82 - 73)^2}$$

$$K1, K1 = \sqrt{(-1)^2 + (14)^2 + (-24)^2 + (9)^2}$$

$$K1, K1 = \sqrt{1 + 196 + 576 + 81}$$

$$K1, K1 = \sqrt{854}$$

$$K1, K1 = 29,223$$

$$K2, K1 = \sqrt{(71 - 87)^2 + (69 - 67)^2 + (71 - 90)^2 + (89 - 73)^2}$$

$$K2, K1 = \sqrt{(-16)^2 + (2)^2 + (-19)^2 + (16)^2}$$

$$K2, K1 = \sqrt{256 + 4 + 361 + 256}$$

$$K2, K1 = \sqrt{877}$$

$$K2, K1 = 29,614$$

$$K3, K1 = \sqrt{(89 - 87)^2 + (87 - 67)^2 + (64 - 90)^2 + (68 - 73)^2}$$

$$K3, K1 = \sqrt{(2)^2 + (20)^2 + (-26)^2 + (-5)^2}$$

$$K3, K1 = \sqrt{4 + 400 + 676 + 25}$$

$$K3, K1 = \sqrt{1105}$$

$$K3, K1 = 32,242$$

$$K4, K1 = \sqrt{(77 - 87)^2 + (70 - 67)^2 + (89 - 90)^2 + (73 - 73)^2}$$

$$K4, K1 = \sqrt{(-10)^2 + (3)^2 + (-1)^2 + (0)^2}$$

$$K4, K1 = \sqrt{100 + 9 + 1 + 0}$$

$$K4, K1 = \sqrt{110}$$

$$K4, K1 = 10,488$$

$$K5, K1 = \sqrt{(82 - 87)^2 + (80 - 67)^2 + (83 - 90)^2 + (66 - 73)^2}$$

$$K5, K1 = \sqrt{(-5)^2 + (13)^2 + (-7)^2 + (-7)^2}$$

$$K5, K1 = \sqrt{25 + 169 + 49 + 49}$$

$$K5, K1 = \sqrt{292}$$

$$K5, K1 = 17,088$$

Dan seterusnya, adapun hasil *Eucliden Distance* dapat dilihat sebagai berikut ini :

Tabel 9 *Eucliden Distance*

No.	Siswa	Euclidean
1	Ahkamul Fajri	28,178
2	Ahmad Pauji	25,632

3	Andika Rahadian Febri	28,373
4	Anissa Zahra Setyarini	15,166
5	Aulia Rahma Widyowati	18,221
6	Chaerul Nisa	38,341
7	Dendi	25,884
8	Desi Lidiawati	23,431
9	Dewi Fitriana Sari	15,264
10	Elin Febriyanti Putri	28,723
...	...	...
123	Ghina Nurzanah	18,520
124	Sampel Baru	26,693
125	Sampel Baru	27,037
126	Sampel Baru	27,235

Dari table diatas urukan objek-objek tersebut kedalam kelompok yang mempunyai jarak *euclidian* terkecil untuk mengurutkan, berikut hasil urutan berupa ranking.

Tabel 10 Distance Ranking

No.	Siswa	Euclidean	Rank	Klasifikasi
1	Ahkamul Fajri	28,178	87	Tidak
2	Ahmad Pauji	25,632	61	Tidak
3	Andika Rahadian Febri	28,373	88	Tidak
4	Anissa Zahra Setyarini	15,166	7	Tidak
5	Aulia Rahma Widyowati	18,221	22	Tidak
6	Chaerul Nisa	38,341	123	Tidak
7	Dendi	25,884	63	Tidak
8	Desi Lidiawati	23,431	44	Tidak
9	Dewi Fitriana Sari	15,264	9	Tidak
10	Elin Febriyanti Putri	28,723	91	Tidak
...	...	...	...	...
123	Ghina Nurzanah	18,520	24	Tidak
124	Sampel Sintesis Baru	26,693	73	Tidak
125	Sampel Sintesis Baru	27,037	78	Tidak
126	Sampel Sintesis Baru	27,235	80	Tidak

Dari table diatas mengumpulkan kategori Y (klasifikasi *nearest neighbor*) pada tahap ini hanya mengambil data sesuai dengan jumlah K yang ditentukan memilih 4 data terbaik hasilnya sebagai berikut.

