ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS (MBG) PADA PLATFORM X DENGAN MENGGUNAKAN METODE NAIVE BAYES

Program Studi

S1 Sistem Informasi

Oleh:

MUHAMMAD EGALINGGA ZAINURI

21410100027

FAKULTAS TEKNOLOGI DAN INFORMATIKA

UNIVERSITAS DINAMIKA

2026

ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS (MBG) PADA PLATFORM X DENGAN MENGGUNAKAN METODE NAIVE BAYES

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk menyelesaikan
Program Sarjana

Oleh:

Nama	**: Muh. Egalingga Zainuri**
NIM	**: 21410100027**
Program Studi	**: S1 Sistem Informasi**

FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA
2026

TUGAS AKHIR

ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS (MBG) PADA PLATFORM X DENGAN MENGGUNAKAN METODE NAIVE BAYES

Dipersiapkan dan disusun Oleh

Muhammad Egalingga Zainuri

NIM: 21410100027

Telah diperiksa, dibahas dan disetujui oleh Dewan Pembahas

Pada: 10, Februari 2026

Susunan Dewan Pembahas

Pembimbing

I. I Gusti Ngurah Alit Widana Putra, S.T., M.Eng.
NIDN. 0805058602

Digitally signed by I Gusti Ngurah Alit WIdana Putra
Date: 2026.02.10 15:23:05 +07'00'

II. Sulistiowati, S.Si., M.M.
NIDN. 0719016801

Pembahas

I. Julianto Lemantara, S.Kom., M.Eng.
NIDN. 0722108601

Digitally signed by Julianto Lemantara
Date: 2026.02.12 18:06:19 +07'00'

Tugas Akhir ini telah diterima sebagai salah satu persyaratan

Untuk memperoleh gelar sarjana

Fakultas Teknologi dan Informatika
UNIVERSITAS
Dinamika

Tan Amelia, S.Kom., M.MT., Ph.D.

NIDN. 0728017602

Dekan Fakultas Teknologi dan Informatika

UNIVERSITAS DINAMIKA

"It always seems impossible until it's done"

-Nelson Mandela

Tugas Akhir ini

Saya persembahkan kepada Orang Tua, Keluarga Besar,

Dosen Pembimbing, Dosen Wali,

dan Teman-teman

PERNYATAAN

PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH

Sebagai mahasiswa **Universitas Dinamika**, Saya :

Nama	: **Muhammad Egalingga Zainuri**
NIM	: **21410100027**
Program Studi	: **S1 Sistem Informasi**
Fakultas	: **Teknologi dan Informatika**
Jenis Karya	: **Laporan Tugas Akhir**
Judul Karya	: **ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS (MBG) PADA PLATFORM X DENGAN MENGGUNAKAN METODE NAIVE BAYES**

Menyatakan dengan sesungguhnya bahwa :

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, Saya menyetujui memberikan kepada **Universitas Dinamika** Hak Bebas Royalti Non-Eksklusif (***Non-Exclusive Royalty Free Right***) atas seluruh isi/sebagian karya ilmiah Saya tersebut diatas untuk disimpan, dialihmediakan, dan dikelola dalam bentuk pangkalan data (***database***) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama Saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.
2. Karya tersebut diatas adalah hasil karya asli Saya, bukan plagiat baik sebagian maupun keseluruhan. Kutipan, karya, atau pendapat orang lain yang ada dalam karya ilmiah ini semata-mata hanya sebagai rujukan yang dicantumkan dalam Daftar Pustaka Saya.
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiasi pada karya ilmiah ini, maka Saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada Saya.

Demikan surat pernyataan ini Saya buat dengan sebenar-benarnya.

Surabaya, 6 Januari 2026

SEPULUH RIBU RUPIAH 10000 METERAI TEMPEL 93ANX188765554

Muhammad Egalingga Zainuri
NIM : 21410100027

ABSTRAK

Program Makan Bergizi Gratis (MBG) merupakan kebijakan strategis pemerintah untuk meningkatkan kualitas gizi nasional yang memicu pertumbuhan berbagai sentimen publik berdasarkan faktor terhadap kesehatan generasi masa depan serta kekhawatiran terkait insiden keracunan pangan dan transparansi anggaran. Analisis sentimen dilakukan sebagai instrumen evaluasi kebijakan untuk memetakan persepsi masyarakat secara objektif dan terukur agar masukan publik dapat menjadi dasar perbaikan program. Dalam analisis ini, algoritma *Naive Bayes* berperan sebagai pengklasifikasi teks, sedangkan teknik ADASYN berfungsi untuk menyeimbangkan distribusi data. Hasil penelitian menunjukkan Dari total 2.351 data, sebanyak 76,2% mencerminkan apresiasi dan dukungan masyarakat terhadap pelaksanaan program MBG. Sementara itu, 23,8% sentimen bersifat negatif akibat kekhawatiran publik terhadap kualitas pangan dan distribusi MBG. Hasil evaluasi performa model mendapatkan nilai optimal dengan nilai *Accuracy* sebesar 86%. Penggunaan ADASYN berhasil menangani faktor *imbalance data* sehingga secara efektif meningkatkan sensitivitas model terhadap kelas minoritas (negatif) dengan nilai *recall* dari 42% naik menjadi 83% dan *F-1 score* dari 57% naik menjadi 67%.

Kata Kunci : *Analisis Sentimen, Makan bergizi gratis (MBG), Naive bayes, ADASYN, Paltform X*

KATA PENGANTAR

Puji Syukur ke hadirat Allah SWT yang telah memberikan rahmat dan hidayahnya sehingga penulis dapa menyelesaikan Laporan Tugas Akhir berjudul "Analisis Sentimen Program Makan Bergizi gratis (MBG) pada platform X dengan metode Naive Bayes".

Penyelesaian Laporan Tugas Akhir ini tidak lepas dari bantuan berbagai pihak yang telah memberikan nasihat, kritik, saran dan masukan serta dukungan moral maupun materil kepada penulis. Oleh karena itu Penulis menyampaikan rasa terima kasih atas bantuannya kepada :

1. Orang tua dan keluarga tercinta yang telah memberikan doa, semangat dan dukungan sepenuhnya kepada penulis.
2. Bapak I Gusti Ngurah Alit Widana Putra, S.T., M.Eng. selaku Dosen Pembimbing 1 yang telah memberikan saran dan masukan dalam serta bimbingan dalam menyelesaikan Laporan ini.
3. Ibu Sulistiowati, S.Si., M.M. selaku Dosen Pembimbing 2 yang telah memberikan saran dan masukan dalam serta bimbingan dalam menyelesaikan Laporan ini.
4. Bapak Julianto Lemantara, S.Kom., M.Eng. yang telah menguji dan memberikan masukan dan saran dalam pengerjaan Tugas Akhir ini
5. Teman-teman seperjuangan angkatan 21 yang telah memberikan motivasi dan dorongan untuk menyelesaikan Laporan Tugas Akhir ini.

Penulis menyadari bahwa laporan ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari pembaca. Akhir kata, semoga laporan Tugas Akhir ini dapat memberikan manfaat bagi banyak pihak.

Surabaya, 10 Februari 2026

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

Halaman

DAFTAR LAMPIRAN

Halaman

BAB I
PENDAHULUAN

1.1 Latar Belakang

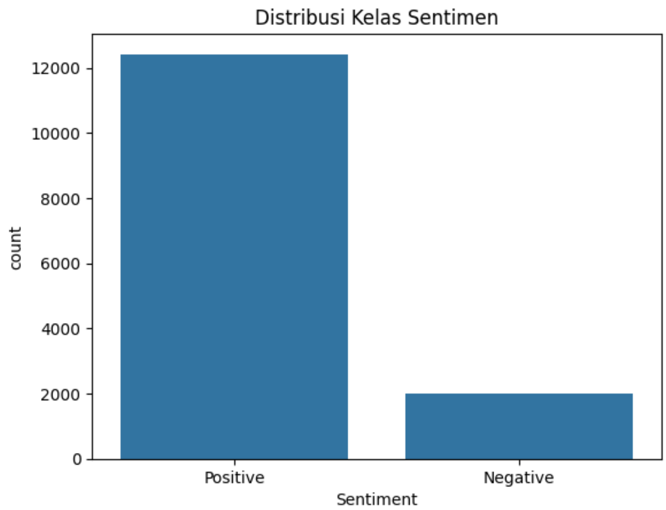
Pergantian pemerintahan di Indonesia sering memunculkan berbagai kebijakan dan program baru yang berfokus pada peningkatan kesejahteraan masyarakat. Program-program prioritas tersebut umumnya dirancang agar memiliki dampak langsung bagi masyarakat luas. Salah satu isu penting yang selalu menjadi perhatian adalah pemenuhan gizi, karena gizi yang baik tidak hanya berpengaruh pada kesehatan individu, tetapi juga pada kualitas sumber daya manusia di masa depan (Victora et al., 2008). Dalam konteks inilah, pemerintah memperkenalkan Program Makan Bergizi Gratis (MBG) yang bertujuan untuk menurunkan angka gizi buruk serta memperbaiki kualitas hidup masyarakat di wilayah tertinggal dan termiskin kelompok yang selama ini kerap luput dari jangkauan layanan dasar negara (Agustini, 2025).

Berdasarkan Peraturan Presiden Nomor 83 Tahun 2024 tentang Badan Gizi Nasional (BGN), pemerintah menegaskan penguatan kelembagaan dan koordinasi penyelenggaraan program nasional di bidang gizi. Implementasi teknis penyediaan dan distribusi makanan bagi penerima manfaat dipertegas melalui Surat Keputusan Deputi Bidang Penyaluran BGN Nomor 2 Tahun 2024 sebagai dasar operasional pelaksanaan program di lapangan. Program Makan Bergizi Gratis (MBG) merupakan kebijakan strategis pemerintah yang bertujuan untuk menyediakan makanan sehat dan bergizi bagi peserta didik serta kelompok masyarakat rentan dalam rangka menekan angka *stunting*, meningkatkan kesehatan, dan mendukung peningkatan kualitas sumber daya manusia Indonesia di masa depan. Selain itu, pelaksanaan MBG juga melibatkan UMKM sektor pangan sebagai penyedia bahan dan pengolah makanan, sehingga tidak hanya berfungsi meningkatkan ketahanan gizi tetapi juga memperkuat ekonomi lokal secara berkelanjutan (Azzahra T, 2025). Program ini mulai dilaksanakan secara serentak di seluruh provinsi pada 6 Januari 2025, sebagai langkah nyata pemerintah dalam memperkuat ketahanan gizi nasional.

Meskipun Program Makan Bergizi Gratis (MBG) bertujuan untuk membantu kualitas kesehatan Indonesia, pelaksanaannya menghadapi masalah yang memicu kekhawatiran publik. Misalnya, menurut laporan Kompas, telah terjadi kasus keracunan makanan di beberapa sekolah sejak awal 2025, dengan dugaan penyebab seperti penyimpanan makanan yang tidak sesuai temperatur, serta masakan yang disiapkan terlalu awal (Herdanto, 2025). Selain itu, DPR dan ahli gizi menyoroti bahwa sejumlah Satuan Pelayanan Pemenuhan Gizi (SPPG) tidak menjalankan fungsi pengawasan gizi secara optimal ini menjadi faktor dalam insiden keracunan tersebut (E-Media DPR RI, 2025). Di sisi anggaran, kritik datang dari pengamat yang menilai bahwa besarnya alokasi (Rp 71 triliun) belum diimbangi dengan efisiensi dan pengawasan yang memadai (Munir, 2025). Cepatnya ekspansi program tanpa pengelolaan lokal yang matang juga dikhawatirkan menyebabkan pemborosan anggaran atau distribusi yang tidak tepat sasaran (termasuk untuk daerah paling rentan) (Hakim, 2025).

Karena berbagai permasalahan yang muncul selama pelaksanaan Program Makan Bergizi Gratis (MBG) muncullah beragam opini publik di media sosial, terutama di Platform X. Masyarakat menyampaikan apresiasi, kritik, keluhan, maupun kekhawatiran secara terbuka melalui unggahan dan komentar. Kondisi ini menunjukkan bahwa kebijakan MBG tidak hanya dinilai dari regulasi dan mekanismenya, tetapi juga dari pengalaman langsung masyarakat sebagai penerima atau pengamat program di lapangan. Oleh karena itu, analisis sentimen menjadi penting untuk memetakan bagaimana publik merespons kebijakan tersebut, mengidentifikasi kritik apa yang paling dominan, serta memahami bagian mana dari program yang dianggap berhasil atau membutuhkan perbaikan. Temuan ini dapat menjadi masukan yang berharga bagi pemerintah dalam memperbaiki komunikasi publik, meningkatkan kualitas pelaksanaan, dan memastikan bahwa tujuan program benar-benar tercapai.

Naive Bayes Classifier menawarkan keunggulan dalam hal kesederhanaan, efisiensi, dan performa yang cukup baik untuk teks pendek seperti cuitan di Platform X. Berdasarkan penelitian (Dwi Anggreny, 2025) Algoritma Naive Bayes merupakan metode klasifikasi teks yang efektif karena bekerja dengan menghitung probabilitas kemunculan kata dalam suatu dokumen untuk menentukan kecenderungan sentimen, baik positif maupun negatif.



Gambar 1.1 Rasio Sentimen Positif dan Negatif

Berdasarkan gambar menunjukkan data didominasi oleh sentimen Positif sehingga mengakibatkan imbalance data. Untuk mengatasi ini diperlukan teknik Oversampling mencegah faktor imbalance data agar tidak bias ke kelas Mayoritas. ADASYN menghasilkan sampel sintetis secara adaptif pada titik-titik kelas minoritas yang paling sulit dipelajari, sehingga model memperoleh representasi yang lebih baik di area keputusan yang kompleks. Pendekatan ini meningkatkan kemampuan generalisasi dan recall pada kelas minoritas, sekaligus mengurangi bias terhadap kelas mayoritas dalam data sentimen sosial media (Zakariah et al., 2023). Dari penelitian Nurhopipah & Magnolia (2022) teknik Oversampling data dengan ADASYN mampu meningkatkan Akurasi model secara signifikan.

