

# **KOMPARASI KINERJA ALGORITMA MACHINE LEARNING UNTUK ANALISIS SENTIMEN OPINI FILM INDONESIA DI TWITTER**

## **TUGAS AKHIR**

Disusun sebagai salah satu syarat untuk kelulusan  
Program Strata 1, Program Studi Teknik Informatika,  
Universitas Pasundan Bandung

Oleh :

Ugi Prasetyo  
NRP.18.3040162



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS PASUNDAN BANDUNG  
MARET 2023**

**LEMBAR PENGESAHAN  
LAPORAN TUGAS AKHIR**

Telah diujikan dan dipertahankan dalam Sidang Sarjana Program Studi Teknik Informatika Universitas Pasundan Bandung, pada hari dan tanggal sidang sesuai berita acara sidang, tugas akhir dari :

Nama : Ugi Prasetyo

Nrp : 18.304.0162

Dengan judul :

**“Komparasi Kinerja Algoritma Machine Learning Untuk Analisis  
Sentimen Opini Film Indonesia Di Twitter”**

Bandung, 18 Maret 2023

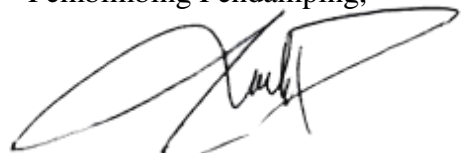
Menyetujui,

Pembimbing Utama,



(Handoko Supeno, S.T, M.T)

Pembimbing Pendamping,



(Asep Somantri, S.T, M.T)

## ABSTRAK

Film adalah media komunikasi audiovisual (suara dan gambar) yang mengandung pesan tersirat yang disampaikan pembuat film kepada sekelompok orang yang menyaksikannya. Analisis sentimen adalah proses perbandingan untuk mengungkap dan mengategorikan pendapat yang diungkapkan dalam teks, untuk menentukan apakah tanggapan terhadap topik tertentu positif, negatif, atau netral. Opini yang disampaikan oleh seseorang bisa mengakibatkan penilaian yang berbeda. Semakin pesat perkembangan teknologi informasi, semakin banyak yang menulis opini mereka mengenai sebuah produk. Algoritma machine learning seperti Naïve Bayes, Support Vector Machine, dan Logistik Regression dapat digunakan peneliti untuk melakukan komparasi algoritma machine learning untuk analisis sentimen opini film Indonesia.

Penelitian ini bertujuan untuk mengetahui cara kerja machine learning dalam menganalisis sentimen berupa opini terhadap suatu film, menganalisis sentimen pada opini film Indonesia menggunakan metode Naïve Bayes, Support Vector Machine, dan Logistik Regression.

Dengan mengkomparasi 3 metode di penelitian ini didapat kesimpulan bahwa algoritma Support Vector Machines mendapatkan tingkat akurasi yang paling tinggi. Nilai akurasi metode Logistik Regression, Naïve Bayes dan Support Vector yaitu Algoritma Support Vector Machines (SVM) sebesar 0,8725, Algoritma Naïve Bayes sebesar 0,655, Algoritma Logistik Regression sebesar 0,8425.

**Kata Kunci** : Algoritma Machine Learning, Analisis Sentimen, Opini, Komparasi, Naïve Bayes, Support Vector Machine, Logistik Regression, Twitter.



