

# PROPOSAL SKRIPSI



**Analisis Sentimen Pengguna Aplikasi *Jobstreet* di Indonesia pada *Google Play Store* Menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*: Perbandingan Akurasi Berdasarkan Rating Bintang**

Oleh:

**RANDI AFIF  
NIM 101210072**

**PROGRAM STUDI TEKNOLOGI INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
INSTITUT TEKNOLOGI DAN SAINS NAHDLATUL ULAMA  
PEKALONGAN  
2024**

## PROPOSAL SKRIPSI

Analisis Sentimen Pengguna Aplikasi *Jobstreet* di Indonesia pada *Google Play Store* Menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*:  
Perbandingan Akurasi Berdasarkan Rating Bintang

Oleh:

RANDI AFIF  
NIM 101210072

PROGRAM STUDI TEKNOLOGI INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
INSTITUT TEKNOLOGI DAN SAINS NAHDLATUL ULAMA  
PEKALONGAN  
2024

## HALAMAN PERSETUJUAN PEMBIMBING

*Assalaamu'alaikum Wr. Wb.*

Setelah saya meneliti dan mengadakan perbaikan seperlunya, bersama ini saya kirim naskah proposal skripsi saudara:

Nama : Randi Afif  
NIM : 101210072  
Program Studi : Teknologi Informasi  
Judul : Analisis Sentimen Pengguna Aplikasi *Jobstreet* di Indonesia pada *Google Play Store* Menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*: Perbandingan Akurasi Berdasarkan Rating Bintang

Proposal Skripsi ini telah disetujui oleh dosen pembimbing dan siap untuk dipertahankan dihadapan Dewan Penguji proposal skripsi program Sarjana Strata 1 (S-1) Fakultas Sains dan Teknologi Institut Teknologi dan Sains Nahdlatul Ulama Pekalongan.

Demikian harap menjadikan maklum.

*Wassalaamu'alaikum Wr. Wb.*

Pekalongan, 25 Oktober 2024

Dosen Pembimbing,

Nur Hayati, M.Kom.  
NIDN 0605109001

## DAFTAR ISI

PROPOSAL SKRIPSI .....	i
PROPOSAL SKRIPSI .....	ii
HALAMAN PERSETUJUAN .....	iii
DAFTAR ISI .....	iv
DAFTAR TABEL .....	v
DAFTAR GAMBAR .....	vi
DAFTAR SINGKATAN .....	vii
DAFTAR SIMBOL .....	viii
DAFTAR LAMPIRAN .....	ix
BAB I PENDAHULUAN .....	1
A. Latar Belakang Penelitian .....	1
B. Rumusan Masalah .....	2
C. Tujuan Penelitian .....	3
D. Batasan Penelitian .....	3
E. Manfaat Penelitian .....	4
BAB II LANDASAN TEORI .....	6
A. Tinjauan Pustaka .....	6
1. Sentimen Analisis .....	6
2. Natural Language Processing (NLP) .....	7
3. TF-IDF .....	8
4. Naive Bayes .....	9
5. Support Vector Machine (SVM) .....	10
6. Linux Mint .....	16
7. Python .....	17
8. Visual Studio Code .....	18
9. LibreOffice Suite .....	20
10. Google Play Store .....	21
11. Jobstreet .....	23
B. Penelitian yang Relevan .....	24
BAB III METODE PENELITIAN .....	17
A. Sumber Data .....	27
B. Alat dan Bahan .....	27
C. Tahapan Penelitian .....	29
DAFTAR PUSTAKA .....	34
LAMPIRAN .....	38

## DAFTAR TABEL

Tabel 2.1.	Persamaan dan Perbedaan antara Penelitian Penulis	
	Penelitian Relevan .....	24
Tabel 3.1	Spesifikasi Alat dan Bahan .....	27
Tabel 3.2	<i>Timeline</i> Penelitian .....	30

## **DAFTAR GAMBAR**

Gambar 2.1	Skema Model Support Vector Machines .....	11
Gambar 3.1	Tahapan Penelitian .....	29

## DAFTAR SINGKATAN

<i>NB</i>	: <i>Naive Bayes</i>
<i>SVM</i>	: <i>Support Vector Machine</i>
<i>KNN</i>	: <i>K-Nearest Neighbors</i>
<i>ML</i>	: <i>Machine Learning</i>
<i>AI</i>	: <i>Artificial Intelligence</i>
<i>TP</i>	: <i>True Positive</i>
<i>FP</i>	: <i>False Positive</i>
<i>TN</i>	: <i>True Negative</i>
<i>FN</i>	: <i>False Negative</i>
<i>F1</i>	: <i>F1 Score</i>
<i>R</i>	: <i>Recall</i>
<i>P</i>	: <i>Precision</i>

## DAFTAR SIMBOL

$P(C X)$	: Probabilitas bahwa kelas CC terjadi untuk teks XX
$P(X C)$	: Probabilitas bahwa teks XX terjadi untuk kelas CC
$P(C)$	: Probabilitas kelas CC
$P(X)$	: Probabilitas teks XX
C	: Kelas dari sentimen (positif dan negatif netral)
X	: Teks atau fitur ulasan (kata-kata dalam ulasan)



## DAFTAR LAMPIRAN

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang Masalah**

Perkembangan teknologi digital yang pesat telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk sektor ketenagakerjaan. Salah satu perubahan yang menonjol adalah transformasi proses rekrutmen dari yang sebelumnya manual menjadi berbasis teknologi digital melalui platform online. Di Indonesia, salah satu platform pencarian kerja yang paling populer adalah *Jobstreet*, sebuah aplikasi yang menyediakan berbagai informasi lowongan pekerjaan dan memungkinkan perusahaan untuk merekrut talenta-talenta berkualitas. *Jobstreet* telah menjadi jembatan penting antara pencari kerja dan perusahaan, terutama di era digital ini.

*Jobstreet* menawarkan kemudahan bagi pengguna, ulasan dan penilaian pengguna di *Google Play Store* menunjukkan berbagai variasi pengalaman, mulai dari kepuasan hingga ketidakpuasan. Pengguna biasanya memberikan ulasan dalam bentuk teks serta rating bintang, yang mencerminkan opini mereka tentang aplikasi. Data ini sangat penting karena dapat memberikan gambaran mengenai pengalaman pengguna, yang menjadi tolak ukur utama dalam pengembangan aplikasi lebih lanjut.

Namun, ulasan yang diberikan pengguna sering kali bervariasi dalam jumlah dan substansi, sehingga sulit untuk dilakukan analisis sentimen secara manual. Oleh karena itu, diperlukan pendekatan otomatis yang dapat menganalisis dan mengklasifikasikan ulasan pengguna berdasarkan sentimen yang mereka sampaikan. Analisis sentimen, yang merupakan bagian dari pemrosesan bahasa alami (Natural Language Processing/NLP), menjadi solusi yang efektif untuk mengidentifikasi apakah ulasan bersifat positif dan negatif.

Dalam analisis sentimen, pemilihan algoritma yang tepat sangat penting. Penelitian sebelumnya telah menunjukkan bahwa algoritma *Naive Bayes* dan *SVM* memiliki kelebihan masing-masing dalam klasifikasi sentimen.

Misalnya, *Naive Bayes* sering kali lebih cepat dalam pelatihan dan prediksi, sementara *SVM* dapat memberikan hasil yang lebih baik dalam dataset yang kompleks (Wijanarto & Brilianti, 2020). Dalam penelitian ini, Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* menjadi dua algoritma pembelajaran mesin yang akan digunakan untuk analisis sentimen. Keduanya memiliki pendekatan yang berbeda dalam mengklasifikasikan data teks, namun telah terbukti efektif dalam berbagai penelitian terkait analisis sentimen. Penelitian ini bertujuan untuk membandingkan tingkat akurasi kedua metode tersebut dalam menganalisis sentimen ulasan pengguna aplikasi *Jobstreet* di Indonesia, dengan mempertimbangkan rating bintang sebagai faktor pembanding utama.

Melalui analisis ini, diharapkan dapat ditemukan metode yang paling efektif untuk memahami sentimen pengguna aplikasi *Jobstreet*, sehingga pengembang aplikasi dapat lebih responsif dalam mengatasi masalah yang dihadapi pengguna dan meningkatkan kualitas layanan aplikasi secara keseluruhan.

## **B. Rumusan Masalah**

Berdasarkan latar belakang yang telah terurai tersebut, peneliti membuat rumusan masalah sebagai berikut:

1. Bagaimana klasifikasi sentimen ulasan pengguna aplikasi *Jobstreet* di Indonesia melalui *Google Play Store* berdasarkan rating bintang menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*?
2. Bagaimana perbandingan akurasi antara Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* dalam menganalisis sentimen ulasan pengguna aplikasi *Jobstreet* di Indonesia?
3. Apa faktor-faktor yang menyebabkan perbedaan akurasi antara Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* dalam klasifikasi sentimen ulasan pengguna aplikasi *Jobstreet*?

### C. Tujuan Penelitian

Berikut adalah tujuan penelitian dari peneliti berdasarkan rumusan masalah sebelumnya:

1. Mengklasifikasikan sentimen ulasan pengguna aplikasi *Jobstreet* di Indonesia melalui *Google Play Store* berdasarkan rating bintang menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*.
2. Menganalisis perbandingan tingkat akurasi antara Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* dalam menganalisis sentimen ulasan pengguna aplikasi *Jobstreet* di Indonesia.
3. Mengidentifikasi faktor-faktor yang mempengaruhi perbedaan akurasi antara Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* dalam klasifikasi sentimen ulasan pengguna aplikasi *Jobstreet*.

