

Analisis Sentimen Pengguna Indosat Pada *Platform X* Menggunakan *Naïve Bayes* dan *Support Vector Machine*

Puput Tri Santika¹, Islamiyah², Amin Padmo Azam Masa*³, Riftika Rizawanti⁴

^{1,2,3}Program Studi Sistem Informasi, Fakultas Teknik, Universitas Mulawarman
e-mail: ¹puputtri.sn@gmail.com, ²islamiyahunmul@gmail.com, *³aminpadmo@ft.unmul.ac.id,
⁴riftikariza@ft.unmul.ac.id

Abstrak

X merupakan platform media sosial berbasis mikroblog yang memungkinkan penggunaanya berinteraksi secara pribadi maupun publik, sehingga menjadikannya saluran utama bagi masyarakat dalam mengekspresikan opini terhadap berbagai layanan, termasuk layanan telekomunikasi. Indosat Ooredoo sebagai salah satu perusahaan penyedia layanan telekomunikasi terbesar di Indonesia memanfaatkan platform *X* untuk menyampaikan informasi layanan sekaligus menerima keluhan dan saran dari pelanggan. Namun, volume data tweet yang besar dan tidak terstruktur memerlukan pendekatan komputasional yang mampu mengklasifikasikan sentimen secara otomatis dan efisien. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan efektivitas algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* dalam menganalisis sentimen pengguna Indosat pada platform *X*. Data yang digunakan berjumlah 1.000 tweet. Tahapan penelitian meliputi preprocessing teks, pelabelan data ke dalam kelas positif dan negatif, pembobotan fitur menggunakan *TF-IDF*, serta penanganan ketidakseimbangan kelas menggunakan teknik *oversampling SMOTE*. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa *Naïve Bayes* menghasilkan akurasi 78%, presisi 78%, dan recall 77%, sedangkan *SVM* menghasilkan akurasi 72%, presisi 73%, dan recall 72%. *Naïve Bayes* terbukti memiliki kinerja lebih baik dibandingkan *SVM* dalam mengklasifikasikan sentimen pengguna Indosat. Hasil analisis secara keseluruhan menunjukkan dominasi sentimen negatif pada ulasan pengguna. Penelitian ini diharapkan dapat membantu Indosat memahami persepsi pelanggan berbasis data serta menjadi referensi bagi pengembangan sistem analisis sentimen pada domain layanan telekomunikasi.

Kata kunci— Analisis Sentimen, Indosat, *Naïve Bayes*, *Support Vector Machine*, Platform *X*

1. PENDAHULUAN

X adalah sebuah layanan sosial media dan juga *microblog* yang memungkinkan penggunaanya untuk berinteraksi secara pribadi atau publik [1]. *X* adalah salah satu platform media sosial yang paling terkenal dan diminati di seluruh dunia, termasuk di Indonesia. Perusahaan-perusahaan besar banyak yang memanfaatkan *X* untuk keperluan bisnis. Perusahaan telekomunikasi di Indonesia yaitu PT Indosat yang juga dikenal sebagai Indosat merupakan salah satu yang memanfaatkan media sosial *X* untuk keperluan bisnisnya. Indosat Ooredoo adalah salah satu perusahaan penyedia layanan telekomunikasi dan jaringan yang ada di Indonesia [2]. Indosat memanfaatkan sosial media *X* sebagai tempat untuk memberikan informasi kepada pelanggannya mengenai promosi dan layanan terbarunya. Indosat juga memanfaatkan *X* sebagai tempat untuk menerima keluhan, kritik, saran, serta untuk menjaga komunikasi antara Indosat dengan pelanggan untuk menjaga kualitas layanan dari Indosat. Namun, ternyata terdapat beberapa tweet keluhan, kritik, serta saran dari pengguna melalui akun *X* Indosat. Oleh karena itu, analisis

