

ANALISIS SENTIMEN PADA ULASAN APLIKASI EHADRAH DI GOOGLE PLAYSTORE MENGGUNAKAN SUPPORT VECTOR MACHINE

Ayu Basirotul Muzayyanah^{1*}, Ratri Enggar Pawening², Zainal Arifin³
^{1,2,3}Fakultas Teknik, Universitas Nurul Jadid, Karanganyar, Paiton, Probolinggo, Jawa Timur, Indonesia
Email: ^{1*} ayuayunda489@gmail.com , ²enggar.r@unuja.ac.id , ³ zainal@unuja.ac.id
(*coresponding author)

Abstrak- Aplikasi Ehadrah adalah sebuah aplikasi untuk mendengarkan dan mengakses konten hadrah, seni musik tradisional islami. Jumlah pengguna aplikasi Ehadrah semakin meningkat di era digital saat ini karena popularitasnya yang terus bertambah. Namun, ulasan pengguna sering tidak sesuai dengan rating yang ditampilkan di *Google Play Store*, menciptakan kesenjangan yang dapat menghambat pengembangan aplikasi lebih lanjut. Oleh karena itu, diperlukan analisis sentimen untuk memahami sentimen pengguna secara lebih mendalam. Data ulasan diperoleh melalui web scraping menggunakan *Google Play Store API* dan kemudian diproses melalui tahapan *preprocessing* seperti *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Pemodelan menggunakan teknik *Support Vector Machine (SVM)* dengan membandingkan kedua *kernel SVM* yaitu *Linear* dan *RBF* untuk mengklasifikasikan sentimen dalam ulasan pengguna. Tujuan penelitian ini adalah untuk mengevaluasi sentimen pengguna terhadap aplikasi eHadrah di *Google Play Store* dengan menggunakan algoritma SVM. Hasil penelitian menunjukkan bahwa SVM dengan *kernel Linear* lebih unggul dibandingkan *kernel RBF*. SVM dengan *Kernel Linear* menghasilkan akurasi 95.46%, *precision* 81.83%, *recall* 55.61%, dan *f-measure* 62.82%, sementara *kernel RBF* menghasilkan akurasi 94.15%, *precision* 58.03%, *recall* 40.33%, dan *f-measure* 43.24% dengan menggunakan 976 data ulasan.

Kata Kunci: Analisis Sentimen, Ehadrah, Google Play Store, Support Vector Machine

Abstract- The Ehadrah app is an application for listening to and accessing hadrah content, a traditional Islamic musical art. The number of users of the Ehadrah app is increasing in the current digital era due to its growing popularity. However, user reviews often do not match the ratings displayed on the Google Play Store, creating a gap that can hinder further development of the app. Therefore, sentiment analysis is required to understand user sentiment more deeply. Review data is obtained through web scraping using the Google Play Store API and then processed through preprocessing stages such as case folding, tokenizing, stopwords removal, and stemming. Modeling using Support Vector Machine (SVM) technique by comparing the two SVM kernels namely Linear and RBF to classify sentiment in user reviews. The purpose of this research is to evaluate user sentiment towards the eHadrah application on the Google Play Store using the SVM algorithm. The results showed that SVM with Linear kernel is superior to RBF kernel. SVM with Linear Kernel produces 95.46% accuracy, 81.83% precision, 55.61% recall, and 62.82% f-measure, while RBF kernel produces 94.15% accuracy, 58.03% precision, 40.33% recall, and 43.24% f-measure using 976 review data.

Keywords: Sentiment Analysis, Ehadrah, Google Play Store, Support Vector Machine

1. PENDAHULUAN

Di era digital saat ini, aplikasi telah menjadi komponen integral dalam kehidupan sehari-hari. Berbagai kategori aplikasi sangat fokus pada pengalaman pengguna, dan jumlah pengguna aplikasi seluler terus bertambah, khususnya di platform distribusi terkemuka seperti *Google Play Store*. *Google Play Store* merupakan toko aplikasi resmi untuk ponsel pintar Android, yang menampung banyak aplikasi yang memenuhi berbagai kebutuhan dan preferensi pengguna. *Google Play Store* memiliki sistem penilaian dan ulasan yang tersedia memungkinkan pengguna untuk memberikan masukan dan evaluasi terhadap aplikasi yang telah mereka gunakan[1]. Maka dari itu, penting untuk memahami tingkat kepuasan konsumen terhadap aplikasi tersebut.

