

**PERBANDINGAN ALGORITMA *SUPPORT VECTOR*
MACHINE DAN *NAÏVE BAYES* PADA ANALISIS
SENTIMEN ULASAN CANDI BOROBUDUR**

SKRIPSI

Oleh:

Muhammad Rayhan

202110715241



PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS BHAYANGKARA JAKARTA RAYA

2025

LEMBAR PERSETUJUAN PEMBIMBING

Judul Tugas akhir : Perbandingan Algoritma *Support Vector Machine*
Dan *Naïve Bayes* Pada Analisis Sentimen Ulasan
Candi Borobudur

Nama Mahasiswa : Muhammad Rayhan

Nomor Pokok Mahasiswa : 202110715241

Program Studi/Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 24 Juli 2025

Jakarta, 10 Juli 2025

MENYETUJUI,

Pembimbing I



Adi Muhajirin, S.Kom, M.Kom, M.M

NIDN. 0318038501

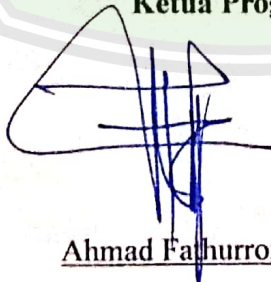
Pembimbing II



Aida Fitriyani, S.Kom., M.M.S.I.

NIDN. 0302078508

Ketua Program Studi



Ahmad Fathurrozi, S.E., M.M.S.I.

NIP. 2012486

LEMBAR PENGESAHAN

Judul Tugas akhir : PERBANDINGAN ALGORITMA SUPPORT
VECTOR MACHINE DAN NAÏVE BAYES
PADA ANALISIS SENTIMEN ULASAN
CANDI BOROBUDUR

Nama Mahasiswa : Muhammad Rayhan

Nomor Pokok Mahasiswa : 202110715241

Program Studi/Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Akhir : 24 Juli 2025

Jakarta, 31 Juli 2025

MENGESAHKAN,

Ketua Tim Penguji : Muhammad Khaerudin, S.Kom, M.Kom

NIDN : 0413066604

Penguji I : Siti Setiawati, S.Pd, M.Pd

NIDN : 0313107904

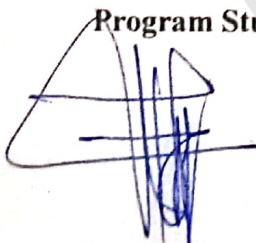
Penguji II : Adi Muhajirin, S.Kom, M.Kom, M.M

NIDN : 0318038501

MENGETAHUI,

Ketua

Program Studi Informatika



Ahmad Fathurrozi, S.E., M.M.S.I.
NIP: 2012486

Dekan

Fakultas Ilmu Komputer



Dr. Robertus Suraji, S.S., M.A
NIP: 1904406



LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Muhammad Rayhan
NPM : 202110715241
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Perbandingan Algoritma Support Vector Machine Dan
Naïve Bayes Pada Analisis Sentimen Ulasan Candi
Borobudur

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan hasil karya saya sendiri dan benar keasliannya. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Bekasi, 2 Agustus 2025

Penulis



Muhammad Rayhan

ABSTRAK

Muhammad Rayhan. 202110715241. Perbandingan Algoritma *Support Vector Machine* Dan *Naïve Bayes* Pada Analisis Sentimen Ulasan Candi Borobudur. Bekasi: Fakultas Ilmu Komputer . Universitas Bhayangkara Jakarta Raya. 2025

