

**ANALISIS SENTIMEN PENGGUNAAN ARTIS KOREA
SEBAGAI BRAND AMBASSADOR PRODUK LOKAL
INDONESIA MENGGUNAKAN NAÏVE BAYES, SVM, DAN
RANDOM FOREST**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Michele Stepanie
00000055477

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**ANALISIS SENTIMEN PENGGUNAAN ARTIS KOREA
SEBAGAI BRAND AMBASSADOR PRODUK LOKAL
INDONESIA MENGGUNAKAN NAÏVE BAYES, SVM, DAN
RANDOM FOREST**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer

Michele Stepanie

00000055477

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Michele Stepanie

Nomor Induk Mahasiswa 00000055477

Program Studi : Sistem Informasi

Skripsi dengan judul:

ANALISIS SENTIMEN PENGGUNAAN ARTIS KOREA SEBAGAI BRAND
AMBASSADOR PRODUK LOKAL INDONESIA MENGGUNAKAN NAÏVE
BAYES, SVM, DAN RANDOM FOREST

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 10 Juni 2025



Michele Stepanie

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN PENGGUNAAN ARTIS KOREA SEBAGAI BRAND AMBASSADOR
PRODUK LOKAL INDONESIA MENGGUNAKAN NAÏVE BAYES, SVM, DAN RANDOM FOREST

Oleh

Nama : Michele Stepanie
NIM : 00000055477
Program Studi : Sistem Informasi
Fakultas : Teknik & Informatika

Telah diujikan pada hari Rabu, 12 Juni 2025

Pukul 13.00 s.d 15.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang

Penguji

Wella, S.Kom., M.MSI
0305119101

Johan Setiawan, S.Kom., M.M., MBA
327106402

Pembimbing

Suryasari, S.Kom., M.T
0323088301

Ketua Program Studi Sistem Informasi

Ririn Ikana Desanti, S.Kom., M.Kom.

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Michele Stepanie

NIM 00000055477

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : ANALISIS SENTIMEN PENGGUNAAN ARTIS KOREA
SEBAGAI BRAND AMBASSADOR PRODUK LOKAL
INDONESIA MENGGUNAKAN NAÏVE BAYES, SVM, DAN
RANDOM FOREST

Menyatakan dengan sesungguhnya bahwa saya bersedia* (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☒ Lainnya, pilih salah satu:
 - ☒ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 12 Juni 2025



Michele Stepanie

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

(Kata Pengantar dapat dikembangkan dan harus meliputi ucapan rasa syukur, tujuan pembuatan tugas akhir, ucapan terima kasih, dan harapan pada hasil Tugas Akhir ini.)

Mengucapkan terima kasih kepada:

1. Bapak Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi, Universitas Multimedia Nusantara.
4. Ibu Suryasari, S.Kom., M.T., selaku Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi sehingga terselesainya tugas akhir ini.
5. Orang tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman-teman saya yang sudah memberikan dukungan dalam menyelesaikan penelitian ini.
7. Kepada NCT Dream dan lagu-lagunya yang sudah memberikan inspirasi kepada penulis sehingga penelitian ini bisa diselesaikan oleh penulis

Semoga karya ilmiah ini dapat bermanfaat dan digunakan sebaik-baiknya sebagai sumber informasi ataupun sumber inspirasi, bagi para pembaca.

Tangerang, 12 Juni 2025



(Michele Stepanie)

ANALISIS SENTIMEN PENGGUNAAN ARTIS KOREA SEBAGAI BRAND AMBASSADOR PRODUK LOKAL INDONESIA MENGGUNAKAN NAÏVE BAYES, SVM, DAN RANDOM FOREST

Michele Stepanie

ABSTRAK

Budaya populer Korea Selatan yang dikenal sebagai *Korean Wave* atau *Hallyu* telah menyebar secara global dan memperoleh perhatian besar di Indonesia. Dampaknya terlihat dari meningkatnya penggunaan figur publik asal Korea sebagai duta merek untuk produk-produk lokal. Strategi promosi ini dinilai mampu menarik minat konsumen Indonesia, terutama pada *platform* media sosial seperti X. Namun, fenomena tersebut menimbulkan reaksi beragam, termasuk tanggapan positif, netral, maupun negatif dari masyarakat.

Oleh sebab itu, penting dilakukan analisis sentimen untuk memahami pandangan publik terhadap tren tersebut. Penelitian ini menggunakan tiga algoritma pembelajaran mesin, yaitu *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest*, dalam mengelompokkan komentar dari pengguna media sosial X yang dikumpulkan selama periode 2021 hingga 2025. Tahapan analisis dilakukan berdasarkan metode CRISP-DM yang mencakup pemahaman konteks, eksplorasi data, pembersihan data, pembuatan model, evaluasi, dan penyampaian hasil akhir. Data dianalisis menggunakan pendekatan text mining dan diklasifikasikan ke dalam tiga jenis sentimen: positif, negatif, dan netral. Evaluasi performa dilakukan dengan mengukur akurasi, presisi, *recall*, serta nilai *F1-Score*.

