

Turnitin _

Proposal Athallah Hazel 221080200135

 논문 및 과제 검사 (검사시 DB 미저장 Originality Check - No Repository)

Document Details

Submission ID

trn:oid::3618:145231594

Submission Date

Jul 9, 2026, 2:57 PM GMT+7

Download Date

Jul 9, 2026, 3:04 PM GMT+7

File Name

Proposal Athallah Hazel 221080200135.doc

File Size

780.0 KB

29 Pages**5,675 Words****37,085 Characters**

13% Overall Similarity

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

Filtered from the Report

- Bibliography
- Quoted Text
- Small Matches (less than 10 words)

Top Sources

- 10%  Internet sources
 - 5%  Publications
 - 9%  Submitted works (Student Papers)
-

Top Sources

- 10% Internet sources
- 5% Publications
- 9% Submitted works (Student Papers)

Top Sources

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

1	Internet	ejournal.itn.ac.id	1%
2	Student papers	Universitas Muhammadiyah Sidoarjo on 2024-03-27	<1%
3	Internet	archive.umsida.ac.id	<1%
4	Internet	www.journal.irpi.or.id	<1%
5	Internet	ojs3.unpatti.ac.id	<1%
6	Student papers	Universitas Islam Bandung on 2026-04-16	<1%
7	Internet	eprints.walisongo.ac.id	<1%
8	Student papers	Universitas Budi Luhur on 2025-07-10	<1%
9	Student papers	Universitas Maritim Raja Ali Haji on 2025-02-28	<1%
10	Internet	repositori.telkomuniversity.ac.id	<1%
11	Internet	repository.binadarma.ac.id	<1%

12	Internet	dspace.uui.ac.id	<1%
13	Internet	eprints.akakom.ac.id	<1%
14	Internet	ejournal.akakom.ac.id	<1%
15	Internet	www.receh.in	<1%
16	Student papers	UNIVERSITAS BUDI LUHUR on 2026-02-06	<1%
17	Student papers	Universitas Budi Luhur on 2026-01-25	<1%
18	Student papers	Universitas Putera Batam on 2026-01-31	<1%
19	Internet	ejournal.unesa.ac.id	<1%
20	Internet	repository.unmuhjember.ac.id	<1%
21	Student papers	Syntax Corporation on 2025-12-03	<1%
22	Student papers	Universitas Budi Luhur on 2025-07-13	<1%
23	Internet	journal.stmiki.ac.id	<1%
24	Internet	repository-penerbitlitnus.co.id	<1%
25	Internet	repository.umnaw.ac.id	<1%

26	Publication	Khadijah, Imam Saputra. "Analisis Sentimen Pengguna Aplikasi Gojek Mengguna...	<1%
27	Student papers	Telkom University on 2026-01-21	<1%
28	Student papers	UPN Veteran Yogyakarta on 2026-06-24	<1%
29	Student papers	Universitas Budi Luhur on 2025-07-10	<1%
30	Student papers	Universitas Islam Indonesia on 2026-05-12	<1%
31	Student papers	Universitas Islam Negeri Sumatera Utara on 2023-10-10	<1%
32	Student papers	Universitas Pendidikan Indonesia on 2014-05-30	<1%
33	Internet	jurnal.itg.ac.id	<1%
34	Internet	www.waspada.id	<1%
35	Publication	Muhammad Irvan Maulana, Savana Putra Aditama, Harun Al Rosyid. "Komparasi...	<1%
36	Internet	choliqbaya.blogspot.com	<1%
37	Internet	e-journal.hamzanwadi.ac.id	<1%
38	Internet	repository.uin-suska.ac.id	<1%
39	Internet	repository.umy.ac.id	<1%

40

Internet

www.mdpi.com

<1%

PROPOSAL SKRIPSI

ANALISIS SENTIMEN REVIEW AWAL PEMBENTUKAN PROGRAM KOPERASI DESA MERAH PUTIH DENGAN METODE NAIVE BAYES DAN LOGISTIC REGRESSION



Muhammad Athallah Hazel Orlando
221080200135

DOSEN PEMBIMBING

Nuril Lutvi Azizah, S.Si., M.Si

PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH SIDOARJO
2026

HALAMAN PENGESAHAN

ANALISIS SENTIMEN KOMENTAR YOUTUBE REVIEW AWAL TERHADAP PEMBENTUKAN PROGRAM KOPERASI DESA MERAH PUTIH DENGAN METODE NAIVE BAYES DAN LOGISTIC REGRESSION

Muhammad Athallah Hazel Ornando
221080200135

Sidoarjo, 1 Juli 2026

Mengetahui
Ketua Program Studi
Informatika

Menyetujui,
Dosen Pembimbing

Ade Eviyanti, S.Kom., M.Kom
NIDN. 0724057803

Nuril Lutvi Azizah, S.Si., M.Si
NIDN. 0729048901

BAB I

PENDAHULUAN

1.1 Latar Belakang

Program strategis pemerintah dalam bentuk program Koperasi Desa Merah Putih, memiliki tujuan untuk meningkatkan ekonomi kerakyatan berbasis desa, didasarkan pada Instruksi Presiden Nomor 9 Tahun 2025 tentang Percepatan Pembentukan Koperasi Desa/Kelurahan Merah Putih., dengan target pendirian, pengembangan, dan revitalisasi 80.000 koperasi di desa dan kelurahan [1]. Kopdes Merah Putih diarahkan sebagai lembaga ekonomi yang dapat menjalankan layanan kebutuhan pokok, simpan pinjam, logistik, pergudangan, apotek, klinik, serta penguatan rantai distribusi lokal. Dengan demikian, koperasi diposisikan sebagai instrumen pemerataan pembangunan, kemandirian ekonomi, dan peningkatan daya tawar masyarakat desa [2]

Penelitian Darwis, Siskawati, dan Abidin menunjukkan bahwa algoritma Naive Bayes dapat diterapkan dalam analisis sentimen review data Twitter BMKG Nasional dengan mengklasifikasikan opini masyarakat ke dalam tiga kelas, yaitu positif, negatif, dan netral, serta menghasilkan akurasi sebesar 69,97% [3]. Ridwan menjelaskan bahwa Naive Bayes merupakan algoritma pembelajaran mesin untuk masalah klasifikasi yang banyak digunakan pada klasifikasi teks berdimensi tinggi, termasuk analisis sentimen dan penyaringan spam, karena sederhana, efektif, serta mampu membangun model secara cepat [4]. Haq, Octiva, Ayuliana, Nuryanto, dan Suryadi menerapkan Naive Bayes untuk mengidentifikasi hoaks di media sosial dengan tahapan prapemrosesan berupa tokenisasi, ekstraksi fitur, pembersihan teks, dan TF-IDF, sehingga menunjukkan bahwa metode ini sesuai untuk pengolahan data teks dari media sosial [5]. Watratan, Puspita, dan Moeis menjelaskan bahwa Naive Bayes memiliki kelebihan berupa perhitungan yang cepat, efisien, mudah dipahami, dapat digunakan untuk klasifikasi biner maupun multikelas, tetapi memiliki keterbatasan karena asumsi independensi antarvariabel dapat memengaruhi akurasi ketika terdapat hubungan antarfitur [6]. Berdasarkan penelitian terdahulu tersebut, Naive Bayes layak digunakan dalam penelitian analisis sentimen

komentar YouTube terhadap pembentukan Program Koperasi Desa Merah Putih karena komentar publik merupakan data teks tidak terstruktur yang perlu diklasifikasikan ke dalam sentimen positif, negatif, dan netral.

