



**ANALISIS SENTIMEN ULASAN APLIKASI TERHADAP PEMBARUAN
VERSI DENGAN METODE LEXICON –
MACHINE LEARNING**

SKRIPSI

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

LARAS WATI

NIM 2211016120013

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2026

SKRIPSI

ANALISIS SENTIMEN ULASAN APLIKASI TERHADAP PEMBARUAN VERSI DENGAN METODE LEXICON – MACHINE LEARNING

Oleh:

LARAS WATI

NIM. 2211016120013

Telah dipertahankan di depan Dosen Penguji pada tanggal 23 Juni 2026.

Susunan Dosen Penguji:

Pembimbing I



Muhammad Itqan Mazdadi, S.Kom., M.Kom.

NIP. 199006122019031013

Dosen Penguji I



Radityo Adi Nugroho, S.T., M.Kom.

NIP. 198212042008011006

Pembimbing II



Muliadi, S.Kom., M.Cs.

NIP. 197804222010121002

Dosen Penguji II

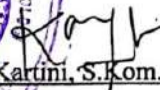


Fatma Indriani, S. T., M. I. T., Ph. D.

NIP. 198404202008122004

Banjarbaru, 08 Juli 2026

Koordinator Program Studi Ilmu Komputer



Dwi Kartini, S.Kom., M.Kom.

NIP. 198704212012122003

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 10 Juni 2025

Yang Menyatakan,

A handwritten signature in black ink, appearing to be 'Laras Wati', with a horizontal line extending to the right.

Laras Wati

NIM. 2211016120013

ABSTRAK

ANALISIS SENTIMEN ULASAN APLIKASI TERHADAP PEMBARUAN VERSI DENGAN METODE LEXICON – MACHINE LEARNING

(Oleh: Laras Wati; Pembimbing: Muhammad Itqan Mazdadi, S.Kom., M.Kom. dan Muliadi, S.Kom., M.Cs.; 2026; 56 halaman)

Setiap pembaruan versi aplikasi dapat memperoleh respons berbeda dari pengguna yang tercermin melalui ulasan pada tiap versi. Ulasan pengguna di Google Play Store memuat opini dan pengalaman terhadap aplikasi dalam bentuk teks yang besar dan tidak terstruktur, sehingga analisis manual kurang efektif dan diperlukan pendekatan otomatis berbasis *Natural Language Processing* (NLP). Metode *Lexicon* dan *Machine Learning* digunakan untuk membantu klasifikasi sentimen secara otomatis. Penelitian ini bertujuan untuk membandingkan performa algoritma *Machine Learning*, yaitu *Support Vector Machine* (SVM), *Naïve Bayes*, dan *Random Forest* dalam klasifikasi sentimen ulasan aplikasi Gojek, serta menganalisis perubahan sentimen pada setiap pembaruan versi aplikasi. Tahapan penelitian meliputi preprocessing data, pelabelan sentimen otomatis menggunakan kamus *InSet Lexicon*, pembobotan kata menggunakan TF-IDF, klasifikasi sentimen menggunakan algoritma SVM, *Naïve Bayes*, dan *Random Forest*, serta analisis perubahan sentimen berdasarkan versi aplikasi. Dataset bersumber dari Kaggle dan terdiri atas 45.881 ulasan pengguna Gojek versi 5.0.0 hingga 5.40.1. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki performa terbaik dengan *accuracy* sebesar 94,4% dan *F1-score (Macro)* sebesar 0,943. Nilai tersebut lebih tinggi dibandingkan *Random Forest* sebesar 0,917 dan *Naïve Bayes* sebesar 0,762. Selain itu, hasil analisis menunjukkan adanya perubahan sentimen positif, negatif, dan netral yang berbeda, sehingga menggambarkan perubahan persepsi pengguna pada setiap versi aplikasi.

Kata Kunci: Analisis Sentimen, Pembaruan Versi Aplikasi, *Lexicon*, *Machine Learning*.

