



Analisis Sentimen terhadap Ulasan Aplikasi Canva di Play Store dengan Menggunakan Pendekatan Lexicon dan Algoritma Decision Tree

Wulandari Wulandari, Nofiyani Nofiyani, Yesi Puspita Dewi

Penerapan Database Life Cycle Dalam Merancang Basis Data Relasional Pengolahan Data Teks
Nofiyani Nofiyani, Yesi Puspita Dewi

Implementasi Sistem Informasi Gudang Berbasis Desktop pada PT. Sistem Aksesindo Perdana Cabang Surabaya

David Saputra Octadianto Soedargo, Ariel Kristianto, Hermawan Andika

Optimasi Penggunaan Enkripsi Homomorfik untuk Keamanan dan Privasi Data Kesehatan (Studi Literatur)

Muhammad Harun Yahya, Chrisdiego Chrisanto, Abdu Afin Arsy

Analisis Sentimen Terhadap Komentar Video Short YouTube Mengenai Paylater Menggunakan Algoritma Bert

Fakhrezi Fajris, Imelda Imelda

Implementasi Metode Jaringan Saraf Tiruan Backpropagation Untuk Memprediksi Harga Emas

Nico Nico, Pipin Farida Ariyani

Pembuatan GH3RTZ Framework untuk Proses Reconnaissance dan Deteksi Cela Keamanan

Muhamad Givari Ramadan, Mardi Hardjianto

Rancang Bangun Sistem Presensi Asisten Laboratorium Komputer Berbasis Internet Of things (IoT)

Danang Ade Muktiawan, Adimas Aglasia, Bayu Nugroho

**ASOSIASI PERGURUAN TINGGI INFORMATIKA & ILMU KOMPUTER
(APTIKOM) WILAYAH 3**

Sekretariat Redaksi :

DRPM Universitas Budi Luhur

Jl. Raya Ciledug, Petukangan Utara, Jakarta Selatan 12260

Telp. 021.5853753 Fax. 021.5869225

Jurnal Ticom: Technology of Information and Communication

Pelindung

Dr. Didi Rosiyadi, M.Kom.
(Ketua APTIKOM Provinsi DKI Jakarta)

Penanggung Jawab

Prof. Dr. Achmad Nizar Hidiyanto, M.Kom.

Ketua Dewan Editor

Dr. Mohammad Syafrullah, M.Kom., M.Sc.

Anggota Dewan Editor

Dr. Agus Subekti, M.T.
Dr. Rusdah, S.Kom., M.Kom.
Dr. Yan Everhard, M.T.
Tri Wahyu Widyaningsih, S.Kom., M.M.S.I
Dra. Andiani, M.Kom.
Nani Tachjar, S.Kom., M.T.
Samsinar., S.Kom, M.Kom.
Agnes Aryasanti, S.Kom, M.Kom.

Alamat Sekretariat Redaksi
DRPM Universitas Budi Luhur

Jl. Raya Ciledug No.99, RT.10/RW.3, Petukangan Utara
Pesanggrahan, Kota Jakarta Selatan, DKI Jakarta 12260

E-mail: jurnalticom@budiluhur.ac.id

Jurnal Ticom: Technology of Information and Communication adalah jurnal ilmiah dalam bidang teknologi informasi dan komunikasi (TIK) yang diterbitkan oleh Asosiasi Pendidikan Tinggi Informatika dan Ilmu Komputer (APTIKOM) Provinsi DKI Jakarta. Jurnal TICOM terbit 3 kali dalam satu tahun yaitu: September, Januari dan Mei.

KATA PENGANTAR

Dengan mengucapkan puji syukur kehadiran Tuhan Yang Maha Esa atas berkat, rahmat dan karunia-Nya sehingga Jurnal Ilmiah Ticom Volume 13 Nomor 2 Januari 2025 dapat terbit sesuai yang direncanakan.

Jurnal penelitian ini terbit sebagai bentuk kepedulian APTIKOM DKI Jakarta dalam meningkatkan mutu penelitian dan publikasi yang dilakukan oleh Dosen, Peneliti, Mahasiswa ataupun Praktisi di perguruan tinggi. Pada Jurnal TICOM Volume 13 Nomor 2 Januari 2025 memuat artikel pada topik, Text Mining, Information System, Kriptografi, AI, Network Security, Semoga Jurnal TICOM dapat menjadi referensi bagi para peneliti di Indonesia dan meningkatkan kualitas dari publikasi penelitian di Indonesia.

Seluruh Dewan Redaksi Jurnal Ticom mengucapkan terima kasih kepada penulis sebagai penyumbang artikel ilmiah, karena tanpa sumbangan artikel ilmiah dan penelitian dari penulis maka mustahil jurnal ilmiah TICOM dapat diterbitkan, terima kasih juga kepada semua pihak yang selalu memberikan dukungan kepada Jurnal TICOM sehingga dapat hingga saat ini.

