

## BAB II

### TINJAUAN LITERATUR

#### 2.1 Analisis Sentimen

Analisis sentimen merupakan salah satu teknik dalam bidang Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi, mengolah, dan mengklasifikasikan opini atau perasaan yang terdapat dalam suatu teks. Analisis sentimen dapat digunakan untuk mengetahui kecenderungan opini masyarakat terhadap suatu objek, topik, produk, layanan, maupun informasi tertentu. Menurut Ligthart, (Ligthart, Catal and Tekinerdogan, 2021) analisis sentimen merupakan kajian komputasional terhadap opini, sentimen, emosi, dan sikap yang terdapat dalam teks yang dihasilkan oleh pengguna. Perkembangan media digital menyebabkan semakin banyaknya konten yang berisi opini masyarakat sehingga analisis sentimen dapat digunakan untuk mengidentifikasi kecenderungan sentimen dalam data tersebut. Analisis sentimen dapat membantu mengubah data teks yang tidak terstruktur menjadi informasi yang dapat digunakan untuk mengetahui kecenderungan pendapat pengguna. Dalam penelitian ini, analisis sentimen digunakan untuk mengetahui kecenderungan opini masyarakat terhadap informasi berita Kompas.com yang dipublikasikan melalui platform Instagram. Komentar masyarakat diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Sentimen positif, yaitu komentar yang menunjukkan dukungan, persetujuan, apresiasi, atau tanggapan yang memiliki kecenderungan positif. Sentimen negatif, yaitu komentar yang menunjukkan kritik, ketidaksetujuan, kekecewaan, atau tanggapan yang memiliki kecenderungan negatif. Sentimen netral, yaitu komentar yang tidak menunjukkan kecenderungan positif maupun negatif secara jelas. Hasil klasifikasi tersebut kemudian digunakan untuk mengetahui kecenderungan respons masyarakat terhadap informasi berita yang menjadi objek penelitian.

## 2.2 Media Online

Media online merupakan media penyampaian informasi yang memanfaatkan jaringan internet sehingga informasi dapat disampaikan dan diakses oleh masyarakat secara cepat. Media online memungkinkan proses penyebaran berita berlangsung secara lebih luas dan tidak terbatas pada ruang dan waktu. Perkembangan teknologi digital menyebabkan media berita tidak hanya menyampaikan informasi melalui situs web, tetapi juga memanfaatkan berbagai platform media sosial untuk menjangkau masyarakat. Penyebaran informasi melalui media sosial juga memungkinkan masyarakat memberikan tanggapan secara langsung terhadap berita yang diterima. Dalam penelitian ini, media online menjadi bagian penting karena objek informasi yang dianalisis berasal dari Kompas.com, yang kemudian menyebarkan konten beritanya melalui platform Instagram.

## 2.3 Kompas.Com

Kompas.com merupakan salah satu media berita online di Indonesia yang menyediakan berbagai informasi aktual kepada masyarakat. Informasi yang disajikan mencakup berbagai bidang, seperti politik, ekonomi, teknologi, sosial, budaya, dan bidang lainnya. Perkembangan media digital membuat Kompas.com tidak hanya menyampaikan berita melalui situs web, tetapi juga memanfaatkan media sosial sebagai sarana untuk mendistribusikan informasi kepada masyarakat. Salah satunya adalah melalui platform Instagram dengan akun @kompascom. Pemanfaatan Instagram sebagai media distribusi berita memungkinkan masyarakat memberikan tanggapan secara langsung melalui kolom komentar. Komentar tersebut dapat menjadi sumber data untuk mengetahui respons masyarakat terhadap informasi yang disampaikan. Dalam penelitian ini, Kompas.com digunakan sebagai sumber informasi, sedangkan komentar pengguna pada unggahan Instagram @kompascom digunakan sebagai data untuk melakukan analisis sentimen.

