p-ISSN : 2406-7733
e-ISSN : 2597-9922

ANALISIS SENTIMEN PENGGUNA APLIKASI QPON PADA ULASAN PLAY STORE MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)

Muhammad Hanif[1], Garno[2], Asep Jamaludin[3]
Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang
E-mail: 2210631170084@student.unsika.ac.id[1], garno@staff.unsika.ac.id[2], asep.jamaludin@staff.unsika.ac.id[3]

Abstrak - Ulasan pengguna menjadi sumber informasi penting untuk menilai kualitas layanan, namun jumlah komentar yang besar menyulitkan pengembang dalam memahami opini secara manual. Aplikasi Qpon sebagai platform penyedia promo dan *voucher* belanja menerima banyak ulasan pada Google Play Store yang mencerminkan pengalaman pengguna terhadap layanan aplikasi. Penelitian ini menganalisis sentimen pengguna terhadap aplikasi Qpon dengan melakukan klasifikasi ulasan untuk mengidentifikasi kecenderungan opini positif dan negatif. Data penelitian diperoleh dari 2.304 ulasan pengguna yang terdiri dari 945 ulasan positif dan 1.359 ulasan negatif. Pelabelan sentimen dilakukan berdasarkan nilai penilaian pengguna dan diverifikasi oleh ahli bahasa untuk memastikan kesesuaian konteks. Data kemudian melalui proses pemilihan data, pembersihan teks, normalisasi kata, serta pembobotan kata menggunakan metode TF-IDF sebelum dilakukan klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Hasil pengujian menunjukkan model mampu mengklasifikasikan sentimen dengan akurasi sebesar 92,98% pada pembagian data latih dan uji 90:10 menggunakan kernel *Linear* dan *Sigmoid*, serta 92,97% pada pembagian 80:20 menggunakan kernel *RBF*. Analisis kosakata menunjukkan sentimen positif didominasi kata yang berkaitan dengan diskon, promo, dan kemudahan penggunaan, sedangkan sentimen negatif didominasi keluhan terkait masuk akun, penggunaan *voucher*, pembayaran, dan pengembalian dana. Hasil penelitian ini menunjukkan bahwa metode yang digunakan efektif untuk mengidentifikasi opini pengguna dan dapat dimanfaatkan sebagai bahan evaluasi peningkatan kualitas layanan aplikasi.

Kata Kunci: Analisis Sentimen, *Knowledge Discovery in Database* (KDD), *Support Vector Machine*, Ulasan Pengguna, Qpon

I. PENDAHULUAN

Perkembangan teknologi digital memberikan dampak signifikan terhadap kehidupan masyarakat modern, khususnya meningkatnya penggunaan aplikasi *mobile* di Indonesia (Ardianto et al., 2024). Aplikasi *mobile* kini menjadi bagian penting dalam aktivitas sehari-hari, mulai dari perbankan digital, belanja daring, hingga hiburan. Seiring meningkatnya adopsi layanan digital, kebutuhan untuk memahami pengalaman serta opini pengguna menjadi semakin penting untuk mendukung peningkatan kualitas layanan (Wahyudin, 2025).

Platform distribusi aplikasi seperti Google Play Store menyediakan ruang bagi pengguna untuk memberikan ulasan dan *rating* terhadap aplikasi (Nurwahidah et al., 2023). Ulasan tersebut merupakan sumber data tidak terstruktur yang berharga untuk menggambarkan kepuasan pengguna, kinerja aplikasi, serta aspek layanan yang perlu diperbaiki (Yani et al., 2025). Berdasarkan Statista (2024), jumlah unduhan aplikasi mobile global mencapai 136 miliar unduhan pada tahun 2024, sementara laporan data.ai menunjukkan lebih dari 257 miliar aktivitas unduhan termasuk pembaruan dan instalasi ulang. Kondisi ini menandakan bahwa tantangan pengembang aplikasi tidak hanya menarik pengguna baru, tetapi juga mempertahankan kepuasan dan loyalitas pengguna melalui pemantauan opini pengguna (Ceci, 2024).

Namun, volume ulasan yang sangat besar membuat pengembang kesulitan mengekstraksi informasi secara manual (Abidin, 2025). Oleh karena itu, diperlukan analisis otomatis menggunakan analisis sentimen, yaitu metode yang mengubah teks ulasan menjadi informasi kuantitatif sehingga dapat diinterpretasikan secara objektif (Sugiarto & Utomo, 2025). Analisis sentimen memungkinkan pengembang memahami faktor-faktor yang memengaruhi persepsi pengguna terhadap kualitas layanan secara lebih mendalam (Sariatin & Ekawati, 2023).