Analisis sentimen menjadi metode yang menarik untuk membaca opini masyarakat secara menyeluruh. Dalam penelitian ini, metode Naive Bayes dan Logistic Regression dipilih karena keduanya termasuk algoritma pembelajaran mesin yang banyak digunakan untuk klasifikasi teks [3]. Naive Bayes memiliki keunggulan pada kesederhanaan model, efisiensi komputasi, dan kemampuan bekerja dengan baik pada data teks berbasis probabilitas kata. Logistic Regression unggul dalam interpretabilitas, stabilitas performa, dan kemampuan untuk memodelkan hubungan antara fitur teks dengan kelas sentimen.

Meskipun memiliki tujuan yang ideal, pembentukan Kopdes Merah Putih tidak terlepas dari berbagai permasalahan implementatif [7]. Pembentukan koperasi dalam jumlah besar berpotensi menghadapi kendala kesiapan sumber daya manusia, pemahaman tata kelola koperasi, administrasi kelembagaan, transparansi pengelolaan, serta kesesuaian model usaha dengan potensi lokal desa. Penelitian tentang pembentukan koperasi ini juga menemukan kendala berupa rendahnya literasi kelembagaan, keterbatasan administrasi, dan ketidaksesuaian petunjuk teknis dengan kondisi lokal [8]. Kondisi tersebut dapat memunculkan opini publik yang beragam, mulai dari dukungan, keraguan, hingga kritik. Salah satu ruang digital yang merekam respons awal masyarakat adalah kolom komentar YouTube pada konten review program [2].

Penelitian Ramadhani dan Suryono menunjukkan bahwa Logistic Regression relevan digunakan dalam analisis sentimen karena metode ini termasuk algoritma klasifikasi teks yang dapat digunakan untuk mengelompokkan opini masyarakat ke dalam kategori sentimen, serta pada studi sentimen metaverse menghasilkan akurasi 95% setelah optimasi SMOTE dengan nilai precision 94%, recall 93%, dan F1-score 95% [9]. Moelyono, Setiawan, dan Wijaya menjelaskan bahwa Logistic Regression termasuk metode klasifikasi dalam supervised learning yang dapat dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score, sehingga sesuai digunakan untuk mengukur performa model klasifikasi secara kuantitatif [10]. Berdasarkan penelitian terdahulu tersebut, Logistic

Regression layak digunakan dalam penelitian analisis sentimen komentar YouTube terhadap pembentukan Program Koperasi Desa Merah Putih karena komentar publik merupakan data teks tidak terstruktur yang perlu diklasifikasikan secara terukur ke dalam sentimen positif, negatif, dan netral.

1.2 Rumusan Masalah

Berdasarkan penjelasan pada latar belakang diatas, maka rumusan masalah pada penelitian ini adalah bagaimana hasil performa metode Naive Bayes dan Logistic Regression dalam analisis sentimen komentar Youtube terhadap awal pembentukan koperasi merah putih?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah, sistematis, dan tidak meluas dari fokus kajian yang telah ditentukan, maka diperlukan batasan masalah sebagai ruang lingkup penelitian. Batasan masalah pada penelitian ini adalah:

1. Penelitian ini hanya menggunakan data komentar dari platform YouTube yang membahas program Koperasi Merah Putih.
2. Data komentar yang digunakan merupakan komentar berbahasa Indonesia.
3. Sentimen komentar diklasifikasikan pada penelitian ini terdiri atas tiga kategori, yaitu sentimen positif, sentimen negatif, dan netral.
4. Proses pengambilan data dilakukan menggunakan teknik web scraping pada komentar YouTube.
5. Data komentar diambil dari platform YouTube menggunakan 4 video yang di upload oleh channel berita, yang terdiri dari satu video Kompas.com, satu video Tempodotco, satu video dari Hersubeno Point, dan satu video dari Fristian Greic Media.
6. Jumlah dataset yang didapat kurang lebih sejumlah 11000 data comments dengan 1 video diberikan batasan 5000 komentar.
7. Dataset merupakan data komentar yang hanya bertipe text tidak termasuk emoji.

22

1.4 Tujuan

Tujuan penelitian ini adalah untuk mengetahui hasil performa metode Naïve Bayes dan Logistic Regression dalam analisis sentimen komentar Youtube terhadap terhadap awal pembentukan koperasi merah putih.

BAB II

METODOLOGI

2.1. Penelitian Terdahulu

Penelitian terdahulu menjadi bagian penting dalam penyusunan penelitian ini karena dapat memberikan gambaran mengenai pendekatan, metode, dan hasil yang telah dicapai oleh peneliti sebelumnya. Melalui kajian penelitian terdahulu, peneliti dapat mengetahui sejauh mana suatu metode telah digunakan, bagaimana tingkat keberhasilannya, serta kelemahan apa saja yang masih dapat dikembangkan.

Penelitian pertama, yang dilakukan oleh Bagus Ramadhani dan Ryan Randi Suryono pada tahun 2024 [9], menganalisis 6.728 komentar masyarakat tentang metaverse pada media sosial X. Penelitian ini menggunakan teknik text mining yang menggunakan perbandingan Naive Bayes dan Logistic Regression. Setelah preprocessing dan pelabelan, data menunjukkan ketidakseimbangan kelas, dengan 5.799 sentimen positif dan 795 sentimen negatif. Setelah SMOTE, kedua model menunjukkan peningkatan, tetapi Naive Bayes memperoleh akurasi 90% dan Logistic Regression 91%, tetapi performa kelas negatif masih rendah. Setelah SMOTE, Logistic Regression menjadi model terbaik dengan akurasi 95%, ketepatan 94%, recall 93%, dan skor F1.

Kedua, penelitian yang dilakukan oleh Meisa Monica dan Asto Purwanto tahun 2026 [11], penelitian ini mengevaluasi kinerja Naïve Bayes dan Logistic Regression untuk klasifikasi sentimen komentar YouTube berbahasa Indonesia pada konteks layanan digital. Data diperoleh melalui crawling komentar menggunakan kata kunci Tokopedia, Shopee, Gojek, OVO, dan Traveloka, lalu diproses melalui cleaning, normalisasi, tokenisasi, stopword removal, stemming, serta pembobotan TF-IDF. Distribusi sentimen menunjukkan komentar positif paling dominan sebesar 40,7%, diikuti netral 35,2% dan negatif 24,1%. Hasil evaluasi memperlihatkan Naïve Bayes meningkat dari akurasi 67,01% menjadi 70,96% setelah SMOTE. Logistic Regression menunjukkan performa lebih baik, dengan akurasi 84,27% pada model normal, 84,88% setelah Grid Search, dan

84,76% setelah SMOTE. Model terbaik adalah Logistic Regression dengan Grid Search.