### D. Batasan Penelitian

Agar penelitian ini bisa fokus dan tidak menyimpang, maka peneliti melakukan pembatasan masalah sebagai berikut:

1. Sumber Data Ulasan: Data ulasan yang digunakan dalam penelitian ini hanya berasal dari platform *Google Play Store* dan tidak mencakup ulasan dari platform lain seperti App Store atau forum daring lainnya.
2. Konteks Aplikasi: Penelitian ini hanya fokus pada aplikasi *Jobstreet*, yang merupakan aplikasi investasi, trading, dan kripto, sehingga hasil analisis sentimen mungkin tidak dapat digeneralisasi untuk aplikasi lain di kategori yang berbeda.
3. Algoritma yang Digunakan: Penelitian ini hanya membandingkan dua metode pembelajaran mesin, yaitu *Naive Bayes* dan *Support Vector Machine (SVM)*, sehingga tidak membahas metode lain yang mungkin lebih efektif dalam analisis sentimen.

4. Negara Ulasan: Penelitian ini hanya menganalisis ulasan di negara Indonesia dan mengabaikan ulasan dari negara lain yang mungkin ada di *Google Play Store*.
5. Akurasi Model: Tingkat akurasi yang diperoleh dari model *Naive Bayes* dan *SVM* bergantung pada teknik praproses data dan parameter yang dipilih selama pelatihan model, sehingga hasil mungkin berbeda jika parameter yang digunakan berubah.

#### **E. Manfaat Penelitian**

Berikut manfaat yang bisa didapat dari penelitian ini:

1. Manfaat Akademis: Penelitian ini dapat memperkaya literatur tentang penerapan Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* dalam analisis sentimen, khususnya pada ulasan aplikasi fintech. Hasil penelitian ini dapat dijadikan referensi bagi penelitian selanjutnya yang ingin mengeksplorasi lebih jauh metode lain atau aplikasi dalam domain yang berbeda.
2. Manfaat Praktis bagi Pengembang Aplikasi: Hasil analisis sentimen ini dapat membantu pengembang aplikasi *Jobstreet* memahami lebih baik persepsi dan kepuasan pengguna terhadap layanan yang mereka tawarkan. Dengan memahami sentimen pengguna, pengembang dapat mengidentifikasi aspek-aspek yang perlu diperbaiki atau ditingkatkan untuk memberikan pengalaman yang lebih baik kepada pengguna.
3. Manfaat bagi Industri Fintech: Penelitian ini dapat memberikan wawasan kepada pelaku industri fintech mengenai pentingnya analisis ulasan pengguna sebagai alat evaluasi kinerja aplikasi. Dengan menggunakan pendekatan berbasis Machine learning seperti *Naive Bayes* dan *SVM*, perusahaan dapat mengotomatiskan proses pemantauan ulasan, yang akan mempercepat pengambilan keputusan terkait perbaikan atau pengembangan fitur baru.

4. Manfaat bagi Pengguna Aplikasi: Dengan adanya peningkatan kualitas aplikasi yang didasarkan pada hasil analisis sentimen, pengguna aplikasi *Jobstreet* akan mendapatkan pengalaman yang lebih baik dalam menggunakan aplikasi, baik dari segi fungsionalitas, performa, maupun kepuasan secara keseluruhan.
5. Manfaat bagi Pengambil Keputusan: Hasil penelitian ini dapat digunakan oleh pengambil keputusan di perusahaan untuk merumuskan strategi pengembangan produk berdasarkan data yang dihasilkan dari analisis sentimen. Dengan demikian, perusahaan dapat merespons lebih cepat terhadap keluhan dan kebutuhan pengguna.

## **BAB II**

### **LANDASAN TEORI**

#### **A. Tinjauan Pustaka**

##### **1. Analisis Sentimen**

Analisis sentimen merupakan proses yang penting dalam memahami opini pengguna terhadap produk dan layanan, terutama dalam konteks aplikasi belanja online. Dalam penelitian yang dilakukan oleh Subowo et al., algoritma Bidirectional Long Short Term Memory (BiLSTM) digunakan untuk menganalisis sentimen berbasis aspek pada aplikasi belanja online, seperti Cicilana. Metode ini memanfaatkan teknik pemrosesan bahasa alami (Natural Language Processing/NLP) untuk mengidentifikasi dan mengkategorikan opini yang dinyatakan dalam teks, dengan tujuan untuk menentukan sikap penulis terhadap suatu subjek, apakah positif dan negatif (Subowo et al., 2022). Dalam konteks yang lebih luas, analisis sentimen merupakan cabang dari text mining yang bertujuan untuk mengekstrak emosi dan opini dari teks. Metode *SVM* telah terbukti efektif dalam berbagai penelitian analisis sentimen, termasuk analisis terhadap aplikasi lain seperti Shopee dan MyPertamina, di mana penelitian-penelitian tersebut menunjukkan bahwa *SVM* mampu memberikan akurasi yang tinggi dalam klasifikasi sentimen (Wahyudi & Kusumawardana, 2021). Analisis sentimen berguna untuk memahami persepsi pengguna terhadap sebuah produk atau layanan secara lebih mendalam.

Beberapa teknik analisis sentimen yang digunakan secara luas termasuk:

1. Pendekatan berbasis leksikon, yang mengandalkan kata-kata yang dikaitkan dengan polaritas tertentu (positif/negatif).
2. Pendekatan berbasis pembelajaran mesin (Machine learning), yang melibatkan pelatihan model untuk mengklasifikasikan sentimen berdasarkan data yang telah diberi label.

Dalam penelitian ini, pembelajaran mesin menjadi pendekatan utama untuk mengklasifikasikan sentimen ulasan aplikasi *Jobstreet*.

Analisis sentimen merupakan teknik yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks, yang dapat bersifat positif dan negatif. Dalam konteks aplikasi investasi, analisis ini sangat penting karena dapat memberikan wawasan tentang kepuasan pengguna dan area yang perlu diperbaiki. Penelitian sebelumnya menunjukkan bahwa analisis sentimen dapat digunakan untuk mengevaluasi berbagai jenis produk dan layanan, termasuk aplikasi mobile, dengan akurasi yang tinggi (Adhi Putra, 2021).

## 2. *Natural Language Processing (NLP)*

*Natural Language Processing (NLP)* adalah cabang ilmu komputer yang berfokus pada interaksi antara komputer dan bahasa manusia. NLP memungkinkan komputer untuk memahami, menganalisis, dan menghasilkan teks dengan cara yang bermanfaat. (Syahroni, 2023) memberikan gambaran menyeluruh tentang bagaimana analisis sentimen, sebuah teknik utama dalam *Natural Language Processing (NLP)*, dapat digunakan secara efektif dalam konteks dasbor untuk mengevaluasi umpan balik siswa dalam bahasa Indonesia. Susanto et al., (2022) memberikan gambaran menyeluruh tentang analisis sentimen dalam konteks pemrosesan bahasa alami (NLP), khususnya berfokus pada penerapan algoritma *Naive Bayes* untuk mengklasifikasikan ulasan pengguna terhadap aplikasi. Dalam analisis sentimen, NLP digunakan untuk memproses ulasan teks, termasuk langkah-langkah seperti tokenisasi, penghapusan stop words, dan stemming.

Proses pra-proses data dalam NLP sangat penting untuk meningkatkan performa model pembelajaran mesin. Beberapa teknik yang sering digunakan dalam pra-proses data ulasan meliputi:

1. Tokenisasi, yaitu memecah teks menjadi unit-unit yang lebih kecil, seperti kata atau frasa.



2. Stop word removal, yaitu menghilangkan kata-kata umum yang tidak bermakna dalam konteks analisis sentimen.
3. Stemming atau lemmatization, yaitu mengubah kata-kata menjadi bentuk dasarnya untuk menyederhanakan analisis.

Teknik-teknik ini akan diterapkan dalam penelitian ini untuk memproses data ulasan pengguna aplikasi *Jobstreet* sebelum dilakukan analisis sentimen. Akurasi dan keandalan yang tinggi dalam mengklasifikasikan sentimen dapat dicapai dengan dukungan pengembangan NLP yang tangguh dalam konteks Indonesia (Tri Sanudin et al., 2024).

### 3. *TF-IDF*

*TF-IDF* (*Term Frequency-Inverse Document Frequency*) adalah salah satu teknik dasar dalam pengolahan teks yang digunakan untuk menilai pentingnya sebuah kata dalam dokumen tertentu, relatif terhadap sekumpulan dokumen lain. Term Frequency - Invers Document Frequency (*TF-IDF*) adalah metrik empiris yang menunjukkan seberapa penting suatu istilah bagi suatu dokumen dalam kumpulan dokumen (Karo Karo et al., 2023). Pada dasarnya, *TF-IDF* digunakan dalam information retrieval untuk membantu mesin pencari menentukan relevansi dokumen dengan kueri yang diajukan pengguna. *TF-IDF* mengkombinasikan dua metrik penting:

- Term Frequency (TF): Mengukur frekuensi kemunculan sebuah kata dalam dokumen. Semakin sering sebuah kata muncul dalam dokumen, semakin besar nilai TF. Rumus umum untuk TF adalah:

$$TF(t, d) = \frac{\text{Jumlah Kemunculan Kata } t \text{ dalam dokumen } d}{\text{Total Kata dalam Dokumen } d}$$

- Inverse Document Frequency (IDF): Mengukur seberapa umum atau jarang sebuah kata muncul di seluruh dokumen. Kata-kata yang sering muncul di banyak dokumen memiliki nilai IDF rendah, sementara kata yang jarang muncul di dokumen lain memiliki nilai IDF tinggi. Rumus umum untuk IDF adalah:

$$IDF(t) = \log\left(\frac{N}{DF(t)}\right)$$

di mana  $N$  adalah total jumlah dokumen, dan  $DF(t)$  adalah jumlah dokumen yang mengandung kata  $t$ .

