TUGAS AKHIR

SENTIMENT ANALYSIS OF PUBLIC PERCEPTION ON PRESIDENT PRABOWO SUBIANTO'S LEADERSHIP ON X AND YOUTUBE PLATFORMS USING DEEP LEARNING

11421024	**Dwi Abelliva Kamny**
11421042	**Loeis Avreldo Lubis**
11421047	**Ruth Karolina Sirait**

PROGRAM STUDI SARJANA TERAPAN TEKNOLOGI REKAYASA PERANGKAT LUNAK
FAKULAS VOKASI
INSTITUT TEKNOLOGI DEL

2025

TUGAS AKHIR

SENTIMENT ANALYSIS OF PUBLIC PERCEPTION ON PRESIDENT PRABOWO SUBIANTO'S LEADERSHIP ON X AND YOUTUBE PLATFORMS USING DEEP LEARNING

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Terapan Komputer (S.Tr.Kom)

11421024	**Dwi Abelliva Kamny**
11421042	**Loeis Avreldo Lubis**
11421047	**Ruth Karolina Sirait**

PROGRAM STUDI SARJANA TERAPAN TEKNOLOGI REKAYASA PERANGKAT LUNAK
FAKULTAS VOKASI
INSTITUT TEKNOLOGI DEL

2025

PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya kami sendiri. Seluruh sumber informasi yang dikutip atau dirujuk telah kami cantumkan dengan benar.

Nama : Dwi Abelliva Kamny
NIM : 11421024
Tanda Tangan :

Tanggal : 01 Juli 2025

Nama : Loeis Avreldo Lubis
NIM : 11421042
Tanda Tangan :

Tanggal : 01 Juli 2025

Nama : Ruth Karolina Sirait
NIM : 11421047
Tanda Tangan :

Tanggal : 01 Juli 2025

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

1	Nama	:	Dwi Abelliva Kamny
	NIM	:	11421024
2	Nama	:	Loeis Avreldo Lubis
	NIM	:	11421042
3	Nama	:	Ruth Karolina Sirait
	NIM	:	11421047
	Program Studi	:	Teknologi Rekayasa Perangkat Lunak
	Judul Tugas Akhir	:	*Sentiment Analysis of Public Perception on President Prabowo Subianto's Leadership on X and Youtube Platforms Using Deep Learning*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai salah satu syarat untuk memperoleh gelar Sarjana Terapan Komputer (S.Tr.Kom) pada Program Teknologi Rekayasa Perangkat Lunak, Fakultas Vokasi, Institut Teknologi Del.

DEWAN PENGUJI

Pembimbing	:	Oppir Hutapea, S.Tr.Kom., M.Kom	( )
Penguji 1	:	Ardiles Sinaga, S.T., M.T.	( )
Penguji 2	:	Yohanssen Pratama, S.Si, M.T	( )

Ditetapkan : Laguboti
Tanggal : 01 Juli 2025

PERNYATAAN PERSETUJUAN PUBLIKASI

Kami yang bertanda tangan di bawah ini, sebagai sivitas akademik Institut Teknologi Del:

1.	Nama	:	Dwi Abelliva Kamny
	NIM	:	11421024
2.	Nama	:	Loeis Avreldo Lubis
	NIM	:	11421042
3.	Nama	:	Ruth Karolina Sirait
	NIM	:	11421047
	Program Studi	:	Teknologi Rekayasa Perangkat Lunak
	Fakultas	:	Fakultas Vokasi
	Jenis Karya	:	Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Institut Teknologi Del **Hak Bebas Royalti Non Eksklusif** atas karya ilmiah saya yang berjudul: ***Sentiment Analysis of Public Perception on President Prabowo Subianto's Leadership on X and Youtube Platforms Using Deep Learning*** beserta seluruh perangkat yang menyertainya. Dengan Hak Non Eksklusif ini, Institut Teknologi Del berhak menyimpan, mengalihmediakan, mengubah format, serta mempublikasikannya dalam bentuk pangkalan data apapun, selama tetap mencantumkan nama kami sebagai penulis/ dan Hak Cipta.

Demikian pernyataan ini kami buat dengan sebenarnya.

Laguboti, 01 Juli 2025

Yang menyatakan

Dwi Abelliva Kamny	Loeis Avreldo Lubis	Ruth Karolina Sirait

ABSTRAK

Nama	:	1. Dwi Abelliva Kamny 2. Loeis Avreldo Lubis 3. Ruth Karolina Sirait
Program Studi	:	Teknologi Rekayasa Perangkat Lunak
Judul	:	*Sentiment Analysis of Public Perception on President Prabowo Subianto's Leadership on X and Youtube Platforms Using Deep Learning*

Media sosial kini menjadi ruang utama bagi masyarakat dalam menyampaikan opini terhadap kebijakan pemerintahan, membuka peluang pemanfaatan data digital sebagai sumber informasi untuk memahami persepsi publik secara sistematis. Penelitian ini mengembangkan sistem analisis otomatis berbasis *deep learning* untuk mengidentifikasi sentimen dan topik dalam komentar masyarakat Indonesia terhadap empat kebijakan awal Presiden Prabowo Subianto, yaitu Program Makan Bergizi Gratis, PPN 12%, Efisiensi Anggaran, dan Danantara. Komentar dikumpulkan dari platform X dan YouTube menggunakan teknik *crawling* dan *scraping*, lalu diproses melalui tahap pra-pemrosesan, pelabelan berbasis leksikon, serta pembagian menjadi data latih, validasi, dan uji. Model IndoBERT diterapkan dengan pendekatan *multi-task learning* untuk mengklasifikasikan sentimen positif, negatif, netral dan topik. Untuk menangani ketidakseimbangan data, digunakan metode SMOTE, sedangkan evaluasi dilakukan berdasarkan akurasi klasifikasi pada data uji dan eksperimen *hyperparameter*. Hasil menunjukkan bahwa model memiliki performa baik dengan akurasi 81,38% untuk klasifikasi topik dan 73,95% untuk klasifikasi sentimen. Analisis juga mengungkap bahwa persepsi publik terhadap keempat kebijakan cenderung negatif, terutama pada Efisiensi Anggaran dan PPN 12% yang dinilai paling membebani masyarakat. Sentimen negatif mendominasi sebesar 56,3%, dengan rincian pada Efisiensi Anggaran (63,4%), PPN 12% (58,3%), Makan Bergizi Gratis (55,0%), dan Danantara (50,9%). Penelitian ini menyarankan pengembangan model yang lebih sensitif terhadap komentar ambigu serta perluasan cakupan data untuk meningkatkan generalisasi.

Kata Kunci: Analisis sentimen; *Deep Learning*; IndoBERT; Kebijakan Publik; *Multi-task learning*

ABSTRACT

Name	:	1. Dwi Abelliva Kamny 2. Loeis Avreldo Lubis 3. Ruth Karolina Sirait
Study Program	:	*Software Engineering Technology*
Title	:	*Sentiment Analysis of Public Perception on President Prabowo Subianto's Leadership on X and Youtube Platforms Using Deep Learning*

Social media has become a primary space for the public to express opinions about government policies, opening opportunities to utilize digital data as a source of information for systematically understanding public perception. This study develops an automated analysis system based on deep learning to identify sentiments and topics within Indonesian public comments on four of President Prabowo Subianto's early policies: the Free Nutritious Meal Program, the 12% VAT, Budget Efficiency, and Danantara. Comments were collected from X and YouTube platforms using crawling and scraping techniques, followed by pre-processing, lexicon-based labeling, and splitting into training, validation, and testing datasets. The IndoBERT model was implemented using a multi-task learning approach to classify sentiment (positive, negative, neutral) and policy topics. To address data imbalance, the SMOTE method was employed, and model evaluation was conducted based on classification accuracy on the test set and hyperparameter tuning experiments.
Results show that the model performs well, achieving an accuracy of 81.38% for topic classification and 73.95% for sentiment classification. The analysis also reveals that public perception toward the four policies tends to be negative, particularly regarding Budget Efficiency and the 12% VAT, which are perceived as most burdensome. Negative sentiment dominated with 56.3%, distributed as follows: Budget Efficiency (63.4%), 12% VAT (58.3%), Free Nutritious Meal (55.0%), and Danantara (50.9%). This study recommends developing models that are more sensitive to ambiguous comments and expanding data coverage to improve generalizability.

Keywords: *Sentiment Analysis, IndoBERT, Social Media, Multitask Learning, Deep learning.*

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan Tugas Akhir yang berjudul *"Sentiment Analysis of Public Perception on President Prabowo Subianto's Leadership on X and Youtube Platforms Using Deep Learning*". Penulis menyadari bahwa tanpa bantuan, dukungan, dan bimbingan dari berbagai pihak, baik selama masa perkuliahan maupun dalam proses penyusunan Tugas Akhir ini, penyelesaian karya ini akan sangat sulit dicapai. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Oppir Hutapea, S.Tr.Kom., M.Kom., selaku dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing serta mengarahkan penulis dalam menyusun Tugas Akhir ini.
2. Bapak Ardiles Sinaga, S.T., M.T. dan Bapak Yohanssen Pratama, S.Si., M.T., selaku dosen penguji yang telah memberikan kritik dan saran yang membangun dalam menyusun Tugas Akhir ini.
3. Orang tua dan keluarga penulis yang senantiasa memberikan doa, semangat, dan dukungan tanpa henti.
4. Teman-teman dan sahabat yang telah membantu serta memberikan motivasi selama proses penyusunan Tugas Akhir ini.

Penulis berharap Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang berkepentingan serta menjadi kontribusi yang berarti dalam pengembangan ilmu pengetahuan, khususnya di bidang teknologi.

Laguboti, 01 Juli 2025

Penulis 1	Penulis 2	Penulis 3
Dwi Abelliva Kamny	Loeis Avreldo Lubis	Ruth Karolina Sirait

DAFTAR ISI

PERNYATAAN ORISINALITAS **3**

HALAMAN PENGESAHAN **4**

PERNYATAAN PERSETUJUAN PUBLIKASI **5**

ABSTRAK **6**

ABSTRACT **7**

KATA PENGANTAR **9**

DAFTAR ISI **10**

DAFTAR TABEL **15**

DAFTAR GAMBAR **16**

BAB I PENDAHULUAN **19**

1.1 Latar Belakang 19

1.2 Rumusan Masalah 22

1.3 Tujuan 23

1.4 Lingkup 23

1.5 Pendekatan 23

1.6 Sistematika Penyajian 25

1.7 Hipotesis 26

BAB II STUDI LITERATUR **28**

2.1 Media Sosial 28

2.2 Platform Media X 29

2.3 Platform Media YouTube 29

2.4 Analisis Sentimen 30

2.5 Pendekatan dan Model Analisis Sentimen 32

2.5.1 Naive Bayes 32

2.5.2 Support Vector Machine (SVM) 32

2.5.3 Convolutional Neural Network (CNN) 33

2.5.4 Long Short-Term Memory (LSTM) 34
2.5.5 Pembelajaran Transfer (BERT) 34
2.5.6 Alasan Pemilihan IndoBERT 35
2.6 Deep Learning 36
2.7 Data Pre-Processing 38
2.7.1 Data Cleaning 38
2.7.2 Normalisasi 39
2.7.3 Tokenization 40
2.7.4 Padding & Truncation 40
2.7.5 Encoding to Tensor 40
2.8 Unsupervised 41
2.9 Supervised 41
2.10 Pengumpulan Data 42
2.11 Ekonomi dan Kebijakan Terkait 43
2.11.1 Efisiensi Anggaran 43
2.11.1.1 Data Media Sosial X Program Efisiensi Anggaran 45
2.11.1.2 Data Media Sosial YouTube Program Efisiensi Anggaran 46
2.11.1.3 Data Media Daring Program Efisiensi Anggaran 49
2.11.2 Kenaikan PPN 12% 52
2.11.2.1 Data Media Sosial X Program PPN 12% 53
2.11.2.2 Data Media Sosial Youtube Program PPN 12% 56
2.11.2.3 Data Media Daring Program PPN 12% 60
2.11.3 Program Makan Bergizi Gratis 61
2.11.3.1 Data Media Sosial X Program Makan Bergizi Gratis 62
2.11.3.2 Data Media Sosial YouTube Program Makan Bergizi Gratis 66
2.11.3.3 Data Media Daring Program Makan Bergizi Gratis 69
2.11.4 Danantara 71
2.11.4.1 Data Media Sosial X Program Danantara 72

2.11.4.2 Data Media Sosial YouTube Program Danantara 74
2.11.4.3 Data Media Daring Program Danantara 77
2.12 Lexicon Based 79
2.13 Arsitektur Model IndoBERT Multitask 82
2.14 Hugging Face 84
2.15 Bidirectional Encoder Representations 84
2.15.1 IndoBERT 87
2.16 Hyperparameter Tuning 88
2.17 Confusion Matrix 88
2.18 Review of Related Work 89
BAB III METODOLOGI 91
3.1 Analisis dan Justifikasi 91
3.1.2 Perbandingan 10 Kebijakan Prabowo Berdasarkan Trends 93
3.1.3 Program Makan Bergizi Gratis 97
3.1.3.1 Data Google Trends Program Makan Bergizi Gratis 98
3.1.4 Program PPN 12% 101
3.1.4.1 Data Google Trends Program PPN 12% 102
3.1.5 Program Danantara 106
3.1.5.1 Data Google Trends Program Danantara 107
3.1.6 Program Efisiensi Anggaran 110
3.1.6.1 Data Google Trends Program Efisiensi Anggaran 111
3.1.7 Tagar Platform X Sebagai Indikator Kebijakan Presiden Prabowo 114
3.1.7.1 Kebijakan Danantara 115
3.1.7.2 Kebijakan Makan Bergizi Gratis 116
3.1.7.3 Kebijakan PPN 12% 118
3.1.7.4 Kebijakan Efisiensi Anggaran 119
3.2 Pendekatan Penelitian 120
3.2.1 Model IndoBERT 121

3.3 Metodologi Pengumpulan Data.....122
3.3.1 Studi Literatur 122
3.3.2 Crawling Data 123
3.3.3 Scraping Data 127
3.4 Tahap Pengolahan Data.....129
3.4.1 Tahapan Translate 129
3.4.2 Case Folding.....130
3.4.3 Data Cleaning.....130
3.4.4 Normalisasi.....130
3.4.5 Tokenizing.....133
3.4.6 Stopword 133
3.4.7 Stemming 134
3.5 Lexicon Based Labelling.....135
3.6 Tahapan Split.....137
3.7 Tahapan Encoded 138
3.8 Tahapan SMOTE.....139
3.9 Tahapan Modelling 140
3.9.1 Tahapan Persiapan Dataset.....141
3.9.2 Tahapan Tokenization dan Word Embedding dengan IndoBERT..142
3.9.3 Tahapan Arsitektur Model Multi-Task IndoBERT.....142
3.9.4 Tahapan Pelatihan Model.....144
3.9.5 Tahapan Hyperparameter Tuning.....145
3.9.6 Evaluasi Model.....146
3.9.7 Tahapan Implementasi 146
BAB IV IMPLEMENTASI 147
4.1 Lingkungan Implementasi.....147
4.2 Pengolahan Data.....147
4.2.1 Implementasi Pengumpulan Data.....148

4.2.2 Implementasi Preprocessing.....150
4.2.3 Implementasi Labeling.....152
4.2.4 Implementasi Data Splitting.....154
4.2.5 Implementasi Mapping.....155
4.2.6 Implementasi SMOTE.....156
4.2.7 Implementasi Modeling.....157
4.2.7.1 Implementasi Load Dataset.....157
4.2.7.2 Implementasi Tokenization.....158
4.2.7.3 Implementasi Dataset Class.....158
4.2.7.4 Implementasi Pelatihan dan Tuning.....160
4.2.7.5 Implementasi Evaluation.....163
4.2.8 Implementasi Akhir.....165
BAB V HASIL.....168
5.1 Hasil Analisis Sentimen Setiap Program.....168
5.1.1 Program Anggaran.....168
5.1.2 Program Danantara.....169
5.1.3 Program PPN 12%.....169
5.1.4 Program Makanan Bergizi Gratis.....170
5.1.5 Keempat Kinerja.....170
5.2 Confusion Matrix.....172
5.2.1 Confusion Matrix Sentiment.....172
5.2.2 Confusion Matrix Topic.....174
5.3 Evaluasi.....176
5.4 Perbandingan Model Tanpa SMOTE dan Dengan SMOTE.....180
5.4.1 Latar Belakang dan Tujuan Perbandingan.....180
5.4.2 Eksperimen Tanpa SMOTE.....181
5.4.3 Eksperimen Dengan SMOTE dan Tuning Hyperparameter.....182
5.4.4 Analisis.....184

5.5 Implementasi 186
5.6 Pengujian Hipotesis Berdasarkan Hasil Analisis Sentimen 187
BAB VI PEMBAHASAN 189
6.1 Analisis Hasil Sentimen pada Setiap Program 189
6.2 Keefektifan Model BERT dalam Analisis Sentimen 190
6.3 Rekomendasi untuk Penelitian Lanjutan 191
6.4 Tantangan dan Solusi 191
BAB VII KESIMPULAN DAN SARAN 192
7.1 Kesimpulan 192
7.2 Saran 193
DAFTAR PUSTAKA 195

DAFTAR TABEL

Tabel 2.1 Perbandingan Model Sentimen Bahasa Indonesia(2024) 36
Tabel 2.2 Rumus confusion matrix 89
Tabel 2.3 Review of Related Work.....90
Tabel 3.1 Dasar Permasalahan 10 Kebijakan.....93
Tabel 3.2 Jumlah Komentar Hasil Crawling Berdasar Topik 125
Tabel 3.3 Hasil Crawling Berdasarkan Topik.....126
Tabel 3.4 Jumlah Komentar Hasil Scraping Berdasar Topik.....128
Tabel 3.5 Hasil Scraping Berdasarkan Topik.....129
Tabel 3.6 Normalisasi Kata Tidak Baku 133
Tabel 4.1 Hasil Preprocessing.....152
Tabel 4.2 Hasil Labeling 154
Tabel 4.3 Hasil Data Split 155
Tabel 4.4 Hasil Pelatihan dan Tuning Model.....164
Tabel 5.1 Classification Report - Sentimen 178
Tabel 5.2 Classification Report - Topik.....180
Tabel 5.3 Hasil Pelatihan Model Tanpa SMOTE 183
Tabel 5.4 Hasil dengan SMOTE dan Tuning Hyperparameter.....185

DAFTAR GAMBAR

Gambar 2.1 Hubungan Antara DL, ML, dan AI 38
Gambar 2.2 Unggahan tweet @Strategi_Bisnis.....46
Gambar 2.3 Kanal KOMPAS TV - Efisiensi Anggaran Pendidikan 47
Gambar 2.4 Kanal CNN Indonesia - ASN Dukung Efisiensi Anggaran ...48
Gambar 2.5 Kanal Kompas.com - Kebijakan Efisiensi Anggaran.....49
Gambar 2.6 Artikel dari Reuters 51
Gambar 2.7 Artikel dari KOMPAS.com.....52
Gambar 2.8 Tweet @KemenkeuR - Kebijakan PPN 12%.....54
Gambar 2.9 SINDONEWS.com - Penolakan PPN 12%.....55
Gambar 2.10 Kanal Metro TV - PPN 12% 57
Gambar 2.11 Kompas TV - Dampak PPN 12%.....58
Gambar 2.12 Kompas TV - Pemberlakuan PPN 12% 59
Gambar 2.13 Artikel dari KOMPAS.com.....60
Gambar 2.14 Artikel dari CNBC Indonesia.....61
Gambar 2.15 @tempodotco - Kasus keracunan program MBG 63
Gambar 2.16 @CNNIndonesia - Sorotan kasus keracunan MBG 64
Gambar 2.17 @kompascom - Sorotan Internasional terhadap MBG 65
Gambar 2.18 CNN Indonesia - Mitra MBG Tidak Dibayar RP1 M.....66
Gambar 2.19 Liputan 6 - Prabowo & Bill Gates Tinjau MBG 67
Gambar 2.20 Tempodotco - Program MBG 68
Gambar 2.21 Artikel dari The Straits Times.....69
Gambar 2.22 Artikel dari The Guardian 70
Gambar 2.23 @kompascom- Tentang Danantara.....73
Gambar 2.24 MetroTV - Sistem Danantara 75
Gambar 2.25 CNN Indonesia - Apa itu Danantara?.....76
Gambar 2.26 Unggahan - Risiko Tinggi Proyek Danantara 77