Namun, ulasan yang diberikan pengguna sering kali tidak mencerminkan penilaian secara keseluruhan. Hal ini menimbulkan kebutuhan untuk memahami tingkat kepuasan konsumen secara lebih mendalam melalui analisis sentimen[2]. Analisis sentimen adalah teknik yang digunakan untuk mengekstraksi informasi dari data teks, menggunakan metode pemrosesan bahasa alami (NLP) dan pembelajaran mesin untuk mendeteksi dan mengkategorikan pandangan atau emosi dalam teks tertulis[3]. Teknik ini memungkinkan kita untuk menentukan apakah ulasan pengguna cenderung positif, negatif, atau netral[4].

Salah satu aplikasi yang membutuhkan analisis sentimen mendalam adalah eHadrah, aplikasi yang menyediakan konten musik hadrah, seni musik tradisional Islami. Berdasarkan data dari *Google Play Store* per 7 Januari 2024, eHadrah telah diunduh lebih dari 1 juta kali dan memiliki rating 4.3 dengan total 796 ulasan pengguna yang terdokumentasi. Ulasan ini beragam, mulai dari masukan positif hingga keluhan dan saran untuk pengembangan aplikasi [5]. Meskipun eHadrah memiliki rating yang cukup tinggi yaitu 4.3, ulasan yang diberikan pengguna menunjukkan beragam pendapat yang signifikan. Banyak ulasan yang mengandung keluhan atau saran yang menunjukkan adanya aspek yang perlu diperbaiki dalam aplikasi. Diskrepansi ini menimbulkan

pertanyaan mengenai sejauh mana rating angka mencerminkan kepuasan pengguna secara keseluruhan [6]. Ulasan pengguna di Google Play Store sering kali tidak terstruktur, membuat sulit untuk mengidentifikasi masalah spesifik dan area yang perlu pengembangan. Pengguna mungkin memberikan rating tinggi namun mencantumkan keluhan di dalam ulasan teks, atau sebaliknya, memberikan rating rendah dengan alasan yang kurang jelas. Ketidajelasan ini menyulitkan pengembang dalam mendapatkan wawasan yang akurat dari umpan balik pengguna[7]. mengingat banyaknya jumlah ulasan pengguna yang tidak terstruktur dan tidak terkategori, diperlukan suatu teknik untuk memperoleh data evaluasi, yaitu dengan melakukan analisis sentimen terhadap kumpulan data ulasan pengguna[8]

Penelitian ini bertujuan untuk mengatasi kesenjangan antara ulasan pengguna dan penilaian aplikasi yang tersedia. Dengan menggunakan teknik analisis sentimen, ulasan pengguna dapat dikategorikan menjadi sentimen positif, negatif, atau netral, yang memberikan wawasan lebih dalam mengenai pengalaman pengguna dan area yang memerlukan perbaikan. Selain itu, penelitian ini akan membandingkan dua *kernel* pada algoritma (SVM), yaitu *kernel* Linear dan *kernel Radial Basis Function* (RBF), untuk menentukan *kernel* mana yang memberikan akurasi tertinggi dalam analisis sentimen [9].

Tujuan utama dari penelitian ini adalah untuk mengevaluasi sentimen pengguna terhadap aplikasi eHadhrah di *Google Play Store* dengan menggunakan algoritma SVM. Penelitian ini akan membandingkan akurasi dari *kernel* Linear dan RBF pada algoritma SVM untuk menentukan *kernel* yang paling efektif dalam mengklasifikasikan sentimen ulasan pengguna. Hasil penelitian ini diharapkan dapat memberikan wawasan bagi pengembang aplikasi dalam meningkatkan kualitas aplikasi berdasarkan umpan balik pengguna yang lebih akurat dan terstruktur [10].

Dalam penulisan yang dilakukan oleh Andriani Nurian dan Betha Nurina Sari pada tahun 2023 yang bertajuk “Analisis Sentimen *Review* Pengguna Aplikasi *Google Play* Menggunakan *Naïve Bayes*” meneliti tentang analisis sentimen yang sudah ada. Tujuan dari analisis ini ialah untuk memahami sentimen dan perspektif pengguna mengenai aplikasi Dana, dan untuk menawarkan wawasan berharga kepada pengembang untuk meningkatkan kualitas layanan dan fitur yang ditawarkan. Berdasarkan hasil analisis, lebih dari 60% pengguna memberikan penilaian yang baik terhadap aplikasi Dana. Di sisi lain, ulasan negatif hanya berjumlah sekitar 12%, sedangkan peringkat netral berjumlah sekitar 28% [1].

