

SURAT KETERANGAN BEBAS PLAGIARISME



Universitas Nusantara PGRI Kediri
UPT. PERPUSTAKAAN, PUBLIKASI DAN INOVASI
Alamat: Kampus 1, Jl. KH. Ahmad Dahlan No.76 Kota Kediri 64112
Telp. (0354) 771576, (0354) 771503, (0354) 771495, Fax. (0354) 771576
Website: <http://ppi.unpkediri.ac.id/> Email: perpustakaan@unpkediri.ac.id

SURAT KETERANGAN BEBAS SIMILARITY

Ketua UPT Perpustakaan, Publikasi dan Inovasi Universitas Nusantara PGRI Kediri menerangkan bahwa mahasiswa dengan identitas berikut:

Nama Mahasiswa : Eka Fauziah
NPM : 2213030030
Program Studi : S1-Sistem Informasi

Judul Karya Ilmiah:

“Naïve Bayes dan Lexicon Sentiment untuk Deteksi Makna Mengenai Kebijakan Tunjangan DPR RI”

Dinyatakan sudah memenuhi syarat batas maksimal 30% *similarity* sesuai dengan ketentuan yang berlaku pada setiap subbab naskah Laporan **Tugas Akhir/Skripsi/Tesis** yang disusun.

Demikian Surat Keterangan ini kami berikan untuk dapat dipergunakan seperlunya.



Kediri, 09 Juni 2026
Ka UPT PPI,



Dr. Abdul Aziz Hunaifi, M.A

Turnitin LLCC

2213030030_EKA FAUZIAH

 SI-14

Document Details

Submission ID

trn:oid:::2945:392192447

Submission Date

Jun 9, 2026, 11:26 AM GMT+7

Download Date

Jun 9, 2026, 11:28 AM GMT+7

File Name

2213030030_EKA FAUZIAH.pdf

File Size

4.5 MB

51 Pages

11,865 Words

78,714 Characters

29% Overall Similarity

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

Filtered from the Report

- Bibliography

Top Sources

- 15%  Internet sources
- 15%  Publications
- 26%  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

- 15% Internet sources
- 15% Publications
- 26% Submitted works (Student Papers)

Top Sources

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

1	Publication	Paulina Gorat Frans, Frans Steven Pakpahan, Sardo Pardingotan Sipayung. "Anal...	<1%
2	Student papers	Universitas Pancasila on 2026-03-05	<1%
3	Internet	openjournal.unpam.ac.id	<1%
4	Internet	ejurnal.stmik-budidarma.ac.id	<1%
5	Student papers	Universitas Budi Luhur on 2026-02-07	<1%
6	Student papers	Universitas Pancasila on 2026-05-08	<1%
7	Student papers	Universitas Pertahanan RI on 2026-01-05	<1%
8	Publication	Dedi Irawan, Yuyu Sri Rahayu, Mohamad Faroz. "KLASIFIKASI SENTIMEN PELAN...	<1%
9	Publication	Adinda Purnamasari, Rini Astuti, Khaerul Anam, Gifthera Dwilestari, Mulyawan. ...	<1%
10	Student papers	UNIVERSITAS BUDI LUHUR on 2026-01-30	<1%
11	Internet	ejournal.itn.ac.id	<1%

12	Publication	Muhammad Chairul Azmi, Muhammad Ikhsan. "Analisis Sentimen Wacana Publik..."	<1%
13	Student papers	Universitas Putera Batam on 2026-01-31	<1%
14	Publication	Saiful Nur Budiman, Sri Lesanti, Erwan. "Analisis Sentimen Berdasarkan Hasil Re..."	<1%
15	Student papers	UPN Veteran Yogyakarta on 2026-05-05	<1%
16	Internet	ejurnal.provisi.ac.id	<1%
17	Internet	repository.unpkediri.ac.id	<1%
18	Student papers	UNIVERSITAS BUDI LUHUR on 2026-01-22	<1%
19	Student papers	Universitas Budi Luhur on 2026-01-25	<1%
20	Student papers	Universitas Islam Indonesia on 2026-05-12	<1%
21	Student papers	Universitas Putera Batam on 2026-01-29	<1%
22	Student papers	Universitas Pamulang on 2025-10-21	<1%
23	Internet	journal.ipm2kpe.or.id	<1%
24	Student papers	Universitas Pamulang on 2026-05-04	<1%
25	Student papers	Universitas Putera Batam on 2026-02-03	<1%

26	Internet	docplayer.info	<1%
27	Student papers	Universitas Putera Batam on 2026-02-02	<1%
28	Publication	Muhammad Irvan Maulana, Savana Putra Aditama, Harun Al Rosyid. "Komparasi...	<1%
29	Student papers	Universitas Muhammadiyah Purwokerto on 2024-11-26	<1%
30	Student papers	Universitas Brawijaya on 2017-06-21	<1%
31	Student papers	Universitas Putera Batam on 2026-02-02	<1%
32	Student papers	Sriwijaya University on 2026-05-04	<1%
33	Student papers	Universitas Budi Luhur on 2025-07-12	<1%
34	Internet	jurnal.polibatam.ac.id	<1%
35	Publication	Muhamad Azril, Giatika Crisnawati. "Klasifikasi Sentimen Komentar Pengguna Tik...	<1%
36	Student papers	Sekolah Teknik Elektro & Informatika on 2025-07-23	<1%
37	Student papers	Universitas Muslim Indonesia on 2024-12-13	<1%
38	Internet	repositorii.urindo.ac.id	<1%
39	Internet	repository.ub.ac.id	<1%

40	Publication	Emil Fatra, Thiara Tri Funny Manguma, Mastura. "Analisis Sentimen Media Sosial ..."	<1%
41	Student papers	Universitas Khairun on 2026-05-18	<1%
42	Internet	ojs.stmik-banjarbaru.ac.id	<1%
43	Internet	proceeding.unpkediri.ac.id	<1%
44	Internet	"IMPLEMENTASI ALGORITMA NAÏVE BAYES MENGGUNAKAN ISEAR UNTUK KLASIFI..."	<1%
45	Student papers	Fakultas ISIP on 2026-02-24	<1%
46	Student papers	UPN Veteran Yogyakarta on 2026-04-02	<1%
47	Student papers	UPN Veteran Yogyakarta on 2026-04-30	<1%
48	Internet	jurnal.uinsu.ac.id	<1%
49	Student papers	Universitas Muslim Indonesia on 2025-07-16	<1%
50	Student papers	Universitas Putera Batam on 2026-02-03	<1%
51	Student papers	Universitas Jember on 2025-12-08	<1%
52	Student papers	Universitas Khairun on 2026-05-18	<1%
53	Internet	repository.itpln.ac.id	<1%

54	Publication	Ina Najiyah, Miftahul Rizal. "Analisis Sentimen Media Sosial X Terhadap Kebijakan...	<1%
55	Student papers	Politeknik Statistika STIS on 2025-07-05	<1%
56	Student papers	UIN Syarif Hidayatullah Jakarta on 2025-06-12	<1%
57	Student papers	Universitas Negeri Manado on 2025-04-13	<1%
58	Internet	jurnal.soratekno.co.id	<1%
59	Internet	openlibrarypublications.telkomuniversity.ac.id	<1%
60	Publication	Bayu Setiawan, M.Najamuddin Dwi Miharja, Edora. "Analisis Sentimen Restoran ...	<1%
61	Publication	Evip Athirah Salsabila, Mellysa Pratama, Putri Wahyuni, Vina Purwaningrum, Mufi...	<1%
62	Publication	Murtiwiayati Murtiwiayati, Risha Fitriawhalia Arini, Leli Safitri, Santi Widiyanti, Iwa...	<1%
63	Student papers	Universitas Budi Luhur on 2025-07-16	<1%
64	Student papers	Universitas Jember on 2025-06-11	<1%
65	Student papers	Universitas Muhammadiyah Jember on 2025-12-18	<1%
66	Student papers	Universitas Muhammadiyah Makassar on 2026-06-07	<1%
67	Internet	jpti.journals.id	<1%

68	Internet	sitasi.upnjatim.ac.id	<1%
69	Publication	Alfiansyah Hasibuan, Djami Olii, Bryan Chelvan Walujan, Bobby Putra Situmorang, ...	<1%
70	Publication	Marastrika Farhan Nur Haiban, Nur Sultan Salahuddin. "Ekstraksi Wawasan Dari ...	<1%
71	Student papers	Universitas Islam Riau on 2026-04-06	<1%
72	Student papers	Universitas Tarumanagara on 2025-12-23	<1%
73	Internet	etheses.uin-malang.ac.id	<1%
74	Internet	www.stmik-budidarma.ac.id	<1%
75	Student papers	Ciputra University on 2026-02-13	<1%
76	Publication	Fatkhurrohman Fatkhurrohman, Bangkit Indarmawan Nugroho, Nurul Fadillah. "...	<1%
77	Publication	Nugroho, Sapto Sri. "Manajemen Kurikulum Terpadu Berbasis Karakter Religius d...	<1%
78	Student papers	Universitas Brawijaya on 2019-07-03	<1%
79	Student papers	Universitas Muhammadiyah Purwokerto on 2025-02-27	<1%
80	Student papers	Universitas Muhammadiyah Purwokerto on 2025-05-09	<1%
81	Student papers	Universitas Pancasila on 2026-03-04	<1%

82	Internet	publikasi.dinus.ac.id	<1%
83	Student papers	Universitas Katolik Widya Mandala Surabaya on 2025-12-03	<1%
84	Student papers	Universitas Pancasila on 2025-08-07	<1%
85	Internet	dspace.uui.ac.id	<1%
86	Student papers	Sultan Agung Islamic University on 2025-07-08	<1%
87	Student papers	itera on 2021-05-11	<1%
88	Internet	journal.ubm.ac.id	<1%
89	Student papers	Institut Bisnis dan Teknologi Indonesia (INSTIKI) on 2025-06-10	<1%
90	Student papers	Universitas Brawijaya on 2017-07-06	<1%
91	Student papers	Universitas Brawijaya on 2018-12-12	<1%
92	Student papers	Universitas Darussalam Gontor on 2026-01-25	<1%
93	Student papers	Universitas Jember on 2021-06-28	<1%
94	Student papers	Universitas Muhammadiyah Purwokerto on 2025-07-10	<1%
95	Student papers	Universitas Muhammadiyah Sukabumi on 2025-07-16	<1%

96	Student papers	Universitas Muslim Indonesia on 2025-02-19	<1%
97	Student papers	Universiti Tun Hussein Onn Malaysia on 2026-01-31	<1%
98	Internet	repository.akba.ac.id	<1%
99	Publication	Khyntia Miranda, Ryan Randy Suryono. "Analisis Sentimen Pinjaman Online: Studi...	<1%
100	Student papers	Universitas Brawijaya on 2017-07-08	<1%
101	Student papers	Universitas Muslim Indonesia on 2025-08-19	<1%
102	Internet	eprints.ums.ac.id	<1%
103	Internet	id.123dok.com	<1%
104	Internet	repository.unpas.ac.id	<1%
105	Internet	repository.usni.ac.id	<1%
106	Internet	subset.id	<1%
107	Publication	Abidin Abidin. "ANALISIS SENTIMEN ULASAN APLIKASI SIREKAP MENGGUNAKAN ...	<1%
108	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2024-01-21	<1%
109	Publication	M. Khaerul Ihsan, Muhammad Maulana, Tanwir Tanwir, Abi Mas'ud, Naufal Hanif,...	<1%

110	Student papers	Universitas Brawijaya on 2021-03-30	<1%
111	Student papers	Universitas Muhammadiyah Jember on 2026-04-21	<1%
112	Student papers	Universitas Sumatera Utara on 2026-06-04	<1%
113	Internet	dinastipub.org	<1%
114	Internet	eprints.itn.ac.id	<1%
115	Internet	ojs.amikom.ac.id	<1%
116	Publication	Danang Arbian Sulisty, Erik Setiadi. "Analisis Sentimen Kebijakan Makan Bergizi ...	<1%
117	Publication	Dedi Junaedi, Faisal Salistia. "Dampak Kawasan dan Status Negara Terhadap Pert...	<1%
118	Student papers	Telkom University on 2026-02-07	<1%
119	Student papers	UNIVERSITAS BUDI LUHUR on 2025-08-23	<1%
120	Student papers	UNIVERSITAS BUDI LUHUR on 2025-09-18	<1%
121	Student papers	UPN Veteran Jakarta on 2025-01-30	<1%
122	Student papers	Universitas Brawijaya on 2021-04-08	<1%
123	Student papers	Universitas Muhammadiyah Purwokerto on 2024-07-17	<1%

124	Student papers	Universitas Muslim Indonesia on 2025-08-14	<1%
125	Student papers	Universitas Pendidikan Ganesha on 2024-02-05	<1%
126	Student papers	Universitas Pendidikan Ganesha on 2024-02-10	<1%
127	Internet	www.jurnal.stkipggritlungagung.ac.id	<1%
128	Internet	123dok.com	<1%
129	Publication	Alvina Lia, Abdul Rahim, Taghfirul Azhima Yoga Siswa. "ANALISIS SENTIMEN APLI...	<1%
130	Publication	Gayus Gregorius Ferdinand Djema, Ozzi Suria. "Public Sentiment Analysis of Dan...	<1%
131	Publication	Herdiesel Santoso, Wahyuni Nareswari. "Analisis Sentimen Berbasis Aspek Pada ...	<1%
132	Publication	Kaswili Sriwenda Putri, Iwan Rizal Setiawan, Agung Pambudi. "ANALISIS SENTIME...	<1%
133	Publication	Muhammad Reyan, Franindya Purwaningtyas. "ANALISIS SENTIMEN ULASAN APL...	<1%
134	Student papers	STT PLN on 2022-08-11	<1%
135	Student papers	Sriwijaya University on 2020-03-05	<1%
136	Student papers	Sultan Agung Islamic University on 2024-08-26	<1%
137	Publication	Tarwoto, Rizki Nugroho, Najmul Azka, Wakhid Sayudha Rendra Graha. "Analisis S...	<1%

138	Student papers	UIN Raden Intan Lampung on 2026-05-13	<1%
139	Student papers	Universitas Brawijaya on 2019-07-05	<1%
140	Student papers	Universitas Djuanda on 2025-07-15	<1%
141	Student papers	Universitas Komputer Indonesia on 2025-10-23	<1%
142	Student papers	Universitas Muslim Indonesia on 2025-07-31	<1%
143	Student papers	Universitas Pancasila on 2025-08-07	<1%
144	Student papers	Universitas Pendidikan Ganesha on 2025-07-24	<1%
145	Student papers	Universitas Putera Batam on 2025-07-28	<1%
146	Student papers	Universitas Putera Batam on 2026-02-02	<1%
147	Student papers	Universitas Putera Batam on 2026-02-03	<1%
148	Student papers	Universitas Tarumanagara on 2025-12-19	<1%
149	Student papers	Universitas Tarumanagara on 2025-12-23	<1%
150	Student papers	University of Southern Mississippi on 2023-02-21	<1%
151	Internet	digilib.uin-suka.ac.id	<1%

152	Internet	eprints.perbanas.ac.id	<1%
153	Internet	journal.artei.or.id	<1%
154	Internet	journal.irpi.or.id	<1%
155	Internet	journal.nurulfikri.ac.id	<1%
156	Internet	repository.mercubuana.ac.id	<1%
157	Internet	repository.unissula.ac.id	<1%
158	Internet	www.coursehero.com	<1%
159	Internet	www.yumpu.com	<1%
160	Publication	Afifah Nurul Izzati. "ANALISIS SENTIMEN HASIL PUTUSAN MK TERKAIT SENGKETA ...	<1%
161	Publication	Azka Bima Aditya, Syafri Samsudin, Winahyu Pandu Rizki, Mahir Mahendra, Arif S...	<1%
162	Student papers	Institut Bisnis dan Teknologi Indonesia (INSTIKI) on 2025-05-30	<1%
163	Student papers	Institut Bisnis dan Teknologi Indonesia (INSTIKI) on 2025-09-16	<1%
164	Publication	Junaedi, Alexius Hendra Gunawan, Verri Kuswanto, Jonathan. "Tinjauan Support ...	<1%
165	Publication	Rahmawati Rahmawati, Nurhikma Arifin, Wawan Firgiawan. "Perbandingan SVM ...	<1%

