Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

ANALISIS SENTIMEN PADA MEDIA SOSIAL TERHADAP LAYANAN SAMSAT DIGITAL NASIONAL DENGAN SUPPORT VECTOR MACHINE

Anindya Sasi Kirana[1*], Rusdah[2], Ririt Roeswidiah[3], Ahmad Pudoli[3]
[1]Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia
[2]Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia
[3]Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia
Email: [1*]anindyasasik@gmail.com, [2]rusdah@budiluhur.ac.id, [3]ririt@budiluhur.ac.id, [3]ahmad.pudoli@budiluhur.ac.id
(* : coressponding author)

Abstrak-Pengguna kendaraan bermotor mengalami perkembangan pesat setiap tahunnya. Peningkatan kendaraan berkontribusi terhadap sumber penerimaan negara yaitu pajak. SAMSAT merupakan lembaga negara yang berwenang mengatur pajak kendaraan bermotor (PKB). Seiring perkembangan teknologi, SAMSAT berinovasi melalui aplikasi SIGNAL dimana masyarakat diberikan kemudahan untuk melakukan pembayaran pajak kendaraan bermotor dengan aman melalui telepon seluler. Media sosial seperti *Instagram* dan *X* berpotensi besar untuk mengumpulkan data dalam memahami reaksi publik terhadap aplikasi SIGNAL. Komentar pada media sosial terhadap aplikasi SIGNAL menimbulkan pro dan kontra dari masyarakat, oleh karena itu perlu dilakukannya analisis sentimen melalui pendekatan *text mining* menggunakan algoritma *Support Vector Machine* (SVM) dengan mengikuti metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM). Penelitian dilakukan melalui beberapa tahap termasuk pengumpulan data, *preprocessing*, pemodelan dengan algoritma *Support Vector Machine* (SVM), hingga evaluasi dengan *confusion matrix*. Data penelitian bersumber dari komentar media sosial *Instagram* sejak 20 September 2023 s/d. 16 April 2024 sebanyak 3.543 record dan 1.335 komentar pada media sosial *X* sejak tanggal 31 Mei 2023 s/d. 27 Maret 2024 dengan kata kunci "aplikasi SIGNAL". Setelah tahap *preprocessing* data yang digunakan berkurang menjadi 3.911 karena terdapat ulasan duplikat dan ulasan yang tidak relevan. berdasarkan 3.911 data menghasilkan 773 komentar *positif*, 1991 *negatif* dan 1147 komentar *neutral*. Penelitian ini bertujuan untuk mengidentifikasi sentimen masyarakat terhadap pelayanan SIGNAL melalui media sosial *Instagram* dan *X* serta membuat model klasifikasi sentimen menggunakan *Support Vector Machine* (SVM) dan kernel terbaik yang nantinya diharapkan dapat menjadi perbaikan bagi pengembang. Untuk kebutuhan pemodelan penelitian menyiapkan dataset dua kelas sentimen dan tiga kelas sentimen. Berdasarkan penerapan model, *Support Vector Machine* (SVM) dengan kernel *linear* menghasilkan nilai lebih baik dibandingkan model Naïve Bayes dan KNN dengan nilai akurasi 0.88, presisi 0.88, *recall* 0.81 dan AUC 0.92 menggunakan *10-fold cross validation* pada data latih dan data uji.

Kata Kunci: analisis sentimen, CRISP-DM, penambangan teks, support vector machine, sinyal.

Abstract- *Motor vehicle users experience rapid growth every year. The increasing number of vehicles contributes to one of the state revenues: taxes. SAMSAT is a state institution with the authority to regulate motor vehicle tax (PKB). As technology develops, SAMSAT innovates through the SIGNAL application, which allows people to make motor vehicle tax payments safely via cell phone. Social media such as Instagram and X have great potential for collecting data to understand public reactions to the SIGNAL application. Comments on social media regarding the SIGNAL application raise pros and cons from the public; therefore, it is necessary to carry out sentiment analysis through a text mining approach using the Support Vector Machine (SVM) algorithm following the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. This research was carried out through several stages: data collection, preprocessing, modeling with the Support Vector Machine (SVM), and evaluation with a confusion matrix. Data in the research were collected from Instagram social media comments from September 20, 2023, until. 16 April 2024 as many as 3,543 records and 1,335 comments on X's social media from 31 May 2023 until March 27, 2024, with the keyword "SIGNAL application". After the preprocessing stage, the data used was reduced to 3,911 because there were duplicate and irrelevant reviews. based on 3,911 data, it produced 773 positive comments, 1991 negative, and 1147 neutral comments. This research aims to identify public sentiment towards SIGNAL services via social media, such as Instagram. We prepared a dataset of two and three sentiment classes for research modeling needs. Based on the application of the model, a Support Vector Machine (SVM) with a linear kernel produces better scores than the Naïve Bayes and KNN models with accuracy values of 0.88, precision of 0.88, recall of 0.81, and AUC of 0.92 using a 10-fold cross-validation on training data and test data.*

***Keywords**: CRISP-DM, Sentiment Analysis, Support Vector Machine, Signal, Text Mining.*

1. PENDAHULUAN

Pengguna kendaraan bermotor mengalami perkembangan pesat setiap tahunnya. Melalui Badan Pusat Statistik tahun 2022, jumlah kendaraan bermotor telah meningkat secara signifikan [1]. Peningkatan jumlah kendaraan ini berperan dalam menambah sumber penerimaan negara melalui pajak. Setiap pemilik kendaraan bermotor wajib dikenakan Pajak Kendaraan Bermotor (PKB) [2]. SAMSAT merupakan lembaga negara yang berwenang mengatur PKB dengan tujuan memastikan setiap warga negara yang memiliki kendaraan patuh terhadap PKB. SAMSAT

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

melakukan pembaruan dalam pelayanannya dengan menghadirkan aplikasi SIGNAL untuk memberi masyarakat kemudahan dan keamanan dalam melakukan pembayaran PKB melalui telepon seluler.