Perhitungan *TF-IDF* dilakukan dengan mengalikan kedua nilai ini:

$$TF-IDF(t, d) = Tf(t, d) \times IDF(t)$$

Hasilnya menunjukkan seberapa penting suatu kata dalam dokumen tersebut dibandingkan dengan dokumen lain. Metode *TF-IDF* digunakan untuk memberikan bobot pada setiap kata (Yunanda et al., 2022). Pada perhitungan bobot, *TF-IDF* pertama-tama menghitung nilai TF untuk setiap kata, yang menunjukkan seberapa sering kata tersebut muncul dalam dokumen tertentu. Kemudian, nilai tersebut digabungkan dengan nilai IDF yang menunjukkan seberapa jarang kata tersebut muncul di seluruh koleksi dokumen (Saputra et al., 2024).

#### 4. *Naive Bayes*

*Naive Bayes* adalah algoritma pembelajaran mesin berbasis probabilitas yang sering digunakan untuk tugas klasifikasi teks, termasuk analisis sentimen. Analisis sentimen merupakan teknik yang penting dalam memahami opini publik terhadap aplikasi, terutama dalam konteks ulasan pengguna di platform seperti *Google Play Store*. Dalam hal ini, algoritma *Naive Bayes* telah terbukti efektif dalam mengklasifikasikan sentimen dari ulasan tersebut. Metode ini menggunakan pendekatan probabilistik untuk menentukan apakah suatu ulasan bersifat positif dan negatif, berdasarkan kata-kata yang terkandung di dalamnya (Irnawati dan Solecha, 2022). *Naive Bayes* tidak hanya efektif dalam analisis sentimen, tetapi juga dalam

memberikan informasi yang berguna bagi pengembang aplikasi untuk meningkatkan kualitas layanan mereka (Susanto et al., 2022). Prinsip Dasar *Naive Bayes* adalah *Naive Bayes* bekerja dengan menghitung probabilitas bahwa teks tertentu termasuk ke dalam kelas tertentu (positif dan negatif), berdasarkan frekuensi kata-kata yang ada dalam teks. Rumus dasar *Naive Bayes* adalah sebagai berikut:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Dimana:

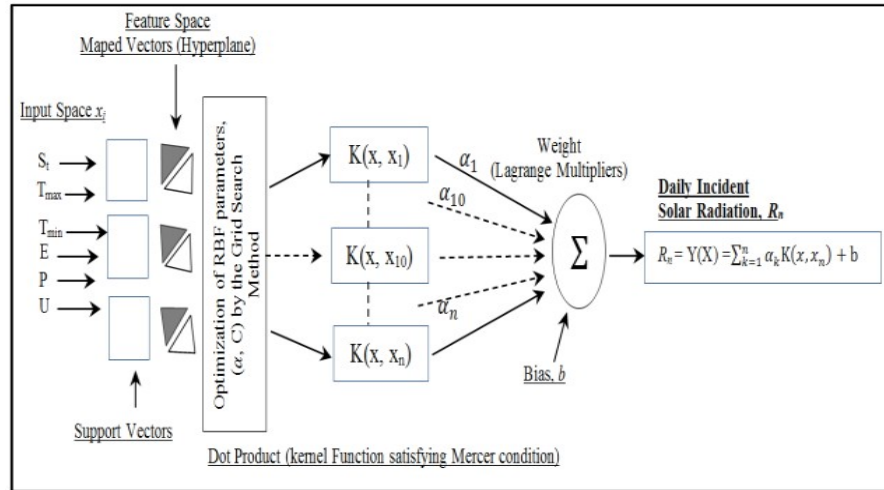
1.  $P(C|X)$  adalah probabilitas bahwa kelas CC terjadi untuk teks XX,
2.  $P(X|C)$  adalah probabilitas bahwa teks XX terjadi untuk kelas CC,
3.  $P(C)$  adalah probabilitas kelas CC,
4.  $P(X)$  adalah probabilitas teks XX.

Dalam konteks dashboard analisis sentimen, implementasi *Naive Bayes* dapat dilakukan dengan memanfaatkan data ulasan yang diambil dari *Google Play Store*. Data ini kemudian diproses melalui tahapan preprocessing, di mana teknik seperti *TF-IDF* dapat digunakan untuk mengekstrak fitur dari teks (Fibriyanti Arminda et al., 2023).

##### 5. *Support Vector Machine (SVM)*

*Support Vector Machine (SVM)* adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. *SVM* bekerja dengan mencari hyperplane yang memisahkan data dalam ruang fitur ke dalam dua atau lebih kelas yang berbeda. Analisis sentimen merupakan teknik yang penting dalam memahami opini pengguna terhadap aplikasi, dan salah satu algoritma yang paling efektif untuk tujuan ini adalah *Support Vector Machine (SVM)*. *SVM* telah terbukti memberikan akurasi yang tinggi dalam berbagai studi analisis

sentimen, termasuk analisis terhadap aplikasi seperti Google Meet dan Pospay, di mana *SVM* menunjukkan kinerja yang lebih baik dibandingkan algoritma lain seperti *Naive Bayes* dan K-Nearest Neighbors (Angraina dan Putri, 2022, Safryda Putri dan Ridwan, 2023, Oktafani dan Prasetyaningrum, 2022). Berikut contoh gambaran skema model *SVM*:



Gambar 2.1 Skema Model *Support Vector Machines*

Gambar 1 adalah gambaran skematik dari alur *SVM* untuk mengklasifikasikan pencahayaan matahari harian. Variabel prediktor yang digunakan sebagai input meliputi jam matahari ( $S_t$ ), suhu maksimum ( $T_{max}$ ), suhu minimum ( $T_{min}$ ), evaporasi ( $E$ ), presipitasi ( $P$ ), dan kecepatan angin ( $U$ ). Semua variabel ini digunakan untuk estimasi harian dari radiasi matahari global yang masuk ( $R_n$ ). Dalam hal ini, *SVM* berusaha menemukan batas optimal (hyperplane) yang dapat memisahkan data berdasarkan variabel-variabel prediktor tersebut, sehingga dapat memberikan estimasi yang akurat untuk radiasi matahari global. Dengan cara ini, *SVM* dapat membantu dalam memahami bagaimana berbagai faktor meteorologis berkontribusi terhadap radiasi matahari yang diterima di permukaan bumi (Hayati et al., 2019).

Dalam konteks *Support Vector Machine (SVM)*, terdapat beberapa jenis atau varian yang dapat digunakan tergantung pada jenis data dan masalah klasifikasi, yaitu sebagai berikut:

a. Linear *SVM*

1. Deskripsi: Linear *SVM* digunakan ketika data dapat dipisahkan secara linear, yaitu jika data dapat dipisahkan oleh garis lurus (untuk kasus dua dimensi) atau hyperplane (untuk dimensi lebih tinggi).
2. Cocok untuk: Data yang dapat dipisahkan dengan jelas tanpa perlu transformasi ke ruang fitur yang lebih tinggi.
3. Contoh kasus: Penggunaan Linear *SVM* sering kali cukup pada masalah dengan dimensi data yang rendah atau ketika data dapat dipisahkan dengan baik oleh garis lurus.

Contoh penggunaan: Klasifikasi teks dengan jumlah fitur yang besar tetapi distribusi data linier (misalnya, analisis sentimen sederhana).

b. Non-linear *SVM*

1. Deskripsi: Non-linear *SVM* digunakan ketika data tidak dapat dipisahkan secara linear. Pada kasus ini, *SVM* menggunakan kernel trick untuk memetakan data ke ruang fitur berdimensi lebih tinggi, di mana data tersebut dapat dipisahkan secara linear.
2. Cocok untuk: Data yang kompleks dengan distribusi yang non-linear, di mana pemisahan tidak dapat dilakukan oleh hyperplane linier.
3. Contoh kasus: Banyak aplikasi seperti deteksi pola pada data biologis, citra, dan klasifikasi teks multikategori memerlukan Non-linear *SVM*.

Jenis Kernel Non-linear yang Umum:

1. Polynomial Kernel: Memperluas ruang fitur menggunakan polinomial dari derajat tertentu.

2. Radial Basis Function (RBF) Kernel: Menciptakan hyperplane non-linear menggunakan fungsi Gaussian. Kernel ini sering digunakan karena fleksibilitasnya untuk menangani banyak jenis data.
3. Sigmoid Kernel: Berfungsi seperti fungsi aktivasi dalam neural network.

Contoh penggunaan: Klasifikasi gambar atau pola wajah yang memerlukan pemisahan non-linear.

c. *SVM* untuk Klasifikasi Multikelas

1. Deskripsi: *SVM* secara alami dirancang untuk masalah biner (dua kelas). Untuk menangani masalah multikelas (lebih dari dua kelas), dua pendekatan umum digunakan:
  - a. One-vs-One (OvO): Membagi masalah multikelas menjadi beberapa masalah biner dengan mengambil setiap pasangan kelas, lalu membangun *SVM* untuk masing-masing pasangan.
  - b. One-vs-All (OvA): Untuk setiap kelas, *SVM* dibangun untuk membedakan antara satu kelas dengan semua kelas lainnya. Hasilnya kemudian digunakan untuk mengklasifikasikan data baru.
2. Cocok untuk: Klasifikasi yang melibatkan banyak kategori, seperti pengklasifikasian produk dalam berbagai kategori atau klasifikasi ulasan sebagai positif, negatif, dan netral.