Tabel 11 Distance Rangkai 4 Terbaik

No	Siswa	Distance	Ranking
94	Nadhirah Putri Nasution	10,488	1
75	Afdhalul Dean Pratama	13,115	2
55	Mohamad Fadli Afdoli Nur	13,711	3
18	Muhamad Ulfi Mudriansyah	13,820	4

Dari table diatas dengan menggunakan kategori mayoritas maka dapat hasil klafikasi dari keempat data terbaik.

Tabel 12 Hasil Klasifikasi

No	Siswa	Distance	Ranking	Kelas
94	Nadhirah Putri Nasution	10,488	1	Lulus
75	Afdhalul Dean Pratama	13,115	2	Lulus
55	Mohamad Fadli Afdoli Nur	13,711	3	Lulus
18	Muhamad Ulfi Mudriansyah	13,820	4	Lulus

Dari table diatas data terbaik adalah data ke 94, 75, 55, dan 18 dari data tersebut terdapat 4 data “Lulus” dan 0 data “Tidak Lulus” sehingga mayoritas data “Lulus” lebih tinggi. Jadi hasil klafikasi dapat dilihat sebagai berikut.

Tabel 13 Hasil Detail

No.	Siswa	A1	A2	A3	A3	Kelas
1	Faisal Viqri	87	67	90	73	Lulus

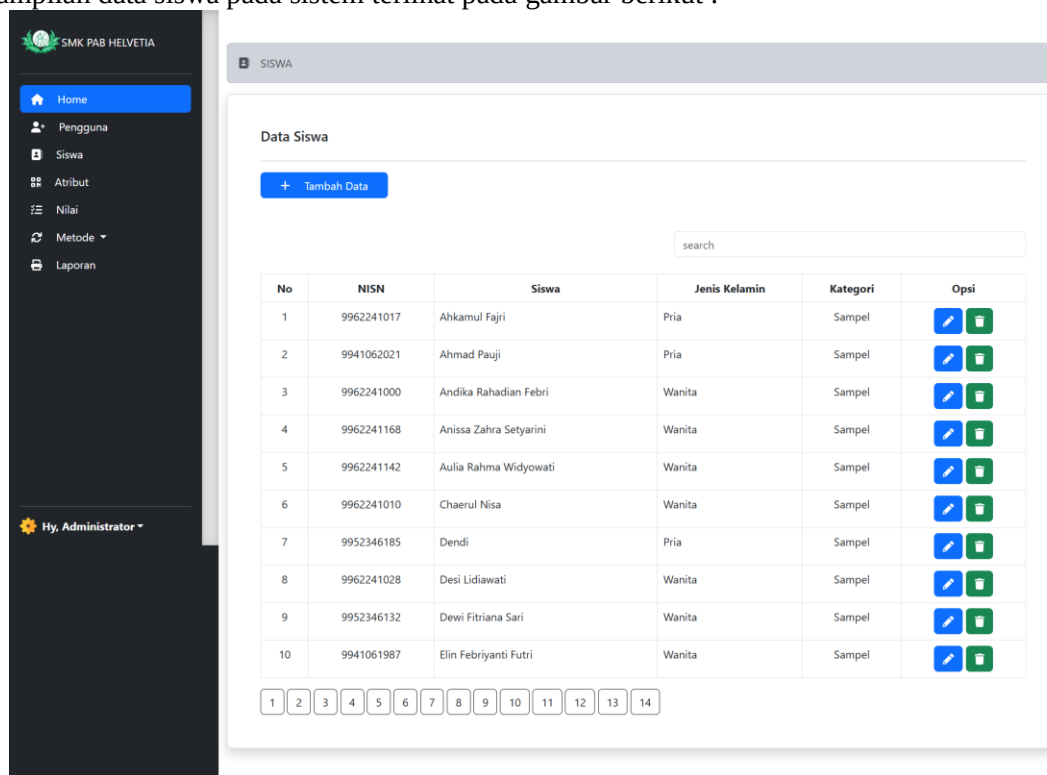
Dapat disimpulkan mayoritas untuk siswa atas nama “Faisal Viqri” masuk dalam kategori “Lulus” yang dapat dilihat pada tabel diatas.