Salah satu aplikasi yang menarik untuk dikaji adalah Qpon, yaitu platform diskon gaya hidup yang menyediakan *voucher* dan kupon digital untuk berbagai kebutuhan seperti makanan, hiburan, dan layanan (Qpon, 2025). Berdasarkan data Play Store (2025), Qpon memiliki *rating* rata-rata 3,8 bintang dari sekitar 6.800 ulasan pengguna, lebih rendah dibandingkan kompetitornya seperti Indomaret Poinku (4,4), Shopback (4,2), dan Ultra Voucher (3,9). Hal tersebut menunjukkan bahwa tingkat kepuasan pengguna terhadap aplikasi Qpon belum optimal. Beberapa ulasan negatif menunjukkan keluhan terkait kesulitan klaim *voucher*, masalah

p-ISSN : 2406-7733
e-ISSN : 2597-9922

akun, serta lambatnya layanan pelanggan, sedangkan ulasan positif menyoroti kemudahan penggunaan dan manfaat penghematan. Perbedaan opini tersebut menimbulkan kebutuhan analisis berbasis data karena analisis manual terhadap ribuan ulasan tidak efektif dan berpotensi bias (Satrio, 2025). Oleh karena itu, diperlukan pendekatan otomatis berbasis machine learning untuk mengklasifikasikan opini pengguna secara cepat dan akurat (Himawan, 2025).

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *Support Vector Machine* (SVM) efektif untuk analisis sentimen. Huda et al. (2025) memperoleh akurasi 84,74% pada analisis sentimen aplikasi Pluang, sedangkan Bazar et al. (2025) memperoleh akurasi 93,12% pada aplikasi BNI Mobile, lebih tinggi dibandingkan *K-Nearest Neighbor.* Hasil tersebut membuktikan bahwa SVM mampu memberikan klasifikasi sentimen yang stabil dan akurat pada data teks berbahasa Indonesia.

Meskipun demikian, penelitian analisis sentimen pada aplikasi diskon digital seperti Qpon masih jarang dilakukan, padahal sektor ini memiliki interaksi pengguna yang tinggi dan tingkat kepuasan pelanggan yang kompleks. Oleh karena itu, penelitian ini mengusulkan penerapan algoritma SVM untuk menganalisis sentimen ulasan pengguna aplikasi Qpon di Google Play Store guna memperoleh gambaran persepsi pengguna secara komprehensif.

Penelitian ini bertujuan menganalisis opini pengguna terhadap aplikasi Qpon dan mengklasifikasikan ulasan menjadi sentimen positif dan negatif menggunakan *Support Vector Machine*. Hasil penelitian diharapkan dapat menjadi bahan evaluasi bagi pengembang dalam meningkatkan kualitas layanan serta kepuasan pengguna di tengah persaingan aplikasi digital yang semakin kompetitif.

II. TINJAUAN PUSTAKA

A. Analisis Sentimen

Analisis sentimen merupakan teknik dalam *text mining* yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini atau pandangan pengguna terhadap suatu objek ke dalam kategori sentimen tertentu, seperti positif dan negatif. Analisis ini termasuk bagian dari *text analytics* yang berfokus pada pengenalan sikap emosional pengguna terhadap produk, layanan, atau organisasi (Liu, 2022).

B. Qpon

Qpon adalah aplikasi diskon gaya hidup yang membantu pengguna menemukan *voucher* dan kupon hemat untuk berbagai kebutuhan seperti makanan, minuman, belanja, dan hiburan. Qpon berada di bawah PT Qpon Digital Indonesia dan ditujukan bagi pengguna di Indonesia untuk memudahkan pencarian promo secara digital (Qpon, 2025).

C. *Support Vector Machine* (SVM)

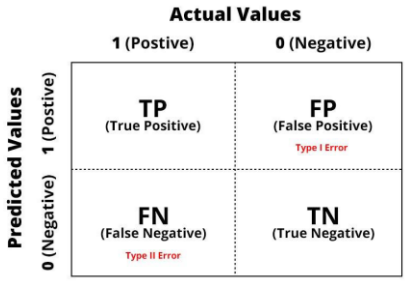
Support Vector Machine (SVM) merupakan algoritma pembelajaran terawasi (*supervised learning*) yang digunakan untuk klasifikasi dengan mencari *hyperplane* optimal sebagai pemisah antar kelas data. SVM bekerja dengan memaksimalkan margin antar kelas sehingga memiliki kemampuan generalisasi yang baik (Putri & Athoillah, 2021).

D. *Term Frequency-Inverse Document Frequency* (TF-IDF)

TF-IDF merupakan metode pembobotan kata yang digunakan untuk mengukur tingkat kepentingan suatu kata dalam dokumen berdasarkan frekuensi kemunculan dan distribusinya dalam kumpulan dokumen. TF (*Term Frequency*) mengukur frekuensi kemunculan kata dalam dokumen, sedangkan IDF (*Inverse Document Frequency*) mengukur seberapa jarang kata tersebut muncul dalam seluruh dokumen. Representasi TF-IDF mengubah teks menjadi bentuk numerik sehingga dapat diproses oleh algoritma klasifikasi seperti SVM.

E. *Confusion Matrix*

Confusion matrix merupakan metode evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi dan label aktual. *Confusion matrix* menunjukkan kelas yang diprediksi di bagian atas matriks dan kelas yang diamati di sebelah kiri untuk memfasilitasi perbandingan antara hasil klasifikasi sistem dan hasil klasifikasi yang sebenarnya. Berikut confusion matrix dalam bentuk 2 × 2 ditunjukan pada Gambar 1.



