

CEK TURNITIN

SKRIPSI ROCKY SAPUTRA (403221010086) SI

 SATSET DIGITAL 2

Document Details

Submission ID

trn:oid:::3618:151105990

Submission Date

Sep 9, 2026, 3:05 PM GMT+7

Download Date

Sep 9, 2026, 3:14 PM GMT+7

File Name

SKRIPSI ROCKY SAPUTRA (403221010086) SI.pdf

File Size

2.1 MB

79 Pages

12,707 Words

88,649 Characters

55% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Top Sources

- 38%  Internet sources
- 41%  Publications
- 50%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 38% Internet sources
- 41% Publications
- 50% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	UPN Veteran Yogyakarta on 2026-07-21	3%
2	Publication	Sri Damayanti, Husna Sari, Bobby Ardiansyah, Dicky Apdillah. "Analisis Sentimen ...	2%
3	Publication	Khadijah, Imam Saputra. "Analisis Sentimen Pengguna Aplikasi Gojek Mengguna...	1%
4	Internet	ejournal.itn.ac.id	1%
5	Internet	jurnal.muaraedukasi.id	<1%
6	Internet	proceeding.unpkediri.ac.id	<1%
7	Publication	Muhammad Hanif, Garno, Asep Jamaludin. "ANALISIS SENTIMEN PENGGUNA APL...	<1%
8	Internet	conference.usm.ac.id	<1%
9	Student papers	Universitas Putera Batam on 2026-07-27	<1%
10	Internet	repository.uin-suska.ac.id	<1%
11	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-02	<1%

12	Student papers	Universitas Pancasila on 2026-03-02	<1%
13	Internet	etheses.uin-malang.ac.id	<1%
14	Student papers	Universitas Putera Batam on 2026-07-25	<1%
15	Internet	journal.irpi.or.id	<1%
16	Internet	journal.likmi.ac.id	<1%
17	Internet	jurnal.wicida.ac.id	<1%
18	Publication	Akbar Supia Dirja, Lana Aulia Salsabila, M. Akhdan Putra Hermawan, Cika Jelita, H...	<1%
19	Internet	publikasi.dinus.ac.id	<1%
20	Student papers	Universitas Budi Luhur on 2026-07-02	<1%
21	Internet	www.ojs.unkriswina.ac.id	<1%
22	Internet	jurnal.amikom.ac.id	<1%
23	Student papers	Universitas Sangga Buana YPKP on 2026-07-23	<1%
24	Internet	repository.umsu.ac.id	<1%
25	Internet	repository.unisi.ac.id	<1%

26	Student papers	Universitas Muslim Indonesia on 2024-10-18	<1%
27	Internet	repository.ar-raniry.ac.id	<1%
28	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-31	<1%
29	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-14	<1%
30	Student papers	Universitas Islam Indonesia on 2026-04-20	<1%
31	Student papers	Universitas Budi Luhur on 2025-07-13	<1%
32	Student papers	Universitas Muhammadiyah Palembang on 2025-08-06	<1%
33	Internet	e-journal.stmik-bnj.ac.id	<1%
34	Student papers	Universitas Putera Batam on 2026-07-28	<1%
35	Internet	subset.id	<1%
36	Student papers	Universitas Sangga Buana YPKP on 2026-07-31	<1%
37	Internet	jurnal.satyaterrabhinneka.ac.id	<1%
38	Student papers	Universitas Katolik Musi Charitas on 2026-08-13	<1%
39	Student papers	Universitas Pamulang on 2025-10-21	<1%

40	Student papers	Universitas Pancasila on 2026-03-05	<1%
41	Internet	ejournal.unidayan.ac.id	<1%
42	Internet	eprints.amikompurwokerto.ac.id	<1%
43	Student papers	East Union High School on 2026-07-02	<1%
44	Student papers	Universitas Muslim Indonesia on 2025-07-16	<1%
45	Student papers	Universitas Pamulang on 2026-07-21	<1%
46	Student papers	UPN Veteran Yogyakarta on 2026-08-18	<1%
47	Student papers	Universitas Islam Riau on 2025-07-17	<1%
48	Publication	Pamungkas, Yudi Ribut. "Dampak pembangunan Abdul Rasyid Tower terhadap p...	<1%
49	Student papers	Universitas Islam Bandung on 2026-08-26	<1%
50	Publication	M. Saleh, Dwi Yuli Prasetyo. "Analisis Sentimen Konsumen Terhadap Usaha Food ...	<1%
51	Student papers	Universitas Bengkulu on 2025-07-09	<1%
52	Student papers	Universitas Sangga Buana YPKP on 2026-07-31	<1%
53	Student papers	Universitas Muhammadiyah Jember on 2026-01-27	<1%

54	Student papers	Universitas Muhammadiyah Makassar on 2026-01-14	<1%
55	Student papers	Universitas Putera Batam on 2026-02-02	<1%
56	Internet	www.ejournal.itn.ac.id	<1%
57	Student papers	Sriwijaya University on 2026-04-28	<1%
58	Student papers	Telkom University on 2026-08-07	<1%
59	Student papers	UPN Veteran Yogyakarta on 2026-04-02	<1%
60	Student papers	UPN Veteran Yogyakarta on 2026-07-02	<1%
61	Student papers	Universitas Pamulang on 2026-07-11	<1%
62	Internet	digilib.uin-suka.ac.id	<1%
63	Publication	Yola Tirta Nurfajriya, Pinkan Prilia Mahmudah, Hana Aurora Aida Rahma, Hanif Z...	<1%
64	Internet	sejurnal.com	<1%
65	Student papers	Universitas Tarumanagara on 2025-12-15	<1%
66	Student papers	Universitas Pancasila on 2026-02-21	<1%
67	Publication	Eti Kurniawati, Ade Irma Purnamasari, Irfan Ali, Rudi Kurniawan, Odi Nurdiawan. ...	<1%

68	Publication	Siti Noraisyah, Fitri Yunita. "Analisis Sentimen Penggunaan ChatGPT di Kalanga...	<1%
69	Student papers	Sultan Agung Islamic University on 2025-07-23	<1%
70	Internet	publikasiilmiah.unwahas.ac.id	<1%
71	Internet	sostech.greenvest.co.id	<1%
72	Internet	ejurnal.methodist.ac.id	<1%
73	Internet	journal.trunojoyo.ac.id	<1%
74	Internet	mail.rumahjurnal.or.id	<1%
75	Internet	www.jurnalitp.web.id	<1%
76	Publication	Nur Azizah Harahap, Puspita Wanny, Khairunnisa Khairunnisa. "Analisis Sentime...	<1%
77	Publication	Muhammad Nabil, Samsudin Samsudin. "Analisis Sentimen Publik Program Maka...	<1%
78	Student papers	Universitas Budi Luhur on 2026-02-07	<1%
79	Student papers	Universitas Muhammadiyah Makassar on 2026-06-07	<1%
80	Student papers	Universitas Sangga Buana YPKP on 2026-08-10	<1%
81	Publication	Jim Maxwell Manampiring, Jenianus Halawa, Jefiri Zai, Ika Kurniawati, Riza Fahlap...	<1%

82	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2025-08-26	<1%
83	Publication	Randi Afif Afif, Aji Supriyanto, Rr. Fitri Damaryanti, Wahyu Prasetya Adi. "Analisis ...	<1%
84	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-14	<1%
85	Internet	journals.upi-yai.ac.id	<1%
86	Publication	Abidin Abidin. "ANALISIS SENTIMEN ULASAN APLIKASI SIREKAP MENGGUNAKAN ...	<1%
87	Publication	Gita Tiara Aftita, Raissa Amanda Putri. "ANALISIS SENTIMEN ULASAN PENGGUNA ...	<1%
88	Publication	Mawadatul Maulidah, Angga Ardiansyah, Suleman Suleman, Lina Putri Gemilang,...	<1%
89	Student papers	Universitas Budi Luhur on 2026-07-09	<1%
90	Student papers	Universitas Islam Sumatera Utara on 2026-04-20	<1%
91	Student papers	Universitas Putera Batam on 2026-01-30	<1%
92	Student papers	UNIVERSITAS BUDI LUHUR on 2026-01-22	<1%
93	Student papers	UPN Veteran Yogyakarta on 2026-06-02	<1%
94	Student papers	UPN Veteran Yogyakarta on 2026-06-29	<1%
95	Student papers	Universitas Trunojoyo on 2026-07-07	<1%

96	Internet	ejournal.stmik-time.ac.id	<1%
97	Student papers	UPN Veteran Yogyakarta on 2026-05-05	<1%
98	Publication	Zendra Utama, Pahrizal Pahrizal. "PENERAPAN METODE CNN UNTUK ANALISIS SE...	<1%
99	Publication	Niken Zeliana Putri, Martanto Martanto, Arif Rinaldi Dikananda, Ahmad Rifa'i. "Se...	<1%
100	Student papers	Universitas Budi Luhur on 2025-07-11	<1%
101	Student papers	Universitas Muslim Indonesia on 2025-06-23	<1%
102	Internet	repositori.buddhidharma.ac.id	<1%
103	Student papers	Universitas Bengkulu on 2025-08-28	<1%
104	Student papers	Universitas Muslim Indonesia on 2026-07-02	<1%
105	Publication	M Fikri Setiawan, Elvi Rahmi. "Analisis Sentimen Masyarakat terhadap Kebijakan ...	<1%
106	Publication	Rossiana Ginting, Satria Prayudi, Fahsi Mahmoud, Handaru Jati et al. "Machine Le...	<1%
107	Publication	Sajid Mahattir Aulia Rahman, Tommy Bustomi, Ahmad Fahrijal Pukeng. "Analisis ...	<1%
108	Student papers	Universitas Budi Luhur on 2025-07-12	<1%
109	Internet	cibangsa.com	<1%

110	Internet	journal.peradaban.ac.id	<1%
111	Publication	M. Rizki Mahrifan, Fitri Yunita. "Analisis Sentimen pada Aplikasi MyXL Mengguna...	<1%
112	Publication	Shinta Nilam Sari Muslim, Firman Nurdyansyah, Aviv Yuniar Rahman. "PERBANDI...	<1%
113	Student papers	Universitas Pamulang on 2026-07-20	<1%
114	Publication	Mochammad Dayke Xavier Lumankun, Ghofar Taufik. "Algoritma Naïve Bayes un...	<1%
115	Internet	ejournal.indo-intellectual.id	<1%
116	Student papers	Universitas Bina Sarana Informatika on 2026-02-12	<1%
117	Publication	Junaidy, Muhammad Fauzan, Roberto Kaban. "Analisis Sentimen Publik Terhada...	<1%
118	Student papers	LL DIKTI IX Turnitin Consortium Part III on 2026-07-19	<1%
119	Student papers	Universitas Budi Luhur on 2026-07-10	<1%
120	Student papers	Universitas Pamulang on 2026-07-28	<1%
121	Internet	jurnal.polibatam.ac.id	<1%
122	Publication	Adinda Purnamasari, Rini Astuti, Khaerul Anam, Gifthera Dwilestari, Mulyawan. ...	<1%
123	Publication	Dody Pradipta, Eko Widodo, David Fahmi A, M. Iqbal Izzul Haq. "Klasifikasi Sentim...	<1%

124	Publication	Muhammad Abil Fadli, Tukino Tukino, Elfina Novalia, April Lia Hananto. "KLASIFIK...	<1%
125	Publication	Paulina Gorat Frans, Frans Steven Pakpahan, Sardo Pardingotan Sipayung. "Anal...	<1%
126	Student papers	Universitas Khairun on 2026-05-18	<1%
127	Publication	Muhammad Diki Hendriyanto, Azhari Ali Ridha, Ultach Enri. "Analisis Sentimen UI...	<1%
128	Student papers	Universitas Pamulang on 2026-06-22	<1%
129	Student papers	Universitas Pamulang on 2026-08-03	<1%
130	Internet	e-journal.ivet.ac.id	<1%
131	Publication	Rizal Dwi Rizki Apriliansyah, Rini Astuti, Willy Prihartono, Ryan Hamonangan. "PE...	<1%
132	Publication	Shazifa Azhari, Nining Rahaningsih, Raditya Danar Dana, Mulyawan .. "PENINGKA...	<1%
133	Student papers	Universitas Bengkulu on 2025-10-17	<1%
134	Student papers	Universitas Islam Indonesia on 2026-05-12	<1%
135	Student papers	Universitas Katolik Widya Mandala Surabaya on 2025-12-09	<1%
136	Student papers	Universitas Djuanda on 2025-07-15	<1%
137	Student papers	Universitas Riau on 2023-09-24	<1%

138	Internet	eprints.umg.ac.id	<1%
139	Internet	jpti.journals.id	<1%
140	Publication	Awang Herjunie Nurdy, Abdul Rahim, Arbansyah. "Analisis Sentimen Ulasan Gam..."	<1%
141	Publication	Mahima Cinta Hage, Pratyaksa Ocsa Nugraha Saian. "ANALISIS SENTIMEN ULASA..."	<1%
142	Publication	Muhamad Azril, Giatika Crisnawati. "Klasifikasi Sentimen Komentar Pengguna Tik..."	<1%
143	Student papers	Sriwijaya University on 2023-07-20	<1%
144	Student papers	UNIVERSITAS BUDI LUHUR on 2026-01-30	<1%
145	Student papers	Universitas Muhammadiyah Sumatera Utara on 2025-08-21	<1%
146	Student papers	Universitas Putera Batam on 2025-07-28	<1%
147	Internet	ejurnal.seminar-id.com	<1%
148	Internet	repository.itpln.ac.id	<1%
149	Publication	Rizkiana, Annisa. "Manajemen Kepala Sekolah Dalam Meningkatkan Profesionali..."	<1%
150	Student papers	Universitas Pamulang on 2026-07-22	<1%
151	Internet	jurnal.itg.ac.id	<1%

152	Internet	repository.unsri.ac.id	<1%
153	Publication	Ichsani Mursidah, Remi Sanjaya, Bambang Yulianto, Dhian Sweetania, Puji Sularsi...	<1%
154	Publication	Steve Nwaiwu, Nipat Jongsawat. "Assessing Transformers and Traditional Models...	<1%
155	Student papers	UNIVERSITAS BUDI LUHUR on 2026-02-06	<1%
156	Student papers	UPN Veteran Yogyakarta on 2026-08-03	<1%
157	Student papers	Universitas Katolik Musi Charitas on 2026-08-12	<1%
158	Student papers	Universitas Muslim Indonesia on 2024-12-09	<1%
159	Student papers	Universitas Muslim Indonesia on 2025-08-16	<1%
160	Internet	journal.aptii.or.id	<1%
161	Internet	journal.itny.ac.id	<1%
162	Internet	journal.nurulfikri.ac.id	<1%
163	Publication	Budi, Eka Setia. "Metode Fuzzy Logic Untuk Penentuan Level Tindakan Pemelihar...	<1%
164	Student papers	IAKN Palangka Raya on 2026-06-14	<1%
165	Internet	eproceeding.instiki.ac.id	<1%

166	Publication	Andriani Nurian. "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE PL...	<1%
167	Publication	Cahyani, Elvana. "Pengaruh pemberian gel ekstrak bunga telang terhadap ekspr...	<1%
168	Publication	Indrasari, Luthfiyah. "Pengaruh Intellectual Capital Terhadap Nilai Perusahaan D...	<1%
169	Publication	M. Audi Alfaridzy, Elin Haerani, Jasril -, Lola Oktavia. "KLASIFIKASI SENTIMEN MAS...	<1%
170	Publication	Maharani, Ifa Agustin. "Analisis Potensi Likuifaksi dan Daya Dukung Fondasi Tan...	<1%
171	Publication	Muhammad Azis Sularso, Jupron Jupron. "Pengaruh ketidakseimbangan data ter...	<1%
172	Publication	Parabelem Tinno Dolf Rompas, Shella Sasoeng, Gladly Caren Rorimpandey. "Decis...	<1%
173	Publication	Tarwoto, Rizki Nugroho, Najmul Azka, Wakhid Sayudha Rendra Graha. "Analisis S...	<1%
174	Student papers	Universitas Budi Luhur on 2025-07-16	<1%
175	Student papers	Universitas Islam Indonesia on 2026-02-12	<1%
176	Student papers	Universitas Pamulang on 2026-07-13	<1%
177	Student papers	Universitas Pamulang on 2026-07-29	<1%
178	Student papers	Universitas Pancasila on 2026-02-21	<1%
179	Student papers	Universitas Putera Batam on 2026-02-02	<1%

180	Internet	archive.umsida.ac.id	<1%
181	Internet	ebook.umpwr.ac.id	<1%
182	Internet	journal.artei.or.id	<1%
183	Internet	jpi.eng.unila.ac.id	<1%
184	Internet	prosiding.aritekin.or.id	<1%
185	Publication	Annisa Saninah, Willy Prihartono, Cep Lukman Rohmat. "ANALISIS SENTIMEN UL...	<1%
186	Publication	Ennianti Hasibuan, Elmo Allistair Heriyanto. "ANALISIS SENTIMEN PADA ULASAN ...	<1%
187	Publication	Heppy Nur Asavia Ginasputri, Atika Dwi Saputri, Siti Nurasriyanti Wahid, Ariska Fi...	<1%
188	Publication	Kasih Delayana, Setia Mangiring Marpaung, Syukur Nimei Iman Gulo, Sardo Sipay...	<1%
189	Publication	Rivan Kurnia, Teguh Nurhadi Suharsono. "Sentimen Analisis Review Pada Aplikasi...	<1%
190	Publication	Sugeng Setyabudi, Enny Aryanny. "Sentiment Analysis of Lazada Marketplace Use...	<1%
191	Student papers	Universitas Budi Luhur on 2025-07-11	<1%
192	Student papers	Universitas Pamulang on 2026-08-05	<1%
193	Student papers	Universitas Pancasila on 2026-02-23	<1%

