PERBANDINGAN *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN BANJIR KOTA BEKASI PADA MEDIA SOSIAL *X*

SKRIPSI

Oleh:
Nabila Rahmadhinta
202110715240

PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2025

LEMBAR PERSETUJUAN PEMBIMBING

Judul Proposal Tugas Akhir	:	Perbandingan Naïve Bayes dan Support Vector Machine dalam Analisis Sentimen Banjir Kota Bekasi pada Media Sosial X
Nama Mahasiswa	:	Nabila Rahmadhinta
Nomor Pokok Mahasiswa	:	202110715240
Program Studi/Fakultas	:	Informatika/Ilmu Komputer

Jakarta, 05 Agustus 2025
MENYETUJUI,

Pembimbing I	Pembimbing II
Wowon Priatna, S.T., M.TI	Rafika Sari, S.Si.,M.Si.
NIDN: 0429118007	NIDN: 0329098902

Ketua Program Studi

Ahmad Fathurrozi, S.E., M.M.S.I
NIP: 2012786

Program Studi Informatika
Fakultas Ilmu Komputer
Universitas Bhayangkara Jakarta Raya
2025

Dipindai dengan CamScanner

LEMBAR PENGESAHAN

Judul Tugas Akhir	: Perbandingan Naïve Bayes dan Support Vector Machine dalam Analisis Sentimen Banjir Kota Bekasi pada Media Sosial X
Nama Mahasiswa	: Nabila Rahmadhinta
Nomor Pokok Mahasiswa	: 202110715240
Program Studi/Fakultas	: Informatika / Ilmu Komputer
Tanggal Lulus Ujian Tugas Akhir	: 29 Juli 2025

Jakarta, 05 Agustus 2025

MENGESAHKAN,

Ketua Tim Penguji	:	Adi Muhajirin, S.Kom., M.Kom., M.M	.....
		NIDN: 0318038501	
Penguji I	:	Dr. Dra. Tyastuti Sri Lestari, M.M	.....
		NIDN: 0327036701	
Penguji II	:	Wowon Priatna, S.T., M.T.I	.....
		NIDN: 0429118007	

MENGETAHUI,

Ketua	**Dekan**
Program Studi Informatika	**Fakultas Ilmu Komputer**
Ahmad Fathurrozi, S.E., M.M.S.I	Dr. Robertus Suraji, S.S., M.A
NIP: 2012486	NIP: 1904406

Dipindai dengan CamScanner

UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama	:	Nabila Rahmadhinta
NPM	:	202110715240
Program Studi	:	Informatika
Fakultas	:	Ilmu Komputer
Judul Tugas Akhir	:	Perbandingan Naïve Bayes Dan Support Vector Machine Dalam Analisis Sentimen Banjir Kota Bekasi Pada Media Sosial X

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 mana pun.

Bekasi, 05 Agustus 2025
Penulis,



Nabila Rahmadhinta

Dipindai dengan CamScanner

ABSTRAK

NABILA RAHMADHINTA, 202110715240. Perbandingan *Naïve Bayes* dan *Support Vector Machine* Dalam Analisis Sentimen Banjir Kota Bekasi Pada Media Sosial *X*. Fakultas Ilmu Komputer. Bekasi: Universitas Bhyangkara Jakarta Raya. 2025.

Analisis sentimen memainkan peran penting dalam memahami opini publik terkait peristiwa bencana, terutama melalui media sosial seperti *X*. Platform ini sering digunakan oleh masyarakat untuk mengekspresikan pandangan mereka mengenai pengelolaan banjir, keluhan warga, dan apresiasi terhadap pemerintah. Dalam konteks ini, komentar pengguna terkait banjir yang terjadi di Kota Bekasi pada tahun 2025 dianalisis untuk mengidentifikasi tren sentimen yang muncul, sekaligus membandingkan kinerja algoritma *Naïve Bayes* dan *Support Vector Machine* dalam mengklasifikasi sentimen tersebut. Sebanyak 2.141 komentar dikumpulkan melalui teknik *crawling* dan dibagi menjadi 80% data pelatihan dan 20% data pengujian. Proses analisis mencakup tahapan pembersihan teks, tokenisasi, normalisasi, serta pembobotan kata menggunakan metode TF-IDF. Hasil pengujian menunjukkan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 80%, dengan hasil presisi 41%, recall 60%, f1-score 49%. Sedangkan, algoritma SVM menghasilkan akurasi lebih besar 8% dari *Naïve Bayes*, yaitu sebesar 88% dengan hasil presisi 65%, recall 49%, f1-score 56%. Hasil ini menunjukkan bahwa SVM lebih efektif dalam mengklasifikasi opini publik terkait bencana banjir yang kompleks dan emosional.