Sejak diluncurkannya aplikasi SIGNAL ke publik, pengguna layanan dapat menyampaikan komentarnya baik keluhan maupun pujian melalui media sosial resmi *Instagram* @samsatdigital dan *X*. Pada komentar laman sosial media *Instagram* resmi @samsatdigital beberapa pengguna menyampaikan keluhan terkait lambatnya pelayanan yang diberikan Selain itu, banyak pula pengguna yang merasa diuntungkan dengan kehadiran aplikasi SIGNAL. Pro dan kontra ini tentu saja dapat mempengaruhi pelayanan yang diberikan sehingga perlu dilakukannya analisis sentimen terhadap aplikasi SIGNAL guna mengetahui pendapat atau opini pengguna mengandung sentimen positif atau negatif.

Penelitian terdahulu terkait analisis sentimen mengenai aplikasi SIGNAL pernah dilakukan. Pada penelitian [3] menggunakan algoritma *Naïve Bayes* dengan sumber data ulasan *Google Play Store* menghasilkan nilai akurasi 63.61%, presisi 92.19% dan *recall* 61.52%. Selain itu, penelitian mengenai analisis sentimen dengan algoritma KNN pernah dilakukan oleh [4] dengan hasil pengujian dan evaluasi menggunakan nilai K=3 untuk setiap kata kunci menunjukkan bahwa kata kunci "Ganjar Pranowo" menghasilkan akurasi sebesar 77%, presisi 77%, dan recall 100%. Sementara itu, kata kunci "Prabowo Subianto" memberikan akurasi 97%, presisi 87%, dan recall 100%. Untuk kata kunci "Anies Baswedan," diperoleh akurasi 67%, presisi 44%, dan recall 42%. Kemudian berdasarkan penelitian oleh [5] menggunakan algoritma *Support Vector Machine* (SVM) dengan *Google Play Store* sebagai sumber data melakukan prediksi tiga kernel. Menghasilkan akurasi *Support Vector Machine* (SVM) kernel Linier 96.2%, kernel RBF 94.12%, dan kernel *polynomial* 85.5%. Dilakukan juga evaluasi dengan KFold linier mencapai 97.65%, KFold RBF 97.86%, dan KFold *polynomial* 69.36%. Penelitian terkait analisis sentimen lainnya juga pernah dilakukan oleh [6] dengan mengkomparasi beberapa algoritma klasifikasi. Dalam penelitiannya, dilakukan komparasi 3 performa algoritma klasifikasi diantaranya *Naive Bayes, Support Vector Machine, K-Nearest Neighbour*, dan *feature selection* algoritma *Particle Swarm Optimization* (PSO). Perbandingan ketiga algoritma tersebut menghasilkan nilai akurasi terbaik adiperoleh algoritma PSO berbasis *Support Vector Machine* (SVM) 78.55% dan AUC sebesar 0.853 dengan algoritma PSO berbasis *Support Vector Machine* (SVM).

Penelitian ini bertujuan untuk mengidentifikasi sentimen masyarakat melalui media sosial *Instagram* dan X terhadap pelayanan yang diberikan oleh aplikasi SIGNAL serta membuat model klasifikasi sentimen masyarakat terhadap pelayanan aplikasi SIGNAL dengan menggunakan *Support Vector Machine* (SVM) dan kernel terbaik yang diharapkan dari hasil dapat berguna untuk bahan evaluasi dalam meningkatkan mutu pelayanan SIGNAL. Penelitian ini memanfaatkan media sosial *Instagram* @samsatdigital dan *X* sebagai sumber data utama untuk diteliti. Data yang diperoleh akan diproses dengan teknik *text mining* untuk dilakukan menganalisis sentimen degan algoritma klasifikasi. Penelitian sebelumnya menunjukkan bahwa masing-masing algoritma memiliki performa yang bervariasi dalam analisis sentimen maka diperlukannya perbandingan algoritma untuk memberikan pandangan yang lebih luas mengenai kinerja metode dalam berbagai jenis data. Oleh karena itu, penelitian ini membandingkan ketiga algoritma klasifikasi diantaranya, *Support Vector Machine* , *Naïve Bayes*, dan *K-Nearest Neighbour* dalam mengidentifikasi algoritma yang paling optimal untuk mengklasifikasikan opini pengguna layanan SIGNAL.

2. METODE PENELITIAN

Metode penelitian merupakan langkah yang menjelaskan secara detail mengenai penerapan metode yang dipilih. Penggunaan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) digunakan dalam penelitian ini. Metode CRISP-DM merupakan metode yang diakui komprehensif untuk proses pengembangan proyek industri dan serta menjadi metode yang banyak digunakan dalam proyek penambangan data [7] .Gambar 1 menunjukkan implementasi CRISP-DM dalam penelitian. Dibawah ini merupakan penjelasan lebih lanjut mengenai tahapan penelitian pada Gambar 1.