4. Tampilan Hasil

Berikut ini dijelaskan tentang tampilan hasil dari perancangan sistem yang dibangun dapat dilihat sebagai berikut :

a. Tampilan Data Siswa

Tampilan data siswa pada sistem terlihat pada gambar berikut :

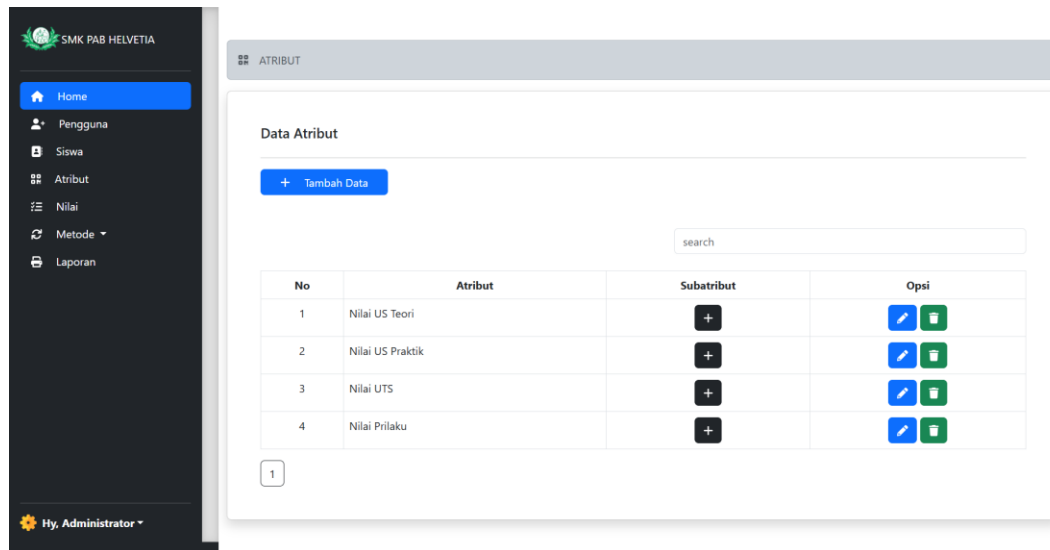


Gambar 2 Tampilan Data Siswa

Tampilan Data Siswa merupakan menu yang hanya dapat diakses oleh admin dan wali kelas untuk mengelola informasi siswa. Admin dan Wali kelas memiliki kemampuan untuk menambah, mengubah, dan menghapus data siswa sesuai kebutuhan. Data yang ditampilkan mencakup beberapa *field* utama, yaitu NISN, Nama Siswa, Jenis Kelamin, dan Kategori. Tampilan ini dirancang dengan antarmuka yang jelas dan mudah digunakan, memungkinkan admin dan wali kelas untuk mengelola data siswa secara efisien. Fitur seperti pencarian dan penyaringan data juga disediakan untuk memudahkan navigasi dan pengelolaan informasi siswa.

b. Tampilan Data Atribut

Tampilan data atribut pada sistem terlihat pada gambar berikut :