194	Publication	Afriyanto Afriyanto, Usman Usman. "Analisis Sentimen pada Google Play Store Pe...	<1%
195	Publication	Dina Siti Nurrochmah, Nining Rahaningsih, Raditya Danar Dana, Cep Lukman Roh...	<1%
196	Publication	Indah Sartika, Imam Saputra. "Analisis Sentimen Masyarakat Terhadap Pembatas...	<1%
197	Publication	Nanda Dwi Husna Sadikin, Sari Susanti. "Analisis Sentimen Publik Terhadap Kamp...	<1%
198	Student papers	UPN Veteran Yogyakarta on 2025-12-22	<1%
199	Student papers	Universitas Islam Indonesia on 2026-07-30	<1%
200	Student papers	Universitas Islam Riau on 2025-10-31	<1%
201	Student papers	Universitas Muhammadiyah Makassar on 2026-06-25	<1%
202	Student papers	Universitas Muslim Indonesia on 2025-08-08	<1%
203	Student papers	Universitas Pamulang on 2026-05-21	<1%
204	Student papers	Universitas Pancasila on 2025-08-07	<1%
205	Publication	Wanda Putra Ramadhan, Didi Juardi. "ANALISIS SENTIMEN ULASAN APLIKASI BTN...	<1%

44

25

**ANALISIS SENTIMEN PENGGUNA TERHADAP
APLIKASI DOLA DI PLAY STORE MENGGUNAKAN
ALGORITMA NAIVE BAYES**

SKRIPSI



Disusun oleh

ROCKY SAPUTRA

403221010086

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN IILMU KOMPUTER
UNIVERSITAS ISLAM INDRAGIRI**

TEMBILAHAN

2026

**ANALISIS SENTIMEN PENGGUNA TERHADAP
APLIKASI DOLA DI PLAY STORE MENGGUNAKAN
ALGORITMA NAIVE BAYES**

**USER SENTIMENT ANALYSIS OF THE DOLA APP
ON THE PLAY STORE USING THE NAIVE BAYES
ALGORITHM**

SKRIPS



Disusun oleh

ROCKY SAPUTRA

403221010086

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS ISLAM INDRAGIRI**

TEMBILAHAN

2026

PERSETUJUAN**ANALISIS SENTIMEN PENGGUNA TERHADAP APLIKASI DOLA DI
PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES**

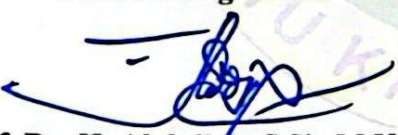
Yang dipersiapkan dan disusun oleh

ROCKY SAPUTRA

403221010086

Telah disetujui oleh Dosen Pembimbing
pada tanggal 17 Desember 2025

Pembimbing Utama


Prof. Dr. H. Abdullah, S.Si., M.Kom
NIDN. 1008037001

Pembimbing Pendamping


Fitri Yunita, S.Si., M.Kom
NIDN. 1012089001

Dekan Fakultas Teknik dan Ilmu Komputer




Dr. Bayu Rianto, S.Si., M.Kom
NIPY. 1691 09 321

PENGESAHAN**ANALISIS SENTIMEN PENGGUNA TERHADAP APLIKASI DOLA DI
PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES**

Dipersiapkan dan disusun oleh

ROCKY SAPUTRA
403221010086

Telah Diuji dan Dipertahankan di depan Dewan Penguji
pada Tanggal 24 Juli 2026

Susunan Dewan Penguji

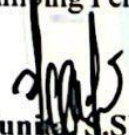
Pembimbing Utama


Prof. Dr. H. Abdullah, S.Si., M.Kom
NIDN. 1008037001

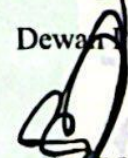
Dewan Penguji I


Usman, ST., M.Kom
NIDN. 1017078301

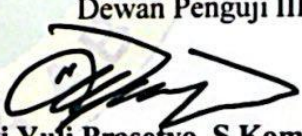
Pembimbing Pendamping


Fitri Yunita, S.Si., M.Kom
NIDN. 1012089001

Dewan Penguji II

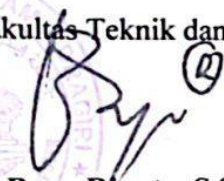

Samsudin, S.Kom., M.Kom
NIDN. 1009098501

Dewan Penguji III


Dwi Yuli Prasetyo, S.Kom., M.Kom
NIDN. 1020078602

Tugas Akhir ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

Dekan Fakultas Teknik dan Ilmu Komputer


Dr. Bayu Rianto, S.Si., M.Kom
NIPY. 1691 09 321

Ka. Prodi Sistem Informasi


Fitri Yunita, S.Si., M.Kom
NIPY. 11590 10 293

25

PERNYATAAN

Saya yang bertanda tangan di bawah ini menyatakan bahwa skripsi yang saya susun merupakan hasil penelitian, pemikiran, dan karya saya sendiri. Seluruh proses penyusunan skripsi dilakukan dengan menjunjung tinggi etika akademik. Setiap pendapat, teori, data, maupun hasil penelitian yang berasal dari karya orang lain telah dicantumkan sumbernya secara jelas sesuai dengan kaidah penulisan ilmiah yang berlaku.

149

44

Saya juga menyatakan bahwa skripsi ini tidak pernah diajukan untuk memperoleh gelar akademik pada perguruan tinggi mana pun. Apabila di kemudian hari ditemukan adanya pelanggaran terhadap pernyataan ini, termasuk penggunaan karya orang lain tanpa mencantumkan sumber yang semestinya, saya bersedia menerima segala konsekuensi akademik sesuai dengan ketentuan yang berlaku.

43

Demikian pernyataan ini saya buat dengan sebenar-benarnya, tanpa adanya paksaan dari pihak mana pun, untuk dapat dipergunakan sebagaimana mestinya.

Tembilahan 12 Mei 2026
Penulis

Rocky Saputra
403221010086

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi DOLA berdasarkan ulasan pada Google Play Store menggunakan algoritma Naive Bayes. Data penelitian diperoleh melalui web scraping sebanyak 5.000 ulasan, kemudian diklasifikasikan menjadi sentimen positif, negatif, dan netral berdasarkan rating pengguna. Tahapan penelitian meliputi preprocessing, pembobotan TF-IDF, oversampling, klasifikasi Naive Bayes, dan evaluasi menggunakan accuracy, precision, recall, F1-score, serta Confusion Matrix. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi ulasan pengguna. Model Naive Bayes memperoleh accuracy sebesar 82,37%, precision 0,83, recall 0,82, dan F1-score 0,82. Hasil tersebut menunjukkan bahwa algoritma Naive Bayes cukup efektif dalam mengklasifikasikan sentimen pengguna terhadap aplikasi DOLA dan dapat menjadi bahan evaluasi bagi pengembang dalam meningkatkan kualitas layanan aplikasi.

Kata kunci : Analisis Sentimen, DOLA, Google Play Store, Naive Bayes, TF-IDF.

ABSTRACT

This study aims to analyze user sentiment regarding the DOLA application based on Google Play Store reviews using the Naive Bayes algorithm. Research data was obtained through web scraping, yielding 5,000 reviews, which were then classified into positive, negative, and neutral sentiments based on user ratings. The research stages included preprocessing, TF-IDF weighting, oversampling, Naive Bayes classification, and evaluation using accuracy, precision, recall, F1-score, and a confusion matrix. The results indicate that positive sentiment dominates user reviews. The Naive Bayes model achieved an accuracy of 82.37%, precision of 0.83, recall of 0.82, and an F1-score of 0.82. These results demonstrate that the Naive Bayes algorithm is effective in classifying user sentiment toward the DOLA application and can serve as a basis for developers to improve the application's service quality.

Keywords: Sentiment Analysis, DOLA, Google Play Store, Naive Bayes, TF-IDF.

KATA PENGANTAR

Alhamdulillah rabbil 'alamin, segala puji syukur penulis ucapkan ke hadirat Allah SWT, atas berkat rahmat, hidayah, serta karunia-Nya. Berkat kesehatan, kekuatan sehinggapenulis dapat menyelesaikan skripsi yang berjudul “*Analisis Sentimen Pengguna Terhadap Aplikasi Dola di Play Store Menggunakan Algoritma Naive Bayes*”. Penyusunan skripsi ini bertujuan mengidentifikasi pemanfaatan teknologi kecerdasan buatan dalam mengolah data teks, khususnya dalam menganalisis opini pengguna terhadap suatu aplikasi sehingga dapat menghasilkan informasi yang bermanfaat bagi pengembangan layanan digital.

Pesatnya perkembangan teknologi informasi mendorong pemanfaatan Artificial Intelligence (AI) dalam pengolahan data teks. Salah satu penerapannya adalah analisis sentimen terhadap ulasan pengguna aplikasi pada Google Play Store untuk memperoleh informasi mengenai kepuasan maupun keluhan pengguna. Penelitian ini memanfaatkan algoritma Naive Bayes untuk mengelompokkan ulasan pengguna aplikasi DOLA ke dalam kategori sentimen positif, negatif, dan netral berdasarkan isi ulasan yang diberikan. Penulis menyadari masih terdapat keterbatasan dalam penyusunan skripsi ini. Oleh sebab itu, masukan yang bersifat membangun diharapkan dapat menjadi bahan perbaikan pada penelitian berikutnya. Semoga penelitian dapat memberikan manfaat dan kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang Analisis Sentimen dan data mining. Selama proses penyusunan skripsi ini, penulis memperoleh banyak dukungan, arahan, serta bantuan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Bapak Dr. H. Najamuddin, Lc., M.A, selaku Rektor Universitas Islam Indragiri, atas fasilitas dan dukungan yang diberikan sehingga penyusunan proposal ini dapat berjalan dengan baik.
2. Bapak Dr. Bayu Rianto, S.Si., M.Kom, selaku Dekan Fakultas Teknik dan Ilmu Komputer, atas bimbingan dan arahnya selama penulis menempuh pendidikan.
3. Ibu Fitri Yunita, S.Si., M.Kom, selaku Ketua Program Studi Sistem Informasi sekaligus pembimbing Pendamping, yang dengan sabar memberikan penjelasan, koreksi, dan arahan sehingga proposal tersebut dapat terselesaikan dengan baik.
4. Bapak Prof. Dr. Abdullah, S.Si., M.Kom selaku Pembimbing utama yang telah meluangkan waktunya untuk memberikan arahan, bimbingan, serta ilmuilmu yang bermanfaat dari awal hingga selesainya skripsi ini.
5. Seluruh dosen, staf, dan civitas akademika Program Studi Sistem Informasi Universitas Islam Indragiri yang telah memberikan ilmu dan pengalaman berharga bagi penulis.
6. Kedua orang tua dan keluarga yg tersayang, atas cinta, doa, semangat, serta pengorbanan dan dukungan yang begitu besar sehingga penulis dapat menyelesaikan pendidikan hingga tahap ini.
7. Teman-teman seperjuangan, sahabat diskusi, serta rekan-rekan mahasiswa atas dukungan yang telah diberikan, kebersamaan, atas motivasi selama penyusunan proposal ini.
8. Terimakasih buat semua pihak yang tidak bisa disebutkan satu persatu atas segala bantuan, do'a dan dukungannya selama proses penyusunan skripsi ini.

9. Dan Terimakasih kepada diri saya sendiri, Rocky Saputra, yang sudah mampu berusaha keras, berjuang dan sampai mampu berada dititik saat ini.

Penulis menyadari bahwa proposal ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi penyempurnaan penelitian selanjutnya. Semoga proposal tersebut dapat memberikan manfaat bagi pembaca, khususnya dalam pengembangan Analisis *Sentimen* dan penerapan kecerdasan buatan pada aplikasi digital.

Akhir kata, semoga Allah SWT senantiasa memberikan kemudahan dan keberkahan dalam setiap proses menuntut ilmu dan penyelesaian penelitian ini.

Tembilahan 12 Mei 2026

Rocky Saputra

NIM : 403221010086

DAFTAR ISI

ABSTRAK	x
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
1.6 Sistematika Penelitian	6
BAB II TINJAUAN LITERATUR.....	9
2.1 Analisis Sentimen	9
2.2 Algoritma Naive Bayes.....	10
2.3 Aplikasi Dola AI	11
2.4 Google Play Store sebagai Sumber Data.....	12
2.5 Penelitian Terdahulu.....	13
2.6 Rangkuman.....	27
BAB III	28

METODE PENELITIAN.....	28
3.1. Kerangka Penelitian	28
3.2. Identifikasi Masalah	30
3.3. Tinjauan Literatur	30
3.4. Pengumpulan Data	30
3.5. <i>Preprocessing</i>	31
3.6. Transformasi <i>TF-IDF</i>	32
3.7. Data Mining Algoritma <i>Naive Bayes</i>	32
3.8. Evaluasi	32
3.9. <i>Word Cloud</i>	34
BAB IV	35
HASIL DAN PEMBAHASAN.....	35
4.1. Hasil.....	35
4.1.1 <i>Data Collection</i>	36
4.1.2 <i>Preprocessing</i> Data.....	38
4.2 Pembagian data.....	42
4.3 Transformasi Data (<i>TF-IDF</i>).....	43
4.4 Klasifikasi <i>Naive Bayes</i>	46
4.5 Evaluasi model	47
4.5.1 Evaluasi <i>Cross Validation</i>	51
4.5.2 Pengujian Skenario Pembagian Data	51

39	4.6 Perbandingan Metode Naive Bayes dan Indobert	52
	4.7 <i>Word Cloud</i>	53
	4.8 Pembahasan	56
27	4.8.1 Perbandingan Hasil Penelitian dengan Literatur	57
	BAB V	58
	5.1 Kesimpulan	58
	5.2 Saran	59
	DAFTAR PUSTAKA	xiii

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	13
Tabel 4. 1 Pembagian Data	52

DAFTAR GAMBAR

30	Gambar 3. 1 Kerangka Penelitian	29
	Gambar 4. 1 Proses Pengumpulan Data	36
	Gambar 4. 2 Kode Program <i>Data Collection</i>	37
	Gambar 4. 3 Kode Program <i>Data Selection</i>	37
	Gambar 4. 4 Kode Program <i>Data Labeling</i>	38
	Gambar 4. 5 Kode Program <i>Case Folding</i>	39
	Gambar 4. 6 Kode Program <i>Cleaning</i>	40
51	Gambar 4. 7 Kode Program <i>Tokenization</i>	40
	Gambar 4. 8 Kode Program <i>Stopword Removal</i>	41
	Gambar 4. 9 Kode Program <i>Stemming</i>	41
	Gambar 4. 10 Kode Program <i>TF-IDF</i>	45
	Gambar 4. 11 Kode Program <i>Oversampling</i>	45
	Gambar 4. 12 Kode Program <i>Train-Test Split</i>	47
24	Gambar 4. 13 Hasil Confusion Matrix	49
	Gambar 4. 14 Hasil Akurasi	50
	Gambar 4. 15 Hasil Evaluasi Model pada setiap kelas sentimen	51
168	Gambar 4. 16 Hasil Perbandingan Metode Naive Bayes dan IndoBERT	53
	Gambar 4. 17 Hasil <i>Word Cloud</i>	54
38	Gambar 4. 18 Hasil <i>Word CLOUD</i> Positif	55
	Gambar 4. 19 Hasil <i>Word Cloud</i> Negatif	55
	Gambar 4. 20 Hasil <i>Word Cloud</i> Netral	56

80

BAB 1

PENDAHULUAN

1.1 Latar Belakang

159

Kemajuan teknologi informasi telah mengubah berbagai aspek kehidupan, termasuk cara masyarakat memperoleh informasi, berkomunikasi, dan memanfaatkan layanan digital dalam aktivitas sehari-hari. Meningkatnya akses internet turut mempercepat penggunaan berbagai aplikasi berbasis perangkat seluler yang memudahkan pengguna memperoleh layanan secara praktis. Seiring dengan meningkatnya penggunaan aplikasi digital, jumlah data yang dihasilkan juga semakin besar dan beragam. Data-data tersebut dapat memberikan informasi yang bermanfaat apabila diolah dan dianalisis dengan baik. Salah satu platform yang menyediakan berbagai aplikasi digital yaitu *Google Play Store*, yang memungkinkan pengguna menggunakan aplikasi, pengguna dapat memberikan penilaian dalam bentuk rating serta ulasan sebagai bentuk umpan balik terhadap kualitas layanan Informasi yang terdapat pada ulasan pengguna dapat dimanfaatkan untuk memahami pengalaman penggunaan aplikasi sekaligus mengidentifikasi tingkat kepuasan maupun permasalahan yang dirasakan pengguna [1].

202

1

1

50

17

Seiring dengan meningkatnya penggunaan aplikasi mobile, jumlah ulasan diberikan oleh pengguna pada platform seperti *Google Play Store* juga semakin banyak. Ulasan tersebut biasanya berisi berbagai pendapat, pengalaman, kritik, maupun saran terhadap kinerja dan kualitas suatu aplikasi. Informasi yang terkandung dalam ulasan pengguna tersebut penting bagi pengembang aplikasi untuk mengetahui tingkat kepuasan pengguna serta mengidentifikasi berbagai

6

permasalahan yang mungkin terjadi pada aplikasi yang mereka kembangkan. Namun, banyaknya jumlah ulasan yang tersedia seringkali menyulitkan proses analisis apabila dilakukan secara manual. Oleh karena itu, diperlukan suatu metode yakni mampu mengolah dan menganalisis data ulasan pengguna secara otomatis sehingga bisa menghasilkan informasi yang lebih efektif dan efisien[2].