ABSTRACT

SENTIMENT ANALYSIS OF APPLICATION REVIEWS TOWARDS VERSION UPDATES USING THE LEXICON – MACHINE LEARNING METHOD

(By: Laras Wati; Supervisor: Muhammad Itqan Mazdadi, S.Kom., M.Kom. and Muliadi, S.Kom., M.Cs.; 2026; 56 pages)

Each application version update may receive different responses from users, as reflected in reviews for each version. User reviews on Google Play Store contain opinions and experiences in the form of large and unstructured text data, making manual analysis less effective and requiring an automated Natural Language Processing (NLP) approach. Lexicon and Machine Learning methods are used to support automatic sentiment classification. This study compares the performance of Support Vector Machine (SVM), Naïve Bayes, and Random Forest in classifying sentiment in Gojek application reviews, and analyzes sentiment changes across application versions. The research stages include preprocessing, automatic sentiment labeling using InSet Lexicon, TF-IDF word weighting, sentiment classification using SVM, Naïve Bayes, and Random Forest, and version-based sentiment analysis. The dataset was obtained from Kaggle and consists of 45,881 Gojek user reviews from versions 5.0.0 to 5.40.1. The results show that SVM achieved the best performance, with an accuracy of 94.4% and a Macro F1-score of 0.943, higher than Random Forest at 0.917 and Naïve Bayes at 0.762. In addition, the analysis shows differences in positive, negative, and neutral sentiments, which can describe changes in user perceptions across application versions.

Keywords: *Sentiment Analysis, Application Version Updates, Lexicon, Machine Learning.*

PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Ulasan Aplikasi terhadap Pembaruan Versi dengan Metode *Lexicon – Machine Learning*” untuk memenuhi syarat dalam menyelesaikan pendidikan S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW beserta sahabat, keluarga, dan pengikut beliau hingga *yaumul qiama*.

Pada lembar ini penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga tersayang, yaitu ayah, ibu, kakak, dan adik yang senantiasa memberikan doa, dukungan, semangat, serta kepercayaan kepada penulis, sekaligus menjadi sumber kekuatan bagi penulis untuk terus berjuang mengejar pendidikan setinggi mungkin dan menyelesaikan skripsi ini.
2. Bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom. selaku dosen pembimbing utama dan Bapak Muliadi, S.Kom., M.Cs. selaku dosen pembimbing pendamping, yang senantiasa membimbing, membantu, dan meluangkan waktu dalam proses penyelesaian skripsi ini.
3. Ibu Dwi Kartini, S.Kom., M.Kom. selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM dan Bapak Muliadi, S.Kom., M.Cs. selaku dosen pembimbing akademik, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
4. Seluruh dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini sangat bermanfaat.

5. Lembaga Etos ID yang telah memberikan dukungan melalui program beasiswa dan pembinaan selama masa studi sehingga penulis dapat menyelesaikan perkuliahan dan penyelesaian skripsi ini.
6. Teman seperjuangan dari masa perkuliahan sampai tugas akhir ini dan seluruh teman-teman angkatan Nvidia'22 Program Studi Ilmu Komputer.
7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini masih jauh dari sempurna dan penulis mengharapkan bantuan berupa kritik dan saran yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 15 Juni 2025



Laras Wati

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK.....	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	
2.1 Kajian Terdahulu.....	6
2.2 Landasan Teori.....	12
2.2.1 Analisis Sentimen.....	12
2.2.2 Machine Learning.....	12
2.2.3 Lexicon.....	13
2.2.4 Teknik Sampel Evaluasi Kesesuaian Label.....	14
2.2.5 Ekstraksi Fitur.....	15
2.2.6 <i>Support Vector Machine</i> (SVM).....	15
2.2.7 <i>Naïve Bayes</i>	16
2.2.8 <i>Random Forest</i>	16
2.2.9 Evaluasi Model.....	17
2.2.10 Uji Signifikansi Perbandingan Model.....	20

