

# Optimalisasi Random Forest dan Support Vector Machine dengan Hyperparameter GridSearchCV untuk Analisis Sentimen Ulasan PrimaKu

Titik Misriati, Riska Aryanti\*

Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia

Email: <sup>1</sup>titik.tmi@bsi.ac.id, <sup>2,\*</sup>riska.rts@bsi.ac.id

Email Penulis Korespondensi: riska.rts@bsi.ac.id

Submitted: 14/06/2024; Accepted: 31/07/2024; Published: 31/07/2024

**Abstrak**—Aplikasi PrimaKu telah menjadi pelopor dalam bidang kesehatan digital sejak tahun 2017. Melalui aplikasi ini orang tua dapat memantau kesehatan dan tumbuh kembang anak secara rutin dan berkesinambungan. PrimaKu juga memiliki aliansi formal dengan Ikatan Dokter Anak Indonesia (IDAI) untuk mempromosikan kesehatan anak di Indonesia. Aplikasi ini dapat diunduh melalui Google Play Store. Google Play Store memiliki fitur yang memungkinkan pengguna meninjau aplikasi sebelum mengunduh. Analisis sentimen digunakan untuk membedakan ulasan positif dan negatif dari pengguna yang sudah memberikan tinjauan sehingga dapat dilakukan evaluasi terhadap layanan yang diberikan. Penelitian ini bertujuan untuk melakukan analisis sentimen ulasan pengguna terhadap aplikasi PrimaKu menggunakan algoritma Random Forest (RF) dan Support Vector Machine (SVM) dengan pembobotan TF-IDF. Optimalisasi dilakukan menggunakan hyperparameter untuk meningkatkan kinerja model algoritma Random Forest dan SVM. Data yang digunakan terdiri dari 2.293 ulasan paling relevan yang dikumpulkan dari Google Play Store. Model yang paling efektif untuk Random Forest dan Support Vector Machine dipilih dengan menyesuaikan hyperparameter menggunakan GridSearchCV. Hasil penelitian ini menunjukkan bahwa Random Forest mempunyai tingkat keberhasilan yang lebih tinggi dalam mengklasifikasikan data ulasan pengguna PrimaKu dengan akurasi sebesar 89%, presisi sebesar 88%, recall sebesar 81%, dan F1-Score sebesar 85%.

**Kata Kunci:** Random Forest; Support Vector Machine; Gridsearchcv; Hyperparameter; Analisis Sentimen

**Abstract**—PrimaKu App has been a pioneer in the field of digital health since 2017. Through this application, parents can regularly and continuously monitor their children's health and development. PrimaKu also has a formal alliance with the Indonesian Pediatric Association (IDAI) to promote child health in Indonesia. This application can be downloaded through the Google Play Store. Google Play Store has a feature that allows users to review the app before downloading. Sentiment analysis is used to distinguish between positive and negative reviews by users who have provided reviews so that an evaluation of the services provided can be made. This research aims to conduct sentiment analysis of user reviews of the PrimaKu application using Random Forest (RF) and Support Vector Machine (SVM) algorithms with TF-IDF weighting. Optimization was performed using hyperparameters to improve the performance of the Random Forest and SVM algorithms. The data used consisted of the 2,293 most relevant reviews collected from the Google Play Store. The most effective models for the Random Forest and Support Vector Machine were selected by adjusting the hyperparameters using GridSearch CV. The results of this study show that Random Forest has a higher success rate in classifying PrimaKu user review data, with an accuracy of 89%, precision of 88%, recall of 81%, and F1-Score of 85%.

**Keywords:** Random Forest; Support Vector Machine; Gridsearchcv; Hyperparameter; Sentiment Analysis

## 1. PENDAHULUAN

Di era digital saat ini, aplikasi seluler telah menjadi bagian tak terpisahkan dalam kehidupan sehari-hari. PrimaKu merupakan aplikasi populer di kalangan orang tua dan keluarga yang menawarkan beragam fitur untuk memantau kesehatan dan tumbuh kembang anak. Seiring bertambahnya jumlah pengguna, analisis sentimen ulasan pengguna menjadi sangat penting untuk memahami kepuasan, keluhan, dan kebutuhan pengguna. Analisis sentimen adalah proses mengidentifikasi dan mengklasifikasikan opini yang diungkapkan dalam teks, terutama untuk menentukan apakah sikap penulis terhadap topik tertentu bersifat positif, negatif, atau netral [1].

Salah satu tantangan utama dalam analisis sentimen adalah memilih dan mengoptimalkan algoritma yang paling efektif untuk mengklasifikasikan teks ulasan secara akurat. Metode Random Forest dan Support Vector Machine (SVM) adalah dua algoritma yang sering digunakan dalam tugas-tugas klasifikasi teks karena keefektifannya dan efisiensinya [2]. Namun, masing-masing algoritma memiliki kelebihan dan kekurangan yang berbeda, dan kinerja algoritma dapat bervariasi tergantung pada karakteristik data yang digunakan. Oleh karena itu, diperlukan penelitian lebih lanjut untuk mengoptimalkan kedua metode ini dalam konteks analisis sentimen ulasan aplikasi PrimaKu.

