



BERITA ACARA SIDANG PENDADARAN TUGAS AKHIR

S/UBL/FTI/0749/VII/26

Pada hari ini, Rabu 15 Juli 2026 telah dilaksanakan Ujian Sidang Pendadaran Tugas Akhir sebagai berikut:

Judul: IMPLEMENTASI NAIVE BAYES, SVM DAN LOGISTIC REGRESSION PADA ANALISIS SENTIMEN PENGGUNA X TERHADAP DEVELOPER CAPCOM

Nama : Yoga Dwi Pamungkas
NIM : 2211501305
Dosen Pembimbing : Dolly Virgian Shaka Yudha Sakti, S.Kom., M.Kom.

Berdasarkan penilaian pada Presentasi + Demo, Penulisan, Penguasaan Materi, Penguasaan Program maka Mahasiswa tersebut di atas dinyatakan:

LULUS

dengan nilai angka : **84** huruf : **A-**

Mahasiswa tersebut di atas wajib menyerahkan hasil perbaikan tulisan Tugas Akhir dalam bentuk terjilid sesuai dengan Panduan Perbaikan Tugas Akhir, selambat-lambatnya Rabu 29 Juli 2026.

Panitia Penguji:

- | | |
|-------------|---|
| 1 Ketua | Dr. Ir. Achmad Solichin, S.Kom., M.T.I. |
| 2 Anggota | Painem, S.Kom., M.Kom. |
| 3 Moderator | Dolly Virgian Shaka Yudha Sakti, S.Kom., M.Kom. |

Keterangan:

Nilai Huruf: A:85-100 A-:80-84,99 B+:75-79,99 B:70-74,99 B-:65-69,99 C:60-64,99 D:40-59,99 E-:0-39,99



UNIVERSITAS BUDI LUHUR

FAKULTAS TEKNOLOGI INFORMASI

Kartu Bimbingan Tugas Akhir

NIM: 2211501305

Nama: Yoga Dwi Pamungkas

Pembimbing: Dolly Virgiana Shaka Yudha Sakti, S.Kom., M.Kom.

No.	Tanggal	Materi
1	02-04-2026	Perkenalan diri kepada dosen dan mengajukan judul tugas akhir
2	06-04-2026	Mengkonfirmasi Kembali Terkait Judul Tugas Akhir
3	20-04-2026	Berkonsultasi Terkait Data Crawling Yang didapat
4	23-04-2026	Perubahan Judul Tugas Akhir Karena Data Terlalu Sedikit Yang Didapat
5	24-04-2026	Akun X/Twitter Terkena ban karena terlalu banyak mengambil data secara langsung
6	28-04-2026	Mendapatkan dataset sebanyak 400+
7	12-05-2026	Bimbingan bersama dengan mahasiswa lain dan mengkonfirmasi kembali terkait judul tugas akhir
8	29-05-2026	Konsultasi kepada dosen pembimbing tugas akhir terkait Website dan dataset sebelum diberikan kepada pakar
9	04-06-2026	Bertemu dan menunjukkan hasil website beserta revisiannya

**IMPLEMENTASI *NAÏVE BAYES*, SVM DAN *LOGISTIC REGRESSION*
PADA ANALISIS SENTIMEN PENGGUNA X TERHADAP
DEVELOPER CAPCOM**

TUGAS AKHIR



Oleh :

Yoga Dwi Pamungkas

NIM : 2211501305

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

JAKARTA

2026

**IMPLEMENTASI *NAÏVE BAYES*, SVM DAN *LOGISTIC REGRESSION*
PADA ANALISIS SENTIMEN PENGGUNA X TERHADAP
DEVELOPER CAPCOM**

**Diajukan untuk memenuhi salah satu persyaratan
Memperoleh gelar Sarjana Komputer (S.Kom)**

TUGAS AKHIR



Oleh :

Yoga Dwi Pamungkas

NIM : 2211501305

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

JAKARTA

2026



LEMBAR PENGESAHAN

Nama : Yoga Dwi Pamungkas
Nomor Induk Mahasiswa : 2211501305
Program Studi : Teknik Informatika
Bidang Peminatan : Cyber Security
Jenjang Studi : Strata 1
Judul : IMPLEMENTASI NAIVE BAYES, SVM DAN LOGISTIC REGRESSION
PADA ANALISIS SENTIMEN PENGGUNA X TERHADAP DEVELOPER
CAPCOM