166	Publication	Rani Nooraeni, Heny Dwi Sariyanti, Aulia Fikri Fadhilah Iskandar, Siti Fatimatul M...	<1%
167	Student papers	UIN Sultan Syarif Kasim Riau on 2021-02-26	<1%
168	Student papers	UIN Sunan Ampel Surabaya on 2023-03-14	<1%
169	Student papers	UPN Veteran Yogyakarta on 2026-05-04	<1%
170	Student papers	Universitas Brawijaya on 2018-01-02	<1%
171	Student papers	Universitas Budi Luhur on 2023-07-05	<1%
172	Student papers	Universitas Islam Indonesia on 2026-04-20	<1%
173	Student papers	Universitas Islam Riau on 2025-10-31	<1%
174	Student papers	Universitas Jember on 2026-04-06	<1%
175	Student papers	Universitas Klabat on 2019-06-26	<1%
176	Student papers	Universitas Kristen Satya Wacana on 2025-12-03	<1%
177	Student papers	Universitas Muhammadiyah Palembang on 2025-08-06	<1%
178	Student papers	Universitas Muhammadiyah Purwokerto on 2024-12-24	<1%
179	Student papers	Universitas Muhammadiyah Purwokerto on 2025-05-27	<1%

180	Student papers	Universitas Putera Batam on 2025-07-26	<1%
181	Student papers	Universitas Putera Batam on 2026-02-03	<1%
182	Student papers	Universitas Tarumanagara on 2025-12-11	<1%
183	Student papers	Universitas Tarumanagara on 2026-01-02	<1%
184	Internet	cdn.juris.id	<1%
185	Internet	ejournal.uhb.ac.id	<1%
186	Internet	etd.repository.ugm.ac.id	<1%
187	Internet	id.m.wikipedia.org	<1%
188	Internet	journal.ilmudata.co.id	<1%
189	Internet	mafiadoc.com	<1%
190	Internet	managementworld.online	<1%
191	Internet	ocs.unud.ac.id	<1%
192	Internet	prosiding-senada.upnjatim.ac.id	<1%
193	Internet	repository.uin-suska.ac.id	<1%

194	Internet	repository.upnjatim.ac.id	<1%
195	Internet	ujdsd.ppj.unp.ac.id	<1%
196	Publication	Achmad Hakim Qoirul Haq, Harminto Mulyo, Adi Sucipto. "Optimasi Kinerja Algor...	<1%
197	Publication	Ayudha Kusuma Ramadhani. "Perbandingan Algoritma K-Nearest Neighbor dan ...	<1%
198	Publication	Aziz Perdana, Arief Hermawan, Donny Avianto. "Analisis Sentimen Terhadap Isu P...	<1%
199	Publication	Ivan Aditya Nugraha, Irving Vitra Paputungan. "Analisis Sentimen Video Review S...	<1%
200	Publication	Lisyo Hileria Sarumpaet, Ryan Randy Suryono. "Analisis Sentimen Publik Program...	<1%
201	Publication	Luban Abdi Susanto. "KOMPARASI MODEL SUPPORT VECTOR MACHINE DAN K-NE...	<1%
202	Publication	Mukhtar Nashir, Dian Ade Kurnia, Yudhistira Arie Wijaya, Ade Irma Purnama Sa...	<1%
203	Publication	Oktaviani Oktaviani, Endah Kurniasari, Rosdiana Rosdiana, Faturachman Kurniaw...	<1%
204	Student papers	Sriwijaya University on 2021-07-08	<1%
205	Student papers	Telkom University on 2026-06-08	<1%
206	Student papers	Universitas Brawijaya on 2021-04-28	<1%
207	Student papers	Universitas Negeri Manado on 2025-05-05	<1%

208	Student papers	Universitas Tarumanagara on 2025-12-17	<1%
209	Internet	journal.um-surabaya.ac.id	<1%
210	Student papers	Ajou University Graduate School on 2024-01-11	<1%
211	Student papers	FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM on 2026-01-06	<1%
212	Student papers	Fakultas Teknik on 2025-08-07	<1%
213	Student papers	Politeknik Negeri Bandung on 2017-08-04	<1%
214	Student papers	Sekolah Global Jaya on 2017-11-02	<1%
215	Student papers	Telkom University on 2025-12-27	<1%
216	Student papers	UNIVERSITAS BUDI LUHUR on 2025-08-19	<1%
217	Student papers	UPN Veteran Yogyakarta on 2026-02-27	<1%
218	Student papers	Universitas Amikom on 2023-07-12	<1%
219	Student papers	Universitas Budi Luhur on 2026-01-31	<1%
220	Student papers	Universitas Islam Riau on 2025-07-17	<1%
221	Student papers	Universitas Kristen Wira Wacana Sumba on 2025-11-03	<1%

222	Student papers	Universitas Muhammadiyah Jember on 2026-01-17	<1%
223	Student papers	Universitas Pamulang on 2026-05-21	<1%
224	Student papers	Universitas Pancasila on 2026-02-21	<1%
225	Student papers	Universitas Pancasila on 2026-03-02	<1%
226	Student papers	Universitas Putera Batam on 2021-03-16	<1%
227	Student papers	Universitas Putera Batam on 2026-02-02	<1%
228	Student papers	Universitas Tarumanagara on 2026-01-04	<1%

BAB I

PENDAHULUAN

Bab ini membahas latar belakang penelitian mengenai analisis sentimen opini publik terhadap kebijakan tunjangan DPR RI pada media sosial X (*Twitter*). Selain itu, bab ini juga memuat batasan masalah, rumusan masalah, tujuan penelitian, serta manfaat penelitian sebagai dasar dan arah pelaksanaan penelitian.

A. Latar Belakang Masalah

Perkembangan teknologi informasi telah memberikan perubahan besar dalam kehidupan masyarakat, khususnya dalam hal berkomunikasi, menukar informasi, dan menyampaikan pendapat di lingkungan publik. Adanya berbagai platform media sosial seperti Twitter (X), Facebook, Instagram, dan platform digital lainnya memungkinkan masyarakat untuk menyampaikan pendapat mereka mengenai berbagai topik yang sedang hangat dibicarakan, mulai dari isu sosial, ekonomi, hingga kebijakan politik. Aktivitas itu menghasilkan banyak data berupa teks yang bisa digunakan untuk memahami bagaimana masyarakat memandang dan bereaksi terhadap suatu peristiwa atau kebijakan tertentu. Fahrudin dan timnya (2022) menyatakan bahwa data dari media sosial bisa dipakai untuk menggambarkan opini masyarakat dengan menganalisis perasaan dan sikap orang-orang, sehingga memberikan pemahaman yang lebih jelas tentang bagaimana masyarakat merespons sesuatu.

Isu yang sering diperbincangkan oleh masyarakat adalah soal kebijakan tunjangan yang diberikan kepada anggota Dewan Perwakilan Rakyat Republik Indonesia (DPR RI). Kebijakan ini memunculkan beragam tanggapan dari masyarakat. Beberapa orang melihat tunjangan itu sebagai bentuk bantuan untuk meningkatkan kinerja lembaga legislatif, tetapi ada pihak lain yang merasa kebijakan tersebut tidak cocok dengan kondisi sosial dan ekonomi masyarakat. Berbagai respons itu sering dikirimkan lewat media sosial dengan berbagai bentuk seperti komentar, percakapan, atau postingan yang menunjukkan perasaan dan pandangan masyarakat terhadap kebijakan yang diterapkan. Menurut Wijaya dan Pratiwi (2021), media sosial, terutama Twitter, telah menjadi sumber data yang sangat penting untuk mengamati dan menganalisis pendapat masyarakat terhadap berbagai kebijakan yang diambil pemerintah.

Pendapat yang disampaikan oleh masyarakat lewat media sosial tidak hanya menunjukkan apakah mereka setuju atau tidak setuju, tetapi juga menyampaikan perasaan

seperti marah, kecewa, mendukung, berharap, atau bahkan menyindir. Maka itu, dibutuhkan pendekatan yang dapat memproses dan menerjemahkan data teks secara teratur agar informasi yang terdapat di dalamnya bisa dipahami secara jelas dan adil. Salah satu cara yang sering dipakai adalah analisis sentimen, yaitu metode yang bertujuan untuk mengetahui dan mengelompokkan pendapat berdasarkan jenis polaritas, seperti positif, negatif, atau netral. Asri dkk. (2022) mengatakan bahwa penggabungan metode klasifikasi dengan pendekatan berbasis leksikon bisa membantu dalam mengidentifikasi sentimen serta memberikan gambaran yang lebih terstruktur mengenai distribusi opini publik. Selain itu, Putri dkk. (2022) menyatakan bahwa metode klasifikasi berhasil memberikan tingkat akurasi yang memadai dalam pemrosesan data teks, sehingga dapat membantu dalam mengambil keputusan.

Analisis sentimen adalah bagian dari Natural Language Processing (NLP) yang bertujuan memahami arti dan pendapat yang terdapat dalam teks. Dengan cara ini, peneliti bisa memahami bagaimana pendapat masyarakat berkembang terhadap suatu isu berdasarkan data yang sudah ada. Saron Tandiapa dan Caren Rorimpandey (2024) mengatakan bahwa kemajuan media digital memberikan kesempatan besar untuk melakukan penelitian analisis sentimen, karena opini publik bisa diperoleh secara langsung dan dalam jumlah yang besar melalui berbagai platform daring.

Dalam penerapannya, metode Naïve Bayes Classifier sering digunakan sebagai salah satu algoritma untuk mengklasifikasikan sentimen. Algoritma ini menggunakan perhitungan yang tidak terlalu rumit, cepat, dan bisa mengolah banyak teks sekaligus. Dimas Sanjaya dan timnya (2022) menjelaskan bahwa Naïve Bayes Classifier adalah metode berbasis probabilitas yang bekerja dengan mengasumsikan bahwa setiap fitur dalam data saling independen, sehingga metode ini sangat efektif digunakan dalam berbagai penelitian klasifikasi teks, termasuk analisis sentimen terhadap isu-isu sosial dan politik.

Selain menggunakan pendekatan berbasis machine learning, analisis sentimen juga bisa dilakukan dengan metode berbasis kamus. Pendekatan ini menggunakan daftar kata yang sudah memiliki tingkat perasaan tertentu untuk mengetahui arah perasaan dalam sebuah teks. Menurut Saron Tandiapa dan Caren Rorimpandey (2024), metode berbasis leksikon mampu mengenali makna emosional dalam teks dengan baik karena penentuan sentimen dilakukan berdasarkan ciri-ciri kata yang digunakan oleh penulis.

Meskipun begitu, menggunakan satu metode saja seringkali tidak cukup untuk memahami tingkat kesulitan dan kompleksitas bahasa yang digunakan di media sosial.

159 214 Bahasa yang digunakan di media sosial biasanya tidak resmi, menggunakan singkatan, campuran bahasa, bahkan ada unsur sarkasme yang bisa memengaruhi hasil pengklasifikasian. Oleh karena itu, diperlukan pendekatan yang lebih lengkap dengan menggabungkan metode pembelajaran mesin dan metode berbasis kamus agar hasil analisis menjadi lebih tepat dan memberi informasi yang lebih baik.

86 Beberapa penelitian sebelumnya menunjukkan bahwa menggabungkan kedua metode tersebut dapat meningkatkan kualitas analisis sentimen. Amrullah dan Solichin (2024) menyatakan bahwa model Multinomial Naïve Bayes yang digunakan untuk menganalisis komentar publik dari program Mata Najwa berhasil mencapai tingkat akurasi sebesar 90,67%. Sementara itu, penelitian yang dilakukan oleh Fitriono dan timnya pada tahun 2025 menunjukkan bahwa menggabungkan metode berbasis kamus dengan algoritma pembelajaran mesin mampu mencapai tingkat akurasi hingga 96,49% dalam menganalisis ulasan produk digital. Penelitian tersebut juga menekankan bahwa ciri khas bahasa Indonesia di media sosial, seperti penggunaan kata yang tidak baku, bahasa daerah, dan ucapan sindir, masih menjadi hambatan dalam pengelompokan otomatis.

Berdasarkan penelitian-penelitian sebelumnya, terlihat bahwa kebanyakan studi masih hanya fokus pada klasifikasi polaritas sentimen, tanpa memperhatikan lebih dalam aspek emosional yang mendasari pendapat masyarakat. Padahal, memahami perasaan yang terkandung dalam suatu pendapat bisa memberikan informasi yang lebih dalam tentang bagaimana masyarakat memandang sebuah kebijakan. Kondisi itu menunjukkan adanya kesempatan untuk melakukan penelitian yang menggabungkan analisis sentimen dan analisis emosi, khususnya terkait isu kebijakan politik yang sedang menjadi perhatian masyarakat.

3 14 128 46 Berdasarkan hal tersebut, penelitian ini dilakukan untuk menganalisis opini masyarakat terhadap kebijakan tunjangan anggota DPR RI dengan menggabungkan teknik Naïve Bayes Classifier dan analisis sentimen berbasis kamus. Metode ini diharapkan bisa tidak hanya mengetahui perasaan masyarakat, tetapi juga memahami lebih dalam arti emosional yang terkandung dalam pendapat publik. Oleh karena itu, penelitian ini diharapkan bisa membantu dalam mengembangkan studi tentang analisis sentimen dalam bahasa Indonesia sekaligus memberikan gambaran yang lebih lengkap mengenai bagaimana masyarakat bereaksi terhadap kebijakan pemerintah di masa kini yang semakin digital.

B. Batasan Masalah

Untuk memastikan penelitian ini tetap terarah dan fokus, batasan-batasan berikut diterapkan:

1. Data yang dipakai hanya berasal dari media sosial X (*Twitter*), yang berisi pendapat dan respons masyarakat terkait kebijakan tunjangan DPR RI. Analisis *sentiment*
2. Analisis sentimen dibatasi dalam tiga kategori utama, yaitu positif, negatif, dan netral.
3. Metode yang digunakan mencakup algoritma *Naïve Bayes Classifier* dengan pendekatan berbasis kamus (*Lexicon-Based*).
4. Bahasa yang dianalisis hanya berupa bahasa Indonesia.
5. Penilaian kinerja model dilakukan dengan menggunakan beberapa metrik, yaitu akurasi, presisi, *recall*, dan *F1-score*.

C. Rumusan Masalah

Berdasarkan masalah yang telah dijelaskan pada latar belakang, rumusan masalah dalam penelitian ini sebagai berikut:

Seberapa efektif metode *Naïve Bayes Classifier* dan *Lexicon Sentiment* dalam mengklasifikasikan *sentiment* serta mengidentifikasi makna emosional pada opini publik terhadap kebijakan tunjangan DPR RI berdasarkan data opini di platform X (*Twitter*)?, dan efektifkah integrasi antara metode *Naïve Bayes Classifier* dan *Lexicon Sentiment* dalam meningkatkan akurasi klasifikasi *sentiment* dibandingkan penggunaan masing-masing metode secara terpisah?

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk:

1. Menerapkan metode *Naïve Bayes Classifier* dan *Lexicon Sentiment* untuk mengklasifikasikan *sentiment* opini publik terkait kebijakan tunjangan DPR RI menjadi kategori positif, negatif, dan netral.
2. Mengidentifikasi kekuatan dan kelemahan penerapan kedua metode tersebut dalam pengolahan data opini publik dari media sosial.
3. Menyajikan hasil analisis *sentiment* berupa pola dan tren persepsi publik terhadap kebijakan tunjangan DPR RI yang diperoleh dari media sosial.

4. Memberikan rekomendasi teknis dan metodologis untuk perkembangan penelitian selanjutnya

A. Manfaat Penelitian

Berdasarkan tujuan penelitian yang sudah ditentukan, penelitian ini diharapkan bisa memberikan beberapa manfaat seperti berikut:

1. Penelitian ini membantu memperkaya pemahaman tentang analisis sentimen dalam bahasa Indonesia dengan menerapkan metode *Naïve Bayes Classifier* dan kamus sentimen dalam konteks sosial-politik, terutama dalam memahami opini masyarakat terhadap kebijakan tunjangan Dewan Perwakilan Rakyat Republik Indonesia.
2. Penelitian ini dapat digunakan sebagai acuan bagi para akademisi untuk memahami cara kerja pendekatan berbasis pembelajaran mesin dan pendekatan berbasis kamus dalam mengenali sikap dan perasaan dari teks media sosial.
3. Hasil penelitian memberikan gambaran bagaimana analisis sentimen bisa efektif digunakan pada data bahasa Indonesia yang memiliki struktur bahasa yang tidak sederhana, seperti penggunaan bahasa santai, singkatan, campuran kata, serta ekspresi emosional yang tidak standar, seperti yang dikemukakan dalam penelitian
4. Penelitian ini juga memberikan dasar konsep dan metode untuk pengembangan penelitian selanjutnya, termasuk pengembangan model hibrida atau pendekatan lanjutan seperti *aspect-based sentiment analysis (ABSA)*, agar lebih memahami makna dan dimensi dari opini publik terhadap kebijakan pemerintah secara mendalam.
5. Secara nyata, hasil penelitian ini bisa digunakan sebagai bahan pertimbangan dalam memahami persepsi publik terhadap kebijakan tunjangan DPR RI secara objektif dan berdasarkan data, serta mendorong peningkatan transparansi dan kualitas komunikasi publik, sesuai dengan temuan
6. Penelitian ini juga berkontribusi untuk pengembangan sistem analisis teks dan media sosial dalam bidang teknologi informasi, khususnya sebagai dasar dalam pembuatan aplikasi analisis opini publik otomatis yang relevan dengan isu-isu sosial-politik.
7. Dari sudut pandang sosial dan akademis, penelitian ini membantu memaparkan bahwa opini publik di media sosial dapat dianalisis secara ilmiah dan terukur,

45

sehingga mendorong diskusi publik yang lebih rasional, transparan, dan berbasis data terhadap kebijakan pemerintah.

