

ANALISIS SENTIMEN ULASAN APLIKASI *COFFEE SHOP* MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*

SITA AFREA SYAMSIYAH



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Analisis Sentimen Ulasan Aplikasi *Coffee Shop* Menggunakan Algoritma *Support Vector Machine*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2026

Sita Afrea Syamsiyah
G14190082



ABSTRAK

SITA AFREA SYAMSIYAH. Analisis Sentimen Ulasan Aplikasi *Coffee Shop* Menggunakan Algoritma *Support Vector Machine*. Dibimbing oleh PIKA SILVIANTI dan SEPTIAN RAHARDIANTORO.

Analisis sentimen adalah metode untuk mengklasifikasikan opini dan pola menarik dari kumpulan data berbentuk teks. Analisis sentimen dapat diterapkan dalam berbagai hal, salah satunya pada aplikasi *coffee shop*. Seiring perkembangan zaman, banyak perusahaan di berbagai bidang yang menggunakan aplikasi *online* untuk memudahkan aktivitas transaksinya. Permintaan akan layanan berkualitas tinggi meningkat seiring dengan meningkatnya popularitas aplikasi. Setiap pengguna aplikasi ini dapat memberikan ulasan mengenai aplikasi ini di situs *Google Play Store*. Informasi data tekstual dari *review* pengguna pada aplikasi di *Google Play Store* dapat digunakan sebagai bahan evaluasi dengan menggunakan bantuan analisis sentimen dengan membagi data menjadi kelas positif, netral dan negatif untuk mendapatkan nilai akurasi yang optimal. Oleh karena itu, penelitian ini ingin melihat pola ulasan data sentimen dan mengidentifikasi perbedaan performa algoritma *Support Vector Machine* dengan menggunakan kernel *linear*, kernel *polynomial*, dan kernel *Radial Basis Function* (RBF) pada tiga fitur n-gram (unigram, bigram, dan trigram). Pemodelan SVM menggunakan kernel *linear* pada fitur unigram menghasilkan performa terbaik dalam mengklasifikasikan sentimen dengan nilai akurasi dan *macro average F1-score* yang didapat sebesar 77% dan 53%. Secara keseluruhan, kernel *linear* dan RBF pada ketiga fitur n-gram memberikan hasil klasifikasi terbaik. Sebaliknya, kernel *polynomial* memberikan hasil klasifikasi terendah. Fitur unigram memberikan hasil klasifikasi terbaik dan fitur trigram memberikan hasil klasifikasi terendah untuk ketiga kernel.

Kata kunci: Analisis sentimen, Kernel, n-gram, *Review*, *Support Vector Machine*.



ABSTRACT

SITA AFREA SYAMSIYAH. Sentiment Analysis of Coffee Shop Application Reviews Using the Support Vector Machine Algorithm. Supervised by PIKA SILVIANTI and SEPTIAN RAHARDIANTORO.

Sentiment analysis is a method for classifying opinions and interesting patterns from textual data sets. Sentiment analysis can be applied in various ways, one of which is in coffee shop applications. As time goes by, many companies in various fields use online applications to facilitate their transaction activities. The demand for high-quality services increases along with the increasing popularity of applications. Every user of this application can provide a review of this application on the Google Play Store site. Textual data information from user reviews on applications in the Google Play Store can be used as evaluation material using sentiment analysis by dividing the data into positive, neutral and negative classes to obtain optimal accuracy values. Therefore, this study aims to examine the pattern of sentiment data reviews and identify differences in the performance of the Support Vector Machine algorithm using linear kernels, polynomial kernels, and Radial Basis Function (RBF) kernels on three n-gram features (unigrams, bigrams, and trigrams). SVM modeling using linear kernels on unigram features produces the best performance in classifying sentiment with accuracy values and macro average F1-scores obtained of 77% and 53%. Overall, linear kernels and RBF on the three n-gram features provide the best classification results. In contrast, the polynomial kernel provided the lowest classification results. The unigram feature provided the best classification results, and the trigram feature provided the lowest classification results for all three kernels.

Keywords: Kernel, n-gram, Review, Sentiment Analysis, Support Vector Machine



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

ANALISIS SENTIMEN ULASAN APLIKASI *COFFEE SHOP* MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*

SITA AFREA SYAMSIYAH

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Sarjana Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**

@Hak cipta milik IPB University

IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
1. Aulia Rizki Firdawanti, S.Stat., M.Si.



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University



Judul Skripsi : Analisis Sentimen Ulasan Aplikasi *Coffee Shop* Menggunakan
Algoritma *Support Vector Machine*

Nama : Sita Afrea Syamsiyah

NIM : G14190082

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Pika Silvianti, S.Si., M.Si

Pembimbing 2:

Dr. Septian Rahardiantoro, S.Stat., M.Si

Diketahui oleh

Ketua Program Studi:

Dr. Bagus Sartono, S.Si., M.Si.

