

Analisis Sentimen Ulasan Aplikasi Brompit Di Google Play Store Menggunakan Algoritma Support Vector Machine

Oleh:

Rifqi Yusril Muslikhin,

Ika Ratna Indra Astutik

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

Februari, 2023



Pendahuluan

Google Play Store adalah salah satu tempat untuk mendownload berbagai macam aplikasi berbasis android. Pada Google Play Store terdapat fitur untuk memberikan penilaian terhadap aplikasi, yaitu fitur rating dan ulasan dari pengguna. Tetapi kebanyakan rating yang diberikan pengguna tidak cocok dengan ulasan sehingga susah bagi pengguna untuk menentukan penilaian terhadap aplikasi. Dengan menggunakan text mining, ulasan pengguna di Google Play Store bisa diklasifikasikan menjadi ulasan positif dan negatif yang nantinya bisa menjadi pertimbangan bagi pengguna baru yang akan menggunakan aplikasi ini.



Pertanyaan Penelitian (Rumusan Masalah)

- Bagaimana penerapan algoritma Support Vector Machine dalam klasifikasi analisis sentimen ?
- Bagaimana menentukan hasil akurasi yang dihasilkan algoritma Support Vector Machine



Metode

01

Pengumpulan Data

Pengumpulan data ulasan dilakukan dengan menggunakan google play scrapper. Data diambil dari ulasan aplikasi brompit di google play store yang berjumlah 3359 data ulasan.

02

Pelabelan Data

Pelabelan Data Dilakukan menggunakan library vader sentiment di Bahasa pemrograman python

03

Text PreProcessing

Cleaning, Case Folding, Tokenizing, Normalisasi, Filtering, Stemming

04

Pembobotan Kata

Pembobotan Term Frequency Inverse Document Frequency (TF-IDF)



Metode

05

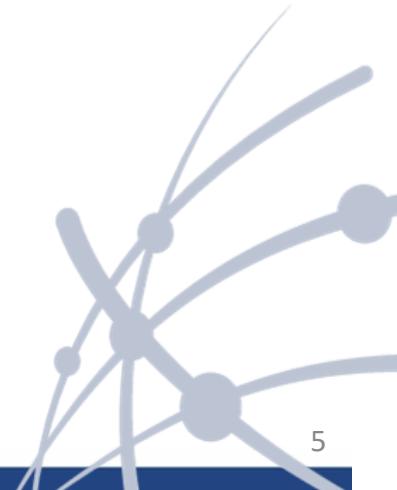
Algoritma SVM

Klasifikasi Menggunakan
Algoritma SVM

06

Pengujian

K-Fold Cross
Validation, Confusion
Matrix



Hasil

Berdasarkan hasil pengujian yang dilakukan dapat disimpulkan bahwa algoritma support vector machine terbukti dapat digunakan untuk klasifikasi data. Kernel RBF mendapatkan akurasi tertinggi dibanding kernel lainnya dengan besar akurasi 84%. Hasil confusion matriks menunjukkan bahwa akurasi dari algoritma support vector machine dalam penelitian ini sebesar 84%. Akurasi tertinggi dalam penelitian ini didapatkan menggunakan metode k-fold cross validation dengan besar akurasi 87,6% yang didapatkan dari iterasi fold ke-8 dari pengujian k-fold cross validation.



Pembahasan

Hasil Preprocessing

Cleaning

Ulasan Sebelum <i>Cleaning</i>	Ulasan Sesudah <i>Cleaning</i>	Kelas	Username
Praktis, professional, dan user friendly..	Praktis professional dan user friendly	Positif	Laila sari
Tidak bisa melakukan booking service... Layar jadi full putih/ngeblank	Tidak bisa melakukan booking service Layar jadi full putihngeblank	Negatif	Oktavia Putri

Case Folding

Ulasan Sebelum <i>Case Folding</i>	Ulasan Sesudah <i>Case Folding</i>	Kelas	Username
Praktis professional dan user friendly	praktis professional dan user friendly	Positif	Laila sari
Tidak bisa melakukan booking service Layar jadi full putihngeblank	tidak bisa melakukan booking service layar jadi full putihngeblank	Negatif	Oktavia Putri



