

# **Perbandingan Support Vector Machine (SVM) dan Naïve Bayes pada Analisis Sentimen**

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada  
Jurusan Teknik Informatika



Oleh :

Karina  
NIM : 09021181621009

**Jurusan Teknik Informatika  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2021**

**LEMBAR PENGESAHAN SKRIPSI**

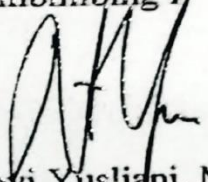
**PERBANDINGAN *SUPPORT VECTOR MACHINE (SVM)* DAN  
*NAÏVE BAYES* PADA ANALISIS SENTIMEN**

Oleh :

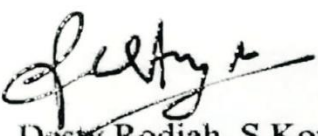
**Karina**  
**NIM : 09021181621009**

Palembang, 5 Juli 2021

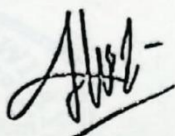
Pembimbing I

  
Novi Yusliani, M.T.  
NIP. 198211082012122001

Pembimbing II

  
Desty Rodiah, S.Kom., M.T.  
NIP. 198912212020122011

Mengetahui,  
Ketua Jurusan Teknik Informatika

  
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

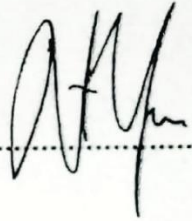
## TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Sabtu tanggal 31 Juli 2021 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Karina  
NIM : 09021181621009  
Judul : Perbandingan *Support Vector Machine (SVM)* dan *Naïve Bayes* Pada Analisis Sentimen

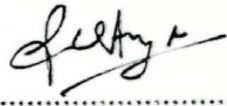
### 1. Pembimbing I

Novi Yusliani, M.T.  
NIP. 198211082012122001

  
.....

### 2. Pembimbing II

Desty Rodiah, M.T.  
NIP. 198912212020122011

  
.....

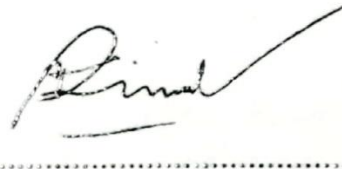
### 3. Penguji I

Dian Palupi Rini, M.Kom., Ph.D.  
NIP. 197802232006042002


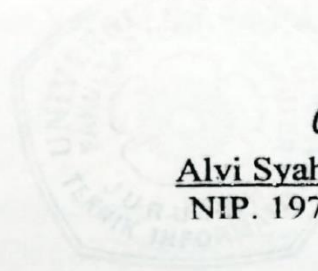
  
.....

### 4. Penguji II

Mastura Diana Marieska, M.T.  
NIP. 198603212018032001

  
.....

Mengetahui,  
Ketua Jurusan Teknik Informatika

  
Alvi Syahrini Utami, M.Kom  
NIP. 197812222006042003

## HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Karina  
NIM : 09021181621009  
Program Studi : Teknik Informatika  
Judul Skripsi : Perbandingan *Support Vector Machine (SVM)* dan *Naïve Bayes* pada Analisis Sentimen  
Hasil Pengecekan *Software Ithenticate/Turnitin* : 17 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, Agustus 2021



Karina  
NIM. 09021181621009

## MOTTO DAN PERSEMBAHAN

- *"Hai orang-orang yang beriman, jadikanlah sabar dan shalat sebagai penolongmu, sesungguhnya Allah beserta orang-orang yang sabar."*

*(Q.S. Al-Baqarah : 153)*

**Kupersembahkan Karya Tulis ini kepada :**

- **Orang Tuaku**
- **Kakak dan Mbakku**
- **Dosen Pembimbing**
- **Teman Seperjuanganku IF REG A 2016**
- **Fakultas Ilmu Komputer, Universitas Sriwijaya**

**COMPARISON OF SUPPORT VECTOR MACHINE (SVM) AND NAIVE  
BAYES IN SENTIMENT ANALYSIS**

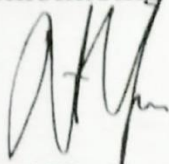
**Karina  
09021181621009**

**ABSTRACT**

*Youtube is a social media that provides many shows, one of which is movie Trailers. A film Trailer is a short Trailer from a film that is used as a promotional medium. Through the comments on the movie Trailer, it can be seen whether the film that will be shown is good to watch or not. So a system is needed to analyze comments to be positive and negative. In this study, the Support Vector Machine and Naïve Bayes methods were used. SVM works by defining a hyperplane that maximizes the margin between two different classes. Meanwhile, Naïve Bayes is a method that is simple and easy to process by applying probability theory to find the greatest probability of classification. Through the advantages of each method, the researcher makes a comparison to see the best accuracy of the two methods in conducting sentiment analysis on Youtube comment data. The results of the test show that SVM has a better performance with an accuracy value of 86% while Naïve Bayes has an accuracy of 46%.*