BAB II

KAJIAN TEORI

Bab ini berisi teori-teori dan konsep yang mendukung penelitian, meliputi analisis sentimen, media sosial dan opini publik, metode *Naïve Bayes Classifier*, *Lexicon Sentiment*, *TF-IDF*, *Evaluation Matrix*, serta metode *CRISP-DM*. Selain itu, bab ini juga memuat kajian penelitian terdahulu yang relevan sebagai landasan pengembangan penelitian.

A. Kajian Teori

1. Analisis *Sentiment*

Analisis sentimen adalah bagian dari Natural Language Processing (NLP) yang menggerakkan proses mengenali dan mengelompokkan pendapat yang terdapat dalam data berupa teks. Dengan cara ini, sebuah teks bisa dianalisis untuk mengetahui sikap atau pandangan yang ingin disampaikan oleh penulis, seperti perasaan positif, negatif, atau netral. Abdillah dan Hasan (2023) menyatakan bahwa analisis sentimen digunakan untuk mengetahui apakah pendapat dalam teks memiliki makna positif, netral, atau negatif berdasarkan isi teks tersebut. Dengan cara ini, metode ini bisa membantu mengubah banyak data teks menjadi informasi yang lebih sederhana dan mudah dipahami.

Perkembangan teknologi digital dan media sosial membuat jumlah data berupa teks yang dibuat oleh masyarakat meningkat banyak. Berbagai situs online memudahkan orang untuk memberikan pendapat, masukan, atau respons terhadap suatu topik dengan cepat dan secara terbuka. Kondisi ini membuat analisis sentimen semakin sering digunakan karena mampu membantu dalam memproses data dalam jumlah besar dengan lebih efisien dibandingkan cara analisis manual. Rakib dan Aris (2024) mengatakan bahwa analisis sentimen adalah bagian dari text mining yang menggunakan algoritma pembelajaran mesin untuk mengenali pola-pola pendapat dan perasaan yang terdapat dalam data teks. Dengan cara itu, informasi yang sebelumnya tersebar dalam jumlah banyak bisa diringkas menjadi gambaran umum mengenai arah pendapat masyarakat.

Dalam penerapannya, analisis sentimen tidak hanya digunakan untuk mengetahui apakah suatu pendapat bersifat positif atau negatif, tetapi juga untuk memahami bagaimana masyarakat melihat berbagai topik yang sedang tren. Menurut

Murniyati dan timnya tahun 2022, analisis sentimen adalah gabungan dari text mining dan data mining yang digunakan untuk mengekstrak informasi penting dari data berbentuk teks, sehingga bisa diketahui arah opini yang terdapat di dalamnya. Hasil dari analisis itu biasanya dibagi menjadi tiga kelompok besar, yaitu positif, negatif, dan netral. Pengelompokan ini membantu para peneliti dan orang yang membuat keputusan untuk memahami bagaimana masyarakat merespons terhadap suatu produk, layanan, atau kebijakan tertentu.

Fadli dan tim peneliti (2026) menjelaskan bahwa analisis sentimen memainkan peran penting dalam memahami pendapat masyarakat mengenai berbagai isu yang muncul di ruang digital, terutama yang berkaitan dengan kebijakan publik. Hasil analisis sentimen bisa menjadi sumber informasi yang jelas untuk memahami bagaimana masyarakat umumnya merespons sebuah kebijakan. Maka, metode ini bukan hanya bisa digunakan untuk mengklasifikasikan opini, tetapi juga membantu dalam proses evaluasi dan pengambilan keputusan yang didasarkan pada data. Dalam penelitian ini, analisis sentimen digunakan untuk memahami tanggapan umum masyarakat terhadap kebijakan tunjangan DPR RI yang sedang ramai dibicarakan di media sosial, sehingga dapat dilihat tren opini yang ada di kalangan masyarakat.

2. Media Sosial dan Opini Publik

Media sosial adalah alat komunikasi digital yang memungkinkan pengguna berbagi informasi, pendapat, dan pengalaman dengan cepat kepada banyak orang. Perkembangan media sosial membuat masyarakat lebih mudah berinteraksi dan menyampaikan pendapat mengenai berbagai isu yang sedang terjadi. Melalui media sosial, setiap orang bisa memberikan komentar langsung mengenai suatu peristiwa, produk, layanan, atau kebijakan pemerintah. Kondisi tersebut membuat media sosial menjadi salah satu sumber data yang sangat potensial untuk digunakan dalam penelitian terkait opini publik dan analisis sentimen.

Salah satu platform yang sering digunakan dalam penelitian analisis sentimen adalah X (Twitter). Platform ini memudahkan pengguna untuk memberikan pendapat berupa tulisan secara langsung dan cepat, sehingga dapat menunjukkan bagaimana masyarakat merespons sebuah isu dengan waktu yang lebih singkat. Abdillah dan Hasan (2023) menyebutkan bahwa X (Twitter) sering dipakai sebagai sumber data dalam penelitian analisis sentimen karena menyediakan banyak opini publik dan mudah diakses. Temuan ini sesuai dengan penelitian Fahrudin dan timnya pada tahun 2022,

yang menunjukkan bahwa Twitter menjadi satu dari media yang efektif untuk memantau perubahan opini masyarakat, karena mampu menunjukkan berbagai pandangan yang muncul langsung dalam bentuk teks.

Menggunakan media sosial sebagai sumber data memberikan beberapa manfaat dibandingkan cara pengumpulan data yang biasa dilakukan. Dengan menggunakan data dari media sosial, peneliti bisa melihat bagaimana masyarakat merespons suatu kebijakan tanpa perlu melakukan survei atau wawancara langsung. Wijaya dan Pratiwi (2021) menjelaskan bahwa data dari Twitter bisa digunakan untuk memahami bagaimana masyarakat merespons kebijakan publik dengan lebih efisien dan juga cepat. Selain itu, banyaknya data tersebut memungkinkan para peneliti mendapatkan gambaran yang lebih luas dan beragam mengenai pendapat masyarakat.

Meskipun begitu, menggunakan media sosial sebagai sumber data dalam penelitian juga menghadapi beberapa kesulitan. Bahasa yang digunakan oleh orang-orang di media sosial biasanya tidak mengikuti aturan bahasa yang baku dan sering kali menggunakan singkatan, kata-kata yang tidak resmi, campuran bahasa, atau kalimat yang bersifat sarkastik. Liza Wikarsa dan tim peneliti tahun 2022 menyatakan bahwa sifat-sifat bahasa dapat memengaruhi cara menganalisis sentimen karena arti sebuah teks tidak selalu bisa difahami secara langsung hanya dari kata-kata yang digunakan. Akibatnya, proses pemrosesan data perlu diperhatikan dengan baik agar hasil analisis yang didapatkan tetap tepat dan sesuai.

kan melalui media sosial X (Twitter). Untuk mengatasi permasalahan tersebut, diperlukan tahapan preprocessing dan metode analisis yang sesuai. Asri et al. (2022) menegaskan bahwa proses preprocessing berperan penting dalam membersihkan dan menyiapkan data sebelum dilakukan analisis lebih lanjut. Tahapan ini membantu mengurangi noise pada data sehingga informasi yang terkandung dalam teks dapat dipahami dengan lebih baik oleh sistem. Oleh karena itu, penggunaan metode analisis sentimen yang mampu mengidentifikasi polaritas sekaligus memahami konteks emosional menjadi penting untuk menghasilkan analisis yang lebih akurat terhadap opini masyarakat yang disampaikan melalui media sosial X (Twitter).

3. Metode *Naïve Bayes Classifier*

Naïve Bayes Classifier (NBC) adalah salah satu algoritma pengelompokan yang sering digunakan dalam bidang pembelajaran mesin, terutama untuk memproses data berbentuk teks. Metode ini menggunakan Teorema Bayes dengan pendekatan berbasis

probabilitas untuk memperkirakan seberapa besar kemungkinan suatu data masuk ke dalam kelas tertentu. Menurut Cinta Azzaria dan timnya pada tahun 2023, metode Naïve Bayes menganggap bahwa setiap fitur yang digunakan dalam proses pengklasifikasian tidak saling memengaruhi satu sama lain. Meskipun mengandalkan asumsi yang mudah, metode ini tetap bisa memberikan hasil yang cukup baik dalam berbagai penelitian mengenai klasifikasi teks.

Dalam analisis sentimen, metode Naïve Bayes digunakan untuk mengelompokkan data ke dalam kategori seperti sentimen positif, negatif, atau netral berdasarkan pola kata yang terdapat dalam dokumen tersebut. Setiap kata akan diberi nilai kemungkinan muncul di setiap kelas sentimen, lalu sistem memilih kelas yang memiliki nilai kemungkinan tertinggi sebagai hasil dari klasifikasi tersebut. Cara tersebut membuat Naïve Bayes menjadi salah satu metode yang sering digunakan karena cara menghitungnya cukup mudah dan bisa diterapkan pada banyak data teks. Fadli dan tim peneliti (2026) menjelaskan bahwa metode Naïve Bayes memiliki kelebihan dalam memproses data teks yang memiliki banyak variasi kata. Selain itu, metode ini juga bisa bekerja dengan baik ketika digabungkan dengan cara mewakili fitur teks, seperti TF-IDF.

Selain itu, metode ini juga memiliki performa komputasi yang cukup baik, sehingga sering dipakai dalam penelitian analisis perasaan di media sosial. Kemampuan itu membuat Naïve Bayes menjadi salah satu algoritma yang sering digunakan untuk menganalisis opini publik yang berasal dari berbagai platform digital. Penelitian yang dilakukan oleh Murniyati dan tim pada tahun 2022 menunjukkan bahwa metode Naïve Bayes sering digunakan dalam proses klasifikasi sentimen karena mampu memberikan hasil yang cukup baik dalam mengenali sikap atau pendapat yang dimiliki oleh pengguna. Oleh karena itu, metode ini masih sering digunakan dalam berbagai penelitian analisis sentimen karena menawarkan keseimbangan antara kemudahan dalam penerapannya, kecepatan dalam prosesnya, dan kemampuan dalam mengklasifikasikan data. Dalam penelitian ini, klasifikasi opini masyarakat tentang kebijakan tunjangan DPR RI dilakukan menggunakan metode Naïve Bayes Classifier, dengan hasil klasifikasi berupa sentimen positif, negatif, atau netral.

Dengan rumus naive bayes seperti berikut

$$P(C|X) \propto P(C) \prod_{i=1}^n P(x_i | C) \quad (2.1)$$

Keterangan:

- a. $P(C|X)$ = Probabilitas posterior, yaitu peluang dokumen X termasuk ke dalam kelas sentimen C.
- b. $P(C)$ = Probabilitas prior dari kelas C, yaitu peluang awal suatu kelas sebelum mempertimbangkan fitur yang ada pada dokumen.
- c. $P(x_i|C)$ = Probabilitas kemunculan fitur atau kata x_i pada kelas C.
- d. $\prod$ = Operasi perkalian seluruh probabilitas fitur yang terdapat dalam dokumen.
- e. x_i = Fitur atau kata ke-i yang terdapat dalam dokumen.
- f. n = Jumlah fitur atau kata yang digunakan dalam proses klasifikasi.
- g. C = Kelas sentimen yang akan diprediksi, seperti positif, negatif, atau netral.
- h. X = Dokumen atau teks yang akan diklasifikasikan.

Rumus Naïve Bayes digunakan untuk menghitung probabilitas suatu dokumen termasuk ke dalam kelas sentimen tertentu berdasarkan kata-kata yang terkandung di dalamnya. Metode ini bekerja dengan mengalikan probabilitas setiap kata terhadap suatu kelas sentimen, kemudian membandingkan hasil probabilitas pada seluruh kelas yang tersedia. Kelas dengan nilai probabilitas terbesar akan dipilih sebagai hasil klasifikasi. Pendekatan ini didasarkan pada asumsi bahwa setiap fitur atau kata bersifat independen terhadap fitur lainnya, sehingga proses perhitungan dapat dilakukan secara lebih sederhana dan efisien.

Dalam analisis sentimen, setiap kata yang muncul pada dokumen akan memberikan kontribusi terhadap penentuan kelas sentimen. Semakin tinggi probabilitas kemunculan kata-kata tersebut pada suatu kelas, maka semakin besar kemungkinan dokumen diklasifikasikan ke dalam kelas yang sama. Oleh karena itu, Naïve Bayes banyak digunakan pada klasifikasi teks karena memiliki proses komputasi yang relatif cepat serta mampu menghasilkan performa yang baik pada data berukuran besar maupun data teks pendek seperti tweet.

Selain itu di dalam eksperimen nantinya akan digunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*. Teknik ini membantu menonjolkan kata-kata yang memiliki makna penting dalam klasifikasi sentimen dan banyak digunakan dalam penelitian analisis opini publik. Rumus TF-IDF dinyatakan sebagai berikut

$$TFIDF(v, d) = \frac{TF(w, d)}{DF(w)} \log \quad (2.2)$$

Keterangan:

- a. $TF-IDF(w, d)$ = Bobot kata (*term*) w pada dokumen d .

- b. $TF(w,d)$ = Jumlah kemunculan kata w dalam dokumen d (*Term Frequency*).
- c. $DF(w)$ = Jumlah dokumen yang mengandung kata w (*Document Frequency*).
- d. $\log$ = Fungsi logaritma yang digunakan untuk mengurangi pengaruh kata yang muncul pada banyak dokumen.
- e. w = Kata (*term*) yang dianalisis.
- f. d = Dokumen yang dianalisis.

TF-IDF digunakan untuk memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam suatu dokumen dan tingkat penyebarannya pada seluruh dokumen. Semakin sering suatu kata muncul dalam dokumen tertentu namun semakin jarang muncul pada dokumen lain, maka nilai TF-IDF yang dihasilkan akan semakin besar. Sebaliknya, kata yang muncul pada hampir semua dokumen akan memiliki nilai bobot yang lebih kecil karena dianggap kurang mampu membedakan isi antar dokumen

4. Metode *Lexicon Sentiment*

Metode Lexicon Sentiment merupakan salah satu pendekatan dalam analisis sentimen yang digunakan untuk menentukan kecenderungan opini berdasarkan kamus sentimen yang telah disusun sebelumnya. Pada metode ini, setiap kata dalam kamus memiliki nilai polaritas tertentu, seperti positif maupun negatif, yang digunakan untuk menilai sentimen suatu teks. Harwenda et al. (2025) menjelaskan bahwa proses analisis dilakukan dengan mencocokkan kata-kata yang terdapat dalam teks dengan daftar kata pada kamus sentimen, kemudian menjumlahkan nilai yang diperoleh untuk menentukan kecenderungan sentimen secara keseluruhan.

Berbeda dengan pendekatan machine learning yang memerlukan data latih, metode Lexicon Sentiment dapat digunakan secara langsung tanpa melalui proses pelatihan model. Pendekatan ini memanfaatkan kumpulan kata yang telah memiliki nilai sentimen sehingga mampu memberikan hasil klasifikasi berdasarkan karakteristik kata yang muncul dalam suatu teks. Menurut Murniyati et al. (2022), pendekatan lexicon banyak digunakan dalam penelitian analisis sentimen karena mampu membantu proses identifikasi opini secara otomatis dan menghasilkan pengelompokan sentimen ke dalam kategori positif, negatif, maupun netral. Dengan karakteristik tersebut, metode lexicon menjadi salah satu pendekatan yang cukup populer dalam pengolahan data opini berbasis teks.

Selain digunakan untuk menentukan polaritas sentimen, metode lexicon juga dapat membantu mengidentifikasi makna emosional yang terkandung dalam suatu opini. Fadli et al. (2026) menjelaskan bahwa pendekatan lexicon dapat digunakan sebagai dasar pelabelan sentimen sebelum dilakukan proses klasifikasi menggunakan algoritma machine learning. Penggunaan kamus sentimen memungkinkan sistem mengenali kecenderungan emosi yang muncul pada teks sehingga hasil analisis menjadi lebih mudah diinterpretasikan. Oleh karena itu, metode lexicon sering dimanfaatkan sebagai pelengkap pendekatan machine learning untuk menghasilkan analisis yang lebih komprehensif.

Meskipun memiliki beberapa keunggulan, metode Lexicon Sentiment masih memiliki keterbatasan dalam memahami konteks bahasa yang kompleks, seperti ironi, sarkasme, atau kalimat yang memiliki makna ganda. Liza Wikarsa et al. (2022) menyatakan bahwa pendekatan ini cenderung lebih efektif ketika digunakan bersama metode lain yang mampu memahami pola data secara lebih mendalam. Oleh karena itu, dalam penelitian ini metode Lexicon Sentiment digunakan sebagai pelengkap Naïve Bayes Classifier untuk membantu mengidentifikasi sentimen sekaligus memberikan gambaran mengenai makna emosional yang terkandung dalam opini masyarakat terhadap kebijakan tunjangan DPR RI.