Random Forest adalah metode dalam analisis yang terdiri dari beberapa pohon keputusan sebagai classifier [3]. Random forest juga sering disebut dengan metode ensemble learning yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting. Di sisi lain, Support Vector Machine (SVM) adalah metode pembelajaran mesin yang kuat yang bekerja dengan menemukan hyperplane optimal yang memisahkan kelas-kelas dalam data [4]. SVM terkenal karena kinerjanya yang baik dalam data dengan dimensi tinggi dan kemampuannya untuk menangani data non-linear melalui kernel trick. Meskipun

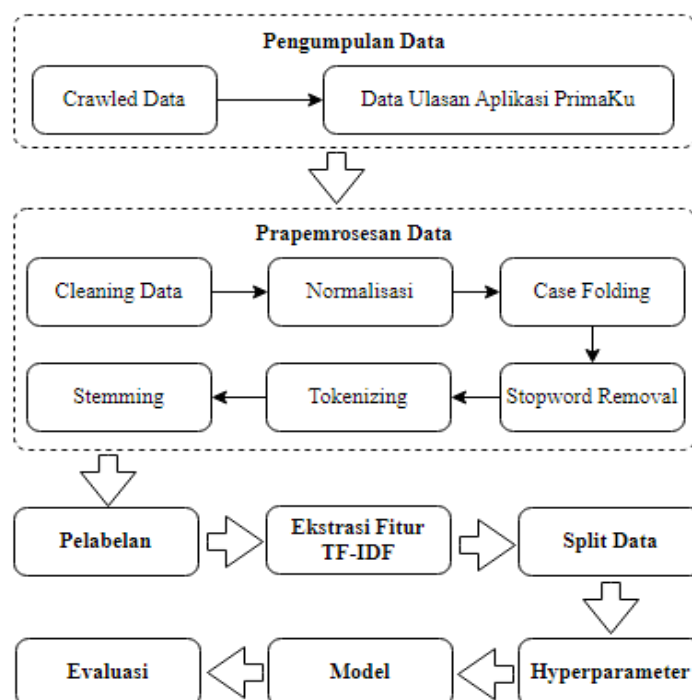
demikian, SVM bisa menjadi lambat saat menangani dataset yang sangat besar, yang menjadi tantangan dalam penerapan pada data ulasan aplikasi yang terus berkembang.

Beberapa penelitian sebelumnya telah dilakukan dalam analisis sentimen aplikasi PrimaKu [5][6] namun dari beberapa penelitian tersebut belum melakukan optimasi pada hyperparameter. Hyperparameter tuning sangat penting dilakukan karena performa random forest dan SVM sangat tergantung dari hyperparameter yang ditetapkan. Hyperparameter yang sering digunakan adalah dengan teknik GridSearchCV. Pengaturan hyperparameter, khususnya melalui GridSearchCV, memiliki peran penting dalam mengoptimalkan kinerja algoritma pembelajaran mesin seperti Random Forest dan Support Vector Machine [7][8][9]. GridSearchCV memungkinkan eksplorasi sistematis berbagai kombinasi hyperparameter, yang pada akhirnya meningkatkan akurasi dan efisiensi dalam tugas-tugas klasifikasi teks [10][11]. Penelitian menunjukkan bahwa penyetelan parameter untuk SVM dan RF dapat secara signifikan meningkatkan kinerja klasifikasi, terutama ketika dikombinasikan dengan teknik seperti pemilihan fitur TF-IDF [12]. Selain itu, studi mendalam tentang penyetelan parameter untuk SVM menekankan pentingnya menemukan rentang optimal hyperparameter untuk mencapai model yang berkinerja baik di berbagai kumpulan data. Dengan menggunakan GridSearchCV dan penyetelan hyperparameter, sehingga dapat menyempurnakan model RF dan SVM dengan lebih efektif, yang pada akhirnya meningkatkan kemampuan prediksi dan kinerja keseluruhan model. Dengan cara menguji secara sistematis berbagai kombinasi hyperparameter, GridSearchCV membantu dalam menyempurnakan model dan meningkatkan akurasi prediksi. Metode ini sangat berguna saat penyetelan manual dilakukan dalam banyak kombinasi parameter yang mungkin terjadi, karena metode ini mengidentifikasi konfigurasi yang paling optimal secara otomatis dan efisien untuk model yang digunakan [13].

Beberapa penelitian sebelumnya yang telah dilakukan dengan menggunakan hyperparameter tuning dengan teknik GridSearchCV menunjukkan bahwa dengan menggunakan teknik ini dapat meningkat kinerja model algoritma [14][15]. Oleh karena itu, penelitian ini bertujuan untuk mengoptimalkan penggunaan random forest dan SVM dalam analisis sentimen ulasan aplikasi PrimaKu, serta mengevaluasi kinerja kedua model tersebut dengan menggunakan teknik GridSearchCV, yang merupakan metode pencarian parameter yang paling efisien dan otomatis, untuk mencari parameter terbaik bagi kedua model ini. GridSearchCV memungkinkan untuk menentukan kombinasi parameter yang paling optimal secara sistematis melalui validasi silang. Melalui penelitian ini, diharapkan dapat diperoleh model-model yang memiliki kinerja terbaik dalam mengklasifikasikan sentimen ulasan aplikasi PrimaKu. Hal ini akan membantu pengembang dalam memahami umpan balik pengguna mereka dengan lebih baik, serta meningkatkan pengalaman pengguna secara keseluruhan.

