

## **ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER**

**Ernianti Hasibuan<sup>a</sup>, Elmo Allistair Heriyanto<sup>b</sup>**

<sup>a</sup> Fakultas Ilmu Komputer dan Teknologi Informasi/ Sistem Informasi, [ernianti@staff.gunadarma.ac.id](mailto:ernianti@staff.gunadarma.ac.id), Universitas Gunadarma

<sup>b</sup> Fakultas Ilmu Komputer dan Teknologi Informasi/ Sistem Informasi, [work.elmoallistair@gmail.com](mailto:work.elmoallistair@gmail.com), Universitas Gunadarma

### **ABSTRACT**

Sentiment analysis or opinion mining is a study that analyzes people's opinions, thoughts and impressions on various topics, subjects, and products or services. The development of social media makes public opinion data available which can be found easily on the internet. The large volume of data causes the need for an automatic system to classify the data based on different aspects because classifying data manually is a time-consuming process. In this study, sentiment analysis will be carried out with a machine learning-based approach using the Naive Bayes algorithm using user review data on the Amazon Shopping application on the Google Play Store. The classification results using the four Naive Bayes algorithms produce an average accuracy of 82.15%, precision of 72.25%, recall of 83.49%, and f1-score of 77.41%. Multinomial NB produces the best accuracy among the four Naive Bayes algorithms used, which is 86.74%. The values of precision, recall, and f1-score are 78.82%, 85.90%, and 82.21%, respectively.

**Keywords:** sentiment analysis, amazon shopping, machine learning, naive bayes classifier

### **ABSTRAK**

Analisis sentimen atau penggalian opini adalah studi yang menganalisis pendapat, pemikiran, dan kesan orang mengenai berbagai topik, subjek, dan produk atau layanan. Perkembangan sosial media membuat tersedianya data-data opini publik tersebut yang dapat ditemukan dengan mudah di internet. Besarnya volume data tersebut menyebabkan adanya kebutuhan akan sistem otomatis untuk mengklasifikasi data tersebut berdasarkan aspek yang berbeda dikarenakan pengklasifikasian data secara manual merupakan proses yang memakan waktu. Pada penelitian ini akan dilakukan analisis sentimen dengan pendekatan berbasis machine learning menggunakan algoritma Naive Bayes menggunakan data ulasan pengguna pada aplikasi Amazon Shopping yang terdapat pada Google Play Store. Hasil klasifikasi menggunakan keempat algoritma Naive Bayes menghasilkan akurasi rata-rata sebesar 82.15%, precision sebesar 72.25%, recall sebesar 83.49%, dan f1-score sebesar 77.41%. Multinomial NB menghasilkan akurasi terbaik diantara keempat algoritma Naive Bayes yang digunakan, yaitu sebesar 86.74%. Nilai precision, recall, dan f1-score berturut-turut adalah 78.82%, 85.90%, dan 82.21%.

**Kata Kunci:** analisis sentimen, amazon shopping, machine learning, naive bayes classifier..

### **1. PENDAHULUAN**

Analisis sentimen atau penggalian opini adalah studi yang menganalisis pendapat, pemikiran, dan kesan orang mengenai berbagai topik, subjek, dan produk atau layanan. Yang bertujuan untuk mengotomatisasi tugas mengidentifikasi opini dan sentimen yang orang-orang ungkapkan, dan kemudian mengklasifikasikan polaritas sentimennya [1]. Sejak awal tahun 2000, analisis sentimen telah berkembang menjadi salah satu area penelitian paling aktif dalam Natural Language Processing [2], aktivitas industri seputar analisis sentimen juga berkembang secara pesat. Pengimplementasian analisis sentimen telah menyebar ke banyak domain, mulai dari bisnis, layanan kesehatan, pemilihan umum, analisis produk, hingga riset pasar.

Ulasan merupakan salah satu bentuk penilaian seseorang terhadap suatu produk atau jasa. Analisis sentimen dapat membantu kita untuk mengetahui opini pelanggan melalui ulasan yang mereka tuliskan

---

mengenai suatu produk atau jasa. Ulasan tersebut dapat menjadi informasi berharga untuk pelanggan lain. Misalnya, sebelum membeli suatu produk, kebanyakan orang cenderung mencari melalui ulasan tentang produk tersebut yang akan membantu mereka dalam membuat pilihan. Berdasarkan survey yang dilakukan oleh brightlocal, 79% konsumen mengatakan mereka mempercayai ulasan online lebih dari rekomendasi pribadi dari teman atau keluarga[3].

Terdapat beberapa pendekatan yang umum digunakan untuk melakukan analisis sentimen, antara lain dengan menggunakan machine learning, berbasis leksikon, hybrid dan teknik lainnya seperti aspect-based approach, transfer learning, dan multimodal sentiment analysis[4]. Pada penelitian ini akan dilakukan klasifikasi sentimen dengan pendekatan berbasis machine learning menggunakan algoritma Naive Bayes menggunakan data ulasan pengguna pada aplikasi Amazon Shopping yang terdapat pada Google Play Store. Pendekatan menggunakan machine learning dipilih karena kemampuannya untuk melakukan analisa dengan cepat, sehingga dapat mengurangi waktu dan upaya untuk mengklasifikasi ulasan. Adapun tujuan yang ingin dicapai adalah mengetahui performa yang dihasilkan algoritma Naive Bayes dalam melakukan klasifikasi sentimen.

Beberapa penelitian serupa telah dilakukan sebelumnya, diantaranya adalah melakukan analisa sentimen menggunakan Naive Bayes untuk mengkategorikan ulasan positif dan negatif dari pelanggan Amazon atas kategori produk yang berbeda seperti ponsel, aksesoris, alat musik dan barang elektronik[5]. Penelitian tersebut mampu menghasilkan akurasi rata-rata di atas 90% dengan mencoba beberapa simulasi yang berbeda seperti cross validation, rasio data latih-uji, dan teknik ekstraksi fitur yang berbeda. Selanjutnya pada penelitian lainnya [6], dibangun sebuah sistem yang dapat mengklasifikasikan opini ulasan produk menjadi sentimen positif dan negatif dengan memanfaatkan rating. Dataset yang digunakan adalah data ulasan bahan makanan and makanan gourmet dari Amazon sebanyak 50.000 data. Hasil klasifikasi sistem dengan menerapkan metode labelling average menggunakan algoritma Multinomial Naive Bayes mampu mengklasifikasikan sentimen dengan akurasi sebesar 80,48%.