Banyaknya ulasan yang dihasilkan pengguna membuat proses analisis secara manual menjadi kurang efisien. Oleh karena itu, diperlukan metode komputasi yang mampu mengelompokkan opini pengguna secara otomatis berdasarkan isi ulasan. Salah satu teknik yakni dapat digunakan adalah analisis *sentimen*, yaitu metode dalam pengolahan teks yang bertujuan untuk mengekstrak, mengidentifikasi, maupun mengklasifikasikan opini atau pendapat yang terdapat dalam suatu teks ke dalam kategori *sentimen* seperti positif, negatif, dan netral.

Analisis *Sentimen* banyak digunakan dalam berbagai bidang seperti e-commerce, media sosial, maupun layanan digital untuk mengetahui persepsi pengguna terhadap suatu produk atau layanan. Dengan memanfaatkan teknik analisis *sentimen*, data ulasan pengguna dengan berjumlah besar dapat dianalisis secara lebih sistematis sehingga menghasilkan informasi yakni dapat membantu pengembang dalam mengevaluasi dan meningkatkan kualitas aplikasi[3].

Dalam proses analisis *sentimen*, diperlukan suatu metode klasifikasi mampu mengelompokkan data teks ke dalam kategori *sentimen* yang telah ditentukan, seperti positif, negatif, maupun netral. Salah satu algoritma yang banyak digunakan dalam penelitian Analisis *Sentimen* adalah *Naïve Bayes*, yaitu metode klasifikasi berbasis probabilitas yaitu mengacu pada Teorema Bayes untuk menentukan kemungkinan suatu data termasuk ke dalam kelas tertentu. Algoritma

ini banyak diterapkan karena memiliki keunggulan dalam kesederhanaan perhitungan, efisiensi dalam pengolahan data teks, serta kemampuan menghasilkan performa yang cukup baik meskipun menggunakan dataset yang relatif sederhana. Selain itu, *Naïve Bayes* juga efektif digunakan pada data berukuran besar dan tidak terstruktur seperti data media social. Dengan memanfaatkan algoritma ini, data berupa ulasan atau opini pengguna dapat diklasifikasikan secara otomatis sebagai kategori *sentimen* yang sesuai, sehingga dapat membantu dalam memahami kecenderungan opini publik pada suatu topik, produk, atau layanan secara lebih cepat, sistematis, dan mendukung pengambilan keputusan berbasis data[4].

Berbagai penelitian sebelumnya menunjukkan bahwa *Analisis Sentimen* berdasarkan ulasan pengguna pada aplikasi mobile di *Google Play Store* dapat memberikan informasi penting mengenai tingkat kepuasan pengguna serta berbagai permasalahan yang terdapat pada aplikasi tersebut. Ulasan pengguna sering kali memuat berbagai keluhan seperti kesulitan untuk mengakses aplikasi, gangguan teknis, maupun keterbatasan fitur yang tersedia. Oleh karena itu, *Analisis Sentimen* menjadi salah satu pendekatan sehingga efektif untuk memahami persepsi pengguna terhadap suatu aplikasi. Dengan memanfaatkan metode klasifikasi seperti *Naive Bayes*, data ulasan yaitu diperoleh dari *Google Play Store* dapat dianalisis untuk mengetahui kecenderungan *sentimen* pengguna serta memberikan rekomendasi bagi pengembang untuk meningkatkan kualitas layanan aplikasi[5].

Berdasarkan permasalahan tersebut, penelitian ini dilakukan dengan tujuan sebagai menganalisis *sentimen* pengguna terhadap aplikasi DOLA berdasarkan

ulasan yang terdapat pada platform *Google Play Store* dengan menggunakan algoritma *Naive Bayes*. Melalui penelitian ini, ulasan pengguna akan diproses dan diklasifikasikan secara otomatis terhadap kategori *sentimen* positif, negatif, dan netral sehingga dapat memberikan gambaran mengenai persepsi serta tingkat kepuasan pengguna terhadap aplikasi DOLA. Dengan adanya *Analisis Sentimen* ini, diharapkan dapat diketahui kecenderungan opini pengguna terhadap kinerja, fitur, maupun kualitas layanan yang diberikan oleh aplikasi tersebut. Hasil dari penelitian tersebut diharapkan dapat memberikan manfaat secara teoritis maupun praktis. Secara teoritis, penelitian ini diharapkan dapat menambah referensi dan wawasan dalam bidang text mining khususnya pada penerapan metode *Naive Bayes* sebagai *Analisis Sentimen* terhadap data ulasan pengguna aplikasi. Sementara secara praktis, hasil penelitian ini diharapkan dapat menjadi bahan evaluasi dan masukan bagi pihak pengembang aplikasi DOLA dalam meningkatkan kualitas layanan, memperbaiki kekurangan aplikasi, serta meningkatkan kepuasan pengguna di masa yang akan datang.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang sebelumnya, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Belum adanya analisis yang mendalam mengenai *sentimen* pengguna terhadap aplikasi Dola berdasarkan ulasan di *Google Play Store*.
2. Belum adanya penerapan algoritma *Naive Bayes* secara khusus untuk mengklasifikasikan *sentimen* pada ulasan pengguna aplikasi Dola.
3. Belum adanya pengujian akurasi algoritma *Naive Bayes* dalam menganalisis dan mengidentifikasi *sentimen* pengguna terhadap aplikasi Dola di *Play Store*.

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan tidak menyimpang dari tujuan, maka ditetapkan batasan masalah sebagai berikut:

1. Penelitian ini hanya berfokus pada Analisis *Sentimen* pengguna terhadap aplikasi Dola yang diambil dari ulasan pada platform *Google Play Store*.
2. Data yang digunakan terbatas pada ulasan berbahasa Indonesia dan diperoleh dalam rentang waktu tertentu sesuai kebutuhan penelitian.
3. Metode klasifikasi *sentimen* yang digunakan dalam penelitian ini dibatasi pada penerapan algoritma *Naive Bayes*, tanpa membandingkannya dengan algoritma lain.
4. Kategori *sentimen* ini digunakan dalam penelitian yang dibatasi menjadi tiga kelas, yaitu positif, negatif, dan netral.
5. Evaluasi performa model hanya dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk menilai tingkat keberhasilan algoritma *Naive Bayes* dalam klasifikasi *sentimen*.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini untuk menggambarkan hasil yang telah direncanakan. Antara lain yakni:

1. Mengetahui dan menganalisis *sentimen* pengguna terhadap aplikasi Dola berdasarkan ulasan yang terdapat di *Google Play Store*.
2. Mengklasifikasikan ulasan pengguna terhadap kategori *sentimen* positif, negatif, dan netral menggunakan algoritma *Naive Bayes* meliputi tahapan *preprocessing*, pelabelan data, dan pembobotan *TF-IDF*.

3. Mengevaluasi kinerja model *Naive Bayes* terhadap mengklasifikasikan *sentimen* pengguna menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk mengetahui tingkat efektivitas model.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini:

1. Memberikan gambaran mengenai *sentimen* pengguna aplikasi Dola berdasarkan ulasan di *Google Play Store* sehingga dapat digunakan sebagai acuan dalam memahami persepsi pengguna.
2. Menyediakan hasil klasifikasi *sentimen* yang dapat membantu pengembangan aplikasi dalam mengevaluasi kualitas layanan dan meningkatkan fitur yang dianggap kurang oleh pengguna.
3. Menjadi referensi terhadap peneliti atau pihak lain yang ingin melakukan penelitian terkait analisis *sentimen*, khususnya yang menggunakan algoritma *Naive Bayes* dan metode *TF-IDF* sebagai pendekatan pengolahan teks.

1.6 Sistematika Penelitian

Sistematika penulisan yang digunakan pada penelitian ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Pada bab I ini terdapat latar belakang masalah yang menjadi latar belakang penelitian ini dilakukan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II : TINJAUAN LITERATUR

Bab II ini terdapat 10 jurnal penelitian terdahulu yang berkaitan dengan Analisis Sentimen dan algoritma Naive Bayes untuk dibandingkan dengan penelitian yang dilakukan. Pada bagian akhir disajikan rangkuman hasil tinjauan penelitian terdahulu sebagai dasar penguatan teori dan identifikasi celah penelitian.

BAB III : METODE PENELITIAN

Bab ini menguraikan metode yang diterapkan dalam penelitian menggunakan pendekatan kuantitatif deskriptif dengan algoritma Naive Bayes untuk melakukan klasifikasi sentimen. Tahapan penelitian meliputi penyusunan kerangka penelitian, identifikasi permasalahan, pengambilan data ulasan aplikasi DOLA melalui teknik web scraping pada Google Play Store, seleksi data, preprocessing teks, pembobotan menggunakan TF-IDF, pelabelan sentimen, pembentukan model Naive Bayes, serta evaluasi kinerja model berdasarkan nilai akurasi, presisi, recall, dan F1-score.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian dan pembahasan proses Analisis Sentimen pengguna aplikasi DOLA di Google Play Store menggunakan algoritma Naive Bayes. Pada bab ini dijelaskan tahapan penelitian dimulai dari proses pengumpulan data ulasan menggunakan teknik web scraping, data Selection, pelabelan sentimen, serta preprocessing data yang meliputi case folding, cleaning, tokenization, stopwords removal, dan stemming. Selanjutnya, data teks ditransformasikan menerapkan metode TF-IDF sebagai bentuk numerik untuk kemudian dilakukan proses klasifikasi menggunakan algoritma Naive Bayes. Setelah itu, dilakukan pembagian data sebagai data training dan data testing

dengan beberapa skenario pembagian data untuk mengetahui pengaruh perbedaan proporsi data terhadap kinerja model. Evaluasi model dilakukan menggunakan *Confusion Matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta pengujian menggunakan metode *k-fold cross validation* untuk memperoleh hasil evaluasi yang lebih stabil. Pada tahap akhir, dilakukan visualisasi *Word Cloud* guna menampilkan kata-kata yang paling sering muncul pada setiap kategori *sentimen* sehingga dapat memberikan gambaran mengenai kecenderungan opini pengguna terhadap aplikasi DOLA.

BAB V : KESIMPULAN DAN SARAN

Pada bab ini merupakan bagian akhir dari penelitian yang berisi kesimpulan dan saran berdasarkan hasil Analisis *Sentimen* pengguna aplikasi DOLA menggunakan algoritma *Naive Bayes*. Kesimpulan berisi rangkuman hasil penelitian, termasuk performa model klasifikasi yang diperoleh berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* serta gambaran kecenderungan *sentimen* pengguna terhadap aplikasi DOLA. Selanjutnya, bagian saran berisi rekomendasi ini dapat dipakai sebagai bahan pengembangan penelitian berikutnya, seperti penggunaan jumlah data yang lebih besar, penerapan metode klasifikasi lain, maupun peningkatan tahapan *preprocessing* agar menghasilkan performa Analisis *Sentimen* yang lebih optimal.

12

BAB II

TINJAUAN LITERATUR

Bab ini mengulas ringkasan penelitian terdahulu yang berkaitan dengan Analisis Sentimen terhadap ulasan pengguna aplikasi dengan memanfaatkan metode klasifikasi dalam bidang text mining. Tinjauan literatur dilakukan untuk mengetahui hasil penelitian sebelumnya serta melihat efektivitas metode yang digunakan dalam mengolah data ulasan pengguna. Berdasarkan hasil kajian dari beberapa penelitian sebelumnya, algoritma Naive Bayes sering dipilih untuk analisis sentimen karena prosesnya sederhana, memiliki waktu komputasi yang cepat, serta mampu memberikan hasil klasifikasi dengan tingkat akurasi yang baik dalam membedakan sentimen positif, negatif, dan netral. Metode ini bekerja berdasarkan konsep probabilitas dari Teorema Bayes, yaitu menghitung kemungkinan suatu dokumen atau teks termasuk ke dalam kelas tertentu berdasarkan kata-kata yang muncul di dalamnya.

2.1 Analisis Sentimen

Analisis Sentimen adalah salah satu cabang dari text mining yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini, perasaan, atau pendapat yang terdapat dalam suatu teks [6]. Analisis Sentimen dimanfaatkan untuk memahami bagaimana pendapat pengguna terhadap suatu produk atau layanan berdasarkan ulasan yang mereka berikan. Melalui proses ini, opini pengguna dapat dikelompokkan ke dalam sentimen positif, negatif, atau netral sehingga memudahkan proses evaluasi terhadap produk atau layanan tersebut.

Dalam praktiknya, Analisis Sentimen sering digunakan untuk menganalisis ulasan pengguna pada media sosial, forum diskusi, maupun platform aplikasi

174

5

20

76

1

17

119

14

68

digital. Hasil Analisis *Sentimen* biasanya diklasifikasikan ke dalam beberapa kategori seperti *sentimen* positif, negatif, dan netral. Penerapan analisis sentimen memungkinkan data ulasan pengguna diproses secara otomatis meskipun jumlahnya cukup banyak. Hasil analisis tersebut memberikan gambaran mengenai kecenderungan pendapat pengguna sehingga dapat dimanfaatkan sebagai bahan evaluasi untuk meningkatkan kualitas produk atau layanan.

2.2 Algoritma *Naive Bayes*

Naive Bayes adalah salah satu algoritma klasifikasi berbasis probabilitas yang menggunakan konsep Bayes' Theorem [7]. Algoritma tersebut bekerja dengan menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan fitur atau kata-kata yang terdapat dalam dokumen.

Disebut *naive* karena algoritma ini mengasumsikan bahwa setiap fitur atau kata dalam dokumen bersifat independen satu sama lain, meskipun dalam kenyataannya hubungan antar kata bisa saja saling berkaitan. Meskipun memiliki asumsi yang sederhana, algoritma *Naive Bayes* terbukti efektif dalam berbagai penelitian, terutama dalam klasifikasi data teks seperti spam detection, analisis *sentimen*, dan kategorisasi dokumen.

- a) Keunggulan algoritma *Naive Bayes* antara lain:
- b) Proses perhitungan yang sederhana dan cepat
- c) Mampu menangani data dalam jumlah besar
- d) Memberikan hasil klasifikasi teks yang cukup akurat
- e) Mudah diimplementasikan

Dalam penelitian ini, algoritma *Naive Bayes* digunakan untuk mengklasifikasikan ulasan pengguna aplikasi Dola ke dalam kategori *sentimen*

positif, negatif, dan netral berdasarkan kata-kata yang muncul dalam ulasan tersebut. Hasil klasifikasi ini diharapkan dapat memberikan gambaran mengenai persepsi pengguna terhadap aplikasi Dola.

2.3 Aplikasi Dola AI

Dola AI adalah aplikasi asisten berbasis kecerdasan buatan yang dimanfaatkan untuk memudahkan pengguna dalam mengelola jadwal, membuat pengingat kegiatan, serta mengatur aktivitas sehari-hari secara lebih terorganisir dengan memanfaatkan teknologi *Artificial Intelligence* dan pemrosesan bahasa alami. Melalui berbagai fitur yang tersedia, aplikasi ini memungkinkan pengguna untuk membuat agenda, mengatur jadwal pertemuan, serta memperoleh pengingat otomatis terhadap kegiatan yang telah dijadwalkan. Sebagai aplikasi yang tersedia pada platform Google *Play Store*, Dola AI juga menerima berbagai ulasan dan penilaian dari pengguna sesuai dengan pengalaman yang dirasakan pengguna saat menggunakan aplikasi tersebut. Ulasan ini dapat dimanfaatkan sebagai sumber data dalam penelitian Analisis *Sentimen* untuk mengetahui persepsi pengguna terhadap kualitas dan kinerja aplikasi Dola AI.

Aplikasi DOLA AI pertama kali diperkenalkan pada tahun 2023 oleh SPRING (SG) PTE, LTD, dengan nama Cici sebagai aplikasi asisten berbasis kecerdasan buatan yang memanfaatkan teknologi *Artificial Intelligence (AI)* dan *Natural Language Processing (NLP)*. Aplikasi ini dikembangkan untuk membantu pengguna dalam menjawab pertanyaan, menyusun tulisan, mengelola pekerjaan, serta mendukung aktivitas belajar dan produktivitas sehari-hari [8].

Seiring dengan perkembangan fitur dan peningkatan kemampuan layanan, pada akhir tahun 2025 pengembang melakukan rebranding dengan mengubah

nama aplikasi dari Cici menjadi DOLA. Berdasarkan informasi resmi pada riwayat pembaruan aplikasi, perubahan tersebut hanya merupakan pergantian identitas aplikasi, sedangkan berbagai fitur utama tetap dipertahankan dan akan terus dikembangkan melalui pembaruan berikutnya [8].

Setelah berganti nama menjadi DOLA, aplikasi terus dikembangkan dengan penambahan berbagai fitur berbasis AI, seperti pembuatan teks, bantuan belajar, analisis dokumen, pembuatan gambar, serta asisten produktivitas. Perkembangan tersebut menjadikan DOLA sebagai salah satu aplikasi AI yang memperoleh banyak pengguna dan ulasan di Google Play Store, sehingga layak dijadikan objek dalam penelitian analisis sentiment [8].