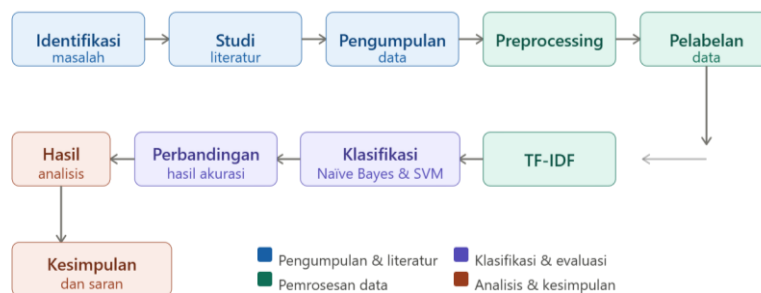
sentimen diperlukan untuk mengetahui permasalahan yang dialami oleh pelanggan dan mengetahui sentimen pengguna, agar masalah serupa tidak terulang kembali di kemudian hari. Penelitian tentang analisis sentimen opini melalui media sosial khususnya X (Twitter dulu dikenal) pernah dilakukan menggunakan *Support Vector Machine* dan *Naïve Bayes* yang terbukti memberikan nilai akurasi atau performa yang baik dalam melihat sentimen terhadap vaksin Covid-19 pada. Hasil akurasi dari algoritma *Support Vector Machine* adalah 90,47% dan algoritma *Naïve Bayes* adalah 88,64% dengan selisih akurasi 1,83% [3]. Beberapa penelitian lainnya membandingkan kedua metode ini dengan hasil yang bervariasi tergantung domain dan dataset yang digunakan. Hasil penelitian menunjukkan bahwa model klasifikasi SVM memiliki akurasi sebesar 70,08% dengan nilai presisi 82,55%, recall 61,27%, dan F1-score 64,76%, sedangkan model *Naïve Bayes* memiliki akurasi sebesar 62,90% dengan nilai presisi 81,03%, recall 48,32%, dan F1-score 46,83% [4].

Berdasarkan hasil penelitian tersebut, dapat disimpulkan bahwa model SVM memiliki performa yang lebih baik dalam melakukan analisis sentimen di media sosial dibandingkan *Naïve Bayes*. Namun demikian, dalam penelitian lain, akurasi *Naïve Bayes* lebih unggul di angka 91,48% dibandingkan SVM, sehingga kedua metode menunjukkan potensi yang signifikan dalam analisis sentimen [5]. *Analisis Sentimen* terhadap brand indosat pernah dilakukan menggunakan metode *Naïve Bayes Classifier* dengan hasil akurasi sebesar 83,91% yang melibatkan 1.000 data tweet [6]. Selanjutnya, SVM dan *Naïve Bayes* diterapkan pada bidang kesehatan untuk aplikasi layanan kesehatan pemerintah dengan jumlah 1046 data [7] dan penelitian lainnya [8]–[10]. Dari penelitian terdahulu terlihat bahwa belum ada perbandingan *Naïve Bayes* dan SVM khusus untuk Indosat pada platform X dengan validasi metode yang lebih akurat untuk domain sentimen pelanggan telekomunikasi berbahasa Indonesia. Oleh karena itu, penelitian ini bertujuan untuk menerapkan metode *Naïve Bayes* dan *Support Vector Machine* dalam analisis sentimen pengguna Indosat di media sosial X dan mengevaluasi nilai tingkat akurasi dari kedua metode tersebut dalam mengidentifikasi sentimen positif dan negatif.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Tahapan penelitian terlihat pada Gambar 1, dimana tahapan ini dimulai dari identifikasi masalah hingga penarikan kesimpulan. Setiap tahapan saling mendukung untuk menghasilkan penelitian yang valid, reproducible, dan dapat dipertanggungjawabkan secara ilmiah.



Gambar 1. Diagram Alur Penelitian

2.2. Analisis Sentimen

Analisis Sentimen adalah studi yang melibatkan bahasa pemrograman pikiran, bahasa komputer, dan analisis teks untuk mengidentifikasi opini dan emosi yang tersembunyi dalam sebuah teks. Tujuannya adalah untuk mengungkapkan sentimen yang terdapat dalam teks

tersebut, seperti kelas positif, negatif, atau netral terhadap suatu produk atau opini. Teknik ini memungkinkan pengungkapan emosi dan tindakan yang terkandung dalam teks secara otomatis melalui media sosial atau platform lainnya [11].

2.3. *Text Mining*

Text Mining merupakan teknik pengolahan data dengan cara menganalisis teks dari yang tidak terstruktur menjadi terstruktur. Salah satu tahapan penting dalam *text mining* adalah proses pengambilan data terlebih dahulu, yang kemudian harus melalui tahap preprocessing sebelum dapat dilakukan klasifikasi atau analisis lebih lanjut. Proses ekstraksi informasi dalam *text mining* memungkinkan untuk melakukan analisis sentimen secara emosional guna mengidentifikasi apakah pernyataan tersebut bersifat positif atau negatif [12].