Contoh penggunaan: Klasifikasi teks dalam beberapa kategori seperti ulasan produk dengan banyak level rating (1 sampai 5 bintang).

d. *SVM* untuk Regresi (Support Vector Regression - SVR)

1. Deskripsi: Support Vector Regression (SVR) adalah variasi dari *SVM* yang digunakan untuk masalah regresi. SVR mencoba menemukan

fungsi yang tidak hanya memprediksi label tetapi juga meminimalkan kesalahan dalam prediksi dengan margin tertentu.

2. Cocok untuk: Masalah prediksi nilai (kontinu), seperti harga saham, suhu, atau volume penjualan.
3. Contoh kasus: Prediksi suhu berdasarkan data cuaca atau prediksi harga rumah.

Contoh penggunaan: Prediksi harga saham atau penjualan berdasarkan fitur-fitur tertentu seperti waktu, volume transaksi, dan tren pasar.

e. Proximal *SVM*

1. Deskripsi: Proximal *Support Vector Machine* (*PSVM*) merupakan variasi dari *SVM* yang mendekati solusi dengan cara meminimalkan jarak dari data ke hyperplane, bukan dengan memaksimalkan margin. *PSVM* biasanya lebih cepat daripada *SVM* standar.
2. Cocok untuk: Situasi di mana pemisahan antara dua kelas tidak jelas, dan klasifikasi harus lebih fleksibel.
3. Contoh kasus: Klasifikasi teks atau citra di mana margin antar kelas mungkin tidak signifikan, tetapi jarak dari hyperplane lebih dipertimbangkan.

f. Least Squares *SVM* (*LS-SVM*)

1. Deskripsi: *LS-SVM* adalah modifikasi dari *SVM* yang menggunakan least squares cost function untuk menemukan hyperplane yang memisahkan data. Hal ini membuat perhitungan menjadi lebih sederhana, dengan mengubah masalah menjadi sistem persamaan linier.
2. Cocok untuk: Aplikasi di mana performa tinggi dan kecepatan diperlukan, misalnya dalam deteksi pola atau klasifikasi yang harus dilakukan secara real-time.

3. Contoh kasus: Pengenalan pola wajah atau suara.

g. Structured *SVM*

1. Deskripsi: Structured *SVM* digunakan untuk tugas-tugas klasifikasi yang tidak hanya menghasilkan label tunggal tetapi label yang berstruktur, seperti urutan kata, parsing bahasa alami, atau segmentasi gambar. Structured *SVM* memperhitungkan hubungan antar elemen dalam output yang dihasilkan.
2. Cocok untuk: Masalah klasifikasi dengan output yang rumit seperti pengenalan citra atau pemrosesan bahasa alami.
3. Contoh kasus: Penandaan bagian-bagian kalimat dalam Named Entity Recognition (NER) atau part-of-speech tagging dalam pemrosesan bahasa alami.

h. Transductive *SVM*

1. Deskripsi: Transductive *SVM* (*TSVM*) digunakan untuk masalah semi-supervised learning, di mana hanya sebagian dari data yang diberi label. *TSVM* berusaha memprediksi label untuk data yang tidak berlabel secara bersamaan saat melatih model dengan data berlabel.
2. Cocok untuk: Ketika ada data yang tidak diberi label dalam jumlah besar, dan klasifikasi masih harus dilakukan.
3. Contoh kasus: Dalam analisis teks, *TSVM* dapat digunakan untuk membantu mengklasifikasikan teks tanpa label, berdasarkan data yang telah dilabeli.

Dengan memanfaatkan *SVM* dan teknik visualisasi yang efektif, dashboard dapat memberikan informasi yang berharga bagi pengembang aplikasi untuk melakukan perbaikan yang diperlukan (Wahyu Sejati et al., 2023). Kekuatan *SVM* dalam analisis sentimen terletak pada kemampuannya untuk menangani data yang tidak terstruktur, seperti teks, serta mengurangi



risiko overfitting pada data yang memiliki dimensi tinggi, seperti ulasan pengguna aplikasi.

#### 6. *Linux Mint*

*Linux Mint*, sebagai distribusi sistem operasi berbasis Ubuntu, menawarkan antarmuka yang ramah pengguna dan stabilitas yang tinggi, menjadikannya pilihan ideal bagi pengguna yang beralih dari Windows. Keunggulan ini sangat penting dalam konteks penelitian analisis sentimen, terutama dalam pemrosesan data dan pengolahan teks. *Linux Mint* mendukung berbagai alat pemrograman, termasuk *Python*, yang merupakan bahasa pemrograman utama dalam analisis data dan pembelajaran mesin. Dengan pustaka seperti *Scikit-learn*, pengguna dapat dengan mudah menerapkan algoritma seperti *Naive Bayes* dan *Support Vector Machine (SVM)* untuk analisis sentimen (Safryda Putri dan Ridwan, 2023, Widodo et al., 2022).

Dalam penelitian ini, penggunaan *Linux Mint* dapat memberikan keuntungan dalam hal efisiensi pemrosesan data. Sistem operasi ini memungkinkan pengguna untuk menjalankan berbagai skrip *Python* yang diperlukan untuk mengumpulkan dan menganalisis data ulasan dari *Google Play Store*. Selain itu, *Linux Mint* juga dikenal karena kemudahan dalam menginstal dan mengkonfigurasi alat analisis data, yang sangat penting untuk penelitian yang melibatkan pengolahan data besar (Abu Samah et al., 2022, Fibriyanti Arminda et al., 2023).

Lebih lanjut, penelitian yang dilakukan oleh Widodo et al. menunjukkan bahwa penggunaan algoritma *Naive Bayes* dalam analisis sentimen dapat memberikan hasil yang akurat dalam mengklasifikasikan ulasan pengguna (Widodo et al., 2022). Penelitian ini sejalan dengan temuan lain yang menunjukkan efektivitas *SVM* dalam analisis sentimen, yang juga dapat diimplementasikan dengan mudah di *Linux Mint* (Safryda Putri dan Ridwan, 2023, Wahyudi dan Kusumawardana, 2021). Dengan demikian, *Linux Mint* tidak hanya memberikan lingkungan yang stabil dan ramah

pengguna, tetapi juga mendukung berbagai alat dan pustaka yang diperlukan untuk melakukan analisis sentimen secara efisien.

Secara keseluruhan, *Linux Mint* merupakan pilihan yang sangat baik untuk penelitian analisis sentimen, terutama dalam konteks aplikasi seperti *Jobstreet*, di mana pengguna dapat dengan mudah mengakses dan menganalisis data ulasan menggunakan algoritma pembelajaran mesin yang kuat. Keberadaan alat seperti *Python* dan pustaka analisis data di *Linux Mint* memungkinkan peneliti untuk melakukan eksplorasi data yang mendalam dan menghasilkan wawasan yang berharga dari data yang tersedia (Erawati et al., 2023, Maulana et al., 2023).

## 7. *Python*

*Python* telah menjadi bahasa pemrograman yang sangat populer di kalangan peneliti dan pengembang, terutama dalam bidang analisis data dan machine learning. Dalam konteks penelitian ini, *Python* menawarkan berbagai pustaka yang sangat mendukung proses analisis data, pemrosesan bahasa alami, dan pelatihan model machine learning. Pustaka-pustaka ini, seperti *NLTK*, *Sastrawi*, *Scikit-learn*, *Pandas*, *NumPy*, *Matplotlib*, dan *Seaborn*, memberikan alat yang diperlukan untuk mengolah dan menganalisis data ulasan pengguna secara efisien (Maulana et al., 2023, Widodo et al., 2022).

Pustaka *NLTK* (*Natural Language Toolkit*) dan *Sastrawi* sangat penting untuk pemrosesan bahasa alami, khususnya dalam konteks bahasa Indonesia. *NLTK* menyediakan berbagai fungsi untuk pemrosesan teks, termasuk penghapusan stopwords dan stemming, yang sangat diperlukan untuk membersihkan data ulasan sebelum analisis dilakukan (Akbar dan Nirwana Samrin, 2023). *Sastrawi*, yang khusus dirancang untuk bahasa Indonesia, memberikan dukungan tambahan dalam pengolahan teks, sehingga peneliti dapat lebih mudah menangani nuansa bahasa lokal (Widodo et al., 2022).

*Scikit-learn* adalah pustaka yang sangat berguna untuk implementasi algoritma machine learning, termasuk *Naive Bayes* dan *SVM*. Pustaka ini menyediakan berbagai fungsi untuk pemisahan data (train-test split), evaluasi model, dan ekstraksi fitur menggunakan metode seperti TF-IDF, yang memungkinkan peneliti untuk mendapatkan hasil analisis yang akurat dan dapat diandalkan (Elistiana et al., 2023, Widodo et al., 2022). Selain itu, *Pandas* dan *NumPy* memungkinkan pengelolaan dan manipulasi data yang efisien, yang sangat penting ketika menangani dataset besar yang diambil dari *Google Play Store* (Rahmany et al., 2020).