Gambar 1. *Confusion Matrix*

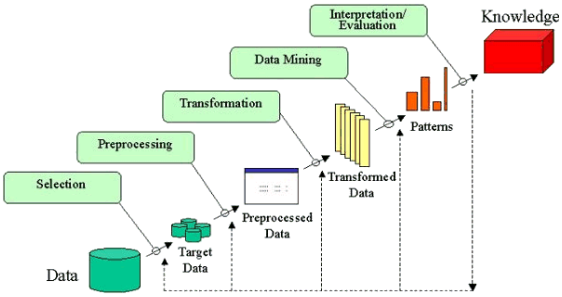
Confusion matrix terdiri dari empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dari komponen tersebut dapat dihitung metrik evaluasi seperti *accuracy, precision, recall,* dan *F1-*

p-ISSN : 2406-7733
e-ISSN : 2597-9922

score untuk menilai performa model secara menyeluruh.

F. *Knowledge Discovery in Database* (KDD)

Knowledge Discovery in Database (KDD) adalah proses eksplorasi dan analisis data untuk memperoleh pengetahuan yang bermanfaat. Metode ini terdiri dari enam tahap utama, yaitu *data selection, data preprocessing, data transformation, data mining, evaluation,* dan *knowledge.* Metode dipilih karena bersifat ilmiah dan mendukung eksplorasi data secara iteratif, sehingga sesuai untuk penelitian berbasis *machine learning* seperti analisis sentimen (Shimaoka et al., 2024).



Gambar 2. Tahapan Metode KDD

G. Penelitian Terkait

Penelitian oleh Pratama et al., (2025) menerapkan algoritma *Support Vector Machine* (SVM) untuk analisis sentimen pada ulasan aplikasi LinkAja. Dataset yang digunakan berjumlah 2.000 ulasan yang terdiri dari 944 sentimen positif dan 1.056 sentimen negatif. Model terbaik diperoleh pada skenario pembagian data 70:30 dengan akurasi 88%, *precision* 90%, *recall* 83%, dan *F1-score* 87%. Visualisasi *word cloud* menunjukkan kata dominan "aplikasi" dan "saldo" pada sentimen positif, sedangkan sentimen negatif didominasi kata "aplikasi" dan "pakai". Analisis akar masalah menunjukkan beberapa penyebab ketidakpuasan pengguna seperti layanan pelanggan kurang responsif, aplikasi sering mengalami *crash* pada perangkat rendah, serta kendala koneksi dan *bug system*.

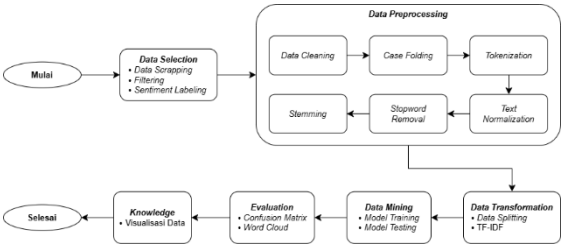
Penelitian oleh Nurian et al. (2024) melakukan analisis sentimen terhadap pelayanan Disdukcapil Karawang menggunakan algoritma *Naïve Bayes.* Data yang diperoleh terdiri dari 146 sentimen positif, 189 negatif, dan 961 netral. Model terbaik diperoleh pada pembagian data 90:10 dengan akurasi 69%, *precision* 82%, dan *recall* 48%. Pengujian menggunakan ROC menghasilkan nilai AUC sebesar 0,79 yang menunjukkan model memiliki kualitas klasifikasi kategori baik dalam mengevaluasi opini masyarakat terhadap layanan public.

Penelitian oleh (Sudrajat et al., 2021) menunjukkan bahwa algoritma *Support Vector Machine* (SVM) menghasilkan tingkat akurasi 93%, sedangkan *Naïve Bayes* hanya mencapai 89%. Dengan demikian, SVM memberikan performa yang lebih baik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi KAI *Access*, sehingga dapat dijadikan acuan bagi pengembang untuk mengetahui kelebihan dan kekurangan aplikasinya.

Meskipun demikian, sebagian besar penelitian sebelumnya berfokus pada sektor perbankan, layanan publik, dan hiburan. Penelitian pada aplikasi diskon digital seperti Qpon masih terbatas. Oleh karena itu, penelitian ini menghadirkan kebaruan dengan menganalisis sentimen pengguna aplikasi Qpon menggunakan SVM untuk memberikan gambaran persepsi pengguna pada sektor layanan kupon digital secara lebih spesifik.

III. METODE PENELITIAN

Metode penelitian yang diterapkan adalah *Knowledge Discovery in Database* (KDD), yang terdiri dari enam tahap, yaitu Data *Selection, Data Preprocessing, Data Transformation, Data Mining, Evaluation,* dan *Knowledge.* Untuk memudahkan pemahaman alur penelitian, keterangan lebih lanjut dapat dilihat pada gambar 3.



Gambar 3. Alur Penelitian

1. *Data Selection*

Tahap *data selection* merupakan proses pengumpulan dataset yang relevan dengan penelitian. Data diperoleh melalui teknik *scraping* ulasan pengguna aplikasi Qpon pada Google Play Store menggunakan *library* google-play-scraper dengan bahasa pemrograman Python. Kemudian dilakukan penyaringan berdasarkan periode penelitian dan pemilihan atribut kolom yang relevan secara langsung merepresentasikan opini pengguna.