2.4 Google Play Store sebagai Sumber Data

Google Play Store adalah layanan distribusi aplikasi berbasis Android yang dikelola oleh Google. Platform ini memudahkan pengguna dalam menemukan, mengunduh, memasang, dan memperbarui berbagai aplikasi sesuai kebutuhan. Selain itu, pengguna juga dapat memberikan rating dan ulasan sebagai bentuk umpan balik terhadap aplikasi yang digunakan., memberikan penilaian berupa rating, serta menuliskan ulasan berdasarkan pengalaman mereka saat menggunakan aplikasi tersebut. Fitur ulasan pada Google Play Store memungkinkan pengguna menyampaikan pengalaman, kritik, dan saran setelah menggunakan suatu aplikasi. Informasi tersebut dapat dimanfaatkan sebagai bahan evaluasi untuk mengetahui kualitas aplikasi dari sudut pandang pengguna. [9], sehingga dapat dimanfaatkan dalam penelitian Analisis Sentimen untuk mengetahui tingkat kepuasan pengguna serta mengidentifikasi berbagai permasalahan yang terdapat pada aplikasi yang digunakan.

2.5 Penelitian Terdahulu

Berikut disajikan beberapa penelitian terdahulu yang relevan dengan topik penelitian ini:

Tabel 2. 1 Penelitian Terdahulu

NO	Nama Peneliti	Judul Penelitian	Pembahasan
(1)	(2)	(3)	(4)
1.	Riki Roska Rismansyah, Aso Sudiarjo, Teuku Mufizar (2025)	Analisis <i>Sentimen</i> Ulasan Shopee pada Google Play Store Menggunakan Algoritma <i>Naive Bayes</i>	Penelitian ini menganalisis sentimen pengguna aplikasi Shopee berdasarkan ulasan di Google Play Store menggunakan algoritma Naive Bayes. Sebanyak 2.000 data ulasan dikumpulkan melalui teknik web scraping dan diproses menggunakan tahapan preprocessing. Hasil penelitian memperoleh akurasi sebesar 83%, dengan sentimen negatif lebih dominan dibandingkan sentimen positif. Temuan ini dapat dimanfaatkan sebagai bahan evaluasi untuk meningkatkan kualitas layanan Shopee.

Tabel 2.1 Penelitian Terdahulu (Lanjutan)

NO	Nama Peneliti	Judul Penelitian	Pembahasan
(1)	(2)	(3)	(4)
2.	Divia Azarine, Nining Rahaningsih, Raditya Danar Dana (2025)	Analisis Sentimen Terhadap Pengguna Aplikasi Jamsostek Mobile pada Google Play Store Menggunakan Metode Naive Bayes	Penelitian tersebut mengkaji sentimen pengguna aplikasi JMO berdasarkan ulasan yang tersedia di Google Play Store dengan menerapkan algoritma Naive Bayes. Tahapan penelitian diawali dengan pengumpulan data ulasan, kemudian dilakukan preprocessing yang meliputi pembersihan teks, penghapusan stopword, stemming, serta pembobotan kata menggunakan metode TF-IDF sebelum proses klasifikasi dilakukan. Berdasarkan hasil pengujian, model memperoleh akurasi sebesar 86,06%. Distribusi sentimen menunjukkan bahwa ulasan positif lebih mendominasi (74,1%) dibandingkan ulasan negatif (25,9%).
3.	Tia Septiani Gumilar, Rini Astuti, Yudhistira Arie Wijaya (2024)	Analisis Sentimen Ulasan Aplikasi Lita di Play Store Menggunakan Algoritma Naive Bayes	Penelitian ini menganalisis sentimen ulasan pengguna aplikasi Lita di Google Play Store menggunakan algoritma Naive Bayes. Data yang digunakan sebanyak 1000 ulasan yang diproses melalui tahapan preprocessing seperti cleaning, Tokenizing, dan stemming, serta pembobotan kata menggunakan metode TF-IDF. Hasil penelitian menunjukkan bahwa sebagian besar ulasan pengguna bersentimen positif dengan tingkat akurasi model mencapai 95%.

Tabel 2.1 Penelitian Terdahulu (Lanjutan)

NO	Nama Peneliti	Judul Penelitian	Pembahasan
(1)	(2)	(3)	(4)
4.	Wahyudi, Rudi Kurniawan, Yudhistira Arie Wijaya (2024)	Analisis <i>Sentimen</i> Pengguna Terhadap Aplikasi Blu BCA di Playstore Menggunakan Algoritma <i>Naïve Bayes</i>	Penelitian ini menganalisis sentimen pengguna aplikasi Blu BCA berdasarkan ulasan yang tersedia di Google Play Store dengan menerapkan algoritma Naive Bayes. Data ulasan diproses melalui beberapa tahapan, yaitu web scraping, preprocessing, pembobotan kata menggunakan metode TF-IDF, hingga proses klasifikasi sentimen. Pengujian model menghasilkan akurasi sebesar 85,31%, dengan nilai precision 84,55%, recall 93,69%, dan F1-score 88,88%. Hasil tersebut menunjukkan bahwa algoritma Naive Bayes mampu memberikan performa yang baik dalam mengklasifikasikan sentimen pada ulasan pengguna aplikasi Blu BCA.
5.	Gery Riansyah Ramadhan, Castaka Agus Sugianto (2024)	Ulasan Aplikasi DANA di Google Play Store Menggunakan Algoritma <i>Naïve Bayes</i>	Penelitian ini menganalisis <i>sentimen</i> ulasan pengguna aplikasi DANA di Google Play Store menggunakan algoritma <i>Naïve Bayes</i> . Data yang digunakan sebanyak 2000 ulasan berbahasa Indonesia yang diperoleh melalui <i>web scraping</i> . Proses penelitian meliputi tahapan <i>preprocessing</i> , pembobotan kata menggunakan <i>TF-IDF</i> , klasifikasi, dan evaluasi menggunakan <i>Confusion Matrix</i> . Hasil penelitian menunjukkan model memiliki performa yang baik dengan akurasi 86%, <i>precision</i> 86%, <i>recall</i> 98%, dan <i>F1-score</i> 92%.

Tabel 2.1 Penelitian Terdahulu (Lanjutan)

NO (1)	Nama Peneliti (2)	Judul Penelitian (3)	Pembahasan (4)
6.	Rema Sekar Nuraini, Dwi Krisbiantoro, Mohammad Imron, Pungkas Subarkah (2025)	Analisis Sentimen Pengguna Aplikasi Agoda di Google Play Store Menggunakan Algoritma Naive Bayes	Penelitian ini menganalisis ulasan pengguna aplikasi Agoda di Google Play Store menggunakan algoritma Naive Bayes dengan 7.885 data hasil scraping. Proses analisis meliputi tahapan <i>preprocessing</i> , pelabelan menerapkan InSet Lexicon, serta pembobotan <i>TF-IDF</i> sebelum dilakukan klasifikasi. Hasil penelitian menunjukkan akurasi terbaik sebesar 91,9%, sehingga Naive Bayes dinilai efektif dalam mengklasifikasikan sentimen pengguna secara akurat.
7.	Suci Nurhikmah, Romi Ramadani, Gandung Triyono (2025)	Analisis Sentimen pada Ulasan Aplikasi Wondr di Play Store dengan Metode Naive Bayes	Penelitian ini mengkaji sentimen pengguna aplikasi Wondr melalui ulasan yang tersedia di Google Play Store. Proses klasifikasi dilakukan dengan menerapkan algoritma Naive Bayes yang diawali dengan tahapan <i>preprocessing</i> dan pembobotan kata sebelum data dianalisis. Hasil pengujian memperoleh akurasi sebesar 82%, meskipun kemampuan model dalam mengenali sentimen netral masih perlu ditingkatkan. Temuan tersebut menunjukkan bahwa analisis sentimen dapat dimanfaatkan untuk memberikan gambaran mengenai tanggapan pengguna terhadap layanan mobile banking Wondr.

Tabel 2.1 Penelitian Terdahulu (Lanjutan)

NO	Nama Peneliti	Judul Penelitian	Pembahasan
(1)	(2)	(3)	(4)
8.	Rachmawati Oktaria Mardiyanto, Kusri, dan Ferry Wahyu Wibowo (2023)	Analisis Sentimen Pengguna Aplikasi Bank Syariah Indonesia Menggunakan Algoritma Support Vector Machine (SVM)	Penelitian ini menganalisis ulasan pengguna aplikasi BSI Mobile yang diambil dari Google Play Store. Data yang digunakan berjumlah 13.568 ulasan yang telah melalui proses preprocessing seperti case folding, filtering, stopwords removal, tokenization, dan stemming. Metode yang digunakan adalah algoritma Support Vector Machine (SVM) dengan pembagian data 70% data training dan 30% data testing. Hasil pengujian menunjukkan bahwa metode SVM mampu mengklasifikasikan sentimen pengguna dengan nilai akurasi sebesar 85,87%.
9.	Hamzah Naufal Zuhdi, Budi Prasetyo (2025)	Analisis Sentimen pada Ulasan Aplikasi iPusnas di Google Play Store Menggunakan Naive Bayes Classifier	Penelitian tersebut menganalisis 700 ulasan pengguna iPusnas di Google Play Store meliputi algoritma Naive Bayes. Data diproses melalui tahap preprocessing dan dibagi dengan rasio 80:20. Hasil penelitian menunjukkan akurasi sebesar 58% dengan dominasi sentimen positif sebesar 75,1%, sehingga metode ini dapat dimanfaatkan untuk analisis sentimen dan evaluasi kualitas layanan.

Tabel 2.1 Penelitian Terdahulu (Lanjutan)

NO	Nama Peneliti	Judul Penelitian	Pembahasan
(1)	(2)	(3)	(4)
10.	Elvin Alan Pratama, Aji Primajaya, Iqbal Maulana (2024)	Penerapan Algoritma <i>Support Vector Machine</i> Terhadap Analisis Sentimen Pada Aplikasi LinkAja	Penelitian ini menerapkan algoritma <i>Support Vector Machine</i> (SVM) untuk menganalisis sentimen ulasan pengguna aplikasi LinkAja di Google Play Store. Dataset diperoleh melalui proses scraping sebanyak 2.000 ulasan. Proses penelitian meliputi tahapan Knowledge Discovery in Database (KDD) meliputi data Selection, preprocessing, transformation, data mining, evaluation, dan knowledge. Hasil penelitian menunjukkan model terbaik diperoleh pada pembagian data 70:30 dengan nilai accuracy 88%, precision 90%, recall 83%, dan F1-score 87%.

Penelitian oleh [10] yang berjudul Analisis Sentimen ulasan pengguna aplikasi Shopee pada Google Play Store menggunakan algoritma Naive Bayes berbasis Python. Penelitian ini bertujuan untuk mengeksplorasi opini pengguna terhadap aplikasi Shopee melalui Analisis Sentimen dari sekitar 1.300 ulasan yang dikumpulkan menggunakan teknik *web scraping*. Data yang diperoleh kemudian melalui tahapan *preprocessing*, meliputi pembersihan teks, *tokenisasi*, dan penghapusan *stopwords* untuk meningkatkan kualitas data sebelum dilakukan proses klasifikasi *sentimen*.

Berdasarkan hasil pengujian, algoritma Naive Bayes memperoleh tingkat akurasi sekitar 86%, yang menunjukkan kemampuan model dalam mengklasifikasikan sentimen ulasan pengguna dengan baik. Distribusi hasil klasifikasi memperlihatkan bahwa sentimen negatif lebih banyak dibandingkan sentimen positif. Kondisi tersebut mengindikasikan masih terdapat berbagai keluhan pengguna terhadap layanan aplikasi yang menjadi objek penelitian. Penggunaan data ulasan secara real-time memberikan gambaran yang lebih aktual mengenai persepsi pengguna, meskipun penelitian ini hanya berfokus pada satu platform dan belum membandingkan performa Naive Bayes dengan algoritma klasifikasi lainnya.

Penelitian selanjutnya oleh [11] bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi *Jamsostek Mobile* (JMO) pada *Google Play Store* menggunakan algoritma *Naive Bayes*. Data diperoleh melalui teknik *web scraping* pada periode September hingga Oktober 2024, yang kemudian diproses melalui tahapan *Knowledge Discovery in Database* (KDD), meliputi seleksi data, *preprocessing* seperti *text cleaning*, penghapusan *stopwords*, normalisasi, dan *stemming*, serta transformasi data menggunakan metode *TF-IDF*. Selanjutnya, data diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral dengan pembagian data latih dan data uji sebesar 80:20 serta menggunakan teknik *cross-validation* untuk pengujian model.

Berdasarkan hasil klasifikasi, sentimen positif mendominasi dengan persentase 74,1%, sedangkan sentimen negatif sebesar 25,9%, sementara jumlah sentimen netral relatif sedikit. Model Naive Bayes menghasilkan akurasi sebesar 86,06% dengan nilai *precision*, *recall*, dan *F1-score* yang baik pada kategori

sentimen positif dan negatif. Namun, performa klasifikasi pada sentimen netral masih belum optimal karena jumlah data pada kategori tersebut jauh lebih sedikit dibandingkan kategori lainnya. Hasil tersebut menunjukkan bahwa algoritma Naive Bayes mampu memberikan performa yang baik dalam mengklasifikasikan sentimen pengguna, meskipun ketidakseimbangan distribusi data masih memengaruhi kemampuan model dalam mengenali kelas sentimen netral.

Penelitian selanjutnya oleh [12] bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Lita di Google Play Store menggunakan algoritma Naive Bayes. Data yang digunakan sebanyak 1000 ulasan yang diambil secara acak, kemudian diproses melalui tahapan *preprocessing* yang meliputi *cleaning*, *Tokenizing*, dan *stemming* untuk menghilangkan kata-kata yang tidak relevan. Selanjutnya, dilakukan pembobotan menggunakan metode *TF-IDF* untuk mengubah data teks menjadi representasi numerik sebelum dilakukan proses klasifikasi ke dalam dua kategori sentimen, yaitu positif dan negatif. Proses pengujian model dilakukan dengan pembagian data latih dan data uji dengan berbagai skenario, seperti 80:20, untuk mengukur performa model.

Berdasarkan hasil pengujian, algoritma Naive Bayes memperoleh tingkat akurasi sebesar 95%, dengan nilai precision dan recall yang juga mencapai 95%. Hasil klasifikasi menunjukkan bahwa mayoritas ulasan pengguna aplikasi Lita termasuk dalam kategori sentimen positif, yang mengindikasikan bahwa pengguna memberikan penilaian yang baik terhadap kualitas aplikasi. Meskipun demikian, proses klasifikasi hanya dilakukan pada dua kategori sentimen, yaitu positif dan negatif, sehingga hasil yang diperoleh masih dapat dikembangkan melalui

penggunaan kategori sentimen yang lebih beragam maupun penerapan dataset yang lebih luas.

Penelitian selanjutnya oleh [13] yang berjudul “Analisis Sentimen Pengguna terhadap Aplikasi Blu BCA di Play Store Menggunakan Algoritma Naive Bayes”. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Blu BCA serta mengidentifikasi persepsi pengguna terhadap aspek kinerja, keamanan, dan privasi aplikasi. Metode yang digunakan adalah Analisis Sentimen dengan algoritma Naive Bayes, dengan tahapan pengumpulan data melalui teknik web scraping dari Google Play Store, kemudian dilakukan preprocessing data yang meliputi data cleaning, case folding, stopword removal, Tokenizing, dan stemming, serta pembobotan menggunakan metode TF-IDF. Selanjutnya, data dibagi menjadi data latih dan data uji dengan proporsi 80:20 dan dievaluasi menggunakan Confusion Matrix untuk mengukur performa model.

Berdasarkan hasil pengujian, algoritma Naive Bayes memperoleh nilai akurasi sebesar 85,31% dengan nilai precision, recall, dan F1-score yang menunjukkan kinerja klasifikasi yang baik. Hasil analisis juga memperlihatkan bahwa sebagian besar ulasan pengguna memiliki sentimen positif. Meskipun demikian, masih ditemukan beberapa ulasan negatif yang berkaitan dengan performa aplikasi, kendala saat login, serta aspek keamanan. Temuan tersebut menunjukkan bahwa algoritma Naive Bayes mampu mengidentifikasi kecenderungan sentimen pengguna secara efektif dan menghasilkan informasi yang dapat dimanfaatkan sebagai bahan evaluasi dalam pengembangan layanan aplikasi perbankan digital.

Penelitian selanjutnya oleh [14] berjudul “Analisis Sentimen Ulasan Aplikasi DANA di Google Play Store Menggunakan Algoritma Naïve Bayes” yang bertujuan untuk menganalisis Sentimen pengguna serta mengetahui kecenderungan opini terhadap penggunaan aplikasi dompet digital DANA. Metode penelitian yang digunakan adalah Knowledge Discovery from Data (KDD) yang meliputi tahapan seleksi data, preprocessing, transformasi, data mining, dan evaluasi. Data dikumpulkan melalui teknik web scraping dari Google Play Store sebanyak 2.000 ulasan berbahasa Indonesia, kemudian dilakukan preprocessing yang mencakup cleaning, case folding, tokenisasi, stopwords removal, dan stemming. Selanjutnya dilakukan pembobotan kata menggunakan TF-IDF serta proses klasifikasi menggunakan algoritma Naïve Bayes.

Berdasarkan hasil klasifikasi, sentimen positif menjadi kategori yang paling dominan dengan jumlah 1.476 ulasan, sedangkan sentimen negatif tercatat sebanyak 443 ulasan dan sentimen netral sebanyak 81 ulasan. Model klasifikasi menggunakan algoritma Naïve Bayes menghasilkan nilai akurasi 86%, dengan precision 86, recall 98%, dan F1-score 92%. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan sentimen pengguna. Ulasan yang termasuk dalam kategori negatif umumnya berkaitan dengan kendala saat login, kegagalan transaksi, serta permasalahan pada proses pengisian saldo, sehingga aspek-aspek tersebut menjadi masukan yang dapat dipertimbangkan dalam pengembangan dan peningkatan kualitas layanan aplikasi.