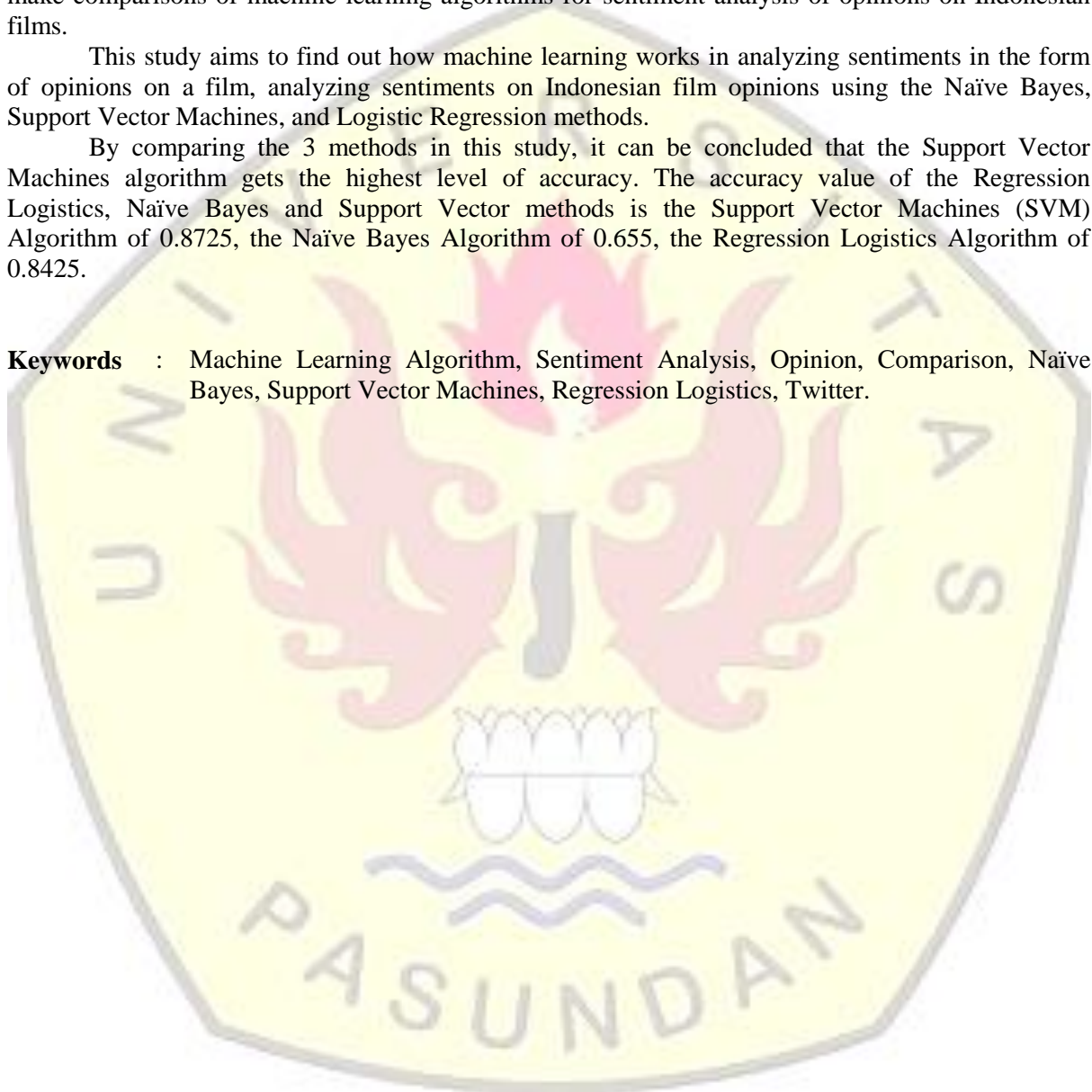
## ABSTRACT

Film is an audiovisual communication medium (sound and image) which contains an implied message that the filmmaker conveys to a group of people who witness it. Sentiment analysis is a benchmarking process for uncovering and categorizing opinions textualized in texts, to determine whether the responses to a given topic are positive, negative, or neutral. Opinions conveyed by someone can result in different judgments. The more rapid the development of information technology, the more people write their opinions about a product. Machine learning algorithms such as Naïve Bayes, Support Vector Machines, and Regression Logistics can be used by researchers to make comparisons of machine learning algorithms for sentiment analysis of opinions on Indonesian films.

This study aims to find out how machine learning works in analyzing sentiments in the form of opinions on a film, analyzing sentiments on Indonesian film opinions using the Naïve Bayes, Support Vector Machines, and Logistic Regression methods.

By comparing the 3 methods in this study, it can be concluded that the Support Vector Machines algorithm gets the highest level of accuracy. The accuracy value of the Regression Logistics, Naïve Bayes and Support Vector methods is the Support Vector Machines (SVM) Algorithm of 0.8725, the Naïve Bayes Algorithm of 0.655, the Regression Logistics Algorithm of 0.8425.

**Keywords** : Machine Learning Algorithm, Sentiment Analysis, Opinion, Comparison, Naïve Bayes, Support Vector Machines, Regression Logistics, Twitter.



