

Penerapan Metode SMOTE dan K-Nearest Neighbor untuk Klasifikasi Tingkat Kepuasan Masyarakat terhadap Penyiaran Radio Most FM Medan

Application of the SMOTE and K-Nearest Neighbor Methods for Classifying Public Satisfaction Levels towards Most FM Medan Radio Broadcasting

Apprilia Faradita^{a,1}, Fitriana Harahap^{b,2}

^{a,b}Sistem Informasi, Universitas Potensi Utama, Medan 20241, Sumatera Utara, Indonesia
Email: fapprilia03@gmail.com¹, fitrianaharahap1@gmail.com²

ABSTRAK

Radio Most FM merupakan salah satu stasiun radio terkemuka di Medan yang terus berupaya meningkatkan kualitas program siaran agar tetap relevan di tengah persaingan dengan media digital. Namun, analisis tingkat kepuasan pendengar menghadapi kendala berupa ketidakseimbangan data survei, di mana kategori tertentu memiliki jumlah data yang jauh lebih sedikit dibandingkan kategori lainnya. Hal ini menyebabkan hasil analisis kurang representatif, sehingga menyulitkan manajemen untuk merancang strategi peningkatan kualitas siaran. Penelitian ini bertujuan untuk mengatasi permasalahan tersebut dengan menerapkan metode Synthetic Minority Oversampling Technique (SMOTE) untuk menangani ketidakseimbangan data, serta algoritma K-Nearest Neighbors (KNN) untuk mengklasifikasikan tingkat kepuasan pendengar. SMOTE digunakan untuk menghasilkan data sintetis pada kategori minoritas, sehingga data survei menjadi seimbang. Selanjutnya, KNN digunakan untuk mengklasifikasikan tingkat kepuasan berdasarkan dataset yang telah diseimbangkan. Hasil penelitian menunjukkan bahwa kombinasi SMOTE dan KNN mampu memberikan hasil klasifikasi yang lebih akurat dan representatif. Informasi yang dihasilkan membantu manajemen dalam memahami tingkat kepuasan pendengar secara menyeluruh, sehingga dapat digunakan sebagai dasar untuk merancang strategi peningkatan kualitas program siaran.

Kata Kunci : Data Mining, Klasifikasi, Tingkat Kepuasan, Penyiaran Radio, Metode SMOTE dan KNN.

ABSTRACT

Radio Most FM is one of the leading radio stations in Medan that continues to improve the quality of its broadcast programs to remain relevant amidst competition with digital media. However, the analysis of listener satisfaction levels faces obstacles in the form of survey data imbalance, where certain categories have much less data than other categories. This causes the analysis results to be less representative, making it difficult for management to design strategies to improve broadcast quality. This study aims to overcome this problem by applying the Synthetic Minority Oversampling Technique (SMOTE) method to handle data imbalance, and the K-Nearest Neighbors (KNN) algorithm to classify listener satisfaction levels. SMOTE is used to generate synthetic data in the minority category, so that the survey data is balanced. Furthermore, KNN is used to classify satisfaction levels based on the balanced dataset. The results of the study show that the combination of SMOTE and KNN is able to provide more accurate and representative classification results. The information generated helps management in understanding the level of listener satisfaction as a whole, so that it can be used as a basis for designing strategies to improve the quality of broadcast programs.

Keywords : Data Mining, Classification, Satisfaction Level, Radio Broadcasting, SMOTE and KNN Methods.

Disubmit: Info Artikel : Diterima :

Copyright © 2025 - senadimu. All rights reserved.

1. PENDAHULUAN

Saat ini perkembangan teknologi dan informasi sudah banyak beralih dari sistem yang manual menjadi sistem berteknologi dan disebut juga sebagai sistem terkomputerisasi [1]. Komputer merupakan alat pengolah data elektronik yang bekerja dandikendalikan oleh sekelompok instruksi-instruksi yang disebut program, digunakan untuk membantu manusia agar pekerjaannya menjadi lebih mudah, cepat dan akurat [2]. Radio Most Fm adalah salah satu stasiun radio yang telah lama menjadi bagian penting dari media penyiaran di Kota Medan. Dengan berbagai program yang inovatif, stasiun ini berupaya memberikan informasi, hiburan, dan edukasi kepada masyarakat. Dalam perjalanannya, Most FM terus berupaya menjaga relevansi dan meningkatkan kualitas siarannya di tengah persaingan dengan platform digital yang semakin populer, seperti layanan streaming musik dan podcast. Sebagai bagian dari identitas kota, Most FM juga berperan dalam menyampaikan isu-isu lokal dan memperkuat keterlibatan masyarakat Medan melalui penyiaran yang interaktif dan menarik.

Selama ini, analisis terhadap kepuasan masyarakat masih dilakukan secara manual, yang memerlukan waktu dan biaya yang cukup besar dalam pengolahannya. Pendekatan manual ini sering kali menjadi tantangan bagi pihak manajemen untuk mendapatkan hasil analisis yang cepat dan efisien. Selain itu, kepuasan masyarakat terhadap penyiaran radio dipengaruhi oleh beberapa faktor penting, seperti kualitas suara, variasi dan relevansi program siaran, kemampuan penyiar radio dalam berinteraksi dengan pendengar, pengalaman mendengarkan secara keseluruhan, serta tingkat loyalitas pendengar terhadap stasiun radio. Namun, data survei yang terkumpul sering kali tidak merata, dengan jumlah responden yang tidak seimbang di setiap kategori tersebut. Akibatnya, hasil analisis cenderung kurang representatif untuk menggambarkan kepuasan masyarakat secara keseluruhan, sehingga menyulitkan pihak manajemen dalam merumuskan langkah strategis yang tepat untuk meningkatkan kualitas penyiaran.