Penelitian selanjutnya oleh [15] berjudul “Analisis Sentimen Pengguna Aplikasi Agoda di Google Play Store Menggunakan Algoritma Naïve Bayes” yang bertujuan untuk menganalisis sentimen pengguna serta mengevaluasi

97 kualitas layanan aplikasi Agoda berdasarkan ulasan pengguna. Metode penelitian
yang digunakan meliputi tahapan pengumpulan data, *preprocessing*, pelabelan
13 data menggunakan pendekatan lexicon-based, pembobotan *TF-IDF*, pembagian
139 data latih dan data uji, serta proses klasifikasi menggunakan algoritma *Naïve Bayes*. Data
dikumpulkan melalui teknik *web scraping* dari *Google Play Store*
2 sebanyak 7.885 ulasan berbahasa Indonesia, kemudian dilakukan *preprocessing*
yang mencakup *cleaning*, *case folding*, *Tokenizing*, *normalization*, *stopword*
removal, dan *stemming* untuk meningkatkan kualitas data sebelum proses
klasifikasi dilakukan.

2 Hasil penelitian menunjukkan bahwa *sentimen* negatif lebih mendominasi
dibandingkan *sentimen* positif, dengan jumlah 4.733 ulasan negatif dan 2.962
ulasan positif. Model yang dibangun mampu mencapai akurasi terbaik sebesar
104 91,9% pada skenario pembagian data tertentu, yang menunjukkan bahwa
24 algoritma *Naïve Bayes* efektif dalam mengklasifikasikan *sentimen* pengguna.
22 Selain itu, ditemukan bahwa keluhan pengguna umumnya berkaitan dengan
masalah pembatalan, pengembalian dana, proses pemesanan, serta ketidakpuasan
terhadap harga dan layanan. Penelitian ini menegaskan bahwa Analisis *Sentimen*
99 dapat memberikan wawasan penting bagi pengembang dalam meningkatkan
189 kualitas layanan aplikasi serta dapat dijadikan acuan untuk penelitian selanjutnya
dalam pengembangan metode analisis *sentimen*.

41 Penelitian selanjutnya oleh [16] berjudul “Analisis *Sentimen* pada Ulasan
10 Aplikasi Wondr di *Play Store* dengan Metode *Naïve Bayes*” yang bertujuan untuk
menganalisis persepsi serta opini pengguna terhadap aplikasi mobile banking
109 Wondr sebagai layanan digital dari Bank Negara Indonesia (BNI). Metode

penelitian yang digunakan adalah kuantitatif dengan pendekatan Analisis Sentimen menggunakan algoritma Naive Bayes Classifier yang diimplementasikan melalui bahasa pemrograman Python. Data penelitian dikumpulkan melalui teknik web scraping dari Google Play Store menggunakan Google Colab sebanyak 27.124 ulasan pengguna, kemudian dilakukan tahapan preprocessing yang meliputi cleaning, tokenization, case folding, stopword removal, filtering token by length, dan stemming. Selanjutnya data diberi label ke dalam tiga kategori sentimen (positif, negatif, dan netral), dilakukan pembobotan menggunakan TF-IDF, serta evaluasi model menggunakan Confusion Matrix dengan metrik akurasi, precision, recall, dan F1-score.

Berdasarkan hasil klasifikasi, sentimen positif menjadi kategori yang paling dominan dengan persentase 52,9%, diikuti sentimen negatif sebesar 39,9%, sedangkan sentimen netral hanya mencapai 7,2%. Model Naive Bayes menghasilkan tingkat akurasi sebesar 82%, yang menunjukkan kemampuan yang cukup baik dalam mengelompokkan sentimen pengguna berdasarkan ulasan di Google Play Store. Meskipun demikian, kemampuan model dalam mengenali sentimen netral masih belum optimal karena jumlah data pada kategori tersebut relatif lebih sedikit dibandingkan kategori lainnya. Hasil analisis juga menunjukkan bahwa sebagian besar pengguna memberikan tanggapan positif terhadap aplikasi Wondr, namun masih terdapat beberapa ulasan yang menyoroti kendala pada sistem, pengalaman penggunaan, serta fitur aplikasi yang memerlukan penyempurnaan.

Penelitian selanjutnya oleh [17] berjudul “Analisis Sentimen Pengguna Aplikasi Bank Syariah Indonesia dengan Menggunakan Algoritma Support Vector

Machine (SVM)” yang bertujuan untuk menganalisis sentimen serta tingkat kepuasan pengguna terhadap aplikasi BSI Mobile di Google Play Store. Metode penelitian yang digunakan adalah kuantitatif dengan pendekatan Analisis Sentimen menggunakan algoritma Support Vector Machine (SVM). Data penelitian dikumpulkan melalui teknik web scraping dari Google Play Store pada periode 1 Februari 2021 hingga 10 Agustus 2022, kemudian dilakukan pelabelan secara manual oleh lima ahli bahasa sehingga diperoleh sebanyak 13.568 data ulasan. Data tersebut selanjutnya melalui tahapan preprocessing yang meliputi case folding, filtering, stopword removal, tokenization, dan stemming, serta dibagi menjadi data training sebesar 70% dan data testing sebesar 30%.

Berdasarkan hasil pengujian, model Support Vector Machine (SVM) memperoleh nilai akurasi sebesar 85,87%, dengan precision sebesar 87,309% dan recall sebesar 86,958%. Hasil klasifikasi menunjukkan bahwa sentimen positif lebih banyak dibandingkan sentimen negatif, yang mengindikasikan bahwa mayoritas pengguna memberikan penilaian yang baik terhadap aplikasi BSI Mobile. Meskipun demikian, masih ditemukan sejumlah ulasan yang mengungkapkan adanya gangguan sistem serta penurunan performa aplikasi. Temuan tersebut menunjukkan bahwa metode SVM memiliki kemampuan yang baik dalam mengklasifikasikan sentimen pengguna sekaligus memberikan informasi mengenai aspek layanan yang masih perlu ditingkatkan.

Penelitian selanjutnya oleh [18] dengan judul “Analisis Sentimen pada Ulasan Aplikasi iPusnas di Google Play Store Menggunakan Naive Bayes Classifier.” Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan sentimen ulasan pengguna aplikasi iPusnas yang diperoleh dari Google Play Store

menggunakan algoritma *Naive Bayes*. Data yang digunakan sebanyak 700 ulasan pengguna yang telah melalui tahapan *preprocessing*, seperti *case folding*, *tokenization*, dan *stemming*, kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20.

Model yang dibangun selanjutnya dievaluasi menggunakan *Confusion Matrix* untuk mengukur kinerja klasifikasi. Hasil penelitian menunjukkan bahwa sebagian besar ulasan pengguna memiliki *sentimen* positif sebesar 75,1%, yang mengindikasikan tingkat kepuasan pengguna yang cukup tinggi. Selain itu, algoritma *Naive Bayes* menghasilkan akurasi sebesar 58% dengan nilai *precision* 69%, *recall* 81%, dan *F1-score* 75%. Penelitian ini menunjukkan bahwa metode *Naive Bayes* cukup efektif dalam mengklasifikasikan *sentimen* ulasan pengguna serta dapat dimanfaatkan sebagai bahan evaluasi dalam meningkatkan kualitas layanan aplikasi iPusnas.

Penelitian selanjutnya oleh [19] berjudul “Penerapan Algoritma *Support Vector Machine* terhadap Analisis *Sentimen* pada Ulasan Aplikasi LinkAja” yang bertujuan untuk menganalisis *sentimen* pengguna serta menginterpretasikan hasilnya sebagai rekomendasi dalam meningkatkan kualitas layanan aplikasi LinkAja. Metode penelitian yang digunakan adalah Knowledge Discovery in Database (KDD) dalam text mining, yang meliputi tahapan data *Selection*, *preprocessing*, transformation, data mining, evaluation, dan knowledge. Data dikumpulkan melalui teknik *web scraping* dari Google Play Store sebanyak 2.000 ulasan pengguna, kemudian dilakukan *preprocessing* yang mencakup *cleaning*, *case folding*, *Tokenizing*, *normalization*, *stopword removal*, dan *stemming*. Selanjutnya, dilakukan pelabelan data menggunakan *lexicon InSet*, serta

pembobotan menggunakan *TF-IDF* sebelum proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

Berdasarkan hasil klasifikasi, sentimen negatif lebih dominan dengan jumlah 1.056 ulasan, sedangkan sentimen positif tercatat sebanyak 944 ulasan. Performa terbaik model *Support Vector Machine* (SVM) diperoleh pada skenario pembagian data 70:30, dengan nilai akurasi 88%, precision 90%, recall 83%, dan F1-score 87%. Analisis terhadap ulasan pengguna juga menunjukkan bahwa sebagian besar keluhan berkaitan dengan kurang optimalnya layanan pelanggan, aplikasi yang sering mengalami crash pada perangkat dengan spesifikasi rendah, kendala koneksi, serta tampilan antarmuka yang dinilai kurang nyaman digunakan. Hasil tersebut menunjukkan bahwa metode SVM mampu mengklasifikasikan sentimen pengguna dengan baik sekaligus mengidentifikasi aspek-aspek layanan yang masih memerlukan perbaikan.

2.6 Rangkuman

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi DOLA di Google *Play Store* menggunakan algoritma *Naive Bayes*. Data penelitian diperoleh melalui teknik *web scraping* sebanyak 5.000 ulasan pengguna, kemudian dilakukan proses data *Selection* dan data labeling menjadi tiga kategori sentimen, yaitu positif, netral, dan negatif, berdasarkan nilai rating pengguna.

Sebelum proses klasifikasi, data melalui tahapan *preprocessing* yang meliputi *case folding*, *cleaning*, *tokenization*, *stopword removal*, dan *stemming* untuk membersihkan serta menormalkan teks. Selanjutnya, data diubah ke dalam bentuk numerik menggunakan metode *TF-IDF*, kemudian dilakukan oversampling untuk

menyeimbangkan distribusi kelas sehingga model dapat belajar dengan lebih optimal.

Proses klasifikasi dilakukan menggunakan algoritma *Naive Bayes* dengan pembagian data 80% sebagai data latih dan 20% sebagai data uji, serta evaluasi menggunakan *k-fold cross-validation* untuk memperoleh hasil yang lebih stabil. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *Confusion Matrix* untuk mengukur kinerja klasifikasi.

Hasil penelitian menunjukkan bahwa model mampu mencapai akurasi sebesar 82,37%, yang menandakan bahwa algoritma *Naive Bayes* cukup efektif dalam mengklasifikasikan *sentimen* pengguna aplikasi DOLA. Nilai evaluasi juga menunjukkan kinerja yang baik, dengan *weighted average precision* sebesar 0,83, *recall* sebesar 0,82, dan *F1-score* sebesar 0,82. Selain itu, hasil *Confusion Matrix* memperlihatkan bahwa sebagian besar data berhasil diprediksi dengan benar, meskipun masih terdapat beberapa kesalahan klasifikasi antar kelas *sentimen*.

Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi *preprocessing*, *TF-IDF*, dan algoritma *Naive Bayes* mampu memberikan hasil klasifikasi yang cukup baik dalam Analisis *Sentimen* ulasan pengguna aplikasi DOLA. Hasil penelitian ini diharapkan dapat menjadi bahan evaluasi bagi pengembang untuk meningkatkan kualitas layanan aplikasi serta menjadi referensi bagi penelitian selanjutnya di bidang analisis *sentimen*.

BAB III

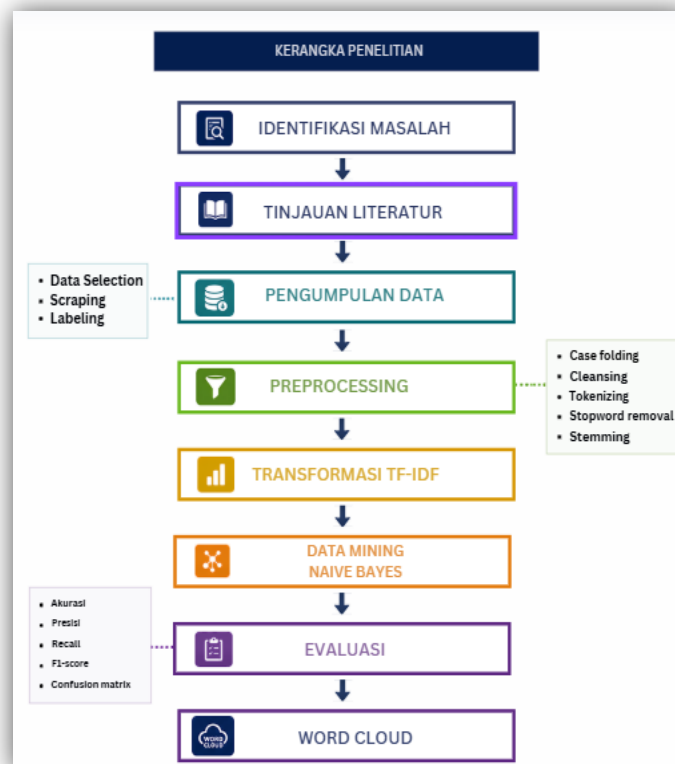
METODE PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian berjudul “Analisis Sentimen Pengguna terhadap Aplikasi Dola di Play Store Menggunakan Algoritma Naive Bayes.” Penelitian ini menggunakan pendekatan kuantitatif deskriptif, karena berfokus pada pengumpulan dan analisis data numerik untuk menggambarkan kecenderungan sentimen pengguna terhadap aplikasi Dola. Data yang digunakan berasal dari ulasan pengguna di Google Play Store, yang dikumpulkan melalui proses web scraping dan kemudian diolah menggunakan teknik text mining. Dalam proses pengembangannya, penelitian ini menggunakan tahapan metode Naive Bayes yang meliputi pengumpulan data, preprocessing (cleaning, Tokenizing, dan stopwords removal), pembobotan TF-IDF, pelabelan data, pelatihan model, dan evaluasi hasil. Setiap tahapan dilakukan secara berurutan untuk memastikan hasil analisis yang akurat dan terukur.

3.1. Kerangka Penelitian

Kerangka penelitian ini dimulai dari tahap identifikasi masalah untuk menentukan kebutuhan utama dalam menganalisis sentimen pengguna terhadap aplikasi Dola di Google Play Store. Selanjutnya dilakukan pengumpulan data berupa ulasan pengguna sebagai dasar untuk proses analisis. Data yang diperoleh kemudian melalui tahap preprocessing dan transformasi TF-IDF agar siap digunakan dalam pemodelan. Setelah itu, algoritma Naive Bayes diterapkan untuk melakukan klasifikasi sentimen positif, negatif, atau netral. Tahap berikutnya adalah evaluasi model untuk memastikan tingkat akurasi, presisi, recall, dan F1-

score yang dihasilkan sudah memadai. Seluruh rangkaian tahapan ini membentuk alur lengkap kerangka penelitian yang ditunjukkan pada gambar berikut :



Gambar 3. 1 Kerangka Penelitian

Kerangka penelitian ini menggambarkan alur sistematis yang digunakan untuk menganalisis sentimen pengguna terhadap aplikasi Dola di Play Store menggunakan algoritma Naive Bayes. Penelitian dimulai dari identifikasi masalah untuk menentukan fokus kajian, kemudian dilanjutkan dengan tinjauan literatur sebagai dasar teori dan metodologi. Tahap berikutnya adalah pengumpulan data, yaitu mengambil ulasan pengguna dari Google Play Store dan menyeleksi data agar hanya ulasan valid yang digunakan. Data yang diperoleh selanjutnya

dianalisis melalui beberapa tahap, dimulai dari *preprocessing* untuk membersihkan teks, pelabelan data, serta transformasi *TF-IDF* untuk mengubah teks menjadi bentuk numerik. Setelah itu, dilakukan pembangunan model *Naive Bayes* untuk mengklasifikasikan *sentimen* menjadi positif, negatif, atau netral. Tahap akhir adalah evaluasi, yaitu menilai performa model menggunakan metrik seperti akurasi dan *Confusion Matrix*. Seluruh tahapan tersebut dilakukan secara berurutan agar penelitian berjalan terarah dan menghasilkan Analisis *Sentimen* yang akurat.

3.2. Identifikasi Masalah

Tahap pertama adalah mengidentifikasi masalah utama yang akan diteliti. Pada penelitian ini, masalah yang ditemukan adalah kebutuhan untuk mengetahui bagaimana *sentimen* pengguna terhadap aplikasi Dola berdasarkan ulasan yang terdapat pada *Google Play Store*. Identifikasi masalah membantu peneliti memahami tujuan penelitian dan menentukan pendekatan yang tepat untuk menyelesaikannya.

3.3. Tinjauan Literatur

Tahap ini dilakukan untuk mengumpulkan berbagai teori, konsep, dan penelitian sebelumnya yang berkaitan dengan analisis *sentimen*, algoritma *Naive Bayes*, *TF-IDF*, serta teknik *preprocessing* teks. Tinjauan literatur berfungsi sebagai landasan teoritis dan metodologis dalam pelaksanaan penelitian sehingga analisis yang dilakukan memiliki dasar yang kuat.

3.4. Pengumpulan Data

Pada tahap ini, data ulasan pengguna aplikasi DOLA dikumpulkan dari *Google Play Store* menggunakan teknik web scraping. Data yang berhasil diperoleh

kemudian melalui proses data selection untuk menyaring ulasan yang relevan, valid, dan berbahasa Indonesia sehingga sesuai dengan kebutuhan penelitian. Selanjutnya, setiap ulasan diberi label sentimen berdasarkan nilai rating yang diberikan pengguna, yaitu sentimen positif, netral, dan negatif. Proses pelabelan tersebut bertujuan untuk menghasilkan data berlabel yang siap digunakan sebagai data pelatihan dan pengujian pada proses klasifikasi menggunakan algoritma Naive Bayes.