Selanjutnya setiap ulasan diberi label sentimen positif dan negatif secara manual sesuai konteks komentar dan divalidasi oleh ahli bahasa untuk menjaga kesesuaian makna.

2. *Data Preprocessing*

Tahap *data preprocessing* bertujuan membersihkan data teks dari berbagai *noise* agar siap dianalisis. Ulasan Play Store memiliki karakteristik tidak terstruktur serta mengandung emoji, URL, simbol, dan bahasa tidak baku sehingga perlu dilakukan pembersihan data.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

a) *Data cleaning*: menghapus URL, emoji, *mention*, dan karakter non-alfanumerik.
b) *Case folding*: mengubah huruf menjadi huruf kecil.
c) *Tokenization*: memecah kalimat menjadi kata.
d) *Text normalization*: mengubah kata tidak baku menjadi baku.
e) *Stopword removal*: menghapus kata umum yang tidak bermakna
f) *Stemming*: mengubah kata berimbuhan menjadi kata dasar.

3. *Data Transformation*

Pada tahap data transformation, dataset terlebih dahulu dibagi menjadi data pelatihan (*data training*) dan data pengujian (*data testing*). Pembagian dilakukan dengan beberapa skenario perbandingan data, seperti 90:10, 80:20, dan 70:30 untuk memperoleh performa model terbaik. Setelah proses *splitting,* pembobotan kata dilakukan menggunakan metode TF-IDF. Pembobotan hanya dibangun dari *data training*, kemudian model pembobotan tersebut diterapkan pada *data testing*. Prosedur ini bertujuan menghindari data leakage sehingga model tidak memperoleh informasi dari data uji pada saat pelatihan. Melalui proses ini, teks ulasan diubah menjadi representasi numerik berupa matriks fitur yang dapat diproses oleh algoritma klasifikasi.

4. *Data Mining*

Tahap *data mining* merupakan proses utama klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM). Model dilatih menggunakan *data training* yang telah direpresentasikan dengan TF-IDF untuk membentuk *hyperplane* pemisah antara sentimen positif dan negatif. Setelah pelatihan, model diuji menggunakan *data testing* untuk mengetahui kemampuan generalisasi model dalam mengklasifikasikan ulasan pengguna baru.

5. *Evaluation*

Tahap *evaluation* digunakan untuk mengukur kinerja model klasifikasi. Evaluasi dilakukan menggunakan *confusion matrix* yang menghasilkan metrik *accuracy, precision, recall,* dan *F1-score.* Nilai-nilai tersebut digunakan untuk menilai tingkat ketepatan model dalam memprediksi sentimen pengguna secara keseluruhan.

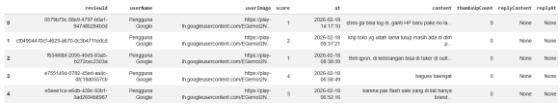
6. *Knowledge*

Tahap terakhir adalah *knowledge,* yaitu interpretasi hasil analisis sentimen untuk memperoleh informasi yang bermanfaat. Pada tahap ini dilakukan visualisasi distribusi sentimen dan penyusunan *insight* terkait persepsi pengguna terhadap kualitas layanan aplikasi Qpon. Hasil interpretasi ini diharapkan dapat menjadi bahan evaluasi bagi pengembang dalam meningkatkan kualitas layanan serta pengalaman pengguna aplikasi Qpon.

IV. HASIL DAN PEMBAHASAN

1. *Data Selection*

Pengumpulan data dilakukan melalui teknik *scraping* terhadap ulasan pengguna aplikasi Qpon pada platform Google Play Store menggunakan parameter pengurutan *Sort.NEWEST*. Hal ini karena memperoleh ulasan terbaru sehingga data yang dikumpulkan merepresentasikan kondisi aplikasi secara aktual pada periode penelitian.



Gambar 4. Hasil *Scraping* Data

Berdasarkan proses *scraping* tersebut, diperoleh hasil sebanyak 4.535 data ulasan pengguna. *Dataset* mentah yang dihasilkan terdiri dari beberapa atribut yang masih bersifat mentah dan belum seluruhnya relevan untuk analisis sentimen. Oleh karena itu, dilakukan proses penyaringan (*filtering*) berdasarkan periode penelitian, yaitu Juli hingga Oktober 2025, serta pemilihan atribut yang relevan terhadap tujuan penelitian.

Setelah proses *filtering* periode dan seleksi atribut dilakukan, jumlah data yang digunakan dalam penelitian ini menjadi 2.304 ulasan pengguna.

	date	time	score	content
0	2025-10-30	23:55:25	4	return nya lamaaaa
1	2025-10-30	18:30:15	5	mantap
2	2025-10-30	15:06:26	1	Pelayanan CS Ribet dan Sangat Tidak Ramah
3	2025-10-30	13:12:55	5	byk diskon nya tbh mantap ja
4	2025-10-30	12:17:48	5	mantap

Gambar 5. Hasil *Filtering* Periode dan Atribut

Selanjutnya, dilakukan proses pelabelan sentimen berdasarkan nilai rating (*score*) yang diberikan pengguna. Untuk menjaga validitas makna dan meminimalkan kesalahan interpretasi, hasil pelabelan kemudian divalidasi oleh ahli bahasa.