Fanka Angelina Larasati, Dian Eka Ratnawati, dan Buce Trias Hanggara melakukan penelitian lebih lanjut dengan judul “Analisis Sentimen Tinjauan Penerapan Dana Menggunakan Metode *Random Forest*”. Penulisan ini menggunakan teknik *random forest* untuk mengklasifikasikan sentimen ke dalam tiga kategori: positif, negatif, dan netral. Selain itu, metrik evaluasi seperti akurasi, perolehan, presisi, dan pengukuran-f juga digunakan. Percobaan dilakukan dengan menggunakan 1.354 titik data untuk menguji dampak jumlah pohon dan kedalaman pohon. Data dibagi menjadi beberapa kelas, dengan masing-masing kelas berisi 250 titik data. Nilai *presisi*, *recall*, *F1-Score*, dan *akurasi* ditentukan melalui pengujian dan analisis. Data latih dibandingkan dengan data uji dengan menggunakan *rasio* 80%:20%. *Presisi*, *recall*, *F1-Score*, dan *akurasi* semuanya ditemukan sebesar 84%. Temuan ini diperoleh dengan menggunakan kedalaman pohon 65 dan total 400 pohon[11].

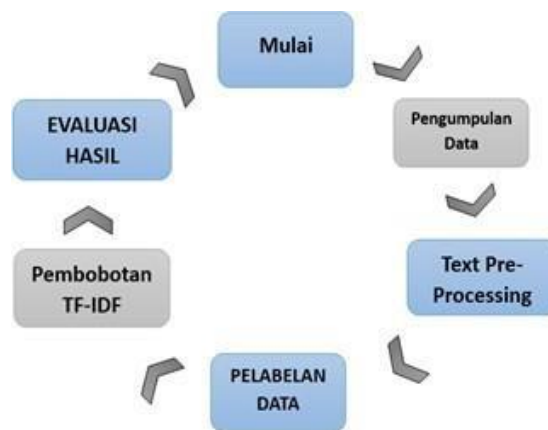
Mengetahui emosi pengguna aplikasi mypertamina yang dapat dikategorikan menjadi dua golongan, yaitu baik dan negatif. Kumpulan data terdiri dari 5.722 titik data dengan label positif dan negatif. Analisis sentimen mypertamina menghasilkan *akurasi* 87%, *presisi* 86%, *recall* 90%, dan *f1-score* 87% dengan menggunakan algoritma NBC [12].

Banyaknya evaluasi pengguna memerlukan penerapan metodologi yang dapat mengklasifikasikan ulasan menjadi baik atau negatif. Teknik klasifikasi yang digunakan ialah *Stochastic Gradient Descending* (SGD). Temuan kategorisasi Tokopedia menunjukkan *akurasi* 84%, *presisi* 87%, dan *recall* 90%. Untuk *Shopee*, hasilnya menunjukkan *akurasi* 66%, *presisi* 65%, dan *memori* 66% [13].

Meskipun berbagai penelitian telah dilakukan menggunakan metode yang berbeda, masih terdapat kesenjangan dalam memahami sentimen pengguna secara mendalam pada aplikasi tertentu seperti eHadhrah. Oleh karena itu, penelitian ini berfokus pada penggunaan algoritma SVM dengan dua jenis *kernel* untuk memberikan analisis yang lebih akurat dan relevan bagi pengembangan aplikasi eHadhrah[9].

2. METODE PENELITIAN

Metodologi penulisan yang digunakan dalam penulisan ini meliputi pengumpulan data, pengolahan teks, penerapan pembobotan TF-IDF, dan evaluasi hasil yang diperoleh. Gambar 1 menggambarkan tahapan penulisan berurutan yang akan dijalankan.



Gambar 1. Tahapan Penulisan

2.1 Pengumpulan Data

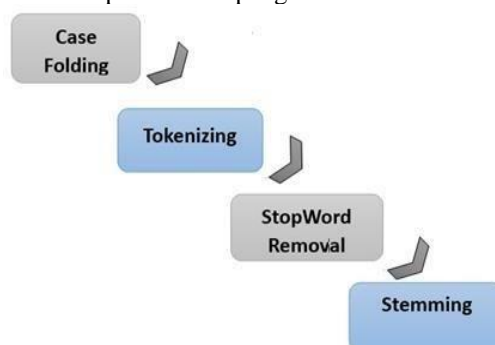
Pengumpulan data untuk penelitian ini dilakukan dengan menggunakan teknik *web scraping* pada *Google Play Store*. *Scraping* dilakukan dengan *google-play-scraper* menggunakan *Python*[5]. *Scraping* mengacu pada ekstraksi informasi dari ulasan pengguna di platform seperti *Google Play Store*, *Apple App Store*, atau situs lain yang menawarkan ulasan pengguna [10]. Data dikumpulkan pada tanggal 7 Januari 2024 hingga tanggal 09 Maret 2024, dan frase pencarian yang diperlukan dimasukkan dalam bentuk *com.unuja.ft.ehadrah* untuk mendapatkan jumlah data yang maksimal. Data yang dikumpulkan berkaitan dengan penilaian terhadap aplikasi ehadrah yang tersedia di *Google Play Store*. Data yang didapatkan berjumlah 796 data dan disimpan dalam format CSV. Tabel 1 menampilkan hasil data *scraping*.

