

Penerapan Algoritma *Machine Learning* untuk Klasifikasi Pesan Media Sosial pada Perguruan Tinggi

Dewi Ariani ¹, Rully Pramudita ^{2,*}

^{1,2} Teknik Informatika; Universitas Bina Insani; Jl. Raya Siliwangi No.6, RT.001/RW.004, Sepanjang Jaya, Kec. Rawalumbu, Kota Bekasi, Jawa Barat 17114; e-mail: [1dewiariani185@gmail.com](mailto:dewiariani185@gmail.com), [2rullypramudita@binainsani.ac.id](mailto:rullypramudita@binainsani.ac.id)

* Korespondensi: e-mail: rullypramudita@binainsani.ac.id

Diterima: 27 Februari 2026; Review: 26 Mei 2026; Disetujui: 11 Juni 2026

Cara sitasi: Ariani D, Pramudita R. 2026. Penerapan Algoritma *Machine Learning* Untuk Klasifikasi Pesan Media Sosial Pada Perguruan Tinggi. *Information System for Educators and Professionals*. Vol 11(1): 59-70.

Abstrak: Di era digital, media sosial seperti *Direct Message (DM)* Instagram dan WhatsApp seringkali menjadi sarana utama komunikasi antara perguruan tinggi dan calon mahasiswa. Universitas Bina Insani menerima banyak pesan setiap harinya yang berisi berbagai pertanyaan terkait pendaftaran, informasi_kampus, biaya, kelas_karyawan, jurusan, beasiswa dan lainnya. Namun, pesan-pesan tersebut masih dikelola secara manual tanpa pengelompokan topik yang terstruktur, sehingga menyulitkan admin dalam mengidentifikasi kebutuhan informasi yang paling sering diajukan. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi pesan media sosial menggunakan algoritma *machine learning* guna mengelompokkan pesan berdasarkan topik secara otomatis. Dataset yang digunakan terdiri dari 350 pesan yang dikumpulkan dari Instagram dan WhatsApp, yang telah diberi label ke dalam beberapa kategori. Tahapan pengolahan data meliputi pra-pemrosesan teks, ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency (TF-IDF)*, serta proses klasifikasi menggunakan algoritma *Support Vector Machine (SVM)*. Hasil pengujian menunjukkan bahwa model SVM mampu melakukan klasifikasi pesan dengan tingkat akurasi sebesar 61%. Untuk meningkatkan performa pada kelas dengan jumlah data yang tidak seimbang, diterapkan pendekatan gabungan antara model machine learning dan *rule-based* berbasis kata kunci, yang menghasilkan peningkatan akurasi hingga 70%. Sistem yang dikembangkan kemudian diimplementasikan dalam bentuk dashboard interaktif berbasis Streamlit untuk memudahkan pengguna dalam mengunggah data, melihat hasil klasifikasi, dan memantau distribusi topik pesan.

Kata kunci: Klasifikasi Pesan, *Machine Learning*, Media Sosial, Streamlit, *Support Vector Machine*

Abstract: In the digital era, social media platforms such as Instagram Direct Messages (DM) and WhatsApp have become the primary communication channels between higher education institutions and prospective students. Universitas Bina Insani receives a large number of messages daily containing various inquiries related to admission information, tuition fees, study programs, scholarships, and campus services. However, these messages are still managed manually without structured topic classification, making it difficult for administrators to identify the most frequently requested information. This study aims to develop a social media message classification system using machine learning algorithms to automatically group messages based on their topics. The dataset used in this study consists of 350 messages collected from Instagram and WhatsApp, which were manually labeled into several categories. The data processing stages include text preprocessing, feature extraction using *Term Frequency–Inverse Document Frequency (TF-IDF)*, and message classification using the *Support Vector Machine (SVM)* algorithm.

The experimental results show that the SVM model achieved a classification accuracy of 61%. To improve performance on imbalanced classes, a hybrid approach combining machine learning and keyword-based rule-based methods was applied, resulting in an accuracy improvement of up

to 70%. The developed system was then implemented in the form of an interactive dashboard based on Streamlit, enabling users to upload data, view classification results, and monitor the distribution of message topics more effectively.

Keywords: Message classification, Machine Learning, Social Media, Streamlit, Support Vector Machine.

1. Pendahuluan

Di era digital ini, Instagram *Direct Message* (DM) dan WhatsApp seringkali menjadi sarana komunikasi utama dengan calon mahasiswa dan masyarakat umum bagi kampus Universitas Bina Insani. Melalui platform tersebut, masuk berbagai pesan yang berisi pertanyaan terkait informasi Penerimaan Mahasiswa Baru (PMB), biaya pendidikan, program studi, beasiswa, kelas karyawan, serta layanan kampus lainnya [1]. Pesan-pesan ini mencerminkan kebutuhan informasi bagi customer dan menjadi sumber data yang potensial untuk dianalisis lebih lanjut. Sejumlah penelitian menunjukkan bahwa pesan teks pada media sosial dapat dimanfaatkan sebagai data yang valid untuk memahami kebutuhan dan minat pengguna secara langsung, karena bersifat spontan dan berasal dari pengalaman nyata pengguna.