Ketiga, penelitian yang dilakukan oleh Andre Septa Wijaya dkk. Tahun 2026 [12], penelitian ini akan melakukan analisis sentimen komentar YouTube tentang edukasi kesehatan mental pada tahun 2026 dengan menggunakan representasi fitur IndoBERT dan algoritma Logistic Regression dan Naive Bayes. Dr_tirta, Kok Bisa, dan Alodokter kanal menggunakan YouTube Data API untuk mengumpulkan data. Dari 2.918 komentar awal, 1.848 di antaranya adalah valid, dibagi menjadi sentimen positif, netral, dan negatif, setelah diproses dan difilter. Untuk mengatasi ketidakseimbangan kelas, SMOTE diterapkan pada 80% data latih dan 20% data uji. Dengan menggunakan IndoBERT berukuran 768, fitur teks diekstraksi. Hasil evaluasi menunjukkan bahwa Logistic Regression unggul dengan akurasi 78%, sedangkan Naive Bayes kalah dengan akurasi 56%.

2.2. Ringkasan Penelitian Terdahulu

Ringkasan penelitian terdahulu disusun untuk memberikan gambaran mengenai kajian-kajian sebelumnya yang relevan dengan topik penelitian. Bagian ini memuat metode, objek, hasil, serta keterbatasan penelitian sebelumnya sebagai dasar untuk menentukan posisi, arah, dan kebaruan penelitian yang dilakukan. Ringkasan penelitian terdahulu dapat dilihat pada Tabel 2.1.

Tabel 2.1 Ringkasan Penelitian Terdahulu

No	Judul dan Peneliti	Pembahasan	Metode	Kekurangan dan Kelebihan
1.	Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse [9]	Penelitian ini membandingkan Naive Bayes dan Logistic Regression untuk sentimen metaverse pada 6.728 komentar X. Data tidak seimbang diatasi dengan SMOTE. Setelah optimasi, Logistic Regression menjadi model terbaik dengan akurasi 95%,	Algoritma Naïve Bayes dan Logistic Regression	Kelebihan: Menggunakan 6.728 data X, membandingkan Naive Bayes dan Logistic Regression, menerapkan TF-IDF serta SMOTE, sehingga evaluasi lebih kuat pada data sentimen tidak seimbang metaverse publik.

		precision 94%, recall 93%, dan F1-score 95%, mengungguli Naive Bayes.		Kekurangan: Penelitian hanya memakai dua algoritma, data terbatas pada media sosial X, serta belum menguji model lain seperti SVM, Random Forest, atau deep learning berbasis konteks.
2.	Evaluasi komparatif kinerja algoritma naïve bayes dan logistic regression dalam klasifikasi sentimen komentar youtube berbahasa indonesia [11]	Studi ini menemukan bahwa Logistic Regression lebih efektif daripada Naive Bayes dalam klasifikasi sentimen komentar YouTube layanan digital. Sentimen positif paling dominan adalah 40,7%, dan Logistic Regression dengan Grid Search memiliki tingkat akurasi terbaik 84,88%. Setelah SMOTE, Naive Bayes mencapai tingkat akurasi tertinggi 70,96%.	Algoritma Naïve Bayes dan Logistic Regression	Kelebihan: Studi membandingkan Naive Bayes dan Logistic Regression, menggunakan TF-IDF, Grid Search, SMOTE, dan banyak kata kunci layanan digital untuk memberikan evaluasi model yang lebih akurat dan menyeluruh. Kekurangan: Penelitian bergantung pada pelabelan lexicon, konteks layanan digital masih umum, dan belum menguji model transformer yang lebih kuat untuk memahami bahasa tidak baku YouTube informal.
3.	Sentiment classification of health education youtube comments using	Penelitian ini menunjukkan bahwa kombinasi IndoBERT embedding dan	Logistic Regression and Naïve Bayes	Kelebihan: Menggunakan IndoBERT kontekstual, SMOTE, data

	indobert embeddings with logistic regression and naïve bayes [12]	Logistic Regression lebih efektif untuk klasifikasi sentimen komentar YouTube edukasi kesehatan mental. Dari 1.848 komentar valid, model Logistic Regression mencapai akurasi 78%, sedangkan Naive Bayes hanya 56%, meskipun keduanya dibantu SMOTE.		YouTube kesehatan nyata, serta membandingkan Logistic Regression dan Naive Bayes sehingga evaluasi model lebih modern dan terukur secara komparatif jelas. Kekurangan: Data masih terbatas pada tiga kanal kesehatan, kelas sentimen tidak seimbang, Naive Bayes lemah, dan belum melakukan fine-tuning penuh IndoBERT untuk meningkatkan performa klasifikasi multikelas signifikan.
1	4. Analisis Sentimen Komentar Youtube Tentang Prediksi Resesi Ekonomi Tahun 2023 Menggunakan Algoritme Sentiment Analysis Of Youtube Comments On Prediction Of Economic Recession In 2023 Using The Naïve Bayes [13].	Penelitian ini menganalisis sentimen 500 komentar YouTube pada video Raymond Chin tentang prediksi resesi ekonomi 2023 menggunakan Naive Bayes, dengan tahapan proses crawling, pelabelan pakar, preprocessing, pembobotan Term Frequency, dan 10-Fold Cross Validation. Hasil menunjukkan	Naïve Bayes	Kelebihan: Preprocessing lengkap dan 10-Fold Cross Validation membuat data komentar lebih bersih terstruktur serta pengujian model lebih stabil dalam klasifikasi sentimen berbasis YouTube secara sistematis. Kekurangan: Data hanya 500 komentar dari satu video memakai satu algoritma dan

		sentimen positif 40%, netral 34%, negatif 26%, sehingga opini dominan positif secara umum pada video tersebut.		belum menyajikan metrik evaluasi lengkap seperti precision recall F1-score secara terperinci dalam kesimpulan.
5.	Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier [14]	Studi ini menggunakan Naive Bayes Classifier untuk menganalisis sentimen dari 1009 komentar YouTube yang menyoroti Anies Baswedan sebagai calon presiden potensial pada tahun 2024. Scraping, pelabelan, preprocessing, TF-IDF, dan confusion matrix adalah komponen proses. Hasil menunjukkan 610 komentar positif dan 399 negatif. Akurasi terbaik tanpa penghapusan stopword adalah 79%, sedangkan pengujian Naive Bayes mencapai 78%.	Naïve Bayes Classifier	Kelebihan: Penelitian ini menggunakan data YouTube yang relevan, langkah-langkah preprocessing yang jelas, pelabelan menggunakan tiga dosen bahasa, dan pengujian menggunakan matriks kekacauan dan variasi rasio data latihan. Kekurangan: Penelitian belum membandingkan kinerja dengan algoritma lain, hanya menggunakan Naive Bayes dan mengumpulkan hanya 1009 komentar, dengan klasifikasi hanya positif-negatif.
6.	Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode Naïve Bayes [15].	Penelitian tersebut menemukan bahwa opini pengguna YouTube mengenai pemindahan Ibu Kota Negara sebagian besar bersifat netral. Dari 7.728 komentar awal, proses	Naïve Bayes	Kelebihan: Penelitian menggunakan data YouTube cukup besar, preprocessing lengkap, pelabelan VADER, serta SMOTE untuk mengatasi