Visualisasi hasil analisis juga merupakan aspek penting dalam penelitian ini. Pustaka *Matplotlib* dan *Seaborn* memungkinkan peneliti untuk membuat grafik dan visualisasi yang jelas, yang dapat membantu dalam memahami tren sentimen dari waktu ke waktu serta akurasi model yang diterapkan (Han dan Kwak, 2023). Dengan menggunakan alat-alat ini, peneliti dapat mengumpulkan ulasan dari *Google Play Store*, mengkategorikannya berdasarkan rating bintang, dan menerapkan algoritma *Naive Bayes* dan *SVM* untuk menentukan model mana yang memberikan hasil akurasi terbaik (Elistiana et al., 2023, Widodo et al., 2022).

Secara keseluruhan, penggunaan *Python* dan pustaka-pustakanya dalam penelitian analisis sentimen ini tidak hanya mempermudah proses pengolahan data, tetapi juga meningkatkan keakuratan dan efisiensi analisis yang dilakukan. Dengan dukungan alat yang tepat, peneliti dapat menghasilkan wawasan yang berharga mengenai sentimen pengguna terhadap aplikasi *Jobstreet*, yang pada gilirannya dapat digunakan untuk meningkatkan kualitas layanan aplikasi tersebut (I. Maulana et al., 2023, Widodo et al., 2022).

#### 8. *Visual Studio Code*

*Visual Studio Code (VSCode)* telah menjadi salah satu alat yang paling banyak digunakan dalam pengembangan perangkat lunak dan analisis data, berkat antarmuka yang sederhana dan dukungan untuk berbagai bahasa

pemrograman. Keunggulan *VSCode*, termasuk kemudahan integrasi dengan ekstensi, dukungan debugging, dan terminal bawaan, sangat mendukung proses pengembangan dan eksperimen dalam analisis data (Rahmany et al., 2020). Dalam konteks penelitian ini, *VSCode* dapat berfungsi sebagai alat utama bagi peneliti untuk menulis, mengelola, dan menjalankan kode Python yang diperlukan untuk memproses data ulasan (Widodo et al., 2022).

Penggunaan *VSCode* dalam penelitian ini memungkinkan peneliti untuk melakukan analisis sentimen dengan lebih efisien. Dengan kemampuan *VSCode* untuk mendukung berbagai ekstensi Python, peneliti dapat memanfaatkan pustaka seperti *scikit-learn* untuk menerapkan algoritma *Naive Bayes* dalam analisis sentimen (Elistiana et al., 2023). Selain itu, *VSCode* menyediakan fitur debugging yang memungkinkan peneliti untuk mengidentifikasi dan memperbaiki kesalahan dalam kode mereka dengan cepat, sehingga meningkatkan akurasi dan keandalan model yang dikembangkan (Rahmany et al., 2020). Dengan demikian, *VSCode* tidak hanya berfungsi sebagai editor kode, tetapi juga sebagai lingkungan pengembangan yang mendukung penelitian yang lebih mendalam dan analisis yang lebih akurat.

Dalam penelitian ini, peneliti memanfaatkan *VSCode* untuk mengelola dataset yang diambil dari *Google Play Store*, yang berisi ulasan pengguna aplikasi *Jobstreet*. Dengan menggunakan *VSCode*, peneliti dapat dengan mudah melakukan preprocessing data, seperti penghapusan stop words dan stemming, yang merupakan langkah penting dalam analisis sentimen (Widodo et al., 2022). Selain itu, *VSCode* memungkinkan peneliti untuk melakukan visualisasi data, yang sangat penting untuk memahami pola dalam data ulasan dan untuk mempresentasikan hasil analisis secara efektif (Han & Kwak, 2023). Dengan demikian, *VSCode* berperan penting dalam meningkatkan efisiensi dan efektivitas penelitian analisis sentimen ini.

Secara keseluruhan, *VSCode* adalah alat yang sangat berguna dalam konteks penelitian analisis sentimen, terutama dalam pengembangan dan penerapan algoritma seperti *Naive Bayes*. Kemampuannya untuk mendukung berbagai bahasa pemrograman, integrasi dengan pustaka analisis data, dan fitur debugging yang kuat menjadikannya pilihan ideal bagi peneliti yang ingin melakukan analisis data secara mendalam dan akurat (Elistiana et al., 2023; Widodo et al., 2022).

#### 9. *LibreOffice Suite*

*LibreOffice Suite*, sebagai perangkat lunak produktivitas open-source, menyediakan berbagai fitur yang sangat berguna dalam konteks analisis data, termasuk analisis sentimen terhadap aplikasi seperti *Jobstreet*. Dalam penelitian yang berfokus pada analisis sentimen pengguna aplikasi *Jobstreet* di *Google Play Store*, LibreOffice dapat berfungsi sebagai alat bantu yang efektif untuk mengelola dan memproses data yang diperoleh dari ulasan pengguna.

Pertama, dalam hal pengelolaan data, *LibreOffice Calc* dapat digunakan untuk membersihkan dan memformat dataset yang diperoleh dari scraping ulasan di *Google Play Store*. Penelitian sebelumnya menunjukkan bahwa penggunaan *spreadsheet* seperti Calc memungkinkan peneliti untuk mengorganisir data dengan lebih baik, termasuk rating, teks ulasan, dan label sentimen, sehingga memudahkan dalam pengolahan lebih lanjut (Susanto et al., 2022, Widodo et al., 2022). Selain itu, kemampuan Calc untuk mengonversi data ke berbagai format, seperti CSV, sangat penting untuk integrasi dengan perangkat lunak analisis statistik atau bahasa pemrograman seperti Python yang sering digunakan dalam analisis sentimen (Valadez Estrada, 2023).

Kedua, setelah pemrosesan data awal, *LibreOffice Writer* dapat digunakan untuk menyusun laporan hasil analisis. Dalam konteks penelitian ini, peneliti dapat menyusun laporan yang mencakup perbandingan akurasi antara algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*, serta

interpretasi hasil berdasarkan rating bintang yang diperoleh dari ulasan pengguna (Gishella Septania Al-Husna et al., 2024, W. Wahyudi et al., 2024). Penulisan laporan yang terstruktur dengan baik sangat penting untuk menyampaikan temuan penelitian secara efektif kepada audiens yang lebih luas.

Lebih lanjut, analisis sentimen yang dilakukan pada aplikasi *Jobstreet* dapat memanfaatkan algoritma seperti *Naive Bayes* dan *SVM*, yang telah terbukti efektif dalam mengklasifikasikan sentimen dari ulasan pengguna (Gishella Septania Al-Husna et al., 2024, W. Wahyudi et al., 2024). Penelitian sebelumnya menunjukkan bahwa kedua algoritma ini dapat memberikan wawasan yang berharga tentang kepuasan pengguna dan area yang perlu diperbaiki dalam aplikasi (Gishella Septania Al-Husna et al., 2024, W. Wahyudi et al., 2024). Dengan demikian, *LibreOffice Suite* tidak hanya berfungsi sebagai alat untuk pengelolaan data, tetapi juga sebagai platform untuk dokumentasi dan pelaporan yang mendukung keseluruhan proses penelitian.

Secara keseluruhan, *LibreOffice Suite* menawarkan fitur-fitur yang sangat mendukung dalam pengelolaan dan analisis data, yang sangat relevan dalam konteks penelitian analisis sentimen pengguna aplikasi *Jobstreet* di *Google Play Store*. Penggunaan alat ini dapat meningkatkan efisiensi dan efektivitas dalam proses penelitian, dari pengumpulan data hingga penyusunan laporan akhir.

#### 10. *Google Play Store*

*Google Play Store* berfungsi sebagai platform distribusi digital yang memungkinkan pengguna Android untuk mengunduh berbagai aplikasi, termasuk aplikasi pencarian kerja seperti *Jobstreet*. Dalam konteks analisis sentimen pengguna aplikasi *Jobstreet* di Indonesia, *Google Play Store* menjadi sumber data yang sangat penting. Pengguna dapat memberikan ulasan dan rating yang mencerminkan pengalaman mereka dengan aplikasi tersebut, yang pada gilirannya dapat memberikan wawasan berharga tentang

persepsi pengguna terhadap aplikasi pencarian kerja ini (Raja dan Tileng, 2022, Widodo et al., 2022).

Analisis sentimen merupakan teknik yang digunakan untuk mengevaluasi opini dan emosi yang terkandung dalam ulasan pengguna. Dalam penelitian ini, algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* digunakan untuk mengklasifikasikan ulasan ke dalam kategori positif dan negatif. Penelitian sebelumnya menunjukkan bahwa algoritma *Naive Bayes* efektif dalam analisis sentimen, terutama dalam konteks aplikasi mobile, termasuk aplikasi *Jobstreet* (Fibriyanti Arminda et al., 2023, Widodo et al., 2022). Selain itu, *SVM* juga telah terbukti memberikan hasil yang baik dalam klasifikasi sentimen, seperti yang ditunjukkan dalam penelitian yang menganalisis aplikasi lain di *Google Play Store* (Angraina dan Putri, 2022, R. Wahyudi dan Kusumawardana, 2021).

Penggunaan data dari *Google Play Store* untuk analisis sentimen tidak hanya membantu dalam memahami persepsi pengguna, tetapi juga memberikan informasi yang dapat digunakan oleh pengembang untuk meningkatkan kualitas aplikasi. Dengan lebih dari 10 juta unduhan, aplikasi *Jobstreet* memiliki potensi untuk mengumpulkan sejumlah besar data ulasan yang dapat dianalisis untuk mendapatkan wawasan tentang fitur yang disukai atau tidak disukai oleh pengguna (R. Maulana et al., 2023; Raja & Tileng, 2022; Widodo et al., 2022).