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, *Support Vector Machine*, Banjir, TF-IDF.

ABSTRACT

NABILA RAHMADHINTA, 202110715240. *Comparison of Naïve Bayes and Support Vector Machine in Sentiment Analysis of Flooding in Bekasi City on Social Media X. Faculty of Computer Science. Bekasi: University of Bhayangkara Jakarta Raya. 2025.*

Sentiment analysis plays a crucial role in understanding public opinion regarding disaster events, especially through social media platforms such as X. This platform is often used by the public to express their views on flood management, residents' complaints, and appreciation toward the government. In this context, user comments related to the 2025 flood in Bekasi City were analyzed to identify emerging sentiment trends, while also comparing the performance of the Naïve Bayes and Support Vector Machine algorithms in classifying these sentiments. A total of 2,141 comments were collected using a crawling technique and divided into 80% training data and 20% testing data. The analysis process included text cleaning, tokenization, normalization, and word weighting using the TF-IDF method. The testing results showed that the Naïve Bayes algorithm achieved an accuracy of 80%, with a precision of 41%, recall of 60%, and F1-score of 49%. Meanwhile, the SVM algorithm achieved 8% higher accuracy than Naïve Bayes, reaching 88%, with a precision of 65%, recall of 49%, and F1-score of 56%. These results indicate that SVM is more effective in classifying public opinions related to complex and emotionally charged flood disasters.

***Keywords**: Sentiment Analysis, Naïve Bayes, Support Vector Machine, Flood, X, TF-IDF.*

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

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

Nama	: Nabila Rahmadhinta
NPM	: 202110715240
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-Esklusif (*Non-Exclusive Royalty-Free Right*)**, atas karya ilmiah saya yang berjudul:

"Perbandingan Algoritma *Naïve Bayes* dan *Support Vector Machine* Dalam Analisis Sentimen Banjir Kota Bekasi Pada Sosial Media *X*"

Beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-ekslusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin 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 sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 05 Agustus 2025

Yang Menyatakan



Nabila Rahmadhinta

Dipindai dengan CamScanner

KATA PENGANTAR

Segala puji syukur bagi Allah SWT, atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Perbandingan *Naïve Bayes* Dan *Support Vector Machine* Dalam Analisis Sentimen Banjir Kota Bekasi Pada Media Sosial *X*" ini dengan baik. Penulis menyadari bahwa tanpa bantuan dari berbagai pihak, penyusunan skripsi ini tidak akan dapat terlaksana dengan baik. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah S.W.T yang telah meridhai untuk menulis penelitian tugas akhir ini hingga selesai.
2. Bapak Irjen Pol (Pum) Prof. Dr. Drs. H. Bambang Karsono, S.H., M.M., Ph.D.,D.Crim (Hc) selaku Rektor Universitas Bhayangkara Jakarta.
3. Bapak Dr. Robertus Suraji, S.S., M.A., selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
4. Bapak Ahmad Fathurrozi, S.E., M.M.S.I., selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
5. Bapak Wowon Priatna, S.T., M.T.I., selaku Dosen Pembimbing I, atas bimbingan, arahan dan pembelajaran yang telah diberikan sehingga dapat menyelesaikan skripsi ini.
6. Ibu Rafika Sari S.Si., M. Si., selaku Dosen Pembimbing II, atas arahan dan pembelajaran yang telah diberikan sehingga dapat menyelesaikan skripsi ini.

7. Kedua orang tua saya yang telah memberikan segala dukungan dan pengorbanan yang tiada henti selama ini, yang selalu mendoakan setiap harinya agar diberikan kemudahan dalam menyelesaikan skripsi ini.
8. Teman-teman seperjuangan saya pada program studi Informatika Universitas Bhayangkara Jakarta Raya, terutama Grup Cemara yang selalu memberikan dukungan dan selalu membantu dalam menyelesaikan skripsi ini.

Penelitian skripsi yang saya susun ini mungkin jauh dari kesempurnaan, untuk itu saya mohon maaf apabila terdapat kekurangan dan kesalahan baik dari isi maupun tulisan. Oleh karena itu, saya mengharapkan kritik dan saran yang membangun untuk memperbaiki kekurangan-kekurangan yang ada. Semoga penelitian ini bermanfaat, baik bagi saya maupun bagi orang yang membacanya.

