



**ANALISIS DATA RESPONS PUBLIK PENGGUNA FITUR
LIVE STREAMING TIKTOK SEBAGAI MEDIA HIBURAN
DI INDONESIA DENGAN MENGGUNAKAN
ALGORITMA *NAIVE BAYES***

SKRIPSI

**ANISA FADILAH SAPUTRI
NIM. 2110512010**

**PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
2025**



**ANALISIS DATA RESPONS PUBLIK PENGGUNA FITUR
LIVE STREAMING TIKTOK SEBAGAI MEDIA HIBURAN
DI INDONESIA DENGAN MENGGUNAKAN
ALGORITMA *NAIVE BAYES***

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh
Gelar Sarjana Komputer**

**ANISA FADILAH SAPUTRI
NIM. 2110512010**

**PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
2025**

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya sendiri serta sumber referensi yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Anisa Fadilah Saputri
NIM : 2110512010
Tanggal : 7 Juli 2025

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan ini maka saya bersedia dituntut dan diproses sesuai dengan ketentuan berlaku.

Jakarta, 7 Juli 2025



10000
REPUBLIK INDONESIA
METEORAL
TEMPERATUR
AF18AMX359743962

Anisa Fadilah Saputri

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya bertanda tangan di bawah ini :

Nama : Anisa Fadilah Saputri
NIM : 2110512010
Fakultas : Ilmu Komputer
Program Studi : S1 Sistem Informasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalti Free Right) atas karya ilmiah saya yang berjudul:

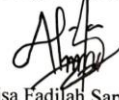
Analisis Data Respons Publik Pengguna Fitur *Live Streaming* TikTok Sebagai Media Hiburan di Indonesia dengan Menggunakan Algoritma *Naive Bayes*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih data/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Dengan demikian pernyataan ini saya buat dengan sebenar-benarnya

Dibuat di : Jakarta
Pada tanggal : 7 Juli 2025

Yang Menyatakan



Anisa Fadilah Saputri

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Judul : Analisis Data Respons Publik Pengguna Fitur *Live Streaming* TikTok
Sebagai Media Hiburan di Indonesia dengan Menggunakan Algoritma
Naive Bayes
Nama : Anisa Fadilah Saputri
NIM : 2110512010
Program Studi : SI Sistem Informasi

Disetujui oleh :

Penguji 1:
Dr. Bambang Saras Yulistiawan, S.T., M.Kom

Penguji 2:
Novi Trisman Hadi, S.Pd., M.Kom

Pembimbing 1:
Dr. Tjahjanto, S.Kom., M.M

Pembimbing 2:
Sarika, S.Kom, M.Kom

Diketahui oleh:

Koordinator Program Studi:
Anita Muliawati, S.Kom., MTI.
NIP. 197005212021212002

Dekan Fakultas Ilmu Komputer:
Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM
NIP. 197605082003121002

Tanggal Ujian Tugas Akhir :
03 Juli 2025



ANALISIS DATA RESPONS PUBLIK PENGGUNA FITUR *LIVE STREAMING* TIKTOK SEBAGAI MEDIA HIBURAN DI INDONESIA DENGAN MENGGUNAKAN ALGORITMA *NAIVE BAYES*

