

**PENERAPAN ALGORITMA *DECISION TREE* UNTUK
ANALISIS SENTIMEN ULASAN APLIKASI VIDIO DI SITUS
*GOOGLE PLAY***

SKRIPSI

DHEA AYU SEPTIANI
20190040074



**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN
SUKABUMI
JULI 2023**

**PENERAPAN ALGORITMA *DECISION TREE* UNTUK
ANALISIS SENTIMEN ULASAN APLIKASI VIDIO DI SITUS
*GOOGLE PLAY***

SKRIPSI

*Diajukan Untuk Memenuhi Salah Satu Syarat Dalam Menempuh
Gelar Sarjana Teknik Informatika*

DHEA AYU SEPTIANI

20190040074



**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN
SUKABUMI
JULI 2023**

PERNYATAAN PENULIS

JUDUL : PENERAPAN ALGORITMA *DECISION TREE* UNTUK
ANALISIS SENTIMEN ULASAN APLIKASI VIDIO DI SITUS
GOOGLE PLAY

NAMA : DHEA AYU SEPTIANI

NIM 20190040074

“Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Teknik Informatika saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.



Penulis

PERSETUJUAN SKRIPSI

JUDUL : PENERAPAN ALGORITMA *DECISION TREE* UNTUK
ANALISIS SENTIMEN ULASAN APLIKASI VIDIO DI SITUS
GOOGLE PLAY

NAMA : DHEA AYU SEPTIANI

NIM 20190040074

Skripsi ini telah diperiksa dan disetujui
Sukabumi, Juli 2023

Ketua Program Studi
Teknik Infomatika,

Pembimbing,

Anggun Fergina, M.Kom

NIDN. 0407029301

Ivana Lucia Kharisma, M.Kom

NIDN. 0429038002



PENGESAHAN SKRIPSI

JUDUL : PENERAPAN ALGORITMA *DECISION TREE* UNTUK
ANALISIS SENTIMEN ULASAN APLIKASI VIDIO DI SITUS
GOOGLE PLAY

NAMA : DHEA AYU SEPTIANI

NIM 20190040074

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji
pada Sidang Skripsi tanggal 11 Juli 2023. Menurut pandangan kami,
Skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan
gelar Sarjana Sarjana Komputer (S.Kom)

Sukabumi, Juli 2023

Pembimbing I

Ivana Lucia Kharisma, M.Kom
NIDN. 0429038002

Ketua Penguji

Somantri, S.T, M.Kom
NIDN. 0419128801



Pembimbing II

Anggun Fergina, M.Kom
NIDN. 0407029301

Ketua Program Studi

Anggun Fergina, M.Kom
NIDN. 0407029301

Dekan Fakultas Teknik Komputer dan Desain

Ir. Paikun, S.T., M.T., IPM.,Asean Eng.
NIDN. 0402037401

ABSTRAK

Aplikasi berbasis *video streaming* menjadi jenis aplikasi paling banyak digunakan di dunia. *Video On Demand* merupakan sistem interaktif yang memungkinkan kita memilih konten video yang akan ditonton. Vidio adalah portal *online* atau situs web *streaming* video yang didirikan pada tahun 2014. Situs web ini memungkinkan pengguna untuk menonton dan menikmati berbagai video dan layanan lain. Namun, berdasarkan ulasan di *Google Play*, Vidio mendapatkan *rating* rata-rata hanya sebesar 3.7 dari 623.000 lebih total ulasan. Hal tersebut yang mendorong dilakukannya penelitian ini. Data yang dikumpulkan adalah sebanyak 1000 pada rentang waktu 2 Februari 2023 – 19 Februari 2023. Data tersebut diklasifikasikan ke dalam sentimen positif dan negatif menggunakan algoritma *Decision Tree*. Berdasarkan 3 skenario pembagian data, didapatkan akurasi terbesar diperoleh dari pembagian data 80% data latih dan 20% data uji yaitu sebesar 97.3%. sedangkan pada skenario pembagian data 70:30, akurasinya 96.8%, dan pembagian data 90:10 akurasinya sebesar 96.8%. Dari akurasi yang telah diperoleh, untuk evaluasi pengujian model, penelitian ini menggunakan *Confusion Matrix*. Agar prediksi dari model yang telah dilatih agar tersedia untuk orang lain, penelitian ini melakukan *model deployment* menggunakan Streamlit.



