

# ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SIREKAP PADA PLAY STORE MENGGUNAKAN ALGORITMA RANDOM FOREST CLASSIFER

Muhamad Fajar Yudhistira Herjanto<sup>1</sup>, Carudin<sup>2</sup>

<sup>1,2</sup>Universitas Singaperbangsa Karawang, Jl.HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp.(0267) 641177

---

## Riwayat artikel:

Received: 11 Maret 2024

Accepted: 30 Maret 2024

Published: 2 April 2024

## Keywords:

Analisis Sentimen, SIREKAP, Random Forest Classifier, Ulasan

## Correspondent Email:

fajaryudhistiraherjanto@gmail.com

**Abstrak.** Pemilihan umum di Indonesia merupakan aspek krusial dari demokrasi yang memerlukan partisipasi warga negara. Dalam era digital, Badan Penyelenggara Pemilu meluncurkan aplikasi SIREKAP untuk membantu mempercepat proses rekapitulasi pemilihan umum. Tujuan penelitian ini menganalisis sentimen ulasan aplikasi SIREKAP pada Play Store untuk memperoleh pemahaman yang lebih mendalam mengenai pandangan dan evaluasi pengguna. Metode yang digunakan adalah *knowledge discovery in database (KDD)* meliputi tahap *data selection, preprocessing, transformation, data mining* dan *evaluation*. Algoritma yang digunakan untuk klasifikasi adalah Random Forest Classifier terhadap 5000 data ulasan. Hasil penelitian menunjukkan akurasi 74%, presisi 75%, recall 74%, dan f1-score 74%. Selain itu hasil dari pelabelan data menunjukkan ulasan negatif mendapatkan terbanyak sebanyak 4002 ulasan dibandingkan dengan ulasan positif hanya 762 ulasan dan 234 ulasan netral. Kesimpulannya aplikasi SIREKAP perlu ditingkatkan kualitasnya untuk mengatasi permasalahan yang sering ditemui pengguna berdasarkan ulasan pengguna. Hasil penelitian ini dapat menjadi evaluasi bagi pengembang untuk meningkatkan pengalaman pengguna.

**Abstract.** General elections in Indonesia are a crucial aspect of democracy that require the participation of citizens. In the digital era, the General Election Commission launched the SIREKAP application to help expedite the process of election recapitulation. The purpose of this study is to analyze the sentiment of SIREKAP application reviews on the Play Store to gain a deeper understanding of user views and evaluations. The method used is *knowledge discovery in the database (KDD)*, including stages such as *data selection, preprocessing, transformation, data mining, and evaluation*. The classification algorithm used is the random forest classifier applied to 5000 review data. The research findings show an accuracy of 74%, precision of 75%, recall of 74%, and an f1-score of 74%. Additionally, the data labeling results indicate that negative reviews are the most prevalent, comprising 4002 reviews compared to only 762 positive reviews and 234 neutral reviews. In conclusion, the SIREKAP application needs to improve its quality to address the issues frequently encountered by users based on their reviews. The results of this study can serve as an evaluation for developers to enhance user experience.

## 1. PENDAHULUAN

Pemilihan umum adalah bentuk konkret dari demokrasi prosedural di mana warga negara memiliki peran penting dalam menentukan jalannya pemerintahan. Ini merupakan aspek krusial dari demokrasi yang harus dilakukan dengan cara yang demokratis pula. Melalui pemilihan umum, negara Indonesia secara langsung melibatkan seluruh rakyatnya untuk berpartisipasi dalam menetapkan arah dan kebijakan pemerintahan suatu negara[1].

Di era digital saat ini Badan Penyelenggara Pemilu meluncurkan aplikasi SIREKAP dengan tujuan durasi masa rekapitulasi lebih cepat, menghindari kesalahan dan kelelahan petugas, Mesin bekerja tanpa emosi dan kepentingan parsial sehingga hasilnya lebih kredibel demikian pula penerapan teknologi informasi [2]. Aplikasi SIREKAP sebagai salah satu aplikasi yang tersedia di Play Store, juga tidak terkecuali dari ulasan pengguna. Ulasan merupakan bentuk respon dari pengguna aplikasi seperti kepuasan atau kekecewaan pengguna atau opini terhadap aplikasi tersebut[3]. Analisis sentimen pada aplikasi SIREKAP menjadi penting untuk mengetahui seberapa baik persepsi dan respon pengguna terhadap aplikasi tersebut. Terutama karena aplikasi SIREKAP menjadi alat yang vital dalam proses pemilihan umum, memiliki peran krusial dalam memfasilitasi partisipasi warga negara dalam proses demokrasi.

