

CEK TURNITIN

Skripsi adif turnitin

 CEK 1

Document Details

Submission ID**trn:oid:::3618:152713602****Submission Date****Sep 23, 2026, 9:24 PM GMT+7****Download Date****Sep 23, 2026, 9:27 PM GMT+7****File Name****Skripsi adif turnitin.docx****File Size****1.7 MB****77 Pages****12,775 Words****85,406 Characters**

49% Overall Similarity

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

Filtered from the Report

- Bibliography

Top Sources

- 35%  Internet sources
- 34%  Publications
- 45%  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

- 35% Internet sources
- 34% Publications
- 45% Submitted works (Student Papers)

Top Sources

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

1	Internet	repository.unisi.ac.id	2%
2	Student papers	UPN Veteran Yogyakarta on 2026-07-21	1%
3	Internet	ejurnal.seminar-id.com	<1%
4	Student papers	Universitas Putera Batam on 2026-07-25	<1%
5	Student papers	Universitas Gunadarma on 2026-07-17	<1%
6	Internet	ejournal.itn.ac.id	<1%
7	Internet	proceeding.unpkediri.ac.id	<1%
8	Student papers	UPN Veteran Yogyakarta on 2026-06-29	<1%
9	Publication	Akbar Supia Dirja, Lana Aulia Salsabila, M. Akhdan Putra Hermawan, Cika Jelita, H...	<1%
10	Student papers	IAKN Palangka Raya on 2026-06-14	<1%
11	Student papers	Universitas Putera Batam on 2026-07-27	<1%

12	Internet	jurnal.polibatam.ac.id	<1%
13	Student papers	UPN Veteran Yogyakarta on 2026-08-18	<1%
14	Internet	jurnal.fikom.umi.ac.id	<1%
15	Student papers	UPN Veteran Yogyakarta on 2026-04-02	<1%
16	Internet	jpti.journals.id	<1%
17	Student papers	Universitas Pamulang on 2026-07-13	<1%
18	Internet	www.ejournal.itn.ac.id	<1%
19	Student papers	Universitas Sangga Buana YPKP on 2026-07-24	<1%
20	Internet	repository.unas.ac.id	<1%
21	Student papers	Universitas Putera Batam on 2026-07-28	<1%
22	Student papers	Universitas Pamulang on 2026-07-21	<1%
23	Student papers	Universitas Putera Batam on 2026-02-02	<1%
24	Internet	ojs.stmik-banjarbaru.ac.id	<1%
25	Publication	Nurhabsah Nurhabsah, Nurdin Nurdin. "PERBANDINGAN ALGORITMA REGRESI LI...	<1%

26	Internet	journal.irpi.or.id	<1%
27	Internet	repository.uinsaizu.ac.id	<1%
28	Student papers	Universitas Sangga Buana YPKP on 2026-07-20	<1%
29	Internet	kc.umn.ac.id	<1%
30	Student papers	Universitas Pamulang on 2026-07-28	<1%
31	Internet	eskripsi.usm.ac.id	<1%
32	Internet	repository.usni.ac.id	<1%
33	Internet	jurnal.pustakagalerimandiri.co.id	<1%
34	Internet	repository.upnvj.ac.id	<1%
35	Student papers	Universitas Bengkulu on 2025-08-28	<1%
36	Internet	ejournal.raharja.ac.id	<1%
37	Publication	Zendra Utama, Pahrizal Pahrizal. "PENERAPAN METODE CNN UNTUK ANALISIS SE...	<1%
38	Publication	Sajid Mahattir Aulia Rahman, Tommy Bustomi, Ahmad Fahrijal Pukeng. "Analisis ...	<1%
39	Student papers	Universitas Islam Riau on 2025-11-12	<1%

40	Student papers	Universitas Muhammadiyah Jember on 2026-01-27	<1%
41	Student papers	Universitas Sangga Buana YPKP on 2026-07-28	<1%
42	Publication	Septiana Septiana, Dwi Krisbiantoro, Suliswaningsih Suliswaningsih. "Classificatio...	<1%
43	Student papers	Telkom University on 2026-06-30	<1%
44	Student papers	Universitas Pamulang on 2026-05-21	<1%
45	Student papers	Universitas Pamulang on 2026-07-11	<1%
46	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2025-08-26	<1%
47	Student papers	Universitas Islam Sumatera Utara on 2026-04-20	<1%
48	Publication	Sri Damayanti, Husna Sari, Bobby Ardiansyah, Dicky Apdillah. "Analisis Sentimen ...	<1%
49	Student papers	Universitas Muslim Indonesia on 2026-07-02	<1%
50	Internet	archive.umsida.ac.id	<1%
51	Internet	jurnal.bsi.ac.id	<1%
52	Student papers	Sriwijaya University on 2021-10-28	<1%
53	Student papers	Telkom University on 2026-08-07	<1%

54	Student papers	Universitas Sangga Buana YPKP on 2026-07-31	<1%
55	Internet	openjournal.unpam.ac.id	<1%
56	Internet	repository.uin-suska.ac.id	<1%
57	Publication	Istiana Ruswita, M. Yhogha Ismail Ibn Ibrahim. "Analisis Sentimen Ulasan Digital ...	<1%
58	Student papers	Universitas Islam Syekh-Yusuf Tangerang on 2026-07-28	<1%
59	Internet	www.mdpi.com	<1%
60	Publication	Siti Khairunnisa Ratu Rahmah, Muhammad Ali Nur Hidayat, Jamaludin Hakim, Mu...	<1%
61	Student papers	Universitas Islam Indonesia on 2026-05-12	<1%
62	Student papers	Universitas Katolik Widya Mandala Surabaya on 2025-12-03	<1%
63	Internet	subset.id	<1%
64	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-14	<1%
65	Student papers	Universitas Bengkulu on 2025-07-09	<1%
66	Student papers	Universitas Katolik Musi Charitas on 2026-08-13	<1%
67	Student papers	Universitas Pancasila on 2026-03-04	<1%

68	Internet	journal.ubm.ac.id	<1%
69	Internet	repository.upnjatim.ac.id	<1%
70	Publication	Selvi Nuriana, Fitri Yunita. "Prediksi Stunting Menggunakan Random Forest dan S...	<1%
71	Student papers	Sriwijaya University on 2025-11-24	<1%
72	Student papers	Universitas Bengkulu on 2025-10-17	<1%
73	Publication	Ahmad Miftakhudin, Nugroho Adhi Santoso, Bayu Aji Santoso. "Komparasi Algotir...	<1%
74	Publication	Muhamad Azril, Giatika Crisnawati. "Klasifikasi Sentimen Komentar Pengguna Tik...	<1%
75	Publication	Octavia Salwa Dzaky Fadhillah, Jajam Haerul Jaman, Carudin Carudin. "PERBANDI...	<1%
76	Student papers	Universitas Katolik Musi Charitas on 2026-08-12	<1%
77	Publication	Karyono, Slamet. "Model Peningkatan Kinerja Sumber Daya Manusia Melalui Kom...	<1%
78	Student papers	Universitas Muhammadiyah Sumatera Utara on 2025-08-21	<1%
79	Student papers	Universitas Putera Batam on 2026-02-02	<1%
80	Internet	repository.unissula.ac.id	<1%
81	Publication	Kezia Cantika Saroinsong, Irene R. H. T. Tangkawarow, Efraim R. S. Moningkey. "R...	<1%

82	Publication	Siti Hajar, Nurdin Nurdin. "PERBANDINGAN KLASIFIKASI ALGORITMA DECISION T...	<1%
83	Student papers	Universitas Muhammadiyah Jember on 2026-01-17	<1%
84	Internet	etheses.uin-malang.ac.id	<1%
85	Internet	journal.ilmudata.co.id	<1%
86	Student papers	Universitas Muhammadiyah Palembang on 2025-08-06	<1%
87	Student papers	Universitas Muslim Indonesia on 2025-12-12	<1%
88	Publication	Harliana Harliana, Tito Prabowo, Ady Alzhava Nuary. "Perbandingan Logistic Reg...	<1%
89	Student papers	Institut Bisnis dan Teknologi Indonesia (INSTIKI) on 2025-06-08	<1%
90	Student papers	Institut Teknologi Sumatera on 2026-09-18	<1%
91	Student papers	LL DIKTI IX Turnitin Consortium Part III on 2026-07-21	<1%
92	Student papers	Universitas Islam Indonesia on 2025-09-17	<1%
93	Student papers	Universitas Islam Riau on 2026-08-21	<1%
94	Internet	aira.or.id	<1%
95	Internet	conference.usm.ac.id	<1%

96	Internet	ejurnal.lkparyaprima.id	<1%
97	Internet	ejurnal.stmik-budidarma.ac.id	<1%
98	Internet	sistemasi.ftik.unisi.ac.id	<1%
99	Internet	www.ejournal.unma.ac.id	<1%
100	Internet	www.hostjournals.com	<1%
101	Publication	Ade Toti Febrian, Purwadi Purwadi, Adam Prayogo Kuncoro. "Comparative Analys...	<1%
102	Publication	Muhammad Basyrah Diyanulhaq, Saeffurohman Saeffurohman. "KLASIFIKASI SE...	<1%
103	Student papers	Politeknik Manufaktur Negeri Bangka Belitung on 2026-07-10	<1%
104	Student papers	Universitas Bina Sarana Informatika on 2026-02-20	<1%
105	Internet	repository.dinamika.ac.id	<1%
106	Publication	Aulia Puspa, Aviarini Indrati. "Implementasi Algoritma Random Forest Classifier ...	<1%
107	Publication	Fadila Ananda Kartika Hidayat, Neny Sulistianingsih, Rifqi Hammad. "Perbanding...	<1%
108	Student papers	Institut Teknologi Sumatera on 2025-09-22	<1%
109	Publication	Janwar Iqbal Kurniawan, Mohammad Imron, Shodiq Khalidy. "Systematic Review:...	<1%

110	Publication	Rinda Fitriana Azmi, Galih Hendro Martono, Neny Suliastianingsih. "Komparasi Al...	<1%
111	Student papers	UPN Veteran Yogyakarta on 2026-03-02	<1%
112	Student papers	Universitas Islam Riau on 2025-07-18	<1%
113	Internet	hostjournals.com	<1%
114	Internet	journal.nurulfikri.ac.id	<1%
115	Internet	jurnalsinta.id	<1%
116	Publication	Putri Widya Sari Sari, Firmansyah Firmansyah, Abdul Rahman Kadafi Kadafi. "PER...	<1%
117	Publication	Sri Intan Ramayani Purba, Anita, Yohana Argentina Br Hutauruk. "PERBANDINGA...	<1%
118	Student papers	Telkom University on 2026-04-27	<1%
119	Student papers	Telkom University on 2026-06-08	<1%
120	Student papers	Universitas Muslim Indonesia on 2025-08-13	<1%
121	Student papers	Universitas Negeri Surabaya The State University of Surabaya on 2026-06-10	<1%
122	Student papers	Universitas Pancasila on 2026-03-05	<1%
123	Student papers	Universitas Sangga Buana YPKP on 2026-07-28	<1%

124	Publication	Dwi Nanda Agustia, Ryan Randy Suryono. "Comparison of Naïve Bayes, Random ...	<1%
125	Student papers	Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur II on 2025-08-20	<1%
126	Publication	Githa Alfiansyah, Neny Sulistianingsih, I Nyoman Switrayana. "Perbandingan Kin...	<1%
127	Publication	Herdiesel Santoso, Wahyuni Nareswari. "Analisis Sentimen Berbasis Aspek Pada ...	<1%
128	Publication	Iqbar Sajidan Widiyanto, Yudhi Raymond Ramadhan, Yudhi Raymond Ramadhan, ...	<1%
129	Publication	Irwansyahputra Irwansyahputra, Nurdin Nurdin. "PERBANDINGAN ALGORITMA ...	<1%
130	Publication	Nadya Mayla Putri Rahmadani, Muhamad Rizaludin. "Perbandingan Kinerja Naïve...	<1%
131	Student papers	UPN Veteran Yogyakarta on 2026-05-04	<1%
132	Student papers	UPN Veteran Yogyakarta on 2026-05-05	<1%
133	Student papers	UPN Veteran Yogyakarta on 2026-09-03	<1%
134	Student papers	Universitas Djuanda on 2025-07-15	<1%
135	Student papers	Universitas Muslim Indonesia on 2024-12-09	<1%
136	Student papers	Universitas Muslim Indonesia on 2025-02-07	<1%
137	Student papers	Universitas Putera Batam on 2026-07-24	<1%

138	Internet	teknosi.fti.unand.ac.id	<1%
139	Publication	I Putu Gede Abdi Sudiatmika, Dewa Ayu Mas Putriari Nusantara, Wayan Andrika P...	<1%
140	Student papers	Sriwijaya University on 2022-01-19	<1%
141	Student papers	UPN Veteran Yogyakarta on 2026-06-26	<1%
142	Student papers	Universitas Putera Batam on 2026-02-03	<1%
143	Student papers	Universitas Putera Batam on 2026-02-03	<1%
144	Internet	dspace.uui.ac.id	<1%
145	Internet	journal.darmajaya.ac.id	<1%
146	Internet	journal2.uad.ac.id	<1%
147	Publication	Siti Aiwestopa Riyandona, Nining Rahaningsih, Raditya Danar Dana, - Mulyawan. ...	<1%
148	Student papers	UIN Raden Intan Lampung on 2026-07-03	<1%
149	Student papers	Universitas Islam Riau on 2026-06-19	<1%
150	Student papers	Universitas Pancasila on 2026-02-21	<1%
151	Internet	docplayer.info	<1%

152	Internet	iconline.ipleiria.pt	<1%
153	Internet	jurnalsyntaxadmiration.com	<1%
154	Internet	pppm.stis.ac.id	<1%
155	Internet	repository.unja.ac.id	<1%
156	Internet	repository.yudharta.ac.id	<1%
157	Publication	Basiroh, Nancy Octyajati, Riana Rahmawati Dewi. "PENDAMPINGAN PEMBELAJAR..."	<1%
158	Student papers	Institut Teknologi Sumatera on 2026-03-10	<1%
159	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2021-08-23	<1%
160	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2024-07-15	<1%
161	Publication	Miftakhul Rahman, Mantri Kromo Fandith Fili, Wardianto. "Analisis Hasil Rekapit..."	<1%
162	Publication	Muhammad Hanif, Garno, Asep Jamaludin. "ANALISIS SENTIMEN PENGGUNA APL..."	<1%
163	Publication	Muhammad Irvan Maulana, Savana Putra Aditama, Harun Al Rosyid. "Komparasi..."	<1%
164	Student papers	UPN Veteran Yogyakarta on 2025-12-22	<1%
165	Student papers	UPN Veteran Yogyakarta on 2026-06-03	<1%

166	Student papers	Universitas Islam Riau on 2025-09-03	<1%
167	Publication	Yola Tirta Nurfajriya, Pinkan Prilia Mahmudah, Hana Aurora Aida Rahma, Hanif Z...	<1%
168	Internet	eprints.uns.ac.id	<1%
169	Internet	jpi.eng.unila.ac.id	<1%
170	Internet	jurnal.satyaterrabhinneka.ac.id	<1%
171	Internet	jurnal.unived.ac.id	<1%
172	Internet	repo.undiksha.ac.id	<1%
173	Publication	Devi Sugianti, Ari Putra, Anas Syaifudin, Rizqi Wijonarko. "Algoritma K Means dan...	<1%
174	Publication	Dinda Mariyani, Ade Irma Purnamasari, Irfan Ali, Odi Nurdiawan, Rudi Nurdiawa...	<1%
175	Publication	Randi Afif Afif, Aji Supriyanto, Rr. Fitri Damaryanti, Wahyu Prasetya Adi. "Analisis ...	<1%
176	Student papers	Sekolah Teknik Elektro & Informatika on 2026-07-17	<1%
177	Student papers	Universitas Putera Batam on 2026-07-25	<1%
178	Student papers	Universitas Sangga Buana YPKP on 2026-07-21	<1%
179	Internet	journal.global.ac.id	<1%

180	Internet	repository.ub.ac.id	<1%
181	Internet	repository.ulb.ac.id	<1%
182	Internet	sipora.polije.ac.id	<1%
183	Publication	Annisa Maizaliyanti, Khothibul Umam, Wenty Dwi Yuniarti, Maya Rini Handayani. ...	<1%
184	Publication	I Putu Arya Putra, I Made Agus Wirawan, I Gede Aris Gunadi. "Improving Butterfl...	<1%
185	Student papers	Institut Teknologi Sumatera on 2026-09-22	<1%
186	Publication	M. Rizki Mahrifan, Fitri Yunita. "Analisis Sentimen pada Aplikasi MyXL Mengguna...	<1%
187	Publication	M. Saleh, Dwi Yuli Prasetyo. "Analisis Sentimen Konsumen Terhadap Usaha Food ...	<1%
188	Publication	Muhammad Reyan, Franindya Purwaningtyas. "ANALISIS SENTIMEN ULASAN APL...	<1%
189	Publication	Siti Noraisyah, Fitri Yunita. "Analisis Sentimen Penggunaan ChatGPT di Kalanga...	<1%
190	Student papers	Sriwijaya University on 2021-12-03	<1%
191	Student papers	Universitas Dinamika on 2026-06-11	<1%
192	Student papers	Universitas Islam Riau on 2025-07-17	<1%
193	Student papers	Universitas Islam Riau on 2025-10-31	<1%

194	Student papers	Universitas Islam Riau on 2025-12-29	<1%
195	Student papers	Universitas Islam Riau on 2026-06-19	<1%
196	Student papers	Universitas Pamulang on 2026-07-29	<1%
197	Student papers	Universitas Sanata Dharma on 2026-06-09	<1%
198	Student papers	Universitas Sangga Buana YPKP on 2026-07-28	<1%
199	Internet	core.ac.uk	<1%
200	Internet	ejournal.um-sorong.ac.id	<1%
201	Internet	ejurnal.itats.ac.id	<1%
202	Internet	ejurnal.methodist.ac.id	<1%
203	Internet	journal.stmikjayakarta.ac.id	<1%
204	Internet	repository.nusaputra.ac.id	<1%
205	Student papers	East Union High School on 2026-06-30	<1%
206	Student papers	Perpustakaan on 2025-08-26	<1%
207	Publication	Syifa Fauzia Tahir, Castaka Agus Sugianto. "OPTIMASI NAIVE BAYES MENGGUNAK...	<1%

208	Student papers	UIN Syarif Hidayatullah Jakarta on 2023-06-05	<1%
209	Student papers	Universitas Budi Luhur on 2026-01-27	<1%
210	Student papers	Universitas Kristen Duta Wacana on 2026-05-07	<1%
211	Student papers	Universitas Muslim Indonesia on 2025-08-27	<1%
212	Student papers	Universitas Pamulang on 2026-08-05	<1%
213	Student papers	Universitas Riau on 2023-02-06	<1%
214	Internet	cdn.juris.id	<1%
215	Internet	journal.amikindonesia.ac.id	<1%
216	Publication	Alidin M, Soni Soni. "Systematic Literature Review Perbandingan Algoritma Tree-...	<1%
217	Publication	Annisa Saninah, Willy Prihartono, Cep Lukman Rohmat. "ANALISIS SENTIMEN UL...	<1%
218	Publication	Christine Eirene, Dian Syafitri, Neny Sulistianingsih, Khasnur Hidjah, Hairani Ha...	<1%
219	Student papers	Institut Teknologi Sumatera on 2026-08-12	<1%
220	Publication	Kartarina ,, Hairul Imam. "Deep Learning Verifikasi Kemiripan Wajah Dengan Arsi...	<1%
221	Publication	Lady Agustin Fitriana, Ipin Sugiyarto, Umi Faddillah. "Penerapan Model LSTM dan...	<1%