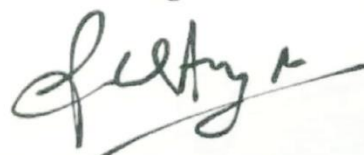
*Keywords : Sentiment analysis, Comments, Youtube, Support Vector Machine, Naive Bayes*

Pembimbing I



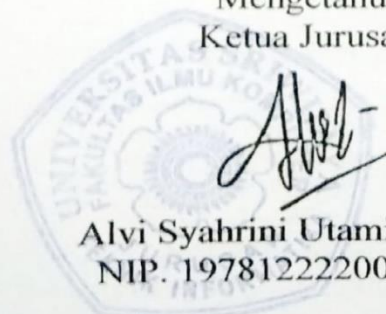
Novi Yustiani, M.T.  
NIP.198211082012122001

Palembang, Agustus 2021  
Pembimbing II



Desty Rodiah, M.T.  
NIP. 198912212020122011

Mengetahui,  
Ketua Jurusan,



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

# PERBANDINGAN *SUPPORT VECTOR MACHINE (SVM)* DAN *NAÏVE BAYES* PADA ANALISIS SENTIMEN

Karina  
09021181621009

## ABSTRAK

*Youtube* merupakan media sosial yang menyediakan banyak tayangan salah satunya trailer film. Trailer film ialah cuplikan singkat dari film yang digunakan sebagai media promosi. Melalui komentar pada trailer film, bisa dilihat apakah film yang akan ditayangkan baik untuk ditonton atau tidak. Sehingga dibutuhkan sistem untuk menganalisa komentar menjadi positif dan negatif. Dalam penelitian ini, digunakan metode *Support Vector Machine* dan *Naïve Bayes*. *SVM* bekerja dengan mendefinisikan hyperplane yang memaksimalkan margin antara dua kelas yang berbeda. Sedangkan, *Naïve Bayes* merupakan metode yang sederhana dan mudah diproses dengan menerapkan teori probabilitas untuk mencari peluang kemungkinan terbesar dari klasifikasi. Melalui kelebihan masing-masing metode peneliti melakukan perbandingan untuk melihat akurasi terbaik dari kedua metode dalam melakukan analisis sentimen pada data komentar *Youtube*. Hasil dari pengujian menunjukkan bahwa *SVM* memiliki kinerja yang lebih baik dengan nilai akurasi sebesar 86% sedangkan *Naïve Bayes* memiliki akurasi sebesar 46%.

Kata Kunci : Analisis sentimen, Komentar, *Youtube*, *Support Vector Machine*, *Naïve Bayes*

Pembimbing I

Novi Yustiani, M.T.  
NIP.198211082012122001

Palembang, Agustus 2021  
Pembimbing II

Desty Rodiah, M.T.  
NIP. 198912212020122011

Mengetahui,  
Ketua Jurusan,



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## KATA PENGANTAR



Puji syukur kepada Allah SWT atas segala rahmat dan karunia-Nya penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 Program Studi Teknik Informatika pada Fakultas Ilmu Komputer, Universitas Sriwijaya.

Penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah memberikan bantuan dan dukungan baik dalam materil maupun moril selama proses pembuatan tugas akhir ini. Adapun penulis merujuk secara khusus pihak yang telah membantu sebagai berikut :

1. Ayahku Sumarlan dan Ibuku Rustinah, selaku orang tuaku tersayang. Kakakku Gunawan Wibisono, dan Mbakku Setyorini yang selalu memberikan dukungan penuh. Serta keponakan-keponakanku yang selalu menjadi teman serta penghibur dikala sulit.
2. Bapak Jaidan Jauhari, S.pd., M.T. selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika.
4. Bapak Rifkie Primartha, M.T. selaku pembimbing akademik yang telah memberikan bimbingan serta arahan dalam urusan akademik.
5. Ibu Novi Yusliani, M.T. dan Ibu Desty Rodiah, S.Kom.,M.T. selaku pembimbing I dan pembimbing II yang telah memberikan arahan, bimbingan, kritik dan saran kepada penulis untuk menyelesaikan tugas akhir.
6. Ibu Dian Palupi Rini, M.Kom, Ph.D dan Ibu Mastura Diana Marieska, M.T selaku Dosen Penguji I dan Dosen Penguji II yang telah memberikan masukan dan arahan kepada penulis dalam menyempurnakan tugas akhir ini.
7. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika yang telah memberikan ilmu kepada penulis selama masa perkuliahan.