Anisa Fadilah Saputri

ABSTRAK

Fitur live streaming TikTok merupakan layanan interaktif yang memungkinkan pengguna dapat melakukan siaran langsung dan berinteraksi secara real-time dengan audiens. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap fitur live streaming TikTok untuk mengetahui persepsi publik, mengetahui cara membangun model sentimen dengan menggunakan algoritma *naive bayes*, serta dapat memberikan rekomendasi yang relevan kepada pengembang untuk meningkatkan kualitas fitur. Data diperoleh dari hasil kuesioner yang telah disebarluaskan kepada pengguna fitur live streaming TikTok di Indonesia, dengan total data 416 opini yang kemudian dikategorikan ke dalam dua jenis sentimen yaitu positif dan negatif. Proses analisis dilakukan dengan melalui tahap preprocessing data dan labelling data, sebanyak 241 (57.9%) sentimen positif serta 175 (42.1%) sentimen negatif. Kemudian Proses klasifikasi dilakukan dengan menggunakan Multinomial Naive Bayes, serta dilakukan perbandingan dengan algoritma *K-Nearest Neighbors (KNN)* dan *Support Vector Machine (SVM)*. Data dibagi dalam proporsi 60:40, 70:30, dan 80:20. Hasil evaluasi menunjukkan algoritma Multinomial Naive Bayes memberikan performa terbaik dalam klasifikasi sentimen pengguna fitur live streaming TikTok. Performa tertinggi dicapai pada rasio data latih dan uji 80:20, dengan accuracy 90%, precision 91%, recall 89%, dan f1-score 90%. Algoritma KNN dan SVM menunjukkan performa stabil di kisaran accuracy 86%–88% tanpa peningkatan signifikan. Penelitian ini menunjukkan *Naive Bayes* paling optimal untuk klasifikasi sentimen dan efektif dalam memberikan rekomendasi strategis untuk meningkatkan kualitas pengalaman pengguna pada fitur live streaming TikTok.

Kata Kunci: Analisis Sentimen, *Live Streaming*, TikTok, *Naive Bayes*, Klasifikasi

**ANALYSIS OF PUBLIC RESPONSE DATA FROM USERS OF THE
TIKTOK LIVE STREAMING FEATURE AS AN ENTERTAINMENT
MEDIUM IN INDONESIA USING THE
NAIVE BAYES ALGORITHM**

Anisa Fadilah Saputri

ABSTRACT

The TikTok live streaming feature is an interactive service that allows users to broadcast live and interact with audiences in real-time. This study aims to analyze user sentiment toward the TikTok live streaming feature to understand public perception, build a sentiment model using the Naive Bayes algorithm, and provide relevant recommendations for developers to improve the feature's quality. Data were obtained from a questionnaire distributed to users of the TikTok live streaming feature in Indonesia, resulting in a total of 416 opinions categorized into two types of sentiment: positive and negative. The analysis process involved data preprocessing and labeling, resulting in 241 (57.9%) positive sentiments and 175 (42.1%) negative sentiments. Sentiment classification was performed using the Multinomial Naive Bayes algorithm and compared with K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) algorithms. Data were split into 60:40, 70:30, and 80:20 train-test ratios. The evaluation results showed that the Multinomial Naive Bayes algorithm consistently provided the best performance for classifying user sentiment toward the TikTok live streaming feature. The highest performance was achieved at the 80:20 train-test ratio, with 90% accuracy, 91% precision, 89% recall, and 90% f1-score. Meanwhile, the KNN and SVM algorithms showed stable performance in the 86%–88% accuracy range without significant improvement. This study demonstrates that the Naive Bayes algorithm is the most optimal choice for sentiment classification and effective in providing strategic recommendations to enhance the user experience of TikTok's live streaming feature.

Keywords: *Sentiment Analysis, Live Streaming, TikTok, Naive Bayes, Classification*

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa atas limpahan rahmat, anugerah, dan hidayah-Nya sehingga penulis masih diberi kesehatan dan kekuatan untuk dapat menyelesaikan penelitian tugas akhir. Penelitian ini dibuat sebagai salah satu syarat untuk menyelesaikan Tugas Akhir Skripsi pada Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Dalam menyelesaikan penelitian tugas akhir ini, tentu banyak dukungan serta doa dari berbagai pihak sehingga selesainya penelitian ini. Dalam kesempatan ini, penulis ingin mengucapkan terima kasih kepada pihak yang telah membantu pelaksanaan dan penyusunan laporan akhir ini, khususnya kepada:

1. Allah SWT dengan segala nikmat, rahmat, berkat, serta karunia-Nya.
2. Kedua orang tua, kakak, dan adik yang selalu memberikan dukungan, doa, semangat, serta motivasi dan juga materil.
3. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM. Selaku dekan Fakultas Ilmu Komputer.
4. Ibu Anita Muliawati, S. Kom., M. TI selaku Ketua Program Studi S1 Sistem Informasi dan selaku Dosen Pembimbing Akademik.
5. Bapak Dr. Tjahjanto, S. Kom, M.M selaku Dosen Pembimbing 1 atas arahan serta masukan selama penelitian.
6. Ibu Sarika, S. Kom., M. Kom selaku Dosen Pembimbing 2 atas arahan dan masukan selama penelitian.
7. Teman-teman yang selama ini membantu dan mendukung penulis selama proses pendidikan dan penelitian.
8. Pihak *Agency* dan *Family* TikTok yang telah membantu dan mendukung selama proses penelitian.

Jakarta, Juni, 2024



Anisa Fadilah Saputri

DAFTAR ISI

PERNYATAAN ORISINALITAS.....	i
PERNYATAAN PERSETUJUAN PUBLIKASI	ii
LEMBAR PENGESAHAN	iii
ABSTRAK	iv
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Tujuan Penelitian.....	6
1.4. Ruang Lingkup Penelitian.....	6
1.5. Manfaat Penelitian	6
1.6. Luaran Yang Diharapkan.....	7
1.7. Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	9
2.1. Analisis.....	9
2.2. TikTok	9
2.3. Fitur <i>Live Streaming</i> TikTok	10
2.4. <i>Text Mining</i>	11
2.5. <i>Text Preprocessing</i>	12
2.5.1 <i>Cleaning</i>	12
2.5.2 <i>Case Folding</i>	12
2.5.3 <i>Normalization</i>	13
2.5.4 <i>Stopword Removal</i>	13
2.5.5 <i>Tokenizing</i>	13
2.5.6 <i>Stemming</i>	13
2.5.7 <i>Term Frequency-Inverse Document Frequency</i>	14
2.6. <i>Naïve Bayes Classifier</i>	14
2.7. Evaluasi	15
2.8. Pemrograman Python	17
2.9. Kajian Literatur	17
BAB III METODE PENELITIAN.....	21
3.1. Tempat dan Waktu Penelitian.....	21

3.1.1 Tempat Penelitian	21
3.1.2 Waktu Penelitian.....	21
3.2. Diagram Alir Penelitian.....	22
3.2.1. Identifikasi Masalah	23
3.2.2. Studi Literatur.....	23
3.2.3. Pengumpulan Data.....	24
3.2.4. <i>Preprocessing</i> Data.....	24
3.2.5. <i>Translate</i> Data	26
3.2.6. <i>Labelling</i> Data	27
3.2.7. Pembobotan Data.....	27
3.2.8. Pembagian Data.....	27
3.2.9. Penerapan Algoritma <i>Naive Bayes Classifier</i>	28
3.2.10. Evaluasi.....	28
3.2.11. Perbandingan Algoritma	28
3.2.12. Pembuatan Dashboard Visualisasi	29
3.2.13. <i>Insight</i> dan Rekomendasi.....	29
3.3. Alat Bantu Penelitian	30
3.3.1 <i>Hardware</i>	30
3.3.2 <i>Software</i>	30
3.4. Jadwal Pelaksanaan.....	31
BAB IV HASIL DAN PEMBAHASAN.....	32
4.1. Pengumpulan Data	32
4.2. Tahap <i>Preprocessing</i>	36
4.2.1. <i>Cleaning</i> Data.....	36
4.2.2. <i>Case Folding</i>	37
4.2.3. <i>Normalization</i>	39
4.2.4. <i>Stopword Removal</i>	40
4.2.5. <i>Tokenizing</i>	42
4.2.6. <i>Stemming</i>	43
4.3. <i>Translate</i> Data	45
4.4. Labeling Data	46
4.5. Pembobotan Data (TF-IDF)	49
4.6. Pembagian Data	51
4.6.1. Pembagian Data Rasio 60:40.....	52
4.6.2. Pembagian Data Rasio 70:30.....	52
4.6.3. Pembagian Data Rasio 80:20.....	53
4.7. Penerapan Model Algoritma	53
4.7.1. Algoritma <i>Naive Bayes</i>	54
4.7.2. Algoritma <i>K-Nearest Neighbors</i> (KNN)	55