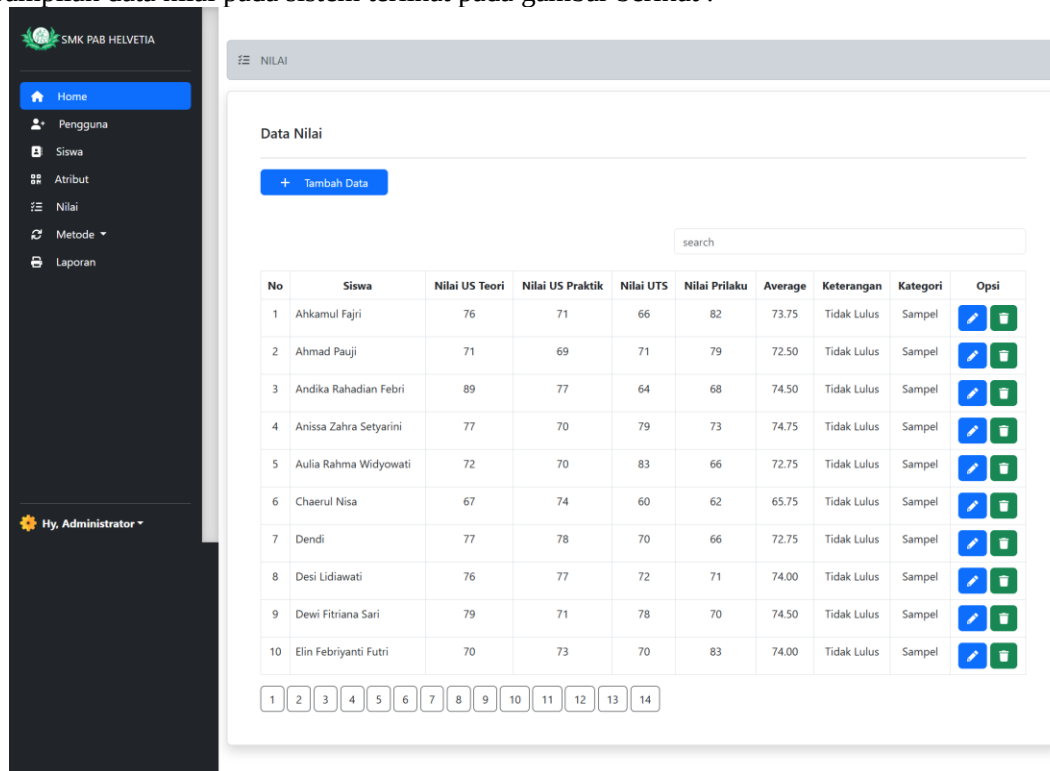


Gambar 3 Tampilan Data Atribut

Tampilan Data Atribut adalah menu khusus yang hanya dapat diakses oleh admin untuk mengelola kriteria penilaian siswa. Admin dapat menambah, mengedit, atau menghapus data atribut sesuai kebutuhan sistem. Data yang ditampilkan terdiri dari dua field utama: Atribut (kategori penilaian utama) dan Subatribut (rincian kriteria penilaian). Antarmuka dirancang sederhana namun fungsional, memungkinkan admin mengelola parameter penilaian dengan efisien. Fitur pencarian disediakan untuk memudahkan admin menemukan data tertentu.

c. Tampilan Data Nilai

Tampilan data nilai pada sistem terlihat pada gambar berikut :

