



# **ANALISIS SENTIMEN PADA ULASAN PRODUK DI ECOMMERCE DENGAN METODE NAÏVE BAYES**

Tugas Akhir  
diajukan untuk melengkapi  
persyaratan mencapai  
gelar sarjana

NAMA : ASSAMI MUZAKI  
NPM : 201943502262

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN ILMU KOMPUTER  
UNIVERSITAS INDRAPRASTA PGRI  
2023**

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang**

*E-commerce* atau perdagangan elektronik telah mengalami pertumbuhan yang pesat dalam beberapa tahun terakhir. Semakin banyak konsumen yang membeli produk melalui platform *e-commerce* seperti Amazon, Alibaba, Tokopedia, Shopee, dan sebagainya. Dalam konteks ini, ulasan produk oleh pengguna menjadi salah satu sumber informasi yang penting bagi calon pembeli dalam membuat keputusan pembelian.

Ulasan produk di *e-commerce* dapat mempengaruhi keputusan pembelian konsumen. Konsumen cenderung mempercayai ulasan dari pengguna sebelumnya untuk menilai kualitas, keandalan, dan kepuasan produk yang ditawarkan. Oleh karena itu, analisis sentimen terhadap ulasan produk di *e-commerce* menjadi penting untuk membantu para penjual atau pedagang dalam memahami bagaimana konsumen merespons produk mereka.

Namun, menganalisis ulasan produk di *e-commerce* secara manual dapat menjadi tugas yang rumit dan memakan waktu, terutama jika jumlah ulasan yang harus dianalisis sangat besar. Oleh karena itu, diperlukan pendekatan yang efisien dan akurat untuk mengolah data ulasan produk yang ada.

Dalam konteks tersebut, Metode *naive bayes* menjadi pilihan yang potensial karena dapat mengatasi kendala tersebut. Metode ini dapat mengklasifikasikan sentimen ulasan produk secara otomatis dengan mengandalkan probabilitas kondisional dari fitur-fitur yang ada dalam teks ulasan. Dengan menggunakan Metode *Naive bayes*, proses analisis sentimen terhadap ulasan produk di *e-commerce* dapat dilakukan secara efisien dan akurat, mengurangi ketergantungan pada analisis manual yang rumit dan memakan waktu.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan baru dalam pemahaman sentimen konsumen terhadap ulasan produk di *e-commerce*, serta dapat digunakan sebagai panduan bagi para penjual atau pedagang dalam mengambil keputusan yang lebih baik dalam meningkatkan kualitas produk, kepuasan pelanggan, dan reputasi di platform *e-commerce*. Penelitian ini juga diharapkan dapat menjadi referensi bagi peneliti lain yang tertarik dalam bidang analisis sentimen, terutama dalam konteks *e-commerce*.

## **B. Identifikasi Masalah**

Berdasarkan latar belakang di atas, dapat diidentifikasi beberapa masalah dalam Analisis Sentimen Terhadap Ulasan Produk di *E-commerce* Menggunakan Metode *Naive bayes*, yaitu:

1. Masalah Efisiensi dan Akurasi: Menganalisis ulasan produk di *e-commerce* secara manual dapat menjadi tugas yang rumit dan memakan waktu,

terutama jika jumlah ulasan yang harus dianalisis sangat besar. Oleh karena itu, diperlukan pendekatan yang efisien dan akurat untuk mengolah data ulasan produk yang ada. Bagaimana mengimplementasikan dan mengoptimalkan Metode *Naive bayes* untuk mengatasi kendala efisiensi dan akurasi dalam analisis sentimen terhadap ulasan produk di *e-commerce*?

2. Masalah Data Tidak Seimbang (Imbalanced Data): Data ulasan produk di *e-commerce* seringkali tidak seimbang, di mana jumlah ulasan positif, negatif, dan netral tidak merata. Hal ini dapat mempengaruhi kinerja Metode *Naive bayes* dalam mengklasifikasikan sentimen ulasan produk. Bagaimana mengatasi masalah data tidak seimbang dalam analisis sentimen menggunakan Metode *Naive bayes* di konteks *e-commerce*?
3. Masalah Pengaturan Parameter: Metode *Naive bayes* memiliki parameter yang perlu diatur, seperti smoothing parameter dalam model *Naive bayes*. Bagaimana mengoptimalkan pengaturan parameter Metode *Naive bayes* untuk memperoleh hasil yang optimal dalam analisis sentimen terhadap ulasan produk di *e-commerce*?
4. Masalah Aplikabilitas Hasil Penelitian: Bagaimana hasil penelitian ini dapat digunakan sebagai panduan bagi para penjual atau pedagang dalam meningkatkan kualitas produk, kepuasan pelanggan, dan reputasi di platform *e-commerce*? Bagaimana hasil penelitian ini dapat menjadi

referensi bagi peneliti lain yang tertarik dalam bidang analisis sentimen, terutama dalam konteks *e-commerce*?

### **C. Batasan Masalah**

Berdasarkan rumusan masalah dan batasan kemampuan peneliti, perlu diadakan batasan-batasan masalah agar penelitian lebih terarah. Batasan-batasan tersebut diantaranya:

1. Batasan Data: Penelitian ini akan membatasi penggunaan data ulasan produk di *e-commerce* dari satu platform atau beberapa platform tertentu. Data yang digunakan dapat berupa teks ulasan produk dalam bahasa tertentu, misalnya bahasa Indonesia saja.

Batasan Metode: Penelitian ini akan berfokus pada penggunaan metode Naive Bayes sebagai algoritma klasifikasi dalam analisis sentimen terhadap ulasan produk di *e-commerce*. Metode lain yang digunakan dalam analisis sentimen, seperti Support Vector Machine (SVM) atau Deep Learning, tidak akan dibahas secara mendalam dalam penelitian ini.

2. Batasan Jumlah Kategori Sentimen: Penelitian ini akan membatasi analisis sentimen pada tiga kategori sentimen utama, yaitu positif, negatif, dan netral. Kategori sentimen lainnya, seperti campuran antara positif dan negatif, tidak akan dipertimbangkan dalam penelitian ini.