Untuk melakukan analisis sentimen memerlukan algoritma yang dapat memproses teks dan mengklasifikasikannya kedalam kategori sentimen yang sesuai. Algoritma Random Forest Classifier telah terbukti efektif untuk melakukan klasifikasi otomatis. Pada penelitian sebelumnya oleh [4], penggunaan algoritma Random Forest Classifier telah terbukti memberikan hasil akurasi tertinggi. Jumlah pohon yang terbentuk dalam algoritma ini mempengaruhi tingkat akurasi secara signifikan. Dengan demikian, Diharapkan bahwa penggunaan Random Forest Classifier dalam penelitian ini mampu menghasilkan analisis sentimen yang akurat dan dapat dipercaya.

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan yang diberikan oleh pengguna aplikasi SIREKAP di platform Play Store, dengan memanfaatkan algoritma Random Forest. Dengan melakukan

analisis sentimen ini, diharapkan dapat memperoleh pemahaman yang lebih mendalam mengenai pandangan dan evaluasi pengguna terhadap aplikasi SIREKAP.

## 2. TINJAUAN PUSTAKA

### 2.1. Analisis Sentimen

Analisis sentimen adalah proses mengidentifikasi arah polaritas teks dalam dokumen atau jenis dokumen lainnya, dengan fokus pada pengelompokan sentimen sebagai positif, negatif, atau netral. Pendekatan ini melibatkan pengumpulan data dengan merinci dan mengorganisir teks ke dalam entitas tertentu seperti layanan, produk, individu, fenomena, dan objek lainnya. Metode ini bertujuan untuk menganalisis pemikiran dan perasaan individu dengan maksud menilai sejauh mana mereka menyetujui atau tidak setuju terhadap topik yang dibahas [5].

### 2.2. Web Scraping

Web scraping adalah teknik pengambilan otomatis dokumen semi-terstruktur dari internet, terutama dari halaman web yang menggunakan markup seperti HTML. Tujuannya adalah untuk mengekstrak data tertentu dari halaman tersebut. Dengan memanfaatkan teknik ini, informasi dapat diambil dari situs tanpa perlu intervensi manual. Web scraping memungkinkan pengumpulan data yang terfokus, mempermudah pencarian informasi yang diinginkan dengan efisiensi waktu [6].

### 2.3. Random Forest Classifier

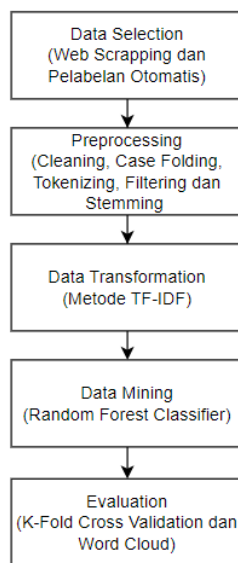
Random Forest Classifier adalah evolusi dari algoritma decision tree, di mana setiap pohon keputusan dilatih menggunakan data individu, dan setiap atribut dipilih secara acak pada setiap pembentukan pohon. Algoritma ini memiliki keunggulan karena mampu meningkatkan akurasi hasil, terutama dalam menghadapi data yang hilang, serta memiliki ketahanan terhadap data outlier dan efisiensi dalam pengelolaan penyimpanan data. Selain itu, proses seleksi fitur yang cermat dapat mengidentifikasi fitur-fitur terbaik, yang secara keseluruhan dapat meningkatkan performa model klasifikasi[7].

### 2.4. K-Fold Cross Validation

K-Fold adalah suatu teknik dalam data mining yang digunakan untuk mendapatkan akurasi optimal dengan membagi dataset menjadi subset data uji dan latih secara iteratif. Teknik ini melibatkan pembagian dataset menjadi partisi-partisi secara acak sebanyak nilai K yang ditentukan. Pada setiap iterasi, satu partisi dijadikan data uji, sementara partisi lainnya menjadi data latih. Proses ini diulangi K kali hingga selesai sesuai dengan jumlah K yang digunakan dalam penelitian [8].

