



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 : Mochamad Bressandy Syach Putra
NPM : 2213020033
Program Studi : S1-Teknik Informatika

Judul Karya Ilmiah:

“ANALISIS SENTIMEN TERHADAP PROYEK PEMBANGUNAN JALAN TOL KEDIRI-TULUNGAGUNG
MENGUNAKAN PENDEKATAN LEXICON BASED”

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, 18 Juni 2026
Ka UPT PPI,



Dr. Abdul Aziz Hunaifi, M.A



Turnitin LLCC

2213020033_Mochamad Bressandy Syach Putra

 TI-09

Document Details

Submission ID

trn:oid::2945:394042465

Submission Date

Jun 18, 2026, 3:42 PM GMT+7

Download Date

Jun 18, 2026, 3:45 PM GMT+7

File Name

2213020033_Mochamad Bressandy Syach Putra.pdf

File Size

5.0 MB

46 Pages

7,827 Words

51,589 Characters

26% Overall Similarity

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

Filtered from the Report

- Bibliography

Top Sources

- 13%  Internet sources
- 10%  Publications
- 22%  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

- 13% Internet sources
- 10% Publications
- 22% Submitted works (Student Papers)

Top Sources

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

1	Student papers	Transylvania University on 2024-07-25	<1%
2	Student papers	Universitas Putera Batam on 2026-02-02	<1%
3	Internet	digilib.unila.ac.id	<1%
4	Student papers	Universitas Budi Luhur on 2026-02-07	<1%
5	Internet	digilib.eng.unhas.ac.id	<1%
6	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2025-08-26	<1%
7	Student papers	Universitas Khairun on 2026-05-18	<1%
8	Internet	repository.unissula.ac.id	<1%
9	Internet	repository.unpkediri.ac.id	<1%
10	Student papers	Universitas Putera Batam on 2026-01-31	<1%
11	Student papers	Universitas Budi Luhur on 2025-07-16	<1%

12	Student papers	SDM Universitas Gadjah Mada on 2021-10-08	<1%
13	Internet	eprints.walisongo.ac.id	<1%
14	Publication	Rossiana Ginting, Satria Prayudi, Fahsi Mahmoud, Handaru Jati et al. "Machine Le...	<1%
15	Student papers	Universitas Pamulang on 2026-06-09	<1%
16	Internet	ejurnal.univbatam.ac.id	<1%
17	Internet	cdn.juris.id	<1%
18	Internet	repository.iainpurwokerto.ac.id	<1%
19	Student papers	Universitas Khairun on 2026-06-09	<1%
20	Student papers	Universitas Tarumanagara on 2025-12-22	<1%
21	Student papers	Universitas Tarumanagara on 2025-12-29	<1%
22	Internet	ejournal.itats.ac.id	<1%
23	Internet	repository.unp.ac.id	<1%
24	Student papers	UPN Veteran Yogyakarta on 2026-06-03	<1%
25	Internet	digilibadmin.unismuh.ac.id	<1%

26	Publication	Sekar Cinta Amaria, Nurtriana Hidayati. "ANALISIS PROGRAM MAKAN BERGIZI G...	<1%
27	Student papers	UNIVERSITAS BUDI LUHUR on 2025-08-26	<1%
28	Student papers	Universitas Brawijaya on 2017-07-20	<1%
29	Internet	jurnal.kdi.or.id	<1%
30	Student papers	Universitas Islam Indonesia on 2026-05-12	<1%
31	Internet	repository.upstegal.ac.id	<1%
32	Student papers	Universitas Pamulang on 2022-06-14	<1%
33	Internet	es.scribd.com	<1%
34	Student papers	Telkom University on 2026-05-04	<1%
35	Student papers	Universitas Negeri Padang on 2025-07-21	<1%
36	Student papers	Universitas Muhammadiyah Palembang on 2025-08-06	<1%
37	Student papers	Universitas Pancasila on 2026-02-21	<1%
38	Internet	repository.umb.ac.id	<1%
39	Internet	repository.upi.edu	<1%

40	Publication	Iwan Kurniawan. "Peningkatan Pelayanan Publik: Pengelolaan Ekowisata Di Daer...	<1%
41	Student papers	Universitas Islam Lamongan on 2021-06-16	<1%
42	Student papers	Universitas Pamulang on 2026-05-04	<1%
43	Student papers	Universitas Putera Batam on 2026-01-31	<1%
44	Publication	Yudha Prastya. "HYBRID CNN-SVM UNTUK ANALISIS SENTIMEN KLUB TIM NASIO...	<1%
45	Publication	Ayu Oktaviyanti. "ANALISIS SENTIMEN ULASAN PELECEHAN DI MEDIA SOSIAL TIK...	<1%
46	Student papers	Universitas Brawijaya on 2017-07-06	<1%
47	Student papers	Universitas Sangga Buana YPKP on 2026-04-09	<1%
48	Internet	indojurnal.com	<1%
49	Internet	repository.ipb.ac.id	<1%
50	Publication	Siti Aulia Rahmadhani, Lia Dwi Rusanti, Harun Al Rosyid. "Klasifikasi Sentimen Ko...	<1%
51	Student papers	Universitas Islam Riau on 2026-04-30	<1%
52	Student papers	Universitas Negeri Semarang - iTh on 2024-07-14	<1%
53	Internet	journal.amikveteran.ac.id	<1%

54	Internet	repository.uinsu.ac.id	<1%
55	Student papers	Universitas Ibn Khaldun on 2024-08-31	<1%
56	Publication	Zahrotun Ni'mah -, Bambang Irawan -, Nur Ariesanto Ramdhan -. "ANALISIS SENT...	<1%
57	Internet	depaulatrading.com	<1%
58	Internet	journal.umkendari.ac.id	<1%
59	Internet	kc.umh.ac.id	<1%
60	Internet	p3m.sinus.ac.id	<1%
61	Internet	repository.wicida.ac.id	<1%
62	Student papers	Udayana University on 2026-01-08	<1%
63	Student papers	Universitas Indonesia on 2025-05-13	<1%
64	Student papers	Universitas Islam Negeri Sumatera Utara on 2026-01-23	<1%
65	Student papers	Universitas Islam Riau on 2026-03-11	<1%
66	Student papers	Universitas Islam Riau on 2026-04-07	<1%
67	Student papers	Universitas Negeri Surabaya The State University of Surabaya on 2026-06-04	<1%

68	Student papers	Universitas Putera Batam on 2026-02-02	<1%
69	Internet	ashabulikhwan.wordpress.com	<1%
70	Internet	eprints.uns.ac.id	<1%
71	Internet	tutukscorpio.wordpress.com	<1%
72	Student papers	FAKULTAS ILMU KOMPUTER on 2026-01-08	<1%
73	Publication	Sarjono. "Kepuasan Kerja Guru Pada Lembaga Paud Islam di Kabupaten Cilacap (...)	<1%
74	Student papers	Universitas Islam Riau on 2026-02-24	<1%
75	Student papers	Universitas Sangga Buana YPKP on 2026-02-23	<1%
76	Internet	openjournal.unpam.ac.id	<1%
77	Internet	pt.scribd.com	<1%
78	Internet	repository.umsu.ac.id	<1%
79	Internet	repository.unj.ac.id	<1%
80	Internet	www.scribd.com	<1%
81	Publication	Asrul Hidayat, Nurwahid Syam, Sri Asfirawati Halik. "Pengembangan Sistem Infor...	<1%

82	Publication	Hafizha Nurul Qolby, Rangga Gelar Guntara, Syti Sarah Maesaroh. "Comparative...	<1%
83	Publication	Muhammad Reza Maulana, Agung Nugroho, Asep Suprianto. "Analisis Sentimen ...	<1%
84	Student papers	UIN Raden Intan Lampung on 2026-01-20	<1%
85	Student papers	UIN Sultan Syarif Kasim Riau on 2020-08-05	<1%
86	Student papers	Universitas Brawijaya on 2019-02-21	<1%
87	Student papers	Universitas Muslim Indonesia on 2025-08-29	<1%
88	Student papers	Universitas Negeri Medan on 2021-09-13	<1%
89	Student papers	Xavier University on 2024-07-15	<1%
90	Internet	etheses.uin-malang.ac.id	<1%
91	Internet	pdfcoffee.com	<1%
92	Internet	scholar.archive.org	<1%
93	Publication	Al Khaidar Author. "ANALISIS SENTIMEN DI INSTAGRAM TERHADAP MENTERI KEU...	<1%
94	Publication	Anjana Haya Atha Zharifa, Erik Iman Heri Ujjianto. "Analisis Sentimen Publik di Tw...	<1%
95	Publication	Danang Arbian Sulisty, Erik Setiadi. "Analisis Sentimen Kebijakan Makan Bergizi ...	<1%

96	Student papers	Fakultas Ekonomi Universitas Indonesia on 2015-07-02	<1%
97	Student papers	Lynn University on 2024-10-14	<1%
98	Publication	Ni Made Ayu Juli Astari, Dewa Gede Hendra Divayana, Gede Indrawan. "Analisis ...	<1%
99	Publication	Taufik Ramlan Alfiansyah, Audy Abdillah Hidayat, Alfarezi Hidayat Pratama, Agil ...	<1%
100	Student papers	UIN Raden Intan Lampung on 2025-11-26	<1%
101	Student papers	Universitas Amikom on 2022-01-10	<1%
102	Student papers	Universitas Brawijaya on 2017-01-04	<1%
103	Student papers	Universitas Brawijaya on 2017-01-24	<1%
104	Student papers	Universitas Brawijaya on 2018-01-03	<1%
105	Student papers	Universitas Brawijaya on 2023-09-01	<1%
106	Student papers	Universitas Islam Riau on 2025-11-03	<1%
107	Student papers	Universitas Muhammadiyah Sumatera Utara on 2025-08-21	<1%
108	Student papers	Universitas Muslim Indonesia on 2025-08-19	<1%
109	Student papers	Universitas Pamulang on 2025-10-21	<1%

110	Student papers	Universitas Putera Batam on 2026-02-02	<1%
111	Student papers	Universitas Putera Batam on 2026-02-03	<1%
112	Student papers	Universitas Sangga Buana YPKP on 2026-02-20	<1%
113	Internet	bennyapriandy.blogspot.com	<1%
114	Internet	core.ac.uk	<1%
115	Internet	dspace.uui.ac.id	<1%
116	Internet	edukasi.kompas.com	<1%
117	Internet	informatika.uin-suka.ac.id	<1%
118	Internet	journal.piksi.ac.id	<1%
119	Internet	kumparan.com	<1%
120	Internet	repository.itpln.ac.id	<1%
121	Internet	repository.uin-suska.ac.id	<1%
122	Publication	Gagas Ezhar Rahmayadi, Deny Haryadi, R. Bhima Danniswara, Valentino Kenny S...	<1%
123	Student papers	UPN Veteran Yogyakarta on 2025-10-29	<1%

124	Student papers	UPN Veteran Yogyakarta on 2026-03-02	<1%
125	Student papers	Universitas Brawijaya on 2019-05-17	<1%
126	Student papers	Universitas Muhammadiyah Prof. Dr. Hamka on 2026-04-07	<1%
127	Publication	AHMAD ABDILLAH HISYAM -. "ANALISIS SENTIMEN PERSEPSI PUBLIK TERHADAP ...	<1%
128	Publication	Abdulrahman Tuasamu, Rapanca Cahya Gumilang -, Mohammad Akmal Fachrian ...	<1%
129	Publication	Bayu Samodera, Kartini Kartini, Muhammad Muharrom Al Haromainy. "IMPLEME...	<1%
130	Publication	Doni Winarso, Yanda Noor Yudha, Syahril. "Analisis Sentimen Masyarakat Pada T...	<1%
131	Publication	O'neal Efrata Madao, Akhmad Irsyad, Muhammad Rivani Ibrahim. "ANALISIS SEN...	<1%
132	Student papers	STT PLN on 2023-07-26	<1%
133	Student papers	Universitas Brawijaya on 2019-07-08	<1%
134	Student papers	Universitas Budi Luhur on 2026-01-23	<1%
135	Student papers	Universitas Dian Nuswantoro on 2022-11-18	<1%
136	Student papers	Universitas Khairun on 2026-05-18	<1%
137	Student papers	Universitas Tarumanagara on 2025-12-23	<1%

138

Internet

journal.ipm2kpe.or.id

<1%

BAB I PENDAHULUAN

A. Latar Belakang

Pembangunan infrastruktur merupakan salah satu prioritas utama pemerintah Indonesia dalam memperkuat konektivitas wilayah dan meningkatkan pertumbuhan ekonomi nasional. Salah satu bentuk nyata dari upaya tersebut adalah pembangunan jalan tol yang berfungsi menghubungkan pusat-pusat ekonomi baru, mempercepat mobilitas masyarakat, dan menurunkan biaya logistik (Kementerian PUPR, 2023). Keberhasilan implementasi proyek strategis ini tidak hanya diukur dari penyelesaian fisik, tetapi juga dari tingkat penerimaan dan dukungan masyarakat yang terdampak secara langsung maupun tidak langsung.