Tabel 1. Hasil *Scraping* Data

<i>userName</i>	<i>score</i>	<i>at</i>	<i>content</i>
Wahyu Hidayat	5	5/11/2024 1:38:56 PM	Aku mau belajar aplikasi eHadrah karena aku di posisi bus
Hasan Andika	5	10/22/2023 8:25:19 AM	Wahhh aplikasinya bagus sekali, sayasangat suka, mantap bos ku, EHadrah paling keren.
Dmsadyp	5	9/28/2023 4:19:36 PM	eHadrah mantap untuk developer updatelagi apknya supaya tambah keren.
MUHAMMAD SAYYIDINA AKBAR	5	9/17/2023 4:50:18 AM	Bagus banget nih Apk nya bisa buat latihan hadroh, matep sihh Btw kerja baguseHadrah banyak banyakin buat update ya makasih eHadrah

2.2 Text pre-processing

Pra-pemrosesan teks mengacu pada manipulasi dan analisis teks menggunakan algoritma dan pendekatan komputer. Tujuannya ialah untuk memahami, menganalisis, dan memanipulasi data yang dicakup oleh bahasa tertulis. Pemrosesan mencakup banyak tugas. Pada tahap *text pre-processing*, teks yang diambil akan menjalani ekstraksi data untuk mengidentifikasi informasi yang relevan dan menghilangkan materi yang tidak relevan [14]. Gambar 2 menampilkan hasil pengolahan data.



Gambar 2. Hasil Processing Data

Gambar 2 mengilustrasikan langkah-langkah berurutan yang terlibat dalam pra-pemrosesan teks, yang meliputi *case folding*, *tokenisasi*, penghapusan *stopword*, *stemming*, dan analisis sentimen. Proses pemrosesan teks melibatkan langkah-langkah berikut:

1. *Case folding* ialah mengubah semua huruf dalam teks tertentu menjadi huruf kecil. Tujuannya ialah untuk menjamin keseragaman penggunaan huruf kapital dalam teks.
2. *Tokenisasi* ialah tindakan membagi teks menjadi bagian-bagian kecil yang disebut token, yang dapat berupa kata, frasa, atau bahkan karakter individual. Tujuan utamanya ialah untuk memfasilitasi proses analisis teks[15].
3. Reduksi *stopword* melibatkan penghapusan istilah-istilah yang sering digunakan yang tidak memiliki nilai informasi signifikan dalam analisis teks. Kata-kata tersebut terdiri dari “dan”, “atau”, “ialah”, dan seterusnya.
4. *Stemming* ialah tata cara mereduksi kata menjadi bentuk dasarnya yang disebut *stem*. Hal ini dicapai dengan memperpendek bagian akhir sebuah kata untuk mencapai bentuk akar kata yang lebih sering digunakan.

2.3 Pelabelan Data

Pelabelan data melibatkan pemberian label atau anotasi pada data yang belum diproses untuk menemukan informasi terkait yang dapat digunakan untuk pelatihan model pembelajaran mesin. Pelabelan data sangat penting dalam pembelajaran yang diawasi, karena model dilatih menggunakan data berlabel untuk membuat prediksi atau mengelompokkan data. Proses pelabelan data ini menggunakan *Vader Lexicon*. *Vader Lexicon* ialah kompilasi kata-kata yang telah diberi nilai sentimen, yang dapat berupa baik, negatif, atau netral. *Vader Lexicon* sangat mahir dalam menganalisis konten yang berasal dari media sosial, yang sering kali menyertakan emotikon, bahasa gaul, dan akronim[16].

2.4 Pembobotan TF-IDF

TF-IDF ialah akronim yang singkatan dari "*Term Frekuensi-Invers Dokumen Frekuensi*". Analisis dan pengelompokan teks menggunakan teknik ini. Metode ini menilai signifikansi suatu kata dalam suatu dokumen atau kumpulan dokumen dengan mengalikan frekuensi kemunculan kata tersebut dalam dokumen (TF) dengan nilai kebalikan dari frekuensi kemunculan kata tersebut di seluruh kumpulan dokumen (IDF). Metodologi ini memfasilitasi identifikasi kata-kata yang paling relevan dalam konteks atau pokok bahasan tertentu. Pentingnya sebuah kata dalam sebuah dokumen atau sekelompok dokumen meningkat seiring dengan meningkatnya nilai TF-IDF-nya. Rumus TF-IDF berasal dari kutipan [17] dan diwakili oleh persamaan 1.

$$TF.IDF_{std}(t) = \frac{tf_d^t}{d} \times \log_{df}^n t \quad (1)$$

t : kata atau *term* yang sedang dihitung nilainya.

tf_d^t : *Term Frequency*, yaitu jumlah kemunculan kata t dalam dokumen d .

n : Total jumlah dokumen dalam korpus.