Kata kunci: *Video Streaming, Vidio, Decision Tree, Confusion Matrix, Streamlit*

ABSTRACT

Video streaming-based applications are becoming the most widely used type of application in the world. Video On Demand is an interactive system that allows us to choose the video content to watch. Vidio is an online portal or video streaming website established in 2014. The website allows users to watch and enjoy various videos and other services. However, based on reviews on Google Play, Vidio received an average rating of only 3.7 from more than 623,000 total reviews. This is what prompted this research. The data collected is as many as 1000 in the period of February 2, 2023 – February 19, 2023. The data is classified into positive and negative sentiment using the Decision Tree algorithm. Based on 3 data sharing scenarios, the greatest accuracy was obtained from data sharing of 80% of training data and 20% of test data, which was 97.3%. While in the 70:30 data sharing scenario, the accuracy is 96.8%, and the 90:10 data sharing is 96.8% accuracy. From the accuracy that has been obtained, for the evaluation of model testing, this study uses the Confusion Matrix. In order for predictions from the trained model to be available to others, this study conducts model deployment using Streamlit.

Keywords: Video Streaming, Vidio, Decision Tree, Confusion Matrix, Streamlit



KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah SWT, berkat rahmat dan karunia-Nya akhirnya skripsi “Penerapan Algoritma *Decision Tree* untuk Analisis Sentimen Ulasan Aplikasi Vidio di Situs *Google Play*” dapat diselesaikan dengan baik.

Maksud dan tujuan penulisan skripsi ini adalah sebagai gambaran terhadap penelitian yang dilakukan. Skripsi ini juga adalah tugas akhir untuk memenuhi syarat memperoleh gelar Sarjana Komputer (S.Kom) pada program studi Teknik Informatika Universitas Nusa Putra.

Sehubungan dengan itu, penulis menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Rektor Universitas Nusa Putra Sukabumi, Dr. Kurniawan S.T, M.Si, M.M.
2. Wakil Rektor I Bidang Akademik Universitas Nusa Putra Sukabumi, Bapak Anggy Pradiftha Junfithrana, S.Pd, M.T.
3. Kepala Program Studi Teknik Informatika Universitas Nusa Putra Sukabumi, Ibu Anggun Fergina, M.Kom.
4. Dosen Pembimbing I Universitas Nusa Putra Sukabumi, Ibu Ivana Lucia Kharisma, M.Kom, yang telah meluangkan waktu, tenaga, dan pikiran untuk mengarahkan penyelesaian skripsi ini.
5. Dosen Pembimbing II Universitas Nusa Putra Sukabumi, Ibu Anggun Fergina, M.Kom, yang juga telah meluangkan waktu untuk mengarahkan penulisan skripsi dengan baik.
6. Dosen Penguji, yang telah memberikan saran dan masukan guna membuat skripsi ini lebih baik.
7. Para Dosen Program Studi Teknik Informatika Universitas Nusa Putra Sukabumi, yang telah berjasa memberikan ilmu selama 4 tahun kuliah.
8. Orang tua dan keluarga yang selalu memberikan doa dan dukungan dalam penyelesaian skripsi ini.
9. Teman-teman seperjuangan, terutama Rismi Nurlaely dan Tesly Navida.



Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, oleh karena itu kritik dan saran yang membangun dari berbagai pihak sangat diharapkan demi perbaikan.

Sukabumi, 27 Juli 2023

Penulis



HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Nusa Putra, saya yang bertanda tangan di bawah ini:

Nama : Dhea Ayu Septiani
NIM : 20190040074
Program Studi : Teknik Informatika
Jenis karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Nusa Putra **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty- Free Right*) atas karya ilmiah saya yang berjudul:

PENERAPAN ALGORITMA *DECISION TREE* UNTUK ANALISIS SENTIMEN ULASAN APLIKASI VIDIO DI SITUS *GOOGLE PLAY*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Nusa Putra berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Sukabumi
Pada tanggal : 27 Juli 2023
Yang menyatakan,

DHEA AYU SEPTIANI

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
PERNYATAAN PENULIS.....	iii
PERSETUJUAN SKRIPSI	iv
PENGESAHAN SKRIPSI	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
HALAMAN PERNYATAAN PUBLIKASI	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian dan Manfaat Penelitian	4
1.5 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Penelitian Terkait.....	6
2.2 Landasan Teori.....	8
2.3 Kerangka Pemikiran.....	14
BAB III METODOLOGI PENELITIAN	17
3.1 Tahapan Penelitian.....	17
3.2 Penjelasan Alur Tahapan Penelitian	18
3.2.1 Identifikasi Masalah.....	19
3.2.2 Perumusan Masalah	19
3.2.3 Pengumpulan Data	19
3.2.4 <i>Preprocessing</i>	21
3.2.5 Pembobotan Kata	24
3.2.6 <i>Split Data</i>	24
3.2.7 Analisis dan Hasil	25
3.2.8 <i>Model Deployment</i>	26
BAB IV HASIL DAN PEMBAHASAN	28