Dalam penelitian ini, pendekatan yang sistematis dan berbasis data sangat penting untuk mendapatkan hasil yang akurat. Dengan menerapkan algoritma yang tepat dan menggunakan data yang relevan dari *Google Play Store*, analisis sentimen dapat memberikan gambaran yang jelas tentang bagaimana pengguna merasakan aplikasi *Jobstreet*, serta area yang perlu diperbaiki untuk meningkatkan pengalaman pengguna (Fibriyanti Arminda et al., 2023, R. Maulana et al., 2023, Widodo et al., 2022).

## 11. *Jobstreet*

*Jobstreet* merupakan platform pencarian kerja yang sangat dikenal di Asia Tenggara, termasuk Indonesia, dan menyediakan berbagai informasi penting bagi pencari kerja, seperti profil perusahaan dan kisaran gaji (Widodo et al., 2022). Dalam penelitian ini, algoritma *Naive Bayes* dan Support Vector Machine (*SVM*) akan digunakan untuk mengklasifikasikan sentimen ulasan menjadi kategori positif, negatif, atau netral. Kedua algoritma ini telah terbukti efektif dalam analisis sentimen di berbagai aplikasi, termasuk aplikasi *Jobstreet* itu sendiri (Gishella Septania Al-Husna et al., 2024, Widodo et al., 2022).

Penggunaan algoritma *Naive Bayes* dalam analisis sentimen telah banyak dibahas dalam literatur, di mana penelitian sebelumnya menunjukkan bahwa algoritma ini mampu mencapai akurasi yang baik dalam mengklasifikasikan sentimen dari ulasan pengguna (Fibriyanti Arminda et al., 2023, Tri Sanudin et al., 2024). Selain itu, *SVM* juga dikenal sebagai metode yang kuat dalam analisis sentimen, dengan beberapa studi menunjukkan bahwa *SVM* sering kali memberikan hasil yang lebih baik dibandingkan dengan *Naive Bayes* dalam konteks tertentu (Gishella Septania Al-Husna et al., 2024, R. Wahyudi dan Kusumawardana, 2021). Oleh karena itu, perbandingan akurasi antara kedua algoritma ini dalam konteks aplikasi *Jobstreet* diharapkan dapat memberikan wawasan yang berharga bagi pengembang aplikasi untuk meningkatkan pengalaman pengguna.

Hasil dari analisis sentimen ini diharapkan tidak hanya memberikan gambaran yang lebih jelas tentang opini pengguna lain terhadap aplikasi *Jobstreet*, tetapi juga dapat memengaruhi keputusan pencari kerja dalam menggunakan aplikasi ini. Dengan memahami sentimen yang terkandung dalam ulasan pengguna, pengembang dapat melakukan perbaikan yang diperlukan untuk meningkatkan fitur dan layanan aplikasi (Gishella Septania Al-Husna et al., 2024, Putri et al., 2023, Widodo et al., 2022).



Penelitian ini juga akan memberikan kontribusi pada pengembangan metode analisis sentimen yang lebih baik, dengan mempertimbangkan aspek-aspek yang mempengaruhi akurasi klasifikasi sentimen (Gishella Septania Al-Husna et al., 2024, Pratmanto et al., 2023).

## B. Penelitian yang Relevan

Ada berbagai rujukan yang diambil dalam penelitian ini guna memperkuat penelitian ini. Selain untuk memperkuat penelitian, rujukan-rujukan ini adalah penelitian yang sudah pernah dilakukan dan menghasilkan berbagai macam kesimpulan yang dapat dijadikan bahan perbandingan ketika akan melakukan penelitian. Tentunya mengambil penelitian ini tidak hanya perlu memperhatikan isi, namun juga apa yang telah diteliti sebelumnya oleh peneliti yang lain. Oleh karena itu, berikut beberapa penelitian terdahulu yang memperkuat penelitian ini serta persamaan dan perbedaannya:

**Tabel 2.1**  
**Persamaan dan Perbedaan antara Penelitian Penulis dengan Penelitian yang Relevan**

No	Penelitian yang Relevan	Persamaan	Perbedaan
1.	Maulana, Voutama, dan Ridwan (2023), berjudul "Analisis Sentimen Ulasan Aplikasi MyPertamina pada <i>Google Play Store</i> menggunakan Algoritma NBC"	Kedua studi bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna di <i>Google Play Store</i> , baik pada aplikasi MyPertamina maupun <i>Jobstreet</i> . Keduanya juga sama-sama menggunakan algoritma klasifikasi.	Konteks Aplikasi: Meskipun kedua penelitian fokus pada aplikasi mobile, aplikasi MyPertamina lebih terkait dengan layanan bahan bakar, sementara <i>Jobstreet</i> adalah aplikasi yang berfokus pada pencarian kerja. Hal ini mempengaruhi jenis sentimen yang diungkapkan dalam

			ulasan. Ulasan MyPertamina mungkin lebih banyak berfokus pada masalah teknis dan kemudahan penggunaan, sedangkan ulasan <i>Jobstreet</i> lebih mungkin mencakup aspek pengalaman pencarian kerja, layanan dukungan, dan informasi lowongan.
2.	Syahroni dan Abd Wahab (2023) menulis jurnal berjudul "Analisis Sentimen Komentar Mahasiswa Terhadap Dosen Mata Kuliah Pada Aplikasi SIMAT"	Penelitian jurnal terkait dengan penelitian ini sama-sama menggunakan algoritma <i>Naive Bayes</i> digunakan sebagai salah satu pendekatan untuk menganalisis data.	Perbedaan dalam konteks penerapan analisis sentimen. Penelitian jurnal terkait fokus pada lingkungan pendidikan, mengamati interaksi antara mahasiswa dan dosen untuk meningkatkan kualitas pengajaran. Sebaliknya, penelitian ini lebih berorientasi pada industri dan pengalaman pengguna terhadap layanan yang disediakan oleh aplikasi, dengan tujuan meningkatkan kepuasan pengguna. Keduanya menggunakan umpan balik sebagai dasar analisis, tetapi konteks dan fokusnya berbeda.
3.	Oktafani dan Prasetyaningrum (2022) menulis jurnal berjudul "Implementasi <i>Support Vector Machine</i> Untuk Analisis Sentimen Komentar Aplikasi"	Kesamaan antara kedua penelitian ini adalah pentingnya pemahaman sentimen pengguna sebagai salah satu indikator utama kualitas aplikasi. Dalam aplikasi tanda tangan digital maupun aplikasi	Fokus penelitian kedua jurnal berbeda. Jurnal terkait berfokus pada analisis sentimen komentar pengguna terhadap aplikasi tanda tangan digital, sedangkan penelitian ini lebih spesifik pada analisis

	Tanda Tangan Digital”	<i>Jobstreet</i> , pengembang dapat memanfaatkan hasil analisis sentimen untuk mengidentifikasi masalah, memperbaiki fitur aplikasi, dan merancang strategi pengembangan lebih lanjut yang sesuai dengan harapan pengguna.	sentimen berdasarkan rating bintang pengguna aplikasi <i>Jobstreet</i> di <i>Google Play Store</i> . Hal ini menunjukkan bahwa konteks aplikasi yang dianalisis memiliki pengaruh pada metodologi yang digunakan dan hasil yang dicapai.
4.	Pratmanto, Imaniawan, dan Maarif (2023) menulis jurnal berjudul "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Identitas Kependudukan Digital Dengan Metode <i>Naive Bayes</i> Dan K-Nearest"	Kedua penelitian menggunakan <i>Naive Bayes</i> sebagai salah satu metode klasifikasi utama dalam analisis sentimen. Metode ini populer dalam pengolahan teks dan analisis sentimen karena kesederhanaannya dan keefektifannya dalam banyak kasus.	Penelitian peneliti membandingkan antara <i>Naive Bayes</i> dan <i>Support Vector Machine (SVM)</i> , dua metode yang sangat umum digunakan dalam klasifikasi teks.  Sebaliknya, Pratmanto et al. membandingkan <i>Naive Bayes</i> dengan K-Nearest Neighbors (KNN), yang fokus pada klasifikasi berdasarkan kedekatan antar data.
5.	Angraina dan Putri (2022) menulis jurnal berjudul "Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma <i>Support Vector Machine</i> "	Keduanya menggunakan analisis sentimen untuk mengklasifikasikan opini pengguna menjadi kategori sentimen (positif dan negatif ). Hal ini menunjukkan bahwa kedua penelitian menitikberatkan pada pemahaman bagaimana pengguna merasakan aplikasi yang mereka gunakan.	Penelitian Angraina dan Putri berfokus pada aplikasi Google Meet, sementara penelitian peneliti meneliti aplikasi <i>Jobstreet</i> , platform pencari kerja. Fokus berbeda ini dapat mempengaruhi pola sentimen dan ulasan pengguna yang dihasilkan karena tujuan dan jenis interaksi pengguna pada kedua aplikasi tersebut tidak sama.

### BAB III

## METODOLOGI PENELITIAN



#### A. Sumber Data

Data yang digunakan dalam penelitian ini adalah ulasan pengguna aplikasi *Jobstreet* yang diambil dari *Google Play Store*. Ulasan ini terdiri dari teks ulasan dan rating bintang yang diberikan oleh pengguna. Pengambilan data akan dilakukan secara otomatis menggunakan *Google Play Store Scraper* sejumlah 3000 data dengan kolom *Score*, *Content*, dan *Date*.