222	Publication	Nanda Dwi Husna Sadikin, Sari Susanti. "Analisis Sentimen Publik Terhadap Kamp...	<1%
223	Student papers	Otto-von-Guericke-Universität Magdeburg on 2026-07-29	<1%
224	Publication	Salma Difa Aristawidya, Tabita Nurtirta Purwita Sari, Hanun Syahidah Ulfya, Adin...	<1%
225	Student papers	Syiah Kuala University on 2026-04-01	<1%
226	Student papers	Telkom University on 2026-03-27	<1%
227	Student papers	UNIVERSITAS BUDI LUHUR on 2026-01-22	<1%
228	Student papers	UPN Veteran Yogyakarta on 2026-05-29	<1%
229	Student papers	UPN Veteran Yogyakarta on 2026-07-20	<1%
230	Publication	Umar Rakhman, Rini Astuti, Willy Prihartono, Ryan Hanomangan. "PENGUKURAN ...	<1%
231	Student papers	Universitas Budi Luhur on 2025-07-11	<1%
232	Student papers	Universitas Diponegoro on 2018-11-04	<1%
233	Student papers	Universitas Esa Unggul on 2026-02-01	<1%
234	Student papers	Universitas Muhammadiyah Makassar on 2026-08-27	<1%
235	Student papers	Universitas Pamulang on 2026-06-22	<1%

236	Student papers	Universitas Pamulang on 2026-07-20	<1%
237	Student papers	Universitas Pancasila on 2025-08-07	<1%
238	Student papers	Universitas Pancasila on 2026-02-23	<1%
239	Student papers	Universitas Pendidikan Indonesia on 2026-06-11	<1%
240	Student papers	Universitas Sanata Dharma on 2026-06-10	<1%
241	Student papers	Universitas Sumatera Utara on 2024-12-23	<1%
242	Publication	Yeni Kustiyahningsih, Yohan Permana. "Penggunaan Latent Dirichlet Allocation (...)	<1%
243	Internet	blog.ub.ac.id	<1%
244	Internet	conference.ut.ac.id	<1%
245	Internet	ejournal.uksw.edu	<1%
246	Internet	jurnal.itg.ac.id	<1%
247	Internet	jurnal.stkipppgritulungagung.ac.id	<1%
248	Internet	marisasugangga.wordpress.com	<1%
249	Internet	penerbitadm.pubmedia.id	<1%

250	Internet	polresmitra.blogspot.com	<1%
251	Internet	repository.upbatam.ac.id	<1%
252	Internet	www.suara.com	<1%
253	Internet	www.unisbank.ac.id	<1%
254	Publication	Fastabiqul Khusna, Khothibul Umam, Siti Nur'aini, Maya Rini Handayani. "EVAL...	<1%
255	Publication	Ina Najiyah, Miftahul Rizal. "Analisis Sentimen Media Sosial X Terhadap Kebijakan...	<1%
256	Publication	Muhammad Rafid Pratama, Maya Rini Handayani, Wenty Dwi Yuniarti, Khothibul...	<1%
257	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-02	<1%
258	Student papers	Universitas Budi Luhur on 2026-07-02	<1%
259	Student papers	Universitas Islam Riau on 2026-07-21	<1%
260	Student papers	Universitas Pamulang on 2026-06-09	<1%
261	Student papers	Universitas Pendidikan Indonesia on 2025-08-11	<1%
262	Student papers	Universitas Tarumanagara on 2025-12-19	<1%
263	Student papers	itera on 2025-02-09	<1%

264	Publication	Adryan Syah, Firman Nurdyansyah, Aviv Yuniar Rahman. "ANALISIS SENTIMEN A...	<1%
265	Publication	Anak Agung Ayu Ratna Wulandari, K Suar Wibawa, N W Emmy Rosiana Dewi. "Co...	<1%
266	Student papers	Defense University on 2024-07-03	<1%
267	Publication	Diah Angraina, Aprillia Putri. "Analisis Sentimen Pengguna Aplikasi Google Meet ...	<1%
268	Publication	Muhammad Rioardian Syaputa, Muhammad Arifin, Diana Laily Fithri. "Klasifikasi ...	<1%
269	Publication	Muhammad Wisnu Nugroho. "Analisis Performa Algoritma Random Forest dalam...	<1%
270	Publication	Surya Viari Tampubolon, Danang Arbian Sulistyo. "Analisis Sentimen Ulasan Gam...	<1%
271	Student papers	UPN Veteran Yogyakarta on 2026-06-02	<1%
272	Student papers	Udayana University on 2026-05-25	<1%
273	Student papers	Universitas Brawijaya on 2020-06-22	<1%
274	Student papers	Universitas Islam Riau on 2025-12-10	<1%
275	Student papers	Universitas Jember on 2022-07-17	<1%
276	Student papers	Universitas Katolik Widya Mandala Surabaya on 2025-12-09	<1%
277	Student papers	Universitas Muslim Indonesia on 2025-08-16	<1%

278	Student papers	Universitas Putera Batam on 2025-07-28	<1%
279	Student papers	Universitas Sangga Buana YPKP on 2026-07-29	<1%
280	Student papers	Universitas Sangga Buana YPKP on 2026-07-29	<1%
281	Student papers	Universitas Sangga Buana YPKP on 2026-08-10	<1%
282	Student papers	Universitas Tarumanagara on 2025-12-15	<1%

6

ANALISIS SENTIMEN ULASAN APLIKASI JOBSTREET PADA GOOGLE PLAY STORE MENGGUNAKAN SMOTE DENGAN ALGORITMA XGBOOST DAN RANDOM FOREST

SKRIPSI



DISUSUN OLEH :

MUHAMAD ADIF

403221010072

1

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS ISLAM INDRAGIRI
2026

PERSETUJUAN

PENGESAHAN

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, Tugas Akhir ini merupakan karya saya sendiri (Asli), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu instusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis dan/ atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka

Segala sesuatu yang terkait dengan naskah dan karya yang dibuat adalah menjadi tanggung jawab saya pribadi.

Tembilahan 24 Juli 2026

Muhamad Adif

Nim : 403221010072

KATA PENGANTAR

“Data bukan sekadar angka, tetapi jejak cerita manusia yang menunggu untuk dipahami.”

Puji syukur ke hadirat Allah Swt, tuhan yang maha Esa atas rahmat, kasih, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Ulasan Aplikasi JobStreet di Google Play Store Menggunakan SMOTE dengan Algoritma XGBoost dan Random Forest”.

Perjalanan saat menyusun skripsi ini bukanlah hal yang mudah. Berbagai tantangan, keraguan, dan keterbatasan sempat menjadi bagian dari proses yang harus dilalui. Namun, di balik setiap proses penyusunan selalu ada dukungan, doa, dan semangat dari orang-orang terdekat yang menjadi energi bagi penulis untuk terus melangkah hingga sampai sekarang ini. Skripsi ini bukan hanya sekadar karya ilmiah, tetapi juga menjadi bukti dari proses belajar, perjuangan, dan ketekunan yang telah dilalui.

Tanpa mengurangi rasa hormat penulis berterima kasih kepada rekan-rekan yang terlibat dalam penyusunan skripsi, yaitu,:

1. Bapak Dr. H. Najamuddin, Lc., M.A, sebagai Rektor Universitas Islam Indragiri.
2. Dr. Bayu Rianto, S.Si., M.,Kom selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Islam Indragiri,.
3. Ibu Fitri Yunita, S.SI., M.Kom, selaku Ketua Program Studi Sistem Informasi.
4. Bapak Samsudin, S.Kom., M.Kom., selaku pembimbing pertama yang berjasa membimbing, mengarahkan dan memberi masukan dengan sabar dalam penulisan skripisi ini.

5. Bapak Muh Rasyid Ridha, S.SI., M.Kom., selaku pembimbing penulisan skripsi ini.
6. Para dosen penguji yang telah meluangkan waktu, memberikan kritik, saran, serta masukan yang sangat berarti demi penyempurnaan skripsi ini.
7. Bapak Sudirman dan Ibu Eliani selaku kedua orangtua yang sangat saya cintai, inti dari segala perjuangan saya, dan menjadi alasan saya mengapa saya sampai ke titik ini, sekali lagi terima kasih banyak atas ketulusan, kesabaran serta dukungannya selama ini yang tidak akan mampu semuanya diutarakan oleh kata.
8. Semua pihak yang terlibat membantu secara langsung maupun tidak, tidak bisa saya sebutkan satu persatu.
9. Terakhir, penulis mempersembahkan skripsi ini spesial untuk orang yang selalu bertanya kapan kamu wisuda? Dan kapan skripsimu selesai? Wisuda hanya seremonial akhir setelah melewati berbagai proses, terlambat lulus atau tidak, lulus tepat waktu bukanlah sebuah kejahatan dan bukanlah sebuah aib. Alangkah kerdilnya jika kecerdasan seseorang diukur dari cepatnya wisudanya. Bukankah sebaik-baiknya skripsi adalah skripsi yang selesai entah itu tepat waktu maupun tidak.

Penulis menyadari skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis dengan rendah hati menerima segala kritik dan saran yang membangun sebagai bahan perbaikan di masa yang akan datang.

Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat, tidak hanya bagi pengembangan ilmu pengetahuan, tetapi juga sebagai langkah

kecil yang dapat memberikan kontribusi nyata, khususnya dalam pemanfaatan teknologi untuk memahami kebutuhan dan pengalaman pengguna.

Tembilahan, 24 juli 2026

Penulis

Muhamad Adif

DAFTAR ISI

HALAMAN JUDUL

PERSETUJUAN

PENGESAHAN

PERNYATAAN

KATA PENGANTAR v

DAFTAR ISI vii

DAFTAR TABEL ix

DAFTAR GAMBAR x

INTISARI xi

ABSTRACT xii

BAB I PENDAHULUAN 1

1.1 Latar Belakang 1

1.2 Rumusan Masalah 4

1.3 Batasan Penelitian 4

1.4 Tujuan Penelitian 5

1.5 Manfaat Penelitian 5

1.6 Sistematika Penelitian 6

BAB II TINJAUAN LITERATUR 8

2.1 Analisis Sentimen 8

2.2 Text Mining 10

2.3 Pelabelan Sentimen 10

2.4	Preprocessing Teks	11
2.5	TF-IDF (Term Frequency-Inverse Document Frequency)	11
2.6	SMOTE.....	12
2.7	XGBoost.....	12
2.8	Random Forest.....	12
2.9	Confusion Matrix.....	13
2.10	Evaluasi Model	13
2.11	Penelitian Terdahulu.....	14
2.12	Rangkuman.....	19
BAB III METODE PENELITIAN		22
3.1	Jenis Penelitian	22
3.2	Kerangka Penelitian.....	22
3.3	Teknik Pengumpulan Data	23
3.4	Seleksi Data	25
3.5	Pelabelan Sentimen	25
3.6	Preprocessing Teks.....	25
3.7	Ekstraksi Fitur Menggunakan TF-IDF	26
3.8	Pembagian Data Training dan Testing	26
3.9	Penerapan SMOTE.....	27
3.10	Klasifikasi Menggunakan XGBoost.....	27
3.11	Klasifikasi Menggunakan Random Forest.....	28
3.12	Evaluasi Model.....	30
3.13	Perbandingan Hasil Algoritma	30
3.14	Perangkat dan Tools Penelitian	31
BAB IV HASIL PENELITIAN		33
4.1	Hasil Pengumpulan data	33
4.2	Hasil Seleksi Data.....	35
4.3	Hasil Pelabelan Sentimen	36
4.4	Hasil Preprocessing Data.....	37
4.4.1	Cleaning	38
4.4.2	Case Folding	38
4.4.3	Tokenizing	39
4.4.4	Stop Removal.....	40
4.4.5	Stemming	41
4.5	Hasil Ekstraksi Fitur TF-IDF.....	42
4.6	Hasil Pembagian Data	44

4.7	Hasil Penerapan SMOTE.....	45
4.8	Hasil Klasifikasi XGBoost	47
4.9	Hasil Klasifikasi Random Forest.....	50
4.10	Analisis Performa Algoritma.....	53
4.11	Pembahasan Hasil Sentimen.....	56
BAB V KESIMPULAN DAN SARAN.....		60
5.1	Kesimpulan.....	60
5.2	Saran.....	62
DAFTAR PUSTAKA		64

DAFTAR TABEL

Tabel 2. 1	Penelitian Terdahulu	14
Tabel 3. 1	Format data awal.....	24
Tabel 3. 2	Tahapan Seleksi Data.....	25
Tabel 3. 3	Pelabelan Awal	25
Tabel 3. 4	Tahapan Proprocessing	26
Tabel 3. 5	Skenario Pembagian Data Training dan Testing.....	27
Tabel 3. 6	Distribusi Data Sebelum dan Sesudah Penerapan SMOTE	27
Tabel 3. 7	Tahapan Klasifikasi Menggunakan XGBoost.....	28
Tabel 3. 8	Tahapan Klasifikasi Menggunakan Random Forest	29
Tabel 3. 9	Metrik Evaluasi Model	30
Tabel 3. 10	Rencana Perbandingan Hasil Evaluasi XGBoost dan Random Forest	30
Tabel 3. 11	Perangkat Keras	31
Tabel 3. 12	Perangkat Lunak	31
Tabel 4. 1	Pembagian Data	45
Tabel 4. 2	Perbandingan Performa Algoritma	53

DAFTAR GAMBAR

Gambar 3. 1 Kerangka Penelitian	23
Gambar 4. 1 Data Ulasan Hasil Pengumpulan Data	34
Gambar 4. 2 Tampilan Data Ulasan Aplikasi Jobstreet	34
Gambar 4. 3 Hasil Seleksi Data	35
Gambar 4. 4 Hasil Pelabelan Sentimen.....	36
Gambar 4. 5 Diagram Distribusi Ulasan Aplikasi Jobstreet	37
Gambar 4. 6 Hasil Tahap Cleaning.....	38
Gambar 4. 7 Hasil Tahap Case Folding	39
Gambar 4. 8 Hasil Tahap Tokenizing	40
Gambar 4. 9 Hasil Tahap Stopword Removal	40
Gambar 4. 10 Hasil Tahap Stemming dan Hasil Akhir Preprocessing.....	41
Gambar 4. 11 Hasil Matriks TF-IDF	43
Gambar 4. 12 Hasil Ekstraksi Fitur TF-IDF	43
Gambar 4. 13 Hasil Distribusi Data Training	46
Gambar 4. 14 Distribusi Data Training Sebelum dan Sesudah SMOTE	46
Gambar 4. 15 Confusion Matrix XGBoost	48
Gambar 4. 16 Confusion Matrix XGBoost	48
Gambar 4. 17 Hasil Classification Report Algoritma XGBoost.....	49
Gambar 4. 18 Confusion Matrix Random Forest.....	50
Gambar 4. 19 Confusion Matrix Random Forest.....	51
Gambar 4. 20 Classification Report Random Forest	52
Gambar 4. 21 Grafik Perbandingan Performa Algoritma XGBoost dan Random Forest.....	54
Gambar 4. 22 WordCloud Sentimen Positif	57
Gambar 4. 23 WordCloud Sentimen Negatif.....	57

INTISARI

Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi JobStreet pada Google Play Store dengan menerapkan teknik oversampling SMOTE serta membandingkan performa algoritma XGBoost dan *Random Forest*. Penelitian ini dilatarbelakangi oleh tingginya jumlah ulasan pengguna yang menghasilkan data teks tidak terstruktur, serta adanya ketidakseimbangan jumlah data pada kelas sentimen yang berpotensi menurunkan kinerja model klasifikasi. Data penelitian diperoleh melalui proses *scraping* sebanyak 9.000 ulasan, kemudian setelah tahap seleksi diperoleh 6.510 data yang digunakan dalam proses analisis. Pelabelan sentimen dilakukan berdasarkan rating pengguna, kemudian data diproses melalui tahapan *preprocessing*. Selanjutnya, data diekstraksi menggunakan metode TF-IDF, dibagi menjadi data latih dan data uji dengan rasio 80:20, serta diterapkan teknik SMOTE pada data latih untuk mengatasi ketidakseimbangan kelas. Hasil penelitian menunjukkan bahwa algoritma XGBoost menghasilkan nilai *accuracy* sebesar 84,79% dengan *F1-score* sebesar 0,848, sedangkan algoritma *Random Forest* memperoleh *accuracy* sebesar 84,56% dengan *F1-score* sebesar 0,846. Berdasarkan hasil tersebut, XGBoost menunjukkan performa yang sedikit lebih baik dibandingkan *Random Forest* dalam klasifikasi sentimen ulasan aplikasi JobStreet. Selain itu, hasil analisis menunjukkan bahwa sentimen positif didominasi oleh kemudahan dalam mencari lowongan pekerjaan, sedangkan sentimen negatif lebih banyak berkaitan dengan kendala teknis, seperti masalah login dan kesulitan dalam penggunaan aplikasi.

Kata Kunci: Analisis Sentimen, JobStreet, SMOTE, XGBoost, *Random Forest*.

ABSTRACT

This study aims to analyze the sentiment of JobStreet application user reviews on Google Play Store by applying the Synthetic Minority Over-sampling Technique (SMOTE) and comparing the performance of the XGBoost and Random Forest algorithms. The study is motivated by the large number of user reviews, which generate unstructured textual data, as well as the imbalance between sentiment classes that may reduce the performance of classification models. The data were collected through a web scraping process, resulting in 9,000 reviews. After the data selection stage, 6,510 reviews were retained for analysis. Sentiment labels were assigned based on user ratings, and the data then underwent several preprocessing. Furthermore, the text data were transformed into numerical features using the TF-IDF method. The dataset was divided into training and testing sets with an 80:20 ratio, and SMOTE was applied to the training data to address class imbalance. The experimental results show that the XGBoost algorithm achieved an accuracy of 84.79% with an F1-score of 0.848, while the Random Forest algorithm achieved an accuracy of 84.56% with an F1-score of 0.846. These findings indicate that XGBoost performed slightly better than Random Forest in classifying the sentiment of JobStreet application reviews. In addition, positive reviews were generally associated with the ease of searching for job opportunities, whereas negative reviews mainly highlighted technical issues, such as login problems and application usability.

Keywords: Sentiment Analysis, JobStreet, SMOTE, XGBoost, Random Forest.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini perkembangan teknologi digital telah mengubah proses pencarian kerja menjadi lebih praktis melalui aplikasi berbasis mobile. Salah satu aplikasi yang viral di Indonesia dan banyak digunakan ialah JobStreet, yang menyediakan layanan pencarian lowongan, pengiriman lamaran, pengelolaan profil, hingga rekomendasi pekerjaan. Berdasarkan Google Play Store, aplikasi JobStreet memiliki rating sekitar 4,6, lebih dari 426 ribu ulasan, dan telah diunduh lebih dari 10 juta kali. Banyaknya ulasan tersebut menunjukkan tingginya aktivitas pengguna serta menghasilkan data opini berbentuk teks tidak terstruktur yang dapat dianalisis menggunakan text mining dan Natural Language Processing (NLP). Analisis ulasan pengguna penting dilakukan untuk mengetahui tingkat kepuasan, kritik, dan kendala yang dialami pengguna secara lebih sistematis. Oleh karena itu, penelitian ini berfokus pada ulasan pengguna aplikasi JobStreet dengan output berupa sentimen positif dan negatif terhadap pengalaman penggunaan aplikasi[1].