2.4. *Term Frequency-Inverse Document Frequency (TF-IDF)*

Term Frequency-Inverse Document Frequency (TF-IDF) adalah algoritma pembobotan yang menetapkan nilai bobot kata berdasarkan seberapa sering kata tersebut muncul dalam dokumen. Tujuannya adalah untuk memberi nilai bobot pada kata-kata sesuai dengan seberapa sering kata-kata tersebut muncul dalam dokumen. Proses klasifikasi dokumen dapat memanfaatkan fitur-fitur kata yang telah diberi nilai bobot [13].

2.5. *Oversampling SMOTE*

Salah satu permasalahan yang sering dihadapi dalam klasifikasi sentimen analisis adalah ketidakseimbangan kelas pada dataset. Data yang tidak seimbang dapat menyebabkan hasil negatif pada hasil klasifikasi, karena jumlah data yang tidak seimbang antara kelas mayoritas dan kelas minoritas cenderung membuat kelas mayor tampil lebih baik daripada kelas minor. *Synthetic Minority Oversampling Technique* (SMOTE) merupakan teknik dalam mengatasi ketidakseimbangan data dengan cara meningkatkan jumlah data sampel pada kelas minoritas. Data yang ditambahkan adalah data sintetik yang dihasilkan dari replikasi data sampel kelas minoritas [14], [15].

2.6. *Naïve Bayes Classifier*

Thomas Bayes yaitu seorang fisikawan Inggris yang menemukan algoritma *Naïve Bayes* yang merupakan pendekatan klasifikasi yang menggunakan teknik statistik dan probabilitas [16]. Metode *Naïve Bayes Classifier* (NBC) adalah teknik klasifikasi dalam *text mining* yang dapat digunakan dalam analisis sentimen. Cara ini dikenal efektif dari segi akurasi dan kemampuan prediksi data. *Naïve Bayes* adalah sebuah metode statistik yang mampu memprediksi nilai probabilitas bahwa sampel data yang diberikan akan diklasifikasikan ke dalam *class* tertentu [16]. Metode ini digunakan dalam penelitian untuk mengklasifikasikan data dari *X* dan membaginya ke dalam *class* positif dan negatif.

2.7. *Support Vector Machine*

Support Vector Machine (SVM) adalah salah satu metode pembelajaran terarah yang dapat menganalisis data dan menemukan pola yang diperlukan untuk analisis klasifikasi. SVM menggunakan fungsi hyperplane pada data untuk menentukan batas-batas antar kelas. Perhitungan hyperplane bertujuan untuk menemukan garis pembatas yang optimal sehingga data dapat dibagi dengan tepat menjadi dua kelas [17].

2.8. *Confusion Matrix*

Confusion Matrix merupakan sebuah alat (*tools*) visualisasi yang umum digunakan dalam supervised learning. Tiap kolom dari matriks ini mewakili contoh-contoh dari kelas yang diprediksi, sementara tiap baris mencerminkan kejadian pada kelas sebenarnya [18]. *Confusion matrix* digunakan untuk mengukur tingkat akurasi dari sistem klasifikasi. Hal ini melibatkan

perhitungan jumlah data yang diklasifikasikan dengan benar dan yang salah berdasarkan algoritma yang digunakan. *Confusion matrix* merupakan proses evaluasi berbentuk matriks yang memberikan gambaran tentang seberapa tepat klasifikasi terhadap kelas-kelas yang diuji [19]

Tabel 1. Struktur *Confusion Matrix*

	<i>Actual Positive</i>	<i>Actual Negatif</i>
<i>Predicted Positive</i>	TP	FP
<i>Predicted Negative</i>	FN	TN

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan dan Pembagian Data

Pengumpulan data dari *X* dilakukan dengan cara mengumpulkan data *tweet* berdasarkan pencarian menggunakan kata kunci ‘Indosat’, data *tweet* diambil dari tanggal 1 September 2023 sampai 30 Desember 2023 data yang diambil adalah berbahasa indonesia berjumlah 1000 *tweet*. Data ini diambil langsung dari database *Twint* berdasarkan kata kunci pencarian yang ditentukan, dan hasilnya disimpan dalam format *Comma Separated Values (CSV)*. Proses pengumpulan data dari media sosial *X* menggunakan teknik *crawling* dengan bantuan *library snsrape*. *Snsrape* adalah sebuah pustaka yang akan digunakan untuk mengambil data dari media sosial *X*, termasuk *Twitter*. Proses *crawling* data dilakukan dengan menggunakan bahasa pemrograman *Python* di *Google Colaboratory notebook*. Gambar 2 menampilkan hasil *crawling* data.