#### B. Alat dan Bahan

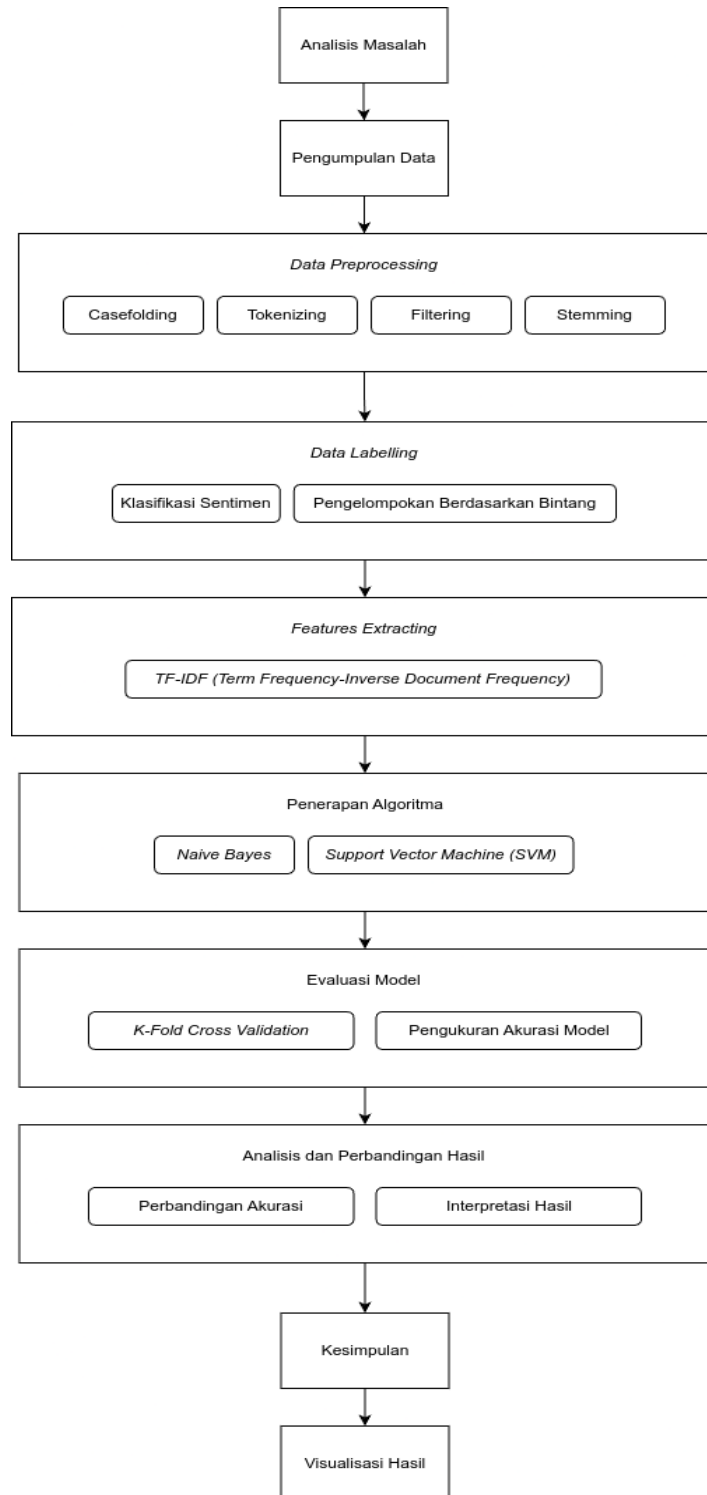
Penelitian ini menggunakan alat dan bahan guna penunjang dalam menyelesaikan penelitian tentang “Analisis Sentimen Pengguna Aplikasi *Jobstreet* di Indonesia pada *Google Play Store* Menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*: Perbandingan Akurasi Berdasarkan Rating Bintang”. Berikut alat dan bahan yang digunakan dalam penelitian ini:

**Tabel 3.1**  
**Spesifikasi Alat dan Bahan**

No	Jenis	Gambar	Keterangan
1.	<i>Linux Mint</i> 21 Cinnamon		<i>Linux Mint</i> mendukung alat-alat seperti <i>Python</i> serta pustaka untuk data science dan analisis sentimen ( <i>Scikit-learn</i> untuk <i>Naive Bayes</i> dan <i>SVM</i> ).
2.	<i>Python</i> 3.10.12		<i>Python</i> menyediakan alat dan pustaka yang memudahkan peneliti untuk mengolah data, melakukan pra-pemrosesan teks, serta melatih model algoritma <i>Naive Bayes</i> dan

			<i>SVM</i> untuk analisis sentimen.
3.	<i>Visual Studio Code</i> 1.94.1		<i>VSCode</i> dapat menjadi alat utama bagi peneliti untuk menulis, mengelola, dan menjalankan kode <i>Python</i> yang dibutuhkan untuk memproses data ulasan.
4.	<i>LibreOffice Suite</i> 24.8.1.2		<i>LibreOffice Suite</i> digunakan sebagai alat bantu untuk menangani dokumen atau data hasil scraping dari <i>Google Play Store</i> diekspor dalam CSV.
5.	<i>Google Play Store</i>		<i>Google Play Store</i> menjadi sumber data yang penting karena pengguna dapat memberikan ulasan dan rating untuk aplikasi tersebut.
6.	<i>Jobstreet</i>		<i>Jobstreet</i> adalah platform pencarian kerja yang sangat dikenal di Asia Tenggara, termasuk Indonesia. Melalui aplikasi ini, pengguna dapat dengan mudah mencari lowongan pekerjaan, mengirimkan lamaran, dan mengelola profil mereka secara online. Penelitian ini bertujuan untuk mengevaluasi bagaimana pengguna merasakan aplikasi <i>Jobstreet</i> berdasarkan ulasan yang mereka tinggalkan.

### C. Tahapan Penelitian



Gambar 3.1 Tahapan Penelitian

Penelitian ini dimulai dari 15 Oktober 2024 dan direncanakan selesai pada bulan Maret 2025. Rincian kegiatan penelitian dan waktu pelaksanaannya dijelaskan dalam tabel berikut:

**Tabel 3.2**  
***Timeline Penelitian***

No	Kegiatan	Bulan ke:					
		1	2	3	4	5	6
1	Pencarian dan penetapan penelitian						
2	Pengumpulan Data						
3	<i>Data Preprocessing</i>						
4	<i>Data Labelling</i>						
5	<i>Features Extracting</i>						
6	Penerapan Algoritma						
7	Evaluasi Model						
8	Analisis dan Perbandingan Hasil						
9	Kesimpulan						
10	Visualisasi Hasil						

Tahapan penelitian dari penelitian ini adalah sebagai berikut:

#### 1. Analisis Masalah

Pada tahap awal, peneliti menemukan bahwa peringkat bintang di *Google Play Store* tidak cukup menggambarkan sentimen orang Indonesia terhadap aplikasi *Jobstreet*, karena ulasannya bersifat global dan peringkat netral (bintang 3) menambah kebingungan. Oleh karena itu, diperlukan analisis sentimen yang lebih mendalam dengan menghapus nilai netral, sehingga dapat memfokuskan pada sentimen positif dan negatif untuk memahami kepuasan atau ketidakpuasan pengguna Indonesia dengan lebih akurat.

#### 2. Pengumpulan Data

Pengumpulan data dalam penelitian ini yaitu mengumpulkan ulasan pengguna aplikasi *Jobstreet* dari *Google Play Store*. Data ulasan meliputi teks ulasan, rating bintang, dan tanggal ulasan.

#### 3. Data Preprocessing

Tahap membersihkan data menggunakan teknik casefolding, tokenizing, filtering, stemming. Berikut penjelasannya:

- a. Casefolding: Menghapus tanda baca yang tidak dibutuhkan.
- b. Tokenisasi: Memecah teks ulasan menjadi kata-kata.
- c. Penghapusan Stop Words: Menghapus kata-kata umum yang tidak berkontribusi pada analisis sentimen.
- d. Stemming atau Lemmatization: Mengubah kata-kata menjadi bentuk dasar mereka untuk menyederhanakan analisis.

#### 4. Data Labelling

Berikut proses *Data Labelling* dalam penelitian ini:

- a. Klasifikasi Sentimen: Setiap ulasan akan dikategorikan berdasarkan sentimen positif dan negatif dengan mempertimbangkan isi teks dan rating bintang.



- b. Pengelompokan Berdasarkan Bintang: Data dibagi ke dalam kelompok berdasarkan rating bintang untuk perbandingan akurasi antara berbagai kelompok.

#### 5. *Features Extracting*

*Features Extracting* dalam penelitian ini menggunakan teknik seperti *TF-IDF* (*Term Frequency-Inverse Document Frequency*) untuk mengubah teks menjadi representasi numerik yang dapat digunakan oleh algoritma pembelajaran mesin.

#### 6. Penerapan Algoritma

Berikut penerapan algoritma dalam penelitian ini:

- a. *Naive Bayes*: Menerapkan algoritma *Naive Bayes* untuk klasifikasi sentimen ulasan.
- b. *Support Vector Machine (SVM)*: Menerapkan algoritma *SVM* untuk klasifikasi sentimen yang sama.