Terima kasih dan selamat membaca

Jakarta, Januari 2025

Ketua Dewan Editor
Jurnal TICOM

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI.....	ii
Analisis Sentimen terhadap Ulasan Aplikasi Canva di Play Store dengan Menggunakan Pendekatan Lexicon dan Algoritma Decision Tree	
Wulandari Wulandari, Nofiyani Nofiyani, Yesi Puspita Dewi	57-63
Penerapan Database Life Cycle Dalam Merancang Basis Data Relasional Pengolahan Data Teks	
Nofiyani Nofiyani, Yesi Puspita Dewi.....	64-68
Implementasi Sistem Informasi Gudang Berbasis Desktop pada PT. Sistem Aksesindo Perdana Cabang Surabaya	
David Saputra Octadianto Soedargo, Ariel Kristianto, Hermawan Andika.....	69-73
Optimasi Penggunaan Enkripsi Homomorfik untuk Keamanan dan Privasi Data Kesehatan (Studi Literatur)	
Muhammad Harun Yahya, Chrisdiego Chrisanto, Abdu Afin Arsy	74-78
Analisis Sentimen Terhadap Komentar Video Short YouTube Mengenai Paylater Menggunakan Algoritma Bert	
Fakhrezi Fajris, Imelda Imelda	79-84
Implementasi Metode Jaringan Saraf Tiruan Backpropagation Untuk Memprediksi Harga Emas	
Nico Nico, Pipin Farida Ariyani	85-91
Pembuatan GH3RTZ Framework untuk Proses Reconnaissance dan Deteksi Cela Keamanan	
Muhamad Givari Ramadan, Mardi Hardjianto	92-97
Rancang Bangun Sistem Presensi Asisten Laboratorium Komputer Berbasis <i>Internet of Things</i> (IoT)	
Danang Ade Muktiawan, Adimas Aglasia, Bayu Nugroho	98-103

Analisis Sentimen terhadap Ulasan Aplikasi Canva di Play Store dengan Menggunakan Pendekatan Lexicon dan Algoritma Decision Tree

Wulandari^{1*}, Nofiyani², Yesi Puspita Dewi³

^{1,2,3}Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, DKI Jakarta, Indonesia
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260. DKI Jakarta
E-mail: ¹wulandari@budiluhur.ac.id, ²nofiyani@budiluhur.ac.id, ³yesi.puspitadewi@budiluhur.ac.id
(*: corresponding author)

Abstrak— Aplikasi Canva merupakan aplikasi yang digunakan untuk desain grafis seperti pembuatan presentasi online, poster, infografis, dan kebutuhan desain. Analisis sentimen aplikasi Canva dilakukan dengan mengumpulkan ulasan dari pengguna aplikasi tersebut. Ulasan pengguna yang dikumpulkan melalui teknik scraping menggunakan google colab dan dianalisis untuk mengidentifikasi sentimen positif, netral, dan negatif. Data yang diperoleh kemudian diolah pada tahapan preprocessing data, seperti cleaning, tokenisasi, normalisasi, dan pelabelan otomatis menggunakan metode lexicon Vader. Tujuan pada penelitian ini yaitu untuk menganalisis ulasan pengguna aplikasi Canva di Play Store dengan menggunakan pendekatan lexicon dan algoritma Decision Tree. Model klasifikasi Decision Tree digunakan untuk mengolah data yang telah dilabeli dengan hasil evaluasi menunjukkan akurasi sebesar 87%. Performa terbaik dicapai pada kelas "2" (positif) dengan nilai precision 90%, recall 89%, dan F1-score 89%, yang mencerminkan keseimbangan yang baik antara precision dan recall. Hasil penelitian ini membuktikan bahwa algoritma Decision Tree efektif dalam mengklasifikasikan sentimen ulasan. Dengan memanfaatkan hasil analisis ini, pengembang aplikasi Canva dapat memperoleh wawasan berharga untuk meningkatkan kualitas layanan mereka, baik dalam menghadapi kebutuhan pengguna maupun mengatasi keluhan yang muncul. Penelitian ini memberikan kontribusi nyata dalam pemanfaatan analisis sentimen untuk mendukung pengembangan produk berbasis data.

Kata Kunci: analisis sentimen, lexicon, Decision Tree, Play Store, Canva.

Abstract— The Canva app is an application used for graphic design such as creating online presentations, posters, infographics, and design needs. Sentiment analysis of the Canva application is done by collecting reviews from users of the application. User reviews are collected through scraping techniques using google colab and analyzed to identify positive, neutral, and negative sentiments. The data obtained is then processed at the data preprocessing stage, such as cleaning, tokenization, normalization, and automatic labeling using the Vader lexicon method. The purpose of this research is to analyze user reviews of the Canva application on the Play Store using a lexicon approach and the Decision Tree algorithm. The Decision Tree classification model was used to process the labeled data with the evaluation results showing an accuracy of 87%. The best performance was achieved in class "2" (positive) with 90% precision, 89% recall, and 89% F1-score, reflecting a good balance between precision and recall. The results of this study prove that the

Decision Tree algorithm is effective in classifying the sentiment of reviews. By utilizing the results of this analysis, Canva app developers can gain valuable insights to improve their service quality, both in dealing with user needs and addressing complaints that arise. This research makes a real contribution to the utilization of sentiment analysis to support data-driven product development.

