

# Analisis Opinion Mining Pada Topik Chatgpt di Aplikasi X Dengan Pendekatan Algoritma SVM Berbasis Lexicon

Darisva Prismala<sup>1</sup>, I Kadek Dwi Nuryana<sup>2</sup>

<sup>1,2</sup> Sistem Informasi, Teknik Informatika, Universitas Negeri Surabaya

[ldarisva.19092@mhs.unesa.ac.id](mailto:ldarisva.19092@mhs.unesa.ac.id)

[dwinuryana@unesa.ac.id](mailto:dwinuryana@unesa.ac.id)

**Abstrak**— Perkembangan teknologi kecerdasan buatan, khususnya dalam bentuk chatbot seperti ChatGPT, telah memberikan dampak signifikan pada berbagai aspek kehidupan, termasuk interaksi di platform media sosial seperti aplikasi X. Analisis sentimen terhadap opini pengguna sangat penting untuk memahami sikap dan reaksi mereka, yang dapat memberikan wawasan berharga bagi pengembang untuk meningkatkan kualitas produk dan layanan. Penelitian ini bertujuan untuk menganalisis opini masyarakat terhadap topik ChatGPT di aplikasi X menggunakan algoritma Support Vector Machine (SVM) berbasis lexicon. Hasil penelitian menunjukkan bahwa algoritma SVM berbasis lexicon dapat secara efektif mengklasifikasikan sentimen *Tweet* dengan tingkat akurasi yang tinggi. Pelabelan data menggunakan kamus VADER Lexicon menunjukkan hasil sebagai berikut: 431 sentimen positif, 1307 sentimen netral, dan 115 sentimen negatif. Sementara itu, hasil klasifikasi menggunakan algoritma SVM menunjukkan 188 sentimen positif, 1264 sentimen netral, dan 1593 sentimen negatif. Akurasi pengujian menggunakan *Cross Validation* menghasilkan akurasi sebesar 78.06%, dengan precision 53.72% untuk sentimen positif, 76.74% untuk sentimen netral, dan 81.98% untuk sentimen negatif. Recall untuk masing-masing kelas adalah 23.43% (positif), 74.22% (netral), dan 99.92% (negatif).

**Kata Kunci**— Sentimen Analisis, ChatGPT, VADER Lexicon, Klasifikasi, Support Vector Machine.

## I. PENDAHULUAN

Perkembangan teknologi saat ini, khususnya kecerdasan buatan semakin mempengaruhi ke setiap aspek kehidupan sehari-hari. Kemajuan teknologi memungkinkan sumber daya pembelajaran menjadi lebih dari sekadar terpusat atau berfokus pada pendidik. Teknologi juga dapat memiliki orientasi yang lebih luas dan menggunakan alat (*as a tools*) untuk mempercepat pencarian sumber belajar yang berbasis luas (*broad-based learning*) [1].

Saat ini ChatGPT merupakan salah satu teknologi yang berkembang. Diterbitkan pada 30 November 2022, tujuannya adalah untuk memudahkan individu menulis dengan cepat dan mengakses informasi. Open AI mendirikan ChatGPT. Sejak saat itu, laboratorium Open AI telah mencapai kemajuan signifikan dalam kemajuan teknologi AI dan telah menciptakan beberapa produk pembelajaran mesin untuk masyarakat umum, seperti DALL-E dan ChatGPT. Open AI mengembangkan

*Generative Pre-Trained Transformer* (GPT), sebuah kecerdasan buatan yang dapat menghasilkan teks respons yang hampir tidak terdeteksi oleh manusia. ChatGPT adalah *chatbot* cerdas yang dapat membalas permintaan teks dari pengguna untuk memecahkan masalah produktivitas, menjawab pertanyaan dasar, dan menyelesaikan pekerjaan [2].

Menurut [3], Model bahasa yang dikenal sebagai ChatGPT, yang diciptakan oleh OpenAI, memiliki kekuatan untuk mengubah sepenuhnya cara manusia berkomunikasi dengan teknologi. Meningkatkan Natural Language Processing (NLP) dan Natural Language Understanding (NLU) di berbagai aplikasi adalah salah satu potensi penggunaan ChatGPT yang paling menarik. Secara khusus, chatbot, asisten virtual, dan antarmuka percakapan lainnya dapat didukung oleh ChatGPT.

Sejak penggunaannya secara luas, pembelajaran mesin telah berdampak pada sejumlah industri, memungkinkannya aktivitas seperti analisis pencitraan medis dan prediksi cuaca resolusi tinggi. Chatbot ini dibuat dengan OpenAI, sehingga dapat merespons pertanyaan seperti manusia. Untuk memanfaatkan ChatGPT, pengguna harus terlebih dahulu mendaftar akun OpenAI gratis. Data yang dihasilkan pengguna dapat digunakan oleh sistem untuk meningkatkan algoritma pelatihannya [4].

Namun, untuk memastikan penggunaan ChatGPT saat ini khususnya di bidang pendidikan tidak membuat penggunanya terlena dan mengikis kemampuan kritisnya. Maka pengguna sendiri harus diberikan pemahaman yang kokoh mengenai nilai-nilai moral, dengan tujuan agar pengguna ChatGPT mampu membatasi diri dalam menggunakan kemampuan AI, karena segala perkerjakan akan mudah selesai hanya dengan bertanya pada ChatGPT [5].

Sangat penting untuk memahami bagaimana sentimen atau perasaan pengguna terhadap program chat generatif ini. Analisis sentimen ini bertujuan untuk mengidentifikasi, serta memahami sudut pandang atau opini, sikap, dan emosi yang terdapat dalam teks ulasan. Pengembang aplikasi dapat memperoleh informasi penting untuk meningkatkan kualitas layanan atau mengatasi masalah terkini dengan menganalisis sentimen [6]. Analisis sentimen juga merupakan cara untuk menyimpulkan opini atau pendapat masyarakat yang diunggah melalui media sosial X ke dalam kategori positif, netral, dan negatif [7]. Dalam penelitian ini menggunakan metode

klasifikasi *Support Vector Machine* (SVM) berbasis *Lexicon Based* dengan kamus *VADER Lexicon*.

*Lexicon-based* digunakan untuk melakukan pelabelan data dengan cara mengelompokkan data ke dalam label kelas positif, netral dan negatif berdasarkan kata dan skor kata yang terdapat pada kamus. *Lexicon Based* digunakan karena memiliki hasil sensitivitas yang tinggi [8]. *VADER* (*Valance Dictionary and Sentiment Reasoner*) digunakan untuk menganalisis data dan menghasilkan kelas dengan polaritas positif, netral, dan negatif serta skor gabungan atau total. *VADER* sentiment *Lexicon* mempunyai 7500 kata yang mencakup sentimen terkait dengan [6].

*Support Vector Machine* (SVM) digunakan karena dianggap sebagai metode yang tepat untuk melakukan klasifikasi teks [8]. Dan menurut penelitian sejenis yang terdahulu terkait SVM menjelaskan bahwa algoritma SVM merupakan metode yang memiliki nilai akurasi yang paling baik.

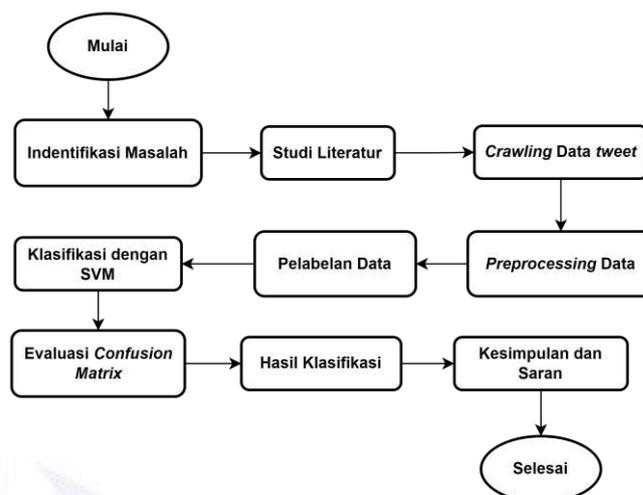
Beberapa penelitian sejenis terdahulu yang berkaitan dengan analisis sentimen, diantaranya: pada penelitian Muhammad Edo Setya Winata (2021) [9] membahas mengenai perbandingan klasifikasi terhadap analisis sentimen twitter mengenai vaksin sinovac, dengan membandingkan SVM, KNN, dan *Naïve Bayes*. Diperoleh hasil akurasi bahwa algoritma SVM mendapatkan hasil lebih besar dibandingkan dengan yang lain dengan hasil akurasi sebesar 87%, KNN sebesar 84% dan *Naïve Bayes* sebesar 83%.

