

**Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Polusi  
Di Indonesia Menggunakan InSet Lexicon**



**Oleh :**

**Rafi Aditya Mahendra**

**(20081010208)**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"**

**JAWA TIMUR**

**2023**

# **Bab 1**

## **Pendahuluan**

### **1.1 Latar Belakang**

Polusi udara di Indonesia merupakan tantangan serius yang dapat berdampak negatif pada kesehatan masyarakat dan keberlanjutan lingkungan. Salah satu aspek yang perlu diperhatikan dalam memahami kompleksitas isu ini adalah persepsi dan tanggapan masyarakat. Dalam era digital, sosial media, khususnya Twitter, telah menjadi wadah bagi masyarakat untuk menyuarakan pandangan mereka terkait isu-isu lingkungan, termasuk polusi udara.

Polusi udara, yang melibatkan pelepasan zat-zat berbahaya ke atmosfer, menjadi fokus perhatian global dan nasional. Dampaknya tidak hanya dirasakan secara fisik melalui peningkatan penyakit pernapasan dan kesehatan, tetapi juga melalui pengaruhnya terhadap kualitas hidup dan keberlanjutan lingkungan. Sosial media memiliki peran penting dalam membentuk opini dan memahami pandangan masyarakat. Twitter, sebagai platform berita cepat dan forum diskusi terbuka, mencerminkan berbagai pandangan dan sentimen terkait isu-isu seperti polusi udara.

Dalam konteks ini, persoalan praktis yang muncul adalah bagaimana analisis sentimen pengguna Twitter dapat memberikan wawasan mendalam tentang bagaimana masyarakat merespons dan merasakan dampak polusi udara di Indonesia. Informasi ini dapat digunakan sebagai dasar bagi pengambil kebijakan untuk merancang langkah-langkah mitigasi yang lebih efektif dan meningkatkan kesadaran masyarakat.

Penelitian ini menggunakan dataset yang dikumpulkan dari Twitter melalui proses "tweet harvest". Data yang diperoleh melalui metode ini mencakup sejumlah besar tweet yang mencantumkan kata-kata kunci terkait polusi udara di Indonesia. Penggunaan dataset ini diharapkan dapat memberikan representasi yang lebih luas terhadap opini dan sentimen masyarakat di platform Twitter.

Permasalahan yang ingin diatasi melalui penelitian ini adalah kurangnya pemahaman mendalam terhadap sentimen masyarakat terkait polusi udara di Indonesia. Dengan menganalisis data dari Twitter, diharapkan dapat teridentifikasi pola-pola umum dalam tanggapan dan pandangan masyarakat terkait isu ini.

Penelitian terkait analisis sentiment dengan menggunakan InSet Lexicon pernah dilakukan oleh (Desi Musfiroh, 2021) yang membahas pendekatan berbasis leksikal untuk memperoleh hasil klasifikasi sentimen dari data tweet pada topik "kuliah daring" menggunakan InSet Lexicon (Indonesian Sentiment Lexicon) yang berasal dari penelitian sebelumnya. Penelitian ini difokuskan untuk melakukan analisis sentimen/opini masyarakat khususnya para pengguna twitter

terhadap topik “kuliah daring” melalui data yang diambil dari twitter dengan pengolahan data menggunakan bahasa pemrograman Phyton.

Pemilihan topik ini didasari oleh kebutuhan akan pemahaman yang lebih baik tentang bagaimana masyarakat Indonesia merespons polusi udara, serta keinginan untuk memberikan kontribusi konkret terhadap pengambilan kebijakan dan kesadaran masyarakat. Analisis sentimen di Twitter dianggap sebagai metode yang relevan dan responsif untuk mencapai tujuan tersebut.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang di atas, rumusan masalah penelitian ini dapat dirumuskan sebagai berikut :

- Bagaimana sentimen pengguna Twitter terkait polusi udara di Indonesia?
- Apa saja isu-isu spesifik yang sering muncul dalam percakapan terkait polusi udara di Twitter?
- Bagaimana intensitas percakapan tentang polusi udara berkembang seiring waktu di platform Twitter?
- Apakah terdapat perbedaan sentimen antara wilayah geografis di Indonesia terkait polusi udara?