## 2.4 Media Sosial

Media sosial merupakan platform digital yang memungkinkan pengguna untuk membuat, menyebarkan, dan bertukar informasi serta melakukan interaksi dengan pengguna lainnya. Media sosial memiliki karakteristik interaktif karena pengguna tidak hanya berperan sebagai penerima informasi, tetapi juga dapat memberikan tanggapan terhadap informasi yang diterima. Perkembangan teknologi informasi dan komunikasi telah menciptakan ekosistem digital yang memungkinkan masyarakat untuk mengakses dan menyebarkan informasi secara cepat melalui media online. Media sosial tidak hanya berfungsi sebagai penyedia informasi, tetapi juga sebagai ruang interaktif yang memungkinkan audiens untuk memberikan respons secara langsung terhadap konten yang disajikan. Menurut (Harjono, Choirunnisa and Febriana, 2025) media berita digital memberikan akses informasi secara real-time serta membuka ruang interaksi antara media dan masyarakat melalui berbagai platform digital. Menurut (Harjono, Choirunnisa and Febriana, 2025) media sosial berfungsi sebagai ruang komunikasi dua arah yang memungkinkan interaksi antara pengguna dan penyedia informasi, sehingga menciptakan dinamika opini publik yang dapat diamati secara real-time.

## 2.5 Instagram

Instagram merupakan salah satu platform media sosial yang memungkinkan pengguna untuk membuat, membagikan, dan memperoleh berbagai informasi dalam bentuk foto, video, maupun teks. Instagram menyediakan berbagai fitur yang mendukung aktivitas komunikasi dan interaksi antarpengguna, seperti feed, Instagram Story, Reels, komentar, tanda suka, dan Direct Message. Menurut (Uddin, Maharani and Baren, 2024) berbagai fitur yang tersedia pada Instagram

dapat mendukung aktivitas komunikasi dan interaksi pengguna. Keberadaan fitur tersebut membuat Instagram tidak hanya digunakan sebagai media berbagi konten, tetapi juga sebagai sarana komunikasi dan penyebaran informasi. Dalam perkembangannya, Instagram juga dimanfaatkan oleh media massa sebagai salah satu sarana penyebaran berita. Media berita dapat mengunggah informasi dalam bentuk gambar maupun video yang kemudian dapat memperoleh tanggapan masyarakat melalui kolom komentar. Komentar pengguna Instagram memiliki karakteristik yang beragam. Pengguna dapat menggunakan bahasa informal, singkatan, kata tidak baku, bahasa gaul, kesalahan penulisan (typo), simbol, emoji, maupun campuran bahasa. Karakteristik tersebut menyebabkan komentar Instagram perlu melalui tahap preprocessing sebelum digunakan dalam proses analisis sentimen. Dalam penelitian ini, Instagram digunakan sebagai platform sumber data, sedangkan komentar masyarakat pada unggahan berita akun @kompascom digunakan sebagai data penelitian. Data komentar tersebut selanjutnya diproses melalui tahapan preprocessing, pembobotan TF-IDF, dan klasifikasi menggunakan Support Vector Machine (SVM).

## **2.6 Opini Publik**

Opini publik merupakan pendapat, pandangan, atau tanggapan yang diberikan masyarakat terhadap suatu persoalan atau informasi tertentu. Dalam lingkungan digital, opini publik dapat terlihat melalui berbagai bentuk interaksi pengguna, salah satunya melalui komentar pada media sosial. Kolom komentar pada media sosial dapat menjadi ruang bagi masyarakat untuk menyampaikan pendapat, kritik, dukungan, maupun tanggapan terhadap suatu informasi. Banyaknya komentar yang muncul dapat menunjukkan adanya keberagaman respons

masyarakat terhadap suatu isu. Dalam penelitian ini, komentar masyarakat terhadap berita Kompas.com dipandang sebagai bentuk data tekstual yang dapat digunakan untuk mengetahui kecenderungan opini masyarakat. Opini tersebut kemudian diklasifikasikan menjadi sentimen positif, negatif, dan netral menggunakan metode analisis sentimen.