Pada penelitian Rizki Wahyudi & Gilang Kusumawardhana (2021) [10] membahas mengenai analisis sentimen terkait review Aplikasi Grab di Google Play Store Menggunakan SVM. Hasil dari analisis menggunakan SVM menghasilkan akurasi 85,54% dan Hasil Review positif yang paling sering diulas adalah "ovo", sedangkan review negatif yang paling sering diulas adalah "driver".

Selanjutnya pada penelitian Iis Siti Aisah, et.al (2023) [11] membahas mengenai algoritma SVM untuk analisis sentimen ulasan aplikasi Al Qur'an digital. Hasil penelitian adalah sebesar 85.11%, hal ini menunjukkan bahwa algoritma SVM memiliki akurasi yang tinggi serta dapat digunakan untuk memahami respons pengguna terhadap aplikasi Al-Qur'an digital.

Berdasarkan latar belakang di atas, penelitian ini untuk melakukan analisis sentimen pada topik ChatGPT dan menilai ketepatan kategorisasi dari klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) berbasis *Lexicon Based* dengan kamus *VADER Lexicon*.

## II. METODE PENELITIAN



Gbr 1. Flowchart proses pemrograman Python *Tweet* harvest

### A. Identifikasi Masalah

Adapun masalah yang diidentifikasi peneliti, yaitu:

- 1) Bagaimana hasil klasifikasi sentimen pengguna X pada topik ChatGPT dengan algoritma *Support Vector Machine* berbasis *Lexicon Based*?
- 2) Bagaimana hasil akurasi algoritma *Support Vector Machine* berbasis *Lexicon Based*?

### B. Studi Literatur

Studi literatur merupakan serangkaian kegiatan untuk mencari literatur, melokalisasi, dan menganalisis dokumen yang berhubungan dengan masalah yang akan diteliti. Menurut [12], dokumen studi literatur dapat berupa teori-teori yang berkaitan dengan penelitian maupun hasil penelitian terdahulu yang membahas permasalahan yang akan diteliti. Pada tahap ini, peneliti mempelajari dan memahami literatur yang berasal dari jurnal ilmiah, buku dan peneliti terdahulu sebagai referensi peneliti dalam melaksanakan penelitian.

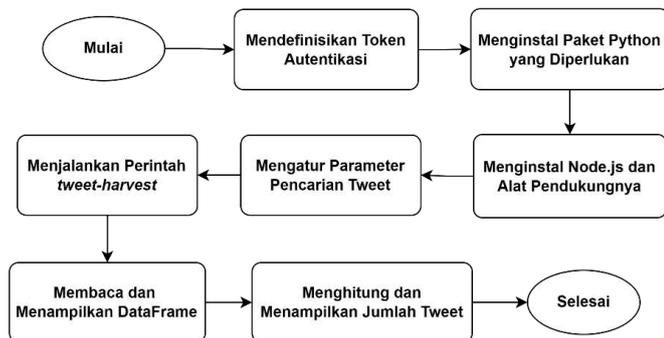
### C. Crawling Data Tweet

Tahapan crawling data *Tweet* dari akun media sosial X dengan bantuan bahasa pemrograman Python yang bernama *Tweet* harvest [13].

Dalam proses crawling memerlukan beberapa hal yang harus diperhatikan seperti: auth token, jumlah data, kata kunci, bahasa, dan tanggal *Tweet* yang ingin diambil. Pertama, untuk mendapatkan izin akses API twitter diperlukan auth token dari sebuah akun X. Auth token berisi kode seperti berikut: "6be547c485818d\*\*\*\*\*c1cffe52e73dcb86". Kedua, jumlah data *Tweet* yang akan diambil adalah 1000 *Tweet*. Ketiga, kata kunci yang akan digunakan untuk mengambil data adalah "chatgpt". Keempat, bahasa yang digunakan untuk crawling adalah bahasa Indonesia, dimana data *Tweet* yang dihasilkan nanti akan berbahasa Indonesia. Terakhir, tanggal data *Tweet* yang akan diambil adalah pada rentang waktu "tahun 2021-01-01 sampai 2023-09-30". Data yang dihasilkan berupa teks yang berisi *Tweet* pengguna terkait pengalaman

menggunakan ChatGPT. Data *Tweet* yang telah didapatkan kemudian akan disimpan dalam file berekstensi CSV.

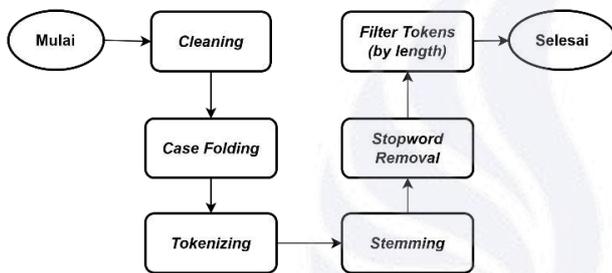
Gbr 2. merupakan flowchart proses pemrograman Python *Tweet harvest* [13].



Gbr 2. Flowchart proses pemrograman Python *Tweet harvest*

#### D. Teknik Analisis Data

##### 1) Preprocessing Data



Gbr 3. Tahapan Preprocessing

Pada tahapan ini data yang berhasil di *crawling* merupakan data mentah yang banyak mengandung noise. Data mentah tersebut harus dilakukan pembersihan melalui tahapan preprocessing, jika tidak melakukan tahapan ini maka akan menurunkan tingkat klasifikasi dan performansi dari sistem yang dihasilkan. Berikut tahapan-tahapan dalam preprocessing data [14]:

1.1 *Cleaning*, Tahap *cleaning* akan membersihkan data pada kolom komentar yang ganda, serta atribut yang tidak penting. Atribut tidak penting berupa Mention (@), URL, hastag/tagar (#), dan karakter simbol (!?%&:;“,,- dan lain-lain) yang ada dalam data.

1.1.1 *Replace Mention*, Proses *replace mention* digunakan untuk membersihkan teks dari Mention atau tag yang dimulai dengan simbol “@” dalam data *Tweet* menggunakan operator “*Replace Mention*” di Rapidminer.

1.1.2 *Replace URL*, Proses *replace URL* digunakan untuk membersihkan teks dari URL yang ada dalam data *Tweet* menggunakan operator “*Replace URL*” di Rapidminer.

1.1.3 *Replace Hastag*, Proses *replace hastag* digunakan untuk membersihkan teks dari hastag yang ada dalam sebuah data menggunakan operator “*Replace Hastag*”.

1.1.4 *Replace symbol*, Proses *replace symbol* digunakan untuk membersihkan teks dari simbol atau karakter

yang ada dalam sebuah data *Tweet* menggunakan operator “*Replace symbol*” di Rapidminer.

Dalam *cleaning* juga dilakukan proses dengan operator “*Filter Examples*” yang akan membersihkan data dari baris-baris yang tidak memiliki nilai atau kosong dalam atribut “*full\_text*”. Dan operator “*Remove Duplicates*” yang akan menghapus duplikat dari data.

1.2 *Case Folding/Transform Cases*, Proses *Case Folding* digunakan untuk mengubah semua huruf menjadi huruf kecil karena jika ada kata yang sama namun huruf tidak seragam maka akan dianggap kata yang berbeda. Proses menyeragamkan huruf menggunakan operator “*Transform cases*” pada Rapidminer dengan parameter (*transform to: “lower case”*).

1.3 *Tokenizing*, Proses *Tokenizing* dilakukan untuk memisahkan opini yang awalnya berupa kalimat menjadi satuan kata serta menghilangkan angka dengan menggunakan operator “*Tokenize*” pada Rapidminer dengan parameter (*mode: non letters*).