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Rabu 15 Juli 2026

Tim Penguji:

Ketua : Dr. Ir. Achmad Solichin, S.Kom., M.T.I.
Anggota : Painem, S.Kom., M.Kom.
Pembimbing : Dolly Virgian Shaka Yudha Sakti, S.Kom., M.Kom.
Ketua Program Studi : Dr. Indra, S.Kom., M.T.I.

ABSTRAK

Implementasi *Naïve Bayes*, SVM dan *Logistic Regression* pada Analisis Sentimen Pengguna X terhadap Developer Capcom

Oleh : Yoga Dwi Pamungkas (2211501305)

Capcom merupakan salah satu perusahaan pengembang video game yang dikenal melalui judul-judul populer seperti *Resident Evil*, *Monster Hunter*, dan *Street Fighter*. Banyaknya komentar, kritik, dan apresiasi dari pengguna media sosial terhadap produk yang dikembangkan menyebabkan volume data terus meningkat sehingga analisis secara manual menjadi kurang efisien. Oleh karena itu, penelitian ini menerapkan teknik *text mining* untuk menganalisis kecenderungan sentimen pengguna media sosial X terhadap *developer* game Capcom. Data penelitian dikumpulkan melalui proses *crawling* menggunakan *Tweet Harvest* pada platform X, kemudian melalui tahapan *preprocessing* yang meliputi *case folding*, *cleansing*, normalisasi *slangword*, *stopword removal*, *stemming*, dan *tokenizing*. Selanjutnya, pembobotan kata dilakukan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), dan klasifikasi sentimen dilakukan menggunakan tiga algoritma, yaitu *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Logistic Regression*. Evaluasi model dilakukan menggunakan *Stratified 5-Fold Cross Validation* dengan parameter *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Hasil pengujian pada dataset asli menunjukkan *Naïve Bayes* memperoleh *Accuracy* tertinggi sebesar 70,90%, SVM memperoleh *Precision* tertinggi sebesar 52,17%, sedangkan *Logistic Regression* memperoleh *Recall* dan *F1-Score* tertinggi masing-masing sebesar 47,35% dan 47,40%. Berdasarkan *F1-Score* sebagai indikator utama, *Logistic Regression* ditetapkan sebagai algoritma dengan performa terbaik dalam penelitian ini. Hasil pengujian juga menunjukkan bahwa distribusi jumlah data pada setiap kelas sentimen memengaruhi kinerja model, di mana dataset dengan distribusi kelas yang lebih seimbang menghasilkan kinerja klasifikasi yang lebih stabil dibandingkan dataset dengan distribusi tidak seimbang. Selain itu, penelitian ini menghasilkan sebuah sistem berbasis web yang mampu melakukan proses *crawling*, *preprocessing*, pembobotan TF-IDF, pelabelan data, klasifikasi sentimen, dan evaluasi model secara terintegrasi.

Kata Kunci : *Text mining*, Analisis Sentimen, *Naïve Bayes*, *Support Vector Machine*, *Logistic Regression*, Media Sosial X