Salah satu proyek strategis yang sedang dikembangkan di Jawa Timur adalah Pembangunan Jalan Tol Kediri–Tulungagung, yang berfungsi sebagai penghubung utama menuju Bandar Udara Internasional Dhoho Kediri serta memperkuat akses ke kawasan selatan Jawa Timur. Proyek ini diharapkan mampu meningkatkan aktivitas ekonomi wilayah dan mempercepat pemerataan pembangunan (Bappenas, 2024). Namun, di sisi lain, proyek tersebut juga menimbulkan beragam dinamika sosial seperti persoalan pembebasan lahan, potensi dampak lingkungan, serta persepsi masyarakat terhadap transparansi dan manfaat proyek.

Dalam era digital, media sosial telah menjadi ruang publik virtual di mana masyarakat dapat mengekspresikan pandangan dan perasaan mereka terhadap berbagai isu, termasuk kebijakan dan proyek pembangunan (Valenzuela dkk, 2023). Platform seperti Youtube tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga sebagai indikator opini publik yang dinamis dan real-time. Fenomena ini menjadikan media sosial sebagai sumber data yang sangat potensial untuk memahami persepsi dan sentimen masyarakat secara lebih luas (Hasan & Islam, 2023).

5 Beberapa penelitian sebelumnya telah menerapkan metode analisis sentimen untuk memahami opini publik melalui media sosial. Todingbua (2023) menggunakan pendekatan gabungan *Lexicon-Based* dan *Deep Learning* (LSTM) untuk menganalisis sentimen pengguna Twitter terkait perkembangan *Metaverse* di Indonesia. Hasilnya menunjukkan bahwa metode *Lexicon-Based* tepat dalam mengklasifikasikan opini publik di media sosial. Sementara itu, Dahur dan Albanna (2021) menerapkan metode *Lexicon-Based* untuk menilai sentimen masyarakat terhadap kebijakan New Normal di Indonesia dan menemukan bahwa opini publik didominasi oleh sentimen negatif.

7 Pendekatan *Lexicon-Based*, yaitu metode analisis sentimen yang mengidentifikasi kecenderungan opini masyarakat (positif, negatif, atau netral) berdasarkan kata-kata yang digunakan dalam teks. Pendekatan ini dipilih karena mampu menganalisis data dalam jumlah besar tanpa memerlukan proses pelatihan model yang kompleks, sehingga dinilai sesuai untuk menggambarkan persepsi publik terhadap proyek pembangunan jalan tol.

13 Dengan demikian, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap proyek pembangunan Jalan Tol Kediri–Tulungagung menggunakan data dari media sosial dengan metode *Lexicon-Based*. Hasil penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai persepsi publik terhadap proyek tersebut serta menjadi bahan pertimbangan bagi pemerintah dan pelaksana proyek dalam merumuskan strategi komunikasi publik yang lebih efektif dan responsif terhadap aspirasi masyarakat.

35 B. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, dapat diidentifikasi masalah sebagai berikut:

1. Belum adanya penerapan metode *Lexicon-Based Sentiment Analysis* yang secara khusus digunakan untuk menganalisis opini masyarakat di media sosial terkait proyek pembangunan Jalan Tol Kediri–

Tulungagung, sehingga kecenderungan sentimen positif, negatif, dan netral belum dapat diketahui secara jelas.

2. Belum tersedianya sistem analisis sentimen yang terstruktur, termasuk rancangan arsitektur sistem, untuk mengolah dan memetakan opini masyarakat dari media sosial secara otomatis.

C. Rumusan Masalah

Berdasarkan latar belakang dan batasan masalah yang telah diuraikan sebelumnya, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan metode *Lexicon-Based Sentiment Analysis* dalam menganalisis opini masyarakat terhadap proyek tersebut?
2. Bagaimana merancang arsitektur sistem analisis sentimen berbasis *Lexicon-Based* untuk memetakan opini masyarakat terhadap proyek pembangunan Jalan Tol Kediri–Tulungagung?

D. Batasan Masalah

Berdasarkan Latar Belakang dan Identifikasi Masalah terdapat batasan masalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini hanya berasal dari platform media sosial Youtube.
2. Pengambilan data (proses *crawling*) dilakukan pada komentar yang tersedia di YouTube hingga bulan Desember 2025.
3. Analisis sentimen hanya akan dilakukan pada data teks yang menggunakan Bahasa Indonesia.
4. Penelitian ini secara spesifik menggunakan pendekatan *Lexicon-Based* untuk klasifikasi sentiment.
5. Fokus utama analisis adalah sentimen masyarakat yang secara eksplisit membahas proyek pembangunan Jalan Tol Kediri-Tulungagung.
6. Penelitian ini tidak akan melakukan analisis mendalam terhadap latar belakang demografis (seperti usia, jenis kelamin, lokasi, atau pekerjaan) dari akun pengguna media sosial yang mengeluarkan opini.

E. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan sebelumnya, maka tujuan yang ingin dicapai melalui penelitian ini adalah sebagai berikut:

1. Untuk menerapkan metode *Lexicon-Based Sentiment Analysis* dalam menganalisis opini masyarakat terhadap proyek pembangunan Jalan Tol Kediri–Tulungagung.
2. Untuk merancang arsitektur sistem analisis sentimen berbasis *Lexicon-Based* guna memetakan opini masyarakat terhadap proyek pembangunan Jalan Tol Kediri–Tulungagung.

F. Manfaat Penelitian

Berdasarkan Latar Belakang terdapat manfaat dan kegunaan penelitian, seagai berikut:

a. Manfaat Teoritis

Secara teoritis, penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang analisis sentimen dan pengolahan bahasa alami (*Natural Language Processing*) untuk Bahasa Indonesia. Penelitian ini juga memperkaya pemahaman tentang peran media sosial sebagai ruang publik digital serta menawarkan alternatif metodologis dalam mengkaji persepsi publik terhadap kebijakan atau proyek pemerintah.

b. Manfaat Praktis

Secara praktis, hasil penelitian ini dapat menjadi bahan pertimbangan bagi pemerintah dan pelaksana proyek dalam memahami opini publik serta meningkatkan strategi komunikasi dan mitigasi risiko sosial. Selain itu, penelitian ini dapat memberikan masukan berbasis data bagi masyarakat dan organisasi sipil untuk menyampaikan aspirasi secara konstruktif. Bagi akademisi, penelitian ini dapat dijadikan referensi atau acuan untuk penelitian selanjutnya dalam topik serupa.

BAB II

TINJAUAN PUSTAKA

A. Teori dan Penelitian Terdahulu

1. Landasan Teori

a. Analisis Sentimen

Analisis Sentimen atau opinion mining merupakan bagian dari *Natural Language Processing* (NLP) yang bertujuan untuk mengenali opini, emosi, dan sikap masyarakat terhadap suatu isu berdasarkan teks digital (Hasan & Islam, 2023). Metode ini digunakan untuk mengklasifikasikan opini menjadi kategori positif, negatif, atau netral. Penerapan analisis sentimen banyak digunakan dalam bidang bisnis, politik, dan sosial untuk memahami persepsi publik. Dengan pendekatan ini, data tidak terstruktur dapat diubah menjadi informasi yang terukur dan bermanfaat (Yadav & Kumar, 2022).

Analisis sentimen dapat dilakukan pada tingkat dokumen, kalimat, atau aspek tertentu dari objek yang diteliti (Rizvi et al., 2023). Analisis tingkat dokumen menilai keseluruhan teks, sedangkan tingkat aspek berfokus pada fitur tertentu. Proses ini memungkinkan peneliti mengetahui bagian mana yang menimbulkan opini positif atau negatif. Dengan demikian, analisis sentimen dapat memberikan wawasan yang lebih mendalam mengenai pandangan masyarakat.

Terdapat dua pendekatan utama dalam analisis sentimen, yaitu *machine learning* dan *lexicon-based*. Pendekatan *machine learning* menghasilkan akurasi tinggi, tetapi memerlukan data latih berlabel dalam jumlah besar (Khan et al., 2023). Sementara itu, pendekatan *lexicon-based* menggunakan kamus kata yang telah diberi nilai polaritas positif atau negatif tanpa pelatihan data (Shah & Patel, 2022). Pendekatan ini lebih efisien untuk data besar seperti media sosial.

Namun, analisis sentimen juga memiliki tantangan seperti ambiguitas kata dan kesulitan mendeteksi sarkasme. Dalam konteks Bahasa Indonesia, penggunaan bahasa tidak formal dan singkatan menjadi kendala tersendiri. Kondisi ini memerlukan penyesuaian metode dan leksikon agar hasilnya tetap akurat. Oleh karena itu, pengembangan leksikon lokal sangat penting untuk meningkatkan performa analisis (Wijaya & Panjaitan, 2024).

b. Pendekatan Berbasis Leksikon (Lexicon-Based Approach)

Pendekatan berbasis leksikon merupakan metode analisis sentimen yang mengandalkan *sentiment lexicon*, yaitu kamus kata yang telah diberi skor polaritas positif, negatif, atau netral (Birjali et al., 2021). Setiap kata dalam teks dicocokkan dengan daftar kata yang ada dalam leksikon. Skor sentimen dari setiap kata kemudian dijumlahkan untuk menentukan kecenderungan keseluruhan teks. Dengan cara ini, sistem dapat mengukur arah opini publik terhadap suatu isu (García-Méndez et al., 2022).

Metode ini memiliki keunggulan dalam efisiensi dan transparansi hasil analisis. Tidak seperti *machine learning*, pendekatan ini tidak memerlukan proses pelatihan yang kompleks. Hasilnya dapat langsung dijelaskan dengan melihat kata yang menyumbang nilai sentimen tertentu. Karena itu, metode ini ideal untuk data berskala besar atau ketika data latih tidak tersedia (Haddi et al., 2021).

Kelemahan utama pendekatan ini adalah ketidakmampuannya memahami konteks, sarkasme, atau makna ganda suatu kata. Misalnya, kata “dingin” bisa positif dalam konteks pendingin ruangan, namun negatif dalam konteks makanan. Untuk mengatasi hal ini, peneliti biasanya menambahkan aturan linguistik seperti valence shifters yang mengubah polaritas kata. Aturan seperti “tidak baik” atau “sangat

bagus” membantu meningkatkan ketepatan hasil analisis (Haddi et al., 2021).

Dalam konteks Bahasa Indonesia, penggunaan leksikon lokal seperti *InSet Lexicon* menjadi penting untuk meningkatkan akurasi hasil (Putra & Wibawa, 2023). Bahasa Indonesia di media sosial memiliki banyak kata tidak baku dan bahasa gaul yang sulit dipahami oleh leksikon umum. Pendekatan berbasis leksikon menjadi solusi yang efisien karena cepat dan mudah diterapkan. Oleh sebab itu, penelitian ini menggunakan pendekatan ini untuk menganalisis opini publik terhadap proyek Jalan Tol Kediri–Tulungagung (Nugroho & Bachtiar, 2024).

c. Media Sosial

Media sosial merupakan platform digital yang memungkinkan pengguna berinteraksi, membuat, dan membagikan informasi secara real-time (Carr & Hayes, 2021). Ciri khas media sosial adalah sifatnya yang interaktif, cepat, dan terbuka bagi siapa saja. Media sosial kini menjadi ruang publik digital yang memfasilitasi masyarakat untuk mengekspresikan pandangan mereka terhadap berbagai isu sosial (Fuchs, 2022). Media sosial sering digunakan sebagai sumber data untuk memahami opini publik.