8. Staff Jurusan Teknik Informatika beserta staff Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi dan akademik selama masa perkuliahan.
9. Sahabat sekaligus teman kosanku, Anis dan Fadhillah. Terima kasih sudah bersedia menemani dan mau berbagi keluh kesah selama masa kos dan perkuliahan ini. Semoga Allah melancarkan urusan kita dan senantiasa mengeratkan persahabatan kita.
10. Sahabat-sahabat ku selama masa perkuliahan, Dina dan Tini. Terima kasih sudah senantiasa menemani selama masa perkuliahan, memberikan semangat serta motivasi dalam proses penulisan tugas akhir ini.
11. Teman-teman Teknik Informatika Reguler 2016, terima kasih atas jalinan pertemanannya selama ini, dan senantiasa berbagi informasi bersama. Semoga Allah memudahkan untuk menyelesaikan studi S-1 ini.
12. Semua pihak lainnya yang tidak dapat disebutkan satu persatu yang telah berperan bagi penulis dalam menyelesaikan tugas akhir, Terima Kasih banyak atas semuanya.

Semoga tugas akhir ini dapat bermanfaat bagi penulis khususnya maupun pembaca pada umumnya. Serta dapat menjadi referensi dan rujukan bagi hal-hal yang bermanfaat. Penulis menyadari bahwa dalam proses penyelesaian tugas akhir ini, terdapat beberapa penjelasan yang kurang sempurna. Oleh karena itu penulis mengharapkan kritik dan saran yang membangun agar tugas akhir ini menjadi karya tulis yang sempurna supaya terciptanya bekal pengetahuan yang baik bagi peneliti di masa depan.

Palembang, Agustus 2021

Karina

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
TANDA LULUS UJIAN SIDANG SKRIPSI .....	iii
MOTTO DAN PERSEMBAHAN .....	v
ABSTRACT .....	vi
ABSTRAK .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	x
DAFTAR TABEL .....	xiv
DAFTAR GAMBAR .....	xvi
DAFTAR LAMPIRAN .....	xvii
BAB I PENDAHULUAN .....	I-18
1.1 Pendahuluan .....	I-18
1.2 Latar Belakang Masalah .....	I-18
1.3 Rumusan Masalah .....	I-21
1.4 Tujuan Penelitian .....	I-21
1.5 Manfaat Penelitian .....	I-22
1.6 Batasan Masalah .....	I-22
1.7 Sistematika Penulisan .....	I-23
1.7.1 Bab I. Pendahuluan .....	I-23
1.7.2 Bab II. Kajian Literatur .....	I-23
1.7.3 Bab III. Metode Penelitian .....	I-23
1.7.4 Bab IV. Pengembangan Perangkat Lunak .....	I-23
1.7.5 Bab V. Hasil Dan Analisa Penelitian .....	I-24
1.7.6 Bab VI. Kesimpulan Dan Saran .....	I-24
1.8 Kesimpulan .....	I-24
BAB II KAJIAN LITERATUR .....	II-
2.1 Pendahuluan .....	II-
2.2 Landasan Teori .....	II-

2.2.1 Analisis Sentimen.....	II-
2.2.2 Prapengolahan Teks .....	II-
2.2.3 Pembobotan Term Frequency - Inverse Document Frequency (TF-IDF).....	II-
2.2.4 Naïve Bayes.....	II-
2.2.5 Support Vector Machine (SVM).....	II-
2.2.6 Pengukuran Hasil Klasifikasi .....	II-
2.2.7 Rational Unified Process (RUP) .....	II-
2.3 Penelitian Lain yang Relevan.....	II-
2.4 Kesimpulan.....	II-
BAB III METODOLOGI PENELITIAN.....	III-
3.1 Pendahuluan .....	III-
3.2 Unit Penelitian.....	III-
3.3 Data .....	III-1
3.3.1 Jenis dan Sumber Data .....	III-
3.3.2 Metode Pengumpulan Data .....	III-
3.4 Tahapan Penelitian .....	III-
3.4.1 Menetapkan Kerangka Kerja/ Framework .....	III-
3.4.2 Menetapkan Kriteria Pengujian.....	III-
3.4.3 Menetapkan Format Data Pengujian .....	III-
3.4.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-
3.4.5 Melakukan Pengujian Penelitian.....	III-
3.4.6 Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-
3.5 Metode Pengembangan Perangkat Lunak Rational Unified Process ...	III-
3.5.1 Fase Insepsi .....	III-
3.5.2 Fase Elaborasi .....	III-
3.5.3 Fase Konstruksi .....	III-
3.5.4 Fase Transisi.....	III-
3.6 Manajemen Proyek Penelitian.....	III-
3.7 Kesimpulan.....	III-