Dengan memanfaatkan Naive Bayes serta ADASYN, penelitian ini dapat memberikan solusi untuk memetakan sentimen masyarakat terhadap Program Makan Bergizi Gratis secara cepat, akurat, dan terukur, sehingga hasilnya dapat menjadi masukan berharga bagi evaluasi dan pengembangan kebijakan publik.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah dari proposal Tugas Akhir (TA) ini yaitu :

1. Bagaimana persepsi masyarakat di media sosial *Platform* X terhadap Program Makan Bergizi Gratis yang dicanangkan pemerintah ?
2. Bagaimana model Naive Bayes Classifier dan ADASYN Menganalisis sentimen positif, atau negative masyarakat terhadap Program Makan Bergizi Gratis pada *Platform* X ?

1.3 Batasan Masalah

Adapun batasan masalah dalam melakukan penelitian ini adalah sebagai berikut:

1. Data merupakan cuitan berbahasa Indonesia
2. Data yang digunakan merupakan cuitan dari tanggal 30 Juli 2025 sampai dengan 10 Oktober 2025
3. Sentimen hanya berupa sentimen positif dan negatif.
4. Pelabelan data hanya menggunakan Lexicon.

1.4 Tujuan

Berdasarkan uraian rumusan masalah, maka tujuan yang dicapai adalah :

1. Untuk memetakan persepsi masyarakat di *Platform* X terhadap Program Makan Bergizi Gratis (MBG) yang dicanangkan pemerintah.
2. Untuk menerapkan metode Naive Bayes Classifier dan ADASYN dalam melakukan analisis sentimen positif atau negatif masyarakat terhadap Program Makan Bergizi Gratis pada *Platform* X.

1.5 Manfaat

Dari penelitian ini dapat memberikan manfaat :

1. Sebagain acuan penelitian dalam penerapan Naïve Bayes dalam menganalisis sentiment public
2. Referensi untuk penelitian serupa atau pengembangan lebih lanjut
3. Menjadi masukan terhadap program pemerintahan berdasarkan opini public di media sosial.

BAB II
LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian terdahulu menjadi dasar penting dalam pengembangan penelitian ini. Sejumlah studi relevan dijadikan rujukan oleh penulis untuk memperluas serta memperdalam kajian yang dilakukan.

Tabel 2.1 Tabel Penelitian Terdahulu

No	Judul	Metode Penelitian	Hasil	Akurasi
1	ANALISIS MULTIDIMENSIONAL SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN BERGIZI GRATIS PADA MEDIA SOSIAL X	Text Mining, KNIME	Hasil akurasi dari kalsifikasi Naive Bayes menunjukkan bahwa Sentimen masyarakat terhadap MBG didominasi oleh sentimen negatif sebanyak 54.89%	Hasil pengolongan sentimen Positif 17,29%, Negatif 54.89% dan Netral 27.82%.
	Oleh : Putriyekti et al. (2025)			
	Perbedaan : Penelitian ini bertujuan untuk menggolongkan sentimen Positif, Negatif dan netral lalu memahami aspek yang mempengaruhi sentimen tersebut dengan menggunakan *Platform* KNIME.			
2	Sentiment Analysisi of Free Meal Program for School Students Using Algoritma Naive Bayes	Naive Bayes	Berdasalkan hasil dari analisis dari X dan data dari Sekolah, Naive bayes menunjukkan performa yang baik dari kedua data, namun data dari X menunjukkan hasil yang lebih baik .	Data X : 91%, Data Sekolah : 83.33%
	Oleh : Dwi Anggreny (2025)			
	Perbedaan : Penelitian ini membandingkan Akurasi metode Naive Bayes dalam menganalisis Sentimen dari Data X (Cuitan dari twitter) dan Data sekolah (Intervew di tempat dan Kuisioner)			
3	Evaluasi algoritma KNN dan Naive Bayes untuk analisis sentimen kebijakan program makan bergizi gratis	Naive Bayes, KNN	Perbandingan dari Metode Naive bayes dan KNN dalam analisi sentimen publik terhadap MBG menunjukkan Naive bayes memiliki kinerja lebih baik	Naive Bayes : Accuracy 79.61%, F1-score 77.33%. KNN : 73.30%, F1-score 67.37%
	Oleh : Sakina et al. (2025)			
	Perbedaan :			

No	Judul	Metode Penelitian	Hasil	Akurasi
	Penelitian ini membandingkan Metode Naive Bayes dengan K-Nearest Neighbor dalam menganalisi sentimen program Makan bergizi gratis.			
4	Sentiment Analysis of Jobstreet Application Reviews on Google Play Store Using Support Vector Machine Algorithm with Adaptive Synthetic	SVM, ADASYN	Hasil Penelitian Analisis sentimen dengan SVM dan ADASYN menunjukkan peningkatan nilai confusion matriks	Sebelum ADASYN : Accuracy 89.08%, setelah ADASYN 89.95%
	Oleh : (Shantika & Abidin, 2025)			
	Perbedaan : Penelitian ini menganalisis sentimen pada review Jobstreet Aplication pada Playstore dengan membandingkan Performa SVM dengan ADASYN dalam meningkatkan kemampuan analisis data.			

2.2 Platform X

Platform X (dahulu Twitter) merupakan media sosial *microblogging* yang memfasilitasi penyampaian opini secara singkat dan cepat melalui fitur tweet, *retweet*, dan *hashtag*. Sifat pesan yang ringkas serta alur informasi yang berlangsung *real-time* menjadikan Platform X sebagai ruang publik digital yang strategis untuk menyampaikan pandangan, berdiskusi, dan membentuk opini publik terhadap isu sosial. Menurut Thoriq et al. (2021) Twitter memiliki potensi besar sebagai sumber data untuk analisis sentimen karena keterbukaan dan volume datanya yang tinggi.

Selain sebagai sarana komunikasi personal, Platform X juga berperan penting dalam ranah sosial dan politik. Faqih Azhar et al. (2025) menjelaskan bahwa penggunaan Twitter meningkatkan partisipasi masyarakat dalam diskusi kebijakan publik. Media sosial ini menjadi kanal komunikasi kebijakan yang efektif antara pemerintah dan masyarakat.

2.3 Program Makan Bergizi gratis (MBG)

Program Makan Bergizi Gratis (MBG) adalah kebijakan pemerintah yang menyediakan makanan bergizi bagi siswa sekolah maupun ibu hamil, dengan tujuan meningkatkan status gizi, mengurangi malnutrisi dan stunting, serta mendukung kualitas pendidikan dan kesehatan masyarakat. Program ini diluncurkan secara resmi Januari 2025, menggunakan alokasi anggaran negara, mekanisme distribusi melalui dapur-dapur khusus (SPPG) dan penyusunan menu bergizi agar mencukupi kebutuhan gizi dasar (Ga' et al., 2025).

2.4 Analisis Sentimen

Analisis sentimen merupakan proses mengidentifikasi dan mengklasifikasikan opini, emosi, atau sikap seseorang terhadap suatu objek ke dalam kategori positif atau negatif berdasarkan data teks. Tujuan utamanya adalah memahami persepsi publik terhadap isu tertentu, baik dalam konteks kebijakan, produk, maupun layanan. Menurut Satya Marga et al. (2021) analisis sentimen berperan penting dalam menilai opini masyarakat secara objektif melalui data media sosial seperti Twitter atau X. Selain itu, Hasibuan & Serdano (2022) menjelaskan bahwa Pendekatan machine learning dalam analisis sentimen memungkinkan pengolahan data teks berukuran besar secara efisien dengan tingkat akurasi yang tinggi, terutama dalam mengidentifikasi tanggapan masyarakat.

2.5 *Text Mining*

Text Mining merupakan proses mengekstraksi informasi atau pola pengetahuan dari data teks yang tidak terstruktur dengan bantuan teknik komputasi seperti *Natural Language Processing* (NLP) dan *Machine Learning*. Proses ini meliputi tahapan *text preprocessing* seperti *tokenizing*, *stopword removal*, dan *stemming*, yang kemudian dilanjutkan dengan transformasi teks menjadi data numerik untuk analisis lebih lanjut, misalnya klasifikasi atau analisis sentimen (Fauziyyah, 2020).

Menurut (Amrullah et al., 2020), Text Mining berperan penting dalam mengolah opini masyarakat di media sosial karena mampu mengidentifikasi kecenderungan sentimen publik secara cepat dan akurat, sekaligus menjadi dasar pengambilan keputusan berbasis data dalam berbagai bidang seperti kebijakan publik, pemasaran, dan riset sosial.

2.6 *Lexicon Based*

Pelabelan berbasis leksikon merupakan metode komputasional untuk mengidentifikasi sentimen teks secara otomatis dengan memanfaatkan kamus kata yang telah memiliki bobot polaritas positif, atau negatif. Setiap kata dalam teks dicocokkan dengan leksikon kemudian skor polaritasnya diakumulasi untuk menentukan label sentimen akhir dokumen (Novfuja et al., 2025).

Pendekatan leksikon efektif mengatasi keterbatasan data berlabel dalam penelitian machine learning karena lebih objektif dan efisien untuk dataset berskala besar dibandingkan pelabelan manual yang memakan waktu dan rentan terhadap bias subjektif. (Hill et al., 2025).

$$Total\ Score = \sum (Skor\ Kata_i) \tag{1}$$

Ketentuan :

Label Positif jika Total Score ≥ 0

Label Negatif jika Total Score < 0

2.7 *Term Frequence – Inverse Document Frequency (TF-IDF)*

Pembobotan TF-IDF digunakan untuk menentukan seberapa besar tingkat kepentingan setiap kata dalam sebuah dokumen. Melalui metode ini, teks dapat diubah menjadi representasi numerik sehingga dapat diolah secara komputasional. Pendekatan TF-IDF juga sering membantu meningkatkan ketepatan dalam proses analisis teks. Nilai bobot pada TF-IDF diperoleh dari kombinasi dua komponen utama, yaitu *Term Frequency* (TF) sebagai ukuran frekuensi kemunculan kata dan *Inverse Document Frequency* (IDF) sebagai ukuran tingkat kelangkaannya dalam keseluruhan dokumen.

2.7.1 TF

Mengukur seberapa sering suatu kata (*term*) muncul dalam satu dokumen. Jika sebuah kata sering muncul dalam sebuah dokumen, maka kata itu dianggap lebih penting untuk dokumen tersebut.

$$tf(t,d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}} \tag{2}$$

Keterangan :

t = Kata tertentu

d = dokumen tertentu

$f_{t,d}$ = Jumlah kemunculan kata (t) dalam dokumen (d)

$\sum_{t'} f_{t',d}$ = Total keseluruhan kata dalam dokumen

$tf(t,d)$ = nilai TF untuk kata (t) pada dokumen (d)

2.7.2 IDF

Mengukur betapa pentingnya suatu kata dalam seluruh kumpulan dokumen (corpus). Jika sebuah kata muncul di banyak dokumen, IDF-nya rendah (kata dianggap umum). Jika sebuah kata jarang muncul di corpus, IDF-nya tinggi (lebih informatif)

$$idf(t,D) = log\frac{N}{n_t} \tag{3}$$

Keterangan :

t = Kata tertentu

$D =$ Kumpulan seluruh dokumen

N = Total jumlah dokumen dalam Korpus (D)

n_t = Total dokumen yang mengandung kata (t)

log = Logaritma, digunakan untuk meredam skala

$idf(t,D)$ = Nilai IDF untuk Kata (t) pada Korpus (D)

2.7.3 TF-IDF

Menggabungkan TF dan IDF untuk menghasilkan bobot yang menunjukkan Seberapa penting sebuah kata dalam sebuah dokumen dibandingkan dengan dokumen lain. Kata yang sering muncul di dokumen tersebut (TF tinggi) tetapi jarang di dokumen lain (IDF tinggi) menandakan bobot TF-IDF paling tinggi.

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \tag{4}$$

Keterangan :

t = Kata tertentu

$d =$ dokumen tertentu

$D =$ Kumpulan seluruh dokumen

$tf(t, d)$ = nilai TF untuk kata (t) pada dokumen (d)

$idf(t, D)$ = Nilai IDF untuk Kata (t) pada Korpus (D)

$tfidf(t, d, D)$ = Bobot akhir TF-IDF

2.8 *Adaptive Synthetic Sampling (ADASYN)*

Adaptive Synthetic Sampling (ADASYN) adalah teknik *oversampling* yang dikembangkan untuk mengatasi masalah ketidakseimbangan kelas dalam dataset. Metode ini secara adaptif membuat data sintetis baru pada kelas minoritas, terutama di sekitar titik-titik data yang sulit dipelajari oleh model. Berbeda dengan metode seperti SMOTE yang menambah data secara acak, ADASYN menyesuaikan jumlah data sintetis berdasarkan tingkat kesulitan klasifikasi di wilayah tertentu.

Tahapan proses ADASYN terbagi menjadi beberapa langka, Yaitu :

2.8.1 MenghitungRasio Class Imbalance

$$r = \frac{m_s}{m_l} \tag{5}$$

Penjelasan :

r = Rasio kelas antar minoritas dan mayoritas

m_s = Jumlah sampel Minoritas

m_l = Jumlah Sampel mayoritas

Semakin kecil nilai r, makan nilai *Imbalance* semakin besar.

2.8.2 Menentukan jumlah data Sintetis yang ingin dihasilkan

$$G = (m_l - m_s) \times \beta \tag{6}$$

Penjelasan :

G = Total jumlah data sintetis

m_l = Jumlah data kelas Mayoritas

m_s = Jumlah data kelas Minoritas

β = Balancing factor, Nilainya antara 0 dan 1. ß =1 berarti kumpulan data yang seimbang sempurna.