YouTube sebagai salah satu platform media sosial memiliki peran besar dalam membentuk opini publik. Pengguna dapat memberikan respons melalui komentar, likes, serta berbagai bentuk interaksi lain yang mencerminkan pandangan mereka terhadap isu tertentu (Valenzuela et al., 2023). Hal ini menjadikan YouTube sebagai sumber data yang relevan untuk penelitian yang memanfaatkan analisis sentimen berbasis teks.

YouTube memiliki fitur komentar yang bersifat terbuka, sehingga memudahkan peneliti dalam mengumpulkan data secara sistematis. Struktur komentar yang terorganisir mulai dari

komentar utama hingga balasan mendukung proses analisis sentimen karena memberikan konteks diskusi yang lebih jelas. Dengan demikian, platform ini dapat digunakan untuk menelusuri kecenderungan opini publik secara terukur.

Melalui data komentar YouTube, penelitian ini dapat menganalisis opini masyarakat secara *real-time* menggunakan pendekatan *Lexicon-Based*. Pendekatan ini memungkinkan pemetaan sentimen secara cepat dan objektif, sehingga hasilnya dapat memberikan wawasan yang lebih mendalam mengenai persepsi publik terhadap suatu isu tertentu (Zhang & Chen, 2023).

2. Kajian Pustaka

Guna mengkaji secara mendalam pada penelitian maka peneliti mengacu pada penelitian terdahulu sebagai berikut :

Penelitian dengan judul “Perbandingan *Algoritma Naive Bayes* dan *Logistic Regression* untuk Analisis Sentimen Pilgub Jatim 2024” yang dilakukan oleh Fajar Wahyuardha Putra, dkk (2024) menggunakan pendekatan klasifikasi *supervised* dengan *algoritma Naive Bayes* dan *Logistic Regression*. Tahapan penelitian meliputi prapemrosesan data berupa pembersihan teks, tokenisasi, *stopword removal*, dan *stemming*, kemudian dilakukan evaluasi menggunakan metrik akurasi, *F1-score*, serta analisis efisiensi waktu. Hasil penelitian menunjukkan bahwa *Logistic Regression* menghasilkan akurasi sebesar 87%, sedikit lebih unggul dibandingkan *Naive Bayes* yang memperoleh akurasi 85%, serta memiliki nilai *F1-score* yang lebih tinggi. Persamaan penelitian ini dengan penelitian yang penulis susun terletak pada fokus kajian, yaitu sama-sama menganalisis sentimen opini masyarakat di media sosial. Adapun perbedaannya terletak pada metode analisis sentimen yang digunakan, di mana penelitian tersebut menerapkan *algoritma supervised learning*, sedangkan penelitian yang penulis susun menggunakan pendekatan *lexicon-based*.

Pada penelitian yang berjudul “*Evaluasi Lexicon-Based dan Deep Learning pada Analisis Sentimen Pengguna Twitter Terkait Perkembangan Metaverse di Indonesia*” yang dilakukan oleh Rendy Juniarta Todingbua (2023) menggunakan dua pendekatan, yaitu *Lexicon-Based* dan model *deep learning* LSTM. Data penelitian diperoleh dari tweet pengguna Twitter dengan kata kunci yang berkaitan dengan *metaverse*. Hasil penelitian menunjukkan bahwa model LSTM memiliki performa terbaik dengan tingkat akurasi mencapai 93%, sedangkan pendekatan gabungan LSTM dan *Lexicon-Based* menghasilkan akurasi sebesar 91%. Persamaan penelitian ini dengan penelitian yang penulis susun terletak pada fokus penelitian, yaitu analisis sentimen masyarakat di media sosial. Perbedaanannya terletak pada metode yang digunakan, karena penelitian tersebut mengombinasikan *lexicon-based* dengan *deep learning*, sementara penelitian yang penulis susun hanya menggunakan pendekatan *lexicon-based*.

Penelitian yang dilakukan oleh Muhammad Akmal Fauzan (2022) dengan judul “*Penerapan Sentimen Leksikon Indonesia pada Analisis Sentimen Mengenai Opini Masyarakat di Twitter Terhadap Kebijakan PPKM Darurat Menggunakan Algoritma Maximum Entropy*” menggunakan algoritma *Maximum Entropy* sebagai metode klasifikasi. Penelitian ini membandingkan dua jenis dataset, yaitu dataset dengan pelabelan manual dan dataset dengan pelabelan otomatis menggunakan *InSet Lexicon*. Hasil penelitian menunjukkan bahwa model dengan pelabelan manual memperoleh akurasi sebesar 86,2%, sedangkan model dengan pelabelan otomatis memperoleh akurasi sebesar 71,1%. Persamaan penelitian ini dengan penelitian yang penulis susun terletak pada fokus kajian, yaitu analisis sentimen masyarakat terhadap kebijakan atau proyek strategis pemerintah di media sosial. Perbedaanannya terletak pada metode analisis yang digunakan, di mana penelitian tersebut menggunakan algoritma *Maximum Entropy*,

sedangkan penelitian yang penulis susun menggunakan pendekatan *lexicon-based*.

Penelitian berjudul “*Analisis Sentimen Masyarakat pada Media Sosial Twitter Terhadap Penerapan New Normal di Indonesia Menggunakan Metode Lexicon Based*” yang dilakukan oleh Arnoldus Janssen Dahur dan Isa Albanna (2021) menerapkan metode *Lexicon-Based* dalam mengklasifikasikan sentimen. Dari total 1.338 tweet yang dianalisis, hasil penelitian menunjukkan bahwa terdapat 563 tweet dengan sentimen positif, 751 tweet dengan sentimen negatif, dan 24 tweet dengan sentimen netral, yang setara dengan 42,1% sentimen positif, 56,1% sentimen negatif, dan 1,8% sentimen netral. Persamaan penelitian ini dengan penelitian yang penulis susun terletak pada fokus penelitian, yaitu analisis sentimen masyarakat terhadap kebijakan atau proyek strategis pemerintah di media sosial. Perbedaannya terletak pada konteks kajian, di mana penelitian tersebut membahas penerapan kebijakan *New Normal*, sedangkan penelitian yang penulis susun membahas proyek pembangunan jalan tol.

Penelitian yang dilakukan oleh Satria Adi Nugraha (2025) dengan judul “*Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara*” menggunakan metode *Lexicon-Based Sentiment Analysis* untuk mengklasifikasikan opini publik. Hasil analisis menunjukkan bahwa sentimen masyarakat terhadap Danantara cenderung beragam dengan dominasi sentimen negatif sebesar 44,9%, sentimen positif sebesar 44,1%, dan sisanya merupakan sentimen netral sebesar 14%. Persamaan penelitian ini dengan penelitian yang penulis susun terletak pada fokus kajian, yaitu analisis sentimen masyarakat di media sosial terhadap isu atau proyek strategis. Perbedaannya terletak pada konteks penelitian, di mana penelitian tersebut membahas fenomena Danantara, sedangkan penelitian yang penulis susun mengkaji sentimen masyarakat terhadap proyek pembangunan jalan tol.

B. Kerangka Berpikir



Gambar 2.1 Kerangka Berpikir

31

BAB III METODE PENELITIAN

A. Desain Penelitian

1. Jenis Penelitian

Penelitian ini merupakan penelitian yang termasuk dalam jenis Desain Pengembangan (*Development Research*). Penelitian ini bertujuan mengembangkan dan merancang sebuah sistem analisis sentimen menggunakan metode *Lexicon Based* atau pendekatan berbasis kamus leksikon, di mana skor sentimen dihitung berdasarkan bobot kata-kata positif dan negatif yang terdapat dalam kamus leksikon yang telah ditetapkan.

10

46

77

2. Variabel Penelitian

Dalam penelitian ini, variabel independen (bebas) adalah teks opini masyarakat yang berasal dari media sosial yang telah melalui proses crawling sesuai dengan kata kunci terkait proyek pembangunan Jalan Tol Kediri–Tulungagung. Variabel ini mencakup berbagai bentuk ekspresi bahasa pengguna seperti kalimat formal, bahasa tidak baku, singkatan, hingga campur kode, yang menjadi input utama dalam proses analisis sentimen. Selain itu, lexicon sentimen yang digunakan dalam pendekatan *Lexicon-Based* juga termasuk bagian dari variabel independen karena memengaruhi hasil klasifikasi melalui nilai polaritas yang diberikan pada setiap kata.

30

10

7

Variabel dependen (terikat) adalah hasil klasifikasi sentimen berupa kategori positif, negatif, atau netral yang dihasilkan dari proses analisis menggunakan metode *Lexicon-Based*. Hasil ini dipengaruhi langsung oleh variasi data teks dan struktur lexicon yang digunakan dalam proses pengolahan. Variabel kontrol dalam penelitian ini meliputi periode waktu pengambilan data, bahasa teks (dibatasi pada Bahasa Indonesia), platform media sosial yang digunakan, serta lexicon yang diterapkan agar proses pengukuran tetap konsisten.

3. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini dikumpulkan melalui proses *web crawling* atau *scraping* dari beberapa platform, yaitu Youtube. Pengumpulan data dilakukan dengan menggunakan kata kunci atau hashtag yang relevan dengan proyek pembangunan Jalan Tol Kediri–Tulungagung. Setiap unggahan yang diambil akan melalui proses seleksi awal untuk memastikan bahwa konten tersebut benar-benar terkait dengan topik penelitian. Selain itu, penelitian ini membatasi waktu pengambilan data pada rentang tertentu agar data yang diperoleh tetap relevan dengan kondisi terbaru. Semua data yang terkumpul disimpan dalam format teks dan kemudian melalui tahap cleaning sebelum dianalisis menggunakan pendekatan *Lexicon-Based Sentiment Analysis*.

B. Instrumen Penelitian

Untuk mendukung jalanya penelitian ini, peneliti menggunakan perangkat instrumen sebagai berikut ini:

1. Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan dalam penelitian ini adalah seperangkat komputer pribadi (PC) atau laptop dengan spesifikasi yang memadai untuk menjalankan proses *Natural Language Processing* (NLP), dengan metode *Lexicon Based*.

2. Perangkat Lunak (*Software*)

Pada penelitian ini perangkat lunak yang digunakan adalah sistem operasi yang digunakan dapat berupa Windows, untuk mendukung instalasi library *Natural Language Processing* (NLP). Bahasa pemrograman yang digunakan adalah *Python* karena memiliki ekosistem library yang kaya dan stabil untuk kebutuhan pengolahan teks, sehingga memudahkan implementasi analisis sentimen *Lexicon-Based*.

Selain itu, tools pendukung juga diperlukan untuk menjalankan dan mengembangkan seluruh proses analisis. *Jupyter Notebook* atau *Google Colab* digunakan sebagai lingkungan kerja utama untuk

mengeksekusi kode, karena mendukung eksperimen interaktif dan dokumentasi proses secara sistematis. Untuk proses *crawling data* media sosial, digunakan API atau *scraper tools* yang sesuai dengan platform yang menyediakan data. Sementara itu, *Matplotlib* atau *Seaborn* dimanfaatkan untuk membuat visualisasi grafik hasil analisis sentimen. Editor kode seperti *Visual Studio Code* (VSCoDe) digunakan untuk pengembangan skrip yang lebih kompleks, memastikan proses analisis berjalan lancar.