1.4 *Stemming*, *Stemming* digunakan untuk merubah setiap ejaan kata yang disingkat dan memiliki imbuhan, sehingga sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Misalnya, kata-kata seperti “*keberadaan*” dan “*berada*” akan distem menjadi bentuk dasar “*ada*”.

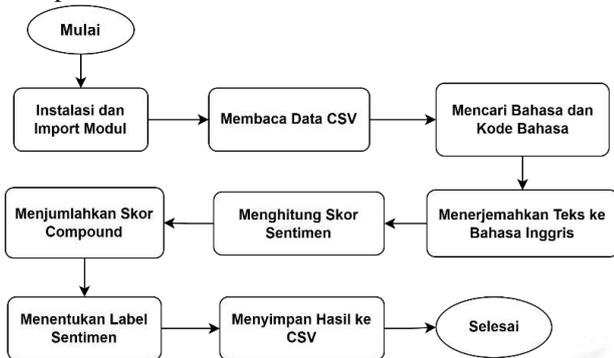
1.5 *Stopwords Removal*, Proses *Stopword Removal* digunakan untuk menghapus kata-kata penghubung seperti kata penghubung “*dan*” “*dengan*” “*ke*” “*di*” dan lainnya dan kata yang tidak mengandung sentimen dari teks. Proses ini menggunakan operator “*Stopword(Dictionary)*” pada Rapidminer. Dalam prosesnya operator “*Stopword(Dictionary)*” menggunakan kamus atau dokumen yang diunduh pada Kaggle dan ditambah dengan memodifikasi kamus sesuai dataset yang diolah.

1.6 *Filter tokens (by length)*, digunakan untuk menghilangkan kata yang terlalu singkat atau terlalu panjang. Dalam proses operator “*Filter Tokens (by Length)*” akan memfilter token berdasarkan panjangnya, yaitu hanya token dengan panjang antara 3 dan 25 karakter yang akan dipertahankan.

##### 2) Pelabelan Data

Proses pelabelan data sentimen ini menggunakan VADER *Lexicon* pada data *Tweet* dimulai dengan menginstal modul-modul yang diperlukan, yaitu (*googletrans*) dan VADER *Lexicon*. Data *Tweet* kemudian dibaca dari file CSV. Kemudian, data *Tweet* akan diterjemahkan dari bahasa Indonesia ke bahasa Inggris. Setelah teks *Tweet* diterjemahkan, VADER *Lexicon* digunakan untuk menghitung skor sentimen. VADER menghasilkan empat jenis skor untuk setiap teks: negatif (*neg*), netral (*neu*), positif (*pos*), dan komposit (*compound*). Kriteria pelabelan ditentukan berdasarkan nilai *compound*. Jika nilai *compound* lebih besar atau sama dengan 0.5, teks dianggap memiliki sentimen positif. Jika nilai *compound* antara -0.5 dan 0.5, teks

dianggap memiliki sentimen netral. Jika nilai *compound* kurang dari atau sama dengan -0.5, teks dianggap memiliki sentimen negatif. Terakhir hasil sentimen akan disimpan ke dalam file CSV baru untuk analisis lebih lanjut. Proses pelabelan data *Tweet* dengan kamus VADER *Lexicon* ditampilkan dalam Gbr 4.



Gbr 4. Proses pelabelan data *Tweet* dengan kamus VADER *Lexicon*

### 3) Klasifikasi Support Vector Machine

Proses klasifikasi SVM dalam penelitian ini menggunakan bantuan *software* Rapidminer. Berikut tahapan-tahapan klasifikasi SVM pada Rapidminer:

- 3.1 *Input Data*, Memasukkan data yang sudah dilabeli ke operator “*Read Excel*” di Rapidminer.
- 3.2 *Set Role*, Selanjutnya menentukan peran atribut sebagai “label” dalam dataset dengan operator “*Set Role*” di Rapidminer.
- 3.3 *Nominal to text*, Data yang tersedia masih berupa nominal dan harus diubah ke dalam bentuk teks.
- 3.4 *TF-IDF*, Selanjutnya data akan melalui tahap pembobotan dengan metode TF-IDF.
- 3.5 *SMOTE-Upsampling*, Jumlah data positif, netral, dan negatif tidak seimbang, sehingga perlu dilakukan penyeimbangan data menggunakan metode *SMOTE Upsampling*.
- 3.6 *Cross Validation*, *Cross Validation* dilakukan dengan bantuan operator “*Cross Validation*” pada. “*Cross Validation*” dilakukan pada fold ke 10. Proses *training* dan *testing* adalah sebagai berikut:

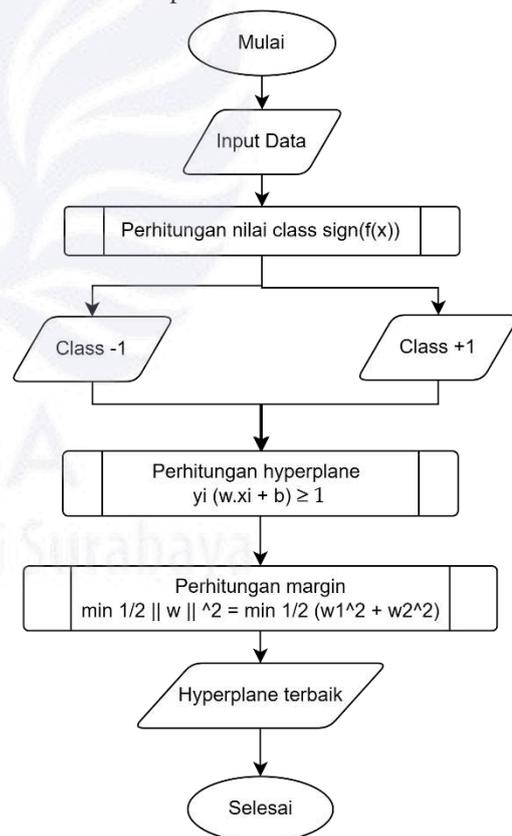
3.6.1 Sub-Proses Pelatihan (*Training*), Menggunakan operator “*Classification by Regression*” yang berisi model “*Support Vector Machine (SVM)*” dengan kernel linear. langkah-langkah klasifikasi algoritma SVM sebagai berikut [15]:

- a. Data yang telah diinputkan akan dicari nilai kelasnya yaitu negatif (kelas -1) dan positif (kelas +1):
  - Kelas -1 (negatif), dimana kelas yang dianggap negatif atau kelas minoritas.
  - Kelas +1 (positif), dimana kelas yang dianggap positif atau kelas mayoritas.
- b. Menentukan fungsi kernel yang digunakan dalam algoritma SVM. Pada penelitian ini

menggunakan fungsi kernel Linier. Fungsi kernel ini menghitung *dot product* dari dua vektor input.

- c. Menentukan parameter yang optimal untuk fungsi kernel yang digunakan. Pada penelitian menggunakan parameter *C (Regularization)*.
- d. Membangun model SVM dengan fungsi kernel. Proses ini melibatkan pemecahan masalah optimasi untuk menemukan *hyperplane* terbaik yang memaksimalkan margin antara kelas positif, netral dan negatif.
- e. Setelah model SVM dilatih diperlukan evaluasi performa model menggunakan data uji. Hal ini bertujuan untuk mengetahui keakuratan algoritma SVM menggunakan confusion matrix.
- f. Menganalisis prediksi klasifikasi dari uji akurasi confusion matrix untuk mengetahui pada bagian mana model melakukan kesalahan dan seberapa baik kinerjanya dalam berbagai situasi.

Berdasarkan langkah-langkah klasifikasi algoritma SVM diatas dapat diGbrkan dalam Flowchart Support Vector Machine pada Gbr 5 dibawah.



Gbr 5. Flowchart algoritma Support Vector Machine

3.6.2 Sub-Proses Evaluasi (*Testing*), Menggunakan operator “*Apply Model*” untuk menerapkan model yang telah dilatih pada data

pengujian. Dan menggunakan operator “Performance” untuk mengevaluasi kinerja model klasifikasi dengan berbagai metrik evaluasi seperti akurasi, error klasifikasi, dan lainnya.