2.1 *Business Understanding (Pemahaman Bisnis)*

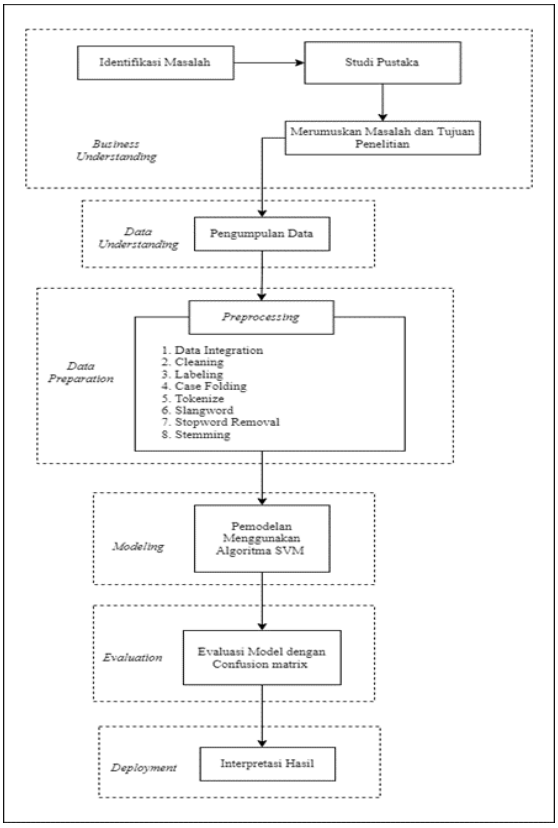
Langkah awal yang dilakukan adalah mengidentifikasi permasalahan. Permasalahan yang diangkat adalah bagaimana sentimen dari pengguna media sosial *Instagram* dan *X* mengenai pelayanan aplikasi SIGNAL berdasarkan komentar pengguna. Kemudian mencari informasi mengenai samsatdigital, serta membaca jurnal dan buku yang berkaitan dengan penelitian. Untuk melengkapi tahapan diatas, langkah yang dilakukan adalah menetapkan tujuan yang ingin dicapai, yaitu mengidentifikasi opini pengguna layanan terhadap aplikasi SIGNAL.

2.2 *Data Understanding* (Pemahaman Data)

Proses *scraping* dilakukan untuk mengumpulkan data komentar *Instagram* @samsatdigital dengan bantuan *tools Apify* sejak tanggal 20 September 2023 s/d. 16 April 2024 menghasilkan 7 atribut dan 3.543 *record*. Sedangkan pengambilan data media sosial X dilakukan dengan *crawling* menggunakan bahasa pemrograman *python* dengan kata

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

kunci ‘aplikasi SIGNAL’ dilakukan sejak tanggal 31 Mei 2023 s/d. 27 Maret 2024 menghasilkan 12 atribut dan 1.335 *record*.



Gambar 1. Tahapan Penelitian

2.3 *Data Preparation*

Perancangan penelitian dimulai dari tahap *preprocessing* karena sebuah teks tidak dapat diproses secara langsung oleh algoritma [8]. Untuk dapat menggunakan data, perlu dilakukan tahap *preprocessing* dimana data mentah diolah menjadi data yang siap digunakan [9]. Berikut proses yang dilakukan pada tahap *preprocessing:*

a. *Data integration*, pada proses ini data yang berhasil didapatkan dari komentar media sosial *Instagram* dan *X* akan digabungkan menjadi satu file csv.
b. *Cleaning*, pembersihan data dengan *data reduction* dan *data cleaning*. Tujuannya untuk menghilangkan duplikasi data serta karakter seperti html, tanda baca *(punctuations)*, angka, dan ruang kosong (ruang kosong) [10].
c. *Labeling*, proses pembelian label sentimen secara manual oleh pakar ahli bahasa dengan label *positif, negatif, neutral*.
d. *Case folding*, pada proses ini seluruh huruf pada data akan disamakan bentuknya menjadi huruf kecil [11].
e. *Tokenizing*, proses ini melakukan pemecahan pada kalimat menjadi satuan kata[12].
f. *Slangword*, proses ini mengubah kata gaul *(slang)* menjadi kata baku sesuai KBBI [13].
g. *Stopword Removal*, betujuan untuk menghilangkan kata yang tidak memiliki makna [14].
h. *Stemming*, proses ini mengubah kata ke bentuk dasar dengan menghilangkan imbuhan[15].
i. *Term Weighting* (TF-IDF), merupakan perhitungan nilai *Term Frequency* (TF), *Document Frequency* (DF), dan *Inverse Document Frequency* (IDF) dengan tujuan melakukan perhitungan bobot untuk mengevaluasi nilai setiap kata terhadap dokumen [16].
j. Penentuan data latih dan data uji dengan metode *split data* dan *cross validation* dalam mencari komposisi dataset terbaik.

2.4 *Modeling* (Pemodelan)

Tahap pemodelan dilakukan dengan mengeksplorasi beberapa algoritma yaitu, *Support Vector Machine* (SVM) dengan tiga kernel (*linear, polynomial*, RBF), *Naïve Bayes* dan KNN. Melalui komparasi model diharapkan dapat menemukan algoritma dengan performa terbaik.