Teknologi data mining menawarkan solusi untuk menganalisis tingkat kepuasan masyarakat secara lebih mendalam dan representative. Data mining adalah proses menganalisis set data yang besar untuk menemukan pola, korelasi, atau anomali yang bermakna dan sebelumnya tidak diketahui [3]. Inti dari data mining adalah mengubah tumpukan data mentah (raw data) menjadi informasi berharga (knowledge) yang dapat digunakan untuk pengambilan Keputusan [4]. Data mining menggabungkan teknik dari beberapa disiplin ilmu, seperti statistika, kecerdasan buatan (AI), machine learning, dan sistem basis data [5]. Tujuannya tidak hanya mendeskripsikan data masa lalu, tetapi juga untuk memprediksi tren dan perilaku di masa depan, serta mengelompokkan data ke dalam kategori tertentu [6]. Data mining merupakan komponen kunci dalam Business Intelligence (BI) dan ilmu data, yang membantu perusahaan dalam pemasaran, penjualan, manajemen risiko, dan strategi bisnis lainnya [7].

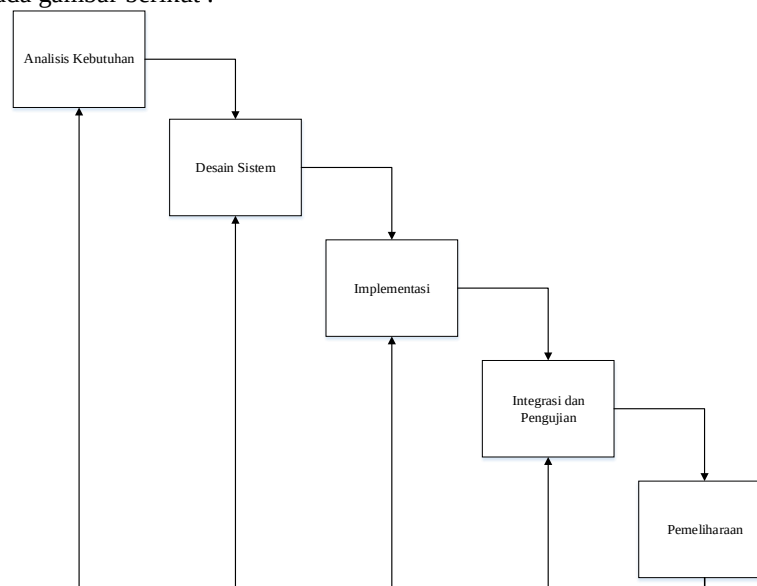
Penelitian ini menggunakan dua metode utama, yaitu Synthetic Minority Oversampling Technique (SMOTE) dan K-Nearest Neighbors (KNN). K-NN adalah metode klasifikasi yang mengelompokkan data point baru berdasarkan mayoritas kelas dari sejumlah data tetangga terdekatnya (ditentukan oleh nilai K) [8]. K-NN tidak membuat model generalisasi dari data pelatihan selama fase pelatihan. Sebaliknya, ia menyimpan semua data dan hanya melakukan perhitungan (seperti jarak Euclidean) ketika ada data baru yang perlu diklasifikasikan, sehingga disebut "lazy learner" [9]. SMOTE adalah teknik yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset dengan cara menghasilkan contoh sintetis untuk kelas minoritas, bukan hanya menduplikasi data yang sudah ada [10]. Cara kerjanya adalah dengan memilih secara acak sampel dari kelas minoritas, menemukan K-tetangga terdekatnya, lalu membuat data sintetis baru di sepanjang ruas garis yang menghubungkan sampel asli dengan tetangganya tersebut [11].

Penelitian "Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN". Hasil penelitian menunjukkan bahwa penggunaan SMOTE mampu mengatasi imbalance class dengan menaikkan nilai akurasi rata-rata sebesar 9.97% [12]. Penelitian "Prediksi Diabetes Menggunakan Algoritma K-Nearest (KNN) Teknik SMOTE-ENN". Hasil penelitian menunjukkan bahwa dengan menerapkan optimasi SMOTE-ENN pada dataset penelitian dapat

meningkatkan hasil akurasi prediksi menggunakan model algoritma KNN dengan pembagian data training 70% dan testing 30% [13]. Penelitian “Penerapan Synthetic Minority Oversampling Technique (SMOTE) untuk Imbalance Class pada Data Text Menggunakan KNN”. tian ini yaitu untuk mengetahui performansi penanganan ketidakseimbangan kelas pada data review aplikasi Tokopedia menggunakan algoritma kNN dan SMOTE [14]. Penelitian “Analisis Pengaruh SMOTE terhadap Kinerja Model KNN untuk Prediksi Risiko Stroke”. Temuan ini menegaskan bahwa pendekatan yang diimplementasikan berkinerja dengan baik dalam mendeteksi risiko stroke, meskipun set data yang digunakan memiliki ketidaksimbangan data di dalamnya [15].

