



ANALISIS SENTIMEN PENGGUNA APLIKASI FEMALE DAILY MENGUNAKAN *SUPPORT VECTOR MACHINE*: EVALUASI LEKSIKON SENTISTRENGHTID DAN INSET

NOVITA RAMADHANI MURTI



DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024



@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 mengumunkan 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 Pengguna Aplikasi Female Daily Menggunakan *Support Vector Machine*: Evaluasi Leksikon Sentistrengthid dan InSet” 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 laporan akhir ini.

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

Bogor, Agustus 2024

Novita Ramadhani Murti
G14180066

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 mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

NOVITA RAMADHANI MURTI. Analisis Sentimen Pengguna Aplikasi Female Daily Menggunakan *Support Vector Machine*: Evaluasi Leksikon Sentistrenghid dan InSet. Dibimbing oleh PIKA SILVIANTI dan AAM ALAMUDI.

Female Daily merupakan sebuah aplikasi kecantikan yang menyediakan wadah bagi sesama penggunanya untuk berdiskusi tentang produk kecantikan. Aplikasi ini dapat diunduh salah satunya melalui Google Play dimana penggunanya dapat memberikan ulasan yang berupa komentar positif, netral, maupun negatif. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Female Daily, serta mengidentifikasi perbedaan performa penggunaan kamus dari leksikon sentistrenghid dan InSet untuk proses pelabelan data pada algoritma *Support Vector Machine*. Data yang digunakan merupakan data ulasan hasil *scrapping* dari Google Play dalam rentang waktu Januari 2019 hingga Desember 2023. *Grid search* digunakan untuk mencari kombinasi fungsi kernel dan *hyperparameter* (C , γ , P) yang optimal. Pemodelan dilakukan menggunakan metode *multi-class Support Vector Machine* (SVM) dengan menerapkan algoritma *One Against All* (OAA). Model SVM pertama dibentuk menggunakan data hasil pelabelan sentistrenghid dengan fungsi kernel RBF serta nilai $C = 5$ dan $\gamma = 'scale'$. Selanjutnya, model SVM kedua dibentuk menggunakan data hasil pelabelan InSet dengan fungsi kernel linear serta nilai $C = 1$. Hasil evaluasi kedua model menggunakan data uji memperlihatkan nilai akurasi model sebesar 78% dan 76%. Selanjutnya, kedua model diterapkan ke seluruh data lalu dievaluasi menggunakan sentimen nilai *rating*. Akurasi yang didapatkan yaitu sebesar 65% dan 57%.

Kata kunci: analisis sentimen, *grid search*, InSet, sentistrenghid, SVM

ABSTRACT

NOVITA RAMADHANI MURTI. Sentiment Analysis of Female Daily Application User Using Support Vector Machine: Evaluation of Sentistrengthtid and InSet Lexicon. Supervised by PIKA SILVIANTI and AAM ALAMUDI.

Female Daily is a beauty application that provides a forum for fellow users to discuss beauty products. This application can be downloaded via Google Play, where users can provide reviews in the form of positive, neutral, or negative comments. This research aims to conduct sentiment analysis on the reviews of Female Daily application users and to identify the differences between the performance of using the dictionary from sentistrengthtid and InSet lexicon for data labeling process in the Support Vector Machine algorithm. The data used is review data scraped from Google Play from January 2019 to December 2023. Grid search is used to find the optimal combination of kernel function and hyperparameters (C , γ , P). Modeling was done using the multi-class Support Vector Machine (SVM) method by applying the One Against All (OAA) algorithm. The first SVM model was formed using data from the sentistrengthtid labeling with the RBF kernel function and the values $C = 5$ and $\gamma = \text{'scale'}$. Then the second SVM model was formed using data from the InSet labeling with the Linear kernel function and the value $C = 1$. Evaluation results of the two models using test data showed model accuracy values of 78% and 76%. Furthermore, when the two models were applied to all data and then evaluated using the sentiment of rating value, the accuracy obtained was 65% and 57%.

Keywords: grid search, InSet, sentiment analysis, sentistrengthtid, SVM



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 2024¹
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 PENGGUNA APLIKASI FEMALE DAILY MENGUNAKAN *SUPPORT VECTOR MACHINE*: EVALUASI LEKSIKON SENTISTRENGHTID DAN INSET

NOVITA RAMADHANI MURTI

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

**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@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.



**Judul Laporan : Analisis Sentimen Pengguna Aplikasi Female Daily
Menggunakan *Support Vector Machine*: Evaluasi Leksikon
Sentistrengthid dan InSet**

Nama : Novita Ramadhani Murti
NIM : G14180066

Disetujui oleh

Pembimbing 1:
Pika Silvianti, S.Si., M.Si.



Pembimbing 2:
Ir. Aam Alamudi, M.Si.