Keyword— sentiment analysis, lexicon, Decision Tree, Play Store, Canva.

I. PENDAHULUAN

Dalam era digital, aplikasi berbasis teknologi menjadi bagian penting kehidupan masyarakat dalam kesehariannya. Salah satunya aplikasi Canva yang dikenal populer, sebuah platform desain grafis yang diciptakan pada tahun 2012 oleh pengusaha Australia Melanie Perkins[1]. Canva merupakan *software* yang menyediakan berbagai produk desain online, termasuk membuat power point, membuat resume, membuat poster atau brosur dan masih banyak lagi fungsi dari aplikasi Canva. Seiring dengan popularitasnya, Canva mendapatkan banyak ulasan dari pengguna di platform seperti Google Play Store. Ulasan ini membahas kelebihan dan kekurangan pembuatan desain dengan Canva dan bagaimana Canva dapat membantu pengguna[1]. Berdasarkan ulasan tersebut memungkinkan platform tersebut untuk melakukan perbaikan dan pengembangan yang lebih tepat.

Analisis sentimen adalah metode yang digunakan untuk mengidentifikasi dan mengumpulkan informasi subjektif contoh opini, emosi, atau perasaan dari sebuah kalimat atau teks[2]. Analisis sentimen bertujuan untuk mengevaluasi tingkat kepuasan user terhadap sebuah produk. Namun, banyaknya ulasan yang tersedia membuat analisis manual menjadi tidak praktis. Untuk itu diperlukan metode otomatis untuk menganalisis sentimen pengguna secara efisien. Analisis sentimen merupakan studi komputasi yang bertujuan untuk memahami opini, sentimen, dan emosi yang disampaikan melalui teks. Dalam konteks ini, teks dalam kalimat akan mengelompokkan untuk mengetahui pendapat yang dikemukakan apakah bersifat positif, negatif atau netral[3]. Salah satu tantangan dalam analisis sentimen adalah labeling data yang belum diberi anotasi. Untuk mengatasi ini, pendekatan dalam menggunakan lexicon-based. Pendekatan lexicon-based melakukan pelabelan pada data training, untuk

mengurangi pelabelan secara personal karena pelabelan dilakukan secara otomatis oleh sistem[4].

Model klasifikasi yang diterapkan dalam penelitian ini menggunakan Algoritma Decision Tree, model decision tree dipilih karena merupakan metode statistik yang sederhana, memiliki tingkat ketepatan yang tinggi, serta menghasilkan tingkat kesalahan yang minimal dalam proses klasifikasi. [5] sehingga dengan menggabungkan kedua metode ini diharapkan dapat meningkatkan hasil akurasi yang baik [4]. Tujuan penelitian ini adalah untuk menerapkan pendekatan lexicon pada labeling data sentimen ulasan pengguna aplikasi Canva di Play Store, melatih model Decision Tree untuk mengklasifikasikan sentimen ulasan berdasarkan data yang telah dilabeli dan mengevaluasi performa model dan memberikan rekomendasi berdasarkan hasil analisis. Melalui penelitian ini, diharapkan dapat diperoleh wawasan yang berguna bagi pengembang aplikasi Canva untuk meningkatkan kualitas produk mereka berdasarkan ulasan pengguna.

II. METODE PENELITIAN

Tahapan dalam penelitian ini menunjukkan tahapan dan rancangan penelitian dari tahap awal sampai akhir. Gambar 1 berikut menunjukkan tahapan penelitian.



Gambar. 1 Tahapan Penelitian

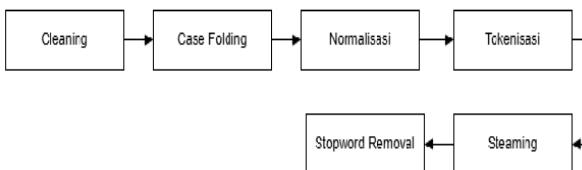
Keterangan:

1. Pengumpulan Data

Dataset dalam penelitian ini merupakan ulasan Aplikasi Canva pada playstore. Teknik yang digunakan yaitu scraping ulasan memakai salah satu library *google-play-scraper* pada Python. Dataset dibatasi 2000 ulasan yang paling relevan.

2. Preprocessing

Pada tahap preprocessing dataset yang masih mentah kemudian diolah dengan cara menghilangkan data yang tidak perlu dan tidak relevan ke dalam format yang dibutuhkan. Tahapan preprocessing tersebut diantaranya *cleaning*, *case folding*, *normalisasi*, *tokenize*, *stopword removal*, *stemming data*.