2. METODE

Model *Waterfall* adalah salah satu metode pengembangan perangkat lunak yang menggunakan pendekatan sistematis dan berurutan. Dalam model ini, setiap fase pengembangan perangkat lunak harus diselesaikan sepenuhnya sebelum melanjutkan ke fase berikutnya. Fase-fase dalam model *Waterfall* biasanya meliputi analisis kebutuhan, desain sistem dan perangkat lunak, implementasi (*coding*), pengujian (*testing*), *deployment* (implementasi sistem), dan pemeliharaan (*maintenance*). Proses dalam model *Waterfall* digambarkan menyerupai aliran air terjun, di mana setiap langkah secara linear mengalir ke langkah berikutnya. Fase pertama dimulai dengan pengumpulan dan analisis kebutuhan, di mana kebutuhan dari pengguna atau klien didokumentasikan secara rinci. Selanjutnya, desain sistem dan perangkat lunak dilakukan untuk merancang struktur teknis yang akan diimplementasikan. Pada tahap implementasi, kode program dibuat berdasarkan desain yang telah dirancang yang dapat dilihat pada gambar berikut :



Gambar 1 Diagram Waterfall

1. Analisis Kebutuhan

Tahap ini berfokus pada identifikasi kebutuhan penelitian. Data yang diperlukan dikumpulkan melalui survei pendengar Most FM, yang mencakup tingkat kepuasan terhadap berbagai aspek siaran seperti konten, durasi, dan interaksi dengan penyiar. Kebutuhan lain adalah memilih metode data mining (SMOTE dan KNN) yang relevan untuk mengklasifikasikan tingkat kepuasan masyarakat, serta memastikan dataset yang akan digunakan cocok untuk diolah.

2. Desain Sistem

Pada tahap ini, dilakukan perancangan alur kerja pengolahan data. Perancangan mencakup langkah-langkah seperti *preprocessing* data (pembersihan data dan format ulang, penerapan SMOTE untuk menyeimbangkan *dataset*, dan penggunaan algoritma KNN untuk klasifikasi. Alur

kerja ini didokumentasikan secara detail untuk memastikan langkah-langkah penelitian dapat direplikasi.

3. Implementasi

Tahap implementasi dilakukan sebagai alat utama untuk data mining. Dataset survei berbentuk kuesioner, kemudian proses *preprocessing* dilakukan untuk memastikan data dalam format yang sesuai. Selanjutnya, SMOTE diterapkan untuk menyeimbangkan dataset dengan menambahkan data sintetis pada kelas minoritas. Setelah itu, algoritma KNN diterapkan untuk klasifikasi tingkat kepuasan pendengar.

4. Integrasi dan Pengujian

Hasil klasifikasi dievaluasi untuk memastikan akurasi dan keandalan. Pengujian dilakukan dengan menggunakan metode validasi silang (*cross-validation*) yang tersedia. Selain itu, hasil klasifikasi dibandingkan dengan data sebenarnya untuk menilai performa algoritma KNN. Jika hasil tidak memuaskan, parameter algoritma dioptimalkan untuk meningkatkan akurasi.

5. Pemeliharaan

Tahap pemeliharaan berfokus pada pengelolaan dataset dan alur kerja penelitian. Dataset diperbarui secara berkala untuk mencerminkan kondisi terkini tingkat kepuasan pendengar Most FM. Selain itu, peneliti dapat terus mengeksplorasi metode atau algoritma lain untuk meningkatkan hasil penelitian di masa mendatang.

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, bahan penelitian yang digunakan mencakup data *primer* dan *sekunder* yang relevan untuk mengukur tingkat kepuasan pendengar Radio Most FM. Data *primer* diperoleh melalui survei yang disebarkan kepada pendengar, dengan menggunakan kuesioner yang dirancang untuk mengukur berbagai aspek kepuasan, seperti seperti kualitas siaran, program siaran, penyiar radio, interaksi dengan pendengar, pengalaman mendengarkan, dan loyalitas pendengar. Data survei yang terkumpul kemudian direkam dalam format kuesioner sebagai dasar analisis. Selain itu, penelitian ini juga memanfaatkan bahan *sekunder* berupa literatur, jurnal, dan referensi terkait penerapan metode data mining, khususnya *Synthetic Minority Oversampling Technique* (SMOTE) dan *K-Nearest Neighbors* (KNN), yang dijadikan acuan dalam penyusunan kerangka teori dan metodologi. Dengan memanfaatkan bahan-bahan tersebut, penelitian diharapkan mampu menghasilkan analisis yang komprehensif dan representatif mengenai tingkat kepuasan masyarakat terhadap penyiaran Radio Most FM.

1. Atribut

Dalam penelitian ini, atribut merujuk pada variabel-variabel yang digunakan untuk mengukur berbagai aspek kepuasan pendengar terhadap penyiaran Radio Most FM. Atribut-atribut tersebut mencakup faktor-faktor penting dan dipilih berdasarkan relevansi dan kontribusinya terhadap persepsi kepuasan. Tabel berikut menyajikan penjelasan rinci mengenai masing-masing atribut, termasuk definisi dan skala pengukuran yang digunakan dalam penelitian. Informasi ini akan menjadi dasar dalam analisis data menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) dan *K-Nearest Neighbors* (KNN) untuk menghasilkan model klasifikasi yang akurat dan representatif. Adapun beberapa atribut yang digunakan dalam penelitian ini dapat dijelaskan pada tabel berikut :