3. Dataset

Dataset dalam penelitian ini berupa kumpulan teks (posting atau komentar) dari platform Youtube yang mengandung opini masyarakat terkait proyek Jalan Tol Kediri–Tulungagung. Data diperoleh melalui proses web *scrapping/crawling* dan menjadi data mentah yang diolah dengan kata kunci tertentu seperti "Tol Kediri", "Tol Kediri–Tulungagung", dan frasa lain. Dataset ini berisi data tidak terstruktur (*unstructured text*) yang kemudian diolah dan dibersihkan agar dapat diproses oleh sistem analisis sentimen.

4. Analisis Hasil

a. Pengujian Fungsional

Pengujian fungsional bertujuan memastikan setiap modul atau fungsi dalam sistem berjalan sesuai rancangan, terutama pada tahap *text preprocessing* (seperti *case folding*, *tokenizing*, dan *stopword removal*) dan modul perhitungan skor polaritas berdasarkan kamus leksikon, untuk memastikan hasil skor sentimen (positif/negatif/netral) dihitung dengan benar pada setiap teks.

b. Pengujian Non-Fungsional

Pengujian non-fungsional berfokus pada efisiensi waktu pemrosesan data (*time efficiency*) dan skalabilitas sistem dalam mengolah sejumlah besar data (*dataset*) dari media sosial.

Pengujian ini memastikan bahwa alat yang dibangun dapat melakukan analisis sentimen secara cepat dan stabil meskipun volume data yang dianalisis sangat besar.

C. Tempat dan Jadwal Penelitian

1. Tempat Penelitian

Tempat utama dalam penelitian ini adalah lingkungan virtual dari platform Youtube yang menjadi sumber data opini masyarakat. Penelitian ini menggunakan proses pengambilan data dengan *crawling* atau *scraping* yang dilakukan secara daring. Sementara itu, untuk proses komputasi, pengembangan sistem *Lexicon Based*, dan analisis data akan dilaksanakan di Laboratorium Komputasi atau Pusat Penelitian di institusi peneliti.

2. Jadwal Penelitian

Jadwal penelitian ini berlangsung selama 6 bulan, yang secara detail dijelaskan dalam tabel jadwal penelitian berikut.

Tabel 3.1 Jadwal Penelitian

Jadwal Penelitian	Bulan Ke																							
	1				2				3				4				5				6			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Analisis Kebutuhan																								
Pengumpulan Data																								
Perancangan Sistem																								
Pembuatan Program Sistem																								
Pengujian Sistem																								
Implementasi dan																								

Jadwal Penelitian	Bulan Ke																							
	1				2				3				4				5				6			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Evaluasi Sistem																								
Laporan																								

D. Objek Penelitian/Subjek Penelitian

1. Analisis Kebutuhan Sistem

Analisis kebutuhan sistem pada penelitian ini adalah:

a. Kebutuhan Fungsional

Sistem mampu melakukan proses *crawling* atau pengambilan data teks dari media sosial menggunakan kata kunci tertentu terkait proyek Jalan Tol Kediri–Tulungagung. Data yang diperoleh kemudian harus melalui tahapan *preprocessing*. Sistem juga dapat menjalankan analisis sentimen menggunakan metode *Lexicon-Based* untuk mengklasifikasikan teks ke dalam kategori positif, negatif, atau netral. Hasil klasifikasi ditampilkan dalam bentuk angka maupun visual agar mudah dipahami pengguna, dan sistem juga harus menyediakan fitur penyimpanan serta ekspor hasil analisis untuk keperluan dokumentasi atau pelaporan.

b. Kebutuhan NonFungsional

Sistem memiliki kecepatan pemrosesan yang baik meskipun menangani data dalam jumlah besar, sehingga analisis tetap berjalan. Stabilitas sistem perlu dijaga agar tidak mudah mengalami error saat mengolah data teks yang sifatnya tidak terstruktur dan bervariasi. Antarmuka pengguna juga dirancang secara sederhana dan intuitif sehingga mudah dipahami bahkan oleh pengguna yang tidak memiliki latar belakang teknis mendalam. Keamanan data turut menjadi pertimbangan penting, terutama pada proses pengambilan data dari media sosial untuk memastikan sistem tetap mematuhi etika dan menjaga privasi.

c. Kebutuhan Data

Data yang dikumpulkan berupa teks mentah (*unstructured text*) yang diperoleh dari platform Youtube dengan menggunakan kata kunci tertentu. Data harus berada dalam rentang waktu yang telah ditentukan agar analisis dapat menggambarkan kondisi opini publik secara akurat pada periode tersebut. Jumlah data yang digunakan perlu cukup besar dengan idealnya ribuan karena teks media sosial cenderung singkat dan sangat bervariasi, sehingga diperlukan volume yang besar untuk menghasilkan analisis sentimen yang stabil dan representatif.

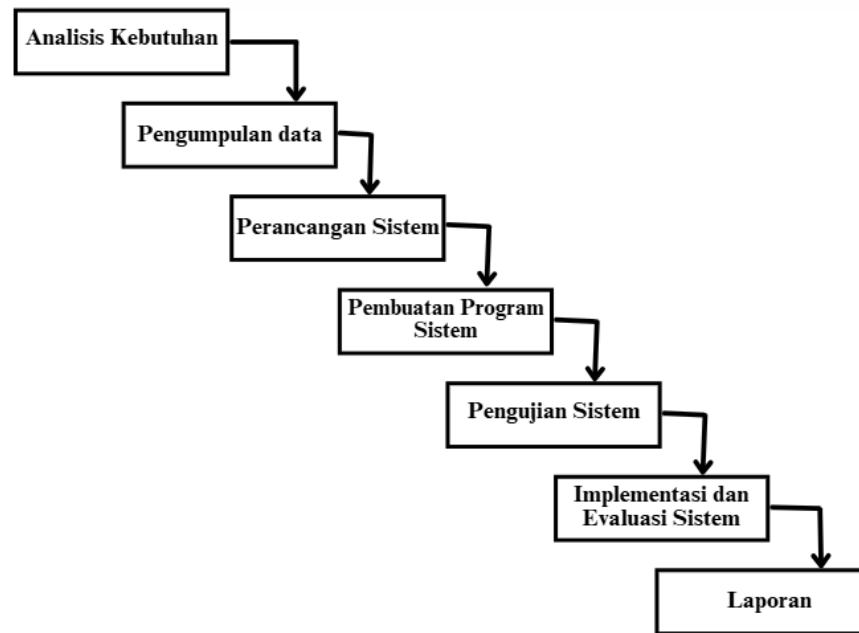
2. Objek Penelitian

Objek penelitian ini adalah data tekstual yang berupa opini, komentar dan postingan publik di media sosial yang secara spesifik membahas dan mengandung kata kunci terkait 'Proyek Pembangunan Jalan Tol Kediri'. Objek ini merupakan data mentah yang akan diproses. Objek penelitian secara konseptual adalah skor polaritas sentimen yang dihasilkan dari pengolahan data tekstual tersebut menggunakan metode *Lexicon Based*, di mana skor ini akan diklasifikasikan menjadi sentimen positif, negatif, atau netral, untuk kemudian disajikan melalui *website* yang dikembangkan.

3. Subjek Penelitian

Subjek penelitian ini adalah masyarakat pengguna media sosial (termasuk individu, kelompok, atau akun publik) yang telah mempublikasikan opini atau tanggapan terkait proyek pembangunan jalan tol Kediri di platform Youtube dalam periode waktu pengambilan data.

E. Prosedur Penelitian



Gambar 3.1 Prosedur Penelitian

Prosedur yang digunakan adalah waterfall dan prosedurnya pada gambar 3.1:

1. Analisis Kebutuhan

Pada tahap ini dilakukan proses identifikasi kebutuhan sistem, meliputi kebutuhan fungsional, non-fungsional, dan kebutuhan data. Analisis dilakukan dengan mengkaji masalah utama penelitian, menentukan tujuan sistem, menetapkan fitur wajib, serta merumuskan metode yang akan digunakan untuk memproses data sentimen dari media sosial. Tahap ini memastikan bahwa sistem yang akan dibangun relevan dengan permasalahan penelitian.

2. Pengumpulan Data

Pengumpulan Data dilakukan melalui teknik *crawling* (penambangan data) pada platform Youtube dengan menggunakan *Application Programming Interface* (API) atau teknik *scraping* yang bertujuan untuk mendapatkan teks opini publik terkait Proyek Pembangunan Jalan Tol Kediri. Data mentah yang berhasil

17 dikumpulkan kemudian melalui tahapan penting berupa *text preprocessing*, yang mencakup *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, untuk memastikan data siap diolah oleh algoritma *Lexicon Based*.

102 3. Perancangan Sistem

112 Pada tahap ini meliputi perancangan arsitektur sistem (*system architecture*), desain basis data (*Database Design*) untuk penyimpanan data mentah dan hasil sentimen, serta perancangan antarmuka pengguna (*User Interface/User Experience - UI/UX*) *website*.
118 Perancangan ini penting untuk memastikan bahwa sistem yang dibangun nantinya dapat menyajikan informasi analisis sentimen secara efektif dan mudah dipahami oleh pengguna.
63

4. Pembuatan Program Sistem

Pada tahap Pembuatan Program Sistem, dilakukan dengan mengembangkan sistem menggunakan bahasa Python. Pengerjaan meliputi pembuatan modul *crawling*, *preprocessing teks*, proses *lexicon-based*, perhitungan skor polaritas, klasifikasi sentimen, penyimpanan data, serta dashboard visualisasi. Merancang *frontend website* sebagai antarmuka pengguna untuk menampilkan *dashboard* dan visualisasi sentimen secara interaktif. Proses ini memastikan bahwa semua fungsionalitas yang dirancang dapat berjalan sesuai spesifikasi.

5. Pengujian Sistem

Pada pengujian sistem dibagi menjadi dua jenis utama: Pengujian Fungsional (memastikan setiap modul *crawling*, *preprocessing*, dan klasifikasi bekerja dengan benar) dan Pengujian Non-Fungsional (mengukur Akurasi, Presisi, dan *Recall* model *Lexicon Based* dengan membandingkan hasilnya dengan data *gold standard*). Jika ditemukan ketidaksesuaian atau *bug*, sistem akan diperbaiki dan diuji ulang (iterasi) hingga mencapai kinerja yang dapat diterima.

6. Implementasi dan Evaluasi Sistem

Setelah sistem lulus pengujian, dilakukan Implementasi dan Evaluasi Sistem. Implementasi mencakup penempatan *website* pada lingkungan server (*deployment*) agar dapat diakses oleh pengguna. Evaluasi sistem dilakukan untuk menilai kelayakan dan manfaat *website* dalam menyajikan informasi sentimen publik secara nyata.

7. Penyusunan Laporan

Tahap terakhir adalah Penyusunan Laporan, yaitu dokumentasi tertulis dari seluruh proses penelitian, mulai dari studi literatur, analisis kebutuhan, pengembangan sistem, hingga hasil pengujian dan evaluasi. Laporan mencakup pembahasan mendalam mengenai hasil analisis sentimen masyarakat terhadap Proyek Pembangunan Jalan Tol Kediri yang didukung oleh data visualisasi dari *website* yang telah dikembangkan.

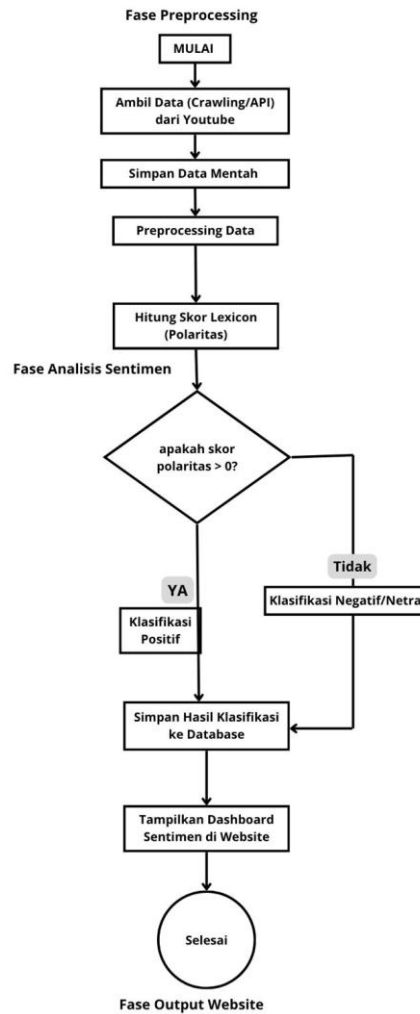
F. Teknik Analisis Data

1. Desain Sistem

a. Arsitektur Sistem

1). Flowchart

Alur program *website* analisis sentimen pada gambar 3.2 dimulai dengan Fase *Preprocessing*, yang diawali ketika sistem dieksekusi (MULAI). Program kemudian menjalankan proses otomatis Ambil Data (*Crawling/API*) dari Media Sosial untuk mengumpulkan teks opini terkait Proyek Tol Kediri.