## 2. METODOLOGI PENELITIAN

Tahapan penelitian dilakukan sesuai dengan Gambar 1, dimulai dari pengumpulan data, prapemrosesan data, pelabelan, ekstraksi fitur, split data, hyperparameter, pemodelan dengan algoritma random forest dan SVM, dilanjutkan dengan evaluasi terhadap model.



**Gambar 1.** Tahapan Penelitian

**a. Pengumpulan Data**

Tahapan pertama dalam penelitian ini adalah pengumpulan data ulasan dari aplikasi PrimaKu. Data ulasan ini dikumpulkan melalui scraping[3] dari platform di mana aplikasi PrimaKu tersedia yaitu Google Play Store. Setiap ulasan akan mencakup teks ulasan, rating, dan metadata lainnya seperti tanggal ulasan dan identitas pengguna, dapat dilihat pada tabel 1. Data ulasan sebanyak 2293 dengan empat atribut, yaitu:

1. **userName**

Atribut ini menyimpan nama pengguna yang memberikan ulasan terhadap aplikasi PrimaKu. Nama pengguna ini bisa berupa nama asli atau nama samaran (username) yang digunakan di platform tempat ulasan tersebut diposting.

2. **Score**

Atribut ini menyimpan nilai rating yang diberikan oleh pengguna terhadap aplikasi PrimaKu. Rating ini berupa angka dalam skala 1 hingga 5, di mana 1 menunjukkan ketidakpuasan yang sangat besar dan 5 menunjukkan kepuasan yang sangat tinggi.

3. **At**

Atribut ini menyimpan tanggal dan waktu ketika ulasan tersebut diposting. Informasi ini penting untuk melacak kapan ulasan tersebut diberikan dan untuk analisis tren sentimen dari waktu ke waktu.

4. **Content**

Atribut ini menyimpan teks ulasan yang ditulis oleh pengguna. Teks ulasan ini berisi opini atau feedback dari pengguna mengenai pengalaman mereka dengan aplikasi PrimaKu.

**Tabel 1.** Data Ulasan Aplikasi PrimaKu

userName	score	at	content
Aulia	5	2024-05-24	Update sudah bisa ubah data pertumbuhan 👍
Rahmawati		03:27:41	
upi supiyatin	3	2024-04-29	baru mau coba aplikasi saya kasih 3 dlu sementara
		05:42:39	
lusi handayani	1	2024-02-10	Tidak pernah di berikan kode Verifikasi.. Sudah sering sekali
		14:56:37	mencoba tapi hasilnya nihil
Didin Nuryati	5	2023-10-12	Sebagai portofolio kesehatan dan tumbang anak yang berguna
		04:38:04	sampai dewasa kelak

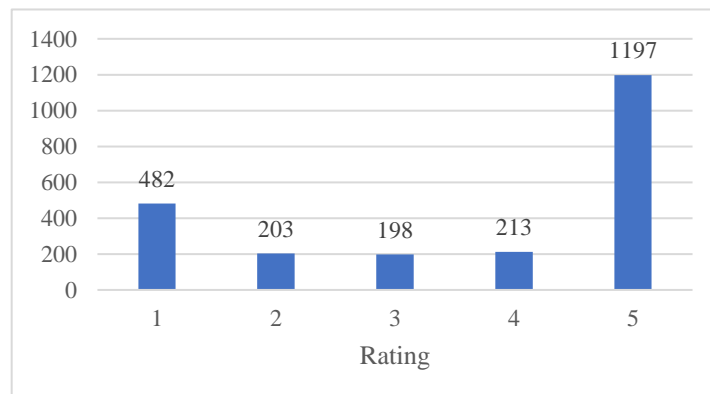
**b. Prapemrosesan Data**

Setelah data terkumpul, langkah berikutnya adalah pra-pemrosesan data. Tahapan ini meliputi beberapa langkah, yaitu:

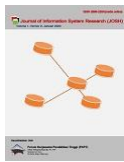
1. Cleaning data dilakukan dengan penghapusan karakter atau simbol yang tidak diinginkan, seperti tanda baca, angka, dan karakter khusus, dilakukan untuk mengurangi noise dalam data [16].
2. Normalisasi data yang dilakukan berupa menormalkan kata-kata gaul [17] dan menghapus kata yang mengandung maksimal tiga huruf [18].
3. Case folding mengubah teks ulasan menjadi huruf kecil (lower case) [19][20]
4. Stopword removal untuk menghapus kata sambung seperti dan, yang.
5. Tokenisasi untuk memisahkan teks ulasan menjadi token-token kata [21].
6. Stemming untuk mengubah kata ke bentuk dasarnya untuk menyederhanakan analisis [22].

**c. Pelabelan**

Untuk melakukan analisis sentimen, ulasan-ulasan tersebut perlu diberi label sebagai positif, negatif, atau netral. Labeling dapat dilakukan menggunakan metode semi-otomatis dengan memanfaatkan skor rating sebagai panduan. Ulasan dengan rating 4-5 dianggap positif, rating 1-3 dianggap negatif. Gambar 2 menunjukkan detail rating dari aplikasi PrimaKu.