Dikarenakan ulasan pengguna yang begitu banyak di Google Play Store menyebabkan informasi di dalamnya sulit dianalisis secara manual karena setiap ulasan memuat pengalaman, kritik, saran, maupun keluhan terkait performa aplikasi. Pada aplikasi JobStreet, ulasan tidak hanya berisi kepuasan pengguna terhadap kemudahan mencari pekerjaan, tetapi juga mencakup kendala fitur, proses login, pembaruan profil, hingga kesulitan saat melamar kerja. Selain itu, perbedaan jumlah ulasan positif dan negatif dapat menyebabkan ketidakseimbangan data dalam proses klasifikasi sentimen. Kondisi ini berpengaruh terhadap performa

75 model karena algoritma cenderung lebih fokus mempelajari kelas mayoritas dibandingkan kelas minoritas. Penelitian terdahulu oleh Marzuqi dkk. menganalisis sentimen ulasan JobStreet dengan membandingkan algoritma Naïve Bayes, Support Vector Machine, Logistic Regression, dan Random Forest. Penelitian tersebut menggunakan 5.000 ulasan, menerapkan preprocessing dan TF-IDF, serta menghasilkan akurasi tertinggi pada algoritma Naïve Bayes sebesar 85,19%, sedangkan Random Forest memperoleh akurasi 82,60%[1][2].

262 Meskipun penelitian sebelumnya telah membahas analisis sentimen ulasan JobStreet, masih terdapat peluang pengembangan pada penanganan ketidakseimbangan data dan penggunaan algoritma klasifikasi yang lebih optimal. Penelitian Marzuqi dkk belum menerapkan SMOTE untuk mengatasi ketidakseimbangan data maupun membandingkan performa XGBoost dan Random Forest setelah proses penyeimbangan data dilakukan. Padahal, beberapa penelitian menunjukkan bahwa kombinasi SMOTE dan XGBoost dapat meningkatkan performa klasifikasi, terutama pada kelas minoritas. Oleh karena itu, penelitian ini menerapkan SMOTE dengan algoritma XGBoost dan Random Forest untuk meningkatkan performa klasifikasi sentimen ulasan aplikasi JobStreet[2][3][4].

216 14 Penelitian ini penting karena hasil sentimen dapat menjadi acuan informasi untuk developer jobstreet untuk mengevaluasi kualitas layanan aplikasi berdasarkan pengalaman pengguna secara langsung. Melalui klasifikasi sentimen, pengembang dapat mengetahui fitur yang paling banyak mendapat apresiasi maupun keluhan, seperti proses login, pencarian lowongan, pengiriman lamaran, dan rekomendasi pekerjaan. Selain manfaat praktis tersebut, penelitian ini juga diharapkan berkontribusi secara akademik dalam penerapan metode penanganan

data tidak seimbang pada analisis sentimen berbasis machine learning. Penerapan SMOTE bersama algoritma XGBoost dan Random Forest diharapkan menghasilkan model pengelompokan yang lebih optimal dan seimbang pada setiap golongan sentimen. Selain itu, Berdasarkan observasi awal terhadap 1.000 ulasan aplikasi JobStreet, ditemukan bahwa distribusi sentimen positif lebih dominan dibanding sentimen negatif sehingga berpotensi menyebabkan ketidakseimbangan kelas pada proses klasifikasi. Dengan demikian, penerapan SMOTE bersama algoritma XGBoost dan Random Forest diharapkan mampu menghasilkan klasifikasi sentimen yang lebih seimbang dan akurat[3][4].

Pemilihan aplikasi JobStreet sebagai objek penelitian didasarkan pada perannya sebagai platform pencarian kerja yang digunakan secara luas di Indonesia maupun Asia. Google Play Store dipilih sebagai sumber data karena menyediakan ulasan pengguna dalam jumlah besar, bersifat terbuka, dan memuat pengalaman langsung pengguna terhadap aplikasi. Besarnya jumlah ulasan tersebut menjadikan data relevan untuk dianalisis menggunakan pendekatan text mining dan machine learning karena mampu merepresentasikan pengalaman pengguna secara langsung terhadap kualitas layanan aplikasi. Berdasarkan hal tersebut, penelitian berjudul “Analisis Sentimen Ulasan Aplikasi JobStreet pada Google Play Store Menggunakan SMOTE dengan Algoritma XGBoost dan Random Forest” dipilih untuk mengidentifikasi kecenderungan sentimen pengguna, menangani ketidakseimbangan data, serta mengevaluasi/membandingkan performa algoritma XGBoost dan Random Forest dalam klasifikasi sentimen[1][2].

1.2 Rumusan Masalah

Bersandarkan latar belakang tersebut, disimpulkan melahirkan beberapa pertanyaan yaitu:

1. Bagaimana proses analisis sentimen pada ulasan aplikasi JobStreet di Google Play Store?
2. Bagaimana proses penerapan teknik SMOTE dalam mengatasi ketidakseimbangan data pada proses analisis sentimen ulasan aplikasi JobStreet?
3. Bagaimana hasil performa algoritma XGBoost dan Random Forest setelah di SMOTE dalam melakukan klasifikasi sentimen pada ulasan aplikasi JobStreet berdasarkan accuary, precision, recall dan f1-score?

1.3 Batasan Penelitian

Agar penelitian terarah dan tidak melebar, ruang lingkup dibatasi sebagai berikut:

1. Jumlah data yang digunakan dalam penelitian ini dibatasi 9000 ulasan.
2. Proses pengambilan data dilakukan menggunakan teknik scraping tanpa melibatkan data pribadi pengguna.
3. Analisis hanya ada kelas positif dan negatif.
4. Penelitian ini tidak membahas pengembangan aplikasi atau sistem berbasis antarmuka (UI/UX), melainkan hanya berfokus pada analisis data teks.

1.4 Tujuan Penelitian

Adapun Tujuan penelitian ini dirumuskan berdasarkan permasalahan yang telah ditetapkan agar tetap selaras dan mudah dipahami dalam alur penelitian berikut:

1. Menganalisis sentimen ulasan aplikasi JobStreet di Google Play Store melalui tahapan *preprocessing*.
2. Menerapkan teknik oversampling SMOTE untuk mengatasi ketidakseimbangan data pada ulasan aplikasi Jobstreet.
3. Mengevaluasi algoritma XGBoost dan *Random Forest* dalam klasifikasi sentimen ulasan aplikasi JobStreet.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat yang jelas dan terukur, baik secara teoritis maupun praktis:

1. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *data mining*, *text mining*, dan *machine learning*, khususnya dalam analisis sentimen menggunakan algoritma XGBoost dan *Random Forest*.
2. Menjadi referensi bagi peneliti selanjutnya dalam mengembangkan penelitian analisis sentimen dengan penerapan teknik oversampling SMOTE pada data ulasan aplikasi digital.
3. Memberikan informasi mengenai performa algoritma XGBoost dan *Random Forest* dalam klasifikasi sentimen sebagai bahan pertimbangan dalam pemilihan metode *machine learning* pada penelitian serupa.

1.6 Sistematika Penelitian

Penulisan penelitian ini disusun secara sistematis untuk memastikan alur pembahasan yang kuat dan runtut, terdapat 5 bab utama yang saling terhubung:

BAB I PENDAHULUAN

Menampilkan latar belakang penelitian yang berkaitan dengan perkembangan platform pencarian kerja berbasis digital, pentingnya peran ulasan pengguna dalam mengevaluasi kualitas layanan aplikasi, serta permasalahan yang muncul dalam proses analisis sentimen pada review aplikasi JobStreet di Google Play Store. Selain itu, pada bab ini juga dijelaskan rumusan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan yang digunakan sebagai pedoman dalam penyusunan skripsi.

BAB II TINJAUAN LITERATUR

Menguraikan berbagai teori dan konsep yang menjadi landasan penelitian, meliputi analisis sentimen, text mining, tahapan preprocessing data teks, metode pembobotan TF-IDF, teknik oversampling SMOTE, algoritma XGBoost, algoritma Random Forest, serta metode evaluasi model klasifikasi. Selain itu, bab ini juga memuat kajian terhadap penelitian terdahulu yang relevan sebagai referensi pendukung dan dasar dalam penyusunan kerangka pemikiran penelitian.

BAB III METODE PENELITIAN

Cara/tahapan metode penelitian yang diterapkan, dimulai dari proses pengumpulan data ulasan aplikasi JobStreet pada Google Play Store, preprocessing data teks, proses pelabelan sentimen, transformasi data menggunakan metode TF-IDF, pembagian dataset, penerapan teknik SMOTE untuk menangani ketidakseimbangan data, pelatihan model menggunakan algoritma XGBoost dan

Random Forest, hingga tahap evaluasi performa model menggunakan confusion matrix, accuracy, precision, recall, dan F1-score.

BAB IV HASIL PENELITIAN

Bab ini menyajikan hasil penelitian beserta pembahasannya berdasarkan seluruh tahapan yang telah dilakukan. Pembahasan mencakup hasil pengumpulan data ulasan, tahapan preprocessing, distribusi data sebelum dan sesudah penerapan SMOTE, hasil pengujian model menggunakan algoritma XGBoost dan Random Forest, evaluasi performa model, serta analisis perbandingan kinerja kedua algoritma dalam melakukan klasifikasi sentimen pada ulasan aplikasi JobStreet.

BAB V PENUTUP

berupa kesimpulan hasil penelitian berdasarkan rumusan masalah dan tujuan penelitian yang telah ditetapkan. Dan juga saran untuk penelitian selanjutnya agar hasil yang diperoleh dapat menjadi lebih optimal serta memiliki cakupan penerapan yang lebih luas.

BAB II

TINJAUAN LITERATUR

Bab ini membahas landasan teori yang digunakan sebagai dasar dalam pelaksanaan penelitian. Teori-teori yang disajikan meliputi konsep Natural Language Processing (NLP), teknik preprocessing teks, metode ekstraksi fitur menggunakan TF-IDF, SMOTE, serta algoritma klasifikasi XGBoost dan random forest. Selain itu, pada bab ini juga dibahas penelitian terdahulu yang relevan sebagai referensi dan pembandingan dalam mendukung penelitian yang dilakukan. Dengan adanya tinjauan pustaka ini, diharapkan penelitian memiliki dasar teoritis yang kuat serta arah yang jelas dalam proses analisis dan pembahasan.

2.1 Aplikasi JobStreet

JobStreet adalah aplikasi pencarian kerja yang cukup familiar di Indonesia yang telah beroperasi lebih dari 20 tahun. Aplikasi ini menyediakan layanan pencarian lowongan pekerjaan, pengiriman lamaran, pengelolaan profil, hingga rekomendasi pekerjaan berbasis kecerdasan buatan. JobStreet beroperasi di beberapa negara Asia, termasuk Indonesia, Malaysia, Singapura, Filipina, dan Thailand, serta tergabung dalam ekosistem SEEK, grup perusahaan lapangan kerja online terbesar di kawasan Asia Pasifik.

Berdasarkan data resmi dari laman Jobstreet, platform ini telah terhubung dengan lebih dari 42 juta kandidat dan 1,35 juta perusahaan di Asia. Sementara itu, berbagai sumber pemberitaan menyebutkan bahwa JobStreet memiliki lebih dari 40 juta pengguna secara global. Di Indonesia secara khusus, Jobstreet by SEEK memiliki lebih dari 20 juta kandidat terdaftar dan lebih dari setengah juta perusahaan yang bergabung, mulai dari korporasi besar hingga pelaku UMKM.

Platform ini juga telah mengiklankan lebih dari 1 juta lowongan pekerjaan dalam satu tahun terakhir melalui Gerakan #NextMillionJobs.

Tingkat efektivitas JobStreet dalam membantu pencari kerja juga tercermin dari data yang menunjukkan bahwa 1 dari 3 pencari kerja di Indonesia berhasil mendapatkan pekerjaan melalui platform ini. Selain itu, berdasarkan survei independen, Jobstreet by SEEK menguasai pangsa pasar tiga kali lipat dibandingkan pesaing terdekatnya di Indonesia. Data ini menunjukkan bahwa JobStreet bukan hanya aplikasi dengan jumlah ulasan terbanyak di Google Play Store, tetapi juga merupakan platform yang sangat vital dalam ekosistem ketenagakerjaan digital di Indonesia. Hal ini memperkuat urgensi dan relevansi penelitian analisis sentimen terhadap ulasan penggunanya.

2.2 Analisis Sentimen

suatu tahapan yang dilakukan untuk mendeteksi, mengolah, serta mengelompokkan opini atau pendapat yang terdapat dalam data berbentuk teks ke dalam kategori sentimen tertentu, seperti positif, negatif, maupun netral. Pada penelitian ini, analisis sentimen dimanfaatkan untuk mengetahui kecenderungan penilaian pengguna terhadap aplikasi JobStreet berdasarkan ulasan yang diberikan melalui Google Play Store[4][5].

Ulasan tersebut umumnya memuat berbagai bentuk tanggapan pengguna, mulai dari pengalaman yang memuaskan, kritik terhadap fitur aplikasi, keluhan mengenai layanan, hingga saran untuk pengembangan sistem. Melalui analisis sentimen, data teks yang awalnya tidak terstruktur dapat diolah menjadi informasi yang lebih sistematis dan mudah dipahami sehingga dapat digunakan sebagai bahan evaluasi terhadap kualitas aplikasi.[23]

2.3 Text Mining

Text mining merupakan proses pengolahan data berbasis teks yang bertujuan untuk menemukan informasi, pola, maupun pengetahuan tertentu dari sekumpulan dokumen. Pada penelitian ini, text mining dimanfaatkan untuk mengubah ulasan pengguna aplikasi JobStreet yang masih berupa data teks menjadi bentuk data relasional sehingga dapat diproses menggunakan algoritma machine learning.

Secara umum, tahapan text mining yang diterapkan dalam penelitian ini mencakup proses pengumpulan data, preprocessing teks, ekstraksi fitur, penanganan ketidakseimbangan data, proses klasifikasi, hingga evaluasi performa model yang digunakan. Melalui tahapan tersebut, data ulasan dapat dianalisis secara lebih sistematis untuk menghasilkan informasi yang relevan terhadap sentimen pengguna aplikasi JobStreet[5][6].

2.4 Pelabelan Sentimen

Pelabelan sentimen merupakan proses pemberian kategori sentimen pada setiap data ulasan yang dianalisis. Kategori sentimen tersebut umumnya dibedakan menjadi tiga kelas, yaitu positif, negatif, dan netral. Pada ulasan di Google Play Store, tahapan pelabelan dapat dijalankan berlandaskan nilai rating yang diberikan pengguna maupun melalui penentuan label secara manual. Dalam penelitian ini, rating digunakan sebagai acuan awal dalam proses pelabelan, di mana rating 1-2 dikategorikan sebagai sentimen negatif, rating 3 sebagai sentimen netral, dan rating 4-5 sebagai sentimen positif. Mesk demikian, proses uji secara manual tetap dilakukan untuk memastikan bahwa label sentimen yang diberikan telah sesuai dengan isi dan makna dari ulasan pengguna aplikasi JobStreet[2][5].

2.5 Preprocessing Teks

Preprocessing teks merupakan tahap awal dalam pengolahan data berbasis teks yang dilakukan untuk membersihkan serta mempersiapkan data sebelum masuk ke proses ekstraksi fitur dan klasifikasi. Tahapan ini memiliki peran penting karena ulasan pengguna umumnya masih mengandung berbagai elemen yang tidak relevan, seperti penggunaan huruf kapital, tanda baca, angka, emoji, URL, kata tidak baku, hingga kata-kata universal yang kurang memiliki nilai. Pada penelitian ini, proses preprocessing yang diterapkan meliputi case folding, cleaning, tokenizing, normalisasi kata, stopwords removal, dan stemming[5]. Melalui tahapan tersebut, data teks diharapkan menjadi lebih terstruktur dan lebih optimal untuk diproses dalam analisis sentimen.

Adapun tahapan preprocessing yang digunakan dalam penelitian ini:

1. Cleaning adalah proses membersihkan teks dari karakter yang tidak diperlukan.
2. Case folding adalah proses mengubah seluruh huruf menjadi huruf kecil.
3. Normalisasi adalah proses mengubah kata tidak baku atau singkatan menjadi kata baku.
4. Stopword removal adalah proses menghapus kata-kata umum yang kurang memiliki makna penting dalam proses klasifikasi.
5. Stemming adalah proses mengubah kata berimbuhan menjadi bentuk kata dasar.

2.6 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan metode pembobotan kata yang digunakan untuk mengukur tingkat pentingnya suatu kata dalam sebuah dokumen. Metode ini memberikan nilai bobot tinggi pada kata yang

sering muncul dalam satu dokumen, tetapi memiliki frekuensi rendah pada dokumen lainnya. Dalam library Scikit-learn, proses tersebut dapat dilakukan menggunakan TfidfVectorizer, yaitu fitur yang berfungsi mengubah kumpulan dokumen teks menjadi matriks berbasis bobot TF-IDF[7][5].

2.7 SMOTE

SMOTE atau *Synthetic Minority Over-sampling Technique* merupakan metode oversampling yang digunakan untuk mengatasi permasalahan ketidakseimbangan data pada proses klasifikasi. Berbeda dengan metode oversampling konvensional yang hanya menduplikasi data pada kelas minoritas, SMOTE menghasilkan data sintetis baru berdasarkan karakteristik sampel dari kelas minoritas tersebut. Pendekatan ini membuat distribusi data menjadi lebih seimbang sehingga model dapat mempelajari pola pada kelas minoritas dengan lebih optimal[8][9][10].

2.8 XGBoost

XGBoost atau *Extreme Gradient Boosting* merupakan algoritma *machine learning* yang dikembangkan berdasarkan metode *gradient boosting* dengan proses pembentukan model secara bertahap. Setiap model yang dibangun bertujuan untuk memperbaiki kesalahan prediksi dari model sebelumnya sehingga kinerja prediksi dapat meningkat secara optimal. XGBoost dikenal memiliki efisiensi komputasi yang baik, mendukung fitur regularisasi untuk mengurangi risiko *overfitting*, serta mampu menangani data dalam jumlah besar dan berdimensi tinggi[4][11][12].

2.9 Random Forest

Random Forest merupakan algoritma *machine learning* berbasis ensemble learning yang bekerja dengan membangun banyak *decision tree* untuk

menghasilkan prediksi akhir. Setiap pohon dibentuk dari sampel data dan fitur yang dipilih secara acak, kemudian hasil seluruh pohon digabungkan melalui voting untuk menentukan kelas akhir. Algoritma ini banyak digunakan dalam klasifikasi karena mampu menangani data berdimensi tinggi, stabil, dan dapat mengurangi risiko overfitting dibandingkan decision tree tunggal[2][5]. Dalam penelitian ini, Random Forest digunakan untuk mengklasifikasikan sentimen ulasan aplikasi JobStreet.

2.10 Confusion Matrix

Confusion matrix merupakan tabel evaluasi yang digunakan untuk menampilkan hasil prediksi dari suatu model klasifikasi dengan membandingkan label aktual dan label hasil prediksi model kesalahan prediksi[13].

2.11 Evaluasi Model

Evaluasi model dilakukan untuk mengukur tingkat kinerja algoritma dalam proses klasifikasi sentimen. Pada penelitian ini, metrik evaluasi yang digunakan meliputi accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk menunjukkan persentase jumlah prediksi yang benar dibandingkan dengan seluruh data yang diuji. Precision menggambarkan tingkat ketepatan model dalam memprediksi suatu kelas tertentu. Recall digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh data yang termasuk ke dalam kelas tertentu. Sementara itu, F1-score merupakan nilai rata-rata harmonik antara precision dan recall, sehingga metrik ini dinilai efektif untuk digunakan pada data dengan distribusi kelas yang tidak seimbang[14][15].