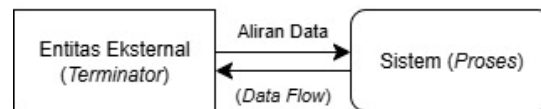
Gambar 3.2 Diagram Alur Program

Data mentah yang terkumpul selanjutnya Simpan Data Mentah ke *Database*, sebelum dilakukan Preprocessing Data (pembersihan, *tokenizing*, dan *stemming*). Setelah data siap, alur berpindah ke Fase Analisis Sentimen, di mana sistem akan Hitung Skor *Lexicon* dengan mencocokkan setiap kata dengan kamus leksikon untuk mendapatkan total nilai Polaritas. Skor polaritas ini menjadi dasar keputusan: jika Skor Polaritas > 0 (Ya), teks dikelompokkan sebagai Klasifikasi Positif; jika Tidak (Skor ≤ 0), teks dikelompokkan sebagai Klasifikasi Negatif/*Netral*. Hasil klasifikasi sentimen ini kemudian memasuki Fase *Output Website* dengan perintah Simpan

Hasil Klasifikasi ke *Database* final. Tahap terakhir, sistem akan Tampilkan *Dashboard* Sentimen ke *Website*, menyajikan visualisasi dan ringkasan sentimen kepada pengguna, sebelum program mencapai titik SELESAI.

2). Data Flow Diagram

a) Data Flow Diagram Level 0

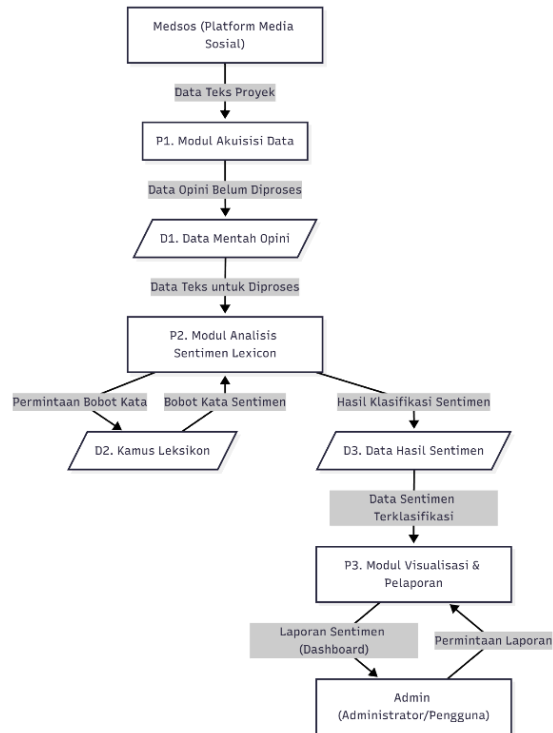


Gambar 3.3 DFD Level 0

Gambar 3.3 menampilkan Sistem Analisis Sentimen Proyek Jalan Tol Kediri sebagai satu proses sentral yang terhubung dengan tiga entitas eksternal. Platform Media Sosial (seperti Youtube) bertindak sebagai sumber data, menyediakan aliran Data Teks Proyek Tol Kediri kepada sistem. Entitas kedua, Administrator/Pengguna *Website*, mengirimkan Permintaan Laporan/Akses kepada sistem. Sebagai timbal balik, sistem menyediakan aliran data Laporan Sentimen (Visualisasi), yang berisi *dashboard* dan ringkasan hasil analisis sentimen, kembali kepada Administrator/Pengguna.

b) Data Flow Diagram Level 1

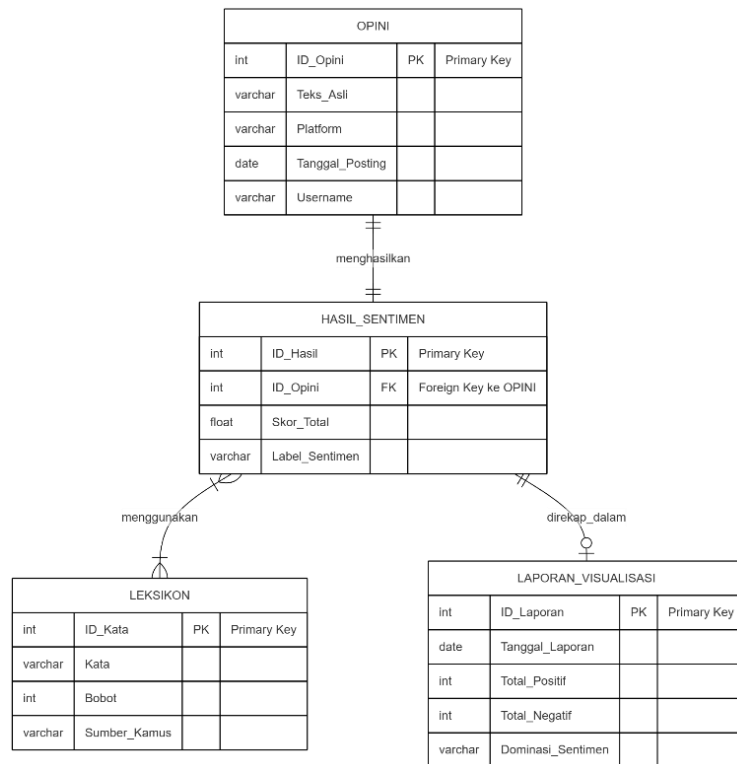
Berdasarkan Gambar 3.4 DFD Level 1 membagi sistem menjadi tiga modul proses utama, dimulai dengan P1. Modul Akuisisi Data. Modul ini menerima Data Teks Proyek Tol Kediri dari entitas eksternal Medsos (Platform Media Sosial), kemudian memproses data tersebut sebelum menyimpannya sebagai Data Opini Belum Diproses di D1.



Gambar 3.4 DFD Level 1

Data Mentah Opini. Selanjutnya, data mengalir dari D1 sebagai Data Teks untuk Diproses menuju P2. Modul Analisis Sentimen *Lexicon*. Proses P2 ini bersifat sentral, di mana ia mengakses D2. Kamus Leksikon dengan mengirimkan Permintaan Bobot Kata dan menerima Bobot Kata Sentimen sebagai *input* algoritmanya. Setelah klasifikasi selesai, P2 menyimpan hasilnya sebagai Hasil Klasifikasi Sentimen ke D3. Data Hasil Sentimen. Alur terakhir dikendalikan oleh P3. Modul Visualisasi & Pelaporan, yang menerima Permintaan Laporan dari Admin (Administrator/Pengguna). P3 kemudian mengakses D3 untuk mengambil Data Sentimen Terklasifikasi dan menyampaikannya kembali kepada Admin dalam bentuk Laporan Sentimen (*Dashboard*).

3). Entity Relation Diagram



Gambar 3.5 Entity Relationship Diagram

Diagram ERD sistem analisis sentimen ini terdiri dari empat entitas utama yang merepresentasikan struktur data sistem. Entitas sentralnya adalah OPINI, yang menyimpan data teks mentah dari media sosial, diidentifikasi secara unik oleh Primary Key (ID_Opini), serta memiliki atribut seperti Teks_Aslil, Platform, dan Username.

Entitas HASIL_SENTIMEN menyimpan hasil klasifikasi yang diidentifikasi oleh ID_Hasil. Entitas ini terhubung ke OPINI melalui relasi "menghasilkan" dengan ID_Opini sebagai Foreign Key, memastikan bahwa setiap hasil sentimen merujuk pada satu opini sumber. Entitas LEKSIKON menyimpan kamus kata yang digunakan algoritma Lexicon Based, dengan ID_Kata sebagai Primary Key dan atribut penting seperti Bobot (nilai polaritas) dan Sumber_Kamus.

Relasi antara HASIL_SENTIMEN dan LEKSIKON adalah "menggunakan", yang menunjukkan bahwa proses menghasilkan satu hasil sentimen melibatkan penggunaan banyak kata dari kamus leksikon. Terakhir, entitas LAPORAN_VISUALISASI (yang diidentifikasi oleh ID_Laporan) berfungsi untuk merekap data agregat, seperti Total_Positif dan Total_Negatif. Entitas ini terhubung ke HASIL_SENTIMEN melalui relasi "direkap_dalam", menunjukkan bahwa banyak hasil sentimen individual digabungkan untuk menghasilkan satu data ringkasan laporan berdasarkan periode waktu tertentu.

b. Desain Antarmuka Pengguna (*UI/UX Design*)

Pada penelitian ini menggunakan desain menu website, yaitu:

1) Halaman Utama

The diagram illustrates the layout of the Main Page (Halaman Utama). It features a central content area with the following components:

- A title box labeled **JUDUL**.
- An information box labeled **INFORMASI**.
- A text input field labeled **Link Youtube**.
- A button labeled **Mulai Analisis**.

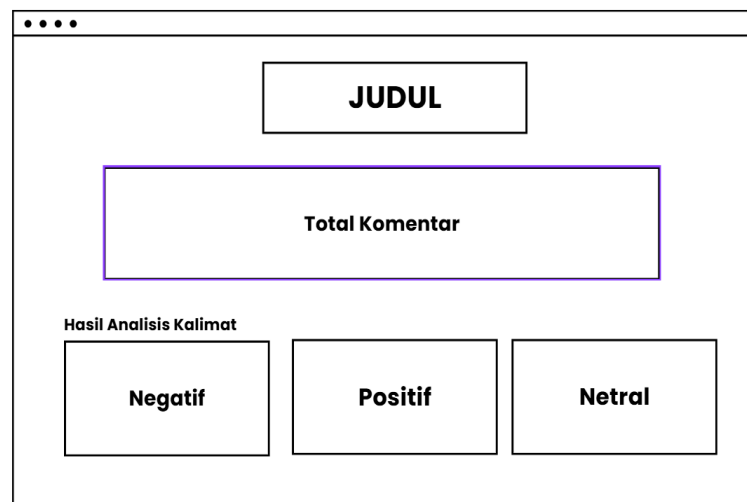
The layout is enclosed in a rectangular frame with a header bar at the top containing four dots.

Gambar 3. 6 Halaman Utama

Pada gambar tersebut merupakan halaman utama yang terdiri dari beberapa komponen utama yaitu judul website di bagian atas sebagai identitas sistem, diikuti oleh bagian Informasi yang berfungsi untuk menampilkan deskripsi atau petunjuk singkat penggunaan sistem. Pada bagian tengah terdapat kolom Link YouTube yang digunakan untuk memasukkan tautan video yang akan

dianalisis, khususnya sebagai sumber data komentar. Di bawahnya terdapat tombol Mulai Analisis yang berfungsi untuk menjalankan proses analisis setelah pengguna memasukkan link video. Secara keseluruhan, desain antarmuka ini dibuat sederhana dan terstruktur agar memudahkan pengguna dalam memasukkan data serta memahami alur penggunaan sistem analisis yang disediakan.