Tabel 1 Atribut

No	Atribut
1	Kualitas Siaran
2	Program Siaran
3	Penyiar Radio
4	Interaksi dengan Pendengar
5	Pengalaman Mendengarkan
6	Loyalitas Pendengar

2. Class

Dalam penelitian ini, data kepuasan pendengar diklasifikasikan ke dalam beberapa kelas yang mencerminkan perbedaan persepsi dan tingkat kepuasan terhadap penyiaran Radio Most FM. Klasifikasi ini meliputi kategori seperti "Sangat Puas", "Puas", "Ragu-Ragu", "Tidak Puas" dan "Sangat Tidak Puas", dimana setiap kelas didasarkan pada kriteria penilaian yang telah ditetapkan. Tabel berikut menyajikan penjelasan rinci mengenai masing-masing kelas, sehingga memudahkan analisis data dengan metode *Synthetic Minority Oversampling Technique* (SMOTE) dan *K-Nearest Neighbors* (KNN) untuk menghasilkan model klasifikasi yang akurat dan representatif. Adapun beberapa *class* yang digunakan dalam penelitian ini dapat dijelaskan pada tabel berikut :

Tabel 2 Atribut

No	Class
1	Sangat Puas
2	Puas
3	Ragu-Ragu
4	Tidak Puas
5	Sangat Tidak Puas

3. Responden

Jumlah responden atau sampel yang diharapkan 100% mewakili populasi adalah jumlah anggota populasi itu sendiri. Untuk penelitian jumlah populasi yang terlalu banyak akan dijadikan sampel dengan harapan jumlah sampel yang dapat mewakili populasi yang ada. Untuk menentukan ukuran sampel bisa menggunakan Rumus Slovin

Rumus Slovin adalah sebagai berikut:

$$n = \frac{N}{1 + (N \times e^2)} \dots \dots \dots (1)$$

Dimana:

n : Ukuran sampel

N : Populasi

Sebagai contoh penghitungan Rumus Slovin:

Populasi responden pendengar pada bulan januari 2025 berjumlah 200 pendengar, maka sampel yang ambil sebagai penelitian jika menggunakan rumus Slovin dengan tingkat kepercayaan 95%, dan tingkat *error* 5% adalah:

$$N = 200 \text{ orang (jumlah populasi/responden)}$$

$$e^2 = (5\%)/100 = 0,05$$

$$e^2 = 0,05 \text{ dikuadratkan } (0,05 \times 0,05 = 0,0025)$$

$$n = \frac{200}{1 + (200 \times 0,05^2)}$$

$$n = \frac{200}{1 + (200 \times 0,0025)} = \frac{200}{1,5}$$

$$= 133,3 \text{ dibulatkan } \underline{133} \text{ responden.}$$

4. Kuesioner

Dalam penelitian ini, kuesioner disusun sebagai alat utama untuk mengumpulkan data *primer* mengenai tingkat kepuasan pendengar Radio Most FM. Kuesioner dirancang secara sistematis dengan mempertimbangkan berbagai aspek penting yang memengaruhi persepsi pendengar, seperti kualitas suara, keberagaman program, interaksi penyiar, pengalaman mendengarkan, dan loyalitas pendengar. Setiap pertanyaan menggunakan skala penilaian (Sangat Puas, Puas, Ragu-Ragu, Tidak Puas, dan

Sangat Tidak Puas) guna memberikan gambaran yang lebih terukur dan objektif terhadap kepuasan pendengar. Dengan demikian, kuesioner ini diharapkan dapat menghasilkan data yang representatif dan mendalam, yang nantinya akan menjadi dasar dalam analisis dan perumusan strategi peningkatan kualitas penyiaran di Radio Most FM. Setiap responden akan mengisi jawaban pada kuesioner yang menyediakan pilihan skala sebagai berikut: 5 = Sangat Puas, 4 = Puas, 3 = Ragu-Ragu, 2 = Tidak Puas, dan 1 = Sangat Tidak Puas. Hasil rekapitulasi kuesioner yang telah dikumpulkan dari 133 responden dapat dilihat pada tabel berikut:

Tabel 3 Hasil Kuesioner

R	A1			A2					A3			A4			A5			A6		
1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
2	5	4	5	2	2	3	4	4	3	2	5	2	3	3	5	4	2	2	5	4
3	4	5	5	3	3	2	2	4	3	2	5	3	3	2	5	4	3	1	5	3
4	4	5	5	4	3	5	5	5	4	4	5	4	5	5	5	5	4	3	5	4
5	4	4	4	4	2	5	5	5	5	3	3	5	2	3	5	2	2	2	5	5
6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
7	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
8	5	4	4	4	4	5	5	5	4	5	5	5	4	5	5	5	5	5	5	3
9	4	4	5	4	3	2	2	5	2	2	5	5	2	2	2	5	2	3	4	5
10	5	5	5	3	4	2	4	2	3	3	5	2	4	2	4	3	5	3	5	2
11	5	5	4	3	5	5	4	2	5	2	2	4	3	4	5	5	5	2	5	3
12	4	5	5	3	3	3	3	2	5	2	4	5	3	4	2	4	4	4	5	5
13	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
14	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
15	4	5	5	2	3	2	5	2	3	3	5	4	4	4	5	2	2	1	4	5
16	4	4	5	5	2	3	5	5	5	2	2	4	3	4	2	3	4	2	4	4
17	4	4	5	2	4	5	2	3	4	4	3	2	2	5	2	5	2	3	5	2
18	4	4	5	5	4	2	4	2	2	3	4	3	2	3	5	5	2	3	4	5
19	5	4	4	5	5	2	4	2	5	4	3	4	3	2	2	5	3	1	5	4
20	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
130	2	1	1	2	1	1	1	1	1	2	2	2	1	2	2	2	2	2	1	1
131	2	1	1	2	1	1	2	1	1	2	2	1	2	1	1	2	1	1	2	2
132	2	2	1	1	2	2	1	1	1	2	2	1	1	2	2	2	2	1	1	2
133	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2

