

Penerapan Support Vector Machine untuk Analisis Sentimen Pengguna X terhadap IndiHome, Biznet, dan Starlink

Zhevin Alfian*, M Afdal, Rice Novita, Zarnelly Zarnelly

Fakultas Sains dan Teknologi, Prodi Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹*12150311690@students.uin-suska.ac.id, ²m.afdal@uin-suska.ac.id, ³rice.novita@uin-suska.ac.id,

⁴zarnelly@uin-suska.ac.id

Email Penulis Korespondensi: 12150311690@students.uin-suska.ac.id

Submitted: 23/05/2025; Accepted: 01/09/2025; Published: 02/09/2025

Abstrak—Penelitian ini bertujuan untuk menganalisis sentimen pengguna media sosial X terhadap tiga penyedia layanan internet di Indonesia yaitu, IndiHome, Biznet, dan Starlink. Analisis difokuskan pada lima variabel utama, yaitu kecepatan internet, kestabilan jaringan, harga dan paket layanan, kualitas layanan pelanggan, serta ketersediaan dan jangkauan wilayah. Sebanyak 4.500 data dikumpulkan melalui crawling, kemudian diproses dengan teknik text mining dan algoritma Support Vector Machine (SVM), serta diimbangi menggunakan metode Random Oversampling. Hasil evaluasi menunjukkan bahwa IndiHome secara konsisten mencatatkan performa terbaik, dengan nilai akurasi mencapai 90% pada variabel kualitas layanan pelanggan, serta rata-rata akurasi keseluruhan di atas 85% pada seluruh variabel. Biznet umumnya berada di posisi kedua, dengan akurasi berkisar antara 63% hingga 80%. Starlink menempati posisi terbawah, meskipun mencatatkan hasil kompetitif seperti akurasi 82% pada variabel kecepatan internet. Penerapan Random Oversampling meningkatkan akurasi klasifikasi model sebesar rata-rata 6–12% dibandingkan model tanpa oversampling. Penelitian ini memberikan kontribusi strategis dalam memahami opini publik terhadap layanan internet dan dapat menjadi referensi dalam perbaikan mutu layanan berbasis data.

Kata Kunci: Analisis Sentimen; Biznet; IndiHome; Random Oversampling; Starlink; Support Vector Machine; Text Mining; X

Abstract—This study aims to analyze user sentiment on the social media platform X toward three major internet service providers in Indonesia, IndiHome, Biznet, and Starlink. The analysis focuses on five key variables: internet speed, network stability, pricing and service packages, customer service quality, and coverage availability. A total of 4,500 data points were collected through data crawling, then processed using text mining techniques and the Support Vector Machine (SVM) algorithm, with data imbalance addressed through the Random Oversampling method. Evaluation results show that IndiHome consistently demonstrated the best performance, achieving an accuracy of up to 90% in the customer service quality variable, and an overall average accuracy above 85% across all variables. Biznet generally ranked second, with accuracy ranging from 63% to 80%. Starlink placed lowest overall, although it still recorded competitive results, such as 82% accuracy in the internet speed variable. The application of Random Oversampling improved the model's classification accuracy by an average of 6–12% compared to the non-oversampling model. This study offers strategic insights into public perception of internet services and can serve as a reference for improving service quality based on data-driven user feedback.

Keywords: Sentiment Analysis; Biznet; IndiHome; Random Oversampling; Starlink; Support Vector Machine; Text Mining; X

1. PENDAHULUAN

Perkembangan teknologi digital dan internet telah mengubah banyak hal dalam kehidupan sehari-hari, termasuk dalam cara masyarakat mengakses informasi dan berkomunikasi [1]. Internet telah menjadi kebutuhan penting dalam segala sisi kehidupan, mulai dari pendidikan, pekerjaan, hiburan, hingga komunikasi [2]. Pada tahun 2024, pengguna internet di Indonesia mencapai sekitar 221,5 juta jiwa atau 79,5% dari total populasi sebesar 282,4 juta. Angka ini mengalami peningkatan 1,4% dibandingkan tahun sebelumnya yang tercatat sebesar 78,19% [3]. Di Indonesia, sejumlah penyedia layanan internet seperti Starlink, IndiHome, dan Biznet bersaing dalam menyediakan konektivitas terbaik bagi pengguna.

IndiHome adalah layanan internet berbasis fiber optik milik PT Telkom Indonesia dengan jangkauan luas [4]. Biznet dikenal sebagai penyedia internet broadband yang fokus pada kecepatan dan stabilitas koneksi, baik untuk kebutuhan rumah tangga maupun bisnis [5]. Starlink menawarkan internet berbasis *Low-Earth orbit* (LEO) yang menjangkau daerah terpencil tanpa infrastruktur kabel [6]. Masing-masing provider memiliki keunggulan dan tantangan tersendiri dalam bersaing sebagai penyedia layanan internet di Indonesia.

Rilisnya Starlink di Indonesia menimbulkan sejumlah permasalahan yang cukup kompleks, terutama dalam konteks persaingan pasar dan regulasi. Kehadiran Starlink sebagai penyedia internet berbasis satelit yang mampu menjangkau daerah terpencil tanpa infrastruktur kabel memberikan keunggulan signifikan dibandingkan provider lokal seperti IndiHome dan Biznet. Hal ini memunculkan kekhawatiran mengenai potensi dominasi pasar oleh perusahaan asing, yang dapat mengancam keberlangsungan penyedia layanan [7]. Di tengah persaingan layanan tersebut, respons publik yang beredar di platform media sosial menjadi salah satu indikator penting untuk mengukur persepsi dan kepuasan pengguna.