2) Halaman Hasil



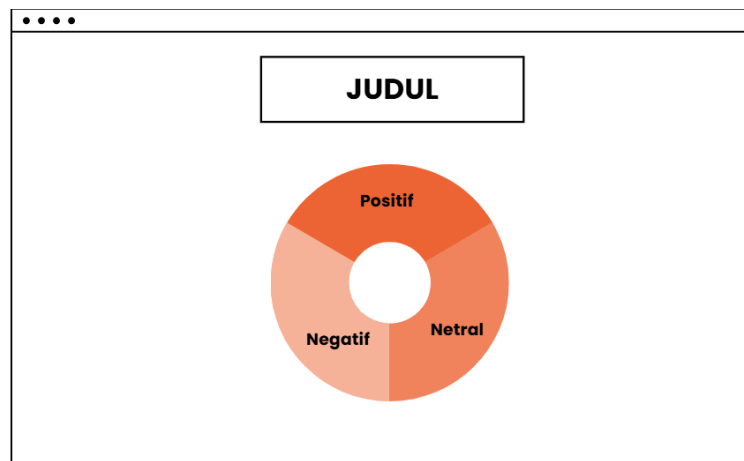
Gambar 3. 7 Halaman Hasil

Gambar tersebut menampilkan rancangan antarmuka halaman hasil analisis yang digunakan untuk menyajikan ringkasan keluaran dari proses analisis sentimen. Pada bagian atas terdapat judul sebagai identitas halaman, kemudian di bagian tengah ditampilkan Total Komentar yang menunjukkan jumlah keseluruhan komentar yang berhasil dianalisis dari sumber data. Di bawahnya terdapat bagian Hasil Analisis Kalimat yang terbagi menjadi tiga kategori utama, yaitu Negatif, Positif, dan Netral, yang masing-masing berfungsi untuk menampilkan jumlah atau ringkasan komentar berdasarkan klasifikasi sentimennya. Desain antarmuka ini disusun secara sederhana dan

terstruktur agar pengguna dapat dengan mudah memahami distribusi sentimen komentar serta memperoleh gambaran umum terhadap opini masyarakat dari data yang dianalisis.

3) Halaman Grafik

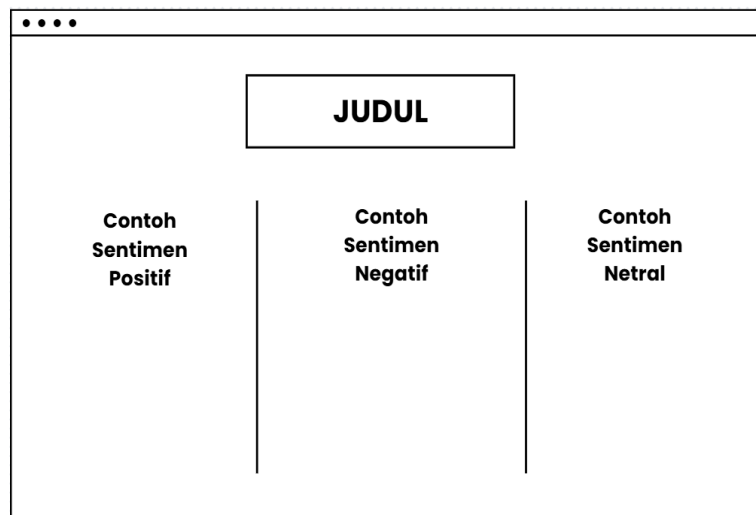
Halaman hasil grafik ini menampilkan visualisasi distribusi sentimen dalam bentuk diagram yang membagi opini masyarakat ke dalam tiga kategori, yaitu positif, negatif, dan netral. Grafik ini berfungsi untuk memberikan gambaran cepat mengenai kecenderungan persepsi publik terhadap isu yang dianalisis, sehingga pembaca dapat dengan mudah memahami proporsi masing-masing sentimen. Desain grafik dibuat sederhana dan informatif dengan penempatan judul di bagian atas sebagai identitas halaman, sementara bagian tengah halaman difokuskan pada diagram agar hasil analisis terlihat lebih jelas dan terpusat. Tampilan ini dirancang untuk membantu pengguna dalam menginterpretasikan hasil analisis sentimen secara intuitif tanpa harus membaca data mentah secara detail.



Gambar 3. 8 Halaman Grafik

4) Halaman Komentar Lain

Halaman ini digunakan untuk menampilkan contoh komentar dari kategori sentimen tertentu berdasarkan hasil klasifikasi sistem. Area halaman dibagi menjadi tiga kolom utama: sentimen positif, sentimen negatif, dan sentimen netral. Setiap kolom berfungsi untuk menampilkan daftar komentar asli dari pengguna media sosial yang termasuk ke dalam kategori tersebut. Halaman ini membantu pengguna melihat konteks opini publik secara langsung, sekaligus memvalidasi kualitas hasil analisis yang dilakukan oleh sistem.



Gambar 3. 9 Halaman Komentar Lain

2. Simulasi Proses Penyelesaian Masalah

a. Data Input

Data *Input* utama dalam simulasi ini adalah teks opini mentah yang diambil dari media sosial (Youtube) terkait Proyek Pembangunan Jalan Tol Kediri. Data ini diperoleh melalui proses *crawling* dengan kata kunci spesifik. Data *input* ini akan melalui tahapan *preprocessing* sebelum dianalisis. Berikut Adalah contoh data input mentah dari ID Opini.

Tabel 3.2 Contoh Data Input Mentah

ID Opini	Teks Asli	Platform
001	Wah, proyek Tol Kediri bagus banget! Pasti macet berkurang dan ekonomi maju pesat. #TolKediri	Youtube
002	Pembebasan lahan tol Kediri ini sangat merugikan warga kecil. Keadilan dimana?	Youtube
003	Pembangunannya lancar jaya. Semoga cepat selesai ya.	Youtube

b. Gambaran Proses

Simulasi proses berpusat pada algoritma Lexicon Based setelah tahapan *preprocessing* (membersihkan teks, *stemming*). Prosesnya melibatkan pencocokan setiap kata dalam teks yang sudah bersih dengan Kamus Leksikon yang telah ditentukan bobotnya (+1 untuk positif, -1 untuk negatif). Skor sentimen (S) dihitung menggunakan rumus dasar:

$$S = \sum(\text{Skor Kata Positif}) - \sum(\text{Skor Kata Negatif}) \dots (3.1)$$

Keterangan:

- 1). Jika $S > 0 \rightarrow$ Sentimen Positif
- 2). Jika $S < 0 \rightarrow$ Sentimen Negatif
- 3). Jika $S = 0 \rightarrow$ Sentimen Netral

Hasil perhitungan skor total ini kemudian digunakan untuk menentukan label sentimen (Positif, Negatif, atau Netral).

Perhitungan :

Teks hasil preprocessing:

"proyek tol kediri bagus banget pasti macet kurang ekonomi maju pesat"

Langkah pertama yaitu identifikasi kata yang ada dalam kamus leksikon:

- 1). bagus $\rightarrow +1$
- 2). maju $\rightarrow +1$

- 3). cepat (tidak muncul pada teks ini)
- 4). merugikan (tidak ada)
- 5). gila (tidak ada)

Langkah kedua yaitu hitung skor kata positif dan negatif:

- 1). Skor positif: bagus (+1), maju (+1), pesat (dianggap positif $\rightarrow +1$) dengan Total skor positif = 3
- 2). Skor negatif: tidak ada dengan Total skor negatif = 0

Langkah ketiga yaitu hasil perhitungan:

$$S = \sum(\text{Skor Kata Positif}) - \sum(\text{Skor Kata Negatif}) \dots (3.2)$$

$$= 3 - 0 = +3$$

Dengan kesimpulan menghasilkan sentiment positif.

Tabel 3.3 Contoh Kamus Leksikon Sederhana

Kata	Bobot
Bagus	+1
Maju	+1
Cepat	+1
Gila	-1
Merugikan	-1
Keadilan	0 (Netral)

c. Data Output

Data Output dari proses ini adalah hasil klasifikasi sentimen dari setiap opini, termasuk skor polaritas total yang dihasilkan oleh algoritma. Hasil ini kemudian di agregasi dan disimpan dalam *database* untuk ditampilkan dalam bentuk *dashboard* pada *website*.

Tabel 3.4 Contoh Data Output (Hasil Klasifikasi)

ID Opini	Teks Hasil <i>Preprocessing</i>	Skor Positif	Skor Negatif	Skor Total (S)	Label Sentimen
001	proyek tol kediri bagus banget pasti macet kurang	3	0	+3	Positif

ID Opini	Teks Hasil <i>Preprocessing</i>	Skor Positif	Skor Negatif	Skor Total (S)	Label Sentimen
	ekonomi maju pesat				
002	bebas lahan tol kediri sangat rugi warga kecil adil dimana	0	2	-2	Negatif
003	bangun lancar jaya semoga cepat selesai	2	0	+2	Positif

d. Pengujian

Pengujian dilakukan untuk memvalidasi kinerja model *Lexicon Based* sebelum diintegrasikan ke *website*. Pengujian utama adalah Pengujian Akurasi Model. Metode yang digunakan adalah membandingkan hasil klasifikasi otomatis (*predicted label*) dari sistem dengan hasil klasifikasi manual (*true label* atau *gold standard*) yang dilakukan oleh *annotator* manusia. Metrik evaluasi yang dipakai adalah Akurasi, Presisi, dan Recall berdasarkan matriks kebingungan (*confusion matrix*). Selain itu, dilakukan pengujian fungsional pada *website* untuk memastikan modul *crawling* dan tampilan *dashboard* berjalan dengan benar.

e. Analisis

Berdasarkan simulasi data pada Tabel 3 diatas dan hasil pengujian menyeluruh, dilakukan analisis untuk menjawab tujuan penelitian. Jika Akurasi model pengujian tinggi (misalnya > 80%), dapat disimpulkan bahwa model *Lexicon Based* yang dibangun efektif dalam mengklasifikasikan opini publik terkait Proyek Tol Kediri. Analisis akhir pada keseluruhan data akan menunjukkan persentase dominan. Misalnya, jika mayoritas hasil adalah Positif, maka disimpulkan bahwa penerimaan masyarakat pada media sosial terhadap proyek tersebut adalah baik, meskipun

perlu diidentifikasi topik negatif spesifik (misalnya, *merugikan* atau *macet*) yang diangkat oleh segmen negatif. Analisis ini memberikan wawasan bagi pengambil kebijakan.

9

BAB IV IMPLEMENTASI DAN HASIL

A. Hasil Penelitian

1. Implementasi Desain Sistem

a. Implementasi Lembar Kerja

8

1) Halaman Utama



Gambar 4. 1 Halaman Utama

Pada gambar 4.1 ditampilkan halaman input serta hasil detail dari proses analisis sentimen. Pengguna dapat memasukkan link video YouTube pada kolom yang tersedia, kemudian menekan tombol Mulai Analisis untuk menjalankan proses. Setelah proses selesai, sistem menampilkan notifikasi bahwa analisis berhasil dilakukan. Selain itu, ditampilkan juga tabel yang berisi daftar komentar beserta hasil klasifikasi sentimennya. Tabel memberikan informasi secara rinci mengenai bagaimana setiap komentar dikategorikan ke dalam sentimen positif, negatif, atau netral, sehingga pengguna dapat melakukan evaluasi terhadap hasil analisis yang diberikan oleh sistem.

2

34

41

2) Halaman Hasil



Gambar 4. 2 Halaman Hasil

Pada gambar 4.2 ditampilkan hasil rekapitulasi dari proses analisis sentimen yang telah dilakukan. Sistem menunjukkan total komentar yang dianalisis sebanyak 19 komentar, dengan rincian 11 komentar positif, 4 komentar negatif, dan 4 komentar netral. Dari hasil tersebut terlihat bahwa sentimen netral mendominasi dari sentimen positif dan negatif. Hal ini menunjukkan bahwa sebagian besar komentar yang dianalisis tidak mengandung opini yang kuat, melainkan lebih bersifat informatif atau berupa pertanyaan.