3. Batasan Jenis Data: Penelitian ini akan membatasi data ulasan yang digunakan yakni hanya berupa ulasan text. Ulasan lain seperti gambar, rating dan lainnya tidak akan di gunakan untuk perhitungan dalam penelitian ini.

Batasan-batasan di atas akan membantu dalam merujuk area yang spesifik dan terfokus dalam penelitian, sehingga penelitian dapat dilakukan dengan cakupan yang jelas dan hasil yang relevan dapat diperoleh.

#### **D. Rumusan Masalah**

Berdasarkan latar belakang dan batasan masalah di atas, maka rumusan masalah untuk topik ini adalah:

1. Bagaimana metode *Naive bayes* dapat diterapkan untuk melakukan analisis sentimen terhadap ulasan produk di *e-commerce*?
2. Seberapa akurat dan efisien metode *Naive bayes* dalam melakukan analisis sentimen terhadap ulasan produk di *e-commerce*?
3. Bagaimana hasil analisis sentimen menggunakan metode *Naive bayes* dapat membantu pemilik bisnis *e-commerce* dalam pengambilan keputusan terkait produk dan layanan yang ditawarkan?

Rumusan masalah tersebut dapat digunakan sebagai panduan untuk menjalankan penelitian yang mendalam dan komprehensif mengenai

penggunaan metode *Naive bayes* dalam analisis sentimen terhadap ulasan produk di *e-commerce*.

#### **E. Tujuan Penelitian**

Berdasarkan topik "Analisis Sentimen Pada Ulasan Produk di *E-commerce* dengan Metode Naïve Bayes", tujuan dari penelitian skripsi tersebut adalah:

1. Menjelaskan penerapan metode *Naive bayes* dalam melakukan analisis sentimen terhadap ulasan produk di *e-commerce*, termasuk langkah-langkah yang diperlukan untuk mengimplementasikan metode ini dalam memproses data ulasan produk di platform *e-commerce*.
2. Mengevaluasi akurasi dan efisiensi metode *Naive bayes* dalam melakukan analisis sentimen terhadap ulasan produk di *e-commerce* dengan menggunakan data uji yang relevan. Tujuan ini melibatkan pengukuran tingkat akurasi dan efisiensi metode *Naive bayes* dalam mengklasifikasikan sentimen positif, negatif, atau netral pada ulasan produk di *e-commerce*.
3. Menganalisis hasil analisis sentimen menggunakan metode *Naive bayes* dan mengidentifikasi bagaimana hasil tersebut dapat membantu pemilik bisnis *e-commerce* dalam pengambilan keputusan terkait produk dan layanan yang ditawarkan. Tujuan ini melibatkan interpretasi dan analisis terhadap output dari metode *Naive bayes* dalam konteks bisnis *e-commerce*, sehingga dapat

memberikan wawasan yang berguna bagi pemilik bisnis dalam pengambilan keputusan yang lebih informasional dan tepat.

Tujuan penelitian tersebut bertujuan untuk menggali pemahaman yang lebih dalam tentang penggunaan metode *Naive bayes* dalam analisis sentimen terhadap ulasan produk di *e-commerce*, serta memberikan kontribusi pada pengembangan pengetahuan dalam bidang ini.

#### **F. Manfaat Penelitian**

Adapun kegunaan dari penelitian skripsi dengan topik "Analisis Sentimen Pada Ulasan Produk di *E-commerce* dengan Metode Naïve Bayes" adalah sebagai berikut:

1. Memberikan wawasan yang lebih baik kepada pemilik bisnis *e-commerce* tentang bagaimana konsumen merespons produk mereka melalui analisis sentimen terhadap ulasan produk. Hal ini dapat membantu pemilik bisnis dalam mengambil keputusan yang lebih tepat dalam meningkatkan kualitas produk, kepuasan pelanggan, dan reputasi produk di platform *e-commerce*.
2. Menyediakan alat analisis yang efisien dan akurat bagi pemilik bisnis *e-commerce* untuk mengolah data ulasan produk yang ada, yang dapat menghemat waktu dan upaya dalam melakukan analisis secara manual.
3. Menyediakan pemahaman yang lebih dalam tentang kinerja metode *Naive bayes* dalam melakukan analisis sentimen terhadap ulasan produk di *e-*



*commerce*, yang dapat menjadi referensi bagi peneliti atau praktisi dalam bidang analisis sentimen dan *e-commerce*.

4. Menambahkan pengetahuan dalam bidang analisis sentimen, terutama dalam konteks *e-commerce*, yang dapat berkontribusi pada pengembangan metode analisis sentimen dan penelitian lebih lanjut di bidang ini.
5. Memberikan solusi dalam pengambilan keputusan bagi pemilik bisnis *e-commerce* dalam meningkatkan reputasi produk dan penjualan di lingkungan *e-commerce* yang semakin kompetitif.

## **G. Sistematika Penulisan**

Untuk memberikan mempermudah melihat dan mengetahui pembahasan yang ada pada skripsi ini, maka diperlukan sebuah gambaran berupa sistematika yang akan menjadi sebuah pedoman dan kerangka penulisan skripsi. Adapun sistematika penulisannya sebagai berikut :

### **BAB I      PENDAHULUAN**

Terdiri dari latar belakang masalah, identifikasi masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, sistematika penulisan.

### **BAB II    LANDASAN TEORI, PENELITIAN YANG RELEVAN DAN KERANGA BERFIKIR**

Bab ini berisi mengenai uraian landasan teori, penelitian yang relevan, kerangka berfikir, dan membahas berbagai teori-teori yang berkaitan dengan penulisan skripsi

### **BAB III METODE PENELITIAN**

Bab ini berisi mengenai waktu dan tempat penelitian, desain penelitian, metode pengumpulan data, langkah langkah pengembangan sistem