$\log_{df}^n t$: *Invers Document Frequency*, mengukur seberapa umum atau jarang kata t di seluruh dokumen dalam korpus.

$TF.IDF_{std}(t)$: Nilai TF-IDF standar untuk kata t .

2.5 Evaluasi Hasil

Pada tahap ini, dua jenis *kernel* linear dan RBF akan diterapkan dalam teknik *Support Vector Machine* untuk klasifikasi sentimen pada data tinjauan. Mengukur metrik *akurasi*, *presisi*, *recall*, dan *f-measure* akan memungkinkan seseorang menilai kinerja setiap *kernel* [18]. Persamaan 2 memegang rumus *kernel* linier; persamaan 3 memegang rumus *kernel* RBF.

a) *Kernel* Linear

$$k(x_i, x) = x_i \cdot x \quad (2)$$

$k(x_i, x)$: Fungsi *kernel* antara dua vektor fitur x_i dan x .

$x_i \cdot x$: Hasil perkalian dot *inner product* antara dua vektor fitur x_i dan x . Dalam konteks ini, *kernel* linear sebenarnya hanya mengukur kesamaan linear antara dua vektor.

b) *Kernel* RBF

$$k(x_i, x) = \exp\left(-\frac{\|x_i - x\|}{2\sigma^2}\right) \tag{3}$$

$k(x_i, x)$: Fungsi *kernel* antara dua vektor fitur x_i dan x .

$\exp\left(-\frac{\|x_i - x\|}{2\sigma^2}\right)$: Fungsi eksponensial yang dihitung berdasarkan jarak antara x_i dan x .

$\|x_i - x\|$: Norma *Euclidean* kuadrat (atau jarak *Euclidean* kuadrat) antara dua vektor fitur x_i dan x . Ini mengukur seberapa jauh dua vektor tersebut dalam ruang fitur.

σ : Parameter yang menentukan "lebar" atau "skala" dari fungsi *kernel*. Nilai ini mempengaruhi seberapa jauh pengaruh dari setiap titik data meluas.

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi analisis, hasil pelaksanaan, pengujian, dan topik pembahasan pada bagian ini. Bagian ini juga mencakup penjelasan bergambar grafis yang didukung oleh tabel dan alat bantu visual lainnya.

3.1 *Text Pre-Processing*

Mengacu pada langkah-langkah penulisan yang disebutkan sebelumnya, bagian ini akan membahas hasil penulisan yang telah dilakukan. Kumpulan data awalnya akan diproses sebelum pengujian dimulai. Dikenal dengan *Case Folding*, pada tahap ini hanya huruf besar ke huruf kecil yang akan diubah. Tabel 2 menunjukkan hasil *Case Folding*.

Tabel 2. Hasil *Case Folding*

<i>Full text</i>	<i>Case Folding</i>
@Fathur media udah bagus banget min, tolong kembangin lagi b..	udah bagus banget min, tolongkembangin lagi b..
@Zikry Ahmad Aplikasinya udah bagus dan mantap min tolong..	aplikasinya udah bagus danmantap min,,tolong ..
@Fahrizal Mansyur Project Saya habis ngiringin majelisan pake aplikasi i..	saya habis ngiringin majelisanpake aplikasi i..
@Fadhil Al-Faqir Pokoknya harus diupdate trus! Fasilitasnya dip..	pokoknya harus diupdate trus!fasilitasnya dip..
@Raisyahardiyantirusli Raisya Bagus banget sih aku suka kalau mau main atau..	bagus banget sih aku sukakalau mau main atau ..

Untuk *processing* selanjutnya yaitu *Tokenizing* dimana setelah melakukan Hasil proses *case folding*, kalimat akan diproses dengan menguraikannya menjadi token-token atau kata-kata. Pada tahap ini akan dibantu dengan *library* NLTK. Tabel 3 menunjukkan hasil *tokenizing*.