3.5. Preprocessing

Tahapan *preprocessing* meliputi *case folding*, *cleaning*, *Tokenizing*, *stopword removal*, dan *stemming* untuk meningkatkan kualitas data teks sebelum proses klasifikasi [20].

- 1) Case Folding: Mengubah seluruh huruf pada teks menjadi huruf kecil agar penulisan kata menjadi seragam.
- 2) Cleaning: Menghapus karakter yang tidak diperlukan, seperti angka, simbol, tanda baca, emoji, dan tautan (*URL*), sehingga data menjadi lebih bersih.
- 3) Tokenizing: Memisahkan setiap kalimat menjadi kumpulan kata (*token*) untuk memudahkan proses analisis.
- 4) Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna penting menggunakan daftar *stopword* bahasa Indonesia sehingga hanya kata yang relevan yang dipertahankan.
- 5) Stemming: Mengubah kata berimbuhan menjadi bentuk kata dasar agar variasi kata menjadi lebih konsisten.

Tahapan *preprocessing* tersebut dilakukan untuk meningkatkan kualitas data teks sebelum proses pembobotan TF-IDF dan klasifikasi menggunakan algoritma Naive Bayes.

3.6. Transformasi *TF-IDF*

Setelah seluruh tahapan *preprocessing* selesai dilakukan, setiap dokumen direpresentasikan ke dalam bentuk numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Pembobotan ini memberikan nilai pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen serta tingkat kepentingannya terhadap keseluruhan kumpulan dokumen sehingga kata yang lebih informatif memperoleh bobot yang lebih tinggi[21].

3.7. Data Mining Algoritma *Naive Bayes*

Dataset yang telah direpresentasikan menggunakan TF-IDF selanjutnya digunakan sebagai masukan bagi algoritma Naive Bayes. Model dilatih menggunakan data latih untuk mempelajari pola hubungan antara fitur dan kategori sentimen, kemudian digunakan untuk memprediksi sentimen pada data uji berdasarkan probabilitas setiap kelas

3.8. Evaluasi

Model yang telah dibangun kemudian dievaluasi untuk mengetahui tingkat akurasinya. Evaluasi dilakukan menggunakan beberapa metrik, yaitu:

- 1) Akurasi: tingkat ketepatan prediksi model.
- 2) Presisi: ketepatan prediksi pada satu kelas tertentu.
- 3) *Recall*: kemampuan model menangkap data positif.
- 4) *F1-score*: keseimbangan presisi dan *recall*.
- 5) *Confusion Matrix*: tabel visual prediksi benar dan salah.

Perhitungan metrik evaluasi dilakukan berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) yang diperoleh dari *Confusion Matrix*. Adapun rumus yang digunakan adalah sebagai berikut:

a. Akurasi : $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots (1)$

Keterangan:

- TP (*True Positive*) = data positif yang diprediksi positif
- TN (*True Negative*) = data negatif yang diprediksi negatif
- FP (*False Positive*) = data negatif yang diprediksi positif
- FN (*False Negative*) = data positif yang diprediksi negatif

Nilai akurasi dihitung menggunakan persamaan (1) untuk mengetahui tingkat ketepatan model dalam mengklasifikasikan data secara keseluruhan berdasarkan perbandingan jumlah prediksi yang benar terhadap total seluruh data pengujian.

b. Presisi : $Precision = \frac{TP}{TP+FP} \times 100\% \dots (2)$

Keterangan:

- TP = *True Positive*
- FP = *False Positive*

Nilai presisi dihitung menggunakan persamaan (2) untuk mengetahui tingkat ketepatan model dalam memberikan prediksi positif berdasarkan perbandingan jumlah prediksi positif yang benar terhadap total seluruh prediksi positif.

c. Recall : $Recall = \frac{TP}{TP+FN} \times 100\% \dots (3)$

Keterangan:

- TP = *True Positive*
- FN = *False Negative*

59 Nilai *recall* dihitung menggunakan persamaan (3) untuk mengetahui tingkat keberhasilan model dalam menemukan seluruh data positif yang sebenarnya berdasarkan perbandingan jumlah prediksi positif yang benar terhadap total data positif sebenarnya.

d. *F1-score* :
$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots (4)$$

Keterangan:

- 32 • *Precision* = tingkat ketepatan prediksi positif.
- *Recall* = tingkat keberhasilan model dalam menemukan data positif.

9 24 Nilai *F1-score* dihitung menggunakan persamaan (4) untuk mengetahui keseimbangan antara nilai *precision* dan *recall* dalam mengevaluasi kinerja model klasifikasi.

Evaluasi ini bertujuan untuk menilai seberapa baik model dalam melakukan klasifikasi *sentimen* serta mengetahui kekuatan dan kelemahan algoritma yang digunakan.

43 Selain menggunakan pembagian data training dan testing, evaluasi model juga 29 dilakukan menggunakan metode *k-fold cross-validation* dengan nilai $k=5$ untuk memperoleh hasil evaluasi yang lebih stabil serta mengurangi potensi bias akibat pembagian data.

3.9. Word Cloud

77 Sebagai pendukung hasil analisis, dilakukan pembuatan visualisasi *Word Cloud* untuk menggambarkan kata-kata yang paling sering muncul dalam ulasan pengguna. *Word Cloud* membantu menampilkan pola umum atau topik yang 1 sering dibahas pengguna, sehingga dapat menjadi informasi tambahan dalam memahami persepsi publik terhadap aplikasi Dola.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini membahas hasil penelitian dan pembahasan mengenai Analisis Sentimen pengguna terhadap aplikasi DOLA yang terdapat pada Google Play Store dengan menggunakan algoritma Naive Bayes. Data yang digunakan dalam penelitian ini berupa ulasan atau komentar yang diberikan oleh pengguna aplikasi DOLA. Ulasan tersebut dikumpulkan kemudian diolah melalui beberapa tahapan *preprocessing* seperti *case folding*, *Tokenizing*, *stopword removal*, dan *stemming* agar data teks dapat diproses dengan baik dalam proses analisis.

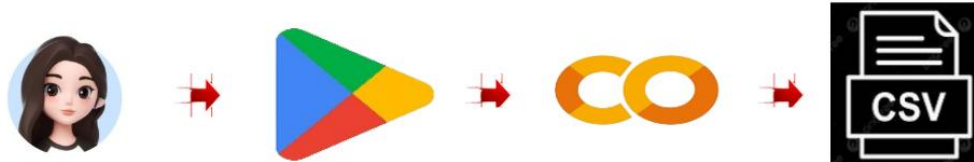
Selanjutnya, data yang telah diproses digunakan dalam proses klasifikasi *sentimen* menggunakan algoritma Naive Bayes untuk mengelompokkan ulasan pengguna ke dalam kategori *sentimen* positif, negatif, dan netral. Hasil dari proses klasifikasi tersebut kemudian dianalisis untuk mengetahui kecenderungan *sentimen* pengguna terhadap aplikasi DOLA serta dilakukan evaluasi model menggunakan beberapa metrik seperti *accuracy*, *precision*, *recall*, dan *Confusion Matrix* untuk mengetahui tingkat kinerja model yang digunakan.

4.1. Hasil

Hasil penelitian ini diperoleh melalui beberapa tahapan, yaitu pengumpulan data, *preprocessing*, transformasi data menggunakan *TF-IDF*, serta klasifikasi menggunakan algoritma Naive Bayes. Data yang digunakan berupa ulasan pengguna aplikasi DOLA pada Google Play Store yang kemudian diklasifikasikan ke dalam tiga kategori *sentimen*, yaitu positif, negatif, dan netral.

4.1.1 Data Collection

Data pada penelitian ini diperoleh menggunakan teknik *web scraping* dari *Google Play Store*. Proses ini dilakukan untuk mengumpulkan ulasan pengguna aplikasi DOLA secara otomatis sehingga dapat digunakan sebagai data penelitian.



Gambar 4. 1 Proses Pengumpulan Data

Gambar 4.1 memperlihatkan proses pengambilan data ulasan pengguna aplikasi DOLA menggunakan library *google-play-scraper* pada Google Colab. Proses tersebut berhasil memperoleh data ulasan yang selanjutnya digunakan sebagai dataset penelitian. Data yang telah dikumpulkan kemudian disimpan dalam format CSV sehingga lebih mudah diproses pada tahapan berikutnya.

Hasil dari proses ini menunjukkan bahwa data ulasan pengguna berhasil dikumpulkan dengan baik dan siap digunakan dalam tahap *preprocessing* serta analisis *sentimen*. Dengan demikian, tahap pengumpulan data telah berhasil menyediakan dataset yang relevan dan terstruktur sebagai dasar dalam proses Analisis *Sentimen* pengguna aplikasi DOLA.

a) Data Collection

Proses pengambilan data ulasan pengguna dilakukan menggunakan library *google-play-scraper* dengan bahasa pemrograman Python.:

```

from google_play_scraper import reviews, Sort

result, continuation_token = reviews(
    'com.larus.wolf',
    lang='id',
    country='id',
    sort=Sort.NEWEST,
    count=5000
)
  
```



```

print('Jumlah ulasan yang berhasil diambil:', len(result))
  
```

```

Jumlah ulasan yang berhasil diambil: 5000
  
```

Gambar 4. 2 Kode Program *Data Collection*

Pada Gambar 4.2, Hasil yang diperoleh menunjukkan bahwa sistem berhasil mengumpulkan 5000 data ulasan pengguna yang siap digunakan pada tahap analisis selanjutnya. Dengan demikian, proses pengumpulan data berjalan dengan baik dan menghasilkan dataset yang cukup untuk analisis *sentimen*.

b) *Data Selection*

Pada tahap data *Selection*, dilakukan pemilihan atribut yang relevan dari hasil pengumpulan data. Data hasil scraping diubah ke dalam bentuk *DataFrame* menggunakan library *pandas*, kemudian dipilih atribut yang digunakan dalam penelitian yaitu *userName*, *score*, *at*, dan *content*. Atribut tersebut selanjutnya diubah menjadi *username*, *rating*, *tanggal_ulasan*, dan *review* untuk mempermudah proses analisis. Berikut merupakan kode program yang digunakan:

```
data = pd.DataFrame(result)
data = data[['userName', 'score', 'at', 'content']]
data.columns = ['username', 'rating', 'tanggal_ulasan', 'review']
print(f"Jumlah baris di DataFrame setelah pembuatan: {len(data)}")
display(data.head())
```

```
Jumlah baris di DataFrame setelah pembuatan: 5000
```

	username	rating	tanggal_ulasan	review
0	Pengguna Google	5	2026-05-02 13:36:10	sangat membantu
1	Pengguna Google	5	2026-05-02 13:32:58	gacor Wak
2	Pengguna Google	5	2026-05-02 13:32:51	terbaik
3	Pengguna Google	1	2026-05-02 13:31:17	jangan pakai dola dia itu ngk bener

Gambar 4. 3 Kode Program *Data Selection*

Pada Gambar 4.3, Memperlihatkan hasil tersebut menunjukkan bahwa data telah terorganisir dengan baik, dengan atribut yang sesuai kebutuhan penelitian, yaitu *username*, *rating*, *tanggal_ulasan*, dan *review*. Dengan demikian, proses data

Selection berhasil menunjukan dataset yang terstruktur dan siap digunakan pada tahap analisis selanjutnya.

c) **Data Labeling**

Pada tahap ini dilakukan proses pelabelan *sentimen* terhadap data ulasan berdasarkan nilai rating yang diberikan pengguna.

Rating 4–5 : Dikategorikan sebagai *Sentimen* Positif

Rating 3 : Dikategorikan sebagai *Sentimen* Netral

Rating 1–2 : Dikategorikan sebagai *Sentimen* Negatif[22] .

```
def label_sentimen(rating):  
    if rating >= 4:  
        return "positif"  
    elif rating == 3:  
        return "netral"  
    else:  
        return "negatif"  
  
data['sentimen'] = data['rating'].apply(label_sentimen)  
  
display(data.head())
```

	username	rating	tanggal_ulasan	review	sentimen
0	Pengguna Google	5	2026-05-02 13:36:10	sangat membantu	positif
1	Pengguna Google	5	2026-05-02 13:32:58	gacor Wak	positif
2	Pengguna Google	5	2026-05-02 13:32:51	terbaik	positif
3	Pengguna Google	1	2026-05-02 13:31:17	jangan pakai dola dia itu ngk bener	negatif
4	Pengguna Google	5	2026-05-02 13:23:59	dola ini membantu aku banget 🥰	positif

Gambar 4. 4 Kode Program Data Labeling

Pada Gambar 4.4 memperlihatkan hasil tersebut menunjukkan bahwa data ulasan telah berhasil diklasifikasikan ke dalam kategori *sentimen* positif, negatif, dan netral. Dengan demikian, tahap data labeling berhasil mempersiapkan data yang siap digunakan dalam proses klasifikasi menggunakan algoritma *Naive Bayes*.

4.1.2 **Preprocessing Data**

Sebelum memasuki tahap analisis, data ulasan terlebih dahulu melalui proses preprocessing untuk menghasilkan data teks yang lebih bersih dan

konsisten. Tahapan ini dilakukan agar data siap digunakan pada proses pembobotan TF-IDF dan klasifikasi menggunakan algoritma Naive Bayes. Adapun tahapan preprocessing yang diterapkan meliputi case folding, cleaning, tokenization, stopwords removal, dan stemming.

1. Case folding

Case Folding dilakukan dengan mengubah seluruh huruf pada data ulasan menjadi huruf kecil (lowercase). Tahap ini bertujuan untuk menyeragamkan penulisan kata sehingga perbedaan penggunaan huruf kapital tidak memengaruhi proses pengolahan dan analisis data pada tahapan berikutnya.

```
def case_folding(text):  
    return text.lower()  
  
data['case_fold'] = data['review'].apply(case_folding)  
  
display(data.head())
```

Gambar 4. 5 Kode Program Case Folding

Pada Gambar 4.5, Proses ini berhasil menunjukkan bahwa data teks menjadi lebih konsisten dan siap digunakan pada tahap preprocessing selanjutnya. Dengan demikian, proses Case folding berhasil menormalkan data teks sehingga dapat meningkatkan kualitas analisis sentimen.

2. Cleaning

Cleaning dilakukan dengan menghapus karakter yang tidak diperlukan, seperti angka, simbol, tanda baca, dan karakter lainnya yang tidak berpengaruh terhadap proses analisis sentimen. Setelah tahap ini, data ulasan menjadi lebih bersih sehingga siap digunakan pada proses berikutnya.

```
import string  
import re  
  
def cleaning(text):  
    text = text.lower()  
    text = re.sub(r'http\S+', '', text)  
    text = re.sub(r'\d+', '', text)  
    text = text.translate(str.maketrans('', '', string.punctuation))  
    return text.strip()  
  
data['clean'] = data['review'].apply(cleaning)
```

Gambar 4. 6 Kode Program *Cleaning*

Gambar 4.6 Memperlihatkan hasil proses *cleaning*, yaitu penghapusan karakter yang tidak diperlukan, seperti angka, simbol, tanda baca, dan karakter khusus lainnya. Setelah proses ini dilakukan, data ulasan menjadi lebih bersih sehingga siap digunakan pada tahap berikutnya.

3. Tokenization

Tokenization dilakukan dengan memisahkan setiap ulasan menjadi kumpulan kata (token). Proses ini menghasilkan token yang selanjutnya digunakan pada tahapan analisis berikutnya sehingga setiap kata dapat diproses secara terpisah oleh sistem.

```
nltk.download('punkt_tab')

data['token'] = data['clean'].apply(word_tokenize)

display(data.head())
```

[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt_tab.zip.

	username	rating	tanggal_ulasan	review	sentimen	case_fold	clean	token
0	Pengguna Google	5	2026-05-02 13:36:10	sangat membantu	positif	sangat membantu	sangat membantu	[sangat, membantu]
1	Pengguna Google	5	2026-05-02 13:32:58	gacor Wak	positif	gacor wak	gacor wak	[gacor, wak]
2	Pengguna Google	5	2026-05-02 13:32:51	terbaik	positif	terbaik	terbaik	[terbaik]
3	Pengguna Google	1	2026-05-02 13:31:17	jangan pakai dola dia itu ngk bener	negatif	jangan pakai dola dia itu ngk bener	jangan pakai dola dia itu ngk bener	[jangan, pakai, dola, dia, itu, ngk, bener]
4	Pengguna Google	5	2026-05-02 13:23:59	dola ini membantu aku banget 🥰	positif	dola ini membantu aku banget 🥰	dola ini membantu aku banget 🥰	[dola, ini, membantu, aku, banget, 🥰]

Gambar 4. 7 Kode Program *Tokenization*

Pada Gambar 4.7, Hasil dari proses ini menunjukkan bahwa teks ulasan menjadi lebih bersih dan terstruktur sehingga siap digunakan pada tahap selanjutnya. Dengan demikian, proses *cleaning* berhasil menghilangkan noise pada data dan meningkatkan kualitas data untuk analisis *sentimen*.

4. Stopword removal

Stopword removal dilakukan dengan menghapus kata-kata umum yang tidak memiliki pengaruh signifikan terhadap analisis sentimen. Setelah proses ini, hanya kata-kata yang memiliki informasi penting yang dipertahankan untuk digunakan pada tahap berikutnya.

```
stop_words = set(stopwords.words('indonesian'))

data['stopword'] = data['token'].apply(
    lambda x: [word for word in x if word not in stop_words]
)
```

Gambar 4. 8 Kode Program Stopword Removal

Pada Gambar 4.8, Memerlihatkan Hasil dari proses ini menunjukkan bahwa teks menjadi lebih fokus pada kata-kata yang memiliki makna penting. Dengan demikian, proses *stopword removal* membantu meningkatkan kualitas data untuk analisis *sentimen*.