### 3. METODE PENELITIAN

Penelitian ini menggunakan metodologi *Knowledge Discovery in Database (KDD)*. Terdiri dari beberapa tahapan seperti *data selection, preprocessing, transformation, data mining* dan *evaluation* untuk lebih jelas dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

#### 3.1. Data Selection

Pada tahap ini penulis menghimpun sejumlah data yang berperan sebagai dataset. Data ini diperoleh melalui teknik web scraping dari ulasan yang diberikan oleh pengguna aplikasi SIREKAP di platform Play Store.

#### 3.2. Preprocessing

Pada tahap *Preprocessing* penulis mempersiapkan kumpulan data mentah yang sebelumnya telah dikumpulkan agar dapat diolah lebih lanjut. *Preprocessing* melibatkan

beberapa tahapan yaitu *Cleaning, Case Folding, Tokenizing, Filtering dan Stemming*[9].

#### 3.3. Data Transformation

*Data transformation* merupakan suatu proses mengubah data dari format asalnya ke format yang lebih sesuai seperti mengubah data bertipe nominal menjadi data numerik. Tahap ini bertujuan untuk mengubah data yang dipilih agar sesuai dan mempermudah proses data mining.[10]

#### 3.4. Data Mining

*Data Mining* merupakan proses pencarian pola atau informasi penting dalam kumpulan data yang telah dipilih, dengan menggunakan teknik, metode, atau algoritma tertentu yang relevan dengan tujuan yang telah ditentukan [11]. Algoritma Random Forest Classifier dipilih untuk melakukan analisis sentimen karena merupakan teknologi pre-processing dalam klasifikasi fitur yang meningkatkan skalabilitas, akurasi, dan efisiensi, terutama dalam proses klasifikasi teks seperti pohon. Oleh karena itu, Random Forest dianggap efisien dan sederhana, serta sensitif dalam pemilihan fitur, sehingga mudah dan sederhana dalam penggunaannya [12].

#### 3.5. Evaluation

Tahap *evaluation* merupakan langkah terakhir yang dilakukan untuk menilai kinerja algoritma klasifikasi yang digunakan dalam penelitian. Performa model diukur melalui empat metrik, yaitu [13]:

1. *Accuracy* merupakan proporsi dari keseluruhan data yang diprediksi dengan tepat.
2. *Recall* adalah metrik yang menunjukkan seberapa banyak data dari kelas positif yang diprediksi dengan akurat.
3. *Precision* adalah indikator seberapa tepat model dalam memprediksi data dari kelas positif.
4. *F-measure* adalah sebuah nilai yang diperoleh dengan menggabungkan rata-rata dari *precision* dan *recall*.

### 4. HASIL DAN PEMBAHASAN

Hasil dan pembahasan analisis sentimen pada aplikasi SIREKAP menggunakan metode klasifikasi Random Forest sebagai berikut.

### 4.1. Web Scraping

Pengumpulan data menggunakan teknik web scraping pada Play Store dengan mengumpulkan sebanyak 5000 data ulasan pada aplikasi SIREKAP dengan menggunakan skrip sebagai berikut.

```

Web Scraping
result, continuation_token = reviews(
    "id.go.kpu.sirekap2024",
    lang='id',
    country='id',
    sort=Sort.NEWEST,
    count=10000
)
result, _ = reviews(
    "id.go.kpu.sirekap2024",
    continuation_token=continuation_token
)
    
```

Hasil output dari skrip tersebut dapat dilihat pada Gambar 2.

**Gambar 2.** Data hasil web scrapping

Selanjutnya, data diberi label dalam tiga kelompok yaitu positif, netral, dan negatif. Berikut adalah skrip untuk memberi label secara otomatis pada data.

```

Pelabelan
g_df2['label'] = g_df2['rating'].apply(lambda x:
'positif'
if x > 3
else 'negatif'
if x < 3
else 'netral'
)
    
```

Setelah melakukan pelabelan otomatis, didapatkan sebanyak 4004 label negatif, 762 label positif, dan 234 label netral dalam dataset. Untuk menghindari overfitting, penulis membatasi jumlah data yang akan digunakan

dalam penelitian ini, yaitu 500 label negatif dan 500 label positif.

### 4.2. Preprocessing

Selanjutnya merupakan tahap preprocessing, pada tahap ini data yang telah di kumpulkan melalui web scrapping dibersihkan melalui beberapa tahapan yaitu *Data Cleaning*, *Case Folding*, *Tokenizing*, *Filtering* dan *Stemming*.