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah diidentifikasi, tujuan penelitian ini dapat dirumuskan sebagai berikut:

- Untuk Menganalisis Sentimen Pengguna Twitter : Meneliti dan mengidentifikasi sentimen pengguna Twitter terkait polusi udara di Indonesia untuk memahami apakah tanggapan masyarakat cenderung positif, negatif, atau netral.
- Untuk Menemukan Isu-isu Spesifik dalam Percakapan : Mengidentifikasi isu-isu spesifik yang sering muncul dalam percakapan pengguna Twitter terkait polusi udara, sehingga dapat diidentifikasi fokus utama dari perbincangan di media sosial.
- Untuk Menganalisis Perubahan Intensitas Percakapan : Meneliti bagaimana intensitas percakapan tentang polusi udara berkembang seiring waktu di platform Twitter. Hal ini bertujuan untuk memahami apakah isu polusi udara semakin menjadi perhatian atau mengalami fluktuasi seiring berjalannya waktu.
- Untuk Menilai Perbedaan Sentimen antara Wilayah Geografis : Mengidentifikasi dan menganalisis apakah terdapat perbedaan sentimen antara wilayah geografis di Indonesia terkait polusi udara. Hal ini dapat memberikan wawasan tentang bagaimana persepsi masyarakat berbeda-beda di berbagai daerah.

## **1.4 Batasan Masalah**

- Penelitian ini terbatas pada analisis sentimen yang diungkapkan oleh pengguna Twitter saja. Sumber data dari platform lain tidak akan diikutsertakan dalam penelitian ini.
- Penelitian ini akan memfokuskan pada analisis sentimen dalam bahasa Indonesia. Penggunaan bahasa lain atau terjemahan tidak akan termasuk dalam cakupan penelitian.
- Jumlah data yang akan digunakan untuk analisis sentimen dapat dibatasi untuk memastikan kualitas dan akurasi analisis. Pemilihan sampel data yang representatif menjadi pertimbangan utama.
- Analisis sentimen terhadap perubahan intensitas percakapan seiring waktu akan dibatasi pada rentang waktu tertentu, dan tidak mencakup seluruh sejarah percakapan terkait polusi udara di Twitter.
- Ketersediaan data dari Twitter API atau sumber data lainnya dapat menjadi batasan. Keterbatasan ini dapat mempengaruhi jumlah dan jenis data yang dapat digunakan dalam penelitian.

## **1.5 Manfaat Penelitian**

### **A. Manfaat Teoritis**

1. Pengembangan Teori Sentimen : Penelitian ini dapat memberikan kontribusi pada pengembangan teori sentimen dengan menggunakan InSet Lexicon sebagai alat analisis. Hasil penelitian dapat membantu melengkapi dan memperkaya literatur terkait analisis sentimen di lingkungan media sosial.
2. Pemahaman Lebih Mendalam tentang Opini Publik : Penelitian ini dapat memberikan pemahaman lebih mendalam tentang bagaimana opini publik terbentuk dan berkembang di media sosial, khususnya terkait isu lingkungan seperti polusi udara.

### **B. Manfaat Praktis**

1. Hasil penelitian dapat menjadi sumber informasi berharga bagi pihak-pihak yang terlibat dalam pengambilan keputusan kebijakan, terutama dalam merancang langkah-langkah mitigasi polusi udara yang lebih efektif dan responsif terhadap kebutuhan masyarakat.
2. Penelitian ini dapat berperan dalam meningkatkan kesadaran masyarakat terkait dampak polusi udara. Informasi mengenai sentimen masyarakat dapat menjadi dasar untuk merancang kampanye edukasi yang lebih tepat sasaran.
3. Penelitian ini dapat memberikan panduan bagi penelitian-penelitian berikutnya terkait analisis sentimen dalam konteks isu lingkungan di media sosial.

Memperluas pengetahuan ini dapat membantu mengatasi permasalahan yang lebih kompleks di masa depan.

## **Bab II**

### **Tinjauan Pustaka**

#### **2.1 Penelitian Terdahulu**

Pada bagian ini, akan disajikan tinjauan literatur yang relevan dengan penelitian sebelumnya yang berkaitan dengan permasalahan yang sedang diteliti. Hasil-hasil penelitian terdahulu ini akan dijadikan sebagai acuan untuk melaksanakan Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Polusi di Indonesia menggunakan InSet Lexicon. Teori-teori pendukung yang dibahas dalam literatur ini akan menjadi sumber data penunjang untuk menjawab permasalahan penelitian yang diajukan.