Tabel 1. Perbandingan Distribusi Sentimen

Jenis Label	Banyak		Persentase	
	Positif	Negatif	Positif	Negatif
Skor	1025	1279	44,49%	55,51%
Validasi	945	1359	41,02%	58,98%

2. *Data Preprocessing*

Tahapan ini mencakup serangkaian proses untuk mengubah data teks mentah (*raw data*) yang

p-ISSN : 2406-7733
e-ISSN : 2597-9922

diperoleh dari tahap sebelumnya menjadi format data yang bersih dan terstruktur.

a) Data Cleaning

Proses *data cleaning* ini dilakukan dengan membangun sebuah fungsi pembersihan *noise*. Fungsi ini menggunakan pola *regular expression*.

```
def cleaning(text):
    if text is None or not isinstance(text, str):      # Pastikan tipe string
        return text
    text = re.sub(r'https?://\S+|www\.\S+', '', text)  # Hapus URL
    text = re.sub(r'@\w+', '', text)                   # Hapus username (@user)
    text = re.sub("["                                  # Hapus emoji
        u"\U0001F600-\U0001F64F"
        u"\U0001F300-\U0001F5FF"
        u"\U0001F680-\U0001F6FF"
        u"\U0001F700-\U0001F77F"
        u"\U0001F780-\U0001F7FF"
        u"\U0001F800-\U0001F8FF"
        u"\U0001FA00-\U0001FA6F"
        u"\U0001FA70-\U0001FAFF"
        u"\U0001F004-\U0001F0CF"
        u"\U0001F1E0-\U0001F1FF"
        "]+", '', text)
    text = re.sub(r'\d+', '', text)                    # Hapus angka
    text = re.sub(r'[^a-zA-Z\s]', '', text)            # Hapus simbol
    text = re.sub(r'\s+', ' ', text).strip()           # Hapus spasi berlebih
    return text
```

Gambar 6. Kode *Data Cleaning*

Untuk memberikan gambaran perubahan yang terjadi, Tabel 2 menyajikan contoh perbandingan teks sebelum dan sesudah proses *data cleaning*.

Tabel 2. Perbandingan Teks *Data Cleaning*

Sebelum *Cleaning*	Sesudah *Cleaning*
sudah 3 hari nggak bisa login, chat cs jawabannya template mulu 😋	sudah hari nggak bisa login chat cs jawabannya template mulu
bermanfaat sekali.. buat yg low bajet	bermanfaat sekali buat yg low bajet
mantap😘	mantap

b) Case Folding

Proses *case folding* dilakukan menggunakan fungsi Python dengan mengubah seluruh karakter teks menjadi huruf kecil, kemudian diterapkan pada seluruh dataset.

```
def case_folding(text):
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text
```

Gambar 7. Kode *Case Folding*

Untuk melihat pengaruh proses tersebut, Tabel 3 menampilkan contoh perbandingan teks sebelum dan sesudah *case folding*.

Tabel 3. Perbandingan Teks *Case Folding*

Sebelum *Case Folding*	Sesudah *Case Folding*
Pelayanan CS Ribet dan Sangat Tidak Ramah	pelayanan cs ribet dan sangat tidak ramah
Semakin sering mahal	semakin sering mahal
Berguna sekali	berguna sekali

c) Tokenization

Proses ini bertujuan memisahkan kalimat menjadi kumpulan kata tunggal sehingga setiap kata dapat dianalisis secara individual pada tahap pengolahan selanjutnya.

```
def tokenization(text):
    tokens = text.split()
    return tokens
```

Gambar 8. Kode *Tokenization*

Untuk memperjelas perubahan yang terjadi, Tabel 4 menunjukkan contoh perbandingan teks sebelum dan sesudah proses *tokenization*.

Tabel 4. Perbandingan Teks *Tokenization*

Sebelum *Tokenization*	Sesudah *Tokenization*
return nya lamaaaa	[return, nya, lamaaaa]
semakin sering mahal	[semakin, sering, mahal]
berguna sekali	[berguna, sekali]

d) Text Normalization

Tahap ini bertujuan mengubah kata tidak baku, singkatan, dan bahasa tidak formal pada ulasan pengguna menjadi kata baku. Pada proses ini menggunakan kamus normalisasi manual yang divalidasi oleh ahli bahasa.

```
def normalize_text(tokens, kamus):
  normalized_words = []
  kata_diganti = []
  kata_baku = []

  for word in tokens:
      if word in kamus:
          kata_diganti.append(word)
          kata_baru = kamus[word]
          kata_baku.append(kata_baru)
          normalized_words.append(kata_baru)
      else:
          normalized_words.append(word)

  return normalized_words, kata_diganti, kata_baku
```

Gambar 9. Kode *Tokenization*

Tabel 5 menampilkan contoh perubahan kata sebelum dan sesudah proses normalisasi.