Namun, dalam praktiknya, pesan yang diterima masih dikelola secara manual tanpa adanya pengelompokan topik yang terstruktur. Kondisi ini menyebabkan pihak admin kesulitan dalam mengidentifikasi topik apa yang paling sering ditanyakan serta menyusun laporan kuantitatif sebagai bahan evaluasi layanan informasi. Seiring dengan meningkatnya jumlah pengguna media sosial dan calon mahasiswa setiap tahunnya, kebutuhan akan sistem pengelolaan pesan yang efisien dan berbasis data menjadi semakin penting. Oleh karena itu, penerapan teknik klasifikasi teks menjadi solusi yang diharapkan dapat mengelola pesan dalam jumlah besar secara otomatis.

Klasifikasi pesan merupakan salah satu penerapan *text mining* yang bertujuan untuk mengelompokkan dokumen atau pesan ke dalam kategori tertentu berdasarkan isi teksnya. Pendekatan ini telah banyak diterapkan dalam berbagai bidang, seperti layanan pelanggan, manajemen komunikasi digital, dan sistem informasi pendidikan. Dengan memanfaatkan algoritma *machine learning*, pesan dapat diklasifikasikan secara otomatis ke dalam kategori topik tertentu, sehingga membantu institusi dalam memahami pola kebutuhan informasi audiens secara lebih sistematis [2].

Salah satu algoritma *machine learning* yang banyak digunakan dalam klasifikasi teks adalah Support Vector Machine (SVM). SVM dikenal memiliki performa yang baik dalam menangani data teks berdimensi tinggi, terutama ketika dikombinasikan dengan teknik ekstraksi fitur seperti Term Frequency–Inverse Document Frequency (TF-IDF). Beberapa penelitian menunjukkan bahwa SVM mampu menghasilkan tingkat akurasi yang lebih stabil dibandingkan algoritma lain dalam tugas klasifikasi teks pendek, seperti pesan media sosial [3].

Dalam konteks perguruan tinggi, pemanfaatan *machine learning* untuk klasifikasi pesan diharapkan dapat memberikan nilai strategis dalam mendukung pengambilan keputusan berbasis data. Informasi mengenai topik pesan yang paling sering diajukan dapat digunakan sebagai dasar evaluasi layanan informasi, atau mungkin penyusunan *Frequently Asked Questions* (FAQ), serta perencanaan strategi komunikasi yang lebih efektif. Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada klasifikasi teks umum atau ulasan produk, sementara penerapan klasifikasi pesan media sosial secara spesifik pada layanan informasi perguruan tinggi masih relatif terbatas.

Berbeda dengan penelitian klasifikasi teks konvensional yang hanya mengandalkan single classifier, penelitian ini mengusulkan pendekatan hybrid classification framework yang mengintegrasikan machine learning dan rule-assisted hybrid classification framework untuk menangani distribusi kelas yang tidak seimbang pada pesan media sosial perguruan tinggi. Pendekatan ini dirancang untuk meningkatkan robustness klasifikasi pada kategori minoritas yang umumnya sulit dipelajari menggunakan supervised learning murni. Dengan demikian, kontribusi penelitian tidak hanya terletak pada implementasi sistem klasifikasi, tetapi juga pada perancangan strategi hybrid yang lebih sesuai untuk karakteristik data komunikasi institusi pendidikan.

Berdasarkan uraian tersebut, penelitian ini berfokus pada pengembangan sistem klasifikasi pesan media sosial menggunakan algoritma *machine learning* untuk mengelompokkan pesan

dari DM Instagram dan WhatsApp ke dalam beberapa kategori topik. Sistem yang dibangun mengimplementasikan algoritma *Support Vector Machine* (SVM) dengan fitur TF-IDF, serta disajikan dalam bentuk dashboard interaktif berbasis Streamlit. Streamlit dipilih karena mampu menyederhanakan proses pengembangan antarmuka tanpa memerlukan pengaturan kompleks menggunakan HTML, CSS, maupun JavaScript [4]. Kontribusi utama dari penelitian ini adalah menyediakan sistem klasifikasi pesan yang terotomatisasi dan mudah digunakan, sehingga membantu admin perguruan tinggi dalam mengelola pesan masuk secara lebih efektif dan berbasis data.

2. Metode Penelitian

Pendekatan kuantitatif digunakan dalam penelitian ini, dengan fokus pada pengumpulan dan pengolahan data dari pesan masuk melalui Instagram *Direct Message* (DM) dan WhatsApp Universitas Bina Insani. Pendekatan kuantitatif dipilih karena memungkinkan analisis data secara objektif dan terukur, serta mendukung penerapan algoritma *machine learning* dalam proses klasifikasi pesan berdasarkan topik. Secara umum, penelitian ini terdiri dari lima tahapan utama, yaitu: (1) pengumpulan data, (2) pra-pemrosesan data, (3) pelatihan model, (4) pengujian model, dan (5) implementasi sistem [5]. Diagram alur ditunjukkan pada Gambar 1.



Sumber: Hasil Penelitian (2025)

Gambar 1. Metode Penelitian

1. Pengumpulan Data

Pengumpulan data dilakukan dengan mengumpulkan pesan masuk dari akun resmi Instagram dan melakukan *export* data WhatsApp Universitas Bina Insani. Dataset awal terdiri dari 350 pesan yang dikirim oleh calon mahasiswa maupun masyarakat umum dari periode Januari 2025-Mei 2025 sebagai bahan data latih untuk system yang akan dibangun menggunakan streamlit. Setiap pesan telah diberi label secara manual ke dalam kategori topik, seperti pendaftaran, biaya, jurusan, beasiswa, kelas karyawan, dan informasi kampus. Pendekatan ini digunakan karena pesan teks dari media sosial bersifat tidak terstruktur dan beragam, sehingga diperlukan metode manual untuk memberikan label awal yang valid. Dataset yang telah dilabeli kemudian disimpan dalam format CSV agar dapat digunakan dalam tahap pra-pemrosesan dan pemodelan. Pengumpulan data ini juga sejalan dengan praktik penelitian sebelumnya yang menekankan pentingnya dataset terlabel untuk pengembangan model klasifikasi pesan [4], [5].