Gambar.2 Alur Preprocessing

3. Pelabelan

Pada tahap labelling data tidak dilakukan pelabelan manual, pelabelan dilakukan secara otomatis menggunakan kamus lexicon Vader. Proses pelabelan lexicon yaitu memeriksa kata yang mengandung sentimen positif, negatif, atau netral berdasarkan kamus sentimen (lexicon). Frekuensi kemunculan kata-kata ini dalam sebuah dokumen/teks dihitung

untuk menentukan labelnya[6]. Dengan pelabelan otomatis lebih mempercepat proses pelabelan data.

4. Klasifikasi

Pada tahapan klasifikasi metode yang digunakan adalah *Decision tree*. Decision Tree merupakan algoritma yang mampu menghasilkan keputusan melalui pembentukan struktur pohon keputusan[7]. Definisi lain Decision Tree merupakan algoritma pembelajaran yang memanfaatkan struktur berbentuk pohon untuk mengambil keputusan. Pohon keputusan terdiri dari simpul-simpul (nodes) yang merepresentasikan fitur atau atribut, cabang-cabang yang menggambarkan keputusan yang mungkin, dan daun-daun (leaves) yang menunjukkan hasil akhir atau kelas. Algoritma ini bekerja dengan cara memecah dataset menjadi subset berdasarkan aturan tertentu, hingga ditemukan struktur pohon yang optimal untuk memprediksi atau mengklasifikasikan data dengan akurasi tinggi.

5. Evaluasi

Untuk menguji seberapa baik metode yang digunakan pada tahapan klasifikasi maka perlu dilakukan evaluasi, evaluasi pada penelitian ini menggunakan *confusion matrix*. Pada tahap evaluasi bertujuan untuk menilai sejauh mana keberhasilan model yang telah dikembangkan[8]. *Confusion matrix* merupakan salah satu alat analisis prediktif yang berfungsi untuk menampilkan dan membandingkan nilai aktual atau nilai yang sebenarnya dengan nilai yang diprediksi oleh model. Alat ini dapat dimanfaatkan untuk menghasilkan berbagai metrik evaluasi, termasuk perhitungan Nilai Akurasi, Nilai Presisi, Nilai Recall, dan Nilai F1-Score, berikut rumus perhitungannya:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Presisi} \cdot \text{Recall}}{\text{Presisi} + \text{Recall}} : \quad (4)$$

Keterangan:

- a. **True Negative (TN):** Data yang bernilai negatif dan berhasil terdeteksi dengan benar sebagai negatif.
- b. **False Positive (FP):** Data yang bernilai negatif tetapi salah terdeteksi sebagai positif.
- c. **True Positive (TP):** Data yang bernilai positif dan berhasil terdeteksi dengan benar sebagai positif.
- d. **False Negative (FN):** Data yang bernilai positif tetapi salah terdeteksi sebagai negatif.

III. HASIL DAN PEMBAHASAN

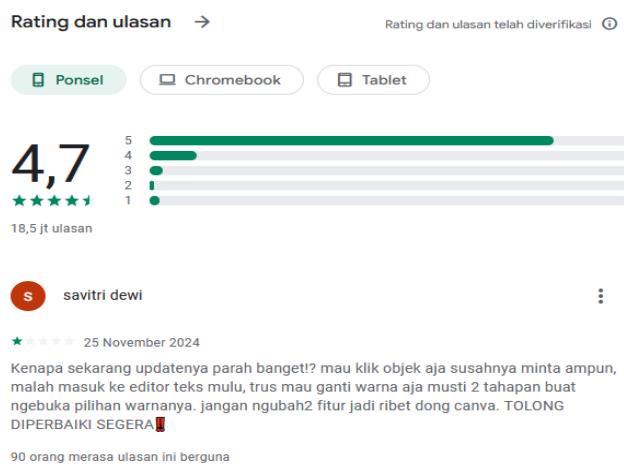
Hasil dan pembahasan pada penelitian dibuat dalam beberapa tahapan sebagai berikut:

A. Pengumpulan Data

Dataset yang dipakai dalam penelitian ini adalah ulasan yang terdapat pada Aplikasi Canva yang tersedia playstore yang merupakan platform resmi aplikasi android[9]. Pengguna aplikasi Canva dapat mendownload di playstore dan dapat berbagi pengalaman dengan memberikan ulasan pada aplikasi tersebut. Berikut tampilan Canva dan ulasan yang ditunjukkan pada gambar 3 dan 4.



Gambar. 3 Tampilan Aplikasi Canva



Gambar 4. Tampilan Jumlah Ulasan dan Rating aplikasi Canva

Teknik yang digunakan untuk menggambil data yaitu *scraping* data memakai library Python yaitu *google-play-scrapers*. Data yang diambil dibatasi 2000 ulasan berdasarkan ulasan yang paling relevan. Gambar 5 menunjukkan gambaran proses pengambilan data.