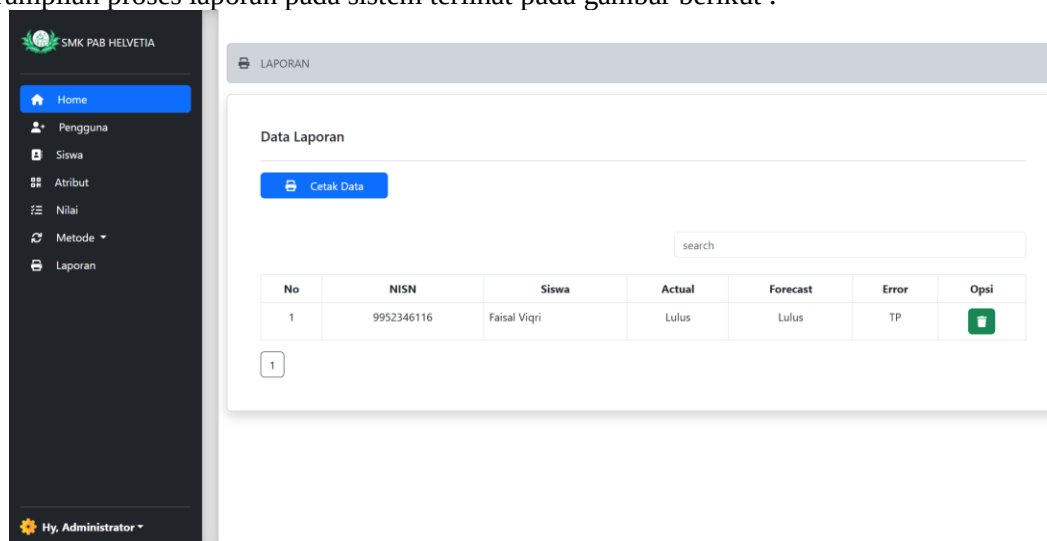


Gambar 4 Tampilan Data Nilai

Tampilan Data Nilai adalah menu khusus admin dan wali kelas untuk mengelola nilai akademik dan perilaku siswa. Admin dan Wali Kelas dapat menambah, mengubah, dan menghapus rekaman nilai sesuai kebutuhan. Menu ini menampilkan field lengkap meliputi: Siswa (nama siswa), Nilai US Teori, Nilai US Praktik, Nilai UTS, Nilai Perilaku, *Average* (rata-rata nilai), Keterangan, dan Kategori (sampel/uji). Tampilan dirancang sistematis untuk memudahkan input, analisis, dan pelacakan perkembangan akademik siswa.

d. Tampilan Laporan Smote dan K-nn

Tampilan proses laporan pada sistem terlihat pada gambar berikut :



Gambar 5 Tampilan Laporan Smote dan K-nn

Tampilan Laporan merupakan menu yang dapat diakses oleh admin dan kepala sekolah untuk melihat hasil analisis klasifikasi siswa. Menu ini menyediakan opsi cetak data untuk menghasilkan dokumen laporan dan hapus data untuk membersihkan rekaman yang tidak diperlukan. Data laporan menampilkan *field* penting meliputi NISN (Nomor Induk Siswa Nasional), Nama Siswa, *Actual* (kategori aktual siswa), *Forecast* (hasil prediksi sistem), dan *Error* (selisih antara prediksi dan kondisi sebenarnya). Tampilan ini dirancang untuk memberikan informasi jelas tentang akurasi sistem klasifikasi, memudahkan evaluasi dan pengambilan keputusan oleh pihak sekolah.

4. KESIMPULAN

Adapun kesimpulan yang dapat diambil oleh penulis adalah sebagai berikut :

1. Penerapan metode kombinasi SMOTE dan K-NN terbukti efektif untuk memprediksi tingkat kelulusan siswa di SMK PAB Helvetia, terutama dalam menangani masalah ketidakseimbangan data antara siswa yang lulus dan tidak lulus.
2. Hasil prediksi yang dihasilkan oleh sistem dapat membantu pihak sekolah dalam mengidentifikasi siswa yang berpotensi tidak lulus lebih dini, sehingga intervensi pembelajaran dapat dilakukan dengan lebih tepat sasaran.
3. Sistem prediksi ini berpotensi meningkatkan kualitas pengambilan keputusan akademik, karena mampu memberikan gambaran yang lebih akurat dan objektif terkait kondisi siswa berdasarkan data historis yang dimiliki sekolah.