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

2.5 *Evaluation* (Evaluasi)

Pada tahap evaluasi model diuji dengan *confusion matrix* dengan *10-fold cross validation*. *Confusion matrix* memberikan hasil evaluasi berupa *classification report*, parameter uji pada penelitian ini yaitu nilai akurasi, presisi, *recall*, serta nilai AUC. Pengukuran kinerja ini bertujuan untuk mengetahui apakah metode ini memiliki performa yang baik serta dapat memenuhi tujuan yang ditetapkan dengan mengklasifikasikan sentimen *positif* dan *negatif*. Rumus untuk menghitung nilai *confusion matrix* yaitu:

a. Akurasi, merupaka rasio antara data yang benar dengan total jumlah data.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

b. Presisi, menilai tingkat akurasi model dalam mengklasifikasikan data yang telah diprediksi, ini menggambarkan seberapa efektif model dalam mengurangi kesalahan saat memprediksi kelas *positif* dari total kelas yang diprediksi dengan benar, serta jumlah sebenarnya dari data yang *positif*.

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

c. *Recall,* menggambarkan data kelas positif yang terprediksi dengan benar dari total keseluruhan data dalam kelas positif.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data pada penelitian ini bersumber dari data primer. Data didapatkan secara langsung dari sumber utama yaitu komentar media sosial *Instagram* dan *X*. Pengumpulan data komentar *Instagram* dilakukan dengan memanfaatkan *tools Apify* meliputi komentar pada data *image* dan data video. Sedangkan pengumpulan data media sosial X dilakukan menggunakan bahasa pemrograman *pyhton* dengan kata kunci 'aplikasi SIGNAL' menggunakan bantuan *tools Google Collab*.

Raw data yang berhasil diperoleh dari proses *scraping* dan *crawling berjumlah* 4.878 komentar. Data media sosial *Instagram* dikumpulkan dari beberapa unggahan resmi milik samsat, @samsatdigital. Unggahan dipilih sejak tanggal 20 September 2023 s/d. 16 April 2024 menghasilkan 3.543 *record*. Sedangkan data media sosial *X* dikumpulkan sejak tanggal 31 Mei 2023 s/d. 27 Maret 2024 dengan bantuan *tools Google Colab* menghasilkan total 1.335 *record*. data

3.2 *Data Preprocessing*

Setelah data berhasil dikumpulkan, berikutnya adalah tahap *preprocessing* yang diterapkan dalam penelitian ini.:

3.2.1 *Data Integration*

Dalam tahap penelitian ini dilakukan penggabungan dari hasil pengumpulan dua dataset diatas dengan hanya menggunakan kolom komentar. Kedua dataset tersebut akan disatukan kedalam file csv menjadi kolom *username* dan *text* mengunakan *tools* Ms.Excel. Tabel 1 merupakan contoh hasil penggabungan data *scraping Instagram* dan *crawling X*.

Tabel 1. Penggabungan Dataset

text
Min kok lemot banget ya aplikasinya @samsatdigital
@wiwinbong saya jg error lemot
Gimana nih min aplikasinya? Dari jam 8 saya coba masih eror

3.2.2 *Cleaning*

Pada tahap ini dilakukan pembersihan pada dataset dengan *data reduction* dan *data cleaning*. *Data reduction* menghapus balasan dari admin dengan *username* 'samsatdigital' dan 'SamsatDigital' karena dianggap tidak relevan

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

serta menghapus duplikasi. Kemudian dilakukan *data cleaning* dengan menghapus *link, mention, hashtag,* emoji, angka, tanda baca, karakter spesial, menghapus baris yang memiliki nilai kosong, menghapus kata tidak berguna seperti "wkwk", "min" dan yang lainnya. Setelah dilakukan pembersihan pada data, data berkurang menjadi 3.911 komentar. Selanjutnya hasil data untuk diserahkan kepada pakar ahli bahasa untuk diberi label sentimen. Tabel 2 berikut merupakan contoh komentar yang telah dilakukan pembersihan pada data.

Tabel 2. *Data Cleaning*

Sebelum *data cleaning*	Setelah *data cleaning*
Min kok lemot banget ya aplikasinya @samsatdigital	kok lemot banget ya aplikasinya
@wiwinbong saya jg error lemot	saya jg error lemot
Gimana nih min aplikasinya? Dari jam 8 saya coba masih eror	Gimana nih aplikasinya? Dari jam 8 saya coba masih eror

3.2.3 *Labeling*

Pemberian sentimen dilakukan oleh pakar ahli bahasa dengan menganalisa setiap komentar menurut kaidah kebahasaan dengan bantuan tools Ms.Excel. Label sentimen pada komentar terdiri dari sentimen negatif, positif dan *neutral*. Tabel 3 dibawah ini merupakan komentar yang sudah diberikan sentimen. Dari hasil pelabelan diperoleh data sebanyak 3911 dengan 1991 data *negatif*, 773 data *positif* dan 1147 data *neutral*. Untuk kebutuhan pemodelan *(modeling)*, penelitian ini juga menyiapkan dataset dengan dua kelas sentimen yaitu *negatif* dan *positif* dengan menghapus data berlabel *neutral*, sebanyak 1147 *record*.

Tabel 3. Komentar Dengan Label

Text	Sentimen
proses penerbitan kak	neutral
live chatnya gabisa, katanya masih ditutup. Gimana sih	negatif
dumb email saya dibalas koq komplain waktu pembayaran sdh kadaluarsa saya lampirkan bukti pembayaran	positif

3.2.4 *Case Folding*

Proses *case folding* bertujuan untuk mengubah bentuk huruf pada sebuah teks menjadi huruf kecil *(lowercase)*, proses ini terlihat pada Tabel 4.