2.12 Penelitian Terdahulu

Berikut Tabel 2.1 menyajikan Penelitian sebelumnya dijadikan sebagai acuan dalam menentukan metode dan pendekatan yang digunakan pada analisis sentimen.

Tabel 2. 1 Penelitian Terdahulu

No	Artikel	Judul	Hasil Penelitian
(1)	(2)	(3)	(4)
1	[2]	Perbandingan Algoritma Machine Learning untuk Sentimen Ulasan Aplikasi JobStreet	Penelitian ini menganalisis sentimen ulasan aplikasi JobStreet di Google Play Store menggunakan algoritma Naïve Bayes, SVM, Logistic Regression, dan Random Forest. Penelitian tersebut relevan karena menggunakan objek yang sama, yaitu ulasan aplikasi JobStreet. Hasil penelitian menunjukkan bahwa Naïve Bayes memperoleh tingkat akurasi tertinggi dibandingkan algoritma lainnya.
2	[16]	Sentiment Analysis of Jobstreet Application Reviews on Google Play Store Using Support Vector Machine Algorithm with Adaptive Synthetic	Penelitian lain membahas analisis sentimen ulasan aplikasi JobStreet menggunakan algoritma SVM dengan teknik oversampling ADASYN. Hasil penelitian menunjukkan bahwa penerapan ADASYN mampu meningkatkan akurasi SVM dari 89,08% menjadi 89,95%. Penelitian ini relevan karena sama-sama menggunakan objek JobStreet serta membahas penanganan data tidak seimbang.
3	[17]	Analisis Sentimen Negatif pada Aplikasi JobStreet Menggunakan Hadoop Distributed File System (HDFS)	Penelitian ini berfokus pada analisis sentimen negatif pada aplikasi JobStreet. Metode yang digunakan berkaitan dengan pengolahan data ulasan dan klasifikasi sentimen. Penelitian ini relevan karena sama-sama menggunakan objek aplikasi JobStreet, tetapi berbeda pada fokus dan metode penelitian.

(1)	(2)	(3)	(4)
4	[4]	Analisis Sentimen Ulasan Pengguna Aplikasi Threads di Google Play Menggunakan Algoritma XGBoost dengan Penguatan SMOTE	Penelitian ini menggunakan algoritma XGBoost dengan teknik SMOTE pada ulasan aplikasi Threads di Google Play. Data direpresentasikan menggunakan TF-IDF dan dianalisis menggunakan XGBoost sebelum dan sesudah penerapan SMOTE. Hasil penelitian menunjukkan bahwa SMOTE mampu meningkatkan performa model, terutama pada nilai recall kelas negatif. Penelitian ini relevan karena sama-sama menggunakan metode XGBoost dan SMOTE.
5	[3]	Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers	Penelitian ini membandingkan algoritma Logistic Regression, SVM, XGBoost, Naïve Bayes, dan Random Forest pada data ulasan Google Reviews Puskesmas di Bali. Penelitian menggunakan metode TF-IDF dan SMOTE untuk mengatasi ketidakseimbangan data. Penelitian ini relevan karena sama-sama menerapkan TF-IDF, SMOTE, XGBoost, dan Random Forest.
6	[5]	Analisis Sentimen Review Aplikasi di Google Play Store Menggunakan Random Forest	Penelitian ini menggunakan algoritma Random Forest untuk mengklasifikasikan ulasan aplikasi di Google Play Store ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Proses klasifikasi dilakukan berdasarkan isi ulasan pengguna untuk mengetahui kecenderungan opini terhadap aplikasi yang dianalisis. Penelitian ini relevan karena sama-sama menggunakan data ulasan dari Google Play Store serta menerapkan algoritma Random Forest dalam analisis sentimen.

(1)	(2)	(3)	(4)
7	[18]	Analisis Sentimen Aplikasi M-Pajak Menggunakan Machine Learning Berbasis Algoritma XGBoost	Penelitian ini menganalisis sentimen ulasan pengguna aplikasi M-Pajak di Google Play Store menggunakan algoritma XGBoost. Proses analisis dilakukan untuk mengidentifikasi kecenderungan opini pengguna terhadap aplikasi berdasarkan ulasan yang diberikan. Penelitian ini relevan karena sama-sama menerapkan algoritma XGBoost dalam analisis sentimen ulasan aplikasi.
8	[19]	Analisis Sentimen Aplikasi Halo BCA di Google Play Store Menggunakan Metode Naive Bayes, Support Vector Machine dan Random Forest	Penelitian ini menganalisis ulasan aplikasi Halo BCA di Google Play Store menggunakan algoritma Naïve Bayes, SVM, dan Random Forest. Hasil penelitian menunjukkan bahwa Random Forest memiliki performa terbaik dibandingkan algoritma lainnya dalam klasifikasi sentimen. Penelitian ini relevan karena sama-sama menggunakan data ulasan dari Google Play Store serta menerapkan algoritma Random Forest.
9	[20]	Analisis Sentimen Traveloka Berdasarkan Ulasan Google Play Store Menggunakan Algoritma Support Vector Machine dan Random Forest	Penelitian ini menganalisis sentimen ulasan aplikasi Traveloka di Google Play Store menggunakan algoritma SVM dan Random Forest. Analisis dilakukan untuk mengetahui kecenderungan opini pengguna terhadap aplikasi berdasarkan ulasan yang diberikan. Penelitian ini relevan karena sama-sama membahas analisis sentimen ulasan aplikasi serta menggunakan Random Forest sebagai salah satu algoritma klasifikasi.

(1)	(2)	(3)	(4)
10	[21]	Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models	Penelitian ini menganalisis sentimen ulasan Tokopedia menggunakan TF-IDF, SMOTE, dan beberapa model machine learning, termasuk Random Forest. Penelitian ini relevan karena menggunakan TF-IDF dan SMOTE dalam proses klasifikasi sentimen..

Berdasarkan Tabel 2.1 Penelitian Terdahulu, dapat diketahui bahwa penelitian mengenai analisis sentimen pada ulasan aplikasi telah banyak dilakukan dengan menggunakan berbagai objek dan metode yang berbeda. Sejumlah penelitian memanfaatkan ulasan pengguna di Google Play Store sebagai sumber data, seperti pada aplikasi JobStreet, Threads, Halo BCA, Traveloka, dan M-Pajak. Kondisi tersebut menunjukkan bahwa ulasan pengguna pada platform digital memiliki potensi yang besar untuk dimanfaatkan dalam mengidentifikasi opini, persepsi, dan pengalaman pengguna terhadap suatu aplikasi.

Penelitian yang memiliki keterkaitan paling kuat dengan penelitian ini adalah penelitian yang dilakukan oleh Marzuqi dkk. serta Shantika dan Abidin karena sama-sama menggunakan ulasan aplikasi JobStreet di Google Play Store sebagai objek penelitian. Marzuqi dkk. melakukan perbandingan beberapa algoritma machine learning, termasuk Random Forest, sedangkan Shantika dan Abidin menerapkan algoritma SVM dengan teknik oversampling ADASYN. Meskipun demikian, kedua penelitian tersebut belum mengombinasikan teknik SMOTE dengan algoritma XGBoost dan Random Forest pada analisis sentimen ulasan aplikasi JobStreet.

Di sisi lain, penelitian yang dilakukan oleh Setiyono dkk serta Budaya dan Suniantara menunjukkan bahwa teknik SMOTE efektif digunakan untuk mengatasi permasalahan ketidakseimbangan data pada analisis sentimen. Hasil penelitian mereka juga membuktikan bahwa algoritma XGBoost dan *Random Forest* mampu memberikan performa yang baik dalam proses klasifikasi sentimen berbasis ulasan pengguna. Namun, objek penelitian yang digunakan berbeda dengan penelitian ini, yaitu aplikasi Threads dan Google Reviews Puskesmas. Oleh karena itu, penelitian ini menawarkan kebaruan melalui penerapan kombinasi SMOTE dengan algoritma XGBoost dan *Random Forest* pada ulasan aplikasi JobStreet di Google Play Store.

Berdasarkan uraian tersebut, penelitian ini berupaya melanjutkan sekaligus mengembangkan penelitian terdahulu dengan mengintegrasikan beberapa pendekatan yang relevan. Penelitian menggunakan data ulasan aplikasi JobStreet, melakukan pelabelan sentimen positif dan negatif, menerapkan tahapan *preprocessing* teks, memanfaatkan TF-IDF sebagai metode ekstraksi fitur, serta menggunakan SMOTE untuk mengatasi ketidakseimbangan data. Selanjutnya, proses klasifikasi sentimen dilakukan dengan algoritma XGBoost dan *Random Forest*. Melalui pendekatan tersebut, penelitian ini diharapkan dapat mengidentifikasi algoritma yang memiliki performa terbaik berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*.

Berdasarkan hasil kajian penelitian terdahulu, dapat disimpulkan bahwa analisis sentimen terhadap ulasan aplikasi JobStreet telah banyak dilakukan, namun masih terdapat peluang pengembangan terutama dalam aspek penanganan ketidakseimbangan data dan pemilihan algoritma klasifikasi. Penelitian sebelumnya pada objek JobStreet umumnya menggunakan algoritma seperti *Naïve*

Bayes, SVM, Logistic Regression, Random Forest, serta teknik oversampling ADASYN. Sementara itu, penerapan kombinasi SMOTE dengan algoritma XGBoost dan Random Forest pada analisis sentimen ulasan aplikasi JobStreet masih jarang ditemukan dalam penelitian yang telah ada. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan menerapkan teknik SMOTE guna mengatasi ketidakseimbangan data serta membandingkan performa algoritma XGBoost dan Random Forest dalam proses klasifikasi sentimen ulasan aplikasi JobStreet.

2.13 Rangkuman

Secara umum, penelitian terdahulu, Berdasarkan penelitian terdahulu yang telah dikaji, dapat diketahui bahwa analisis sentimen pada ulasan aplikasi telah banyak dilakukan dengan memanfaatkan berbagai algoritma machine learning, seperti Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, dan XGBoost. Selain itu, beberapa penelitian juga telah menerapkan teknik oversampling, seperti SMOTE dan ADASYN, untuk mengatasi permasalahan ketidakseimbangan data. Penelitian yang menggunakan ulasan aplikasi JobStreet sebagai objek kajian juga telah dilakukan, namun sebagian besar masih berfokus pada penggunaan algoritma tertentu dan belum mengombinasikan teknik SMOTE dengan algoritma XGBoost dan Random Forest dalam satu penelitian.

Penelitian yang dilakukan oleh Marzuqi dkk dan Shantika & Abidin menggunakan objek yang sama, yaitu ulasan aplikasi JobStreet pada Google Play Store, tetapi metode yang diterapkan berbeda dengan penelitian ini. Di sisi lain, penelitian Setiyono dkk serta Budaya dan Suniantara menunjukkan bahwa penerapan SMOTE mampu meningkatkan kinerja model klasifikasi pada data yang

tidak seimbang. Meskipun demikian, perbedaan objek penelitian menyebabkan hasil yang diperoleh belum dapat menggambarkan kondisi sentimen pengguna aplikasi JobStreet secara spesifik.

Berdasarkan kajian tersebut, masih terdapat peluang untuk mengembangkan penelitian dengan menerapkan teknik SMOTE pada data ulasan aplikasi JobStreet yang berpotensi memiliki distribusi kelas sentimen yang tidak seimbang. Selain itu, penggunaan algoritma XGBoost dan *Random Forest* pada objek penelitian yang sama memungkinkan dilakukan perbandingan performa untuk mengetahui algoritma yang lebih efektif dalam mengklasifikasikan sentimen pengguna.

Secara teoritis, penelitian ini didasarkan pada konsep analisis sentimen yang merupakan bagian dari *text mining* untuk mengidentifikasi opini pengguna berdasarkan data berbentuk teks. Sebelum dilakukan analisis, data ulasan terlebih dahulu melalui tahapan *preprocessing* yang meliputi *cleaning*, *case folding*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Tahapan tersebut bertujuan untuk membersihkan serta menstandarkan data sehingga lebih siap digunakan pada proses pengolahan berikutnya.

Data yang telah melalui proses *preprocessing* kemudian dikonversi ke dalam bentuk numerik menggunakan metode TF-IDF. Selanjutnya, teknik SMOTE diterapkan untuk menyeimbangkan distribusi data sentimen sebelum dilakukan proses klasifikasi. Tahap klasifikasi menggunakan dua algoritma *machine learning*, yaitu XGBoost dan *Random Forest*. Hasil klasifikasi kemudian dievaluasi menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan model yang memiliki performa terbaik.

41 Berdasarkan penelitian terdahulu dan landasan teori yang digunakan,
29 kebaruan penelitian ini terletak pada penerapan teknik SMOTE yang
37 dikombinasikan dengan algoritma XGBoost dan *Random Forest* untuk analisis
18 sentimen ulasan aplikasi JobStreet pada Google Play Store. Penelitian ini
diharapkan dapat memberikan gambaran mengenai sentimen pengguna terhadap
aplikasi JobStreet sekaligus mengidentifikasi algoritma yang memiliki performa
paling optimal dalam proses klasifikasi sentimen.

BAB III

METODE PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian analisis sentimen ulasan aplikasi JobStreet pada Google Play Store menggunakan SMOTE dengan algoritma XGBoost dan Random Forest. Metodologi penelitian disusun secara sistematis mulai dari jenis penelitian, sumber data, pengumpulan data, preprocessing teks, ekstraksi fitur TF-IDF, penerapan SMOTE, proses klasifikasi menggunakan XGBoost dan Random Forest, hingga evaluasi model.

3.1 Jenis Penelitian

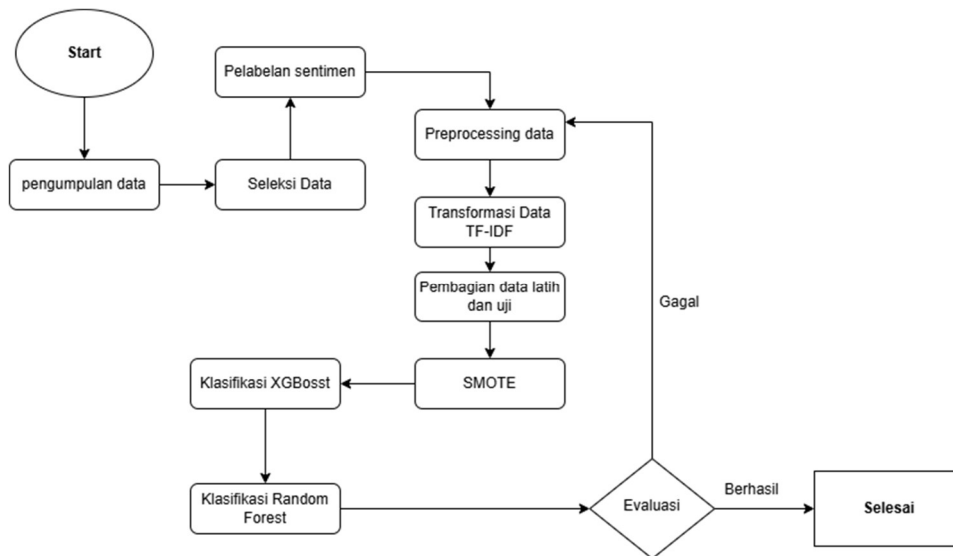
Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen. Pendekatan kuantitatif digunakan karena data ulasan pengguna diolah dalam bentuk angka melalui proses ekstraksi fitur, sedangkan pendekatan eksperimen digunakan untuk menerapkan teknik SMOTE serta algoritma XGBoost dan Random Forest dalam klasifikasi sentimen.

Penelitian ini bertujuan menganalisis sentimen ulasan aplikasi JobStreet pada Google Play Store ke dalam kelas positif dan negatif, serta membandingkan performa algoritma XGBoost dan Random Forest berdasarkan nilai accuracy, precision, recall, dan F1-score.

3.2 Kerangka Penelitian

menggambar alur penelitian secara sistematis, mulai dari pengumpulan data, pelabelan sentimen, preprocessing, transformasi data menggunakan TF-IDF, pembagian data latih dan data uji, penerapan SMOTE, hingga klasifikasi menggunakan algoritma XGBoost dan Random Forest. Selanjutnya, hasil

klasifikasi dievaluasi untuk mengetahui performa masing-masing algoritma.



Gambar 3. 1 Kerangka Penelitian

3.3 Teknik Pengumpulan Data

Sebelum menjelaskan teknik pengumpulan data, berikut disampaikan gambaran umum mengenai objek penelitian, yaitu aplikasi JobStreet, beserta panduan penggunaannya oleh pengguna, sebagai konteks terhadap data ulasan yang menjadi bahan analisis dalam penelitian ini. JobStreet merupakan aplikasi pencarian kerja yang mempertemukan pencari kerja dengan perusahaan penyedia lowongan melalui layanan pencarian lowongan, pengiriman lamaran, dan pengelolaan profil secara daring.

Secara umum, panduan penggunaan aplikasi JobStreet oleh pengguna meliputi tahapan sebagai berikut:

1. Mengunduh dan memasang (*install*) aplikasi JobStreet melalui Google Play Store pada perangkat pengguna.
2. Melakukan pendaftaran akun (*registrasi*) atau masuk (*login*) menggunakan alamat email maupun akun media sosial.

3. Melengkapi profil pengguna, seperti data diri, riwayat pendidikan, dan pengalaman kerja, serta mengunggah *curriculum vitae* (CV).
4. Melakukan pencarian lowongan pekerjaan berdasarkan kata kunci, lokasi, maupun kategori bidang pekerjaan yang diinginkan.
5. Memilih lowongan yang sesuai dan mengirimkan lamaran kerja (*apply*) secara langsung melalui aplikasi.
6. Memantau status lamaran serta menerima notifikasi terkait perkembangan proses rekrutmen.
7. Memberikan ulasan dan rating terhadap aplikasi JobStreet melalui Google Play Store berdasarkan pengalaman penggunaan.

Tahapan penggunaan aplikasi tersebut menjadi dasar bagi peneliti dalam memahami konteks ulasan yang diberikan pengguna, baik yang berkaitan dengan kemudahan pencarian kerja maupun kendala teknis yang dialami selama menggunakan aplikasi

Teknik pengumpulan data dilakukan dengan mengambil ulasan pengguna aplikasi JobStreet dari Google Play Store menggunakan teknik web scraping dengan bantuan library Python, seperti google-play-scraper.

Data yang dikumpulkan berupa teks ulasan, rating, dan tanggal ulasan. Setelah itu, dilakukan seleksi data untuk menghapus data kosong, duplikat, dan ulasan yang tidak relevan.

Tabel 3. 1 Format data awal

No	Review	Rating	Tanggal
(1)	(2)	(3)	(4)
1	Aplikasi Sangat Membantu Kerja	5	2026
2	Susah Login	2	2025....

3.4 Seleksi Data

Seleksi data dilakukan untuk memastikan data yang digunakan sesuai dengan kebutuhan penelitian. Data yang tidak lengkap, kosong, duplikat, atau tidak menggunakan bahasa Indonesia dapat dihapus.