## 2. TINJAUAN PUSTAKA

### 2.1. *Machine Learning dan Natural Language Processing*

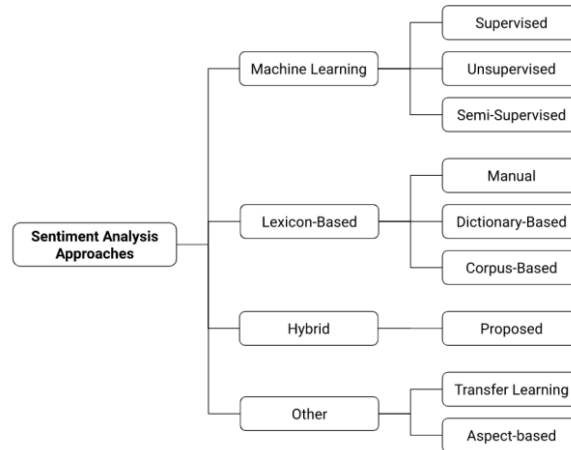
*Machine learning* dapat didefinisikan secara luas sebagai metode komputasi yang menggunakan pengalaman untuk meningkatkan kinerja atau membuat prediksi yang akurat [7]. Pengalaman tersebut mengacu pada informasi yang biasanya berbentuk data elektronik yang dikumpulkan dan tersedia untuk analisis. Data ini bisa dalam bentuk digitalisasi perangkat pelatihan yang sudah berlabel, atau jenis informasi lain yang diperoleh melalui interaksi dengan lingkungan. *Machine learning* merupakan bagian dari disiplin ilmu kecerdasan bagian (AI) dapat diaplikasikan dengan sangat luas, beberapa diantaranya adalah klasifikasi teks atau dokumen, *natural language processing*, *speech processing*, *computer vision*, dan lain-lain. Tujuan praktis utama dari *machine learning* adalah menghasilkan prediksi yang akurat dan merancang algoritma yang efisien dan kuat untuk menghasilkan prediksi tersebut bahkan untuk permasalahan skala besar.

Natural Language Processing (NLP) merupakan pemrosesan komunikasi manusia berbasis mesin yang bertujuan untuk mengajarkan mesin bagaimana cara memproses dan memahami bahasa manusia, sehingga memungkinkan saluran komunikasi yang mudah antara manusia dan mesin [8]. NLP telah banyak diterapkan di berbagai area, salah satunya adalah analisis sentimen, analisis sosial media, *chatbots*, penerjemah, ekstraksi informasi dan lain-lain. NLP dan *machine learning* merupakan disiplin ilmu yang saling terkait. Dalam banyak kasus, output dari teknik NLP seperti *text preprocessing* digunakan sebagai input untuk *machine learning*, dan sebaliknya seperti menerapkan *supervised machine learning* ke konstruksi kamus untuk mengidentifikasi *entity* dengan NLP. Aplikasi terkait NLP dibangun menggunakan sejumlah besar data yang disebut corpus. Corpus didefinisikan sebagai kumpulan materi bahasa alami tertulis atau lisan, disimpan di dalam komputer, dan digunakan untuk mengetahui bagaimana bahasa digunakan [9]. Beberapa manfaat penggunaan corpus antara lain membantu beberapa analisis statistik seperti distribusi frekuensi dan memvalidasi aturan linguistik seperti menemukan tata bahasa (*grammar*) yang salah. Aplikasi NLP terkadang menggunakan satu corpus sebagai masukan, atau menggunakan lebih dari satu corpus yang disebut corpora.

Sedangkan *text mining*, juga disebut *text analytics*, adalah teknik yang mengubah data tidak terstruktur menjadi data terstruktur menggunakan bantuan NLP untuk menyempurnakan hasil analisis yang dilakukan menggunakan algoritma *machine learning*[10]. *Text mining* mengidentifikasi fakta, hubungan, dan pernyataan dalam kumpulan data tekstual yang besar. Informasi yang diekstrak dapat diubah menjadi bentuk terstruktur yang dapat dianalisis lebih lanjut. Penerapan dari *text mining* yang paling umum adalah analisa sentimen, *text categorization*, *text clustering*, dan *document summarization*.

## 2.2. Analisis Sentimen

Analisis sentimen atau juga disebut penggalian opini merupakan bidang studi yang menganalisis pendapat, sentimen, penilaian, sikap, dan emosi seseorang terhadap suatu entitas yang dapat berupa produk, layanan, organisasi, individu, acara, masalah, atau topik yang diungkapkan dalam teks tertulis [2]. Analisis sentimen telah memperluas penelitian NLP secara signifikan karena telah memperkenalkan banyak masalah penelitian yang menantang yang belum pernah dipelajari sebelumnya. Analisis sentimen merupakan masalah analisis semantik, tetapi sangat terfokus dan terbatas karena sistem analisis sentimen tidak perlu sepenuhnya "memahami" setiap kalimat atau dokumen dan hanya perlu memahami beberapa aspek saja, misalnya opini positif dan negatif serta sarasannya. Terdapat beberapa pendekatan untuk melakukan analisis sentimen seperti pada Gambar 1. Pendekatan yang cukup populer dan banyak digunakan adalah *Machine Learning*.