Dari hasil rekapitulasi kuesioner selanjutnya data tersebut dilakukan proses hitung nilai rata-rata setiap atributnya, Adapun contoh perhitungan yang menggunakan rumus berikut :

$$R = \frac{\sum xi}{N} \dots \dots \dots (1)$$

Keterangan:

R = Rata-Rata untuk setiap pilihan jawaban.

xi = nilai skor untuk setiap pilihan jawaban

N = total jumlah responden.

a. Kualitas Siaran

$$R_{1.1} = \frac{5+5+5}{3} = 5,00$$

$$R_{2.1} = \frac{5+4+5}{3} = 4,67$$

$$R_{3.1} = \frac{4+5+5}{3} = 4,67$$

$$R_{4.1} = \frac{4+5+5}{3} = 4,67$$

$$R_{5.1} = \frac{4+4+4}{3} = 4,00$$

b. Program Siaran

$$R_{1.2} = \frac{5+5+5+5+5}{5} = 5,00$$

$$R_{2.2} = \frac{2+2+3+4+4}{5} = 3,00$$

$$R_{3.2} = \frac{3+3+2+2+4}{5} = 2,80$$

$$R_{4.2} = \frac{4+3+5+5+5}{5} = 4,40$$

$$R_{5.2} = \frac{4+2+5+5+5}{5} = 4,20$$

c. Penyiara Radio

$$R_{1.3} = \frac{5+5+5}{3} = 5,00$$

$$R_{2.3} = \frac{3+2+5}{3} = 3,33$$

$$R_{3.3} = \frac{3+2+5}{3} = 3,33$$

$$R_{4.3} = \frac{4+4+5}{3} = 4,33$$

$$R_{5.3} = \frac{5+3+3}{3} = 3,67$$

d. Interaksi dengan Pendengar

$$R_{1.4} = \frac{5+5+5}{3} = 5,00$$

$$R_{2.4} = \frac{2+3+3}{3} = 2,67$$

$$R_{3.4} = \frac{3+3+2}{3} = 2,67$$

$$R_{4.4} = \frac{4+5+5}{3} = 4,67$$

$$R_{5.4} = \frac{5+2+3}{3} = 3,33$$

e. Pengalaman Mendengarkan

$$R_{1.5} = \frac{5+5+5}{3} = 5,00$$

$$R_{2.5} = \frac{5+4+2}{3} = 3,67$$

$$R_{3.5} = \frac{5+4+3}{3} = 4,00$$

$$R_{4.5} = \frac{5+5+4}{3} = 4,67$$

$$R_{5.5} = \frac{5+2+2}{3} = 3,00$$

f. Loyalitas Pendengar

$$R_{1.6} = \frac{5+5+5}{3} = 5,00$$

$$R_{2.6} = \frac{2+5+4}{3} = 3,67$$

$$R_{3.6} = \frac{1+5+3}{3} = 3,00$$

$$R_{4.6} = \frac{3+5+4}{3} = 4,00$$

$$R_{5.6} = \frac{2+5+5}{3} = 4,00$$

Sehingga untuk hasil rata-rata keseluruhan ya dapat dilihat pada tabel berikut :

Tabel 4 Rata-Rata Hasil Kuesioner

R	A1	A2	A3	A4	A5	A6
1	5.00	5.00	5.00	5.00	5.00	5.00
2	4.67	3.00	3.33	2.67	3.67	3.67
3	4.67	2.80	3.33	2.67	4.00	3.00
4	4.67	4.40	4.33	4.67	4.67	4.00
5	4.00	4.20	3.67	3.33	3.00	4.00
6	5.00	5.00	5.00	5.00	5.00	5.00
7	5.00	5.00	5.00	5.00	5.00	5.00

8	4.33	4.60	4.67	4.67	5.00	4.33
9	4.33	3.20	3.00	3.00	3.00	4.00
10	5.00	3.00	3.67	2.67	4.00	3.33
11	4.67	3.80	3.00	3.67	5.00	3.33
12	4.67	2.80	3.67	4.00	3.33	4.67
13	5.00	5.00	5.00	5.00	5.00	5.00
14	5.00	5.00	5.00	5.00	5.00	5.00
15	4.67	2.80	3.67	4.00	3.00	3.33
16	4.33	4.00	3.00	3.67	3.00	3.33
17	4.33	3.20	3.67	3.00	3.00	3.33
18	4.33	3.40	3.00	2.67	4.00	4.00
19	4.33	3.60	4.00	3.00	3.33	3.33
20	5.00	5.00	5.00	5.00	5.00	5.00
...	...	...	...	...	...	...
130	1.33	1.20	1.67	1.67	2.00	1.33
131	1.33	1.40	1.67	1.33	1.33	1.67
132	1.67	1.40	1.67	1.33	2.00	1.33
133	1.00	1.00	1.00	1.00	1.00	1.33

Selanjutnya hitung rata-rata responden berdasarkan atribut, untuk mencari kelas dengan ketentuannya “Sangat Puas = 5”, “Puas =4”, “Ragu-Ragu =3”, “Tidak Puas =2” dan “Sangat Tidak Puas = 1”. Adapun rumus yang digunakan dapat dijelaskan sebagai berikut :

$$R = \frac{\sum xi}{N} \dots\dots\dots (2)$$

Keterangan:

R = Rata-Rata untuk setiap pilihan jawaban.

xi = nilai skor untuk setiap atribut.