#### 4) Evaluasi Confusion Matrix

Confusion matrix digunakan peneliti untuk melakukan evaluasi performa dari algoritma Support Vector Machine. Tabel confusion matrix akan menampilkan dan membandingkan nilai aktual atau nilai sebenarnya dengan nilai hasil prediksi model sehingga diperoleh metrik evaluasi accuracy (akurasi), precision, dan recall.

Adapun model *confusion matrix* dapat dilihat seperti pada tabel 2.2 dibawah ini (Saputro et al., 2018):

**Tabel 2.2 Model Confusion Matrix**

Prediction	Actual		
	Positive	Negative	Neutral
Positive	TP	FP1	FP2
Negative	FNg1	TNg	FNg2
Neutral	FNt1	FNt2	TNt

Keterangan:

- True Positive (TP) : Jumlah data yang bernilai Positif dan diprediksi benar sebagai Positif.
- False Positive (FP) : Jumlah data yang bernilai Negatif/Netral tetapi diprediksi sebagai Positif.
- True Netral (TNt) : Jumlah data yang bernilai Netral dan diprediksi benar sebagai Netral.
- False Netral (FNt) : Jumlah data yang bernilai Negatif/Positif tetapi diprediksi sebagai Netral.
- False Negative (FNg) : Jumlah data yang bernilai Positif/Netral tetapi diprediksi sebagai Negatif.
- True Negative (TNg) : Jumlah data yang bernilai Negatif dan diprediksi benar sebagai Negatif.

#### 4.1 Accuracy

Accuracy adalah rasio antara jumlah prediksi yang benar dengan total jumlah prediksi (Mulyawan, 2024).

$$Accuracy = \frac{TP+TNg+TNt}{TP+FP+TNt+FNt+FNg1+TNt} \times 100\% \quad (2.6)$$

Nilai akurasi dihitung dengan membagi jumlah prediksi positif untuk data positif dan prediksi negatif untuk data negatif dengan jumlah total data dalam dataset.

#### 4.2 Precision

Precision adalah peluang kasus yang diprediksi positif dan keadaan aktualnya termasuk kasus kategori positif (Afifah, 2023).

$$Precision\ Positive = \frac{TP}{TP+FP1+FP2} \times 100\% \quad (2.7)$$

Nilai *precision positive* dihitung dengan membagi nilai prediksi positif untuk data positif dengan jumlah total nilai prediksi positif untuk data positif, prediksi negatif untuk data positif dan prediksi netral untuk data positif.

#### 4.3 Recall

Recall adalah peluang kasus dengan kategori positif dan secara tepat diprediksi positif .

$$Recall\ Positive = \frac{TP}{TP+FNg+FNt1} \times 100\% \quad (2.10)$$

Nilai *precision neutral* dihitung dengan membagi nilai prediksi netral untuk data netral dengan jumlah total nilai prediksi positif untuk data netral, prediksi negatif untuk

### III. PEMBAHASAN

#### A. Pengambilan Data

Dari proses crawling data dari aplikasi X menggunakan bantuan pemrograman python dengan kata kunci “chatgpt” yang telah dilakukan, diperoleh data *Tweet* sebanyak 1960 data. Data *Tweet* diambil berdasarkan rentang waktu pada 01 Januari 2021 hingga 30 September 2023.

#### B. Preprocessing Data

*Preprocessing data* adalah tahap yang penting untuk mengubah data yang awalnya tidak terstruktur menjadi lebih terstruktur. Proses ini menggunakan tools Rapidminer dengan beberapa tahapan, yaitu *cleaning data*, *case folding*, *tokenizing*, *stemming*, *stopword*, dan *filter tokens (by length)*. Berikut adalah tahapan-tahapan *preprocessing data*

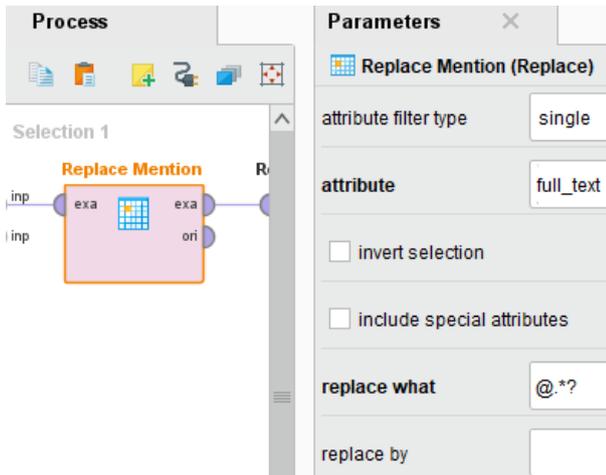
##### 1) Cleaning

*Cleaning* digunakan untuk membersihkan data pada kolom komentar yang ganda, serta atribut yang tidak penting. Atribut tidak penting berupa tanda baca, URL, *Mention* (@), *hashtag/tagar* (#), dan simbol-simbol lainnya. Proses *Cleaning* data menggunakan operator “*Replace*” pada Rapidminer.

TABLE I  
SAMPEL DATA TWEET PENGGUNA CHATGPT

No.	Teks
1.	@Dennysiregar7 Setuju, Industri yg bertumpu pd kreativitas sumber daya manusia adalah industri yg bertumbuh Pekerjaan yg tdk kreatif akan digantikan oleh robot AI seperti ChatGPT belum bisa mengalahkan manusia dlm hal kreativitas bikin aplikasi, misalnya

##### 1.1 Replace Mention (@username)

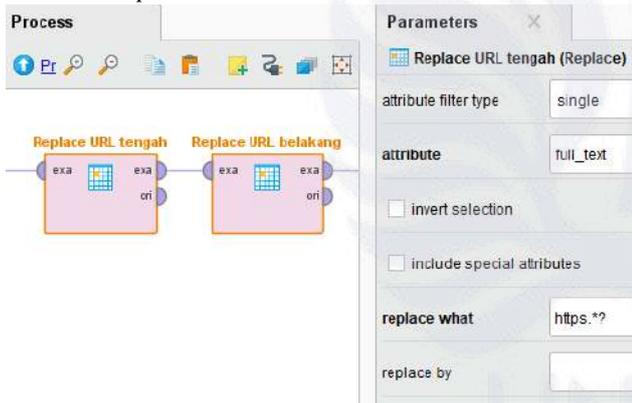


Gbr 6. Proses Replace @ data Tweet

TABEL II  
 HASIL REPLACE @ DATA TWEET

No.	Teks
1.	Setuju, Industri yg bertumpu pd kreativitas sumber daya manusia adalah industri yg bertumbuh Pekerjaan yg tdk kreatif akan digantikan oleh robot AI seperti ChatGPT belum bisa mengalahkan manusia dlm hal kreativitas bikin aplikasi, misalnya

### 1.2 Replace URL

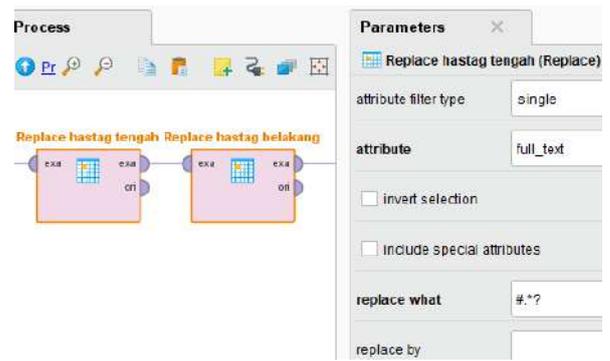


Gbr 7. Proses Replace URL data Tweet

TABEL III  
 HASIL REPLACE URL DATA TWEET

No.	Teks
1.	Setuju, Industri yg bertumpu pd kreativitas sumber daya manusia adalah industri yg bertumbuh Pekerjaan yg tdk kreatif akan digantikan oleh robot AI seperti ChatGPT belum bisa mengalahkan manusia dlm hal kreativitas bikin aplikasi, misalnya