4.7.3. Algoritma <i>Support Vector Machine</i> (SVM)	56
4.8. Evaluasi dan Perbandingan Hasil Kinerja Model	56
4.8.1. Algoritma <i>Naive Bayes</i>	57
4.8.2. Algoritma <i>K-Nearest Neighbors</i> (KNN)	65
4.8.3. Algoritma <i>Support Vector Machine</i> (SVM)	73
4.8.4. Perbandingan Hasil Evaluasi Algoritma	82
4.9. <i>Visualization</i>	84
4.9.1. <i>Visualization</i> Pernyataan <i>Option</i>	85
4.9.2. <i>Visualization</i> Pernyataan Opini atau Ulasan	98
4.10. <i>Insight</i> dan Rekomendasi	107
4.10.1. <i>Insight</i> Algoritma	107
4.10.2. <i>Insight</i> Sentimen Pengguna	110
4.10.3. Rekomendasi	112
BAB V PENUTUP	117
5.1. Kesimpulan	117
5.2. Saran	120
DAFTAR PUSTAKA	121
LAMPIRAN	124

DAFTAR GAMBAR

Gambar 1.1 Diagram Lingkaran Pengguna Pernah mengalami masalah teknis	2
Gambar 1.2 Diagram Permasalahan yang sering dikeluhkan pengguna	3
Gambar 2.1 Pengguna Live Streaming	11
Gambar 3.1 Tahapan Penelitian	22
Gambar 4.1 Kuesioner Form Pribadi	32
Gambar 4.2 Kuesioner Pernyataan Option	33
Gambar 4.3 Kuesioner Pertanyaan Opini atau Ulasan	34
Gambar 4.4 Kode Cleaning Data	36
Gambar 4.5 Kode Case Folding	38
Gambar 4.6 Kode Normalization	39
Gambar 4.7 Kode Stopword Removal	41
Gambar 4.8 Kode Tokenizing	42
Gambar 4.9 Kode Stemming	44
Gambar 4.10 Kode Translate Data	45
Gambar 4.11 Kode Labelling Data	47
Gambar 4.12 Kode Pembobotan Data	50
Gambar 4.13 Kode Pembagian Data dengan berbagai Rasio	51
Gambar 4.14 Kode Naive Bayes Classifier	54
Gambar 4.15 Kode K-Nearest Neighbors (KNN)	55
Gambar 4.16 Kode Support Vector Machine (SVM)	56
Gambar 4.17 Kode Confusion Matrix Naive Bayes	57
Gambar 4.18 Hasil Confusion Matrix Naive Bayes Rasio 60:40	58
Gambar 4.19 Hasil Confusion Matrix Naive Bayes Rasio 70:30	60
Gambar 4.20 Hasil Confusion Matrix Naive Bayes Rasio 80:20	63
Gambar 4.21 Kode Confusion Matrix K-Nearest Neighbors (KNN)	66
Gambar 4.22 Hasil Confusion Matrix K-Nearest Neighbors Rasio 60:40	66
Gambar 4.23 Hasil Confusion Matrix K-Nearest Neighbors Rasio 70:30	69
Gambar 4.24 Hasil Confusion Matrix K-Nearest Neighbors Rasio 80:20	71
Gambar 4.25 Kode Confusion Matrix Support Vector Machine (SVM)	74
Gambar 4.26 Hasil Confusion Matrix Support Vector Machine Rasio 60:40	74