Pertumbuhan ulasan wisatawan secara daring telah membuka peluang baru dalam pemanfaatan data teks untuk mendukung pengambilan keputusan di sektor pariwisata. Volume ulasan pengunjung yang besar di platform daring menyimpan informasi berharga tentang pengalaman wisata, namun sulit dianalisis secara manual karena keterbatasan waktu dan sumber daya. Selain itu, data ulasan cenderung *imbalanced* akibat tidak semua pengunjung meninggalkan ulasan, sehingga dibutuhkan pendekatan analisis yang mampu mengakomodasi kondisi ketidakseimbangan data. Di sisi lain, belum terdapat kejelasan mengenai algoritma *machine learning* yang paling efektif dalam menganalisis sentimen ulasan wisatawan, sehingga perlu dilakukan perbandingan performa antara algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes*. Penelitian ini bertujuan untuk mengklasifikasikan sentimen wisatawan terhadap Candi Borobudur secara otomatis menggunakan pendekatan *machine learning*. Data sebanyak 4.341 ulasan dikumpulkan dari dua platform, yaitu Tiket.com dan Tripadvisor. Proses penelitian mencakup tahap *preprocessing*, ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency (TF-IDF)*, dan klasifikasi sentimen menggunakan dua algoritma *supervised learning*: *Naïve Bayes* dan *Support Vector Machine (SVM)*. Hasil evaluasi menunjukkan bahwa *SVM* memiliki performa lebih stabil dalam menangani data yang *imbalanced*, dengan akurasi sebesar 88%, *precision* positif 91%, *recall* positif 95%, dan *F1-score* positif 93%. Sementara itu, *Naïve Bayes* memiliki akurasi sebesar 83% dan *recall* positif 100%, tetapi lemah dalam mengklasifikasikan sentimen negatif. Temuan ini menunjukkan bahwa *SVM* lebih representatif untuk klasifikasi biner dalam analisis sentimen. Dengan demikian, penerapan algoritma *machine learning* terbukti efektif dalam menganalisis sentimen ulasan wisatawan secara otomatis, serta dapat menjadi dasar objektif dalam pengelolaan destinasi wisata berbasis data.

Kata Kunci : machine learning; analisis sentimen; *Support Vector Machine*; *Naïve Bayes*; *TF-IDF*; ulasan wisatawan; klasifikasi teks.

ABSTRACT

Muhammad Rayhan. 202110715241. *A Comparison Of Support Vector Machine And Naïve Bayes Algorithms In Sentiment Analysis Of Borobudur Temple Reviews.* Bekasi: Faculty of Computer Science . Bhayangkara University Jakarta Raya. 2025

The growth of online tourist reviews has opened up new opportunities for leveraging textual data to support decision-making in the tourism sector. The large volume of visitor reviews on online platforms contains valuable information about tourist experiences, but it is difficult to analyze manually due to time and resource constraints. Additionally, review data tends to be imbalanced, as not all visitors leave reviews, requiring an analytical approach that can accommodate data imbalance. On the other hand, there is no clear consensus on which machine learning algorithm is most effective in analyzing tourist sentiment, making it necessary to compare the performance of Support Vector Machine (SVM) and Naïve Bayes algorithms. This study aims to automatically classify tourist sentiment toward Borobudur Temple using a machine learning approach. A total of 4,341 reviews were collected from two platforms: Tiket.com and Tripadvisor. The research process included preprocessing, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), and sentiment classification using two supervised learning algorithms: Naïve Bayes and Support Vector Machine (SVM). Evaluation results show that SVM performs more consistently in handling imbalanced data, achieving an accuracy of 88%, positive precision of 91%, positive recall of 95%, and a positive F1-score of 93%. Meanwhile, Naïve Bayes achieved an accuracy of 83% and a positive recall of 100% but was weak in classifying negative sentiment. These findings indicate that SVM is more representative for binary classification in sentiment analysis. Thus, the application of machine learning algorithms has proven effective in automatically analyzing tourist reviews and can serve as an objective basis for data-driven tourism destination management.

Keywords : machine learning; sentiment analysis; Support Vector Machine; Naïve Bayes; TF-IDF; tourist reviews; text classification.

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Rayhan
NPM : 202110715241
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif** (*Non-Exclusive Royalty-Free Right*), atas karya ilmiah saya yang berjudul:

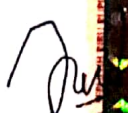
“Perbandingan Algoritma *Support Vector Machine* Dan *Naïve Bayes* Pada Analisis Sentimen Ulasan Candi Borobudur”

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan memublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi.

Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Dibuat di : Bekasi
Pada Tanggal : 2 Agustus 2025
Yang Menyatakan,



Muhammad Rayhan

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas segala limpahan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "**Perbandingan *Support Vector Machine* Dan *Naïve Bayes* pada Analisis Sentimen Ulasan Candi Borobudur**" dengan baik dan tepat waktu. Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan jenjang Strata Satu (S1) pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya. Dalam proses penyusunan skripsi ini, penulis memperoleh banyak ilmu, pengalaman, dan wawasan yang sangat bermanfaat dalam pengembangan diri dan keilmuan.

Penulis menyadari bahwa penyusunan skripsi ini tidak lepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih kepada:

1. Bapak Irjen. Pol. (Purn.) Dr. Drs. H. Bambang Karsono, S.H., M.M., Ph.D., D.Crim., HC, selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Bapak Dr. Robertus Suraji, S.S., M.A., selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I., selaku Ketua Program Studi Informatika.
4. Bapak Adi Muhajirin, S.Kom., M.Kom., M.M., selaku Dosen Pembimbing I dalam penyusunan proposal skripsi Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya.
5. Ibu Aida Fitriyani, S.Kom., S.I., M.M., selaku Dosen Pembimbing II sekaligus Dosen Pembimbing Akademik dalam penyusunan proposal skripsi Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya.
6. Ayah dan Ibu, selaku orang tua yang senantiasa memberikan dukungan moral, spiritual, dan materiil kepada penulis.

7. Seluruh pihak yang telah membantu secara langsung maupun tidak langsung hingga tersusunnya laporan ini.

Penyusunan skripsi ini tentu masih jauh dari sempurna. Oleh karena itu, segala bentuk kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat serta menjadi kontribusi positif dalam pengembangan ilmu pengetahuan, khususnya di bidang Informatika.

Bekasi, 2 Agustus 2025



Muhammad Rayhan



DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN BUKAN PLAGIASI.....	iv
ABSTRAK.....	v
ABSTRACT	vi
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABLE.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Identifikasi Masalah.....	5
1.3 Rumusan Masalah.....	5
1.4 Batasan Masalah	6
1.5 Tujuan Penelitian	6
1.6 Manfaat Penelitian	6
1.7 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	8
2.1 Penelitian Terdahulu	8
2.2 Landasan Teori	12
2.2.1 Data Mining	13
2.2.2 Knowledge Discovery in Database (KDD).....	14

2.2.3	Machine Learning	15
2.2.4	Text Mining	16
2.2.5	Pre-Processing Teks	17
2.2.6	TF-IDF	19
2.2.7	Analisis Sentimen	21
2.2.8	Naïve bayes	21
2.2.9	SVM.....	23
2.2.10	Evaluasi Model (Confusion Matrix,Accuracy,dll.).....	25
2.2.11	Python	28
2.2.12	Google Colab	29
2.2.13	Candi Borobudur.....	30
BAB III METODOLOGI PENELITIAN.....		31
3.1	Objek Penelitian.....	31
3.2	Kerangka Penelitian	32
3.3	Metode Pengumpulan Data.....	33
3.4	Metode Analisis Data.....	34
3.5	Analisis Kebutuhan Sistem.....	38
BAB IV HASIL DAN PEMBAHASAN		39
4.1	Data Selection	39
4.2	Pre-processing.....	40
4.3	Pembagian Data (Data Splitting)	51
4.5	Modelling.....	54
4.5.1	Naïve Bayes	54
4.5.2	SVM.....	60
4.6	Evaluasi.....	66
4.6.1	Hasil Evaluasi Model Naïve Bayes.....	67

4.6.2	Hasil Evaluasi Model SVM	70
4.6.3	Perbandingan Hasil dan Diskusi	74
4.6.4	WordCloud.....	77
BAB 5 PENUTUP		79
5.1	Kesimpulan	79
5.2	Saran	80
DAFTAR PUSTAKA.....		81
LAMPIRAN		91