4.1	Hasil Penelitian	28
4.1.1	Pengumpulan Data	28
4.1.2	<i>Text Preprocessing</i>	31
4.1.3	Pembobotan Kata	35
4.1.4	<i>Split</i> Data	35
4.1.5	Klasifikasi <i>Decision Tree</i>	35
4.1.6	Confusion Matrix	36
4.1.7	Visualisasi	37
4.1.8	<i>Model Deployment</i>	38
4.2	Pembahasan	41
BAB V KESIMPULAN.....		47
5.1	Kesimpulan	47
5.2	Saran	47
DAFTAR PUSTAKA		49



DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	6
Tabel 3.1 Daftar Ulasan dengan Rating.....	20
Tabel 3.2 Daftar Ulasan dengan Rating 5.....	21
Tabel 3.3 Teks Saat Sebelum dan Sesudah Proses Case Folding	22
Tabel 3.4 Teks Saat Sebelum dan Sesudah Proses Cleaning.....	22
Tabel 3.5 Teks Saat Sebelum dan Sesudah Proses Tokenizing.....	22
Tabel 3.6 Teks Saat Sebelum dan Sesudah Proses Stopword.....	23
Tabel 3.7 Teks Saat Sebelum dan Sesudah Proses Normalization	23
Tabel 3.8 Teks Saat Sebelum dan Sesudah Proses Stemming.....	24
Tabel 3.9 Confusion Matrix	25
Tabel 4.1 Jumlah Data Berdasarkan Rating	28
Tabel 4.2 Dataset Hasil Web Scraping.....	29
Tabel 4.3 Dataset yang Sudah Diberikan Label	30
Tabel 4.4 Dataset Sebelum dan Sesudah Tahapan Preprocessing	33
Tabel 4.5 Jumlah Data Berdasarkan Rating Setelah Mengalami Pengurangan	33
Tabel 4.6 Contoh Kalimat Positif.....	38
Tabel 4.7 Contoh Kalimat Negatif	39
Tabel 4.8 Contoh Kalimat Netral	39
Tabel 4.9 Jumlah Sentimen Positif dan Negatif Sebelum Pengurangan.....	42
Tabel 4.10 Jumlah Sentimen Positif dan Negatif Sebelum Pengurangan.....	42



DAFTAR GAMBAR

Gambar 1.1 Penetrasi internet Indonesia.....	1
Gambar 1.2 Alasan Masyarakat Menggunakan Aplikasi Video On Demand	2
Gambar 1.3 Aplikasi <i>Video On Demand</i> Paling Banyak Digunakan Di Indonesia.....	2
Gambar 1.4 Urutan Video On Demand Dengan Jumlah Durasi Nonton Penggunanya.....	3
Gambar 2.1 Ilustrasi Decision Tree	11
Gambar 2.2 Kerangka Pemikiran	15
Gambar 3.1 Tahapan Penelitian	18
Gambar 3.2 Proses Scraping Data di Google Colab.....	19
Gambar 3.3 Dataset yang Sudah Diberi Label	20
Gambar 4.1 Syntax untuk Proses Web Scraping.....	28
Gambar 4.2 Persentase Jumlah Sentimen Positif dan Negatif.....	30
Gambar 4.3 Library tahapan preprocessing.....	31
Gambar 4.4 Syntax untuk Proses Case Folding.....	Error! Bookmark not defined.
Gambar 4.5 Syntax untuk Proses Cleaning	31
Gambar 4.6 Syntax untuk Proses Tokenizing.....	32
Gambar 4.7 Syntax untuk Proses Normalization.....	32
Gambar 4.8 Syntax untuk Proses Stopword	32
Gambar 4.9 Syntax untuk Proses Stemming	33
Gambar 4.10 Diagram Jumlah Data Berdasarkan Rating.....	34
Gambar 4.11 Diagram Jumlah Sentimen Positif dan Negatif.....	34
Gambar 4.12 Syntax untuk Pembobotan Kata.....	35
Gambar 4.13 Syntax Pembagian Data.....	35
Gambar 4.14 Syntax Prediksi.....	36
Gambar 4.15 Syntax Pelatihan Model Decision Tree.....	36
Gambar 4.16 Syntax Confusion Matrix.....	36
Gambar 4.17 Syntax untuk Mengkalkulasi Confusion Matrix.....	37
Gambar 4.18 Wordcloud Sentimen Positif.....	37
Gambar 4.19 Wordcloud Sentimen Negatif	38
Gambar 4.20 Tampilan Streamlit	38
Gambar 4.21 Tampilan Streamlit Terhadap Sentimen Positif	40
Gambar 4.22 Tampilan Streamlit Terhadap Sentimen Negatif.....	40
Gambar 4.23 Tampilan Streamlit Terhadap Sentimen Netral	41
Gambar 4.24 Diagram Jumlah Data Berdasarkan Rating.....	41
Gambar 4.25 Pembagian Data 70:30.....	42
Gambar 4.26 Confusion Matrix Data Uji 30%.....	43
Gambar 4.27 Pembagian Data 80:20.....	44
Gambar 4.28 Confusion Matrix Data Uji 20%.....	44
Gambar 4.29 Pembagian Data 90:10.....	45
Gambar 4.30 Confusion Matrix Data Uji 90%.....	46