Media sosial X merupakan salah satu platform yang paling banyak digunakan masyarakat untuk mengekspresikan opini dan pengalaman mereka terhadap berbagai layanan, termasuk layanan internet [8]. Sentimen yang muncul di platform ini bisa menjadi cerminan kepuasan atau kekecewaan pengguna terhadap suatu produk atau layanan [9]. Informasi tersebut memiliki potensi besar untuk dianalisis sebagai umpan balik bagi perusahaan penyedia layanan internet dalam meningkatkan kualitas layanan mereka [10]. Opini-opini ini dapat diolah dan dianalisis

menggunakan pendekatan *text mining* dan *machine learning*, khususnya analisis sentimen, untuk mengidentifikasi kecenderungan sentimen positif, negatif, atau netral terhadap masing-masing penyedia layanan.

Salah satu algoritma yang efektif dalam klasifikasi teks adalah *Support Vector Machine* (SVM). Algoritma ini dikenal memiliki performa yang baik dalam memisahkan data berdimensi tinggi dan cocok digunakan dalam tugas-tugas klasifikasi [11]. Dengan menggabungkan data dari media sosial dan metode klasifikasi berbasis SVM, penelitian ini bertujuan untuk mengetahui bagaimana persepsi masyarakat Indonesia terhadap layanan Starlink, IndiHome, dan Biznet, serta membandingkan kecenderungan sentimen yang muncul terhadap ketiganya.

Penelitian sebelumnya telah dilakukan untuk menganalisis sentimen pengguna terhadap layanan internet menggunakan pendekatan *machine learning*, salah satunya yang dilakukan oleh Nugroho dan Tedi Kurniadi (2024) [9], yang berjudul “*Sentiment Analysis of Starlink on Twitter Using Support Vector Machine Algorithm*” diperoleh hasil yang menunjukkan bahwa 56,3% orang setuju dengan kehadiran starlink di Indonesia. Sedangkan dari penggunaan algoritma *Support Vector Machine*, *accuracy* yang dihasilkan adalah 76,22%, untuk *Precision* sebesar 77,48% dan untuk *Recall* sebesar 81,38%.

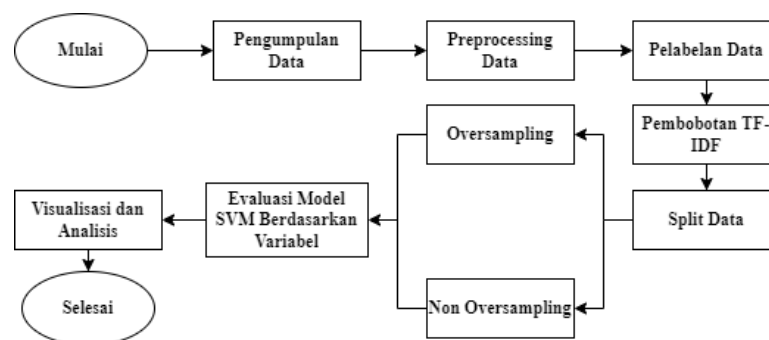
Akbar et al. (2022) [12] melakukan penelitian analisis sentimen terhadap layanan IndiHome dengan menggunakan algoritma *Support Vector Machine* (SVM). Data diambil dari Twitter dan terdiri atas 452 tweet, yang diklasifikasikan secara manual menjadi 338 tweet positif dan 114 tweet negatif. Hasil pengujian model menunjukkan tingkat akurasi mencapai 91,3%, dengan *precision* 100%, *recall* 88%, dan *F1-score* sebesar 93%.

Haranto dan Sari (2019) [5] menerapkan algoritma *Support Vector Machine* (SVM) untuk menganalisis sentimen pengguna Twitter terhadap layanan Telkom dan Biznet. Model dikembangkan menggunakan 500 tweet, masing-masing 250 tweet untuk Telkom dan Biznet. Hasil evaluasi menunjukkan bahwa untuk Telkom, SVM menghasilkan akurasi sebesar 79,6%, *precision* 76,5%, *recall* 72,8%, dan *F1-score* 74,6%. Sementara itu, untuk Biznet diperoleh akurasi 83,2%, *precision* 78,8%, *recall* 71,6%, dan *F1-score* 75%.

Penelitian ini menganalisis dan membandingkan ketiga provider berdasarkan lima variabel utama, yaitu kecepatan internet, kestabilan jaringan, harga dan paket layanan, kualitas layanan pelanggan, serta ketersediaan dan jangkauan wilayah. Penelitian ini diharapkan mampu menggambarkan secara jelas dan lebih objektif mengenai posisi dan citra masing-masing penyedia layanan internet di mata pengguna.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Gambar 1 merupakan tahapan-tahapan yang dilakukan dalam penelitian ini. Penelitian dimulai dengan memperoleh data sentimen dari pengguna X menggunakan metode *crawling data*, yang menghasilkan total 4.500 data. Data ini kemudian dibagi ke dalam lima variabel yang telah ditentukan, di mana masing-masing variabel terdiri dari 900 data. Setelah data dikumpulkan, tahap selanjutnya adalah *preprocessing* yang mencakup proses *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* untuk membersihkan dan menyiapkan data agar siap untuk dianalisis. Selanjutnya, data diberi label sentimen secara manual oleh ahli bahasa untuk memastikan akurasi pelabelan. Kemudian, dilakukan pembobotan menggunakan metode *TF-IDF* guna mengevaluasi tingkat kepentingan suatu kata dalam sebuah dokumen. Setelah pembobotan, data dibagi menjadi data latih dan data uji. Proses pengujian dilakukan dalam dua tahap menggunakan algoritma *Support Vector Machine* (SVM). Tahap pertama adalah pengujian dengan algoritma SVM tanpa metode *oversampling*, sementara tahap kedua menggunakan metode *Random Oversampling* untuk mengatasi masalah ketidakseimbangan data. Terakhir, hasil klasifikasi dievaluasi berdasarkan variabel-variabel yang telah ditentukan dan divisualisasikan untuk dianalisis lebih lanjut sebelum penelitian diselesaikan.