#### 4.2.1 Data Cleaning

Pada tahap *data cleaning*, dilakukan penghapusan elemen-elemen yang tidak relevan seperti simbol dan emotikon. Proses ini bertujuan untuk meningkatkan kualitas data dan membersihkan data dari noise. Berikut skrip dan hasil dari proses *data cleaning* dapat dilihat pada Tabel 1.

```

Data Cleaning
cleaning_text = re.sub(r'[^\a-zA-Z\s]', "", text)
return cleaning_text
    
```

**Tabel 1.** Hasil *Data Cleaning*

Sebelum	Sesudah
Terobosan baru. Selamat untuk Kpu yang aplikasinya gak bisa dipakai oleh iPhone 🤔🤔🤔	Terobosan baru Selamat untuk Kpu yang aplikasinya gak bisa dipakai oleh iPhone
Perbaiki aplikasinya, dong masak gagal login terus.	Perbaiki dong aplikasinya masak gagal login terus
Mohon dibantu agar dapat login.	Mohon dibantu agar dapat login

#### 4.2.2 Case Folding

Pada tahap case folding data yang telah dibersihkan dari noise diubah ke bentuk yang seragam kedalam lowercase. Berikut ini adalah skrip dan hasil dari case folding yang dapat dilihat pada Tabel 2.

```

Case Folding
text_lower = text.lower()
return text_lower
    
```

**Tabel 2.** Hasil *Case Folding*

Sebelum	Sesudah
Terobosan baru Selamat untuk Kpu yang aplikasinya gak	terobosan baru selamat untuk kpu yang aplikasinya gak bisa dipakai oleh iphone

bisa dipakai oleh iPhone	
Perbaiki dong aplikasinya masak gagal login terus	perbaiki dong aplikasinya masak gagal login terus
Mohon dibantu agar dapat login	mohon dibantu agar dapat login

### 4.2.3 Tokenizing

Pada tahap *tokenizing* data dipisahkan dalam bentuk unit-unit kecil yang disebut sebagai token, berikut skrip dan hasil dari tahap *tokenizing* dapat dilihat pada Tabel 3.

<b>Tokenizing</b>
word_token = word_tokenize(text) return word_token

Tabel 3. Hasil Tokenizing

Sebelum	Sesudah
terobosan baru selamat untuk kpu yang aplikasinya gak bisa dipakai oleh iphone	'terobosan', 'baru', 'selamat', 'untuk', 'kpu', 'yang', 'aplikasinya', 'gak', 'bisa', 'dipakai', 'oleh', 'iphone'
perbaiki dong aplikasinya masak gagal login terus	'perbaiki', 'dong', 'aplikasinya', 'masak', 'gagal', 'login', 'terus'
mohon dibantu agar dapat login	'mohon', 'dibantu', 'agar', 'dapat', 'login'

### 4.2.4 Filtering

Pada tahap *filtering* setelah data di tokenisasi langkah berikutnya menghapus kata-kata yang tidak relevan, berikut script dan hasil dari *filtering* dapat dilihat pada Tabel 4.

<b>Filtering</b>
filtered_text = [word for word in word_token if word.lower() not in stop_words]

Tabel 4. Hasil Filtering

Sebelum	Sesudah
'terobosan', 'baru', 'selamat', 'untuk', 'kpu', 'yang', 'aplikasinya', 'gak', 'bisa', 'dipakai', 'oleh', 'iphone'	terobosan selamat kpu aplikasinya gak dipakai iphone
'perbaikan', 'dong', 'aplikasinya', 'masak', 'gagal', 'login', 'terus'	Perbaiki aplikasinya masak gagal login
'mohon', 'dibantu', 'agar', 'dapat', 'login'	mohon dibantu login

### 4.2.5 Stemming

Dalam tahap ini, data diubah menjadi bentuk dasarnya dengan menghapus imbuhan atau akhiran, sehingga hanya kata dasarnya yang tersisa. Berikut adalah skrip dan hasil dari proses stemming dapat dilihat pada Tabel 5.

<b>Stemming</b>
stemmer = PorterStemmer() words = [stemmer.stem(word) for word in words]

Tabel 5. Hasil Stemming

Sebelum	Sesudah
'terobosan', 'selamat', 'kpu', 'aplikasinya', 'gak', 'dipakai', 'iphone'	terobosan selamat kpu aplikasinya gak dipakai iphon
'perbaiki', 'aplikasinya', 'masak', 'gagal', 'login'	perbaiki aplikasinya masak gagal login
'mohon', 'dibantu', 'login'	mohon dibantu login

### 4.3. Data Transformation

Setelah melalui proses *preprocessing*, data kemudian ditransformasikan menjadi representasi vektor dengan menggunakan metode TD-IDF, Representasi vektor yang dihasilkan dari transformasi tersebut akan diproses oleh model. Berikut merupakan script dari metode TD-IDF dan hasil dari perubahan data dapat dilihat pada Gambar 3.