2.8.3 Hitung tingkat kesulita tiap sampel minoritas

$$r_i = \frac{\Delta_i}{K} \tag{7}$$

Penjelasan :

r_i = Tingkat kesulitan klasifikasi untuk data minoritas ke - i

Δ_i = Jumlah KNN mayoritas

K = KNN Terdekatnya

2.8.4 Hitung distribusi normalisasi tingkat kesulitan

$$\delta_i = \frac{r_i}{\sum r_i} \tag{8}$$

Penjelasan :

δ_i = Bobot kesulitan yang sudah dinormalisasi untuk data minoritas ke-i

r_i = Kesulitan lokal data ke-i

$\sum r_i$ = total kesulitan semua sampel minoritas

2.8.5 Tentukan jumlah sampel sintetis yang akan dihasilkan

$$g_i = \delta_i \times G \tag{9}$$

Penjelasan :

g_i = Dumlah data sintetis baru yang akan dihasilkan dari sampel minoritas

δ_i = Bobot Kesulitan

G = Total jumlah data sintetis

2.8.6 Buat data sintetis

$$x_{new} = x_i + (x_{zi} - x_i) \times \lambda \quad (10)$$

Penjelasan :

x_{new} = Data sintetis baru yang dihasilkan

x_i = Titik data minoritas asli

x_{zi} = Salah satu KNN terdekat dari data minoritas asli (x_i)

λ = Nilai acak antara 0 dan 1

Dengan cara ini, ADASYN tidak hanya menyeimbangkan distribusi kelas, tetapi juga membantu model fokus pada area yang paling kompleks dalam ruang fitur, sehingga meningkatkan akurasi dan kemampuan generalisasi pada kasus seperti analisis sentimen dengan distribusi data yang tidak seimbang.

2.9 *K-Nearest Neighbor (KNN)*

K-Nearest Neighbors (KNN) adalah algoritma non-parametrik untuk klasifikasi/regresi yang mengklasifikasikan sebuah sampel berdasarkan label mayoritas dari k tetangga terdekatnya di ruang fitur. Untuk data teks atau vektor berdimensi tinggi (Bagus Trianto et al., 2021).

2.9.1 Euclidean Distance

Euclidean Distance adalah metrik yang mengukur jarak garis lurus terpendek antara dua titik (atau vektor) dalam ruang Euklides (ruang geometris yang kita kenal). Euclidean Distance adalah kasus khusus dari Minkowski Distance di mana parameter p diatur menjadi 2. Dalam konteks KNN, menggunakan Minkowski Distance dengan p=2 akan menghasilkan hasil yang sama persis dengan menggunakan Euclidean Distance sebagai metrik jarak (Flach, 2012).

$$Dis_1 = \sqrt{\sum_{j=1}^{d} (x_j - y_j)^2} \quad (11)$$

Keterangan :

x_j = titik koordinat *term* x ke - 1

y_j = titik koordinat *vector* y ke - 1

Dis_1 = Jarak atau Distance yang dihasilkan dari *vector* x dan y.

$\sum_{j=1}^{d}$ = Penjulahan *vector* dari titik awal hingga batas atas dari penjulahan.

2.10 *Naïve Bayes Classifier*

Naive Bayes adalah kelompok algoritma klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi *naif* bahwa setiap fitur saling independen satu sama lain dalam memprediksi kelas. Setiap kata pada sebuah dokumen/*tweet* dianggap sebagai fitur yang berkontribusi secara independen terhadap probabilitas dokumen termasuk kelas *positif* atau *negatif*. Karena asumsi kesederhanaan ini, algoritma disebut "naif" meskipun dalam praktiknya sering memberikan performa baik pada data teks berdimensi tinggi (Merawati et al., 2021).

Formula yang digunakan dalam algoritma naive bayes yaitu :

$$Posterior = \frac{Likelihood\ .\ Prior}{Evidence} \tag{12}$$

Keterangan :

Posterior : Merupakan probabilitas akhir (hasil yang ingin dicari)

Likelihood : Adalah probabilitas bukti yang muncul dalam kelas tertentu.

Prior : Merupakan probabilitas awal suatu kelas

Evidence : Adalah probabilitas total dari bukti muncul dalam seluruh kelas.

Secara umum rumus Rumus dari *Naive Bayes* antara lain :

$$P(A|B) = \frac{P(B|A)\ .\ P(A)}{P(B)} \tag{13}$$

Keterangan :

$P(A|B)$: Probabilitas suatu kelas A diberikan data fitur X.

$P(B|A)$: Probabilitas munculnya fitur B pada kelas A.

$P(A)$: Probabilitas awal atau prior probability dari kelas A.

$P(B)$: Probabilitas total dari fitur B, berfungsi sebagai konstanta normalisasi.

Rumus ini berfungsi untuk mengklasifikasikan opini publik terhadap Program MBG ke dalam kategori sentimen yang berbeda, dengan dasar logika probabilistik yang kuat dan dapat dibuktikan secara matematis.

2.10.1 Laplace Smoothing

Laplace Smoothing adalah teknik untuk menghindari probabilitas nol dalam Naive Bayes ketika sebuah kata tidak muncul pada kelas tertentu dengan cara menambahkan nilai +1 pada setiap frekuensi kata sehingga semua kata memiliki peluang > 0 (Ataei et al., 2024).

$$P(B|A) = \frac{Count(B|A) + 1}{\sum_i Count(B_i|A) + |v|} \tag{14}$$

Keterangan :

$Count(B|A) =$ Jumlah frekuensi kata B pada kelas A

1 = Nilai Smoothing untuk mencegah probabilitas 0

$\sum_i Count(B_i|A)$ = Jumlah Frekuensi semua kata pada kelas A

$|v|$ = Total Kata

2.11 *Split Validaiton*

Split Validation adalah metode evaluasi model yang membagi dataset menjadi dua bagian, yaitu data latih (*training*) dan data uji (*testing*) untuk mengukur kemampuan model dalam memprediksi data baru. Pembagian umumnya menggunakan rasio 80:20 atau 70:30, dengan tujuan menilai akurasi model dan mencegah *overfitting*. Metode ini sederhana dan efisien, namun hasilnya dapat bervariasi tergantung pembagian data (Oktafiani et al., 2023). Karena itu, penggunaan *stratified split* agar distribusi kelas tetap seimbang antara data latih dan uji (FAHRUDIN et al., 2024).

2.12 *Confusion Matrix*

Confusion Matrix adalah sebuah tabel kontingensi yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan kelas sebenarnya dan kelas yang diprediksi oleh model. Matriks ini menyediakan gambaran terperinci tentang kesalahan klasifikasi bukan hanya seberapa sering model tepat secara keseluruhan, tetapi juga pola kesalahan spesifik (kelas mana yang sering terkelirukan menjadi kelas lain). Dengan kata lain, confusion matrix mempermudah pendeteksian *which* classes the model confuses. Banyak penelitian sistem klasifikasi di Indonesia menggunakan confusion matrix sebagai alat utama evaluasi. Confusion Matrix biasanya disajikan dalam bentuk 2x2 seperti pada tabel berikut.

Tabel 1. Confussion Matrix

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positife (TP)	False Negative (FN)
Actual Negative	False Positife (FP)	True Negative (TN)

Penjelasan :

true positive (TP) : kondisi ketika prediksi dan data aktual positif

true negative (TN) : kondisi dimana prediksi dan data aktual negatif

false positive (FP): kondisi ketika prediksi positif dan data aktual negatif

false negative (FN) : kondisi ketika prediksi negatif dan data aktual positif

Kempat kategori tersebut dapat dimanfaatkan untuk menghitung nilai akurasi, presisi, recall, dan F1-score, yang berfungsi sebagai indikator dalam menilai tingkat keandalan suatu model. Adapun cara pengukuran keandalan model dijelaskan sebagai berikut.

1. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

Menunjukkan proporsi prediksi yang benar dari seluruh sampel. Bermanfaat sebagai ukuran umum, tetapi menyesatkan saat data tidak seimbang.

2. Precision

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

Dari semua instance yang diprediksi positif, berapa proporsi yang benar-benar positif. Penting ketika biaya FP tinggi (mis. false alarm). Banyak penelitian text-classification di Indonesia melaporkan presisi dari confusion matrix untuk menilai kualitas prediksi kelas tertentu.

3. Recall

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

dari seluruh instance positif yang benar, berapa yang berhasil dideteksi? Penting jika biaya FN tinggi. Dalam konteks *Binary Classification, recall* dihitung secara independen untuk setiap kelas dan mencerminkan tingkat sensitivitas model dalam mengidentifikasi seluruh anggota kelas target (Grandini et al., 2020).

4. Specifity

$$Specifity = \frac{TN}{TN + FP} \tag{18}$$

Mengukur kemampuan model mengenali kelas negatif.

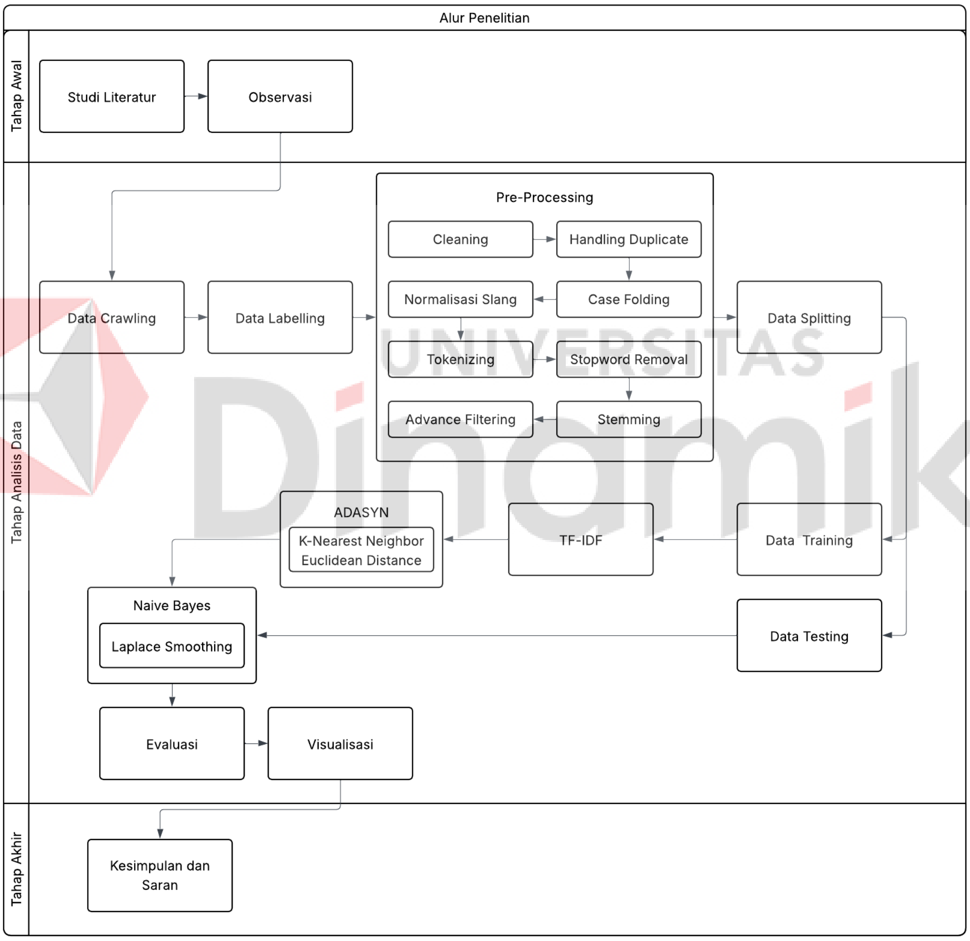
5. F1-Score

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{19}$$

Berguna saat terdapat trade-off antara precision dan recall; biasa dilaporkan penelitian klasifikasi.

BAB III
METODOLOGI PENELITIAN

Pada tahap ini menjelaskan tahap-tahap yang akan dilakukan untuk menyelesaikan penelitian. Tahapan yang digunakan dalam mengerjakan penelitian ini dibagi menjadi 3 tahap antara lain Tahap Awal, Tahap Analisis Data dan Tahap Akhir.



Gambar 3.1 Alur penelitian

3.1 Tahap Awal

Pada tahap ini merupakan tahap awalan yang dilakukan sebelum memulai tahap analisis data.

3.1.1 Studi Literatur

Berdasarkan Penelitian terdahulu, didapatkan keputusan metodologis yang akan diterapkan di dalam penelitian seperti penggunaan Metode Naive Bayes karena sifatnya yang menghitung probabilitas kata dan bagus dalam mengklasifikasi teks pendek, serta ADASYN untuk mengatasi faktor *Imbalance* Data pada data sentimen publik dengan menyeimbangkan kelas minoritas dengan kelas Mayoritas menggunakan teknik *Oversampling* data.

3.1.2 Observasi

Hasil Observasi yang ditemukan bahwa sentimen publik banyak membahas tentang distribusi MBG karena banyaknya diskusi yang membahas pasal keracunan MBG, selain itu banyaknya temuan cuitan yang berulang yang berupa spam dari akun bot/berita. Selain itu Cuitan di *Platform* X memiliki karakteristik data tidak terstruktur yang mengandung banyak *noise* atau elemen yang tidak relevan untuk analisis sentimen, seperti *URL*, *mention* (@), *hashtag* (#), angka, dan tanda baca berlebih dan beberapa cuitan juga merupakan kata slang atau kata gaul (yg, tdk, bgt).

3.2 Tahap Analisis Data

Tahap Analisis data Adalah tahap yang dilakukan untuk memahami kebutuhan data yang diperlukan dalam melakukan penelitian, lalu dilanjutkan dengan pemrosesan data.

3.2.1 Data Crawling

Pada tahap ini dilakukan proses pengambilan data komentar publik dari *Platform* X (Twitter) yang berkaitan dengan Program Makan Bergizi Gratis (MBG). Proses crawling dilakukan menggunakan bahasa pemrograman *Python* melalui *tools Google Colab*. Proses Crawling dilakukan dengan menggunakan *library* Node.js dan *library* Pandas untuk memanipulasi dan mengolah data dalam bentuk *dataframe*.