2.2 Preprocessing Data

Preprocessing merupakan langkah awal untuk mengolah data mentah agar menjadi data yang siap diproses lebih lanjut [13]. Tahapan ini mencakup beberapa proses, yaitu, pembersihan data (*cleaning*), mengubah seluruh teks menjadi

huruf kecil (*case folding*), memisahkan kata-kata dalam kalimat (*tokenizing*), menghapus kata-kata umum yang tidak memiliki makna signifikan (*stopword removal*), serta mengembalikan kata ke bentuk dasarnya (*stemming*) [14].

2.3 TF-IDF (*Term Frequency - Inverse Document Frequency*)

TF-IDF digunakan untuk menghitung pengaruh atau bobot sebuah kata dalam struktur dokumen. Setiap kata dalam dokumen dianalisis menggunakan metode ini untuk memperoleh nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF)[15] *Term Frequency* (TF) adalah aspek yang menentukan bobot sebuah kata dalam sebuah dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Saat memberikan bobot pada sebuah kata, nilai dari jumlah kemunculan kata tersebut harus diperhitungkan. Semakin besar bobot yang diberikan pada kata tersebut, maka semakin besar pula bobot pada dokumen, atau memberikan nilai aplikasi yang lebih tinggi[16]. Sedangkan, *Inverse Document Frequency* (IDF) adalah salah satu aspek yang berfungsi untuk mengurangi dominasi istilah-istilah yang sering muncul dalam dokumen. Hal ini harus dilakukan karena istilah-istilah yang sering muncul dapat dikategorikan sebagai istilah umum, sehingga dianggap kurang penting [17].

$$W_{ij} = tf_{ij} \times \log \left(\frac{N}{df_i} \right) \quad (1)$$

Persamaan (1) mendefinisikan W_{ij} sebagai bobot perhitungan kata i yang berasal dari dokumen j , di mana tf_{ij} merepresentasikan frekuensi kemunculan kata i dalam dokumen j , df_i menunjukkan jumlah dokumen yang memuat kata i , dan N adalah total jumlah dokumen dalam korpus [13].

2.4 Random Oversampling

Random Oversampling adalah salah satu teknik yang dipakai untuk menangani masalah ketimpangan kelas dalam dataset. Ketimpangan kelas muncul saat jumlah data pada satu kelas jauh lebih banyak dibandingkan kelas lainnya, yang dapat menyebabkan model bias terhadap kelas mayoritas [18]. *Random oversampling* dilakukan dengan menambah jumlah data pada kelas minoritas secara acak agar jumlahnya menjadi sama dengan jumlah data pada kelas mayoritas. Proses ini dilakukan dengan menggandakan sampel dari kelas minoritas secara acak hingga tercapai rasio kelas yang diinginkan [19].

2.5 Support Vector Machine (SVM)

SVM atau *Support Vector Machine* berperan sebagai metode klasifikasi dalam machine learning dengan membentuk model dari hasil training untuk memetakan data ke dalam kelas tertentu [11]. SVM mentransformasikan ruang pembelajaran menggunakan fungsi kernel untuk mengatasi masalah linear. Prinsip utamanya adalah menentukan hyperplane terbaik yang menjadi batas pemisah yang digunakan untuk membedakan dua kelas yang ada [12]. Prinsip kerja SVM adalah menemukan hyperplane terbaik yang memisahkan dua kelas dengan margin yang paling luas. Margin diartikan sebagai jarak antara hyperplane dan titik data terdekat, sedangkan titik-titik terdekat tersebut dikenal sebagai support vector [5]. Dalam algoritma SVM, pemilihan kernel memegang peranan krusial karena mengatur ruang representasi data tempat fungsi klasifikasi akan dibentuk. Melalui teknik *kernel trick*, kita dapat memilih jenis kernel yang sesuai seperti *linear*, *rbf*, *polynomial*, atau *sigmoid* untuk menyesuaikan model dengan karakteristik data yang berbeda-beda[20]. Berikut rumus dari persamaan kernel SVM:

a. Kernel *linear* :

$$K(x, y) = x \cdot y \quad (2)$$

Kernel *linear* menggunakan operasi *dot product* antara dua vektor dan digunakan ketika data dapat dipisahkan secara *linear* tanpa perlu transformasi ruang fitur [20].

b. Kernel *rbf* :

$$K(x, y) = e^{-|x-y|^2/2\sigma^2} \quad (3)$$

Kernel *rbf* mengukur kemiripan antara dua titik berdasarkan jarak *Euclidean* dan sangat efektif untuk data yang tidak dapat dipisahkan secara *linear* [20].

c. Kernel *sigmoid* :

$$K(x, y) = \tanh(Kx \cdot y - \delta) \quad (4)$$

Kernel *sigmoid* terinspirasi dari fungsi aktivasi dalam *neural network* dan digunakan untuk pola data yang memiliki distribusi kompleks [20].

d. Kernel *polynomial* :

$$K(x, y) = (x \cdot y + 1)^p \quad (5)$$

Kernel *polynomial* berguna untuk memetakan data ke ruang berdimensi lebih tinggi, sehingga cocok untuk data dengan hubungan *non-linear* yang masih terstruktur. [20].