Pembahasan

Hasil Preprocessing

Tokenizing

Ulasan Sebelum <i>Tokenizing</i>	Ulasan Sesudah <i>Tokenizing</i>	Kelas	Username
praktis professional dan user friendly	['praktis', 'professional', 'dan', 'user', 'friendly']	Positif	Laila sari
tidak bisa melakukan booking service layar jadi full putihngeblank	['tidak', 'bisa', 'melakukan', 'booking', 'service', 'layar', 'jadi', 'full', 'putihngeblank']	Negatif	Oktavia Putri

Normalisasi

Ulasan Sebelum <i>Normalisasi</i>	Ulasan Sesudah <i>Normalisasi</i>	Kelas	Username
['praktis', 'professional', 'dan', 'user', 'friendly']	['praktis', 'professional', 'dan', 'user', 'friendly']	Positif	Laila sari
['tidak', 'bisa', 'melakukan', 'booking', 'service', 'layar', 'jadi', 'full', 'putihngeblank']	['tidak', 'bisa', 'melakukan', 'booking', 'service', 'layar', 'jadi', 'full', 'putih', 'ngeblank']	Negatif	Oktavia Putri

Pembahasan

Hasil Preprocessing

Filtering

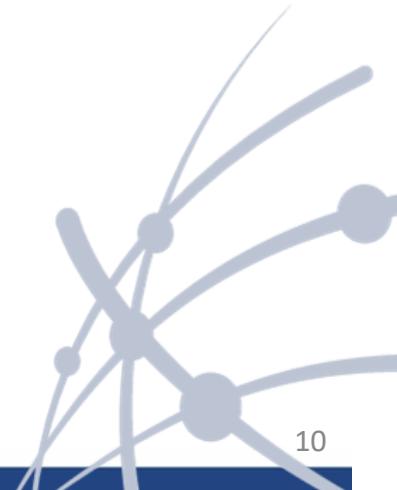
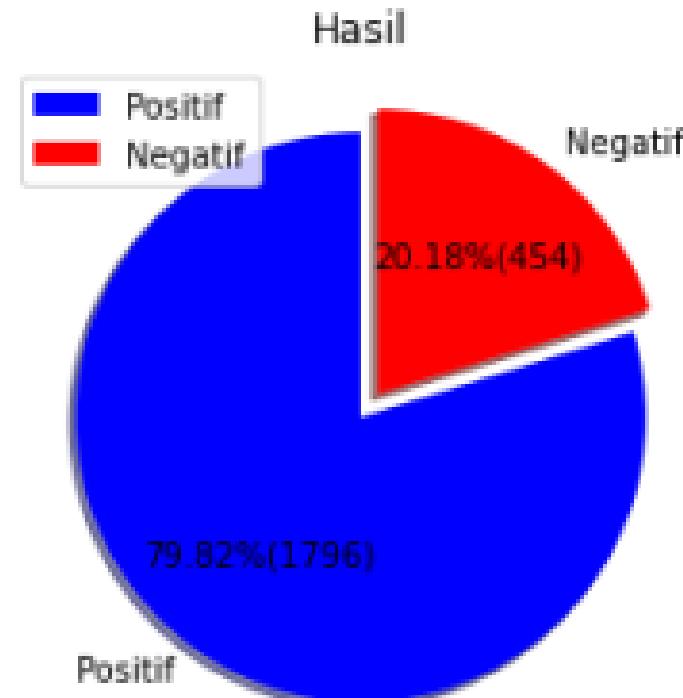
Ulasan Sebelum <i>Filtering</i>	Ulasan Sesudah <i>Filtering</i>	Kelas	Username
<code>['praktis', 'professional', 'dan', 'user', 'friendly']</code>	<code>['praktis', 'professional', 'user', 'friendly']</code>	Positif	Laila sari
<code>['tidak', 'bisa', 'melakukan', 'booking', 'service', 'layar', 'full', 'putih', 'jadi', 'full', 'putihngeblank']</code>	<code>['booking', 'service', 'layar', 'full', 'putih', 'ngeblank']</code>	Negatif	Oktavia Putri