Tabel 4. *Case Folding*

Sebelum *Case Folding*	Sesudah *Case Folding*
Dari tgl april sampai april stnk fisik gw belum dtng juga padahal udh pke pengiriman express Gak lgi gw perpanjang stnk di online	dari tgl april sampai april stnk fisik gw belum dtng juga padahal udh pke pengiriman express gak lgi gw perpanjang stnk di online
udah offline. Bintang 1 banget sih pelayanannya	udah offline. bintang banget sih pelayanannya

3.2.5 *Tokenizing*

Tahap selanjutnya adalah tokenize yaitu menguraikan kalimat pada data komentar menjadi satuan kata. Tabel 5 merupakan perbandingan sebelum proses *tokenize* dan sesudah proses *tokenize*.

Tabel 5. *Tokenize*

Sebelum *Tokenize*	Sesudah *Tokenize*
Dari tgl april sampai april stnk fisik gw belum dtng juga padahal udh pke pengiriman express Gak lgi gw perpanjang stnk di online	['dari', 'tgl', 'april', 'sampai', 'april', 'stnk', 'fisik', 'gw', 'belum', 'dtng', 'juga', 'padahal', 'udh', 'pke', 'pengiriman', 'express', 'gak', 'lgi', 'gw', 'perpanjang', 'stnk', 'di', 'online']

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Sebelum *Tokenize*	Sesudah *Tokenize*
udah offline. Bintang 1 banget sih pelayanannya	['udah', 'offline', 'bintang', 'banget', 'sih', 'pelayanannya']

3.2.6 *Slangword*

pada tahap *slangword* dilakukan pengubahan kata tidak baku menjadi kata baku sesuai KBBI. Untuk menerapkan proses *slangword*, proses sebelumnya akan dilakukan normalisasi untuk memastikan konsistensi teks dapat dilihat pada Tabel 6.

Tabel 6. *Slangword*

Sebelum S*langword*	Sesudah S*langword*
Dari tgl april sampai april stnk fisik gw belum dtng juga padahal udh pke pengiriman express Gak lgi gw perpanjang stnk di online	dari tanggal april sampai april stnk fisik saya belum dtng juga padahal sudah pakai pengiriman express tidak lagi saya perpanjang stnk di daring
udah offline. Bintang 1 banget sih pelayanannya	sudah offline bintang banget sih pelayanannya

3.2.7 *Stopword Removal*

Merupakan proses penghilangan kata yang tidak memiliki makna seperti seperti kata “dari”, “akan”, “atau”, “yang”. Tujuannya mengurangi jumlah kata yang akan digunakan untuk model latih agar menghasilkan dimensi yang lebih efisien. Hasil *stopword removal* dapat dilihat pada Tabel 7.

Tabel 7. *Stopword Removal*

Sebelum *Stopword Removal*	Sesudah *Stopword Removal*
Dari tgl april sampai april stnk fisik gw belum dtng juga padahal udh pke pengiriman express Gak lgi gw perpanjang stnk di online	tanggal april april stnk fisik dtng pakai pengiriman express perpanjang stnk daring
udah offline. Bintang 1 banget sih pelayanannya	offline bintang banget pelayanannya

3.2.8 *Stemming*

Proses *stemming* bertujuan untuk menghilangkan awalan, akhiran, imbuhan, atau kata depan dari suatu kata serta menguraikan kata ke bentuk dasarnya. Proses ini dapat dilihat pada Tabel 8.

Tabel 8. *Stemming*

Sebelum *Stemming*	Sesudah *Stemming*
udah hubungin via live chat tp katanya agent kami tidak berada ditempat semua, ini gmn sih aplikasinya?	hubungin lalu live chat kata agen tidak ada tempat semua bagaimana aplikasi
udah offline. Bintang 1 banget sih pelayanannya	offline bintang banget layan

3.2.9 *Term Weighting* (TF-IDF)

Setelah rangkaian tahapan preprocessing sudah dilakukan, tahapan terakhir yang dilakukan adalah pembobotan term atau term weighting. Pembobotan ini bertujuan mengubah dokumen yang berisikan teks menjadi angka agar dapat memudahkan pembelajaran mesin. Penelitian ini menggunakan kelas *tfidfvectorizer* dari *library sklearn* untuk memberikan bobot pada jumlah kata berdasarkan seberapa sering kata tersebut muncul dalam dokumen.

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

3.3 Penentuan Data Latih dan Data Uji

Setelah menyelesaikan tahap *preprocessing*, data akan dibagi menjadi data latih *(training)* dan data uji *(testing)*. Data latih akan digunakan untuk melatih model klasifikasi sedangkan data uji digunakan untuk proses uji klasifikasi [17]. Penentuan data latih dan data uji pada 3 kelas menggunakan *split data* dengan rasio 60:40, 70:30 dan 80:20 dapat dilihat lebih lengkap dalam Tabel 9.

Tabel 9. Pembagian Data Latih dan Data Uji 3 Kelas

Perbandingan 60:40			
Sentimen	*Training*	*Testing*	Total
Positif	469	304	773
Negatif	1193	784	1977
Netral	662	462	1124
Total	2324	1550	3874
Perbandingan 70:30			
Sentimen	*Training*	*Testing*	Total
Positif	546	227	773
Negatif	1382	595	1977
Netral	783	341	1124
Total	2711	1163	3874
Perbandingan 80:20			
Sentimen	*Training*	*Testing*	Total
Positif	617	156	773
Negatif	1578	399	1977
Netral	904	220	1124
Total	3099	775	3874

Penentuan data latih dan data uji dengan *split data* pada 2 kelas sentimen dapat dilihat lebih lengkap dalam Tabel 10.