Gambar 5. Alur proses scraping data

Berikut code program dan hasil dari proses *scraping* data menggunakan google colab ditunjukkan pada gambar 6.

```
from google_play_scraper import Sort, reviews
result, continuation_token = reviews(
    'com.canva.editor',
    lang='id',
    country='id',
    sort=Sort.MOST_RELEVANT,
    count=2000,
    filter_score_with=None
)
```

Gambar 6. Program untuk scraping data

B. Preprocessing Data

Pada tahapan *preprocessing* dataset yang diperoleh kemudian diproses agar dataset dibersihkan dari symbol - symbol yang tidak diperlukan pada penelitian ini, data yang kurang lengkap, data yang tidak valid, serta atribut yang kurang yaitu:

1) Cleaning

Tahap *cleaning* yaitu data yang dibersihkan membersihkan data dari elemen-elemen yang kurang relevan yaitu menghapus tanda baca, emoji, angka, dan elemen lain yang tidak diperlukan tahap tersebut ditunjukkan pada gambar 7.

content	cleaning
keren banget mudah digunakan	keren banget mudah digunakan
Mantap pula buat editan nya ni. Cuman iklan nya...	Mantap pula buat editan nya ni Cuman iklan nya...
TK bisa login padahal udh daftar pakai akun be...	TK bisa login padahal udh daftar pakai akun be...
Ini kenapa udah seminggu mw edit susah bgt lam...	Ini kenapa udah seminggu mw edit susah bgt lam...
paling bisa diandalkan buat bikin konten sy su...	paling bisa diandalkan buat bikin konten sy su...

Gambar 7. Tahapan Cleaning

2) Case folding

Pada tahap *case folding* semua huruf kapital diubah menjadi huruf kecil, tahap *casefolding* ditunjukkan pada gambar 8.

cleaning	case_folding
keren banget mudah digunakan	keren banget mudah digunakan
Mantap pula buat editan nya ni Cuman iklan nya...	mantap pula buat editan nya ni cuman iklan nya...
TK bisa login padahal udh daftar pakai akun be...	tk bisa login padahal udh daftar pakai akun be...
Ini kenapa udah seminggu mw edit susah bgt lam...	ini kenapa udah seminggu mw edit susah bgt lam...
paling bisa diandalkan buat bikin konten sy su...	paling bisa diandalkan buat bikin konten sy su...

Gambar 8. Tahap case folding

3) Normalisasi

Pada tahap *normalisasi* yaitu mengubah kata tidak baku menjadi kata baku berdasarkan kamus baku yang telah disediakan, tahap normalisasi ditunjukkan pada gambar 9.

case_folding	normalisasi
keren banget mudah digunakan	keren banget mudah digunakan
mantap pula buat editan nya ni cuman iklan nya...	mantap pula buat editan ya nih cuman iklan ya ...
tk bisa login padahal udh daftar pakai akun be...	tak bisa login padahal sudah daftar pakai akun...
ini kenapa udah seminggu mw edit susah bgt lam...	ini kenapa sudah seminggu mau edit susah bange...
paling bisa diandalkan buat bikin konten sv su...	paling bisa diandalkan buat bikin konten saya

Gambar 9. Tahap Normalisasi

4) Tokenization

Pada tahap *tokenization* memecah kalimat menjadi kata, tahap tokenization ditunjukkan pada gambar 10.

normalisasi	tokenize
keren banget mudah digunakan	[keren, banget, mudah, digunakan]
mantap pula buat editan ya nih cuman iklan ya ...	[mantap, pula, buat, editan, ya, nih, cuman, i...]
tak bisa login padahal sudah daftar pakai akun...	[tak, bisa, login, padahal, sudah, daftar, pak...]
ini kenapa sudah seminggu mau edit susah bange...	[ini, kenapa, sudah, seminggu, mau, edit, susah...]

Gambar 10. Tahap Tokenization

5) Stopword Removal

Pada tahap *Stopword Removal* yaitu menghapus kata-kata yang tidak diperlukan dalam proses preprocessing, contoh kata yaitu **yang**, **di**, **ke** dan masih banyak lagi kata terdapat dalam kamus sastrawi yang disediakan oleh python, tahap stopword removal ditunjukkan pada gambar 11.

tokenize	stopword removal
[keren, banget, mudah, digunakan]	[keren, banget mudah]
[mantap, pula, buat, editan, ya, nih, cuman, i...]	[mantap, editan ya, nih, cuman iklan, ya, am..]
[tak, bisa, login, padahal, sudah, daftar, pak...]	[login, daftar pakai, akun belajarid]
[ini, kenapa, sudah, seminggu, ma...]	[seminggu, edit susah, banget loading, ya]

Gambar 11. Tahap Stopword Removal

6) Stemming Data

Pada tahap *Steaming* yaitu mengubah kata menjadi kata dasar, contoh pemimpin => pimpin, tahap *steamming* ditunjukkan pada gambar 12.

stopword removal	stemming_data
[keren, banget, mudah]	keren banget mudah
[mantap, editan, ya, nih, cuman, iklan, ya, am..]	mantap edit ya nih cuman iklan ya ampun
[login, daftar, pakai, akun, belajarid]	login daftar pakai akun belajarid
[seminggu, edit, susah, banget, loading, ya]	minggu edit susah banget loading ya
[diandalkan, bikin, konten, menggunakan, go...]	andal bikin konten guna gonta ganti akun downl...