Gambar 2.27 Artikel dari Reuters 78
Gambar 2.28 Artikel dari MetroTV 79
Gambar 2.29 Arsitektur Multitask Learning 83
Gambar 2.30 Proses Pre-training dan Fine Tuning 86
Gambar 3.1 Trend Google 5 Kebijakan (1).....95
Gambar 3.2 Trend Google 5 Kebijakan (2).....95
Gambar 3.3 Trends “Makan Bergizi Gratis”.....99
Gambar 3.4 Trend “MBG” - 5 Sub Wilayah Teratas 100
Gambar 3.5 Tren “MBG” - Topik Teratas 101
Gambar 3.6 Trends “PPN 12%”.....103
Gambar 3.7 Trend “PPN 12%” - 5 Wilayah Teratas.....105
Gambar 3.8 Trends “PPN 12%” - Topik Teratas 106
Gambar 3.9Trends “Danantara”.....108
Gambar 3.10 Trends “Danantara” - 5 Sub Wilayah Teratas 109
Gambar 3.12 Trends “Efisiensi Anggaran”.....112
Gambar 3.13 Trends “Efisiensi Anggaran” - 5 Sub Wilayah Teratas.....113
Gambar 3.14 Trends “Efisiensi Anggaran” Topik Teratas 114
Gambar 3.15 Trends Tagar #DanantaraBangunNegeri di Platform X....116
Gambar 3.16 Trends Tagar #MakanBergiziBangun di Platform X 117
Gambar 3.17 Trends Tagar #PajakMencekik pada Platform X 118
Gambar 3.18 TrendsTagar #EfisiensiUntukBangsa di Platform X.....119
Gambar 3.19 Tahapan analisis sentimen.....121
Gambar 3.20 Pengumpulan data X dan YouTube.....123
Gambar 3.21 Tahapan Crawling Data.....124
Gambar 3.22 Tahapan Scraping Data 127
Gambar 3.23 Tahapan Preprocessing.....129
Gambar 3.24 Diagram Alur Split 138
Gambar 3.25 Tahapan Modelling.....141

Gambar 4.1 Hasil Crawling Data 149
Gambar 4.2 Hasil Scraping Data 149
Gambar 5.1 Diagram Pada Program Anggaran 168
Gambar 5.2 Diagram Pada Program Danantara 169
Gambar 5.3 Diagram Pada Program PPN 12% 169
Gambar 5.4 Diagram Pada Program Makan Bergizi Gratis 170
Gambar 5.5 Diagram Pada Keempat Program 171
Gambar 5.6 Confusion Matrix Sentiment 172
Gambar 5.7 Confusion Matrix Topic 174
Gambar 5.8 Implementasi Form Analisis 186

BAB I
PENDAHULUAN

Dalam bagian ini, akan dijelaskan latar belakang dari topik tugas akhir, rumusan masalah, tujuan, lingkup, pendekatan dalam penanganan masalah, dan penyajian dokumen penelitian.

1.1 Latar Belakang

Media sosial adalah platform digital yang memainkan peran penting dalam politik dan partisipasi publik. Hal ini memungkinkan pengguna untuk terlibat dalam proses politik, membangun citra politik, dan berpartisipasi dalam kampanye. Platform media sosial seperti Facebook, Twitter, dan YouTube digunakan secara luas untuk kampanye politik dan keterlibatan publik, mempengaruhi hasil pemilu dengan menyediakan ruang bagi opini publik dan analisis sentimen [1]. Di Indonesia, penggunaan media sosial sangat signifikan, dengan platform seperti YouTube yang memiliki basis pengguna yang besar, menjadikannya alat yang efektif untuk komunikasi dan partisipasi politik [2].

Media sosial dapat mempengaruhi proses politik melalui konten yang diunggah, yang dapat membentuk opini publik dan mempengaruhi hasil pemilu. Hal ini menimbulkan kekhawatiran tentang manipulasi informasi dan penyebaran berita palsu yang dapat mempengaruhi persepsi publik. Meskipun media sosial memungkinkan partisipasi politik yang lebih luas, ada kekhawatiran bahwa partisipasi ini mungkin tidak selalu mencerminkan tindakan politik yang nyata. Aktivitas *online* mungkin tidak selalu diterjemahkan menjadi tindakan politik di dunia nyata, seperti pemungutan suara.

Ada tantangan dalam menjaga etika media di platform sosial, terutama terkait dengan penyebaran informasi yang tidak diverifikasi dan bias media. Ini dapat mempengaruhi kepercayaan publik terhadap informasi yang disebarkan melalui media sosial [3]. Media sosial juga dapat digunakan untuk membahas isu-isu sensitif seperti dinasti politik, yang dapat menimbulkan perdebatan tentang penyalahgunaan kekuasaan dan pengaruh politik dalam pemilu. Masalah-masalah

ini menunjukkan bahwa meskipun media sosial memiliki potensi besar untuk meningkatkan partisipasi politik dan keterlibatan publik, ada tantangan signifikan yang perlu diatasi untuk memastikan bahwa pengaruhnya positif dan konstruktif [4].

Kondisi politik di Indonesia yang kian memanas menjelang pemilu 2024 telah melalui tahapan penting yang dijadwalkan sejak akhir tahun 2022. Partai politik pun telah melakukan berbagai manuver untuk bisa memenangkan kontestasi politik ini. Sejauh ini, tiga nama bakal calon presiden telah dideklarasikan oleh masing-masing partai pengusungnya, dan pemberitaan media pun ramai mengangkat isu tentang hasil survei calon presiden sebagai konten menarik yang meningkatkan *traffic*. Tak terlepas dari peran kepemilikan dan konglomerasi media di Indonesia, di mana sebagiannya juga aktif berpolitik, pembingkaian berita hasil *polling* dapat mempengaruhi persepsi publik terhadap para bakal calon presiden yang dideklarasikan [5].

Masyarakat diberikan ruang untuk berperan aktif dan menjadi bagian dari proses demokrasi, meskipun secara substansial, keikutsertaan masih cenderung prosedural dan momentum. Sesuai dengan prinsip demokrasi yang menjadi dasar negara Indonesia, pemilihan umum akan diselenggarakan pada tahun 2024. Dalam pelaksanaan pemilu, terdapat kegiatan kampanye yang bertujuan untuk mempengaruhi suara publik agar memilih kandidat terbaik berdasarkan opini masyarakat. Media sosial Twitter menjadi salah satu sarana untuk menyuarakan pendapat serta berbagi informasi, sehingga dapat berfungsi sebagai platform kampanye tidak langsung.

Setelah Pemilu Presiden 2024, pasangan Prabowo Subianto dan Gibran Rakabuming Raka resmi terpilih sebagai Presiden dan Wakil Presiden Republik Indonesia. Seperti halnya pada masa pemerintahaan sebelumnya, 100 hari pertama menjadi momen yang sangat penting. Masyarakat dan media menaruh perhatian besar pada kebijakan-kebijakan yang diambil di awal masa jabatan, karena dianggap mencerminkan arah pemerintahan ke depan. Selama periode 100 hari tersebut, terdapat beberapa kebijakan yang menarik perhatian publik. Penelitian ini secara khusus memfokuskan pada empat kebijakan utama yaitu Kenaikan Pajak

Pertambahan Nilai (PPN) menjadi 12%, Program Makan Bergizi Gratis untuk siswa, Danantara, dan Kebijakan Efisiensi Anggaran pemerintah.

Keempat kebijakan ini tidak hanya memiliki pengaruh yang luas terhadap kehidupan masyarakat, tetapi juga memunculkan reaksi beragam di media sosial. Misalnya, kebijakan PPN 12% banyak menuai kritik karena dianggap membebani masyarakat kelas menengah ke bawah. Sebaliknya, program Makan Bergizi Gratis mendapatkan banyak dukungan karena bertujuan untuk meningkatkan kesehatan dan gizi anak-anak sekolah. Proyek Danantara dilihat sebagai langkah digitalisasi nasional, sementara kebijakan efisiensi anggaran dinilai penting dalam mencegah pemborosan keuangan negara.

Reaksi masyarakat terhadap keempat kebijakan ini sangat terlihat di media sosial melalui berbagai komentar, unggahan, dan tagar populer seperti #PPN12Persen, #MakanBergizi, #Danantara, dan #EfisiensiAnggaran. Volume dan intensitas percakapan ini menunjukkan bahwa masyarakat sangat memperhatikan langkah-langkah pemerintah sejak awal masa jabatan.

Tingginya volume dan intensitas diskusi di media sosial menunjukkan bahwa masyarakat memperhatikan secara aktif langkah-langkah pemerintahan sejak awal masa jabatan. Namun, karena banyaknya komentar yang muncul, pemahaman terhadap opini publik menjadi tidak mudah. Ragam bahasa, nada, serta sudut pandang yang digunakan membuat analisis secara manual menjadi tidak efektif.

Oleh karena itu, peneliti ini menggunakan pendekatan analisis sentimen untuk memahami opini publik secara sistematis. Model yang digunakan adalah IndoBERT, yaitu model bahasa berbasis BERT yang dikembangkan khusus Bahasa Indonesia. IndoBERT mampu membaca dan mengklasifikasikan komentar menjadi sentimen positif, negatif, atau netral serta mengenali topik-topik yang dibahas dalam komentar seperti PPN 12%, Makan Bergizi Gratis, dan Efisiensi Anggaran.

Data yang relevan untuk analisis sentimen, penelitian ini menerapkan teknik *crawling* data digunakan untuk mengumpulkan opini publik dari platform X dan YouTube. Data yang dikumpulkan berfokus pada komentar pengguna di kedua

platform tersebut yang berkaitan dengan kepemimpinan Prabowo Subianto. Untuk platform X, data komentar diperoleh melalui pengguna API dengan memasukan kata kunci tertentu terkait kepemimpinan Prabowo. Sementara itu, data komentar dari platform YouTube dikumpulkan menggunakan platform *Apify* dengan cara memasukkan API dari suatu konten YouTube yang relevan. Pemanfaatan teknik *crawling* ini memungkinkan pengumpulan data secara otomatis dalam jumlah besar, sehingga menyajikan gambaran persepsi publik yang lebih luas dan objektif.

Dengan demikian, data yang terkumpul dapat diolah dan dianalisis lebih lanjut untuk menghasilkan wawasan yang lebih akurat dan mendalam mengenai sentimen masyarakat terhadap kepemimpinan Prabowo. *Urgensi* penelitian ini juga muncul karena adanya beragam opini di media sosial, baik positif maupun negatif, mengenai kepemimpinan Prabowo Subianto. Publik seringkali mempertanyakan ketepatan kebijakan dan kinerja pemerintah. Tanpa analisis yang sistematis, opini-opini ini dapat menjadi bias dan tidak berdasarkan data yang akurat. Oleh karena itu, penelitian ini bertujuan untuk menyediakan metode yang lebih objektif dalam memahami sentimen publik secara luas.

1.2 Rumusan Masalah

Berikut merupakan rumusan masalah berdasarkan latar belakang tentang analisis sentimen terhadap kepemimpinan Prabowo Subianto di media sosial.

1. Bagaimana persepsi masyarakat terhadap kepemimpinan Prabowo Subianto yang diungkapkan melalui komentar-komentar di media sosial berdasarkan algoritma yang ditentukan?
2. Bagaimana kinerja model IndoBERT dalam melakukan analisis sentimen terhadap teks komentar berbahasa Indonesia?

1.3 Tujuan

Penelitian ini memiliki dua tujuan utama, yaitu :

1. Mengetahui dan menganalisis persepsi masyarakat terhadap kebijakan-kebijakan yang diterapkan pada 100 hari pertama pemerintah Prabowo Subianto dan Gibran Rakabuming Raka, sebagaimana tercermin dalam komentar-komentar masyarakat di media sosial, khususnya yang berkaitan

dengan Program Makan Bergizi Gratis, kenaikan PPN 12%, Danantara, dan Kebijakan Efisiensi Anggaran.

2. Mengevaluasi kinerja model IndoBERT dalam mengklasifikasikan sentimen teks berbahasa Indonesia, guna memperoleh pemetaan opini publik yang lebih objektif, sistematis, dan terukur.

1.4 Lingkup

Lingkup pengerjaan tugas akhir ini adalah sebagai berikut:

1. Data diambil dari platform X dan Youtube.
2. Hanya komentar berbahasa Indonesia yang dianalisis.
3. Pengumpulan data 14 Januari 2024 - 20 Februari 2025.
4. Fokus pada empat kebijakan: Danantara, PPN 12%, Makan Bergizi Gratis, dan Efisiensi Anggaran.
5. Metode pengumpulan : *crawling* dan *scrapping* disimpan di Google Drive.
6. Metode analisis IndoBERT.
7. Bahasa pemrograman Python dan Tools Google Collaboratory.

1.5 Pendekatan

Dalam penelitian ini, metode analisis sentimen dengan menggunakan model BERT diterapkan melalui beberapa tahapan penting. Proses dimulai dengan pengumpulan data yang diambil dari platform X dan YouTube menggunakan teknik *crawling, scraping,* dan API untuk memperoleh komentar-komentar serta persepsi publik terkait kepemimpinan Presiden Prabowo Subianto. Selanjutnya, dilakukan tahap *pre-processing* untuk membersihkan dan mentransformasi data agar siap digunakan dalam pelatihan model. Proses ini mencakup normalisasi, penghapusan elemen yang tidak relevan, serta penyesuaian format data bila diperlukan.

Tahapan berikutnya adalah pelabelan data (*data labeling*), yaitu proses mengklasifikasikan komentar ke dalam kategori sentimen seperti positif, negatif, atau netral. Proses pelabelan ini dapat dilakukan secara manual maupun dengan bantuan pendekatan berbasis kamus (*lexicon-based*) untuk mempercepat otomatisasi. Setelah data diberi label, dilakukan pembagian dataset menjadi tiga bagian, yaitu data latih, data validasi, dan data uji, masing-masing dengan proporsi 70%, 20%, dan 10%. Pembagian ini bertujuan agar model dapat belajar dari data

latih dan diuji pada data yang tidak pernah dilihat sebelumnya untuk mengukur kemampuan generalisasi model.

Tahap berikutnya adalah konversi label ke dalam bentuk numerik (*mapping*), agar dapat dikenali oleh model dalam proses pelatihan. Label sentimen dan label topik yang semula berbentuk teks diubah menjadi angka sesuai kategori masing-masing. Karena distribui data sentimen dalam dataset tidak simbang dengan dominasi pada kelas negatif dilakukan teknik *oversampling* menggunakan SMOTE (*Synthetic Minority Oversampling Technique*) secara eksklusif pada data latih. Hal ini bertujuan untuk mencegah bias model terhadap kelas mayoritas, sementara data validasi dan data uji tetap dibiarkan dalam bentuk aslinya untuk menjaga objektivitas evaluasi.

Penerapan model BERT dalam penelitian ini dilakukan dengan menggunakan IndoBERT, yaitu varian BERT dalam penelitian ini dilakukan dengan menggunakan IndoBERT, yaitu varian BERT yang telah dilatih khusus untuk Bahasa Indonesia. Model ini memiliki arsitektur *transformer* dua arah yang mampu memahami konteks kata dalam kalimat secara lebih mendalam. Proses pelatihan mencakup penyetelan *hyperparameter* seperti *learning rate, batch size,* dan jumlah *epoch* untuk memperoleh konfigurasi terbaik. Selama proses pelatihan, data validasi digunakan untuk memantau kinerja model secara berkala, dan strategi *early stopping* diterapkan agar proses berhenti otomatis apabila model tidak menunjukkan peningkatan performa setelah beberapa *epoch* berturut-turut.

Maka dari itu, dilakukannya evaluasi terhadap performa model dengan menggunakan data uji. Evaluasi ini bertujuan untuk mengukur kemampuan model dalam mengklasifikasikan sentimen dan topik pada data baru. Hasil prediksi kemudian dianalisis menggunakan *confusion matrix* yang menjadi dasar dalam perhitungan metrik evaluasi seperti *accuracy, precision, recall,* dan *F1-score.*

1.6 Sistematika Penyajian

Laporan ini disusun secara sistematis, ke dalam beberapa bab yang saling berkaitan untuk memberikan gambaran alur penelitian secara utuh dan menyeluruh.

Bab I Pendahuluan berisi latar belakang penelitian, perumusan masalah, tujuan, ruang lingkup, pendekatan, sistematika penyajian laporan, dan hipotesis penelitian. Bab ini memberikan gambaran menyeluruh mengenai konteks, urgensi, dan arah penelitian yang dilakukan.

Bab II Studi Literatur menguraikan landasan teori dan penelitian terdahulu yang relevan, termasuk pembahasan mengenai media sosial, platform X dan YouTube, konsep dan pendekatan analisis sentimen, teknik *deep learning, pre-processing* data, serta model IndoBERT yang menjadi fokus utama penelitian. Bab ini juga mencakup pembahasan metode *supervised , unsupervised , lexicon based,* dan *framework Hugging Face* yang digunakan dalam penelitian.

Bab III Metodologi menjelaskan tahapan penelitian secara detail, mulai dari analisis dan justifikasi setiap kebijakan yang diteliti, metode pengumpulan data melalui *crawling* dan *scrapping*, tahapan *pre-processing* data seperti *cleaning,* normalisasi, tokenisasi, hingga proses *labeling,* pembagian data, penerapan SMOTE, pembangunan model IndoBERT multitask, tahapan pelatihan model, evaluasi, hingga implementasi.

Bab IV Implementasi memaparkan bagaimana metode dan tahapan dalam penelitian diterapkan secara teknis, mulai dari proses pengumpulan dan pengolahan data, penerapan *labeling,* pembagian data, *modeling* dengan IndoBERT, hingga tahap evaluasi dan implementasi akhir model.

Bab V Hasil berisi pemaparan hasil analisis sentimen pada masing-masing kebijakan, *confusion matrix,* evaluasi model, ringkasan hasil, serta pelaporan implementasi penelitian.

Bab VI Kesimpulan dan Saran merangkum keseluruhan hasil penelitian, menjawab pertanyaan penelitian, serta memberikan saran atau rekomendasi untuk penelitian di masa mendatang.

1.7 Hipotesis

Hipotesis dalam penelitian ini dibangun berdasarkan persepsi publik terhadap empat kebijakan utama Presiden Prabowo Subianto, yaitu Program Makan Bergizi Gratis, Efisiensi Anggaran, Danantara, dan Kenaikan PPN 12%. Pada kebijakan Program Makan Bergizi Gratis, diasumsikan bahwa masyarakat memberikan tanggapan positif, terutama dari kalangan siswa dan orang tua, karena program ini dinilai mampu meningkatkan kualitas gizi dan mendukung proses belajar anak-anak di sekolah. Namun demikian, tidak sedikit pula masyarakat yang menunjukkan asumsi negatif dengan alasan bahwa program ini akan membebani Anggaran Pendapatan dan Belanja Negara (APBN) serta berisiko sulit direalisasikan secara merata

Pada kebijakan Efisiensi Anggaran, hipotesis yang diajukan adalah bahwa publik cenderung mengapresiasi langkah ini karena dianggap sebagai bentuk pengelolaan keuangan negara yang lebih sehat dan efisien. Namun, terdapat pula kekhawatiran bahwa efisiensi tersebut dapat berdampak pada penurunan kualitas pelayanan publik apabila tidak dijalankan dengan strategi yang matang.

Selanjutnya, pada kebijakan pembentukan Danantara, diasumsikan bahwa masyarakat menaruh harapan terhadap inisiatif ini sebagai langkah strategis untuk mendorong pertumbuhan ekonomi nasional melalui pengelolaan dana investasi. Akan tetapi, muncul pula kekhawatiran publik terhadap potensi penyalahgunaan wewenang dan risiko korupsi, jika Danantara tidak dikelola secara transparan dan akuntabel.