		preprocessing, pelabelan VADER, dan filtering menghasilkan data yang tidak seimbang. Naive Bayes tanpa SMOTE belum optimal mengenali kelas minoritas, sedangkan SMOTE meningkatkan performa dengan balanced accuracy 76,01%, AUC 0,8711, dan G-mean 0,8089.		imbalance data sehingga performa Naive Bayes meningkat. Kekurangan: Data didominasi sentimen netral, sumber hanya dari tiga video YouTube, serta belum membandingkan Naive Bayes dengan algoritma lain secara lebih luas.
7.	Analisis sentimen komentar video mobil listrik di platform youtube dengan metode Naive Bayes	Penelitian ini memeriksa 1517 data komentar YouTube tentang mobil listrik menggunakan Naive Bayes. Akurasi yang ditingkatkan dengan SMOTE Upsampling meningkat dari 50,70% menjadi 70,69%, dengan precision 43,64% dan recall 39,48%. Hasil menunjukkan bahwa SMOTE membantu mengatasi ketidakseimbangan kelas dan meningkatkan kinerja klasifikasi.	Naïve Bayes	Kelebihan: Studi ini menggunakan dataset yang cukup besar dan menggunakan KDD, RapidMiner, Naive Bayes, dan SMOTE Upsampling, yang meningkatkan akurasi secara signifikan dari 50,70% menjadi 70,69%. Kekurangan: Studi ini masih memiliki tingkat akurasi dan recall yang rendah, dan belum membandingkan Naive Bayes dengan algoritma klasifikasi lain yang berkaitan dengan mobil listrik.
8.	Klasifikasi	Studi ini	Natural	Kelebihan:

	Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten 2024 [16]	menggunakan NLP untuk mengolah 1.729 komentar YouTube debat Pilkada Banten 2024. SVM memiliki akurasi 73% yang lebih tinggi, sedangkan Naive Bayes memiliki 71% dan lebih akurat. Sentimen negatif tertinggi pada 49%, diikuti oleh sentimen negatif pada 35% dan sentimen netral pada 16%.	Language Processing (NLP).	Penelitian menggunakan NLP, visualisasi sentimen publik, membandingkan SVM dan Naive Bayes, dan 1.729 komentar dari beberapa video YouTube. Kekurangan: Model masih lemah mengenali sentimen netral, serta menghadapi kendala bahasa ambigu, ironi, dan sarkasme dalam komentar politik di YouTube.
9.	Perbandingan sent2vec tf-idf logistic regression dan word2vec cnn pada hasil sentiment analysis youtube comment [17]	Penelitian ini membandingkan Sent2vec TF-IDF Logistic Regression dan Word2vec CBOW CNN pada komentar YouTube. Hasil menunjukkan Word2vec CBOW CNN lebih unggul dengan akurasi 88,44%, precision 94,38%, recall 88,6%, F1-score 90,81%, dan ROC 95,2%, dibanding Logistic Regression yang akurasinya 84,35%.	TF-IDF Logistic Regression dan Word2vec CNN	Kelebihan: Penelitian membandingkan dua model representasi teks menggunakan data besar komentar YouTube, serta menyajikan metrik lengkap: accuracy, precision, recall, F1-score, dan ROC secara terukur komparatif jelas. Kekurangan: Penelitian hanya menggunakan satu sumber video YouTube, pelabelan berbasis kamus sederhana, serta belum menguji dataset lain atau model terbaru berbasis

				transformer untuk validasi lebih luas.
23	10. Efektivitas logistic regression dalam analisis sentimen berbahasa Indonesia pada komentar youtube tentang isu ketenagakerjaan [18]	Penelitian ini mengklasifikasikan sentimen komentar YouTube isu ketenagakerjaan dengan Logistic Regression. Dari 1.020 komentar berlabel, sentimen negatif paling dominan. Model berbasis TF-IDF, SMOTE, dan GridSearchCV memperoleh akurasi 44%, dengan performa terbaik pada kelas negatif melalui F1-score 0,55.	Logistic Regression	Kelebihan: Penelitian memberikan dasar terukur untuk analisis sentimen ketenagakerjaan berbahasa Indonesia pada komentar publik digital melalui penggunaan data nyata YouTube, pelabelan manual, TF-IDF, SMOTE, dan GridSearchCV. Kekurangan: Model tidak memiliki performa kelas yang baik (positif dan netral), hanya memiliki satu anotator yang dilabelkan, dan belum menggunakan model deep learning kontekstual untuk meningkatkan generalisasi.
20 35	11. Analisis Sentimen Komentar Youtube Masterchef Indonesia Menggunakan Algoritma Support Vector Machine Dan Gaussian Naïve Bayes [19]	Penelitian ini menunjukkan SVM lebih unggul dibanding Gaussian Naïve Bayes dalam klasifikasi sentimen komentar Masterchef Indonesia. Setelah oversampling, SVM mencapai akurasi 85%, presisi 84%, dan recall 89%,	Algoritma Support Vector Machine Dan Gaussian Naïve Bayes	Kelebihan: Menggunakan data nyata komentar YouTube, pelabelan ahli NLP, dua algoritma pembandingan, TF-IDF, K-Fold Cross Validation, serta oversampling yang terbukti meningkatkan

		sedangkan Gaussian Naïve Bayes meningkat menjadi akurasi 72%, presisi 68%, dan recall 93%.		performa model secara jelas terukur. Kekurangan: Data hanya berasal dari satu topik Masterchef Indonesia, kelas sentimen tidak seimbang, dan penelitian belum mengeksplorasi tuning mendalam atau algoritma lain sebagai pembandingan tambahan lanjutan.
12	Analisis sentimen komentar youtube terhadap kebijakan dedi mulyadi dengan metode machine learning (studi kasus : kenakalan remaja) [20]	Penelitian ini menemukan bahwa komentar YouTube tentang kebijakan Dedi Mulyadi didominasi oleh sentimen positif (52,4% dan 47,6% masing-masing) dan negatif (47,6% masing-masing). Model terbaik adalah Logistic Regression dengan akurasi 0,84, sedangkan Ensemble Majority Voting tetap stabil pada 0,82. Naive Bayes mendapatkan 0,80, sementara terendah KNN adalah 0,60 0,61.	Machine Learning	Kelebihan: Studi ini membuat evaluasi sentimen kebijakan publik lebih stabil dan terukur secara komparatif dengan menggunakan data YouTube nyata, preprocessing lengkap, TF-IDF, tiga algoritma, dan kelompok suara mayoritas. Kekurangan: Penelitian hanya menggunakan dua kelas sentimen, bergantung pada lexicon otomatis, memiliki KNN yang rendah, dan belum menguji data dari platform atau masalah kebijakan lainnya.
13.	Analisis sentimen pada komentar youtube dalam	Studi tersebut menganalisis komentar YouTube	Esemble Machine Learning	Kelebihan: Penelitian ini menggunakan data