Gambar 12. Tahap Steaming

C. Pelabelan Data

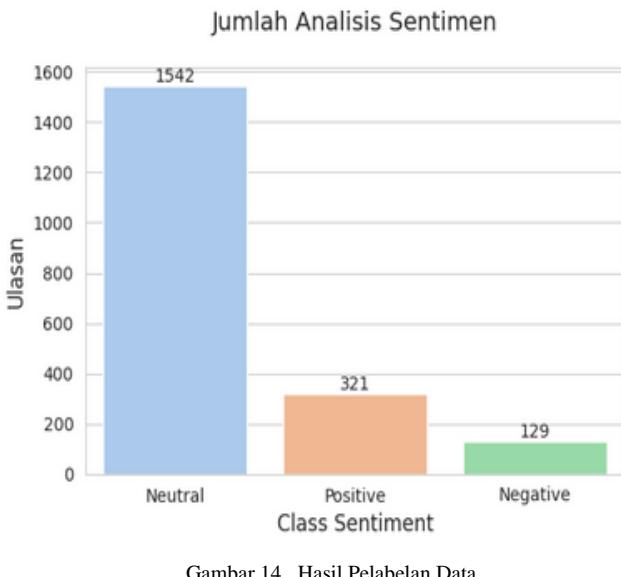
Pelabelan data pada penelitian ini dilakukan menggunakan pelabelan Lexicon Vader, di mana nilai *compound* Vader merupakan hasil kombinasi atau rata-rata dari bobot sentimen. Jika nilai *compound* ≥ 0.5 , maka ulasan tersebut dikategorikan sebagai sentimen positif. Jika nilai *compound* = 0, ulasan termasuk sentimen netral, sedangkan jika *compound* ≤ -0.5 maka ulasan diklasifikasikan sebagai sentimen negatif[10]. Dari 2000 data ulasan, hasil pelabelannya menunjukkan bahwa 321 ulasan memiliki sentimen positif, 129 ulasan memiliki sentimen negatif, dan 1.542 ulasan termasuk dalam sentimen netral. Code program dan hasil pelabelan ini ditampilkan pada gambar 13 dan 14.

```

score = data["Compound"].values
sentiment = []
for i in score:
    if i >= 0.05 :
        sentiment.append('Positive')
    elif i <= -0.05 :
        sentiment.append('Negative')
    else:
        sentiment.append('Neutral')
data["Sentiment"] = sentiment
data.head()
    
```

Gambar 13. Code Program Pelabelan Data

Berikut merupakan grafik jumlah analisis sentimen yang telah terlabeli, dari hasil pelabelan tersebut didapatkan jumlah kalimat netral 1.542 kata, positif 321 kata dan 129 kata. Grafik jumlah kata terlabeli ditunjukkan pada gambar 14.



Gambar 14. Hasil Pelabelan Data

D. Klasifikasi

Pada tahap klasifikasi algoritma yang digunakan adalah Decision Tree. Algoritma Decision Tree terdiri dari node dan cabang, di mana setiap node merepresentasikan atribut atau fitur yang digunakan untuk klasifikasi, sedangkan setiap cabang menunjukkan nilai atau kategori yang mungkin diambil oleh simpul tersebut[11][12]. Berikut rumus Decision Tree[12][13] ditunjukkan pada gambar 15.

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (5)$$

Penjelasan :

- S = Himpunan kasus
- A = Atribut
- N = Jumlah dari partisi S
- Pi = Proporsi dari Si terhadap S

$$Gain(S, A) = Entropy(S) \sum_{k=0}^n \frac{|S_k|}{|S|} + Entropy(S_k) \quad (6)$$

Penjelasan:

- S = himpunan permasalahan
- A = Atribut
- n = jumlah sebagian atribut A
- |Si| = jumlah masalah pada partisi ke-i
- |S| = jumlah masalah dalam

Berikut Code Program klasifikasi menggunakan Decission Tree, program ditunjukkan pada gambar 16.

```

# Membuat model Decision Trees
model = DecisionTreeClassifier(random_state=42)

# Melatih model pada data latih
model.fit(X_train, y_train)

# Melakukan prediksi pada data uji
predictions = model.predict(X_test)\n
    
```