Tabel 3. Hasil *Tokenizing*

<i>Case Folding</i>	<i>Tokenizing</i>
udah bagus banget min, tolongkembangin lagi b..	[udah, bagus, banget, min, ,, tolong, kembangin..
aplikasinya udah bagus danmantap min,,tolong ..	[aplikasinya, udah, bagus, dan, mantap, min, ,,
saya habis ngiringin majelisanpake aplikasi i..	[saya, habis, ngiringin, majelisan, pake, apli..
pokoknya harus diupdate trus!fasilitasnya dip..	[pokoknya, harus, diupdate, trus, !, fasilitas..
bagus banget sih aku suka kalaumaui main atau ..	[bagus, banget, sih, aku, suka, kalau, mau, ma..

Selanjutnya kita akan melanjutkan dengan penghapusan *stopword*. Dalam segmen penghapusan *stopword*, hanya kata-kata yang tidak penting yang dipilih dan dihilangkan. Perpustakaan NLTK memberikan bantuan pada langkah ini. Hasil eliminasi *stopword* ditampilkan pada Tabel 4.

Tabel 4. Hasil *StopWord*

<i>Tokenizing</i>	<i>Stopword removal</i>
[udah, bagus, banget, min, ,,tolong, kembang..	[udah, bagus, banget, min, ,, tolong, kembang..
[aplikasinya, udah, bagus, dan,mantap, min, ,..	[aplikasinya, udah, bagus, mantap, min, ,, ,to..
[saya, habis, ngiringin,majelisan,	[habis, ngiringin, majelisan, pake, aplikasi, ..
pake, apli..	
[pokoknya, harus, diupdate,trus, !,	[pokoknya, diupdate, trus, !,
fasilitas..	fasilitasnya, di..
[bagus, banget, sih, aku, suka,kalau, mau, ma..	[bagus, banget, sih, suka, main, gabut, nyanyi..

Selanjutnya yaitu *stemming*, proses *stemming* dilakukan perubahan kata yang berimbuhan menjadi kata dasar. Tahap ini dibantu dengan *library* Sastrawi dan Swifter. Tabel 5 menunjukkan hasil *stemming*.

Tabel 5. Hasil *Steming*

<i>Stopword removal</i>	<i>Stemming</i>
[udah, bagus, banget, min, ,,tolong, kembang..	[udah, bagus, banget, min, , tolong, kembangin..
[aplikasinya, udah, bagus, mantap, min, ,, ,to..	[aplikasi, udah, bagus, mantap, min, , tolong,..
[habis, ngiringin, majelisan, pake, aplikasi, ..	[habis, ngiringin, majelis, pake, aplikasi, , ..
[pokoknya, diupdate, trus, !, fasilitasnya, di..	[pokok, diupdate, trus, , fasilitas, banyak, l..
[bagus, banget, sih, suka, main, gabut, nyanyi..	[bagus, banget, sih, suka, main, gabut, nyanyi..

3.2 Pelabelan Data

Tahapan selanjutnya menentukan lapisan output yaitu untuk mengetahui hasil positif, negatif dan juga netral dengan menggunakan kamus *vader lexicon*. Gambar 3 menunjukkan hasil *code vader lexicon*.

```
!pip install nltk
import nltk
from nltk.sentiment import SentimentIntensityAnalyzer
import pandas as pd

nltk.download('punkt')
nltk.download('vader_lexicon')
```

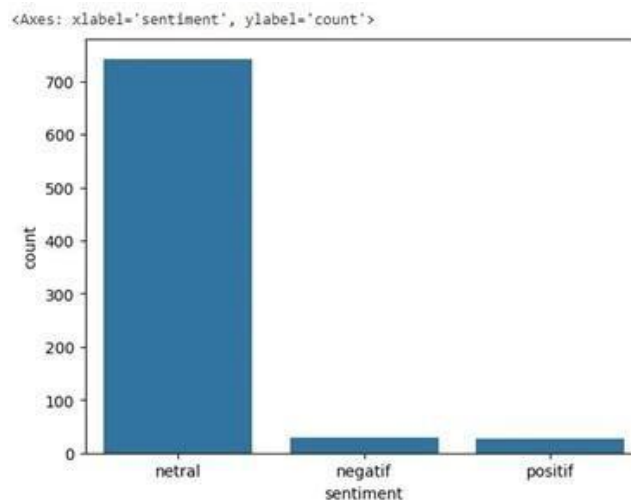
Gambar 3. *Library Vader Lexicon*

Tujuan dari *code vader lexicon* pada Gambar 3 ialah untuk melakukan analisis sentimen yang cepat dan akurat pada teks. *Vader lexicon* sangat efektif dalam menganalisis teks dari media sosial karena kemampuannya untuk mengenali sentimen positif, negatif, dan netral dalam teks serta mengukur intensitas sentimen. Hasil dari pelabelan sentimen dapat dilihat pada Gambar 4, di mana nilai sentimen yang lebih besar dari 0 menunjukkan hasil yang positif, nilai yang kurang dari 0 menunjukkan hasil yang negatif, dan nilai sentimen yang sama dengan 0 menunjukkan hasil yang netral.