3. HASIL DAN PEMBAHASAN

3.1 Tahap Pengumpulan Data

Data dalam penelitian ini dikumpulkan dengan mengekstraksi sentimen pengguna dari platform media sosial X yang membahas layanan internet dari IndiHome, Biznet, dan Starlink menggunakan metode *Crawling Data*. Data yang dikumpulkan difokuskan pada lima variabel utama, yaitu kecepatan internet, stabilitas jaringan, harga dan paket layanan, kualitas layanan pelanggan, serta ketersediaan dan jangkauan wilayah. Untuk memastikan representasi yang seimbang dan mendalam, setiap variabel dikumpulkan sebanyak 900 data yang kemudian dibagi lagi menjadi 3 dengan masing-masing provider menggunakan 300 data per variabel, sehingga total data yang dianalisis mencapai 4.500 data. Data yang digunakan memiliki rentang waktu dari tanggal 19 mei 2024 sampai dengan 31 maret 2025.

3.2 Tahap Preprocessing

Proses *Preprocessing* digunakan untuk merapikan data mentah menjadi format yang dapat diproses untuk tahapan selanjutnya. Tahapan *Preprocessing* terdiri dari *Cleaning*, *Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Stemming*. Tahapan *Preprocessing* dapat dilihat pada Tabel 1.

Tabel 1. *Preprocessing* Data Sentimen (Sampel Dataset Starlink)

Tahapan	Hasil
Data Awal	@SantuySaja89 @Starlink Saya pakai di gunung tinggi puncak bukit Barisan namanya gunung Sae sae.. Bisa tuh pakai 2 unit Starlink dengan 84 Device.. 35 KK Kencang kali untuk download dan nonton.
<i>Cleaning</i>	Saya pakai di gunung tinggi puncak bukit Barisan namanya gunung Sae sae Bisa tuh pakai unit Starlink dengan Device KK Kencang kali untuk download dan nonton
<i>Case Folding</i>	saya pakai di gunung tinggi puncak bukit barisan namanya gunung sae sae bisa tuh pakai unit starlink dengan device kk kencang kali untuk download dan nonton
<i>Tokenizing</i>	['saya', 'pakai', 'di', 'gunung', 'tinggi', 'puncak', 'bukit', 'barisan', 'namanya', 'gunung', 'sae', 'sae', 'bisa', 'tuh', 'pakai', 'unit', 'starlink', 'dengan', 'device', 'kk', 'kencang', 'kali', 'untuk', 'download', 'dan', 'nonton']
<i>Stopword Removal</i>	['pakai', 'gunung', 'puncak', 'bukit', 'barisan', 'namanya', 'gunung', 'sae', 'sae', 'tuh', 'pakai', 'unit', 'starlink', 'device', 'kk', 'kencang', 'kali', 'download', 'nonton']
<i>Stemming</i>	['pakai', 'gunung', 'puncak', 'bukit', 'baris', 'nama', 'gunung', 'sae', 'sae', 'tuh', 'pakai', 'unit', 'starlink', 'device', 'kk', 'kencang', 'kali', 'download', 'nonton']

3.3 Tahap TF-IDF

TF-IDF adalah metode yang berguna untuk menghitung seberapa penting atau krusial suatu kata dalam sebuah dokumen relatif terhadap kumpulan dokumen lainnya.. Sampel hasil TF-IDF dalam penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. Hasil TF-IDF (Sampel Dataset IndiHome pada Variabel Kecepatan Internet)

No.	indihome	internet	lambat	lancar	lemot	wifi	...
1	0,050	0,160	0,367	0,000	0,000	0,206	...
2	0,071	0,000	0,141	0,000	0,091	0,000	...
3	0,056	0,000	0,000	0,299	0,144	0,443	...
4	0,000	0,000	0,000	0,268	0,155	0,262	...
5	0,084	0,000	0,324	0,000	0,000	0,218	...
...
300	0,077	0,156	0,168	0,000	0,175	0,000	...

3.4 Tahap Pelabelan Data

Pelabelan data dilakukan menggunakan metode pakar dimana dataset diberi label secara manual oleh ahli Bahasa Indonesia. Pelabelan bertujuan untuk memberi label positif, negatif, dan netral pada data. Dalam penelitian ini data diberikan label berdasarkan 5 variabel utama yang digunakan. Pelabelan data berdasarkan variabel dapat dilihat pada Tabel 3.

Tabel 3. Pelabelan Data Berdasarkan Variabel

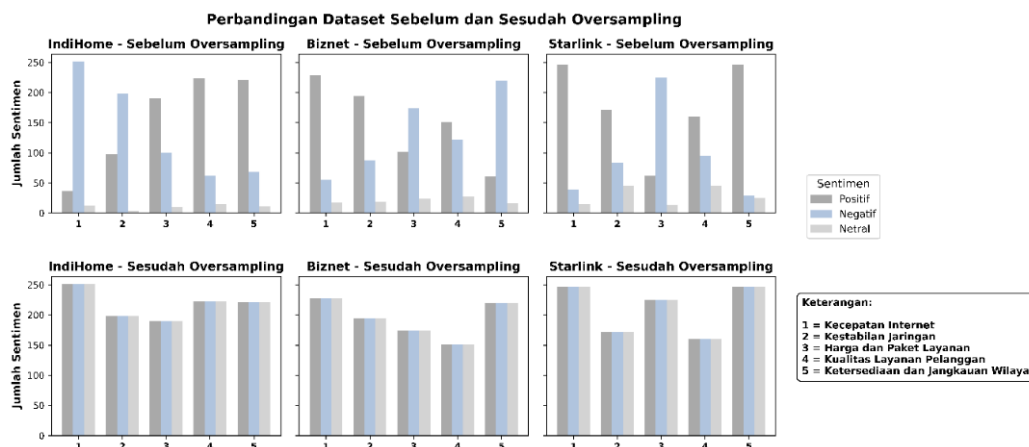
Variabel	IndiHome			Biznet			Starlink		
	Positif	Negatif	Netral	Positif	Negatif	Netral	Positif	Negatif	Netral
Kecepatan Internet	37	251	12	228	55	17	246	39	15
Kestabilan Jaringan	98	198	4	194	87	19	171	84	45
Harga dan Paket Layanan	190	100	10	102	174	24	62	225	13
Kualitas Layanan Pelanggan	223	62	15	151	122	27	160	95	45