BAB I

PENDAHULUAN

1.1 Latar Belakang

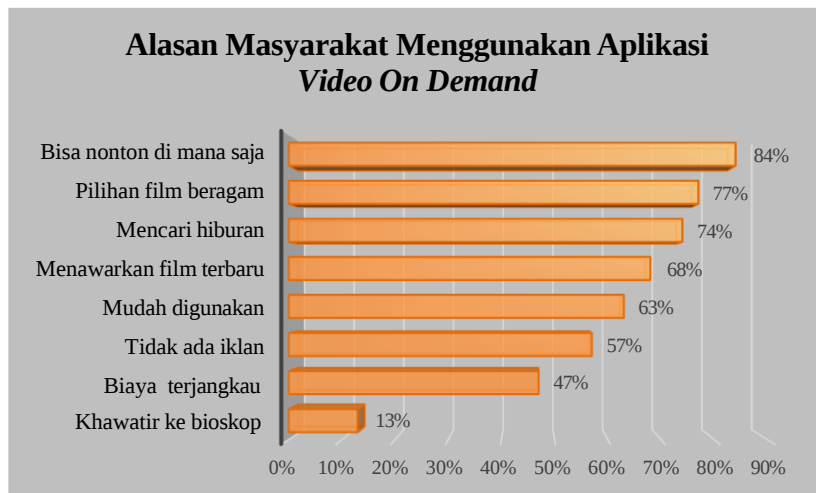
Jaringan interaktif yang digerakkan oleh aplikasi telah membantu kita, bisnis yang kita jalani serta keluarga tetap terhubung. *Video streaming, social networking* dan *web* adalah tiga kategori aplikasi teratas. Aplikasi berbasis *video streaming* menjadi jenis aplikasi paling banyak digunakan di dunia berdasarkan lalu lintas hilir (*downstream*) yang persentasenya mencapai 48,9% dari total lalu lintas hilir internet [1]. Menariknya, Indonesia menjadi negara dengan total waktu yang dihabiskan untuk *streaming* tertinggi di dunia [2]. Fenomena ini tentunya dipengaruhi oleh berkembangnya internet di Indonesia. Berdasarkan survei yang dilakukan oleh APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) [3], penetrasi pengguna internet di Indonesia mengalami perkembangan menjadi 77,02% pada 2021-2022 dari 73,70% pada 2019-2020 seperti yang tertera pada Gambar 1.1.



Gambar 1.1 Penetrasi internet Indonesia

(selanjutnya disebut VOD) adalah sistem interaktif yang memungkinkan kita memilih konten video yang akan ditonton. Berbeda dengan televisi yang harus menunggu konten sesuai jadwal tayang, dengan VOD kita bisa menikmati, mengunduh dan memilih konten yang kita inginkan kapan saja dan di perangkat apa saja [4]. Berdasarkan temuan survei yang dirilis oleh Populix [5], alasan paling tinggi masyarakat Indonesia menggunakan layanan VOD adalah “Bisa

nonton di mana saja” dengan persentase mencapai 84%, seperti yang terdapat pada Gambar 1.2.



Gambar 1.2 Alasan Masyarakat Menggunakan Aplikasi Video On Demand

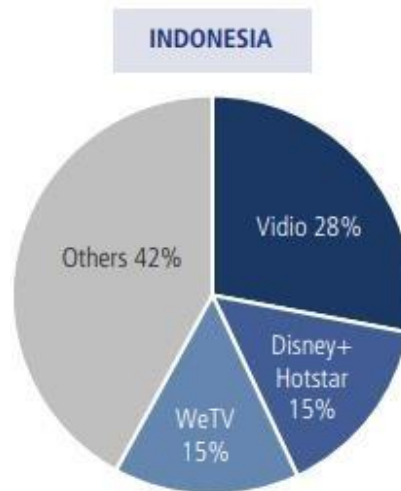
Di dalam survei tersebut juga menunjukkan bahwa Netflix adalah aplikasi VOD yang paling banyak digunakan di Indonesia disusul Disney+, Youtube, dan Viu. Membuktikan bisa bersaing dengan platform VOD dari luar negeri, Vidio menjadi platform VOD lokal yang juga paling banyak digunakan oleh masyarakat Indonesia disusul oleh Mola TV dan GoPlay, seperti terdapat pada Gambar 1.3.