DAFTAR ISI

Lembar Pengesahan.....	i
ABSTRAK	ii
Surat Pernyataan Tidak Plagiat	iii
Surat Pernyataan Penggunaan AI Generatif	iv
KATA PENGANTAR.....	v
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
DAFTAR SIMBOL	ix
DAFTAR ALGORITMA	x
DAFTAR ISI.....	xi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan.....	2
1.5 Manfaat	3
1.5.1 Manfaat Teoritis.....	3
1.5.2 Manfaat Praktis.....	3
1.6 Sistematika Penulisan.....	3
BAB II LANDASAN TEORI	5
2.1 Pengertian Text Mining	5
2.2 Seleksi Data.....	5
2.3 Preprocessing	5
2.3.1 Case Folding.....	6
2.3.2 Cleansing	6
2.3.3 Slangword.....	6
2.3.4 Stopword.....	6
2.3.5 Stemming.....	6
2.3.6 Tokenizing	6
2.4 Term Frequency-Inverse Document Frequency (TF-IDF).....	7
2.4.1 Term Frequency (TF).....	7
2.4.2 Document Frequency (DF).....	7
2.4.3 Inverse Document Frequency (IDF).....	7
2.4.4 Pembobotan TF-IDF	8
2.5 Naïve Bayes	8
2.5.1 Teorema Bayes	8
2.5.2 Naïve Bayes untuk Klasifikasi Teks	8
2.5.3 Kelebihan dan Kekurangan Naïve Bayes	9
2.6 Support Vector Machine (SVM)	9
2.6.1 Konsep Hyperplane	9
2.6.2 Cara Kerja Support Vector Machine.....	10
2.6.3 Fungsi Keputusan SVM	10
2.7 Logistic Regression	10
2.7.1 Fungsi Sigmoid.....	10
2.7.2 Cara Kerja Logistic Regression.....	11
2.8 K-Fold Cross Validation.....	11
2.8.1 Ilustrasi K-Fold Cross Validation	11
2.8.2 Kelebihan K-Fold Cross Validation.....	12

2.8.3 Penerapan Pada Penelitian.....	12
2.9 Confusion Matrix	12
2.9.1 Accuracy	12
2.9.2 Precision	13
2.9.3 Recall.....	13
2.9.4 F1-Score	13
2.9.5 Penerapan Pada Penelitian.....	13
2.10 Penelitian Terdahulu.....	13
BAB III METODOLOGI PENELITIAN	16
3.1 Data Penelitian	16
3.2 Metode Pembandingan.....	18
3.3 Tahap Penelitian	19
3.4 Implementasi Metode Penelitian.....	20
3.4.1 Crawling Data.....	20
3.4.2 Implementasi Seleksi Data	21
3.4.3 Implementasi Preprocessing.....	21
3.4.4 Implementasi TF-IDF	23
3.4.5 Implementasi Stratified 5-Fold Cross Validation	24
3.4.6 Implementasi Naïve Bayes	25
3.4.7 Implementasi Support Vector Machine (SVM).....	26
3.4.8 Implementasi Logistic Regression	26
3.4.9 Implementasi Prebandingan Hasil Evaluasi Model.....	26
3.4.10 Implementasi Penentuan Metode Terbaik	26
3.4.11 Implementasi Analisis Sentimen Menggunakan Metode Terbilang	26
3.4.12 Rancangan Menu	27
3.4.13 Rancangan Layar	27
BAB IV HASIL DAN PEMBAHASAN.....	35
4.1 Implementasi Sistem	35
4.1.1 Lingkungan Implementasi Sistem	35
4.2 Implementasi Metode.....	37
4.2.1 Implementasi Crawling Data.....	37
4.2.2 Implementasi Preprocessing.....	39
4.2.3 Implementasi Pelabelan Data Pakar	57
4.2.4 Implementasi Term Frequency-Inverse Document Frequency (TF-IDF)	60
4.2.5 Implementasi Stratified K-Fold Cross Validation	62
4.2.6 Implementasi Pemodelan Naïve Bayes	64
4.2.7 Implementasi Pemodelan SVM.....	66
4.2.8 Implementasi Pemodelan Logistic Regression.....	68
4.2.9 Implementasi Evaluasi Model	70
4.2.10 Implementasi Visualisasi Data.....	73
4.2.11 Implementasi Hasil Sentimen.....	76
4.3 Analisis Hasil Penelitian	77
4.3.1 Analisis Skenario Pengujian Dataset	77
4.4 Flowchart.....	79
4.4.1 Flowchart Keseluruhan.....	79
4.4.2 Flowchart Preprocessing	80
4.4.3 Flowchart Pembagian Data Stratified 5-Fold Cross Validation.....	81
4.5 Algoritma.....	82
4.5.1 Algoritma Preprocessing	82
4.5.2 Algoritma Stratified 5 Fold Cross Validation	82