Gambar 1. Pendekatan Analisis Sentimen [11]

## 2.3. Ekstraksi Fitur

Ekstraksi fitur adalah proses pengurangan dimensi di mana kumpulan awal data mentah direduksi menjadi kelompok yang lebih mudah dikelola untuk diproses. Ekstraksi fitur mencoba untuk mengekstrak satu set fitur (dimensi) yang sama sekali baru dari pola data daripada memilih fitur dari atribut yang ada [12]. Model *machine learning* cenderung hanya memahami data numerik. Oleh karena itu, data teks perlu diubah menjadi bentuk numerik. Dalam kasus ini, data ulasan yang dibersihkan akan direpresentasikan menjadi bentuk numerik menggunakan metode *Bag-of-Words* dan TF-IDF. Pembuatan *features* dalam teks dapat dilakukan menggunakan metode *Bag-of-Words* dan TF-IDF. *Bag-of-Words* merupakan representasi yang mengubah teks menjadi vektor dengan panjang tetap dengan menghitung berapa kali setiap kata muncul dalam dokumen, proses ini juga sering disebut sebagai vektorisasi. *Term Frequency-Inverse Document Frequency* (TF-IDF) menggunakan semua token dalam kumpulan data sebagai kosakata. Frekuensi kemunculan suatu token dari kosakata dalam setiap dokumen terdiri dari *Term Frequency* (TF) dan jumlah dokumen di mana token muncul akan menentukan nilai *Inverse Document Frequency* (TF-IDF).

## 2.4. Naïve Bayes Classifier

Naive Bayes adalah algoritma probabilistik yang biasanya digunakan untuk masalah klasifikasi. Naive Bayes Classifier bukanlah algoritma tunggal, melainkan kumpulan algoritma klasifikasi di mana semuanya memiliki prinsip yang sama berdasarkan Teorema Bayes [13] yang menggambarkan probabilitas suatu peristiwa berdasarkan pengetahuan sebelumnya atau probabilitas tertentu lainnya yang diketahui dari peristiwa itu. Pengklasifikasi Naive Bayes yang digunakan dalam penelitian ini adalah: 1) Multinomial Naive Bayes, model ini bekerja pada konsep *term frequency* yang berarti berapa kali sebuah kata muncul dalam dokumen. Multinomial NB memiliki kemampuan untuk mengklasifikasikan data yang tidak dapat direpresentasikan secara numerik, 2) Bernoulli Naive Bayes, model ini berguna jika vektor fitur yang digunakan adalah biner (yaitu 0 dan 1). Contoh paling umum adalah memeriksa apakah sebuah kata akan muncul dalam dokumen atau tidak (0 jika tidak, 1 jika iya). Bernoulli NB dapat memberikan hasil yang lebih baik dalam kasus di mana menghitung frekuensi kemunculan suatu kata kurang penting, 3) Gaussian Naive Bayes, model ini digunakan ketika fitur memiliki nilai kontinu. Gaussian NB mengasumsikan bahwa semua fitur mengikuti distribusi gaussian yaitu distribusi normal, 4) Complement Naive Bayes, model ini dirancang untuk mengoreksi “asumsi berat” yang dibuat oleh pengklasifikasi Multinomial Naive Bayes yang tidak

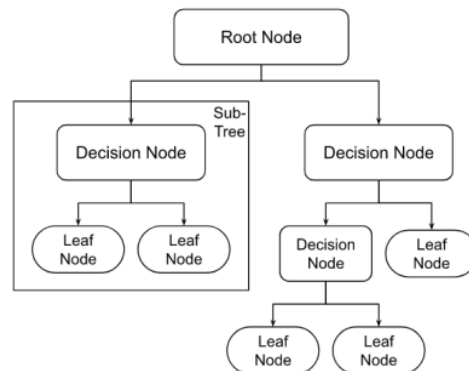
bekerja dengan baik pada kumpulan data yang tidak seimbang. Pengklasifikasi ini tidak menghitung probabilitas item milik kelas tertentu, melainkan menghitung probabilitas item milik semua kelas.

## 2.5. Algoritma Pembandingan Lain

Algoritma pembandingan digunakan sebagai upaya mencari model dengan akurasi terbesar sekaligus sebagai perbandingan dengan algoritma Naive Bayes yang digunakan. Algoritma yang dipilih adalah, sebagai deskripsi berikut.

### 2.5.1. Decision Tree

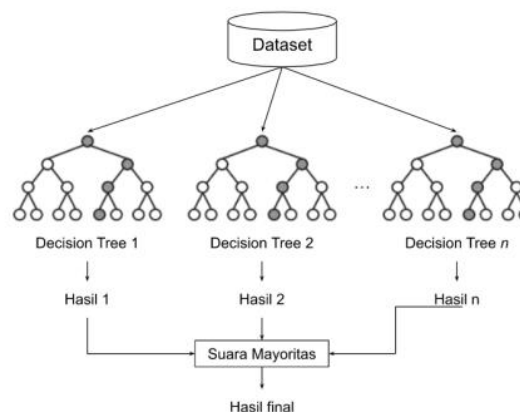
*Decision Tree*, model ini memiliki struktur seperti pohon di mana simpul teratas dianggap sebagai akar pohon yang secara rekursif dibagi pada rangkaian simpul keputusan dari akar sampai simpul terminal atau simpul keputusan tercapai [14] seperti ilustrasi pada Gambar 2.. Bagian atas *decision tree* dikenal sebagai *root node* yang sesuai dengan variabel prediktor terbaik.



Gambar 2. Ilustrasi Decision Tree [15]

### 2.5.2. Random Forest

*Random Forest*, model ini menggunakan beberapa pohon untuk menghitung rata-rata (regresi) atau suara mayoritas (klasifikasi) di simpul daun terminal saat membuat prediksi [14] seperti yang diilustrasikan pada Gambar 3. Pohon klasifikasi dan regresi dibangun menggunakan set data pelatihan yang dipilih secara acak dan subset acak dari variabel prediktor untuk hasil pemodelan [16].



Gambar 3. Ilustrasi Random Forest [17]

### 2.5.3. Logistic Regression

*Logistic Regression*, model *logistic regression* menguji hubungan antara satu atau lebih faktor independen untuk meramalkan variabel data dependen. Berbeda dengan *linear regression*, *logistic regression* memprediksi hasil sebagai probabilitas kelas default [18].