2. Pemrosesan Data



Sumber: Hasil Penelitian (2025)

Gambar 2. Diagram Alir Tahap Pra-pemrosesan Teks

Tahap pra-pemrosesan dilakukan untuk menyiapkan data agar dapat diolah oleh algoritma *machine learning* sebagaimana dijelaskan pada Gambar 2. Seluruh ulasan terlebih dahulu diubah menjadi huruf kecil (*lowercasing*) untuk menyeragamkan format. Karakter yang tidak relevan, seperti angka, tanda baca, serta spasi berlebih, dihapus agar tidak mengganggu proses analisis. Selain itu, kata-kata umum yang tidak memiliki bobot penting dalam penentuan kategori, misalnya “yang”, “dan”, atau “dengan”, dihapus dengan menggunakan daftar *stopword* Bahasa Indonesia dari pustaka NLTK. Lalu, teks dipecah menjadi token (*tokenization*) dan disusun kembali menjadi representasi teks bersih. Tahapan berikutnya adalah ekstraksi fitur dengan *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk mengubah data teks ke dalam bentuk vektor numerik

3. Pelatihan Model

Pada tahap pelatihan, digunakan algoritma *Support Vector Machine* (SVM) sebagai metode utama untuk klasifikasi pesan berdasarkan topik. SVM dipilih karena performanya yang baik dalam membedakan data antar kelas dan mampu menangani data berdimensi tinggi yang dihasilkan dari TF-IDF. Selain SVM, pengujian awal juga dilakukan menggunakan *Naive Bayes* dan *Logistic Regression* untuk membandingkan performa model. Model kemudian dilatih menggunakan Python di Google Colab, yang menyediakan lingkungan komputasi berbasis cloud untuk eksperimen *machine learning* [6],[7]. Untuk mengatasi ketidakseimbangan jumlah pesan pada beberapa kategori minoritas, diterapkan pendekatan gabungan antara *machine learning* dan *rule-based* berbasis kata kunci [8]. Metode ini memungkinkan pesan dengan kata kunci spesifik, seperti “beasiswa”, “biaya”, atau “kelas karyawan”, langsung dikategorikan, sehingga meningkatkan akurasi keseluruhan.

4. Pengujian Model

Kinerja model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan F1-score. Hasil pengujian dibandingkan untuk menentukan algoritma terbaik [9],[10]. Evaluasi dilakukan menggunakan data uji yang telah dipisahkan dari data pelatihan dengan rasio 80:20. Pemilihan empat metrik tersebut bertujuan untuk memberikan gambaran performa model secara menyeluruh, tidak hanya dari sisi akurasi, tetapi juga keseimbangan antara kemampuan model dalam mengenali kelas positif, negatif, dan netral. *Precision* digunakan untuk mengukur ketepatan prediksi terhadap kelas yang benar, sedangkan *recall* mengukur sejauh mana model mampu menangkap seluruh data dari setiap kelas. Nilai F1-score menjadi indikator rata-rata harmonis dari *precision* dan *recall*, sehingga sangat berguna ketika distribusi data tidak seimbang.

5. Implementasi Sistem

Model terbaik diintegrasikan ke dashboard Streamlit untuk menampilkan hasil klasifikasi pesan secara interaktif [11]. Sistem ini dibangun dengan pendekatan *Agile Development* yang memungkinkan proses iteratif berupa pengujian, evaluasi, dan perbaikan sistem [12]. Dashboard Streamlit dikembangkan menggunakan bahasa pemrograman Python dan dihubungkan langsung dengan model *machine learning* hasil pelatihan untuk melakukan klasifikasi pesan berdasarkan input pengguna [11],[13]. Melalui antarmuka yang sederhana, pengguna dapat memasukkan teks ulasan baru atau memuat dataset dalam format CSV untuk dianalisis secara otomatis. Hasil analisis kemudian ditampilkan dalam bentuk visualisasi grafik batang dan tabel distribusi katego, sehingga memudahkan interpretasi data. Selain itu, sistem juga menampilkan ringkasan proporsi kategori pesan.

3. Hasil dan Pembahasan

3.1. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 350 pesan yang dikumpulkan dari akun resmi Dataset untuk klasifikasi pesan masuk berasal dari kumpulan pertanyaan mahasiswa melalui Whatsapp dan Instagram. Dataset ini berjumlah 350 dari kurun waktu Januari 2025 – Mei 2025 dan telah dikategorikan secara manual ke dalam beberapa kategori seperti pendaftaran, informasi_kampus, biaya, kelas_karyawan, jurusan, beasiswa dan lainnya. Meskipun ukuran dataset relatif terbatas, penelitian ini berfokus pada domain-specific classification pada konteks komunikasi perguruan tinggi, di mana ketersediaan data terlabel secara manual masih menjadi tantangan. Oleh karena itu, penelitian ini diposisikan sebagai exploratory implementation study untuk mengevaluasi efektivitas pendekatan hybrid pada lingkungan institusi pendidikan. Berikut merupakan contoh lima entri dari dataset yang telah dikumpulkan dan diberi label secara manual:

Tabel IV. 1

Dataset Klasifikasi Pesan

No	Pesan Asli	Kategori
1	Ka mau tanya jurusan di bina insani ada apa aja	Jurusan
2	Per semester brp	Biaya
3	Bisa nanya info kelas karyawan	Kelas_Karyawan
4	Pendaftaran nya sampe kapan gelombang pertama	Pendaftaran
5	Selamat pagi, saya Herning lulusan SMK tahun 2024. Saya mau bertanya apakah masih ada program beasiswa KIP tahun 2025? Jika masih ada waktunya kapan yaa.. Terima kasih 🙏	Beasiswa
...