	turnamen mplseason13 dengan metodeensemble machine learning [21]	terkait MPL Season 13 dengan metode ensemble machine learning. Setelah preprocessing, data berkurang dari 6.424 menjadi 5.185 komentar, dengan dominasi sentimen positif. Model hard voting memberikan akurasi tertinggi sebesar 86,70%, dan EVOS tercatat sebagai tim paling banyak didukung.		YouTube nyata, preprocessing lengkap, SMOTE, banyak algoritma, dan kelompok voting hard-soft untuk analisis popularitas tim yang lebih akurat. Kekurangan: Studi ini hanya membahas MPL Season 13, kelas sentimen positif-negatif terbatas, dan validasi manual mungkin subjektif. Itu juga belum menguji konteks e-sports lain secara menyeluruh pada berbagai dataset.
14.	Analysis of apriori and k-nearest neighbor (knn) algorithm in recommending appropriate learning method [22]	Penelitian ini menunjukkan integrasi Apriori dan KNN efektif merekomendasikan metode pembelajaran sesuai karakteristik mahasiswa. Dari 150 responden, metode digital paling diminati sebesar 84,29%, sedangkan ceramah 15,70%. KNN optimal pada k=5 dengan akurasi 84,29%, precision 0,86, recall 0,83, dan F1-score 0,84, sehingga mendukung pembelajaran adaptif berbasis data	Algoritma Apriori dan K- Nearest Neighbour KNN)	Kelebihan: Penelitian ini mengintegrasikan Apriori dan KNN dengan menggunakan data siswa nyata, normalisasi, validasi k-fold, dan metrik lengkap. Ini membuat saran metode pembelajaran lebih terukur, fleksibel, dan relevan untuk institusi. Kekurangan: Dataset hanya terdiri dari 150 responden, konteksnya terbatas pada mahasiswa

		secara terarah dan inklusif.		informatika, dan ada ketidakseimbangan metode ceramah di kelas. Selain itu, belum diuji pada institusi atau mata kuliah lain secara luas.
--	--	------------------------------	--	---

2.3. Analisis GAP

Berdasarkan kajian penelitian terdahulu oleh Bagus Ramadhani dan Ryan Randi Suryono Tahun 2024 [9], analisis sentimen dilakukan dengan membandingkan algoritma Naive Bayes dan Logistic Regression pada opini masyarakat mengenai metaverse di media sosial X, dengan 6.728 data.

GAP penelitian ini terletak pada belum banyaknya kajian yang secara khusus menganalisis sentimen komentar YouTube terhadap pembentukan Program Koperasi Desa Merah Putih menggunakan perbandingan metode Naive Bayes dan Logistic Regression. Penelitian ini diharapkan dapat memberikan gambaran kecenderungan opini publik, sekaligus mengetahui algoritma yang lebih efektif dalam mengklasifikasikan sentimen masyarakat terhadap program koperasi desa tersebut. Adapun jumlah data yang digunakan 11.408 data dari media sosial YouTube.

Pada penelitian sebelumnya, platform X dipilih karena dianggap sebagai media sosial yang sering digunakan masyarakat untuk menyampaikan opini secara langsung, singkat, cepat, dan responsif terhadap isu teknologi [9].

Dengan demikian, penelitian tersebut berbeda dengan penelitian yang akan dilakukan. Penelitian tentang Koperasi Desa Merah Putih lebih tepat menggunakan komentar YouTube karena review awal program pemerintah disampaikan melalui video berita, podcast, atau diskusi publik di YouTube. Komentar YouTube dapat merepresentasikan tanggapan masyarakat secara kontekstual, terutama pada isu kebijakan publik dan program ekonomi desa.

BAB III

METODOLOGI PENELITIAN

3.1. Lokasi dan Waktu Penelitian

Penelitian ini dilakukan secara online dengan data komentar pengguna yang dikumpulkan dari platform YouTube. Untuk mengumpulkan komentar secara langsung dari halaman yang relevan, teknik web scraping digunakan untuk pengumpulan data. Data dikumpulkan dari 10 Juni 2026 hingga 4 Juli 2026. Setiap langkah penelitian mulai dari pengumpulan data, pengolahan data, hingga analisis dilakukan secara mandiri menggunakan perangkat komputer dan koneksi internet.

3.2. Alat

Software :

Windows 10
Google Collab
Microsoft Edge
Microsoft Word 2021
Microsoft Excel 2021
Python 3.11

Hardware :

Laptop MacBookAir7
Processor Intel Core i5 (1.80GHz)
Besar Memory 8 GB
Kapasitas Penyimpanan 128 GB

3.3. Variabel Penelitian

- a. Variabel Bebas (Independen): Komentar teks dari pengguna diplatform Youtube.
- b. Variabel Terikat (Dependen): Sentimen kepuasan pengguna (positif, negatif, dan netral).

3.4. Model yang Digunakan

Model yang digunakan untuk menganalisa sentimen pada penelitian ini adalah:

1. Model pertama, Naive Bayes merupakan algoritma probabilistik yang mengelompokkan sentimen komentar atau ulasan dengan menghitung peluang kemunculan setiap kata, sehingga opini dapat diprediksi sebagai positif, negatif, atau netral secara otomatis dan terarah.
2. Model kedua, Logistic Regression merupakan metode klasifikasi sentimen yang mengolah kombinasi fitur kata dalam komentar menjadi nilai probabilitas, kemudian menentukan kelas positif, negatif, atau netral berdasarkan peluang tertinggi, misalnya memakai fungsi sigmoid atau softmax sebagai dasar pengambilan keputusan model secara otomatis tepat.

3.5. Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan pada penelitian ini adalah:

Scrapping: Data komentar YouTube dikumpulkan melalui teknik web scrapping untuk memperoleh opini pengguna yang relevan sebagai bahan analisis sentimen dalam penelitian.

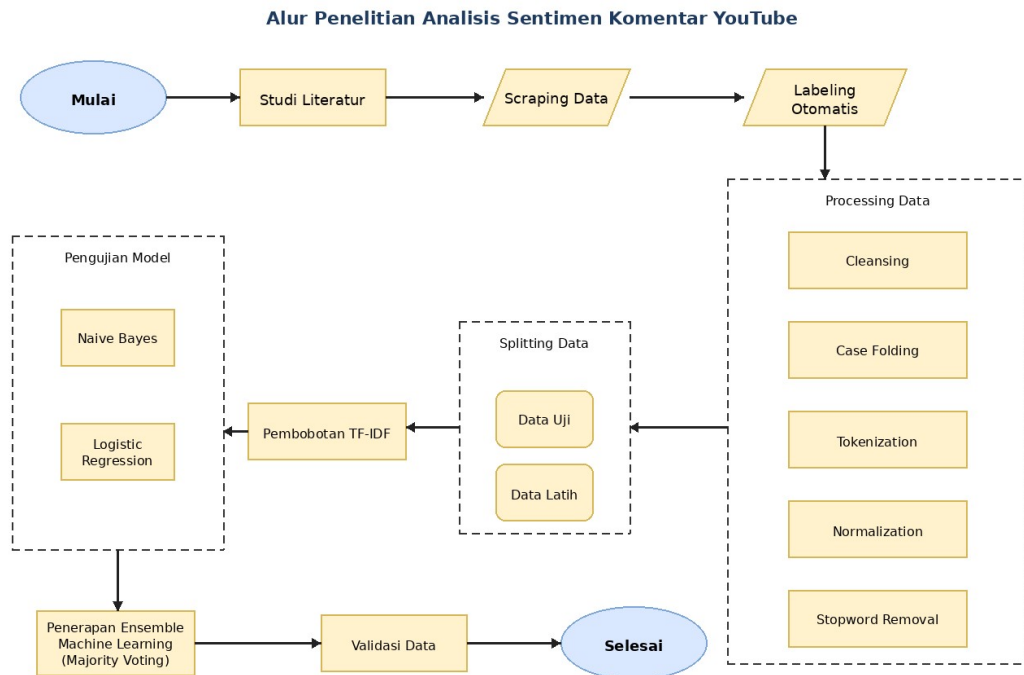
Preprocessing Data: Data dibersihkan melalui tahapan cleansing, case folding, tokenization, dan stopwords removal agar teks menjadi lebih rapi, seragam, serta siap dianalisis.