### 1.3 Replace hastag (#)

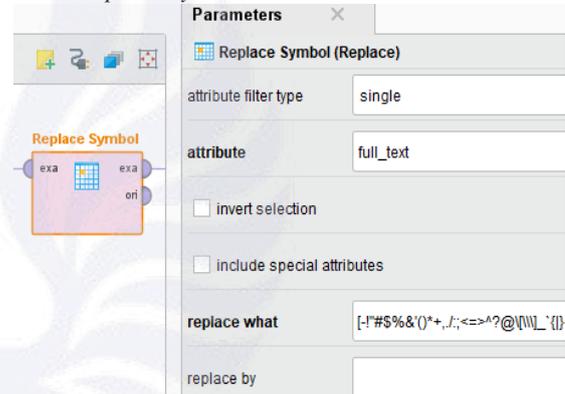


Gbr 8. Proses Replace hastag # data Tweet

TABEL IV  
 HASIL REPLACE HASTAG # DATA TWEET

No.	Teks
1.	Setuju, Industri yg bertumpu pd kreativitas sumber daya manusia adalah industri yg bertumbuh Pekerjaan yg tdk kreatif akan digantikan oleh robot AI seperti ChatGPT belum bisa mengalahkan manusia dlm hal kreativitas bikin aplikasi, misalnya

### 1.4 Replace symbol

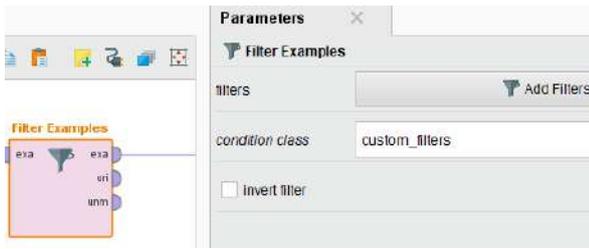


TABEL V  
 HASIL REPLACE HASTAG # DATA TWEET

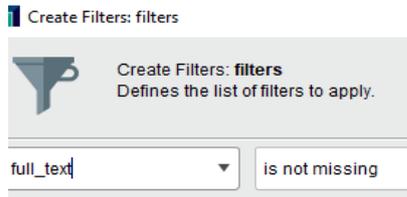
No.	Teks
1.	Setuju Industri yg bertumpu pd kreativitas sumber daya manusia adalah industri yg bertumbuh Pekerjaan yg tdk kreatif akan digantikan oleh robot AI seperti ChatGPT belum bisa mengalahkan manusia dlm hal kreativitas bikin aplikasi misalnya

Dalam *cleaning* juga dilakukan tahapan menghapus data yang kosong dan menghapus data duplikat dengan operator “*Filter Example*” dan “*Remove Duplicate*” di Rapidminer. Hasil akhir dari proses *cleaning* dan setelah data duplikat dihilangkan, sehingga data saat ini yang tersisa sebanyak 1932 data.

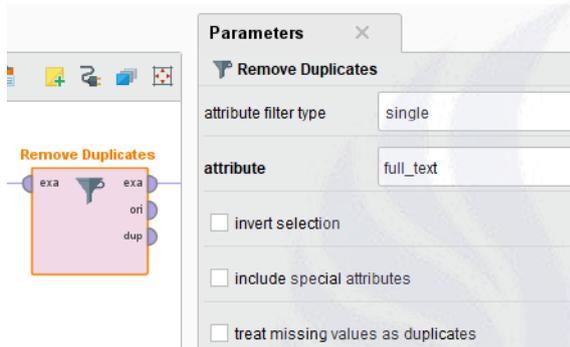
Pada Gbr (10, 11, dan 12) merupakan proses dari operator “*Filter Example*” dan “*Remove Duplicate*”:



Gbr 10. Proses filter examples



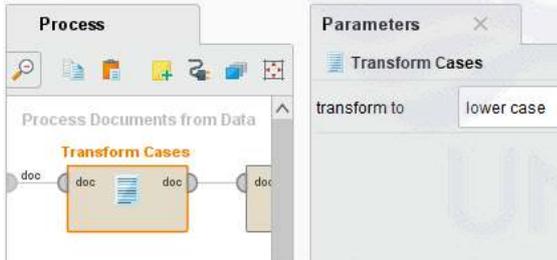
Gbr 11. Proses custom filter examples



Gbr 12. Proses remove duplicates

2) *Case Folding/Transform Case*

Proses *Case Folding* digunakan untuk mengubah semua huruf menjadi huruf kecil.



Gbr 13. Proses Case Folding/ transform case data Tweet

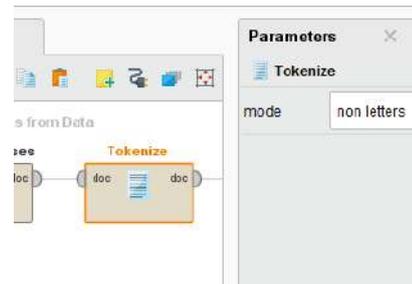
TABLE VI

HASIL CASE FOLDING/ TRANSFORM CASE DATA TWEET

No.	Teks
1.	setuju industri yg bertumpu pd kreativitas sumber daya manusia adalah industri yg bertumbuh pekerjaan yg tdk kreatif akan digantikan oleh robot ai seperti chatgpt belum bisa mengalahkan manusia dlm hal kreativitas bikin aplikasi misalnya

3) *Tokenizing*

Proses *Tokenizing* dilakukan untuk memisahkan opini menjadi satuan kata serta menghilangkan angka.



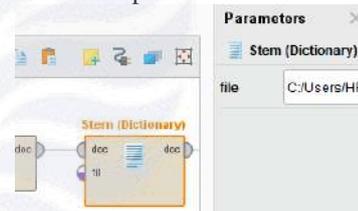
Gbr 14. Proses Tokenizing data Tweet dengan Rapidminer

TABLE VII  
HASIL TOKENIZING DATA TWEET

No.	Teks	Token
1.	setuju industri yg bertumpu pd kreativitas sumber daya manusia adalah industri yg bertumbuh pekerjaan yg tdk kreatif akan digantikan oleh robot ai seperti chatgpt belum bisa mengalahkan manusia dlm hal kreativitas bikin aplikasi misalnya	'setuju' 'industri' 'yg' 'bertumpu' 'pd' 'kreativitas' 'sumber' 'daya' 'manusia' 'adalah' 'industri' 'yg' 'bertumbuh' 'pekerjaan' 'yg' 'tdk' 'kreatif' 'akan' 'digantikan' 'oleh' 'robot' 'ai' 'seperti' 'chatgpt' 'belum' 'bisa' 'mengalahkan' 'manusia' 'dlm' 'hal' 'kreativitas' 'bikin' 'aplikasi' 'misalnya'

4) *Stemming*

Pada tahap *Stemming* mengubah kata-kata menjadi bentuk dasar atau kata dasar setiap kata.



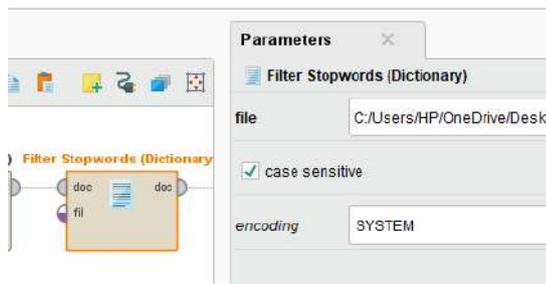
Gbr 15. Proses Stemming data Tweet

TABLE VIII  
HASIL STEMING DATA TWEET

No.	Tokenize	Stemming
1.	'setuju' 'industri' 'yg' 'bertumpu' 'pd' 'kreativitas' 'sumber' 'daya' 'manusia' 'adalah' 'industri' 'yg' 'bertumbuh' 'pekerjaan' 'yg' 'tdk' 'kreatif' 'akan' 'digantikan' 'oleh' 'robot' 'ai' 'seperti' 'chatgpt' 'belum' 'bisa' 'mengalahkan' 'manusia' 'dlm' 'hal' 'kreativitas' 'bikin' 'aplikasi' 'misalnya'	'setuju' 'industri' 'yang' 'tumpu' 'pada' 'kreativitas' 'sumber' 'daya' 'manusia' 'adalah' 'industri' 'yang' 'tumbuh' 'kerja' 'yang' 'tidak' 'kreatif' 'akan' 'ganti' 'oleh' 'robot' 'ai' 'seperti' 'chatgpt' 'belum' 'bisa' 'kalah' 'manusia' 'dalam' 'hal' 'kreativitas' 'bikin' 'aplikasi' 'misal'

5) *Stopword Removal*

Proses *Stopword Removal* digunakan untuk menghapus kata-kata penghubung atau kata yang tidak mengandung sentimen dari teks.