### **BAB IV ANALISIS SISTEM BERJALAN DAN RANCANGAN SISTEM YANG DIUSULKAN**

Bab ini berisi mengenai uraian profil perusahaan, struktur organisasi perusahaan, proses bisnis sistem berjalan, aturan bisnis sistem berjalan, dekomposisi fungsi sistem, analisis masukan(Input), proses dan keluaran(output) sistem berjalan, diagram aliran data (DAD) sistem berjalan (Diagram Konteks, Nol, Rinci), analisis permasalahan, alternatif penyelesaian masalah, aturan bisnis sistem yang diusulkan, dekomposisi fungsi sistem diusulkan, rancangan masukan, proses, keluaran, diagram aliran data (DAD) sistem yang diusulkan (Diagram Konteks, Nol, Rinci), Kamus data sistem diusulkan, spesifikasi proses sistem yang diusulkan, bagan terstruktur sistem diusulkan, spesifikasi modul yang diusulkan, perancangan basis data sistem yang diusulkan, rancangan layar, rancangan form

masukan data, dan rancangan keluaran, tampilan dan penjelasan layer, tampilan format masukan, dan tampilan keluaran

## **BAB V    SIMPULAN DAN SARAN**

Bab ini berisi mengenai simpulan dan sarandari penulis yang di peroleh dari penulisan, dan penyusunan skripsi ini agar pembaca dapat memperbaiki dan mengembangkan skripsi ini lebih baik lagi

## **BAB II**

# **LANDASAN TEORI, PENELITIAN YANG RELEVAN DAN KERANGA BERFIKIR**

### **A. Landasan Teori**

Bab ini berisi teori-teori yang berkaitan dengan analisis sentimen terhadap ulasan produk di *e-commerce* dengan metode *naive bayes*.

#### **1. Text Mining**

*Text mining* merupakan sebuah teknik atau metode yang digunakan untuk mengekstrak informasi yang berguna dari teks atau dokumen yang tersimpan dalam bentuk digital. Tujuan utama dari *text mining* yaitu untuk mengidentifikasi pola, tren, dan hubungan dalam data teks yang besar dan kompleks.

Allahyari et al. (2017) dalam kalusnya menyebutkan

*Text mining* atau *knowledge discovery from text* (KDT) - yang pertama kali diperkenalkan oleh Fledman dkk. - merujuk pada proses ekstraksi informasi berkualitas tinggi dari teks (baik yang terstruktur seperti data RDBMS, semi-terstruktur seperti XML dan JSON, maupun tidak terstruktur seperti dokumen kata, video, dan gambar).

Proses ini melibatkan penggunaan algoritma dan teknik analisis data untuk menggali informasi yang bernilai dari berbagai jenis sumber teks. Teks yang digunakan dalam *text mining* dapat berupa data yang sudah terstruktur, seperti data yang disimpan dalam database terkait, maupun data yang masih semi-terstruktur atau bahkan tidak terstruktur, seperti dokumen teks, video,

atau gambar. Proses *text mining* melibatkan ekstraksi informasi yang berarti dan bermanfaat dari teks tersebut, dengan menggunakan berbagai teknik dan metode analisis data, seperti *natural language processing* (NLP), ekstraksi fitur, atau analisis statistik.

*Text mining* dapat digunakan di berbagai bidang seperti bisnis, ilmu pengetahuan, pemasaran, dan bidang lainnya. Contohnya, pada bidang bisnis, teknik *text mining* bisa dimanfaatkan untuk menganalisis umpan balik pelanggan seperti ulasan produk atau komentar yang diberikan oleh pelanggan untuk mengidentifikasi tren atau pola yang relevan. Selain itu, pada bidang ilmu pengetahuan, *text mining* dapat digunakan untuk mengeksplorasi informasi dari teks ilmiah seperti artikel jurnal, guna menemukan pengetahuan baru atau mengungkapkan hubungan antara konsep yang berbeda.

Dengan demikian, *text mining* merupakan suatu pendekatan yang kuat untuk menggali informasi yang berguna dari berbagai jenis sumber teks, yang dapat digunakan dalam berbagai bidang untuk mengambil wawasan yang bernilai dan mendukung pengambilan keputusan yang lebih baik.

## **2. Machine Learning**

Dalam kulusanya Zhou (2021) menyatakan bahwa “*Machine learning* adalah cabang dari kecerdasan buatan yang berfokus pada pengembangan teknik dan algoritma yang memungkinkan komputer untuk belajar dan

meningkatkan kinerjanya secara otomatis melalui pengalaman.”. Tujuan utama dari *machine learning* yaitu memungkinkan komputer untuk belajar dari data input yang diberikan dan menghasilkan output yang tepat atau memprediksi output yang sesuai tanpa instruksi langsung yang spesifik.

*Machine learning* adalah teknologi yang memungkinkan komputer untuk mempelajari pola-pola dalam data dan meningkatkan kinerjanya seiring waktu. Pada *machine learning*, data input digunakan untuk melatih model atau algoritma mesin, yang kemudian dapat digunakan untuk memprediksi hasil dari data input baru yang belum pernah dilihat sebelumnya. Selama proses pelatihan model mesin, parameter atau bobot pada model dioptimalkan agar dapat menghasilkan prediksi yang lebih akurat pada data baru.

Terdapat beberapa teknik dalam *machine learning*, seperti *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. Dalam *supervised learning*, model mesin dilatih menggunakan data input dan output yang telah diberikan sebelumnya. Pada *unsupervised learning*, model mesin dilatih untuk menemukan pola-pola dalam data input tanpa menggunakan output yang telah diberikan. Sedangkan pada *reinforcement learning*, model mesin dilatih dengan memberikan penghargaan atau hukuman berdasarkan keputusan yang diambil, sehingga model mesin dapat mempelajari tindakan yang menghasilkan hasil yang optimal.

*Machine learning* sangat penting dalam berbagai aplikasi seperti deteksi penipuan, klasifikasi dokumen, dan bahkan pengenalan wajah. Dengan *machine learning*, komputer dapat belajar secara otomatis dan meningkatkan kinerjanya seiring waktu sehingga dapat menghasilkan hasil yang lebih akurat dan efisien.

### **3. Natural Language Processing**

Dalam Kalimatnya Linddy (2001) menjelaskan.

*Natural Language Processing* (NLP) adalah kumpulan teknik komputasi yang didasarkan pada teori untuk menganalisis dan merepresentasikan teks yang muncul secara alami pada satu atau beberapa level analisis bahasa dengan tujuan mencapai pemrosesan bahasa yang mirip dengan manusia untuk berbagai tugas atau aplikasi.