No	Pesan Asli	Kategori
350	Lokasi kampus dimana kak?	Informasi_kampus

Sumber: Hasil Penelitian (2025)

3.2. Hasil Pra-pemrosesan Data

Pra-pemrosesan dilakukan untuk membersihkan pesan teks agar dapat dikenali dengan baik oleh algoritma klasifikasi. Adapun tahapan yang dilakukan adalah sebagai berikut:

```
stop_words = set(stopwords.words('indonesian'))

def bersih_teks(teks):
    teks = teks.lower()
    teks = re.sub(r'^a-zA-Z\s', '', teks)
    teks = teks.strip()
    teks = ' '.join([kata for kata in teks.split() if kata not in stop_words])
    return teks
```

Sumber: Hasil Penelitian (2025)

Gambar 3
Pra-pemrosesan Data Klasifikasi

- 1) *Lowercasing*
Semua huruf diubah menjadi huruf kecil untuk menghindari perbedaan makna akibat kapitalisasi.
- 2) Penghapusan Simbol dan Angka
Teks dibersihkan dari karakter selain huruf dan spasi, seperti angka, tanda baca, dan simbol lainnya.
- 3) Penghapusan Spasi Berlebih
Spasi kosong pada awal atau akhir teks dihapus.
- 4) *Stopword Removal*
Kata-kata umum yang tidak memberikan informasi penting, seperti “yang”, “dengan”, “dan”, dihapus menggunakan daftar stopwords Bahasa Indonesia dari pustaka *nlTK*.
- 5) Ekstraksi Fitur
Setelah dibersihkan, data diubah ke dalam bentuk numerik menggunakan TF-IDF Vectorizer agar bisa diproses lebih lanjut oleh algoritma *machine learning*.

```
tfidf_vectorizer = TfidfVectorizer(
    ngram_range=(1,2),
    max_df=0.85,
    min_df=3,
    sublinear_tf=True,
    stop_words=list(stop_words)
)
```

Sumber: Hasil Penelitian (2025)

Gambar 4
TF-IDF Klasifikasi

Setelah dilakukan tahapan prapemrosesan, setiap komentar diubah menjadi bentuk yang lebih sederhana agar dapat diproses oleh algoritma *machine learning*. Berikut merupakan contoh hasil prapemrosesan dari lima data pertama:

Tabel IV. 2
Hasil Pemrosesan Klasifikasi

No	Pesan Asli	Kategori	Hasil Pra-pemrosesan
1	Ka mau tanya jurusan di bina insani ada apa aja	Jurusan	tanya jurusan bina insani
2	Per semester brp	Biaya	semester brp
3	Bisa nanya info kelas karyawan	Kelas_Karyawan	nanya info kelas karyawan
4	Pendaftaran nya sampe kapan gelombang pertama	Pendaftaran	pendaftaran sampe kapan gelombang pertama
5	Selamat pagi, saya Herning lulusan SMK tahun 2024. Saya mau bertanya apakah masih ada program beasiswa KIP tahun 2025? Jika masih ada waktunya kapan yaa.. Terima kasih 🙏	Beasiswa	selamat pagi herning lulusan smk bertanya program beasiswa kip waktunya kapan terima kasih
..
350	Lokasi kampusnya dimana kak?	Informasi_kampus	lokasi kampus dimana

Sumber: Hasil Penelitian (2025)

Tabel di atas menunjukkan bagaimana sistem prapemrosesan bekerja dalam menyederhanakan teks dari pesan asli yang cenderung tidak terstruktur menjadi bentuk yang lebih bersih dan konsisten.

Setelah proses tersebut, data teks diubah menjadi representasi numerik menggunakan metode TF-IDF (*Term Frequency - Inverse Document Frequency*). Rumus TF-IDF dapat dituliskan sebagai berikut:

$$TF\text{-}IDF(t,d) = TF(t,d) \times \log \frac{N}{DF(t)} \quad (1)$$

Keterangan:

- TF(t,d): jumlah kemunculan term (*t*) pada dokumen (*d*)
- DF(t): jumlah dokumen yang mengandung term (*t*)
- N: total jumlah dokumen

Implementasi TF-IDF dilakukan menggunakan pustaka *scikit-learn* dengan Python, sebagaimana ditunjukkan pada Gambar 3.

```
def clean_text(text):
    text = text.lower()
    text = re.sub(r'\d+', '', text)
    text = text.translate(str.maketrans('', '', string.punctuation))
    text = text.strip()
    stop_words = set(stopwords.words('indonesian'))
    tokens = text.split()
    tokens = [w for w in tokens if w not in stop_words]
    return ' '.join(tokens)

x = TfidfVectorizer().fit_transform(df["clean_komentar"])
y = LabelEncoder().fit_transform(df["label"])
```

Sumber: Hasil Penelitian (2025)

Gambar 5 Pipeline TF-IDF

Dengan demikian, teks ulasan yang sudah dibersihkan dapat direpresentasikan dalam bentuk numerik melalui TF-IDF sebagaimana ditunjukkan pada Tabel 3, sehingga siap digunakan pada tahap pelatihan model *machine learning*.