## DAFTAR ISI

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR .....	i
ABSTRAK.....	ii
ABSTRACT.....	iii
KATA PENGANTAR .....	iv
DAFTAR ISI.....	v
DAFTAR TABEL.....	vii
DAFTAR GAMBAR.....	viii
DAFTAR LAMPIRAN.....	x
DAFTAR ISTILAH .....	xi
DAFTAR SIMBOL .....	xii
BAB 1 PENDAHULUAN .....	1-1
1.1 Latar Belakang .....	1-1
1.2 Identifikasi Masalah.....	1-2
1.3 Tujuan Tugas Akhir .....	1-2
1.4 Lingkup Tugas Akhir.....	1-2
1.5 Metodologi Tugas Akhir .....	1-3
1.6 Sistematika Penulisan .....	1-5
BAB 2 TINJAUAN PUSTAKA .....	2-1
2.1 Landasan Teori.....	2-1
2.1.1 Analisis Sentimen .....	2-1
2.1.2 Logistik Regression.....	2-1
2.1.3 Support Vector Machine (SVM).....	2-2
2.1.4 Naïve Bayes .....	2-4
2.1.5 Text Processing.....	2-5
2.1.6 Confusion Matrix .....	2-6
2.1.7 TF-IDF .....	2-7
2.1.8 Validasi Snscraper .....	2-8
2.2 Penelitian Terdahulu .....	2-8
BAB 3 SKEMA PENELITIAN.....	3-1
3.1 Alur Penyelesaian Tugas Akhir .....	3-1
3.2 Peralatan Yang Digunakan.....	3-3
3.3 Analisis Sebab Akibat.....	3-4
3.3.1 Solusi Masalah .....	3-4
3.4 Kerangka Berfikir Teoritis .....	3-5
3.4.1 Gambaran Produk TA .....	3-5
3.4.2 Skema Analisis.....	3-6



3.5. Profile Penelitian.....	3-8
3.5.1. Objek Penelitian.....	3-8
<b>BAB 4 PENGUMPULAN DAN PENGOLAHAN DATA .....</b>	<b>4-1</b>
4.1. Rancangan Model Pengumpulan dan Pengolahan Data.....	4-1
4.2. Pengumpulan Data .....	4-2
4.3. Preprocessing .....	4-4
4.3.1. Case Folding .....	4-6
4.3.2. Tokenizing .....	4-8
4.3.3. Filtering.....	4-9
4.3.4. Stemming .....	4-9
4.4. Visualisasi .....	4-10
<b>BAB 5 HASIL DAN PEMBAHASAN .....</b>	<b>5-1</b>
5.1. Data Sentimen Film .....	5-1
5.2. Algoritma Support Vector Machine (SVM) .....	5-2
5.3. Algoritma Naïve Bayes.....	5-5
5.4. Algoritma Logistik Regression .....	5-9
5.5. Komparasi Algoritma Machine Learning Perbandingan 3 Algoritma.....	5-12
<b>BAB 6 KESIMPULAN.....</b>	<b>6-1</b>
6.1. Kesimpulan .....	6-1
6.2. Saran .....	6-1
<b>DAFTAR PUSTAKA .....</b>	<b>.....</b>

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pesatnya perkembangan dunia teknologi informasi dan komunikasi tidak terlepas dari penyedia layanan web yang menyediakan berbagai informasi. Informasi tersebut menghasilkan data tambahan, terutama dalam bentuk data tekstual, yang dapat digunakan sebagai sumber potensial untuk penemuan lebih lanjut. Contohnya adalah data teks yang diambil dari Twitter.

Film adalah salah satu media komunikasi yang menampilkan serangkaian gambar bergerak dengan suatu jalan cerita yang dimainkan oleh para pemeran yang diproduksi untuk menyampaikan suatu pesan kepada penonton.

Film adalah media komunikasi audiovisual (suara dan gambar) yang mengandung pesan tersirat yang disampaikan pembuat film kepada sekelompok orang yang menyaksikannya (Effendy, 1986). Pesan – pesan yang terkandung dalam film mempengaruhi cara berpikir penonton tentang tujuan film tersebut. Menonton tampaknya mampu memikat penonton untuk mengikuti alur dan mempengaruhi persepsi penonton untuk menyampaikan pendapatnya tentang alur, latar, penokohan, dan ending cerita [KUR19].

Opini merupakan suatu pendapat, tanggapan, pandangan, atau hasil pikiran seseorang dalam menjelaskan atau menyikapi suatu hal tetapi sifatnya tidak objektif dan belum pasti kebenarannya. Opini yang disampaikan oleh seseorang bisa mengakibatkan penilaian yang berbeda. Semakin pesat perkembangan teknologi informasi, semakin banyak yang menulis opini mereka mengenai sebuah produk atau layanan pada media sosial.