3.6. Metode Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis sentimen dengan dukungan machine learning agar proses pengolahan data opini dapat dilakukan secara sistematis, objektif, terukur, dan akurat.

3.7. Alur Penelitian

Alur penelitian analisis sentimen terhadap tanggapan awal masyarakat terhadap koperasi desa merah putih menggunakan Naïve Bayes dan Logistic Regression



Gambar 3.1 Rancangan Alur Penelitian

3.7.1 Studi Literatur

Pada tahap studi literatur, peneliti menelusuri dan mengumpulkan referensi relevan dari berbagai sumber, seperti jurnal, buku, artikel ilmiah, serta informasi daring untuk mendukung landasan teori penelitian secara sistematis dan terpercaya. Metode yang digunakan untuk menganalisa sentimen komentar Youtube tentang pembentukan koperasi desa merah putih pada penelitian ini adalah Naïve Bayes dan Logistic Regression.

Naïve Bayes adalah algoritma klasifikasi probabilistik berbasis Teorema Bayes yang memprediksi kelas data berdasarkan peluang kemunculan fitur dengan asumsi setiap fitur saling independen secara sederhana. Adapun rumus Naïve Bayes adalah [9]:

$$P(c/x) = \frac{P(c/x)P(c)}{P(x)}$$

Keterangan:

P : Probabilitas atau peluang terjadinya suatu kejadian

c : Kelas atau kategori yang akan diprediksi, misal positif, negatif atau netral.

x : Data atau fitur yang diamati, misal kata atau komentar yang akan diklasifikasikan

Adapun uraian dari masing masing komponen variabel pada rumus diatas adalah:

$P(c|x)$: Peluang data x termasuk ke dalam kelas c setelah data diamati.

$P(x|c)$: Peluang munculnya data/fitur x jika diketahui kelasnya adalah c. Disebut likelihood.

$P(c)$: Peluang awal suatu kelas c sebelum melihat data x. Disebut prior probability.

$P(x)$: Peluang munculnya data/fitur x secara umum. Disebut evidence

Sedangkan Logistis Regression merupakan algoritma klasifikasi probabilistik yang memodelkan hubungan fitur dan kelas menggunakan fungsi logistik untuk menghasilkan peluang prediksi pada kategori tertentu secara terukur otomatis. Adapun rumus Logistic Regression adalah [9]:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k)}}$$

Keterangan:

P : Nilai probabilitas hasil prediksi.

Y : Kelas atau hasil yang diprediksi. Biasanya berupa dua kategori, misalnya positif atau negatif, ya atau tidak, atau sentimen positif ataupun sentimen negatif.

e : Bilangan Euler dengan nilai sekitar 2,718. Digunakan untuk mengubah hasil regresi menjadi nilai probabilitas.

b : Koefisien regresi (regression coefficient) atau sering juga disebut parameter model.

x : Variable independen atau fitur, bisa juga disebut data masukan yang digunakan untuk memprediksi variabel Y.

k : Jumlah variabel (number of features or predictors), menunjukkan banyaknya variabel independen (X) yang digunakan dalam model Logistic Regression.

Adapun uraian dari masing masing komponen variabel pada rumus diatas adalah:

$P(Y)$: Probabilitas atau peluang terjadinya kelas/kejadian Y.

e: Bilangan Euler, nilainya sekitar 2,718.

b_0 : Konstanta atau intercept, yaitu nilai awal model.

$b_1, b_2, \dots, b_k$: Koefisien regresi untuk setiap variabel bebas.

$X_1, X_2, \dots, X_k$: Variabel bebas atau fitur yang digunakan untuk memprediksi Y.

k: Jumlah variabel bebas atau fitur dalam model.

3.7.2 Scrapping Data

Penelitian ini memperoleh data komentar terkait review awal pembentukan koperasi desa merah putih dari platform Youtube dengan menggunakan metode web scrapping melalui Google Collab. Data didapatkan dengan periode Maret 2026 sampai dengan Juni 2026, dengan total 11.408 ulasan yang disimpan dalam format XLSX.

Tabel 3.1 Hasil Scrapping Data

video_id	author	published_at	text	like_count
H5A5TvKztwM	@ariantiutomo2805	2026-02-10T21:05:19Z	Ujung2nya lahan korupsi lg....	182
H5A5TvKztwM	@sisonajeg517	2026-02-10T17:15:35Z	* Lokasi tak Setrategis * SDM tak siap * Harga Tak bisa bersaing	32
H5A5TvKztwM	@just-for-entertain798	2026-02-10T16:49:20Z	Koperasi nggak bisa dipaksa dari atas..!!!	27
H5A5TvKztwM	@Iskandar65-oe3pb	2026-02-10T16:47:18Z	Kopera si Indone sia itu harusn ya dari rakyat oleh rakyat dan untuk rakyat	77
H5A5TvKztwM	@wellas-ow6mo	2026-02-10T16:46:30Z	Masalah nya uang dari pajak rakyat... 😞😞😞	22

3.7.3 Labeling Otomatis

Data komentar terhadap pembentukan Koperasi Merah Putih dikategorikan ke dalam tiga kelas sentimen, yaitu negatif, netral, dan positif. Pada penelitian ini, proses *labeling* dilakukan secara otomatis menggunakan metode *Lexicon-Based Approach*, di mana nilai label diberikan berdasarkan kecenderungan makna atau opini yang terdapat dalam komentar setelah dicocokkan dengan kamus sentimen bahasa Indonesia [23].

Inisiasi penentuan label pada sistem klasifikasi ini didasarkan pada akumulasi skor akhir (*Skor_Total*) dari setiap kata bermakna yang ditemukan dalam satu komentar. Aturan (*rule-based*) pembobotan kata diinisiasikan sebagai berikut:

- a. Setiap kata yang tergolong dalam kamus positif diberi bobot +1.
- b. Setiap kata yang tergolong dalam kamus negatif diberi bobot -1.
- c. Setiap kata yang tidak terdaftar dalam kedua kamus (kata benda, kata depan, kata hubung) diberi bobot 0.