	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation_id_str
0	Fri Dec 29 23:59:32 +0000 2023	1740885377831821717	@anmusaa Mohon maaf kekan kak. terkait post.	0	0	0	0	in	98255721575608320	1740701729241604158
1	Fri Dec 29 23:59:07 +0000 2023	1740885188335428753	@tanyakani @indosat sisa sikitlah nih min mur...	0	1	0	0	in	539516662	1740876872265945422
2	Fri Dec 29 23:59:07 +0000 2023	1740885186790015419	@convomf Bgh 20k nih niler indosat	0	0	0	0	in	1464286685244628994	1740701729241604158
3	Fri Dec 29 23:57:02 +0000 2023	1740884861512048771	@tanyakani @indosatm3 tolong dong indosat g.	0	5	4	253	in	1655537315332736000	1740876872265945422
4	Fri Dec 29 23:56:56 +0000 2023	1740884637948326709	@tanyakani @indosat ihut ni wahai	0	1	0	1	in	1411960807642963969	1740876872265945422

Gambar 2. Hasil *Crawling* Data

Gambar 2 merupakan hasil dari proses setelah pengambilan data yang melakukan *remove duplicates* untuk menghapus data berganda atau baris yang berulang dengan tujuan untuk mendapatkan list satu data saja sehingga dari 1002 data menjadi 977 data.

3.2 Preprocessing

Preprocessing adalah proses mengubah data yang tidak terstruktur menjadi data terstruktur. Berikut adalah langkah-langkah dari teks *preprocessing*:

1) *Cleaning*

Proses *cleaning* digunakan untuk menghilangkan kata-kata yang tidak diperlukan seperti @ (*at*) untuk tautan, *URL* atau situs web seperti *http*, *url*, *www*, # (*hashtag*), *RT* (*retweet*), dan emoji. Hasil *cleaning* dapat dilihat pada Tabel 2.

Tabel 2. *Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
@convomf Plis aku dah pake indosat sejak tahun 2012 riil nambah mahal skrng dY© mau ganti provider udh nyaman jg nomornya pake ini	Plis aku dah pake indosat sejak tahun riil nambah mahal skrng mau ganti provider udh nyaman jg nomornya pake ini

2) *Case Folding*

Case Folding adalah proses pengubahan huruf besar menjadi huruf kecil dengan tujuan agar tidak ada perbedaan makna antara huruf besar dan huruf kecil. Hasil *case folding* dapat dilihat pada Tabel 3.

Tabel 3. *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
Plis aku dah pake indosat sejak tahun riil nambah mahal skrng mau ganti provider udh nyaman jg nomornya pake ini	plis aku dah pake indosat sejak tahun riil nambah mahal skrng mau ganti provider udh nyaman jg nomornya pake ini

3) *Tokenizing*

Tokenizing yaitu proses memisahkan teks ulasan menjadi kata-kata individual yang tidak saling terkait. *Tokenizing* berguna untuk menghitung frekuensi kemunculan kata dalam suatu dokumen. Hasil dari *tokenizing* dapat dilihat pada Tabel 4.

Tabel 4. *Tokenizing*

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
plis aku dah pake indosat sejak tahun riil nambah mahal skrng mau ganti provider udh nyaman jg nomornya pake ini	['plis', 'aku', 'dah', 'pake', 'indosat', 'sejak', 'tahun', 'riil', 'nambah', 'mahal', 'skrng', 'mau', 'ganti', 'provider', 'udh', 'nyaman', 'jg', 'nomornya', 'pake', 'ini']

4) *Filtering*

Filtering atau *Stopword removal* adalah proses menghapus kata yang tidak penting dan kata yang tidak berarti. Pada langkah ini, digunakan library NLTK untuk menghapus kata-kata yang terdaftar sebagai *stopword* dalam NLTK. Hasil dari *filtering* dapat dilihat pada Tabel 5.

Tabel 5. *Filtering*

Sebelum <i>Filtering</i>	Sesudah <i>Filtering</i>
['plis', 'aku', 'dah', 'pake', 'indosat', 'sejak', 'tahun', 'riil', 'nambah', 'mahal', 'skrng', 'mau', 'ganti', 'provider', 'udh', 'nyaman', 'jg', 'nomornya', 'pake', 'ini']	['plis', 'dah', 'pake', 'indosat', 'riil', 'nambah', 'mahal', 'skrng', 'ganti', 'provider', 'udh', 'nyaman', 'jg', 'nomornya', 'pake']

5) *Stemming*

Stemming yaitu proses untuk mengubah kata menjadi kata dasar. Pada langkah ini, menggunakan bantuan pustaka Sastrawi. Hasil dari *stemming* dapat dilihat pada Tabel 6.