Analisis sentimen adalah proses perbandingan untuk mengungkap dan mengategorikan pendapat yang diungkapkan dalam teks, untuk menentukan apakah tanggapan terhadap topik tertentu positif, negatif, atau netral. Analisis sentimen dapat membedakan antara teks yang diperoleh dari sebuah data yang merupakan topik positif maupun topik negatif sehingga dapat dibedakan dan dilihat hasilnya hanya dengan memasukan kata kunci maupun kalimat yang ingin diketahui apakah merupakan topik positif maupun negatif, sebagai contoh ulasan buruk masuk kedalam kategori topik negatif dan kepuasan penonton atau pengguna yang merupakan topik positif. Analisis sentimen juga disebut pertambangan adalah bidang studi yang menganalisis pendapat, sentimen, penilaian, dan penilaian emosional pada entitas atribut yang dibahas dalam makalah. Analisis sentimen adalah studi komputasi tentang pendapat, sikap, dan emosi orang terhadap suatu entitas. Entitas dapat mewakili individu, acara, atau topik [MED14].

Satria Wira Yudha dan Mochamad Wahyudi melakukan penelitian tentang komparasi algoritma yang diujicobakan di beberapa dataset baik yang berbahasa Inggris dan Jerman, SVM mendapatkan hasil yang terbaik dengan akurasi 91,92% dan AUC 0,981 pada dataset berbahasa Inggris, sedangkan pada dataset berbahasa Jerman menghasilkan akurasi 90,96% dan AUC 0,978.

Iwan Sinanto dan Ahlijati Nuraminah melakukan penelitian tentang Komparasi algoritma feature selection pada analisis sentiment review film dengan Hasil perbandingan algoritma, SVM mendapatkan hasil terbaik dengan akurasi sebesar 81,10% dan AUC 0,904. Hasil perbandingan pemilihan fitur, information gain mendapatkan hasil terbaik dengan akurasi rata-rata 84,57% dan AUC rata-rata 0,899. Hasil integrasi algoritma klasifikasi terbaik dan algoritma pemilihan fitur terbaik menghasilkan akurasi sebesar 81,50% dan AUC sebesar 0,929.

Pada penelitian ini Analisis Sentimen dilakukan untuk melihat dan mengambil informasi sebuah opini seseorang di twitter, untuk mengetahui apakah opini itu masuk kategori positif, atau negatif dan untuk menguji akurasi analisis sentimen dalam penelitian ini menggunakan metode Logistik Regression, Naïve Bayes, dan Support Vector Machine. Penelitian ini bertujuan untuk menentukan algoritma terbaik dalam analisis sentimen pada opini film indonesia menggunakan Logistic Regression, Naïve Bayes, dan Support Vector Machine.

## 1.2 Identifikasi Masalah

Identifikasi Masalah dari penelitian Tugas Akhir ini sebagai berikut :

- a. Bagaimana cara kerja *Machine Learning* dalam menganalisis sentimen berupa Opini pada suatu Film?
- b. Bagaimana dapat menganalisis sentimen menggunakan metode *Logistik Regression, Naïve Bayes, Support Vector Machine* pada Opini Film?
- c. Berapa besar nilai akurasi metode *Logistik Regression, Naïve Bayes* dan *Support Vector Machine*?
- d. Bagaimana dapat mengetahui algoritma terbaik dalam menganalisis sentimen?

## 1.3 Tujuan Tugas Akhir

Adapun tujuan Tugas Akhir ini sebagai berikut :

- a. Mengetahui cara kerja *machine learning* dalam menganalisis sentimen berupa opini terhadap suatu film.
- b. Mengetahui sentimen tentang Opini dari suatu film.
- c. Menganalisis sentimen dengan menggunakan metode *Logistik Regression, Naïve Bayes*, dan *Support Vector Machine*.
- d. Menganalisis komentar positif dan negatif terhadap Opini pada film.
- e. Mengkomparasi beberapa algoritma untuk menganalisis sentimen dari sebuah film.

## 1.4 Lingkup Tugas Akhir

Lingkup Tugas Akhir dalam penulisan ini dimaksudkan agar pembahasan dan penulisan laporan dapat dilakukan secara terarah dan mencapai sasaran antara lain :

- a. Analisis yang dilakukan oleh peneliti lebih berfokus pada sentimen atau opini pada suatu film.
- b. Data yang digunakan menggunakan opini tentang film Indonesia.