BAB IV PENGEMBANGAN PERANGKAT LUNAK .....	IV-
4.1 Pendahuluan .....	IV-
4.2 Fase Insepsi .....	IV-
4.2.2 Pemodelan Bisnis .....	IV-
4.2.2 Kebutuhan Sistem .....	IV-
4.2.3.1 Fitur Penerima Data Input .....	IV-
4.2.3.2 Fitur Analisis Sentimen dengan SVM.....	IV-
4.2.3.3 Fitur Analisis Sentimen dengan Naïve Bayes .....	IV-
4.2.3 Analisis dan Desain.....	IV-
4.2.3.1 Analisis Kebutuhan Perangkat Lunak.....	IV-
4.2.3.2 Analisis Data .....	IV-
4.2.3.3 Analisis Prapengolahan Data .....	IV-
4.2.3.4 Analisis Proses Klasifikasi Menggunakan SVM .....	IV-
4.2.3.5 Analisis Proses Klasifikasi Menggunakan Naïve Bayes.....	IV-
4.2.3.6 Analisis Hasil Klasifikasi.....	IV-
4.2.4 Implementasi .....	IV-
4.3 Fase Elaborasi .....	IV-
4.3.1 Pemodelan Bisnis .....	IV-
4.3.3.1 Perancangan Data.....	IV-
4.3.3.2 Perancangan Antarmuka .....	IV-
4.3.2 Kebutuhan Sistem .....	IV-
4.3.3 Analisis dan Perancangan .....	IV-
4.3.3.1 Diagram Aktivitas .....	IV-
4.3.3.2 Diagram Alur .....	IV-
4.4 Fase Konstruksi.....	IV-
4.4.1 Kebutuhan Sistem .....	IV-
4.4.2 Diagram Kelas.....	IV-
4.4.3 Implementasi .....	IV-
4.4.3.1 Implementasi Kelas .....	IV-
4.4.3.2 Implementasi Antarmuka .....	IV-
4.5 Fase Transisi .....	IV-
4.5.1 Pemodelan Bisnis .....	IV-
4.5.2 Kebutuhan .....	IV-
4.5.3 Analisis dan Perancangan .....	IV-

4.5.3.1 Rencana Pengujian .....	IV-
4.5.4 Implementasi .....	IV-
4.6 Kesimpulan .....	IV-
BAB V HASIL DAN ANALISIS PENELITIAN.....	V-
5.1 Pendahuluan .....	V-
5.2 Data Hasil Penelitian.....	V-
5.2.1 Konfigurasi Percobaan .....	V-
5.2.1.1 Data Hasil Konfigurasi I .....	V-
5.2.1.2 Data Hasil Konfigurasi II.....	V-
5.2.1.3 Perbandingan Data Hasil Konfigurasi.....	V-
5.2.1.4 Data Hasil Kuisisioner .....	V-
5.3 Analisis Hasil Penelitian .....	V-
5.3.1. Analisis Rata-rata Hasil Klasifikasi .....	V-
5.3.2. Analisis Waktu Komputasi .....	V-
5.4 Kesimpulan .....	V-
BAB VI KESIMPULAN DAN SARAN .....	VI-
6.1 Pendahuluan.....	VI-
6.2 Kesimpulan .....	VI-
6.3 Saran .....	VI-
DAFTAR PUSTAKA .....	xviii

## DAFTAR TABEL

<b>Tabel II-1.</b> <i>Confusion Matrix</i> Untuk 2 Kelas .....	II-11
<b>Tabel III-1.</b> Contoh Data Komentar yang Dikumpulkan .....	III-1
<b>Tabel III-2.</b> Rancangan Tabel <i>Confusion Matrix</i> Hasil Analisis Sentimen ..	III-8
<b>Tabel III-3.</b> Rancangan Tabel Hasil Pengujian.....	III-9
<b>Tabel III-4.</b> Rancangan Tabel Analisis Klasifikasi.....	III-10
<b>Tabel III-5.</b> Tabel <i>WBS</i> dari penelitian yang akan dilakukan.....	III-14
<b>Tabel IV-1.</b> Kebutuhan Fungsional.....	IV-3
<b>Tabel IV-2.</b> Kebutuhan Non Fungsional .....	IV-3
<b>Tabel IV-3.</b> Teks Komentar .....	IV-6
<b>Tabel IV-4.</b> Tahap <i>Cleaning</i> .....	IV-6
<b>Tabel IV-5.</b> Tahap <i>Case Folding</i> .....	IV-7
<b>Tabel IV-6.</b> Tahap <i>Tokenizing</i> .....	IV-7
<b>Tabel IV-7.</b> Tahap <i>Filtering</i> .....	IV-8
<b>Tabel IV-8.</b> Hasil <i>Stemming</i> dari Data Komentar .....	IV-8
<b>Tabel IV-9.</b> Hasil Perhitungan Bobot <i>Term</i> dengan <i>TF-IDF</i> .....	IV-9
<b>Tabel IV-10.</b> Hasil Perhitungan Probabilitas <i>Term</i> .....	IV-14
<b>Tabel IV-11.</b> Definisi Aktor .....	IV-17
<b>Tabel IV-12.</b> Definisi <i>Use Case</i> .....	IV- 18
<b>Tabel IV-13.</b> Skenario <i>Use Case</i> Menerima Data Input .....	IV-19
<b>Tabel IV-14.</b> Skenario <i>Use Case</i> Melakukan Analisis Sentimen Menggunakan Algoritma <i>SVM</i> .....	IV-20
<b>Tabel IV-15.</b> Skenario <i>Use Case</i> Melakukan Analisis Sentimen Menggunakan Algoritma <i>Naïve Bayes</i> .....	IV-22
<b>Tabel IV-16.</b> Implementasi Kelas .....	IV-35
<b>Tabel IV-17.</b> Rencana Pengujian <i>Use Case</i> Menerima Data <i>Input</i> .....	IV-38
<b>Tabel IV-18.</b> Rencana Pengujian <i>Use Case</i> Melakukan Analisis Sentimen	