Tabel 6. *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
['plis', 'dah', 'pake', 'indosat', 'riil', 'nambah', 'mahal', 'skrng', 'ganti', 'provider', 'udh', 'nyaman', 'jg', 'nomornya', 'pake']	plis dah pake indosat riil nambah mahal skrng ganti provider udh nyaman jg nomor pake

6) *Normalisasi*

Normalisasi adalah proses mengubah kata-kata yang tidak sesuai dengan ejaan atau kata tidak baku berdasarkan basis data kamus. Tahap *normalisasi* dilakukan untuk standarisasi kata yang memiliki makna sama dengan melakukan perubahan penulisan suatu kata yang disingkat dan tidak baku agar memiliki arti yang seragam. Hasil *normalisasi* dapat dilihat pada Tabel 7.

Tabel 7. *Normalisasi*

Sebelum <i>Normalisasi</i>	Sesudah <i>Normalisasi</i>
plis dah pake indosat riil nambah mahal skrng ganti provider udh nyaman jg nomor pake	convert plis dah pakai indosat riil nambah mahal skrng tidak provider udh nyaman juga nomor pake

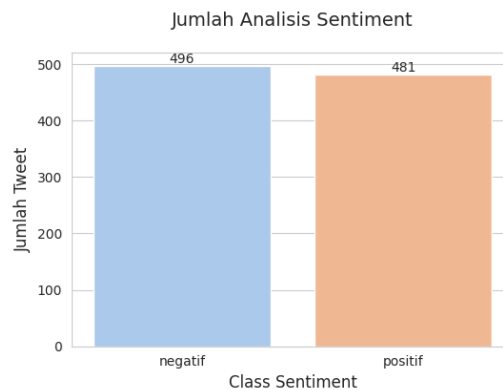
3.3 *Pelabelan*

Proses *labeling* yang dilakukan yaitu untuk menentukan kelas dari setiap ulasan atau komentar dalam pembagian data kedalam kelompok jenis sentimen. Dalam penelitian ini, dilakukan uji model dengan kumpulan data yang telah dibagi menjadi dua kelompok, yaitu kelompok sentimen positif dan sentimen negatif. Pada tahap ini pemberian label (*labeling*) dilakukan pada semua *tweet* yang akan digunakan sebagai data testing dan data *training* pada

implementasi algoritma NBC dan *Support Vector Machine*. Hasil pelabelan dapat dilihat pada Tabel 8. Visualisasi hasil pelabelan terlihat pada Gambar 3.

Tabel 8. Hasil *Labeling*

<i>Tweet</i>	<i>Label</i>
mohon maaf kak kait tawar promo nomor beda ya kak sifat random acak tawar promo cek aplikasi myim umb ya kak terima kasih gea	Positif
indosat murahh	Positif
iya najis banget indosat mahal banget babi	Negatif
indosat lihat mimin hormat mahal karna hamba uang paspasan	Negatif



Gambar 3. Visualisasi Jumlah Sentimen

3.4 Pembobotan Kata (TF-IDF)

Bobot TF-IDF digunakan untuk menilai pentingnya kata di dalam dokumen. *Term Frequency* (TF) yaitu semakin tinggi istilah yang muncul pada dokumen, semakin sering istilah tersebut muncul, dan semakin tinggi nilai bobot istilah tersebut. Tabel 9 merupakan sample hasil perhitungan TF-IDF.

Tabel 9. TF-IDF

<i>Term</i>	<i>tf</i>			<i>df</i>	<i>D/df</i>	<i>idf</i>	<i>w</i>		
	<i>d1</i>	<i>d2</i>	<i>d3</i>				<i>d1</i>	<i>d2</i>	<i>d3</i>
indosat	1	1	1	3	1.00	0.00	0.00	0.00	0.00
rumah	1	0	0	1	3.00	0.48	0.48	0.00	0.00
mati	2	0	0	1	3.00	0.48	0.95	0.00	0.00
lampu	1	0	0	1	3.00	0.48	0.48	0.00	0.00
ikut	1	0	0	1	3.00	0.48	0.48	0.00	0.00
indosat	0	1	1	2	1.50	0.18	0.00	0.18	0.18
pagi	0	1	0	1	3.00	0.48	0.00	0.48	0.00
sinyal	0	1	0	1	3.00	0.48	0.00	0.48	0.00
jelek	0	1	0	1	3.00	0.48	0.00	0.48	0.00
banget	0	1	1	2	1.50	0.18	0.00	0.18	0.18
ganggu	0	1	0	1	3.00	0.48	0.00	0.48	0.00
kerja	0	1	0	1	3.00	0.48	0.00	0.48	0.00
indosat	0	0	1	1	3.00	0.48	0.00	0.00	0.48
paketnya	0	0	1	1	3.00	0.48	0.00	0.00	0.48
mahal	0	0	2	1	3.00	0.48	0.00	0.00	0.95
banget	0	0	1	1	3.00	0.48	0.00	0.00	0.48