## **2.7 Text Mining**

Text mining merupakan proses untuk memperoleh informasi dan pola tertentu dari data berbentuk teks yang tidak terstruktur. Menurut (Sarker, 2021), machine learning dapat digunakan untuk mengolah data dalam jumlah besar dan menemukan pola berdasarkan data yang tersedia. Data teks yang berasal dari media sosial memiliki jumlah dan karakteristik yang beragam sehingga sulit dianalisis secara manual. Oleh karena itu, diperlukan teknik komputasi untuk mengolah data tersebut menjadi informasi yang dapat dianalisis. Dalam penelitian ini, text mining digunakan untuk mengolah komentar masyarakat pada Instagram @kompascom. Data komentar yang diperoleh kemudian dibersihkan dan diubah menjadi bentuk yang dapat diproses oleh algoritma klasifikasi. Tahapan pengolahan teks dalam penelitian ini meliputi cleaning, case folding, normalisasi, tokenisasi, stopword removal, dan stemming

## **2.8 Preprocessing**

Preprocessing merupakan tahap awal dalam pengolahan data teks yang bertujuan membersihkan dan mempersiapkan data sebelum dilakukan proses analisis. Tahap ini penting karena komentar media sosial memiliki banyak variasi penulisan dan mengandung berbagai unsur yang tidak diperlukan dalam proses klasifikasi.

Tahapan preprocessing yang digunakan dalam penelitian ini adalah sebagai berikut.

### **2.8.1 Cleaning**

Cleaning merupakan proses membersihkan teks dari unsur-unsur yang tidak diperlukan, seperti URL, angka, simbol, emoji, karakter khusus, dan *mention* pengguna. Tahap ini dilakukan agar data yang digunakan dalam proses klasifikasi lebih bersih dan relevan.

### **2.8.2 Case Folding**

*Case Folding* merupakan proses mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*). Proses ini bertujuan menyeragamkan bentuk penulisan sehingga kata dengan huruf kapital dan huruf kecil tidak dianggap sebagai dua kata yang berbeda.

### **2.8.3 Normalisasikan Kata**

Normalisasi merupakan proses mengubah kata tidak baku, singkatan, atau istilah tertentu menjadi bentuk kata yang lebih standar. Tahap ini penting karena komentar Instagram banyak menggunakan bahasa informal dan singkatan.

### **2.8.4 Tokenisasi**

Tokenisasi merupakan proses memecah suatu kalimat menjadi bagian-bagian kata atau *token*. Tokenisasi diperlukan agar setiap kata dapat diproses pada tahap selanjutnya.

### **2.8.5 Stopword removal**

*Stopword Removal* merupakan proses menghapus kata-kata umum yang dianggap tidak memberikan kontribusi penting terhadap proses analisis. Penghapusan *stopword* bertujuan mengurangi kata yang tidak diperlukan sehingga data menjadi lebih efektif untuk diproses.

### **2.8.6 Stemming**

*Stemming* merupakan proses mengubah kata berimbuhan menjadi bentuk kata dasarnya. Dengan *stemming*, beberapa bentuk kata yang memiliki kata dasar sama dapat diseragamkan sehingga membantu proses pembobotan dan klasifikasi teks.

### **2.9 Pelabelan Sentimen**

Pelabelan *sentimen* merupakan proses memberikan kategori atau label terhadap data teks berdasarkan kecenderungan sentimennya. Dalam penelitian ini, data komentar dikategorikan menjadi tiga kelas, yaitu positif, negatif, dan netral. Pelabelan awal pada penelitian ini menggunakan pendekatan *lexicon-based*, yaitu menentukan sentimen berdasarkan kata-kata yang terdapat dalam kamus sentimen. Skor sentimen kemudian digunakan untuk menentukan kategori komentar. Data yang telah diberikan label selanjutnya digunakan sebagai data untuk melatih dan menguji model *Support Vector Machine*.

### **2.10 Term Frequency - Inverse Document Frequency(TF-IDF)**

TF-IDF merupakan metode pembobotan yang digunakan untuk mengubah data teks menjadi bentuk numerik berdasarkan tingkat kepentingan suatu kata dalam dokumen. TF-IDF terdiri dari dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). Term Frequency menunjukkan seberapa sering suatu kata muncul dalam dokumen,

sedangkan Inverse Document Frequency memberikan bobot berdasarkan tingkat kemunculan kata tersebut pada keseluruhan dokumen. Pembobotan TF-IDF digunakan karena algoritma SVM membutuhkan data dalam bentuk numerik. Oleh karena itu, komentar yang telah melalui tahap preprocessing diubah menjadi vektor numerik menggunakan TF-IDF sebelum digunakan sebagai input model SVM. Dalam penelitian ini, TF-IDF diterapkan terhadap 2.083 data komentar bersih yang diperoleh setelah proses preprocessing dan penghapusan data kosong.