Diketahui oleh

Ketua Departemen Statistika:
Dr. Bagus Sartono, S.Si, M.Si.
NIP 19780411 2005011002



Tanggal Ujian:
2 Agustus 2024

Tanggal Lulus:

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.

PRAKATA

Puji dan syukur kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Desember 2022 sampai bulan Agustus 2024 ini ialah analisis sentimen, dengan judul “Analisis Sentimen Pengguna Aplikasi Female Daily Menggunakan *Support Vector Machine*: Evaluasi Leksikon Sentistrengthtid dan InSet”.

Terima kasih penulis ucapkan kepada para pihak-pihak yang telah berperan sejak penulis menempuh studi sampai proses penulisan karya ilmiah ini berakhir, diantaranya:

1. Kedua orang tua dan keluarga yang telah memberikan doa dan dukungan kepada penulis.
2. Ibu Pika Silvianti, S.Si, M.Si. dan Bapak Ir. Aam Alamudi, M.Si. selaku dosen pembimbing skripsi yang telah membimbing dan memberikan arahan terkait penulisan karya ilmiah ini.
3. Bapak Dr. Ir. I Made Sumertajaya, M.Si., dan Ibu Akbar Rizki, M.Si., selaku moderator kolokium dan seminar yang telah memberikan saran dan masukan untuk memperbaiki karya ilmiah ini.
4. Bapak Gerry Alfa Dito, S.Si., M.Si. selaku dosen penguji luar komisi pembimbing yang telah memberikan saran dan masukan untuk memperbaiki karya ilmiah ini.
5. Seluruh dosen dan staff akademik Departemen Statistika IPB yang telah menyampaikan ilmunya dan membantu selama proses perkuliahan hingga menyelesaikan studi.
6. Fetri, Detia, Dea, Mila, Cemi, dan Cika yang telah memberikan doa, dukungan, motivasi, dan semangat sehingga penulis mampu menyelesaikan karya ilmiah ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2024

Novita Ramadhani Murti



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
TINJAUAN PUSTAKA	3
2.1 Analisis Sentimen	3
2.2 <i>Sentiment Scoring</i>	3
2.3 Pembobotan Kata	3
2.4 <i>Support Vector Machine</i>	4
2.5 <i>Hyperparameter tuning</i>	8
2.6 Evaluasi Model	9
III METODE	11
3.1 Bahan dan Data	11
3.2 Prosedur Penelitian	11
IV HASIL DAN PEMBAHASAN	18
4.1 Visualisasi Data	18
4.2 Pelabelan Sentimen	20
4.3 TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)	20
4.4 Pembagian Data	20
4.5 <i>Hyperparameter Tuning</i>	21
4.6 Pemodelan Klasifikasi dan Evaluasi Performa Model	22
4.7 Hasil Prediksi	23
V SIMPULAN	25
DAFTAR PUSTAKA	26
RIWAYAT HIDUP	28

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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	Matriks konfusi	10
2	Peubah penelitian	11
3	Contoh hasil pengambilan data	12
4	Contoh proses <i>cleaning data</i>	13
5	Contoh proses normalisasi kata	13
6	Contoh proses <i>stopword removal</i>	13
7	Contoh proses <i>stemming</i>	14
8	Contoh proses tokenisasi	14
9	Contoh perhitungan <i>sentiment scoring</i>	15
10	Contoh perhitungan bobot setiap <i>term</i>	16
11	Matriks hasil TF-IDF	16
12	Rincian hasil pelabelan	20
13	Rincian pembagian data latih dan data uji	21
14	Kombinasi fungsi kernel dan <i>hyperparameter</i>	21
15	Hasil <i>grid search</i> (pelabelan <i>sentistrengthtid</i>)	22
16	Hasil <i>grid search</i> (pelabelan InSet)	22
17	Rincian model SVM yang dibuat	23
18	Hasil performa model SVM yang telah dibuat	23

DAFTAR GAMBAR

1	Ilustrasi <i>Support Vector Machine</i> (Kecman 2005)	5
2	Ilustrasi fungsi kernel (Schölkopf dan Smola 2018)	7
3	Ilustrasi pendekatan OAA	8
4	Ilustrasi pengaruh nilai <i>Cost</i> (C) (Ben-hur dan Weston 2010)	8
5	Ilustrasi pengaruh nilai P (Ben-hur dan Weston 2010)	9
6	Ilustrasi pengaruh nilai γ (Ben-hur dan Weston 2010)	9
7	Diagram alir penelitian	11
8	Diagram alir proses <i>sentiment scoring</i>	15
9	Proporsi ulasan antara <i>rating</i> 1 hingga 5	18
10	Klasifikasi sentimen menggunakan <i>rating</i>	18
11	Awan kata ulasan positif	19
12	Awan kata ulasan netral	19
13	Awan kata ulasan negatif	20
14	Matriks konfusi model 1 (SVM+ <i>sentistrengthtid</i>)	24
15	Matriks konfusi model 2 (SVM+InSet)	24