4.6 Implementasi Antarmuka Sistem	83
4.6.1 Halaman Dashboard	83
4.6.2 Halaman Crawling Data	84
4.6.3 Halaman Upload Dataset Crawling	84
4.6.4 Halaman Preprocessing	85
4.6.5 Halaman Pengambilan Dataset Label Pakar	85
4.6.6 Halaman Upload Dataset Label Pakar	86
4.6.7 Halaman TF-IDF	86
4.6.8 Halaman Stratified K-Fold	87
4.6.9 Halaman Hasil Prediksi	88
4.6.10 Halaman Evaluasi Model	90
4.6.11 Halaman Visualisasi Data	91
4.6.12 Halaman Hasil Sentimen	91
BAB V PENUTUP	92
5.1 Kesimpulan	92
5.2 Saran	93
DAFTAR PUSTAKA	94
LAMPIRAN	95

DAFTAR PUSTAKA

- Atmajaya, D., Febrianti, A., & Darwis, H. (2023). Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter. *The Indonesian Journal of Computer Science*, 12(4), 2173–2181.
- Bachtiar, A., Hawali, M. J., Simbolon, N. C., Maulana, D. A., & Anggraini, R. A. (2026). Analisis Sentimen Ulasan Free Fire Menggunakan Naive Bayes Logistic Regression. *Journal of Information System Management (JOISM)*, 7(2).
- Gunawan, B., Sastypratiwi, H., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 4(2), 113–118.
- Harahap, E. D., & Kurniawan, R. (2024). Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat (TAPERA) Pada Aplikasi X Menggunakan Metode Naive Bayes. *Jurnal Teknik Informatika UNIKA Santo Thomas (JTIUST)*, 9(1), 166–175.
- Hidayat, T., Siddiq, M. J., Jayasri, S., Suhendi, A., & Rizky, R. (2024). Analisis Sentimen Opini Masyarakat Terhadap Pilkada 2024 di Media Sosial Twitter Menggunakan Algoritma Naive Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). New York: Wiley.
- Indriyani, F. A., Fauzi, A., & Faisal, S. (2023). Analisis Sentimen Aplikasi TikTok Menggunakan Algoritma Naive Bayes dan Support Vector Machine. *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, 10(2), 176–184.
- Maulana, N. A., & Darwis, D. (2025). Perbandingan Metode SVM dan Naive Bayes untuk Analisis Sentimen pada Twitter tentang Obesitas di Kalangan Gen Z. *Jurnal Pendidikan dan Teknologi Indonesia (JPTI)*, 5(3), 655–666.
- Prajela, A., Syafriandi, S., Permana, D., & Fitria, D. (2024). Twitter Data Sentiment Analysis 2024 Presidential Candidate Using Algorithm Naive Bayes Classifier By Methods K-Fold Cross Validation. *UNP Journal of Statistics and Data Science*, 2(1), 99–104.
- Pratiwi, R. H., Dairoh, D., Afidah, D. I., Rifa, Q., & Gian, A. F. (2021). Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM). *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 4(1), 40–46.
- Putra, J. A., Dharmawan, A., & Gondohanindijo, J. (2024). Sentimen Analisis Aplikasi Digitalent Mobile Menggunakan Naive Bayes dan SVM Dengan Ekstraksi Fitur TF-IDF. *INTECOMS: Journal of Information Technology and Computer Science*, 7(4), 1139–1148.
- Rachmatullah, M. M., Irawan, B., Faqih, A., Pratama, D., & Kurnia, D. A. (2026). Analisis Komparatif Multinomial Naive Bayes dan Logistic Regression untuk Klasifikasi Sentimen Ulasan Pengguna Aplikasi TIX ID. *JSI (Jurnal Sistem Informasi) Universitas Suryadarma*, 13(1), 62–70.
- Saputra, J., Maryani, L., Rahmadden, R., Wulandari, D., & Eka, W. (2025). Analisis Performa Naive Bayes dan SVM Terhadap Sentimen Teks Media Sosial Dengan Word2Vec dan SMOTE. *Jurnal INSTEK*, 10(1).
- Septiani, D., & Isabela, I. (2022). Analisis Term Frequency Inverse Document Frequency (TF-IDF) dalam Temu Kembali Informasi pada Dokumen Teks. *SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia*, 1(2), 81–88.