Salah satu penelitian yang membahas tentang Analisis Sentimen pengguna Twitter dengan menggunakan Inset Lexicon dilakukan oleh (Desi Musfiroh, 2021) dengan judul “Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon”. Penelitian tersebut membahas mengenai pendekatan berbasis leksikal untuk memperoleh hasil klasifikasi sentimen dari data tweet pada topik “kuliah daring” menggunakan InSet Lexicon (Indonesian Sentiment Lexicon) yang berasal dari penelitian sebelumnya. Penelitian ini difokuskan untuk melakukan analisis sentimen/opini masyarakat khususnya para pengguna twitter terhadap topik “kuliah daring” melalui data yang diambil dari twitter dengan pengolahan data menggunakan bahasa pemrograman Python. InSet lexicon disusun oleh Fajri Koto dan Gemal Y. Rahmaningtyas pada penelitian sebelumnya dengan menggunakan kata-kata yang dikumpulkan dari twitter sebagai media sosial yang umum digunakan di Indonesia. InSet lexicon dibangun untuk mengidentifikasi opini tertulis dan mengkategorikannya menjadi opini positif atau negatif yang bisa digunakan untuk menganalisis sentimen publik terhadap topik, acara, atau produk tertentu.

Dalam konteks penelitian ini, tinjauan literatur terhadap penelitian-penelitian sebelumnya yang relevan sangat penting untuk memperoleh pemahaman yang lebih mendalam tentang penggunaan InSet Lexicon dalam analisis sentimen pengguna Twitter. Salah satu penelitian terkait adalah yang dilakukan oleh Desi Musfiroh pada tahun 2021,

berjudul "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon."

Penelitian tersebut berfokus pada pendekatan berbasis leksikal untuk mengklasifikasikan sentimen dari data tweet yang berkaitan dengan perkuliahan daring di Indonesia. Dengan menggunakan InSet Lexicon yang telah dikembangkan sebelumnya oleh Fajri Koto dan Gemal Y. Rahmaningtyas, penelitian tersebut berhasil mengidentifikasi opini positif dan negatif yang terkandung dalam data tweet mengenai kuliah daring. InSet Lexicon, yang diperoleh dari kata-kata yang dikumpulkan dari pengguna Twitter, menjadi alat analisis sentimen yang efektif dalam mengkategorikan pendapat masyarakat.

Penelitian ini memberikan dasar yang kuat untuk penggunaan InSet Lexicon dalam konteks analisis sentimen terhadap isu tertentu, seperti polusi udara di Indonesia. Dengan melihat hasil dan temuan yang telah dicapai oleh penelitian terdahulu, penelitian ini dapat memanfaatkan metode dan temuan tersebut sebagai landasan untuk melaksanakan analisis sentimen pengguna Twitter terkait polusi udara menggunakan InSet Lexicon. Melalui pendekatan yang serupa, diharapkan penelitian ini dapat memberikan wawasan yang berharga tentang pandangan dan perasaan masyarakat terhadap isu lingkungan yang krusial ini.

## **2.2 Twitter**

Twitter, sebagai salah satu platform media sosial paling populer di dunia, memainkan peran penting dalam merekam dan mengekspresikan opini masyarakat terkait berbagai isu, termasuk masalah lingkungan seperti polusi udara di Indonesia. Dalam konteks penelitian ini, tinjauan pustaka akan menggali penelitian dan literatur terkait peran Twitter dalam analisis sentimen serta bagaimana platform ini menjadi sumber data yang relevan untuk memahami pandangan masyarakat terhadap isu lingkungan. Twitter menyediakan platform terbuka untuk pengguna berbagi pendapat, pemikiran, dan reaksi mereka terhadap berbagai topik. Menurut (Java, et al., 2007), Twitter menjadi "real-time information network" yang mencerminkan opini dan sentimen masyarakat secara langsung. Oleh karena itu, penggunaan Twitter sebagai sumber data dapat memberikan pandangan instan tentang opini masyarakat terkait polusi di Indonesia. Penelitian sebelumnya, seperti yang dilakukan oleh (Pang, et al., 2008), telah menunjukkan bahwa analisis sentimen di media sosial, termasuk Twitter, dapat memberikan

pemahaman yang mendalam tentang pandangan masyarakat. Penggunaan teknik-teknik analisis sentimen, seperti pendekatan berbasis leksikal dan machine learning, membuka pintu untuk mengidentifikasi dan mengkategorikan sentimen pengguna terkait isu lingkungan.