Tahaan seleksi data meliputi:

Tabel 3. 2 Tahapan Seleksi Data

Tahapan (1)	Keterangan (2)
Menghapus data kosong	Menghapus ulasan yang tidak memiliki isi teks
Menghapus duplikat	Menghapus ulasan yang sama agar tidak memengaruhi hasil model
Memilih bahasa Indonesia	Memilih ulasan berbahasa Indonesia
Memilih atribut penting	Menggunakan atribut review, rating, tanggal, dan label

3.5 Pelabelan Sentimen

Pelabelan sentimen dilakukan untuk menentukan kategori sentimen dari setiap ulasan. Pada penelitian ini, label sentimen dapat dibagi menjadi 2 kategori, positif dan negatif. Pelabelan awal dapat dilakukan berdasarkan rating pengguna sebagai berikut:

Tabel 3. 3 Pelabelan Awal

Rating (1)	Label Sentimen (2)
1-3	Negatif
4-5	Positif

Tahap ini bertujuan untuk memastikan kualitas data yang digunakan dalam penelitian.

3.6 Preprocessing Teks

Preprocessing data merupakan tahap pengolahan teks agar data ulasan menjadi lebih bersih dan siap digunakan dalam proses klasifikasi. Data ulasan dari

Google Play Store umumnya masih mengandung huruf kapital, tanda baca, angka, emoji, URL, simbol, dan kata tidak baku. Berikut preprocessing dalam penelitian ini meliputi:

Tabel 3. 4 Tahapan Proprocessing

No	Tahapan	Keterangan
(1)	(2)	(3)
1	Case folding	Mengubah seluruh huruf menjadi huruf kecil
2	Cleaning	Menghapus angka, tanda baca, simbol, emoji, URL, dan karakter tidak penting
3	Tokenizing	Memecah kalimat menjadi kata atau token
5	Stopword removal	Menghapus kata umum yang kurang bermakna
6	Stemming	Mengubah kata berimbuhan menjadi kata dasar

3.7 Ekstraksi Fitur Menggunakan TF-IDF

Setelah preprocessing selesai, data teks diubah ke dalam bentuk numerik agar dapat diproses oleh algoritma machine learning. Pada penelitian ini, metode yang digunakan adalah TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF digunakan untuk memberikan bobot pada kata berdasarkan frekuensi kemunculan dan tingkat kepentingannya dalam dokumen. Semakin penting suatu kata dalam membedakan dokumen, maka semakin besar bobot yang diberikan. Hasil proses TF-IDF berupa matriks numerik yang digunakan sebagai input bagi algoritma XGBoost dan Random Forest. Dokumentasi Scikit-learn menjelaskan bahwa TfidfVectorizer digunakan untuk mengubah kumpulan dokumen menjadi matriks fitur TF-IDF[7].

3.8 Pembagian Data Training dan Testing

Data yang telah melalui proses ekstraksi fitur kemudian dibagi menjadi data training dan data testing. Data training digunakan untuk melatih model, sedangkan data testing digunakan untuk menguji performa model pada data yang belum pernah digunakan sebelumnya. Pembagian data dapat menggunakan skenario berikut :

Tabel 3. 5 Skenario Pembagian Data Training dan Testing

Skenario	Data Training	Data Testing
(1)	(2)	(3)
1	80%	20%

Dalam penelitian ini, data dibagi menjadi 80% data training dan 20% data testing. Pembagian ini umum digunakan karena menyediakan data training yang cukup untuk proses pembelajaran model serta data testing untuk evaluasi performa.

3.9 Penerapan SMOTE

SMOTE (Synthetic Minority Over-sampling Technique) digunakan untuk mengatasi ketidakseimbangan data, yaitu kondisi ketika jumlah data pada setiap kelas sentimen tidak seimbang, seperti kelas positif yang lebih banyak dibandingkan kelas negatif. SMOTE bekerja dengan menghasilkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang. Dokumentasi imbalanced-learn menjelaskan bahwa SMOTE merupakan teknik oversampling untuk menghasilkan sampel sintetis pada kelas minoritas[9].

Dalam penelitian ini, SMOTE diterapkan hanya pada data training agar data testing Pembagian dilakukan dengan rasio tertentu, misalnya 80:20. tetap merepresentasikan kondisi data asli dan menghindari kebocoran data. Contoh distrubsi SMOTE:

Tabel 3. 6 Distribusi Data Sebelum dan Sesudah Penerapan SMOTE

Sentimen	Sebelum SMOTE	Sesudah SMOTE
(1)	(2)	(3)
Positif	2.000	2.000
Negatif	500	2.000

3.10 Klasifikasi Menggunakan XGBoost

XGBoost (Extreme Gradient Boosting) merupakan algoritma machine learning berbasis gradient boosting yang bekerja dengan membangun model secara

bertahap untuk memperbaiki kesalahan model sebelumnya. Dalam penelitian ini, XGBoost digunakan untuk mengklasifikasikan ulasan aplikasi JobStreet ke dalam kelas sentimen positif dan negatif menggunakan data training yang telah diseimbangkan dengan SMOTE.

implementasinya dilakukan melalui library XGBoost Python dengan parameter bawaan (`n_estimators=100`, `learning_rate=0.3`, `max_depth=6`, `subsample=1.0`), karena fokus penelitian adalah membandingkan performa XGBoost dengan Random Forest setelah SMOTE, bukan mengoptimalkan parameter.

Model dilatih menggunakan data yang telah diproses dengan TF-IDF dan SMOTE, melalui fungsi `XGBClassifier()` dan metode `.fit()`. Setelah pelatihan, model diuji dengan data uji dan dievaluasi menggunakan *confusion matrix* serta *classification report* untuk memperoleh nilai akurasi, presisi, *recall*, dan F1-score

Dokumentasi XGBoost menjelaskan bahwa XGBoost merupakan library gradient boosting yang dirancang secara efisien, fleksibel, dan portabel untuk berbagai tugas machine learning[22].

Tabel 3. 7 Tahapan Klasifikasi Menggunakan XGBoost

Tahapan	Keterangan
(1)	(2)
Input	Data training hasil TF-IDF dan SMOTE
Proses	Pelatihan model XGBoost
Output	Prediksi sentimen positif dan negatif
Evaluasi	Accuracy, precision, recall, F1-score, confusion matrix

3.11 Klasifikasi Menggunakan Random Forest

Random Forest merupakan algoritma ensemble learning yang membangun banyak decision tree untuk menghasilkan prediksi akhir. Setiap tree memberikan

prediksi, kemudian hasil akhir ditentukan berdasarkan voting. Dalam penelitian ini, Random Forest digunakan untuk mengklasifikasikan sentimen ulasan aplikasi JobStreet menggunakan data training yang telah melalui proses TF-IDF dan SMOTE.

Implementasinya menggunakan Scikit-learn versi 1.5.0 dengan parameter bawaan (`n_estimators=100`, `max_depth=None`, `min_samples_split=2`, `min_samples_leaf=1`), karena fokus penelitian adalah perbandingan performa algoritma, bukan optimasi parameter. Hasil yang diperoleh merepresentasikan performa dasar Random Forest pada data ulasan JobStreet.

Pelatihan dilakukan dengan `RandomForestClassifier()` dan metode `.fit()` pada data yang telah diproses TF-IDF dan SMOTE. Setelah itu, model diuji dengan data uji. Evaluasi performa menggunakan *confusion matrix* dan *classification report* untuk mengukur akurasi, presisi, *recall*, dan F1-score

Dokumentasi Scikit-learn menjelaskan bahwa Random Forest merupakan meta estimator yang membangun sejumlah decision tree pada berbagai sub-sampel dataset untuk meningkatkan akurasi prediksi dan mengurangi overfitting[13].

Tabel 3. 8 Tahapan Klasifikasi Menggunakan Random Forest

Tahapan (1)	Keterangan (2)
Input	Data training hasil TF-IDF dan SMOTE
Proses	Pelatihan model Random Forest
Output	Prediksi sentimen positif dan negatif
Evaluasi	Accuracy, precision, recall, F1-score, confusion matrix

Dokumentasi Scikit-learn menjelaskan bahwa Random Forest merupakan meta estimator yang membangun sejumlah decision tree pada berbagai sub-sampel dataset untuk meningkatkan akurasi prediksi dan mengurangi overfitting[13]

3.12 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui performa algoritma XGBoost dan Random Forest dalam mengklasifikasikan sentimen ulasan aplikasi JobStreet. Proses evaluasi menggunakan data testing yang tidak melalui SMOTE agar hasil pengujian tetap mencerminkan kondisi data asli.

Tabel 3. 9 Metrik Evaluasi Model

Metrik	Keterangan
(1)	(2)
Accuracy	Mengukur jumlah prediksi benar dari seluruh data
Precision	Mengukur ketepatan model dalam memprediksi suatu kelas
Recall	Mengukur kemampuan model menemukan seluruh data pada suatu kelas
F1-score	Mengukur keseimbangan antara precision dan recall
Confusion matrix	Menampilkan jumlah prediksi benar dan salah pada setiap kelas

Scikit-learn menyediakan fungsi `classification_report` untuk menampilkan precision, recall, F1-score, dan support, serta fungsi `confusion_matrix` untuk mengevaluasi hasil klasifikasi berdasarkan label aktual dan label prediksi[15]. Selain itu, digunakan juga confusion matrix untuk melihat hasil klasifikasi secara lebih detail.

Tabel 3. 10 Rencana Perbandingan Hasil Evaluasi XGBoost dan Random Forest

Algoritma	Accuracy	Precision	Recall	F1-score
(1)	(2)	(3)	(4)	(5)
XGBoost + SMOTE	...	...	...	...
Random Forest + SMOTE	...	...	...	...

3.13 Perbandingan Hasil Algoritma

Setelah model XGBoost dan Random Forest diuji, hasil evaluasi kedua algoritma dibandingkan untuk menentukan algoritma dengan performa terbaik dalam klasifikasi sentimen ulasan aplikasi JobStreet.

Perbandingan dilakukan berdasarkan metrik accuracy, precision, recall, F1-score, dan confusion matrix. Apabila terdapat perbedaan hasil antar metrik, maka F1-score dijadikan pertimbangan utama karena mampu menunjukkan keseimbangan antara precision dan recall, terutama pada data tidak seimbang.

Penelitian ini mengimplementasikan dua algoritma klasifikasi, yaitu XGBoost dan Random Forest, masing-masing menggunakan library Python dengan parameter bawaan. XGBoost versi 2.0.3 dikonfigurasi dengan `n_estimators=100`, `learning_rate=0.3`, `max_depth=6`, dan `subsample=1.0`, sedangkan Random Forest dari Scikit-learn versi 1.5.0 menggunakan `n_estimators=100`, `max_depth=None`, `min_samples_split=2`, dan `min_samples_leaf=1`. Karena tujuan penelitian adalah membandingkan performa algoritma pasca-SMOTE, bukan mengoptimalkan model, maka parameter default pada kedua algoritma digunakan sebagai tolok ukur awal yang objektif.

3.14 Perangkat dan Tools Penelitian

Penelitian ini dapat dilakukan menggunakan perangkat keras dan perangkat lunak pada tabel berikut:

Tabel 3. 11 Perangkat Keras

Komponen	Spesifikasi
(1)	(2)
Laptop/PC	HP
RAM	4 GB
Penyimpanan	128 GB
Sistem Operasi	Windows

Tabel 3. 12 Perangkat Lunak

Tools	Fungsi
(1)	(2)
Python	Bahasa pemrograman utama

Google Colab / Jupyter Notebook	Lingkungan pengolahan data
Pandas	Pengolahan dataset
NumPy	Operasi numerik
Scikit-learn	TF-IDF, split data, evaluasi model, Random Forest
Imbalanced-learn	Penerapan SMOTE
XGBoost	Klasifikasi menggunakan XGBoost
Matplotlib / Seaborn	Visualisasi data dan hasil evaluasi

BAB IV

HASIL PENELITIAN

Membahas hasil penelitian berdasarkan tahapan metodologi yang telah dikerjakan, mulai dari pengumpulan dan seleksi data ulasan aplikasi JobStreet di Google Play Store, pelabelan sentimen, preprocessing teks, hingga ekstraksi fitur menggunakan TF-IDF. Selanjutnya, data dibagi menjadi data training dan testing serta diterapkan metode SMOTE untuk mengatasi untuk menyeimbangkan data. Proses klasifikasi dilakukan menggunakan algoritma XGBoost dan Random Forest, kemudian dievaluasi untuk mengetahui performa masing-masing model dalam menentukan algoritma yang terbaik.

4.1 Hasil Pengumpulan data

Data dalam penelitian ini terdapat ulasan pengguna aplikasi JobStreet pada Google Play Store melalui teknik scraping menggunakan python. Jumlah data awal yang berhasil dikumpulkan sebanyak 9000 opini.

Pada tahap ini, data hasil scraping masih berupa data mentah yang belum melalui proses seleksi, pelabelan, dan preprocessing. Oleh karena itu, data tersebut digunakan sebagai dasar untuk tahap seleksi data dan pelabelan sentimen. Ulasan tersebut berisi berbagai pendapat pengguna mengenai pengalaman penggunaan aplikasi JobStreet, seperti penilaian positif, keluhan, kritik, dan saran.

	username	rating	tanggal_ulasan	review
0	Pengguna Google	2	2026-05-15 12:11:48	ga bisa buka
1	Pengguna Google	2	2026-05-15 04:01:20	pas lagi ngetik mau kirim lamaran, tiba tiba k...
2	Pengguna Google	5	2026-05-15 03:35:30	Bagus
3	Pengguna Google	5	2026-05-15 02:08:29	ok....
4	Pengguna Google	5	2026-05-15 01:59:27	Ok

Gambar 4. 1 Data Ulasan Hasil Pengumpulan Data

Berdasarkan Gambar 4.1 data ulasan hasil pengumpulan data, data ulasan yang dikumpulkan terdiri dari beberapa atribut, seperti nama pengguna, rating, tanggal ulasan, dan isi ulasan. Pada penelitian ini, atribut utama yang digunakan dalam proses analisis sentimen adalah isi ulasan, karena teks ulasan berisi opini pengguna terhadap aplikasi JobStreet. Atribut lain seperti rating dan tanggal digunakan sebagai informasi pendukung untuk memahami karakteristik data yang diperoleh.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	username, rating, tanggal_ulasan, review													
2	Pengguna Google, 2, 2026-05-15 04:01:20, "pas lagi ngetik mau kirim lamaran, tiba tiba keluar sendiri aplikasinya"													
3	Pengguna Google, 5, 2026-05-15 03:35:30, Bagus													
4	Pengguna Google, 5, 2026-05-15 02:08:29, ok....													
5	Pengguna Google, 5, 2026-05-15 01:59:27, Ok													
6	Pengguna Google, 1, 2026-05-15 00:30:29, "aplikasi Tokai, udah lamar cuma ada pemberitahuan sudah kadaluarsa, mau cari kerja bukan cari pemberitahuan"													
7	Pengguna Google, 1, 2026-05-14 18:01:11, GK berpungsi siap di daftarkan malah di suruh daftar lagu													
8	Pengguna Google, 5, 2026-05-14 15:18:32, keren sangat membantu													
9	Pengguna Google, 2, 2026-05-14 14:00:07, good													
10	Pengguna Google, 5, 2026-05-14 12:46:47, sngat membantu													
11	Pengguna Google, 5, 2026-05-14 11:50:52, bagus banget gais													
12	Pengguna Google, 5, 2026-05-14 10:51:25, mencari pekerjaan dengan mudah dengan aplikasi Jobstreet..													
13	Pengguna Google, 5, 2026-05-14 10:42:39, mantap													
14	Pengguna Google, 5, 2026-05-14 10:29:12, baik													
15	Pengguna Google, 5, 2026-05-14 10:22:10, sangat membantu mencari pekerjaan													
16	Pengguna Google, 5, 2026-05-14 10:20:19, nice													
17	Pengguna Google, 5, 2026-05-14 10:07:14, "info lokernya update terus, jadi mudah mencari kerja."													

Gambar 4. 2 Tampilan Data Ulasan Aplikasi Jobstreet

Gambar 4.2 tampilan data ulasan aplikasi jobstreet, menunjukkan tampilan sebagian data ulasan aplikasi JobStreet yang berhasil dikumpulkan dari Google Play Store. Data tersebut masih berupa data mentah sehingga perlu melalui beberapa tahap pengolahan sebelum digunakan dalam proses klasifikasi sentimen.

4.2 Hasil Seleksi Data

Seleksi data dilakukan untuk memilih data yang layak digunakan dalam proses analisis sentimen. Data hasil pengumpulan dari Google Play Store masih berupa data mentah sehingga perlu diperiksa terlebih dahulu untuk memastikan tidak terdapat data kosong maupun data duplikat yang dapat memengaruhi hasil penelitian.

Pada penelitian ini, seleksi dilakukan pada kolom review karena berisi teks ulasan pengguna sebagai data utama analisis sentimen. Data dengan nilai kosong pada kolom review dihapus karena tidak dapat dianalisis, sedangkan data duplikat dihapus agar tidak terdapat ulasan yang sama secara berulang dalam dataset.



...	Keterangan	Jumlah Data
0	Total data awal	9000
1	Data kosong	0
2	Data duplikat	2490
3	Data setelah seleksi	6510

Gambar 4.3 Hasil Seleksi Data

Berdasarkan Gambar 4.3 hasil seleksi data, jumlah data awal yang diperoleh dari Google Play Store sebanyak 9.000 data. Hasil pemeriksaan menunjukkan tidak terdapat data kosong pada kolom review, namun ditemukan 2.490 data duplikat. Data duplikat tersebut kemudian dihapus untuk menghindari pengulangan ulasan dalam proses analisis. Setelah proses seleksi dilakukan, jumlah data yang digunakan dalam penelitian ini menjadi 6.510 data. Data hasil seleksi tersebut selanjutnya digunakan pada tahap pelabelan sentimen. Tahap seleksi data dilakukan

agar dataset yang digunakan lebih bersih dan representatif untuk analisis sentimen ulasan aplikasi JobStreet.

4.3 Hasil Pelabelan Sentimen

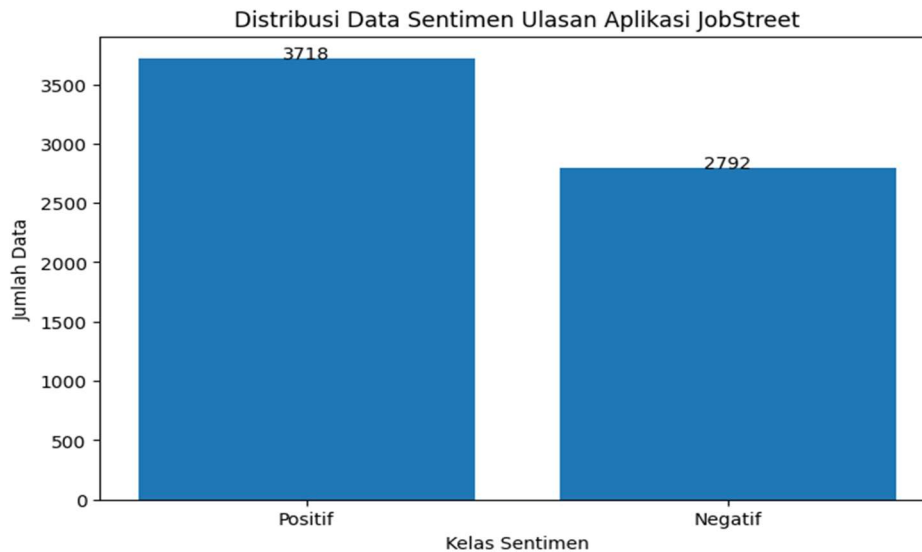
Setelah data ulasan aplikasi JobStreet dikumpulkan, tahap selanjutnya adalah pelabelan sentimen untuk memberikan kategori pada setiap ulasan sebelum proses klasifikasi. Dalam penelitian ini, sentimen dibagi menjadi dua kelas, yaitu positif dan negatif.