3.3. Hasil Pelatihan dan Evaluasi Model

Tahap ini memaparkan hasil pengujian berbagai algoritma untuk klasifikasi pesan masuk dari Instagram dan WhatsApp. Empat skenario diuji, yakni:

1. Naive Bayes (NB)
2. Support Vector Machine (SVM)
3. Logistic Regression (LR)

4. SVM + Rule-Based Keyword

Pada pengujian awal, Naive Bayes menghasilkan akurasi sebesar 62%. Model ini cenderung mampu mengenali kategori pesan yang memiliki jumlah data banyak, seperti Pendaftaran dan Biaya, tetapi performanya menurun drastis untuk kategori minoritas, misalnya Loker_Dosen atau Layanan_Kampus. Hal ini menunjukkan bahwa NB kurang optimal untuk dataset dengan distribusi kategori yang tidak seimbang. Gambar 6 adalah hasil menggunakan algoritma *naïve bayes*:

	precision	recall	f1-score	support
beasiswa	0.83	0.71	0.77	7
biaya	0.40	0.67	0.50	12
informasi_kampus	1.00	0.83	0.91	6
jadwal_kuliah	0.78	1.00	0.88	7
jurusan	0.75	0.43	0.55	7
kelas_karyawan	1.00	1.00	1.00	5
lainnya	1.00	0.43	0.60	7
layanan_kampus	0.50	0.25	0.33	8
pendaftaran	0.38	0.50	0.43	10
accuracy			0.62	69
macro avg	0.74	0.65	0.66	69
weighted avg	0.68	0.62	0.62	69

Sumber: Hasil Penelitian (2025)

Gambar 6
Classification Report Algoritma Naïve Bayes

Pengujian dengan *Support Vector Machine (SVM)* menunjukkan akurasi awal 61%, sedikit lebih baik dalam memisahkan kategori pesan, namun tetap kurang stabil untuk kategori dengan jumlah data sedikit.

	precision	recall	f1-score	support
beasiswa	0.75	0.86	0.80	7
biaya	0.80	0.67	0.73	12
informasi_kampus	1.00	1.00	1.00	6
jadwal_kuliah	0.70	1.00	0.82	7
jurusan	0.60	0.43	0.50	7
kelas_karyawan	0.67	0.80	0.73	5
kerjasama	0.60	1.00	0.75	3
layanan_kampus	0.60	0.38	0.46	8
loker_dosen	0.00	0.00	0.00	1
pendaftaran	0.29	0.20	0.24	10
umum	0.00	0.00	0.00	3
accuracy			0.61	69
macro avg	0.55	0.58	0.55	69
weighted avg	0.62	0.61	0.60	69

Sumber: Hasil Penelitian (2025)

Gambar 7
Classification Report Algoritma Support Vector Machine

Selanjutnya, Logistic Regression (LR) diuji dengan hasil akurasi 65%, menunjukkan performa serupa dengan SVM, namun masih kurang memadai pada kategori minoritas.

	precision	recall	f1-score	support
beasiswa	0.86	0.86	0.86	7
biaya	0.75	0.50	0.60	12
informasi_kampus	1.00	1.00	1.00	6
jadwal_kuliah	0.70	1.00	0.82	7
jurusan	0.60	0.43	0.50	7
kelas_karyawan	1.00	1.00	1.00	5
lainnya	0.44	0.57	0.50	7
layanan_kampus	0.42	0.62	0.50	8
pendaftaran	0.43	0.30	0.35	10
accuracy			0.65	69
macro avg	0.69	0.70	0.68	69
weighted avg	0.66	0.65	0.64	69

Sumber: Hasil Penelitian (2025)

Gambar 8

Classification Report Algoritma Logistic Regression

Untuk mengatasi ketidakseimbangan data antar kategori, diterapkan pendekatan gabungan SVM dengan rule-based keyword. Metode ini memanfaatkan daftar kata kunci khusus untuk masing-masing kategori, sehingga pesan yang tidak dapat dikenali dengan baik oleh model SVM tetap dapat diklasifikasikan secara akurat. Contohnya, kata "beasiswa" secara otomatis diarahkan ke kategori Beasiswa, sementara kata "kelas karyawan" diarahkan ke kategori Kelas_Karyawan. Pesan yang tidak terdeteksi melalui keyword kemudian diproses menggunakan model SVM.

```
keyword_rules = {
    'beasiswa': ['kip', 'beasiswa'],
    'biaya': ['biaya', 'bayar', 'pembayaran', 'ukt'],
    'loker_dosen': ['lamar dosen', 'loker', 'lowongan dosen'],
    'kelas_karyawan': ['kelas karyawan', 'karyawan'],
    'pendaftaran': ['pendaftaran', 'daftar', 'buka gel'],
    'jadwal_kuliah': ['jadwal', 'waktu kuliah'],
}
```

Sumber: Hasil Penelitian (2025)

Gambar 9

Menerapkan keyword rules

Pendekatan hybrid ini meningkatkan akurasi hingga 70%, sekaligus memperbaiki precision dan recall untuk kategori minoritas.