## 2.2 InSet Lexicon

InSet Lexicon, yang dikembangkan oleh Fajri Koto dan Gemal Y. Rahmaningtyas pada tahun 2015, telah menjadi alat yang penting dalam analisis sentimen bahasa Indonesia. Lexicon ini terdiri dari daftar kata-kata yang diberi label emosional, memungkinkan identifikasi sentimen positif, negatif, dan netral dalam teks. Penggunaan InSet Lexicon memberikan dasar struktural yang kuat untuk menganalisis opini dan tanggapan masyarakat di media sosial, termasuk Twitter. Koto dan Rahmaningtyas membangun InSet Lexicon dengan mengumpulkan kata-kata dari Twitter, menciptakan sebuah sumber daya leksikal yang spesifik untuk bahasa Indonesia. Langkah ini menjadi relevan dalam konteks penelitian terkait polusi udara di Indonesia di mana pemahaman sentimen masyarakat dalam bahasa lokal menjadi krusial. Penelitian oleh Desi Musfiroh pada tahun 2021, yang berjudul "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," menjadi contoh nyata penerapan InSet Lexicon. Penelitian ini menunjukkan keberhasilan InSet Lexicon dalam mengklasifikasikan sentimen pengguna Twitter terkait topik tertentu. Hal ini memberikan landasan kuat untuk menggunakan InSet Lexicon dalam analisis sentimen terkait polusi udara. Keunggulan InSet Lexicon tidak hanya terletak pada kemampuannya mengklasifikasikan sentimen, tetapi juga pada kemampuannya bekerja dengan bahasa Indonesia. Pemilihan kata-kata yang spesifik untuk konteks budaya dan bahasa Indonesia menjadi kelebihan dalam memahami nuansa sentimen yang mungkin sulit diidentifikasi oleh lexicon bahasa asing.

Kata	Bobot
selisih	-4
sengsara	-5
awas	-3
iri	-4
coretan	-1
wewenang	2
sahabat	4
warisan	3
menarik	4
motivasi	3

Gambar 1. 1 Contoh Daftar Kata Dalam InSet Lexicon



## **2.3 Lexicon-based Approach**

Pemrosesan Bahasa Alami (Natural Language Processing atau NLP) merupakan cabang ilmu komputer yang berfokus pada pengembangan algoritma dan model untuk memahami, memproses, dan menghasilkan bahasa manusia. Dalam konteks ini, pendekatan berbasis lexicon atau leksikon menjadi salah satu metode yang penting untuk meningkatkan pemahaman dan analisis bahasa. Lexicon merujuk pada kumpulan kata-kata dan informasi terkait, seperti makna, kelas kata, dan hubungan semantik antar kata. Pendekatan berbasis lexicon mengintegrasikan pengetahuan leksikal dalam pemrosesan bahasa. Lexicon mencakup informasi tentang makna kata, kata-kata sinonim, antonim, serta relasi semantik lainnya. Dengan menggunakan lexicon, sistem NLP dapat meningkatkan kualitas analisis semantik dan mendekati pemahaman konteks secara lebih akurat. Salah satu penerapan utama pendekatan berbasis lexicon adalah dalam analisis sentimen. Lexicon menyediakan basis pengetahuan tentang kata-kata yang mengandung nuansa emosional. Misalnya, sebuah kata dengan makna positif seperti "bahagia" akan memiliki nilai sentimen positif. Dengan memanfaatkan lexicon, analisis sentimen dapat dilakukan secara efisien dan dapat diterapkan dalam berbagai domain, termasuk media sosial, ulasan produk, dan berita. Pendekatan berbasis lexicon menjadi komponen kunci dalam pengembangan sistem NLP yang mampu memahami bahasa manusia dengan lebih baik. Dengan memanfaatkan informasi leksikal, pendekatan ini meningkatkan kemampuan sistem untuk melakukan analisis semantik, pengenalan entitas, analisis sentimen, dan tugas-tugas NLP lainnya. Meskipun pendekatan ini memiliki kelebihan, penggunaannya juga harus memperhitungkan batasan dan tantangan, seperti kekurangan lexicon dalam menangani kata-kata baru atau perubahan dalam penggunaan kata.