Tabel 10. Pembagian Data Latih dan Data Uji 2 Kelas

Perbandingan 60:40			
Sentimen	*Training*	*Testing*	Total
Positif	456	317	773
Negatif	1194	783	1977
Total	1650	1100	2750
Perbandingan 70:30			
Sentimen	*Training*	*Testing*	Total
Positif	533	240	773
Negatif	1392	585	1977
Total	1925	825	2750
Perbandingan 80:20			
Sentimen	*Training*	*Testing*	Total
Positif	614	159	773
Negatif	1586	391	1977
Total	2200	550	2750

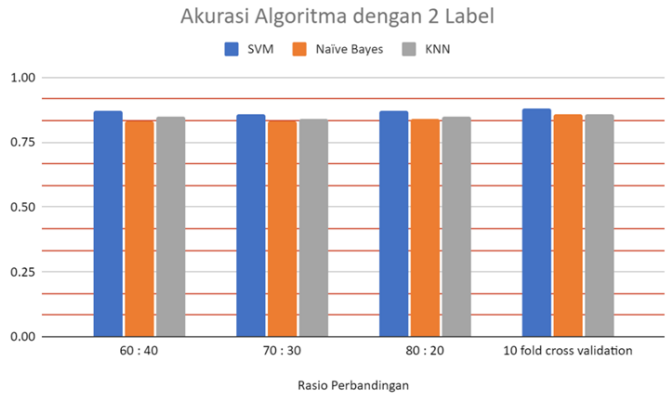
3.4 *Modeling*

Pada tahap *modeling* akan dilakukan komparasi algoritma klasifikasi yaitu *Support Vector Machine* (SVM), Naïve Bayes dan KNN. Proses ini dilakukan dengan teknik *split data* pada kedua kelas. Selain itu, pemodelan juga dilakukan dengan menggunakan *cross* validation dengan membagi data secara acak kedalaam 10 lipatan *(10-fold cross validation)*.

3.4.1 Komparasi Model

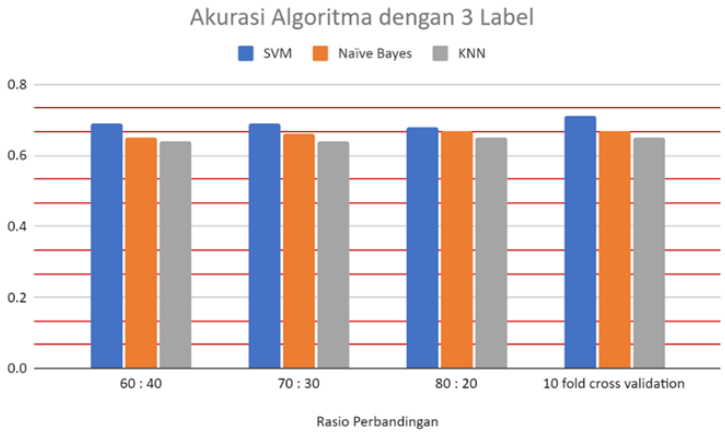
Hasil pemodelan menggunakan *split data* dengan rasio perbandingan 60:40, 70:30 dan 80:20 dan pemodelan dengan *10-fold cross validation* label dua kelas dapat dilihat pada Gambar 2.

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index



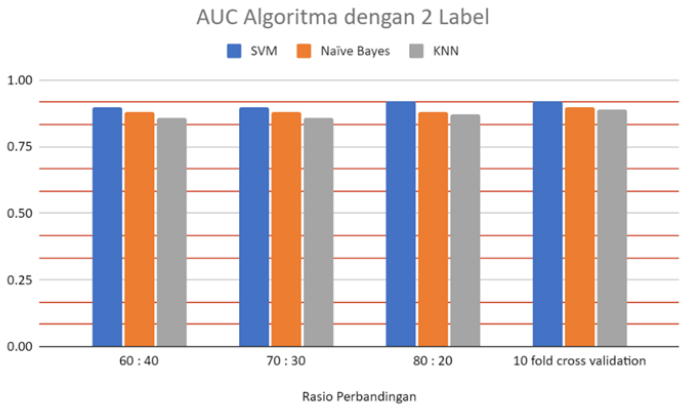
Gambar 2. Visualisasi Akurasi Label Dua Kelas

Gambar 3 merupakan visualisasi perbandingan nilai akurasi dengan *split data* dan 10-*fold cross validation* pada label tiga kelas.



Gambar 3. Visualisasi Akurasi Label Tiga Kelas

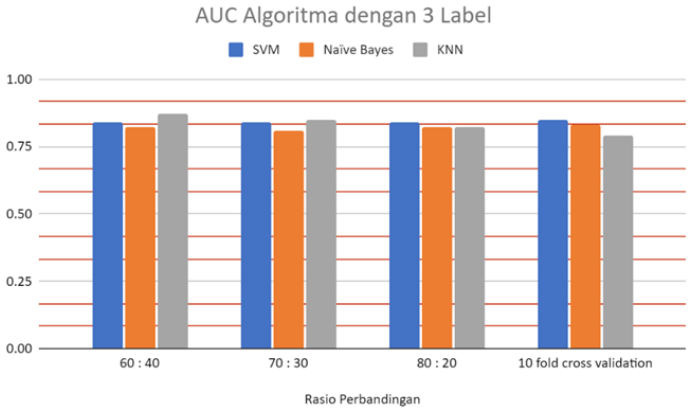
Dari sudut pandang akurasi, terlihat algoritma *Support Vector Machine* (SVM) dengan dua kelas sentimen memiliki nilai *accuracy* tertinggi diantara algoritma *Naïve Bayes* dan KNN dengan nilai akurasi 0.88 menggunakan *10-fold cross validation*.
Selanjutnya dilakukan perbandingan nilai AUC dari proses pemodelan dengan teknik *split data* dan *10 fold cross validation*. Gambar 4 merupakan visualisasi nilai AUC dengan dua kelas sentimen.