**Gambar 2.** Rating Aplikasi PrimaKu



- d. Ekstaksi Fitur TF-IDF  
Langkah berikutnya adalah ekstraksi fitur dari teks ulasan. Metode yang digunakan adalah Term Frequency-Inverse Document Frequency (TF-IDF), yang mengukur seberapa penting sebuah kata dalam satu dokumen relatif terhadap seluruh korpus. Fitur-fitur ini kemudian akan digunakan sebagai input untuk model klasifikasi [23].
- e. Split Data  
Data yang telah diproses kemudian dibagi menjadi dua set, yaitu set pelatihan (training set) dan set pengujian (testing set). Pembagian dilakukan dengan rasio 80:20, di mana sebagian besar data digunakan untuk melatih model dan sisanya untuk menguji kinerja model [24].
- f. Hyperparameter  
Hyperparameter merupakan variabel yang nilainya mempengaruhi proses pembelajaran dan mempengaruhi parameter model [25]. GridSearchCV adalah teknik yang digunakan untuk mengoptimalkan hyperparameter. Teknik ini mempersiapkan algoritma untuk setiap kombinasi hyperparameter terbaik. Kumpulan nilai hyperparameter optimal dari GridSearchCV kemudian digunakan dalam model Random Forest dan Support Vector Machine.
- g. Model  
Pada tahap ini, dua model yang digunakan, yaitu:
  1. Random Forest  
Random forest merupakan evolusi dari algoritma decision tree, dimana setiap pohon keputusan dilatih menggunakan data terpisah dan setiap atribut dipilih secara acak setiap kali pohon tersebut dilatih. Kelebihan algoritma ini adalah kemampuan dalam meningkatkan akurasi hasil terutama dalam menangani data yang hilang, serta kemampuan menahan bias data yang salah dan efisiensi pengelolaan penyimpanan data. Selain itu, proses pemilihan fitur yang cermat dapat mengidentifikasi fitur terbaik, sehingga meningkatkan kinerja model klasifikasi secara keseluruhan [26].
  2. Support Vector Machine (SVM)  
Model pembelajaran mesin yang mencari hyperplane optimal untuk memisahkan kelas-kelas dalam data. Pelatihan SVM dilakukan dengan menggunakan kernel linear atau non-linear, tergantung pada karakteristik data [27][28].
- h. Evaluasi  
Model yang telah dilatih dan dioptimalkan kemudian diuji menggunakan data pengujian. Metode evaluasi yang digunakan meliputi:
  - a. Confusion Matrix  
Confusion matrix digunakan untuk melihat distribusi prediksi benar dan salah dari model.
  - b. Akurasi  
Akurasi dilihat berdasarkan persentase ulasan yang diklasifikasikan dengan benar.
  - c. Precision, Recall, dan F1-Score  
Untuk mengevaluasi kinerja model dalam menangani kelas-kelas minoritas.
  - d. ROC-AUC  
Untuk mengukur kemampuan model dalam membedakan antara kelas positif dan negatif.

### 3. HASIL DAN PEMBAHASAN

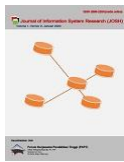
#### 3.1 Hasil

- a. Prapemrosesan Data  
Prapemrosesan data ulasan aplikasi PrimaKu dilakukan dengan cleaning data, normalized data, case folding, stopword removal, tokenizing, dan stemming.
  1. Cleaning Data  
Cleaning data dilakukan untuk membersihkan data yang mengandung emoji, angka, link, maupun simbol. Hasil cleaning data pada Tabel 2 menunjukkan data sebelum dilakukan cleaning data dan setelah dilakukan cleaning data.

**Tabel 2.** Cleaning Data

Sebelum Cleaning Data	Sesudah Cleaning Data
Update sudah bisa ubah data pertumbuhan 📈 baru mau coba aplikasi saya kasih 3 dlu sementara Tidak pernah di berikan kode Verifikasi.. Sudah sering sekali mencoba tapi hasilnya nihil Sebagai portofolio kesehatan dan tumbang anak yang berguna sampai dewasa kelak	Update sudah bisa ubah data pertumbuhan baru mau coba aplikasi saya kasih sementara Tidak pernah di berikan kode Verifikasi Sudah sering sekali mencoba tapi hasilnya nihil Sebagai portofolio kesehatan dan tumbang anak yang berguna sampai dewasa kelak

2. Normalized Data



Data ulasan dinormalisasi untuk dihilangkan kata slang dan kata yang terdiri dari 1-3 huruf. Proses normalisasi ini penting untuk memastikan kualitas dan kejelasan ulasan yang dikumpulkan. Kata slang sering kali bersifat informal dan bisa menimbulkan kebingungan dalam analisis data, sementara kata yang sangat pendek seperti "di", "ke", atau "dan" sering kali tidak memberikan informasi yang signifikan dalam konteks ulasan aplikasi, seperti pada tabel 3. Dengan menghilangkan kata-kata ini, ulasan menjadi lebih fokus dan bermakna, sehingga analisis dapat dilakukan dengan lebih akurat dan hasilnya lebih dapat diandalkan.