Variabel	IndiHome			Biznet			Starlink		
	Positif	Negatif	Netral	Positif	Negatif	Netral	Positif	Negatif	Netral
Ketersediaan dan Jangkauan Wilayah	221	68	11	61	220	16	246	29	25

Tabel 3 menunjukkan bahwa sentimen pengguna terhadap IndiHome didominasi oleh respons negatif pada variabel Kecepatan Internet (251 data), namun mendapat respons positif tertinggi pada Kualitas Layanan Pelanggan (223) dan Ketersediaan serta Jangkauan Wilayah (221). Starlink memperoleh respons positif tertinggi pada variabel Ketersediaan Wilayah (246), disusul oleh Kecepatan Internet dengan jumlah yang sama (246), menandakan kepuasan pengguna terhadap aspek ini. Namun, Starlink juga menerima banyak keluhan pada variabel Harga dan Paket Layanan (225) serta Kestabilan Jaringan (84). Sentimen netral paling banyak muncul pada Starlink, terutama pada variabel Kestabilan Jaringan (45) dan Kualitas Layanan Pelanggan (45), menunjukkan adanya ambiguitas penilaian pengguna pada aspek tersebut. Di sisi lain, Biznet menunjukkan distribusi sentimen yang lebih seimbang, meskipun memperoleh cukup banyak sentimen negatif pada Harga dan Paket Layanan (174).

3.5 Tahap Pembagian Data dan Penerapan *Oversampling*

Sebelum dilakukan proses pelatihan dan evaluasi menggunakan algoritma *Support Vector Machine* (SVM), data akan di split atau dibagi menjadi 2 bagian, yaitu data latih dan data uji, dengan perbandingan data latih 80% dan data uji 20%. Selanjutnya, akan dilakukan pengujian performa model dengan dua pendekatan, yaitu menggunakan data tanpa *Oversampling* dan dengan data yang telah diolah menggunakan teknik *Random Oversampling*. Perbandingan data sebelum dan sesudah dilakukan *Oversampling* dapat dilihat pada Gambar 2.



Gambar 2. Data Sentimen Sebelum dan Sesudah *Oversampling*

3.6 Evaluasi Model Pertama (*Non Oversampling*)

Skenario evaluasi model pertama dilakukan dengan menguji dataset sentimen berdasarkan 5 variabel utama menggunakan algoritma SVM tanpa *Oversampling*. Pengujian ini dilakukan untuk mengetahui hasil performa model dalam menangani data yang tidak seimbang.

3.6.1 Kecepatan Internet

Pengujian dataset menggunakan SVM tanpa *Oversampling* dilakukan pada variabel “Kecepatan Internet” untuk mengevaluasi performa model dalam kondisi data yang tidak seimbang.

Tabel 4. Evaluasi Model SVM *Non Oversampling* pada Variabel Kecepatan Internet

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	87%	88%	87%	87%
Biznet	78%	76%	78%	77%
Starlink	83%	77%	83%	80%

Tabel 4 menunjukkan hasil evaluasi model SVM tanpa *oversampling* pada variabel “Kecepatan Internet”. IndiHome memperoleh hasil terbaik dengan *accuracy* 87%, *presicion* 88%, *recall* 87%, dan *F1-score* 87%. Starlink menyusul dengan *accuracy* 83%, *presicion* 77%, *recall* 83%, dan *F1-score* 80%. Sementara itu, Biznet mencatat nilai terendah dengan *accuracy* 78%, *presicion* 76%, *recall* 78%, dan *F1-score* 77%.

3.6.2 Kestabilan Jaringan

Pengujian dataset menggunakan SVM tanpa *Oversampling* dilakukan pada variabel “Kestabilan Jaringan” untuk mengevaluasi performa model dalam kondisi data yang tidak seimbang.

Tabel 5. Evaluasi Model SVM *Non Oversampling* pada Variabel Kestabilan Jaringan

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	78%	76%	78%	76%
Biznet	63%	61%	63%	62%
Starlink	57%	57%	57%	55%

Tabel 5 memperlihatkan hasil evaluasi model SVM *Non Oversampling* pada variabel “Kestabilan Jaringan”. Provider IndiHome menunjukkan performa terbaik dengan *accuracy* 78%, *presicion* 76%, *recall* 78%, dan *F1-score* 76%. Biznet menyusul dengan *accuracy* 63%, *presicion* 61%, *recall* 63%, dan *F1-score* 62%. Starlink memiliki performa terendah dengan seluruh metrik bernilai 57% untuk *accuracy*, *presicion*, *recall*, dan *F1-score* sebesar 55%.

3.6.3 Harga dan Paket Layanan

Pengujian dataset menggunakan SVM tanpa *Oversampling* dilakukan pada variabel “Harga dan Paket Layanan” untuk mengevaluasi performa model dalam kondisi data yang tidak seimbang.