beasiswa	0.78	1.00	0.88	7
biaya	0.77	0.83	0.80	12
informasi_kampus	1.00	1.00	1.00	6
jadwal_kuliah	0.78	1.00	0.88	7
jurusan	0.60	0.43	0.50	7
kelas_karyawan	0.67	0.80	0.73	5
kerjasama	0.75	1.00	0.86	3
layanan_kampus	1.00	0.38	0.55	8
loker_dosen	0.25	1.00	0.40	1
pendaftaran	0.44	0.40	0.42	10
umum	0.00	0.00	0.00	3
accuracy			0.70	69
macro avg	0.64	0.71	0.64	69
weighted avg	0.70	0.70	0.67	69

Sumber: Hasil Penelitian (2025)

Gambar 10

Classification Report Hybrid SVM + Rule Based

Dengan metode ini, sistem mampu mengenali seluruh kategori pesan secara lebih merata dan konsisten, sehingga hasil klasifikasi lebih dapat diandalkan untuk keperluan pengolahan pesan rutin di kampus.

Tabel 4.
Performa Algoritma Terbaik

Model	Akurasi	Precision (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)
<i>Naive Bayes</i>	61%	74%	65%	66%
<i>Logistic Regression</i>	65%	69%	70%	68%
<i>SVM</i>	61%	55%	58%	55%
<i>SVM + Rule- Based</i>	70%	64%	71%	64%

Sumber: Hasil Penelitian (2025)

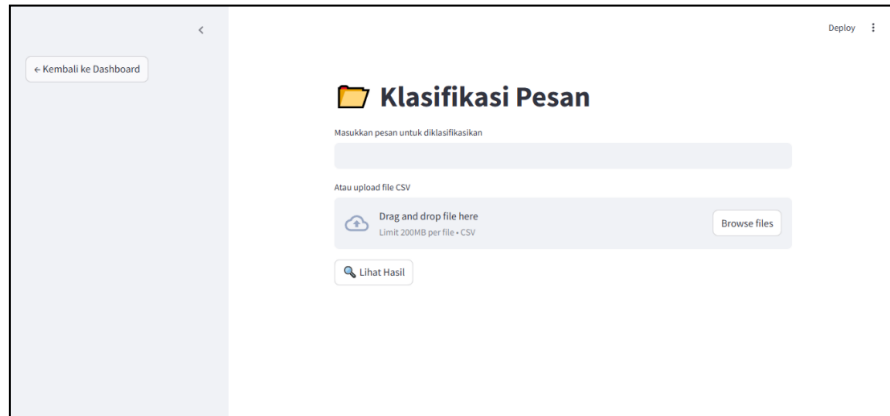
Hasil evaluasi menunjukkan adanya perbedaan performa yang cukup signifikan antar model klasifikasi yang digunakan. Naive Bayes menghasilkan tingkat akurasi sebesar 61% dengan precision macro average sebesar 74%, menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam menghasilkan prediksi yang tepat pada kelas tertentu, namun masih terbatas dalam mengenali keseluruhan distribusi kategori secara merata. Support Vector Machine (SVM) menunjukkan performa yang relatif lebih rendah dengan accuracy sebesar 61%, precision 55%, recall 58%, dan F1-score 55%, yang mengindikasikan bahwa model masih mengalami kesulitan dalam menangani distribusi data yang tidak seimbang antar kategori.

Logistic Regression menghasilkan performa yang lebih stabil dibandingkan kedua model sebelumnya dengan accuracy sebesar 65%, recall sebesar 70%, serta F1-score sebesar 68%. Hasil ini menunjukkan bahwa Logistic Regression memiliki kemampuan generalisasi yang lebih baik dalam mengenali berbagai kategori pesan dibandingkan model machine learning lainnya.

Pendekatan Hybrid SVM + Rule-Based menghasilkan performa terbaik dengan accuracy mencapai 70% dan recall tertinggi sebesar 71%. Peningkatan ini menunjukkan bahwa integrasi keyword-based routing mampu membantu sistem mengenali kategori minoritas yang sebelumnya sulit dipelajari oleh model machine learning murni. Pendekatan hybrid memungkinkan pesan dengan pola tertentu untuk dikategorikan lebih awal melalui aturan berbasis kata kunci, sementara pesan yang lebih kompleks tetap diproses menggunakan model SVM. Temuan ini menunjukkan bahwa kombinasi machine learning dan rule-based lebih efektif untuk menangani klasifikasi pesan media sosial yang memiliki karakteristik data tidak seimbang dan variasi bahasa yang tinggi.

3.4. Implementasi Dashboard Streamlit

Sistem yang dibangun diimplementasikan dalam bentuk dashboard interaktif menggunakan Streamlit yang terlihat pada Gambar 9. Dashboard ini dibuat agar pengguna bisa dengan mudah melihat hasil klasifikasi pesan tanpa harus memahami proses teknis di baliknya. Melalui tampilan yang sederhana, seluruh proses analisis mulai dari input data hingga visualisasi hasil dapat dilakukan secara praktis. *Framework* Streamlit dipilih karena mampu menampilkan hasil prediksi secara langsung setelah model dijalankan, serta mendukung interaksi pengguna dengan sistem tanpa perlu konfigurasi tambahan yang kompleks. Dashboard ini memiliki beberapa fitur utama, antara lain:

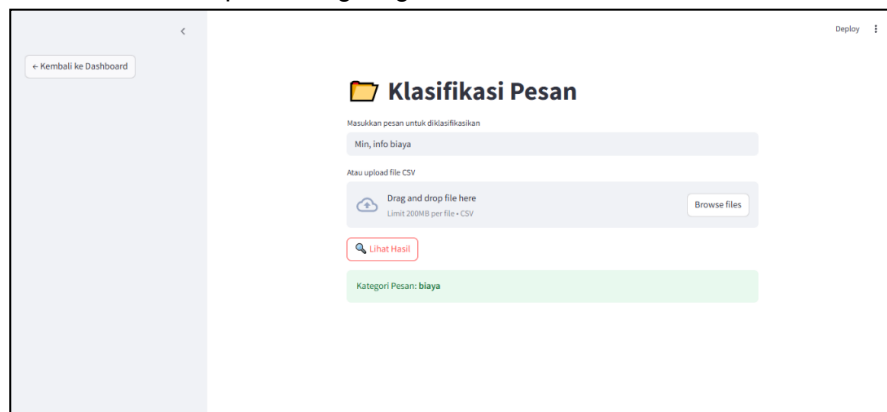


Sumber: Hasil Penelitian (2025)

Gambar 11. Tampilan Halaman Ujicoba Klasifikasi Pesan

a. Input Ulasan Baru

Pengguna dapat memasukkan data pesan baru melalui file CSV dengan kolom minimal “pesan” dan “platform” (Instagram/WhatsApp). Sistem memungkinkan analisis satu per satu pesan maupun banyak data sekaligus. Setiap pesan teks secara otomatis akan melewati pra-pemrosesan dan vektorisasi TF-IDF, sementara pesan yang berbentuk non-teks (misal hanya emoji atau angka) akan diproses menggunakan rule-based keyword detection. Hasil klasifikasi kemudian ditampilkan langsung di dashboard.

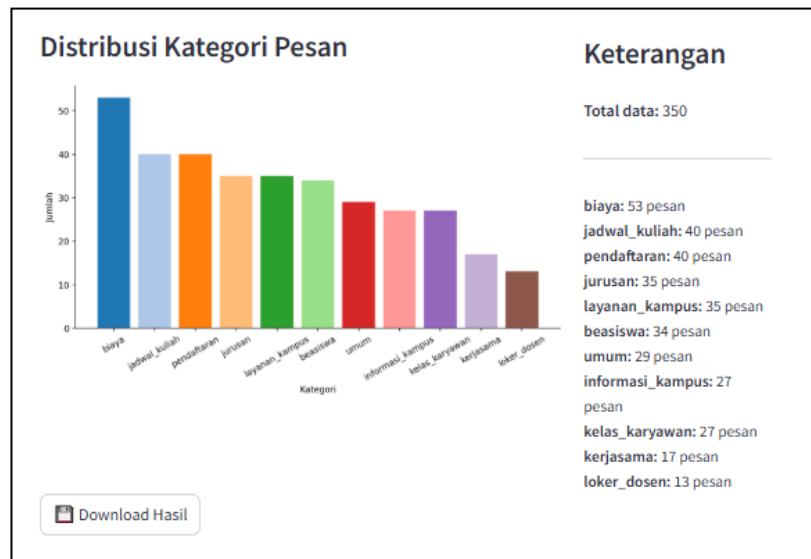


Sumber: Hasil Penelitian (2025)

Gambar 12. Tampilan Hasil Ujicoba Klasifikasi Pesan

b. Visualisasi Kategori Pesan dan Ringkasan Kategori

Dashboard menampilkan distribusi kategori pesan dalam bentuk grafik batang dan tabel. Visualisasi ini mempermudah admin kampus untuk mengetahui topik pesan yang paling sering ditanyakan, misalnya jumlah pertanyaan mengenai pendaftaran, biaya, jurusan, beasiswa, kelas karyawan, atau informasi kampus. Tabel hasil klasifikasi juga dapat diunduh dalam format CSV untuk dokumentasi atau laporan.



Sumber: Hasil Penelitian (2025)

Gambar 13. Tampilan Hasil Ujicoba Klasifikasi Pesan

c. Export Laporan

Sistem juga dilengkapi dengan fitur *export laporan* yang memungkinkan pengguna mengunduh hasil analisis dalam format CSV. Melalui fitur ini, pengguna dapat menyimpan rekapitulasi hasil klasifikasi pesan untuk keperluan dokumentasi dan evaluasi lanjutan. Selain itu, laporan yang dihasilkan dapat digunakan sebagai bahan pertimbangan dalam rapat evaluasi layanan maupun publikasi internal kampus. Dengan adanya fungsi ekspor ini, sistem tidak hanya berperan sebagai alat analisis, tetapi juga sebagai media pendukung pengambilan keputusan berbasis data yang lebih terukur dan berkelanjutan sebagaimana digambarkan pada Gambar 5

Seluruh 350 pesan yang dikumpulkan dari media sosial diproses secara otomatis oleh sistem klasifikasi. Pesan teks dianalisis menggunakan algoritma SVM yang dipadukan dengan pendekatan rule-based, sementara pesan non-teks dikategorikan berdasarkan kata kunci atau pola khusus. Dataset ini berfungsi sebagai data latih bagi model, sehingga sistem mampu mempelajari pola setiap kategori pesan.

Hasil pengolahan ditampilkan pada dashboard Streamlit, yang memungkinkan admin untuk:

1. Mengunggah pesan baru dan melihat klasifikasi secara real-time,
2. Memantau distribusi kategori pesan,
3. Menangani pesan dari berbagai format dan platform secara konsisten tanpa harus membaca satu per satu.