Tabel 3. Normalized Data

Sebelum Normalized Data	Sesudah Normalized Data
Update sudah bisa ubah data pertumbuhan baru mau coba aplikasi saya kasih sementara Tidak pernah di berikan kode Verifikasi Sudah sering sekali mencoba tapi hasilnya nihil Sebagai portofolio kesehatan dan tumbang anak yang berguna sampai dewasa kelak	Update sudah bisa ubah data pertumbuhan baru coba aplikasi saya kasih sementara Tidak pernah berikan kode Verifikasi Sudah sering sekali mencoba tapi hasilnya nihil Sebagai portofolio kesehatan tumbang anak yang berguna sampai dewasa kelak

3. Case Folding

Case folding membantu mengurangi kesalahan dalam pengenalan kata yang sama namun dengan penulisan yang berbeda, seperti "PrimaKu" dan "primaku". Tabel 4 menunjukkan dengan melakukan case folding, analisis ulasan aplikasi menjadi lebih akurat karena teks yang diproses sudah berada dalam format yang seragam.

Tabel 4. Case Folding

Sebelum Case Folding	Sesudah Case Folding
Update sudah bisa ubah data pertumbuhan baru coba aplikasi saya kasih sementara Tidak pernah berikan kode Verifikasi Sudah sering sekali mencoba tapi hasilnya nihil Sebagai portofolio kesehatan tumbang anak yang berguna sampai dewasa kelak	update sudah bisa ubah data pertumbuhan baru coba aplikasi saya kasih sementara tidak pernah berikan kode verifikasi sudah sering sekali mencoba tapi hasilnya nihil sebagai portofolio kesehatan tumbang anak yang berguna sampai dewasa kelak

4. Stopword Removal

Stopword adalah kata-kata umum yang sering muncul dalam teks tetapi memiliki sedikit atau tidak ada nilai analisis, seperti "sudah", "bisa", "tetapi". Tabel 5 menunjukkan bahwa menghapus stopwords dapat meningkatkan efisiensi dan akurasi pemrosesan teks dengan mengurangi jumlah kata yang harus dianalisis, sehingga memungkinkan fokus pada kata-kata yang lebih bermakna dan relevan.

Tabel 5. Stopword Removal

Sebelum Stopword Removal	Sesudah Stopword Removal
update sudah bisa ubah data pertumbuhan baru coba aplikasi saya kasih sementara tidak pernah berikan kode verifikasi sudah sering sekali mencoba tapi hasilnya nihil sebagai portofolio kesehatan tumbang anak yang berguna sampai dewasa kelak	update ubah data pertumbuhan coba aplikasi kasih kode verifikasi mencoba hasilnya nihil portofolio kesehatan tumbang anak berguna dewasa kelak

5. Tokenisasi

Tokenisasi dilakukan untuk memisahkan kata pada ulasan aplikasi PrimaKu sehingga lebih mudah dipahami, tabel 6 menunjukkan data sebelum dan setelah dilakukan tokenisasi.

Tabel 6. Tokenisasi

Sebelum Tokenisasi	Sesudah Tokenisasi
update ubah data pertumbuhan coba aplikasi kasih kode verifikasi mencoba hasilnya nihil portofolio kesehatan tumbang anak berguna dewasa kelak	['update', 'ubah', 'data', 'pertumbuhan'] ['coba', 'aplikasi', 'kasih'] ['kode', 'verifikasi', 'mencoba', 'hasilnya', 'nihil'] ['portofolio', 'kesehatan', 'tumbang', 'anak', 'berguna', 'dewasa', 'kelak']

6. Stemming

Stemming data ulasan aplikasi PrimaKu dilakukan dengan mengubah kata-kata ke bentuk kata dasar untuk meningkatkan kualitas data [29] dan keakuratan analisis sentimen [30], seperti pada tabel 7.

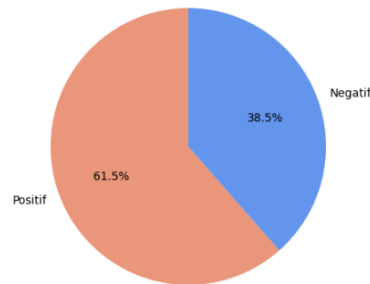


**Tabel 7. Stemming**

Sebelum Stemming	Sesudah Stemming
['update', 'ubah', 'data', 'pertumbuhan']	update ubah data tumbuh
['coba', 'aplikasi', 'kasih']	coba aplikasi kasih
['kode', 'verifikasi', 'mencoba', 'hasilnya', 'nihil']	kode verifikasi coba hasil nihil
['portofolio', 'kesehatan', 'tumbang', 'anak', 'berguna', 'dewasa', 'kelak']	portofolio sehat tumbang anak guna dewasa kelak

**b. Pelabelan**

Dalam analisis sentimen ulasan aplikasi PrimaKu, pelabelan data berdasarkan rating untuk membedakan antara sentimen positif dan negatif. Proses pelabelan ini melibatkan penggunaan rating yang diberikan oleh pengguna pada ulasan. Ulasan dengan rating 4 dan 5 dianggap mencerminkan sentimen positif. Rating ini menunjukkan bahwa pengguna merasa puas atau sangat puas dengan aplikasi, yang tercermin dalam komentar-komentar yang memuji fitur, kegunaan, dan manfaat aplikasi. Sebaliknya, ulasan dengan rating 3, 2, dan 1 dianggap mencerminkan sentimen negatif. Rating ini menunjukkan ketidakpuasan atau masalah yang dihadapi oleh pengguna, seperti bug, kekurangan fitur, atau kinerja aplikasi yang tidak memuaskan. Gambar 3 menunjukkan presentase sentimen ulasan aplikasi PrimaKu.



