



---

## ANALISIS SENTIMEN ULASAN APLIKASI DEEPL PADA GOOGLE PLAY DENGAN METODE *SUPPORT VECTOR MACHINE* (SVM)

Septi Putri Azzahra<sup>1(\*)</sup>, Yohanes Agung Apriyanto<sup>2</sup>, Andri Wijaya<sup>3</sup>

<sup>1</sup>Universitas Katolik Musi Charitas, Palembang

<sup>2</sup>Universitas Katolik Musi Charitas, Palembang

<sup>3</sup>Universitas Katolik Musi Charitas, Palembang

---

### Abstract

*In the era of globalization and technological advances, the growth in the use of mobile applications is experiencing rapid development. Understanding the level of user satisfaction and dissatisfaction with mobile applications is very important. One of the very popular online translator applications on the Google Play Store is DeepL which uses artificial intelligence technology to translate languages. DeepL has become a very useful tool for individuals and organizations in overcoming language barriers. This research aims to conduct sentiment analysis on user reviews of the DeepL application on the Google Play Store platform, using the Support Vector Machine classification method. The Support Vector Machine (SVM) classification method is one of the classification approaches in the realm of guided learning in data mining. The advantage of SVM lies in its ability to handle nonlinear and high-dimensional input data by utilizing kernel functions. Data collection was carried out using Web Scraping techniques using Python. The results of this research show an accuracy rate of 91%, which reflects how well the Support Vector Machine (SVM) model is at classifying data. In this research, it is known that the model has a high level of precision, especially in the 'Positive' category of 94%, which shows the model's ability to accurately recognize data that is included in positive sentiment.*

---

**Kata Kunci:** *Sentiment Analysis, DeepL, SVM, Python*

Juli - Desember 2023, Vol 4 (2) : hlm 59 - 66  
©2023 Institut Teknologi dan Bisnis Ahmad Dahlan.  
All rights reserved.

## PENDAHULUAN

Di era globalisasi dan kemajuan teknologi, Penggunaan aplikasi *mobile* semakin berkembang pesat, sehingga menjadi esensial dalam membantu memahami tingkat kepuasan pengguna maupun ketidakpuasan pengguna saat berinteraksi dengan aplikasi. Google menyediakan *platform* bernama *Play Store*. Salah satu fitur yang dapat ditemukan di *Play Store* adalah fitur penilaian dan ulasan, di mana pengguna dapat memberikan pendapat mereka terkait aplikasi yang telah mereka gunakan.

Di *Google Play Store*, terdapat berbagai aplikasi penerjemah online yang memiliki tingkat penggunaan serta jumlah unduhan yang sangat signifikan. Menurut (Gestani,2019), proses penerjemahan diakui sebagai tahap yang paling rumit di antara semua langkah dalam pembelajaran bahasa. Hal ini mencerminkan pentingnya penerjemahan dan kebutuhan pengguna akan aplikasi yang dapat memenuhi kebutuhan mereka.

Salah satu aplikasi penerjemah online yang populer di *Google Play Store* adalah DeepL, yang merupakan aplikasi penerjemah bahasa yang menggunakan teknologi kecerdasan buatan. DeepL menjadi alat yang sangat berguna bagi individu dan organisasi dalam mengatasi tantangan bahasa. Namun, dalam era perkembangan teknologi yang cepat, sangat penting untuk tidak hanya mengukur sejauh mana suatu aplikasi teknologi dapat memenuhi kebutuhan pengguna, melainkan juga mengevaluasi tingkat kepuasan dan persepsi pengguna terhadap aplikasi.

Tujuan dari penelitian ini adalah untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi DeepL di *platform Google Play Store*, menggunakan metode klasifikasi *Support Vector Machine (SVM)*. Metode klasifikasi *Support Vector Machine (SVM)* merupakan salah satu metode klasifikasi dalam ranah *supervised learning* di data mining. Keunggulan SVM terletak pada kemampuannya untuk melakukan pembagian linear pada data input yang bersifat nonlinear dan mempunyai dimensi yang besar, dengan memanfaatkan fungsi kernel. Menurut (Hartmann,2018) sebagaimana disebutkan banyak peneliti telah mengungkapkan bahwa penggunaan SVM dalam klasifikasi teks memberikan hasil yang lebih akurat.

Dalam penelitian yang dilakukan oleh Dea Oktavia, Yudhi Raymond Ramadhan, dan Minarto berjudul "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma *Support Vector Machine (SVM)*," tujuan penelitian ini adalah untuk menganalisis sentimen pengguna Twitter terkait penerapan sistem e-Tilang. Metode yang digunakan adalah algoritma *Support Vector Machine (SVM)*, dan hasil analisis menunjukkan bahwa nilai sentimen dari data tweet mencapai 74,20%, dengan tingkat ketepatan (*precision*) sebesar 83,33%, dan tingkat keberhasilan menemukan kembali informasi (*recall*) sebesar 5,28%. Berdasarkan hasil penelitian dengan menggunakan algoritma SVM, sentimen yang terungkap di media sosial Twitter terkait penerapan e-Tilang dapat dikategorikan sebagai netral.