Menggunakan Algoritma <i>SVM</i> .....	IV-39
<b>Tabel IV-19.</b> Rencana Pengujian Use Case Melakukan Analisis Sentimen	
Menggunakan Algoritma <i>Naïve Bayes</i> .....	IV-39
<b>Tabel IV-20.</b> Pengujian Use Case Menerima Data <i>Input</i> .....	IV-40
<b>Tabel IV-21.</b> Pengujian Use Case Melakukan Analisis Sentimen	
Menggunakan Algoritma <i>SVM</i> .....	IV-41
<b>Tabel IV-22.</b> Rencana Pengujian Use Case Melakukan Analisis Sentimen	
Menggunakan Algoritma <i>Naïve Bayes</i> .....	IV-42
<b>Tabel V-1.</b> <i>Confusion Matrix</i> Klasifikasi <i>SVM</i> .....	V-2
<b>Tabel V-2.</b> Data Hasil Evaluasi Klasifikasi <i>SVM</i> .....	V-3
<b>Tabel V-3.</b> <i>Confusion Matrix</i> Klasifikasi <i>Naïve Bayes</i> .....	V-4
<b>Tabel V-4.</b> Data Hasil Evaluasi Klasifikasi <i>Naïve Bayes</i> .....	V-5
<b>Tabel V-5.</b> Perbandingan Rata-rata Hasil Klasifikasi .....	V-7
<b>Tabel V-6.</b> Data Hasil Akumulasi Pengisian Kuisisioner oleh Partisipan .....	V-8

## DAFTAR GAMBAR

<b>Gambar II-1.</b> Ilustrasi <i>Naive Bayes</i> .....	II-4
<b>Gambar II-2.</b> <i>Hyperplane Support Vector Machine (SVM)</i> .....	II-8
<b>Gambar II-3.</b> <i>Arsitektur Rational Unified Process (RUP)</i> .....	II-13
<b>Gambar III-1.</b> Proses Pemilihan Komentar .....	III-3
<b>Gambar III-2.</b> Format Kuisisioner .....	III-4
<b>Gambar III-3.</b> Diagram Alir Sistem Analisis Sentimen .....	III-6
<b>Gambar IV-1.</b> Diagram <i>Use Case</i> .....	IV-17
<b>Gambar IV-2.</b> Rancangan Antar Muka .....	IV-24
<b>Gambar IV-3.</b> Diagram Aktivitas Menerima Data <i>Input</i> .....	IV-26
<b>Gambar IV-4.</b> Diagram Aktivitas Melakukan Analisis Sentimen Menggunakan Algoritma <i>SVM</i> .....	IV-27
<b>Gambar IV-5.</b> Diagram Aktivitas Melakukan Analisis Sentimen Menggunakan Algoritma <i>Naive Bayes</i> .....	IV-28
<b>Gambar IV-6.</b> Diagram Sequence Menerima Data <i>Input</i> .....	IV-29
<b>Gambar IV-7.</b> Diagram Sequence Melakukan Analisis Sentimen Menggunakan Algoritma <i>SVM</i> .....	IV-30
<b>Gambar IV-8.</b> Diagram Sequence Melakukan Analisis Sentimen Menggunakan Algoritma <i>Naive Bayes</i> .....	IV-31
<b>Gambar IV-9.</b> Diagram Kelas .....	IV-34
<b>Gambar IV-10.</b> Tampilan Antar Muka Halaman Utama Perangkat Lunak .	IV-37
<b>Gambar V-1.</b> Grafik Hasil Pengujian 10 Fold Menggunakan <i>SVM</i> .....	V-4
<b>Gambar V-2.</b> Grafik Hasil Pengujian 10 Fold Menggunakan <i>Naive Bayes</i> .	V-6
<b>Gambar V-3.</b> Data Hasil Pengisian Kuisisioner .....	V-8
<b>Gambar V-4.</b> Diagram Perbandingan Rata-rata Hasil Klasifikasi <i>SVM</i> dan <i>Naive Bayes</i> .....	V-20
<b>Gambar V-5.</b> Diagram Perbandingan Rata-rata Waktu Komputasi Metode Analisis Sentimen .....	V-22