Tabel 3.1 Data t*ext Tweet*

No.	*Tweet*
1.	Program Makan Bergizi Gratis Menjamin Setiap Anak Mendapatkan Asupan yang Sehat dan Seimbang. #MBG #MakanBergiziGratis #AnakSehatBangsaKuat #GiziUntukSemua
2.	Melalui Program Makan Bergizi Gratis setiap anak Indonesia dijamin mendapatkan asupan yang sehat dan seimbang setiap harinya. #MBG #MakanBergiziGratis #AnakSehatBangsaKuat #GiziUntukSemua
3.	Sufmi Dasco menyoroti kasus keracunan masal dalam program makan bergizi gratis termasuk insiden di Kupang. Ia mendesak BGN untuk bertindak tegas guna mencegah terulangnya kejadian serupa dan memastikan keamanan program https://t.co/jXD5gXxtBH
4.	@CNNIndonesia Makan Bergizi Gratis katanya. Puluhan sampai ratusan anak keracunan. Pidato bilang sukses fakta bilang gagal. Kalau kelalaian bisa dihukum ini jelas pasal utama.,-5.0,Negative

Data yang berhasil dikumpulkan kemudian disimpan dalam format .csv (*comma-separated values*) agar dapat diolah pada tahap berikutnya. Hasil crawling yang diperoleh sebanyak 14.414 data dan mencakup text komentar dari *user*.

3.2.2 Data Labelling

Proses labelling data ini merupakan proses dalam mengkalasifikasi *tweet* dari pengguna yang bersifat Positif atau Negatif. Proses pelabelan data dilakukan secara otomatis dengan menggunakan *Lexicon*, teknik ini digunakan untuk melakukan Klasifikasi sentimen berbasis daftar kata yang sudah didefinisikan sebelumnya sebagai kata positif atau negatif

Tabel 3.2 Data Tweet yang sudah dilabeli

Tweet	Sentiment
Program Makan Bergizi Gratis Menjamin Setiap Anak Mendapatkan Asupan yang Sehat dan Seimbang. #MBG #MakanBergiziGratis #AnakSehatBangsaKuat #GiziUntukSemua https://t.co/3KxbeWMv4Q	Positive
Melalui Program Makan Bergizi Gratis setiap anak Indonesia dijamin mendapatkan asupan yang sehat dan seimbang setiap harinya. #MBG #MakanBergiziGratis #AnakSehatBangsaKuat #GiziUntukSemua https://t.co/7j1VJBwWT9	Positive
Ayo dukung program Makan Bergizi Gratis guna terciptanya generasi Indonesia Emas 2045 #KawalMBG #IndonesiaEmas2045 #ManfaatMBG #DukungMBG https://t.co/hPXSAD4aDZ	Positive
Kami hanya ingin belajar yang layak dari Pendidikan Gratis bukan Makan Bergizi Gratis mas Broo @prabowo https://t.co/ZWqw9WsSfJ	Negative
@txtdrimedia MBG = makan bergizi gratis □□ MBG = makan belatung gratis MBG = makan beracun gratis	Negative

3.2.3 Data Pre-Processing

Selanjutnya akan dilakukan tahap Pre-Processing data agar data siap digunakan sebelum dilakukan pemodelan. Tahap Pre-Processing antara lain :

1. Cleaning

Pada tahap Cleansing, data akan dibersikan dari berbagai elemen-elemen tidak jelas atau noise yang menganggu kualitas data seperti URL, Mention (@), Tanda Baca, Angka, Spasi berlebih, *hashtag* (#), Emoticon dan sebagainya.

Tabel 3.3 Data Tweet setelah Cleaning

Tweet
Program Makan Bergizi Gratis Menjamin Setiap Anak Mendapatkan Asupan yang Sehat dan Seimbang
Melalui Program Makan Bergizi Gratis setiap anak Indonesia dijamin mendapatkan asupan yang sehat dan seimbang setiap harinya
Ayo dukung program Makan Bergizi Gratis guna terciptanya generasi Indonesia Emas
Ayo dukung program Makan Bergizi Gratis guna terciptanya generasi Indonesia Emas
Badan Gizi Nasional BGN menyampaikan permintaan maaf secara terbuka menyusul insiden keracunan massal dalam program Makan Bergizi Gratis MBG yang terjadi di Nusa Tenggara Timur NTT

2. Handling Duplicate Data

Tahap ini dilakukan untuk mengidentifikasi dan menghapus data *tweet* yang memiliki isi teks sama persis (*duplicate rows*) yang terjadi akibat *retweet* atau spam dari *bot*. Penghapusan duplikasi perlu untuk mencegah bias pada model Naive Bayes, memastikan model tidak "menghapal" kalimat yang sama berulang kali, serta mengurangi beban komputasi.

Tabel 3.4 data yang teridentifikasi Duplikat

Tweet
Ayo dukung program makan bergizi gratis
Ayo dukung program makan bergizi gratis
Ayo dukung program makan bergizi gratis
Ayo dukung program makan bergizi gratis

3. Case Folding

Proses Case Folding merupakan proses pengubahan *tweet* yang memiliki huruf besar (*Upperase*) menjadi huruf kecil (*LowerCase*). Hal ini dilakukan agar menghinfari kata yang sama namun terdeteksi berbeda karena Kapitalisasi.

Tabel 3.5 Hasil setelah tahap Case Folding

Tweet
program makan bergizi gratis menjamin setiap anak mendapatkan asupan yang sehat dan seimbang
melalui program makan bergizi gratis setiap anak indonesia dijamin mendapatkan asupan yang sehat dan seimbang setiap harinya
ayo dukung program makan bergizi gratis guna terciptanya generasi indonesia emas

4. Normalisasi *Slang*

Setelah huruf diseragamkan, dilakukan proses pengubahan kata-kata tidak baku, singkatan (seperti "yg", "bgt", "gak"), dan bahasa gaul (*slang*) menjadi bentuk baku sesuai kaidah Bahasa Indonesia yang baik dan benar. Proses ini dilakukan dengan mencocokkan setiap kata dalam *tweet* dengan kamus normalisasi yang diambil dari data Kaggle kamus Slang Indonesia. Tujuannya adalah untuk menyatukan variasi kata yang memiliki makna sama (misalnya: "sy", "aq", "gue" menjadi "saya") sehingga pembobotan TF-IDF menjadi lebih akurat dan fokus.

Tabel 3.6 Normalisasi Komentar Slang

Tweet Sebelum	Tweet Sesudah
ini gara gara program makan siang gratis ga bergizi	ni gara gara program makan siang gratis tidak bergizi
yang penting kan makan bergizi gratis nder yg lain mah gak penting apa itu analisis kebijakan	yang penting kan makan bergizi gratis nder yang lain mah tidak penting apa itu analisis kebijakan

5. Tokenizing

Tahap Tokenizing bertujuan untuk memecah kalimat menjadi kata-kata atau token berdasarkan spasi dari kalimat itu agar mempermudah pemrosesan teks dan analisis. Proses tokenisasi dilakukan dengan dengan Regex Tokenizer dari *library* re berdasarkan prinsip pencocokan pola (pattern matching) untuk memisahkan teks menjadi unit-unit diskrit yang berupa token.

Tabel 3.7 Tabel hasil Tokenizing

Tweet
program,makan,bergizi,gratis,menjamin,setiap,anak,mendapatkan,asupan,yang,sehat,dan,seimbang
melalui,program,makan,bergizi,gratis,setiap,anak,indonesia,dijamin,mendapatkan,asupan,yang,sehat,dan,seimbang,setiap,harinya
ayo,dukung,program,makan,bergizi,gratis,guna,terciptanya,generasi,indonesia,emas

6. Stopword Removal

Tahap ini bertujuan untuk menghapus atau menghilangkan kata yang secara umum tidak memiliki makna penting dalam sebagian kalimat sehingga dapat dikelola dengan mudah oleh sistem dalam melakukan analisis. Proses *Stopword Removal* akan dilakukan dengan Sastrawi Stopword dari *library* Sastrawi yang menyediakan kamus kata henti ("yang", "dan", "di") kemudian memproses input teks untuk menghilangkan kata-kata dari kamus tersebut.

Tabel 3.8 Hasil setelah Stopword Removal

Tweet
program,makan,bergizi,gratis,menjamin,anak,mendapatkan,asupan,sehat,seimbang
melalui,program,makan,bergizi,gratis,anak,indonesia,dijamin,mendapatkan,asupan,sehat,seimbang,harinya
ayo,dukung,program,makan,bergizi,gratis,terciptanya,generasi,indonesia,emas

7. Stemming

Tahap ini digunakan untuk mengubah kata yang berimbuhan menjadi kata dasarnya (Contoh : Memakan menjadi Makan). Proses stemming akan dilakukan dengan Sastrawi Stemming dari *library* Sastrawi yang bekerja berdasarkan serangkaian aturan Nazief & Adriani Algorithm yang telah dimodifikasi.

Tabel 3.9 Data setelah Proses Stemming

Tweet
program,makan,gizi,gratis,jamin,anak,dapat,asupan,sehat,imbang
lalu,program,makan,gizi,gratis,anak,indonesia,jamin,dapat,asupan,sehat,imbang,hari
ayo,dukung,program,makan,gizi,gratis,cipta,generasi,indonesia,emas

8. Advance Filtering

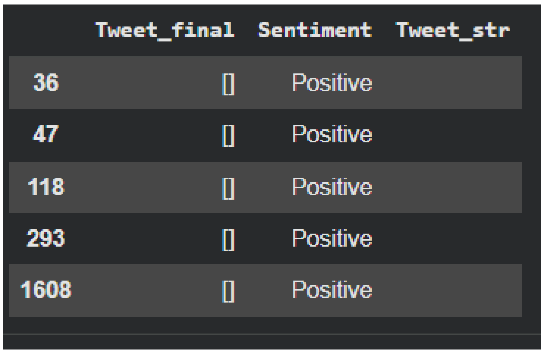
Tahap ini dilakukan beberapa proses seperti penghapusan terhadap kata-kata entitas yang menjadi topik utama pencarian data, yaitu "makan", "bergizi", dan "gratis" di mana kata-kata tersebut memiliki frekuensi kemunculan yang sangat dominan di hampir seluruh dokumen (*high frequency*) namun memiliki daya pembeda (*discriminative power*) yang rendah dalam menentukan sentimen.

Tabel 3.10 menghapus kata tidak penting

Tweet Sebelum	Tweet Sesudah
marsel,asyerem,semua,sekolah,nabire,terima,program,makan,gizi,gratis,januari,baca,lengkap	marsel,asyerem,semua,sekolah,nabire,terima,januari,baca,lengkap

program,makan,gizi,gratis,mbg,dorong,putar,ekonomi, lokal,lalu,libat,umkm,tani,tempat makan,gizi,gratis,cipta,generasi,unggul	dorong,putar,ekonomi,lokal,lalu,libat, umkm,tani,tempat cipta,generasi,unggul

Lalu dilakukan proses untuk mengidentifikasi dan menghapus baris data yang bernilai kosong (*null/NaN*).



	Tweet_final	Sentiment	Tweet_str
36	[]	Positive	
47	[]	Positive	
118	[]	Positive	
293	[]	Positive	
1608	[]	Positive	

Gambar 3.2 Data Null

3.2.4 Data Splitting

Data dibagi menjadi dua bagian utama, yaitu data Training dan data Testing dengan rasio 80:20 dengan 80% untuk pelatihan dan 20% untuk pengujian, sehingga model seperti ADASYN punya lebih banyak data Training dan membuat sampel sintetis yang lebih baik (Dawson et al., 2023).

1. Data Training

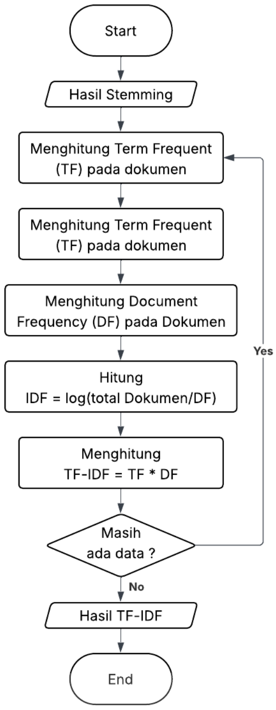
Data Training digunakan untuk membangun dan melatih model klasifikasi. Pada tahap ini, model akan mempelajari pola sentimen dan membentuk batas keputusan berdasarkan probabilitas tiap kelas.

2. Data Testing

Data Testing adalah bagian kecil dari dataset 20%, yang tidak digunakan saat pelatihan agar dapat mengukur kemampuan model pada data yang benar-benar baru.

3.2.5 TF-IDF

Pada tahap ini akan dilakukan pembobotan kata dengan TF-IDF kedalam bentuk *vector* dengan menhitung bobot kemunculan kata pada dokumen.



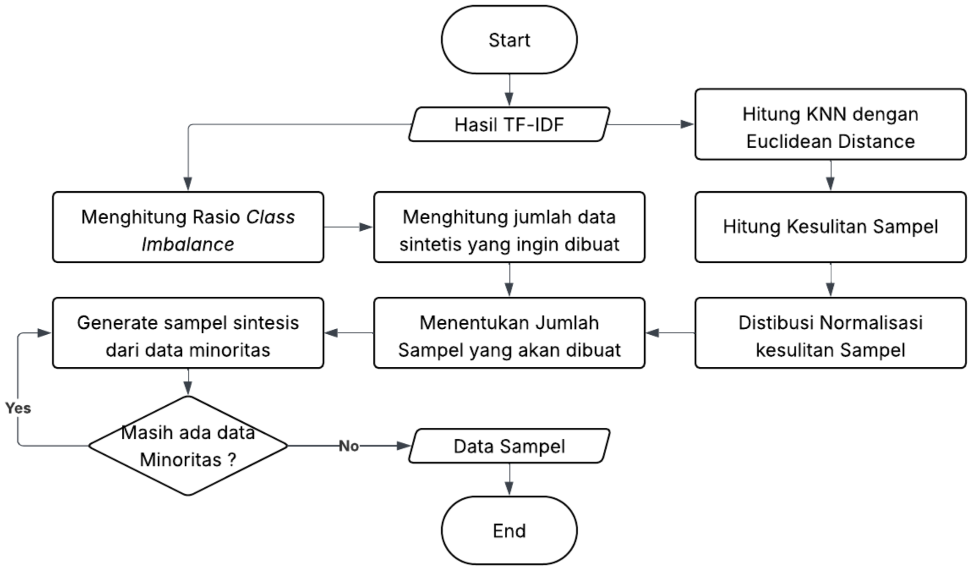
Gambar 3.3 Flowchart TF-IDF

Sumber : Karima & Mutiara (2023)

TF-IDF menghasilkan vektor numerik dengan memberikan bobot tinggi pada kata-kata yang diskriminatif (kunci sentimen) dan bobot rendah pada kata-kata umum, mengubah data teks yang tidak terstruktur menjadi format yang dapat diolah oleh algoritma. Proses ini dilakukan dengan menggunakan *Library Scikit-learn* (SKLearn) yang menyediakan kelas *Tfidfvectorizer* untuk mengubah data kedalam *vector* dan *Tfidftransformer* untuk mengubah nilai matriks yang sudah ada dengan menerapkan perhitungan bobot dan normalisasi. *Library* Pandas juga dipakai untuk mengelola dan menampilkan data teks serta hasil matriks dalam format yang *dataframe*.