N = total jumlah atribut.

$$R_1 = \frac{5,00+5,00+5,00+5,00+5,00+5,00}{6} = 5,00$$

$$R_2 = \frac{4,67+3,00+3,33+2,67+3,67+3,50}{6} = 3,50$$

$$R_3 = \frac{4,67+2,80+3,33+2,67+4,00+3,00}{6} = 3,41$$

$$R_4 = \frac{4,67+4,40+4,33+3,67+4,67+4,00}{6} = 4,46$$

$$R_5 = \frac{4,00+4,20+3,67+3,33+3,00+4,00}{6} = 3,70$$

Sehingga untuk hasil rata-rata kelas ya dapat dilihat pada tabel berikut :

Tabel 5 Rata-Rata Kelas

R	Rata-Rata	Kelas
1	5.00	Sangat Puas
2	3.50	Ragu-Ragu
3	3.41	Ragu-Ragu
4	4.46	Puas
5	3.70	Ragu-Ragu
6	5.00	Sangat Puas
7	5.00	Sangat Puas
8	4.60	Puas
9	3.42	Ragu-Ragu
10	3.61	Ragu-Ragu
11	3.91	Ragu-Ragu
12	3.86	Ragu-Ragu
13	5.00	Sangat Puas
14	5.00	Sangat Puas
15	3.58	Ragu-Ragu

16	3.56	Ragu-Ragu
17	3.42	Ragu-Ragu
18	3.57	Ragu-Ragu
19	3.60	Ragu-Ragu
20	5.00	Sangat Puas
...	...	...
130	1.53	Sangat Tidak Puas
131	1.46	Sangat Tidak Puas
132	1.57	Sangat Tidak Puas
133	1.06	Sangat Tidak Puas

5. SMOTE

SMOTE (*Synthetic Minority Over-sampling Technique*) adalah salah satu teknik yang digunakan dalam pemrosesan data untuk menangani masalah ketidakseimbangan kelas (*class imbalance*) dalam dataset. Ketidakseimbangan kelas terjadi ketika salah satu kelas (biasanya kelas minoritas) memiliki jumlah sampel yang jauh lebih sedikit dibandingkan kelas lainnya (kelas mayoritas). Hal ini dapat menyebabkan model *machine learning* cenderung bias terhadap kelas mayoritas dan kurang akurat dalam memprediksi kelas minoritas. Tahap pertama hitung jumlah setiap masing-masing kelas, Adapun hasilnya dapat dilihat pada tabel berikut :

Tabel 6 Jumlah Kelas

Sangat Puas	Puas	Ragu-Ragu	Tidak Puas	Sangat Tidak Puas
15	18	87	3	10

Pada tabel di atas, ditentukan kelas minoritas. Pada tabel tersebut, kelas yang memiliki nilai 0 tidak dihitung sehingga minoritas adalah kelas "Tidak Puas" karena jumlahnya lebih sedikit dibandingkan kelas yang lainnya. Oleh karena itu, kelas "Tidak Puas" akan dijadikan sebagai kelas minoritas. Selanjutnya, kelompokkan setiap kelas minoritas yang hasilnya dapat dilihat pada tabel di bawah ini:

Tabel 7 Kelas Minoritas

R	A1	A2	A3	A4	A5	A6	Kelas
116	2.00	2.00	2.00	2.00	2.00	2.00	Tidak Puas
117	2.00	2.00	2.00	2.00	2.00	2.00	Tidak Puas
118	2.00	2.00	2.00	2.00	2.00	2.00	Tidak Puas

Selanjutnya tentukan sampel acak dari beberapa kelas minoritas yang dapat dilihat pada tabel berikut :

Tabel 8 Sampel Acak

R	A1	A2	A3	A4	A5	A6	Kelas
133	1.00	1.00	1.00	1.00	1.00	1.33	Sangat Tidak Puas

Kemudian hitung jarak setiap kelas menggunakan jarak K=3 yang diperjelas pada rumus dibawah ini :

$$d(k_i, k_a) = \sqrt{\sum_{i=1}^n (k_i - k_a)^2} \dots \dots \dots (3)$$

Keterangan :

k, k = dua titik di ruang-n *Euclidean*.

k_i, k_a = *vektor Euclidean*, dimulai dari titik awal.

n = ruang-n, untuk atribut dengan nilai kategori.

$$K_{116.133} = \sqrt{(2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,33)^2} = 2,333$$

$$K_{117.133} = \sqrt{(2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,33)^2} = 2,333$$

$$K_{118,133} = \sqrt{(2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,00)^2 + (2,00-1,33)^2} = 2,333$$

Sehingga untuk hasil jarak yang menggunakan K=3 dapat dilihat pada tabel berikut :