Hasil pengujian menunjukkan bahwa sebagian besar komentar masyarakat bersifat positif terhadap pemanfaatan artis Korea sebagai duta produk lokal. Dari ketiga algoritma yang diuji, *Random Forest* menghasilkan tingkat akurasi tertinggi dalam mengklasifikasikan sentimen. Penemuan ini memberikan masukan strategis bagi pemilik merek lokal untuk mempertimbangkan kolaborasi dengan figur internasional. Selain itu, studi ini memberikan kontribusi dalam penerapan algoritma klasifikasi untuk analisis opini dalam bidang pemasaran digital.

Kata kunci: Analisis Sentimen, Duta Merek, *Korean Wave*, *Naïve Bayes*, *Random Forest*, *SVM*

SENTIMENT ANALYSIS OF THE USE OF KOREAN ARTISTS AS BRAND AMBASSADORS OF LOCAL INDONESIAN PRODUCTS USING THE NAÏVE BAYES, SVM AND RANDOM FOREST

Michele Stepanie

ABSTRACT (English)

South Korea's popular culture, widely known as the Korean Wave or Hallyu, has spread globally and garnered significant attention in Indonesia. One of its impacts is the growing use of Korean public figures as brand ambassadors for local products. This promotional strategy is considered effective in attracting the interest of Indonesian consumers, especially on social media platforms such as X. However, this phenomenon has sparked various responses from the public, ranging from positive and neutral to negative sentiments.

Therefore, sentiment analysis is essential to understand public perception regarding this trend. This study employs three machine learning algorithms—Naïve Bayes, Support Vector Machine (SVM), and Random Forest—to classify user comments from platform X collected between 2021 and 2025. The analysis follows the CRISP-DM methodology, which includes business understanding, data exploration, data cleaning, modeling, evaluation, and result presentation. The data is processed using a text mining approach and categorized into three sentiment classes: positive, negative, and neutral. Model performance is evaluated using accuracy, precision, recall, and the F1-Score.

The results indicate that the majority of public comments reflect positive sentiment toward the use of Korean celebrities as brand ambassadors for local products. Among the three tested algorithms, Random Forest achieved the highest accuracy in sentiment classification. These findings offer strategic insights for local brand owners considering collaborations with international figures. Moreover, this research contributes to the application of classification algorithms for opinion analysis in the context of digital marketing.

Keywords: *Brand Ambassador, Korean Wave, Naïve Bayes, Random Forest, Sentiment Analysis, SVM*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	2
HALAMAN PENGESAHAN.....	4
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	5
KATA PENGANTAR.....	6
ABSTRAK.....	7
ABSTRACT (<i>English</i>).....	8
DAFTAR ISI	9
DAFTAR TABEL.....	11
DAFTAR RUMUS	14
DAFTAR LAMPIRAN	15
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	5
1.4 Tujuan dan Manfaat Penelitian	5
1.4.1 Tujuan Penelitian	5
1.4.2 Manfaat Penelitian	5
1.5 Sistematika Penulisan.....	6
BAB II LANDASAN TEORI.....	7
2.1 Penelitian Terdahulu	7
2.2 Teori Penelitian.....	12
2.2.2 X.....	13
2.2.3 Analisis Sentimen	13
2.3.4 Framework dan algoritma penelitian	14
2.3.1 CRISP-DM	14
2.4 Tools dan Software Penelitian	21
2.4.2 Google Colab.....	22
BAB III METODOLOGI PENELITIAN.....	23
3.1 Gambaran Umum Objek Penelitian	23
3.2 Metode Penelitian.....	23
3.3 Teknik Pengumpulan Data	25
3.4 Teknik Analisis Data	25
BAB IV.....	29
4.1 Business Understanding	29
4.2 Data Understanding	29

4.3	Data Preparation	32
4.3.1	Case Folding dan Cleaning Data.....	35
4.3.2	Tokenization dan Lemmatization	36
4.3.3	Stopword Removing	37
4.3.4	Stemming.....	38
4.3.5	Labeling.....	40
4.3.6	Splitting Data.....	42
4.3.7	TF-IDF Vectorization	43
4.3.8	SMOTE	43
4.4	Modelling.....	45
4.4.1	Naive Bayes.....	45
4.4.2	Support Vector Machine (SVM).....	46
4.4.3	Random Forest.....	46
4.5	Evaluation.....	47
4.5.1	Naïve Bayes	47
4.5.2	Support Vector Machine (SVM).....	51
4.5.3	Random Forest.....	55
4.5.4	Distribusi Jumlah Sentimen.....	58
4.5.5	5 Kata Teratas.....	59
4.5.6	10 Kata yang sering muncul	59
4.5.7	World Cloud Analysis.....	60
4.5.8	Perbandingan Akurasi Model.....	61
4.6	Hasil Penelitian.....	62
4.7	Pembahasan Hasil Penelitian	63
BAB V SIMPULAN DAN SARAN.....		66
5.1	Simpulan	66
5.2	Saran	66
DAFTAR PUSTAKA		68
LAMPIRAN		75