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] Sari, P. J., & Kurniawan, H. (2024). Penerapan algoritma naive bayes untuk mengetahui kepuasan masyarakat terhadap layanan administrasi kantor lurah tanjung mulia hilir kecamatan medan deli. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(2), 540-551.
- [2] Kurniawan, H., & Azmi, C. (2021). Bobot 1000 Butir dan kualitas benih tujuh lot varietas cabai open pollinated (OP). *Proceedings: peningkatan produktivitas pertanian era society*, 5.
- [3] Kurniawan, H. (2023). Measurement Of Customer Satisfaction Using Fuzzy Service Quality Method At PT. ABC. *Journal of Intelligent Decision Support System (IDSS)*, 6(3), 168-177.
- [4] Sitorus, E. P. A., & Kurniawan, H. (2023). Penerapan Metode Double Exponential Smoothing Dalam Memprediksi Stok Opname Produk Susu Berbasis WEB (Studi Kasus: PT. Beringin Suka Sehat). *Jurnal Rekayasa Sistem (JUREKSI)*, 1(1), 295-307.
- [5] Yusfrizal, M. Sovina, F. A. Harahap, D. Adhar, W. Fahrozi and H. Kurniawan, "Decision Support System for Determining the Amount of Production Based on Inventory Data and the Amount of Demand using the Tsukamoto Method," 2021 IEEE 7th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Bandung, Indonesia.
- [6] Pravita, N. J., & Puspita, K. (2024). Penerapan Data Mining Dengan Menggunakan Metode K-Nearest Neighbor (KNN) Dalam Memprediksi Penjualan Produk Pada CV. Sukses Harapan Baru. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(3 A), 1689-1702.
- [7] Pratama, D. T., & Sovina, M. (2024). Pelaksanaan Metode K-Nearest Neighbor (K-NN) Dalam Sistem Pendukung Keputusan Penerima Beasiswa Pada SMK Swasta PAB 2 Helvetia Berbasis Web. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(3), 1005-1017.
- [8] Pulungan, M., & Leman, D. (2024). Penerapan Data Mining Dalam Prediksi Penjualan Sparepart Sepeda Motor Terlaris Menggunakan Metode K-Nearest Neighbor (KNN) Pada CV. Solid Mitra Cemerlang. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(3 A), 2056-2069.
- [9] Nurdian, R. A., Ridwan, M., & Yusuf, A. (2022). Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru. *Jurnal Teknik Informatika dan Sistem Informasi*, 8(1), 24-32.
- [10] Erlin, E., Desnelita, Y., Nasution, N., Suryati, L., & Zoromi, F. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(3), 677-690.
- [11] Firmansyah, A. S., Aziz, A., & Ahsan, M. (2023). Optimasi K-Nearest Neighbor Menggunakan Algoritma Smote Untuk Mengatasi Imbalance Class Pada Klasifikasi Analisis Sentimen. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3341-3347.
- [12] Astuti, F. D., & Lenti, F. N. (2021). Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN. *JUPITER: Jurnal Penelitian Ilmu dan Teknologi Komputer*, 13(1), 89-98.

-
- [13] Pramayasa, K., Maysanjaya, I. M. D., & Indradewi, I. G. A. A. D. (2023). Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE. *SINTECH (Science and Information Technology) Journal*, 6(2), 89-98.
 - [14] Wafiq, M., Muflikhah, L., & Furqon, M. T. (2022). Klasifikasi Kelayakan Calon Kreditur menggunakan Metode Syntethic Minority Over-Sampling Technique (SMOTE) dan K-Nearest Neighbor (KNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(11), 5390-5395.
 - [15] Chamzah, S. M., Lestandy, M., Kasan, N., & Nugraha, A. (2022). Penerapan Synthetic Minority Oversampling Technique (SMOTE) untuk Imbalance Class pada Data Text Menggunakan KNN. *Syntax: Jurnal Informatika*, 11(02), 56-67.
 - [16] Yuni, Y. F., & Kurniawan, H. (2023). Sistem Informasi Nilai Akademik Siswa Berbasis Web Studi Kasus: SMP IT Al Bayan. *Jurnal Rekayasa Sistem (JUREKSI)*, 1(3), 1160-1172.