<b>Metode TD-IDF</b>
X_preprocessed = X.apply(preprocessing_text) vectorizer = TfidfVectorizer() X_tfidf = vectorizer.fit_transform(X_preprocessed)

(0, 993)	0.8634238983111571
(0, 69)	0.5044791094041107
(1, 1282)	0.7241784061741855
(1, 993)	0.6896126710197663
(2, 190)	0.7760736364395099
(2, 707)	0.6306423002174492

Gambar 3. Hasil Data Transformation

### 4.4. Data Mining

Kemudian penulis menginialisasikan model Random Forest yang nantinya digunakan untuk membuat prediksi berdasarkan fitur-fitur yang terdapat didalam data yang telah di transformasikan sebelumnya.





**DAFTAR PUSTAKA**

- [1] D. Handita and L. Dewi A, “Penerapan Pemilu Online Berbasis Aplikasi Smarthphone di Era Pandemi Covid-19”, *Prosiding Seminar Nasional Desain Sosial*, pp. 848-851, 2021.
- [2] Fitriyah, F. Herawati, N. R., and Herawati. R, “Challenges and Obstacles KPU in The Use of E-Recapitulation SIREKAP in The 2024 General Elections”, *ICISPE (International Conference Proceeding)*, pp. 1-7, 2022.
- [3] A. Erfina, E. S. Basyrah *et al*, “Analisis Sentimen Aplikasi Pembelajaran Online di Play Store Pada Masa Pandemi Covid-19 Menggunakan Algoritma Support Vector Machine (SVM)”, *SEMANIF (Seminar Nasional Informatika)*, pp. 146-152, 2020.
- [4] I. Agustina, Y. Mulyani *et al*, “Analisis Pengembangan Model Prediksi Kesuksesan Kickstarter Menggunakan Algoritma Backpropagation dan Random Forest”, *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 10, no. 3, pp. 173-182, 2022.
- [5] Dedi P., M. Afdal, Mustakim and I. Permana, “Analisis Sentimen Terhadap Peminangan Ibu Kota Negara Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbors”, *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp.1306-1314, 2023.
- [6] D. F. Sari, A. Kusjani, D. Kurniawati and I. Setiawan, “Pencarian Data Quick Count Pilpres dengan Teknik Web Scraping”, *JIRK (Journal of Innovation Research and Knowledge)*, vol.3, no. 5, pp. 1025-1033, 2023.
- [7] O. Manullang, C. Prianto and N. H. Harani, “Analisis Sentimen untuk Menampilkan Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest”, *JIF (Jurnal Ilmiah Informatika)*, vol. 11, no. 2, pp. 159-169, 2023.
- [8] Nurainun, E. Haerani, F. Syafria and L.Oktavia,”Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation”, *JoSYC (Journal of Computer System and Informatics)*, vol. 4, no. 4, pp. 578-586, 2023.
- [9] R. Q. Rohmansa, N. Pratiwi and M. J. Palepa,”Analisis Sentimen Ulasan Pengguna Aplikasi Discord Menggunakan Metode K-Nearest Neighbor”, *JIFI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol.9, no.1, pp. 330-348, 2024.
- [10] S. Pujiono, R. Astuti and F. M. Basysyar, ”Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering”, *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol.8, no.1, pp. 615-620, 2024.
- [11] S. Asyuti and A. Aji. S,”Data Mining Dalam Penggunaan Presensi Karyawan Dengan Cluster Means”, *Jurnal Ilmiah Sains Teknologi dan Informasi*, vol.1, no.1, pp. 01-10, 2023.
- [12] Farras and L. Trisnawati, “Analisis Sentimen Pendapat Masyarakat Terhadap Aksi People Power (22 Mei 2019) Melalui Media Sosial Twitter Menggunakan Algoritma Random Forest”, *Jurnal Teknik Informatika*, vol. 3, no. 3, pp. 1-12, 2023.
- [13] F. Nufairi, N. Pratiwi and F. Herlando, ”Analisis Sentimen Pada Ulasan Aplikasi Threads di Google Playstore Menggunakan Algoritma Support Vector Machine”, *JIFI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 1, pp. 339-348, 2024.