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 TABEL ..... **xiv**

DAFTAR GAMBAR ..... **xvi**

DAFTAR LAMPIRAN ..... **xviii**

BAB I PENDAHULUAN **1**

1.1 Latar Belakang 1

1.2 Identifikasi Masalah 5

1.3 Rumusan Masalah 5

1.4 Batasan Masalah 5

1.5 Tujuan Penelitian 6

1.6 Manfaat Penelitian 6

1.7 Sistematika Penulisan 7

BAB II TINJAUAN PUSTAKA **9**

2.1 *State of The Art* 9

2.2 *Data Mining* 13

2.3 *Machine Learning* 15

2.4 *Text Mining* 15

2.5 *Knowledge Discovery in Databases* (KDD) 17

2.6 *Text Preprocessing* 18

2.7 *Term Frequency – Inverse Document Frequency* (TF-IDF) 20

2.8 *Naïve Bayes* 22

2.9 *Support Vector Machine* 24

2.10 Pengujian Sistem 25

2.11 *Tools* Pemrograman 27

2.12 Analisis Sentimen 28

2.13 *Lexicon Based* 28

2.14 *X* 29

2.15 BNPB 30

BAB III METODOLOGI PENELITIAN 32

3.1 Objek Penelitian 32

3.2 *Data Selection* 32

3.3 Metode Pengumpulan Data 34

3.4 Kerangka Penelitian 34

3.5 Data Penelitian 36

3.5.1 *Preprocessing* 37

3.6 Analisis Data 42

3.6.1 *Transformation* 42

3.7 *Data Mining* 43

3.8 *Evaluation* 45

3.9 Analisis Kebutuhan *Hardware* dan *Software* 45

3.9.1 Spesifikasi *Hardware* 45

3.9.2 Spesifikasi *Software* 45

BAB IV 46

HASIL DAN PEMBAHASAN 46

4.1 Hasil 46

4.2 Pembahasan 47

4.3 *Data Selection* 47

4.3.1 Pengambilan Data 47

4.3.2 *Labeling* Data 49

4.4 *Preprocessing* 51

4.4.1 *Case Folding* 51

4.4.2 *Cleaning* 53

4.43 Tokenizing 55

4.4.4 *Normalizing* 56

4.4.5 *Filtering / Stopword* 57

4.4.6 *Stemming* 59

4.5 *Transformation* 60

4.5.1 Mengubah Dataset Menjadi *Numpy Array* 60

4.5.2 *Splitting Data* 62

4.5.3 Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) 62

4.6 *Data Mining* 70

4.6.1 Implementasi Algoritma *Naïve Bayes* 71

4.6.2 Implementasi Algoritma *Support Vector Machine* 76

4.7 *Evaluation* 79

4.7.1 *Evaluation Naïve Bayes* 79

4.7.2 *Evaluation Support Vector Machine* 82

4.8 *Knowladge Presentation* 85

BAB V KESIMPULAN DAN SARAN 90

5.1 Kesimpulan 90

5.3 Saran 91

DAFTAR PUSTAKA 92

LAMPIRAN 102

DAFTAR TABEL

Tabel 2. 1 *State of The Art* 9

Tabel 2. 2 *Confusion Matrix* 26

Tabel 4. 1 Hasil Perbandingan Algoritma NB dan SVM..... 46

Tabel 4. 2 Hasil *Crawling* dan *Labeling* 50

Tabel 4. 3 Hasil Tahapan *Case Folding*..... 52

Tabel 4. 4 Hasil Tahapan *Cleaning*..... 54

Tabel 4. 5 Hasil Tahapan *Tokenizing*..... 55

Tabel 4. 6 Hasil Tahapan *Normalizing* 57

Tabel 4. 7 Hasil Tahapan *Stopwords* 58

Tabel 4. 8 Hasil Tahapan *Stemming*..... 59

Tabel 4. 9 Pembuatan *Word Vector Manual*..... 64

Tabel 4. 10 Perhitungan TF dan DF..... 65

Tabel 4. 11 Perhitungan IDF..... 67

Tabel 4. 12 Perhitungan Bobot..... 67

Tabel 4. 13 Sampel Data *Train* 71

Tabel 4. 14 Sampel Data *Test* 71

Tabel 4. 15 Penentuan Jumlah *Term* Pada Dokumen 71

Tabel 4. 16 Perhitungan *Prior Probability* 72

Tabel 4. 17 Sampel Data *Test* 74