Pada penelitian ini, sentimen dibagi menjadi 2 kelas, yaitu positif dan negatif. Sentimen positif menunjukkan kepuasan atau pengalaman baik pengguna terhadap aplikasi JobStreet, sedangkan sentimen negatif menunjukkan keluhan, kritik, atau ketidakpuasan pengguna.

...	username	rating	tanggal_ulasan	review	sentiment
0	Pengguna Google	2	2026-05-15 12:11:48	ga bisa buka	Negatif
1	Pengguna Google	2	2026-05-15 04:01:20	pas lagi ngetik mau kirim lamaran, tiba tiba k...	Negatif
2	Pengguna Google	5	2026-05-15 03:35:30	Bagus	Positif
3	Pengguna Google	5	2026-05-15 02:08:29	ok....	Positif
4	Pengguna Google	5	2026-05-15 01:59:27	Ok	Positif

Gambar 4. 4 Hasil Pelabelan Sentimen

Berdasarkan Gambar 4.4 hasil pelabelan sentimen, ulasan dengan rating tinggi seperti rating 4-5 diberi label positif, sedangkan rating rendah seperti 1- 3 diberi label negatif. Hasil pelabelan ini digunakan sebagai target kelas dalam proses klasifikasi menggunakan algoritma XGBoost dan Random Forest. Setelah pelabelan selesai, data memiliki tambahan kolom sentiment yang berisi kategori sentimen setiap ulasan.



Gambar 4. 5 Diagram Distribusi Ulasan Aplikasi Jobstreet

Berdasarkan Gambar 4.2, diagram distribusi sentimen ulasan aplikasi JobStreet, menunjukkan perbedaan jumlah data pada setiap kelas. Kondisi ini menandakan adanya ketidakseimbangan data (imbalanced data) yang dapat memengaruhi proses klasifikasi karena algoritma machine learning cenderung lebih fokus pada kelas mayoritas. Oleh karena itu, pada tahap selanjutnya digunakan metode SMOTE untuk menyeimbangkan jumlah data pada kelas minoritas agar model dapat mempelajari pola setiap kelas sentimen dengan lebih baik. Data tersebut kemudian digunakan pada tahap seleksi data dan preprocessing sebelum dilakukan ekstraksi fitur menggunakan TF-IDF.

4.4 Hasil Preprocessing Data

Setelah proses pelabelan sentimen selesai, tahap berikutnya adalah preprocessing data. Tahap ini bertujuan untuk membersihkan dan menyeragamkan data teks agar siap digunakan pada proses ekstraksi fitur. Data ulasan dari Google Play Store masih berupa teks mentah yang umumnya mengandung huruf kapital, tanda baca, angka, simbol, spasi berlebih, dan kata tidak baku.

Pada penelitian ini, preprocessing dilakukan pada kolom review yang berisi teks ulasan pengguna aplikasi JobStreet. Hasil preprocessing disimpan ke dalam beberapa kolom baru, yaitu clean, case_fold, tokenizing, stopword, stemming.

Tahapan preprocessing yang dilakukan meliputi cleaning, case folding, tokenizing, normalisasi, stopwords removal, dan stemming.

4.4.1 Cleaning

Cleaning merupakan tahap pembersihan teks dari karakter yang tidak diperlukan, seperti URL, angka, tanda baca, simbol, dan spasi berlebih. Proses ini dilakukan agar teks menjadi lebih bersih dan tidak mengganggu proses analisis.

	username	rating	tanggal_ulasan	review	sentiment	clean
0	Pengguna Google	2	2026-05-15 12:11:48	ga bisa buka	Negatif	ga bisa buka
1	Pengguna Google	2	2026-05-15 04:01:20	pas lagi ngetik mau kirim lamaran, tiba tiba k...	Negatif	pas lagi ngetik mau kirim lamaran tiba tiba ke...
2	Pengguna Google	5	2026-05-15 03:35:30	Bagus	Positif	bagus
3	Pengguna Google	5	2026-05-15 02:08:29	ok....	Positif	ok
4	Pengguna Google	5	2026-05-15 01:59:27	Ok	Positif	ok

Gambar 4. 6 Hasil Tahap Cleaning

Berdasarkan Gambar 4.6 hasil tahap Cleaning, kolom review menunjukkan ulasan asli sebelum preprocessing, sedangkan kolom clean menunjukkan hasil setelah proses cleaning. Contohnya, ulasan “Bagus” berubah menjadi “bagus”, “ok....” menjadi “ok”, dan “Ok” menjadi “ok”. Perubahan tersebut menunjukkan bahwa tanda baca telah dihapus dan teks dibuat lebih seragam.

4.4.2 Case Folding

Tahap preprocessing selanjutnya adalah case folding, yaitu proses mengubah seluruh huruf pada teks ulasan menjadi huruf kecil. Tahap ini dilakukan untuk menyeragamkan penulisan kata agar tidak terdapat perbedaan antara huruf kapital

dan huruf kecil. Pada penelitian ini, case folding dilakukan pada kolom clean yang telah melalui proses cleaning. Hasilnya kemudian disimpan pada kolom case_fold.

	username	rating	tanggal_ulasan	review	sentiment	clean	case_fold
0	Pengguna Google	2	2026-05-15 12:11:48	ga bisa buka	Negatif	ga bisa buka	ga bisa buka
1	Pengguna Google	2	2026-05-15 04:01:20	pas lagi ngetik mau kirim lamaran, tiba tiba k...	Negatif	pas lagi ngetik mau kirim lamaran tiba tiba ke...	pas lagi ngetik mau kirim lamaran tiba tiba ke...
2	Pengguna Google	5	2026-05-15 03:35:30	Bagus	Positif	bagus	bagus
3	Pengguna Google	5	2026-05-15 02:08:29	ok....	Positif	ok	ok
4	Pengguna Google	5	2026-05-15 01:59:27	Ok	Positif	ok	ok

Gambar 4. 7 Hasil Tahap Case Folding

Berdasarkan Gambar 4.7 Hasil Tahap Case Folding, dapat dilihat bahwa kolom clean merupakan hasil dari tahap cleaning, sedangkan kolom case_fold merupakan hasil setelah dilakukan proses case folding. Pada tahap ini, seluruh teks ulasan diubah menjadi huruf kecil. Sebagai contoh, ulasan “Bagus” berubah menjadi “bagus”, dan ulasan “Ok” berubah menjadi “ok”.

4.4.3 Tokenizing

Tahap selanjutnya dalam preprocessing adalah tokenizing, yaitu proses memecah teks ulasan menjadi kata atau token agar setiap kata dapat diproses secara terpisah pada tahap berikutnya. Pada penelitian ini, tokenizing dilakukan pada kolom case_fold yang melalui proses case folding. Hasil tokenizing disimpan pada kolom token, sehingga ulasan yang sebelumnya berbentuk kalimat berubah menjadi kumpulan kata dalam bentuk list.

	username	rating	tanggal_ulasan	review	sentiment	clean	case_fold	token
0	Pengguna Google	2	2026-05-15 12:11:48	ga bisa buka	Negatif	ga bisa buka	ga bisa buka	[ga, bisa, buka]
1	Pengguna Google	2	2026-05-15 04:01:20	pas lagi ngetik mau kirim lamaran, tiba tiba k...	Negatif	pas lagi ngetik mau kirim lamaran tiba tiba ke...	pas lagi ngetik mau kirim lamaran tiba tiba ke...	[pas, lagi, ngetik, mau, kirim, lamaran, tiba, ke, ...]
2	Pengguna Google	5	2026-05-15 03:35:30	Bagus	Positif	bagus	bagus	[bagus]
3	Pengguna Google	5	2026-05-15 02:08:29	ok....	Positif	ok	ok	[ok]
4	Pengguna Google	5	2026-05-15 01:59:27	Ok	Positif	ok	ok	[ok]

Gambar 4. 8 Hasil Tahap Tokenizing

Berdasarkan Gambar 4.7, kolom case_fold menunjukkan hasil case folding, sedangkan kolom token menunjukkan hasil tokenizing. Contohnya, ulasan “ga bisa buka” diubah menjadi [ga, bisa, buka], “bagus” menjadi [bagus], dan “ok” menjadi [ok].

4.4.4 Stop Removal

Tahap selanjutnya dalam preprocessing adalah **stopword removal**, yaitu proses menghapus kata-kata umum yang dianggap kurang memiliki makna penting dalam analisis sentimen, seperti “yang”, “dan”, “di”, “ke”, “dari”, dan “untuk”. Pada penelitian ini, **stopword removal** dilakukan pada data yang telah melalui tahap **preprocessing** sebelumnya. Hasil sebelum penghapusan stopwords ditampilkan pada kolom **sebelum_stopword**, sedangkan hasil setelah proses **stopword removal** ditampilkan pada kolom **sesudah_stopword**.

	review	sebelum_stopword	sesudah_stopword
0	ga bisa buka	ga bisa buka	ga bisa buka
1	pas lagi ngetik mau kirim lamaran, tiba tiba k...	pas lagi ngetik mau kirim lamaran tiba tiba ke...	pas lagi ngetik mau kirim lamaran tiba tiba ke...
2	Bagus	bagus	bagus
3	ok....	ok	ok
4	Ok	ok	ok

Gambar 4. 9 Hasil Tahap Stopword Removal

Berdasarkan Gambar 4.9 Hasil Tahap Stopword Removal, kolom review menunjukkan ulasan asli, kolom sebelum_stopword menunjukkan data sebelum penghapusan stopwords, dan kolom sesudah_stopword menunjukkan hasil setelah proses stopwords removal.

4.4.5 Stemming

Tahap selanjutnya dalam preprocessing adalah stemming, yaitu proses mengubah kata berimbuhan menjadi kata dasar agar kata dengan makna sama memiliki bentuk yang seragam. Pada penelitian ini, stemming dilakukan setelah tahap stopwords removal. Data yang telah melalui penghapusan stopwords kemudian diproses menjadi kata dasar, dan hasilnya disimpan pada kolom stem.

	review	clean	case_fold	token	hasil_stopword	stem	final
0	ga bisa buka	ga bisa buka	ga bisa buka	[ga, bisa, buka]	ga bisa buka	[ga, bisa, buka]	ga bisa buka
1	pas lagi ngetik mau kirim lamaran, tiba tiba k...	pas lagi ngetik mau kirim lamaran tiba tiba ke...	pas lagi ngetik mau kirim lamaran tiba tiba ke...	[pas, lagi, ngetik, mau, kirim, lamaran, tiba, ...]	pas lagi ngetik mau kirim lamaran tiba tiba ke...	[pas, lagi, ngetik, mau, kirim, lamar, tiba, t...]	pas lagi ngetik mau kirim lamar tiba tiba kelu...
2	Bagus	bagus	bagus	[bagus]	bagus	[bagus]	bagus
3	ok....	ok	ok	[ok]	ok	[ok]	ok
4	Ok	ok	ok	[ok]	ok	[ok]	ok

Gambar 4. 10 Hasil Tahap Stemming dan Hasil Akhir Preprocessing

Berdasarkan Gambar 4.9, data ulasan telah melalui beberapa tahap preprocessing, yaitu clean, case_fold, token, hasil_stopword, stem, hingga final. Dapat dilihat bahwa ulasan pada kolom review mengalami perubahan setelah melalui proses preprocessing. Contohnya, ulasan “Bagus” berubah menjadi “bagus”, “ok....” menjadi “ok”, sedangkan “ga bisa buka” tetap menjadi “ga bisa buka” karena tidak mengalami perubahan pada proses cleaning, stopwords removal,

dan stemming. Kolom stem menunjukkan hasil stemming dalam bentuk kumpulan kata dasar.

Sebagai contoh, pada ulasan “pas lagi ngetik mau kirim lamaran, tiba tiba k...”, kata “lamaran” berubah menjadi “lamar”. Sementara itu, kata seperti “bagus” dan “ok” tidak mengalami perubahan karena sudah berbentuk kata dasar. Secara keseluruhan, preprocessing berhasil mengubah data ulasan mentah menjadi teks yang lebih bersih, seragam, dan terstruktur. Tahapan cleaning, case folding, tokenizing, stopwords removal, dan stemming dilakukan secara berurutan agar data siap digunakan pada proses ekstraksi fitur TF-IDF dan klasifikasi sentimen menggunakan algoritma XGBoost dan Random Forest.

4.5 Hasil Ekstraksi Fitur TF-IDF

Setelah melalui tahap preprocessing, data ulasan diekstraksi menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Tahap ini bertujuan mengubah data teks menjadi bentuk numerik agar dapat diproses oleh algoritma machine learning. Data pada kolom final yang merupakan hasil akhir preprocessing diubah menjadi vektor numerik menggunakan TF-IDF. Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculan dan tingkat kepentingannya dalam keseluruhan dokumen. Kata yang sering muncul pada suatu ulasan tetapi jarang muncul pada ulasan lain akan memiliki bobot lebih tinggi.

Pada penelitian ini, hasil ekstraksi fitur TF-IDF membentuk matriks berukuran (6510, 6250). Angka 6510 menunjukkan jumlah data ulasan, sedangkan 6250 menunjukkan jumlah fitur atau kata unik yang terbentuk dari seluruh data.

	aa	aamiinnn	aamiin	aamiinnn	aamin	aangat	aarrghhh	aatas	abai	abal	...	yuk	yups	yutub	zaman	zariah	zemoga	zonk	zoom	zoon
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows x 6250 columns

Gambar 4. 11 Hasil Matriks TF-IDF

Berdasarkan Gambar 4.11, hasil TF-IDF ditampilkan dalam bentuk matriks, di setiap baris menampilkan data ulasan dan setiap kolom merepresentasikan fitur kata. Nilai pada setiap sel menunjukkan bobot TF-IDF suatu kata pada ulasan tertentu. Sebagian besar nilai matriks bernilai 0.0 karena tidak semua kata muncul pada setiap ulasan. Jika suatu kata tidak muncul, maka nilai TF-IDF bernilai 0. Sebaliknya, kata yang muncul akan memiliki bobot sesuai tingkat kepentingannya.

Untuk mengetahui kata dengan bobot tertinggi, dilakukan perhitungan rata-rata bobot TF-IDF pada setiap fitur. Hasilnya dapat dilihat pada Gambar 4.12 berikut:

	Kata/Fitur	Rata-rata Bobot TF-IDF
0	kerja	0.054216
1	sangat	0.047237
2	bantu	0.045125
3	cari	0.039001
4	aplikasi	0.035512
5	saya	0.029199
6	bagus	0.029003
7	bisa	0.027988
8	mudah	0.025509
9	jobstreet	0.022635

Gambar 4. 12 Hasil Ekstraksi Fitur TF-IDF

81 Berdasarkan Gambar 4.12 Hasil Ekstraksi Fitur TF-IDF, bobot TF-IDF Tertinggi, kata dengan rata-rata bobot TF-IDF tertinggi adalah “kerja” dengan nilai 0.054216, diikuti kata “sangat” sebesar 0.047237, “bantu” sebesar 0.045125, “cari” sebesar 0.039001, dan “aplikasi” sebesar 0.035512. Kemunculan kata seperti “kerja”, “bantu”, “cari”, “aplikasi”, “bagus”, “mudah”, dan “jobstreet” menunjukkan bahwa ulasan pengguna banyak membahas fungsi utama aplikasi sebagai media pencarian kerja. Kata-kata tersebut memiliki peran penting karena sering muncul dan berkaitan langsung dengan topik penelitian.

9 Hasil ekstraksi fitur TF-IDF kemudian digunakan sebagai input pada tahap klasifikasi. Matriks TF-IDF yang terbentuk digunakan untuk proses pembagian data training dan testing sebelum diterapkan pada algoritma XGBoost dan Random Forest. Melalui proses ini, data teks berhasil diubah menjadi data numerik yang dapat diproses oleh model machine learning.

4.6 Hasil Pembagian Data

230 Setelah data ulasan diubah menjadi bentuk numerik menggunakan TF-IDF, 47 tahap selanjutnya adalah pembagian data menjadi data training dan data testing. 64 Pembagian ini dilakukan untuk memisahkan data pelatihan model dan data 63 pengujian model. Pada penelitian ini, pembagian data menggunakan rasio 80% data training dan 20% data testing. Data training digunakan untuk melatih model klasifikasi, sedangkan data testing digunakan untuk menguji kemampuan model 15 pada data yang belum pernah dipelajari sebelumnya. Dari total 6.510 data hasil 52 ekstraksi fitur TF-IDF, diperoleh 5.208 data training dan 1.302 data testing.

Tabel 4. 1 Pembagian Data

Data	Persentase	Jumlah Data
(1)	(2)	(3)
Training	80%	5.208
Testing	20%	1.302
Total	100%	6.510

Berdasarkan Tabel 4.1 Pembagian data, data training berjumlah 5.208 data atau 80% dari total data, sedangkan data testing berjumlah 1.302 data atau 20%. Dengan demikian, total data yang digunakan sebanyak 6.510 data. Pembagian data dilakukan secara acak agar data training dan testing memiliki variasi yang baik. Data training digunakan untuk membangun model klasifikasi menggunakan algoritma XGBoost dan Random Forest, sedangkan data testing digunakan untuk mengevaluasi performa model.

Pada tahap ini, data testing tidak melalui proses penyeimbangan agar tetap merepresentasikan distribusi data asli. Proses penyeimbangan menggunakan metode SMOTE hanya diterapkan pada data training agar evaluasi model tetap objektif.

4.7 Hasil Penerapan SMOTE

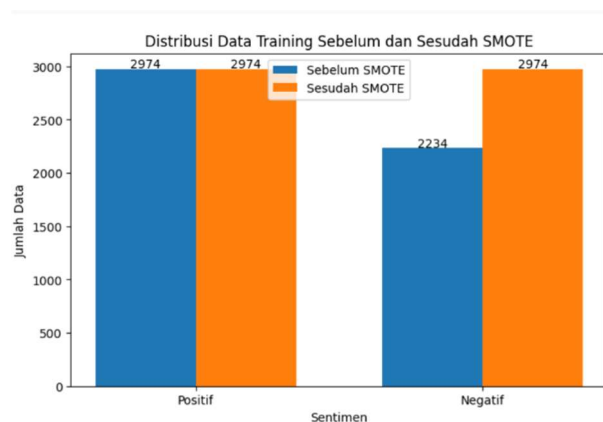
Setelah data dibagi menjadi data training dan data testing, tahap berikutnya adalah penerapan SMOTE (Synthetic Minority Oversampling Technique). Metode ini digunakan untuk mengatasi ketidakseimbangan jumlah data pada kelas sentimen. Ketidakseimbangan data dapat memengaruhi performa model karena algoritma cenderung lebih fokus pada kelas mayoritas dibandingkan kelas minoritas. Oleh karena itu, penyeimbangan data dilakukan agar model dapat mempelajari pola setiap kelas sentimen secara lebih baik.

Pada penelitian ini, SMOTE hanya diterapkan pada data training, sedangkan data testing tetap menggunakan distribusi data asli agar hasil evaluasi model lebih objektif. Hasil distribusi data training sebelum dan sesudah penerapan SMOTE dapat dilihat pada Gambar 4.13 berikut.