	content	sentiment_score	sentiment
0	udah bagus banget min, tolong kembangin lagi b...	0.3400	positif
1	Aplikasinya udah bagus dan mantap min,,tolong ...	0.0000	netral
2	Saya habis ngiringin majelisan pake aplikasi i...	-0.2732	negatif
3	Pokoknya harus diupdate trus! Fasilitasnya dip...	0.0000	netral
4	Bagus banget sih aku suka kalau mau main atau ...	-0.4215	negatif

Gambar 4. Hasil Dari Pelabelan Sentimen

Total keseluruhan dari setiap kelas sentimen (positif, negatif, netral) dalam dataset ditampilkan dalam visualisasi menggunakan diagram batang menggunakan *library seaborn*. Gambar 5 menunjukkan hasil *class distribution*.



Gambar 5. Class Distribution

Grafik pada Gambar 5 mengilustrasikan hasil analisis sentimen berdasarkan kuantitas ulasan aplikasi Ehadrah, yang diklasifikasikan ke dalam tiga kategori berbeda: positif, negatif, dan netral. Hasil sentimen netral lebih tinggi dibandingkan dengan sentimen positif dan negatif, sedangkan hasil sentimen positif dan negatif sama.

3.3 Pembobotan TF-IDF

Berikut ini ialah prosedur pembobotan kata dengan TF-IDF. Pada tahap ini kata-kata diberi bobot berdasarkan hasil *stemming* dengan menggunakan pendekatan *Term Inverse Document Frekuensi* (TF-IDF). Metode TF-IDF digunakan untuk menentukan frekuensi kemunculan suatu kata dalam suatu dokumen. Level ini difasilitasi oleh perpustakaan *Sklearn*. Gambar 6 menampilkan hasil pembobotan TF-IDF.

```
Kata: bagus, Kemunculan: 78.93210344018661
Kata: banget, Kemunculan: 26.792505750281478
Kata: aplikasi, Kemunculan: 23.18112702589807
Kata: suara, Kemunculan: 23.055003965923294
Kata: mantap, Kemunculan: 19.85438106470518
Kata: game, Kemunculan: 17.85415804381713
Kata: keren, Kemunculan: 13.11431866741933
Kata: hadroh, Kemunculan: 13.11175180900071
Kata: suka, Kemunculan: 12.34561534493543
Kata: apk, Kemunculan: 12.293454350490723
```

Gambar 6. Hasil Pembobotan TF-IDF

Dari hasil pembobotan kata pada Gambar 6 dengan menggunakan TF-IDF, kata “bagus” paling banyak muncul di dalam data ulasan ehadrah dan menghasilkan kemunculan sejumlah 78, kata “banget” menghasilkan kemunculan 26, kata “aplikasi” menghasilkan kemunculan 23, kata “suara” menghasilkan kemunculan 23, kata “mantap” menghasilkan kemunculan 19, kata “game” menghasilkan kemunculan 17, kata “keren” menghasilkan kemunculan 13, kata “hadroh” menghasilkan kemunculan 13, kata “suka” menghasilkan kemunculan 12, dan kata “apk” menghasilkan kemunculan 12.

3.4 Evaluasi Hasil

Langkah selanjutnya dalam proses ini melibatkan pengklasifikasian ulasan pengguna menggunakan algoritma SVM, dengan bantuan dari modul *Scikit-Learn*. Teknik klasifikasinya menggunakan dataset latih dan dataset uji dengan rasio 80% hingga 20%. Percobaan dilakukan menggunakan *kernel* RBF dan *kernel* Linier. Tabel 6 menampilkan hasil penilaian RBF dan model Linier.

Tabel 6. Hasil Evaluasi SVM

<i>Kernel</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
RBF	94,15%	58,03%	40,33%	43,24%
Linear	95,46%	81,83%	55,61%	62,82%

Tabel 6 menampilkan metrik kinerja untuk kernel yang berbeda. *Kernel* Linear mencapai akurasi 95,46%, *presisi* sebesar 81,83%, *recall* sebesar 55,61%, dan *f-measure* sebesar 62,82%. Sedangkan *kernel* RBF mencapai akurasi 94,15%, *presisi* 58,03%, *recall* 40,33%, dan *f-measure* 43,24% dengan menggunakan 796 data. Dan hasil yang lebih baik yaitu *kernel* Linear.