Tabel 6. Evaluasi Model SVM *Non Oversampling* pada Variabel Harga dan Paket Layanan

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	85%	84%	85%	84%
Biznet	67%	70%	67%	68%
Starlink	73%	71%	73%	72%

Tabel 6 menunjukkan hasil evaluasi model SVM *Non Oversampling* pada variabel “Harga dan Paket layanan”. IndiHome mencatat hasil tertinggi dengan *accuracy* 85%, *precision* 84%, *recall* 85%, dan *F1-score* 84%. Starlink menyusul dengan *accuracy* 73%, *precision* 71%, *recall* 73%, dan *F1-score* 72%. Sementara itu, Biznet memperoleh hasil paling rendah, yaitu *accuracy* 67%, *precision* 70%, *recall* 67%, dan *F1-score* 68%.

3.6.4 Kualitas Layanan Pelanggan

Pengujian dataset menggunakan SVM tanpa *Oversampling* dilakukan pada variabel “Kualitas Layanan Pelanggan” untuk mengevaluasi performa model dalam kondisi data yang tidak seimbang.

Tabel 7. Evaluasi Model SVM *Non Oversampling* pada Variabel Kualitas Layanan Pelanggan

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	87%	87%	87%	84%
Biznet	77%	79%	77%	75%
Starlink	70%	66%	70%	67%

Tabel 7 menunjukkan hasil evaluasi model SVM *Non Oversampling* terhadap variabel “Kualitas Layanan Pelanggan”. Dari hasil evaluasi, IndiHome mendapatkan performa terbaik dengan nilai *accuracy*, *precision*, dan *recall* sebesar 87%, serta *F1-score* sebesar 84%. Biznet menempati posisi kedua dengan *accuracy* 77%, *precision* 79%, *recall* 77%, dan *F1-score* 75%. Sementara itu, Starlink memperoleh nilai evaluasi terendah, dengan *accuracy* dan *recall* sebesar 70%, *precision* sebesar 66%, serta *F1-score* 66%.

3.6.5 Ketersediaan dan Jangkauan Wilayah

Pengujian dataset menggunakan SVM tanpa *Oversampling* dilakukan pada variabel “Ketersediaan dan Jangkauan Wilayah” untuk mengevaluasi performa model dalam kondisi data yang tidak seimbang.

Tabel 8. Evaluasi Model SVM *Non Oversampling* pada Variabel Ketersediaan dan Jangkauan Wilayah

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	82%	81%	82%	81%
Biznet	80%	80%	80%	77%
Starlink	72%	66%	72%	61%

Tabel 8 memperlihatkan hasil evaluasi model SVM *Non Oversampling* terhadap variabel “Ketersediaan dan Jangkauan Wilayah”. IndiHome kembali menunjukkan performa terbaik dengan *accuracy* sebesar 82%, *precision* 81%, *recall* 82%, dan *F1-score* 81%. Biznet menyusul dengan hasil yang cukup kompetitif, yakni *accuracy*, *precision*, dan *recall* sebesar 80%, serta *F1-score* 77%. Di sisi lain, Starlink mencatat performa terendah, dengan *accuracy* dan *recall* sebesar 72%, *precision* 66%, serta *F1-score* 61%.

3.7 Evaluasi Model Kedua (Oversampling)

Skenario evaluasi model kedua dilakukan dengan menguji dataset sentimen berdasarkan 5 variabel utama menggunakan algoritma SVM dengan metode *Random Oversampling*. Pengujian ini dilakukan untuk mengatasi ketidakseimbangan data antar kelas dengan menyeimbangkan jumlah data positif, negatif, dan netral.

3.7.1 Kecepatan Internet

Pengujian dataset menggunakan SVM dengan metode *Random Oversampling* dilakukan pada variabel “Kecepatan Internet” untuk mengevaluasi performa model ketika data sudah diseimbangkan.

Tabel 9. Evaluasi Model SVM dengan *Oversampling* pada Variabel Kecepatan Internet

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	90%	89%	90%	90%
Biznet	79%	74%	79%	76%
Starlink	82%	79%	82%	80%

Tabel 9 menampilkan hasil evaluasi model SVM dengan metode *Random Oversampling* terhadap variabel “Kecepatan Internet”. IndiHome mencatatkan kinerja terbaik dengan *accuracy*, *precision*, *recall*, dan *F1-score* masing-masing sebesar 90%, 89%, 90%, dan 90%. Starlink berada di posisi kedua dengan *accuracy* 82%, *precision* 79%, *recall* 82%, dan *F1-score* 80%. Sementara itu, Biznet memperoleh *accuracy* 79%, *precision* 74%, *recall* 79%, dan *F1-score* 76%.

3.7.2 Kestabilan Jaringan

Pengujian dataset menggunakan SVM dengan metode *Random Oversampling* dilakukan pada variabel “Kestabilan Jaringan” untuk mengevaluasi performa model ketika data sudah diseimbangkan.

Tabel 10. Evaluasi Model SVM dengan *Oversampling* pada Variabel Kestabilan Jaringan

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	87%	86%	87%	86%
Biznet	69%	65%	69%	66%
Starlink	62%	57%	62%	58%

Tabel 10 menyajikan hasil evaluasi model SVM dengan metode *Random Oversampling* terhadap variabel “Kestabilan Jaringan”. IndiHome kembali menunjukkan performa tertinggi dengan *accuracy* 87%, *precision* 86%, *recall* 87%, dan *F1-score* 86%. Biznet berada di posisi kedua dengan *accuracy* dan *recall* sebesar 69%, *precision* 65%, serta *F1-score* 66%. Starlink mencatat performa terendah dengan *accuracy* 62%, *precision* 57%, *recall* 62%, dan *F1-score* 58%.