**Gambar 3. Sentimen Ulasan Aplikasi PrimaKu**

**c. Visualisasi Data**

Visualisasi data ulasan aplikasi Primaku menggunakan wordcloud untuk menampilkan kata-kata yang paling sering muncul dalam ulasan pengguna sebagai acuan analisis ulasan pengguna [31]. Visualisasi data dapat dilihat pada Gambar 4. Kata yang paling dominan adalah 'aplikasi', muncul sebanyak 743 kali, menunjukkan bahwa pengguna sering merujuk langsung ke aplikasi saat memberikan ulasan. Ini bisa mencakup berbagai aspek seperti fitur, antarmuka, atau pengalaman umum dengan aplikasi tersebut.



**Gambar 4. Wordcloud Ulasan Aplikasi PrimaKu**

Kata 'anak' muncul sebanyak 615 kali, yang mengindikasikan bahwa aplikasi ini sangat berfokus pada anak-anak, dalam hal penggunaan atau tujuan pengembangan. 'Data' muncul 567 kali, menandakan pentingnya fitur pengelolaan data yang ditawarkan kepada pengguna.

Kata 'bantu' muncul 484 kali, menyoroti bagaimana aplikasi ini dianggap membantu oleh pengguna. Ini bisa mencakup bantuan dalam hal tumbuh kembang anak. Kata 'bagus', dengan frekuensi 471 kali, menunjukkan banyak ulasan positif mengenai kualitas aplikasi.

Kata 'update' muncul 454 kali, menunjukkan bahwa pengguna sering membahas pembaruan aplikasi, yang bisa mencakup penambahan fitur baru, perbaikan bug, atau peningkatan kinerja. 'Kembang' dan 'tumbuh', masing-masing muncul 320 dan 315 kali, menandakan bahwa aplikasi ini berkaitan erat dengan perkembangan dan pertumbuhan anak.

Kata 'login' muncul 312 kali, mengindikasikan bahwa ada banyak diskusi mengenai proses masuk ke aplikasi, yang bisa berarti pengalaman pengguna terkait login sangat penting. Terakhir, kata 'prima', muncul 273 kali, merujuk pada nama aplikasi, menunjukkan pengakuan terhadap identitas atau branding aplikasi.

Secara keseluruhan, visualisasi ini memberikan gambaran jelas tentang apa yang dianggap penting oleh pengguna dalam aplikasi Primaku, dengan fokus pada aspek kegunaan, kualitas, pembaruan, dan dampak pada perkembangan anak.

### 3.2 Pembahasan

Pembagian data yang digunakan adalah 80% sebagai data pelatihan dan 20% sebagai data pengujian. Setelah dilakukan pembagian data, dilakukan pengolahan data dengan GoogleColab dan bahasa pemrograman Python dengan model klasifikasi model klasifikasi menggunakan algoritma Random Forest dan Support Vector Machine.

Skenario pengujian dapat dilihat pada Tabel 7. Pengujian ini dilakukan dengan melakukan proses pemodelan menggunakan algoritma Random Forest dan Support Vector Machine. Proses pengujian dilakukan pada algoritma tanpa hyperparameter dan algoritma dengan hyperparameter Grid Search CV.

**Tabel 7.** Parameter Terbaik pada Random Forest

Pengujian	Model
1	Random Forest, Support Vector Machine
2	Random Forest dengan hyperparameter, Support Vector Machine dengan hyperparameter

Pemodelan dioptimalkan dengan menggunakan hyperparameter Grid Search CV. Metode ini digunakan untuk mendapatkan kombinasi parameter dengan model terbaik. Kombinasi parameter yang dihasilkan pada proses optimasi dapat merepresentasikan nilai-nilai yang menentukan pelatihan model pada Tabel 9 dan Tabel 10. Tabel 9 menunjukkan pengaturan hyperparameter terbaik untuk Random Forest dengan Grid Search CV menghasilkan `n_estimators` 150, `max_features` None, `max_depth` 9, dan `max_leaf_nodes` 9.

**Tabel 9.** Parameter Terbaik pada Random Forest

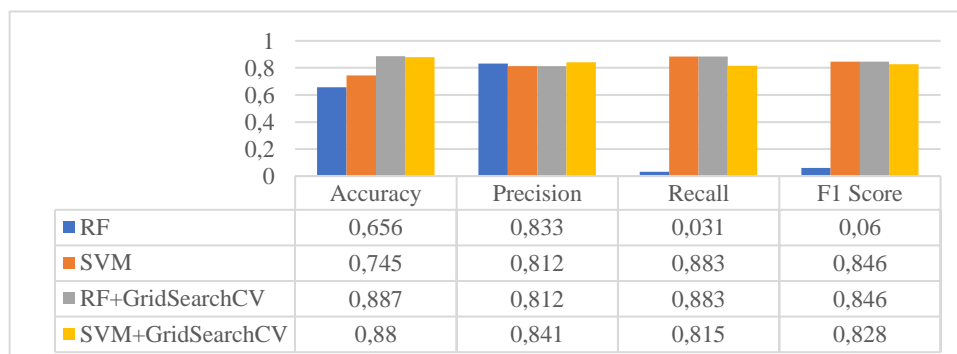
Hyperparameter	Values	Parameter
<code>n_estimators</code>	[25, 50, 100, 150]	150
<code>max_features</code>	'sqrt', 'log2', None	None
<code>max_depth</code>	[3, 6, 9]	9
<code>max_leaf_nodes</code>	[3, 6, 9]	9