4. KESIMPULAN

Penelitian ini membahas analisis sentimen pada ulasan aplikasi Ehadrah di Google Play Store menggunakan metode Support Vector Machine (SVM). Data ulasan diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa sentimen netral mendominasi, dengan proporsi sentimen positif dan negatif yang lebih sedikit dan hampir seimbang. Dalam analisis lebih lanjut, kata-kata yang sering muncul dalam ulasan positif adalah “baik,” “menyenangkan,” dan “hebat,” sedangkan dalam ulasan negatif sering muncul kata “kurang,” “tolong,” dan “tolong.” Ulasan netral sering mengandung kata “aplikasi ini,” “buat,” “bass,” dan “saran”.

Untuk mendapatkan hasil yang akurat, digunakan teknik SVM dengan pemilihan fitur TF-IDF. Dalam pengujian, kernel linier mencapai akurasi sebesar 95,46%, presisi 81,83%, recall 55,61%, dan f-measure 62,82%. Kernel RBF mencapai akurasi 94,15%, presisi 58,03%, recall 40,33%, dan f-measure 43,24%. Dari hasil tersebut, dapat disimpulkan bahwa penggunaan kernel linier dalam SVM memberikan kinerja yang lebih baik dibandingkan kernel RBF dalam analisis sentimen ulasan aplikasi Ehadrah. Penelitian ini menyarankan penggunaan metodologi alternatif untuk meningkatkan akurasi dan presisi analisis sentimen di masa depan.

UCAPAN TERIMAKASIH

Kami mengucapkan terima kasih yang setulus-tulusnya kepada semua pihak yang telah memberikan bantuan dan upayanya hingga terselesaikannya makalah ini. Terima kasih saya sampaikan kepada Ibu Ratri Enggar Pawening dan Bapak Zainal Arifin atas bimbingannya, saran-sarannya yang bermanfaat, serta ilmu dan kompetensinya yang luas.

DAFTAR PUSTAKA

- [1] A. Nurian, “Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes,” *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 829–835, 2023, doi: 10.23960/jitet.v11i3s1.13348.
- [2] T. Tinaliah and T. Elizabeth, “Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 4, pp. 3436–3442, 2022, doi: 10.35957/jatisi.v9i4.3586.
- [3] A. D. Adhi Putra, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [4] A. Novantika, “Analisis sentimen ulasan pengguna aplikasi video conference google meet menggunakan metode svm dan logistic regression,” *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 808–813, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [5] R. Tineges, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM),” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [6] M. R. U. Pulungan, D. E. Ratnawati, and B. Rahayudi, “Analisis Sentimen Ulasan Aplikasi PeduliLindungi dengan Metode Random Forest,” *J. Pengemb. Teknol. Inf. dan Ilmu Komun.*, vol. 6, no. 9, pp. 4378–4385, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/11582/5142>
- [7] D. Nurwahidah, G. Dwilestari, N. Dienwati Nuris, and R. Narasati, “Analisis Sentimen Data Ulasan Pengguna Aplikasi Google Kelas Pada Google Play Store Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3673–3678, 2024, doi: 10.36040/jati.v7i6.8245.
- [8] I. Maulana, W. Apriandari, and A. Pambudi, “Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi Mypertamina Menggunakan Support Vector Machine,” *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 172–181, 2023, doi: 10.36080/idealism.v6i2.3022.
- [9] D. Safryda Putri and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi Pospay Dengan Algoritma Support

- Vector Machine,” *J. Ilm. Inform.*, vol. 11, no. 01, pp. 32–40, 2023, doi: 10.33884/jif.v11i01.6611.
- [10] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, “Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc,” *Fakt. Exacta*, vol. 14, no. 2, p. 64, 2021, doi: 10.30998/faktorexacta.v14i2.9697.
- [11] B. T. H. Fanka Angelina Larasati, D. Eka Ratnawati, “Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest,” ... *Teknol. Inf. dan ...*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] R. Maulana, A. Voutama, and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC,” *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42–48, 2023, doi: 10.54914/jtt.v9i1.609.
- [13] W. Kurnia, “Sentimen Analisis Aplikasi E-Commerce Berdasarkan Ulasan Pengguna Menggunakan Algoritma Stochastic Gradient Descent,” *J. Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 138–143, 2023.
- [14] M. M. Maarif and N. Setiyawati, “Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine,” *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 454, 2024, doi: 10.35889/progresif.v20i1.1614.
- [15] S. Fathoniah and C. Rozikin, “Analisis Sentimen Masyarakat terhadap Teroris dalam Media Sosial Twitter menggunakan NLP,” *J. Ilm. Wahana Pendidik.*, vol. 2022, no. 13, pp. 412–419, 2022.
- [16] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, “Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile,” *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [17] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naive Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [18] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, “Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm,” *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022.