Terakhir, terkait kebijakan kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%, diasumsikan bahwa sebagai masyarakat memahami tujuan kebijakan ini sebagai upaya peningkatan penerimaan negara untuk menopang pembangunan. Meski demikian, sebagian besar masyarakat, khususnya dari kalangan menengah ke bawah, menunjukkan resistensi karena kebijakan ini dianggap memperberat beban ekonomi rumah tangga.

Hipotesis-hipotesis ini akan diuji melalui analisis sentimen menggunakan model *IndoBERT* terhadap komentar-komentar masyarakat yang diperoleh dari platform media sosial X dan YouTube, guna mengidentifikasikan kecenderungan opini publik terhadap kebijakan-kebijakan tersebut secara kuantitatif dan objektif.

BAB II
STUDI LITERATUR

Dalam era digital yang semakin berkembang, media sosial telah menjadi sarana utama bagi masyarakat untuk berinteraksi, berbagai informasi, dan mengemukakan pendapat, termasuk ke dalam ranah politik. Studi literatur ini membahas mengenai berbagai pendekatan dalam analisis sentimen guna memahami bagaimana teknologi dapat digunakan untuk mengukur persepsi publik masyarakat terhadap kebijakan dan sikap kepemimpinan presiden.

2.1 Media Sosial

Media sosial adalah platform atau media digital yang memungkinkan pengguna untuk membuat, berbagi, dan berinteraksi dengan konten serta berkomunikasi dengan pengguna lain secara online. Media sosial mencakup berbagai jenis situs dan aplikasi seperti *Facebook*, *Twitter*, *Instagram*, dan lainnya yang digunakan untuk berbagai tujuan,termasuk komunikasi, hiburan , pemasaran, dan penyebaran informasi [6].

Istilah "media sosial" mengacu pada berbagai layanan berbasis internet dan seluler yang memungkinkan pengguna untuk berpartisipasi dalam pertukaran daring, menyumbangkan konten yang dibuat pengguna, atau bergabung dengan komunitas daring. Media sosial adalah *platform* digital yang memungkinkan pengguna untuk membuat, berbagi, dan bertukar informasi, ide, dan konten dalam bentuk berbagi, dan bertukar informasi, ide, dan konten dalam berbagai bentuk, seperti teks, gambar, dan video. Media sosial memainkan peran penting dalam kehidupan politik dan sosial saat ini, memungkinkan partisipasi politik dan sosial yang lebih luas dan efektif. Platform seperti *Facebook, Twitter,* YouTube dan *Instagram* digunakan untuk berbagai tujuan, termasuk kampanye politik, komunikasi massa, dan membangun citra publik [1].

2.2 Platform Media X

Popularitas situs jejaring sosial telah meningkat pesat dalam beberapa tahun terakhir. Jumlah pengguna Twitter dilaporkan sekitar 40 juta pada tahun 2009 dan meningkat menjadi 330 juta pada tahun 2019. Pengguna Twitter biasanya mengekspresikan emosi dan pandangan pengguna melalui teks sepanjang 140 karakter (yang kini telah diperpanjang menjadi 280 karakter) atau membagikan gambar dan video, yang disebut sebagai "*tweet*" [7].

Namun, pada tahun 2023, Twitter telah berubah nama menjadi X, menandai perubahan besar dalam identitas dan arah pengembangannya. Sebagai platform media sosial *microblogging* yang populer, X tetap memiliki jutaan pengguna yang berbagai pendapat dan pemikiran pengguna melalui status secara *real-time.* Saat ini, platform ini memiliki lebih dari 100 juta pengguna aktif, dengan jumlah unggahan yang sebelumnya disebut "tweet" dapat mencapai hingga 500 juta per hari [8].

Karena jangkauan dan popularitas media sosial yang tinggi di seluruh dunia, organisasi juga menggunakan media sosial ini untuk merencanakan dan memobilisasi acara protes dan demokrasi publik. Selain itu, X memiliki peran penting dalam membentuk narasi dan pandangan publik terhadap berbagai isu yang muncul dengan menyebarkan informasi dan pandangan publik secara cepat dan luas. Platform ini juga berfungsi sebagai sarana untuk menanyang narasi sekuritisasi dan mempromosikan perspektif alternatif. Melalui penggunaan tag (#) yang terkait dengan suatu peristiwa atau kebijakan pemerintah, X dapat menjadi alat untuk memobilisasi dukungan atau protes [9].

2.3 Platform Media YouTube

Popularitas YouTube menjadikannya sebagai situs web kedua yang paling banyak dikunjungi di dunia, hanya dikalahkan oleh Google. Kemudahan penggunaan dan fungsionalitas platform ini memungkinkan berbagai jenis video untuk dibagikan oleh pengguna dari berbagai tingkat keahlian di seluruh dunia. Pengguna juga dapat berinteraksi melalui fitur suka (*like*), tidak suka (*dislike*), dan komentar pada video.

Diluncurkan pada tahun 2005, YouTube telah memainkan peran penting dalam meningkatkan komunikasi berbasis video. Media sosial telah menjadi sumber utama untuk pembaruan berita secara *real-time* di seluruh dunia, sementara YouTube tetap menjadi platform utama untuk berbagi video [10]. YouTube adalah salah satu platform media sosial yang terkenal dan populer di kalangan pengguna. Platform ini memungkinkan pengguna untuk mengunggah, berbagi, dan menonton video secara *online*. Selain itu, pengguna dapat berinteraksi melalui komentar pada video yang diunggah [11]. YouTube juga menyediakan berbagai mekanisme sosial seperti voting, rating, favorit, berbagi, dan komentar negatif untuk mengukur opini dan pandangan pengguna terhadap sebuah video [1]**.**

Sebagai platform video terbesar di dunia, YouTube menampilkan beragam konten, termasuk video musik, promosi produk, vlog, ulasan, dan video edukasi. Namun, komentar YouTube terkadang mengandung spam, yang dapat mengganggu pengalaman pengguna serta menimbulkan risiko keamanan, seperti *phising* dan penyebaran *malware*[11].

Kreator konten dapat membuat kanal (*channel*) YouTube sebagai tempat untuk mengorganisasi dan menampilkan video pengguna. Pengguna lain dapat memilih untuk berlangganan kanal tersebut agar mendapatkan pemberitahuan setiap kali video baru diunggah. Dengan kemampuannya untuk menjangkau audiens luas dan fitur interaktifnya, YouTube berperan sebagai sarana penting dalam menyebarkan informasi, membangun komunitas, dan mendukung berbagai bentuk komunikasi digital.

2.4 Analisis Sentimen

Analisis sentimen adalah teknik yang digunakan untuk mengidentifikasi dan memproses kecenderungan emosional, perasaan, dan perspektif seseorang dari data teks. Teknik ini memainkan peran penting dalam berbagai bidang, seperti analisis ulasan produk, pemantauan opini publik, dan prediksi perilaku konsumen. Dengan analisis sentimen, perusahaan dan organisasi dapat lebih memahami sentimen publik dan membuat keputusan yang lebih tepat. Analisis sentimen melibatkan studi kompuasional terhadap teks, gambar, dan ucapan [11].

Analisis sentimen berguna untuk memperoleh dengan cepat dalam memahami keseluruhan gagasan dengan menggunakan sejumlah besar data teks dan akan membantu dalam memahami pendapat pengguna. Analisis sentimen juga dikenal sebagai opini *mining*, yang berarti menemukan atau mengidentifikasi opini, pandangan, kesan, emosi, dan peran yang bersifat positif, negatif, atau netral dalam teks[12].

Analisis sentimen adalah proses mengetahui, memahami, memproses, dan mengekstraksi data secara otomatis untuk memperoleh informasi sentimen dalam kalimat dan teks. Ini dapat diartikan memiliki berbagai jenis, termasuk analisis sentimen berbasis aspek, deteksi emosi, dan analisis sentimen tingkat lanjut. Analisis sentimen berbasis aspek digunakan untuk menentukan aspek yang mempengaruhi dan mengevaluasi pelanggan. Deteksi emosi bertujuan untuk menentukan emosi dari sebuah pesan atau teks, seperti kesedihan, kemarahan, kebahagiaan, positif, dan negatif. Sedangkan analisis sentimen tingkat lanjut biasanya digunakan dalam bidang *e-commerce* untuk penilaian yang lebih spesifik[1].

Namun, analisis sentimen sering digunakan secara tidak tepat ketika seseorang mengacu pada polaritas klasifikasi, yang sebaliknya merupakan sub tugas yang ditujukan untuk mengekstraksi sentimen positif, negatif, atau netral (juga disebut polaritas) dari teks. Meskipun opini juga bisa memiliki polaritas netral (misalnya, "Saya tidak tahu apakah saya menyukai film ini atau tidak. Saya harus menontonnya dengan tenang."), sebagian besar pekerjaan dalam analisis sentimen biasanya hanya mengasumsikan sentimen positif dan negatif untuk kesederhanaan.Tergantung pada bidang aplikasi, beberapa nama digunakan untuk analisis sentimen (misalnya, penambangan opini, ekstraksi opini, sentimen penambangan, analisis subjektivitas, analisis pengaruh, analisis emosi, dan penambangan tinjauan)[15].

2.5 Pendekatan dan Model Analisis Sentimen

Dalam pendekatan analisis sentimen digunakan untuk memahami opini masyarakat terhadap kebijakan pemerintah Presiden Prabowo Subianto berdasarkan data

komentar di media sosia. Untuk menentukan model yang paling tepat, diperlukan kajian literatur terhadap berbagai pendekatan populer dalam klasifikasi sentimen. Kajian ini bertujuan untuk membandingkan kelebihan dan kekurangan masing-masing model, sehingga dapat memberikan landasan ilmiah atas pemilihan model *IndoBERT* sebagai metode utama dalam penelitian ini. Berikut adalah lima model popular yang sering digunakan dalam analisis sentimen menurut studi literatur terkini, khususnya dalam konteks Bahasa Indonesia.

2.5.1 Naive Bayes

Algoritma *naive bayes* adalah metode klasifikasi data yang menggunakan teorema *Bayes* untuk menghitung probabilitas suatu kelas data berdasarkan probabilitas fitur yang terkait dengan kelas tersebut. Metode ini dianggap "*naive*" karena adanya independensi antar fitur, meskipun dalam situasi nyata, fitur-fitur dapat saling berhubungan. *Naive Bayes* digunakan untuk klasifikasi teks, analisis sentimen, dan tugas lain yang melibatkan kategori data [13].

Naive Bayes merupakan metode klasifikasi berbasis probabilistik yang menggunakan Teorema *Bayes* dengan asumsi independensi antar fitur. Meskipun asumsi ini tidak selalu realistis, algoritma ini terbukti cukup efektif dalam berbagai tugas klasifikasi teks, termasuk analisis sentimen. Keunggulan utamanya adalah kesederhanaan, kecepatan pelatihan, dan kinerja yang cukup baik pada data berukuran kecil. Dalam studi Setiawan dan Yuniarti(2023), algoritma ini diterapkan untuk menganalisis sentimen terhadap seleksi CPNS 2023 di Twitter dan memberikan hasil yang cukup baik sebagai baseline model [13].

2.5.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan pada tahun 1992 oleh *Boser, Guyon, dan Vapink* dalam konferensi COLT-92. SVM merupakan metode pembelajaran yang digunakan untuk tugas klasifikasi dan regresi, dan termasuk dalam kelompok *classifier linier* umum. Secara sederhana, *SVM* adalah alat prediksi untuk klasifikasi dan regresi yang memanfaatkan teori *machine learning* guna meningkatkan akurasi prediksi secara optimal sekaligus meminimalkan risiko *overfitting* terhadap data.

Dalam konteks analisis sentimen, SVM telah terbukti sebagai salah satu pendekatan yang paling handal dan banyak digunakan. SVM merupakan salah satu teknik *supervised learning* yang paling populer dalam analisis sentimen berbasis teks, karena kemampuannya menangani data berdimensi tinggi, serta ketepatannya dalam membedakan polaritas sentimen secara konsisten. Studi tersebut menekankan bahwa SVM sangat efektif ketika dikombinasikan dengan teknik representasi teks seperti *TF-IDF*, *n-gram,* atau *word embeddings* [14].

SVM dikenal kuat terhadap dimensi tinggi dan bekerja efektif pada dataset yang memiliki margin pemisahan yang jelas. Dalam konteks analisis sentimen, SVM kerap digunakan karena kemampuannya menghasilkan akurasi tinggi pada data teks yang terstruktur, Namun, SVM dapat mengalami kesulitan ketika digunakan pada data sangat besar atau tidak seimbang [14].

2.5.3 Convolutional Neural Network (CNN)

*Convolutional Neural Network (*CNN*)* adalah jenis arsitektur jaringan saraf dalam yang awalnya dirancang untuk pengolahan data citra, namun telah banyak diadaptasi untuk berbagai tugas pemrosesan bahasa alami (*Natural Language Processing*/NLP), termasuk klasifikasi sentimen. CNN bekerja dengan menerapkan filter *konvolusional* untuk mengekstraksi fitur lokal dari input, yang dalam konteks teks, dapat berupa pola kata atau frasa penting

Dalam klasifikasi teks, CNN memiliki kemampuan untuk mengenali pola-pola penting seperti kombinasi kata yang sering muncul pada kelas tertentu. CNN unggul dalam mendeteksi fitur spasial lokal dan dapat diimplementasikan secara efisien karena tidak melibatkan rekursi atau pemrosesan sekuensial seperti RNN, Struktur arsitekturnya yang bersifat paralel juga membuat proses pelatihan relatif cepat dan efisiensi.Namun, keterbatasan utama CNN terletak pada kemampuannya yang kurang dalam memahami konteks jangka panjang. CNN cenderung fokus pada jendela lokal dan tidak memiliki memori untuk mengingat informasi kata sebelumnya, sehingga konteks keseluruhan kalimat dapat terabaikan.

Studi oleh Kim (2020) menunjukkan bahwa CNN dapat digunakan secara efektif untuk tugas klasifikasi kalimat, termasuk analisis sentimen, dengan performa yang kompetitif terhadap model berbasis RNN, CNN juga lebih ringan secara komputasional dibandingkan LSTM atau BERT, menjadikannya pilihan yang baik untuk sistem dengan sumber daya terbatas.

2.5.4 Long Short-Term Memory (LSTM)

*Long Short-Term Memory (*LSTM*)* merupakan salah satu arsitektur *Recurrent Neural Network* (RNN) yang dirancang khusus untuk menangani masalah *long-term dependencies,* yaitu saat model perlu mengingat informasi dari urutan data sebelumnya untuk membuat prediksi yang akurat. Dalam konteks analisis sentimen, LSTM terbukti efektif karena kemampuannya dalam memahami konteks kalimat secara lebih dalam dibandingkan dengan metode tradisional [15].

LSTM dikembangkan oleh *Hochreiter* dan *Schmidhuber* pada tahun 1997 untuk mengatasi masalah *vanishing gradient* yang biasa terjadi pada RNN standar ketika menghadapi urutan data panjang. Struktur LSTM terdiri dari *cell state* (memori internal) dan tiga gerbang *input*, *forget*, dan *output* yang bersama-sama mengendalikan aliran informasi masuk, penyimpanan, dan keluaran. Mekanisme ini memungkinkan LSTM untuk menjaga informasi penting tetap relevan dari waktu ke waktu dan membuang yang tidak berguna. Seiring waktu, arsitektur LSTM terus berevolusi, seperti varian DPG-LSTM yang memadukan struktur dependensi sintaksis dan *Graph Convolutional Network* (GCN) dengan mekanisme perhatian *hybrid* untuk menyoroti kata kunci sintaksis dengan hasil peningkatan akurasi dalam tugas analisis sentimen.

2.5.5 Pembelajaran Transfer (BERT)

Bidirectional Encoder Representation from Transformers (BERT), adalah model pra-latihan berbasis arsitektur *transformer* yang dikembangkan oleh Google dan diperkenalkan pada tahun 2018. BERT dilatih dengan pendekatan *masked language model* dan *next sentence prediction,* yang memungkinkan model memahami konteks kata secara dua arah, yaitu dari kiri ke kanan dan sebaliknya. Keunggulan utama BERT terletak pada kemampuan untuk menangkap makna kontekstual dari

kata dalam suatu kalimat secara lebih akurat dibandingkan dengan model klasik berbasis *word embeddings* statis.

BERT telah merevolusi banyak tugas dalam pemrosesan bahasa alami (*Natural Language Processing /* NLP), seperti klasifikasi teks, ekstraksi entitas, dan analisis sentimen. Salah satu kekuatan utamanya adalah kemampuan untuk diadaptasikan ke berbagai tugas downstream melalui proses fine-tuning dengan sejumlah kecil data berlabel, menjadikan sangat efisien dan fleksibel untuk berbagai aplikasi bahasa.

Dalam konteks Bahasa Indonesia, model BERT telah diadaptasi dan dikembangkan menjadi *IndoBERT* oleh *IndoNLP*. *IndoBERT* merupakan hasil pelatihan dari model BERT menggunakan korpus teks dalam Bahasa Indonesia, sehingga lebih relevan dan akurat untuk tugas-tugas NLP dalam bahasa lokal. Ketersediaan model ini secara terbuka di platform *HuggingFace* memudahkan peneliti dan praktisi dalam mengimplementasikan dan menyesuaikan *IndoBERT* sesuai dengan kebutuhan aplikasi, termasuk dalam analisis sentimen yang memerlukan pemahaman konteks kalimat yang kompleks [16].

2.5.6 Alasan Pemilihan IndoBERT

Model *IndoBERT* dipilih dalam penelitian ini karena merupakan model pelatihan berbasis *transformer* yang telah dioptimalkan untuk Bahasa Indonesia. Dibandingkan dengan pendekatan tradisional seperti *Naive Bayes* atau SVM, *IndoBERT* mampu memahami konteks kata secara lebih akurat dan menangkap nuansa dalam opini masyarakat. Selain itu pendekatan multitasking yang diterapkan dalam penelitian ini (klasifikasi sentimen dan topik simultan) sangat sesuai dengan arsitektur *transformer* seperti *IndoBERT*. Model ini juga memungkinkan integrasi yang efisien antara representasi kontekstual dan pelabelan ganda, pada akhirnya meningkatkan akurasi dan efektivitas klasifikasi.

Tabel 2.1 Perbandingan Model Sentimen Bahasa Indonesia(2024)

Model	Keunggulan	Kelemahan	Cocok Untuk	Sumber
Naive Bayes	Cepat, sederhana	Asumsi independensi fitur tidak realistis	Dataset Kecil	Suardi (2025)

Model	Keunggulan	Kelemahan	Cocok Untuk	Sumber
SVM	Akurat untuk margin jelas	Kurang fleksibel untuk data besar	Data teks terstruktur	Suardi (2025)
CNN	Menangkap pola lokal	Tidak menangkap konteks jangka panjang	Kalimat pendek	Mandhasiya ei at (2024)
LSTM	Menangkap konteks jangka panjang	Pelatihan lambat ,sensitif input	Kalimat kompleks	Mandhasiya ei at (2024)
BERT/ IndoBERT	Kontekstual, performa tinggi	Membutuhkan sumber daya besar	Bahasa alami tingkat lanjut	Mandhasiya ei at (2024)

2.6 Deep Learning

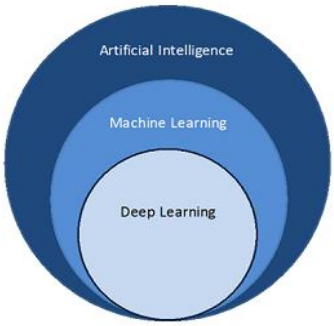
Deep learning adalah jenis khusus dari *Machine Learning* yang memiliki kekuatan fleksibilitas tinggi dengan cara mempelajari representasi dunia sebagai hierarki bertingkat dari konsep dan representasi. Setiap konsep didefinisikan berdasarkan konsep yang lebih sederhana, sementara representasi yang lebih abstrak dihitung dari representasi yang kurang abstrak. Karena *Machine Learning* sendiri merupakan subkategori dari *Artificial Intelligence* (AI), maka *Deep Learning* (DL) juga termasuk dalam cakupan AI. Bidang pembelajaran mendalam terutama berkaitan dengan cara membangun sistem komputer yang mampu berhasil menyelesaikan tugas-tugas yang membutuhkan kecerdasan, sedangkan bidang ilmu saraf komputasi terutama berkaitan dengan membangun lebih akurat model bagaimana otak benar-benar bekerja [17].