93

7

7

32

3) Halaman Grafik



Gambar 4. 3 Halaman Grafik

Pada gambar 4.3 ditampilkan hasil analisis dalam bentuk diagram lingkaran (*pie chart*). Diagram tersebut menunjukkan persentase masing-masing kategori sentimen,

122

yaitu netral sebesar 21,1%, positif sebesar 21,1%, dan negatif sebesar 57,9%. Hasil ini mempermudah pengguna dalam memahami pembagian data secara keseluruhan. Dari grafik tersebut dapat dilihat dengan jelas bahwa sentimen netral sangat mendominasi, sedangkan sentimen positif dan negatif hanya memiliki proporsi yang kecil.

4) Halaman Komentar Lain



Gambar 4. 4 Halaman Komentar Lain

Pada gambar 4.4 ditampilkan halaman komentar lain dari sistem analisis sentimen pembangunan Tol Kediri–Tulungagung. Halaman ini berisi contoh klasifikasi sentimen yang dibagi menjadi tiga kategori, yaitu positif, negatif, dan netral. Setiap kategori ditampilkan beberapa contoh komentar untuk memberikan gambaran kepada pengguna mengenai bagaimana sistem mengelompokkan opini berdasarkan isi teks. Dengan adanya contoh ini, pengguna dapat memahami perbedaan masing-masing jenis sentimen sebelum melakukan analisis secara langsung menggunakan data yang diinput.

b. Keterkaitan Antar Lembar Kerja

Pada website ini terdiri dari 4 halaman yang saling terhubung dan membentuk satu kesatuan alur penggunaan. Setiap halaman memiliki peran dalam mendukung proses sistem analisis sentimen pembangunan Tol Kediri–Tulungagung. Berikut adalah uraian keterkaitan antar halaman :

1) Halaman Utama

Pada halaman ini merupakan bagian utama dari sistem karena menjadi tempat pengguna memasukkan data berupa link video YouTube dan menjalankan proses analisis sentimen. Setelah tombol Mulai Analisis ditekan, sistem akan mengambil komentar, memprosesnya, dan menampilkan hasil klasifikasi dalam bentuk tabel yang berisi komentar beserta label sentimennya. Keterkaitannya dengan halaman lain sangat penting, karena seluruh data yang ditampilkan pada halaman hasil, grafik, dan komentar lain berasal dari proses yang dilakukan pada halaman ini. Dengan kata lain, halaman ini menjadi sumber utama data dalam sistem.

2) Halaman Hasil

Halaman hasil menampilkan proses analisis yang telah dilakukan pada halaman input. Informasi yang disajikan berupa jumlah total komentar serta distribusi sentimen (positif, negatif, dan netral). Keterkaitannya dengan halaman input terletak pada penggunaan data hasil klasifikasi yang telah diproses sebelumnya. Halaman ini berfungsi memberikan gambaran umum kepada pengguna terkait hasil analisis secara ringkas sebelum melihat detail.

3) Halaman Grafik

Halaman grafik merupakan pengembangan dari halaman hasil, di mana data rekapitulasi ditampilkan dalam bentuk visual berupa diagram lingkaran. Keterkaitannya sangat jelas karena grafik dibuat berdasarkan data jumlah sentimen yang ditampilkan pada halaman hasil. Dengan adanya visualisasi ini, pengguna dapat lebih mudah memahami perbandingan antar kategori sentimen secara cepat dan intuitif, terutama dalam melihat dominasi sentimen tertentu.

4) Halaman Komentar Lain

Halaman komentar lain menampilkan contoh atau kumpulan komentar yang telah diklasifikasikan ke dalam kategori sentimen tertentu. Halaman ini berkaitan dengan halaman input karena data yang ditampilkan berasal dari hasil pengambilan dan pengolahan komentar. Selain itu, halaman ini juga mendukung halaman hasil dan grafik dengan memberikan konteks nyata dari data yang telah direkap dan divisualisasikan. Melalui halaman ini, pengguna dapat melihat langsung contoh komentar pada masing-masing kategori sentimen sehingga hasil analisis menjadi lebih mudah dipahami dan tidak hanya berupa angka atau grafik saja.

2. Pengujian Fungsional

Pengujian fungsional dilakukan untuk mengevaluasi performa sistem. Hasilnya adalah sebagai berikut :

a. *White Box Testing*

Pengujian *white box* dilakukan untuk menguji struktur internal dan alur logika program pada sistem analisis sentimen. Pengujian ini berfokus pada bagaimana proses dalam sistem berjalan, khususnya pada tahap *text preprocessing* dan perhitungan skor sentimen berbasis kamus leksikon. Tahapan yang diuji meliputi:

1) Tahapan *Case Folding*

Tahapan pertama dalam *preprocessing* adalah *case folding*, yaitu proses normalisasi teks dengan mengubah seluruh huruf menjadi huruf kecil (*lowercase*). Proses ini dilakukan agar sistem tidak membedakan kata berdasarkan penggunaan huruf kapital sehingga proses pencocokan kata dengan kamus leksikon dapat berjalan. Sistem menerima teks mentah kemudian menjalankan fungsi perubahan huruf secara otomatis.

	1_Data_Asli	2_Case_Folding
0	Remajaku dstu omahku.. Industri kraft aku dsik n...	remajaku dstu omahku industri kraft aku dsik n...
1	Sing WIS cair ngguya ngguyumesam mesem ..si...	sing wis cair ngguya ngguyumesam mesem sing du...
2	Wilayah kota kediri lambat ,terutama wilayah m...	wilayah kota kediri lambat terutama wilayah mr...

Gambar 4. 5 Tahapan Case Folding

2) Tahapan Tokenizing

Pada tahap *tokenizing*, yaitu proses pemecahan kalimat menjadi kata-kata tunggal (*token*). Tujuan dari proses ini adalah agar setiap kata dapat dianalisis secara individual pada tahap analisis sentimen. Sistem memisahkan kata berdasarkan spasi serta menghilangkan tanda baca yang tidak diperlukan.

	2_Case_Folding	3_Tokenizing
0	remajaku dstu omahku industri kraft aku dsik n...	[remajaku, dstu, omahku, industri, kraft, aku, ...
1	sing wis cair ngguya ngguyumesam mesem sing du...	[sing, wis, cair, ngguya, ngguyumesam, mesem, ...
2	wilayah kota kediri lambat terutama wilayah mr...	[wilayah, kota, kediri, lambat, terutama, wila...

Gambar 4. 6 Tahapan Tokenizing

3) Tahapan Stopword Removal

Tahap *stopword removal*, yaitu proses penghapusan kata umum yang tidak memiliki kontribusi terhadap nilai sentimen, seperti kata penghubung atau kata yang sering muncul dalam kalimat sehari-hari. Proses ini bertujuan meningkatkan fokus analisis pada kata yang memiliki makna emosional atau opini.

	3_Tokenizing	4_Stopword_Removal
0	[remajaku, dstu, omahku, industri, kraft, aku, ...	[remajaku, dstu, omahku, industri, kraft, aku, ...
1	[sing, wis, cair, ngguya, ngguyumesam, mesem, ...	[sing, wis, cair, ngguya, ngguyumesam, mesem, ...
2	[wilayah, kota, kediri, lambat, terutama, wila...	[wilayah, kota, kediri, lambat, terutama, wila...

Gambar 4. 7 Tahapan Stopword Removal

Berdasarkan hasil pengujian, alur program berjalan dengan baik tanpa ditemukan kesalahan pada proses pengolahan data maupun perhitungan skor. Selain itu, pengujian terhadap jalur program (*path testing*) menunjukkan bahwa seluruh kemungkinan alur eksekusi dapat berjalan sesuai dengan yang diharapkan. Dengan demikian, dapat disimpulkan bahwa struktur logika dalam sistem

telah berjalan dan mampu menghasilkan output yang sesuai dengan proses yang dilakukan.

b. Black Box Testing

Pengujian *black box* dilakukan untuk menguji fungsi sistem berdasarkan input dan output tanpa melihat struktur kode program. Pengujian ini bertujuan untuk memastikan bahwa setiap fitur pada sistem analisis sentimen dapat berjalan sesuai dengan yang diharapkan.

Tabel 4.1 Black Box Testing

No	Skenario Pengujian	Input	Output yang Diharapkan	Hasil
1.	Input link Youtube	Link video Youtube	Sistem berhasil mengambil komentar	Berhasil
2.	Klik tombol analisis	Klik “Mulai Analisis”	Sistem memproses data	Berhasil
3.	Proses preprocessing	Data komentar	Teks diproses (<i>case folding, tokenizing, stopword removal</i>)	Berhasil
4.	Klasifikasi sentimen	Data hasil preprocessing	Sistem menampilkan label positif, negatif, netral	Berhasil
5.	Tampilan Hasil	Data hasil analisis	Menampilkan jumlah sentimen	Berhasil
6.	Tampilan grafik	Data rekap	Grafik <i>pie chart</i> muncul	Berhasil
7.	Tampilan detail komentar	Data komentar	Menampilkan tabel komentar dan sentimen	Berhasil

3. Pengujian NonFungsional

Pengujian nonfungsional dilakukan untuk mengevaluasi kinerja sistem analisis sentimen yang telah dibangun, khususnya pada aspek efisiensi proses analisis, stabilitas aplikasi, serta kemampuan sistem dalam menangani data dalam jumlah besar.

Pengujian Akurasi Metode Lexicon bertujuan untuk mengetahui performa metode lexicon dalam melakukan klasifikasi sentimen

komentar. Pengujian ini dilakukan dengan membandingkan hasil klasifikasi yang dihasilkan sistem dengan hasil pelabelan manual.

Hasil pelabelan sistem kemudian dibandingkan dengan label manual menggunakan metode *Confusion Matrix* untuk memperoleh nilai evaluasi berupa akurasi, *precision*, dan *recall*. Berdasarkan hasil pengujian, sistem memperoleh nilai akurasi sebesar 86%, yang menunjukkan bahwa metode lexicon mampu melakukan klasifikasi sentimen dengan tingkat ketepatan yang cukup baik.

Dataset yang digunakan berisi sebanyak 1608 data komentar. Setelah dilakukan proses pembersihan data, seperti penghapusan komentar duplikat, jumlah data yang digunakan menjadi 1458 komentar. Selanjutnya melakukan pelabelan manual terhadap seluruh komentar menjadi kategori positif, negatif, dan netral. Selanjutnya, dataset yang sama dianalisis oleh sistem menggunakan metode lexicon.

B. Pembahasan

1. Pengujian Fungsional

Berdasarkan hasil pengujian fungsional, sistem analisis sentimen berbasis *lexicon* yang dikembangkan dapat menjalankan seluruh fungsi sesuai dengan perancangan sistem. Pada pengujian *white box testing*, seluruh alur logika program berjalan dengan baik, mulai dari proses *text preprocessing* yang meliputi *case folding*, *tokenizing*, dan *stopword removal* hingga proses perhitungan skor sentimen berbasis kamus leksikon. Setiap tahapan pengolahan teks berhasil menghasilkan output yang sesuai tanpa ditemukan kesalahan logika program.

Selain itu, pengujian *black box testing* menunjukkan bahwa seluruh fitur pada website dapat digunakan dengan normal oleh pengguna. Proses input link YouTube, pengambilan komentar, analisis sentimen, hingga penampilan hasil dalam bentuk tabel dan grafik dapat berjalan secara otomatis tanpa error. Hal ini menunjukkan bahwa sistem telah memenuhi kebutuhan fungsional pengguna serta mampu digunakan secara langsung sebagai alat analisis sentimen.

Secara keseluruhan, hasil pengujian fungsional membuktikan bahwa sistem yang dibangun telah berjalan stabil, sesuai dengan rancangan.

2. Pengujian NonFungsional

Pada pengujian nonfungsional, evaluasi dilakukan untuk mengetahui kemampuan sistem dalam memproses data komentar serta ketepatan metode *Lexicon-Based* dalam melakukan klasifikasi sentimen. Pengujian dilakukan dengan membandingkan hasil klasifikasi sistem dengan hasil pelabelan manual. Berdasarkan hasil pengujian, sistem memperoleh nilai akurasi sebesar 86%, yang menunjukkan bahwa sebagian besar hasil klasifikasi sistem sudah sesuai dengan pelabelan manual. Hasil ini menunjukkan bahwa metode *lexicon-based* mampu mengenali sentimen komentar masyarakat dengan cukup baik.