Tabel 5. Perbandingan *Text Normalization*

Sebelum Normalisasi	Sesudah Normalisasi
[return, nya, lamaaaa]	[retur, nya, lama]
[gk, bisa]	[tidak, bisa]
[byk, diskon, nya, tbh, mantap, ja]	[banyak, diskon, nya, lebih, mantap, saja]

e) Stopword Removal

Tahapan proses menghapus kata-kata umum yang tidak memiliki makna penting. Proses ini memanfaatkan daftar *stopword* Bahasa Indonesia yang tersedia.

```
def remove_stopwords(tokens):
    return [word for word in tokens if word not in stop_words]
```

Gambar 10. Kode *Stopword Removal*

Tabel 6 menunjukkan contoh perbandingan teks sebelum dan sesudah proses *stopword removal*.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Tabel 6. Perbandingan *Stopword*

Sebelum *Stopword*	Sesudah *Stopword*
[return, nya, lamaaaa]	[returnya]
[banyak, promo]	[promo]
[byk, diskon, nya, tbh, mantap, ja]	[diskon, nya, mantap]

f) Stemming

Tahap ini bertujuan mengubah setiap kata berimbuhan menjadi kata dasar sehingga kata yang memiliki makna sama.

```
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stemming(tokens):
    return [stemmer.stem(token) for token in tokens]
```

Gambar 11. Kode *Stemming*

Tabel 7 menampilkan contoh perbandingan kata sebelum dan sesudah proses *stemming*.

Tabel 7. Perbandingan *Stemming*

Sebelum *Stopword*	Sesudah *Stopword*
[returnnya]	[retur]
[bermanfaat, low, budget]	[manfaat, low, budget]
[tingkatkan, kualitas]	[tingkat, kualitas]

3. *Data Transformation*

Pada penelitian ini, *dataset* terlebih dahulu dibagi menjadi data pelatihan (*data training*) dan data pengujian (*data testing*).

Tabel 8. Jumlah *Data Splitting*

Rasio Pembagian	*Data Training*	*Data Testing*
90:10	2046	228
80:20	1819	455
70:30	1591	683

Setelah proses *splitting*, dilakukan pembobotan kata menggunakan metode TF-IDF. Pembobotan hanya dibangun dari *data training*, kemudian model pembobotan tersebut diterapkan pada *data testing* untuk menghindari terjadinya *data leakage*.

```
vectorizer = TfidfVectorizer(
    max_features=5000
)

X_train1_tfidf = vectorizer.fit_transform(X_train1)
X_test1_tfidf = vectorizer.transform(X_test1)
```

Gambar 11. Kode Proses TF-IDF

Hasil dari proses TF-IDF berupa matriks fitur numerik yang merepresentasikan setiap ulasan dalam bentuk vektor kata.

4. *Data Mining*

Pada tahap ini dilakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi Qpon menggunakan algoritma *Support Vector Machine* (SVM). Pengujian dilakukan dengan membandingkan empat jenis kernel SVM pada seluruh skenario pembagian data.

```
all_results = {}

for kernel in kernels:
    print(f"\n========== Kernel {kernel.upper()} ==========")
    model = SVC(kernel=kernel)

    for name, X_train, X_test, y_train, y_test in scenarios:
        model.fit(X_train, y_train)
        y_pred = model.predict(X_test)
        acc = accuracy_score(y_test, y_pred)

        print(f"{name} = {acc:.4f}")

        # Simpan hasil
        all_results[(kernel, name)] = acc
```

Gambar 11. Kode Proses Klasifikasi SVM

Hasil pengujian menunjukkan bahwa perbedaan jenis kernel dan rasio pembagian data memengaruhi performa klasifikasi model.

Tabel 9. Hasil Akurasi Klasifikasi SVM

Rasio Pembagian	Kernel	*Accuracy*
Skenario 1 (90:10)	*Linear*	92,98%
	RBF	92,54%
	Polynomial	88,60%
	Sigmoid	92,98%
Skenario 2 (80:20)	*Linear*	92,53%
	RBF	92,97%
	Polynomial	87,69%
	Sigmoid	92,09%
Skenario 3 (70:30)	*Linear*	91,65%
	RBF	94,24%
	Polynomial	87,70%
	Sigmoid	91,22%

Berdasarkan hasil tersebut, perbedaan jenis kernel dan rasio pembagian data berpengaruh terhadap nilai akurasi model. Pada skenario 90:10, kernel *Linear* dan *Sigmoid* menghasilkan akurasi tertinggi sebesar 92,98%, sedangkan kernel *Polynomial* memperoleh nilai terendah sebesar 88,60%. Pada skenario 80:20, kernel RBF memberikan akurasi tertinggi sebesar 92,97%, sementara *Polynomial* tetap menunjukkan performa terendah sebesar 87,69%. Adapun pada skenario 70:30, kernel *RBF* kembali menjadi yang terbaik dengan akurasi sebesar 94,24%, yang sekaligus merupakan nilai tertinggi dari seluruh pengujian. Secara umum, kernel *Linear* dan *RBF* menunjukkan performa yang lebih stabil dibandingkan kernel *Polynomial*.