### **2.11 Support Vector Machine (SVM)**

SVM merupakan algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal untuk memisahkan data ke dalam kelas tertentu. Menurut Dang, Moreno-García, dan De la Prieta (2020), Support Vector Machine (SVM) merupakan salah satu metode machine learning yang digunakan dalam klasifikasi sentimen. Dalam penelitian analisis sentimen, SVM digunakan sebagai salah satu metode pembandingan terhadap pendekatan deep learning untuk melakukan klasifikasi polaritas sentimen pada data teks. Dalam implementasinya, analisis sentimen banyak memanfaatkan metode machine learning untuk meningkatkan akurasi klasifikasi. Machine learning memungkinkan sistem untuk mempelajari pola dari data teks dan melakukan prediksi secara otomatis. Menurut (Sarker, 2021), pendekatan machine learning efektif dalam menangani data besar serta mampu meningkatkan performa analisis melalui proses pembelajaran berbasis data. Salah satu metode yang banyak digunakan dalam analisis sentimen adalah Support Vector Machine (SVM).

Menurut (Utomo *et al.*, 2025), metode Support Vector Machine (SVM) mampu digunakan secara efektif dalam mengklasifikasikan sentimen pada data teks media sosial dan menghasilkan performa klasifikasi yang baik. Penelitian ZONasi (Tampubolon and Devega, 2025) juga menunjukkan bahwa metode SVM mampu memberikan tingkat akurasi yang tinggi dalam klasifikasi sentimen komentar Instagram sehingga relevan digunakan dalam analisis opini publik berbasis media sosial. Beberapa penelitian terdahulu menunjukkan bahwa metode SVM memiliki performa yang baik dalam analisis sentimen.

Penelitian oleh (Triully Prasetyo and Atiqah Meutia Hilda, 2024) menunjukkan bahwa SVM mampu mengklasifikasikan sentimen media sosial dengan tingkat akurasi yang tinggi. Selain itu, Penelitian oleh (Tarigan and Yusupa, 2024) menunjukkan bahwa Support Vector Machine (SVM) memiliki performa klasifikasi yang tinggi serta mampu bersaing dengan metode lain seperti Random Forest dalam analisis sentimen. Temuan ini juga didukung oleh penelitian lain yang menyatakan bahwa SVM memiliki kinerja yang kompetitif dibandingkan beberapa algoritma klasifikasi lainnya (Indriani and Davy Wiranata, 2024).

Berdasarkan kajian tersebut, dapat disimpulkan bahwa analisis sentimen merupakan pendekatan yang efektif dalam memahami opini masyarakat terhadap suatu informasi. Pemanfaatan metode SVM dalam analisis sentimen terbukti mampu memberikan hasil klasifikasi yang akurat dan stabil. Oleh karena itu, penelitian ini menggunakan metode SVM untuk menganalisis sentimen masyarakat terhadap berita Kompas.com mengenai pernyataan Menteri Keuangan Purbaya Yudhi Sadewa terkait penurunan omzet Warung Tegal

(Warteg) yang dipublikasikan melalui akun Instagram @kompascom, dengan tujuan untuk memahami persepsi publik secara lebih objektif dan berbasis data.

## 2.12 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kemampuan algoritma SVM dalam melakukan klasifikasi sentimen. Dalam penelitian ini, evaluasi dilakukan menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score.

### 2.12.1 Accuracy

*Accuracy* digunakan untuk mengetahui persentase prediksi yang benar dari keseluruhan data pengujian.

### 2.12.2 Precision

*Precision* menunjukkan tingkat ketepatan model ketika memberikan prediksi terhadap suatu kelas.

### 2.12.3 Recall

*Recall* menunjukkan kemampuan model dalam menemukan data yang sebenarnya termasuk ke dalam suatu kelas.