Stemming

Ulasan Sebelum <i>Stemming</i>	Ulasan Sesudah <i>Stemming</i>	Kelas	Username
<code>['praktis', 'professional', 'user', 'friendly']</code>	<code>['praktis', 'professional', 'user', 'friendly']</code>	Positif	Laila sari
<code>['booking', 'service', 'layar', 'full', 'putih', 'blank']</code>	<code>['booking', 'service', 'layar', 'full', 'putih', 'blank']</code>	Negatif	Oktavia Putri

Pembahasan

Setelah proses text preprocessing, data yang sebelumnya berjumlah 3359 data berkurang menjadi 2250 data.



Pembahasan

Hasil Pengujian Menggunakan Kernel Linear

Perbandingan	Hasil Pengujian
80% : 20%	82,2%
90% : 10%	80%
70% : 30%	81,4%

Hasil Pengujian Menggunakan Kernel Polynomial

Perbandingan	Hasil Pengujian
80% : 20%	82,4%
90% : 10%	83,1%
70% : 30%	83,8%



Pembahasan

Hasil Pengujian Menggunakan Kernel RBF

Perbandingan	Hasil Pengujian
80% : 20%	82%
90% : 10%	84%
70% : 30%	83,7%

Hasil Pengujian Menggunakan Kernel Sigmoid

Perbandingan	Hasil Pengujian
80% : 20%	82,4%
90% : 10%	83,1%
70% : 30%	83,8%

Dari hasil pengujian menggunakan kernel yang ada di algoritma Support Vector Machine, akurasi tertinggi didapatkan menggunakan kernel RBF dengan besar akurasi 84%.



Pembahasan

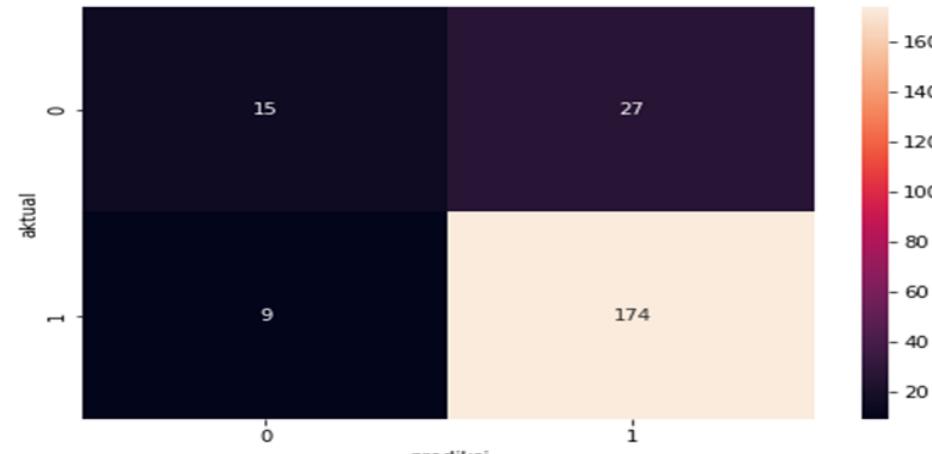
- Pengujian Menggunakan K-Fold Cross Validation

Iterasi ke-	Hasil Pengujian
1	79,8%
2	86,2%
3	83,7%
4	83,7%
5	86,2%
6	83,1%
7	84,6%
8	87,6%
9	85,1%
10	87,1%



Pembahasan

- Pengujian Menggunakan Confusion Matriks



Confusion Matrix :

```
[[ 15 27]
 [ 9 174]]
```

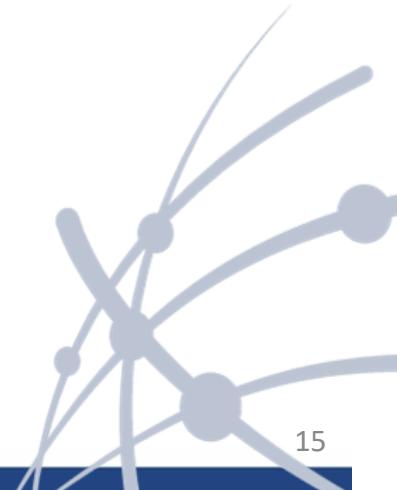
Report Hasil :

	precision	recall	f1-score	support
0	0.62	0.36	0.45	42
1	0.87	0.95	0.91	183
accuracy			0.84	225
macro avg	0.75	0.65	0.68	225
weighted avg	0.82	0.84	0.82	225