Regresi logistik digunakan ketika variabel dependen adalah kategoris, yang merupakan model paling umum untuk data respons biner. Dalam hal kemudahan komputasi, *logistic regression* adalah model terbaik di antara model linier umum untuk data respons biner [19].

**2.5.4. Support Vector Machine**

*Support Vector Machine*, merupakan model yang memanfaatkan titik batas yang memisahkan kelas yang dikenal sebagai vektor pendukung untuk klasifikasi data yang tidak berlabel. Tujuan utama SVM adalah untuk menemukan bidang data optimal yang membagi dua kelas titik data, algoritma ini memberikan akurasi yang lebih baik dalam kasus dimana ruang dimensi tinggi [20].

**2.6. Confusion Matrix**

*Confusion matrix* merupakan teknik untuk meringkas kinerja algoritma klasifikasi yang merupakan matriks berukuran  $n \times n$  yang menunjukkan klasifikasi yang diprediksi dan aktual, di mana  $n$  adalah jumlah kelas yang berbeda. Beberapa istilah yang terdapat dalam *confusion matrix* seperti yang direpresentasikan pada Tabel 1, antara lain:

- *True Positive (TP)* adalah jumlah prediksi positif yang benar;
- *False Positive (FP)* adalah jumlah prediksi positif yang salah;
- *True Negative (TN)* adalah jumlah prediksi negatif yang benar;
- *False Negative (FN)* adalah jumlah prediksi negatif yang salah.

Tabel 1. Representasi *Confusion Matrix*[21]

		PREDICTED CLASS	
		Positive	Negative
ACTUAL CLASS	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Akurasi, *precision*, *recall*, dan *F-1 score* dapat dihitung menggunakan informasi yang diberikan dalam *confusion matrix*. Akurasi merupakan seberapa akurat model dapat mengklasifikasi dengan benar, nilai ini dapat dihitung dengan membagi jumlah prediksi yang benar dengan jumlah total sampel. *Precision* menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model, nilai *precision* yang besar menunjukkan rendahnya nilai *false positive*. *Recall* adalah rasio pengamatan positif yang diprediksi dengan benar dengan semua pengamatan di kelas yang sebenarnya, nilai *recall* yang besar menunjukkan rendahnya nilai *false negative*. Menggabungkan nilai *precision* dan *recall* akan menghasilkan metrik tunggal yang dikenal sebagai *F1-score*, yang merupakan rata-rata harmonik tertimbang dari *precision* dan *recall*. Tabel 2 menunjukkan metrik pengukuran kinerja model *machine learning* beserta formula untuk menghitungnya.

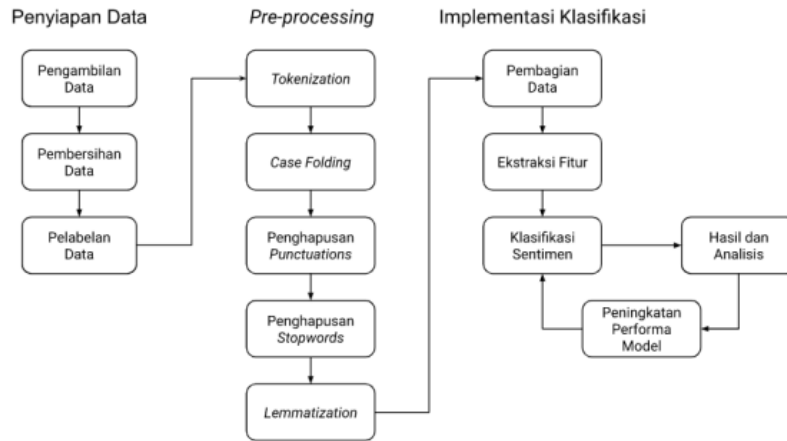
Tabel 2. Formula Metrik Evaluasi [22]

Metrik	Formula
Akurasi	$\frac{TP + TN}{TP + TN + FP + FN}$
<i>Recall</i>	$\frac{TP}{TP + FN}$
<i>Precision</i>	$\frac{TP}{TP + FP}$
<i>F1-score</i>	$2 * \frac{Recall * Precision}{Recall + Precision}$

**3. METODOLOGI PENELITIAN**

Alur penelitian ini dimulai dari penyiapan data, pada tahapan ini dilakukan pengambilan data ulasan aplikasi Amazon Shopping pada Google Play Store, data yang dikumpulkan akan dibersihkan dan diberikan label sentimen. Data yang telah disiapkan selanjutnya akan melalui tahapan preprocessing agar data siap dimasukkan ke dalam model, pada tahapan ini akan dilakukan beberapa proses secara urut. Tahapan

selanjutnya adalah dilakukan implementasi analisis sentimen, data ulasan akan dibagi menjadi set data latih dan uji yang kemudian dilakukan ekstraksi fitur untuk mengubah data mentah tersebut fitur numerik. Klasifikasi sentimen dilakukan menggunakan algoritma Naive Bayes, setelah hasil klasifikasi didapatkan, model kemudian akan dievaluasi dan dilakukan beberapa percobaan untuk meningkatkan performanya. Adapun tahapan penelitian ini digambarkan secara urut pada Gambar 4.



Gambar 4 . Tahapan Penelitian

**Penyiapan data:** Tahapan pertama dalam penelitian ini adalah penyiapan data. Kegiatan yang dilakukan pada tahapan ini antara lain pengambilan data ulasan aplikasi Amazon Shopping yang tersedia di Google Play Store. Data yang didapatkan akan dibersihkan dan diberi label sentimen terlebih dahulu sebelum dilanjutkan ke tahap berikutnya.

**Text pre-processing:** Pre-processing dilakukan untuk membersihkan noise untuk mendapatkan informasi akurat sebanyak mungkin dari teks. Proses-proses yang dilakukan antara lain: *tokenization*, *case folding*, penghapusan *stopwords* dan *punctuation*, *part-of-speech tagging*, dan *lemmatization*.