BAB III METODE PENELITIAN

3.1 Alat Penelitian	21
3.2 Bahan Penelitian	21
3.3 Variabel Penelitian.....	21
3.4 Prosedur Penelitian	22

BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil	26
4.1.1 Pengumpulan Dataset	26
4.1.2 <i>Preprocessing</i>	27
4.1.3 Pelabelan Sentimen.....	33
4.1.4 Pembagian Data dan Ekstraksi Fitur	36
4.1.5 Klasifikasi	38
4.1.6 Evaluasi Model	40
4.1.7 Analisis Versi.....	44
4.2 Pembahasan	46

BAB V PENUTUP

5.1 Kesimpulan.....	53
5.2 Saran	53

DAFTAR PUSTAKA

LAMPIRAN

RIWAYAT PENULIS

DAFTAR TABEL

Tabel 1 Keaslian Penelitian.....	9
Tabel 2 Perancangan Penelitian	11
Tabel 3 Contoh Data Ulasan Aplikasi Gojek.....	27
Tabel 4 Contoh Penerapan <i>Cleaning</i>	28
Tabel 5 Contoh Penerapan <i>Case Folding</i>	29
Tabel 6 Contoh Penerapan Normalisasi	30
Tabel 7 Contoh Penerapan Tokenisasi.....	31
Tabel 8 Contoh Penerapan <i>Stopword Removal</i>	32
Tabel 9 Contoh Penerapan <i>Stemming</i>	33
Tabel 10 Contoh Penerapan Perhitungan Skor Sentimen	34
Tabel 11 Distribusi Label Sentimen.....	34
Tabel 12 Evaluasi Kesesuaian Label Manual dengan Sampel Label InSet	36
Tabel 13 Hasil Pembagian Data	37
Tabel 14 Contoh perhitungan TF-IDF	37
Tabel 15 Evaluasi Model Klasifikasi	40
Tabel 16 Hasil Uji <i>McNemar</i>	43

DAFTAR GAMBAR

Gambar 1 Proses <i>Random Forest</i>	17
Gambar 2 Visualisasi <i>Confusion Matrix Multiclass</i>	18
Gambar 3 Prosedur Penelitian.....	22
Gambar 4 <i>Confusion Matrix SVM</i>	41
Gambar 5 <i>Confusion Matrix Naïve Bayes</i>	41
Gambar 6 <i>Confusion Matrix Random Forest</i>	42
Gambar 7 <i>Crosstab</i> Hasil Uji <i>McNemar</i>	43
Gambar 8 Distribusi Sentimen Ulasan Gojek	46
Gambar 9 Perbandingan Performa Antar Model	48
Gambar 10 Perubahan Sentimen Berdasarkan Versi Aplikasi.....	49
Gambar 11 Kata Dominan pada Sentimen Positif	51
Gambar 12 Kata Dominan pada Sentimen Negatif.....	51
Gambar 13 Kata Dominan pada Sentimen Netral.....	51

DAFTAR LAMPIRAN

- Lampiran 1 Persiapan Program - Source code import library
- Lampiran 2 Data Collection - Source code load dataset
- Lampiran 3 Data Preprocessing – Source Code Cleaning, Case Folding, Normalization, Tokenizing, Stopword, dan Stemming
- Lampiran 4 Labeling – Source Code Lexicon InSet
- Lampiran 5 Data Splitting dan Feature Extraction – Source Code Split data dan TF-IDF
- Lampiran 6 Sentiment Classification – Source Code Algoritma Machine Learning
- Lampiran 7 Model Evaluation – Source Code Accuracy, Precision (Macro), Recall (Macro), dan F1-score (Macro)
- Lampiran 8 Model Evaluation – Source Code Confusion Matrix
- Lampiran 9 Version Analysis – Source Code Distribusi Sentimen Berdasarkan Versi