Gambar 4.27 Hasil Confusion Matrix Support Vector Machine Rasio 70:30.....	77
Gambar 4.28 Hasil Confusion Matrix Support Vector Machine Rasio 80:20.....	79
Gambar 4.29 Kode Library Visualization	85
Gambar 4.30 Visualization Bar Aksesibilitas.....	85
Gambar 4.31 <i>Visualization Bar</i> Panduan Penggunaan.....	86
Gambar 4.32 Visualization Bar Performa Fitur	87
Gambar 4.33 <i>Visualization Bar</i> Stabilitas <i>Fitur</i>	88
Gambar 4.34 Visualization Bar Kecepatan dan Kinerja Fitur.....	89
Gambar 4.35 Visualization Bar Efektivitas Respons Aplikasi.....	90
Gambar 4.36 <i>Visualization Bar</i> Keamanan Pengguna dan Pengawasan Konten...	92
Gambar 4.37 Visualization Bar Efektivitas Sistem Deteksi Fitur.....	93
Gambar 4.38 Visualization Bar Manfaat Fitur	94
Gambar 4.39 Visualization Bar Desain antarmuka (UI)	95
Gambar 4.40 Visualization Bar Kepuasan Pengguna	96
Gambar 4.41 <i>Visualization Dashboard</i> Pernyataan Option	97
Gambar 4.42 Kode Visualization Bar Sentimen	98
Gambar 4.43 Visualization Bar Sentimen.....	99
Gambar 4.44 Kode Visualization Diagram Lingkaran.....	100
Gambar 4.45 Visualization Diagram Lingkaran	100
Gambar 4.46 Kode Visualization Donut	101
Gambar 4.47 Visualizatin Donut.....	102
Gambar 4.48 Kode Visualization Word Cloud Positif	103
Gambar 4.49 Visualization Word Cloud Positif.....	103
Gambar 4.50 Kode Visualization Word Cloud Negatif.....	104
Gambar 4.51 Visualization Word Cloud Negatif	105
Gambar 4.52 Visualization Dashboard Pernyataan Opini.....	106

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	15
Tabel 2.2 Kajian Literatur	18
Tabel 3.1 Jadwal Kegiatan	31
Tabel 4.1 Sampel Data Pernyataan Opini atau Ulasan.....	34
Tabel 4.2 Sampel Data Pernyataan Option	35
Tabel 4.3 Tabel Hasil Cleaning Data.....	37
Tabel 4.4 Hasil Case Folding	38
Tabel 4.5 Hasil Normalization	39
Tabel 4. 6 Hasil Stopword Removal	41
Tabel 4.7 Hasil Tokenizing.....	42
Tabel 4.8 Hasil Stemming	44
Tabel 4.9 Hasil Translate Data	46
Tabel 4. 10 Hasil Labelling Data.....	48
Tabel 4.11 Hasil Pembobotan Data	50
Tabel 4.12 Pembagian Data Rasio 60:40	52
Tabel 4.13 Pembagian Data Rasio 70:30	53
Tabel 4.14 Pembagian Data Rasio 80:20	53
Tabel 4.15 Perbandingan Hasil Evaluasi Berdasarkan Rasio dan Algoritma.....	82
Tabel 4.16 Perbandingan Komputasi Algoritma	108

DAFTAR LAMPIRAN

Lampiran 1. Hasil Data Responden Kuesioner Pra Penelitian.....	124
Lampiran 2. Hasil Data Responden Kuesioner Penelitian	124
Lampiran 3. Kode Hasil Tahap Preprocessing Data	125
Lampiran 4. Kode dan Hasil Tahap Translate Data	129
Lampiran 5. Kode dan Hasil Tahap Labelling Data.....	130
Lampiran 6. Kode dan Hasil Tahap Pembobotan Data	131
Lampiran 7. Sampel Data.....	132
Lampiran 8. Kode dan Hasil Pembagian Data, Model Evaluasi Naive Bayes ...	133
Lampiran 9. Kode dan Hasil Pembagian Data, Model Evaluasi KNN	136
Lampiran 10 Kode dan Hasil T Pembagian Data, Model Evaluasi SVM.....	139
Lampiran 11. Kode dan Hasil Tahap Visualisasi Dashboard	142
Lampiran 12. Turnitin	147