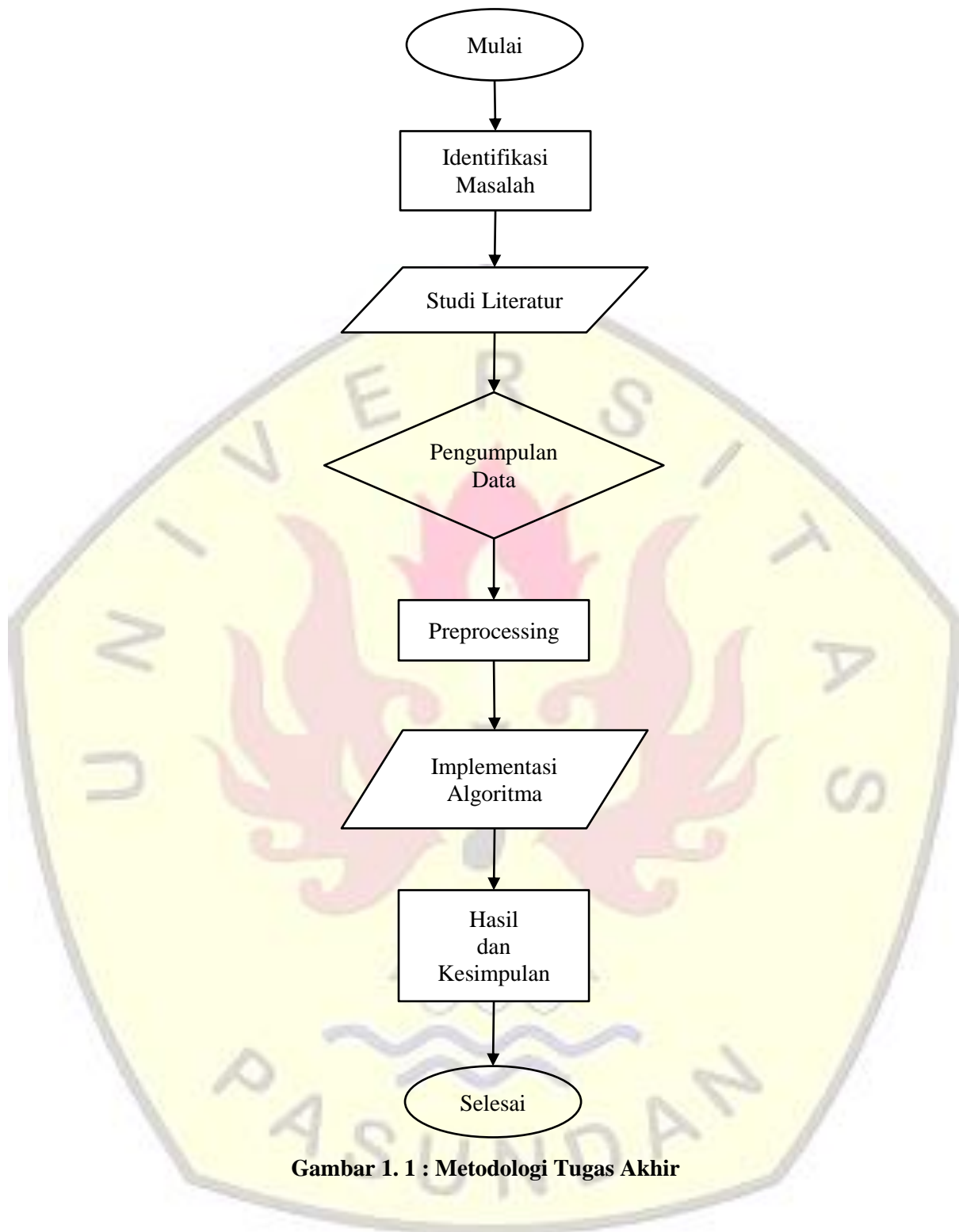


- c. Mengukur tingkat akurasi dari analisis berupa sentimen yang bersifat positif dan negatif pada opini sebuah film.
- d. Tugas Akhir ini hanya menggunakan metode *Logistik Regression*, *Naïve Bayes*, dan *Support Vector Machine* untuk menganalisis sentimen.

### 1.5 Metodologi Tugas Akhir

Bagian ini akan menjelaskan langkah-langkah metodologi penyelesaian yang dilakukan penulis dalam menyelesaikan tugas akhir. Langkah-langkah tersebut dijelaskan dalam bentuk diagram yang dapat dilihat pada gambar 1.1.