	Sentimen	Sebelum SMOTE	Sesudah SMOTE
0	Positif	2974	2974
1	Negatif	2234	2974
2	Total	5208	5948

Gambar 4. 13 Hasil Distribusi Data Training

Berdasarkan Gambar 4.13 Hasil Distribusi Data Training, sebelum penerapan SMOTE, jumlah data training pada kelas positif sebanyak 2.974 data, sedangkan kelas negatif sebanyak 2.234 data. Kondisi ini menunjukkan adanya ketidakseimbangan data karena jumlah kelas positif lebih banyak dibandingkan kelas negatif. Setelah SMOTE diterapkan, jumlah data pada kelas negatif meningkat menjadi 2.974 data, sehingga jumlah data pada kedua kelas menjadi seimbang.



Gambar 4. 14 Distribusi Data Training Sebelum dan Sesudah SMOTE

Berdasarkan Gambar 4.14 Distribusi Data Training Sebelum dan Sesudah SMOTE, sebelum SMOTE jumlah data kelas positif lebih tinggi dibandingkan kelas negatif. Setelah SMOTE diterapkan, distribusi kedua kelas menjadi sama, yaitu masing-masing 2.974 data. SMOTE diterapkan setelah pembagian data karena proses penyeimbangan hanya dilakukan pada data training. Hal ini bertujuan agar data testing tetap merepresentasikan kondisi asli sehingga hasil evaluasi model tetap objektif. Data training yang telah diseimbangkan kemudian digunakan untuk melatih model klasifikasi menggunakan algoritma XGBoost dan Random Forest agar model dapat mengenali pola sentimen secara lebih baik.

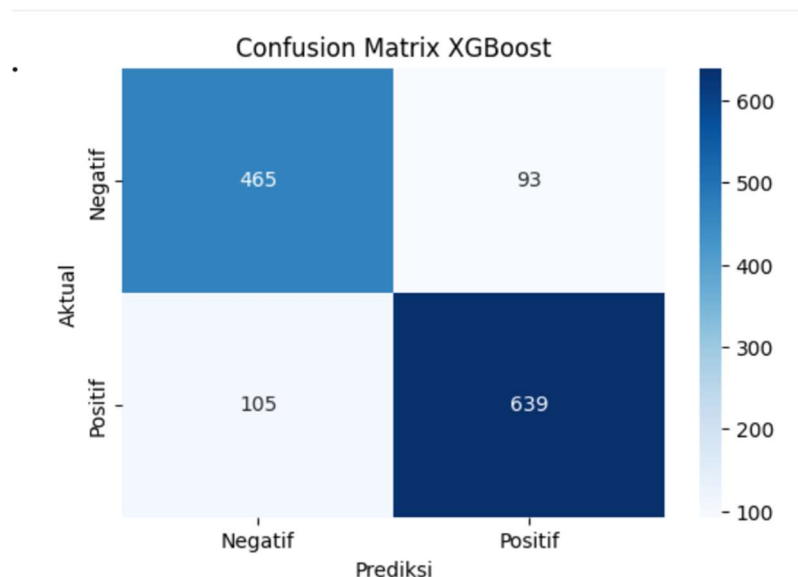
4.8 Hasil Klasifikasi XGBoost

Tahap klasifikasi pertama dilakukan menggunakan algoritma XGBoost. Algoritma ini digunakan untuk mengklasifikasikan sentimen ulasan aplikasi JobStreet ke dalam dua kelas, yaitu positif dan negatif. Data yang digunakan untuk melatih model adalah data training hasil penerapan SMOTE, sedangkan data testing digunakan untuk menguji performa model. Evaluasi model XGBoost dilakukan menggunakan confusion matrix dan classification report. Confusion matrix digunakan untuk mengetahui jumlah data yang berhasil diklasifikasikan dengan benar maupun salah. Sementara itu, classification report digunakan untuk mengetahui nilai precision, recall, F1-score, dan accuracy. Hasil confusion matrix algoritma XGBoost dapat dilihat pada Gambar 4.15 berikut.

		Negatif	Positif
Negatif	465	93	
Positif	105	639	

Gambar 4. 15 Confusion Matrix XGBoost

Berdasarkan Tabel 4.10, terdapat 465 data negatif yang berhasil diprediksi dengan benar sebagai negatif, sedangkan 93 data negatif salah diprediksi sebagai positif. Selanjutnya, terdapat 639 data positif yang berhasil diprediksi dengan benar sebagai positif, sedangkan 105 data positif salah diprediksi sebagai negatif. Hasil tersebut menunjukkan bahwa algoritma XGBoost mampu mengklasifikasikan sebagian besar data testing dengan benar. Jumlah prediksi benar pada diagonal utama, yaitu 465 dan 639, lebih besar dibandingkan jumlah kesalahan prediksi di luar diagonal utama, yaitu 93 dan 105.



Gambar 4. 16 Confusion Matrix XGBoost

Berdasarkan Gambar 4.16 Confusion Matrix XGBoost, dapat dilihat bahwa nilai terbesar terdapat pada kelas positif yang diprediksi sebagai positif, yaitu sebanyak 639 data. Hal ini menunjukkan bahwa model XGBoost cukup baik dalam mengenali ulasan dengan sentimen positif. Selain itu, model juga mampu mengenali sentimen negatif dengan cukup baik, yaitu sebanyak 465 data negatif berhasil diklasifikasikan secara benar. Hasil classification report algoritma XGBoost dapat dilihat pada Gambar 4.17 berikut.

	precision	recall	f1-score	support
Negatif	0.815789	0.833333	0.824468	558.000000
Positif	0.872951	0.858871	0.865854	744.000000
accuracy	0.847926	0.847926	0.847926	0.847926
macro avg	0.844370	0.846102	0.845161	1302.000000
weighted avg	0.848453	0.847926	0.848117	1302.000000

Gambar 4. 17 Hasil Classification Report Algoritma XGBoost

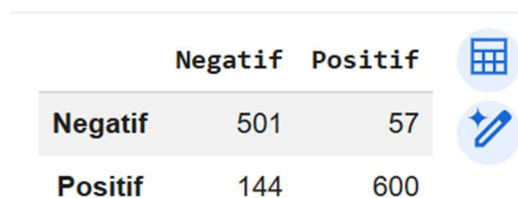
Berdasarkan Tabel 4.11, algoritma XGBoost memperoleh nilai accuracy sebesar 0.847926 atau sekitar 84,79%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan data testing dengan cukup baik. Pada kelas negatif, model memperoleh precision sebesar 0.815789, recall sebesar 0.833333, dan F1-score sebesar 0.824468. Sementara itu, pada kelas positif diperoleh precision sebesar 0.872951, recall sebesar 0.858871, dan F1-score sebesar 0.865854. Nilai F1-score yang lebih tinggi pada kelas positif menunjukkan bahwa model lebih baik dalam mengenali sentimen positif dibandingkan sentimen negatif.

Secara keseluruhan, model memperoleh macro average F1-score sebesar 0.845161 dan weighted average F1-score sebesar 0.848117. Hasil tersebut menunjukkan bahwa algoritma XGBoost memiliki performa yang cukup baik dalam klasifikasi sentimen ulasan aplikasi JobStreet. Hasil klasifikasi ini selanjutnya dibandingkan dengan algoritma Random Forest untuk mengetahui algoritma dengan performa terbaik dalam analisis sentimen ulasan aplikasi JobStreet.

4.9 Hasil Klasifikasi Random Forest

Tahap klasifikasi kedua dilakukan menggunakan algoritma Random Forest untuk membandingkan performanya dengan XGBoost dalam mengklasifikasikan sentimen ulasan aplikasi JobStreet. Penelitian ini menggunakan dua kelas sentimen, yaitu positif dan negatif. Model Random Forest dilatih menggunakan data training hasil SMOTE agar distribusi data menjadi seimbang. Setelah proses pelatihan selesai, model diuji menggunakan data testing tanpa SMOTE agar hasil evaluasi tetap objektif.

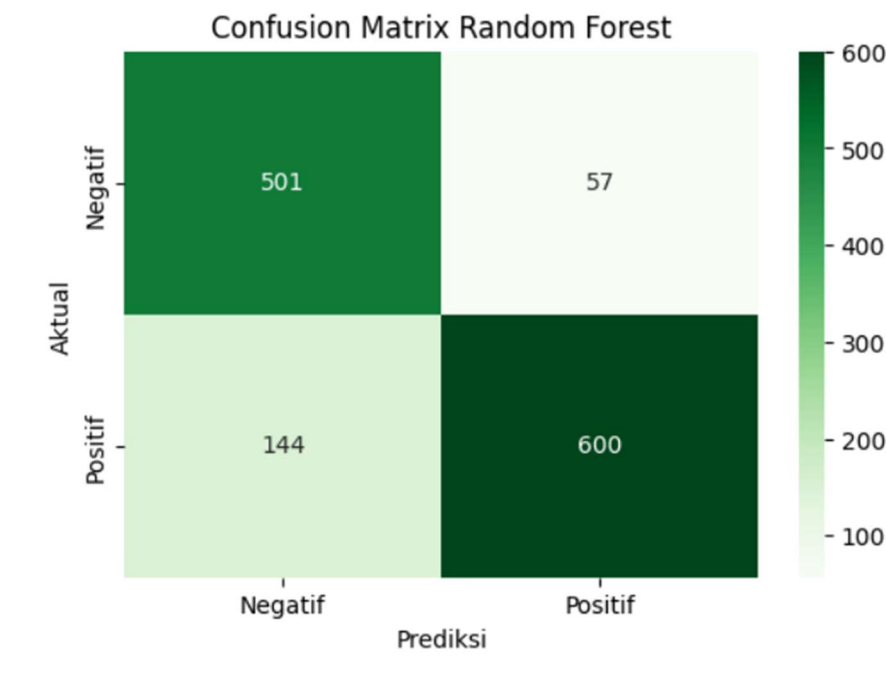
Evaluasi model dilakukan menggunakan confusion matrix dan classification report. Confusion matrix digunakan untuk melihat jumlah prediksi benar dan salah, sedangkan classification report digunakan untuk mengetahui nilai precision, recall, F1-score, dan accuracy.



	Negatif	Positif
Negatif	501	57
Positif	144	600

Gambar 4. 18 Confusion Matrix Random Forest

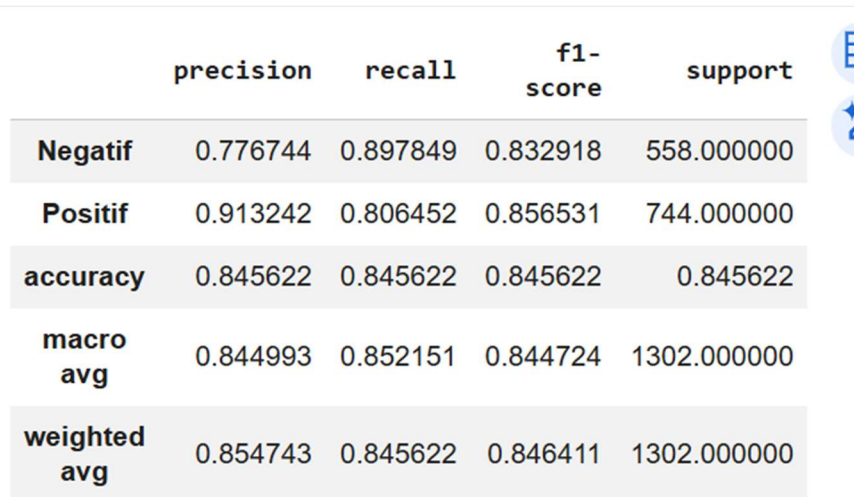
Berdasarkan Gambar 4.18 Confusion Matrix Random Forest, sebanyak 501 data negatif berhasil diprediksi dengan benar, sedangkan 57 data negatif diprediksi sebagai positif. Pada kelas positif, sebanyak 600 data berhasil diprediksi dengan benar, sedangkan 144 data diprediksi sebagai negatif. Hasil tersebut menunjukkan bahwa Random Forest mampu mengklasifikasikan sebagian besar data testing dengan benar. Jumlah prediksi benar pada diagonal utama yaitu 501 dan 600, lebih besar dibandingkan jumlah kesalahan prediksi di luar diagonal utama, yaitu 57 dan 144.



Gambar 4. 19 Confusion Matrix Random Forest

Berdasarkan Gambar 4.19 Confusion Matrix Random Forest, jumlah prediksi terbesar terdapat pada kelas positif yang diprediksi benar sebagai positif, yaitu sebanyak 600 data. Model juga mampu mengenali sentimen negatif dengan cukup baik, yaitu sebanyak 501 data berhasil diklasifikasikan dengan benar. Namun, masih terdapat kesalahan klasifikasi, yaitu 57 data negatif diprediksi sebagai positif

dan 144 data positif diprediksi sebagai negatif. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan beberapa pola ulasan yang mirip.



The image shows a classification report for a Random Forest model. The report is presented in a table with five columns: precision, recall, f1-score, and support. The rows represent different classes and overall averages. The 'Negatif' class has a precision of 0.776744, recall of 0.897849, and f1-score of 0.832918, with a support of 558.000000. The 'Positif' class has a precision of 0.913242, recall of 0.806452, and f1-score of 0.856531, with a support of 744.000000. The 'accuracy' row shows an overall accuracy of 0.845622. The 'macro avg' row shows a macro average f1-score of 0.844724, and the 'weighted avg' row shows a weighted average f1-score of 0.846411. The table is displayed on a screen with a sidebar on the right containing icons for a document and a magnifying glass.

	precision	recall	f1-score	support
Negatif	0.776744	0.897849	0.832918	558.000000
Positif	0.913242	0.806452	0.856531	744.000000
accuracy	0.845622	0.845622	0.845622	0.845622
macro avg	0.844993	0.852151	0.844724	1302.000000
weighted avg	0.854743	0.845622	0.846411	1302.000000

Gambar 4. 20 Classification Report Random Forest

Berdasarkan Gambar 4.20, algoritma Random Forest memperoleh accuracy sebesar 0.845622 atau sekitar 84,56%. Pada kelas negatif, model memperoleh precision sebesar 0.776744, recall sebesar 0.897849, dan F1-score sebesar 0.832918. Tingginya nilai recall menunjukkan bahwa model cukup baik dalam mengenali ulasan negatif. Pada kelas positif, model memperoleh precision sebesar 0.913242, recall sebesar 0.806452, dan F1-score sebesar 0.856531. Nilai precision yang tinggi menunjukkan bahwa sebagian besar prediksi positif yang dihasilkan model sudah tepat.

Secara keseluruhan, Random Forest memperoleh macro average F1-score sebesar 0.844724 dan weighted average F1-score sebesar 0.846411. Hasil ini menunjukkan bahwa algoritma Random Forest memiliki performa yang cukup baik dalam klasifikasi sentimen ulasan aplikasi JobStreet. Dengan itu, algoritma random forest mampu mengklasifikasi sentimen positif dan negatif dengan tingkat akurasi

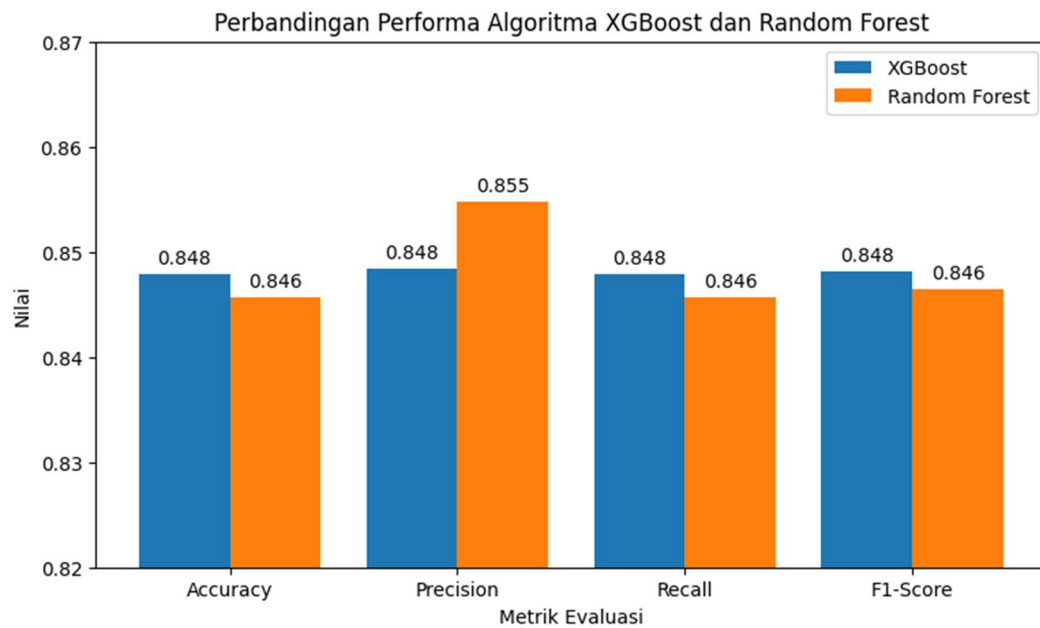
yang tinggi. Hasil klasifikasi Random Forest selanjutnya dibandingkan dengan XGBoost untuk menentukan algoritma dengan performa terbaik berdasarkan accuracy, precision, recall, dan F1-score.

4.10 Analisis Performa Algoritma

Setelah proses klasifikasi menggunakan algoritma XGBoost dan Random Forest selesai, tahap berikutnya adalah membandingkan performa kedua algoritma untuk mengetahui model yang memiliki hasil terbaik dalam klasifikasi sentimen ulasan aplikasi JobStreet. Perbandingan dilakukan berdasarkan metrik accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk mengukur tingkat ketepatan model secara keseluruhan, precision untuk mengukur ketepatan prediksi pada suatu kelas, recall untuk mengukur kemampuan model mengenali seluruh data pada kelas tertentu, dan F1-score untuk melihat keseimbangan antara precision dan recall. Hasil perbandingan performa algoritma XGBoost dan Random Forest dapat dilihat pada Tabel 4.2 berikut.

Tabel 4. 2 Perbandingan Performa Algoritma

Algoritma	Accuracy	Precision	Recall	F1-Score
(1)	(2)	(3)	(4)	(5)
XGBoost	0.847926	0.848453	0.847926	0.848117
Random Forest	0.845622	0.854743	0.845622	0.846411



Gambar 4. 21 Grafik Perbandingan Performa Algoritma XGBoost dan Random Forest

Berdasarkan Gambar 4.21, dapat dilihat bahwa algoritma XGBoost memiliki nilai accuracy, recall, dan F1-score yang sedikit lebih tinggi dibandingkan Random Forest. XGBoost memperoleh accuracy sebesar 0.847926, recall sebesar 0.847926, dan F1-score sebesar 0.848117. Sementara itu, Random Forest memperoleh accuracy sebesar 0.845622, recall sebesar 0.845622, dan F1-score sebesar 0.846411..

Namun, pada nilai precision, Random Forest memperoleh nilai lebih tinggi yaitu 0.854743 dibandingkan XGBoost sebesar 0.848453. Hal ini menunjukkan bahwa Random Forest sedikit lebih unggul dalam menghasilkan prediksi yang tepat, tetapi secara keseluruhan XGBoost tetap lebih baik karena memperoleh nilai accuracy dan F1-score yang lebih tinggi. Hasil tersebut menunjukkan bahwa XGBoost memiliki nilai accuracy, recall, dan F1-score yang sedikit lebih tinggi dibandingkan Random Forest. XGBoost memperoleh accuracy sebesar 84,79%,

61 sedangkan Random Forest sebesar 84,56%, dengan selisih sekitar 0,23%. Pada nilai
5 F1-score, XGBoost juga lebih unggul dengan nilai 0.848117 dibandingkan Random
Forest sebesar 0.846411. Hal ini menunjukkan bahwa XGBoost memiliki
keseimbangan precision dan recall yang lebih baik.