Tabel 9 Nilai Jarak

R	Euclidean	Jarak
116	2.333	1
117	2.333	2
118	2.333	3

Selanjutnya ambil nilai 3 jarak terkecil yang dapat dilihat pada tabel berikut :

Tabel 10 Jarak Terkecil

R	Euclidean	Jarak
116	2.333	1
117	2.333	2
118	2.333	3

Proses mencari sampel sintetis baru dilakukan dengan mengalikan nilai acak antara 0 hingga 1. Untuk perhitungan ini, nilai acak yang digunakan adalah 0,5. Penjelasan lebih detail mengenai perhitungannya dapat dilihat dengan rumus berikut :

$$X_{new} = X_i + (X_k - X_i) \times \delta \dots\dots\dots (4)$$

Keterangan :

X_{new} = data sintetis baru.

X_i = data dari kelas minoritas.

X_k = data dari k tetangga terdekat yang memiliki jarak terdekat dengan X_i .

δ = bilangan acak antara 0 dan 1.

X116	2.00	2.00	2.00	2.00	2.00	2.00
X117	2.00	2.00	2.00	2.00	2.00	2.00
X118	2.00	2.00	2.00	2.00	2.00	2.00
-	1.00	1.00	1.00	1.00	1.00	1.33
X116	1.00	1.00	1.00	1.00	1.00	0.67
X117	1.00	1.00	1.00	1.00	1.00	0.67
X118	1.00	1.00	1.00	1.00	1.00	0.67
x	0,5	0,5	0,5	0,5	0,5	0,5
X116	0.50	0.50	0.50	0.50	0.50	0.33
X117	0.50	0.50	0.50	0.50	0.50	0.33
X118	0.50	0.50	0.50	0.50	0.50	0.33
+	2.67	3.00	3.33	2.67	3.33	4.33
X116	1.50	1.50	1.50	1.50	1.50	1.67
X117	1.50	1.50	1.50	1.50	1.50	1.67
X118	1.50	1.50	1.50	1.50	1.50	1.67

Sehingg hasil sintesis baru atau sampel baru dapat disimpulkan pada tabel berikut :

Tabel III. 11 Sampel Baru

R	A1	A2	A3	A4	A5	A6	Kelas
S.Baru	1.50	1.50	1.50	1.50	1.50	1.67	Sangat Tidak Puas
S.Baru	1.50	1.50	1.50	1.50	1.50	1.67	Sangat Tidak Puas
S.Baru	1.50	1.50	1.50	1.50	1.50	1.67	Sangat Tidak Puas

6. K-NN

K-Nearest Neighbors (K-NN) adalah salah satu algoritma dalam *machine learning* yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan cara memprediksi kelas atau nilai dari suatu data baru berdasarkan "kedekatan" (*similarity*) dengan data-data yang sudah ada dalam dataset. Dataset yang digunakan berjumlah 133 sampel, dan 3 sampel hasil *Synthetic Minority Oversampling Technique*

(SMOTE) sehingga untuk hasil keeluruhan ya 136 sampel dataset atau data latih. Sebelum melakukan tahapan K-NN tentukan terlebih dahulu sampel ujinya. Adapun sampel yang akan di uji dapat di lihat pada tabel berikut.

Tabel 12 Data Uji

R	A1	A2	A3	A4	A5	A6	Kelas
U1	3.67	3.00	4.00	4.00	3.00	2.67	?

Lakukan *Euclidean Distance* yang contoh perhitunganya dapat dijelaskan dengan rumus berikut.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \dots \dots \dots (5)$$

Dimana :

p, q = dua titik di ruang-n *Euclidean*.

qi, pi = *vektor Euclidean*, dimulai dari titik awal.

n = ruang-n, untuk atribut dengan nilai kategori.

$$d_{1,1} = \sqrt{(5,00-3,67)^2 + (5,00-3,00)^2 + (5,00-4,00)^2 + (5,00-4,00)^2 + (5,00-3,00)^2 + (5,00-2,67)^2} = 4,150$$

$$d_{2,1} = \sqrt{(4,67-3,67)^2 + (3,00-3,00)^2 + (3,33-4,00)^2 + (2,67-4,00)^2 + (3,67-3,00)^2 + (3,67-2,67)^2} = 2,160$$

$$d_{3,1} = \sqrt{(4,67-3,67)^2 + (2,80-3,00)^2 + (3,33-4,00)^2 + (2,67-4,00)^2 + (4,00-3,00)^2 + (3,00-2,67)^2} = 2,091$$

$$d_{4,1} = \sqrt{(4,67-3,67)^2 + (4,40-3,00)^2 + (4,33-4,00)^2 + (4,67-4,00)^2 + (4,67-3,00)^2 + (4,00-2,67)^2} = 2,841$$

$$d_{5,1} = \sqrt{(4,00-3,67)^2 + (4,20-3,00)^2 + (3,67-4,00)^2 + (3,33-4,00)^2 + (3,00-3,00)^2 + (4,00-2,67)^2} = 1,971$$

Sehingga hasil keseluruhan dapat dilihat pada tabel berikut :

Tabel 13 Euclidean Distance

R	Euclidean
1	4.150
2	2.160
3	2.091
4	2.841
5	1.971
6	4.150
7	4.150
8	3.267
9	2.065
10	2.261
11	2.682
12	2.294
13	4.150
14	4.150
15	1.263
16	1.732
17	1.428
18	2.482
19	1.536
20	4.150
...	...
130	4.728
131	4.933
132	4.679
133	5.907