Gambar 4. Visualisasi AUC Label Dua Kelas

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Gambar 5 merupakan visualisasi perbandingan nilai AUC dengan *split data* dan *10-fold cross validation* pada label tiga kelas

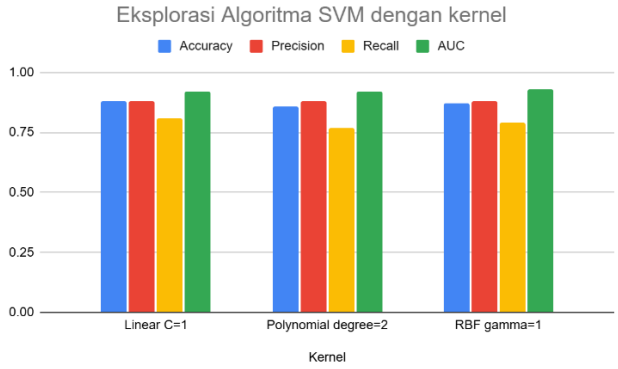


Gambar 5. Visualisasi AUC Label Tiga Kelas

Berdasarkan komparasi nilai *accuracy* dan AUC menggunakan teknik *split data* dan *cross validation* komposisi *dataset* terbaik diperoleh teknik *10-fold cross validation* pada label dua kelas dengan nilai *accuracy* 0.88 dan AUC 0.92 menggunakan algoritma *Support Vector Machine* (SVM).

3.4.2 Penyajian Model terbaik

Setelah dilakukan pemodelan, model terbaik diperoleh oleh algoritma *Support Vector Machine* (SVM) dengan *10-fold cross validation*. Untuk menemukan model terbaik dilakukan eksplorasi algoritma *Support Vector Machine* (SVM) dengan kernel *linear, polynomial* dan RBF dapat dilihat pada Gambar 6.



Gambar 6. Eksplorasi SVM dengan kernel

Berdasarkan **Error! Reference source not found.** diperoleh algoritma *Support Vector Machine* (SVM) kernel *linear* dengan parameter C=1 menghasilkan nilai *accuracy* mencapai 0.88 menunjukkan kinerja yang lebih baik dibandingkan kernel *polynomial* parameter *degree*=2 dengan nilai *accuracy* 0.86 dan kernel RBF parameter *gamma*=1 dengan nilai *accuracy* 0.87. Eksplorasi ketiga kernel *Support Vector Machine* (SVM) ini juga menghasilkan nilai, *precision*, *recall* dan AUC yang relatif tinggi yang menunjukkan bahwa model *Support Vector Machine* (SVM) mampu mengklasifikasikan data dengan baik.

3.5 Pengujian *(Evaluation)*

Hasil pemodelan dengan *10-fold cross validation* menggunakan algoritma *Support Vector Machine* (SVM) kernel *linear* menunjukkan evaluasi model yang baik secara keseluruhan, langkah selanjutnya adalah pengujian dengan *confusion matrix*. Tabel 11 menunjukkan nilai rata-rata dari *confusion matrix* dengan *cross validation* algoritma *Support Vector Machine* (SVM) *kernel linear*.

Idealis: Indonesia Journal Information System
Volume 8, Nomor 1, Januari 2025
ISSN 2684-7280 (*online*)
Halaman 53-63
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Tabel 11. *Confusion Matrix* SVM dengan *Cross Validation*

	True Negatif	*True Positif*	*Class Precision*
Pred Negative	1902	75	0.96
Pred Positive	258	515	0.67
Class Recall	0.88	0.87	

$$\text{Akurasi} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} = \frac{515+1902}{515+1902+258+75} = 0.88$$

$$\text{Presisi} = \frac{\text{TP}}{\text{TP+FP}} = \frac{515}{515+258} = 0.67$$

$$\textit{Recall} = \frac{\text{TP}}{\text{TP+FN}} = \frac{515}{515+75} = 0.87$$

Berdasarkan *confusion matrix* diperoleh nilai presisi 0.96 yang merupakan sentimen negatif dari keseluruhan data komentar yang di prediksi negatif. Sedangkan presisi dengan nilai 0.67 merupakan sentimen positif dari keseluruhan data komentar yang di prediksi positif. *Recall* 0.88 merupakan sentimen yang diprediksi negatif dibandingkan dengan keseluruhan komentar yang sebenarnya negatif. *Recall* 0.87 merupakan sentimen yang diprediksi positif dibandingkan dengan keseluruhan komentar yang sebenarnya positif. Algoritma *Support Vector Machine* (SVM) kernel *linear* menggunakan teknik *10-fold cross validation* berdasarkan Tabel 11 menghasilkan pengujian terbaik dengan nilai akurasi 0.88 dan nilai AUC sebesar 0.92.