3.7.3 Harga dan Paket Layanan

Pengujian dataset menggunakan SVM dengan metode *Random Oversampling* dilakukan pada variabel “Harga dan Paket Layanan” untuk mengevaluasi performa model ketika data sudah diseimbangkan.

Tabel 11. Evaluasi Model SVM dengan *Oversampling* pada Variabel Harga dan Paket Layanan

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	79%	77%	79%	78%
Biznet	74%	71%	74%	71%
Starlink	71%	67%	71%	68%

Tabel 11 menampilkan hasil evaluasi model SVM dengan *Random Oversampling* pada variabel “Harga dan Paket Layanan”. IndiHome memperoleh performa terbaik dengan *accuracy* 79%, *precision* 77%, *recall* 79%, dan *F1-score* 78%. Biznet berada di posisi kedua dengan *accuracy* 74%, *precision* 71%, *recall* 74%, serta *F1-score* 71%. Starlink mencatatkan hasil terendah dengan *accuracy* 71%, *precision* 67%, *recall* 71%, dan *F1-score* 68%.

3.7.4 Kualitas Layanan Pelanggan

Pengujian dataset menggunakan SVM dengan metode *Random Oversampling* dilakukan pada variabel “Kualitas Layanan Pelanggan” untuk mengevaluasi performa model ketika data sudah diseimbangkan.

Tabel 12. Evaluasi Model SVM dengan *Oversampling* pada Variabel Kualitas Layanan Pelanggan

Provider	Accuracy	Precision	Recall	F1-Score
IndiHome	90%	91%	90%	90%
Biznet	79%	79%	79%	79%
Starlink	71%	72%	71%	71%

Tabel 12 menampilkan hasil evaluasi model SVM dengan *Random Oversampling* terhadap variabel “Kualitas Layanan Pelanggan”. IndiHome menunjukkan performa tertinggi dengan *accuracy* 90%, *precision* 91%, *recall* 90%, dan *F1-score* 90%. Biznet menempati posisi kedua dengan seluruh metrik di angka 79%. Sementara itu, Starlink mencatatkan hasil terendah dengan *accuracy* 71%, *precision* 72%, *recall* 71%, dan *F1-score* 71%.

Gambar 5 merepresentasikan sentimen netral, yang terlihat dari kata-kata seperti "layanan", "akun", "paket", "akses", "pengguna", "internet", "waktu", dan "email". Kata-kata ini menunjukkan interaksi layanan pelanggan dalam menangani laporan atau kendala pelanggan secara informatif dan responsif, tanpa ekspresi emosi yang kuat.

4. KESIMPULAN

Penelitian ini mampu mengaplikasikan dengan baik algoritma Support Vector Machine (SVM) untuk menganalisis sentimen pengguna media sosial X terhadap tiga penyedia layanan internet, yaitu IndiHome, Biznet, dan Starlink, berdasarkan lima variabel utama: kecepatan internet, kestabilan jaringan, harga dan paket layanan, kualitas layanan pelanggan, serta ketersediaan dan jangkauan wilayah. Dari hasil evaluasi model SVM baik tanpa oversampling maupun dengan oversampling, dapat disimpulkan bahwa IndiHome konsisten menunjukkan performa terbaik pada seluruh variabel, dengan nilai akurasi mencapai hingga 90%, precision 91%, recall 90%, dan F1-score 90%, terutama pada variabel kualitas layanan pelanggan. Biznet umumnya menempati posisi kedua, dengan akurasi bervariasi antara 63% hingga 80%, tergantung pada variabel yang diuji. Starlink seringkali memiliki performa terendah, dengan akurasi pada beberapa variabel berada di bawah 70%, meskipun pada variabel kecepatan internet dan jangkauan wilayah masih kompetitif dengan akurasi mencapai 82%. Penggunaan metode Random Oversampling terbukti meningkatkan performa model secara signifikan, dengan rata-rata peningkatan akurasi antara 6% hingga 12% dibandingkan model tanpa penanganan ketidakseimbangan data. Secara umum, pendekatan text mining dan machine learning berbasis algoritma SVM dalam penelitian ini memberikan hasil klasifikasi yang akurat dan seimbang, serta dapat dijadikan acuan penting bagi penyedia layanan internet untuk memahami persepsi publik, sekaligus mengevaluasi dan meningkatkan kualitas layanan berdasarkan opini nyata dari pengguna. Penelitian selanjutnya disarankan untuk mengeksplorasi penggunaan algoritma klasifikasi lain guna dibandingkan dengan hasil dari SVM. Selain itu, dapat pula dilakukan analisis sentimen berbasis aspek lain atau penambahan sumber data dari platform yang berbeda agar hasil yang diperoleh lebih kaya dan representatif.