Oleh karena itu, penelitian ini digunakan metode *Support Vector Machine (SVM)* agar dapat mengkategorikan apakah ulasan tersebut bersifat positif, negatif, atau netral serta diharapkan kita bisa memperoleh pemahaman yang lebih menyeluruh mengenai keunggulan dan kelemahan aplikasi DeepL dari sudut pandang pengguna.

## METODE

Dalam penelitian ini, dilakukan proses web scraping di platform Play Store menggunakan bahasa pemrograman Python untuk mengambil link yang terkait dengan aplikasi DeepL. Metode penelitian yang digunakan yaitu metode *Knowledge Discovery from Data (KDD)*. *Knowledge Discovery In Database (KDD)* merujuk pada sejumlah kegiatan yang melibatkan pengumpulan dan pemanfaatan data historis untuk mengenali pola, keteraturan, atau hubungan dalam kumpulan data yang berskala besar (Santoso, 2007). *Knowledge Discovery In Database (KDD)* mempunyai 5 langkah, yakni *Data Selection*, *Preprocessing*, *Data transformation*, *Modeling data mining*, *Interpretation/Evaluation*.



Gambar 1. Langkah Penelitian metode *Knowledge Discovery from Data (KDD)*.

### **Data Selection**

Data dikumpulkan menggunakan *Web Scraping* dengan python, mengenai ulasan pengguna aplikasi DeepL. *Web scraping* adalah metode untuk secara otomatis mengambil informasi dari suatu situs web tanpa perlu melakukan penyalinan manual (A. Yani et al., 2019). Penelitian ini memisahkan dataset menjadi dua bagian, yakni dataset pelatihan (*training data*) dan dataset pengujian (*test data*). Sebanyak 800 data telah dikumpulkan secara total, dengan rincian pembagian sebagai berikut:

1. Dataset pelatihan (*training data*): Terdiri dari 640 data.
2. Dataset pengujian (*test data*): Terdiri dari 160 data.

### **Preprocessing**

*Preprocessing* menjelaskan berbagai jenis proses yang mengolah data mentah sebagai persiapan untuk menjalankan prosedur-prosedur berikutnya. *Preprocessing* dalam data mining bertujuan untuk mengubah struktur data menjadi bentuk yang lebih mudah dan efisien untuk kebutuhan pengguna. Pada fase *preprocessing* dalam penelitian ini, mencakup pembersihan (*cleaning*), normalisasi, dan pemberian label (*labeling*).

### **Data Transformation**

*Data Transformation* bertujuan untuk menyiapkan data agar dapat diolah dalam data mining, pada penelitian ini melibatkan langkah-langkah seperti *Case Folding*, *Uppercase*, *Tokenisasi*, serta *Clears text from character*.

### **Modeling Data Mining**

*Modeling Data mining* yang diterapkan adalah *Support Vector Machine (SVM)* merupakan suatu metode klasifikasi yang digunakan untuk mengelompokkan opini secara

efektif, serta mampu mengkategorikan komentar ke dalam kelompok positif, negative maupun netral.

### ***Interpretation/Evaluation***

Pada tahap *Interpretation/Evaluation* tujuannya adalah menilai seberapa baik kinerja algoritma klasifikasi yang telah diterapkan dalam penelitian ini. Parameter yang digunakan sebagai ukuran performansi mencakup akurasi (*accuracy*), presisi (*precision*), *recall*, dan dukungan (*support*).

## **HASIL DAN PEMBAHASAN**

Hasil dan pembahasan terkait analisis sentimen ulasan pengguna aplikasi DeepL menggunakan metode Support Vector Machine (SVM) diproses melalui tools Anaconda Navigator Jupyter Notebook dan *Google Collab* dengan *library Google Play Scraper* untuk memproses Scraping data.