5. Stemming

Stemming dilakukan dengan mengubah kata yang memiliki imbuhan menjadi bentuk kata dasar. Proses ini menghasilkan bentuk kata yang lebih konsisten sehingga variasi penulisan dapat dikurangi sebelum dilakukan proses klasifikasi.

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

def stemming(text):
    return [stemmer.stem(word) for word in text]

data['stem'] = data['stopword'].apply(stemming)

data['final'] = data['stem'].apply(lambda x: " ".join(x))
```

Pada Gambar 4.9, Tahap ini menunjukkan bahwa variasi kata menjadi lebih sederhana sehingga memudahkan proses analisis. Dengan demikian, proses *stemming* membantu meningkatkan konsistensi data dalam klasifikasi *sentimen*.

4.2 Pembagian data

Sebelum dilakukan proses klasifikasi menggunakan algoritma Naive Bayes, data hasil preprocessing dan pembobotan TF-IDF dibagi menjadi fitur (X) dan label (y). Fitur (X) merupakan representasi numerik dari teks ulasan yang dihasilkan melalui proses pembobotan TF-IDF, sedangkan label (y) berisi kategori sentimen yang terdiri atas positif, negatif, dan netral. Pembagian ini dilakukan sebagai tahap persiapan sebelum model dilatih dan dievaluasi.

Selanjutnya, dilakukan pembagian data menjadi data training dan data testing menggunakan metode `train_test_split` dari library *sklearn*. Pada penelitian ini digunakan beberapa skenario pembagian data, yaitu 80:20, 70:30, dan 60:40, untuk mengetahui pengaruh proporsi data terhadap performa model. Selain itu, parameter *stratify* digunakan untuk menjaga keseimbangan distribusi kelas pada data training dan testing.

Setelah proses pembagian data, dilakukan pelatihan model menggunakan algoritma *Naive Bayes*. Proses pelatihan dilakukan dengan fungsi:

- `model.fit(X_train, y_train)`

Pada tahap ini, model mempelajari pola hubungan antara fitur dan label *sentimen* berdasarkan probabilitas kemunculan kata pada masing-masing kelas.

Selanjutnya, model digunakan untuk melakukan prediksi terhadap data testing menggunakan fungsi:

- `model.predict(X_test)`

Hasil prediksi ini kemudian digunakan pada tahap evaluasi untuk mengukur performa model dalam mengklasifikasikan *sentimen*. Selain itu, evaluasi model juga dilakukan menggunakan metode *k-fold cross-validation* dengan nilai $k=5$ untuk memperoleh hasil evaluasi yang lebih stabil serta mengurangi potensi bias akibat pembagian data.

4.3 Transformasi Data (*TF-IDF*)

Data hasil preprocessing selanjutnya diproses menggunakan metode TF-IDF (Term Frequency–Inverse Document Frequency) untuk mengubah teks ulasan menjadi representasi numerik. Hasil pembobotan tersebut digunakan sebagai fitur masukan dalam proses klasifikasi sentimen menggunakan algoritma Naive Bayes.

1. *Term Frequency (TF)*

Mengukur seberapa sering suatu kata muncul dalam sebuah dokumen.

Semakin sering sebuah kata muncul, semakin tinggi nilai TF-nya.

2. *Inverse Document Frequency (IDF)*

Mengukur seberapa jarang sebuah kata muncul di seluruh dokumen. Kata-kata yang muncul di hampir semua dokumen (seperti “yang”, “adalah”, “dan”) akan memiliki nilai IDF yang rendah karena dianggap kurang penting.

3. Rumus *TF-IDF*

Pembobotan TF-IDF dilakukan menggunakan Persamaan (5) untuk menghasilkan nilai bobot pada setiap kata dalam data ulasan. Bobot yang diperoleh merepresentasikan tingkat kepentingan setiap kata terhadap keseluruhan

dokumen dan selanjutnya digunakan sebagai fitur masukan dalam proses klasifikasi sentimen menggunakan algoritma Naive Bayes.

$$TF-IDF(t,d) = TF(t,d) \times IDF(t) \dots (5)$$

Dimana :

- t = Term (kata)
- d = Dokumen
- $TF(t,d)$ = Frekuensi kata t dalam dokumen d
- $IDF(t) = \text{Log} (N/df_t)$

Dengan N = Total dokumen, dan df_t = Jumlah dokumen yang mengandung kata t

4. Pembobotan TF-IDF

Proses pembobotan TF-IDF menghasilkan nilai pada setiap kata dalam data ulasan yang mencerminkan tingkat kepentingannya terhadap keseluruhan dokumen. Nilai tersebut kemudian digunakan sebagai representasi numerik sebelum data diproses oleh algoritma Naive Bayes.

```
from sklearn.feature_extraction.text import TfidfVectorizer

data['final'] = data['final'].fillna('').astype(str)

tfidf = TfidfVectorizer(
    max_features=5000,
    ngram_range=(1, 2),
    min_df=3,
    max_df=0.9
)

X = tfidf.fit_transform(data['final'])
y = data['sentimen']
```

Gambar 4. 10 Kode Program *TF-IDF*

Pada Gambar 4.10, Hasil dari proses ini menghasilkan representasi data berupa matriks fitur (X) yang berisi bobot setiap kata dalam dokumen, serta label *sentimen* (y) sebagai target klasifikasi.

Pembobotan ini membuat setiap kata memiliki nilai yang mencerminkan tingkat kepentingannya dalam dokumen, sehingga dapat digunakan dalam proses klasifikasi menggunakan algoritma *Naive Bayes*. Dengan demikian, proses *TF-IDF* berhasil mengubah data teks menjadi data numerik yang sudah siap digunakan pada tahap pemodelan.

5. *Oversampling*

Pada tahap ini dilakukan penyeimbangan data (data balancing) menggunakan metode Random Oversampling untuk mengatasi ketidakseimbangan jumlah data pada setiap kelas *sentimen*. Metode ini bekerja dengan menduplikasi data pada kelas minoritas hingga jumlahnya seimbang dengan kelas mayoritas. Proses dilakukan menggunakan library *imbalanced-learn* melalui fungsi `fit_resample()`,

```
from imblearn.over_sampling import RandomOverSampler

ros = RandomOverSampler(random_state=42)

X, y = ros.fit_resample(X, y)
```

Gambar 4. 11 Kode Program *Oversampling*

Pada Gambar 4.11, Berdasarkan hasil tersebut menunjukkan bahwa data pada kelas minoritas telah ditambah sehingga distribusi data menjadi lebih seimbang.

Kondisi ini membuat model dapat belajar dengan lebih baik tanpa bias terhadap kelas tertentu. Proses oversampling membantu menyeimbangkan distribusi data sehingga model dapat belajar lebih optimal.

4.4 Klasifikasi Naive Bayes

Sebelum dilakukan proses klasifikasi menggunakan algoritma Naive Bayes, data yang telah melalui tahap preprocessing dan pembobotan TF-IDF terlebih dahulu dipersiapkan dalam bentuk fitur (X) dan label (y). Fitur (X) merupakan hasil transformasi TF-IDF yang berisi representasi numerik dari teks ulasan, sedangkan label (y) merupakan kategori sentimen yang telah diberikan pada tahap pelabelan, yaitu positif, negatif, dan netral.

Selanjutnya, data dipecah menjadi dua bagian, yaitu data training dan data testing. Data training digunakan untuk melatih model Naive Bayes agar dapat mengenali pola dari data, sedangkan data testing digunakan untuk menguji kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya.

1. Train_Test_Split

Proses pembagian data menggunakan metode train_test_split menghasilkan data training 80% dan data testing 20%. Selain itu, pembagian dilakukan dengan menjaga proporsi distribusi kelas agar tetap seimbang.

```
X_train, X_test, y_train, y_test = train_test_split(  
    X,  
    y,  
    test_size=0.2,  
    random_state=42,  
    stratify=y  
)
```

Pada Gambar 4.12, Berdasarkan hasil tersebut menghasilkan data latih

Gambar 4. 12 Kode Program *Train-Test Split*

proses pelatihan dan evaluasi model. Pembagian ini juga mempertahankan proporsi kelas pada setiap data sehingga distribusi tetap seimbang. Dengan demikian, data telah siap digunakan untuk proses pelatihan dan pengujian model klasifikasi.

2. Algoritma *Naïve Bayes*

Data hasil pembobotan TF-IDF selanjutnya digunakan sebagai masukan pada algoritma Naive Bayes. Model dibangun menggunakan data training untuk mempelajari pola dari setiap kategori sentimen, kemudian diuji menggunakan data testing guna menghasilkan prediksi sentimen pada setiap ulasan. Prediksi yang diperoleh selanjutnya digunakan sebagai dasar dalam proses evaluasi kinerja model menggunakan metrik yang telah ditentukan..

3. Pelatihan model dengan *Naïve Bayes*

Proses pelatihan model dilakukan menggunakan algoritma *Naive Bayes* berdasarkan data training yang telah dipersiapkan sebelumnya. Model dilatih menggunakan data fitur (*X_train*) dan label *sentimen* (*y_train*) untuk mengenali pola pada data. Selanjutnya, model digunakan untuk melakukan prediksi terhadap data testing menggunakan fungsi *predict()*.

Hasil dari proses ini menghasilkan prediksi *sentimen* pada data uji yang akan digunakan pada tahap evaluasi model. Dengan demikian, model *Naive Bayes* berhasil dilatih dan mampu melakukan klasifikasi *sentimen* pada data ulasan pengguna.

4.5 Evaluasi model

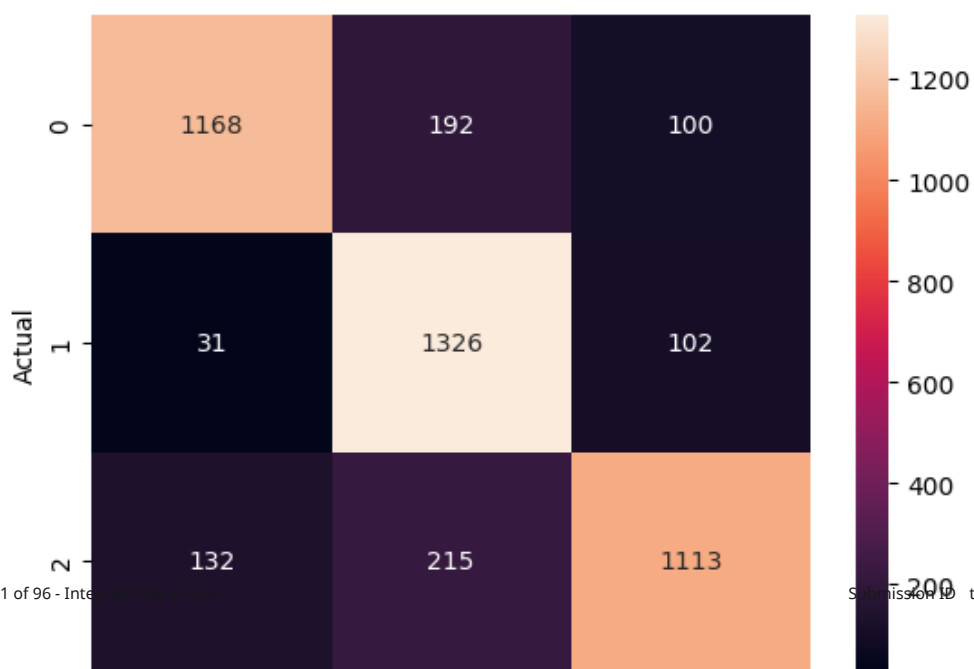
Setelah model klasifikasi dilatih, langkah berikutnya adalah melakukan evaluasi kinerja model. Evaluasi ini bertujuan mengetahui seberapa baik model dalam memprediksi kategori *sentimen* secara akurat.

Proses evaluasi dilakukan menggunakan beberapa metrik utama, yaitu *Confusion Matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk mengukur tingkat ketepatan model secara keseluruhan, sedangkan *precision*, *recall*, dan *F1-score* digunakan untuk mengevaluasi performa model pada masing-masing kelas *sentimen*.

Berdasarkan hasil evaluasi, diperoleh nilai *precision*, *recall*, dan *F1-score* yang relatif seimbang pada setiap kelas *sentimen*. Model menunjukkan performa terbaik pada kelas netral dengan nilai *recall* tertinggi, sedangkan pada kelas positif masih terdapat kesalahan klasifikasi yang ditunjukkan oleh nilai *recall* yang lebih rendah. Hal tersebut menunjukkan bahwa model masih mengalami kesulitan dalam mengenali beberapa data pada kelas tertentu.

1. *Confusion Matrik*

Confusion Matrik merupakan table yang menunjukkan perbandingan label actual dan label prediksi. Untuk klasifikasi tiga kelas (Positif, Negatif, Netral).



Gambar 4. 13 Hasil Confusion Matrix

Pada Gambar 4.13, Confusion Matrix memperlihatkan bahwa sebagian besar data berhasil diklasifikasikan sesuai dengan kelas sebenarnya. Hal tersebut terlihat dari nilai pada diagonal utama, yaitu 1168 data pada kelas 0, 1326 data pada kelas 1, dan 1113 data pada kelas 2. Di sisi lain, masih terdapat beberapa kesalahan klasifikasi yang ditunjukkan oleh nilai di luar diagonal. Sebanyak 192 data pada kelas 0 diprediksi sebagai kelas 1 dan 100 data sebagai kelas 2. Pada kelas 1, terdapat 31 data yang diprediksi sebagai kelas 0 dan 102 data sebagai kelas 2. Sementara itu, pada kelas 2 terdapat 132 data yang diprediksi sebagai kelas 0 serta 215 data sebagai kelas 1. Secara keseluruhan, hasil Confusion Matrix menunjukkan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan sentimen, meskipun masih terjadi kesalahan prediksi pada sebagian data.

Hasil evaluasi menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data sentimen dengan baik. Meskipun demikian, masih ditemukan beberapa kesalahan prediksi, terutama pada kelas yang memiliki karakteristik atau pola yang serupa sehingga memengaruhi ketepatan klasifikasi.

2. Akurasi

Akurasi merupakan metrik evaluasi yang digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan data:

```
from sklearn.metrics import accuracy_score

# Prediksi data testing
y_pred = model.predict(X_test)

# Hitung akurasi
akurasi = accuracy_score(y_test, y_pred)

# Tampilkan hasil
print("Accuracy:", akurasi)
```

Accuracy: 0.8237040420187257

Gambar 4. 14 Hasil Akurasi

Gambar 4.14, Memperlihatkan hasil pengujian bahwa algoritma *Naive Bayes* memperoleh nilai akurasi sebesar 0,8237 atau 82,37%. Hasil tersebut menunjukkan bahwa model mampu mengklasifikasikan *sentimen* ulasan pengguna aplikasi DOLA dengan cukup baik, di mana sebagian besar data uji berhasil diprediksi sesuai dengan kategori *sentimen* yang sebenarnya. Meskipun masih terdapat beberapa kesalahan klasifikasi, nilai akurasi yang diperoleh menunjukkan bahwa model memiliki performa yang cukup baik dan efektif untuk digunakan dalam Analisis *Sentimen* ulasan pengguna aplikasi DOLA di Google *Play Store*.

3. Precision, Recall, dan F1-score

Hasil Classification Report menunjukkan bahwa model memperoleh accuracy sebesar 82,37%, yang mengindikasikan kemampuan model dalam mengklasifikasikan data dengan baik. Pada kelas negatif, model menghasilkan precision sebesar 0,88, recall sebesar 0,80, dan F1-score sebesar 0,84. Kelas netral memiliki precision sebesar 0,77, recall sebesar 0,91, serta F1-score sebesar 0,83. Adapun pada kelas positif, nilai precision mencapai 0,85, recall sebesar 0,76, dan F1-score sebesar 0,80. Secara keseluruhan, nilai weighted average untuk precision, recall, dan F1-score masing-masing sebesar 0,83, 0,82, dan 0,82,

yang menunjukkan bahwa model memiliki kinerja klasifikasi yang cukup konsisten pada seluruh kelas sentimen.

Gambar 4. 15 Hasil Evaluasi Model pada setiap kelas sentimen

Pada Gambar 4.15, Berdasarkan hasil tersebut memperlihatkan bahwa model memiliki performa yang cukup baik dan relatif seimbang dalam mengklasifikasikan ketiga kelas sentimen. Meskipun terdapat sedikit perbedaan nilai antar kelas, model tetap mampu bekerja dengan stabil dalam mengenali pola sentimen pada data ulasan pengguna. Dengan demikian, model Naive Bayes memiliki kinerja yang baik dalam klasifikasi sentimen

4.5.1 Evaluasi Cross Validation

Selain menggunakan pembagian data latih dan data uji, evaluasi model juga dilakukan dengan metode 5-fold cross-validation. Hasil pengujian menghasilkan rata-rata akurasi sebesar 83,06%, yang menunjukkan bahwa model mampu memberikan kinerja yang relatif konsisten pada setiap proses validasi.

4.5.2 Pengujian Skenario Pembagian Data

Kinerja model dievaluasi melalui lima skenario pembagian data latih dan data uji yaitu 70:30, 75:25, 80:20, 85:15, 90:10, Berdasarkan hasil pengujian, diperoleh nilai akurasi sebesar 83,75% pada skenario 70:30, 82,95% pada

skenario 75:25, 83,12% pada skenario 80:20, 83,40% pada skenario 85:15, 83,75% pada skenario 90:10, Dan 82,19%.

Tabel 4. 1 Pembagian Data

NO	Skenario pembagian data	Akurasi	Persentase
1	70% Data Training : 30 % Data Testing	0.8295	82,95%
2	75% Data Training: 25% Data Testing	0.8312	83,12%
3	80% Data Training : 20% Data Testing	0.8340	83,40%
4	85% Data Training : 15% Data Testing	0.8375	83,75%
5	90% Data Training: 10% Data Testing	0.8219	82,19%

Pada Tabel 4.1, Hasil tersebut menunjukkan bahwa perbedaan proporsi pembagian data tidak memberikan pengaruh yang terlalu signifikan terhadap performa model. Namun, skenario 85:15 menghasilkan nilai akurasi tertinggi dibandingkan skenario lainnya, sehingga dapat dikatakan bahwa pembagian data 85:15 merupakan skenario yang paling optimal pada penelitian ini.

4.6 Perbandingan Metode Naive Bayes dan Indobert

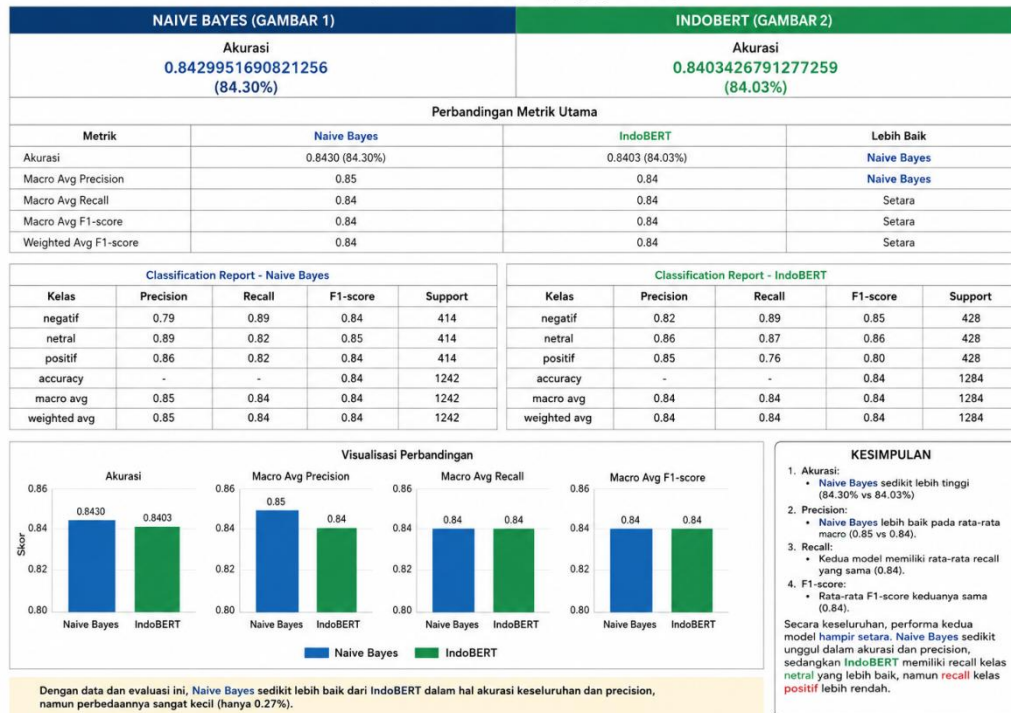
Setelah dilakukan pengujian menggunakan metode Naive Bayes, penelitian ini juga melakukan perbandingan dengan metode IndoBERT menggunakan dataset yang sama. Tujuan perbandingan ini adalah untuk mengetahui metode yang memberikan performa terbaik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi DOLA. Evaluasi dilakukan berdasarkan nilai accuracy, precision, recall, dan F1-score yang diperoleh dari masing-masing model. Selain itu, perbandingan juga mempertimbangkan aspek efisiensi komputasi, seperti

kompleksitas algoritma, waktu pelatihan, dan kebutuhan sumber daya. Hasil

perbandingan kedua metode disajikan pada Gambar 4.6

PERBANDINGAN HASIL NAIVE BAYES DAN INDOBERT

(Data dan Evaluasi dari Output yang Diberikan)



Gambar 4. 16 Hasil Perbandingan Metode Naive Bayes dan IndoBERT

Berdasarkan hasil perbandingan pada dataset yang sama, algoritma Naive Bayes menunjukkan performa yang sedikit lebih baik dibandingkan IndoBERT dengan akurasi sebesar 84,30%, sedangkan IndoBERT memperoleh akurasi 84,03%. Selain menghasilkan akurasi yang lebih tinggi, Naive Bayes juga memiliki proses komputasi yang lebih sederhana, waktu pelatihan yang lebih cepat, serta kebutuhan sumber daya yang lebih rendah. Oleh karena itu, pada penelitian ini Naive Bayes dinilai sebagai metode yang lebih unggul dan lebih sesuai untuk analisis sentimen ulasan aplikasi DOLA.

4.7 Word Cloud

Visualisasi Word Cloud digunakan untuk menampilkan kata-kata yang paling sering muncul pada data ulasan pengguna. Hasil visualisasi tersebut memberikan gambaran mengenai kata-kata dominan yang banyak muncul dalam ulasan aplikasi DOLA.



Gambar 4. 17 Hasil Word Cloud

Pada Gambar 4.17, Ditampilkan Word Cloud untuk sentiment pengguna, Berdasarkan hasil visualisasi, pada sentimen positif didominasi oleh kata-kata seperti “bagus”, “mudah”, dan “membantu” yang menunjukkan kepuasan pengguna terhadap aplikasi. Pada sentimen negatif, kata-kata yang sering muncul antara lain “error”, “lambat”, dan “tidak bisa” yang mengindikasikan adanya kendala dalam penggunaan aplikasi. Sedangkan pada sentimen netral, kata-kata yang muncul cenderung bersifat informatif tanpa menunjukkan emosi yang kuat.

1. Word Cloud Positif



3. *Word Cloud* Netral



4.8 Pembahasan

Submission ID trn:oid:::3618:151105990

antara ketepatan prediksi dan kemampuan model dalam mengenali setiap kelas sentimen.

Performa klasifikasi tersebut diperoleh melalui rangkaian proses yang meliputi preprocessing, pembobotan TF-IDF, penerapan oversampling, dan klasifikasi menggunakan algoritma Naive Bayes. Penerapan oversampling membantu menyeimbangkan distribusi data pada setiap kelas sehingga model dapat mempelajari pola sentimen dengan lebih baik.

Meskipun menghasilkan performa yang cukup baik, hasil Confusion Matrix masih menunjukkan adanya sejumlah prediksi yang tidak sesuai dengan kelas sebenarnya. Kondisi ini diduga dipengaruhi oleh penggunaan kata yang ambigu, bahasa tidak baku, singkatan, serta konteks kalimat yang sulit dikenali oleh model. Untuk penelitian selanjutnya, performa klasifikasi dapat ditingkatkan dengan membandingkan Naive Bayes terhadap algoritma lain, seperti Support Vector Machine (SVM, Random Forest, atau Bidirectional Encoder Representations from Transformers (BERT)), sehingga dapat diketahui metode yang memberikan hasil terbaik pada analisis sentimen ulasan aplikasi.

4.8.1 Perbandingan Hasil Penelitian dengan Literatur

Berdasarkan hasil penelitian yang telah dilakukan, model Naive Bayes dalam penelitian ini menghasilkan nilai akurasi sebesar 83,4%, dengan nilai precision, recall, dan F1-score yang relatif seimbang. Hasil ini menunjukkan bahwa model memiliki performa yang cukup baik dalam melakukan klasifikasi sentimen.

Jika dibandingkan dengan penelitian sebelumnya, algoritma Naive Bayes umumnya menghasilkan tingkat akurasi yang berkisar antara 70% hingga 85%

pada kasus Analisis *Sentimen* berbasis teks. Dengan demikian, hasil yang diperoleh dalam penelitian ini masih berada dalam rentang yang wajar dan kompetitif.

Beberapa penelitian terdahulu menunjukkan bahwa metode lain seperti *Support Vector Machine* (SVM) atau algoritma berbasis deep learning dapat menghasilkan akurasi yang lebih tinggi. Namun, *Naive Bayes* tetap menjadi salah satu metode yang efektif karena memiliki keunggulan dalam kesederhanaan, kecepatan proses, serta kemampuan yang cukup baik dalam mengolah data teks.

Perbedaan hasil akurasi antar penelitian dapat disebabkan oleh beberapa faktor, seperti jumlah dan kualitas dataset, proses *preprocessing*, serta teknik ekstraksi fitur yang digunakan. Dalam penelitian ini, penggunaan *TF-IDF* sebagai metode pembobotan kata turut berkontribusi dalam meningkatkan performa model.

Selain itu, pengujian menggunakan beberapa skenario pembagian data serta metode *cross-validation* menunjukkan bahwa model memiliki performa yang stabil dan tidak bergantung pada satu pembagian data saja. Dengan demikian, dapat disimpulkan bahwa hasil penelitian ini sejalan dengan penelitian sebelumnya dan menunjukkan bahwa algoritma *Naive Bayes* masih relevan dan efektif digunakan dalam analisis *sentimen*, khususnya pada data ulasan pengguna aplikasi.

BAB V

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai Analisis Sentimen ulasan pengguna aplikasi DOLA menggunakan algoritma Naive Bayes, dapat diambil beberapa kesimpulan sebagai berikut:

1. Penelitian ini berhasil menganalisis dan mengklasifikasikan ulasan pengguna aplikasi DOLA pada Google Play Store ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Hasil analisis menunjukkan bahwa sentimen positif mendominasi dibandingkan sentimen negatif dan netral, sehingga dapat disimpulkan bahwa sebagian besar pengguna memberikan tanggapan yang baik terhadap aplikasi DOLA. Proses klasifikasi dilakukan melalui tahapan preprocessing, pembobotan TF-IDF, oversampling, dan penerapan algoritma Naive Bayes.
2. Hasil pengujian menunjukkan bahwa model Naive Bayes memberikan performa yang baik dengan nilai accuracy sebesar 82,37%, precision sebesar 0,83, recall sebesar 0,82, dan F1-score sebesar 0,82. Selain itu, hasil Confusion Matrix menunjukkan bahwa sebagian besar data ulasan berhasil diklasifikasikan dengan benar pada masing-masing kategori sentimen.
3. Berdasarkan hasil penelitian tersebut, algoritma Naive Bayes terbukti efektif dalam menganalisis sentimen pengguna aplikasi DOLA berdasarkan ulasan di Google Play Store. Dengan demikian, seluruh tujuan penelitian telah tercapai, yaitu mengetahui kecenderungan sentimen pengguna, mengklasifikasikan ulasan menggunakan Naive Bayes, serta mengevaluasi kinerja model. Hasil

penelitian ini diharapkan dapat menjadi bahan evaluasi bagi pengembang untuk meningkatkan kualitas layanan dan fitur aplikasi DOLA.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, penelitian selanjutnya disarankan untuk menggunakan jumlah data ulasan yang lebih banyak serta periode pengambilan data yang lebih panjang agar hasil analisis sentimen lebih representatif. Selain itu, penelitian juga dapat diterapkan pada objek atau aplikasi lain sehingga hasil yang diperoleh dapat dibandingkan dan memiliki cakupan yang lebih luas. Peneliti selanjutnya disarankan membandingkan algoritma Naive Bayes dengan algoritma klasifikasi lain, seperti Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest, maupun metode Deep Learning, untuk mengetahui metode yang memberikan performa terbaik. Tahapan preprocessing, terutama pada proses stemming, stopwords removal, normalisasi kata, dan pembersihan teks, juga perlu dioptimalkan agar kualitas data dan hasil klasifikasi semakin baik. Selain menggunakan tiga kategori sentimen (positif, negatif, dan netral), penelitian berikutnya dapat mengembangkan klasifikasi sentimen yang lebih rinci atau menggunakan pendekatan berbasis aspek (aspect-based sentiment analysis) untuk memperoleh informasi yang lebih mendalam. Evaluasi model juga dapat diperluas dengan menggunakan metode validasi dan metrik evaluasi lainnya agar hasil pengujian lebih komprehensif. Hasil penelitian ini diharapkan dapat dimanfaatkan oleh pengembang aplikasi DOLA sebagai bahan evaluasi dalam meningkatkan kualitas layanan, memperbaiki fitur yang masih memperoleh respons negatif, serta mempertahankan dan mengembangkan fitur yang telah mendapatkan respons positif dari pengguna.

DAFTAR PUSTAKA

- [1] N. A. C. A, D. H. Citra, and W. Purnama, “The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store,” vol. 2, no. April, pp. 47–54, 2022.
- [2] D. Fristtikasari, S. Alam, and I. Kurniawan, “Analisis Sentimen Pengguna Aplikasi Kitalulus pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes,” vol. 10, no. 2, pp. 457–473, 2024.
- [3] I. Angga, T. Imam, and Purwadi, “Analisis Sentimen Ulasan Pengguna Aplikasi Sinaga Mobile pada Google Play Store Menggunakan Algoritma Naive Bayes Sentiment Analysis of Sinaga Mobile App User Reviews on Google Play Store Using Naive Bayes Algorithm,” vol. 5, no. 6, pp. 1635–1645, 2025.
- [4] K. Amirulah, S. A. Agus, Muhlis, B. Haqun, and D. Muhamad, “Analisis Sentimen pada Media Sosial dengan Teknik Kecerdasan Buatan Naïve Bayes : Kajian Literatur Review,” vol. 2, no. 6, pp. 1642–1649, 2023.
- [5] V. Nomor, E. Damayanti, N. Rahaningsih, and R. D. Dana, “Analisis Sentimen Pengguna Aplikasi Mobile Jaminan Kesehatan Nasional di Playstore Menggunakan Metode Naive Bayes,” vol. 8, pp. 1–12, 2025.
- [6] M. Metode and N. Bayes, “Jurnal ilmu komputer,” vol. 1, pp. 171–180, 2023.
- [7] S. Raschka, “Introduction and Theory,” pp. 1–20, 2014.
- [8] “An Extant of Natural Language Processing,” vol. 6, no. 2, pp. 1–6, 2024.

- [9] D. Pratmanto, R. Wijianto, and A. Widayanto, “Komparasi K-Nearest Neighbors (KNN) dan Naive Bayes pada Klasifikasi Sentimen Ulasan Aplikasi Tokopedia di Google Play Store,” vol. 5, no. 2, pp. 75–80, 2025.
- [10] R. R. Roska, S. Aso, and M. Teuku, “Analisi Ulasan Shopee Pada Google Play Store Menggunakan Algoritma Naive Bayes”.
- [11] D. Azarine, N. Rahaningsih, and R. D. Dana, “Analisis Sentimen terhadap Pengguna Aplikasi Jamsostek Mobile (JMO) pada Google Play Store Menggunakan Metode Naive Bayes,” vol. 24, no. 1, pp. 13–21, 2025.
- [12] T. S. Gumilar, R. Astuti, and Y. A. Wijaya, “Analisis Ulasan Plikasi Di Play Store Menggunakan ALgoritma Naive Bayes,” vol. 8, no. 1, pp. 543–550, 2024.
- [13] R. Kurniawan and Y. A. Wijaya, “Analisis Sentimen Pengguna terhadap Aplikasi Blu BCA di Play Store Menggunakan Algoritma Naive Bayes,” vol. 8, no. 3, pp. 2511–2517, 2024.
- [14] G. R. Ramadhan *et al.*, “Analisis Sentimen Ulasan Aplikasi DANA di Google Play Store Menggunakan Algoritma Naïve Bayes,” vol. 8, no. 5, pp. 9849–9857, 2024.
- [15] R. S. Nuraini, D. Krisbiantoro, M. Imron, and P. Subarkah, “Analisis Sentimen Aplikasi Agoda Di Google Play Store Menggunakan Algoritma Naive Bayes,” vol. 7, no. 1, 2025.
- [16] S. Nurhikmah, R. Ramadani, and G. Triyono, “Analisis Sentimen pada Ulasan Aplikasi Wondr di Play Store dengan Metode Naïve Bayes,” pp. 1919–1930, 2025, doi: 10.33364/algoritma/v.22-2.2507.

- [17] R. O. Mardiyanto, “Analisis sentimen pengguna aplikasi bank syariah indonesia dengan0menggunakan0algoritma support vector machine (svm),” pp. 9–15.
- [18] H. N. Zuhdi and B. Prasetyo, “Sentiment Analysis On Ipusnas Application Reviews In Google Play Store Using Naive Bayes Classifier Analisis Sentimen pada Ulasan Aplikasi iPusnas di Google Play Store Menggunakan Naive Bayes Classifier,” vol. 5, no. 1, pp. 12–19, 2025.
- [19] F. I. Komputer, U. S. Karawang, J. H. S. Ronggowaluyo, T. Timur, and J. Barat, “Penerapan Algoritma Support Vector Machine terhadap Analisis Sentimen pada Ulasan Aplikasi LinkAja,” vol. 13, no. 3, pp. 1908–1917, 2024.
- [20] L. Hermawan and M. B. Ismiati, “Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval,” vol. 17, no. 2, pp. 188–199, 2020.
- [21] D. Septiani, I. Isabela, F. Teknik, U. N. Jakarta, F. Teknik, and U. N. Jakarta, “Analisis term frequency inverse document frequency (tf-idf) dalam temu kembali informasi pada dokumen teks,” vol. 25, pp. 81–88.
- [22] S. F. Kadir and A. Fairuzabadi, “Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan Logistic Regression,” vol. 4, no. 2, pp. 7940–7945, 2025.