Sebagai bagian dari *Machine Learning, Deep Learning* menggunakan jaringan syaraf tiruan dengan banyak lapisan untuk memodelkan dan memahami data yang kompleks. Arsitektur jaringan syaraf tiruan ini umumnya terdiri dari tiga jenis lapisan utama: lapisan masukan (input layer), satu atau lebih lapisan tersembunyi (hidden layers), dan lapisan keluaran (output layer). Lapisan tersembunyi inilah yang menjadi inti dari '*deep learning*', di mana proses komputasi utama terjadi. Setiap lapisan tersembunyi menerima informasi dari lapisan sebelumnya, melakukan transformasi matematis yang kompleks untuk mengekstraksi fitur-fitur penting dari data, dan meneruskannya ke lapisan berikutnya. Semakin banyak jumlah hidden layer, semakin 'dalam' (*deep*) arsitektur jaringan tersebut dan semakin kompleks pula pola yang dapat dipelajarinya.

Metode ini telah menjadi pendekatan dominan dalam penelitian pemrosesan bahasa alami NLP (*Natural Language Processing*), terutama saat bekerja dengan korupsi berskala besar. Salah satu aspek utama dari *Deep Learning* adalah bahwa lapisan-lapisan fitur tidak dirancang oleh insinyur manusia, melainkan dipelajari secara otomatis dari data melalui prosedur pembelajaran yang bersifat umum. Dengan kata lain, *Deep Learning* mengandalkan arsitektur jaringan saraf yang lebih kompleks dibandingkan metode lain dalam *Machine Learning*, yang menjadikannya unggul dalam berbagai tugas berbasis data besar [18].

Karena berada dalam *Artificial Intelligence*, *Deep Learning* telah mencapai kemajuan besar dalam menyelesaikan berbagai permasalahan yang selama bertahun-tahun sulit dipecahkan oleh komunitas kecerdasan buatan. Teknik ini terbukti sangat efektif dalam pengenalan gambar dan suara, yang merupakan bagian dari pengembangan kecerdasan buatan secara luas. Selain itu, *Deep Learning* sebagai cabang dari *Machine Learning* telah menghasilkan hasil yang sangat menjanjikan dalam berbagai tugas pemahaman bahasa alami, seperti klasifikasi topik, analisis sentimen, pencarian jawaban, dan penerjemahan bahasa [19]. Dengan kemampuannya dalam mengenali pola secara otomatis, *Deep Learning* memainkan peran penting dalam evolusi teknologi AI.

Dapat dilihat pada gambar di bawah ini bahwa *Deep Learning* berada dalam lingkup *Machine Learning*, dan *Machine Learning* berada dalam cakupan *Artificial Intelligence*, menunjukkan bagaimana ketiga teknologi ini saling berkaitan dalam pengembangan sistem kecerdasan buatan.



Gambar 2.1 Hubungan Antara DL, ML, dan AI
Sumber: AI-ML-DL *Difference*

2.7 Data Pre-Processing

Data *Preprocessing* adalah langkah awal dalam klasifikasi teks yang melibatkan pembersihan teks dengan menghilangkan data yang tidak relevan dan bagian yang tidak informatif dari teks asli, seperti *hastag*. Proses ini meningkatkan efisiensi proses klasifikasi. *Preprocessing* juga mencakup langkah-langkah seperti *tokenisasi*, normalisasi (termasuk *URL*, email, tanggal, waktu, dan nomor telepon), serta anotasi huruf kapital dan pengurangan kata yang diperpanjang atau diulang[20].

2.7.1 Data Cleaning

Data *cleaning* adalah langkah pertama dalam *Preprocessing* data yang bertujuan untuk menghilangkan informasi yang tidak relevan dari dataset guna meningkatkan kualitas dan efisiensi analisis selanjutnya. Proses ini biasanya mencakup penghapusan karakter yang tidak diinginkan, serta URL, karakter khusus, dan angka, yang tidak memberikan informasi bermakna untuk tugas seperti analisis sentimen.

Tahapan Data *Cleaning* juga mencegah duplikasi akibat perbedaan kapitalisasi, begitu juga untuk kata-kata umum yang tidak memiliki bobot signifikan dalam analisis, juga dihapus agar fokus tetap pada kata-kata yang lebih berdampak. Koreksi singkatan dalam teks juga merupakan bagian dari proses ini, di mana kata-kata yang disingkat diperluas ke bentuk lengkapnya untuk memastikan konsistensi dan kejelasan data [21].

2.7.2 Normalisasi

Normalisasi teks adalah proses mengubah tulisan tidak formal menjadi bentuk standarnya dalam suatu bahasa. Tujuan normalisasi teks adalah untuk mengurangi variasi leksikon dan ortografis dalam teks, sehingga lebih mudah untuk diproses, dianalisis, dan dipahami. Ini merupakan langkah penting dalam *Preprocessing* dalam *Natural Language Processing* (NLP)[22]. Dengan semakin banyaknya data teks tidak terstruktur yang tersedia dalam berbagai bentuk, seperti media sosial, situs web, dan dokumen, semakin penting untuk melakukan pre pemrosesan dan normalisasi teks guna meningkatkan kualitas dan kegunaannya.

Penanganan bahasa informal, termasuk singkatan, typo, dan bahasa gaul, merupakan aspek krusial dalam normalisasi teks, terutama saat menganalisis data dari media sosial. Singkatan (misalnya, “yg” menjadi “ yang” atau “ dgn” menjadi “dengan”) dan *typo* (kealahan ketik) sangat umum di platform daring, dan jika tidak dinormalisasikan, dapat mengganggu analisis sentimen karena dianggap sebagai kata yang berbeda oleh model.

Selain itu, masalah polisemi (kata dengan banyak makna) juga perlu diperhatikan dalam normalisasi teks. Misalnya, kata “ bisa” dapat berarti “ mampu” atau “ racun ular”. Dalam konteks bahasa informal, *polisemi* bisa lebih kompleks dengan penggunaan kata-kata gaul yang memiliki makna ganda atau tergantung pada konteks kalimatnya. Pendekatan berbasis aturan seringkali makan ganda atau tergantung pada konteks kalimatnya. Pendekatan berbasis aturan seringkali digunakan untuk menangani singkatan dan *typo* dengan memetakan kata-kata *non-standar* ke bentuk standarnya menggunakan kamus atau *lexicon.* Untuk polisemi, penentuan makna yang benar seringkali membutuhkan analisis konteks yang lebih luas dalam kalimat atau bahan paragraf. Namun, dalam konteks analisis sentimen, fokusnya lebih pada polaritas emosional kata, sehingga penanganan polisemi seringkali disederhanakan dengan memilih makna yang paling umum atau relevan dengan konteks sentimen.

2.7.3 Tokenization

Pada proses *tokenization* adalah proses memecah kalimat atau teks menjadi bagian-bagian yang lebih kecil, biasanya berupa kata-kata. proses ini bertujuan untuk menjaga konsistensi dan membantu dalam mengidentifikasi makna dari kata kunci. Langkah ini penting dalam analisis teks karena memungkinkan algoritma untuk memproses dan memahami setiap elemen data dengan lebih efisien [22]. Dalam praktiknya, *tokenization* dapat dilakukan dengan menggunakan karakter spasi untuk memisahkan setiap kata dalam sebuah kalimat atau paragraf.

2.7.4 Padding & Truncation

Dalam analisis sentimen menggunakan BERT, *padding* dan *truncation* digunakan untuk menyesuaikan panjang teks sebelum diproses oleh model. *Padding*

memastikan panjang *input* seragam dengan menambahkan nol ke sekuens yang lebih pendek. Sementara itu, *truncation* memangkas teks yang melebihi batas maksimum token yang dapat diterima BERT, yang berisiko menghilangkan informasi penting dalam analisis sentimen. Pemilihan strategi *padding* dan *truncation* sangat penting untuk memastikan model tetap akurat dan efisien dalam mengklasifikasikan sentimen [22].

2.7.5 Encoding to Tensor

Encoding berbasis *tensor* telah terbukti meningkatkan efisiensi perhitungan dan akurasi model dalam proses retrieval gambar berbasis fitur. Konsep ini dapat diterapkan dalam analisis sentimen menggunakan model BERT, di mana setiap kata dalam teks dikonversi menjadi representasi tensor yang menangkap hubungan semantik antar kata [23]. Metode tensor factorization yang memungkinkan reduksi dimensi tensor fitur yang dihasilkan dari *deep learning*, yang dapat membantu optimalisasi model berbasis transformer seperti BERT. Selain itu, metode tensor *decomposition* dapat diadaptasi untuk meminimalkan kompleksitas komputasi dan meningkatkan efisiensi dalam model NLP seperti BERT, yang bergantung pada representasi tensor dalam memahami hubungan semantik antar kata.

2.8 Unsupervised

Unsupervised learning mengacu pada teknik yang menggunakan data tanpa label untuk pelatihan tanpa bergantung pada jawaban yang sudah diketahui. Ini berbeda dengan *supervised learning* yang membutuhkan sejumlah besar data berlabel. Dalam pendekatan *unsupervised*, model dapat melakukan *pre-training* pada *embedding* kata atau model bahasa menggunakan teks tanpa label, yang kemudian dapat di *fine-tune* untuk tugas-tugas spesifik [16].

Selain pendekatan berbasis *deep learning* modern seperti pre-training BERT, dalam analisis sentimen juga dikenal metode unsupervised tradisional yang lebih sederhana. Dalam analisis sentimen, *unsupervised learning* mengacu pada teknik pembelajaran mesin yang tidak memerlukan data berlabel. Algoritma *unsupervised* mencoba mengidentifikasi pola, struktur, atau pengelompokan alami dalam data

berdasarkan kemiripan fitur [24]. Menurut Feldman (2013), analisis sentimen *unsupervised* banyak digunakan ketika anotasi manual terhadap data terlalu mahal atau tidak tersedia. Dalam pendekatan ini, metode seperti *lexicon-based* dan *clustering* sering diterapkan untuk polaritas sentimen dari teks. Selain itu, Feldman juga menjelaskan bahwa salah satu metode *unsupervised* yang umum adalah dengan menggunakan kamus sentimen (*sentimen lexicon*) yang berisi daftar kata bermuatan positif dan negatif. Skor suatu teks kemudian dihitung berdasarkan frekuensi kata.

2.9 Supervised

Supervised learning merupakan pendekatan pembelajaran mesin di mana algoritma dilatih menggunakan data berlabel, sehingga model dapat mempelajari pemetaan dari *input* ke *output* yang benar[13]. Menurut Zhang, Wang, dan Liu (2018), dalam *supervised learning,* model memerlukan dataset besar yang telah diberi label, seperti komentar yang telah diklasifikasikan ke dalam kategori (positif, negatif, dan netral). Meskipun pendekatan ini menunjukkan peningkatan yang mengesankan dibandingkan dengan *unsupervised, supervised* memiliki kelemahan utama, yakni ketergantungan pada jumlah data berlabel yang besar. Proses pelabelan data secara manual dapat menjadi sangat mahal dan memakan waktu. Meski demikian, *supervised learning* tetap menjadi metode yang domain karena pendekatan ini memungkinkan model untuk mempelajari pemetaan langsung dari *input* ke *output,* yang sering kali menghasilkan kinerja tinggi dalam berbagai tugas, seperti klasifikasi teks, penerjemahan bahasa, dan menjawab pertanyaan.

2.10 Pengumpulan Data

Pengumpulan data adalah proses penting dalam penelitian yang melibatkan pengambilan data dari berbagai sumber untuk dianalisis lebih lanjut. Dalam Konteks media sosial, pengumpulan data sering dilakukan melalui *Application Programming Interface* (API), yang memungkinkan penelitian untuk mengakses dan mengunduh data dari platform seperti *Twitter(X)* atau YouTube. Misalnya, dalam penelitian yang menggunakan data dari *Twitter,* pengumpulan data dapat melibatkan pengguna API untuk mengumpulkan tweet yang relevan dengan topik tertentu, seperti *metaverse*, dengan batasan tertentu pada jumlah *tweet* yang dapat diambil dalam satu kali proses *crawling* [25].

Untuk mendapatkan data yang dibutuhkan untuk analisis suasana hati ini, kami menggunakan teknik data creeper untuk mengumpulkan opini publik dari platform X dan YouTube. Data berkomentar dalam bentuk pengguna platform X dan YouTube yang terkait dengan kepemimpinan Prabowo Subianto. Merangkak teknologi data dari platform X adalah kumpulan komentar pengguna X dengan memasukkan kata kunci untuk mendapatkan komentar tentang kepemimpinan Prabowo. Sementara itu, merangkak teknologi data dari platform YouTube menggunakan platform Apify dengan memasukkan api dari konten YouTube dan mengambil komentar pengguna ke dalam konten YouTube. Dari teknologi ini, pengumpulan data secara otomatis memungkinkan jumlah besar dan karenanya memberikan citra persepsi publik yang lebih luas, obyektif. Metode merangkak memungkinkan Anda untuk memproses dan menganalisis data yang diperoleh agar lebih akurat dan menciptakan kedalaman suasana hati umum terhadap kepemimpinan Prabowo. Data yang dikumpulkan dapat disimpan dalam format yang mudah diolah, seperti *Comma Separated Values (CSV),* untuk analisis lebih lanjut [25].

2.11 Ekonomi dan Kebijakan Terkait

Perkembangan ekonomi suatu negara sangat dipengaruhi oleh kebijakan fiskal dan regulasi yang diterapkan oleh pemerintah. Dalam beberapa tahun terakhir, pemerintah Indonesia telah mengambil langkah-langkah strategis untuk meningkatkan efisiensi anggaran, memperkuat penerimaan negara, serta mendorong pertumbuhan ekonomi dan kesejahteraan masyarakat. Beberapa kebijakan yang diterapkan antara lain:

1. Efisiensi anggaran melalui pemotongan belanja negara sebagaimana tertuang dalam Instruksi Presiden Nomor 1 Tahun 2025, yang bertujuan untuk menghemat pengeluaran negara hingga RP 306,69 triliun.
2. Kenaikan tarif Pajak Pertambahan Nilai (PPN) menjadi 12% sebagaimana diatur dalam Undang-Undang Nomor 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan.
3. Program makan bergizi gratis yang mulai diterapkan sejak Agustus 2024, yang memiliki tujuan dalam meningkatkan kualitas gizi masyarakat, khususnya anak-anak sekolah.

4. Pembentukan Danantara (Daya Anagata Nusantara *Investment Management Agency*), lembaga investasi negara kedua setelah INA, yang diresmikan pada 24 Februari 2025, dengan tujuan mengoptimalkan investasi pemerintah untuk mendukung pertumbuhan ekonomi nasional.

Berikut ini adalah penjelasan lebih lanjut mengenai masing-masing kebijakan dan dampaknya terhadap perekonomian Indonesia.

2.11.1 Efisiensi Anggaran

Anggaran adalah pernyataan mengenai estimasi kinerja yang hendak dicapai selama periode waktu tertentu yang dinyatakan dalam ukuran finansial [26]. Secara keseluruhan, menurut Mahsun (2006) efisiensi anggaran merupakan pengukuran efisiensi mengukur perbandingan tingkat *output* dengan tingkat *input* dari organisasi sektor publik. Tingkat efisiensi diukur dengan cara membandingkan realisasi anggaran belanja langsung dengan total realisasi anggaran belanja.

Pada tahun 2025, pemerintah Indonesia mengeluarkan Instruksi Presiden Nomor 1 Tahun 2025 yang bertujuan melakukan efisiensi anggaran dengan memotong pengeluaran sebesar Rp 306,69 triliun. Langkah ini mencakup pemotongan anggaran kementerian/lembaga sebesar Rp 256,1 triliun dan transfer ke pemerintah daerah sebesar Rp50,595 triliun. Implementasi kebijakan ini menimbulkan berbagai reaksi dan kritik dari berbagai pihak[27].

Sebagai tindak lanjut, beberapa kementerian menerbitkan surat edaran untuk mengatur strategi pengendalian belanja. Misalnya Kementerian Kesehatan mengeluarkan SE Nomor HK.02.02/A/548/2025 yang menetapkan strategi pengendalian belanja dalam rangka efisiensi anggaran tahun 2025 [27]. Langkah-langkah efisiensi anggaran ini menimbulkan tantangan bagi aparatur negara dalam menjalankan tugas dan fungsi. Namun, dengan adaptasi dan inovasi, diharapkan pelayanan publik tetap berjalan efektif meskipun dalam kondisi anggaran yang terbatas.

Selain itu, terdapat penelitian yang membahas efisiensi anggaran di Indonesia. Misalnya, penelitian oleh Oktaviani dan Mulyandani (2022) menganalisis efektivitas dan efisiensi anggaran belanja langsung pada perwakilan BKKBN Provinsi Jawa Barat tahun anggaran 2016-2020. Hasil penelitian menunjukkan bahwa tingkat efektivitas mencapai 83,58% (cukup efektif) dan efisiensi sebesar 77,28% (efisien).

Sebagai bentuk penguatan terhadap uraian kebijakan ini, penelitian ini turut menyajikan bukti-bukti berupa data digital dari berbagai sumber terpercaya. Data tersebut mencakup unggahan dari platform media sosial X dan YouTube, serta artikel dari media daring nasional dan Internasional yang merekam tanggapan publik terhadap kebijakan efisiensi anggaran. Bukti-bukti ini digunakan untuk mengilustrasikan bagaimana kebijakan tersebut dipersepsikan oleh masyarakat secara luas di ruang digital. Dengan pendekatan ini, diharapkan pemahaman terhadap dinamika opini publik menjadi lebih komprehensif dan kontekstual, serta mendukung analisis sentimen yang dilakukan dalam penelitian ini.

2.11.1.1 Data Media Sosial X Program Efisiensi Anggaran

Pembahasan mengenai efisiensi anggaran di platform X menunjukkan adanya perhatian publik yang cukup tinggi. Banyak pengguna memberikan opini mereka terkait kebijakan pemangkasan anggaran, terutama yang berdampak pada sektor publik. Tagar seperti ***#efisiensianggaran*** menjadi perbincangan, dengan respons yang beragam, mulai dari dukungan terhadap pengelolaan anggaran yang lebih disiplin hingga kritik terhadap risiko pemotongan anggaran layanan sosial.

Salah satu representasi diskursus publik mengenai kebijakan efisiensi anggaran juga dapat ditemukan melalui unggahan media sosial, seperti yang ditunjukkan dalam tangkapan layar *tweet* dibawah ini. Dalam unggahan tersebut, ditampilkan tangkapan layar dari konten visual milik CNBC Indonesia yang memuat pernyataan Presiden Prabowo Subianto tentang target penghematan anggaran hingga mencapai Rp 750 triliun dalam tiga tahap kebijakan.

Unggahan ini tidak hanya menyampaikan informasi faktual mengenai rencana efisiensi anggaran yang masif, tetapi juga menyertakan opini dan narasi interpretatif dari pemilik akun, yang menyebutkan bahwa Rp 100 triliun akan dialokasikan untuk program makan siang gratis, sementara sisanya akan digunakan sebagai modal awal bagi lembaga investasi negara *Danantara*. Konten ini menggambarkan dinamika penyebaran informasi kebijakan melalui platform media sosial, yang sering kali mempertemukan sumber informasi institusional (dalam hal ini CNBC Indonesia) dengan interpretasi dan opini publik yang lebih bebas. Interaksi seperti jumlah *retweet*, *like*, dan komentar menunjukkan tingginya atensi publik terhadap isu efisiensi anggaran, serta memberikan indikator awal bagi analisis persepsi publik di ruang digital. Analisis terhadap konten media sosial semacam ini menjadi penting dalam memahami bagaimana kebijakan besar pemerintah diterima, dibicarakan, dan diperdebatkan oleh publik luas di era digital. Hal ini juga memperkuat posisi media sosial sebagai kanal distribusi informasi sekaligus arena konstruksi opini publik terhadap kebijakan fiskal negara.