## DAFTAR LAMPIRAN

<b>Lampiran 1.</b> Kode Program.....	L-1
--------------------------------------	-----

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Pendahuluan**

Bab ini menguraikan tentang latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, sistematika penulisan skripsi dan kesimpulan dari penelitian yang menjadi landasan penelitian.

#### **1.2 Latar Belakang Masalah**

*Youtube* adalah *platform* digital yang paling banyak diakses pengguna internet di Indonesia tahun 2020 (Junawan & Laugu, 2020). Para pengguna memanfaatkan *Youtube* sebagai media untuk berbagi dan melihat tayangan video. Beragam jenis tayangan dapat diunggah dan disaksikan dengan mudah di *Youtube*. Akses yang cepat dan mudah menjadikan *Youtube* banyak digunakan sebagai media promosi untuk meningkatkan jumlah tayangan. Salah satu tayangan yang digunakan untuk media promosi adalah trailer film (Safeii et al., 2017).

Trailer film merupakan cuplikan singkat dari sebuah film. Penonton dapat melihat gambaran hingga kualitas dari film yang akan ditayangkan. Trailer yang bagus akan menarik banyak minat penonton untuk menonton film, begitupun sebaliknya. Melalui penayangan trailer film pada *Youtube*, maka dapat dilihat sentimen publik terhadap trailer film tersebut. Publik akan menyampaikan sentimennya dalam bentuk komentar positif maupun negatif melalui fitur komentar. Melalui komentar yang disampaikan maka bisa dilihat bagaimana respon penonton terhadap film yang akan ditayangkan. Namun, jumlah dan ragam

komentar yang banyak cukup menyulitkan dalam proses menganalisa komentar. Karena itu, dibuatlah pendekatan menggunakan sistem analisis sentimen.

Analisis sentimen merupakan proses pengolahan data untuk mendapatkan kecenderungan suatu data (Saputra et al., 2012). Analisis sentimen mengelompokkan data kedalam kelas seperti kelas positif atau negatif. dengan tujuan agar bisa didapat informasi lebih detail mengenai data tersebut. Analisis sentimen dapat dilakukan dengan pendekatan *Machine Learning* (pembelajaran mesin). *Machine Learning* menggunakan data pelatihan dan data uji untuk melakukan klasifikasi. Data pelatihan digunakan untuk membangun model yang nantinya akan digunakan untuk mengklasifikasikan masukan data ke dalam kelasnya. Kemudian, data uji akan digunakan untuk memvalidasi model yang sudah dibuat dengan memprediksi hasil klasifikasi pada data yang belum memiliki kelas (Neethu & Rajasree, 2013). Algoritma yang banyak digunakan dalam klasifikasi adalah *Support Vector Machine (SVM)* dan *Naïve Bayes*.

*Naïve Bayes* merupakan metode yang menerapkan aturan Bayes dengan asumsi setiap fiturnya bersifat bebas/independen (Webb, 2016). Metode ini menggunakan teori probabilitas dengan mencari probabilitas tertinggi dalam proses klasifikasi pada data latih. *Naïve Bayes* merupakan metode yang sederhana dan mudah diimplementasikan sehingga memiliki waktu pemrosesan yang relatif cepat, serta memiliki tingkat efektifitas yang tinggi (Taheri & Mammadov, 2013).

Selain *Naïve Bayes*, *Support Vector Machine (SVM)* merupakan salah satu metode dalam algoritma pembelajaran mesin. *SVM* bekerja dengan mendefinisikan *hyperplane* (sebuah fungsi yang dapat digunakan untuk pemisah

antar kelas) terpisah yang memaksimalkan margin antara dua kelas yang berbeda. Dalam pemodelan klasifikasi, *SVM* memiliki konsep yang lebih matang dan lebih jelas dibanding algoritma yang lain. Hal ini membuat *SVM* banyak digunakan dalam penelitian (Zhou et al., 2016).