Dengan demikian, sistem tidak hanya memproses pesan secara otomatis, tetapi juga meningkatkan **efisiensi, konsistensi, dan skalabilitas** pengelolaan komunikasi di media sosial.

4. Kesimpulan

Penelitian ini dapat merancang sistem klasifikasi pesan media sosial Universitas Bina Insani dengan pendekatan hybrid, memadukan *machine learning*. Pesan teks diproses menggunakan *Support Vector Machine* (SVM) yang dipadukan *rule-based* untuk menangani pola pesan dan variasi bahasa, sementara pesan non-teks dikategorikan langsung berdasarkan kata kunci atau pola tertentu. Hasil implementasi menunjukkan bahwa model yang dibangun mampu mengelompokkan pesan ke dalam kategori yang telah ditentukan secara konsisten, sekaligus

memungkinkan integrasi data baru secara dinamis. Implementasi sistem dalam bentuk dashboard interaktif berbasis Streamlit mendukung visualisasi distribusi kategori pesan, ringkasan klasifikasi, dan input pesan baru. Dengan demikian, sistem ini mengotomatisasi proses klasifikasi, sehingga admin dapat mengelola berapa banyak pesan dari calon mahasiswa secara lebih efisien tanpa harus membaca setiap pesan satu per satu. Namun, keterbatasan penelitian ini berupa jumlah dataset yang relatif kecil dan aturan *rule-based* yang sederhana menunjukkan perlunya pengembangan lebih lanjut, seperti pengujian pada dataset yang lebih besar, integrasi multi-platform, serta peningkatan kompleksitas aturan untuk pesan non-teks. Secara keseluruhan, penelitian ini menegaskan bahwa penerapan klasifikasi pesan berbasis *machine learning* dapat memberikan kontribusi nyata dalam mendukung pengelolaan komunikasi digital di lingkungan perguruan tinggi. Sistem yang dikembangkan tidak hanya berfungsi sebagai alat klasifikasi, tetapi juga sebagai sarana untuk memahami pola dan kebutuhan informasi dari calon mahasiswa secara lebih terstruktur. Harapannya, penelitian ini dapat mendorong mencakup lebih luas terhadap teknologi kecerdasan buatan dalam pengelolaan komunikasi institusi, sekaligus menjadi dasar pengembangan sistem klasifikasi pesan yang lebih baik untuk kebutuhan Perguruan Tinggi.

Referensi

- [1] L. E. Zuniananta, "Penggunaan Media Sosial sebagai Media Komunikasi Informasi Di Perpustakaan," *J. Ilmu Perpust.*, vol. 10, no. 4, pp. 37–42, 2021.
- [2] M. Raihan, S. Putra Pamungkas, M. Nurul Huda, D. A. Fauzan, A. Hilal Itsna, and M. Al-Hijri, "Sistem Klasifikasi Otomatis Dengan Konsep Machine Learning As A Service (MLaaS) Pada Kasus Pesan Berindikasi Cyberbullying," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 4, no. 3, pp. 252–261, 2022.
- [3] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022.
- [4] G. A. Syafarina and Z. Zaenuddin, "Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi," *Metik J.*, vol. 7, no. 2, pp. 121–125, 2023, doi: 10.47002/metik.v7i2.608.
- [5] M. S. Haris, A. N. Khudori, and W. T. Kusuma, "Perbandingan Metode Supervised Machine Learning untuk Prediksi Prevalensi Stunting di Provinsi Jawa Timur," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 7, p. 1571, 2022, doi: 10.25126/jtiik.2022976744.
- [6] R. Nazar, "Implementasi Pemrograman Python Menggunakan Google Colab," *J. Inform. dan Komput.*, vol. 15, no. 1, pp. 50–56, 2024.
- [7] T. Tan, H. Sama, G. Wijaya, and O. E. Aboagye, "Studi Perbandingan Deteksi Intrusi Jaringan Menggunakan Machine Learning: (Metode SVM dan ANN)," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 152–164, 2023, doi: 10.34010/jati.v13i2.10484.
- [8] A. Supratman, B. Indarmawan Nugroho, and R. Dwi Kurniawan, "Penerapan Metode Rule-Based System untuk Menentukan Jenis Tanaman Pertanian Berdasarkan Ketinggian dan Curah Hujan," *Innov. J. Soc. Sci. Res.*, vol. 4, pp. 7879–7890, 2024.
- [9] H. Leidiyana, T. Misriati, and R. Aryanti, "Klasifikasi Sentimen Terhadap Kebijakan Tapera Menggunakan Komparasi Machine Learning dan SMOTE," *J. Komtika (Komputasi dan Inform.)*, vol. 8, no. 2, pp. 125–135, 2024, doi: 10.31603/komtika.v8i2.12595.
- [10] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [11] A. Putranto, N. L. Azizah, and A. I. Ratna Ika, "Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit," *J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 4, no. 2, pp. 442–452, 2023, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/heart+disease>
- [12] M. Hariz Izzuddin and M. Daniel Krisna Halim Putra, "Implementasi Metodologi Agile dalam Pengembangan Platform Kursus Online pada Platform Eduskill," *J. Educ.*, vol. 06, no. 04, pp. 20158–20166, 2024.
- [13] F. D. Pramakrisna, F. D. Adhinata, and N. A. F. Tanjung, "Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression," *Teknika*, vol. 11, no. 2, pp. 90–97, 2022, doi: 10.34148/teknika.v11i2.466.