3.2.6 ADASYN

Langkah selanjutnya ADASYN akan mencari tetangga terdekat lalu membuat sampel sintetis baru berupa vektor numerik melalui interpolasi antara sampel-sampel minoritas yang sudah ada sehingga model memiliki representasi data yang kuat dan seimbang sebelum digunakan untuk pelatihan dan klasifikasi sentimen.

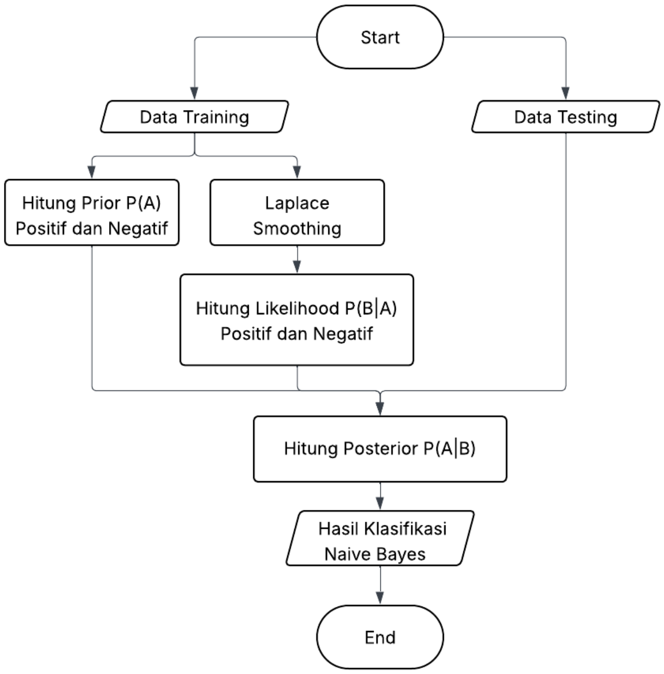


Gambar 3.4 ADASYN Flowchart

Proses ini menggunakan *library imbalanced-learn* (imblearn) yang menyediakan kelas untuk proses ADASYN yang mengimplementasikan algoritma untuk menyeimbangkan dataset. *Library Scikit-learn* juga digunakan untuk mengambil kelas NearestNeighbor yang menyediakan perhitungan tetangga terdekat dengan metrik minkowski untuk membantu ADASYN dalam mencari tetangga terdekat.

3.2.7 Naïve Bayes Classifier

Pemodelan sentimen dilakukan menggunakan algoritma Multinomial Naive Bayes untuk analisis teks berbasis frekuensi kata dengan bantuan Laplace Smoothing untuk menghindari Probabilitas nol dalam data Training.



Gambar 3.5 Naive Bayes Flowchart

Sumber : Barros et al. (2022)

Model dilatih untuk mempelajari probabilitas kemunculan kata terhadap kelas sentimen, lalu dilakukan pengujian untuk mengukur performa klasifikasi. Proses modelling dengan Naive Bayes menggunakan *library Scikit-learn (sknlearn)* yang menyediakan modul *MultinomialNB* untuk data diskrit yang merepresentasikan *term frequency* untuk klasifikasi teks.

3.2.8 Evaluasi

Evaluasi dilakukan untuk mengukur kinerja model dalam mengklasifikasikan sentimen dengan menghitung metrik seperti *accuracy, precision, recall*, dan *F1-score* yang diperoleh dari *confusion matrix*. Selain itu, digunakan teknik *k-fold cross-validation* untuk memastikan performa model stabil dan tidak bergantung pada satu pembagian data saja. Proses ini dilakukan menggunakan modul *sklearn.metrics* yang menyediakan alat untuk mengukur performa model, hasil pengujian confusion matriks lalu divisualisasikan dengan *library seaborn* dan *matplotlib*.

3.2.9 Visualisasi

Visualisasi data dilakukan menggunakan *word cloud* untuk melihat kata-kata yang paling sering muncul dalam percakapan terkait Program Makan Bergizi Gratis (MBG) di *Platform* X. Visualisasi *word cloud* dibagi setiap kelas sentimen (positif atau negatif) agar dapat terlihat perbedaan isu atau opini yang memengaruhi masing-masing sentimen.

3.3 Tahap Akhir

Tahap Akhir merupakan fase penutup dari rangkaian penelitian yang berfungsi untuk merangkum seluruh temuan utama dan memberikan rekomendasi konstruktif.

3.3.1 Kesimpulan dan Saran

Tahap kesimpulan bertujuan untuk merangkum hasil penelitian terkait analisis sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG) di *Platform* X menggunakan metode Naive Bayes Classifier, termasuk temuan mengenai kecenderungan opini publik dan performa model yang dihasilkan.

Sementara itu, tahap saran memberikan rekomendasi bagi penelitian berikutnya maupun bagi pihak terkait, seperti peningkatan metode analisis, penambahan sumber data, serta pemanfaatan hasil sentimen sebagai masukan dalam evaluasi dan pengembangan program MBG agar lebih tepat sasaran.

BAB IV
HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas hasil analisis dari pengelolaan data *Tweet* Sentimen Makan Bergizi gratis yang diambil dari platform X dengan Naive Bayes dan ADASYN.

4.1 Analisis

Tahap analisis diawali dengan mengidentifikasi permasalahan utama yang melatarbelakangi penelitian ini. Berdasarkan observasi awal, ditemukan tingginya volume opini dan reaksi masyarakat terhadap pelaksanaan Program Makan Bergizi Gratis (MBG) di Platform X.

4.2 Data crawling

Data Crawling digunakan dengan menggunakan Python dengan metode Tweet Harvest yang mengambil data tweet berdasarkan akses login kamu (*auth_token*) dari browser untuk melakukan pencarian tweet di X.


```
import subprocess
import time
import os
import pandas as pd
from glob import glob
```

```
# Instalasi paket dasar
!pip install pandas

# Instal Node.js (karena tweet-harvest butuh Node)
!sudo apt-get update
!sudo apt-get install -y nodejs npm
```

Gambar 4.1 *Install library* untuk *crawling*

Gambar 4.1 menampilkan tahap *environment setup* di Google Colab sebelum melakukan *data crawling* yang berfungsi untuk menginstal *library* Python (Pandas) serta mengunduh dan menginstal Node.js versi terbaru (versi 20) ke dalam sistem Linux Google Colab. Perintah-perintah teknis seperti sudo apt-get, curl, dan penambahan *repository* dilakukan untuk memastikan versi Node.js yang terpasang kompatibel dengan alat yang akan digunakan.

```
twitter_auth_token = "2945bfc40d04af6e0e774cbb2b39fc3873bcc33f"

filename = "MBGSentimen.csv"
search_keyword = "(Makan Bergizi Gratis OR MBG) lang:id since:2025-01-06 until:2025-10-01"
limit = 10000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 4.2 Proses pengambilan data

Gambar 4.2 menampilkan script untuk mengatur konfigurasi penting seperti *token* otentikasi, kata kunci pencarian ("Makan Bergizi Gratis" atau "MBG"), serta rentang waktu yang diinginkan, lalu menjalankan perintah npx untuk secara otomatis memanen data *tweet* sesuai kriteria tersebut dan menyimpannya ke dalam file .csv untuk analisis selanjutnya.

	A
1	
2	Program Makan Bergizi Gratis Menjamin Setiap Anak Mendapatkan Asupan yang Sehat dan Seimbang. #MBG #MakanBergiziGratis #AnakSehatBangsaKuat #GiziUntukSemua https://t.co/3KxbeWMv4
3	Melalui Program Makan Bergizi Gratis setiap anak Indonesia dijamin mendapatkan asupan yang sehat dan seimbang setiap harinya. #MBG #MakanBergiziGratis #AnakSehatBangsaKuat #GiziUntukSer
4	Ayo dukung program Makan Bergizi Gratis guna terciptanya generasi Indonesia Emas 2045 #KawalMBG #IndonesiaEmas2045 #ManfaatMBG #DukungMBG https://t.co/hPXSAD4aDZ
5	Ayo dukung program Makan Bergizi Gratis guna terciptanya generasi Indonesia Emas 2045 #KawalMBG #IndonesiaEmas2045 #ManfaatMBG #DukungMBG https://t.co/1ir18lHxC2
6	Badan Gizi Nasional (BGN) menyampaikan permintaan maaf secara terbuka menyusul insiden keracunan massal dalam program Makan Bergizi Gratis (MBG) yang terjadi di Nusa Tenggara Timur (NTT)
7	Kapolri Targetkan Bangun 400 SPPG se-Indonesia Akhir Tahun Ini LAMPUNG - Kapolri Jenderal Listyo Sigit menargetkan bakal membangun 400 Satuan Pelayanan Pemenuhan Gizi (SPPG) Polri se-In
8	Kami hanya ingin belajar yang layak dari Pendidikan Gratis bukan Makan Bergizi Gratis mas Broo @prabowo https://t.co/ZWqw9WsSfJ
9	Polres Sibolga Dukung Program Presiden RI Meningkatkan Makan Bergizi Anak Gratis https://t.co/OlurBra5rM
10	@IckurM @MurtadhaOne1 Perbandingan antara anggaran konsumsi rapat menteri Rp 171 ribu per orang dan bantuan makan bergizi gratis (MBG) Rp 10 ribu per anak memang sering jadi polemik tapi
11	Hasil dari penelitian di Aceh maupun di Papua menunjukkan bahwa siswa yang mendapatkan makan bergizi gratis mengalami peningkatan konsentrasi yang lebih baik ujar Ikeu. / #Makanbergizigratis ht
12	OJK Dukung Program Prioritas Pemerintah dan Sinergi Jaga Stabilitas Sistem Keuangan Sobat OJK OJK mendukung penuh program prioritas pemerintah seperti program makan bergizi gratis serta pro
13	Babinsa Koramil jajaran Kodim 1420/Sidrap melakukan pendampingan pendistribusian program Makan Bergizi Gratis (MBG) di sekolah-sekolah di wilayah empat kecamatan yang ada di Kabupaten Sidr
14	@DivHumas_Polri Makan bergizi gratis? Pak polisi perhatian banget sama kesehatan kita nih
15	Polsek Muara Dua Kawal Ketat Distribusi Program Makan Bergizi Gratis di Sekolah dan Posyandu #polriuntukmasyarakat #bidhumaspoldaaceh #poldaaceh #polripresisi #kapolreslhokseumawe #quickw

Gambar 4.3 Data hasil Crawling

Data yang berhasil diperoleh sebanyak 14.414 data dari periode 30 juli 2025 sampai dengan 10 Oktober 2025. Hasil dataset didapatkan dari hasil crawling beberapa kali karena batas pengambilan data tweet sebanyak 1000 sampai 3000 data.

4.3 Data Labelling

Pelabelan dilakukan secara otomatis dengan menggunakan *Lexicon Based* pada Google Colab dengan menentukan Sentimen positif dan negatif.

```
positive_url = "https://raw.githubusercontent.com/fajri91/InSet/master/positive.tsv"
negative_url = "https://raw.githubusercontent.com/fajri91/InSet/master/negative.tsv"
```

Gambar 4.4 Repository GitHub Kamus Lexicon

Pelabellan dibantu repository github fajri91 yang menyediakan kamus bobot kata positif dan negatif untuk menentukan sentimen pada komentar tweet.

```python
    # Tentukan Label Akhir
    if score >= 0:
        sentiment = 'Positive'
    else:
        sentiment = 'Negative'

    return sentiment

print("Sedang memproses label baru...")

df[['sentiment']] = df['full_text'].apply(lambda x: pd.Series(determine_sentiment(x)))
```

Gambar 4.5 Menyimpan data label

Gambar 4.5 data akan ditambahkan label "*sentiment*" sebagai kolom yang mengidentifikasikan *tweet* tersebut merupakan positif atau negatif berdasarkan hasil skore perhitungan berdasarkan kamus kata positif dan negatif pada repository github fajri91.

	full_text	sentiment
0	Program Makan Bergizi Gratis Menjamin Setiap A...	Positive
1	Melalui Program Makan Bergizi Gratis setiap an...	Positive
2	Ayo dukung program Makan Bergizi Gratis guna t...	Positive
3	Ayo dukung program Makan Bergizi Gratis guna t...	Positive
4	Badan Gizi Nasional (BGN) menyampaikan permint...	Positive

Gambar 4.6 Hasil labelling data

Data yang sudah dilabeli akan disimpan dalam format .csv yang nantinya akan digunakan pada tahap selanjutnya.

4.4 *Pre-Processing*

Tahap preprocessing digunakan untuk mengatasi *noise* pada komentar tweet agar hasil analisis data dapat lebih baik. Proses dilakukan dengan pemrograman python pada google colab.

```
from google.colab import drive
drive.mount('/content/drive')

Show hidden output

df =
pd.read_csv('/content/drive/MyDrive/Skripsi/Tugas Akhir/TA Lingga/Dataset/MBGSentimen_Labelled.csv')

df.rename(columns
          = {'full_text':'Tweet', 'sentiment':'Sentiment'}, inplace = True)
```

Gambar 4.7 Import dataset dari Google Drive

Sebelum melangkah ke tahap pembersihan data (*pre-processing*), dilakukan *load dataset* yang telah melalui proses pelabelan sentimen. Dataset tersebut disimpan dalam penyimpanan *cloud* Google Drive dan diintegrasikan ke dalam lingkungan kerja Google Colab. File dataset yang digunakan adalah MBGSentimen_Labelled.csv yang berisi data *tweet* yang sudah dilabeli.

4.4.1 Cleaning

Tahap selanjutnya adalah pembersihan data (*cleaning*) yang bertujuan untuk menghilangkan elemen-elemen gangguan (*noise*) dalam teks agar hasil analisis lebih akurat.