**Implementasi klasifikasi sentimen:** Data teks yang telah dilakukan text preprocessing selanjutnya digunakan untuk mengklasifikasi sentimen, algoritma yang digunakan adalah Naive Bayes Classifier, hasil analisis akan diklasifikasikan menjadi dua kategori: positif dan negatif. Tahapan ini meliputi pengekstraksian fitur pada data ulasan, pembagian data menjadi set data latih dan uji, kemudian dilakukannya klasifikasi sentimen itu sendiri.

**Evaluasi hasil:** Kinerja algoritma klasifikasi diringkas menggunakan teknik confusion matrix. Hasil analisis sentimen akan dievaluasi menggunakan matrik evaluasi akurasi, *precision*, *recall*, dan *f1-score*.

**Peningkatan kinerja model:** Model yang telah dievaluasi dan diketahui kinerja klasifikasinya kemudian akan dilakukan percobaan untuk meningkatkan kinerja tersebut. Percobaan yang akan dilakukan antara lain menggunakan pembobotan TF-IDF dan n-gram yang berbeda saat melakukan pengekstraksian fitur, serta penggunaan persentase pembagian data yang berbeda.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Pengambilan Data

Data-data ulasan diambil menggunakan google-play-scraper, fungsi yang digunakan adalah `reviews()` yang hanya mengambil data ulasan pada aplikasi yang dituju. Beberapa parameter wajib didefinisikan sebelum mengambil data ulasan, yaitu parameter `id`, `lang`, dan `country`. Pengaturan untuk bahasa dan negara yang digunakan adalah pengaturan default, yaitu bahasa “en” (English) dan negara “us” (United States). Selain itu, data ulasan juga diurutkan berdasarkan tanggal terbaru dituliskannya ulasan. Menggunakan parameter yang diberikan, dihasilkan data ulasan mentah sebanyak 2234 ulasan yang ditulis dalam 6 bulan terakhir saat data diambil, yaitu pada rentang 01 Januari 2022 sampai 06 Juni 2022.

#### 4.2 Pembersihan Data

Proses pembersihan meliputi penghapusan ulasan yang tidak berbahasa Inggris dan ulasan yang terlalu singkat dan tidak bermakna. Kelengkapan data akan dicek, jika terdapat ulasan yang terduplikat ataupun hilang, data tersebut akan dihapus. Kolom yang dibutuhkan untuk penelitian adalah kolom content yang berisi ulasan yang ditulis dan kolom *score* yang berisi rating yang diberikan pengulas. Kolom *score* dibutuhkan karena akan digunakan sebagai pertimbangan dalam memberikan skor sentimen terhadap ulasan yang ditulis dalam proses pelabelan data. Sebanyak 88 data ulasan telah dihapus dalam tahapan pembersihan data sehingga menyisakan data sebanyak 2147 ulasan.

#### 4.3 Pelabelan Data

Proses pelabelan data dilakukan secara manual terhadap 2147 ulasan. Label sentimen yang diberikan adalah negatif (-1) atau positif (1), label ini membentuk representasi dari kelas objek apa yang dimiliki ulasan dan membantu model *machine learning* untuk belajar mengidentifikasi sentimen pada data tanpa label. Sekumpulan aturan berikut digunakan untuk menentukan label sentimen:

- Jika ulasan cenderung negatif, label yang diberikan adalah negatif.
- Jika ulasan cenderung positif, label yang diberikan adalah positif.
- Jika ulasan cenderung netral, jumlah kata positif atau negatif terbanyak menentukan label sentimennya.
- Rating yang diberikan pengulas digunakan sebagai pertimbangan.

Hasil akhir dari proses pelabelan menghasilkan data sebanyak 991 ulasan bersentimen positif dan 1156 bersentimen negatif dengan persentase masing-masing 46.2% dan 53.8% dari keseluruhan data.

#### 4.4 Pemrosesan Teks

Data ulasan mentah melalui serangkaian proses *preprocessing* secara urut sebelum dapat dimasukkan ke dalam model. Tabel 3 berikut memperlihatkan perubahan yang terjadi pada salah satu teks ulasan untuk setiap tahapan.

Tabel 3. Perubahan Teks Ulasan pada Setiap Tahapan Pra-Pemrosesan

Tahapan	Hasil
Teks Awal	“Amazon Prime is absolutely amazing! Great prices. Shipped right to your door!”
Tokenization	["Amazon", "Prime", "is", "absolutely", "amazing", "!", "Great", "prices", ".", "Shipped", "right", "to", "your", "door", "!"]
Case Folding	“amazon prime is absolutely amazing! Great prices. shipped right to your door!”
Punctuations Removal	“amazon prime is absolutely amazing great prices shipped right to your door”
Stopwords Removal	“amazon prime absolutely amazing great prices shipped right door”
Normalization	“amazon prime absolutely amazing great price shipped right door”

Hasil preprocessing kemudian divisualisasikan dalam bentuk *wordcloud* yang memberikan representasikan daftar kata yang sering muncul untuk setiap kelas dalam data ulasan. Gambar 1 memvisualisasikan *word clouds* untuk semua kelas (A), kelas positif (B), dan kelas negatif (C).

#### 4.5 Ekstraksi Fitur

Ekstraksi fitur dilakukan menggunakan metode *Bag-of-Words* dan TF-IDF. Fungsi *CountVectorizer()* digunakan untuk pembuatan fitur menggunakan metode *Bag-of-Words*, sedangkan metode pembobotan TF-IDF dilakukan menggunakan fungsi *TfidfVectorizer()* yang disediakan dalam *library* *sklearn* dengan jumlah *n-gram* yang digunakan adalah 1 (*unigram*). Tabel 2 menunjukkan daftar kata-kata dengan frekuensi kemunculan terbesar yang didapatkan melalui proses *Bag-of-Words*. Sedangkan daftar istilah dengan nilai pembobotan TF-IDF terbesar dalam data ulasan dapat dilihat pada Tabel 4.