Gambar 1.3 Aplikasi Video On Demand Paling Banyak Digunakan Di Indonesia

Hingga saat ini (04/02/2023) Vidio sudah terunduh sebanyak 50 juta dengan *rating* 3.7 dan 623 ribu ulasan, serta menempati urutan pertama pada kategori *top grossing* segmentasi *entertainment* di situs *Google Play*. Selain memiliki konten serial dan film, Vidio juga menayangkan program *live* termasuk

program televisi lokal. Menurut laporan riset yang dilakukan oleh Media Partner Asia [6], Vidio berada di urutan pertama berdasarkan durasi nonton penggunanya seperti pada Gambar 1.4.



Gambar 1.4 Urutan Video On Demand Dengan Jumlah Durasi Nonton Penggunanya

Google Play sendiri adalah sebuah toko *online* untuk pengguna *android* dimana pengguna bisa menemukan aplikasi, *game*, film, acara TV, buku dan konten lainnya (https://play.google.com/intl/id_id/about/howplayworks/). *Google Play* dapat digunakan melalui web, aplikasi *android* (*Play Store*) dan *Google TV*. *Google Play* memiliki fitur yang menyediakan ulasan pengguna yang dapat menunjukkan pengalamannya dalam menggunakan aplikasi. Ulasan pengguna sering digunakan sebagai alat yang efektif untuk menemukan informasi tentang suatu produk atau layanan [7].

Dengan latar belakang yang sudah diuraikan, penelitian ini bertujuan untuk menganalisis ulasan pengguna agar mengetahui kepuasannya dalam menggunakan aplikasi Vidio. Maka dari itu judul dari penelitian ini adalah “PENERAPAN ALGORITMA *DECISION TREE* UNTUK ANALISIS SENTIMEN ULASAN APLIKASI VIDIO DI SITUS *GOOGLE PLAY*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan, rumusan masalah yang akan dikaji dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritma *Decision Tree* untuk analisis sentimen ulasan aplikasi Vidio pada situs *Google Play* dan bagaimana penerapannya pada aplikasi *website*?
2. Bagaimana ulasan aplikasi Vidio dalam kategori positif dan negatif?

1.3 Batasan Masalah

Untuk mencapai sasaran yang diinginkan, batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang akan digunakan adalah data ulasan selama bulan Februari 2023 sebanyak 1000 ulasan mengenai aplikasi Vidio pada situs *Google Play* dengan kategori ‘terbaru’.
2. Ulasan yang diambil merupakan ulasan yang menggunakan Bahasa Indonesia.
3. Algoritma yang digunakan pada penelitian ini adalah *Decision Tree*.

1.4 Tujuan Penelitian dan Manfaat Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui sentimen ulasan aplikasi Vidio pada situs *Google Play* menggunakan algoritma *Decision Tree* dan penerapannya pada aplikasi *website*.
2. Mengetahui ulasan aplikasi Vidio dalam kategori positif dan negatif.

Manfaat pada penelitian ini adalah sebagai berikut:

1. Mengetahui gambaran umum mengenai ulasan pengguna aplikasi Vidio berdasarkan data yang diambil dari situs *Google Play*
2. Menambah pengetahuan khususnya pengklasifikasian menggunakan metode *Decision Tree*
3. Mengetahui tanggapan publik mengenai kelebihan dan kekurangan aplikasi Vidio

1.5 Sistematika Penulisan

Sistematika penulisan bertujuan agar penulisan laporan penelitian dapat tersusun rapi dan mudah dipahami oleh pembaca. Berikut merupakan sistematika penelitian pada penelitian ini:

BAB I PENDAHULUAN

Bab ini berisi Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian dan Manfaat Penelitian serta Sistematika Penelitian.

BAB II TINJAUAN PUSTAKA

Berisi tentang dasar teori berkaitan dengan penelitian, juga memuat penelitian terkait.

BAB III METODOLOGI PENELITIAN

Mengandung uraian mengenai populasi dan sampel penelitian, jenis dan sumber data, variabel penelitian, metode pengumpulan data dan metode analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini mengandung hasil dari tahapan penelitian dan implementasinya ke dalam aplikasi website.

BAB V PENUTUP

Berisi kesimpulan dari hasil penelitian serta saran untuk penelitian selanjutnya.