5. *Evaluation*

Evaluasi dilakukan menggunakan metode *confusion matrix* dengan membandingkan hasil prediksi model terhadap label aktual pada *data testing*.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

```
Classification Report - Skenario 1 - 90:10 - LINEAR:
              precision    recall  f1-score   support

    negative     0.9248    0.9535    0.9389       129
    positive     0.9368    0.8990    0.9175        99

    accuracy                         0.9298       228
   macro avg     0.9308    0.9262    0.9282       228
weighted avg     0.9300    0.9298    0.9296       228
```

Gambar 12. Hasil *Classification Report*

Berdasarkan hasil tersebut, diperoleh nilai *accuracy* sebesar 0,9298 dengan rincian kelas negatif memiliki *precision* 0,9248 dan *recall* 0,9535, sedangkan kelas positif memiliki *precision* 0,9368 dan *recall* 0,8990. Nilai *f1-score* pada kedua kelas menunjukkan keseimbangan performa model dalam mengklasifikasikan sentimen positif maupun negatif.

Tabel 10. Hasil Akurasi Klasifikasi *SVM*

No	Kernel	Akurasi	Presisi	*Recall*	*F1-Score*
1	*Linear*	92,98%	93,00%	92,98%	92,96%
	RBF	92,54%	92,58%	92,54%	92,52%
	Poly	88,60%	89,19%	88,60%	88,43%
	Sigmoid	92,98%	93,00%	92,98%	92,96%
2	*Linear*	92,53%	92,52%	92,53%	92,52%
	RBF	92,97%	92,97%	92,97%	92,95%
	Poly	87,69%	88,38%	87,69%	87,43%
	Sigmoid	92,09%	92,08%	92,09%	92,08%
3	*Linear*	91,65%	91,65%	91,65%	91,65%
	RBF	92,24%	92,23%	92,24%	92,22%
	Poly	87,70%	88,48%	87,70%	87,38%
	Sigmoid	91,22%	91,23%	91,22%	91,22%

Berdasarkan tabel tersebut, performa model dipengaruhi oleh jenis kernel dan rasio pembagian data. Kernel *Linear* dan *RBF* menunjukkan kinerja paling baik, dengan hasil tertinggi pada skenario 90:10 untuk kernel *Linear* dan *Sigmoid* yang mencapai accuracy 92,98%, serta pada skenario 80:20 untuk kernel *RBF* dengan accuracy 92,97%. Sebaliknya, kernel *Polynomial* secara konsisten menghasilkan performa terendah pada seluruh skenario pengujian. Hal ini menunjukkan bahwa kernel *Linear* dan *RBF* lebih mampu memodelkan pola data ulasan pengguna dibandingkan kernel *Polynomial*, sehingga pemilihan kernel menjadi faktor penting dalam meningkatkan kinerja klasifikasi sentimen.

6. *Knowledge*

Tahap *knowledge* dilakukan pengolahan lebih lanjut terhadap hasil klasifikasi guna mengetahui kecenderungan opini pengguna serta aspek yang paling sering dibahas dalam ulasan.



Gambar 13. *Word Cloud* Sentimen Positif

Berdasarkan Gambar 13, kata-kata dominan pada sentimen positif antara lain diskon, mantap, bagus, hemat, mudah, terima kasih, dan promo. Hal ini menunjukkan bahwa ulasan positif umumnya berkaitan dengan manfaat ekonomi, kemudahan penggunaan aplikasi, serta kepuasan terhadap promo yang diberikan. Dengan demikian, pengalaman penggunaan yang lancar dan keuntungan finansial menjadi faktor utama terbentuknya sentimen positif pengguna.



Gambar 13. *Word Cloud* Sentimen Negatif

Berdasarkan Gambar 14, kata-kata dominan pada sentimen negatif meliputi login, tidak, refund, akun, bayar, dan *voucher*. Hal tersebut menunjukkan bahwa keluhan pengguna terutama berkaitan dengan kendala akses akun, penggunaan *voucher*, serta proses pembayaran dan pengembalian dana. Selain itu, kemunculan kata seperti error, gagal, dan kecewa mengindikasikan adanya masalah teknis dan layanan yang memengaruhi persepsi pengguna terhadap kualitas aplikasi.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa kombinasi metode TF-IDF dan algoritma SVM, khususnya dengan kernel *Linear* dan *RBF*, efektif dalam melakukan klasifikasi sentimen ulasan aplikasi dengan tingkat akurasi yang konsisten di atas 92%. Selain menunjukkan stabilitas performa pada berbagai skenario pembagian data, model juga mampu merepresentasikan perbedaan karakteristik linguistik antara sentimen positif dan negatif.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

V. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh beberapa kesimpulan sebagai berikut:

1. Kombinasi metode TF-IDF dan algoritma *Support Vector Machine* mampu melakukan klasifikasi sentimen ulasan pengguna aplikasi Qpon dengan sangat baik, dengan tingkat akurasi yang konsisten di atas 92%.
2. Pemilihan jenis kernel memengaruhi performa model, di mana kernel *Linear* dan *RBF* menunjukkan kinerja terbaik dibandingkan kernel *Polynomial* pada berbagai skenario pembagian data.
3. Hasil analisis menunjukkan bahwa sentimen positif didominasi oleh kata-kata yang berkaitan dengan diskon, promo, dan kemudahan penggunaan aplikasi, sedangkan sentimen negatif didominasi oleh keluhan terkait *login*, *voucher*, pembayaran, dan proses *refund*, sehingga hasil klasifikasi dapat dimanfaatkan sebagai bahan evaluasi kualitas layanan aplikasi.