## **2.4 Analisis Sentimen**

Analisis sentimen merupakan sebuah pendekatan yang digunakan untuk mengevaluasi dan menginterpretasi opini, pendapat, atau perasaan yang terkandung dalam teks, seperti ulasan produk, tweet, atau artikel. Tujuan dari analisis sentimen adalah untuk memahami apakah suatu teks menyatakan sentimen positif, negatif, atau netral terhadap suatu subjek atau topik tertentu. Beberapa metode dan teknik telah dikembangkan dalam literatur untuk

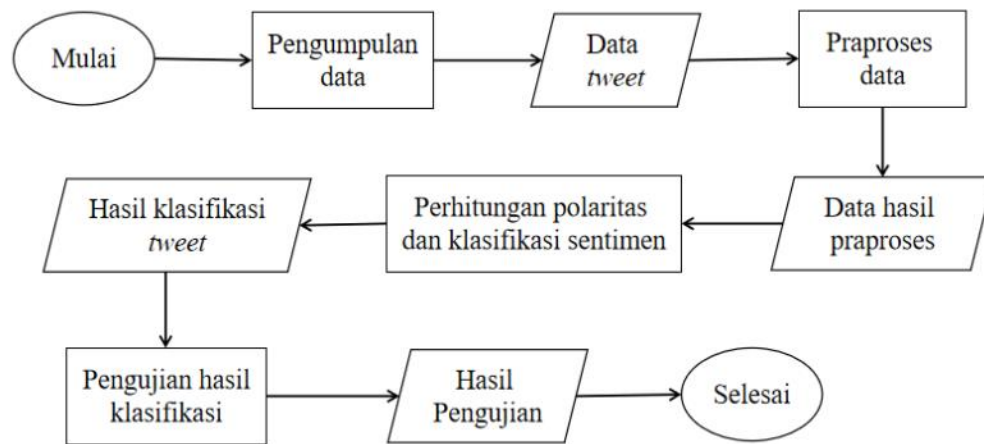
menerapkan analisis sentimen. Pada tinjauan pustaka ini, akan dibahas beberapa konsep dan pendekatan utama yang terkait dengan analisis sentimen. Analisis sentimen adalah proses komputasional untuk menentukan dan mengekstrak opini atau perasaan yang terkandung dalam teks. Sentimen dapat dibagi menjadi positif, negatif, atau netral, dan analisis sentimen bertujuan untuk mengklasifikasikan teks ke dalam kategori ini. lexicon-based approach atau pendekatan berbasis leksikon termasuk salah satu metode analisis sentimen. Metode ini bergantung pada penggunaan leksikon atau kamus kata-kata yang sudah diberi label sentimen (positif, negatif, atau netral) untuk mengevaluasi teks.

## Bab 3

### Metodologi Penelitian

#### 3.1 Prosedur Penelitian

Prosedur yang dilaksanakan dalam penelitian ini terdiri dari beberapa tahapan proses. Dimulai dari tahap pengumpulan data tweet, praproses data, perhitungan polaritas dan klasifikasi sentimen, hingga pengujian hasil klasifikasi.



Gambar 3. 1 Prosedur Penelitian

Adapun tiap tahapan yang ada pada prosedur penelitian akan diproses menggunakan bahasa pemrograman Python melalui Google Collab dengan menggunakan keyword “polusi udara”.

#### 3.2 Pengumpulan Data

Data yang digunakan dalam penelitian berasal dari Twitter, dipilih karena popularitasnya sebagai media sosial yang banyak digunakan oleh masyarakat. Twitter dianggap unggul sebagai saluran informasi dan berita tercepat, terutama ketika ada topik baru yang muncul (trending), sehingga banyak opini masyarakat yang bermunculan di platform ini. Opini-opini tersebut dianggap sebagai sumber data yang sangat relevan untuk melakukan analisis sentimen. Pengumpulan data dari Twitter dilakukan menggunakan Twitter API dengan kata kunci utama "polusi udara", dan jumlah data awal yang diambil dibatasi menjadi 100 tweet. Pembatasan jumlah data ini dilakukan untuk memastikan efisiensi pada langkah-langkah

selanjutnya, menghindari perlambatan pada praproses data, dan mencegah akun terkena pembatasan oleh Twitter. Jumlah 100 tweet dianggap ideal untuk memperoleh keragaman opini. Proses pengumpulan data tweet menggunakan alat scraping Twitter bernama Tweet Harvest (<https://github.com/helmisatria/tweet-harvest>). Data yang diperoleh mencakup beberapa atribut, namun fokus analisis sentimen difokuskan pada atribut tweet.