#### 7. Evaluasi Model

Berikut evaluasi model dalam penelitian ini:

- a. *Cross Validation*: Menggunakan *k-fold cross validation* untuk mengevaluasi kinerja masing-masing model.
- b. Mengukur performa model menggunakan metrik seperti akurasi.

#### 8. Analisis dan Perbandingan Hasil

Berikut analisis dan perbandingan hasil dalam penelitian ini:

- a. Perbandingan Akurasi: Menganalisis perbandingan akurasi antara algoritma *Naive Bayes* dan *SVM* pada setiap kelompok data berdasarkan rating bintang.
- b. Interpretasi Hasil: Menjelaskan bagaimana masing-masing algoritma bekerja pada ulasan dengan rating yang berbeda, dan mengapa ada perbedaan akurasi di antara keduanya.

## 9. Kesimpulan

Menyimpulkan algoritma mana yang memberikan hasil terbaik untuk analisis sentimen pengguna aplikasi *Jobstreet* berdasarkan rating bintang.

## 10. Visualisasi Hasil

Visualisasi hasil menggunakan Diagram Garis (Line Chart). Diagram garis dapat digunakan untuk menunjukkan tren akurasi sepanjang setiap hasil. Tahapan Penelitian ini dirancang untuk memastikan setiap langkah dari pengumpulan data hingga evaluasi model dilakukan secara sistematis dan menghasilkan analisis yang dapat dipertanggungjawabkan.

## DAFTAR PUSTAKA

- Abu Samah, K. A. F., Dinnie Wirakarnain, N. S., Hamzah, R., Mocketar, N. A., Riza, L. S., & Othman, Z. (2022). *A Linear Regression Approach to Predicting Salaries With Visualizations of Job Vacancies: A Case Study of Jobstreet Malaysia*. 11(3), 1130. <https://doi.org/10.11591/ijai.v11.i3.pp1130-1142>
- Akbar, M. N., & Nirwana Samrin, N. (2023). *Analisis Sentimen Komentar Pengguna Aplikasi Threads Pada Google Playstore Menggunakan Algoritma Multinomial Naive Bayes Classifier*. 3(2), 21–29. <https://doi.org/10.24252/jagti.v3i2.67>
- Angraina, D., & Putri, A. (2022). *Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine*. 3(3), 472–478. <https://doi.org/10.37859/coscitech.v3i3.4260>
- Elistiana, K. M., Bagus Adhi Kusuma, N., Subarkah, P., & Awal Rozaq, H. A. (2023). *Improvement of Naive Bayes Algorithm in Sentiment Analysis of Shopee Application Reviews on Google Play Store*. 4(6), 1431–1436. <https://doi.org/10.52436/1.jutif.2023.4.6.1486>
- Erawati, N. L., Adi Jaya, I. K. N., & Asmarajaya, I. K. A. (2023). *Visualisasi Data Pemetaan Nasabah Kredit Pada LPD Desa Adat Kesiman Menggunakan Python*. 2(1), 129–138. <https://doi.org/10.32795/resi.v1i3.4266>
- Fibriyanti Arminda, N., Sulistiyowati, N., & Nur Padilah, T. (2023). *Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo*. 7(3), 1817–1822. <https://doi.org/10.36040/jati.v7i3.7012>
- Gishella Septania Al-Husna, N., Dian Asmarajati, N., Iman Ahmad Ihsannuddin, N., & Rina Mahmudati, N. (2024). *Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi LinkedIn*. 3(2), 139–144. <https://doi.org/10.55123/storage.v3i2.3602>
- Han, S., & Kwak, I.-Y. (2023). *Mastering Data Visualization With Python: Practical Tips for Researchers*. 26(4), 167–175. <https://doi.org/10.7602/jmis.2023.26.4.167>

- Hayati, N., Suryono, S., & Widodo, C. E. (2019). Fog computing uses Radial Basis Function for Power Production Classification Solar Panel Electricity. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC* 2019, 14–19. <https://doi.org/10.1109/ICIC47613.2019.8985702>
- Karo Karo, I. M., Romia, R., Dewi, S., & Fadilah, P. M. (2023). Hoax Detection on Indonesian Tweets using Naïve Bayes Classifier with TF-IDF. *Journal of Information System Research (JOSH)*, 4(3), 914–919. <https://doi.org/10.47065/josh.v4i3.3317>
- Maulana, I., Apriandari, W., & Pambudi, A. (2023). *Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi MyPertamina Menggunakan Support Vector Machine*. 6(2), 172–181. <https://doi.org/10.36080/idealism.v6i2.3022>
- Maulana, R., Voutama, A., & Ridwan, T. (2023). *Analisis Sentimen Ulasan Aplikasi MyPertamina Pada Google Play Store Menggunakan Algoritma NBC*. 9(1), 42–48. <https://doi.org/10.54914/jtt.v9i1.609>
- Oktafani, M., & Prasetyaningrum, P. T. (2022). *Implementasi Support Vector Machine Untuk Analisis Sentimen Komentar Aplikasi Tanda Tangan Digital*. 15(1), 10–19. <https://doi.org/10.33005/sibc.v15i1.4>
- Pratmanto, D., Imaniawan, F. F. D., & Maarif, V. (2023). *Analisis Sentimen Pada Ulasan Pengguna Aplikasi Identitas Kependudukan Digital Dengan Metode Naive Bayes Dan K-Nearest*. 7(2), 155–166. <https://doi.org/10.24912/computatio.v7i2.26322>
- Putri, D. S., Sulistiyowati, N., & Voutama, A. (2023). *Analisis Sentimen Dan Pemodelan Ulasan Aplikasi AdaKami Menggunakan Algoritma SVM Dan KNN*. 9(2), 209–225. <https://doi.org/10.33050/sensi.v9i2.2914>
- Rahmany, M., Zin, A. M., & Sundararajan, E. A. (2020). *Comparing Tools Provided by Python and R for Exploratory Data Analysis*. 4(3), 131. <https://doi.org/10.56327/ijisecs.v4i3.933>
- Raja, J. H. L., & Tileng, K. G. (2022). *Analisis Kualitas Website Jobstreet Menggunakan Metode Webqual 4.0 Dan Importance-Performance Analysis (IPA)*. 8(1), 38–45. <https://doi.org/10.37715/juisci.v8i1.2623>
- Safryda Putri, D., & Ridwan, T. (2023). *Analisis Sentimen Ulasan Aplikasi Pospay Dengan Algoritma Support Vector Machine*. 11(01), 32–40.

<https://doi.org/10.33884/jif.v11i01.6611>

- Saputra, N. A., Aeni, K., & Saraswati, N. M. (2024). Indonesian Hate Speech Text Classification Using Improved K-Nearest Neighbor with TF-IDF-ICSpF. *Scientific Journal of Informatics*, 11(1), 21–30. <https://doi.org/10.15294/sji.v11i1.48085>
- Susanto, E. B., Paminto Agung Christianto, N., Mohammad Reza Maulana, N., & Satriedi Wahyu Binabar, N. (2022). *Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah*. 3(3), 234–241. <https://doi.org/10.37859/coscitech.v3i3.4343>
- Syahroni, A. wahab. (2023). *Analisis Sentimen Komentar Mahasiswa Terhadap Dosen Mata Kuliah Pada Aplikasi SIMAT*. 18(2). <https://doi.org/10.33998/processor.2023.18.2.1447>
- Tri Sanudin, F., Irawan, B., & Bahtiar, A. (2024). *Analisis Sentimen Ulasan Aplikasi Sapawarga Di Playstore Menggunakan Algoritma Naïve Bayes*. 8(1), 170–175. <https://doi.org/10.36040/jati.v8i1.8300>
- Valadez Estrada, R. A. (2023). *Iniciando Con LibreOffice*. 13–38. <https://doi.org/10.61728/ae24260022>
- Wahyu Sejati, N., Ankur Singh Bist, N., & Amirsyah Tambunan, N. (2023). *Pengembangan Analisis Sentimen Dalam Rekayasa Software Engineering Menggunakan Tinjauan Literatur Sistematis*. 2(1), 95–103. <https://doi.org/10.33050/mentari.v2i1.377>
- Wahyudi, R., & Kusumawardana, G. (2021). *Analisis Sentimen Pada Aplikasi Grab Di Google Play Store Menggunakan Support Vector Machine*. 8(2), 200–207. <https://doi.org/10.31294/ji.v8i2.9681>
- Wahyudi, W., Kurniawan, R., & Arie Wijaya, Y. (2024). *Analisis Sentimen Pengguna Terhadap Aplikasi Blu Bca Di Playstore Menggunakan Algoritma Naïve Bayes*. 8(3), 2511–2517. <https://doi.org/10.36040/jati.v8i3.9216>
- Widodo, B. K., Matondang, N. H., & Prasvita, D. S. (2022). *Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Penggunaan Aplikasi Jobstreet*. 21(3), 523–533. <https://doi.org/10.33633/tc.v21i3.6361>
- Wijanarto, W., & Brilianti, S. P. (2020). *Peningkatan Performa Analisis Sentimen Dengan Resampling Dan Hyperparameter Pada Ulasan Aplikasi BNI Mobile*. 9(2), 140–153. <https://doi.org/10.30864/eksplora.v9i2.333>

Yunanda, G., Nurjanah, D., & Meliana, S. (2022). Recommendation System from Microsoft News Data using TF-IDF and Cosine Similarity Methods. *Building of Informatics, Technology and Science (BITS)*, 4(1), 277–284. <https://doi.org/10.47065/bits.v4i1.1670>

## **LAMPIRAN**