Penjelasan mengenai langkah-langkah penyelesaian Tugas Akhir sebagai berikut:

a. Identifikasi Masalah

Identifikasi Masalah merupakan langkah awal yang dilakukan dalam penelitian ini. Pada tahap mengidentifikasi masalah dimaksudkan agar dapat memahami masalah yang akan diteliti, sehingga dalam tahap analisis dan perancangan tidak keluar dari permasalahan yang diteliti.

b. Studi Literatur

Pada tahap studi literatur ini merupakan proses mengetahui dan mempelajari penelitian sebelumnya yang memiliki kesesuaian dengan studi kasus penelitian. Studi literatur terdiri dari landasan teoritis yang memuat teori yang dibutuhkan dalam penelitian dan penelitian yang memiliki kesesuaian dengan studi kasus penelitian. Pada tahap ini, studi literatur digunakan sebagai referensi atau acuan dalam penelitian.

c. Pengumpulan Data

Tahap ini meliputi pengumpulan data, pemilihan aspek, dan perancangan dataset. Perancangan dataset meliputi tahap pelabelan manual terhadap data opini menggunakan aplikasi excel. Dataset dilabelkan kedalam dua kelas sentimen yaitu positif, dan negatif. Positif untuk komentar dengan sentimen positif terhadap aspek, dan negatif untuk ulasan dengan sentimen negatif terhadap aspek.

d. Preprocessing

Setelah dataset disiapkan, maka proses selanjutnya adalah preprocessing. Preprocessing yang dilakukan adalah menghapus mention dan *hashtag* pada teks.

e. Implementasi Algoritma

Pada tahap ini metode yang diterapkan untuk memberikan solusi akan dijelaskan secara bertahap. Tahapan dari implementasi algoritma *Logistik Regression*, *Naïve Bayes*, dan *Support Vector Machine* akan dijelaskan untuk memberikan solusi pada sistem, sehingga memberikan suatu informasi yang bermanfaat.

Secara rinci dalam mengimplemntasikan sistem dibutuhkan perangkat keras, perangkat lunak, dan data perlu dipenuhi untuk mengimplementasikan sistem.

f. Hasil dan Kesimpulan

Pada tahap ini akan dilakukan penarikan kesimpulan dari studi literatur dan hasil dari penelitian. Untuk mendapatkan suatu kesimpulan maka perlu diselesaikan seluruh tahapan perancangan dan analisis metode sehingga dapat menjawab rumusan masalah yang telah diuraikan. Penelitian yang tidak luput dari kekurangan akan membutuhkan saran sehingga segala kesalahan dan kekurangan dalam melakukan penelitian maupun hasil penelitian yang diperoleh dapat diperbaiki dan dipertimbangkan untuk dilakukan penelitian kembali dengan tujuan mengembangkan penelitian.

## 1.6 Sistematika Penulisan

Sistematika penulisan dari Tugas Akhir ini sebagai berikut:

a. BAB 1 Pendahuluan

Pada bab ini menguraikan latar belakang dari penelitian ini, serta memuat identifikasi masalah, tujuan penelitian, lingkup penelitian, metodologi penelitian, dan sistematika dalam penelitian.

b. BAB 2 Tinjauan Pustaka dan Landasan Teori

Pada bab ini memuat beberapa referensi penelitian yang memiliki kesamaan dalam konsep pengerjaan. Pada bab ini penulis membandingkan penelitian tersebut dengan penelitian yang sedang dilakukan berkaitan dengan sentimen analisis dan penggunaan metode *naïve bayes*, *support vector machine*, dan *logistik regression*.

c. BAB 3 Metode Penelitian

Pada bab ini menjelaskan tentang metode penelitian, yaitu langkah-langkah yang dilakukan dalam penelitian.