Berdasarkan akumulasi nilai tersebut, inisiasi fungsi keputusan untuk penulisan label otomatis diklasifikasikan ke dalam tiga kategori berikut:

1. Sentimen Negatif (Label -1): Diinisiasikan jika hasil *Skor_Total* < 0. Komentar ini menunjukkan penolakan, kritik, kekecewaan, atau ketidaksetujuan terhadap pembentukan Koperasi Merah Putih.

Contoh: "Pembentukan koperasi ini hanya akan memperluas celah korupsi (-1) dan menyengsarakan (-1) rakyat kecil." (Total Skor: -2 → Label -1).

2. Sentimen Netral (Label 0): Diinisiasikan jika hasil *Skor_Total* = 0. Komentar ini bersifat informatif, tidak memihak, tidak menunjukkan emosi tertentu, atau memiliki jumlah kata positif dan negatif yang seimbang sehingga saling menghilangkan.

Contoh: "Koperasi Merah Putih mengadakan rapat sosialisasi perdana pada hari Senin di aula kecamatan." (Total Skor: 0 → Label 0).

3. Sentimen Positif (Label 1): Diinisiasikan jika hasil Skor_Total>0. Komentar ini berisi dukungan, persetujuan, harapan baik, atau tanggapan positif terhadap pembentukan Koperasi Merah Putih.

Contoh: "Langkah yang sangat bagus (+1), semoga koperasi ini bisa membantu (+1) perekonomian warga." (Total Skor: +2 → Label 1).

Tabel 3.2 Hasil Labeling Otomatis

Comment	Label
Ujung2nya lahan korupsi lg....	-1
* Lokasi tak Setrategis * SDM tak siap * Harga Tak bisa bersaing	0
Koperasi nggak bisa dipaksa dari atas..!!!	1
Koperasi Indonesia itu harusnya dari rakyat oleh rakyat dan untuk rakyat	1
Masalah nya uang dari pajak rakyat...😞😞😞	0

3.7.4 Preprocessing Data

1. Case Folding

Tahapan ini bertujuan mengubah seluruh komentar YouTube ke dalam huruf kecil agar bentuk teks lebih seragam dan memudahkan proses analisis data selanjutnya secara lebih efektif.

Tabel 3.3 Preprocessing Teks menjadi Case folding

Comment	Label	case_folding
Ujung2nya lahan korupsi lg....	-1	ujung2nya lahan korupsi lg....
* Lokasi tak Setrategis * SDM tak siap * Harga Tak bisa bersaing	0	* lokasi tak setrategis * sdm tak siap * harga tak bisa bersaing
Koperasi nggak bisa dipaksa dari atas..!!!	1	koperasi nggak bisa dipaksa dari atas..!!!
Koperasi Indonesia itu harusnya dari rakyat oleh rakyat dan untuk rakyat	1	koperasi Indonesia itu harusnya dari rakyat oleh rakyat dan untuk rakyat
Masalah nya uang dari pajak rakyat...😞😞😞	0	masalah nya uang dari pajak rakyat...😞😞😞

2. Cleansing

Tahapan ini dilakukan untuk menghapus elemen yang tidak diperlukan dalam data teks, seperti URL, tanda baca, simbol, dan karakter lain yang dapat mengganggu proses analisis.

Tabel 3.4 Preprocessing Case Folding menjadi Cleansing

case_folding	cleansing
ujung2nya lahan korupsi lg....	ujungnya lahan korupsi lg
* lokasi tak setrategis * sdm tak siap * harga tak bisa bersaing	lokasi tak setrategis sdm tak siap harga tak bisa bersaing
koperasi nggak bisa dipaksa dari atas...!!!	koperasi nggak bisa dipaksa dari atas
koperasi Indonesia itu harusnya dari rakyat oleh rakyat dan untuk rakyat	koperasi indonesia itu harusnya dari rakyat oleh rakyat dan untuk rakyat
masalah nya uang dari pajak rakyat...😞😞😞	masalah nya uang dari pajak rakyat

3. Tokenization

Tahapan ini dilakukan untuk memecah teks atau kalimat menjadi bagian-bagian kecil yang disebut “token”, sehingga data lebih mudah diproses pada analisis lanjutan.

Tabel 3.5 Preprocessing Case Folding menjadi Tokenization

cleansing	tokenizing
ujungnya lahan korupsi lg	['ujungnya', 'lahan', 'korupsi', 'lg']
lokasi tak setrategis sdm tak siap harga tak bisa bersaing	['lokasi', 'tak', 'setrategis', 'sdm', 'tak', 'siap', 'harga', 'tak', 'bisa', 'bersaing']
koperasi nggak bisa dipaksa dari atas	['koperasi', 'nggak', 'bisa', 'dipaksa', 'dari', 'atas']
koperasi indonesia itu harusnya dari rakyat oleh rakyat dan untuk rakyat	['koperasi', 'indonesia', 'itu', 'harusnya', 'dari', 'rakyat', 'oleh', 'rakyat', 'dan', 'untuk', 'rakyat']
masalah nya uang dari pajak rakyat	['masalah', 'nya', 'uang', 'dari', 'pajak', 'rakyat']

4. Normalization

Tahap ini dilakukan untuk menyeragamkan bentuk kata dengan mengurangi variasi penulisan yang berbeda namun bermakna sama, sehingga data menjadi konsisten dan mudah diproses lebih akurat.

Tabel 3.6 Preprocessing Tokenizaztion menjadi Normalization

tokenizing	normalization
['ujungny', 'lahan', 'korupsi', 'lg']	['akhirnya', 'lahan', 'korupsi', 'lagi']
['lokasi', 'tak', 'setrategis', 'sdm', 'tak', 'siap', 'harga', 'tak', 'bisa', 'bersaing']	['lokasi', 'tidak', 'strategis', 'sdm', 'tidak', 'siap', 'harga', 'tidak', 'bisa', 'bersaing']
['koperasi', 'nggak', 'bisa', 'dipaksa', 'dari', 'atas']	['koperasi', 'tidak', 'bisa', 'dipaksa', 'dari', 'atas']
['koperasi', 'indonesia', 'itu', 'harusnya', 'dari', 'rakyat', 'oleh', 'rakyat', 'dan', 'untuk', 'rakyat']	['koperasi', 'indonesia', 'itu', 'seharusnya', 'dari', 'rakyat', 'oleh', 'rakyat', 'dan', 'untuk', 'rakyat']
['masalah', 'nya', 'uang', 'dari', 'pajak', 'rakyat']	['masalah', 'uang', 'dari', 'pajak', 'rakyat']

5. Stopword Removal

Proses ini adalah tahapan preprocessing yang dilakukan untuk menghilangkan kata-kata umum yang kurang memiliki makna penting dalam analisis, seperti dan, atau, di, dari, ke, yang, serta kata penghubung lainnya.