Tabel 4. Daftar 10 Istilah dengan nilai TF-IDF terbesar

No	Kata	Frekuensi
1	amazon	1605
2	app	1436
3	order	716
4	item	651
5	time	584
6	prime	423
7	love	392
8	easy	382
9	service	371
10	use	369

#### 4.6 Pembagian Data

Data ulasan akan dibagi menjadi set data *train* dan *test*. Data latih digunakan untuk melatih dan membuat model mempelajari fitur/pola tersembunyi dalam data. Sementara data uji merupakan data terpisah yang digunakan untuk menguji performa model setelah menyelesaikan pelatihan. Data dibagi dengan rasio 80:20, dimana 80% akan digunakan untuk data latih, dan 20% untuk data uji. Setelah melalui tahapan pembagian data, data ulasan yang berjumlah 2147 dibagi menjadi dua bagian, yaitu sebanyak 1717 buah data latih dan 429 buah data uji.

#### 4.7 Klasifikasi Sentimen dengan Naïve Bayes

Pengimplementasian klasifikasi sentimen menggunakan Naive Bayes dimulai dengan mengkompilasi model, hal ini dilakukan mengimpor modul dan kelas yang telah dikompilasi dari sklearn. Daftar pengklasifikasi Naive Bayes yang digunakan adalah BernoulliNB, ComplementNB, GaussianNB, dan MultinomialNB. Setelah model-model yang akan digunakan didefinisikan, langkah selanjutnya adalah melakukan *fitting* dengan memasukkan data latih pada model. Langkah terakhir adalah melakukan prediksi label sentimen menggunakan data uji.

#### 4.8 Pengujian Model

Model yang telah dilatih dan dapat memprediksi sentimen digunakan sebagai penentu untuk pengujian data menggunakan set data uji sebanyak 426 ulasan. Setiap ulasan diprediksi apakah termasuk ke dalam kelas positif atau negatif. Berikut pada Tabel 5 adalah contoh hasil prediksi klasifikasi sentimen pada set data uji yang didapatkan menggunakan algoritma Multinomial Naive Bayes.

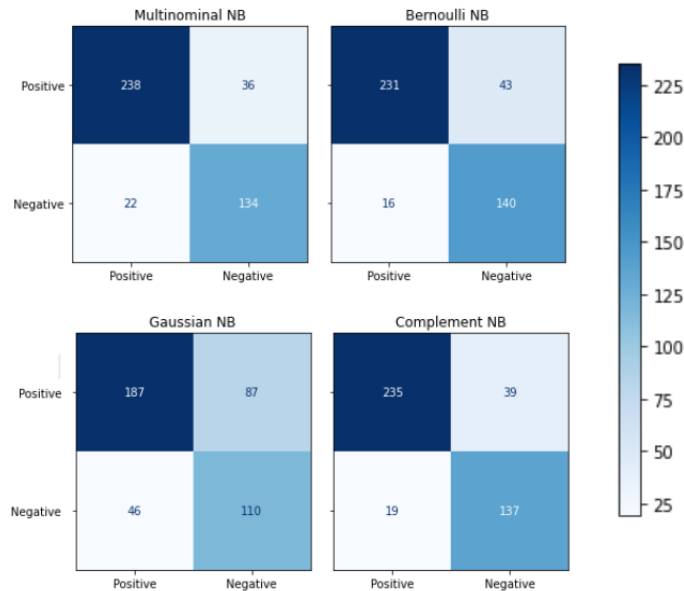
Tabel 5. Hasil Pengujian Menggunakan Set Data Uji

Teks	Prediksi	Aktual
Sucks lately. Both the web page and the app suck after l...	-1	-1
Amazon has made shopping during this time of uncertainty...	1	1
It is the worst, it doesn't even want to open. The app d...	-1	-1
Love it when it works but it keeps signing me out. It do...	1	-1
Amazon shopping is that 'right from the convenience of y...	1	1



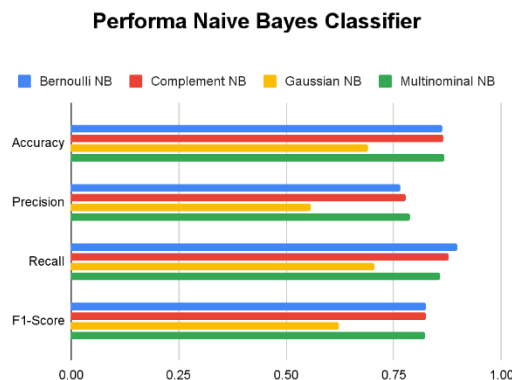
#### 4.9 Evaluasi Hasil

Model yang telah dilatih dan diuji perlu dievaluasi untuk menghitung performa klasifikasi yang dihasilkan. Proses evaluasi dari model dilakukan menggunakan *confusion matrix*. Performa pengklasifikasi dapat dilihat setelah mengetahui besarnya nilai akurasi, *precision*, *recall*, dan *f1-score* yang dihitung menggunakan nilai *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) dari *confusion matrix*. Performa model yang baik ditentukan dengan besarnya nilai-nilai tersebut, semakin besar maka semakin baik. *Confusion matrix* yang dihasilkan dari setiap model dapat dilihat pada Gambar 5.



Gambar 5. Confusion Matrix untuk Setiap Model yang dibandingkan

Nilai TP, TN, FP, FN yang didapatkan dari confusion matrix setiap model selanjutnya digunakan untuk menghitung nilai akurasi, *precision*, *recall*, dan *f1-score*. Setelah melalui proses perhitungan, akurasi tertinggi dari keempat algoritma Naive Bayes yang didapatkan adalah sebesar 86.74% dengan menggunakan Multinomial Naive Bayes. Sedangkan Gaussian Naive Bayes menghasilkan akurasi terkecil, yaitu 69.07%. Rata-rata keempat model Naive Bayes menghasilkan akurasi, *precision*, *recall* dan *f1-score* masing-masing sebesar 82.15%, 72.25%, 83.49% dan 77.41%. Pada Gambar 6 adalah visualisasi perbandingan performa klasifikasi dari keempat model.