```
def clean_text(text):

    if pd.isna(text):
        return ""

    if isinstance(text, list):
        text = " ".join(text)

    # Remove URLs
    text = re.sub(r'https?://\S+|www\.\S+', '', text)

    # Remove mentions (@user)
    text = re.sub(r'@\w+', '', text)

    # Remove hashtags (#tag)
    text = re.sub(r'#\w+', '', text)

    # Remove punctuation
    text = text.translate(str.maketrans('', '', string.punctuation))

    # Remove numbers
    text = re.sub(r'\d+', '', text)

    # Remove multiple spaces with a single space
    text = re.sub(r'\s+', ' ', text).strip()
    return text

print("Applying cleaning to 'Tweet'...")

df['Tweet'] = df['Tweet'].apply(clean_text)
```

Gambar 4. 8 Proses Cleaning dataset

Berdasarkan skrip clean_text yang dijalankan, proses ini secara spesifik menghapus URL atau tautan situs, mention pengguna (kata yang diawali tanda @), serta hashtag (#). Selain itu, fungsi ini juga membersihkan seluruh tanda baca dan angka yang tidak memiliki nilai sentimen, serta merapikan spasi berlebih (*multiple spaces*) agar format teks menjadi seragam dan rapi.

4.4.2 Handling Duplicate

Setelah proses *cleaning*, tahap selanjutnya adalah penanganan data duplikat (*Handling Duplicate*) untuk memastikan tidak ada redundansi data yang dapat menyebabkan bias pada hasil analisis.

```
Duplicate =  df.duplicated(subset = ['Tweet']).sum()
print(f"Data Duplikat pada Clean Tweet = {Duplicate}")
```

```
Data Duplikat pada Clean Tweet = 2610
```

```
#Hapus data Duplicate
df = df.drop_duplicates(subset = ['Tweet'],  keep = 'first')
```

```
df.shape
```

```
(11804, 2)
```

Gambar 4. 9 Proses *Handling Duplicate* Data

Dari total awal sebanyak 14.414 data, sistem mengidentifikasi adanya 2.610 data yang memiliki isi teks sama persis (*duplicate rows*). Data duplikat tersebut kemudian dihapus dari dataset, sehingga menyisakan 11.804 data.

4.4.3 Case Folding

Selanjutnya tahap Case Folding bertujuan untuk menyeragamkan format teks dengan mengubah seluruh huruf kapital menjadi huruf kecil (*lowercase*).

```
df['Tweet'] = df['Tweet'].str.lower()
df.sample(n = 10, random_state = 42)
```

Gambar 4.10 Proses Case Folding

Proses ini dilakukan menggunakan fungsi .str.lower() agar sistem dapat memperlakukan kata yang sama dengan kapitalisasi berbeda (misalnya 'MBG' dan 'mbg') sebagai satu entitas yang sama. Untuk memverifikasi hasil proses ini, ditampilkan 10 sampel data secara acak menggunakan fungsi .sample() dengan parameter random_state=42 guna menjaga konsistensi hasil acakan saat kode dijalankan ulang.

13901	marsel asyerem semua sekolah di nabire akan te...	Positive
676	program makan bergizi gratis mbg mendorong per...	Positive
7905	program makan bergizi gratis mbg di kabupaten ...	Negative
4522	makan bergizi gratis ciptakan generasi unggul	Positive
8419	dalam tiga hari terakhir hampir siswa di kabup...	Negative
10550	pemimpin redaksi cnn indonesia menjelaskan sek...	Positive
4597	makan bergizi gratis kini jadi bagian penting ...	Positive
2443	perwakilan bpkp provinsi riau melakukan permin...	Positive
3122	ini namanya makan bergizi wanna be gratis wkwk...	Positive
6892	program ini hadir bukan sekadar bantuan makana...	Positive

Gambar 4.11 Hasil Case Folding

4.4.4 Normalisasi Slang

Setelah proses penyeragaman huruf, dilakukan tahap Normalisasi Slang untuk mengubah kata-kata tidak baku, singkatan, dan bahasa gaul menjadi bentuk baku sesuai kaidah Bahasa Indonesia yang benar.

	slang	formal
0	aamiin	amin
1	adek	adik
2	adlh	adalah
3	aer	air
4	aiskrim	es krim
5	aj	saja

Gambar 4.12 Dataset kata Slang

Proses pemetaan kata ini menggunakan kamus referensi eksternal yang bersumber dari dataset publik Kaggle berjudul *'Slang Indonesia'* (oleh Sodolanangbjkatio). Penelitian Ilyas Tri Khaqiqi et al. (2023) juga menggunakan Kamus Slang Indonesia dengan tujuan untuk memetakan variasi kata informal yang sering muncul di media sosial X menjadi kata dasar yang seragam, sehingga makna kalimat dapat dikenali dengan lebih akurat oleh algoritma pada tahap selanjutnya.

4.4.5 Tokenizing

Tahap berikutnya adalah Tokenizing, yang bertujuan untuk memecah kalimat dalam setiap *tweet* menjadi potongan kata yang berdiri sendiri atau disebut sebagai token.

```
try:
    nltk.data.find('tokenizers/punkt_tab')
except LookupError:
    nltk.download('punkt_tab')

df['Tweet'] = df['Tweet'].apply(word_tokenize)
df.sample( n = 10, random_state = 42)
```

Gambar 4.13 Proses tokenizing

Proses ini dilakukan menggunakan fungsi word_tokenize dari pustaka NLTK (*Natural Language Toolkit*). Sebelum eksekusi, sistem memastikan paket punkt_tab telah terunduh karena paket tersebut berisi model pre-trained yang diperlukan NLTK untuk mengenali batas-batas kata dan kalimat secara cerdas. Hasil dari proses ini mengubah struktur data teks dari kalimat utuh menjadi daftar kata (*list of strings*).

4.4.6 Stopword removal

Tahapan berikutnya adalah *Stopword Removal*, yang bertujuan mengeleminasi kata-kata umum yang sering muncul namun tidak memberikan kontribusi signifikan terhadap sentimen, seperti kata sambung 'dan', 'yang', atau 'di'.

```
factory = StopWordRemoverFactory()
stop_words_list = factory.get_stop_words()

def remove_stopwords_sastrawi(text_tokens):
    filtered_tokens = []
    for word in text_tokens:
        if word not in stop_words_list:
            filtered_tokens.append(word)
    return filtered_tokens

print("Applying stopword removal to 'Tweet'...")
df['Tweet'] = df['Tweet'].apply(remove_stopwords_sastrawi)
df.sample( n = 10, random_state = 42)
```

Gambar 4.14 Proses *Stoword Removal* dengan Sastrawi

Proses ini memanfaatkan pustaka Sastrawi dengan memanggil kelas StopWordRemoverFactory untuk mendapatkan daftar kata henti standar Bahasa Indonesia. Melalui fungsi remove_stopwords_sastrawi, sistem akan memfilter setiap token; kata-kata yang termasuk dalam daftar *stop words* akan dihapus, sehingga data teks menjadi lebih ringkas dan fokus pada kata-kata inti yang memuat makna.

4.4.7 Stemming

Tahap akhir dari pra-pemrosesan teks adalah *Stemming*, yang bertujuan mengembalikan kata-kata berimbuhan menjadi kata dasarnya (*root word*) agar variasi kata yang memiliki makna sama dapat diseragamkan.

```
tqdm.pandas()

factory = StemmerFactory()

stemmer = factory.create_stemmer()

def stem_tokens(tokens):
    text_to_stem = " ".join(tokens)
    stemmed_text = stemmer.stem(text_to_stem)
    return stemmed_text.split()

print("Applying stemming to 'tokens_filtered'...")
# Terapkan fungsi stem_tokens ke kolom 'tokens_filtered'
df['Tweet'] = df['Tweet'].progress_apply(stem_tokens)
```

Gambar 4.15 Proces Stemming dengan Sastrawi

Berdasarkan fungsi stem_tokens yang dibuat, daftar token terlebih dahulu digabungkan kembali menjadi kalimat utuh, kemudian diproses oleh mesin *stemmer* untuk menghilangkan imbuhan, dan akhirnya dipecah kembali menjadi daftar token dasar. Karena proses ini memakan waktu komputasi yang cukup intensif, eksekusi dilakukan menggunakan fungsi progress_apply untuk memantau kemajuan pemrosesan data.

4.4.8 Advance Filtering

Proses ini dilakukan untuk menyaring kata-kata yang sangat terkait dengan topik utama dan dinilai kurang informatif.

```
unimportant_words = [
    'makan', 'gizi', 'gratis', 'program', 'mbg', 'menu', 'dapur','sppg',
    'bgn', 'anak', 'siswa', 'siswi', 'murid', 'sekolah', 'prabowo', 'subianto',
    'gibran', 'jokowi', 'presiden', 'wapres', 'perintah', 'pemerintah',
    'negara', 'bangsa', 'rakyat', 'masyarakat', 'warga', 'indonesia', 'nasional',
    'desa', 'daerah', 'juta', 'anggaran', 'program'
]

def remove_unimportant_words(tokens):
    return [word for word in tokens if word not in unimportant_words]

print("Applying removal of unimportant words...")
df['Tweet_unword'] = df['Tweet_stem'].apply(remove_unimportant_words)
print("Removal complete.")
```

Gambar 4.16 Proses hapus kata tidak bermakna

Kata-kata tersebut dikategorikan sebagai *unimportant words*, meliputi “makan”, “gizi”, “gratis”, “mbg”, “program”, dan “siang”, dan sebagainya. Penghapusan dilakukan melalui fungsi *remove_unimportant_words* yang memeriksa setiap token dalam dokumen dan mengeliminasi token yang termasuk dalam daftar tersebut. Setelah itu dilakukan pengecekan data bernilai kosong.

```
df_processed = df[df['Tweet_unword'].apply(lambda x: isinstance(x, list) and len(x) > 0)].copy()

df_processed = df_processed[['Tweet_unword', 'Sentiment']]
df_processed.rename(columns={'Tweet_unword': 'Tweet_final'}, inplace=True)
df_processed.reset_index(drop=True, inplace=True)

print("DataFrame baru setelah membersihkan data kosong dan merename kolom:")
display(df_processed.head())
print(f"Bentuk DataFrame baru: {df_processed.shape}")
```

Gambar 4.17 Proses menghapus data null

Didapatkan bahwa data yang benilai kosong sebanyak 49 data. Dari hasil ini dilakukan penghapusan data yang kosong agar tidak menjadi noise yang menganggu frekuensi kata. Setelah penghapusan data yang kosong, jumlah data yang sebelumnya 11.804 sekarang berjumlah 11.755 data.

4.5 Modelling

Tahap *Modelling* merupakan inti dari proses analisis data, di mana data teks yang telah bersih dan terstruktur digunakan untuk melatih sistem agar mampu mengenali pola sentimen secara otomatis.

4.5.1 Data Split

Tahapan pemodelan diawali dengan proses *data splitting* dengan rasio 80:20. Dalam proses ini, parameter *stratify* = *y* diterapkan untuk mempertahankan distribusi kelas sentimen, sehingga proporsi data positif dan negatif pada data latih maupun data uji tetap mencerminkan distribusi kelas pada dataset awal.

```
from sklearn.model_selection import train_test_split

df_final['Tweet_str'] = df_final['Tweet_final'].apply(lambda x: ' '.join(x))

X = df_final['Tweet_str']
y = df_final['Sentiment']

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

print(f"X_train shape: {X_train.shape}")
print(f"X_test shape: {X_test.shape}")
print(f"y_train shape: {y_train.shape}")
print(f"y_test shape: {y_test.shape}")
```

Gambar 4.18 Proses Data Splitting

Hasil eksekusi dari proses *splitting* tersebut menghasilkan distribusi pembagian data dengan memisahkan total dataset menjadi dua subset: 9.443 data latih (80%) yang digunakan algoritma untuk mempelajari pola sentimen, dan 2.361 data uji (20%) yang disisihkan secara khusus untuk memvalidasi akurasi prediksi model terhadap data baru yang belum pernah dilihat sebelumnya.

4.5.2 TF-IDF

Setelah proses pembagian data, dilakukan ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengubah teks mentah menjadi representasi vektor numerik berbobot, sehingga dapat diproses oleh algoritma machine learning. Proses ini diimplementasikan menggunakan TfidfVectorizer dari pustaka Scikit-Learn.

```
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer()

X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)

X_test_tfidf = tfidf_vectorizer.transform(X_test)

print(f"X_train_tfidf shape: {X_train_tfidf.shape}")
print(f"X_test_tfidf shape: {X_test_tfidf.shape}")
```

Gambar 4.19 Proses TF-IDF

Pada data latih (X_train), digunakan fungsi fit_transform untuk membangun kosakata dan menghasilkan matriks TF-IDF, sedangkan pada data uji (X_test) hanya diterapkan fungsi transform.

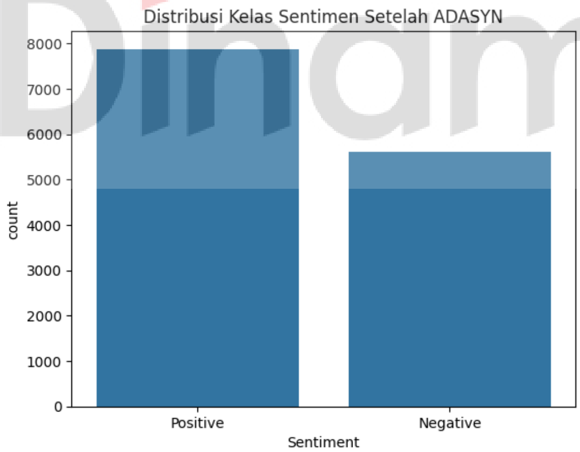
4.5.3 ADASYN

Mengingat adanya ketimpangan jumlah data yang signifikan antara sentimen positif dan negatif, penelitian ini menerapkan teknik ADASYN (*Adaptive Synthetic Sampling*) untuk memperbaiki representasi kelas minoritas.