Gambar 2.2 Unggahan *tweet* @Strategi_Bisnis

2.11.1.2 Data Media Sosial YouTube Program Efisiensi Anggaran

Di platform YouTube, sejumlah video dari media nasional dan kanal independen membahas program efisiensi anggaran secara mendalam. Topik ini dibahas dalam berbagai bentuk, seperti liputan berita, analisis anggaran negara, dan diskusi

kebijakan. Jumlah penayangan dan komentar yang tinggi menunjukkan bahwa publik cukup tertarik untuk memahami lebih lanjut dampak serta urgensi dari kebijakan ini dalam konteks pembangunan nasional.

Gambar dibawah ini merupakan tangkapan layar dari *channel* YouTube KOMPAS TV. Dalam video tersebut, dibahas secara mendalam pernyataan Wakil Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi (Wamendikbudristek) mengenai kebijakan efisiensi anggaran di sektor pendidikan, dan mencerminkan kekhawatiran publik atas dampak langsung dari kebijakan pemotongan anggaran terhadap biaya pendidikan tinggi di Indonesia.

Melalui forum ini, narasumber dari pihak pemerintah memberikan klarifikasi dan penjelasan mengenai konteks kebijakan tersebut serta langkah-langkah mitigasi yang direncanakan. Tayangan ini menunjukkan bagaimana media mainstream memanfaatkan platform YouTube untuk menyampaikan informasi kebijakan secara terbuka, dan sekaligus memberikan ruang bagi masyarakat untuk mengakses, merespons, dan berdialog terhadap isu-isu penting seperti anggaran pendidikan. Jumlah penayangan, komentar warganet, dan interaksi lainnya turut menjadi indikator keterlibatan publik yang signifikan dalam menanggapi isu efisiensi anggaran di sektor pendidikan.

Gambar 2.3 Kanal KOMPAS TV - Efisiensi Anggaran Pendidikan

Gambar dibawah ini merupakan tangkapan layar dari *channel* YouTube resmi CNN Indonesia dalam sebuah program dialog bertema kebijakan efisiensi anggaran negara. Video ini mengusung tajuk “Dukung Efisiensi Anggaran, WFA ASN

Dikedepankan?" dan menampilkan narasumber dari Institut Otonomi Daerah yang membahas potensi penerapan kebijakan *Work From Anywhere* (WFA) bagi Aparatur Sipil Negara (ASN) sebagai strategi penghematan biaya operasional pemerintah.

Diskusi ini memberikan gambaran bahwa efisiensi anggaran tidak hanya diwujudkan melalui pemotongan anggaran langsung, tetapi juga melalui pendekatan inovatif terhadap sistem kerja birokrasi. Kebijakan WFA dianggap dapat menurunkan beban anggaran negara seperti biaya transportasi dinas, konsumsi, dan pemeliharaan fasilitas kantor. CNN Indonesia, melalui platform YouTube, menghadirkan ruang edukatif bagi publik untuk memahami kebijakan tersebut secara lebih luas.

Tayangan ini memperlihatkan bagaimana media digital digunakan untuk memperluas cakupan informasi mengenai kebijakan pemerintah, serta mendorong keterlibatan masyarakat dalam menilai efektivitas dan dampak dari strategi efisiensi yang dijalankan.

Gambar 2.4 Kanal CNN Indonesia - ASN Dukung Efisiensi Anggaran

Gambar dibawah ini merupakan tangkapan layar dari *channel* YouTube resmi KOMPAS.com. Dalam episode ini, Kompas.com menghadirkan narasumber Meidia Waluyadi Askar dalam sebuah diskusi santai namun substansial mengenai isu kenaikan tarif Pajak Pertambahan Nilai (PPN) menjadi 12 persen dan kaitannya dengan efisiensi anggaran negara. Tayangan ini diberi judul "Efisiensi Anggaran

Hasil Tukar Guling PPN 12 Persen?" yang sekaligus menggambarkan pokok bahasan utama terkait kebijakan fiskal terbaru pemerintah.

Pengemasan video dalam format talkshow non-formal dengan latar studio yang nyaman memperlihatkan pendekatan media digital dalam menjangkau audiens muda dan publik umum dengan gaya komunikasi yang lebih ringan namun tetap informatif. Melalui diskusi ini, berbagai argumen mengenai pro dan kontra kebijakan PPN 12 persen diuraikan, termasuk bagaimana dampaknya terhadap struktur penerimaan negara dan efektivitas alokasi anggaran dalam konteks efisiensi fiskal.

Kehadiran program ini di platform YouTube mencerminkan upaya media daring dalam memperluas akses informasi dan meningkatkan literasi kebijakan publik secara inklusif. Tanggapan audiens melalui komentar dan interaksi di kolom diskusi juga menambah dimensi partisipatif yang penting dalam menilai persepsi publik terhadap kebijakan efisiensi anggaran berbasis reformasi perpajakan.

Gambar 2.5 Kanal Kompas.com - Kebijakan Efisiensi Anggaran

2.11.1.3 Data Media Daring Program Efisiensi Anggaran

Selain melalui media sosial, media daring nasional dan lokal juga memainkan peran krusial dalam menyebarkan informasi terkait kebijakan efisiensi anggaran yang diusung pemerintah. Media daring tidak hanya memuat berita langsung dari

pernyataan resmi pemerintah, tetapi juga menyajikan analisis mendalam melalui tajuk opini, wawancara dengan pakar, serta laporan investigatif yang memberi sudut pandang beragam. Narasi yang dibangun oleh media dapat mempengaruhi persepsi publik terhadap urgensi, kelayakan, dan keadilan dari kebijakan tersebut. Oleh karena itu, pemetaan terhadap pemberitaan di media daring dapat memberikan gambaran utuh mengenai bagaimana kebijakan efisiensi anggaran diposisikan dalam ruang publik digital. Sub-bab ini akan memuat rangkuman dan analisis atas sejumlah artikel dan konten pemberitaan dari media daring yang membahas tema efisiensi anggaran, baik dari segi kebijakan pemangkasan anggaran, perubahan sistem kerja aparatur negara, hingga reformasi perpajakan.

Salah satu sumber media daring internasional yang turut mengangkat isu efisiensi anggaran di Indonesia adalah Reuters. Dalam artikel berjudul *"Indonesia plans spending cuts worth $18.84 bln, finance ministry says"* yang diterbitkan pada 23 Januari 2025, Reuters melaporkan pernyataan resmi dari Kementerian Keuangan Indonesia mengenai rencana pemotongan anggaran belanja negara sebesar Rp306,7 triliun atau setara dengan USD 18,84 miliar.

Artikel dibawah ini menyoroti arahan langsung dari Presiden Prabowo Subianto kepada pemerintahannya untuk melakukan penyesuaian belanja negara demi mendukung efisiensi fiskal. Langkah tersebut disebut mewakili sekitar 8% dari total anggaran yang disetujui pada tahun 2025, yang berjumlah Rp3.621,3 triliun. Dalam laporan tersebut, Reuters juga menekankan bahwa efisiensi pengeluaran merupakan salah satu komitmen utama yang kerap disampaikan Presiden dalam berbagai kesempatan.

Pemberitaan dari Reuters ini menunjukkan bahwa kebijakan efisiensi anggaran Indonesia tidak hanya menjadi perhatian media nasional, tetapi juga mendapat sorotan dari media internasional yang kredibel. Hal ini mengindikasikan bahwa kebijakan tersebut memiliki dampak strategis yang dipantau secara global, khususnya dalam konteks stabilitas fiskal, pengelolaan belanja publik, dan kredibilitas ekonomi nasional di mata investor internasional.

Analisis terhadap pemberitaan seperti ini penting dalam memahami bagaimana media luar negeri membingkai kebijakan fiskal Indonesia serta bagaimana narasi tersebut dapat mempengaruhi persepsi internasional terhadap tata kelola pemerintahan dan ekonomi nasional.

Reuters

World Business Markets Sustainability Legal Commentary Technology Investigations

Indonesia plans spending cuts worth $18.84 bln, finance ministry says

By Reuters

January 23, 2025 6:52 PM GMT+7 · Updated January 23, 2025

Aa

Aerial view of Sudirman Business District in Jakarta, Indonesia October 25, 2017. REUTERS/Beawiharta/File Photo Purchase Licensing Rights

JAKARTA, Jan 23 (Reuters) - Indonesia's President Prabowo Subianto has instructed his government to cut spending by 306.7 trillion rupiah ($18.8 billion) this year, a finance ministry spokesperson told Reuters on Thursday.

Prabowo has frequently highlighted the need for efficient spending under his administration. The cuts are equivalent to about 8% of the total approved spending this year of 3,621.3 trillion rupiah.

Gambar 2.6 Artikel dari *Reuters*

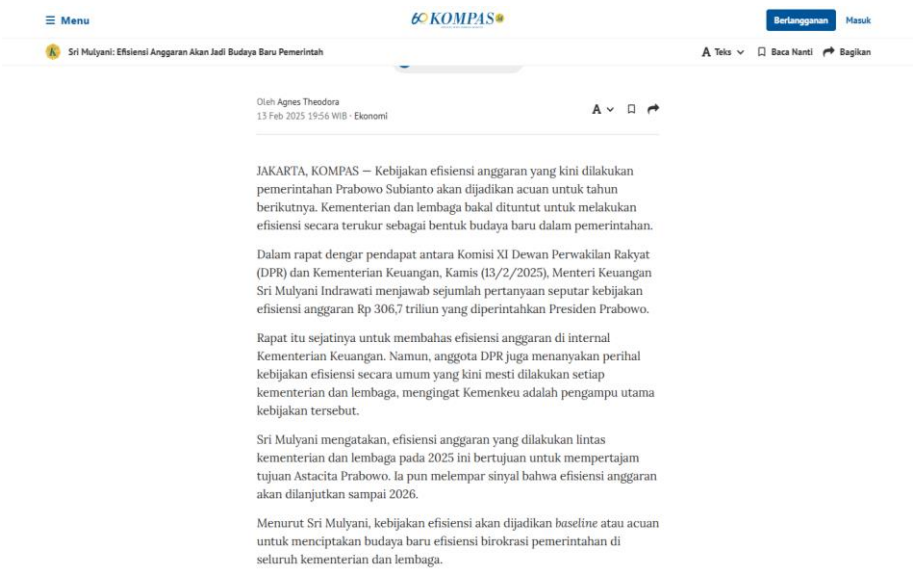
Salah satu contoh pemberitaan dari media daring nasional mengenai kebijakan efisiensi anggaran datang dari Kompas.com melalui artikel berjudul *"Sri Mulyani: Efisiensi Anggaran Akan Jadi Budaya Baru Pemerintah"* yang diterbitkan pada 21 Maret 2024. Artikel ini memuat pernyataan Menteri Keuangan Sri Mulyani Indrawati dalam rapat dengan Komisi XI DPR RI terkait penyusunan anggaran belanja pemerintah tahun 2025.

Dalam laporan ini, Sri Mulyani menegaskan bahwa kebijakan efisiensi anggaran bukan hanya bersifat sementara atau berbasis respons krisis, melainkan akan dijadikan sebagai *budaya baru* dalam tata kelola keuangan negara. Ia menekankan

pentingnya efisiensi dalam setiap tahapan perencanaan dan pelaksanaan anggaran oleh seluruh kementerian dan lembaga, sejalan dengan arahan Presiden Prabowo.

Artikel Kompas.com ini menggambarkan bagaimana kebijakan efisiensi anggaran mulai diposisikan sebagai prinsip tata kelola yang berkelanjutan, bukan sekadar strategi penghematan belaka. Narasi ini menunjukkan bahwa efisiensi mulai di institusionalisasi sebagai bagian dari reformasi birokrasi dan peningkatan akuntabilitas pengelolaan anggaran.

Sebagai media daring nasional dengan jangkauan luas, Kompas.com berperan dalam membentuk opini publik terhadap urgensi dan arah kebijakan fiskal pemerintah. Pemberitaan seperti ini memberi kontribusi penting dalam menyampaikan perkembangan kebijakan kepada masyarakat, sekaligus memperkuat legitimasi pemerintah dalam mendorong efisiensi dan akuntabilitas belanja negara.


Menu

KOMPAS

Berlangganan Masuk

Sri Mulyani: Efisiensi Anggaran Akan Jadi Budaya Baru Pemerintah

Teks Baca Nanti Bagikan

Oleh Agnes Theodora
13 Feb 2025 19:56 WIB · Ekonomi

JAKARTA, KOMPAS — Kebijakan efisiensi anggaran yang kini dilakukan pemerintahan Prabowo Subianto akan dijadikan acuan untuk tahun berikutnya. Kementerian dan lembaga bakal dituntut untuk melakukan efisiensi secara terukur sebagai bentuk budaya baru dalam pemerintahan.

Dalam rapat dengar pendapat antara Komisi XI Dewan Perwakilan Rakyat (DPR) dan Kementerian Keuangan, Kamis (13/2/2025), Menteri Keuangan Sri Mulyani Indrawati menjawab sejumlah pertanyaan seputar kebijakan efisiensi anggaran Rp 306,7 triliun yang diperintahkan Presiden Prabowo.

Rapat itu sejatinya untuk membahas efisiensi anggaran di internal Kementerian Keuangan. Namun, anggota DPR juga menanyakan perihal kebijakan efisiensi secara umum yang kini mesti dilakukan setiap kementerian dan lembaga, mengingat Kemenkeu adalah pengampu utama kebijakan tersebut.

Sri Mulyani mengatakan, efisiensi anggaran yang dilakukan lintas kementerian dan lembaga pada 2025 ini bertujuan untuk mempertajam tujuan Astacita Prabowo. Ia pun melempar sinyal bahwa efisiensi anggaran akan dilanjutkan sampai 2026.

Menurut Sri Mulyani, kebijakan efisiensi akan dijadikan *baseline* atau acuan untuk menciptakan budaya baru efisiensi birokrasi pemerintahan di seluruh kementerian dan lembaga.

Gambar 2.7 Artikel dari *KOMPAS.com*

2.11.2 Kenaikan PPN 12%

Pajak merupakan salah satu hal terpenting dalam upaya pembangunan nasional dan merupakan sumber daya keuangan strategis yang besar bagi suatu negara, sehingga bisa mendukung upaya untuk mendorong pertumbuhan ekonomi dan menciptakan kesejahteraan bagi masyarakat. Berdasarkan data dari Sekretariat Kabinet RI(2023),

pajak menjadi instrumen yang sangat penting untuk perekonomian di Indonesia dan menyumbang sebanyak 64,6% dari sumber penerima negara. Salah satu bentuk penerimaan yang berasal dari pajak adalah PPN. Dikutip dari Kementerian Keuangan *Learning Center*, PPN merupakan pajak yang dipungut atas transaksi penjualan ataupun pembelian suatu barang/jasa di suatu daerah pabean yang sifat pungutnya bertahap di setiap alur produksi maupun distribusi, artinya PPN akan dipungut saat melakukan transaksi atau penyerahan [28].

Berdasarkan amanat Undang-Undang Nomor 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan, pemerintah akan menambah tarif Pajak Pertambahan Nilai menjadi 12% pada tahun depan. BAB 4 Pasal 7 ayat (2) UU HPP secara khusus mengatur tentang PPN yang berubah dari 11% menjadi 12 mulai dari 1 Januari 2025. Dari sisi pemerintah, dengan naiknya tarif PPN ini akan mengoptimalkan penerimaan negara, sehingga dapat mendukung pemulihan ekonomi dan pembangunan.

Untuk mendukung penjelasan mengenai kebijakan kenaikan tarif PPN ini, penelitian ini juga menyertakan berbagai data pendukung dari media digital yang merefleksikan reaksi dan persepsi masyarakat. Data tersebut diperoleh dari platform media sosial X dan YouTube, serta artikel media daring yang melaporkan respons berbagai kalangan terhadap implementasi kebijakan ini. Kehadiran data ini menjadi bukti bahwa kebijakan fiskal seperti PPN tidak hanya dipahami dalam konteks teknokratis, tetapi juga menimbulkan diskusi sosial yang luas. Dokumentasi digital ini akan dianalisis lebih lanjut untuk memetakan sentimen publik terhadap kebijakan PPN 12%, dan memberikan konteks tambahan terhadap dinamika yang muncul di tengah masyarakat secara real-time.

2.11.2.1 Data Media Sosial X Program PPN 12%

Media sosial X menjadi ruang diskusi digital yang mencerminkan respons publik secara real-time terhadap kebijakan pemerintah. Dalam konteks kebijakan kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%, platform ini memperlihatkan tingginya dinamika opini masyarakat, baik dari kalangan akademisi, jurnalis,

ekonom, hingga warga biasa. Topik ini memicu diskusi luas karena dianggap memiliki dampak langsung terhadap daya beli masyarakat dan keberlangsungan sektor usaha kecil menengah.

Berdasarkan penelusuran, beberapa unggahan dengan tingkat interaksi tinggi berasal dari akun-akun publik ternama serta media-media nasional. Selain itu, sejumlah tagar seperti #TolakPPN12Persen, #PPN12%, sempat menjadi trending topic, yang menunjukkan bahwa isu ini bukan hanya hangat, tetapi juga menimbulkan kecemasan sosial yang meluas. Sentimen yang muncul cenderung bersifat negatif, ditandai dengan ekspresi kekecewaan, kritik terhadap beban ekonomi yang ditimbulkan, serta pertanyaan mengenai transparansi dan efektivitas alokasi dana dari kebijakan tersebut.

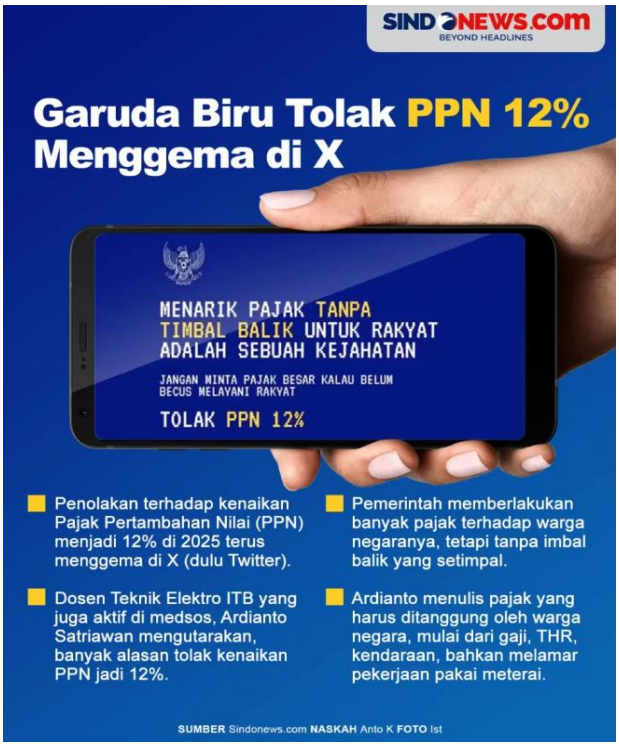
Gambar 2.8 *Tweet @KemenkeuR* - Kebijakan PPN 12%

Gambar di atas merupakan tangkapan layar unggahan dari akun resmi @KemenkeuRI di platform X yang membahas kebijakan baru terkait kenaikan tarif Pajak Pertambahan Nilai (PPN) menjadi 12%. Dalam unggahan tersebut, Kementerian Keuangan menjelaskan bahwa kebijakan ini merupakan bagian dari

paket kebijakan ekonomi yang berlandaskan pada tiga prinsip utama: asas keadilan, keberpihakan kepada masyarakat, dan asas gotong royong.

Unggahan ini juga memuat visualisasi konferensi pers dan infografis berisi penjelasan mengenai nilai-nilai yang mendasari kebijakan PPN 12%. Penekanan diberikan pada perlunya pembiayaan negara yang adil dan kolaboratif, serta pentingnya peran pajak dalam mendukung pembangunan nasional.

Gambar ini menunjukkan bahwa pemerintah aktif memanfaatkan media sosial sebagai sarana komunikasi kebijakan fiskal secara terbuka dan visual. Tingginya intensitas penyampaian pesan publik ini mencerminkan upaya negara untuk mengedukasi sekaligus merespons potensi reaksi masyarakat terhadap perubahan kebijakan perpajakan yang bersifat strategis.