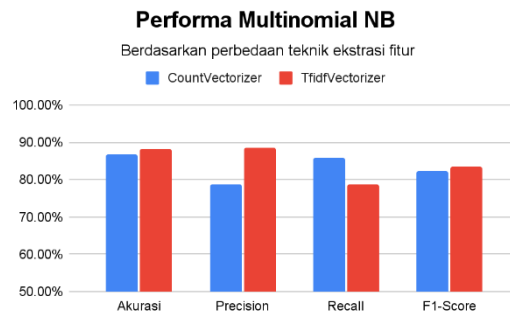


Gambar 6. Perbandingan Performa Klasifikasi Algoritma Naive Bayes

#### 4.10 Percobaan Peningkatan Akurasi Model

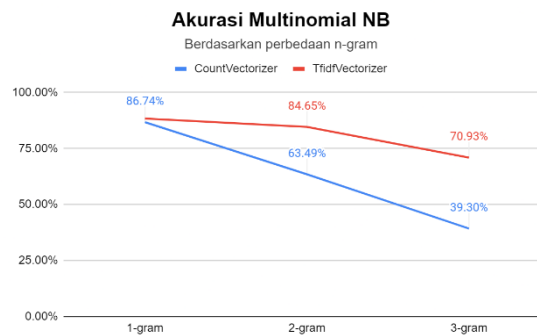
Setelah mendapatkan hasil evaluasi keempat model Naive Bayes, didapatkan model dengan nilai akurasi terbesar, yaitu Multinomial Naive Bayes sebesar 86.74%. Model tersebut kemudian akan dilakukan proses lanjutan untuk percobaan peningkatan akurasinya. Percobaan peningkatan akurasi model yang pertama dilakukan dengan melakukan ekstraksi fitur menggunakan pembobotan TF-IDF. Setelah

membandingkan kedua teknik ekstraksi fitur, didapatkan akurasi Multinomial NB dengan menggunakan seleksi fitur TF-IDF menghasilkan akurasi sebesar 88.37%, dibandingkan metode *Bag-of-Words* yang sebesar 86.74% seperti yang diilustrasikan pada Gambar 7.



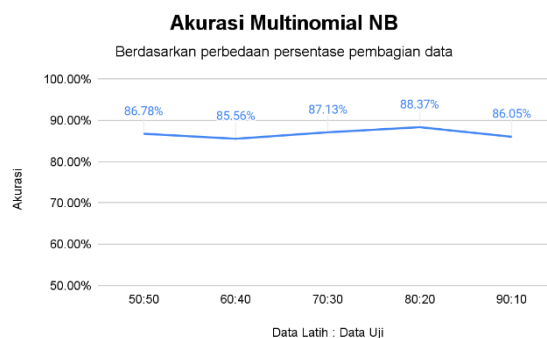
Gambar 7. Perbandingan Performa Teknik Ekstraksi Fitur pada Multinomial NB

Model Multinomial NB diuji kembali menggunakan n-gram yang berbeda saat melakukan ekstraksi fitur, token-token dibagi menjadi *unigram*, *bigram*, dan *trigram*. Hasil akurasi yang didapatkan adalah sebesar 82.7%, 75.2%, dan 54.6% untuk setiap n-gram metode TF-IDF. Sedangkan *Bag-of-Words* menghasilkan akurasi 82.1%, 69.0%, dan 46.2% seperti yang diilustrasikan pada Gambar 8.



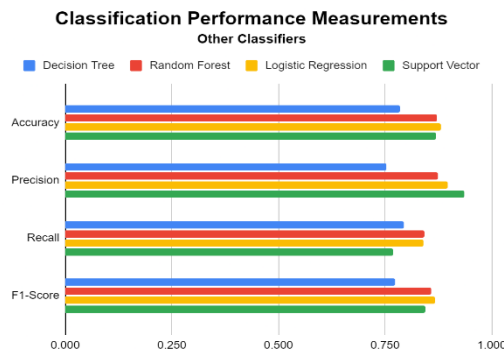
Gambar 8. Akurasi Multinomial NB berdasarkan Perbedaan N-Gram

Persentase pembagian data sebesar 80:20 digunakan dalam proses pembagian set data, yang berarti sebanyak 20% dari keseluruhan data akan digunakan sebagai set data uji, dan 80% sebagai set data latih. Pada percobaan berikut dilakukan percobaan penggunaan persentase pembagian data untuk persentase data uji sebanyak 10%-50% dari keseluruhan data. Hasil akurasi yang didapatkan untuk setiap persentase terdapat pada Gambar 9.



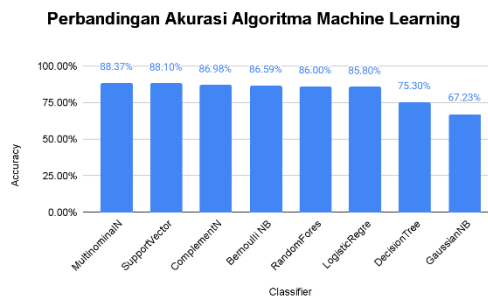
Gambar 9. Akurasi Multinomial NB berdasarkan Perbedaan Persentase Pembagian Data

Pada percobaan selanjutnya dilakukan tugas klasifikasi sentiment yang sama menggunakan algoritma *machine learning* non-bayesian. Teknik ekstraksi fitur yang digunakan adalah pembobotan TF-IDF dengan n-gram sebesar 1, dan persentase pembagian data sebanyak 80:20. Perbandingan performa klasifikasi algoritma non-bayesian yang digunakan digambarkan pada Gambar 10.



Gambar 10. Perbandingan Performa Klasifikasi Algoritma Non-Bayesian

Hasil performa semua algoritma *machine learning* yang digunakan kemudian dibandingkan berdasarkan tingkat akurasi. Menggunakan pengaturan yang sama, dihasilkan perbandingan performa akurasi sebagai berikut dalam Gambar 11.