Dengan tahapan pada metode *Lexicon* sebagai berikut:

- a. Mencocokkan token dalam tweet dengan daftar kata dalam *Lexicon*.
- b. Menjumlahkan skor *sentiment* dari tiap kata.
- c. Menghasilkan skor total per tweet.
- d. Menentukan apakah *sentiment*nya positif, negatif, atau netral.

Metode ini juga digunakan dalam penelitian *Lexicon-NBC* (Liza Wikarsa et al., 2022).

5. Integrasi Naïve Bayes Clasifier dan Lexicon Sentiment

Integrasi metode Naïve Bayes Classifier dengan Lexicon Sentiment merupakan pendekatan yang menggabungkan keunggulan algoritma machine learning dengan analisis berbasis kamus sentimen. Naïve Bayes berperan dalam mengklasifikasikan data berdasarkan perhitungan probabilitas, sedangkan Lexicon Sentiment digunakan untuk mengidentifikasi kecenderungan makna emosional yang terkandung dalam teks. Menurut Rorimpandey (2024), pendekatan integratif ini memungkinkan sistem tidak hanya menentukan kategori sentimen suatu opini, tetapi juga memahami konteks emosional yang melatarbelakangi opini tersebut.

Penggabungan kedua metode tersebut dilakukan karena masing-masing pendekatan memiliki karakteristik yang saling melengkapi. Naïve Bayes memiliki kemampuan yang baik dalam mengenali pola dari data teks dan melakukan klasifikasi secara otomatis, tetapi sering kali masih terbatas dalam menangkap makna emosional yang tersirat. Sebaliknya, Lexicon Sentiment mampu memberikan informasi mengenai polaritas dan emosi berdasarkan kata-kata yang digunakan dalam suatu teks. Dengan memanfaatkan keunggulan keduanya, hasil analisis sentimen dapat menjadi lebih komprehensif dan mudah diinterpretasikan.

Fadli et al. (2026) menjelaskan bahwa pendekatan berbasis lexicon dapat digunakan untuk mendukung proses klasifikasi sentimen dengan memberikan informasi tambahan mengenai kecenderungan opini yang terdapat pada data teks. Sementara itu, Murniyati et al. (2022) menyatakan bahwa analisis sentimen bertujuan untuk mengelompokkan opini ke dalam kategori positif, negatif, dan netral sehingga dapat digunakan untuk memahami respons masyarakat terhadap suatu isu. Kombinasi kedua pendekatan tersebut memungkinkan proses klasifikasi tidak hanya berfokus pada hasil akhir sentimen, tetapi juga memperhatikan makna yang terkandung di dalam opini yang dianalisis.

Pendekatan integrasi ini sangat relevan dalam penelitian yang berkaitan dengan kebijakan publik karena opini masyarakat sering kali mengandung emosi yang tidak disampaikan secara langsung. Respons berupa dukungan, kritik, kekecewaan, maupun ketidakpercayaan dapat muncul melalui pilihan kata tertentu yang digunakan dalam suatu pernyataan. Oleh karena itu, penggunaan Naïve Bayes dan Lexicon Sentiment secara bersamaan diharapkan mampu menghasilkan analisis yang lebih mendalam sehingga dapat memberikan gambaran yang lebih akurat mengenai persepsi masyarakat terhadap kebijakan tunjangan DPR RI.

Dengan tahapan-tahapan sebagai berikut :

- a. *Preprocessing teks*, yang mencakup pembersihan teks, *tokenisasi*, dan normalisasi kata agar sesuai dengan kamus sentimen dan model klasifikasi.
- b. Ekstraksi fitur teks menggunakan TF-IDF, sehingga setiap dokumen direpresentasikan dalam bentuk *vektor numerik*.
- c. Klasifikasi sentimen menggunakan *Naïve Bayes Classifier*, dengan menghitung probabilitas setiap kelas sentimen berdasarkan fitur TF-IDF.

- d. Penentuan skor sentimen berbasis *Lexicon*, dengan menjumlahkan bobot polaritas kata-kata dalam teks berdasarkan kamus sentimen.
- e. Penggabungan hasil *NBC* dan *Lexicon Sentiment*, untuk menentukan kelas sentimen akhir berdasarkan aturan *integrasi* yang telah ditetapkan.

6. Evaluation Matrix

Evaluasi model adalah proses yang dilakukan untuk mengetahui seberapa baik suatu algoritma klasifikasi mampu menghasilkan prediksi yang sesuai dengan situasi nyata. Dalam penelitian analisis sentimen, proses evaluasi digunakan untuk mengetahui sejauh mana model berhasil mengelompokkan data ke dalam kategori sentimen yang sudah ditentukan. Hasil evaluasi digunakan untuk mengetahui apakah model yang digunakan sudah berjalan baik atau masih perlu diperbaiki dalam proses pembuatan model dan pengolahan datanya.

Menurut Perdana et al. (2022), umumnya model klasifikasi dievaluasi dengan menggunakan confusion matrix, yaitu tabel yang berguna untuk membandingkan hasil prediksi model dengan data sebenarnya. Dengan menggunakan confusion matrix, peneliti bisa mengetahui berapa banyak prediksi yang benar dan yang salah dalam setiap kategori klasifikasi. Informasi tersebut kemudian dipakai untuk menghitung berbagai metrik penilaian yang bisa menjelaskan seberapa baik performa model itu. Beberapa ukuran yang biasa digunakan untuk mengevaluasi hasil analisis sentimen adalah akurasi, presisi, recall, dan F1-score. Akurasi digunakan untuk mengetahui seberapa tepat model dalam mengklasifikasikan data secara keseluruhan. Precision menggambarkan kemampuan model dalam memberikan prediksi yang tepat dan sesuai dengan kategori tertentu, sementara recall digunakan untuk menilai seberapa baik model bisa mencakup semua data yang sebenarnya termasuk ke dalam kategori tersebut. F1-score adalah nilai yang didapat dari gabungan antara presisi dan recall, sehingga bisa digunakan untuk menilai seberapa baik model dalam melakukan klasifikasi, baik dalam mengenali objek yang benar maupun tidak melewatkan objek yang seharusnya terdeteksi.

Metrik evaluasi tersebut sudah banyak digunakan dalam penelitian analisis sentimen. Rakib dan Aris (2024) menggunakan beberapa metrik seperti accuracy, precision, recall, dan F1-score untuk mengevaluasi seberapa baik model klasifikasi sentimen yang mereka kembangkan. Abdillah dan Hasan (2023) juga melakukan hal yang sama untuk mengevaluasi seberapa berhasil algoritma dalam mengenali polaritas

sentimen dari data teks. Oleh karena itu, metrik evaluasi tersebut digunakan dalam penelitian ini untuk mengetahui sejauh mana metode Naïve Bayes Classifier mampu mengklasifikasikan opini publik mengenai kebijakan tunjangan DPR RI. Berikut adalah rumus yang digunakan untuk menghitung nilai akurasi, presisi, recall, dan F1-score.

a. *accuracy*

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.3)$$

b. *precision*

$$precision = \frac{TP}{TP+FP} \quad (2.4)$$

c. *recall*

$$Recall = \frac{TP}{TP+FN} \quad (2.5)$$

d. *f1-score*

$$f1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (2.6)$$

Keterangan:

- a. TP (True Positive) adalah jumlah data yang diprediksi positif dan sesuai dengan kondisi sebenarnya.
- b. TN (True Negative) adalah jumlah data yang diprediksi negatif dan sesuai dengan kondisi sebenarnya.
- c. FP (False Positive) adalah jumlah data yang diprediksi positif tetapi sebenarnya termasuk kategori negatif.
- d. FN (False Negative) adalah jumlah data yang diprediksi negatif tetapi sebenarnya termasuk kategori positif.

Accuracy digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan seluruh data yang diuji. Nilai accuracy menunjukkan persentase prediksi yang berhasil dilakukan dengan benar dibandingkan dengan jumlah keseluruhan data. Semakin tinggi nilai accuracy yang diperoleh, maka semakin baik kemampuan model dalam melakukan klasifikasi.

Precision digunakan untuk mengukur tingkat ketepatan prediksi pada suatu kelas tertentu. Metrik ini menunjukkan seberapa banyak data yang diprediksi sebagai suatu kelas benar-benar termasuk ke dalam kelas tersebut. Nilai precision yang tinggi menunjukkan bahwa model mampu meminimalkan kesalahan prediksi positif.

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang sebenarnya termasuk ke dalam suatu kelas. Nilai recall yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar data yang relevan sehingga jumlah kesalahan prediksi negatif dapat diminimalkan.

F1-Score merupakan metrik yang digunakan untuk mengukur keseimbangan antara precision dan recall. Nilai ini diperoleh dari rata-rata harmonik precision dan recall sehingga dapat memberikan gambaran performa model secara lebih menyeluruh. F1-Score umumnya digunakan ketika distribusi data tidak seimbang karena mampu memberikan evaluasi yang lebih objektif dibandingkan hanya menggunakan accuracy

Keempat metrik evaluasi tersebut digunakan secara bersamaan karena masing-masing memiliki fungsi yang berbeda dalam menilai performa model klasifikasi. Accuracy memberikan gambaran umum mengenai tingkat ketepatan prediksi model terhadap seluruh data yang diuji. Namun, nilai accuracy yang tinggi belum tentu menunjukkan bahwa model memiliki performa yang baik pada setiap kelas sentimen. Oleh karena itu, precision dan recall digunakan sebagai metrik tambahan untuk mengevaluasi kualitas prediksi secara lebih rinci. Sementara itu, F1-Score digunakan untuk memberikan ukuran yang lebih seimbang antara precision dan recall sehingga hasil evaluasi model dapat dianalisis secara lebih komprehensif.

Dalam penelitian analisis sentimen, hasil evaluasi model menjadi salah satu indikator penting untuk menentukan tingkat keberhasilan proses klasifikasi. Semakin tinggi nilai accuracy, precision, recall, dan F1-Score yang diperoleh, maka semakin baik kemampuan model dalam mengenali pola sentimen pada data teks. Oleh karena itu, metrik evaluasi tidak hanya digunakan untuk mengukur performa model, tetapi juga sebagai dasar dalam membandingkan hasil penelitian dengan penelitian-penelitian sebelumnya yang menggunakan metode serupa. Pada penelitian ini, metrik accuracy, precision, recall, dan F1-Score digunakan untuk mengevaluasi kinerja metode Naïve Bayes Classifier dalam mengklasifikasikan sentimen opini masyarakat terhadap kebijakan tunjangan DPR RI.

7. CRISP-DM

CRISP-DM adalah sebuah kerangka kerja yang digunakan dalam proses data mining untuk mengelola dan menganalisis data dengan cara yang teratur dan terstruktur. Metode ini masih sering digunakan dalam penelitian saat ini karena mampu memberikan

langkah-langkah yang terstruktur untuk membangun model berbasis data serta membantu peneliti dalam mengelola seluruh proses analisis (Rahman et al., 2022). CRISP-DM memiliki enam tahapan utama, yaitu memahami bisnis, memahami data, mempersiapkan data, membuat model, mengevaluasi hasil, serta menerapkan model tersebut. Enam langkah tersebut membentuk siklus kerja yang saling terkait dalam proses data mining, mulai dari memahami permasalahan sampai menerapkan hasil analisisnya (Sari & Pratama, 2024).

Fase Business Insights berfokus pada menemukan tujuan penelitian dan masalah yang perlu diselesaikan. Selanjutnya dilakukan proses memahami data untuk mengumpulkan dan mengeksplorasi data agar dapat memahami ciri-ciri dan tingkat kualitasnya. Tahap persiapan data mencakup membersihkan data, mengubah bentuk data, dan memilih data yang tepat agar bisa digunakan untuk analisis. Pada tahap pemodelan, algoritma khusus seperti Naïve Bayes Classifier digunakan untuk membuat model yang memperkirakan kategori atau jenis data (Putra et al., 2024).

Evaluasi dilakukan untuk mengecek sejauh mana model bekerja dengan baik menggunakan beberapa ukuran seperti akurasi, presisi, recall, dan F1-skor. Langkah terakhir adalah implementasi, yaitu cara menampilkan hasil dari analisis yang sudah dilakukan dalam bentuk yang bisa dipakai, misalnya dalam bentuk laporan atau gambar data. Penggunaan CRISP-DM dalam penelitian data mining dianggap efektif karena memberikan struktur yang jelas dan bisa diadaptasi sesuai kebutuhan penelitian (Zhang et al., 2022). Dalam penelitian ini, CRISP-DM digunakan sebagai panduan untuk menganalisis pendapat masyarakat mengenai kebijakan pembagian dana DPR Indonesia di platform media sosial X (Twitter). Menggunakan metode ini memastikan penelitian berjalan secara rapi, mulai dari pengumpulan data sampai penjelasan hasilnya, sehingga menghasilkan analisis yang lebih rapi dan dapat dipertanggungjawabkan.

B. Kajian Hasil Penelitian Terdahulu

Tabel 2.1 Kajian Terdahulu

Penulis & Tahun	Judul Penelitian	Dataset	Metode	Hasil/ Kontribusi
(Fachturroziq et al., 2025)	Analisis Sentimen Ulasan Produk Amazon Menggunakan	Amazon	Naive bayes	Akurasi 80% membuktikan algoritma naive bayes sangat baik untuk

	Algoritma Naive Bayes Untuk Prediksi Rating Produk			digunakan dalam analisis sentimen
Klik atau ketuk di sini untuk memasukkan teks.(Adisti et al., 2025)	ANALISIS SENTIMEN UJARAN KEBENCIAN PADA TWEET DI TWITTER	Twitter	Support Vector Machine	Hasil penelitian menunjukkan bahwa analisis sentimen mampu digunakan untuk mengidentifikasi ujaran kebencian pada tweet di Twitter secara efektif.
(Daniati & Utama, 2023)	ANALISIS SENTIMEN DENGAN PENDEKATAN ENSEMBLE LEARNING DAN WORD EMBEDDING PADA TWITTER	kaggle	Ensambl e Learning dan pendekat an Word Embeddi ng	Akurasi 69.2% yang telah meningkatkan akurasi sebesar 1.22%
(M. Alfianto Alamsyah et al., 2025)	Analisis Pengalaman Pengguna Terhadap Aplikasi Panjalu e-Library di Perpustakaan Mastrip Pare	kuesione r	Kuantitat if deskripti f	Dari 6 variabel yang telah di uji 5 diantaranya ada di kategori baik dan satu di kategori cukup menandakan aplikasi beroperasi dengan baik
(Nugroho & Harini, 2024)	Teknik Random Forest untuk	kaggle	Optimali sasi	Menghasilkan akurasi sebesar 97,5%

	Meningkatkan Akurasi Data Tidak Seimbang		Random forest dengan SMOTE	membuktikan bahwa kombinasi tersebut dapat meningkatkan akurasi
(Muhammad Faruqziddan et al., 2025)	Pendekatan BERT Dalam Analisis Sentimen Terhadap Kominfo Di Media Sosial X	Twitter (X)	<i>Bidirectional Encoder Representations from Transformers (BERT)</i>	Akurasi sebesar 84% membuktikan BERT efektif dalam menganalisis opini masyarakat
(Abdillah & Hasan, 2023)	Analisis sentimen calon presiden menggunakan pengklasifikasi <i>Naïve Bayes</i>	Twitter	<i>naïve bayes classifier</i>	Metode naive bayes dapat menunjukkan kandidat calon presiden yang mana yang paling di sukai masyarakat.
(Cinta Azzaria et al., 2023)	Analisis Data Mining dengan Naive Bayes Classifier untuk Peningkatan Penjualan Produk	Data Penjualan	<i>Naive Bayes Classifier</i>	Bahwa Pendekatan tersebut dapat meningkatkan penjualan dengan optimal dengan akurasi 98,98%
(Alviyanti et al., 2024)	Klasifikasi Sentimen Pengguna Aplikasi Livin By Mandiri Pada Playstore Menggunakan Algoritma	kaggle	<i>Naive Bayes Classifier</i>	Metode naive bayes dapat menklasifikasi data dengan baik dengan akurasi 93%