3.5 WordCloud

Word Cloud yaitu menampilkan data kata-kata (*term*) yang paling sering muncul. Semakin sering sebuah kata muncul dalam teks, semakin tinggi frekuensinya dalam analisis sentimen. Gambar 4 merupakan hasil visualisasi *Word Cloud*.

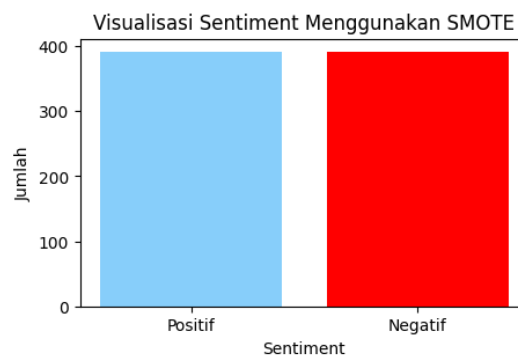


Gambar 4. WordCloud

Berdasarkan Gambar 4 terlihat bahwa hasil data penelitian yang digunakan, kata-kata yang sering muncul yaitu Indosat, Sinyal, Kuota, Pulsa, Paket.

3.6 Oversampling SMOTE

Penelitian ini menggunakan 977 data yang dibagi menjadi data *training* dan data *testing* dengan rasio 80% untuk data *latih* dan 20% untuk data uji, sehingga hasil perbandingan jumlah data tes sebanyak 196 dan data training sebanyak 781. Proses pelatihan dilakukan untuk memungkinkan model mempelajari pola dalam data teks sehingga saat menerima teks baru, sistem dapat memprediksi dengan tepat berdasarkan apa yang telah dipelajarinya sebelumnya. Data uji atau data *testing* digunakan untuk menguji kemampuan model dalam memberikan prediksi, sehingga didapatkan hasil akurasi model. Setelah berhasil melakukan preprocessing data, pelabelan dan pembagian data, kemudian data training dilakukan resampling dengan menggunakan metode SMOTE untuk menyeimbangkan data.

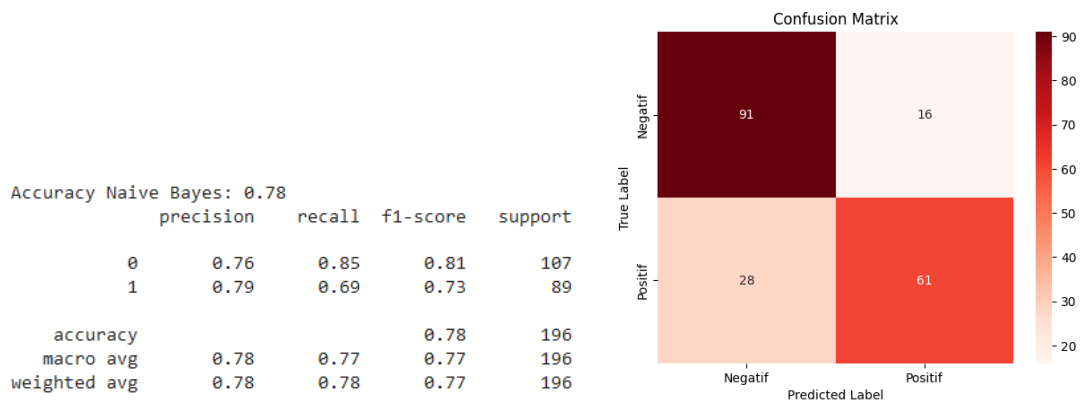


Gambar 5. Oversampling SMOTE

Gambar 5 merupakan *bar chart* yang menampilkan hasil distribusi data sentimen setelah diterapkan teknik oversampling SMOTE. Grafik ini menunjukkan perbandingan jumlah data antara dua kelas sentimen, yaitu kelas positif dan negatif mencapai sekitar 390 data.