Tabel 3.7 Preprocessing Normalization menjadi Stopword Removal

tokenizing	stopword_removal
['akhirnya', 'lahan', 'korupsi', 'lagi']	['akhirnya', 'lahan', 'korupsi', 'lagi']
['lokasi', 'tidak', 'strategis', 'sdm', 'tidak', 'siap', 'harga', 'tidak', 'bisa', 'bersaing']	['lokasi', 'tidak', 'strategis', 'sdm', 'tidak', 'siap', 'harga', 'tidak', 'bisa', 'bersaing']
['koperasi', 'tidak', 'bisa', 'dipaksa', 'dari', 'atas']	['koperasi', 'tidak', 'bisa', 'dipaksa', 'atas']
['koperasi', 'indonesia', 'itu', 'seharusnya', 'dari', 'rakyat', 'oleh', 'rakyat', 'dan', 'untuk', 'rakyat']	['koperasi', 'indonesia', 'seharusnya', 'rakyat', 'rakyat', 'rakyat']
['masalah', 'uang', 'dari', 'pajak', 'rakyat']	['masalah', 'uang', 'pajak', 'rakyat']

2

JADWAL PENELITIAN

Tabel 3.8. Jadwal Penelitian

No	Kegiatan	Waktu				
		1	2	3	4	5
1	Penulisan proposal					
2	Seminar dan bimbingan proposal					
3	Penelitian dan tindakan					
4	Analisis dan bimbingan hasil penelitian					
5	Ujian skripsi					

DAFTAR PUSTAKA

- [1] P. Arthayoga, L. Fanggi, and B. Evangelista, "Kedudukan Hukum Koperasi Merah Putih," vol. 5, no. 1, pp. 1367–1378, 2025.
- [2] P. S. Manajemen, F. Ekonomi, and U. Muhammadiyah, "TRANSFORMASI EKONOMI DESA MELALUI KOPERASI MERAH," vol. 12, 2025.
- [3] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," vol. 15, no. 1, pp. 131–145.
- [4] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," vol. IV, no. September, pp. 15–21, 2020.
- [5] J. M. Polgan *et al.*, "Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial," vol. 13, pp. 2020–2025, 2024.
- [6] A. F. Watratan, A. P. B, D. Moeis, S. Informasi, and S. P. Makassar, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," vol. 1, no. 1, pp. 7–14, 2020.
- [7] S. H. Junaidi, T. Sjah, and K. Budastra, "Mungkinkah Koperasi Merah Putih Berhasil," vol. 4, no. November, pp. 422–434, 2025.
- [8] V. N. Mei, R. Juanda, T. U. Sabilla, W. Kurniawati, and C. Fransisca, "Implementasi Program Koperasi Merah Putih dalam Mendorong Kemandirian Ekonomi Desa manfaat yang merata bagi masyarakat perdesaan (Nugroho & Setyowati , 2023).," no. April, 2026.
- [9] B. Ramadhani and R. R. Suryono, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," vol. 8, no. April, pp. 714–725, 2024, doi: 10.30865/mib.v8i2.7458.
- [10] T. Utary *et al.*, "Komparasi Support Vector Machine dengan Logistic Regression terhadap klasifikasi pada data asma," vol. 9, no. 1, pp. 1–8, 2026.
- [11] M. Monica and A. Purwanto, "Comparative Performance Evaluation of Naïve Bayes and Logistic Regression for Indonesian YouTube Comment Sentiment Classification Evaluasi Komparatif Kinerja Algoritma Naïve Bayes dan Logistic Regression dalam Klasifikasi Sentimen Komentar YouTube Berbahasa Indonesia," vol. 6, no. April, pp. 934–943, 2026.
- [12] A. S. Wijaya, A. Fahmi, Y. Tyas, and C. Pramudi, "Sentiment Classification of Health Education YouTube Comments Using IndoBERT Embeddings with Logistic Regression and Naïve Bayes," vol. 10, no. 3, pp. 3005–3020, 2026.
- [13] A. A. Ningtyas, A. Solichin, and R. Pradana, "ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PREDIKSI RESESI EKONOMI TAHUN 2023 MENGGUNAKAN ALGORITME SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON PREDICTION OF ECONOMIC RECESSION IN 2023 USING THE NAÏVE BAYES," vol. 20, no. 1, pp. 9–16, 2023.
- [14] C. A. Misrun, E. Haerani, M. Fikry, and E. Budianita, "Jurnal Computer

- Science and Information Technology (CoSciTech) naive bayes classifier method,” vol. 4, no. 1, pp. 207–215, 2024.
- [15] S. F. Huwaida, R. Kusumawati, and B. Isnaini, “Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode Naïve Bayes,” vol. 6, no. 1, pp. 26–39, 2024, doi: 10.37905/jji.v6i1.24718.
 - [16] D. Prastyo, I. H. Mursyidin, and D. Irawan, “Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten,” vol. 7, no. 2, 2024, doi: 10.32877/bt.v7i2.1833.
 - [17] A. Maulana, D. Dyantono, and R. E. Putra, “Perbandingan Sent2vec TF-IDF Logistic Regression dan Word2vec CNN pada hasil Sentiment Analysis Youtube Comment,” vol. 05, pp. 63–72, 2023.
 - [18] H. S. Mulyono and U. Saprudin, “Efektivitas Logistic Regression dalam Analisis Sentimen Berbahasa Indonesia pada Komentar YouTube tentang Isu,” vol. 6, no. 3, pp. 1547–1555, 2025.
 - [19] D. Marganingsih, H. Oktavianto, and G. Abdurrahman, “Analisis Sentimen Komentar Youtube Masterchef Indonesia Menggunakan Algoritma Support Vector Machine Dan Gaussian Naïve Bayes Sentiment Analysis of YouTube Comments on Masterchef Indonesia Using Support Vector Machine and Gaussian Naïve Bayes Algorithms,” vol. 5, no. 1, pp. 16–26, 2025, doi: 10.59395/jitp.v5i1.117.
 - [20] R. B. Aprianto, A. Eviyanti, H. Setiawan, I. Ratna, and I. Astutik, “DEDI MULYADI DENGAN METODE MACHINE LEARNING (STUDI KASUS : KENAKALAN REMAJA),” vol. 10, no. 2, pp. 2997–3004, 2026.
 - [21] Z. Y. Burnama, M. A. Rosid, and N. L. Azizah, “Analisis Sentimen Pada Komentar Youtube Dalam Turnamen MPL Season 13 Dengan Metode Ensemble Machine Learning Sentiment Analysis on YouTube Comments in MPL Season 13 Tournament Using Ensemble Machine Learning Method”.
 - [22] A. Eviyanti and N. Ariyanti, “ANALYSIS OF APRIORI AND K-NEAREST NEIGHBOR (KNN) ALGORITHM IN RECOMMENDING APPROPRIATE LEARNING,” vol. 20, no. 1, pp. 557–572, 2026.
 - [23] V. N. Juli, A. Jl, R. Madya, G. Anyar, K. Gn, and J. Timur, “Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM Muhammad Fernanda Naufal Fathoni Eva Yulia Puspaningrum Andreas Nugroho Sihananto menghasilkan F1 sebesar 80 %. Penelitian lain dilakukan (Muhammadi et al ., 2022) dengan VADER sehingga menjadikan peneliti tertarik membuat perbandingan kombinasi lexicon InSet dan VADER dengan SVM pada analisa sentimen Rohingya sosial media X untuk mengetahui manakah performa yang lebih baik .,” vol. 1, no. 3, 2024.