Gambar 2.9 SINDONEWS.com - Penolakan PPN 12%

Gambar di atas merupakan tangkapan layar unggahan *SINDONEWS.com* yang membahas penolakan publik terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%. Dalam unggahan tersebut, terlihat jelas narasi penolakan yang mengemuka di platform X, dengan judul utama “Garuda Biru Tolak PPN 12%

Menggema di X. Visual utama memperlihatkan kutipan keras yang menyebutkan bahwa “menarik pajak tanpa timbal balik untuk rakyat adalah sebuah kejahatan”, yang menjadi representasi dari sentimen negatif terhadap kebijakan tersebut.

Gambar ini menunjukkan bagaimana media daring berperan dalam mengangkat dan menyebarluaskan opini kritis yang berkembang di media sosial. Berdasarkan perspektif sentimen analisis, konten ini dapat dikategorikan sebagai sentimen negatif terhadap kebijakan fiskal pemerintah. Ungkapan-ungkapan emosional, penggunaan diksi tajam, serta banyaknya kutipan dari tokoh kritis menjadi indikator adanya resistensi masyarakat terhadap PPN 12%.

2.11.2.2 Data Media Sosial Youtube Program PPN 12%

Dalam upaya mendalami persepsi publik terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%, platform YouTube dipilih sebagai salah satu sumber data karena menyediakan ruang ekspresi yang lebih panjang dan mendalam. Berbeda dengan platform lain yang lebih mengedepankan unggahan singkat, YouTube memungkinkan masyarakat, kreator konten, serta media untuk mempublikasikan ulasan, diskusi, dan opini dalam bentuk video yang komprehensif.

Penelusuran yang dilakukan menunjukkan bahwa terdapat sejumlah video yang membahas isu ini, baik dari kanal berita, kanal edukasi pajak, hingga kanal pribadi yang menyoroti dampak kebijakan tersebut terhadap sektor usaha dan daya beli masyarakat. Konten-konten tersebut tidak hanya menguraikan informasi kebijakan, namun juga mengundang komentar aktif dari warganet, yang menjadi bagian penting dalam pemetaan sentimen publik.

Data yang diperoleh dari YouTube ini menjadi salah satu elemen pendukung yang merepresentasikan persepsi publik secara lebih luas dan dalam. Oleh karena itu, data ini digunakan dalam proses analisis sentimen untuk mengkaji bagaimana respons masyarakat terhadap kebijakan PPN 12% terbentuk dan tersebar di ruang digital.

Gambar 2.10 Kanal Metro TV - PPN 12%

Gambar diatas menampilkan tangkapan layar dari siaran berita kanal Youtube Metro TV dalam program *Newsline*, yang membahas isu kenaikan tarif Pajak Pertambahan Nilai (PPN) menjadi 12%. Dalam tayangan tersebut, ditampilkan bahwa topik "PPN 12%" menempati posisi *Trending #1* dalam pencarian publik, menandakan tingginya potensi masyarakat terhadap kebijakan fiskal ini di ruang digital.

Tayangan ini menunjukkan bagaimana media televisi nasional ikut serta dalam menyampaikan informasi mengenai isu kebijakan yang sedang berkembang di media sosial. Dengan menyoroti tingginya jumlah pencarian terkait PPN 12%, tayangan ini memberikan gambaran bahwa kebijakan tersebut tidak hanya berdampak pada aspek ekonomi, tetapi juga menimbulkan reaksi dan diskusi luas di ruang publik. Dalam konteks analisis sentimen, perhatian besar masyarakat terhadap topik ini mencerminkan adanya kekhawatiran, pertanyaan, maupun kritik yang perlu dicermati dalam proses evaluasi kebijakan pemerintah.

Gambar 2.11 Kompas TV - Dampak PPN 12%

Gambar di atas menampilkan hasil tangkapan layar dari channel Kompas TV terkait kebijakan kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%. Narasi utama dalam tayangan ini, yang ditampilkan melalui judul "PPN 12 Persen, Keputusan Bijak untuk Rakyat?", menunjukkan bahwa media mencoba menghadirkan ruang reflektif bagi pemirsa untuk menilai apakah kebijakan fiskal tersebut benar-benar membawa manfaat bagi rakyat.

Tayangan ini secara tidak langsung membentuk persepsi publik melalui narasi yang seimbang menampilkan pandangan pemerintah sekaligus membuka ruang untuk pertanyaan kritis. Dalam konteks analisis media dan sentimen publik, unggahan ini dapat mencerminkan sentimen campuran (*mixed sentiment*), di mana terdapat dukungan terhadap langkah pemerintah sebagai bentuk reformasi pajak, namun juga muncul kekhawatiran terhadap dampaknya terhadap daya beli masyarakat. Visual ini memperlihatkan bahwa media arus utama berperan penting dalam memediasi komunikasi antara pembuat kebijakan dan masyarakat, sekaligus memperkuat fungsi kontrol publik terhadap kebijakan fiskal yang diambil pemerintah.

Gambar 2.12 Kompas TV - Pemberlakuan PPN 12%

Video diatas yang menjadi salah satu referensi utama dalam analisis persepsi publik terhadap kebijakan kenaikan PPN 12% diambil dari kanal KOMPASTV, sebuah media arus utama yang kredibel dan memiliki jangkauan audiens luas. Video ini menampilkan pernyataan resmi dari Menteri Keuangan Republik Indonesia terkait pemberlakuan PPN sebesar 12% yang akan efektif mulai 1 Januari 2025. Dalam penyampaiannya, Menteri Keuangan menjelaskan bahwa kebijakan ini mencakup barang dan jasa tertentu, termasuk rumah sakit VIP dan makanan impor, yang menjadi perhatian utama masyarakat.

Video ini telah memperoleh ribuan komentar dari warganet, menandakan tingginya antusiasme dan perhatian publik terhadap isu tersebut. Banyak komentar mengungkapkan kekhawatiran terkait dampak kenaikan PPN terhadap daya beli masyarakat, khususnya di tengah kondisi ekonomi yang masih belum stabil. Beberapa komentar lainnya menyoroti potensi dampak negatif terhadap sektor usaha dan UMKM.

Melalui video ini, terlihat bahwa YouTube menjadi sarana penting dalam menyampaikan informasi kebijakan pemerintah secara luas sekaligus sebagai wadah dialog antara pemerintah dan masyarakat. Komentar-komentar dalam video ini kemudian digunakan sebagai bagian dari data analisis sentimen dalam penelitian

ini, untuk memetakan persepsi publik secara lebih mendalam terhadap kebijakan PPN 12%.

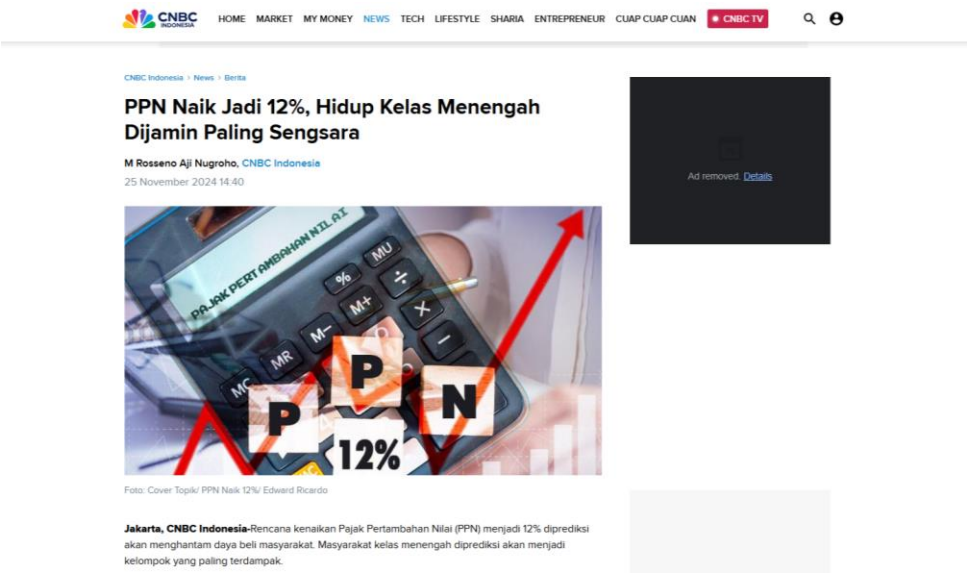
2.11.2.3 Data Media Daring Program PPN 12%

Untuk memperoleh gambaran yang lebih komprehensif mengenai respons publik terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%, peneliti turut menghimpun data dari berbagai media daring nasional yang kredibel. Media daring menjadi salah satu kanal utama dalam penyebaran informasi serta pembentukan opini publik di era digital saat ini. Melalui artikel, wawancara, dan pernyataan resmi yang dimuat di media tersebut, peneliti dapat mengkaji sudut pandang pemerintah, tanggapan ekonom, serta reaksi masyarakat luas terhadap rencana penerapan PPN 12% yang dijadwalkan mulai 1 Januari 2025.

Gambar 2.13 Artikel dari *KOMPAS.com*

Berita di atas ini menyajikan penjelasan dari pemerintah terkait latar belakang kebijakan kenaikan PPN serta proyeksi dampaknya terhadap perekonomian nasional. Terlihat dalam artikel tersebut para pejabat negara tengah memberikan keterangan resmi dalam sebuah konferensi pers yang disiarkan ke publik, menunjukkan bahwa isu ini ditanggapi secara serius dan terbuka oleh pemerintah. Penayangan berita ini di media daring nasional yang kredibel menegaskan bahwa topik PPN 12% menjadi perhatian utama dalam wacana publik, dan media turut

berperan sebagai penghubung antara kebijakan pemerintah dan pemahaman masyarakat.



Gambar 2.14 Artikel dari *CNBC Indonesia*

Berita di atas ini menyajikan penjelasan dari pemerintah terkait latar belakang kebijakan kenaikan PPN serta proyeksi dampaknya terhadap perekonomian nasional. Terlihat dalam artikel tersebut media memberikan keterangan resmi dalam sebuah konferensi pers yang disiarkan ke publik, menunjukkan bahwa isu ini ditanggapi secara serius dan terbuka oleh pemerintah. Penayangan berita ini di media daring nasional yang kredibel menegaskan bahwa topik PPN 12% menjadi perhatian utama dalam wacana publik, dan media turut berperan sebagai penghubung antara kebijakan pemerintah dan pemahaman masyarakat.

2.11.3 Program Makan Bergizi Gratis

Makan gratis merupakan penyediaan makanan tanpa harus membayarnya, dan bagi banyak orang, hal ini menjadi sesuatu yang sangat diidamkan. Mengingat makan adalah kebutuhan pokok setiap hari, maka ketika dilekatkan dengan konsep “gratis”, hal tersebut tentu membawa kebahagiaan tersendiri karena seseorang dapat memenuhi kebutuhan tersebut tanpa beban biaya. Manusia membutuhkan makanan untuk bertahan hidup, hidup tanpa makan merupakan gagasan yang tidak realistis. Sebagai makhluk hidup, manusia berjalan, bernapas, dapat jatuh sakit, merasa lapar,

dan pada akhirnya akan menghadapi kematian. Oleh karena itu, makan bukan hanya kebutuhan, tetapi juga kewajiban demi mempertahankan kehidupan.

Makan Bergizi Gratis (MBG) sebagai kebijakan prioritas Pemerintah Prabowo Subianto-Gibran Rakabumi Raka sudah dilakukan pemerintah kota Tangerang sejak Agustus 2024. Para menteri di era pemerintah Joko Widodo hingga Wapres Gibran Rakabuming Raka sudah meninjau pelaksanaan MBG bagi para murid sekolah di daerah tersebut. Dikutip dari website yang sama, Direktur Eksekutif Indef Esther Sri Astuti juga menemukan, efek program MBG (Makan Bergizi Gratis) ke PDB (Produksi Domestik Bruto) akan terus meningkatkan seiring bertambahnya jumlah peserta dan anggaran yang dialokasikan pemerintah[36].

Untuk memperkuat pembahasan kebijakan ini, penelitian ini menyertakan bukti-bukti digital yang mencerminkan bagaimana masyarakat memberikan tanggapan terhadap Program Makan Bergizi Gratis (MBG). Data yang dikumpulkan meliputi unggahan dari platform media sosial X dan YouTube, serta berita dari media daring nasional dan internasional yang menyoroti berbagai aspek pelaksanaan program ini, baik dari sisi keberhasilan maupun tantangan. Kehadiran data ini memberikan gambaran yang lebih luas dan realistis mengenai bagaimana program MBG diterima oleh publik. Selain itu, data ini juga berperan penting dalam proses analisis sentimen untuk memetakan persepsi masyarakat terhadap kebijakan tersebut secara objektif dan berbasis data.

2.11.3.1 Data Media Sosial X Program Makan Bergizi Gratis

Media sosial menjadi salah satu sumber penting untuk mengamati persepsi publik terhadap isu-isu yang sedang ramai diperbincangkan. Dalam konteks penelitian ini, platform X digunakan untuk mengumpulkan data berupa opini, kritik, dan diskusi publik mengenai program Makan Bergizi Gratis (MBG). Beberapa unggahan dari akun media resmi seperti Kompas, CNN, dan Tempo menunjukkan intensitas perhatian masyarakat terhadap isu ini.

Gambar 2.15 @tempodotco - Kasus keracunan program MBG

Gambar diatas merupakan tangkapan layar unggahan dari akun resmi media @tempodotco di platform X yang menyoroti kejadian keracunan massal akibat konsumsi makanan dari program Makan Bergizi Gratis (MBG) di Cianjur, Jawa Barat. Dalam unggahan tersebut, disebutkan bahwa Dinas Kesehatan Kabupaten Cianjur menetapkan status Kejadian Luar Biasa (KLB) setelah puluhan siswa mengalami gejala keracunan. Gambar ini menunjukkan bahwa kasus ini mendapat perhatian serius dari media nasional dan mencerminkan adanya kekhawatiran publik terhadap keamanan pelaksanaan program MBG.

Unggahan ini juga menandakan bahwa isu MBG telah menjadi topik penting dalam diskursus publik, terutama ketika menyangkut aspek kesehatan dan keselamatan siswa yang menjadi sasaran program. Dengan lebih dari 19.3 ribu tayangan, 324 retweet, dan 895 suka, unggahan ini menunjukkan tingginya interaksi dan keterlibatan warganet terhadap isu tersebut.

Gambar 2.16 @CNNIndonesia - Sorotan kasus keracunan MBG

Gambar diatas memperlihatkan tangkapan layar unggahan dari akun resmi @CNNIndonesia di platform X, yang memuat pemberitaan mengenai sorotan media internasional terhadap kasus keracunan makanan yang menimpa murid sekolah dasar di Sukoharjo dalam program Makan Bergizi Gratis (MBG). Dalam unggahan tersebut, CNN Indonesia menekankan bahwa isu ini telah menarik perhatian media asing, yang menunjukkan bahwa permasalahan ini tidak hanya berdampak secara lokal, tetapi juga mencerminkan citra nasional di mata internasional.

Tingginya jumlah tayangan, tanggapan, dan interaksi yang ditunjukkan pada unggahan ini mengindikasikan bahwa publik sangat responsif terhadap pemberitaan yang memiliki dimensi global. Unggahan ini menjadi bagian penting dalam dokumentasi persepsi publik dan representasi media terhadap pelaksanaan program MBG, khususnya dalam konteks krisis dan pengawasan terhadap kebijakan pemerintah.

Gambar 2.17 @kompascom - Sorotan Internasional terhadap MBG

Gambar diatas menampilkan tangkapan layar dari unggahan akun resmi @kompascom di platform X yang menginformasikan bahwa program Makan Bergizi Gratis (MBG) di Indonesia telah menjadi perhatian sejumlah media internasional, seperti BBC, Malay Mail, *South China Morning Post,* dan *The Guardian*. Dalam unggahan tersebut, Kompas menggarisbawahi bahwa kasus-kasus keracunan massal di sekolah yang diduga berkaitan dengan program MBG memunculkan kekhawatiran publik terhadap aspek keamanan pangan serta transparansi dalam pengelolaan anggaran.

Narasi yang dibangun dalam unggahan ini tidak hanya menggambarkan sisi kritis dari implementasi program MBG, tetapi juga mengajak masyarakat untuk memberikan masukan terhadap evaluasi kebijakan pemerintah. Selain itu, adanya infografik yang mendukung penjelasan memperkuat daya informatif unggahan ini dan menunjukkan upaya media dalam menyampaikan isu dengan pendekatan visual yang mudah dipahami.

Unggahan ini mencerminkan peran media nasional dalam menjembatani isu domestik dengan perhatian global, serta memberikan tekanan terhadap perlunya perbaikan sistematis dalam pelaksanaan program-program sosial berskala nasional seperti MBG.

2.11.3.2 Data Media Sosial YouTube Program Makan Bergizi Gratis

YouTube merupakan salah satu platform media sosial berbasis video yang banyak dimanfaatkan oleh masyarakat untuk menyebarkan dan mengakses informasi secara luas. Dalam konteks kajian ini, YouTube berperan penting sebagai sumber data karena mampu merepresentasikan opini publik, pemberitaan media, serta respons terhadap isu-isu aktual secara visual dan naratif. Isu mengenai program Makan Bergizi Gratis (MBG) yang dicanangkan oleh pemerintah turut menjadi perhatian dalam berbagai unggahan video, baik oleh kanal media resmi maupun konten kreator independen.

Melalui beberapa video unggahan yang dijadikan data dalam penelitian ini, dapat diamati bagaimana dinamika pemberitaan, persepsi masyarakat, serta bentuk kritik maupun dukungan terhadap implementasi program MBG tersebar luas di platform ini. Analisis terhadap konten YouTube ini menjadi penting dalam menggambarkan persepsi publik secara lebih komprehensif, sekaligus melengkapi data dari platform sosial media lainnya seperti X.

Gambar 2.18 CNN Indonesia - Mitra MBG Tidak Dibayar RP1 M

Gambar diatas menampilkan tangkapan layar dari *channel* CNN Indonesia yang menginformasikan tentang permasalahan pembayaran kepada mitra penyedia makanan dalam program Makan Bergizi Gratis (MBG). Dalam tayangan tersebut, narasumber dari pihak terkait memberikan keterangan bahwa terjadi keterlambatan pembayaran sebesar Rp1 miliar kepada mitra penyedia, yang kemudian berujung

pada pelaporan kepada pihak kepolisian. Sementara itu, pihak BGN menyatakan bahwa dana telah ditransfer kepada yayasan yang bersangkutan.

Narasi yang dibangun dalam tayangan ini menyoroti persoalan administratif dan keuangan dalam pelaksanaan program MBG, khususnya terkait akuntabilitas dan transparansi distribusi anggaran. Tayangan ini sekaligus mencerminkan dinamika yang terjadi di lapangan dan memperlihatkan perlunya perbaikan tata kelola dalam implementasi kebijakan sosial berskala besar seperti program MBG.

Gambar 2.19 Liputan 6 - Prabowo & Bill Gates Tinjau MBG

Gambar diatas menampilkan hasil tangkapan layar dari *channel* Liputan 6 yang mendokumentasikan kunjungan Presiden Prabowo Subianto bersama tokoh dunia, Bill Gates, ke salah satu institusi pendidikan di Indonesia. Kunjungan tersebut dilakukan dalam rangka meninjau langsung pelaksanaan program Makan Bergizi Gratis (MBG). Narasi yang ditampilkan dalam tayangan ini menggaris bawahi komitmen pemerintah dan dukungan internasional terhadap program MBG, yang dipandang sebagai upaya strategis untuk meningkatkan kualitas gizi anak-anak Indonesia. Kehadiran tokoh internasional dalam kunjungan ini memperkuat citra positif program di mata publik, sekaligus menunjukkan dimensi diplomatik dan kolaboratif dalam penyelenggaraan program sosial. Visual yang dihadirkan memberikan kesan optimistis terhadap keberlanjutan program MBG dan

membangun narasi bahwa pemerintah aktif memantau langsung pelaksanaan kebijakan di lapangan.