Tabel 4. 18 Hasil Prediksi Data *Test*..... 75

Tabel 4. 19 Sampel Data *Train* 76

Tabel 4. 20 *Vektor Fitur* untuk Tiap Dokumen..... 77

Tabel 4. 21 Penentuan Bobot 77

Tabel 4. 22 *Confusion Matrix Naive Bayes* 79

Tabel 4. 23 *Confusion Matrix* SVM..... 82

DAFTAR GAMBAR

Gambar 1. 1 Grafik Data Bencana Tertinggi di Indonesia..... 2

Gambar 2. 1 Proses *Knowledge Discovery in Database*..... 17

Gambar 2. 2 Ilustrasi Klasifikasi *Support Vector Machine* 24

Gambar 2. 3 Data Pengguna *X* di Indonesia 30

Gambar 3. 1 Kerangka Tahapan *Crawling* Data..... 33

Gambar 3. 2 Kerangka Penelitian 35

Gambar 3. 3 Data Penelitian 36

Gambar 3. 4 Kerangka Tahapan *Preprocessing* 37

Gambar 3. 5 Kerangka Tahapan *Case Folding*..... 38

Gambar 3. 6 Kerangka Tahapan *Cleansing* 39

Gambar 3. 7 Kerangka Tahapan *Tokenizing*..... 39

Gambar 3. 8 Kerangka Tahapan *Normalizing*..... 40

Gambar 3. 9 Kerangka Tahapan *Stopword* 41

Gambar 3. 10 Kerangka Tahapan *Stemming*..... 41

Gambar 3. 11 Kerangka Tahapan KDD..... 42

Gambar 3. 12 Kerangka Tahapan *Naive Bayes*..... 43

Gambar 3. 13 Kerangka Tahapan *Support Vector Machine* 44

Gambar 4. 1 *Script Auth Token* untuk Akses *Twitter*..... 48

Gambar 4. 2 *Script Instal Library* Pandas dan *Node.js*..... 48

Gambar 4. 3 *Script* Untuk *Crawling* Data..... 49

Gambar 4. 4 *Script* untuk *Crawling* Data 49

Gambar 4. 5 *Script* untuk *Labeling*..... 50

Gambar 4. 6 *Script* untuk Tahapan *Case Folding*..... 51

Gambar 4. 7 *Script* untuk Tahapan *Cleaning* 53
Gambar 4. 8 *Script* untuk Tahapan *Tokenizing* 55
Gambar 4. 9 *Script* untuk Tahapan *Normalizing* 56
Gambar 4. 10 *Script* Tahapan untuk *Stopwords* 58
Gambar 4. 11 *Script* Tahapan untuk *Stemming* 59
Gambar 4. 12 *Script* untuk Mengubah Menjadi *Numpy Array* 61
Gambar 4. 13 Hasil Merubah Data Menjadi Kolom Komentar 61
Gambar 4. 14 Hasil Merubah Data Menjadi Kolom Sentimen 61
Gambar 4. 15 *Script* untuk Tahapan *Splitting* Data 62
Gambar 4. 16 Hasil *Splitting* Data 62
Gambar 4. 17 *Script* untuk Tahapan Pembobotan TF IDF 70
Gambar 4. 18 Hasil *Script* Pembobotan Kata 70
Gambar 4. 19 *Script* Tahapan *Modelling Naive Bayes* 76
Gambar 4. 20 *Script* Tahapan *Modelling* SVM 79
Gambar 4. 21 *Script Confusion Matrix Naive Bayes* 81
Gambar 4. 22 Hasil *Confusion Matrix Naive Bayes* 81
Gambar 4. 23 *Script* Tahapan dan Hasil Uji Model *Naïve Bayes* 82
Gambar 4. 24 *Script Confusion Matrix* SVM 83
Gambar 4. 25 Hasil *Confusion Matrix* SVM 84
Gambar 4. 26 *Script* Tahapan dan Hasil Uji Model SVM 84
Gambar 4. 27 *Pie Chart* Hasil *Labeling* Data 85
Gambar 4. 28 Hasil *Netral Wordcloud* 86
Gambar 4. 29 Hasil *Positif Wordcloud* 87
Gambar 4. 30 Hasil *Negatif Wordcloud* 88

DAFTAR LAMPIRAN

Plagiarisme

Biodata Mahasiswa

Kartu Bimbingan Skripsi Pembimbing I

Kartu Bimbingan Skripsi Pembimbing II

Infografis

Hasil Dataset

Dipindai dengan CamScanner