d. BAB 4 Pengumpulan dan Pengolahan Data

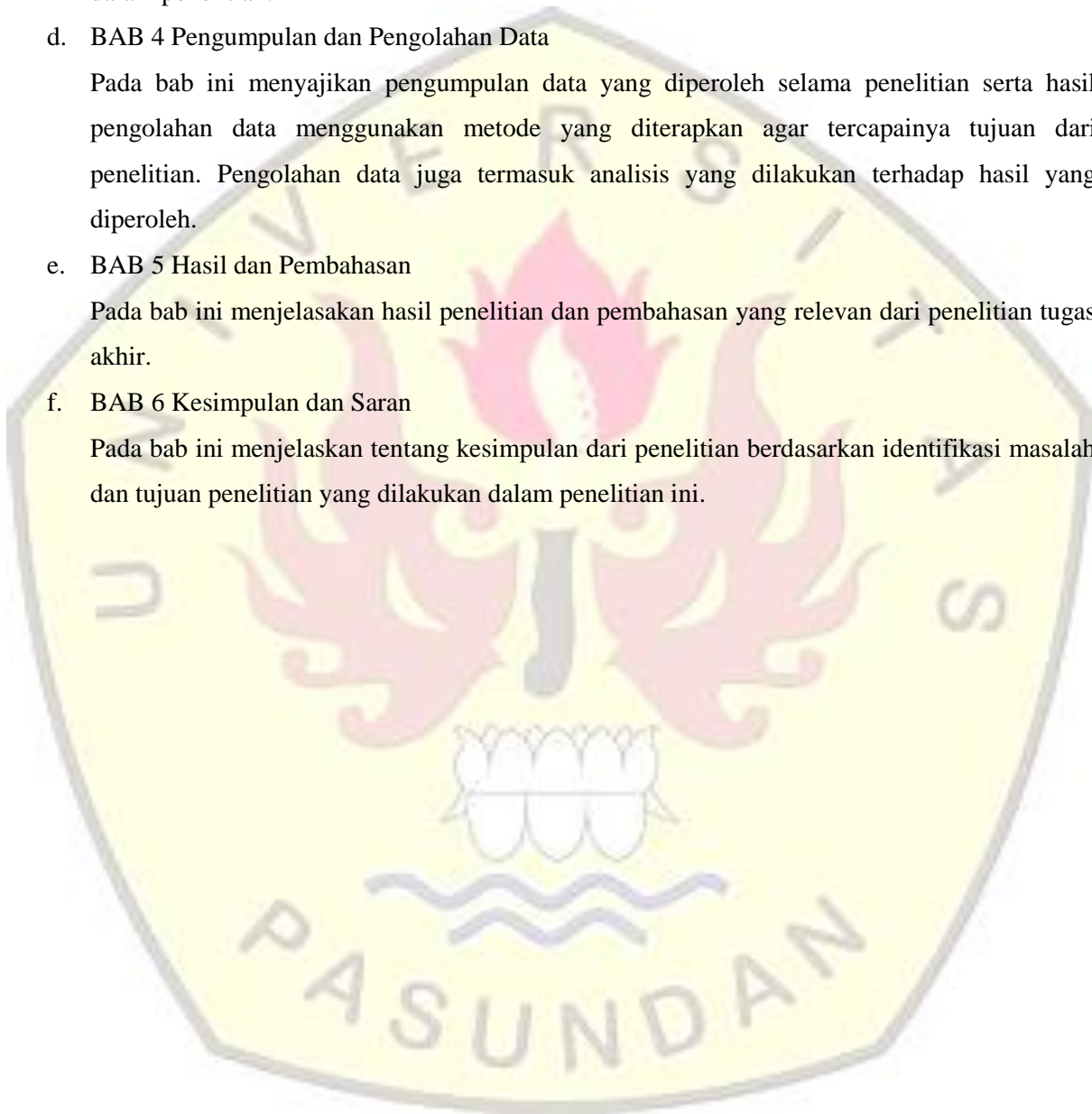
Pada bab ini menyajikan pengumpulan data yang diperoleh selama penelitian serta hasil pengolahan data menggunakan metode yang diterapkan agar tercapainya tujuan dari penelitian. Pengolahan data juga termasuk analisis yang dilakukan terhadap hasil yang diperoleh.

e. BAB 5 Hasil dan Pembahasan

Pada bab ini menjelaskan hasil penelitian dan pembahasan yang relevan dari penelitian tugas akhir.

f. BAB 6 Kesimpulan dan Saran

Pada bab ini menjelaskan tentang kesimpulan dari penelitian berdasarkan identifikasi masalah dan tujuan penelitian yang dilakukan dalam penelitian ini.