Gambar 16. Code Program Klasifikasi Decision Tree

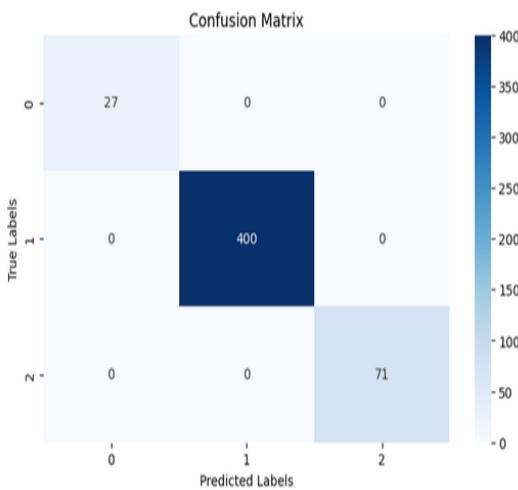
E. Evaluasi

Pada tahap evaluasi, peneliti menggunakan *confusion matrix*. *Confusion matrix* adalah sebuah tabel yang menggambarkan jumlah data uji yang berhasil diklasifikasikan dengan tepat, serta jumlah data uji yang diklasifikasikan secara keliru. [14][15]. Selain itu confucion matrik menyajikan ringkasan hasil prediksi model dibandingkan dengan label aktual dalam berbagai kategori[16]. Melalui analisis terhadap nilai-nilai seperti true positive (TP), false positive (FP), true negative (TN), dan false negative (FN) dalam matriks, praktisi dapat memahami performa model secara mendalam, termasuk mengidentifikasi keunggulan dan kekurangannya[17][16]. Tabel *confusion matrix* bisa dilihat pada tabel 1.

TABEL I
 CONFUSION MATRIX

Nilai Prediksi	Nilai Aktual		
	Label	Positive (1)	Negative (0)
		Positive (1)	TP
	Negative (0)	FN	TN

Kelas atau kategori Hasil evaluasi menggunakan *Confusion matrix* ditunjukan pada gambar 17.



Gambar 7. Hasil *Confusion Matrik*

Nilai akurasi yang didapatkan adalah
0.8721804511278195

Accuracy: 0.8721804511278195

Classification Report:

	precision	recall	f1-score	support
0	0.86	0.83	0.84	130
1	0.86	0.89	0.88	140
2	0.90	0.89	0.89	129
accuracy			0.87	399
macro avg	0.87	0.87	0.87	399
weighted avg	0.87	0.87	0.87	399

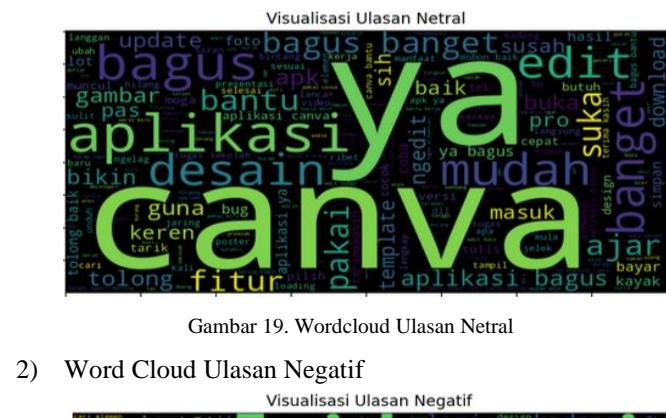
Gambar 18. Nilai Accuracy

Gambar 18 diatas menunjukan hasil evaluasi model klasifikasi Decission Tree dimana kelas "0", "1", dan "2" adalah kategori atau kelas dalam data yang diprediksi oleh model. Hasil *precision* untuk kelas "2" adalah 0.90, berarti 90%, sedangkan recall adalah 0.89, berarti 89% data aktual kelas "1". F1-score memberikan keseimbangan jika terdapat ketimpangan antara precision dan recall. Support Jumlah data aktual untuk setiap kelas. Misalnya, ada 130 data untuk kelas "0". Akurasi keseluruhan model adalah 87% (0.8721804511278195), yang berarti 87% prediksi model benar dari total data. Jadi dapat disimpulkan klasifikasi menggunakan Decision Tree memiliki performa yang cukup baik dengan akurasi 87%, dan performa terbaiknya terlihat pada kelas "2" dengan F1-score 0.89.

F. Visualisasi Word Cloud

Visualisasi Word Cloud setelah proses labeling ditunjukkan pada gambar 19, 20 dan 21.