Gambar 2.20 Tempodotco - Program MBG

Gambar diatas menampilkan tangkapan layar dari unggahan *channel* YouTube *Tempodotco*, yang membahas Program Makan Bergizi Gratis (MBG) dalam segmen *Tempo Explained*. Dalam gambar terlihat seorang tokoh nasional, yakni calon presiden Prabowo Subianto, sedang menyampaikan pernyataan dalam forum resmi debat pemilu yang diselenggarakan oleh Komisi Pemilihan Umum (KPU RI). Narasi yang dibangun dalam video ini bertujuan memberikan penjelasan mendalam mengenai gagasan dan strategi calon presiden terkait program MBG, termasuk tujuan utama, target penerima manfaat, serta mekanisme pelaksanaannya. Tayangan ini tidak hanya berfungsi sebagai dokumentasi pernyataan politik, tetapi juga sebagai bahan edukasi publik yang mengajak masyarakat untuk memahami latar belakang dan urgensi program MBG dalam konteks pembangunan nasional. Melalui visual yang kuat dan pendekatan analitis khas media investigatif seperti Tempo, unggahan ini memperkaya diskursus publik tentang program MBG dari perspektif politik dan kebijakan.

2.11.3.3 Data Media Daring Program Makan Bergizi Gratis

Media daring nasional dan internasional memiliki peran signifikan dalam menyampaikan informasi dan membentuk opini publik terkait isu-isu yang

berkembang di masyarakat. Dalam konteks program Makan Bergizi Gratis (MBG) yang menjadi salah satu kebijakan unggulan Presiden Prabowo Subianto, sejumlah media daring aktif memberitakan perkembangan pelaksanaan program tersebut, termasuk tantangan, kritik, dan dampaknya di lapangan.

Sumber berita daring seperti *The Straits Times*, dan *The Guardian* tidak hanya menyampaikan informasi faktual, namun juga merefleksikan tanggapan masyarakat dan berbagai pemangku kepentingan. Pemberitaan tersebut mencakup sorotan dari media asing, kritik terhadap transparansi program, hingga isu terkait keamanan pangan dan pengelolaan anggaran. Dokumentasi terhadap data media daring ini penting sebagai bagian dari analisis sentimen publik yang dilakukan dalam penelitian ini.

Gambar 2.21 Artikel dari *The Straits Times*

Gambar diatas merupakan tangkapan layar dari artikel yang dipublikasikan oleh *The Straits Times*, salah satu media arus utama di kawasan Asia Tenggara yang berbasis di Singapura. Artikel tersebut berjudul *"Indonesian President's signature free-meal scheme well received but faced early snags"* dan ditulis oleh jurnalis senior Wahyudi Soeriaatmadja. Dalam artikel ini, *The Straits Times* memberikan perhatian terhadap pelaksanaan awal Program Makan Bergizi Gratis (MBG) yang merupakan salah satu program prioritas Presiden Prabowo Subianto.

Secara umum, artikel ini menyoroti bahwa meskipun program MBG memperoleh sambutan positif dari masyarakat karena tujuannya dalam meningkatkan kualitas gizi anak-anak sekolah, terdapat pula sejumlah tantangan yang muncul dalam fase implementasi awal. Beberapa kendala yang dikemukakan antara lain terkait dengan

aspek distribusi makanan, kesiapan logistik, keterbatasan infrastruktur sekolah, serta pengawasan terhadap keamanan pangan. Permasalahan tersebut menunjukkan adanya kebutuhan mendesak untuk penguatan tata kelola dan mekanisme pelaksanaan program secara sistemik.

Sorotan dari media internasional ini mengindikasikan bahwa kebijakan domestik di Indonesia, khususnya yang berkaitan dengan isu sosial dan kesejahteraan anak, memiliki daya tarik bagi komunitas global. Dengan demikian, publikasi ini tidak hanya mencerminkan respons publik internasional terhadap kebijakan pemerintah Indonesia, tetapi juga memberikan tekanan terhadap pentingnya evaluasi berkelanjutan dan peningkatan kualitas pelaksanaan program MBG agar selaras dengan standar pelayanan publik yang efektif dan aman.

Gambar 2.22 Artikel dari *The Guardian*

Gambar di atas merupakan tangkapan layar dari artikel yang diterbitkan oleh *The Guardian*, salah satu media internasional terkemuka berbasis di Inggris. Artikel ini secara kritis mengangkat isu terkait kasus keracunan makanan yang terjadi dalam implementasi awal Program Makan Bergizi Gratis (MBG) di Indonesia program unggulan Presiden Prabowo Subianto yang bertujuan meningkatkan asupan gizi anak-anak usia sekolah.

Dalam artikel tersebut, *The Guardian* menyoroti bahwa insiden keracunan yang menimpa sejumlah siswa sekolah dasar di berbagai daerah tidak hanya menimbulkan keprihatinan publik domestik, tetapi juga mengundang perhatian komunitas internasional terhadap aspek keamanan pangan dalam kebijakan sosial

pemerintah Indonesia. Penggunaan frasa *"mars flagship free meal program"* mengindikasikan bahwa insiden tersebut dianggap mencederai kredibilitas program yang semestinya membawa dampak positif jangka panjang terhadap kesehatan generasi muda.

Lebih jauh, artikel ini mencerminkan adanya ekspektasi global terhadap pelaksanaan program publik berskala besar yang tidak hanya berorientasi pada kuantitas atau jangkauan penerima manfaat, tetapi juga pada kualitas layanan, termasuk standar kebersihan, pengawasan distribusi, dan efektivitas manajemen risiko. Dengan demikian, sorotan dari *The Guardian* mempertegas urgensi evaluasi menyeluruh dan reformulasi mekanisme pengawasan program MBG guna memastikan keberlanjutan dan akuntabilitas kebijakan di mata publik nasional maupun internasional.

2.11.4 Danantara

Daya Anagata Nusantara *Investment Management Agency*, atau Danantara Indonesia, adalah dana kekayaan negara (*sovereign wealth fund*) kedua Indonesia setelah Indonesia *Investment Authority* (INA). Badan ini didirikan untuk mengkonsolidasi dan mengoptimalkan investasi pemerintahan guna mendukung pertumbuhan ekonomi nasional. Nama “Daya Anagata Nusantara” diberikan oleh Presiden Prabowo Subianto, dengan “Daya” berarti energi, “Anagata” berarti masa depan, dan “Nusantara” merujuk pada Negara Kesatuan Republik Indonesia, mencerminkan kekuatan dan potensi masa depan Indonesia. Danantara diresmikan pada 24 Februari 2025 oleh Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming Raka di Istana Merdeka, Jakarta. Badan ini mengelola aset dengan total sekitar US$900 miliar, menjadikannya salah satu dana kekayaan negara terbesar di dunia. Danantara bertujuan untuk mendorong transformasi ekonomi Indonesia dengan menumbuhkan badan *Sovereign Wealth Fund* berskala dunia, mendukung pembangunan nasional, dan menciptakan kemakmuran bagi seluruh rakyat Indonesia. Dari beberapa media sosial, tanggapan masyarakat mengenai Danantara beragam. Sebagian mendukung adanya keberadaan Danantara ini karena dapat membuka lapangan pekerjaan, dan mempercepat pembangunan ekonomi. Namun, ada juga yang mempertanyakan keterbukaan dan potensi penyalahgunaan

dana, mengingat besarnya investasi yang dikelola. Sementara itu, beberapa pihak memilih bersikap netral dengan menunggu hasil nyata dari implementasi Danantara sebelum memberikan penilaian lebih lanjut.

Untuk memberikan gambaran yang lebih komprehensif mengenai respons publik terhadap kebijakan ini, penelitian ini juga menyertakan bukti-bukti dari media digital, seperti unggahan pengguna di platform media sosial X dan YouTube, serta artikel dari media daring nasional maupun internasional. Data tersebut mencerminkan spektrum opini masyarakat terhadap Danantara, mulai dari dukungan antusias hingga keraguan terhadap transparansi dan akuntabilitas pengelolaan dana investasi. Dokumentasi digital ini berfungsi sebagai landasan dalam proses analisis sentimen dan topik, serta memberikan konteks sosial yang lebih kaya terhadap penerimaan publik terhadap kebijakan Danantara. Dengan demikian, data ini menjadi penting dalam mengukur persepsi masyarakat secara obyektif berdasarkan bukti nyata dari ruang digital.

2.11.4.1 Data Media Sosial X Program Danantara

Media sosial X berperan penting sebagai wadah diskusi publik yang cepat dan responsif terhadap isu-isu terkini, termasuk kebijakan Danantara. Topik yang menjadi perbincangan di X seringkali mencerminkan reaksi spontan masyarakat terhadap kebijakan yang baru diperkenalkan. Melalui analisis data percakapan di platform ini, dapat dilihat bagaimana dinamika sentimen, volume perbincangan, serta isu-isu turunan yang muncul dalam konteks Program Danantara. Sub-bab ini akan menyajikan cuplikan data dan insight dari perbincangan di X terkait kebijakan tersebut.

Gambar dibawah ini merupakan tangkapan layar dari unggahan akun resmi Kompas.com di platform media sosial X. Dalam unggahan tersebut, Kompas menyampaikan informasi penting mengenai rencana kerjasama investasi antara Danantara dan Bill Gates, seorang tokoh bisnis dan filantropis ternama asal Amerika Serikat. Kerja sama tersebut dijadwalkan akan dilaksanakan pada bulan September 2025 dan akan mencakup tiga sektor strategis, sebagaimana dilansir oleh situs berita *money.kompas.com*.

Tagar yang digunakan seperti **#Danantara**, **#Investasi**, dan **#BillGates** menunjukkan upaya untuk menjangkau audiens yang lebih luas dan meningkatkan eksposur isu tersebut di ruang digital. Keterlibatan media arus utama seperti Kompas dalam memberikan informasi ini melalui platform X menjadi bukti bahwa kebijakan Danantara tidak hanya dibahas secara formal, tetapi juga menjadi bagian dari pertukaran ide publik di media sosial.

Tingginya potensi *engagement* dari unggahan ini dilihat dari kontennya yang melibatkan tokoh global dan institusi nasional mendukung fakta bahwa kebijakan Danantara telah menarik perhatian luas dan memicu diskusi publik secara daring. Oleh karena itu, unggahan ini sangat relevan sebagai data pendukung dalam menganalisis persepsi publik terhadap program Danantara melalui media sosial X.

Gambar 2.23 @kompascom- Tentang Danantara

2.11.4.2 Data Media Sosial YouTube Program Danantara

YouTube sebagai platform berbagi video memiliki daya tarik yang kuat dalam menyampaikan informasi secara visual dan naratif. Banyak kanal media, analis, maupun tokoh publik yang menggunakan YouTube untuk menjelaskan kebijakan Danantara, baik dalam bentuk berita, podcast, maupun konten edukatif. Melalui kolom komentar dan jumlah penayangan, tingkat ketertarikan serta reaksi

masyarakat terhadap topik ini dapat dinilai. Sub-bab ini akan membahas data dan tren seputar konten-konten YouTube yang berkaitan dengan Program Danantara.

Gambar dibawah ini menampilkan cuplikan dari *channel* MetroTV. Salah satu fokus utama dalam tayangan ini adalah penguatan arah pembangunan ekonomi nasional di bawah pemerintahan baru, termasuk pengenalan dan penguatan sistem Danantara sebagai bagian dari infrastruktur kelembagaan strategis.

Platform YouTube dipilih sebagai media distribusi karena mampu menjangkau audiens yang lebih luas dan heterogen, serta memungkinkan interaksi langsung melalui komentar dan tanggapan pengguna. Tayangan ini mendapatkan engagement yang tinggi, menandakan bahwa wacana seputar Danantara cukup menarik perhatian publik, terutama dalam kaitannya dengan masa depan perekonomian nasional.

Data visual dan naratif dari tayangan ini memperkuat validitas Danantara sebagai topik yang tidak hanya dibicarakan oleh pemerintah dan media, tetapi juga dibahas secara aktif oleh masyarakat melalui kanal digital populer. Dengan demikian, cuplikan ini menjadi elemen pendukung penting dalam kajian analisis sentimen terhadap kebijakan Danantara di media sosial, khususnya YouTube.

Gambar 2.24 MetroTV - Sistem Danantara

Gambar dibawah ini diambil dari video YouTube resmi CNN Indonesia berjudul *"*Mengenal Apa Itu Danantara*"*. Tayangan ini menyajikan laporan informatif

mengenai pendirian Danantara, lembaga baru yang digagas sebagai bentuk *sovereign wealth fund* Indonesia dengan peran strategis dalam pembangunan ekonomi nasional. Video ini bertujuan memberikan pemahaman kepada masyarakat mengenai latar belakang, fungsi, serta visi dari Danantara dalam mendorong investasi dan pengelolaan aset negara yang lebih produktif. Keberadaan tayangan ini di kanal resmi CNN Indonesia menunjukkan tingkat kepentingan dan perhatian media nasional terhadap isu ini. Selain itu, komentar serta tanggapan warganet pada video tersebut mencerminkan munculnya rasa ingin tahu dan diskusi publik yang aktif terhadap konsep Danantara, menjadikannya data relevan dalam pengkajian opini publik melalui platform YouTube.

Gambar 2.25 CNN Indonesia - Apa itu Danantara?

Gambar dibawah ini merupakan tangkapan layar dari kanal YouTube TV Tempo, dalam program *"Jelasin Dong!"* yang secara khusus membahas topik Risiko Tinggi Proyek Danantara. Dalam diskusi ini, para narasumber dari kalangan jurnalis dan pengamat kebijakan membahas secara kritis mengenai potensi dan tantangan dari proyek Danantara, termasuk implikasi politik dan ekonominya.

Program ini mengambil pendekatan dialog terbuka yang bersifat edukatif dan analitis, serta menyasar audiens yang ingin memahami isu kebijakan secara lebih mendalam. Penayangan konten ini di kanal media besar seperti Tempo memperkuat validitas bahwa isu Danantara telah menjadi bagian dari perdebatan publik dan wacana media arus utama. Tingginya partisipasi penonton melalui *views*, *likes*, dan

komentar juga mengindikasikan bahwa konten ini berkontribusi dalam pembentukan opini publik di ranah digital, khususnya YouTube.

Gambar 2.26 Unggahan - Risiko Tinggi Proyek Danantara

2.11.4.3 Data Media Daring Program Danantara

Selain media sosial, media daring seperti portal berita nasional dan lokal juga memegang peranan penting dalam menyebarkan informasi kebijakan Danantara. Media daring sering kali menjadi sumber rujukan utama masyarakat dalam memahami kebijakan pemerintah. Analisis terhadap artikel-artikel berita, tajuk opini, dan laporan khusus di media daring dapat memberikan wawasan lebih mendalam mengenai bagaimana kebijakan ini diposisikan oleh media, serta bagaimana informasi tersebut diterima oleh pembaca. Sub-bab ini akan memuat rangkuman pemberitaan dan narasi media daring mengenai Program Danantara.

Gambar dibawah ini merupakan tangkapan layar dari artikel berita yang diterbitkan oleh ***Reuters***, salah satu media internasional yang kredibel dan berpengaruh. Artikel tersebut berjudul *"Indonesia's new sovereign fund will run with commercial mindset, official says"* dan dipublikasikan pada 28 Februari 2025. Dalam artikel ini, disampaikan bahwa lembaga baru ini akan dijalankan dengan orientasi komersial, bukan sekadar birokratis. Pandangan tersebut menegaskan bahwa Danantara diposisikan sebagai entitas strategis dalam mengelola kekayaan negara secara

produktif, transparan, dan profesional dengan prinsip tata kelola layaknya perusahaan publik.

Pemilihan berita dari Reuters ini memperkuat signifikansi Program Danantara di mata dunia internasional dan menunjukkan bahwa kebijakan ini tidak hanya menjadi perhatian domestik, tetapi juga diamati secara global. Eksistensi Danantara sebagai lembaga pengelola investasi dengan pendekatan korporasi, menunjukkan adanya perubahan paradigma dalam pengelolaan dana publik di Indonesia menuju tata kelola yang lebih modern dan akuntabel. Artikel ini menjadi salah satu bukti bahwa pembentukan Danantara mendapat perhatian dari berbagai kalangan, termasuk investor asing dan pelaku ekonomi internasional.

Dengan mencantumkan artikel dari sumber internasional yang kredibel seperti Reuters, peneliti ingin memberikan gambaran yang lebih komprehensif terhadap persepsi dan ekspektasi global atas arah kebijakan strategis ekonomi Indonesia, khususnya terkait Program Danantara.

Gambar 2.27 Artikel dari *Reuters*

Gambar dibawah ini merupakan dokumentasi dari artikel dan tayangan video berita yang dipublikasikan oleh **MetroTV News**, salah satu media arus utama di Indonesia yang memiliki reputasi tinggi dalam penyajian berita ekonomi dan kebijakan

publik. Artikel berjudul *"Danantara Digadang-Gadang Percepat Pertumbuhan Ekonomi Nasional"* ini diterbitkan pada 27 Februari 2025, dan secara khusus menyoroti peran strategis Danantara sebagai lembaga baru yang dirancang untuk mendorong investasi dan memperkuat pertumbuhan ekonomi nasional.

Dari sisi komunikasi publik, pemberitaan ini mencerminkan upaya pemerintah dalam membangun narasi positif terhadap Danantara, dengan menekankan potensi besar lembaga ini dalam mendongkrak investasi dan menumbuhkan kepercayaan investor. Penyampaian visual dan editorial dari media nasional seperti MetroTV juga memperlihatkan dukungan naratif yang kuat terhadap kebijakan ini, yang turut berpengaruh dalam membentuk opini publik.

Oleh karena itu, data dari media daring nasional ini memperkaya kajian analisis sentimen dengan menambahkan dimensi persepsi yang dibentuk oleh media mainstream. Ini penting untuk memahami bagaimana pesan-pesan strategis mengenai Danantara dikomunikasikan kepada masyarakat luas dan bagaimana respons awal yang tercermin melalui liputan berita tersebut.

Gambar 2.28 Artikel dari *MetroTV*

2.12 Lexicon Based

Metode *Lexicon-Based* merupakan salah satu pendekatan yang umum digunakan dalam analisis sentimen, khususnya pada data yang bersumber dari media sosial. Pendekatan ini menggunakan sebuah kamus kata (*Lexicon*) yang berisi daftar kosakata beserta skor sentimen atau polaritas yang melekat pada masing-masing

kata. Dengan memanfaatkan kamus tersebut, suatu teks dianalisis berdasar akumulasi nilai sentimen dari setiap kata, sehingga dapat diklasifikasikan ke dalam kategori sentimen positif, negatif, atau netral [29].

Prinsip dasar metode *Lexicon Based* adalah terlebih dahulu menyusun kamus kata opini yang memuat kata-kata beserta bobot polaritas sentimen. Kamus ini kemudian digunakan untuk mengidentifikasi dan mengklasifikasikan sentimen dalam teks secara langsung, tanpa perlu melalui proses pelatihan model seperti yang dilakukan pada metode pembelajaran(*machine learning*)[38]. Kelebihan dari pendekatan ini adalah kemampuannya untuk bekerja tanpa data latih berlabel, sehingga sangat berguna pada situasi di mana data berlabel sulit diperoleh.

Dalam penelitian ini, kamus sentimen yang digunakan merupakan hasil kombinasi dari dua sumber utama. Pertama *lexicon* yang disusun secara manual berdasarkan data komentar dan *tweet* yang diperoleh dari platform media sosial X dan Youtube. Data tersebut telah melalui proses kurasi dan validasi untuk memastikan kosakata yang digunakan merepresentasikan opini yang relevan serta sesuai konteks. Untuk mengumpulkan sentimen negatif dan positif, peneliti mengumpulkan data frekuensi dan dibuat ke dalam normalisasi dan peneliti membagi secara merata secara otomatis menjadi 3 set teks yaitu teks yang mengandung emosi positif, seperti kebahagiaan, kesenangan, atau sukacita. Teks selanjutnya yang mengandung emosi negatif, seperti kesedihan, kekecewaan dan kemarahan terakhir, teks objektif yang hanya menyatakan fakta atau tidak mengekspresikan emosi [30].