Seperti yang dilakukan (Budi & Mude, 2020) pada penelitian yang membahas tentang penerapan *Support Vector Machine (SVM)* dan *Naïve Bayes* untuk mengelompokkan ulasan teks. Penelitian ini membandingkan tingkat akurasi dari masing-masing metode terhadap pengelompokkan ulasan teks yang diambil dari platform *Google Play Store*. Hasil dari penelitian ini menunjukkan bahwa *SVM* lebih unggul dibanding *Naïve Bayes* dalam mengelompokkan ulasan teks dengan hasil pengujian akurasi masing-masing sebesar 81,46% dan 75,41%.

Selanjutnya penelitian yang dilakukan (Buntoro, 2017) terhadap data sentimen *Twitter*. Dalam penelitian ini, data sebanyak 300 tweet dalam Bahasa Indonesia diolah pada tahap *preprocessing* dengan tokenisasi, cleaning dan filtering. Kemudian, data dikelompokkan ke dalam kelas sentimennya menggunakan *Lexicon Based*. Selanjutnya, data diklasifikasikan dengan *Naïve Bayes* dan *SVM*. Setelah diklasifikasikan, diperoleh hasil dengan akurasi terbaik menggunakan *Naïve Bayes* berupa rata-rata akurasi mencapai 95%.

Melalui penelitian yang telah dilakukan dapat dilihat bahwa kedua penelitian membandingkan kedua metode yang sama yaitu *SVM* dan *Naïve Bayes*. Namun, pada penelitian yang dilakukan (Budi & Mude, 2020) hasil metode yang lebih unggul yakni *Support Vector Machine (SVM)* dengan akurasi 81,46% dan 75,41%. Sedangkan pada penelitian oleh (Buntoro, 2017), *Naïve Bayes*

mendapatkan hasil lebih unggul dengan nilai rata-rata akurasi 95%. Maka, berdasarkan hasil tersebut, peneliti akan melakukan perbandingan terhadap 2 metode klasifikasi yakni *SVM* dan *Naïve Bayes* terhadap analisis sentimen. Data yang akan digunakan pada penelitian ini merupakan data komentar yang diambil dari *Youtube*. Penelitian ini dilakukan untuk melihat akurasi yang dihasilkan kedua metode untuk mengetahui akurasi yang terbaik dari kedua metode tersebut.

### **1.3 Rumusan Masalah**

Berdasarkan latar belakang sebelumnya, maka rumusan masalah pada penelitian ini antara lain :

1. *Bagaimana mengembangkan sistem analisis sentimen menggunakan metode Support Vector Machine (SVM) ?*
2. *Bagaimana mengembangkan sistem analisis sentimen menggunakan metode Naïve Bayes ?*
3. *Bagaimana tingkat akurasi dari metode Support Vector Machine (SVM) pada analisis sentimen?*
4. *Bagaimana tingkat akurasi dari metode Naïve Bayes pada analisis sentimen?*

### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah sebelumnya, maka tujuan dari penelitian ini antara lain :

1. Menghasilkan sistem *analisis sentimen menggunakan metode Support Vector Machine (SVM)*.

2. Mengetahui sistem *analisis sentimen menggunakan metode Naïve Bayes*.
3. Mengetahui tingkat akurasi dari metode *Support Vector Machine (SVM)* pada analisis sentimen.
4. Mengetahui tingkat akurasi dari metode *Naïve Bayes* pada analisis sentimen.

### **1.5 Manfaat Penelitian**

Manfaat yang dapat diperoleh dari penelitian ini sebagai berikut :

1. Memberi informasi berupa akurasi metode yang terbaik dari *SVM* dan *Naïve Bayes* pada analisis sentimen.
2. Menjadi referensi pada penelitian selanjutnya khususnya mengenai algoritma *SVM* dan *Naïve Bayes*.

### **1.6 Batasan Masalah**

Batasan masalah pada penelitian ini adalah :

1. Data komentar dalam penelitian ini diambil dari Kanal *Youtube* trailer film Indonesia.
2. Data komentar yang diambil berupa komentar dalam Bahasa Indonesia.
3. Data komentar yang dipilih tidak berisi emoji ataupun emotikon.
4. Penelitian akan mengklasifikasikan komentar menjadi kategori positif dan negatif.

## **1.7 Sistematika Penulisan**

Berikut adalah sistematika dari penyusunan skripsi :

### **1.7.1 Bab I. Pendahuluan**

Bab ini membahas mengenai pokok pikiran yang menjadi landasan penelitian, seperti latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

### **1.7.2 Bab II. Kajian Literatur**

Bab ini membahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi analisis sentimen, proses pra-pengolahan, pembobotan *TF-IDF*, algoritma *Support Vector Machine* dan *Naive Bayes*, *Confusion Matrix*, *RUP*, serta beberapa kajian literatur mengenai penelitian lain yang relevan dengan penelitian ini.