```python
from imblearn.over_sampling import ADASYN
from collections import Counter

print('Original training set shape %s' % Counter(y_train))

counts = Counter(y_train)
majority_class = max(counts, key=counts.get)
minority_class = min(counts, key=counts.get)

target_minority_count = int(counts[majority_class] / 2)

adasyn = ADASYN(sampling_strategy={minority_class: target_minority_count}, random_state=42)

X_train_res, y_train_res = adasyn.fit_resample(X_train_tfidf, y_train)

print('Resampled training set shape %s' % Counter(y_train_res))
print(f"X_train_res shape: {X_train_res.shape}")
print(f"y_train_res shape: {y_train_res.shape}")
```

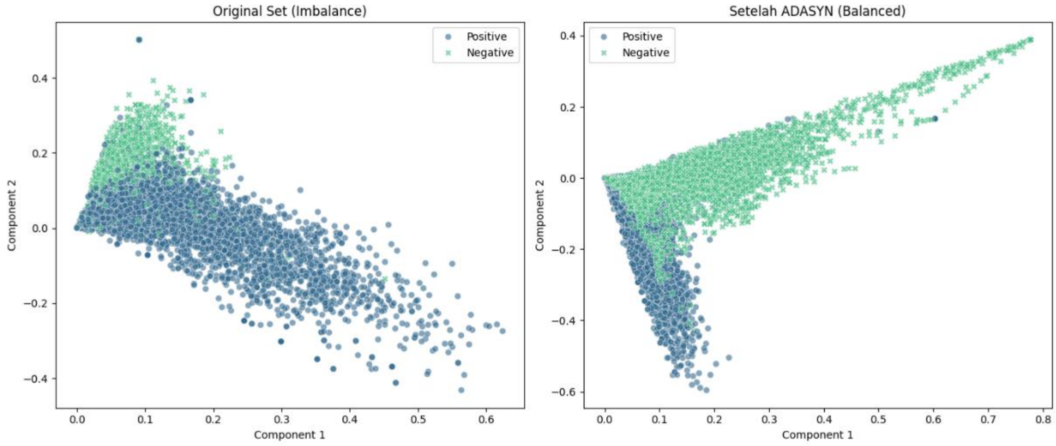
Gambar 4.20 Proses ADASYN

Pada tahap resampling, digunakan *balancing factor* sebesar 0,7, yang berarti jumlah data kelas minoritas ditingkatkan hingga mencapai 70% dari kelas mayoritas. Pendekatan ini berbeda dari skema penyeimbangan penuh (rasio 1:1) dan dipilih untuk meningkatkan representasi kelas minoritas secara proporsional, sekaligus meminimalkan risiko *overfitting* akibat penambahan data sintetik yang berlebihan (Chia, 2025).



Gambar 4.21 Distribusi Kelas setelah ADASYN

Evaluasi terhadap hasil resampling dilakukan melalui visualisasi sebaran data menggunakan teknik reduksi dimensi *Truncated Singular Value Decomposition* (TruncatedSVD) menunjukkan bahwa distribusi data kelas minoritas emenjadi lebih padat dan merata dibandingkan kondisi awal yang timpang, menandakan bahwa proses ADASYN berhasil mengintegrasikan vektor sintetik ke dalam dataset pelatihan secara efektif.



Gambar 4.22 Hasil resampling ADASYN

4.5.4 Naive Bayes

Tahap akhir pemodelan dilakukan melalui klasifikasi sentimen menggunakan algoritma Multinomial Naive Bayes. Proses ini dimulai dengan inisialisasi model, kemudian dilanjutkan dengan pelatihan menggunakan fungsi fit(). Model dilatih menggunakan data latih hasil resampling ADASYN (x_train_res dan y_train_res) agar distribusi kelas menjadi lebih seimbang dan mengurangi bias terhadap kelas mayoritas.

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report

mnb = MultinomialNB()

mnb.fit(X_train_res, y_train_res)

y_pred_train_mnb = mnb.predict(X_train_res)

y_pred_mnb = mnb.predict(X_test_tfidf)
```

Gambar 4.23 Proses Naive Bayes dengan ADASYN

```
mnb_no_adasyn = MultinomialNB()

mnb_no_adasyn.fit(X_train_tfidf, y_train)

y_pred_train_mnb_no_adasyn = mnb_no_adasyn.predict(X_train_tfidf)

y_pred_mnb_no_adasyn = mnb_no_adasyn.predict(X_test_tfidf)
```

Gambar 4.24 proses Naive bayes tanpa ADASYN

Selanjutnya, model digunakan untuk memprediksi sentimen pada data uji dengan membandingkan model baseline tanpa ADASYN dan model yang menerapkan ADASYN. Kedua model dievaluasi menggunakan data uji yang sama (X_test_tfidf) guna menilai secara objektif dampak penerapan ADASYN terhadap peningkatan kinerja klasifikasi dibandingkan model standar.

4.6 *Evaluasi*

Tahap Evaluasi berperan untuk memvalidasi kinerja model secara objektif. Tahap ini bertujuan mengukur sejauh mana model Naive Bayes yang telah dilatih dengan data hasil penyeimbangan ADASYN mampu mengenali sentimen secara tepat pada *data testing*.

4.6.1 Confusion Matrix

Evaluasi kinerja model dilakukan menggunakan confusion matrix untuk menilai akurasi dan kesalahan klasifikasi pada tiap kelas.

Tabel 4.1 Hasil *Accuracy*

		Naive Bayes	**Naive Bayes + ADASYN**
Accuracy	Training	0.91	0.92
	Testing	0.89	0.86

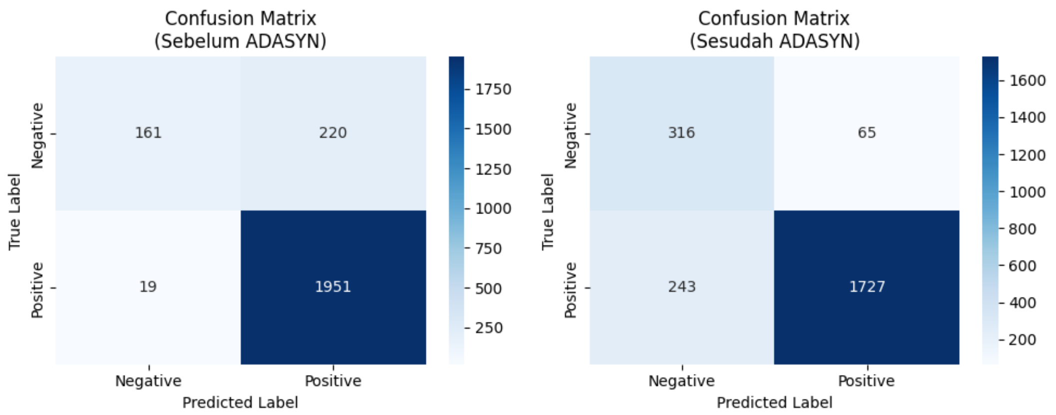
Tabel 4.1 menunjukkan bahwa model Baseline Naive Bayes mencapai akurasi sebesar 89%. Setelah penerapan metode ADASYN nilai *training accuracy* naik menjadi 92% dan *testing accuracy* turun mejadi 86%, hasil ini menyatakan bahwa model memberikan hasil yang *Good Fit*.

Tabel 4.2 Hasil *Classification Report*

		Precision	**Recall**	**F-1 Score**
NB	Positive	0.90	0.99	0.94
	Negative	0.89	0.42	0.57
NB + ADASYN	Positive	0.96	0.88	0.92
	Negative	0.57	0.83	0.67

Tabel 4.2 menampilkan hasil *baseline Naive Bayes* menunjukkan nilai *recall* kelas negatif sebesar 42%, sedangkan kelas positif mencapai 99%. Temuan ini mengindikasi adanya bias terhadap kelas positif, sehingga hampir seluruh data uji

diklasifikasikan sebagai sentimen positif. Setelah implementasi ADASYN, *recall* pada kelas negatif meningkat menjadi 83%, sementara *recall* kelas positif turun menjadi 88%.



Gambar 4.25 Visualisasi Confusion Matrix

Dari hasil confusion matriks, dilakukan perhitungan manual untuk membuktikan hasil yang didapat dalam memvalidasi nilai *Specifity* dari kelas negatif :

Tabel 4.3 Pebandingan Perhitungan Kelas Negatif

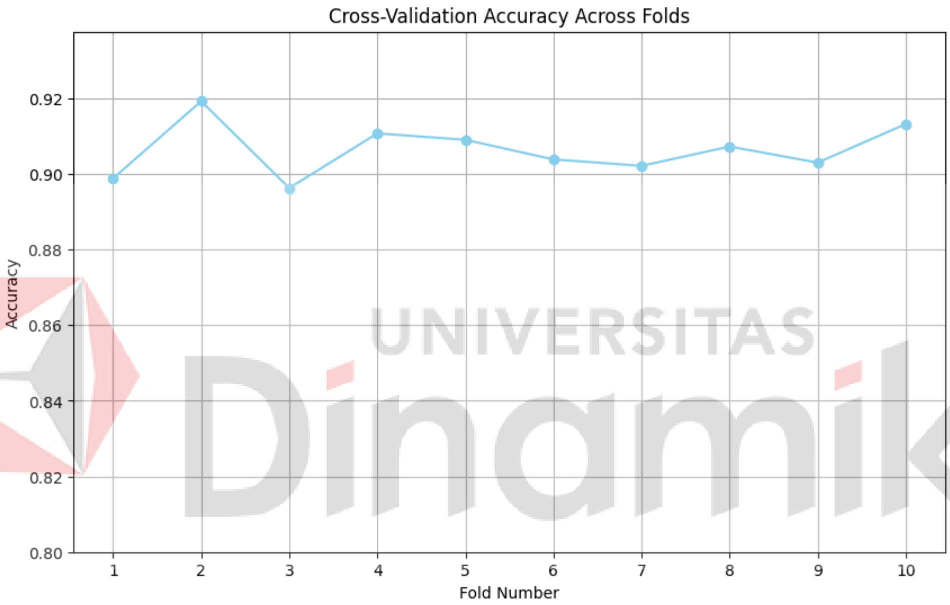
Sebelum ADASYN	Sesudah ADASYN
1. Precision = $\frac{TN}{TN+FN}$ Precision = $\frac{161}{161+19}$ Precision = 0.89	1. Precision = $\frac{TN}{TN+FN}$ Precision = $\frac{316}{316+243}$ Precision = 0.57
2. Specifity = $\frac{TN}{TN+FP}$ Specifity = $\frac{161}{161+19}$ Specifity = 0.42	2. Specifity = $\frac{TN}{TN+FP}$ Specifity = $\frac{316}{316+65}$ = 0.829 Specifity = 0.83
3. F1-Score = 2 $\times \frac{Precision \times Recall}{Precision+Recall}$ F1-Score = 2 $\times \frac{0.89 \times 0.42}{0.89+0.42}$ F1-Score = 0.57	3. F1-Score = 2 $\times \frac{Precision \times Recall}{Precision+Recall}$ F1-Score = 2 $\times \frac{0.57 \times 0.83}{0.57+0.83}$ F1-Score = 0.67

Hasil ini menunjukkan bahwa ADASYN berhasil mereduksi bias akibat *imbalance data* dan meningkatkan sesitifitas model terhadap sentimen negatif yang mengonfirmasi bahwa model mampu mendeteksi 316 data negatif hal ini dibuktikan secara empiris melalui lonjakan nilai *Recall* negatif dari 42% menjadi

83%. Namun, nilai *Precision* negatif turun pada angka 0.57, hal ini merupakan *trade-off* akibat perluasan *decision boundary* model agar lebih agresif dalam menangkap kelas negatif (Klu et al., 2025).

4.6.2 K-Fold Cross Validation

Untuk memastikan konsistensi kinerja model, dilakukan pengujian ulang menggunakan teknik *K-Fold Cross Validation* dengan nilai k=10 di mana model akan dilatih dan diuji secara bergiliran pada setiap *folds*.

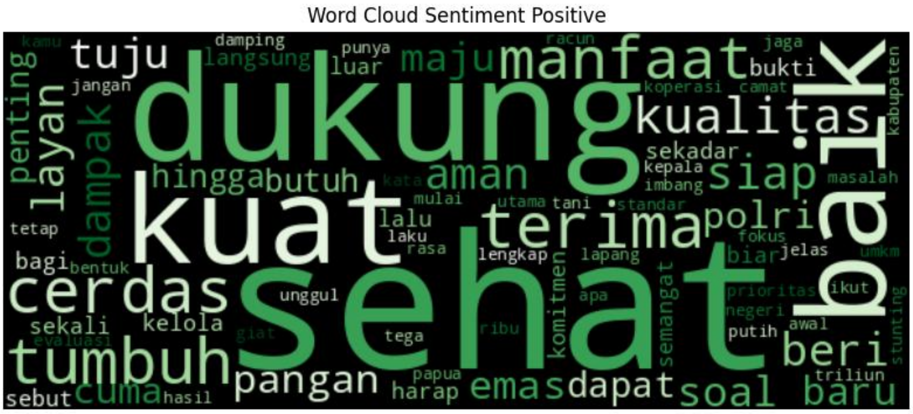


Gambar 4.26 hasil Pengujian *K-Fold Cross validation*

Gambar 4.26 memperlihatkan bahwa akurasi model bergerak cukup stabil pada rentang nilai 0.89 hingga 0.92 dengan *Mean Accuracy* di angka 0.90. Garis grafik tidak menunjukkan lonjakan atau penurunan yang drastis (ekstrem), yang mengindikasikan bahwa model Naive Bayes yang dibangun memiliki tingkat stabilitas yang baik (*robustness*) (Avula et al., 2022).

4.7 Visualisasi

Setelah melalui tahapan klasifikasi dan evaluasi, langkah terakhir dalam analisis ini adalah visualisasi data menggunakan Word Cloud.