NIP 19780411 200501 1 002

Tanggal Ujian:
24 Juli 2026

Tanggal Lulus:

PRAKATA

Puji syukur kehadiran Allah SWT atas segala karunia-Nya sehingga karya ilmiah dengan judul " Analisis Sentimen Ulasan Aplikasi *Coffee Shop* Menggunakan Algoritma *Support Vector Machine* " bisa diselesaikan. Karya ilmiah ini tentunya tidak lepas dari dukungan dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan yang berharga ini penulis ini menyampaikan ucapan terima kasih kepada:

1. Ibu Pika Silvianti, S.Si., M.Si dan Bapak Dr. Septian Rahardiantoro, S.Stat., M.Si selaku dosen pembimbing yang telah senantiasa membimbing, memberi saran, dan masukan dalam penulisan karya ilmiah ini.
2. Ibu Aulia Rizki Firdawanti, S.Stat., M.Si selaku dosen penguji yang telah memberikan kritik dan saran dalam penulisan karya ilmiah ini.
3. Seluruh dosen Program Studi Statistika dan Sains Data IPB yang telah memberikan ilmu dan pembelajaran yang sangat bermanfaat.
4. Seluruh staf TU Program Studi Statistika dan Sains Data IPB yang telah banyak membantu dalam proses administrasi perkuliahan.
5. Bapak Syamsul Falah, Ibu Nurhasanah, Epha, Aliceta, dan Faris selaku orang tua dan keluarga yang senantiasa memberikan do'a, dukungan, nasihat, dan kasih sayang.
6. Teman-teman Statistika IPB angkatan 56 yang selalu memberikan dukungan dan semangat.

Penulis menyadari bahwa dalam penulisan karya ilmiah masih terdapat kekurangan. Namun, penulis berharap semoga karya ilmiah ini bisa memberikan manfaat untuk berbagai pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2026

Sita Afrea Syamsiyah



DAFTAR ISI

DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	2
TINJAUAN PUSTAKA	3
2.1 <i>Support Vector Machine</i> (SVM)	3
2.2 Analisis Sentimen	4
2.3 <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF)	5
2.4 Fungsi Kernel	5
2.5 <i>Text Pre-processing</i>	6
2.6 <i>Confusion Matrix</i>	7
2.7 <i>K-Fold Cross Validation</i>	8
2.8 <i>Synthetic Minority Oversampling Technique</i> (SMOTE)	9
METODE	11
3.1 Data	11
3.2 Prosedur Analisis Data	11
3.3 Eksplorasi Data	14
3.4 Tokenisasi dan Pembobotan TF-IDF	14
3.5 Pembagian data	15
3.6 Penanganan data tidak seimbang	15
3.7 Pemodelan Klasifikasi dan Evaluasi Model	15
HASIL DAN PEMBAHASAN	16
4.1 Pra-proses Data	16
4.2 Fitur N-Gram	18
4.3 Eksplorasi Data	18
4.4 Tokenisasi dan Pembobotan TF-IDF	21
4.5 <i>Synthetic Minority Over-Sampling Technique</i> (SMOTE)	22
4.6 Pemodelan Klasifikasi	23
4.7 Evaluasi Model	23
SIMPULAN DAN SARAN	26
5.1 Simpulan	26
5.2 Saran	26
DAFTAR PUSTAKA	27
LAMPIRAN	30
RIWAYAT HIDUP	32

DAFTAR TABEL

1	Rumus fungsi kernel	6
2	<i>Confusion matrix</i> untuk klasifikasi non-biner	8
3	Perhitungan akurasi, <i>precision</i> , <i>recall</i> , <i>F1-score</i> , <i>Macro Average F1-Score</i> , dan <i>CV-Acc</i>	8
4	Peubah hasil penarikan data yang digunakan	11
5	Hasil scraping data	12
6	Tahap pemilihan data penelitian	13
7	Keputusan kelas sentimen	14
8	Contoh perhitungan <i>sentiment scoring</i>	14
9	Contoh proses case folding	16
10	Contoh proses data <i>cleaning</i>	16
11	Contoh proses normalisasi kata	16
12	Contoh proses <i>stemming</i>	17
13	Contoh proses <i>stopwords removal</i>	17
14	Contoh proses tokenisasi fitur unigram	17
15	Contoh proses tokenisasi fitur bigram dan trigram	17
16	Jumlah term unik pada setiap fitur n-gram	18
17	Proporsi kelas hasil pelabelan berdasarkan rating dan sentiment scoring	19
18	Term penciri kelas sentimen	20
19	Contoh dokumen	21
20	Contoh hasil perhitungan bobot tiap term menggunakan TF-IDF	22
21	Contoh bentuk data hasil TF-IDF	22
22	Komposisi data latih dan data uji	23
23	Proporsi kelas data latih sebelum dan setelah menerapkan SMOTE	23
24	Hyperparameter default yang digunakan	23
25	Performa SVM dengan kernel pada fitur n-gram pada data testing	24
26	Contoh ulasan hasil klasifikasi	25

DAFTAR GAMBAR

1.	<i>Hyperplane Support Vector Machine</i> (Pamuji <i>et al.</i> 2015)	3
2.	Ilustrasi pendekatan OAA (dimodifikasi dari Band (2020))	4
3.	Transformasi data dari <i>input space</i> ke <i>feature space</i> (Nugroho <i>et al.</i> 2003)	5
4.	Ilustrasi <i>k-fold cross validation</i> (Refaeilzadeh <i>et al.</i> 2016)	9
5.	Ilustrasi algoritma SMOTE (Graa dan Rekik 2019)	10
6.	Prosedur analisis data	11
7.	Proporsi ulasan berdasarkan rating 1 hingga 5	18
8.	<i>Word cloud</i> n-gram menurut kelas sentimen	19
9.	<i>Word cloud</i> keseluruhan ulasan	21

DAFTAR LAMPIRAN

1.	Lampiran 1 Performa SVM dengan kernel pada fitur n-gram pada data training	30
2.	Lampiran 2 <i>Confusion matrix</i> hasil klasifikasi model terbaik	31