5.1 Kesimpulan

Dari pengujian yang sudah dilakukan, terdapat beberapa kesimpulan yang dapat diambil dalam analisis sentimen terhadap ulasan aplikasi Vidio di situs *Google Play* dengan menggunakan algoritma *Decision Tree*, diantaranya adalah:

1. Setelah mengalami pengurangan dikarenakan terdapat data kosong, jumlah *dataset* yang sebelumnya 1000, menjadi 940. Adapun dari *dataset* 940, jumlah ulasan kelas positif sebanyak 356 (37.9%), sedangkan untuk kelas negatif sebanyak 584 (62.1%).
2. Berdasarkan perbandingan pembagian data latih dan data uji 70:30, 80:20 dan 90:10, didapatkan akurasi terbaik adalah dengan skenario 80:20, yaitu 97.3% dengan *precision* 0.947 (94.7%), *recall* 0.986 (98.6%), dan *f-1 score* 0.966 (96.6%). Sedangkan pembagian 70:30

mendapatkan akurasi sebesar 96.8% dengan *precision* 0.941 (94.1%), *recall* 0.969 (96.9%), dan *f-1 score* 0.953 (95.3%). Pembagian data 90:10 mendapatkan akurasi sebesar 96.8% dengan *precision* 0.916 (91.6%), *recall* 1 (100%), dan *f-1 score* 0.956 (95.6%).

3. Dengan besarnya akurasi yang dihasilkan oleh ketiga skenario pembagian pengujian data, serta kecilnya angka ketidakakuratan prediksi, maka dapat disimpulkan bahwa algoritma *Decision Tree* berjalan dengan baik pada penelitian ini.
4. Dari hasil klasifikasi yang telah dilakukan, dapat disimpulkan bahwa ulasan aplikasi Vidio mendapat sentimen negatif lebih besar daripada sentimen positif.

5.2 Saran

Berdasarkan hasil analisis, pembahasan, hingga kesimpulan, beberapa saran yang dapat diberikan adalah:



1. Hasil penelitian ini, khususnya sentimen negatif dapat dijadikan bahan evaluasi bagi pihak perusahaan untuk mengembangkan dan memperbaiki aplikasi guna meningkatkan kepuasan dan kepercayaan pengguna.
2. Untuk penelitian selanjutnya, dapat menggunakan algoritma klasifikasi lain, guna melihat perbandingan dengan algoritma yang telah digunakan pada penelitian ini. Pengacakan data untuk latih dan uji dapat dilakukan dengan lebih banyak, menggunakan metode untuk menyeimbangkan data, serta periode ulasan yang akan dikumpulkan dapat dilakukan dengan rentang waktu lebih panjang untuk mendapatkan informasi lebih dalam dan lebih banyak.



DAFTAR PUSTAKA

- [1] Sandvine, "The Mobile Internet Phenomena Report," 2021. Accessed: Jan. 27, 2023. [Online]. Available: https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2021/Phenomena/MIPR%20Q1%202021%2020210510.pdf
- [2] data.ai, "STATE OF MOBILE ," 2022. Accessed: Feb. 27, 2023. [Online]. Available: <https://www.data.ai/en/go/state-of-mobile-2022/>
- [3] Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), "Profil Internet Indonesia," 2022. Accessed: Jan. 23, 2023. [Online]. Available: <https://www.apjii.or.id/gudang-data/standar>
- [4] N. A. Yusuf and Indrawati, "Analisis Faktor yang Memengaruhi Pembentukan Minat Berlangganan di Industri Video-On-Demand di Indonesia.," vol. 3, pp. 161–173, 2019.
- [5] Populix, "Indonesian Video Entertainment on Demand Consumption," Jul. 2022. Accessed: Feb. 03, 2023. [Online]. Available: <https://info.populix.co/report/indonesian-video-entertainment-on-demand-consumption/>
- [6] Media Partner Asia, "Southeast Asia Online Consumer Insights and Analytics." pp. 1–2, Nov. 22, 2022. [Online]. Available: https://media-partners-asia.com/AMPD/Q3_2022/SEA/PR.pdf
- [7] S. A. Saputra, D. Rosiyadi, W. Gata, and S. M. Husain, "Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization," *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, pp. 377–382, 2019.
- [8] N. T. Romadloni, I. Santoso, and S. Budilaksono, "PERBANDINGAN METODE NAIVE BAYES, KNN DAN DECISION TREE TERHADAP ANALISIS SENTIMEN TRANSPORTASI KRL COMMUTER LINE," *Jurnal IKRA-ITH Informatika*, vol. 2, pp. 1–9, 2019.
- [9] M. Syarifuddin, "ANALISIS SENTIMEN OPINI PUBLIK TERHADAP EFEK PSBB PADA TWITTER DENGAN ALGORITMA DECISION TREE-KNN-NAÏVE BAYES," *PPKM Nusa Mandiri*, vol. 15, pp. 87–94, Aug. 2020.
- [10] M. Aufar, R. Andreswari, and D. Pramesti, "Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study," *2020 IEEE International Conference on Data Science and Its Applications (ICoDSA)*, pp. 1–7, Oct. 2020, doi: <https://doi.org/10.1109/ICoDSA50139.2020.9213078>.