NLP bertujuan untuk memahami bahasa manusia yang rumit dan beraneka ragam dengan menggunakan algoritma dan teknik komputasi seperti statistik, pembelajaran mesin, dan pemrosesan bahasa alami (*natural language processing*). NLP memungkinkan mesin untuk menganalisis, memahami, dan menghasilkan bahasa manusia melalui pengolahan teks dan bahasa manusia secara otomatis.

NLP memiliki berbagai aplikasi, seperti terjemahan mesin, pengenalan suara, analisis sentimen, dan pengklasifikasian teks. Semua aplikasi tersebut memungkinkan komputer untuk memproses bahasa manusia dengan cara yang diinginkan oleh pengguna.

Untuk mencapai pemrosesan bahasa yang mirip dengan manusia diperlukan berbagai teknik seperti pengolahan bahasa alami (*natural language processing*), pemodelan bahasa, pengklasifikasian teks, dan *machine learning*. Sebagai contoh, dalam terjemahan mesin, NLP menggunakan teknik pengolahan bahasa alami untuk mengurai kalimat asal dan kemudian membangun kembali kalimat dalam bahasa yang ditargetkan. Dalam pengenalan suara, NLP menggunakan teknik pembelajaran mesin untuk mengenali dan mengonversi suara menjadi teks.

NLP telah menjadi teknologi penting dalam berbagai bidang, termasuk industri, pendidikan, dan kesehatan. Misalnya, NLP digunakan dalam industri untuk menganalisis data pelanggan dan membuat rekomendasi produk. Di bidang pendidikan, NLP digunakan untuk mengembangkan perangkat lunak pembelajaran bahasa asing. Dan di bidang kesehatan, NLP digunakan untuk mengidentifikasi gejala penyakit dari catatan medis pasien. Dalam kesimpulannya, NLP adalah kumpulan teknik komputasi yang didasarkan pada teori untuk menganalisis dan merepresentasikan teks yang muncul secara alami pada satu atau beberapa level analisis bahasa dengan tujuan mencapai pemrosesan bahasa yang mirip dengan manusia untuk berbagai tugas atau aplikasi.

#### **4. Analisis Sentimen**



“*Sentiment analysis* atau analisis sentimen adalah jenis klasifikasi teks yang mengkategorikan teks berdasarkan orientasi sentimen dari opini yang terkandung dalam teks tersebut.” (Devika et al., 2016) Analisis sentimen bertujuan untuk mengidentifikasi perasaan, pandangan, atau evaluasi yang dinyatakan oleh penulis atau pengguna teks, seperti ulasan produk, komentar media sosial, atau feedback pelanggan.

Dengan menggunakan teknik klasifikasi, analisis sentimen mengidentifikasi dan mengkategorikan teks ke dalam kategori sentimen positif, negatif, atau netral berdasarkan isi opini yang terdapat dalam teks.

Dalam analisis sentimen, data teks biasanya diproses menggunakan metode-metode pemrosesan bahasa alami dan teknik pembelajaran mesin untuk mengenali pola dan ciri-ciri kalimat yang mengindikasikan sentimen.

Beberapa metode populer yang digunakan dalam analisis sentimen antara lain metode berbasis aturan, metode berbasis kata kunci, dan metode berbasis mesin pembelajaran, seperti metode *Naive bayes*, *Support Vector Machine* (SVM), dan Jaringan Saraf Tiruan (*Artificial Neural Network/ANN*).

## **5. Text Pre-processing**

*Text Pre-processing* adalah teknik *data mining* yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. “Metode *Pre-processing* memainkan peran yang sangat penting dalam teknik dan

aplikasi *text mining*. Metode ini merupakan langkah pertama dalam proses *text mining*.” (Vijayarani et al., 2015) Tujuan dari *Pre-processing* adalah untuk membersihkan, mengatur, dan mempersiapkan teks agar dapat diolah lebih lanjut dengan akurasi dan efisiensi yang optimal.

Beberapa langkah umum dalam *Pre-processing* teks meliputi penghapusan tanda baca, penghilangan karakter khusus, konversi teks menjadi *lowercase* atau *uppercase*, penghilangan kata-kata yang tidak berarti (*stop words*), penggantian sinonim, pengelompokan kata (*stemming* atau *lemmatization*), serta penghapusan informasi yang tidak relevan atau *noise*. Metode *Pre-processing* yang baik dapat mempengaruhi kualitas dan akurasi analisis *text mining* yang dilakukan selanjutnya.

## 6. Naïve Bayes

Metode *Naive bayes* adalah sebuah algoritma dalam pembelajaran mesin yang digunakan untuk melakukan klasifikasi data ke dalam kategori atau kelas tertentu. “Klasifikasi dengan metode Naïve Bayes secara umum dilakukan dengan pendekatan peluang atau probabilitas. Algoritma ini memprediksi probabilitas di masa depan berdasarkan pengalaman yang sudah ada di masa lalu.” (Sipayung et al., 2016)

Asumsi utama dari metode *Naive bayes* adalah bahwa setiap fitur atau atribut dalam data adalah independen satu sama lain, artinya tidak ada

ketergantungan atau interaksi antara fitur-fitur tersebut. Meskipun dalam kenyataannya hal ini mungkin tidak selalu benar, namun asumsi ini digunakan untuk mempermudah perhitungan dan membuat algoritma menjadi sederhana dan cepat.

Metode naïve bayes berdasarkan teorema bayes memiliki bentuk umum serbagai berikut:

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)}$$

Keterangan:

$P(H|X)$  = Probabilitas hipotesis H pada kondisi X

$P(H)$  = Probabilitas hipotesis H

$P(X|H)$  = Probabilitas X pada kondisi H

$P(X)$  = Probabilitas X

Dalam analisis sentiment, naïve bayes disesuaikan dengan memberikan petunjuk-petunjuk untuk menentukan probabilitas sentiment. Penyesuaian berikut sebagai berikut:

$$P(X|F1 \dots Fn) = \frac{P(F1 \dots Fn|X) \times P(X)}{P(F1 \dots Fn)}$$

Keterangan:

$P(X|F1 \dots Fn)$  = Probabilitas kelas X pada fitur  $F1 \dots Fn$

$P(X)$  = Probabilitas kelas X

$P(F_1 \dots F_n)$  = Probabilitas fitur fitur untuk klasifikasi

$P(F_1 \dots F_n | X)$  = Probabilitas fitur  $F_1 \dots F_n$  pada kelas  $X$

## 7. Evaluasi Kinerja Metode

Dalam klausanya Tharwat (2021) menjelaskan.