3.7 Pemodelan

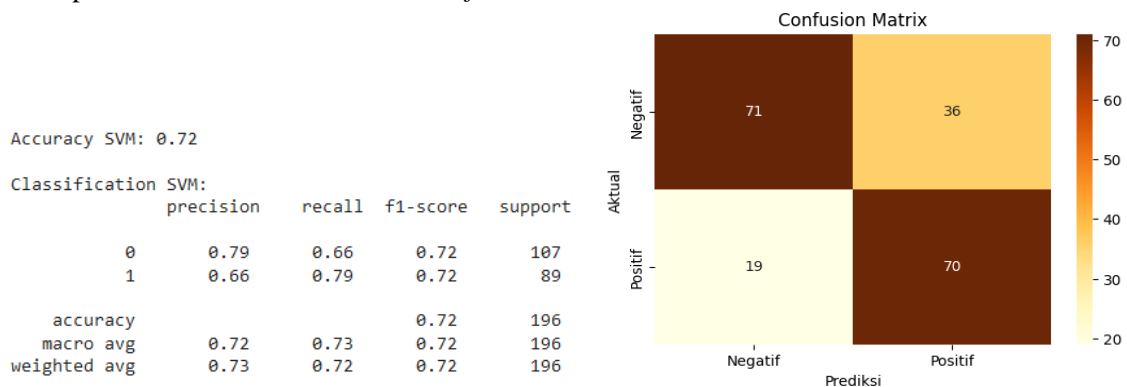
Model *Naïve Bayes Classifier* dibuat dengan menggunakan *MultinomialNB* dalam *library Scikit-learn*. Hasil pengujian dari dataset yang telah dilatih pada Gambar 6 menggunakan model *Naïve Bayes* menunjukkan bahwa akurasi yang dicapai sebesar 78%, presisi sebesar 78%, *recall* sebesar 78%, dan *f1-score* sebesar 77%.



Gambar 6. Confusion Matrix Naive Bayes

Dari Gambar 6 terlihat bahwa hasil evaluasi menggunakan *confusion matrix* didapatkan prediksi benar pada sentimen positif atau *true positif* sebanyak 91 dan prediksi benar pada sentimen negatif atau *true negatif* sebanyak 61. Setelah mendapatkan rasio dengan hasil rata-rata nilai dari akurasi, *precision*, *recall*, dan *f1-score* terbesar, kemudian dilakukan *cross validation* untuk mendapatkan akurasi yang maksimal. Selanjutnya, penelitian ini menerapkan *10-folds cross validation* dengan hasil akurasi, presisi, dan *recall* tertinggi pada n ke-10 sebesar 75,64%, 80,09, 75,64%, dan 74,71%.

Model SVM dibangun menggunakan *library Scikit-learn* dengan menggunakan modul *SVC (Support Vector Classifier)*. Pada metode SVM menggunakan kernel *linear* dengan nilai parameter C terbaik adalah 1. Gambar 7 menampilkan hasil pengujian dari SVM, dimana dataset yang telah dilatih menggunakan model SVM menunjukkan bahwa akurasi yang dicapai sebesar 72%, presisi 72%, dan *recall* 73%, dan *f1-score* 72%.



Gambar 9. Confusion Matrix SVM

Pada Gambar 9 terlihat bahwa hasil evaluasi *confusion matrix* didapatkan prediksi benar pada sentimen positif atau *true positif* sebanyak 70 dan prediksi benar pada sentimen negatif atau *true negatif* sebanyak 71. kemudian dilakukan *cross validation* untuk mendapatkan akurasi yang maksimal. Pada penelitian ini digunakan *10-folds cross validation* dengan hasil akurasi, presisi, dan *recall* tertinggi pada n ke-5 sebesar 76,92%, 76,92%, 76,92%, 76,92% dan 76,92%.

4. KESIMPULAN

Implementasi metode *Naive Bayes* dan *Support Vector Machine* telah berhasil melakukan analisis sentimen opini masyarakat terhadap *provider* Indosat dengan hasil analisis sentimen sebagian besar adalah negatif. Metode *Naive Bayes* memiliki nilai akurasi, presisi, dan *recall* yang

lebih baik masing-masing sebesar 78%, 78% dan 77%. Pada metode SVM menguji dengan kernel *linear* dan parameter C 1 dimana hasil nilai akurasi sebesar 72%, 72% untuk presisi dan *recall* yaitu 73%. Hasil perbandingan menunjukkan bahwa *Naïve Bayes Classifier* memberikan hasil nilai pengujian yang lebih baik dalam setiap perbandingan *rasio* dan direkomendasikan untuk analisis sentimen pengguna layanan komunikasi Indosat di platform X.