### 3.2 Praproses Data

Praproses data diperlukan untuk membersihkan dan mengorganisir data tweet yang telah dikumpulkan, sehingga dapat meningkatkan akurasi klasifikasi sentimen. Proses praproses dalam penelitian ini mencakup cleaning, tokenizing, filtering, dan stemming.

#### 1. Proses Cleaning :

- a. **Remove Punctuation:** Tahap pertama adalah menghilangkan tanda baca, sehingga hanya huruf alfabet yang tetap ada, dan karakter selain huruf dihapus.
- b. **Case Folding:** Melibatkan pengubahan seluruh teks menjadi huruf kecil untuk konsistensi.
- c. **Drop Duplicates:** Tujuan dari langkah ini adalah menghilangkan tweet yang duplikat atau menghapus spam tweet.
- d. **Spelling Correction:** Melibatkan perbaikan ejaan kata untuk memastikan konsistensi dan ketepatan kata-kata.

#### 2. Proses Tokenizing :

Proses ini melibatkan pemotongan string kalimat berdasarkan kata-kata yang menyusunnya. String karakter dipecah menjadi satuan kata untuk analisis lebih lanjut.

#### 3. Proses Filtering :

Tahap ini melibatkan pengambilan kata-kata penting dari hasil tokenisasi. Stopword dihilangkan untuk mengurangi jumlah kata yang disimpan, dengan stopwords adalah kata-kata umum yang dianggap tidak memberikan makna signifikan seperti "itu", "dan", "yang", dan "atau".

#### 4. Proses Stemming :

Proses ini mencakup penghapusan imbuhan kata untuk mengubah setiap kata ke dalam bentuk dasarnya, mempermudah analisis dan meminimalkan variasi morfologi kata.

### 3.3 Perhitungan Polaritas dan Klasifikasi Sentimen

Pada tahap ini, setiap tweet akan mengalami analisis sentimen secara individual. Klasifikasi sentimen untuk setiap tweet dilakukan menggunakan metode pendekatan berbasis lexicon dengan memanfaatkan InSet Lexicon. InSet Lexicon berisi sejumlah kata dalam bahasa Indonesia yang memiliki nilai sentimen positif dan negatif, beserta bobot yang terkait dengan masing-masing kata tersebut. Bobot kata berkisar antara -5 hingga 5, dengan nilai negatif menunjukkan sentimen negatif, dan nilai positif menunjukkan sentimen positif. Setiap kata dalam kalimat tweet akan dipasangkan dengan kata-kata dalam lexicon, dan selanjutnya, dilakukan perhitungan polarity score untuk setiap kalimat. Proses perhitungan polarity score dilakukan dengan menjumlahkan bobot dari kata-kata yang terdeteksi oleh sistem. Selanjutnya, data tweet diklasifikasikan ke dalam jenis sentimen menggunakan algoritma yang telah ditentukan. Algoritma tersebut dapat dijelaskan sebagai berikut :

- a. Jika sentiment score  $> 0$ , maka Sentimen Positif (1).
- b. Jika sentiment score  $< 0$ , maka Sentimen Negatif (2).
- c. Jika sentiment score  $= 0$ , maka Sentimen Netral (3).

Klasifikasi sentimen tweet ke dalam kategori positif, negatif, atau netral ditentukan berdasarkan bobot dari polarity score (sentiment score) yang dihasilkan. Tweet dianggap sebagai kelas positif jika bobot dari polarity score-nya lebih besar dari 0, dianggap sebagai kelas negatif jika bobot dari polarity score-nya lebih kecil dari 0. Sementara itu, tweet dengan polarity score yang sama dengan 0 akan diklasifikasikan sebagai kelas netral.

### 3.4 Pengujian Hasil Klasifikasi

Pada domain klasifikasi, evaluasi akurasi suatu model klasifikasi menjadi aspek kritis yang perlu diperhatikan. Nilai akurasi mencerminkan sejauh mana model klasifikasi dapat memberikan hasil yang baik. Penelitian ini melakukan pengujian akurasi menggunakan teknik cross-validation, di mana dataset dibagi menjadi dua bagian: training set (data latih) dan testing set (data uji). Training set digunakan untuk melatih model, sementara testing set digunakan untuk mengevaluasi kinerja model. Teknik cross-validation dilakukan dengan

sejumlah perulangan (epoch) untuk mencegah overfitting dan overlapping pada data uji. Setelah itu, data uji diproses untuk membuat confusion matrix.