Metode klasifikasi telah diterapkan dalam banyak aplikasi di berbagai bidang ilmu. Ada beberapa cara untuk mengevaluasi algoritma klasifikasi. Analisis metrik-metrik tersebut dan signifikansinya harus diinterpretasikan dengan benar untuk mengevaluasi algoritma pembelajaran yang berbeda. Sebagian besar metrik ini adalah metrik skalar dan beberapa di antaranya adalah metode grafis.

Terdapat banyak Teknik klasifikasi yang telah di terapkan dalam banyak aplikasi dalam berbagai bidang ilmu, termasuk Naïve bayes yang umum digunakan dalam analisis sentiment. Setelah mengimplementasikan metode naïve bayes untuk melakukan klasifikasi sentiment, Langkah selanjutnya yakni melakukan evaluasi kinerja dari model yang telah dibangun

“Hasil olahan kasifikasi dapat dievaluasi untuk mengetahui tingkat akurasi dan efisiensi, beberapa metrik yang perlu diketahui dalam evaluasi model yaitu akurasi, *confussion matrix*, f1-score, presisi, *recall*” (Tharwat, 2021).

Metrik akurasi adalah salah satu metrik evaluasi yang umum digunakan dalam pembelajaran mesin untuk mengukur seberapa akurat suatu model dalam melakukan prediksi kelas atau kategori dari data yang diberikan. Metrik akurasi dihitung dengan membandingkan jumlah prediksi yang

benar (*True Positive* dan *True Negative*) dengan jumlah total data yang dievaluasi.

Matriks *confusion* atau *confusion matrix*, adalah suatu tabel yang digunakan untuk mengevaluasi performa dari suatu model klasifikasi dengan cara membandingkan prediksi model dengan nilai yang sebenarnya dari data yang dievaluasi. Matriks confusion biasanya digunakan untuk masalah klasifikasi dua kelas (*binary*), di mana model klasifikasi memprediksi data menjadi dua kategori yang mungkin, yaitu positif dan negatif (atau benar dan salah). F1-score dan confusion matrix (metrik confusion) saling berhubungan dalam evaluasi performa suatu model klasifikasi. Confusion matrix umumnya terdiri dari empat sel, yaitu *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN).

		Actual Condition		
		FALSE	TRUE	
Predicted Condition	FALSE	TN	FN	Predicted Negative
	TRUE	FP	TP	Predicted Positive
		Actual Negative	Actual Positive	

**Gambar 1.1 Confussion Matrix**

Sumber: <https://benhay.es/posts/demystifying-confusion-matrix/>

Presisi (*precision*) merupakan perbandingan antara jumlah prediksi yang benar positif (*true positive*) dengan jumlah prediksi positif (*true positive* dan *false positive*). Rumus presisi adalah:

$$Presisi = TP \frac{TP}{TP + FP}$$

*Recall*, juga dikenal sebagai sensitivitas atau *true positive rate* (TPR), merupakan perbandingan antara jumlah prediksi yang benar positif (*true positive*) dengan jumlah data positif sebenarnya (*true positive* dan *false negative*). Rumus *recall* adalah:

$$Recall = \frac{TP}{TP + FN}$$

*F1-score* dihitung sebagai harmonic mean antara presisi dan recall. Harmonic mean merupakan suatu jenis rata-rata yang memberikan bobot yang seimbang antara dua nilai yang dihitung. Rumus *F1-score* adalah:

$$F1 - score = 2 * \frac{(Presisi * Recall)}{(Presisi + Recall)}$$

*F1-score* memiliki rentang nilai antara 0 dan 1, dimana nilai 1 menunjukkan performa yang sempurna (presisi dan *recall* yang optimal) dan nilai 0 menunjukkan performa yang buruk (salah satu atau keduanya dari presisi dan *recall* sangat rendah). *F1-score* digunakan untuk menggabungkan presisi dan *recall* menjadi suatu metrik yang tunggal, sehingga dapat memberikan gambaran yang lebih komprehensif tentang performa keseluruhan suatu model klasifikasi.

## 8. Term Frequency – Inverse Document Frequency

TF-IDF merupakan singkatan dari *Term Frequency-Inverse Document Frequency*, yang merupakan teknik pemodelan teks yang digunakan dalam pemrosesan bahasa alami. Tujuan dari teknik ini adalah untuk mengevaluasi seberapa penting suatu kata dalam suatu dokumen atau kumpulan dokumen. Yun-tao et al. (2005) dalam kalsunya menyebutkan.

*Term Frequency-Inverse Document Frequency* (TF-IDF) umumnya digunakan untuk memberi bobot pada setiap kata dalam dokumen teks berdasarkan seberapa unik kata tersebut. Dengan kata lain, pendekatan TF-IDF dapat menangkap relevansi antara kata-kata, dokumen teks, dan kategori tertentu.

Secara sederhana, TF-IDF memperhitungkan dua hal yaitu frekuensi kata dalam dokumen (TF) dan frekuensi kata dalam kumpulan dokumen (IDF). Frekuensi kata dalam dokumen (TF) menghitung seberapa sering suatu kata muncul dalam dokumen tersebut, sementara frekuensi kata dalam kumpulan dokumen (IDF) menghitung seberapa umum kata tersebut dalam kumpulan dokumen. IDF dihitung dengan menghitung logaritma dari jumlah dokumen dalam kumpulan dokumen yang mengandung kata tersebut dibagi dengan total dokumen dalam kumpulan dokumen.

Dengan menggunakan kedua nilai ini, TF-IDF memberikan skor untuk setiap kata dalam suatu dokumen, di mana kata-kata yang muncul lebih sering dalam dokumen tersebut dan lebih jarang muncul dalam kumpulan dokumen lain akan memiliki skor yang lebih tinggi. Dalam hal ini, kata-

kata yang memiliki skor TF-IDF yang lebih tinggi dianggap lebih penting dalam dokumen tersebut.