Sebelum proses klasifikasi dilakukan, sistem terlebih dahulu menjalankan tahap text preprocessing untuk membersihkan data komentar. Proses *preprocessing* meliputi *case folding*, pembersihan karakter, dan penghapusan stopwords. Tahapan ini bertujuan agar teks komentar dapat dikenali oleh kamus leksikon dengan lebih baik.

```
text \
0 Remajaku dstu omahku...😂 Industri kraft aku dsi...
1 Sing WIS cair ngguya ngguyu...mesam mesem ..si...
2 Wilayah kota kediri lambat ,terutama wilayah m...
3 Bismillah Jalan Tol Gresik - Lamongan - Tuban 🍀
4 Sengdurung pencairan sabar lor..mengko lek way...

text_clean
0 remajaku dstu omahku😂 industri kraft aku dsik ...
1 sing wis cair ngguya ngguyumesam mesem sing du...
2 wilayah kota kediri lambat terutama wilayah mr...
3 bismillah jalan tol gresik lamongan tuban 🍀
4 sengdurung pencairan sabar lormengko lek wayah...
```

Gambar 4. 8 Proses *Preprocessing*

Berdasarkan gambar 4.8 terlihat adanya perubahan antara data komentar asli pada kolom *text* dengan hasil preprocessing pada kolom *text_clean*. Pada tahap ini sistem mengubah seluruh huruf menjadi huruf kecil (*case folding*), menghapus karakter yang tidak diperlukan,

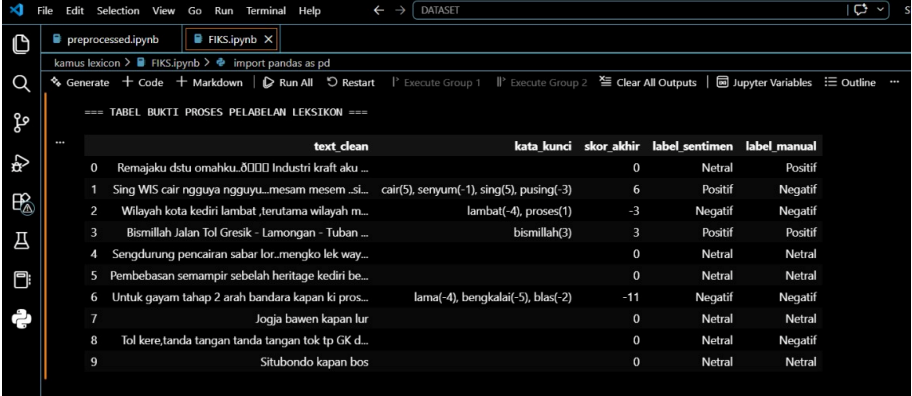
20

serta membersihkan teks agar lebih mudah diproses pada tahap analisis sentimen. Contohnya pada komentar “Remajaku dstu omahku..😂 Industri kraft aku dsik...” setelah *preprocessing* berubah menjadi “remajaku dstu omahku😂 industri kraft aku dsik ...”. Proses ini membuat teks menjadi lebih terstruktur sehingga kata-kata pada komentar dapat dikenali oleh sistem dan dicocokkan dengan kamus leksikon secara lebih akurat.

50

Setelah proses *preprocessing* selesai, sistem melakukan pencocokan kata dengan kamus leksikon positif dan negatif untuk menentukan label sentimen setiap komentar. Proses ini dilakukan dengan menghitung skor sentimen berdasarkan kata yang terdapat pada komentar. Sistem akan mencari kata yang memiliki nilai sentimen pada kamus leksikon, kemudian menghitung total skor untuk menentukan label akhir berupa positif, negatif, atau netral.

134



	text_clean	kata_kunci	skor_akhir	label_sentimen	label_manual
0	Remajaku dstu omahku..😂 Industri kraft aku ...		0	Netral	Positif
1	Sing WIS cair ngguyu ngguyu...mesam mesem ...st...	cair(5), senyum(-1), sing(5), pusing(-3)	6	Positif	Negatif
2	Wilayah kota kediri lambat ,terutama wilayah m...	lambat(-4), proses(1)	-3	Negatif	Negatif
3	Bismillah Jalan Tol Gresik - Lamongan - Tuban ...	bismillah(3)	3	Positif	Positif
4	Sengdunung pencairan sabar lor...mengko lek way...		0	Netral	Netral
5	Pembebasan semampir sebelah heritage kediri be...		0	Netral	Netral
6	Untuk gayam tahap 2 arah bandara kapan ki pros...	lama(-4), bengkalai(-5), blas(-2)	-11	Negatif	Negatif
7	Jogja bawen kapan lur		0	Netral	Netral
8	Tol kere,tanda tangan tanda tangan tok tp GK d...		0	Netral	Negatif
9	Situbondo kapan bos		0	Netral	Netral

Gambar 4. 9 Proses Pelabelan

Pada gambar 4.9 terlihat bahwa sistem berhasil melakukan proses pelabelan sentimen secara otomatis. Pada tabel tersebut ditampilkan beberapa kolom penting seperti *text_clean*, *kata_kunci*, *skor_akhir*, *label_sentimen*, dan *label_manual*. Kolom *kata_kunci* menunjukkan kata yang dikenali oleh sistem beserta bobot sentimennya, misalnya kata “cair(5)” dan “senyum(-1)”. Selanjutnya sistem menghitung total skor pada kolom *skor_akhir* untuk menentukan label sentimen akhir. Jika skor bernilai positif maka komentar dikategorikan sebagai

2

sentimen positif, jika bernilai negatif maka termasuk sentimen negatif, sedangkan skor nol dikategorikan netral.

Pada gambar tersebut juga terlihat perbandingan antara hasil klasifikasi sistem pada kolom `label_sentimen` dengan hasil pelabelan manual pada kolom `label_manual`. Sebagian besar hasil klasifikasi menunjukkan kesesuaian antara sistem dan pelabelan manual, sehingga menunjukkan bahwa metode *lexicon-based* mampu bekerja dengan cukup baik dalam mengenali sentimen komentar masyarakat.

Selanjutnya, hasil pelabelan sistem dibandingkan dengan pelabelan manual menggunakan metode *Confusion Matrix* untuk mengetahui tingkat akurasi sistem. Hasil evaluasi menunjukkan bahwa sistem memperoleh nilai akurasi sebesar 86%.

```

File Edit Selection View Go Run Terminal Help
preprocessed.ipynb FIKS.ipynb X
kamus lexicon > FIKS.ipynb > import pandas as pd
Generate + Code + Markdown Run All Restart Execute Group 1 Execute Group 2
...
=== ANALISIS SENTIMEN SKRIPSI (GOD MODE 86% - BIG DATA EDITION) ===
Pilih Mode:
[1] Cek Validasi (Hitung Akurasi 150 Data)
[2] Analisis Utama (Labeling 1400+ Data Gabungan)
Sedang memuat library Sastrawi...
Membaca file: Validasi_Manual_Skripsi.csv ...
Memproses 150 baris data...

[SUKSES] Hasil disimpan di: HASIL_VALIDASI_REVISI.csv

=====
HASIL AKURASI TERBARU: 86.00%
=====

              precision    recall  f1-score   support

   Negatif         0.96         0.77         0.85         30
    Netral         0.84         0.85         0.85         61
    Positif         0.84         0.92         0.88         59

 accuracy          0.88         0.84         0.86         150
  macro avg         0.88         0.84         0.86         150
 weighted avg         0.86         0.86         0.86         150

```

Gambar 4. 10 Hasil Evaluasi

Berdasarkan hasil pengujian pada Gambar 4.10, sistem analisis sentimen berbasis *Lexicon-Based* memperoleh nilai akurasi sebesar 86% dengan nilai *precision*, *recall*, dan *f1-score* yang cukup baik pada setiap kategori sentimen. Hasil tersebut menunjukkan bahwa sistem mampu melakukan klasifikasi komentar ke dalam kategori positif, negatif, dan netral dengan tingkat ketepatan yang cukup tinggi. Selain itu, nilai *precision* yang tinggi pada kategori negatif menunjukkan bahwa sistem cukup baik dalam mengenali komentar bernada negatif,

42

sedangkan nilai *recall* yang tinggi pada kategori positif menunjukkan bahwa sebagian besar komentar positif berhasil dikenali oleh sistem. Meskipun masih terdapat beberapa kesalahan klasifikasi akibat penggunaan bahasa tidak baku, singkatan, bahasa daerah, maupun konteks kalimat yang ambigu, secara umum metode *lexicon-based* tetap mampu memberikan performa yang stabil dalam proses analisis sentimen komentar masyarakat.

20

Secara keseluruhan, hasil pengujian nonfungsional menunjukkan bahwa sistem mampu menjalankan proses *preprocessing*, pelabelan sentimen, hingga evaluasi klasifikasi dengan baik dan terstruktur. Sistem juga dapat mengolah data komentar dalam jumlah besar secara stabil tanpa mengalami kendala yang berarti selama proses analisis berlangsung. Dengan demikian, sistem yang dikembangkan sudah dapat digunakan sebagai alat bantu untuk menganalisis opini masyarakat terhadap pembangunan Tol Kediri–Tulungagung, serta mampu memberikan informasi sentimen masyarakat secara lebih cepat dan terstruktur.

33

51

BAB V PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah disajikan pada bab sebelumnya, maka dapat diperoleh kesimpulan sebagai berikut:

1. Metode *Lexicon-Based Sentiment Analysis* berhasil diterapkan untuk menganalisis opini masyarakat terhadap pembangunan Tol Kediri–Tulungagung berdasarkan komentar pada platform YouTube. Proses analisis dilakukan melalui tahapan pengumpulan data, preprocessing, pencocokan kata dengan kamus leksikon, serta klasifikasi sentimen ke dalam kategori positif, negatif, dan netral. Hasil pengujian menunjukkan bahwa metode yang digunakan memperoleh nilai akurasi sebesar 86%, sehingga mampu mengidentifikasi kecenderungan sentimen masyarakat dengan cukup baik.
2. Arsitektur sistem analisis sentimen berbasis *Lexicon-Based* berhasil dirancang dan diimplementasikan dalam bentuk *website*. Sistem mampu melakukan pengambilan data komentar YouTube, *preprocessing teks*, klasifikasi sentimen, serta menampilkan hasil analisis dalam bentuk tabel dan grafik secara otomatis. Berdasarkan hasil pengujian fungsional dan nonfungsional, seluruh fitur sistem dapat berjalan sesuai dengan rancangan yang telah dibuat dan mampu mengolah data komentar dalam jumlah besar secara stabil. Dengan demikian, sistem yang dikembangkan dapat digunakan sebagai alat bantu untuk memetakan opini masyarakat terhadap pembangunan Tol Kediri–Tulungagung.

B. Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya, yaitu sebagai berikut:

1. Penelitian selanjutnya dapat mengembangkan kamus leksikon yang digunakan dengan menambahkan kosakata baru, bahasa gaul, singkatan,

serta istilah yang sering digunakan pada media sosial sehingga hasil klasifikasi sentimen dapat menjadi lebih akurat.

2. Sistem analisis sentimen yang telah dibangun dapat dikembangkan dengan menambahkan fitur analisis yang lebih lengkap, seperti pencarian kata kunci yang sering muncul, visualisasi data yang lebih interaktif, serta penyajian hasil analisis secara *real-time*.
3. Penelitian selanjutnya dapat melakukan perbandingan antara metode *Lexicon-Based* dengan metode analisis sentimen lainnya untuk mengetahui tingkat akurasi dan performa yang dihasilkan pada data komentar terkait pembangunan infrastruktur.
4. Sistem yang dikembangkan dapat diterapkan pada objek penelitian lain yang berkaitan dengan opini masyarakat, sehingga manfaat sistem analisis sentimen dapat digunakan pada berbagai topik dan kebutuhan analisis data.