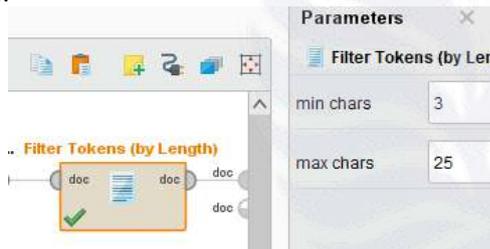
Gbr 26. Proses Stopword Removal data Tweet

TABLE IX  
 HASIL STOPWORD REMOVAL DATA TWEET

No.	Stemming	Stopword
1.	'setuju' 'industri' 'yang' 'tumpu' 'pada' 'keaktivitas' 'sumber' 'daya' 'manusia' 'adalah' 'industri' 'yang' 'tumbuh' 'kerja' 'yang' 'tidak' 'kreatif' 'akan' 'ganti' 'oleh' 'robot' 'ai' 'seperti' 'chatgpt' 'belum' 'bisa' 'kalah' 'manusia' 'dalam' 'hal' 'keaktivitas' 'bikin' 'aplikasi' 'misal'	'setuju' 'industri' 'tumpu' 'keaktivitas' 'sumber' 'daya' 'manusia' 'industri' 'tumbuh' 'kerja' 'kreatif' 'ganti' 'robot' 'ai' 'chatgpt' 'kalah' 'manusia' 'keaktivitas' 'bikin' 'aplikasi'

6) *Filter tokens (by length)*

*Filter tokens (by length)* bertujuan untuk menghilangkan kata yang terlalu singkat atau terlalu panjang antara 3 dan 25 karakter.



Gbr 17. Proses Filter tokens (by length) data Tweet

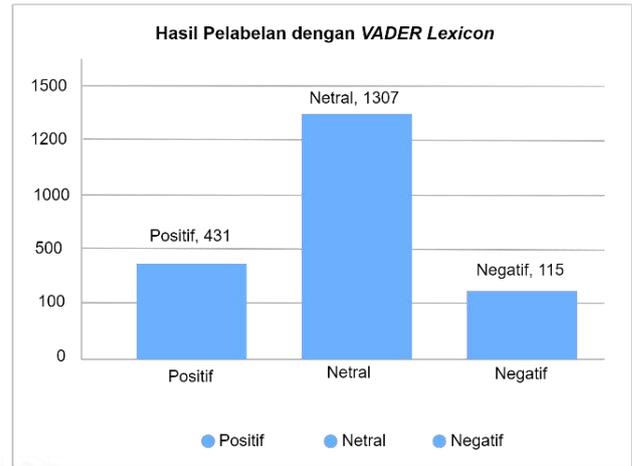
TABLE X  
 HASIL FILTER TOKENS (BY LENGTH) DATA TWEET

No.	Stopword	Filter Token
1.	'setuju' 'industri' 'tumpu' 'keaktivitas' 'sumber' 'daya' 'manusia' 'industri' 'tumbuh' 'kerja' 'kreatif' 'ganti' 'robot' 'ai' 'chatgpt' 'kalah' 'manusia' 'keaktivitas' 'bikin' 'aplikasi'	'setuju' 'industri' 'tumpu' 'keaktivitas' 'sumber' 'daya' 'manusia' 'industri' 'tumbuh' 'kerja' 'kreatif' 'ganti' 'robot' 'chatgpt' 'kalah' 'manusia' 'keaktivitas' 'bikin' 'aplikasi'

Hasil dari *preprocessing* akan disimpan dalam bentuk file berekstensi CSV, dimana nantinya akan lanjut ke tahap pelabelan data menggunakan metode *VADER Lexicon* dengan bahasa pemrograman *python*.

C. *Pelabelan Data*

Pada penelitian ini pelabelan pada tiap data *Tweet* dilakukan dengan metode pelabelan data menggunakan metode kamus *VADER Lexicon* dengan bantuan bahasa pemrograman *Python*. Data yang akan melewati proses pelabelan berasal dari data bersih yang sudah melewati tahapan *preprocessing*.



Gbr 18. Hasil pelabelan menggunakan VADER Lexicon

Dari hasil pelabelan sentimen ChatGPT menggunakan *VADER Lexicon* terdapat pada Gbr 4.42 di atas, dimana diperoleh hasil sebanyak 431 sentimen positif, 1307 sentimen netral, dan 115 sentimen negatif dengan total data 1853 data.

D. *Klasifikasi Support Vector Machine*

1) *Input Data*

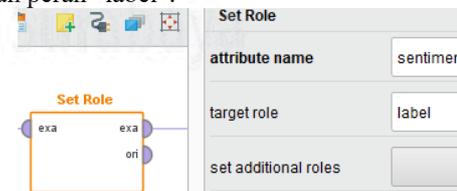
Memasukkan data yang sudah dilabeli menggunakan metode *VADER Lexicon* pada proses sebelumnya menggunakan operator "*Read Excel*" di Rapidminer.



Gbr 19. Proses memasukkan data yang sudah di labeli

2) *Set Role*

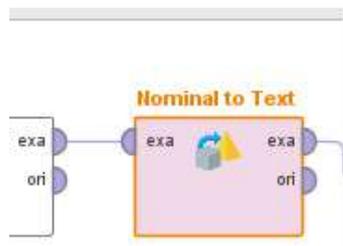
Operator "*Set Role*" digunakan untuk menentukan peran atribut dalam dataset. Dalam penelitian ini atribut yang akan diberikan peran adalah atribut "sentimen" dengan peran "label".



Gbr 20. Proses Set Role attribute

3) *Nominal to text*

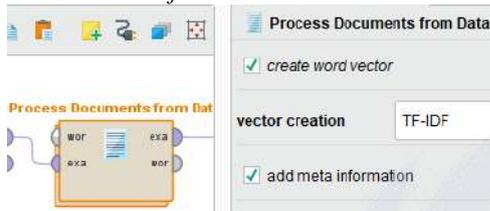
*Nominal to text* adalah operator yang digunakan untuk mengubah atribut bertipe nominal menjadi teks.



Gbr 21. Proses ubah data nominal ke teks

4) TF-IDF

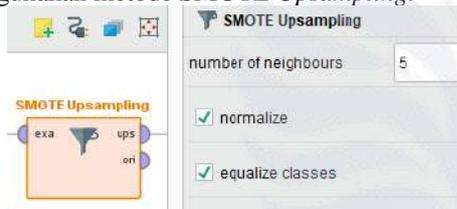
Selanjutnya data dilakukan pembobotan dengan metode TF-IDF. Proses TF-IDF terjadi dalam operator “Process Document from Data”.



Gbr 22. Proses TF-IDF

5) SMOTE-Upsampling

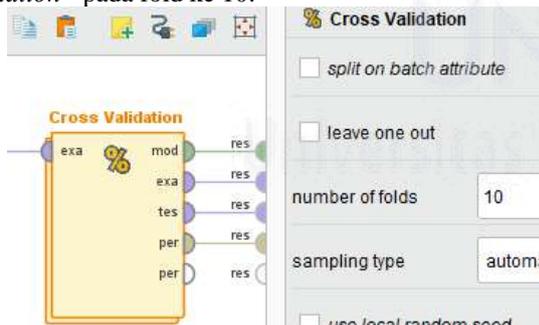
Jumlah data positif, netral, dan negatif tidak seimbang, sehingga perlu dilakukan penyeimbangan data menggunakan metode SMOTE Upsampling.



Gbr 23. Proses SMOTE-Upsampling

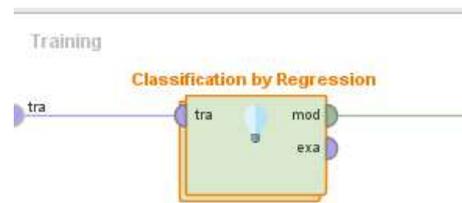
6) Cross Validation

Cross Validation dilakukan dengan operator “Cross Validation” pada fold ke 10.