1) Wordcloud ulasan Netral



Gambar 19. Wordcloud Ulasan Netral

2) Word Cloud Ulasan Negatif



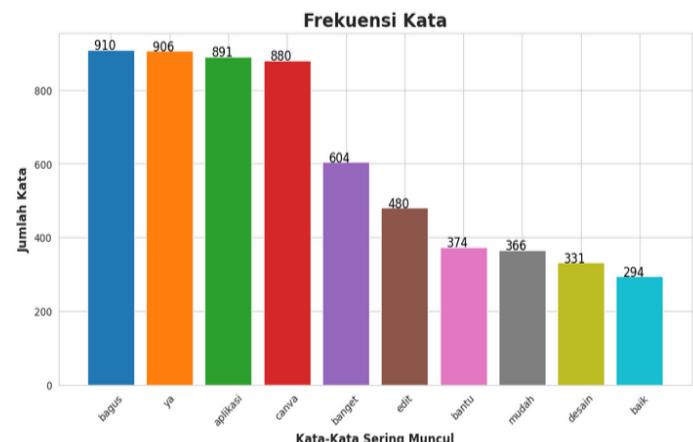
Gambar 20. Wordcloud Ulasan Negatif

3) Word Cloud Ulasan Positif



Gambar 21. Wordcloud Ulasan Netral

Berikut frekuensi kata yang sering muncul dari analisis sentimen pada penelitian ini ditunjukkan pada gambar 22.



Gambar 22. Frekuensi Kata Yang Sering Muncul

IV. KESIMPULAN

Pada penelitian ini bertujuan untuk menganalisis sentimen ulasan yang terdapat pada aplikasi Canva di Play Store menggunakan pendekatan lexicon dan algoritma Decision Tree. Berdasarkan hasil evaluasi model klasifikasi, algoritma Decision Tree menunjukkan performa yang sangat baik dengan akurasi sebesar 87%. Kinerja terbaik tercatat pada kelas “2”, dengan nilai precision sebesar 90%, recall 89%, dan F1-score 89%. Nilai-nilai tersebut menunjukkan adanya keseimbangan yang baik antara precision dan recall, meskipun terdapat sedikit ketimpangan di antara keduanya. Dengan demikian, penerapan algoritma Decision Tree terbukti efektif dalam mengklasifikasikan ulasan aplikasi Canva. Hasil penelitian ini diharapkan dapat memberikan wawasan yang berharga untuk membantu aplikasi Canva meningkatkan kualitas layanannya berdasarkan analisis sentimen yang telah dilakukan.

REFERENSI

- [1] A. P. Gehred, “Canva,” *J. Med. Libr. Assoc.*, vol. 108, no. 2, pp. 338–340, 2020, doi: 10.5195/jmla.2020.923.
- [2] D. Pratmanto and F. F. D. Imanawan, “Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors,” *Comput. Sci.*, vol. 3, no. 2, pp. 110–117, 2023, doi: 10.31294/coscience.v3i2.1917.
- [3] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, “Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square,” *Jurnal*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [4] L. Wikarsa, A. Angdrsesey, and J. D. Kapantow, “Implementasi Metode Naïve Bayes dan Lexicon-Based Approach Untuk Mengklasifikasi Sentimen Netizen Pada Tweet berbahasa Indonesia,” *J. Ilm. Realt.*, vol. 18, no. 1, pp. 15–24, 2022.
- [5] A. D. Wibisono, S. Dadi Rizkiono, and A. Wantoro, “Filtering Spam Email Menggunakan Metode Naïve Bayes,” *TELEFORTECH J. Telemat. Inf. Technol.*, vol. 1, no. 1, 2020, doi: 10.33365/tft.v1i1.685.
- [6] D. W. Seno and A. Wibowo, “Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Dengan Metode Lexicon Based Dan Support Vector Machine,” *J. Ilm. FIFO*, vol. 11, no. 2, p. 144, 2019, doi: 10.22441/fifo.2019.v11i2.004.
- [7] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5,” *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [8] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, “Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [9] H. Hariyadi, D. Firdo, and M. H. Al Rafi, “Implementasi Algoritma Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Ulasan Aplikasi Canva,” *J. Minfo Polgan*, vol. 13, no. 1, pp. 261–269, 2024, doi: 10.33395/jmp.v13i1.13568.
- [10] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, “Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naïve Bayes dalam menganalisis sentimen data ulasan PLN Mobile,” *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [11] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jast20165.
- [12] W. A. Firmansyah, U. Hayati, and Y. Arie Wijaya, “Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naïve Bayes Dan Decision Tree Dengan Teknik Cross Validation,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
- [13] W. Laila, W. Widiarto, A. Wijayanto, and E. Suryani, “Rekomendasi Makanan Pasien Hiperlipidemia Berdasarkan Hasil Klasifikasi Menggunakan Metode Naïve Bayes dan Decision Tree,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 2, p. 328, 2022, doi: 10.26418/jpi.v8i2.56386.
- [14] D. Putra and A. Wibowo, “Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes,” *Pros. Semin.* Nas. Ris. Dan Inf. Sci., vol. 2, pp. 84–92, 2020.
- [15] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [16] A. Nurdin, D. Satria, Y. Kartika, A. Rezha, and E. Najaf, “Klasifikasi Penyakit Daun Tomat Dengan Metode Convolutional Neural Network Menggunakan Arsitektur Inception-V3,” no. 1, pp. 1–6, 2024.
- [17] C. Sammut, *Encyclopedia of Machine Learning*. Springer, 2010. doi: 10.1007/978-0-387-30164-8_731.