4. KESIMPULAN

Berdasarkan pengujian pada penelitian data komentar yang dihasilkan setelah tahap *preprocessing* berjumlah 3.911, terdiri dari 773 komentar *positif*, 1.991 komentar *negatif*, dan 1.147 komentar *neutral*. Data tersebut diperoleh dari laman resmi media sosial *Instagram* @samsatdigital dan media sosial *X* dengan kata kunci 'aplikasi SIGNAL'. Berdasarkan ulasan pengguna di media sosial *Instagram* dan *X* mengenai aplikasi SIGNAL belum berfungsi dengan maksimal, terbukti dari hasil penelitian berupa jumlah komentar negatif jauh lebih banyak dibandingkan dengan komentar positif. Setelah melakukan pengujian algoritma klasifikasi yaitu *Support Vector Machine* (SVM), Naïve Bayes dan KNN diperoleh *Support Vector Machine* (SVM) dengan dua kelas sentimen sebagai algoritma terbaik. Selanjutnya, eksplorasi algoritma *Support Vector Machine* (SVM) dengan kernel *linear*, parameter C=1, dan teknik *10-fold cross validation* menunjukkan performa optimal dalam menganalisis sentimen terhadap layanan aplikasi SIGNAL, dengan nilai *a*kurasi 0,88, presisi0,88, *recall* 0,81, dan AUC 0,92. Disarankan untuk penelitian mendatang agar melibatkan lebih banyak responden dalam proses manual *labeling* untuk memperoleh pendapat yang lebih beragam serta melakukan analisis lebih mendalam terkait aspek-aspek yang perlu dievaluasi oleh pengembang aplikasi SIGNAL.

DAFTAR PUSTAKA

[1] B. P. Statistik, "Perkembangan Jumlah Kendaraan Bermotor Menurut Jenis (Unit), 2021-2022," 2024. https://www.bps.go.id/id/statistics-table/2/NTcjMg==/perkembangan-jumlah-kendaraan-bermotor-menurut-jenis--unit-.html (accessed Mar. 20, 2024).

[2] S. D. Nasional, "Mengapa Bayar Pajak Kendaraan Bermotor?," 2021. https://samsatdigital.id/artikel/mengapa-bayar-pajak-kendaraan-bermotor (accessed Mar. 20, 2024).

[3] D. Wijaya, R. A. Saputra, and F. Irwiensyah, "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 2369–2380, 2024, doi: 10.30865/klik.v4i4.1738.

[4] R. A. Putra and W. Pramusinto, "Sentimen Tweet Pada Elektabilitas Bakal Calon Presiden 2024 Implementation K-Nearest Neighbor (Knn) Sentiment Analysis Of 2024 Presidential Candidates ' Electability," vol. 2, no. September 2023, pp. 985–994, 2024.

[5] S. Kacung, B. C. P. Putra, and D. Cahyono, "Analisis sentimen terhadap layanan samsat digital nasional (signal) menggunakan metode svm 1,2,3," *MNEMONIC*, vol. 7, no. 1, pp. 118–122, 2024.

[6] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *TEKNOINFO*, vol. 14, no. 2, pp. 116–124, 2020, doi: 10.33365/jti.v14i2.679.

[7] S. Mahendra and S. Syofian, "Penggunaan Algoritma Support Vector Machine (SVM) untuk Menganalisis Sentimen dari Ulasan Pelanggan Terhadap Layanan Kurir J&T Express di Google Play Store," *J. Sains Teknol.*

, vol. 13, no. 2, pp. 29–36, 2023, [Online]. Available: https://unsada.e-journal.id/jst/article/view/450
[8] Y. Findawati and M. A. Rosid, *Buku Ajar Text Mining*. Sidoarjo, Jawa TImur: UMSIDA Press, 2020.
[9] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM) Sentiment Analysis On The Discussion Of Relocating I Ndonesia ' S Capital City Using The Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
[10] T. A. Prasetiarini, I. Ernawati, and N. Chamidah, "Analisis Sentimen Media Sosial Twitter Terhadap Maskapai Penerbangan PT Garuda Indonesia (Persero) Tbk Menggunakan Metode Support Vector Machine (SVM)," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 1, no. 2, pp. 637–646, 2020.
[11] R. Darmawan and S. Amini, "Perbandingan Hasil Sentimen Analysis Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor pada Twitter Comparison of Sentiment Analysis Results Using Naïve Bayes and K-Nearest Neighbor Algorithm on Twitter," *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 495–501, 2022, [Online]. Available: https://senafti.budiluhur.ac.id/index.php/
[12] R. A. Fauzianto and Supatman, "Analisis Sentimen Opini Masyarakat Terhadap Tech Winter Pada Twitter Menggunakan Natural Language Processing," *J. Syntax Admiration*, vol. 4, no. 9, pp. 1577–1585, 2023, doi: 10.46799/jsa.v3i9.909.
[13] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
[14] E. P. Sutrisno and S. Amini, "Implementasi Algoritma K-Nearest Neighbor Pada Implementation Of K-Nearest Neighbor Algorithm In Sentiment Analysis Of User Reviews For Digital," *SENAFTI*, vol. 2, pp. 687–695, 2023.
[15] A. D. A. Putra and S. Juanita, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
[16] R. Kosasih and A. Alberto, "Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 1, pp. 134–139, 2021, [Online]. Available: https://doi.org/10.30743/infotekjar.v6i1.3893
[17] J. Adiputra and D. Mahdiana, "Analisis Sentimen Dengan Algoritma Support Vector," *Indones. J. Inf. Syst.*, vol. 6, no. April 2022, pp. 1–8, 2023.