	Naive Bayes			
(Harwenda et al., 2025)	Sentiment Analysis on Government Public Policies: A Systematic Literature Review	literature	<i>Systematic Literature Review (SLR)</i>	Metode yang paling sering di pakai dan efektifif adalah SVM naive bayes dan Random forest
(Perdana et al., 2022)	Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Classifier	Twitter	<i>Naive Bayes Classifier</i>	Hasilnya masyarakat lebih banyak memberikan sentimen positif dengan jumlah 60.442, dan precision 96,1%
(Wijaya Tamora Nonia et al., 2021)	Analisis Sentimen Opini Publik Tentang Undang Undang Cipta Kerja Pada Twitter	Twitter	<i>Naive Bayes Classifier</i>	NBC memberikan hasil yang baik dengan akurasi 89,9%
(Asri et al., 2022)	Pelabelan Otomatis <i>Lexicon</i> Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile	PLN Mobile App	Kamus vader <i>Lexicon</i> dan naive bayes	Berhasil memiliki Tingkat akurasi sebesar 70%
(Fahrudin et al., 2022)	Pemodelan Teks Tweet pada Isu	Twitter	<i>Naive Bayes</i> ,	Akuarsinya yang dihasilkan sebesar

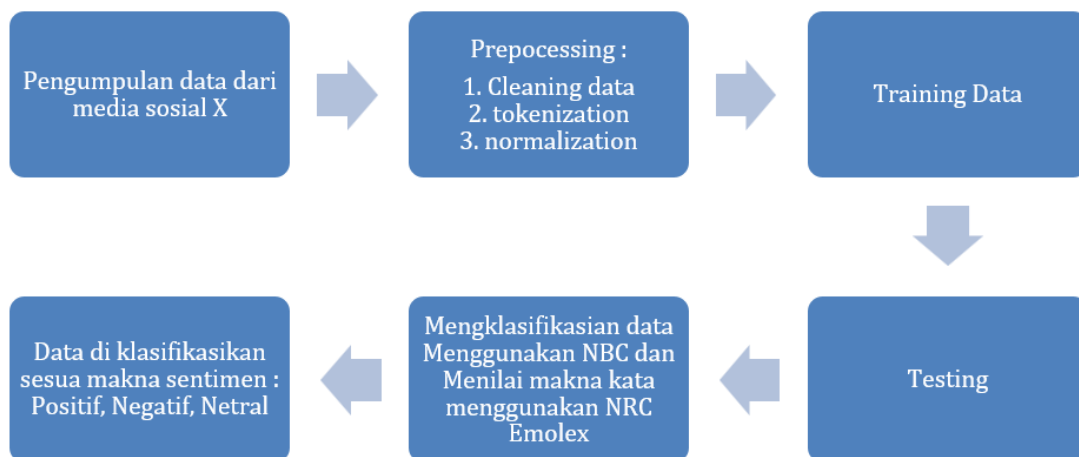
	Pelecehan Seksual Berbasis Analisis Sentimen dan Leksikon Emosi		dan NRC Emotion <i>Lexicon</i>	74% dan dapat mengenali emosi dari tweet.
(Kurnia et al., 2024)	Analisis Perbandingan Algoritma <i>Naïve Bayes</i> dengan K-Nearest Neighbor (KNN) Pada Dataset Mobile Price Classification	klasifikasi harga smartphone menggunakan Jupyter	K-Nearest Neighbor (KNN) dan naive bayes	Hasilnya menunjukkan bahwa KNN memiliki akurasi yang lebih tinggi dibandingkan dengan <i>Naïve Bayes</i> dengan akurasi naive bayes 64,21% dan KNN 66,69%
(Putri et al., 2022)	Perbandingan Kinerja Algoritma K-NN dan NBC Untuk Klasifikasi Penyakit Jantung	kaggle	K-Nearest Neighbor (KNN) dan naive bayes	Pada penelitian ini Naive bayes memiliki akurasi lebih tinggi dibandingkan KNN yaitu 86,17% untuk NBC dan 85,11% untuk KNN
(Faizal et al., 2023)	PERBANDINGAN <i>LEXICON</i> BASED DAN <i>NAÏVE BAYES</i> CLASSIFIER PADA ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP GEMPA TURKI	Twitter	<i>Lexicon</i> Based dan <i>Naive Bayes</i> Classifier	Gabungan kedua metode mendapatkan akurasi 65% sedangkan naive bayes dan <i>Lexicon</i> mendapatkan akurasi sebesar 64% membuktikan bahwa gabungan kedua metode menghasilkan akurasi yang lebih tinggi.

(Yusnitasari & Santoso, 2022)	Sentiment Analysis of Comments on Instagram Gramedia Pustaka Accounts (GPU) Using <i>Lexicon</i> Based and Naive Bayes Classifier Methods	Instagram	<i>Lexicon</i> Based dan <i>Naive Bayes</i> <i>Clasifier</i>	Menghasilkan akurasi sebesar 76.47% dengan dominasi sentimen positif
(Fadli Muhammad et al., 2026)	Analisis Sentimen Isu Redominasi Rupiah Menggunakan <i>Lexicon</i> Based dan <i>Naive Bayes</i>	YouTube	<i>Lexicon</i> Based dan <i>Naive Bayes</i> <i>Clasifier</i>	Memperoleh akurasi 87,56% dengan hasil yang akurat sekaligus interpretable
(Al Khadafi et al., 2022)	PENERAPAN METODE <i>NAIVE BAYES</i> CLASSIFIER DAN <i>LEXICON</i> BASED UNTUK ANALISIS SENTIMEN CYBERBULLYING PADA BPJS	Twitter	<i>Naive Bayes</i> <i>Clasifier</i> dan <i>Lexicon</i> Based	<i>Naive Bayes</i> Classifier menghasilkan akurasi 80%, sedangkan <i>Lexicon</i> Based menghasilkan akurasi 22%. dapat diambil kesimpulan bahwa <i>Naive Bayes</i> Classifier lebih baik dan lebih akurat daripada <i>Lexicon</i> Based.
(Fernanda & Fathoni, 2024)	Perbandingan Performa Labeling <i>Lexicon</i> InSet dan	Twitter	<i>Lexicon</i> Inset dan <i>Lexicon</i>	<i>Lexicon</i> Inset menghasilkan akurasi sebesar 87,83%

	VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM		Vader SVM	sedangan vader SVM menghasilkan akurasi 87,66% membuktikan <i>Lexicon</i> inset lebih unggul.
--	---	--	--------------	---

Berdasarkan banyak penelitian sebelumnya, dapat disimpulkan bahwa metode *Naïve Bayes Classifier* dan *Lexicon Sentiment* memiliki kemampuan analisis sentimen yang baik. Namun, sebagian besar penelitian masih berfokus pada mengklasifikasikan perasaan yang terpolarisasi tanpa melihat lebih dalam makna emosionalnya. Oleh karena itu, penelitian ini berupaya mengatasi keterbatasan tersebut dengan mengintegrasikan kedua metode tersebut sehingga memberikan analisis yang lebih komprehensif dan mendalam.

C. Kerangka Berpikir



Gambar 1. Gambar kerangka berpikir

Pada gambar di atas dipaparkan kerangka berpikir pada Penelitian ini yang dilakukan karena adanya kebutuhan untuk memahami bagaimana masyarakat memandang kebijakan tunjangan dari Dewan Perwakilan Rakyat Republik Indonesia melalui analisis opini publik di media sosial X (*Twitter*). Data dikumpulkan dengan cara mencari kata-kata kunci yang berkaitan dengan topik penelitian. Setelah itu, data diperoleh kemudian diolah dengan langkah-langkah sebagai berikut :

1. *Text Cleaning* (menghapus URL, emotikon, simbol, dan karakter khusus).

2. *Tokenization* (memecah teks menjadi token/kata).
3. *Normalization* mengatur bentuk teks agar sama dengan menggunakan huruf kecil, memperbaiki kata-kata yang tidak benar, dan menghilangkan kata-kata yang diulang-ulang.

Selanjutnya, metode *Naïve Bayes* digunakan untuk mengklasifikasikan teks berdasarkan kemungkinan kemunculan kata-kata tertentu. Sementara itu, metode *Lexicon Sentiment* digunakan untuk mengevaluasi skor emosional dari teks berdasarkan makna kata-kata yang ada dalam kamus *sentiment* bahasa Indonesia dengan menggunakan InSet *Lexicon* (Indonesia Sentiment *Lexicon*). Dengan menggabungkan kedua metode tersebut, diharapkan hasilnya bisa memberikan klasifikasi sentimen yang lebih akurat, sesuai dengan konteks, serta mampu mencerminkan perasaan masyarakat secara nyata. Hasil akhir dari analisis ini akan menunjukkan persentase sentimen positif, negatif, dan netral terhadap kebijakan tunjangan DPR RI. Hasil ini dapat digunakan sebagai bahan pertimbangan atau masukan.

BAB III

METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah CRISP-DM (Cross Industry Standard Process for Data Mining). Metode ini dipilih karena sesuai dengan karakteristik penelitian yang berfokus pada pengolahan data teks dan analisis sentimen. CRISP-DM terdiri dari enam tahapan utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*.

A. Business Understanding (Pemahaman Bisnis)

Studi ini akan menganalisis opini publik tentang kebijakan tunjangan DPR RI yang diperoleh dari media sosial X (*Twitter*). Bagaimana mengidentifikasi sentimen dan makna emosional dalam data teks yang tidak terstruktur adalah masalah utama dalam penelitian ini. Dengan menggunakan metode *Naive Bayes Classifier* dan *Lexicon Sentiment*, tujuan penelitian ini adalah untuk mengklasifikasikan sentimen opini publik dan menemukan emosi yang terkandung dalam teks. Penggunaan metrik evaluasi seperti *accuracy*, *Precision*, *recall*, dan *f1-score* menentukan kriteria keberhasilan penelitian. Ini juga mencakup kemampuan model untuk menemukan secara akurat makna emosional.

B. Data Understanding (Pemahaman Data)

1. Pengumpulan Data

Data yang digunakan merupakan data sekunder yang diperoleh dari platform X (*Twitter*) melalui proses *scraping* berdasarkan kata kunci yang berkaitan dengan kebijakan tunjangan DPR RI.

2. Deskripsi Data

Data berupa teks opini publik dalam bentuk tweet yang bersifat tidak terstruktur dan mengandung variasi bahasa, termasuk bahasa informal, singkatan, serta simbol.

3. Eksplorasi Data Awal

Eksplorasi dilakukan untuk mengetahui karakteristik data, seperti jumlah data, distribusi kata, serta keberadaan noise dalam teks.

4. Verifikasi Kualitas Data

Tahap ini dilakukan untuk memastikan kualitas data, seperti menghapus data duplikat, data kosong, serta data yang tidak relevan dengan topik penelitian.

C. Data Preparation (Persiapan Data)

1. Seleksi Data

Data diseleksi berdasarkan relevansi terhadap topik penelitian, sehingga hanya data yang berkaitan dengan kebijakan tunjangan DPR RI yang digunakan.

2. Pembersihan Data (*Cleaning*)

Proses pembersihan data meliputi:

- a. Menghapus URL, mention, hashtag, dan simbol
- b. Menghilangkan karakter khusus
- c. Menghapus data duplikat

3. Transformasi Data

Transformasi data dilakukan melalui beberapa tahapan, yaitu:

- a. *Case folding*
- b. *Tokenization*
- c. *Stopword removal*
- d. *Stemming*

4. Pembentukan Dataset Modeling

Data yang telah diproses kemudian diubah ke dalam bentuk *numerik* menggunakan pembobotan *TF-IDF* untuk keperluan proses modeling.

D. Modeling

1. Pemilihan Teknik Model

Penelitian ini menggunakan metode *Naïve Bayes Classifier* untuk klasifikasi sentimen dan metode *Lexicon Sentiment* untuk analisis makna emosional.

2. Proses Modeling

Data dibagi menjadi data latih dan data uji. Model *Naïve Bayes* digunakan untuk mengklasifikasikan sentimen, sedangkan metode *Lexicon* digunakan untuk mengidentifikasi emosi dalam teks.

3. Evaluasi Parameter Model

Evaluasi dilakukan untuk mengetahui performa model dalam mengklasifikasikan data berdasarkan hasil prediksi terhadap data uji.

4. Interpretasi Hasil Modeling

Hasil modeling dianalisis untuk mengetahui distribusi sentimen serta jenis emosi yang dominan dalam opini publik.

E. Evaluation (Evaluasi)

1. Evaluasi Model

Evaluasi dilakukan untuk mengetahui performa metode *Naïve Bayes Classifier* dalam mengklasifikasikan sentimen opini publik terkait kebijakan tunjangan DPR RI pada media sosial X (*Twitter*). Proses evaluasi dilakukan menggunakan *confusion matrix* dengan membandingkan hasil prediksi model terhadap data uji yang telah diberi label sentimen.

Adapun metrik evaluasi yang digunakan meliputi:

a. *Accuracy*

Digunakan untuk mengetahui seberapa akurat model dalam mengklasifikasikan data sentimen positif, negatif, dan netral secara keseluruhan. Nilai akurasi menunjukkan seberapa besar data yang berhasil diprediksi dengan tepat dibandingkan semua data yang digunakan untuk pengujian. Semakin besar nilai akurasi, maka semakin baik kemampuan model dalam melakukan pengelompokan secara umum.

b. *Precision*

Digunakan untuk mengecek seberapa tepat model dalam menebak jenis perasaan yang dimaksud. Nilai precision menunjukkan seberapa besar hasil prediksi yang benar dari semua data yang diprediksi dan termasuk ke dalam kategori sentimen tertentu. Metrik ini penting untuk mengetahui seberapa tepat prediksi yang dihasilkan model dalam setiap kategori sentimen.

c. *Recall*

Digunakan untuk mengetahui sejauh mana kemampuan model dalam memahami dan menemukan data yang sesuai dengan kategori sentimen yang dimaksud. Nilai recall menunjukkan seberapa besar data yang berhasil terdeteksi secara tepat dibandingkan dengan semua data yang seharusnya berada dalam kategori tersebut. Semakin tinggi nilai recall, semakin bagus kemampuan model dalam mengenali data pada kelas sentimen tertentu.

d. *F1-score*

Digunakan untuk mengukur keseimbangan antara tingkat akurasi dalam mengenali sentimen yang benar dengan kemampuan sistem dalam menemukan semua sentimen yang ada. F1-score didapatkan dari rata-rata harmonik antara presisi dan recall, sehingga bisa menggambarkan kemampuan model secara lebih lengkap. Semakin besar nilai F1-score, maka semakin baik keseimbangan antara akurasi prediksi dan kemampuan model untuk menemukan data yang tepat.

2. Validasi terhadap Tujuan Penelitian

Tahap ini dilakukan untuk memastikan bahwa metode yang digunakan telah mampu memenuhi tujuan penelitian, yaitu mengklasifikasikan sentimen opini publik dan memahami kecenderungan respons masyarakat terhadap kebijakan tunjangan DPR RI berdasarkan data dari media sosial X (*Twitter*).

3. Tinjauan Proses

Hasil evaluasi dianalisis untuk melihat kelebihan dan keterbatasan model dalam proses klasifikasi sentimen. Selain itu, tahap ini juga digunakan untuk mengetahui apakah tahapan *Preprocessing*, pembobotan *TF-IDF*, dan integrasi metode *Lexicon Sentiment* telah memberikan hasil analisis yang optimal.

F. Deployment (Penerapan)

1. Rencana Penerapan

Hasil penelitian ini dapat digunakan sebagai bahan analisis dalam memahami opini publik terhadap kebijakan pemerintah.

2. Implementasi Hasil

Hasil analisis disajikan dalam bentuk laporan dan visualisasi data yang menunjukkan distribusi sentimen dan emosi masyarakat.

3. Dokumentasi

Seluruh tahapan penelitian didokumentasikan sebagai laporan ilmiah yang dapat dijadikan referensi untuk penelitian selanjutnya.

BAB IV

HASIL DAN PEMBAHASAN

Penelitian ini melakukan analisis sentimen terhadap opini publik mengenai kebijakan tunjangan DPR RI menggunakan metode *Naïve Bayes Classifier*, *Lexicon-Based Sentiment*, dan integrasi kedua metode. Data diperoleh dari media sosial X (Twitter) kemudian diproses melalui tahapan *Preprocessing*, pembobotan TF-IDF, klasifikasi sentimen, dan evaluasi model.

A. Data Hasil Penelitian

1. *Preprocessing* dan Transformasi Data

Sebelum proses klasifikasi dilakukan, data opini publik harus melewati tahap preprocessing yang mencakup cleaning, tokenization, dan normalization. Tahap pembersihan dilakukan dengan menghilangkan URL, simbol, tanda baca, angka, dan karakter khusus sehingga teks menjadi lebih rapi dan teratur. Langkah berikutnya adalah tokenisasi yang digunakan untuk memisahkan kalimat menjadi kata-kata individu agar bisa diproses pada tahap analisis sentimen. Proses normalisasi dilakukan agar kata yang tidak standar dapat diubah menjadi bentuk yang baku, sehingga makna dari teks menjadi lebih konsisten.