### **1.7.3 Bab III. Metode Penelitian**

Bab ini membahas tahapan yang akan dilakukan pada penelitian, seperti pengumpulan data, analisis data, mekanisme pengujian serta perancangan sistem yang akan dibangun. Setiap tahapan penelitian dijelaskan secara rinci dan mengacu pada kerangka kerja.

### **1.7.4 Bab IV. Pengembangan Perangkat Lunak**

Bab ini membahas mengenai analisis dan perancangan perangkat lunak yang akan digunakan sebagai alat penelitian. Proses diawali dengan pengumpulan dan analisa kebutuhan, rancangan dan konstruksi perangkat

lunak serta pengujian untuk memastikan semua kebutuhan pengembangan perangkat lunak sesuai dengan kebutuhan.

#### **1.7.5 Bab V. Hasil Dan Analisa Penelitian**

Bab ini berisi uraian dari hasil pengujian yang telah dilakukan. Setiap hasil akan dianalisis dan diuraikan menggunakan tabel untuk memudahkan mengambil kesimpulan dari penelitian.

#### **1.7.6 Bab VI. Kesimpulan Dan Saran**

Bab ini berisi kesimpulan dari setiap uraian yang terdapat pada bab sebelumnya serta berisi saran yang diberikan berdasarkan hasil dari penelitian.

### **1.8 Kesimpulan**

Kesimpulan yang dapat diperoleh pada bab ini yaitu :

1. Pada bab ini telah diuraikan bahwa akan dilakukan penelitian mengenai perbandingan pada analisis sentimen.
2. Penelitian ini akan membandingkan analisis sentimen terhadap komentar *Youtube* dengan metode *Support Vector Machine (SVM)* dan *Naïve Bayes*.
3. Hasil pengujian berupa tingkat akurasi pada analisis sentimen terhadap komentar *Youtube* dengan menggunakan data uji dari analisis sentimen manual.



## DAFTAR PUSTAKA

- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*.
- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102–4107.
- Anwar, A. (2014). A Review of RUP (Rational Unified Process). *International Journal of Software Engineering*, 5(2), 8–24.
- Budi, L., & Mude, A. (2020). *Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store*. 12(2), 154–161.
- Deng, X., Liu, Q., Deng, Y., & Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, 340–341, 250–261.
- Huang, X., & Wu, Q. (2013). Micro-blog commercial word extraction based on improved TF-IDF algorithm. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*.
- Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Evolusi Volume 4 Nomor 2 - 2016*, 4(2), 20–27.

- Junawan, H., & Laugu, N. (2020). Eksistensi Media Sosial, Youtube, Instagram dan Whatsapp Ditengah Pandemi Covid-19 Dikalangan Masyarakat Virtual Indonesia. *Baitul 'Ulum: Jurnal Ilmu Perpustakaan Dan Informasi*, 4(1), 41–57.
- Kadhim, A. I. (2018). An Evaluation of Preprocessing Techniques for Text Classification. *International Journal of Computer Science and Information Security*, 16(6), 22–32.
- Kurniawan, M. F., Bayes, N., & Bayes, N. (2014). *KOMPARASI ALGORITMA DATA MINING UNTUK KLASIFIKASI PENYAKIT. 1*, 1–8.
- Nawaz, S., Rizwan, M., Yasin, S., Ahmed, M., & Farooq, U. (2020). *Multi-Class Classification of the YouTube Comments using Machine Learning PakJET*. 183–188.
- Nitin Rajvanshi, & K. R. Chowdhary,. (2017). Comparison of SVM and Naïve Bayes Text Classification Algorithms using WEKA. *International Journal of Engineering Research And*, V6(09).
- Safeii, A., Arif, M., Sultan, M. I., & Bahfiarti, T. (2017). *STRATEGI PROMOSI FILM ADA APA DENGAN CINTA 2 MELALUI MEDIA ONLINE*
- Saputra, P. Y., Subhi, D. H., Zain, F., Winatama, A., Informatika, T., Informasi, T., Malang, P. N., Sentimen, A., & Classifier, N. B. (2012). *Implementasi Sentimen Analisis Komentar Kanal Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naive Bayes*. 209–213.

- Taheri, S., & Mammadov, M. (2013). Learning the naive bayes classifier with optimization models. *International Journal of Applied Mathematics and Computer Science*, 23(4), 787–795.
- Van De Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications*, 42(11), 4999–5010.
- Webb, G. I. (2016). Encyclopedia of Machine Learning and Data Mining. *Encyclopedia of Machine Learning and Data Mining, January 2016*.
- Zhou, X., Zhang, X., & Wang, B. (2016). Online support vector machine: A survey. *Advances in Intelligent Systems and Computing*, 382(8), 269–278.