Tabel 10 menunjukkan pengaturan hyperparameter untuk SVM. GridSearchCV menghasilkan 3 hyperparameter yaitu `C` dengan nilai 10, `gamma` 0.01, dan kernel `rbf`. Parameter `C` mengacu pada parameter regularisasi, parameter kernel menentukan jenis kernel yang akan digunakan, dan parameter `gamma` mengacu pada koefisien kernel.

**Tabel 10.** Parameter Terbaik pada Support Vector Machine

Hyperparameter	Values	Parameter
<code>C</code>	[0.1, 1, 10, 100, 1000]	10
<code>gamma</code>	[1, 0.1, 0.01, 0.001, 0.0001]	0.01
kernel	['rbf', 'poly', 'linear']	rbf

Gambar 5 menunjukkan adanya peningkatan dengan penyetelan hyperparameter GridsearchCV memberikan peningkatan terhadap kinerja model, random forest diperoleh nilai akurasi 65,6%, precision 83,3%, recal 3,1% dan nilai f1 score 6% setelah menggunakan teknik GridSearchCV menunjukkan adanya peningkatan sebesar 23,1%, random forest dengan teknik GridSearchCV diperoleh nilai akurasi 88,7%, precision 81,2%, recall 88,3% dan f1 score 84,6%. Dan SVM sebelum menggunakan teknik GridSearchCV diperoleh 74,5%, precision 81,2%, recall 88,3% dan f1 score 84,6% sedangkan SVM dengan menggunakan teknik GridSearchCV menunjukkan adanya peningkatan sebesar 13,5% sehingga diperoleh nilai akurasi sebesar 88%, precision 84,1%, recal 81,5% dan f1 score 82,8%.



**Gambar 5.** Confusion Matrik



## 4. KESIMPULAN

Penelitian menunjukkan bahwa hyperparameter GridSearchCV memberikan dampak dari penyetelan hyperparameter terhadap kinerja model. Pengoptimalan hyperparameter, terutama melalui GridSearchCV dapat meningkatkan akurasi model Random Forest dan Support Vector Machine. Random Forest mengalami peningkatan akurasi sebesar 23,1% dan Support Vector Machine mengalami peningkatan sebesar 13,5%. Random Forest menunjukkan peningkatan yang signifikan dibandingkan Support Vector Machine dalam analisis sentiment ulasan aplikasi PrimaKu.