Teknik TF-IDF sering digunakan dalam berbagai aplikasi pemrosesan bahasa alami seperti klasifikasi dokumen, pengambilan informasi, analisis sentimen, dan pencarian informasi.

## **9. Google Colab**

Dijelaskan oleh Google (2023) dalam *frequently asked questions*.

Colaboratory, atau disingkat "Colab", adalah produk dari Google Research. Colab memungkinkan siapa saja untuk menulis dan menjalankan kode python secara acak melalui browser, dan sangat cocok untuk pembelajaran mesin, analisis data, dan pendidikan. Secara teknis, Colab adalah layanan notebook Jupyter yang di-hosting yang tidak memerlukan pengaturan untuk digunakan, sambil memberikan akses gratis ke sumber daya komputasi termasuk GPU.

Google colab banyak digunakan karena penggunaannya mudah tanpa memerlukan instalasi atau pengaturan yang rumit, hanya perlu memiliki akun Google. Namun, penggunaan sumber daya komputasi pada Google Colab terbatas dan hanya tersedia dalam jumlah terbatas, sehingga pengguna perlu membayar untuk menggunakan sumber daya yang lebih besar.

## **10. E-commerce**

Dalam Wikipedia (2023) disebutkan.

*E-commerce* (electronic commerce) adalah kegiatan pembelian atau penjualan produk secara elektronik melalui layanan online atau Internet.



*E-commerce* menggunakan teknologi seperti mobile commerce, transfer dana elektronik, manajemen rantai pasokan, pemasaran internet, pemrosesan transaksi online, pertukaran data elektronik (EDI), sistem manajemen inventaris, dan sistem pengumpulan data otomatis. *E-commerce* didorong oleh kemajuan teknologi industri semikonduktor, dan merupakan sektor terbesar dalam industri elektronik.

*E-commerce* memberikan kemudahan bagi konsumen untuk membeli produk dari berbagai negara atau wilayah tanpa harus keluar rumah. Selain itu, para pelaku bisnis juga mendapatkan manfaat dari *e-commerce* karena dapat menjangkau pasar yang lebih luas dan memperluas jaringan distribusi produk mereka. Seiring dengan pertumbuhan teknologi dan internet yang semakin pesat, *e-commerce* diprediksi akan terus berkembang dan menjadi salah satu bentuk perdagangan yang dominan di masa depan.

#### **a. Ulasan Produk**

Ulasan produk merupakan pendapat atau komentar yang diberikan oleh pelanggan atau pengguna produk tertentu setelah mereka menggunakannya. Ulasan produk dapat berisi pengalaman pribadi, penilaian, kritik, dan saran untuk produk tersebut.

*E-commerce* juga memiliki fitur yang memungkinkan konsumen memberikan ulasan atau penilaian produk yang telah mereka beli. Ulasan produk di *e-commerce* menjadi sangat penting karena dapat memengaruhi keputusan pembelian calon konsumen. Konsumen seringkali mempercayai ulasan dari pengguna sebelumnya untuk

menilai kualitas, keandalan, dan kepuasan produk yang ditawarkan. Oleh karena itu, analisis sentimen terhadap ulasan produk di *e-commerce* menjadi penting bagi penjual atau pedagang dalam memahami bagaimana konsumen merespons produk mereka.

## **B. Penelitian Yang Relevan**

### **1. Skripsi**

Nama	: Tata Ilham Rais
Universitas	: Universitas Islam Negeri Syarif Hidayatullah Jakarta
Tahun	: 2023
Judul	: Analisis sentimen terhadap komentar video youtube Raiden Shogun - Judgment of Euthymia menggunakan metode majority voting
Masalah	: Bagaimana analisis sentiment terhadap komentar dalam video youtube dapat dilakukan menggunakan metode majority voting?
Tujuan Penelitian	: Melakukan suatu analisis sentiment yang dapat menghasilkan gambaran umum terkait hal-hal yang disukai penonton dari video “Raiden Shogun – Judgement of Euthymia” yang dapat menjadi

masuk untuk pembuatan video atau desain karakter selanjutnya, serta menguji akurasi dari model analisis sentiment majority voting

## 2. Jurnal

Nama Jurnal	: Jurnal Sistem Informasi (JSI)
Nama Penulis	: Evasaria M. Sipayung, Herastia Maharani, Ivan Zefanya
Judul	: Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naïve Bayes Classifier
E-ISSN	: 2355-4614
Volume	: Vol. 8, No. 1, April 2016
Hasil	: Perhitungan kehandalan sistem analisis sentimen dengan metode <i>naïve bayes classifier</i> ini dari 175 data latih dapat dibagi menjadi dua bagian, perhitungan klasifikasi kategori dan perhitungan klasifikasi sentimen, untuk hasil dari akurasi sistem memiliki nilai <i>accuration</i> kategori 77.14% dan untuk

*precision* sentimen 99.12%, *recall* sentimen 72.9%,  
dan *accuration* sentimen 75.42%, sehingga sistem  
mampu mengembalikan dokumen dan kecocokan  
data yang tinggi.

## BAB III

### METODOLOGI PENELITIAN

#### A. Waktu dan Tempat Penelitian

##### 1. Waktu Penelitian

Waktu yang digunakan peneliti untuk penelitian ini dilaksanakan sejak tanggal dikeluarkannya ijin penelitian. Berikut dibawah ini adalah tabel perincian tahapan penelitian berdasarkan waktu dan kegiatan

**Table 3.1 Jadwal Kegiatan Penelitian**

No	Aktifitas	Maret				April				Mei				Juni				Juli			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Tahap Perancangan																				
2.	Pengumpulan Data																				
3.	Tahap Pengolahan Data																				
4.	Analisis dan Interpretasi Data																				
5.	Penulisan Laporan																				