Setelah proses preprocessing selesai, data kemudian diproses lagi dengan cara mengubahnya menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). TF-IDF digunakan untuk memberi bobot pada setiap kata berdasarkan tingkat pentingnya dalam dokumen. Semakin sering suatu kata muncul dalam satu dokumen dan semakin sedikit muncul di dokumen lainnya, maka nilai bobot TF-IDF untuk kata tersebut akan semakin besar. Proses ini bertujuan untuk membantu model memahami kata-kata yang penting dalam menentukan jenis perasaan yang dimaksud. Rumus TF-IDF

seperti yang terdapat pada formula nomor 2.2. Hasil transformasi TF-IDF kemudian digunakan sebagai fitur pada proses klasifikasi menggunakan *Naïve Bayes* Classifier.

	Username	Tanggal	Teks	Label	Tautan Tweet	Tweet_ID	clean_text
0	@FahrudinLie	1.09 PM · 29 Jul 2025	Aku sempat baca berita mau tanya apakah betul...	Netral	https://x.com/FahrudinLie/status/195007614136...	1.950076e+18	baca berita negara perintah pra gib kena pajak...
1	@MamatAlkatiri	5.09 PM · 29 Agt 2025	DPR DAN PEMERINTAH STOP SEOLAH OLAH JADI PAHLA...	Negatif	https://x.com/MamatAlkatiri/status/19613706234...	1.961371e+18	dpr perintah stop olah olah pahlawan gak hati ...
2	@scryoty	5.32 PM · 29 Agt 2025	yups, terlalu toned deaf bngt mereka, kebnykan m...	Negatif	https://x.com/scryoty/status/1961376406090203217	1.961376e+18	yups toned deaf bngt kebnykan makan duit haram a...
3	@roman_kherzakov	5.36 PM · 29 Agt 2025	Ya begitulah, watak seseorang yang punya materi...	Negatif	https://x.com/roman_kherzakov/status/196137742...	1.961377e+18	watak orang materi nilai situ uang bersih hala...
4	@anggasa03640717	5.30 PM · 29 Agt 2025	sepakat, semua terjadi karena kelakuan mereka...	Negatif	https://x.com/anggasa03640717/status/196137588...	1.961376e+18	sepakat laku mulut sampah sahorni indonesia ru...

Gambar 4.1

Gambar 4.1 menunjukkan bagian dari data yang sudah melewati proses Preprocessing sebelum digunakan dalam proses klasifikasi sentimen. Pada tahap ini, teks asli dari media sosial sudah diproses untuk dibersihkan, yaitu dengan mengubah semua huruf menjadi huruf kecil, menghilangkan tanda baca, angka, serta karakter yang tidak penting, dan juga menghapus kata-kata yang dianggap tidak memberikan kontribusi signifikan dalam analisis.

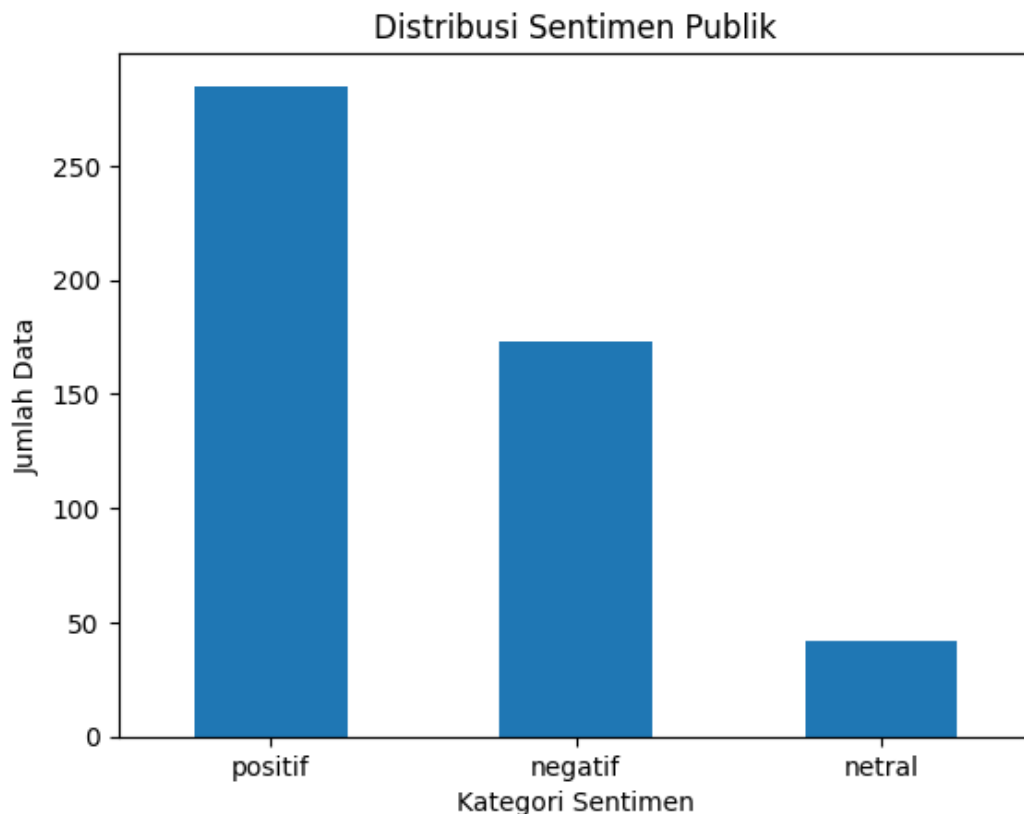
Hasil dari tahapan tersebut dapat dilihat di kolom `clean_text`, yang berisi teks yang sudah diperbaiki dan disederhanakan. Dibandingkan dengan teks asli, isi pada kolom `clean_text` hanya menyisakan kata-kata yang dianggap penting, sehingga memudahkan proses analisis di tahap selanjutnya.

Setelah proses preprocessing selesai, data selanjutnya disiapkan untuk tahap ekstraksi fitur dengan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini bertujuan mengubah teks menjadi angka dengan memberi nilai pada setiap kata sesuai seberapa sering kata tersebut muncul dalam dokumen serta seluruh kumpulan data. Kata-kata yang sering muncul dalam sebuah dokumen namun tidak sering muncul dalam dokumen lain akan mendapat bobot yang lebih besar karena dianggap lebih mampu mewakili isi dokumen tersebut.

Dengan proses Preprocessing dan pemberian bobot TF-IDF, data teks yang tidak teratur berhasil diubah menjadi data yang lebih rapi dan siap digunakan untuk proses klasifikasi sentimen. Tahap ini sangat penting karena kualitas hasil klasifikasi bergantung pada kualitas data yang diperoleh dari proses preprocessing dan ekstraksi fitur.

2. Hasil Klasifikasi Sentimen

Proses klasifikasi dilakukan menggunakan metode *Naïve Bayes Classifier* yang bekerja berdasarkan probabilitas kemunculan kata pada masing-masing kategori sentimen. Metode ini menggunakan *Teorema Bayes* untuk menghitung peluang suatu teks termasuk ke dalam kelas positif, negatif, atau netral. Dengan menggunakan rumus yang tertera pada nomor 2.1 .



Gambar 4.2

Visualisasi distribusi sentimen pada gambar 4.2 menggambarkan sebaran jumlah data pada masing-masing kategori sentimen, yaitu positif, negatif, dan netral dalam bentuk diagram batang. Berdasarkan grafik tersebut, terlihat bahwa sentimen positif merupakan kategori yang paling dominan dengan jumlah sekitar lebih dari 270 data.

Kategori sentimen negatif berada pada urutan kedua dengan jumlah sekitar 170 data, sedangkan kategori netral memiliki jumlah paling sedikit, yaitu kurang dari 50 data. Perbedaan jumlah ini menunjukkan adanya ketimpangan distribusi sentimen, dimana opini positif mendominasi keseluruhan data yang dianalisis.

Dari visualisasi ini dapat disimpulkan bahwa persepsi publik terhadap objek penelitian cenderung positif. Namun demikian, adanya sentimen negatif tetap

20

Submission ID trn:oid::2945:392192447

Namun, yang menarik adalah bagaimana kata-kata tersebut sering muncul berdampingan dengan istilah seperti “*kerja*”, “*jabatan*”, dan “*anggota*”. Hal ini menunjukkan adanya pola evaluatif dari masyarakat, yaitu membandingkan antara besaran tunjangan dengan kinerja atau peran yang dijalankan oleh anggota DPR.

Kata “*kerja*” dan “*kasihan*” juga menjadi indikator penting dalam memahami dinamika opini. Di satu sisi, kata “*kerja*” dapat mencerminkan tuntutan akuntabilitas publik terhadap kinerja DPR. Di sisi lain, kemunculan kata “*kasihan*” menunjukkan adanya sebagian kecil narasi empati atau pembelaan terhadap anggota DPR, meskipun tidak dominan. Dualitas ini mengindikasikan bahwa opini publik tidak sepenuhnya homogen, melainkan mengandung variasi perspektif, meskipun tetap didominasi oleh sudut pandang kritis.

Lebih lanjut, kemunculan kata seperti “*negara*”, “*indonesia*”, dan “*wakil*” menunjukkan bahwa isu ini juga dipahami dalam konteks yang lebih luas, yaitu legitimasi peran DPR sebagai representasi rakyat. Dengan demikian, kritik yang muncul tidak hanya bersifat individual, tetapi juga menyentuh aspek kelembagaan dan kepercayaan publik terhadap sistem pemerintahan.

Secara keseluruhan, *wordcloud* ini mengindikasikan bahwa opini publik terhadap tunjangan DPR RI didominasi oleh narasi evaluatif yang berfokus pada keadilan ekonomi, akuntabilitas kinerja, dan relasi antara pemerintah dengan masyarakat. Temuan ini memperlihatkan bahwa percakapan publik tidak sekadar bersifat reaktif, tetapi juga mengandung dimensi kritis yang mencerminkan kesadaran sosial terhadap isu distribusi sumber daya dan peran representatif lembaga negara.

4. Visualisasi *wordcloud* setiap sentimen

Untuk memperkuat hasil analisis sentimen dan emosi, penelitian ini juga menggunakan visualisasi *wordcloud* guna melihat kata-kata yang paling sering muncul pada setiap kategori sentimen maupun emosi. Visualisasi ini membantu menggambarkan kecenderungan opini publik terhadap isu tunjangan DPR RI berdasarkan dominasi kata yang digunakan dalam tweet. Semakin besar ukuran kata pada *wordcloud*, maka semakin sering kata tersebut muncul dalam data. Dengan demikian, visualisasi ini dapat memberikan gambaran mengenai fokus pembahasan, pola opini, serta konteks yang melatarbelakangi sentimen masyarakat terhadap isu yang diteliti.

a. *Wordcloud* sentimen positif



Visualisasi *wordcloud* pada sentimen positif yang ditunjukkan pada gambar 4.4 menampilkan beberapa kata yang paling sering muncul dalam opini masyarakat terkait kebijakan tunjangan DPR RI. Kata yang terlihat paling dominan adalah “rakyat”, “gaji”, “dp”, “kerja”, “tunjang”, dan “kasih”. Dominasi kata-kata tersebut menunjukkan bahwa sebagian opini publik yang termasuk dalam kategori positif cenderung membahas harapan masyarakat terhadap peningkatan kinerja anggota DPR RI setelah adanya tunjangan.

Selain itu, kata seperti “kerja”, “gaji”, dan “tunjang” menunjukkan bahwa sebagian pengguna media sosial mengaitkan kebijakan tersebut dengan tanggung jawab pekerjaan anggota DPR RI. Beberapa opini positif juga terlihat mengandung apresiasi atau dukungan terhadap kebijakan apabila tunjangan tersebut digunakan untuk meningkatkan kualitas kerja lembaga legislatif.

Secara keseluruhan, *wordcloud* sentimen positif memperlihatkan bahwa opini positif masyarakat lebih banyak berupa harapan, dukungan bersyarat, serta keinginan agar kebijakan tunjangan mampu memberikan dampak yang baik bagi kinerja anggota DPR RI dan kesejahteraan masyarakat.

29



64

142

187

12

Submission ID trn:oid::2945:392192447

berfokus pada hubungan antara anggota DPR RI dan masyarakat sebagai pihak yang terdampak oleh kebijakan tersebut. Hal ini menunjukkan bahwa opini netral lebih banyak digunakan untuk membahas fakta, informasi, maupun diskusi umum terkait isu tunjangan DPR RI.

Secara keseluruhan, *wordcloud* sentimen netral menggambarkan bahwa sebagian masyarakat menggunakan media sosial sebagai sarana untuk berdiskusi dan menyampaikan informasi mengenai kebijakan tunjangan DPR RI tanpa menunjukkan kecenderungan sentimen yang terlalu dominan.

5. Hasil Naïve Bayes

Hasil evaluasi model *Naïve Bayes* ditunjukkan melalui nilai akurasi, precision, recall, dan F1-score.

Tabel 4.1 hasil naive bayes

	Precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	2
Negatif	0.72	1.00	0.84	67
Netral	1.00	0.25	0.40	28
Positif	0.00	0.00	0.00	3
accuracy			0.74	100
macro avg	0.43	0.31	0.31	100
weighted avg	0.76	0.74	0.67	100

Hasil pengevaluasian model klasifikasi sentimen dengan metode Naïve Bayes Classifier (NBC) terlihat pada Tabel 4.1, yang menampilkan nilai precision, recall, f1-score, serta support untuk setiap kategori sentimen. Berdasarkan pengujian terhadap 100 data, model mencapai tingkat akurasi sebesar 0,74 atau 74%. Nilai tersebut menunjukkan bahwa sebagian besar data berhasil dikelompokkan ke dalam kategori sentimen yang tepat oleh model yang telah dibuat.

Jika dilihat per kelas sentimen, terdapat perbedaan dalam kemampuan model untuk mengenali setiap kategori. Dalam kelas negatif yang memiliki 67 data, model mendapatkan nilai precision sebesar 0,72, recall sebesar 1,00, dan f1-score sebesar 0,84. Nilai recall yang mencapai 1,00 menunjukkan bahwa semua data negatif dalam kelompok tersebut berhasil terdeteksi oleh model. Hasil ini menunjukkan bahwa model berhasil memahami pola sentimen negatif dengan cukup baik dibandingkan dengan kategori lainnya.

116 Dalam kelas netral yang menggunakan 28 data, model menghasilkan nilai
precision 1,00, recall 0,25, dan f1-score 0,40. Nilai presisi yang tinggi menunjukkan
24 bahwa prediksi yang dilakukan model terhadap kategori netral umumnya akurat.
Namun, nilai recall yang lebih rendah menunjukkan bahwa masih ada beberapa data
netral yang salah masuk ke dalam kategori sentimen yang tidak tepat. Kondisi itu
menunjukkan bahwa ciri-ciri sentimen netral mirip dengan kategori sentimen lainnya,
sehingga membuat proses mengenali sentimen lebih sulit.

185 Sementara itu, pada kategori positif yang memiliki jumlah data terkecil, yaitu
sebanyak 3 data, model belum berhasil mengenali data secara baik, sehingga
menghasilkan nilai precision, recall, dan f1-score sebesar 0,00. Kondisi yang sama juga
terjadi pada salah satu kelompok data negatif yang hanya memiliki dukungan dari 2
data saja. Hasil ini menunjukkan bahwa ketika jumlah data dalam satu kategori terlalu
sedikit, hal itu bisa mengganggu kemampuan model untuk memahami pola yang
mewakili kelas tersebut.

109 Berdasarkan rata-rata keseluruhan, hasilnya menunjukkan presisi makro sebesar
0,43, recall makro sebesar 0,31, dan skor f1 makro sebesar 0,31. Nilai ini menunjukkan
80 seberapa baik model bekerja dengan memberikan pengaruh yang sama untuk semua
jenis perasaan, tanpa memperhatikan berapa banyak data yang ada di setiap jenis
87 perasaan. Di sisi lain, nilai presisi rata-rata terbobot sebesar 0,76, recall rata-rata
terbobot sebesar 0,74, dan f1-score rata-rata terbobot sebesar 0,67 menunjukkan
bagaimana performa model setelah memperhitungkan jumlah data pada setiap kategori
sentimen.

127 Perbedaan antara rata-rata makro dan rata-rata berbobot menunjukkan bahwa
data dalam penelitian ini tidak merata terdistribusi. Hal itu bisa dilihat dari jumlah data
sentimen negatif yang lebih banyak dibandingkan dengan kategori netral atau positif.
Dalam proses klasifikasi, kondisi ini disebut class imbalance, yaitu situasi di mana
212 jumlah data di setiap kelas berbeda secara signifikan. Distribusi data yang tidak
135 seimbang bisa membuat model lebih cepat mengenali pola pada kelas yang memiliki
data lebih banyak dibandingkan kelas dengan jumlah data lebih sedikit.