### ***Scraping Data***

Dalam langkah ini, dilakukan pengambilan data melalui teknik scraping dengan menyertakan tautan aplikasi DeepL dari Play Store. Output dari proses *scraping* data yang dilakukan dengan mencari kata kunci "DeepL" serta jumlah pencarian yang mencapai 800 data terkait aplikasi DeepL, dapat dilihat di bawah ini melalui hasil outputnya.

	userName	score	at	content
0	AJJUB	1	2023-11-13 00:55:36	No option translate to arabic
1	GAME Bell	5	2023-11-11 16:11:15	goodddd
2	エギエカサトラ	5	2023-11-11 14:22:04	MANTAPUUUU JIWAAA
3	achir acher	4	2023-11-11 09:04:35	The best app ya..👍👍👍
4	winda mayehan	5	2023-11-11 08:58:06	Aplikasinya keren
...	...	...	...	...
795	cLan Dika	5	2022-05-27 11:53:20	良かったアプリ
796	DASTIN NERO	5	2022-05-25 11:07:12	The best
797	Chitato Ryu	5	2022-05-20 02:38:55	Aplikasi penerjemah terakurat yang pernah saya...
798	Bismillah	1	2022-04-30 03:16:46	Indonesia tidak ada
799	SidikCasper_Id	1	2022-03-18 18:51:15	Gajelas gdaa translate ke bahasa Indonesianya

800 rows x 5 columns

Gambar 1. Hasil *Crawling Data*

### ***Labeling data***

Setelah mengumpulkan data penelitian, langkah selanjutnya adalah memberikan labeling data. Hasil pelabelan dapat dilihat di bawah ini melalui outputnya.

	userName	score	at	content	sentiment
0	AJUR	1	2023-11-12 00:55:36	No option translate to english	Negative
1	GAHE Bel	5	2023-11-11 16:11:15	good	Positif
2	エドエカゴフ	5	2023-11-11 14:22:04	MANTAPUUUJIRAAA	Positif
3	amr aher	4	2023-11-11 09:04:35	The best app ya. 🌟🌟🌟	Positif
4	winda mayhan	5	2023-11-11 00:50:06	Aplikasinya keren	Positif
795	slan Dika	5	2023-05-27 11:52:20	👍👍👍👍👍	Positif
796	DASTIN NERO	5	2023-05-26 11:07:12	The best	Positif
797	Chitah Rev	5	2023-05-20 02:30:55	Aplikasi penerjemah terakurat yang pernah saya...	Positif
798	Bamran	1	2023-04-30 03:15:45	Indonesia tidak ada	Negative
799	SitiCasper_16	1	2023-03-18 18:51:15	Gejala gila translate ke bahasa Indonesia...	Negative

500 rows × 6 columns

Gambar 2. Hasil *Labeling Data*

Gambar 2 memperlihatkan data yang telah diperoleh dan telah diberi label 'sentiment'. Jika pada data sebelumnya hanya ada kolom *user*, kolom *score*, kolom *at*, dan kolom *content*, kini telah dibentuk sebuah tabel baru dengan penambahan kolom 'sentiment'. Kolom ini digunakan untuk mengetahui sifat dari teks yang telah diperoleh, apakah bersifat positif, negative, atau netral.

### Preprocessing

Langkah selanjutnya adalah menjalankan tahap *preprocessing*. Contohnya, kita akan menggunakan satu contoh teks untuk proses tersebut. Dimulai dari data asli yang diambil dari hasil *scraping*, proses ini melibatkan *Data Transformation* yang mencakup *Case folding*, *Uppercase*, *Tokenisasi*, *Clears text from character*. dapat disimak table di bawah ini:

Tabel 1. *Preprocessing*

<i>Preprocessing</i>	<i>Content</i>
<b>Data Asli</b>	Bagus sih bagus, cuma kenapa akun aku sering log-out, ya? Kesel pake bangetðŸ™” ntah aku-nya yang salah atau sistemnya. Gatau ah, kesel
<b>Case folding</b>	bagus sih bagus, cuma kenapa akun aku sering log-out, ya? kesel pake bangetðŸ™” ntah aku-nya yang salah atau sistemnya. gatau ah, kesel.
<b>Uppercase</b>	BAGUS SIH BAGUS, CUMA KENAPA AKUN AKU SERING LOG-OUT, YA? KESEL PAKE BANGETðŸ™” NTAH AKU-NYA YANG SALAH ATAU SISTEMNYA. GATAU AH, KESEL.
<b>Tokenisasi</b>	['Bagus', 'sih', 'bagus', ',', 'cuma', 'kenapa', 'akun', 'aku', 'sering', 'log-out', ',', 'ya', '?', 'Kesel', 'pake', 'bangetðŸ™', '"', 'ntah', 'aku-nya', 'yang', 'salah', 'atau', 'sistemnya', ',', 'Gatau', 'ah', ',', 'kesel', '.']
<b>Clears text from character</b>	Bagus sih bagus cuma kenapa akun aku sering logout ya Kesel pake banget ntah akunya yang salah atau sistemnya Gatau ah kesel

### Modelling Data Mining

Selanjutnya melakukan *modeling* data mining dengan menggunakan *Support Vector Machine (SVM)*. Tahap ini dimulai dengan mengimpor modul *pandas*, diikuti dengan mengimpor *train* data dan *test* data dan memperlihatkan sebanyak 5 data sample. Dan Melakukan *TF-IDF Vectorizer* dengan memanfaatkan fitur dari modul *sklearn*, dengan

pengaturan `min_df` sebesar 5 dan `max_df` sebesar 0.8, sebagai langkah persiapan data untuk model SVM. Dapat disimak di bawah ini melalui hasil output menggunakan script yang telah dieksekusi.