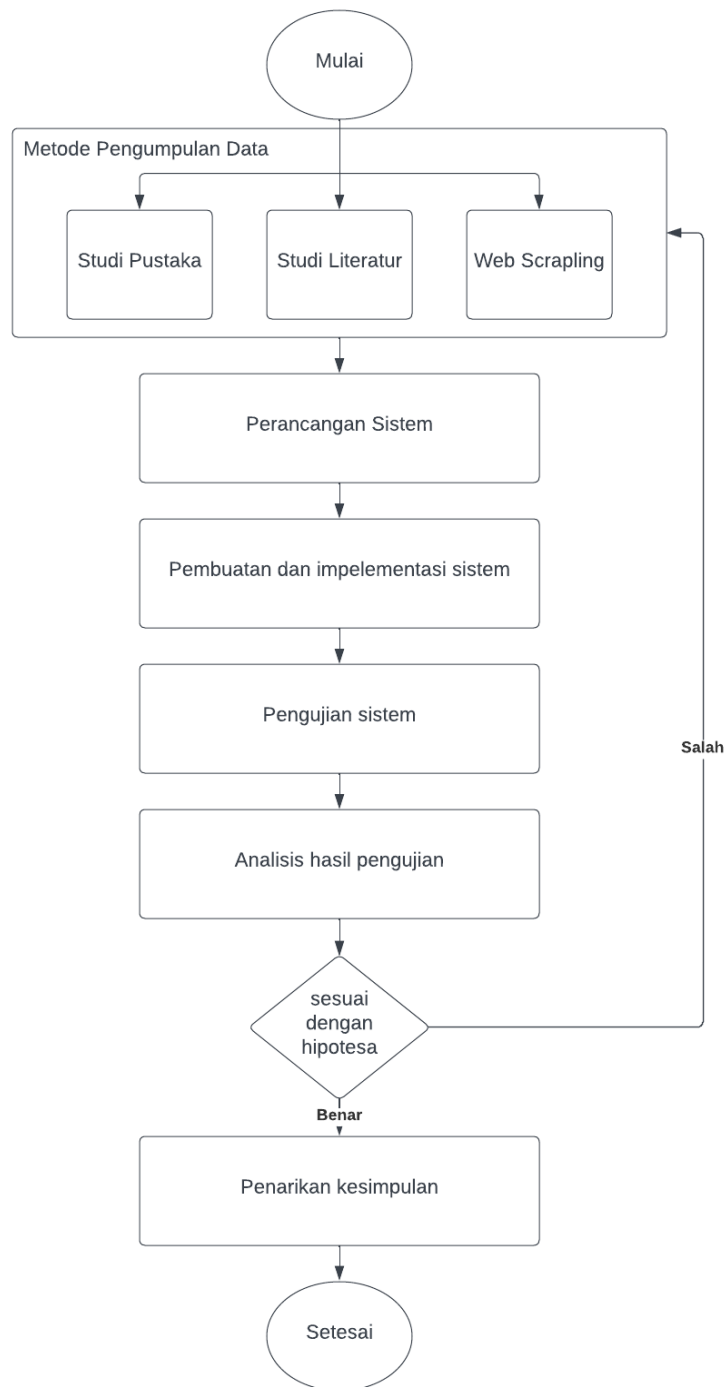
Sumber: Penulis

## **2. Tempat Penelitian**

Dalam penulisan ini penulis menjadikan *E-commerce* Tokopedia sebagai tempat penelitian. Tokopedia adalah salah satu ecommerce terbesar di Indonesia yang memiliki banyak ulasan produk dari berbagai kategori seperti fashion, elektronik, makanan, dan banyak lagi. Sebagai tempat penelitian, penulis memilih Tokopedia karena populernya di kalangan masyarakat Indonesia dan banyaknya produk serta ulasan yang tersedia di platform ini. Selain itu, Tokopedia memiliki fitur yang memungkinkan pengguna untuk memberikan penilaian atau ulasan terhadap produk yang telah dibeli, sehingga penulis dapat dengan mudah mengambil sampel ulasan produk dari platform ini.

## **B. Tahapan Penelitian**

Tahapan penelitian merupakan tahapan yang dilakukan untuk melakukan penelitian mulai dari awal hingga selesai. Penelitian yang dilakukan sesuai dengan tahapan sebagai berikut:



**Gambar 3.1 Tahapan Penelitian**

Sumber: penulis

Berikut adalah penjelasan untuk tahapan penelitian dari tahapan-tahapan yang diberikan:

1. Tahapan pengumpulan data melalui studi pustaka, studi literatur, web scraping

Tahap awal penelitian adalah pengumpulan data melalui studi pustaka dan literatur. Dalam hal ini, penulis akan mengumpulkan informasi tentang teori dan konsep terkait analisis sentimen dan metode Naive Bayes. Selain itu, penulis juga dapat menggunakan web scraping untuk mengambil data ulasan produk dari platform ecommerce seperti Tokopedia, Shopee, atau Lazada. Web scraping merupakan teknik pengambilan data dari halaman web yang dilakukan secara otomatis dengan menggunakan program komputer.

2. Perancangan sistem

Tahapan selanjutnya adalah perancangan sistem yang akan digunakan untuk melakukan analisis sentimen terhadap ulasan produk. Penulis akan merancang sistem yang terdiri dari beberapa komponen seperti proses *Pre-processing* data, proses pelatihan dan pengujian model Naive Bayes, dan proses klasifikasi sentimen.

3. Pembuatan dan implementasi sistem

Setelah perancangan sistem selesai dilakukan, penulis kemudian akan memulai tahap pembuatan dan implementasi sistem. Dalam tahap ini,



penulis akan membuat program komputer atau aplikasi yang dapat melakukan analisis sentimen terhadap ulasan produk menggunakan metode Naive Bayes yang telah dirancang sebelumnya.

#### 4. Pelatihan model Naive Bayes

Tahapan berikutnya adalah pelatihan model Naive Bayes menggunakan data pelatihan yang telah dikumpulkan sebelumnya. Data pelatihan akan digunakan untuk melatih model Naive Bayes agar dapat mengklasifikasikan sentimen dari ulasan produk dengan akurasi yang tinggi.

#### 5. Analisis hasil pengujian

Setelah model Naive Bayes terlatih, tahap selanjutnya adalah melakukan pengujian pada data pengujian. Data pengujian akan digunakan untuk menguji akurasi model Naive Bayes dalam mengklasifikasikan sentimen dari ulasan produk. Hasil pengujian akan dianalisis untuk mengetahui seberapa baik model Naive Bayes dalam melakukan analisis sentimen terhadap ulasan produk.