Saran

Berdasarkan hasil dan keterbatasan penelitian, beberapa saran untuk pengembangan selanjutnya adalah sebagai berikut:

1. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan periode pengambilan data yang lebih panjang agar hasil analisis lebih representatif.
2. Penelitian berikutnya dapat membandingkan metode SVM dengan algoritma klasifikasi lain atau pendekatan *deep learning* untuk memperoleh model dengan performa yang lebih optimal.

DAFTAR PUSTAKA

Abidin, M. A. Z. (2025). Evaluasi Sentimen Ulasan Pengguna CGV Cinemas Indonesia Menggunakan Metode Naïve Bayes dan Support Vector Machine. *Indonesian Journal on Software Engineering (IJSE)*, *11*(1 JUNI), 44–53.

Ardianto, R., Ramdhani, R. F., Dewi, L. O. A., Prabowo, A., Saputri, Y. W., Lestari, A. S., & Hadi, N. (2024). Transformasi digital dan antisipasi perubahan ekonomi global dalam dunia perbankan. *MARAS: Jurnal Penelitian Multidisiplin*, *2*(1), 80–88.

Bazar, I., Wjidi, F., & Cirua, A. A. A. (2025). Analisis Sentimen Ulasan Aplikasi Wondr By Bni Menggunakan Algoritma Svm Dengan Optimasi Kernel Trick. *Jurnal Ilmiah Teknik Dan Ilmu Komputer*, *4*(2), 69–81. https://doi.org/10.55123

Ceci, L. (2024, August 12). *Mobile App Downloads Worldwide 2024 | Statista*. https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/

Himawan, K. (2025). *Analisis Sentimen dan Pemodelan Topik terhadap Ulasan Aplikasi Mobile JKN Menggunakan SVM dan LDA*. Institut Teknologi Kalimantan.

Huda, R., Permata Sari, F., & Auriga, W. (2025). Penerapan Support Vector Machine dalam Analisis Sentimen Ulasan Pluang di Google Play Store Implementation of Support Vector Machine for Sentiment Analysis of Pluang Reviews on Google Play Store. *Technologica*, *4*(2), 219–231.

Liu, B. (2022). *Sentiment Analysis and Opinion Mining*. Springer International Publishing. https://books.google.co.id/books?id=xYhyEAAAQBAJ

Nurian, A., Padilah, T. N., & Garno, G. (2024). Analisis Sentimen Terhadap Pelayanan Disdukcapil Karawang Menggunakan Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2). https://doi.org/10.23960/jitet.v12i2.4178

Nurwahidah, D., Dwilestari, G., Nuris, N. D., & Narasati, R. (2023). Analisis Sentimen Data Ulasan Pengguna Aplikasi Google Kelas Pada Google Play Store Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3673–3678.

Play Store. (2025). *Qpon - Ulasan Pengguna di Google Play*. https://play.google.com/store/apps/details?id=com.qpon.platform&hl=id

Pratama, E. A., Primajaya, A., & Maulana, I. (2025). Penerapan Algoritma Support Vector Machine Terhadap Analisis Sentimen Pada Ulasan Aplikasi LinkAja. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(3).

Putri, R. K., & Athoillah, M. (2021). Support Vector Machine Untuk Identifikasi Berita Hoax Terkait Virus Corona (Covid-19). *Jurnal Informatika: Jurnal Pengembangan IT*, *6*(3), 162–167.

Qpon. (2025). *Tentang Qpon: Ketahui Lebih Dalam Tentang Kami*. https://qpon.id/about

Sariatin, S., & Ekawati, C. (2023). Faktor-faktor yang Mempengaruhi Kepuasan Pelanggan. *Jurnal Informatika Ekonomi Bisnis*, 1258–1261.

Satrio, R. B. (2025). *Analisis Sentimen Ulasan Pengguna Aplikasi AdaKami dengan Pendekatan Hybrid Menggunakan Ensemble*

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Stacking Naive Bayes dan Support Vector Machine. Universitas Amikom Purwokerto.

Shimaoka, A. M., Ferreira, R. C., & Goldman, A. (2024). The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks. *SBC Computing Reviews*, *4*(1), 28–43. https://doi.org/10.5753/reviews.2024.3757

Sudrajat, M. R., Atika, P. D., & Herlawati. (2021). Implementasi Support Vector Machine (SVM) dan Naive Bayes untuk Analisis Sentimen Aplikasi KAI Access. *Jurnal ICT : Information Communication & Technology*, *20*(2), 254–259.

Sugiarto, A. D., & Utomo, M. S. (2025). Analisis Sentimen Ulasan Pengguna Bca Mobile Di Google Play Store Menggunakan Metode Decision Tree. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(5), 8004–8009. https://doi.org/10.36040/jati.v9i5.14969

Wahyudin, H. (2025). Analisis Sentimen terhadap Data Ulasan Aplikasi Mobile Banking di Play Store. *Bandung Conference Series: Statistics*, *5*. https://doi.org/10.29313/bcss.v5i2.20693

Yani, F., Imelda, K., & Kurniadi, N. T. (2025). Analisis Sentimen Pengguna Aplikasi Mobile Legend di Playstore dengan Menggunakan Algoritma Naïve Bayes dan Support Vector Machine (SVM). *Jurnal Teknologi Informasi*, *6*(2). https://doi.org/10.46576/djtechno