DAFTAR TABLE

Tabel 2. 1 Penelitian Terdahulu	8
Tabel 2. 2 Contoh Confusion Matrix pada Klasifikasi Sentimen Dua Kelas	25
Tabel 4. 1 Contoh Data Gabungan.....	39
Tabel 4. 2 Hasil Proses Case Folding Dan Cleansing.....	41
Tabel 4. 3 Hasil Data Normalisasi	43
Tabel 4. 4 Hasil Tokenisasi	44
Tabel 4. 5 Hasil Stemming	45
Tabel 4. 6 Hasil Stopword	47
Tabel 4. 7 Contoh Data Setelah Proses Pelabelan	50
Tabel 4. 8 Contoh Representasi TF-IDF	53
Tabel 4. 9 Ulasan Data Latih	55
Tabel 4. 10 Prediksi Data Latih Naïve Bayes	56
Tabel 4. 11 Ulasan Data Uji.....	58
Tabel 4. 12 Hasil Data Uji	59
Tabel 4. 13 Data Untuk Perhitungan SVM Secara Manual	61
Tabel 4. 14 Hasil Ekstraksi TF-IDF untuk Perhitungan SVM	62
Tabel 4. 15 Confusion Matrix Model Naive Bayes	67
Tabel 4. 16 Hasil Metrik Evaluasi Model Naïve Bayes	70
Tabel 4. 17 Confusion Matrix Model SVM.....	71
Tabel 4. 18 Hasil Metrik Evaluasi Model SVM	73
Tabel 4. 19 Perbandingan Metrik Evaluasi Model Naïve Bayes dan SVM	74

DAFTAR GAMBAR

Gambar 1. 1	Data Pengunjung Wisata Candi Borobudur	2
Gambar 1. 2	Ulasan Pengguna Tripadvisor	3
Gambar 1. 3	Ulasan Pengguna Tiket.com	4
Gambar 2. 1	Tahapan KDD	14
Gambar 2. 2	Preprocessing Teks	18
Gambar 2. 3	Support Vector Machine	23
Gambar 2. 4	Confusion Matrix	26
Gambar 2. 5	Python	28
Gambar 2. 6	Google Colaboratory	29
Gambar 3. 1	Peta Candi Borobudur	31
Gambar 3. 2	Kerangka Penelitian	32
Gambar 3. 3	Scrapping Data	33
Gambar 3. 4	Hasil Scrapping Data Ulasan	34
Gambar 3. 5	Hasil Preprocessing Data	35
Gambar 3. 6	Pembobotan TF-IDF	37
Gambar 4. 1	Kode Case Folding Dan Cleansing	40
Gambar 4. 2	Kode Filtering	41
Gambar 4. 3	Kode Normalisasi	42
Gambar 4. 4	Kode Tokenisasi	43
Gambar 4. 5	Kode Stemming	45
Gambar 4. 6	Kode Stopword Removal	46
Gambar 4. 7	Kode Labelling	48
Gambar 4. 8	Distribusi 3 Kelas	49
Gambar 4. 9	Distribusi 2 Kelas	50
Gambar 4. 10	Kode Split Data	51
Gambar 4. 11	Kode Transformasi TF-IDF	52
Gambar 4. 12	Kode Model Naive Bayes	55
Gambar 4. 13	Kode SVM Linear	60
Gambar 4. 14	Visualisasi Hyperplane SVM Linear	60
Gambar 4. 15	Confusion Matrix Model Naïve Bayes	67

Gambar 4. 16 Laporan Klasifikasi Model Naïve Bayes	68
Gambar 4. 17 Confusion Matrix Model SVM.....	71
Gambar 4. 18 Laporan Klasifikasi Model SVM.....	72
Gambar 4. 19 WordCloud Sentimen Positif	77
Gambar 4. 20 WordCloud Sentimen Negatif.....	78



DAFTAR LAMPIRAN

Biodata Mahasiswa

Kartu Bimbingan Skripsi

Rekomendasi Pembimbing

Hasil Plagiarisme

Dataset