- [11] T. W. Putra, A. Triayudi, and Andrianingsih, “Analisis Sentimen Pembelajaran Daring menggunakan Metode Naïve Bayes, KNN, dan Decision Tree,” *JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, pp. 20–26, Jan. 2022.
- [12] P. D. Batlayeri and W. Gatta, “ANALISIS SENTIMEN PEJUALAN JAFRA DALAM PANDEMI COVID-19 DENGAN ALGORITMA KLASIFIKASI,” *JIRE (Jurnal Informatika & Rekayasa Elektronik)*, vol. 5, pp. 11–18, Apr. 2022.
- [13] M. T. Waluyan and K. D. Hartomo, “Analisis Sentimen Kebutuhan Fast Track Pada Originals Vidio Menggunakan Support Vector Machine,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 3, pp. 2153–2162, Sep. 2022, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/2348>
- [14] Ibnu Daqiqil ID, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*, 1st ed. Riau: UR Press, 2021.
- [15] A. Ma’arif, “BUKU AJAR PEMROGRAMAN LANJUT BAHASA PEMROGRAMAN PYTHON.” pp. 1–62, 2020. [Online]. Available: <http://eprints.uad.ac.id/32743/1/buku%20python.pdf>
- [16] E. V. Rahmadani, S. F. Panevandy, H. Harani, *Algoritma C4.5 dan K-Nearest Neighbors (KNN) untuk Memetakan Matakuliah dan Keterlambatan Kelulusan Mahasiswa*, 1st ed. Bandung: Kreatif Industri Nusantara, 2020. [Online]. Available: https://www.google.co.id/books/edition/Algoritma_C4_5_dan_K_Nearest_Neighbors_K/BGv9DwAAQBAJ?hl=en&gbpv=0&kptab=getbook
- [17] U. D. Arni, “Apa Itu Text Mining?,” *Garuda Cyber Indonesia*, Oct. 25, 2018. <https://garudacyber.co.id/artikel/1254-apa-itu-text-mining> (accessed Feb. 08, 2023).
- [18] A. Sujjada and A. Fergina, “IMPLEMENTASI METODE VECTOR SPACE MODEL UNTUK DETEKSI EMOSI MENGGUNAKAN DATA TEKS TWITTER,” *Restikom (Riset Teknik Informatika dan Komputer)*, vol. 3, no. 3, pp. 116–129, Dec. 2021.
- [19] M. I. Aditama, R. I. Pratama, K. H. U. Wiwaha, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19,” *JIEET (Journal Information Engineering and Educational Technology)*, vol. 04, no. 02, pp. 90–92, 2020, [Online]. Available: <https://journal.unesa.ac.id/index.php/jieet/article/view/11018>
- [20] M. F. Mubaraq and W. Maharani, “Sentiment Analysis on Twitter Social Media towards Climate Change on Indonesia Using IndoBERT Model,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2426, Oct. 2022, doi: 10.30865/mib.v6i4.4570.