37 Secara keseluruhan, hasil pengujian menunjukkan bahwa metode Naïve Bayes
Classifier mampu melakukan klasifikasi sentimen dengan tingkat akurasi yang cukup
12 baik pada data penelitian. Model menunjukkan kemampuan yang lebih stabil dalam
mengenali kategori sentimen yang memiliki banyak data, sedangkan kemampuan pada
kategori dengan data yang sedikit masih bisa ditingkatkan. Hasil penelitian ini

menunjukkan bahwa cara penyusunan data juga berpengaruh terhadap hasil klasifikasi yang didapatkan.

6. Hasil *Lexicon* Sentiment

	count
<i>Lexicon sentiment</i>	
positif	285
negatif	173
netral	42

Tabel 4.2

Hasil klasifikasi sentimen menggunakan metode *Lexicon* yang di paparkan pada tabel 4.2 menunjukkan distribusi sentimen yang terbagi ke dalam tiga kategori, yaitu positif, negatif, dan netral. Berdasarkan hasil perhitungan, sentimen positif mendominasi dengan jumlah sebanyak 285 data, diikuti oleh sentimen negatif sebanyak 173 data, dan sentimen netral sebanyak 42 data.

Dominasi sentimen positif menunjukkan bahwa sebagian besar opini publik terhadap isu tunjangan DPR RI cenderung mengandung kata-kata yang secara leksikal bernilai positif. Hal ini mengindikasikan bahwa dalam banyak teks, terdapat penggunaan kata-kata yang diasosiasikan dengan makna positif seperti “baik”, “senang”, atau kata lain yang terdaftar dalam kamus *Lexicon*. Namun demikian, hasil ini perlu ditafsirkan secara hati-hati karena metode *Lexicon* hanya mengandalkan pencocokan kata tanpa mempertimbangkan konteks kalimat secara keseluruhan.

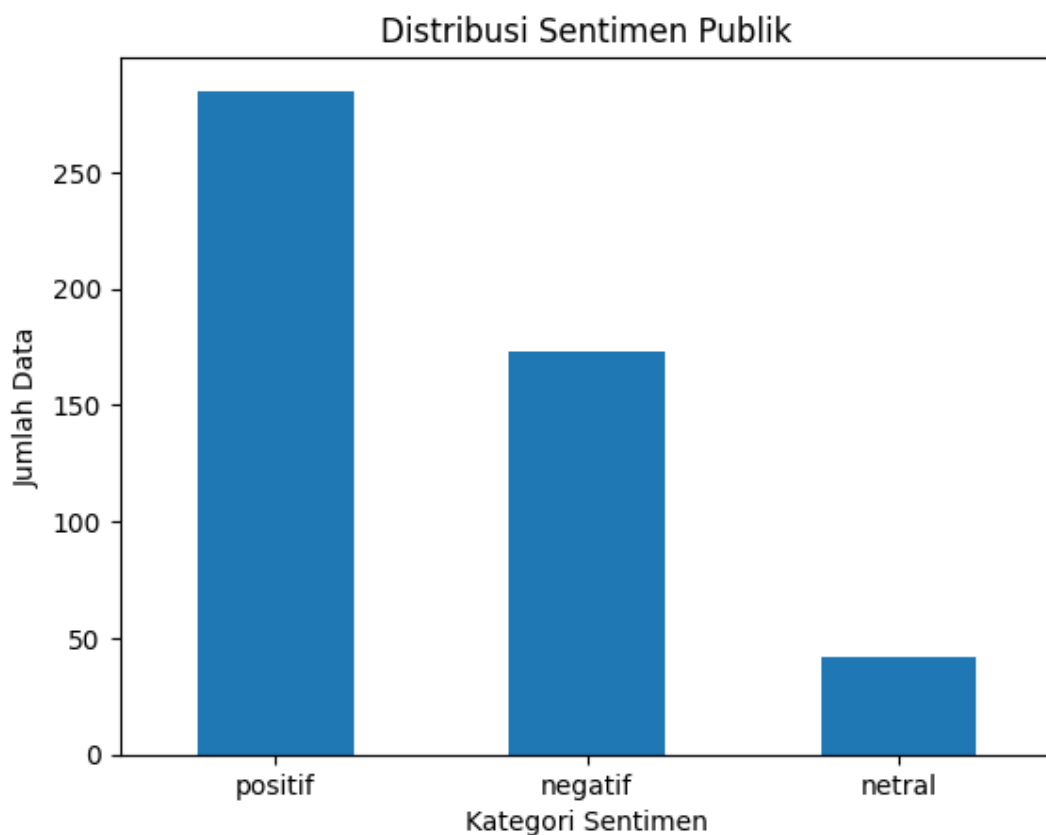
Di sisi lain, jumlah sentimen negatif yang cukup signifikan (173 data) menunjukkan bahwa terdapat pula opini publik yang mengandung ekspresi ketidakpuasan atau kritik terhadap kebijakan tersebut. Hal ini memperlihatkan bahwa isu tunjangan DPR RI tetap menjadi topik yang memunculkan respon beragam di masyarakat.

Sementara itu, jumlah sentimen netral yang relatif sedikit (42 data) mengindikasikan bahwa hanya sebagian kecil data yang tidak mengandung muatan emosional atau cenderung bersifat informatif. Rendahnya jumlah sentimen netral juga dapat disebabkan oleh karakteristik metode *Lexicon* yang cenderung “memaksa” klasifikasi ke arah positif atau negatif ketika menemukan kata-kata yang memiliki skor sentimen, sehingga kategori netral menjadi kurang dominan.

Secara keseluruhan, hasil ini menunjukkan bahwa metode *Lexicon* cenderung menghasilkan distribusi sentimen yang lebih condong ke arah positif. Namun, kecenderungan ini tidak selalu mencerminkan kondisi sebenarnya, karena metode *Lexicon* memiliki keterbatasan dalam memahami konteks, ironi, serta hubungan antar kata dalam kalimat. Oleh karena itu, hasil ini perlu dibandingkan dengan metode lain seperti *Naïve Bayes* untuk memperoleh gambaran yang lebih akurat mengenai sentimen publik.

7. Hasil Integrasi kedua Metode

Integrasi dilakukan dengan menggabungkan hasil *Naïve Bayes* dan *Lexicon sentiment*.



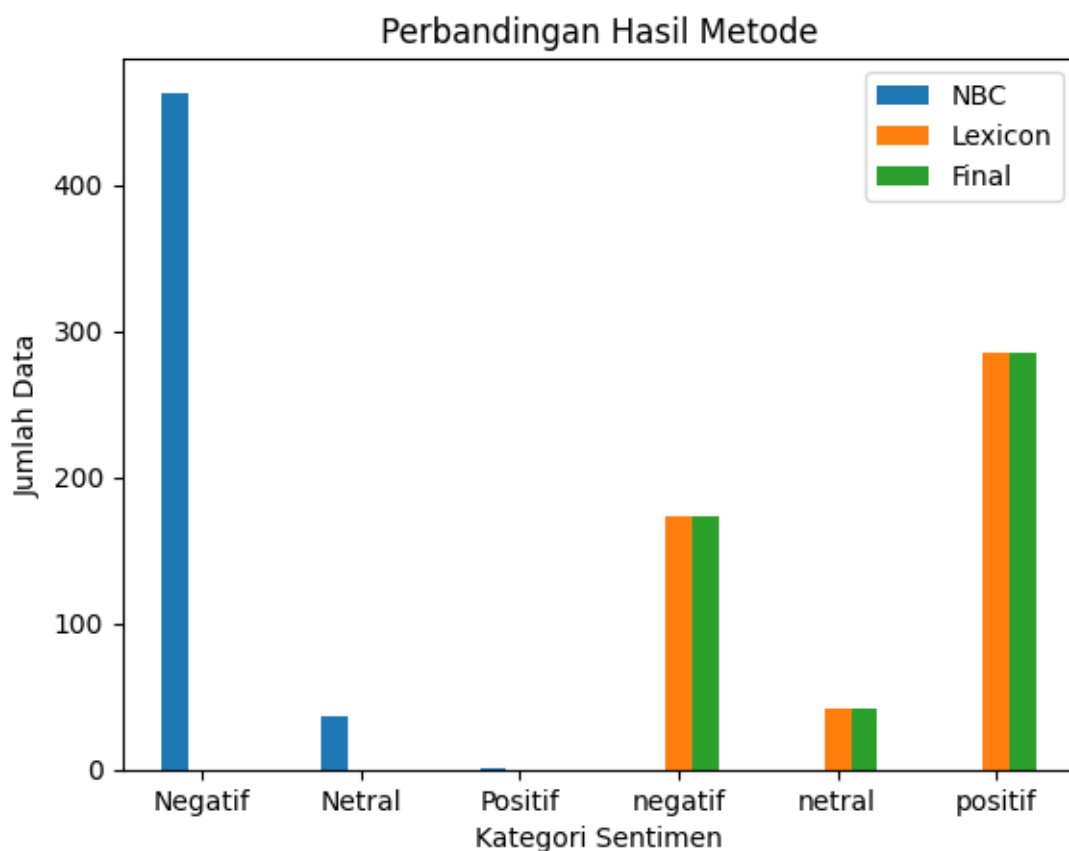
Gambar 4.7

Hasil integrasi data pada gambar 4.7 di atas menunjukkan bahwa setelah proses penggabungan dan klasifikasi sentimen, data berhasil dikelompokkan ke dalam tiga kategori utama, yaitu positif, negatif, dan netral. Proses ini menghasilkan distribusi awal yang memperlihatkan bahwa sebagian besar data memiliki kecenderungan sentimen positif, diikuti oleh sentimen negatif, dan sebagian kecil lainnya termasuk dalam kategori netral. Dominasi sentimen positif mengindikasikan bahwa secara umum data

yang dianalisis memiliki kecenderungan opini yang mendukung atau bernilai apresiatif terhadap objek penelitian.

Sementara itu, keberadaan sentimen negatif menunjukkan masih adanya respon yang bersifat kritis atau tidak puas, meskipun tidak mendominasi. Adapun sentimen netral yang jumlahnya relatif sedikit menunjukkan bahwa hanya sebagian kecil data yang tidak mengandung polaritas emosi yang kuat. Hasil *integrasi* ini menjadi dasar awal dalam memahami karakteristik dataset sebelum dilakukan analisis lebih lanjut pada tahap visualisasi dan interpretasi mendalam.

8. Perbandingan Metode



Gambar 4.8

Gambar 4.8 perbandingan hasil metode menunjukkan distribusi klasifikasi sentimen yang dihasilkan oleh tiga pendekatan, yaitu *Naïve Bayes Classifier* (NBC), metode *Lexicon*, dan hasil final (gabungan/penyesuaian). Berdasarkan grafik tersebut, terlihat adanya perbedaan yang cukup signifikan dalam pola distribusi sentimen antar metode.

Pada metode NBC, data cenderung sangat terpusat pada kelas negatif dengan jumlah yang jauh lebih dominan dibandingkan kelas lainnya. Sementara itu, jumlah data pada kelas netral dan positif relatif sangat sedikit. Hal ini menunjukkan bahwa model NBC memiliki kecenderungan bias terhadap kelas mayoritas, yang dalam hal ini adalah sentimen negatif. Kondisi ini sejalan dengan hasil evaluasi sebelumnya, di mana model mengalami kesulitan dalam mengenali kelas dengan jumlah data yang lebih kecil.

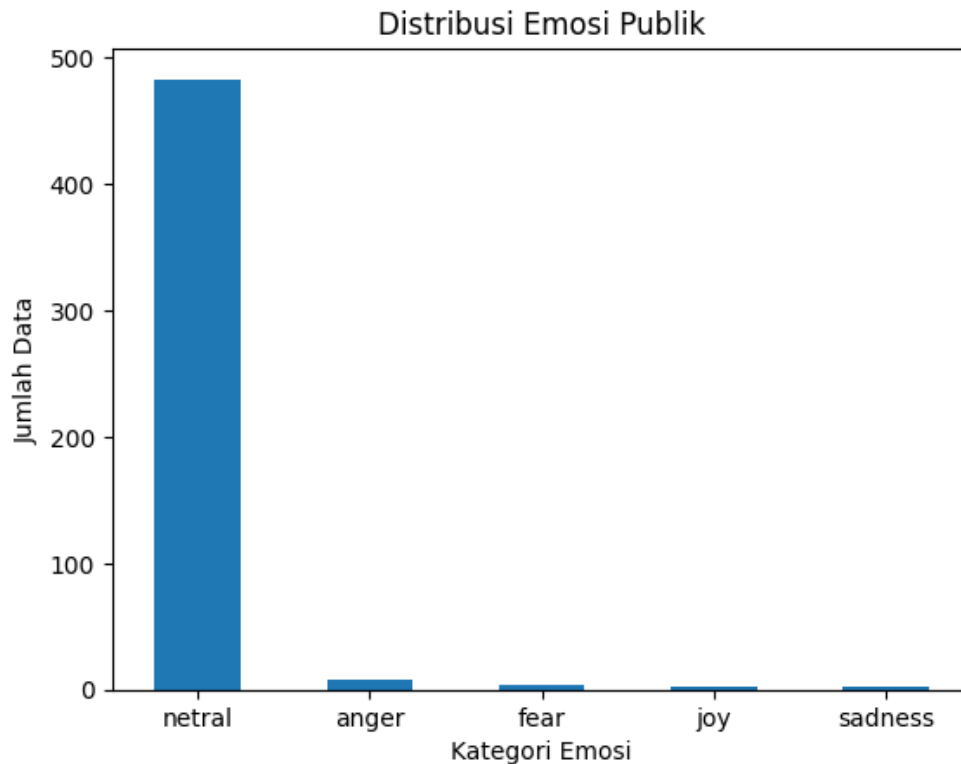
Berbeda dengan NBC, metode *Lexicon* menghasilkan distribusi sentimen yang lebih seimbang antara kelas negatif, netral, dan positif. Terlihat bahwa jumlah data pada masing-masing kategori tidak terlalu timpang, sehingga metode ini mampu memberikan representasi sentimen yang lebih proporsional. Hal ini disebabkan oleh pendekatan *Lexicon* yang berbasis kamus kata, sehingga setiap kata memiliki kontribusi langsung terhadap penentuan polaritas tanpa terlalu dipengaruhi oleh distribusi data latih.

Sementara itu, hasil final menunjukkan pola distribusi yang hampir identik dengan metode *Lexicon*. Hal ini mengindikasikan bahwa dalam proses penggabungan atau penyesuaian, kontribusi metode *Lexicon* lebih dominan dibandingkan NBC. Dengan kata lain, pendekatan *Lexicon* mampu memperbaiki kelemahan NBC, khususnya dalam menangani ketidakseimbangan data dan meningkatkan representasi pada kelas netral dan positif.

Berdasarkan perbandingan tersebut, dapat disimpulkan bahwa metode NBC cenderung unggul dalam mengenali kelas dengan jumlah data besar, namun kurang efektif pada data yang tidak seimbang. Sebaliknya, metode *Lexicon* memberikan hasil yang lebih stabil dan merata antar kelas. Oleh karena itu, penggunaan pendekatan kombinasi (hybrid) menjadi solusi yang tepat untuk memperoleh hasil klasifikasi yang lebih seimbang dan representatif terhadap kondisi sentimen yang sebenarnya.

9. Analisis Makna Emosional

Selain analisis sentimen, dilakukan analisis emosi untuk memahami makna emosional dalam opini publik.



Gambar 4.9

149 Gambar distribusi emosi publik menunjukkan hasil klasifikasi makna emosional dari data yang dianalisis, yang terbagi ke dalam lima kategori utama, yaitu netral, anger (marah), *fear* (takut), *joy* (senang), dan *sadness* (sedih). Berdasarkan grafik tersebut, terlihat bahwa emosi netral memiliki jumlah yang sangat dominan dibandingkan kategori emosi lainnya.