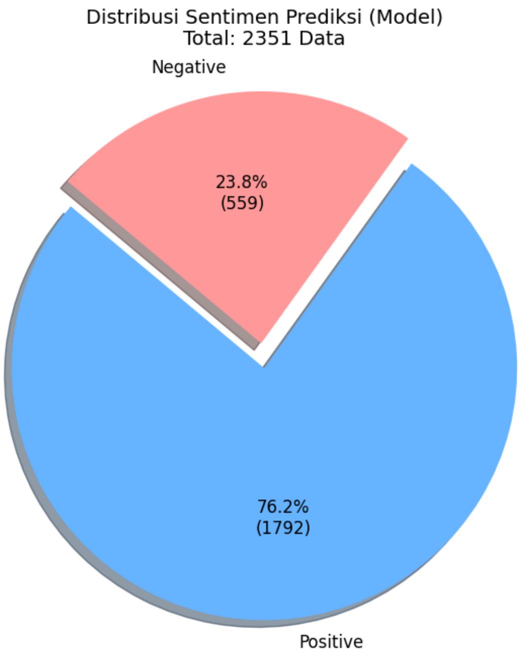


Gambar 4.27 *Wordcloud* Sentimen Positif

Visualisasi *WordCloud* sentimen positif kata yang paling mencolok ialah "sehat", "dukung", "baik", dan "kuat", kata sehat bagaimana pengaruh MBG terhadap kesehatan anak, lalu dukungan terhadap MBG dalam mewujudkan generasi emas, kata baik yang menyampaikan kinerja MBG dalam mencegah stunting agar dapat menjadi fondasi kuat dalam menjunjung pendidikan yang berkualitas.

Gambar 4.28 W*ordcloud* sentimen negatif

Visualisasi WordCloud sentimen negatif secara kontras didominasi oleh kata kata “racun”, “kasus”, dan “korban” yang merefleksikan kekhawatiran masyarakat terhadap insiden keamanan pangan dalam pelaksanaan MBG. Sementara itu, kemunculan kata “korupsi” menunjukkan tuntutan publik akan evaluasi menyeluruh terhadap transparansi anggaran yang digunakan dalam menjalankan program MBG



Gambar 4.29 Distribusi Sentimen

Berdasarkan hasil klasifikasi akhir menggunakan algoritma Naive Bayes, ditemukan dominasi sentimen positif. Visualisasi distribusi sentimen menunjukkan bahwa 76.2% (1.792 data) diklasifikasikan sebagai sentimen positif, sedangkan 23.8% (559 data) teridentifikasi sebagai sentimen negatif. Hasil ini mengindikasikan bahwa secara garis besar, program Makan Bergizi Gratis (MBG) mendapatkan penerimaan dan dukungan yang kuat dari masyarakat. Meskipun demikian, proporsi sentimen negatif tetap menjadi temuan krusial karena merepresentasikan adanya segmen masyarakat yang menyuarakan kritik dan kekhawatiran nyata terhadap implementasi program.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil ayng telah didapatkan, dapat disimpulkanbahwa :

1. Hasil klasifikasi tweet menunjukkan bahwa sentimen publik terhadap program MBG didominasi oleh sentimen positif. Dari total 2.351 data, sebanyak 76.2% mencerminkan apresiasi dan dukungan masyarakat terhadap pelaksanaan program MBG. Sementara itu, 23.8% sentimen bersifat negatif berisi berbagai kekhawatiran publik terkait pelaksanaan program, terutama mengenai kualitas makanan dan proses distribusi sehingga perlunya evaluasi terhadap berjalannya program.
2. Dari hasil evaluasi didapatkan bahwa *Naive Bayes* melakukan klasifikasi sentimen MBG dengan baik dengan *hasil Accuracy* 86%. hasil *k-fold cross validation* menunjukkan model stabil dengan mean Accuracy 90% menandakan kalau model *robustness*. Hasil ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik dan tidak mengalami *overfitting* saat diuji pada data baru.
3. Teknik oversampling menggunakan ADASYN berhasil menangani faktor *Data Imbalance* dalam mencegah model menjadi bias terhadap kelas mayoritas sehingga meningkatkan performa model khusunya *Recall* yang meningkat dari 42% menjadi 84% dan *F-1 Score* dari 57% menjadi 67% pada kelas negatif.
4. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 86%, melampaui capaian Sakina et al. (2025) sebesar 79,61% namun kompetitif dari penelitian Dwi Anggreny (2025) yang mencapai 91%, namun penelitian ini unggul dalam menangani *Imbalance data* melalui penerapan ADASYN, yang meningkatkan recall sentimen negatif dari 42% menjadi 84% sehingga mengurangi model bias ke kelas mayoritas.

5.2 Saran

Saran untuk pengembangan penelitian lebih lanjut antara lain :

1. Menggunakan Metode yang lebih advance seperti SVM, LSTM maupun XGBoost.
2. Penerapan ekstraksi fitur berbasis *word embedding* yang lebih mutakhir, seperti Word2Vec, GloVe, atau FastText.
3. Pelabelan data dengan metode yang lebih baik dalam memahami konteks kalimat, seperti menggunakan teknik labelling dengan IndoBERT.
4. Analisis berdasarkan aspek penelitian memberikan wawasan evaluasi yang jauh lebih detail dan tepat sasaran bagi pembuat kebijakan.

DAFTAR PUSTAKA

Agustini, U. (2025). Efektivitas dan Tantangan Kebijakan Program Makan Bergizi Gratis sebagai Intervensi Pendidikan di Indonesia. *Jurnal Kiprah Pendidikan*, *4*(3), 362–368. https://doi.org/10.33578/kpd.v4i3.p362-368

Amrullah, A. Z., Sofyan Anas, A., Adrian, M., & Hidayat, J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, *2*(1). https://doi.org/10.30812/bite.v2i1.804

Ataei, P., Regula, S., Staegemann, D., & Malgaonkar, S. (2024). Filtering Useful App Reviews Using Naïve Bayes—Which Naïve Bayes? *AI (Switzerland)*, *5*(4), 2237–2259. https://doi.org/10.3390/ai5040110

Avula, N. V. S., Veesam, S. K., Behera, S., & Balasubramanian, S. (2022). *Building Robust Machine Learning Models for Small Chemical Science Data: The Case of Shear Viscosity*. https://doi.org/10.1088/2632-2153/acac01

Azzahra T, A. (2025, January 4). *Makan Bergizi Gratis Mulai Serentak Seluruh Indonesia 6 Januari 2025*. https://news.detik.com/berita/d-7713769/makan-bergizi-gratis-mulai-serentak-seluruh-indonesia-6-januari-2025?utm_source=chatgpt.com

Bagus Trianto, R., Triyono, A., Malita Puspita Arum, D., Komputer, I., Sains dan Kesehatan, F., An Nuur, U., Gajah Mada No, J., Purwodadi, K., & Grobogan, K. (2021). *Kombinasi Metode K-Nearest Neighbor dengan Cosine Similarity untuk Prediksi Serangan Firewall pada Jaringan Komputer*. *6*(4), 672–679. https://doi.org/10.32493/informatika.v6i4.12680

Barros, W. K. P., Barbosa, M. T., Dias, L. A., & Fernandes, M. A. C. (2022). Fully Parallel Proposal of Naive Bayes on FPGA. *Electronics (Switzerland)*, *11*(16). https://doi.org/10.3390/electronics11162565

Chia, L. H. (2025). *Finding the Sweet Spot: Optimal Data Augmentation Ratio for Imbalanced Credit Scoring Using ADASYN*. http://arxiv.org/abs/2510.18252

Dawson, H. L., Dubrule, O., & John, C. M. (2023). Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. *Computers and Geosciences*, *171*. https://doi.org/10.1016/j.cageo.2022.105284

Dwi Anggreny, A. (2025). Sentiment Analysis Of Free Meal Program For School Students Using Algoritma Naive Bayes. In *International Journal of Science*.

E-Media DPR RI. (2025, November 2). *Banyak Kasus Keracunan, Ahli Gizi di SPPG Harus Bekerja Optimal Jaga Kualitas Makanan - E-Media DPR RI*. https://emedia.dpr.go.id/2025/10/02/banyak-kasus-keracunan-ahli-gizi-di-sppg-harus-bekerja-optimal-jaga-kualitas-makanan/?utm_source=chatgpt.com

FAHRUDIN, N. F., PUTRA, K. R., UMAROH, S., & LAUTAN, G. B. (2024). Influence of Data Scaling and Train/Test Split Ratios on LightGBM Efficacy for Obesity Rate Prediction. *MIND Journal*, *9*(2), 220–234. https://doi.org/10.26760/mindjournal.v9i2.220-234

Faqih Azhar, M., Nusantara Bakry, G., Raya Bandung Sumedang, J. K., & Barat, J. (2025). Pengaruh Penggunaan Twitter (X) terhadap Orientasi Pemilih Pemula pada Pemilihan Presiden 2024 The Effect of Twitter Use (X) on the Orientation of Young Voters in the 2024 Presidential Election. *Jurnal Sains Sosial Dan Humaniora)*, *9*(1), 101–112. https://doi.org/10.30595/jssh.v9i1.20385

Fauziyyah, A. K. (2020). ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON. *Jurnal Ilmiah SINUS*, *18*(2), 31. https://doi.org/10.30646/sinus.v18i2.491

Flach, P. (2012). *MACHINE LEARNING : The Art and Science of Algorithms that Make Sense of Data*.

Ga', M. E., Freadyanus Kasi, Y., Didimus, Y., & Epu, A. (2025). PROGRAM MAKAN BERGIZI GRATIS DI KABUPATEN NAGEKEO: TUJUAN DAN TAHAPANNYA. *Www.Jurnal.Umb.Ac.Id*, *5*(2), 84. www.jurnal.umb.ac.id

Hakim, A. M. (2025, August 28). *Program Makan Bergizi Gratis Mulai Retak: Anggaran Besar, Banyak Masalah - Jurnalis TV*. https://jurnalistv.id/nasional/6680/program-makan-bergizi-gratis-mulai-retak-anggaran-besar-banyak-masalah/?utm_source=chatgpt.com

Hasibuan, M. S., & Serdano, A. (2022). Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naive Bayes. *JRST*

(Jurnal Riset Sains Dan Teknologi), *6*(2), 199. https://doi.org/10.30595/jrst.v6i2.15145

Herdanto, Y. M. (2025, May 8). *Urgensi Evaluasi Pelaksanaan Program Makan Bergizi Gratis*. https://www.kompas.id/artikel/urgensi-evaluasi-pelaksanaan-program-makan-bergizi-gratis?utm_source=chatgpt.com

Ilyas Tri Khaqiqi, M., Hanum Harani, N., & Prianto, C. (2023). Performance Analysis and Development of QnA Chatbot Model Using LSTM in Answering Questions. *Indonesian Journal of Computer Science*. https://www.kaggle.com/datasets/sodolanangbjkatio/slang-indonesia,

Karima, S., & Mutiara, A. B. (2023). Comparison of Classification Algorithms for Predicting Indonesian Fake News using Balanced and Imbalanced Datasets. *Faktor Exacta*, *16*(1). https://doi.org/10.30998/faktorexacta.v16i1.16486

Klu, A. K., Affam, M., Ewusi, A., Ziggah, Y. Y., & Brempong, E. A. (2025). APPLICATION OF MACHINE LEARNING IN SOIL LIQUEFACTION PREDICTION: THE CASE OF ACCRA'S ACTIVE SEISMIC ZONES. In *Journal of Ghana Science Association* (Vol. 23, Issue 1).

Merawati, N. L. P., Amrullah, A. Z., & Ismarmiaty. (2021). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. *Jurnal RESTI*, *5*(1), 123–131. https://doi.org/10.29207/resti.v5i1.2587

Munir, B. (2025, April 7). *Refokus Program Makan Bergizi Gratis*. https://www.kompas.id/artikel/refokus-program-makanan-bergizi-gratis?utm_source=chatgpt.com

Nurhopipah, A., & Magnolia, C. (2022). JURNAL PUBLIKASI ILMU KOMPUTER DAN MULTIMEDIA PERBANDINGAN METODE RESAMPLING PADA IMBALANCED DATASET UNTUK KLASIFIKASI KOMENTAR PROGRAM MBKM. *JUPIKOM*, *1*(2).

Oktafiani, R., Hermawan, A., & Avianto, D. (2023). Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning. *Jurnal Sains Dan Informatika*, 19–28. https://doi.org/10.34128/jsi.v9i1.622

Putriyekti, A., Sulistiawati, D., Khairani, N., Rizky, R., Kusuma, A., Keuangan, P., & Stan, N. (2025). ANALISIS MULTIDIMENSIONAL SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN BERGIZI GRATIS PADA MEDIA SOSIAL X. *Integrative Perspectives of Social and Science Journal*, *2*(1), 1004.

Sakina, N., Wajidi, F., & Rasyid, Muh. R. (2025). Evaluasi Algoritma KNN dan Naive Bayes untuk Analisis Sentimen Kebijakan Program Makan Bergizi Gratis. *Jambura Journal of Informatics*, *1*(2), 83–97. https://doi.org/10.37905/jji.v1i2.34418

Satya Marga, N., Rahman Isnain, A., & Alita, D. (2021). Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA). *Abstrak*, *453*(4), 453–463. http://jim.teknokrat.ac.id/index.php/informatika

Shantika, F. S., & Abidin, Z. (2025). Sentiment Analysis of Jobstreet Application Reviews on Google Play Store Using Support Vector Machine Algorithm with Adaptive Synthetic. *Recursive Journal of Informatics*, *3*(1), 99–107. https://doi.org/10.15294/rji.v3i2.11891

Thoriq, E. M., Ratnawati, D. E., & Rahayudi, B. (2021). *Analisis Sentimen Opini Publik pada Media Sosial Twitter terhadap Vaksin Covid-19 menggunakan Algoritma Support Vector Machine dan Term Frequency-Inverse Document Frequency* (Vol. 5, Issue 12). http://j-ptiik.ub.ac.id

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., & Sachdev, S. (2008). Maternal and Child Undernutrition 2 Maternal and child undernutrition: consequences for adult health and human capital. *Www.Thelancet.Com*, *371*. https://doi.org/10.1016/S0140

Zakariah, M., AlQahtani, S. A., & Al-Rakhami, M. S. (2023). Machine Learning-Based Adaptive Synthetic Sampling Technique for Intrusion Detection. *Applied Sciences (Switzerland)*, *13*(11). https://doi.org/10.3390/app13116504