5. SARAN

Penelitian selanjutnya dapat menggunakan algoritma klasifikasi yang lain seperti *KNN* dan *Decision Tree* dengan mengambil data komentar di media sosial lain seperti *facebook*, *instagram* atau pada aplikasi *google play*. Pada proses preprocessing data memiliki peran krusial dalam klasifikasi teks, di mana langkah-langkah yang diambil harus disesuaikan dengan karakteristik khusus dari dataset tersebut.

DAFTAR PUSTAKA

- [1] D. Aryanti, "Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naïve Bayes Classifier," *J. Inf. Syst. Res.*, vol. 3, no. 4, pp. 524–531, 2022, doi: 10.47065/josh.v3i4.1944.
- [2] A. Maghfiroh, Z. Arifin, and Sunarti, "PEMBELIAN (Survei pada Mahasiswa Program Studi Administrasi Binis Tahun Angkatan 2013 / 2014 Fakultas Ilmu Administrasi Universitas Brawijaya Malang Pembeli Indosat Ooredoo)," *J. Adm. Bisnis*, vol. 40, no. 1, pp. 132–140, 2014.
- [3] F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [4] U. R. H. Baba, "Analisa Sentimen Menjelang Pemilihan Umum Presideen 2024 di Indonesia Menggunakan Perbandingan Performa Support Vector Machine (SVM) dan Naïve Bayes," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 3, pp. 11972–11990, 2024.
- [5] T. T. Widowati and M. Sadikin, "Analisis sentimen Twitter terhadap tokoh publik dengan algoritma Naïve Bayes dan Support Vector Machine," *Simetris J. Tek. Mesin, Ind. Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2020.
- [6] L. A. Fitrana, S. Linawati, N. Herlinawati, R. Sa'adah, and S. Seimahuria, "Analisis Sentimen Pengguna Twitter Terhadap Brand Indosat Menggunakan Metode Naïve Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 4291–4297, 2024.
- [7] F. M. Sarimole and K. Kudrat, "Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 783–790, 2024.
- [8] 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," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 6, pp. 1253–1264, 2024.
- [9] E. Hasibuan and E. A. Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *J. Tek. Dan Sci.*,

vol. 1, no. 3, pp. 13–24, 2022.

- [10] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, “Analisis klasifikasi sentimen terhadap sekolah daring pada twitter menggunakan Supervised Machine Learning,” *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, 2021.
- [11] F. Prasetiawan, S. Widiyanesti, and T. Widarmanti, “Analisis Sentimen Mengenai Kualitas Layanan Jasa Ekspedisi Barang Sicepat Di Media Sosial Twitter Sentiment Analysis Regarding Quality of Sicepat Expedition Services On Twitter Social Media,” *e-Proceeding Manag.*, vol. 9, no. 2, pp. 1–14, 2022.
- [12] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, “Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM,” *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [13] N. I. P. Kalingara, O. N. Pratiwi, and H. D. Anggana, “Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi Jne Dan J & T Express Menggunakan Metode Naïve Bayes,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9035–9048, 2021.
- [14] R. Ridwan, E. H. Hermaliani, and M. Ernawati, “Penerapan: Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian,” *Comput. Sci.*, vol. 4, no. 1, pp. 80–88, 2024.
- [15] M. A. Aprihartha, Z. Putrawan, D. Zulhan, and F. A. Nurfaizal, “Algoritma Synthetic Minority Oversampling Technique dan C5. 0 dalam Mengatasi Ketidakseimbangan Data pada Klasifikasi Kelulusan Siswa,” *Upgrad. J. Pendidik. Teknol. Inf.*, vol. 2, no. 1, pp. 1–10, 2024.
- [16] N. Puspitasari, M. Ajay, M. Wati, and A. Septiarini, “Classification for Determining the Level of Drugs Dependence Using the Naïve Bayes Classifier,” *IT J. Res. Dev.*, vol. 9, no. 1, pp. 68–79, 2024.
- [17] E. R. Indriyani, P. Paradise, and M. Wibowo, “Perbandingan Metode Naïve Bayes dan Support Vector Machine Untuk Analisis Sentimen Terhadap Vaksin Astrazeneca di Twitter,” *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1545, 2022, doi: 10.30865/mib.v6i3.4220.
- [18] M. Farid and E. Firdaus, “Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, pp. 1329–1336, 2022, doi: 10.30865/jurikom.v9i5.4774.
- [19] D. A. Wulandari, “Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Naive Bayes Analysis of Twitter Social Media Sentiment on the Public ’ S Reaction To the Drafts of Job Creation Law Using the Classification Method Naive Bayes,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9007–9016, 2021.