19 Namun, pada metrik precision, Random Forest memperoleh nilai lebih tinggi,
yaitu 0.854743, sedangkan XGBoost sebesar 0.848453. XGBoost memang
273 menunjukkan akurasi dan F1-score yang sedikit lebih baik, namun selisihnya hanya
sekitar 0,23% tergolong sangat tipis. Dari perspektif statistik, perbedaan ini tidak
81 cukup berarti, sehingga kedua algoritma dianggap memiliki kinerja yang setara
dalam klasifikasi sentimen. Keunggulan yang dimiliki XGBoost dipengaruhi oleh
mekanisme boosting yang diterapkannya, yaitu proses pembangunan model secara
195 bertahap dengan tujuan memperbaiki kesalahan prediksi pada iterasi sebelumnya.
Pendekatan tersebut memungkinkan model untuk lebih optimal dalam mengenali
74 pola dan karakteristik sentimen yang terdapat pada data ulasan pengguna.

Di sisi lain, Random Forest menggunakan metode bagging, yakni dengan
178 mengombinasikan sejumlah decision tree dan menentukan hasil prediksi
berdasarkan voting mayoritas dari seluruh pohon keputusan yang dibangun.
25 Berdasarkan hasil pengujian dan perbandingan performa yang telah dilakukan,
dapat disimpulkan bahwa algoritma XGBoost memiliki tingkat kinerja yang sedikit
lebih baik dibandingkan Random Forest dalam proses klasifikasi sentimen ulasan
68 aplikasi JobStreet. Keunggulan tersebut terutama terlihat pada perolehan nilai
accuracy dan F1-score yang lebih tinggi dibandingkan algoritma pembanding.

4.11 Pembahasan Hasil Sentimen

Setelah proses klasifikasi sentimen menggunakan algoritma XGBoost dan Random Forest dilakukan, tahap berikutnya adalah membahas hasil analisis sentimen pada ulasan pengguna aplikasi JobStreet. Pembahasan ini bertujuan untuk mengetahui kecenderungan opini pengguna berdasarkan data ulasan yang telah dianalisis. Hasil pelabelan sentimen menunjukkan bahwa ulasan pengguna terbagi ke dalam dua kategori, yaitu sentimen positif dan sentimen negatif. Sentimen positif mencerminkan kepuasan atau pengalaman baik pengguna terhadap aplikasi, sedangkan sentimen negatif menggambarkan keluhan, kritik, maupun ketidakpuasan pengguna.

Berdasarkan hasil ekstraksi fitur menggunakan TF-IDF, beberapa kata dengan bobot tertinggi meliputi “kerja”, “bantu”, “cari”, “aplikasi”, “bagus”, “mudah”, dan “jobstreet”. Kemunculan kata-kata tersebut menunjukkan bahwa mayoritas ulasan berkaitan dengan fungsi utama JobStreet sebagai aplikasi pencarian kerja. Pada sentimen positif, sebagian besar pengguna memberikan ulasan mengenai kemudahan dalam mencari lowongan pekerjaan melalui aplikasi JobStreet. Pengguna menilai aplikasi ini membantu dalam memperoleh informasi pekerjaan yang sesuai dengan kebutuhan. Hal tersebut terlihat dari kata-kata seperti “kerja”, “bantu”, “cari”, “bagus”, dan “mudah” yang sering muncul pada ulasan positif. Beberapa contoh ulasan positif yang ditemukan antara lain seperti “Bagus” dan “Ok”, serta ulasan lain yang menunjukkan penilaian baik terhadap aplikasi. Ulasan tersebut menggambarkan bahwa pengguna merasa puas terhadap layanan dan manfaat aplikasi, khususnya dalam membantu proses pencarian pekerjaan.



224

65



Submission ID trn:oid:::3618:152713602

134 Berdasarkan Gambar 4.23, kata-kata yang dominan pada sentimen negatif antara lain “nyari”, “kerja”, “lamar”, “lamaran”, “tidak”, “buka”, “susah”, “login”, “masuk”, dan “kenapa”. Kata-kata tersebut menunjukkan bahwa ulasan negatif pengguna banyak membahas kendala dalam penggunaan aplikasi, seperti aplikasi sulit dibuka, masalah login, dan kesulitan saat proses melamar pekerjaan. Hal ini menunjukkan bahwa meskipun aplikasi JobStreet membantu pengguna dalam mencari pekerjaan, masih terdapat beberapa masalah teknis yang perlu diperbaiki.

12 Secara keseluruhan, hasil wordcloud menunjukkan bahwa sentimen positif lebih banyak berkaitan dengan manfaat aplikasi dalam membantu pencarian kerja, sedangkan sentimen negatif lebih banyak berkaitan dengan kendala teknis dan pengalaman pengguna yang kurang optimal. Hasil ini mendukung pembahasan sebelumnya bahwa aplikasi JobStreet memiliki manfaat bagi pengguna, tetapi masih perlu peningkatan pada aspek teknis dan kenyamanan penggunaan aplikasi.

4.12 Rangkuman

2 Berdasarkan seluruh tahapan penelitian yang telah dilakukan, diperoleh sebanyak 9.000 data ulasan aplikasi JobStreet dari Google Play Store. Selanjutnya dilakukan proses seleksi data untuk menghapus data kosong dan data duplikat. Hasil seleksi menunjukkan bahwa tidak ditemukan data kosong, sedangkan data duplikat berjumlah 2.490 data. Setelah penghapusan data duplikat dilakukan, jumlah data yang digunakan dalam penelitian menjadi sebanyak 6.510 data ulasan.

37 Tahap berikutnya adalah pelabelan sentimen berdasarkan rating pengguna. Rating 4 dan 5 dikategorikan sebagai sentimen positif, sedangkan rating 1, 2, dan 3 dikategorikan sebagai sentimen negatif. Setelah itu dilakukan preprocessing data yang meliputi cleaning, case folding, tokenizing, stopword removal, dan stemming

2 untuk membersihkan teks ulasan agar siap digunakan pada tahap analisis. Hasil
38 preprocessing kemudian digunakan pada proses ekstraksi fitur menggunakan
metode TF-IDF untuk mengubah teks menjadi bentuk numerik.

79 Data selanjutnya dibagi menjadi data training dan data testing dengan rasio
80% : 20%. Untuk mengatasi ketidakseimbangan data, diterapkan metode SMOTE
pada data training sehingga jumlah data positif dan negatif menjadi seimbang.
55 Proses klasifikasi sentimen dilakukan menggunakan algoritma XGBoost dan
Random Forest. Hasil evaluasi menunjukkan bahwa XGBoost memperoleh
17 accuracy sebesar 84,79% dengan F1-score sebesar 0,848117, sedangkan Random
Forest memperoleh accuracy sebesar 84,56% dengan F1-score sebesar 0,846411.

36 Berdasarkan hasil perbandingan, algoritma XGBoost memiliki performa
sedikit lebih baik dibandingkan Random Forest dalam melakukan klasifikasi
136 sentimen ulasan aplikasi JobStreet. Secara umum, hasil penelitian menunjukkan
bahwa ulasan positif pengguna berkaitan dengan kemudahan mencari pekerjaan dan
manfaat aplikasi, sedangkan ulasan negatif lebih banyak membahas kendala teknis,
136 seperti masalah login dan aplikasi yang sulit digunakan. Hasil penelitian ini
34 diharapkan dapat menjadi masukan bagi pengembang aplikasi JobStreet untuk
4 meningkatkan kualitas layanan dan pengalaman pengguna.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis sentimen ulasan aplikasi JobStreet pada Google Play Store menggunakan metode SMOTE dengan perbandingan algoritma XGBoost dan Random Forest, maka dapat diperoleh beberapa kesimpulan sebagai berikut:

1. Proses analisis sentimen pada ulasan aplikasi JobStreet di Google Play Store dilakukan melalui serangkaian tahapan yang sistematis. Tahap pertama adalah pengumpulan data menggunakan teknik scraping dari Google Play Store yang menghasilkan 9.000 data ulasan. Tahap kedua adalah seleksi data untuk menghapus data kosong dan duplikat, yang menghasilkan 6.510 data ulasan setelah menghapus 2.490 data duplikat. Tahap ketiga adalah pelabelan sentimen berdasarkan rating pengguna, di mana rating 4-5 dikategorikan sebagai sentimen positif dan rating 1-3 dikategorikan sebagai sentimen negatif. Tahap keempat adalah preprocessing data yang meliputi cleaning, case folding, tokenizing, stopwords removal, dan stemming untuk membersihkan dan menstandarkan teks ulasan. Tahap kelima adalah ekstraksi fitur menggunakan metode TF-IDF yang menghasilkan matriks berukuran (6510, 6250), yang menunjukkan terdapat 6.510 data dan 6.250 fitur kata unik. Tahap terakhir adalah pembagian data menjadi data training (80%) dan data testing (20%) untuk proses klasifikasi.
2. Penerapan teknik SMOTE dalam mengatasi ketidakseimbangan data pada proses analisis sentimen ulasan aplikasi JobStreet berhasil menyeimbangkan

jumlah data pada kelas sentimen dalam data training. Sebelum penerapan SMOTE, jumlah data training pada kelas positif adalah 2.974 data, sedangkan kelas negatif sebanyak 2.234 data, yang menunjukkan adanya ketidakseimbangan data. Setelah diterapkan SMOTE, jumlah data pada kelas negatif meningkat menjadi 2.974 data, sehingga jumlah data pada kedua kelas menjadi seimbang masing-masing 2.974 data. SMOTE hanya diterapkan pada data training agar data testing tetap merepresentasikan kondisi data asli dan evaluasi model tetap objektif. Meskipun demikian, teknik ini berpotensi menimbulkan overfitting karena data sintetis yang dihasilkan tidak sepenuhnya mewakili variasi data asli di dunia nyata, terutama jika data minoritas memiliki banyak noise.

3. Performa algoritma XGBoost dan Random Forest setelah diterapkan SMOTE dalam klasifikasi sentimen ulasan aplikasi JobStreet menunjukkan bahwa kedua algoritma memiliki performa yang baik. Algoritma XGBoost memperoleh nilai accuracy sebesar 84,79% (0,847926), precision sebesar 84,85% (0,848453), recall sebesar 84,79% (0,847926), dan F1-score sebesar 84,81% (0,848117). Sementara itu, algoritma Random Forest memperoleh nilai accuracy sebesar 84,56% (0,845622), precision sebesar 85,47% (0,854743), recall sebesar 84,56% (0,845622), dan F1-score sebesar 84,64% (0,846411). Berdasarkan hasil tersebut, algoritma XGBoost memiliki performa yang sedikit lebih baik dibandingkan Random Forest karena memperoleh nilai accuracy dan F1-score yang lebih tinggi. Namun, selisihnya sangat tipis sekitar 0,23% sehingga secara statistik kedua algoritma dianggap memiliki kinerja yang setara dalam klasifikasi sentimen ulasan aplikasi

JobStreet. Secara umum, hasil pembahasan sentimen menunjukkan bahwa ulasan positif pengguna aplikasi JobStreet berkaitan dengan kemudahan mencari pekerjaan dan manfaat aplikasi, sedangkan ulasan negatif lebih banyak membahas kendala teknis seperti masalah login, aplikasi sulit dibuka, dan kesulitan saat melamar pekerjaan. Hasil penelitian ini diharapkan dapat menjadi masukan bagi pengembang aplikasi JobStreet untuk meningkatkan kualitas layanan dan pengalaman pengguna.

4. Hasil pembahasan sentimen menunjukkan bahwa ulasan positif pengguna aplikasi JobStreet umumnya berkaitan dengan kemudahan mencari pekerjaan, manfaat aplikasi, dan proses pencarian lowongan kerja. Sementara itu, ulasan negatif banyak berkaitan dengan kendala teknis, seperti aplikasi sulit dibuka, masalah login, aplikasi keluar sendiri, serta kendala dalam proses melamar pekerjaan.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk penelitian selanjutnya maupun pihak pengembang aplikasi, yaitu sebagai berikut:

1. Penelitian selanjutnya dapat menggunakan jumlah data yang lebih besar dan periode pengambilan data yang lebih luas agar hasil analisis sentimen menjadi lebih representatif terhadap kondisi pengguna aplikasi JobStreet.
2. Penelitian selanjutnya dapat menambahkan kelas sentimen netral agar analisis sentimen menjadi lebih detail. Pada penelitian ini, sentimen hanya dibagi menjadi dua kelas, yaitu positif dan negatif, sehingga ulasan dengan rating 3 dimasukkan ke dalam kategori negatif.

3. Penelitian selanjutnya dapat membandingkan algoritma lain, seperti Support Vector Machine, Naive Bayes, Logistic Regression, LightGBM, atau metode deep learning seperti LSTM dan BERT, untuk mengetahui algoritma mana yang menghasilkan performa terbaik dalam klasifikasi sentimen.
4. Pada tahap preprocessing, penelitian selanjutnya dapat menggunakan kamus normalisasi bahasa Indonesia yang lebih lengkap agar kata-kata tidak baku, singkatan, dan bahasa informal pada ulasan pengguna dapat diproses dengan lebih baik.
5. Penelitian selanjutnya disarankan untuk melakukan optimasi hyperparameter pada algoritma XGBoost dan Random Forest menggunakan metode seperti Grid Search atau Bayesian Optimization. Optimasi hyperparameter dapat meningkatkan performa model secara signifikan dibandingkan penggunaan parameter default seperti yang digunakan dalam penelitian ini.
6. Bagi pengembang aplikasi JobStreet, hasil penelitian ini dapat dijadikan sebagai masukan untuk meningkatkan kualitas aplikasi, terutama pada aspek teknis seperti perbaikan masalah login, aplikasi sulit dibuka, aplikasi keluar sendiri, serta kendala saat pengguna mengirim lamaran kerja.
7. Pengembang aplikasi juga dapat meningkatkan relevansi rekomendasi lowongan kerja agar pengguna mendapatkan informasi pekerjaan yang lebih sesuai dengan kebutuhan, pengalaman, dan lokasi mereka.

DAFTAR PUSTAKA

- [1] J. co. P. Limited, “Jobstreet: Smart job matching,” Google Play Store. Accessed: Jun. 10, 2026. [Online]. Available: <https://play.google.com/store/apps/details?id=com.jobstreet.jobstreet>
- [2] F. Marzuqi, Y. Purba, T. Syaputra, W. Bismi, I. Kurniawati, and R. Fahlapi, “Perbandingan Algoritma Machine Learning Untuk Sentimen Ulasan Aplikasi Jobstreet,” *INFOTECH J.*, vol. 11, no. 2, pp. 416–423, 2025, doi: 10.31949/infotech.v11i2.16609.
- [3] I. G. B. A. Budaya and I. K. P. Suniantara, “Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1077–1086, 2024, doi: 10.57152/malcom.v4i3.1459.
- [4] Bayu Nur Setiyono, Nadia Annisa Maori, and Teguh Tamrin, “Analisis Sentimen Ulasan Pengguna Aplikasi Threads di Google Play Menggunakan Algoritma XGBoost Dengan Penguatan SMOTE,” *J. Inform. Dan Teknologi Komput.*, vol. 5, no. 2, pp. 119–130, 2025, doi: 10.55606/jitek.v5i2.6832.
- [5] M. F. Rahmatullah, P. L. Lokapitasari Belluano, and H. Darwis, “Analisis Sentimen Review Aplikasi di Google Play Store Menggunakan Random Forest,” *LINIER Lit. Inform. dan Komput.*, vol. 2, no. 3, pp. 380–389, 2025, doi: 10.33096/linier.v2i3.3149.
- [6] J. Riset and M. Edukasi, “Analisis Ulasan Aplikasi Dalam Google Play Store Menggunakan,” vol. 3, pp. 1–17, 2026.
- [7] S. Developers, “TfidfVectorizer — scikit-learn documentation,” Scikit-learn. Accessed: Jun. 10, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
- [8] H. Mansourifar and W. Shi, “Deep Synthetic Minority Over-Sampling Technique,” vol. 16, pp. 321–357, 2020, [Online]. Available: <http://arxiv.org/abs/2003.09788>
- [9] S. Kiyohara, T. Miyata, and T. Mizoguchi, “Prediction of grain boundary structure and energy by machine learning,” vol. 18, pp. 1–5, 2015, [Online]. Available: <http://arxiv.org/abs/1512.03502>
- [10] H. Ali *et al.*, “Hybrid TF-IDF and SBERT Feature Engineering with SMOTE-Tomek Resampling for Enhanced Software Requirements Classification,” no. Mcc, pp. 1–19, 2026.
- [11] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Aug, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

- [12] J. Sistem, “Analisis Sentimen Ulasan Aplikasi M-Pajak pada Google Play Store,” vol. 3, no. 1, pp. 11–25, 2025.
- [13] S. Developers, “classification_report — scikit-learn documentation,” Scikit-learn. Accessed: Jun. 10, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html
- [14] P. Bairagi and S. Arora, “A mini-approach for retail banking with transparent recommendation system enabled by explainable AI,” *Discov. Artif. Intell.*, vol. 6, no. 1, 2026, doi: 10.1007/s44163-025-00817-5.
- [15] S. Developers, “f1_score — scikit-learn documentation,” Scikit-learn. Accessed: Jun. 10, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html
- [16] F. S. Shantika and Z. Abidin, “Sentiment Analysis of Jobstreet Application Reviews on Google Play Store Using Support Vector Machine Algorithm with Adaptive Synthetic,” *Recursive J. Informatics*, vol. 3, no. 1, pp. 99–107, 2025, doi: 10.15294/rji.v3i2.11891.
- [17] B. Basiroh and Widya Novita Al Afifah Irwanto, “Analisis Sentimen Negative Pada Aplikasi Jobstreet Menggunakan Hadoop Distributed File System (Hdfs),” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 4, no. 4, pp. 287–293, 2025, doi: 10.55123/storage.v4i4.6338.
- [18] I. P. Gede, A. Sudiatmika, D. Ayu, M. Putriari, and W. A. Putera, “Analisis Sentimen Aplikasi M-Pajak Menggunakan Machine Learning Berbasis Algoritma Xgboost,” vol. 5, no. 1, pp. 2377–2383, 2026.
- [19] S. A. S. Mola, D. L. B. Baun, I. O. Nunes, and M. M. A. R. Sani, “Analisis Sentimen Aplikasi Halo Bca Di Google Play Store Menggunakan Metode Naive Bayes, Support Vector Machine Dan Random Forest,” *HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf.)*, vol. 15, no. 2, pp. 69–79, 2024, doi: 10.52972/hoaq.vol15no2.p69-79.
- [20] A. Sentimen *et al.*, “TA - Siti Rohimah,” 2025.
- [21] H. Barus, I. N. Fajri, and Y. Pristyanto, “Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models,” *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2552–2561, 2025, doi: 10.30871/jaic.v9i5.10524.
- [22] T. et al. Chen, “XGBoost Documentation,” XGBoost Developers. Accessed: Jun. 10, 2026. [Online]. Available: <https://xgboost.readthedocs.io/en/stable/index.html>
- [23] Jobstreet by SEEK, “Tentang SEEK di Asia,” Jobstreet.co.id. [Online]. Available: <https://www.jobstreet.co.id/id/about>. [Accessed: Jul. 28, 2026].