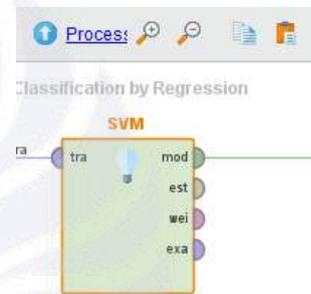
Gbr 24. Proses Cross Validation

Adapun proses didalam operator “Cross Validation”, sebagai berikut:



Gbr 25. Proses didalam Cross Validation (classification by regression)

Gbr 25. merupakan proses didalam *Cross Validation* yaitu (*classification by regression*) digunakan untuk melakukan klasifikasi pada data dengan menggunakan model regresi. Operator ini akan melakukan transformasi target dalam klasifikasi biner, label kelas dikonversi menjadi nilai numerik (kelas positif 0, netral 1, dan negatif 2). Setelah transformasi, model regresi linier akan diterapkan. Selanjutnya proses prediksi dan konversi balik, dimana model akan membuat prediksi dalam bentuk numerik, yang kemudian dikonversi kembali menjadi label kelas.

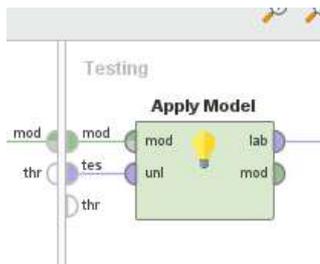


Gbr 26. Proses didalam classification by regression (SVM)

Selanjutnya yaitu melakukan pelatihan klasifikasi model menggunakan SVM dengan kernel linear menggunakan operator “SVM” di Rapidminer pada Gbr 26. Berikut parameter-parameter algoritma SVM:

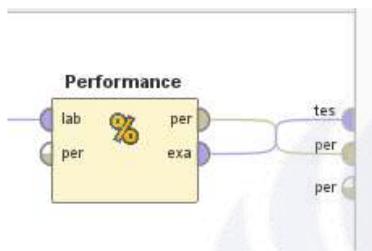
TABLE XI  
 PARAMETER ALGORITMA SUPPORT VECTOR MACHINE

Parameter	
Kernel type	Dot
Kernel cache	200
C	0.0
Convergence epsilon	0.001
Max iterations	100000
L pos	1.0
L neg	1.0
Epsilon	0.0
Epsilon plus	0.0
Epsilon minus	0.0



Gbr 27. Proses didalam Cross Validation (Apply Model)

Gbr 27. merupakan proses *Cross Validation (Apply Model)* di Rapidminer dengan operator “*Apply Model*”. Operator ini digunakan untuk menerapkan model yang telah dilatih pada data baru untuk membuat prediksi atau mengklasifikasikan data tersebut.



Gbr 28. Proses didalam Cross Validation (Performance)

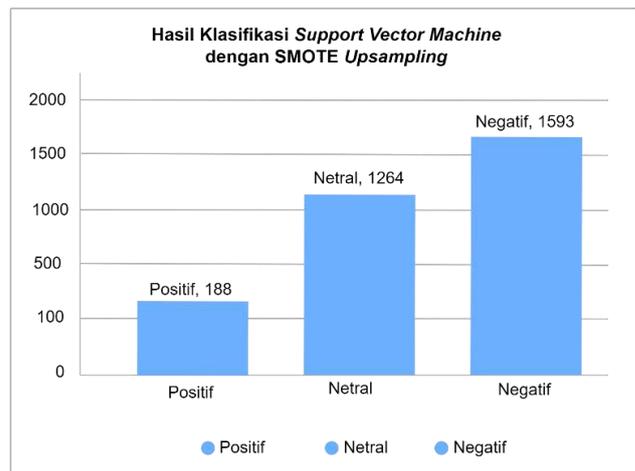
Gbr 28. merupakan proses *Cross Validation (Performance)* yang mengevaluasi kinerja model klasifikasi. Operator “*Performance*” akan menghitung akurasi dari model klasifikasi pada proses sebelumnya. Hasil akhir dari proses ini yaitu akurasi model yang akan ditampilkan dalam bentuk angka atau dalam bentuk grafik.

E. Evaluasi Confusion Matrix dan Hasil Klasifikasi

1) Hasil klasifikasi Support Vector Machine

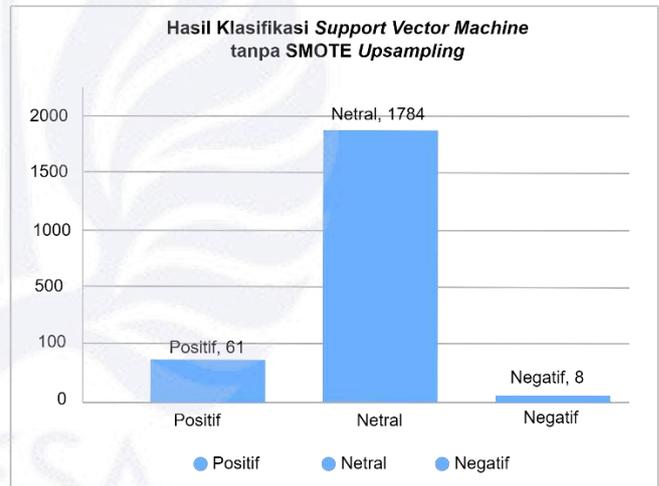
Setelah data selesai diproses dalam tahap *preprocessing* dan pelabelan sentimen dengan VADER Lexicon, serta telah melalui tahap pengklasifikasian dengan algoritma SVM. Selanjutnya adalah hasil klasifikasi, dimana hasil ini berupa klasifikasi *Support Vector Machine* dengan SMOTE *Upsampling* dan tanpa SMOTE.

Hasil klasifikasi SVM dengan SMOTE adalah seperti berikut: untuk sentimen positif sebesar 188, sentimen netral sebesar 1264, dan sentimen negatif sebesar 1593. Grafik hasil klasifikasi sentimen pada ChatGPT menggunakan algoritma *Support Vector Machine* dengan SMOTE dapat dilihat pada Gbr 4.54.



Gbr 29. Grafik hasil klasifikasi Support Vector Machine dengan SMOTE

Hasil klasifikasi SVM dengan SMOTE adalah seperti berikut: untuk sentimen positif sebesar 61, sentimen netral sebesar 1784, dan sentimen negatif sebesar 8. Grafik hasil klasifikasi sentimen pada ChatGPT menggunakan algoritma *Support Vector Machine* tanpa SMOTE dapat dilihat pada Gbr 30.



Gbr 30. Grafik hasil klasifikasi Support Vector Machine tanpa SMOTE

2) Evaluasi Confusion Matrix

Pada tahap ini menampilkan hasil akurasi dari pengujian menggunakan *Cross Validation* dalam bentuk *confusion matrix*.

2.1 Hasil akurasi dengan SMOTE

Evaluasi hasil akurasi dengan SMOTE dalam bentuk *confusion matrix* dapat dilihat dalam tabel XII.

TABLE XII  
 CONFUSION MATRIX SVM DENGAN SMOTE

Accuracy: 78,06% +/- 1,59% (micro average: 78,06%)

Prediction	Actual			Class Precision
	Neutral	Positive	Negative	
Neutral	970	293	1	76,74%
Positive	87	101	0	53,72%
Negative	250	37	1306	81,98%

<b>Class Recall</b>	74,22%	23,43%	99,92%	
---------------------	--------	--------	--------	--

Tabel XII menunjukkan hasil pengujian akurasi klasifikasi *Support Vector Machine* dengan *Cross Validation* terhadap data sentimen ChatGPT dengan *SMOTE Upsampling*, dimana memperoleh nilai akurasi sebesar 78,06%.