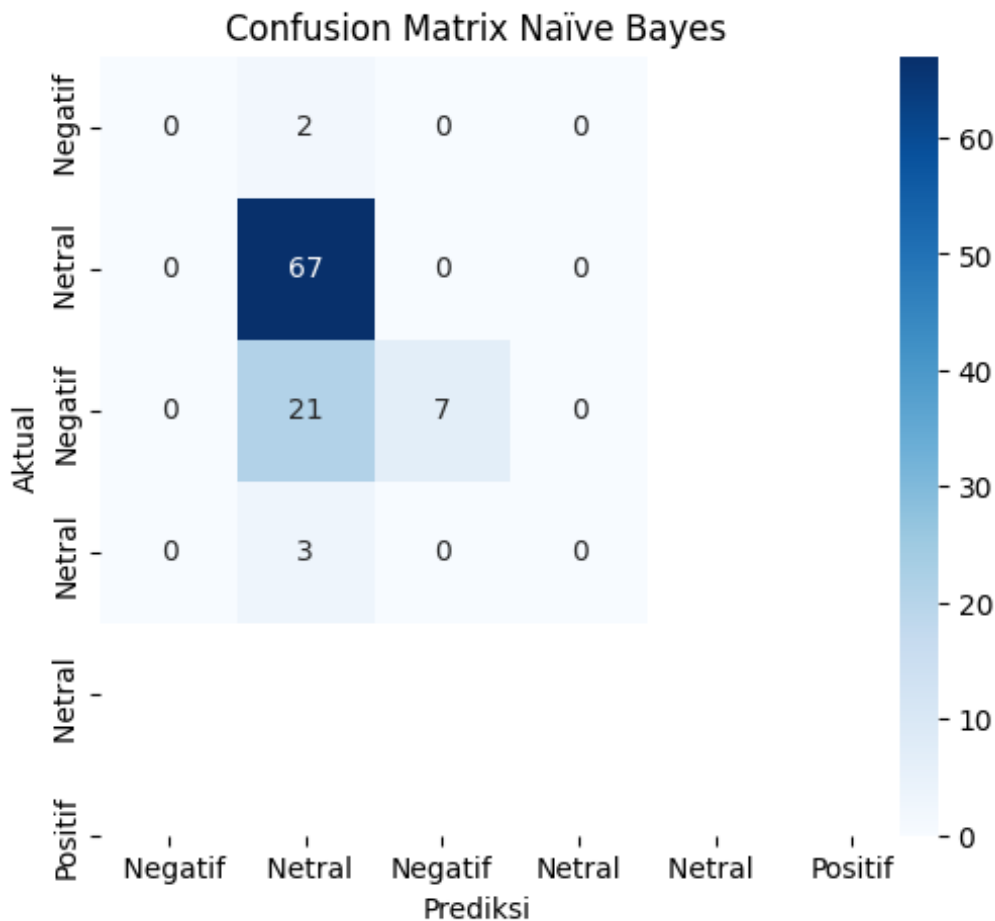
Dominasi emosi netral mengindikasikan bahwa sebagian besar data yang dianalisis tidak mengandung ekspresi emosi yang kuat atau cenderung bersifat informatif dan deskriptif. Hal ini menunjukkan bahwa dalam konteks pembahasan terkait tunjangan DPR RI, banyak pengguna media sosial yang menyampaikan opini secara objektif tanpa disertai ekspresi emosional yang intens.

Sementara itu, emosi anger (marah) muncul sebagai kategori emosi yang paling dominan setelah netral, meskipun jumlahnya masih jauh lebih kecil. Kemunculan emosi ini menunjukkan adanya sebagian kecil masyarakat yang mengekspresikan ketidakpuasan atau kritik terhadap isu yang dibahas. Emosi *fear* (takut), *sadness* (sedih), dan *joy* (senang) juga terdeteksi, namun dengan jumlah yang relatif sangat sedikit. Hal ini mengindikasikan bahwa respon emosional yang bersifat kuat, baik positif maupun negatif, tidak banyak ditemukan dalam data.

Rendahnya jumlah emosi joy menunjukkan bahwa sentimen positif yang disertai ekspresi kebahagiaan atau kepuasan terhadap kebijakan tersebut sangat terbatas. Demikian pula dengan emosi fear dan sadness yang menunjukkan bahwa kekhawatiran maupun kesedihan bukan merupakan respon dominan dari masyarakat terhadap isu ini.

Secara keseluruhan, distribusi emosi ini menunjukkan bahwa persepsi publik terhadap tunjangan DPR RI cenderung berada pada ranah netral dengan sebagian kecil kecenderungan emosi negatif, khususnya kemarahan. Hal ini dapat menjadi indikasi bahwa isu tersebut lebih banyak dipandang sebagai topik diskusi daripada memicu reaksi emosional yang kuat di kalangan masyarakat.

10. Visualisasi Confusion Matrix



Gambar 4.10

Berdasarkan Gambar 4.10, confusion matrix digunakan untuk menggambarkan performa model *Naïve Bayes Classifier* dalam mengklasifikasikan data sentimen ke dalam kategori negatif, netral, dan positif. Visualisasi ini memperlihatkan hubungan antara data aktual dengan hasil prediksi model sehingga dapat diketahui jumlah data yang berhasil maupun gagal diklasifikasikan dengan benar.

Pada confusion matrix terlihat bahwa sebagian besar data berhasil diprediksi ke dalam kategori netral. Hal ini ditunjukkan oleh nilai prediksi yang cukup tinggi pada kategori tersebut dibandingkan kategori lainnya. Model mampu mengenali sebanyak 67 data netral dengan benar sesuai dengan label aktualnya. Hasil ini menunjukkan bahwa model memiliki kecenderungan lebih kuat dalam mengenali pola kata yang termasuk dalam kategori netral dibandingkan kategori sentimen lainnya.

Namun demikian, masih terdapat beberapa kesalahan klasifikasi yang terjadi pada model. Sebagian data aktual pada kategori negatif diprediksi sebagai netral oleh sistem. Kondisi ini menunjukkan bahwa model mengalami kesulitan dalam membedakan kata-kata bernuansa negatif dengan kata-kata yang bersifat netral, terutama pada teks media sosial yang mengandung bahasa informal, singkatan, maupun konteks kalimat yang ambigu. Selain itu, jumlah data positif yang sangat sedikit menyebabkan model kurang mampu mempelajari pola sentimen positif secara optimal.

Ketidakseimbangan jumlah data pada masing-masing kategori sentimen juga memengaruhi performa model. Data netral memiliki jumlah yang lebih dominan dibandingkan kategori positif sehingga model cenderung menghasilkan prediksi ke arah kategori yang memiliki jumlah data lebih besar. Akibatnya, kemampuan model dalam mengenali sentimen positif menjadi kurang maksimal dan menyebabkan beberapa data tidak dapat diklasifikasikan dengan tepat.

Selain dipengaruhi oleh distribusi data yang tidak seimbang, performa model juga dipengaruhi oleh karakteristik data media sosial yang cenderung dinamis. Banyak opini publik menggunakan bahasa tidak baku, singkatan, kata berulang, hingga bentuk sarkasme yang sulit dikenali hanya melalui pendekatan probabilistik. Kondisi tersebut menyebabkan beberapa kata yang sebenarnya mengandung makna emosional tertentu justru diprediksi ke dalam kategori netral karena model hanya melihat pola kemunculan kata berdasarkan data pelatihan.

Di sisi lain, confusion matrix juga menunjukkan bahwa penggunaan *Naïve Bayes* Classifier masih cukup efektif untuk mengidentifikasi kecenderungan umum sentimen publik terhadap kebijakan tunjangan DPR RI. Meskipun belum mampu memberikan klasifikasi yang sempurna pada seluruh kategori, model tetap dapat digunakan untuk memperoleh gambaran awal mengenai opini masyarakat di media sosial. Oleh karena itu, penelitian ini menambahkan pendekatan *Lexicon-Based Sentiment* agar proses klasifikasi tidak hanya mempertimbangkan probabilitas kata, tetapi juga makna

emosional yang terkandung dalam teks sehingga hasil analisis menjadi lebih kontekstual.

Secara keseluruhan, confusion matrix menunjukkan bahwa metode *Naïve Bayes Classifier* cukup baik dalam mengenali pola umum sentimen pada data opini publik. Akan tetapi, model masih memiliki keterbatasan dalam menangani data yang memiliki kemiripan kata antar kategori sentimen dan konteks bahasa media sosial yang kompleks. Dengan adanya integrasi metode *Lexicon*, diharapkan hasil klasifikasi sentimen dapat menjadi lebih akurat dan mampu merepresentasikan opini publik secara lebih menyeluruh.

B. Pembahasan

1. Model Paling Optimal

Model integrasi antara *Naïve Bayes* dan *Lexicon* merupakan pendekatan yang paling optimal dalam penelitian ini. Hal ini disebabkan karena kombinasi kedua metode mampu saling melengkapi kelemahan masing-masing. *Naïve Bayes* sebagai metode berbasis statistik efektif dalam menangkap pola dari data latih, khususnya pada kelas dengan jumlah data besar. Namun, metode ini cenderung mengalami penurunan performa pada data yang tidak seimbang (*imbalanced data*).

Di sisi lain, pendekatan *Lexicon* yang berbasis kamus mampu memberikan interpretasi makna kata secara langsung tanpa bergantung pada distribusi data. Hal ini membuat metode *Lexicon* lebih stabil dalam mengidentifikasi sentimen pada kelas minoritas. Oleh karena itu, integrasi kedua metode menghasilkan performa yang lebih seimbang, terutama dalam memperbaiki bias klasifikasi dan meningkatkan representasi sentimen secara keseluruhan.

2. Evaluasi terhadap Tujuan Penelitian

Evaluasi terhadap tujuan penelitian menunjukkan bahwa pendekatan yang digunakan dalam penelitian ini telah mampu memenuhi tujuan utama penelitian, yaitu melakukan analisis sentimen terhadap opini publik mengenai kebijakan tunjangan DPR RI pada media sosial X (*Twitter*). Penerapan metode *Naïve Bayes Classifier* berhasil digunakan untuk mengklasifikasikan data opini ke dalam kategori sentimen positif, negatif, dan netral berdasarkan pola kata yang terdapat pada teks. Hasil evaluasi model menunjukkan bahwa metode yang digunakan mampu memberikan performa klasifikasi yang cukup baik berdasarkan nilai *accuracy*, *precision*, *recall*, dan *f1-score* yang diperoleh selama proses pengujian model.

Selain melakukan klasifikasi sentimen, penelitian ini juga menerapkan pendekatan *Lexicon Sentiment* untuk membantu memahami konteks emosional yang terkandung dalam opini publik. Penggunaan metode tersebut memberikan hasil yang lebih mendalam karena tidak hanya berfokus pada pengelompokan sentimen, tetapi juga mampu menggambarkan kecenderungan emosi masyarakat terhadap isu tunjangan DPR RI. Hal ini terlihat dari hasil distribusi emosi yang menunjukkan adanya dominasi emosi tertentu dalam data opini publik yang dianalisis.

Visualisasi data yang dilakukan melalui grafik distribusi sentimen dan *wordcloud* juga membantu memperlihatkan pola pembahasan yang paling sering muncul dalam percakapan masyarakat. Beberapa kata yang dominan menunjukkan bahwa opini publik banyak berkaitan dengan isu kesejahteraan, pajak, kinerja anggota DPR, dan kondisi ekonomi masyarakat. Hasil tersebut memperlihatkan bahwa respons masyarakat terhadap kebijakan tunjangan DPR RI tidak hanya berupa dukungan maupun penolakan, tetapi juga mencerminkan berbagai bentuk ekspresi emosional yang berkembang di media sosial.

Berdasarkan keseluruhan hasil pengujian dan analisis yang dilakukan, penelitian ini dapat dikatakan berhasil mencapai tujuan penelitian karena mampu memberikan gambaran mengenai kecenderungan sentimen publik sekaligus memahami konteks opini masyarakat secara lebih mendalam. Integrasi antara metode *Naïve Bayes Classifier* dan *Lexicon Sentiment* juga memberikan hasil analisis yang lebih representatif dibandingkan jika hanya menggunakan satu metode saja dalam proses analisis sentimen.

3. Kesesuaian dengan Business Understanding

Hasil penelitian menunjukkan bahwa opini publik terhadap kebijakan tunjangan DPR RI cenderung didominasi oleh sentimen negatif dan netral, dengan fokus utama pada isu ekonomi seperti pajak, gaji, dan kesejahteraan masyarakat. Selain itu, analisis emosi memperlihatkan bahwa sebagian kecil respon masyarakat mengandung emosi negatif seperti kemarahan, yang berkaitan dengan persepsi ketidakadilan dan ketimpangan sosial.

Temuan ini menunjukkan bahwa masyarakat tidak hanya menilai kebijakan secara administratif, tetapi juga dalam kerangka keadilan sosial dan distribusi sumber daya. Oleh karena itu, hasil penelitian ini sejalan dengan tujuan awal (*business understanding*), yaitu untuk memahami respon dan persepsi masyarakat terhadap kebijakan publik, khususnya yang berkaitan dengan kepentingan ekonomi dan kesejahteraan.

4. Validitas dan Generalisasi

Model yang digunakan dalam penelitian ini menunjukkan performa yang cukup baik pada dataset yang digunakan, khususnya dalam mengidentifikasi sentimen dominan. Namun demikian, terdapat beberapa keterbatasan yang perlu diperhatikan. Salah satunya adalah

kemampuan model yang masih terbatas dalam memahami bahasa informal, singkatan, serta konteks sarkasme yang sering muncul dalam media sosial.

Selain itu, ketidakseimbangan jumlah data antar kelas juga memengaruhi kinerja model, terutama pada kelas dengan jumlah data yang kecil seperti sentimen positif. Oleh karena itu, untuk meningkatkan validitas dan kemampuan generalisasi, model perlu diuji pada dataset yang lebih besar dan beragam agar dapat merepresentasikan kondisi nyata secara lebih akurat.

5. Rekomendasi

Model yang dikembangkan dalam penelitian ini dapat digunakan sebagai dasar dalam sistem pemantauan opini publik berbasis media sosial, khususnya untuk menganalisis respon masyarakat terhadap kebijakan pemerintah secara cepat dan efisien.

Pengembangan lebih lanjut dapat dilakukan dengan:

- a. menambahkan dataset yang lebih besar dan lebih beragam untuk mengurangi bias dan meningkatkan akurasi model,
- b. menggunakan metode *deep learning* seperti LSTM atau transformer untuk menangkap konteks bahasa yang lebih kompleks,
- c. mengembangkan analisis emosi yang lebih mendalam dengan kategori emosi yang lebih beragam,
- d. serta mengintegrasikan deteksi sarkasme dan konteks bahasa informal agar hasil analisis lebih akurat dan kontekstual.

BAB V

PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dengan menggunakan pendekatan CRISP-DM, dapat disimpulkan bahwa penelitian ini berhasil menganalisis pendapat publik mengenai kebijakan tunjangan anggota DPR RI di platform media sosial X (Twitter) melalui metode analisis sentimen dan analisis emosi.

Dalam tahap pemodelan, metode Klasifikasi Naïve Bayes memberikan tingkat akurasi sebesar 0,74 dan skor rata-rata weighted f1 sebesar 0,67. Hasil tersebut menunjukkan bahwa model tersebut mampu dengan cukup baik mengelompokkan sentimen publik ke dalam kategori-kategori yang telah ditentukan. Metode analisis sentimen berbasis kamus ini dapat memberikan distribusi sentimen yang lebih seimbang dan lebih mampu memahami makna emosional dalam teks secara lebih tepat sesuai konteksnya.

Menggabungkan metode Klasifikasi Naïve Bayes dengan Sentimen Berbasis Kamus memberikan hasil analisis sentimen yang lebih akurat karena menggabungkan kelebihan pendekatan statistik dan pendekatan berbasis kamus. Hasil evaluasi dengan menggunakan confusion matrix serta metrik akurasi, presisi, recall, dan f1-score menunjukkan bahwa model yang dibuat cukup baik dalam mengenali sentimen utama dari opini masyarakat mengenai kebijakan tunjangan DPR RI.

Selain itu, hasil analisis perasaan menunjukkan bahwa mayoritas pendapat masyarakat termasuk dalam kategori perasaan netral dan marah. Hasil penelitian ini menunjukkan bahwa kebijakan tunjangan yang diberikan oleh DPR RI mendapat berbagai reaksi dari masyarakat, baik berupa kritik maupun emosi lainnya. Dengan demikian, tujuan penelitian ini untuk mengklasifikasikan sentimen publik dan mengidentifikasi emosi yang ada dalam pendapat masyarakat telah berhasil tercapai.

B. Saran

Berdasarkan hasil penelitian yang sudah didapatkan, ada beberapa rekomendasi yang bisa diberikan untuk penelitian berikutnya. Berikutnya,

117

penelitian bisa menggunakan jumlah data yang lebih banyak dan jangka waktu pengumpulan data yang lebih lama agar bisa menggambarkan pendapat masyarakat secara lebih lengkap.

5

Selanjutnya, penelitian dapat dilakukan untuk membandingkan hasil kerja metode Klasifikasi Naïve Bayes dengan algoritma klasifikasi lainnya, seperti Support Vector Machine (SVM), Random Forest, atau metode berbasis Deep Learning agar mendapatkan hasil klasifikasi yang lebih baik.

148

Ketiga, pembuatan kamus sentimen dan kamus emosi yang lebih cocok dengan konteks bahasa Indonesia, terutama bahasa yang digunakan di media sosial, bisa dilakukan agar hasil analisis sentimen dan emosi lebih akurat. Selain itu, penelitian berikutnya juga bisa mempertimbangkan penggunaan teknik dalam mengolah bahasa yang santai, singkatan, serta kata-kata yang sering digunakan oleh pengguna media sosial.

28

102

130

Akhirnya, hasil penelitian ini diharapkan bisa menjadi bahan acuan bagi para peneliti lain dalam melakukan penelitian tentang analisis sentimen dan analisis emosi terkait isu-isu kebijakan publik, sehingga dapat memberikan pemahaman yang lebih dalam mengenai bagaimana persepsi masyarakat terhadap suatu kebijakan