- [21] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS*, vol. 02, no. 01, pp. 20–28, 2021.
- [22] Yulia and N. Azwanti, "Penerapan Algoritma C4.5 Untuk Memprediksi Besarnya Penggunaan Listrik Rumah Tangga di Kota Batam," *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 02, no. 02, pp. 584–590, 2018, [Online]. Available: <https://jurnal.iaii.or.id/index.php/RESTI/article/view/503>
- [23] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi dengan Pohon Keputusan Berbasis Algoritma C4.5," *PRISMA, Prosiding Seminar Nasional Matematika 3*, pp. 64–71, 2020.
- [24] A. Y. Ikhsanti, Y. Fauziah, and R. I. Perwira, "Implementation of the c4.5 decision tree learning algorithm for sentiment analysis in e-commerce application reviews on google play store," *Computing and Information Processing Letter*, vol. 1, no. 1, pp. 25–30, Nov. 2021.
- [25] Vidio, "Tentang Vidio," Nov. 11, 2021. <https://support.vidio.com/support/solutions/articles/43000026852-tentang-vidio> (accessed Feb. 07, 2023).
- [26] M. Levi, H. N. Palit, and S. Sukrostaningsih, "Perbandingan Performa Tools Web Scraping pada Website dengan Data Statis dan Dinamis," *INFRA*, vol. 8, no. 2, pp. 202–208, 2020, [Online]. Available: <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/10521>
- [27] S. V. Sofwan and S. Cahyani, "PENGARUH PENDAPATAN ASLI DAERAH DAN SISA LEBIH PEMBIAYAAN ANGGARAN TERHADAP BELANJA MODAL BADAN PENGELOLAAN KEUANGAN DAN ASET KOTA BANDUNG PERIODE 2010-2018," *AKURAT | Jurnal Ilmiah Akuntansi*, vol. 11, no. 3, pp. 115–130, 2020.
- [28] A. R. Aulia and A. L. Yulianti, "PENGARUH CITY BRANDING 'A LAND OF HARMONY' TERHADAP MINAT BERKUNJUNG DAN KEPUTUSAN BERKUNJUNG KE PUNCAK, KABUPATEN BOGOR," *JIMEA | Jurnal Ilmiah MEA (Manajemen, Ekonomi, & Akuntansi)*, vol. 3, no. 3, pp. 67–75, 2019.
- [29] H. Nguyen, A. Veluchamy, M. Diop, and R. Iqbal, "Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches," *SMU Data Science Review*, vol. 1, no. 4, 2018.
- [30] B. A. Adli, "ANALISIS SENTIMEN CUSTOMER REVIEW PRODUK TOKOPEDIA MENGGUNAKAN ALGORITMA DECISION TREE," 2022.

- [31] N. K. Widyasanti, I. K. G. D. Putra, and N. K. D. Rusjayanthi, "Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia," *MERPATI*, vol. 6, no. 7, pp. 119–126, Aug. 2018.
- [32] E. Undamayanti, T. I. Hermanto, and I. Kaniawulan, "Analisis Sentimen Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization Terhadap Pelaksanaan Program Merdeka Belajar Kampus Merdeka," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6, no. 2, pp. 916–930, Sep. 2022.
- [33] D. A. N. Krisna and U. Salamah, "PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI BERITA HOAX KESEHATAN DI MEDIA SOSIAL TWITTER," *Jurnal Teknik Informatika Kaputama (JTik)*, vol. 6, no. 2, pp. 836–845, Jul. 2022.
- [34] D. P. Santoso and W. Wibowo, "Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store," *JURNAL SAINS DAN SENI ITS*, vol. 11, no. 2, pp. 190–196, 2022.
- [35] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," *AITI: Jurnal Teknologi Informasi*, vol. 18, no. 2, pp. 174–185, Aug. 2021.
- [36] S. Khairunnisa, Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, pp. 406–414, Apr. 2021.
- [37] I. Mahayani, A. D. R., and M. E. Supriyadi, "Analisis Sentimen Twitter Terhadap Pembayaran ShopeePayLater Pada Aplikasi Belanja Online (Shopee) Menggunakan Metode Lexicon Based dan Naïve Bayes Classifier," *Jurnal Ilmiah KOMPUTASI*, vol. 19, no. 4, pp. 545–558, Dec. 2020.
- [38] R. Kosasih and A. Alberto, "Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor," *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 6, no. 1, pp. 134–139, 2021.
- [39] Delta Sierra, "Algoritma TF — IDF," *Medium*, Feb. 13, 2019. <https://dltsierra.medium.com/algoritma-tf-idf-633e17d10a80> (accessed Jun. 12, 2023).
- [40] S. W. Iriananda, R. P. Putra, and K. S. Nugroho, "ANALISIS SENTIMEN DAN ANALISIS DATA EKSPLORATIF ULASAN APLIKASI

MARKETPLACE GOOGLE PLAYSTORE ,” *The 4th Conference on Innovation and Application of Science and Technology (CIASTECH)*, pp. 473–482, Dec. 2021.

- [41] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *Jurnal Sains Komputer dan Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, Sep. 2021.
- [42] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” *Citec Journal*, vol. 6, no. 1, pp. 1–11, Jan. 2019.
- [43] Terence Shin, “What Does it Mean to Deploy A Machine Learning Model?,” *Towards Data Science*, Apr. 23, 2020.
<https://towardsdatascience.com/what-does-it-mean-to-deploy-a-machine-learning-model-dddb983ac416#:~:text=Deploying%20a%20machine%20learning%20model,input%20and%20return%20an%20output>. (accessed May 21, 2023).
- [44] A. B. Prasetyo and T. G. Laksana, “Optimasi Algoritma K-Nearest Neighbors dengan Teknik Cross Validation Dengan Streamlit (Studi Data: Penyakit Diabetes),” *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 2, pp. 194–204, Dec. 2022.