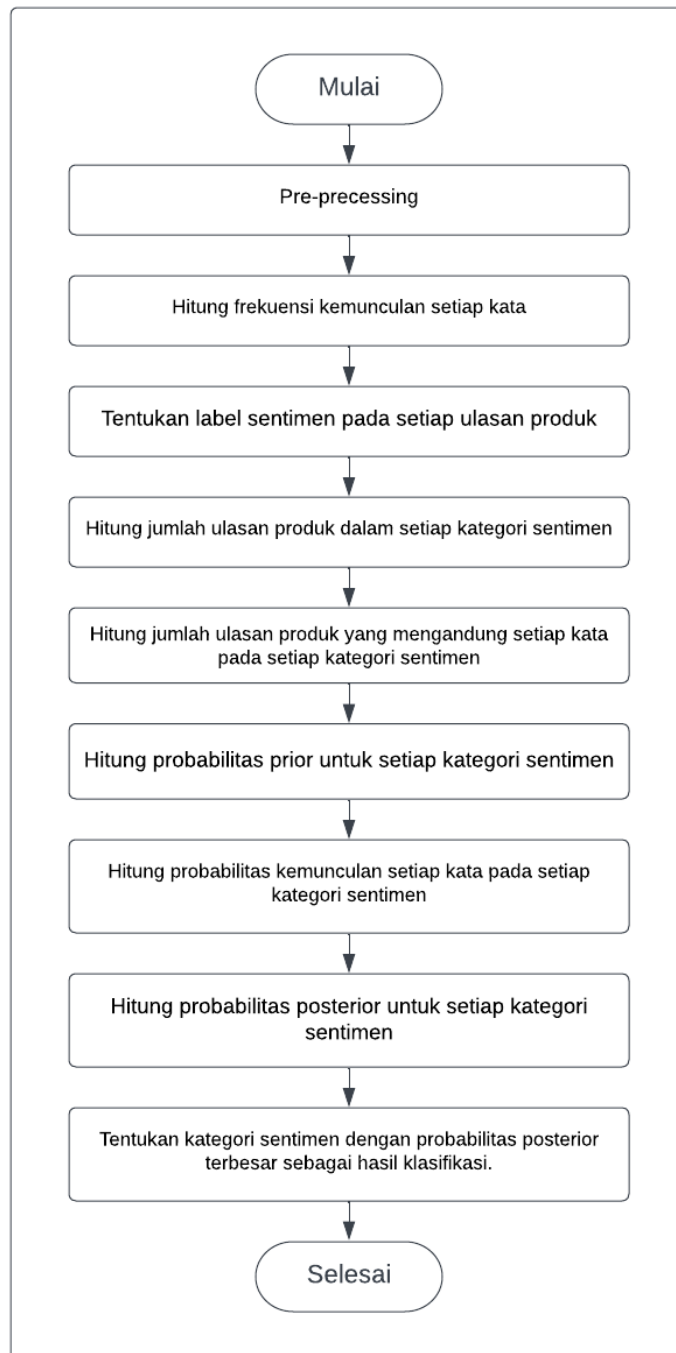
#### 6. Penarikan kesimpulan

Tahap terakhir adalah penarikan kesimpulan dari hasil penelitian yang telah dilakukan. Penulis akan melakukan analisis terhadap hasil pengujian dan menarik kesimpulan mengenai efektivitas dan kegunaan model Naive Bayes dalam melakukan analisis sentimen terhadap ulasan produk. Selain

itu, penulis juga dapat memberikan rekomendasi untuk pengembangan penelitian selanjutnya.

### **C. Algoritma**

Algoritma Naive Bayes merupakan salah satu metode klasifikasi yang dapat digunakan untuk melakukan analisis sentimen pada data ulasan produk. Algoritma ini menghitung probabilitas setiap kategori sentimen berdasarkan kemunculan kata-kata pada ulasan yang dianalisis. Berikut adalah langkah-langkah algoritma *Naive Bayes* dalam analisis sentiment terhadap ulasan produk di *E-Commerce*:



**Gambar 3.2 Diagram Kerja Algoritma**

Sumber: Penulis

Diagram kerja di atas menunjukkan bahwa algoritma Naive Bayes untuk analisis sentimen dimulai dengan membersihkan setiap ulasan produk dari tanda baca, kata-kata yang tidak relevan, dan merubah kata-kata menjadi bentuk dasarnya. Kemudian, algoritma menghitung frekuensi kemunculan setiap kata pada setiap ulasan produk dan menentukan label sentimen (positif, negatif, atau netral) pada setiap ulasan produk.

Selanjutnya, algoritma menghitung jumlah ulasan produk pada setiap kategori sentimen dan jumlah ulasan produk yang mengandung setiap kata pada setiap kategori sentimen. Dari sini, algoritma menghitung probabilitas prior dan probabilitas kemunculan setiap kata pada setiap kategori sentimen.

Setelah itu, algoritma menghitung probabilitas posterior untuk setiap kategori sentimen dan menentukan kategori sentimen dengan probabilitas posterior terbesar sebagai hasil klasifikasi. Terakhir, algoritma menghasilkan hasil klasifikasi sentimen pada setiap ulasan produk.

## **BAB IV**

### **PEMBAHASAN**

## Referensi

- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *ArXiv Preprint ArXiv:1707.02919*.
- Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: a comparative study on different approaches. *Procedia Computer Science*, 87, 44–49.
- Google. (2023). *What is Colaboratory?*  
<https://research.google.com/colaboratory/faq.html>
- Liddy, E. D. (2001). *Natural language processing*.
- Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *JSI: Jurnal Sistem Informasi (E-Journal)*, 8(1).
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). *Pre-processing* techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.
- Wikipedia. (2023). *E-commerce*. <https://en.wikipedia.org/wiki/E-commerce>
- Yun-tao, Z., Ling, G., & Yong-cheng, W. (2005). An improved TF-IDF approach for text classification. *Journal of Zhejiang University-Science A*, 6, 49–55.
- Zhou, Z.-H. (2021). *Machine learning*. Springer Nature.