	userName	score	at	content	sentiment
357	Ediwidodo Satu	5	2023-08-09 07:39:05	Bagus sekali	Positif
74	Sha	3	2023-10-20 10:51:45	Tolong tambah bacaan romaji atau furigana di t...	Netral
5	Zoni Hidayat	5	2023-11-11 01:05:53	I don't know, this translate is really accurat...	Positif
42	Syifa ifa	2	2023-10-30 13:05:23	G ada penerjemahan bahasa Arab nya	Negative
334	10 XI IPA 1 Ari Krisna	5	2023-08-17 14:00:03	perfect translate.	Positif

Gambar 3. Hasil impor *train* data dan *test* data

### Interpretation/Evaluation

Pada Tahap ini model klasifikasi menggunakan metode *Support Vector Machine* (SVM) dengan kernel linear dan dinilai menggunakan modul *scikit-learn* (*sklearn*). Evaluasi model mencakup metrik akurasi (*accuracy*), presisi (*precision*), *recall*, dan dukungan (*support*) untuk setiap kelas, yaitu Negative, Netral, dan Positif. Sehingga mendapatkan hasil *accuracy*, *precision*, *recall*, dan *support* pada gambar di bawah ini :

	precision	recall	f1-score	support
Negative	0.40	0.29	0.33	7
Netral	0.25	0.17	0.20	6
Positif	0.94	0.97	0.95	147
accuracy			0.91	160
macro avg	0.53	0.47	0.50	160
weighted avg	0.89	0.91	0.90	160

Gambar 4. Hasil Evaluasi Model

Hasil evaluasi model menunjukkan tingkat *accuracy* sebesar 91%, mencerminkan seberapa baik model dapat mengklasifikasikan data. Ditemukan bahwa model menunjukkan tingkat presisi yang tinggi, khususnya pada kategori 'Positif' sebesar 94%, menandakan kemampuan model dalam mengenali dengan akurat data yang termasuk dalam *sentiment* positif. Meskipun demikian, tingkat presisi yang lebih rendah untuk kategori 'Negative' dan 'Netral' menunjukkan adanya kesulitan dalam melakukan klasifikasi yang akurat pada kategori-kategori tersebut. Interpretasi hasil ini memberikan wawasan yang lebih mendalam mengenai kelebihan dan kelemahan model *Support Vector Machine* (SVM) dalam mengklasifikasikan setiap kelas dalam situasi pengujian tertentu.

Interpretasi hasil akan dilakukan melalui metode visualisasi kata menggunakan *WordCloud*. *Word cloud* memiliki tujuan untuk secara visual menampilkan data teks, di mana semakin besar suatu kata, frekuensi kemunculan kata tersebut juga semakin tinggi.



- Idris, I.S.K., Mustofa, Y.A., Salihi, I.A., 2023. *Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)*. Jambura J. Electr. Electron. Eng. 5, 32–35.
- Nurian, A., 2023. *ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE PLAY MENGGUNAKAN NAÏVE BAYES*. J. Inform. Dan Tek. Elektro Terap. 11.
- Octaviani, P.A., Wilandari, Y., Ispriyanti, D., 2014. *PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG*. J. Gaussian 3, 811–820.
- Oktavia, D., Ramadhan, Y.R., Minarto, M., 2023. *Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)*. KLIK Kaji. Ilm. Inform. Dan Komput. 4, 407–417.
- Wahyudi, R., Kusumawardhana, G., 2021. *Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine*. J. Inform. 8.
- Wardana, L.A., Baharuddin, B., Nurtaat, L., 2022a. *Kemampuan Mahasiswa melakukan post-editing terhadap Hasil Terjemahan Machine Translation*. J. Ilm. Profesi Pendidik. 7, 53–61.
- Wardana, L.A., Baharuddin, B., Nurtaat, L., 2022b. *Kemampuan Mahasiswa melakukan post-editing terhadap Hasil Terjemahan Machine Translation*. J. Ilm. Profesi Pendidik. 7, 53–61.
- Yani, D.D.A., Pratiwi, H.S., Muhandi, H., 2019. *Implementasi web scraping untuk pengambilan data pada situs marketplace*. JUSTIN J. Sist. Dan Teknol. Inf. 7, 257–262.
- Zen, B.P., Wicaksana, D., Alfidzar, H., 2022. *ANALISIS SENTIMEN TWEET VAKSIN COVID 19 SINOVAH MENGGUNAKAN METODE SUPPORT VECTOR MACHINE*. J. Data Min. Dan Sist. Inf. 3, 21–27.