Confusion Matrix merupakan matriks yang memuat informasi hasil klasifikasi oleh sistem, baik secara aktual maupun prediktif. Dengan mengevaluasi data pada matriks ini, dapat diukur sejauh mana performa suatu model. Confusion matrix untuk klasifikasi dua kelas ditampilkan dalam tabel berikut.

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP (True Positive)	FN (False Negative)
	Negatif	FP (False Positive)	TN (True Negative)

Gambar 3. 2 Confusion Matrix 2 Kelas

Dari analisis nilai yang terdapat dalam kolom matriks, seperti True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP), dapat dihasilkan informasi mengenai nilai akurasi, presisi, recall, dan F-measure.

- a. Akurasi mencerminkan sejauh mana hasil klasifikasi mendekati nilai sesungguhnya. Akurasi dihitung melalui perbandingan antara data yang diklasifikasikan dengan benar dan keseluruhan data.

$$accuracy = \frac{TP + TN}{TP + FP + FN} \times 100\%$$

- b. Presisi merupakan ukuran ketepatan yang mencerminkan seberapa dekat perbedaan nilai pada setiap iterasi atau pengulangan. Dari nilai presisi, kita dapat memahami sejauh mana hasil klasifikasi mendekati informasi yang diminta jika dibandingkan dengan jawaban yang diberikan oleh sistem.

$$precision = \frac{TP}{TP + FP} \times 100\%$$

- c. Recall, atau yang sering disebut sebagai sensitivitas, adalah persentase nilai yang menunjukkan sejauh mana suatu model dapat memprediksi data ke dalam kelas yang sesuai dengan kelas aktualnya.

$$recall = \frac{TP}{TP + FN} \times 100\%$$

- d. F-measure adalah perhitungan yang mendapatkan informasi dengan memadukan nilai recall dan precision.

$$f - measure = \frac{2 \times recall \times precision}{recall + precision} \times 100\%$$

Model klasifikasi menghasilkan prediksi yang benar atau salah, dan hasil ini terlihat dalam Confusion Matrix. Dalam model klasifikasi ini, terdapat tiga kelas sehingga matriks kebingungan yang dihasilkan memiliki ordo 3x3, seperti yang ditunjukkan dalam Gambar 3.3 . Matriks tersebut mencakup data aktual dan data prediksi. Dari Confusion Matrix ini, diperoleh nilai rata-rata akurasi, presisi, recall, dan F-measure.

		Prediksi		
Aktual	Kelas A	AA	AB	AC
	Kelas B	BA	BB	BC
	Kelas C	CA	CB	CC

Gambar 3. 3 Confusion Matrix 3 Kelas

### 3.6 Alat Penelitian

Alat yang akan digunakan dalam penelitian ini terdiri dari 2 macam, yaitu :

1. Perangkat Keras (Hardware) :
  - a. Processor : Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.59 GHz
  - b. RAM : 16GB
  - c. Arsitektur : 64-Bit Operating System

d. Perangkat : Laptop, Mouse, dan Keyboard

2. Perangkat Lunak (Software) :

a. Sistem Operasi : Windows 11

b. Pengolahan Data : Google Colab

c. Tools & Library : Pandas, Nodejs, tweet-harvest

d. Bahasa Pemrograman : Python



## DAFTAR PUSTAKA

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, 56-65.

Pang, B., Lee, L., & Vaithyanathan, S. (2008). Thumbs up? Sentiment Classification using Machine Learning Techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 79-86.

Koto, F., & Rahmaningtyas, G. Y. (2015). InSet Lexicon: An Indonesian Sentiment Analysis Lexicon. *Procedia Computer Science*, 72, 365-372.

Musfiroh, D., Khaira, U., Utomo, P. E. P., & Suratno, T. (2021). Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 24-33.

M. R. Ma'arif, "Analisis Konten Interaksi Pengguna Twitter pada Masa 100 Hari Pertama Pemerintahan Baru DKI Jakarta Menggunakan Text Mining," *Jurnal Pekomnas*, Vol. 3 No. 2, 137-142, 2018

Y. Azhar, "Metode Lexicon Learning Based untuk Identifikasi Tweet Opini Berbahasa Indonesia," *Jurnal Nasional Pendidikan Teknik Informatika*, Vol. 6, No. 3, 2017.