### 2.12.4 F1-Score

F1-score merupakan ukuran yang menggabungkan nilai precision dan *recall*. i ini digunakan untuk melihat keseimbangan antara kemampuan model dalam memberikan prediksi yang tepat dan menemukan data yang relevan.

### 2.12.5 Confusion Matrix

*Confusion matrix* merupakan tabel yang digunakan untuk melihat hasil klasifikasi model dengan membandingkan label aktual dan label hasil prediksi. Pada penelitian ini, *confusion matrix* digunakan untuk melihat

kemampuan SVM dalam mengklasifikasikan tiga kelas sentimen, yaitu positif, negatif, dan netral.

### 2.13 Penelitian Terdahulu

**Tabel 1. Penelitian Terdahulu**

| <b>No</b> | <b>Penelitian</b>   | <b>Tahun</b> | <b>Objek Penelitian</b>                                    | <b>Metode</b> | <b>Hasil</b>                                                                                | <b>Kelemahan</b>                   |
|-----------|---------------------|--------------|------------------------------------------------------------|---------------|---------------------------------------------------------------------------------------------|------------------------------------|
| 1         | Astuti & Ardiansyah | 2025         | Komentar Instagram konten edukasi                          | SVM           | SVM mampu mengklasifikasikan sentimen dengan akurasi yang baik pada data komentar instagram | Tidak fokus pada akun media berita |
| 2         | Saputra & Devaga    | 2024         | Komentar Instagram terkait cyber-bullying                  | SVM           | Mampu mendeteksi komentar negatif/cyber-bullying                                            | Fokus pada isu sosial penting      |
| 3         | Herliana & Muawiyah | 2023         | Komentar Instagram                                         | SVM + PSO     | PSO meningkatkan akurasi klasifikasi                                                        | Objek penelitian masih umum        |
| 4         | Tujahidah et al.    | 2024         | Komentar media sosial (termasuk instagram) terkait pilkada | SVM           | PSO meningkatkan akurasi klasifikasi                                                        | Tidak spesifik instagram saja      |
| 5         | Prastya             | 2024         | Komentar Instagram @timnasindonesia                        | CNN + SVM     | Hybrid model meningkatkan akurasi                                                           | Tidak fokus pada berita/media      |

Berdasarkan berbagai penelitian terdahulu yang telah dianalisis, dapat disimpulkan bahwa metode Support Vector Machine merupakan salah satu algoritma yang efektif dan banyak digunakan dalam analisis sentimen, khususnya pada data teks dari media sosial. Metode ini terbukti mampu menghasilkan tingkat

akurasi yang baik dalam mengklasifikasikan sentimen menjadi kategori positif, negatif, dan netral. Selain itu, platform Instagram menunjukkan potensi yang besar sebagai sumber data dalam penelitian analisis sentimen karena memiliki tingkat interaksi pengguna yang tinggi serta menyediakan data berupa komentar yang mencerminkan opini publik secara langsung. Berbagai penelitian juga menunjukkan bahwa data dari Instagram dapat digunakan untuk menganalisis berbagai isu, seperti sosial, politik, hingga kesehatan.

Namun demikian, sebagian besar penelitian terdahulu masih memiliki beberapa keterbatasan. Banyak penelitian yang hanya berfokus pada isu tertentu seperti cyberbullying, politik, atau kesehatan, tanpa mengaitkannya dengan konten berita dari media massa. Selain itu, objek penelitian yang digunakan umumnya masih bersifat umum dan belum secara spesifik meneliti akun resmi media berita. Di sisi lain, belum banyak penelitian yang mengkaji analisis sentimen terhadap konten berita yang dipublikasikan melalui media sosial, khususnya pada akun resmi media seperti Kompas.com. Padahal, akun media berita memiliki peran penting dalam membentuk opini publik karena tingkat kredibilitas dan jangkauan audiens yang luas.

Oleh karena itu, penelitian ini dilakukan untuk mengisi kekosongan tersebut dengan menganalisis sentimen masyarakat terhadap informasi berita Kompas.com pada platform Instagram menggunakan metode SVM. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang lebih spesifik dan relevan dalam pengembangan analisis sentimen berbasis media sosial.