S.Baru	4.764
S.Baru	4.764
S.Baru	4.764

Setelah *Euclidean Distance*, K data dengan jarak terdekat dipilih sebagai tetangga terdekat. K adalah jumlah tetangga terdekat yang akan digunakan untuk memprediksi kelas atau nilai data baru. Nilai K = 3 yang digunakan, maka algoritma akan mencari 3 data terdekat dari data baru tersebut. Untuk hasilnya dapat dilihat pada tabel berikut :

Tabel 14 Jarak K=3

R	Euclidean	Jarak
81	0.943	1
28	1.000	2
55	1.020	3

Untuk klasifikasi, kelas data baru ditentukan berdasarkan mayoritas kelas dari K tetangga terdekat.

Tabel 15 Data Jarak K=3

R	A1	A2	A3	A4	A5	A6	Kelas
81	4.33	3.00	3.67	3.67	3.33	3.00	Ragu-Ragu
28	4.33	3.00	4.00	4.00	3.67	2.33	Ragu-Ragu
55	4.33	3.20	3.67	4.00	3.00	3.33	Ragu-Ragu

Sehingga dapat disimpulkan kelas mayoritas adalah kelas “Ragu-Ragu” memiliki jumlah 3, sedangkan kelas yang lain tidak muncul. Sehingga pada data yang diuji masuk kedalam kelas Ragu-Ragu”, karena jumlah kemunculan ya lebih banyak dari pada kelas “Ragu-Ragu”.

Tabel 16 Hasil Klasifikasi

R	A1	A2	A3	A4	A5	A6	Kelas
U1	3.67	3.00	4.00	4.00	3.00	2.67	Ragu-Ragu

4. KESIMPULAN

1. Penelitian ini berhasil membangun sistem klasifikasi berbasis website yang mampu mengolah dan menganalisis data survei tingkat kepuasan masyarakat terhadap siaran radio Most FM.
2. Metode SMOTE yang diterapkan dalam proses prapemrosesan data mampu mengatasi ketidakseimbangan jumlah data antara kelas “Sangat Puas”, “Puas”, “Kurang Puas” dan “Tidak Puas”, sehingga meningkatkan representasi data saat dilakukan klasifikasi, dengan nilai K=3 digunakan untuk menentukan kelas dari data baru masuk dalam kelas “Kurang Puas”, karena 2 dari 3 tetangga terdekatnya juga berada dalam kelas “Kurang Puas”, sedangkan 1 sisanya berada dalam kelas “Puas”.
3. Secara khusus, hasil klasifikasi menunjukkan bahwa sistem mampu mengidentifikasi data responden baru ke dalam kelas yang tepat, yaitu kelas “Kurang Puas”, dengan mempertimbangkan nilai rata-rata atribut seperti Kualitas Siaran, Program Siaran, Penyiar, Interaksi, Pengalaman, dan Loyalitas.

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] Puspitasari, C., Jessima, J., Sahara, P., Tambunan, S. N., Andriani, S., & Harahap, F. (2023). Analisa dan Perancangan Sistem Informasi Eksekutif Pemasaran Pada Egg Waffle Bunda. *Jurnal Unitek*, 16(2), 207-216.
- [2] Harahap, F., Kurniawan, H., & Aliyah, S. (2021). Pengenalan Basis Data Microsoft Acces Bagi Siswa Smk Swabina Karya. *Publikasi Pengabdian Masyarakat*, 1(1), 07-12.
- [3]
- [4] Harahap, F., Saragih, N. E., Situmeang, E. D. P., Tuti, E., Ginting, E., & Fahrozi, W. (2022). Implementasi Data Mining dalam Memprediksi Stok Herbal menggunakan Algoritma Apriori. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(2), 1159-1165.
- [5]
- [6] Harahap, F., Saragih, N. E., Siregar, E. T., & Sariangsah, H. (2021). Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Dalam Memprediksi Pembelian Cat. *Jurnal Ilmiah Informatika*, 9(01), 19-23.
- [7] Harahap, F., Fahrozi, W., Adawiyah, R., Siregar, E. T., & Harahap, A. Y. N. (2023). Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes. *Jurnal Unitek*, 16(1), 41-51.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] Damanik, M. H. U., & Hts, D. I. G. (2024). Penerapan Metode KNN Dalam Data Mining Prediksi Penjualan Produk pada CV. Chavasindo Berbasis Web. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(3 A), 2123-2135.
- [12] Normah, N., Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE. *Jurnal Teknik Komputer*, 8(2), 174-180.
- [13] 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.
- [14] 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.
- [15] Amri, Z., Rodi, M., Wathani, M. N., & Bagja, A. (2025). Prediksi Diabetes Menggunakan Algoritma K-Nearest (KNN) Teknik SMOTE-ENN. *Infotek: Jurnal Informatika dan Teknologi*, 8(1), 193-204.
- [16] 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.

- [17] Paramita, C., Simbolon, C. S., Pamungkas, A. S., Triono, J. M., Utomo, E. P. W., & Subhiyakto, E. R. (2025). Analisis Pengaruh SMOTE terhadap Kinerja Model KNN untuk Prediksi Risiko Stroke. *Jurnal Informatika: Jurnal Pengembangan IT*, 10(4), 978-988.