## REFERENCES

- [1] V. K, P. Samuel, B. V. Krishna, and M. J, "Exploration of sentiment analysis in twitter propaganda: a deep dive," *Multimed Tools Appl*, vol. 83, no. 15, pp. 44729–44751, 2024, doi: 10.1007/s11042-023-17383-6.
- [2] F. Rejeki and V. Ayumi, "Analisa Sentimen Mengenai Kenaikan Harga Bbm Menggunakan Metode Naïve Bayes Dan Support Vector Machine," *JSAI : Journal Scientific and Applied Informatics*, vol. 6, no. 1, 2023, doi: 10.36085.
- [3] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] M. Tabany and M. Gueffal, "Sentiment Analysis and Fake Amazon Reviews Classification Using SVM Supervised Machine Learning Model," *Journal of Advances in Information Technology*, vol. 15, no. 1, pp. 49–58, 2024, doi: 10.12720/jait.15.1.49-58.
- [5] T. Tinaliah and T. Elizabeth, "Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 4, 2022, doi: <https://doi.org/10.35957/jatisi.v9i4.3586>.
- [6] R. Aryanti, T. Misriati, and A. Sagiyanto, "Analisis Sentimen Aplikasi Primaku Menggunakan Algoritma Random Forest dan SMOTE untuk Mengatasi Ketidakseimbangan Data," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 218–227, Nov. 2023, doi: 10.47065/josyc.v5i1.4562.
- [7] Y. A. Ali, E. M. Awwad, M. Al-Razgan, and A. Maarouf, "Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity," *Processes*, vol. 11, no. 349, pp. 1–21, Feb. 2023, doi: 10.3390/pr11020349.
- [8] I. Jamaledyn, R. El Ayachi, and B. Mohamed, "Optimization of Machine Learning Algorithms Through Hyperparameter Tuning Applied for the classification of Arabic news," *Res Sq*, 2023, doi: 10.21203/rs.3.rs-2615380/v1.
- [9] J. F. Lundell, "Tuning Support Vector Machines and Boosted Trees Using Optimization Algorithms," *Journal of Data Science*, pp. 1–16, 2023, doi: 10.6339/23-jds1106.
- [10] Md. M. Rahman, A. Rahman, S. Akter, and S. A. Pinky, "Hyperparameter Tuning Based Machine Learning Classifier for Breast Cancer Prediction," *Journal of Computer and Communications*, vol. 11, no. 04, pp. 149–165, 2023, doi: 10.4236/jcc.2023.114007.
- [11] G. N. Ahmad, H. Fatima, Shafiullah, A. Salah Saidi, and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques with and Without GridSearchCV," *IEEE Access*, vol. 10, pp. 80151–80173, 2022, doi: 10.1109/ACCESS.2022.3165792.
- [12] S. George and B. Sumathi, "Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction," 2020. [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [13] K. Rangsiman, C. Fabrizio, and L. Sandra, "Machine Learning-Assisted Discovery of Hidden States in Expanded Free Energy Space," *Journal of Physical Chemistry Letters*, vol. 13, no. 7, pp. 1797–1805, Feb. 2022, doi: 10.1021/acs.jpclett.1c04004.
- [14] S. Rasheed, G. K. Kumar, D. M. Rani, M. V. V. P. Kantipudi, and M. Anila, "Heart Disease Prediction Using GridSearchCV and Random Forest," *EAI Endorsed Trans Pervasive Health Technol*, vol. 10, Jan. 2024, doi: 10.4108/eetpht.10.5523.
- [15] S. K. Jain and K. G. Anil, "Application of Random Forest Regression with Hyper-parameters Tuning to Estimate Reference Evapotranspiration," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 742–750, 2022, [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [16] E. Munson, C. Smith, B. Boehmke, and J. Freels, "Sentiment Analysis of Twitter Data (saotd)," *J Open Source Softw*, vol. 4, no. 34, p. 764, Feb. 2019, doi: 10.21105/joss.00764.
- [17] A. Lertpiya, T. Chalothorn, and P. Buabthong, "How to Progressively Build Thai Spelling Correction Systems?," *IEEE Access*, vol. 11, pp. 72704–72716, 2023, doi: 10.1109/ACCESS.2023.3295004.
- [18] N. Babanejad, H. Davoudi, A. Agrawal, A. An, and M. Papagelis, "The Role of Preprocessing for Word Representation Learning in Affective Tasks," *IEEE Trans Affect Comput*, vol. 15, no. 1, pp. 254–272, Jan. 2024, doi: 10.1109/TAFFC.2023.3270115.
- [19] A. Hanafiah, A. H. Nasution, Y. Arta, R. Wandri, H. O. Nasution, and J. Mardafora, "Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 6, no. 1, pp. 230–236, Mar. 2023, doi: 10.31539/intecom.v6i1.5845.
- [20] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, pp. 407–416, May 2021, doi: 10.30812/matrik.v20i2.1130.
- [21] M. IŞIK and H. DAĞ, "The impact of text preprocessing on the prediction of review ratings," *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, vol. 28, no. 3, pp. 1405–1421, May 2020, doi: 10.3906/elk-1907-46.





- [22] D. Mustikasari, I. Widaningrum, R. Arifin, and W. H. E. Putri, “Comparison of Effectiveness of Stemming Algorithms in Indonesian Documents,” in Proceedings of the 2nd Borobudur International Symposium on Science and Technology (BIS-STE 2020), 2021. doi: 10.2991/aer.k.210810.025.
- [23] R. Aryanti, A. Saryoko, A. Junaidi, S. Marlina, Wahyudin, and L. Nurmalia, “Comparing Classification Algorithm with Genetic Algorithm in Public Transport Analysis,” in Journal of Physics: Conference Series, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012017.
- [24] T. Misriati, R. Aryanti, and A. Sagiyo, “High Accurate Prediction of Heart Disease Classification by Support Vector Machine,” in Proceedings of the 3rd International Conference on Advanced Information Scientific Development, SCITEPRESS - Science and Technology Publications, 2023, pp. 5–9. doi: 10.5220/0012437100003848.
- [25] K. Chong and N. Shah, “Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 12, 2022, doi: 10.14569/IJACSA.2022.0131213.
- [26] O. Manullang, C. Prianto, and N. H. Harani, “Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest,” Jurnal Ilmiah Informatika (JIF), vol. 11, no. 2, pp. 159–169, 2023.
- [27] H. Adams, E. Farnell, and B. Story, “Support vector machines and Radon’s theorem,” Nov. 2020, [Online]. Available: <http://arxiv.org/abs/2011.00617>
- [28] W. W. Sukeiti and S. Surono, “Fuzzy Support Vector Machine Using Function Linear Membership and Exponential with Mahalanobis Distance,” JTAM (Jurnal Teori dan Aplikasi Matematika), vol. 6, no. 2, p. 268, Apr. 2022, doi: 10.31764/jtam.v6i2.6912.
- [29] Z. Yang, “Generating Knowledge-based Explanation for Recommendation from Review,” in Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA: ACM, Jul. 2022, pp. 3494–3494. doi: 10.1145/3477495.3531683.
- [30] V. Yepmo, G. Smits, and O. Pivert, “Anomaly explanation: A review Data and Knowledge Engineering,” Data Knowl Eng, vol. 137, p. 101946, Jan. 2022, doi: 10.1016/j.datak.2021.101946.
- [31] K. Padmanandam, S. P. V. D. S. Bheri, L. Vegesna, and K. Sruthi, “A Speech Recognized Dynamic Word Cloud Visualization for Text Summarization,” in 2021 6th International Conference on Inventive Computation Technologies (ICICT), IEEE, Jan. 2021, pp. 609–613. doi: 10.1109/ICICT50816.2021.9358693.