## DAFTAR PUSTAKA

- [KUR19] KURNIAWAN, Arifin; INDRIATI, Sigit Adinugroho. Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Fiturs. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2019, 2548: 964X.
- [PRAM19] PRAMESWARI, Kartika; SETIAWAN, Erwin Budi. Analisis Kepribadian Melalui Twitter Menggunakan Metode Logistik Regression dengan Pembobotan TF-IDF dan AHP. *eProceedings of Engineering*, 2019, 6.2.
- [KEL22] KELVIN, Kelvin, et al. Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistik Regression Dan Support Vector Machine (SVM). *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2022, 5.2: 47-52.
- [SUM17] SUMARNO, Heny. Komparasi algoritma klasifikasi machine learning pada analisis sentimen film berbahasa Indonesia. *BINA INSANI ICT JOURNAL*, 2017, 4.2: 189-196.
- [SOM22] SOMANTRI, Asep. Perancangan Model Penilaian Kualitas Informasi di Fanpage Facebook. *PasInformatik*, 2022, 1.01: 7-7.
- [MAH22] MAHDY, Khairudin. *ANALISIS SENTIMEN FILM DI TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)*. 2022. PhD Thesis. Fakultas Teknik Unpas.
- [KURN19] KURNIAWAN, Sigit, et al. Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2019, 3.2: 176-183.
- [MAM18] MAULANA, Muhammad Akbar; SETYANTO, Arief; KURNIAWAN, Mei P. Analisis Sentimen Media Sosial Universitas Amikom Yogyakarta Sebagai Sarana Penyebaran Informasi Menggunakan Algoritma Klasifikasi SVM. *Semnasteknomedia Online*, 2018, 6.1: 1-2-7.
- [HIK20] HIKMAWAN, Sisferi; PARDAMEAN, Amsal; KHASANAH, Siti Nur. Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning. *Jurnal Kajian Ilmiah*, 2020, 20.2.
- [SAN22] SANTOSO, Aloysius Kurniawan. ANALISIS SENTIMEN TWITTER BAHASA INDONESIA MENGGUNAKAN PENDEKATAN MACHINE LEARNING. *Jurnal Informatika Kaputama (JIK)*, 2022, 6.2: 129-136.
- [GIF22] GIFARI, Okta Ihza, et al. Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2022, 2.1: 36-40.
- [MED14] MEDHAT, Walaa; HASSAN, Ahmed; KORASHY, Hoda. Sentiment analysis

- algorithms and applications: A survey. *Ain Shams engineering journal*, 2014, 5.4: 1093-1113.
- [SUS21] SUSILO, Ahmed Joko. Analisis Sentimen Ulasan Pengguna Aplikasi Kesehatan (Telemedicine) Sebagai Pendukung Keputusan Dalam Meningkatkan Kualitas Aplikasi Kesehatan. 2021.
- [SAR20] SARASATI, Fajar, et al. Prediksi Klasifikasi Pembangunan Merek Kosmetik Dengan Metode Enbag K-Logres Berdasarkan Keterlibatan Pengguna Facebook. *Indonesian Journal on Software Engineering (IJSE)*, 2020, 6.1: 116-125.
- [FIK20] FIKRI, Mujaddid Izzul; SABRILA, Trifebi Shina; AZHAR, Yufis. Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*, 2020, 10.02: 71-76.
- [SID20] SIDDIK, Muhammad, et al. Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes. *INTECOMS: Journal of Information Technology and Computer Science*, 2020, 3.2: 162-166.
- [HAN21] HANDAYANI, Fitri. Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 2021, 7.3: 329-334.
- [KRI21] KRISTIAWAN, Kristiawan; WIDJAJA, Andreas. Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel. *Jurnal Teknik Informatika dan Sistem Informasi*, 2021, 7.1.
- [LUQ18] LUQYANA, Wanda Athira. *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine*. 2018. PhD Thesis. Universitas Brawijaya.
- [SAN21] SANUSI, HISYAM; STIAWAN, Deris; UBAYA, Huda. *KLASIFIKASI SENTIMEN TERHADAP DATA TEXT JEJARING SOSIAL DENGAN TOPIK PEMBELAJARAN DARING MENGGUNAKAN LOGISTIK REGRESSION*. 2021. PhD Thesis. Sriwijaya University.
- [DAR21] DARUSSALAM, ABDUL GAFUR. *ANALISIS SENTIMEN MENGGUNAKAN TEXT MINING DENGAN METODE NAÏVE BAYES DAN REGRESI LOGISTIK*. 2021. PhD Thesis. Universitas Hasanuddin.
- [LIU20] Liu Bing, "Sentiment Analysis : Mining Opinions, Sentiments, and Emotions" , 2, Cambridge University Press, 2020
- [AIP20] AI Publishing, "Regression Models With Python For Beginners : Theory and Applications of Linear Models and Logistik Model with python from Scratch" , AI Publishing, 2020