Temuan Penting Penelitian

- Algoritma support vector machine terbukti dapat digunakan untuk klasifikasi data.
- Akurasi tertinggi dalam penelitian ini didapatkan menggunakan metode k-fold cross validation dengan besar akurasi 87,6%.
- Sebagian besar ulasan negatif dari pengguna untuk aplikasi ini mengatakan bahwa aplikasi ini masih sering error dan trouble.



Manfaat Penelitian

- Manfaat bagi Penulis

Penulis dapat menambah pengetahuan dan pemahaman dalam topik text mining khususnya algoritma Support Vector Machine. Serta penelitian ini dapat dijadikan sarana untuk menerapkan ilmu dan keterampilan yang telah didapatkan selama masa perkuliahan.

- Manfaat bagi Universitas

Dapat menambah literatur di Perpustakaan serta dapat dijadikan referensi bagi mahasiswa atau calon peneliti lain sebagai bahan acuan dalam mengembangkan dan mengkaji algoritma Support Vector Machine.

- Manfaat bagi Pembaca

Diharapkan dapat digunakan sebagai referensi bagi pembaca yang nantinya akan dikembangkan lebih jauh lagi.



Referensi

- [1] Priyanto, S. A. (2019). Aplikasi Pemesanan Servis Sepeda Motor Online Berbasis Android (Doctoral dissertation, Universitas Panca Marga Probolinggo).
- [2] Mpmhondajatim. (2019). Diakses dari <https://www.mpmhondajatim.com/brompit>.
- [3] Google Play Store. (2021). Diakses dari <https://play.google.com/store/apps/details?id=com.mpm.brompit>.
- [4] Herlinawati, Nuraeni, et al. (2020) "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine." CESS (Journal of Computer Engineering, System and Science) 5.2: 293-298.
- [5] Raymond J Mooney, C. (2006). Machine Learning Text Categorization. Machine Learning Text Categorization, 1–6.
- [6] Deolika, A., Kusrini, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. (JurTI) Jurnal Teknologi Informasi, 3(2), 179-184.
- [7] Luqyana, W. A. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine (Doctoral dissertation, Universitas Brawijaya).
- [8] Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. STRING (Satuan Tulisan Riset dan Inovasi Teknologi), 6(2), 129-136.



Referensi

- [9] Findawati, Y., & Rosid, M. A. (2020). Buku Ajar Text Mining. Umsida Press, 1-123.
- [10] Fauzi, A. (2022). Penerapan Algoritma Text Mining dan Lexrank dalam Meringkas Teks Secara Otomatis. Bulletin of Data Science, 1(2), 65-72.
- [11] Retnawiyati, E., Fatoni, M. M., & Negara, E. S. (2015). Analisis Sentimen pada Data Twitter dengan Menggunakan Text Mining terhadap Suatu Produk. Departemen Teknik Informatika Universitas Bina Darma.
- [12] Pradikdo, A. C., & Ristyawan, A. (2018). Model klasifikasi abstrak skripsi menggunakan text mining untuk pengkategorian skripsi sesuai bidang kajian. Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 9(2), 1091-1098.
- [13] Ling, J., Kencana, I. P. E. N., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. E-Jurnal Matematika, 3(3), 92-99.
- [14] Lina Auliya, Z. (2019). KLASIFIKASI SENTIMEN TERHADAP APLIKASI SHOPEE DENGAN METODE NAÏVE BAYES CLASSIFIER BERDASARKAN ULASAN DI GOOGLE PLAY (Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau).
- [15] Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support vector machine. Proceeding Indones. Sci. Meeiting Cent. Japan



DARI SINI PENCERAHAN BERSEMI