REFERENCES

- [1] M. F. Fachrudin, C. V. Angkoso, and D. A. Fatah, "Analisis Sentimen Pada Sosial Media Twitter Terhadap Kualitas Jaringan Internet Telkomsel Menggunakan Ensemble K-Nearest Neighbour -Support Vector Machine," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1253–1264, Dec. 2024, doi: 10.25126/jtiik.2024118713.
- [2] M. H. A. Sunata, F. Irwiensyah, and F. N. Hasan, "Analisis Sentimen Calon Presiden 2024 di Media Sosial X Menggunakan Naive Bayes dan SMOTE," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 3, p. 1313, Jul. 2024, doi: 10.30865/mib.v8i3.7708.
- [3] Diana Puspitasari and Tata Sutabri, "Analisis Sentimen Berdasarkan pada Twitter (X) terhadap Layanan Indihome Menggunakan Algoritma Support Vector Machine (SVM)," *JUMINTAL: Jurnal Manajemen Informatika dan Bisnis Digital*, vol. 3, no. 2, pp. 58–71, Nov. 2024, doi: 10.55123/jumintal.v3i2.4449.
- [4] H. J. Christanto, "Game Theory Analysis of Indihome and Biznet in the Salatiga Internet Market," *Journal of Information Systems and Informatics*, vol. 6, no. 1, pp. 399–408, Mar. 2024, doi: 10.51519/journalisi.v6i1.677.
- [5] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom dan Biznet," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, Sep. 2019, doi: 10.33480/pilar.v15i2.699.
- [6] J. Garcia, M. Beckerle, S. Sundberg, and A. Brunstrom, "Modeling and predicting starlink throughput with fine-grained burst characterization," *Comput Commun*, vol. 234, Mar. 2025, doi: 10.1016/j.comcom.2025.108090.
- [7] M. Ridwan, T. Al Islami, Daffa, M., P. Rahayu, F. Az Zahra, R., and A. Naerul, Edwin, K., "Oligopoli Telekomunikasi dan Inovasi : Analisis Dampak Masuknya STARLINK bagi Industri Telekomunikasi di Indonesia," *Kampus Akademik Publisng Jurnal Ilmiah Ekonomi Dan Manajemen*, vol. 2, no. 12, pp. 306–312, 2024, doi: 10.61722/jiem.v2i12.3206.
- [8] J. Pebrianto, "Sentiment Analysis Of Service Provider On Twitter Tweet Using Naive Bayes Classifier With PHP," *Journal of Innovation And Future Technology (IFTECH)*, vol. 5, no. 2, pp. 13–23, 2023, doi: 10.47080/iftech.v5i2.2752.
- [9] A. Nugroho and N. Tedi Kurniadi, "Sentiment Analysis of Starlink on Twitter Using Support Vector Machine Algorithm," *Architecture and High Performance Computing*, vol. 6, no. 3, 2024, doi: 10.47709/cnape.v6i3.4348.
- [10] M. Khalil Gibran *et al.*, "Sentiment Analysis of Platform X Users on Starlink Using Naive Bayes," *Instal: Jurnal Komputer*, no. 03, pp. 210–220, 2024, doi: 10.54209/jurnalinstall.v16i03.240.
- [11] H. Jia and S. Shen, "Benders Cut Classification via Support Vector Machines for Solving Two-Stage Stochastic Programs," *INFORMS Journal on Optimization*, vol. 3, no. 3, pp. 278–297, 2021, doi: 10.1287/ijoo.2019.0050.
- [12] M. Nur Akbar and N. Annisa Safitri Yusuf, "Analisis Sentimen Pengguna Indihome dengan Metode Klasifikasi Support Vector Machine (SVM)," *Journal Shift*, vol. 2, no. 1, pp. 14–21, 2022, doi: 10.24252/shift.v2i1.18.
- [13] A. R. Muhammad Fikri, J. Jondri, and W. Astuti, "Sentiment Analysis Against IndiHome and First Media Internet Providers Using Ensemble Stacking Method," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, Sep. 2022, doi: 10.47065/bits.v4i2.1969.
- [14] J. Lu, "Data Analytics Research-Informed Teaching in a Digital Technologies Curriculum," vol. 20, no. 2, pp. 57–123, 2020, doi: 10.1287/ited.2019.0215.
- [15] A. Addiga and S. Bagui, "Sentiment Analysis on Twitter Data Using Term Frequency-Inverse Document Frequency," *Journal of Computer and Communications*, vol. 10, no. 08, pp. 117–128, 2022, doi: 10.4236/jcc.2022.108008.
- [16] F. Hashfi, D. Sugiarto, and I. Mardianto, "Sentiment Analysis of An Internet Provider Company Based on Twitter Using Support Vector Machine and Naïve Bayes Method," *Ultimatics : Jurnal Teknik Informatika*, vol. 14, no. 1, pp. 1–6, 2022, doi: 10.31937/ti.v14i1.2384.



- [17] R. Muliani, A. Solehudin, and A. Jamaludin, “Analisis Sentimen Terhadap Provider XYZ di Twitter Menggunakan Algoritma Naive Bayes Classifier,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 4, pp. 2841–2848, Aug. 2024, doi: 10.36040/jati.v7i4.7191.
- [18] N. Chamidah, D. Widiyanto, H. B. Seta, and A. A. Aziz, “The Impact of Oversampling and Undersampling on Aspect-Based Sentiment Analysis of Indramayu Tourism Using Logistic Regression,” *Revue d’Intelligence Artificielle*, vol. 38, no. 3, pp. 795–804, Jun. 2024, doi: 10.18280/ria.380306.
- [19] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, “Comparison of Random Forest, Naïve Bayes, and Support Vector Machine Algorithms in Analyzing Twitter Sentiment Regarding Public Opinion on the Removal of Honorary Employees,” *Seminar Nasional Official Statistics*, vol. 2022, no. 1, pp. 563–572, 2022, doi: <https://doi.org/10.34123/semnasoffstat.v2022i1.1410>.
- [20] P. M. Susanti, M. Afdal, I. Permana, and A. Marsal, “Klasifikasi Sentimen Pengguna X Terhadap Pemboikotan Produk Pro Israel Menggunakan Algoritma Machine Learning,” *Technology and Science (BITS)*, vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6533.