Adapun perhitungan *precision* pada masing-masing kelas, sebagai berikut:

$$\begin{aligned} \text{Precision Neutral} &= \frac{TP}{FNt1+FNt+TNt} \times 100\% \quad (4.1) \\ &= \frac{970}{970 + 293 + 1} \times 100\% \\ &= \frac{970}{1264} \times 100\% = 76,74\% \end{aligned}$$

$$\begin{aligned} \text{Precision Positive} &= \frac{TP}{TP+FP1+FP2} \times 100\% \quad (4.2) \\ &= \frac{101}{101 + 0 + 87} \times 100\% \\ &= \frac{101}{188} \times 100\% = 53,72\% \end{aligned}$$

$$\begin{aligned} \text{Precision Negative} &= \frac{TNg}{FNg1+TNg+FNg2} \times 100\% \quad (4.3) \\ &= \frac{1306}{1306 + 37 + 250} \times 100\% \\ &= \frac{1306}{1593} \times 100\% = 81,98\% \end{aligned}$$

Dari perhitungan akurasi *precision* dihasilkan sebagai berikut untuk kelas positif sebesar 53,72%, netral sebesar 76,74%, negatif sebesar 81,98%. Kelas negatif mendapatkan nilai paling tinggi, yang berarti model klasifikasi jarang salah dalam memprediksi kelas negatif yang benar-benar negatif.

Adapun perhitungan *recall* pada masing-masing kelas, sebagai berikut:

$$\begin{aligned} \text{Recall Neutral} &= \frac{TNt}{FP2+FNg2+TNt} \times 100\% \quad (4.4) \\ &= \frac{970}{87 + 250 + 970} \times 100\% \\ &= \frac{970}{1307} \times 100\% = 74,22\% \end{aligned}$$

$$\begin{aligned} \text{Recall Positive} &= \frac{TP}{TP+FNg+FNt1} \times 100\% \quad (4.5) \\ &= \frac{101}{101 + 37 + 293} \times 100\% \\ &= \frac{101}{431} \times 100\% = 23,43\% \end{aligned}$$

$$\begin{aligned} \text{Recall Negative} &= \frac{TNg}{FP1+TNg+FNt2} \times 100\% \quad (4.6) \\ &= \frac{1306}{0 + 1306 + 1} \times 100\% \\ &= \frac{1306}{1307} \times 100\% = 99,92\% \end{aligned}$$

Dari perhitungan akurasi *recall* dihasilkan sebagai berikut: untuk kelas positif sebesar 23,43%, netral sebesar

74,22%, negatif sebesar 99,92%. Kelas negatif mendapatkan nilai paling tinggi, yang berarti model klasifikasi jarang salah dalam menemukan semua kelas negatif.

## 2.2 Hasil akurasi tanpa SMOTE

Evaluasi hasil akurasi tanpa SMOTE dalam bentuk *confusion matrix* dapat dilihat dalam tabel XIII.

TABLE XIII  
CONFUSION MATRIX SVM TANPA SMOTE  
Accuracy: 70,16% +/- 1,34% (micro average: 70,16%)

Prediction	Actual			Class Precision
	Neutral	Positive	Negative	
Neutral	1270	403	111	71,19%
Positive	31	28	2	45,90%
Negative	6	0	2	25,00%
<b>Class Recall</b>	97,17%	6,50%	1,74%	

Tabel XIII menunjukkan hasil pengujian akurasi klasifikasi *Support Vector Machine* dengan *Cross Validation* terhadap data sentimen ChatGPT tanpa SMOTE *Upsampling*, dimana memperoleh nilai akurasi sebesar 70,16%.

Adapun perhitungan *precision* pada masing-masing kelas, sebagai berikut:

$$\begin{aligned} \text{Precision Neutral} &= \frac{TP}{FNt1+FNt+TNt} \times 100\% \quad (4.7) \\ &= \frac{1270}{1270 + 403 + 111} \times 100\% \\ &= \frac{1270}{1.784} \times 100\% = 71,19\% \end{aligned}$$

$$\begin{aligned} \text{Precision Positive} &= \frac{TP}{TP+FP1+FP2} \times 100\% \quad (4.8) \\ &= \frac{28}{28 + 2 + 31} \times 100\% \\ &= \frac{28}{61} \times 100\% = 45,90\% \end{aligned}$$

$$\begin{aligned} \text{Precision Negative} &= \frac{TNg}{FNg1+TNg+FNg2} \times 100\% \quad (4.9) \\ &= \frac{2}{2 + 0 + 6} \times 100\% \\ &= \frac{2}{8} \times 100\% = 25,00\% \end{aligned}$$

Dari perhitungan akurasi *precision* dihasilkan sebagai berikut untuk kelas positif sebesar 45,90%, netral sebesar 71,19%, negatif sebesar 25,00%. Kelas netral mendapatkan nilai paling tinggi, yang berarti model klasifikasi jarang salah dalam memprediksi kelas netral yang benar-benar netral.

Adapun perhitungan *recall* pada masing-masing kelas, sebagai berikut:

$$\text{Recall Neutral} = \frac{TNt}{FP2+FNg2+TNt} \times 100\% \quad (4.10)$$





REFERENSI

- [1] Faiz, A., & Kurniawaty, I. (2023). Tantangan Penggunaan ChatGPT dalam Pendidikan Ditinjau dari Sudut Pandang Moral. *Jurnal Ilmu Pendidikan*, 5, 456–463. <https://doi.org/10.31004/edukatif.v5i1.4779>
- [2] Jafar Maulana, M., & Darmawan, C. (2023). PENGGUNAAN CHATGPT DALAM PENDIDIKAN BERDASARKAN PERSPEKTIF ETIKA AKADEMIK. *Bhineka Tunggal Ika: Kajian Teori Dan Praktik Pendidikan PKN*, 10(01), 58–66.
- [3] Aljanabi, M., Ghazi, M., Ali, A. H., & Abed, S. A. (2023). ChatGpt: Open Possibilities. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 62–64. <https://doi.org/10.52866/ijcsm.2023.01.01.0018>
- [4] Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2(4), 100089. <https://doi.org/10.1016/j.tbench.2023>.
- [5] Supriyadi, E. (2022). EKSPLORASI PENGGUNAAN CHATGPT DALAM PENULISAN ARTIKEL PENDIDIKAN MATEMATIKA. *Papanda Journal of Mathematics and Sciences Research*, 1(2).
- [6] Fahriza, M. N., & Riza, N. (2023). ANALISIS SENTIMEN PADA ULASAN APLIKASI CHAT GENERATIVE PRE-TRAINED TRANSFORMER GPT MENGGUNAKAN METODE KLASIFIKASI K-NEAREST NEIGHBOR(KNN). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 2).
- [7] Adi Putra, S., & Wijaya, A. (2023). Analisis Sentimen Artificial Intelligence (AI) Pada Media Sosial Twitter Menggunakan Metode Lexicon Based. *Jurnal Sistem & Teknologi Informasi Komunikasi*, 7(1).
- [8] Syafira, F. (2023). ANALISIS SENTIMEN DAMPAK PERKEMBANGAN ARTIFICIAL INTELLIGENCE (AI) PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN LEXICON BASED. *Fakultas Sains Dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta*.
- [9] Winata, M. E. S. (2021). PERBANDINGAN KLASIFIKASI TERHADAP ANALISIS SENTIMEN TWITTER MENGENAI VAKSIN SINOVAC. *Doctoral Dissertation, Universitas Mercu Buana Jakarta*.
- [10] Wahyudi, R., & Kusumawardhana, G. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *JURNAL INFORMATIKA*, 8(2), 200–207. <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [11] Aisah, I. S., Irawan, B., & Suprpti, T. (2023). ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS SENTIMEN ULASAN APLIKASI AL QUR'AN DIGITAL. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 6).
- [12] Sangadji, & Sopiah. (2010). *Metode Penelitian (Pendekatan Praktis Dalam Penelitian)*.
- [13] Satria, H. (2023, October 17). Cara mendapatkan data twitter yang besar (lebih dari 1500). <https://helimisatria.com/blog/cara-mendapatkan-data-twitter-yang-besar>
- [14] Habibah, N. (2023). ANALISIS SENTIMEN MENGENAI PENGGUNAAN E-WALLET PADA GOOGLE PLAY MENGGUNAKAN LEXICON BASED DAN K-NEAREST NEIGHBOR. *FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU PEKANBARU*, 10(01). <https://ejournal.stmik-budidarma.ac.id/index.php/jurikom/article/view/5429>
- [15] Putri, A. I. (2023). ANALISIS PREDIKSI PENERIMAAN PENGGUNA FITUR SHOPEEFOOD MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *Journal of Emerging Information Systems and Business Intelligence*, 04(03).