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	7
Tabel 2. 2 Confusion Matrix	18
Tabel 4. 1 Variabel pada dataset	31
Tabel 4. 2 Hasil case folding dan cleaning data	36
Tabel 4. 3 Hasil Tokenization dan Leminization	37
Tabel 4. 4 Hasil Stopwords Removing	38
Tabel 4. 5 Hasil Stemming	39
Tabel 4. 6 Hasil Labeling	41
Tabel 4. 7 Hasil Performa Model Naïve Bayes, SVM dan Random Forest	62



DAFTAR GAMBAR

Gambar 1. 1 Jumlah Antusiasme Terhadap Budaya Korea Selatan	1
Gambar 1. 2 Jumlah Penggemar K-pop Terbanyak di Twitter	2
Gambar 1. 3 Artis Korea Yang Menjadi Brand Ambassador Produk Lokal Indonesia.....	2
Gambar 2. 1 Alur Metode CRISP-DM	15
Gambar 3. 1 Tahapan data preprocessing	27
Gambar 4. 1 Memasukkan auth token X dalam melakukan scraping data.....	30
Gambar 4. 2 Scarping Data	30
Gambar 4. 3 Data yang berhasil dikumpulkan.....	30
Gambar 4. 4 Membaca File Excel	32
Gambar 4. 5 Info data tweet artis korea sebagai BA produk lokal	33
Gambar 4. 6 Memeriksa missing value	34
Gambar 4. 7 Kode tahap case folding dan cleaning	35
Gambar 4. 8 Kode Tahap Tokenization dan Leminization.....	36
Gambar 4. 9 Kode Tahapan Stopword Removing	37
Gambar 4. 10 Kode Tahapan Stemming.....	39
Gambar 4. 11 Kode Tahapan Labeling.....	40
Gambar 4. 12 Contoh sample acak 10 baris DataFrame	42
Gambar 4. 13 Kode Tahapan Splitting Data	42
Gambar 4. 14 Kode Tahapan TF-IDF Vectorization.....	43
Gambar 4. 15 Kode Tahapan SMOTE	43
Gambar 4. 16 Distribusi Jumlah Sentimen Sebelum Dilakukan SMOTE	44
Gambar 4. 17 Distribusi Jumlah Sentimen Setelah Dilakukan SMOTE	45
Gambar 4. 18 Latih Model Naïve Bayes	45
Gambar 4. 19 Latih Model SVM.....	46
Gambar 4. 20 Latih Model Random Forest	46
Gambar 4. 21 Evaluasi Naïve Bayes Sebelum SMOTE.....	47
Gambar 4. 22 Evaluasi Naïve Bayes Setelah SMOTE	48
Gambar 4. 23 Confusion Matrix Naïve Bayes Sebelum SMOTE.....	49
Gambar 4. 24 Confusion Matrix Naïve Bayes Setelah SMOTE	50
Gambar 4. 25 Confusion Matrix Naïve Bayes Setelah SMOTE.....	51
Gambar 4. 26 Evaluasi SVM setelah SMOTE.....	52
Gambar 4. 27 Confusion Matrix SVM Sebelum SMOTE.....	53
Gambar 4. 28 Confusion Matrix SVM setelah SMOTE.....	54
Gambar 4. 29 Evaluasi Random Forest Sebelum SMOTE.....	55
Gambar 4. 30 Evaluasi Random Forest Setelah SMOTE	55
Gambar 4. 31 Confusion Matrix Random Forest Sebelum SMOTE	56
Gambar 4. 32 Confusion Matrix Random Forest Setelah SMOTE.....	57
Gambar 4. 33 Piechart Distribusi Jumlah Sentimen.....	58
Gambar 4. 34 Barchart kata teratas	59
Gambar 4. 35 Barchart 10 Kata Yang Sering Muncul	59
Gambar 4. 36 World Cloud Analysis	60

Gambar 4. 37 Perbandingan Akurasi Model Sebelum SMOTE	61
Gambar 4. 38 Perbandingan Akurasi Model Setelah SMOTE	62

DAFTAR RUMUS

Rumus 2. 1Perhitungan Naïve Bayes Classifier (NBC)	16
Rumus 2. 2Perhitungan Accuracy	18
Rumus 2. 3Perhitungan Recall	18
Rumus 2. 4Perhitungan Precision	19
Rumus 2. 5Perhitungan F1-Score.....	19

DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report.....	75
Lampiran B Form Konsultasi Bimbingan	82