Gambar 11. Perbandingan Akurasi Semua Algoritma Machine Learning yang Digunakan

## 5. KESIMPULAN

Data ulasan pengguna terhadap aplikasi Amazon Shopping yang telah diambil berjumlah 2234 ulasan, setelah melalui tahap pembersihan dihapuskan data sebanyak 88 ulasan yang tidak relevan ataupun tidak memenuhi kriteria yang didefinisikan sehingga menyisakan data sebanyak 2147 ulasan. Proses pelabelan data menghasilkan 1156 (53.8%) ulasan bersentimen negatif, dan 991 (46.2%) ulasan bersentimen positif dari keseluruhan ulasan sebanyak 2147. Hasil klasifikasi menggunakan keempat algoritma Naive Bayes menghasilkan akurasi rata-rata sebesar 82.15%, precision sebesar 72.25%, recall sebesar 83.49%, dan f1-score sebesar 77.41%. Multinomial NB menghasilkan akurasi terbaik diantara keempat algoritma Naive Bayes yang digunakan, yaitu sebesar 86.74%. Nilai precision, recall, dan f1-score yang dihasilkan berturut-turut adalah 78.82%, 85.90%, dan 82.21%. Penerapan Ekstraksi fitur menggunakan metode TF-IDF berhasil meningkatkan akurasi model Multinomial NB dari 86.74% menjadi 88.37%. Selain itu, Hasil perbandingan terhadap kedelapan algoritma machine learning yang digunakan menunjukkan Naive Bayes memiliki performa yang bersaing dibandingkan dengan algoritma lainnya. Akurasi tertinggi diperoleh sebesar 88.37% menggunakan Multinomial NB, sedangkan terbesar kedua adalah Support Vector Machine dengan nilai 88.10%.

Penelitian yang telah dilakukan masih jauh dalam kata sempurna, beberapa saran yang dapat penulis sampaikan dalam tulisan ini antara lain: menggunakan data ulasan dengan jumlah yang lebih banyak dalam melatih model machine learning, membuat sistem klasifikasi sentimen dengan antarmuka web maupun desktop agar lebih mudah digunakan dan dipahami pengguna, dan mencoba lebih banyak teknik untuk meningkatkan akurasi model, seperti menerapkan k-Fold Cross-Validation..

## DAFTAR PUSTAKA

- [1] Vanaja, S., Belwal, M. (2018). Aspect-Level Sentiment Analysis on E-Commerce Data. Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018). doi:10.1109/ICIRCA.2018.8597286
- [2] Liu, Bing. (2020). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press. ISBN: 9781108486378

- [3] BrightLocal. (2020). Local Consumer Review Survey 2020. Available Online at: <https://www.brightlocal.com/research/local-consumer-review-survey-2020/> [ April 15th 2022]
- [4] Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. doi:10.1016/j.knosys.2021.107134
- [5] Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. 2018 IEEE International Conference on Innovative Research and Development (ICIRD). doi:10.1109/icird.2018.8376299
- [6] Tama, V. O., Sibaroni, Y., & Adiwijaya. (2019). Labeling Analysis in the Classification of Product Review Sentiments by using Multinomial Naive Bayes Algorithm. *Journal of Physics: Conference Series*, 1192, 012036. doi:10.1088/1742-6596/1192/1/012036
- [7] Mohri, M., Rostamizadeh, A., Talwalkar, A. (2018). *Foundations of Machine Learning* Second Edition. The MIT Press: Cambridge, MA. ISBN 9780262039406
- [8] Borka, K. R., Hora, S., Jain, T., Wambugu, M. (2019). *Deep Learning for Natural Language Processing*. Packt Publishing Ltd: Birmingham. ISBN: 9781838553678.
- [9] Thanaki, J. (2017). *Feature Engineering and NLP Algorithms*. Python Natural Language Processing, 102-172.
- [10] Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, 1(1), 100008. doi:10.1016/j.jjime.2021.100008
- [11] Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. doi:10.1016/j.knosys.2021.107134
- [12] Ghojogh, Benyamin, et al. "Feature selection and feature extraction in pattern analysis: A literature review." arXiv preprint arXiv:1905.02845 (2019).
- [13] Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9, 381-386.
- [14] Kirasich, Kaitlin, Trace Smith, and Bivin Sadler.(2018) "Random forest vs logistic regression: binary classification for heterogeneous datasets." *SMU Data Science Review* 1.3, 2018: 9.
- [15] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.
- [16] Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications*, 134, 93-101.
- [17] Khan, M. A., Memon, S. A., Farooq, F., Javed, M. F., Aslam, F., & Alyousef, R. (2021). Compressive strength of fly-ash-based geopolymers concrete by gene expression programming and random forest. *Advances in Civil Engineering*, 2021.
- [18] Saim, M. M., & Ammor, H. (2022). Comparative study of machine learning algorithms (SVM, Logistic Regression and KNN) to predict cardiovascular diseases. In *E3S Web of Conferences* (Vol. 351, p. 01037). EDP Sciences.
- [19] Golpour, P., Ghayour-Mobarhan, M., Saki, A., Esmaily, H., Taghipour, A., Tajfard, M., ... & Ferns, G. A. (2020). Comparison of support vector machine, naïve Bayes and logistic regression for assessing the necessity for coronary angiography. *International journal of environmental research and public health*, 17(18), 6449.
- [20] Sharma, A., Kaur, S., Memon, N., Fathima, A. J., Ray, S., & Bhatt, M. W. (2021). Alzheimer's patients detection using support vector machine (SVM) with quantitative analysis. *Neuroscience Informatics*, 1(3), 100012.
- [21] Meiyazhagan, J. & S., Sudharsan & Venkatesan, A. & Senthilvelan, M.. (2021). Prediction of Occurrence of Extreme Events using Machine Learning.
- [22] Mousavi, Zarin & Mohammadi Zanjireh, Morteza & Oghbaie, Marzieh. (2020). Applying computational classification methods to diagnose Congenital Hypothyroidism: A comparative study. 18. 10.1016/j.imu.2019.100281.