Kedua, digunakan pula *Indonesia Sentimen Lexicon (InSet),* yaitu sebuah kamus sentimen Bahasa Indonesia yang dikembangkan oleh Fajri Koto, seorang peneliti dan kontributor aktif dalam pengembangan sumber terbuka di platform GitHub melalui repositori dengan nama pengguna *fajri91.* InSet merupakan salah satu lexicon terpercaya yang telah banyak digunakan dalam berbagai penelitian analisis sentimen karena memiliki cakupan kosakata yang luas yang terkalibrasi dengan baik dalam konteks Bahasa Indonesia[31].

Penggabungan kedua sumber tersebut bertujuan untuk memperkaya daftar kosakata sentimen yang digunakan dalam analisis, sehingga mampu meningkatkan cakupan serta akurasi klasifikasi. Strategi ini diharapkan dapat menangkap berbagai ekspresi opini, baik bersifat formal dan informal, yang kerap ditemukan dalam teks-teks di media sosial [31].

Proses pemberian label sentimen terhadap data teks dari *platform* X dan Youtube dilakukan dengan pendekatan *lexicon-based* sebagaimana dijelaskan sebelumnya. Setiap teks dianalisis dengan menjumlahkan nilai sentimen dari setiap kata berdasarkan kamus *lexicon* gabungan.

Perhitungan total skor sentimen positif dan negatif pada suatu teks dilakukan menggunakan rumus berikut:

$$S_{kategori}\ \Sigma_{i=1}^{n} Score_i$$

Diketahui:

$S_{kategori}$: Total skor sentimen kategori dalam teks.

T : Himpunan kata dalam teks yang dianalisis.

n : Jumlah kata dalam teks.

kategori$score_i$: Nilai sentimen positif dari kata ke-i.

Setelah dilakukannya perhitungan terhadap total skor sentimen positif dan negatif dari setiap teks, proses klasifikasi sentimen dilakukan dengan membandingkan kedua nilai tersebut. Jika nilai total skor sentimen positif lebih tinggi dibandingkan skor sentimen negatif, maka teks diklasifikasikan ke dalam kategori sentimen positif begitu sebaliknya, apabila skor negatif memiliki nilai yang lebih dominan, maka teks akan dikategorikan sentimen negatif.

Sementara itu, apabila perbedaan antara skor positif dan negatif relatif kecil atau keduanya memiliki nilai yang rendah (yakni berada di bawah ambang batas tertentu yang ditetapkan), maka teks tersebut diklasifikasikan sebagai sentimen netral. Sentimen netrall ini mengacu ada opini yang tidak menunjukkan kecenderungan

emosional yang signifikan terhadap suatu topik atau isu tertentu, oleh karena itu penting untuk diidentifikasi secara terpisah guna menjaga validasi hasil analisis.

2.13 Arsitektur Model IndoBERT Multitask

IndoBERT merupakan pengembangan dari BERT (*Bidirectional Encoder Representations from Transformers*) yang dilatih secara khusus untuk bahasa Indonesia. IndoBERT memungkinkan pemahaman konteks kata dalam kalimat secara dua arah, menjadikan cocok untuk tugas-tugas pemrosesan bahasa alami dalam bahasa Indonesia[32].

Dalam penelitian ini, IndoBERT digunakan dalam pendekatan multitugas (*multi-task learning*), yaitu metode pelatihan model yang memungkinkan penyelesaian dua tugas klasifikasi secara bersamaan dalam satu arsitektur. Dua tugas tersebut adalah klasifikasi sentimen dan klasifikasi topik kebijakan(PPN 12%, Efisiensi Anggaran, Danantara, Makan Bergizi Gratis).

Arsitektur model *multitask* ini mengadopsi konsep *shared encoder with task specific decoder* di mana *encoder* IndoBERT yang sama digunakan untuk kedua tugas, namun memiliki dua jalur keluaran (*head*) yang berbeda. Proses dimulai dari input berupa teks yang telah di tokenisasi. Tokenisasi menghasilkan tiga jenis input *input_ids, token_type_ids,* dan *attention_mask.* Ketiganya dimasukkan ke dalam *encoder* IndoBERT yang akan memproses dan menghasilkan representasi kontekstual dari seluruh input.

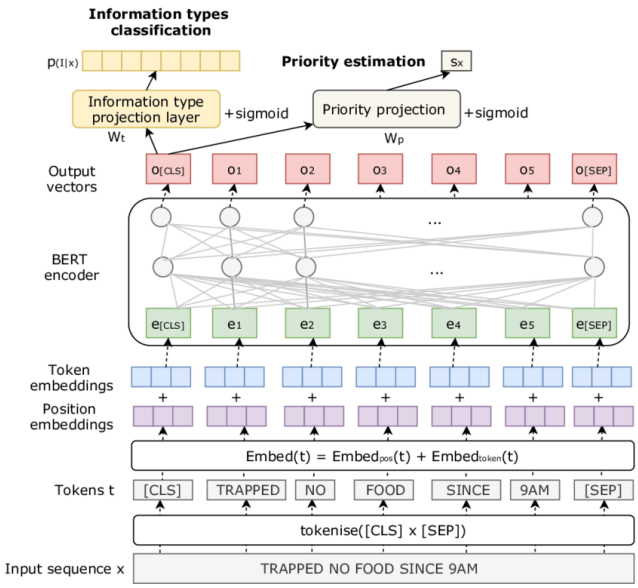
Hasil dari representasi token khusus [CLS] yang dirancang untuk merangkum informasi seluruh kalimat kemudian dilewatkan ke dalam layer *dropout* gun mencegah *overfitting* . Selanjutnya, representasi [CLS] diarahkan ke dua *dense layer* yang masing-masing berfungsi sebagai *classifier:*

1. *Classifier* pertama untuk tugas klasifikasi sentimen (negatif, positif, netral)
2. *Classifier* kedua untuk tugas klasifikasi topik kebijakan (PPN 12%, Efisiensi Anggaran, Danantara, dan Makan Bergizi Gratis)

Setiap *classifier* menggunakan fungsi aktivasi *softmax* untuk menghasilkan distribusi probabilitas atas masing-masing kelas target. Proses pelatihan dilakukan

dengan menghitung dua buah *loss function* menggunakan *CrossEntropyLoss* secara terpisah. kedua nilai *loss* kemudian dijumlahkan dalam pendekatan yang disebut *multitask loss aggregation,* yang memastikan bahwa model belajar kedua tugas secara bersamaan dengan bobot yang seimbang.

Struktur *arsitektur* ini divisualisasikan dalam Gambar 2.29 yang diadaptasi dari penelitian Wang et al. (2021). Gambar tersebut menunjukkan alur pemrosesan input mulai dari *embedding token*, pemrosesan oleh *encoder transformer,* hingga pembagian ke dalam dua output layer terpisah. *Head* pertama menghasilkan prefiksi klasifikasi informasi, sedangkan *head* kedua menghasilkan estimasi prioritas. Meskipun label tugas berbeda, pendekatan arsitektur dan alur pemrosesannya serupa dengan model yang digunakan dalam penelitian ini.



Gambar 2.29 Arsitektur Multitask Learning

Sumber: The overview architecture of transformer-based multi-task learning

Gambar 2.29, Arsitektur *multitask learning* berbasis transformer dengan *shared encoder* dan dua *task-specific projection layer.* Gambar menunjukkan alur representasi token [CLS] dari *encoder* BERT menuju dua output head berbeda untuk dua tugas klasifikasi . Model *multitask* ini merupakan jenis jaringan saraf tiruan (*neural network*) yang menggunakan *arsitektur transformer. Transformer* dirancang untuk mengolah data sekuensial seperti teks melalui mekanisme *self-attention,* yang memungkinkan model menangkap hubungan antar token secara

efisien. Dengan pembagian dua *output head,* model mampu memanfaatkan satu representasi dasar yang sama untuk menyelesaikan dua tugas berbeda secara simultan, meningkatkan efisiensi pelatihan serta akurasi dalam konteks klasifikasi teks bahasa Indonesia.

2.14 Hugging Face

Hugging face adalah perusahaan teknologi dan komunitas *open-source* yang menyediakan platform untuk mengakses, mengembahngkan, dan membagikan model-model AI berbasis transformer seperti BERT, GPT, dan T5 [33]. Fungsi utama *Hugging face* adalah menyediakan pustaka transformer siap pakai, API seragam untuk berbagai model *pre-trained,* serta alat untuk *fine-tuning*, *deployment,* dan distribusi model NLP dan *multimodal* AI [34]. *Hugging face* memungkinkan pengguna untuk mengakses, memodifikasi, dan menyebarluaskan model BERT melalui *transformers,* dan model hub, termasuk tokenisasi, *fine-tuning,* dan *inference. Hugging face* menjadi salah satu platform yang paling banyak digunakan dalam pengembangan model bahasa karena didukung oleh komunitas *open-source* yang aktif, dokumentasi yang komprehensif, serta kemudahan integrasi dengan berbagai *framework* populer PyTorch dan TensorFlow. Keunggulan-keunggulan ini menjadikannya sangat diminati oleh peneliti akademik, pelaku industri, hingga praktisi pendidikan dalam menerapkan dan mengembangkan model berbasis transformer [32].

2.15 Bidirectional Encoder Representations

Saat ini, model pra-pelatihan yang paling banyak digunakan oleh BERT. BERT tergolong salah satu algoritma yang cukup baru, yang diciptakan oleh Jacob Devlin pada tahun 2018. BERT merupakan metode *state-of-the art* dalam pembangunan *language* model dengan pendekatan *deep learning.* BERT menggunakan transformer, sebuah mekanisme attention yang mempelajari hubungan kontekstual antara kata-kata pada sebuah teks. Secara garis besar, transformer mencakup dua mekanisme terpisah, sebuah *encoder* yang berfungsi untuk membaca teks masukan dan sebuah *decoder* yang memproduksi prediksi untuk pekerjaan tertentu. Dikarenakan tujuan dari BERT adalah untuk menghasilkan sebuah *language* model, hanya mekanisme encoder-nya yang dianggap penting. Berlawanan dengan model satu arah dan *direction* yang hanya membaca teks masukan secara sekuensial (*left-*

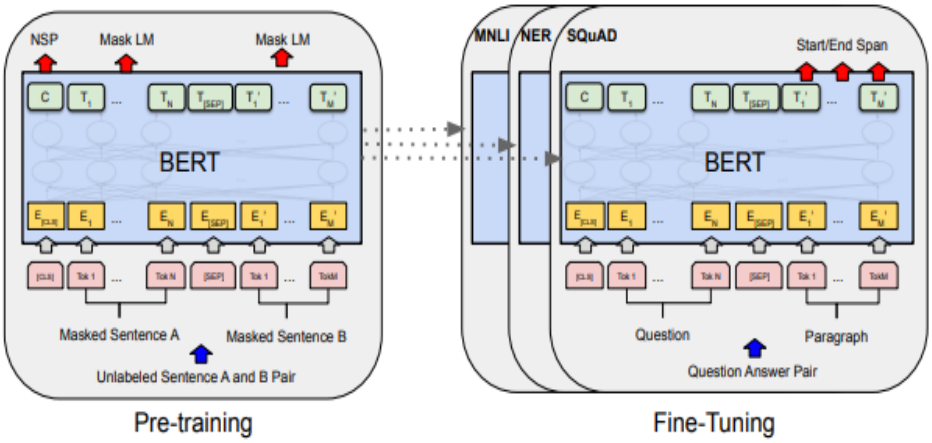
to-right atau *right-to-left),* encoder dari 3 transformer membaca seluruh urutan kata sekaligus. Karakteristik directional dari BERT memungkinkan model untuk mempelajari konteks dari sebuah kata berdasarkan lingkungan.

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model *deep learning* berbasis Transformer yang dikembangkan oleh Google pada tahun 2019. Model ini dirancang untuk memahami representasi kata dalam konteks dua arah secara lebih efektif dibandingkan metode sebelumnya, seperti LSTM. BERT juga mampu memahami bahasa dengan lebih baik dengan memproses kata secara paralel serta belajar dari kedua arah (kiri ke kanan dan kanan ke kiri) secara simultan [35].

Secara historis, bahasa sulit dipahami oleh mesin karena mesin tidak memiliki pemahaman konteks bahasa secara alami. BERT dirancang untuk membantu mesin memahami makna bahasa yang ambigu dalam teks dengan menghasilkan *word embeddings* yang menangkap informasi kontekstual.

Pada BERT terdapat dua tahapan penting dalam melatih model BERT (*Bidirectional Encoder Representations from Transformers)* yaitu tahapan Pra-Pelatihan *(Pre-Training)* dan Penyesuaian (*Fine-tuning).* Pada tahapan ini, tujuan utama adalah mengajarkan BERT pemahaman umum tentang bahasa dari sejumlah besar data teks yang tidak berlabel. Terdapat dua tugas utama yang dilakukan selama *Pre-Training* yaitu ada *Masked Language Model (Mask LM),* dimana pada MLM ini beberapa kata dalam sebuah kalimat secara acak "ditutupi" atau diganti dengan simbol khusus seperti [MASK]. Tugas BERT adalah untuk memprediksi kata-kata yang hilang ini berdasarkan konteks kata-kata lain dalam kalimat. Ini seperti permainan tebak kata, di mana BERT belajar untuk memahami hubungan antar kata dan bagaimana dapat saling melengkapi.

Pada *N*ext *Sentence Prediction,* pada BERT diberikan dua kalimat dan tugasnya adalah memprediksi apakah kalimat kedua merupakan kelanjutan logis dari kalimat pertama. Ini membantu BERT memahami hubungan antar kalimat dan alur wacana. Berikut gambar yang menggambarkan dua tahapan penting tersebut.



Gambar 2.30 Proses Pre-training dan Fine Tuning
Sumber: *Bidirectional Encoder Representations from Transformers (BERT)*

Dalam gambar, dapat melihat dua kalimat masukan ("*Masked Sentence A"* dan *"Masked Sentence B"*) yang di proses oleh arsitektur BERT. Dibagian atas, dapat melihat output untuk tugas Mask LM, dimana BERT mencoba memprediksi kata-kata yang telah di-masked. Untuk tugas NSP, BERT menghasilkan prediksi apakah kedua kalimat tersebut berurutan atau tidak.

Setelah BERT selesai pra-pelatihan dan memiliki pemahaman umum tentang bahasa, dapat menyesuaikannya untuk tugas-tugas spesifik yang ingin diselesaikan. Pada tahap penyesuaian, menggunakan data yang berlabel, yang berarti memiliki contoh masukan dan keluaran yang benar untuk tugas yang diinginkan.

Di sini, masukan BERT terdiri dari sebuah pertanyaan (*Question)* dan sebuah paragraf yang mungkin berisi jawaban atas pertanyaan tersebut. BERT kemudian dilatih untuk mengidentifikasi rentang awal (*Start Span)* dan rentang akhir *(End Span)* dalam paragraf yang merupakan jawaban dari pertanyaan tersebut yang bisa melihat output "*Start/End Span"* di bagian atas arsitektur BERT.

2.15.1 IndoBERT

IndoBERT adalah model BERT yang disesuaikan khusus untuk bahasa Indonesia, yang dirancang untuk memahami konteks bahasa Indonesia secara mendalam dengan pendekatan bidirectional dan tanpa pengawasan penuh. Model ini

digunakan dalam berbagai tugas pemrosesan bahasa alami, termasuk karakteristik bahasa Indonesia [36].

Proses pelatihan IndoBERT menggunakan dua tugas utama, yaitu *Masked Language Modeling(MLM)* dan *Next Sentence Prediction (NSP).* Dalam MLM, Beberapa kata sengaja disembunyikan sehingga model belajar memprediksi kata-kata yang hilang berdasarkan konteksnya. Sedangkan NSP melatih model untuk mengenali apakah dua kalimat berurutan memang saling terkait atau tidak. Kedua mekanisme ini membuat IndoBERT dapat memahami konteks dan hubungan antar kota dan kalimat dengan baik.

Dalam penelitian ini, dikembangkan sebuah model klasifikasi sentimen berbasis arsitektur IndoBERT, yang merupakan turunan dari model BERT(*Bidirectional Encoder Representation from Transformers)* yang telah diadaptasi khusus untuk bahasa Indonesia. Model yang dikembangkan bertujuan untuk mengklasifikasikan opini publik pada media sosial terhadap isu-isu nasional ke dalam kategori sentimen positif, negatif, atau netral.

Proses perancangan model dilakukan melalui dua tahap utama: pre-training data dan fine-tuning IndoBERT. Dataset digunakan merupakan hasil labeling manual dan semi otomatis menggunakan pendekatan *lexicon-based.* Kemudian, data latih tersebut digunakan untuk melatih ulang(*fine-tuning)*model IndoBERT agar mampu menggeneralisasi terhadap data sentimen dengan konteks yang spesifik, yaitu opini terhadap kebijakan nasional.

2.16 Hyperparameter Tuning

Hyperparameter tuning BERT dilakukan guna menemukan kombinasi *hyperparameter* yang optimal untuk meningkatkan kinerja model. Untuk perbandingan model sebelum dan sesudah *hyperparameter tuning,* dipilih metode yang sederhana dan paling umum digunakan yaitu *grid search, random search,* dan *Bayesian optimization. Grid search* dipilih karena mudah diimplementasikan dan cocok untuk semua jenis model, *random search* dipilih karena lebih efisien untuk ruang pencarian besar dan banyak *hyperparameter,* sedangkan *Bayesian*

optimization dipilih karena memberikan optimal global untuk fungsi kotak hitam, sehingga mengurangi evaluasi kesalahan validasi yang dilakukan [37].

Proses *Hyperparameter tuning* dilakukan dengan cara mencoba berbagai kombinasi nilai *hyperparameter* dan mengevaluasi performa model berdasarkan metrik tertentu, seperti akurasi, precision, recall atau loss pada data validasi. Dengan demikian, diperoleh konfigurasi *hyperparameter* yang dapat memaksimalkan kemampuan model dalam mengenali pola dan memprediksi data baru secara akurat.

2.17 Confusion Matrix

Confusion Matrix merupakan sebuah instrumen yang digunakan untuk mengevaluasi sistem klasifikasi yang menghasilkan tabel matriks untuk membandingkan hasil dari prediksi model dengan kelas sebenarnya dari data. Rumus *confusion matrix* ditunjukkan pada tabel berikut,

Tabel 2.2 Rumus *confusion matrix*

Kelas		**Kelas Prediksi**	
		Kelas 1	Kelas 2
Kelas Sebenarnya	Kelas 1	*True Positive(TP)*	*False Negative(FN)*
	Kelas 2	*False Positive(FP)*	*True Negative(TN)*

Keterangan :

1. *True Positive (TP)* = Kondisi data yang diprediksi positif dan nilai sebenarnya positif.
2. *False Positive (FP)* = Kondisi data yang diprediksi positif dan nilai sebenarnya negatif.
3. *False Negative (FN)* = Kondisi data yang diprediksi negatif dan nilai sebenarnya positif.
4. *True Negative (TN)* = Kondisi data yang diprediksi negatif dan nilai sebenarnya negatif.

Berikut merupakan rumus untuk menghitung *accuracy, precision, recall, f1-score.*

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (1)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

$$F1\text{-}Score = 2x \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

2.18 Review of Related Work

Penelitian-penelitian yang telah dilakukan terkait dengan analisis sentimen menggunakan model *Bidirectional Encoder Representations from Transformers* (BERT) dapat dilihat pada tabel berikut .

Tabel 2.3 *Review of Related Work*

No	Judul	Penulis	Tujuan	Algoritma	Hasil	Saran
1	Emotion and Sentiment Analysis of Tweets Using BERT	Andrea Chiorrini et al.	Menganalisis sentimen dan emosi dari data Twitter menggunakan BERT	BERT-Base (uncased & cased)	Akurasi 92% untuk sentimen, 90% untuk emosi	Menggunakan dataset lebih besar untuk meningkatkan performa model
2	A BERT Framework to Sentiment Analysis of Tweets	Bello et al.	Menggabungkan BERT dengan model deep learning lainnya untuk analisis sentimen	BERT + CNN, RNN, BiLSTM	Akurasi 93%, F1-score 95%	Mengeksplorasi model Transformer lain seperti RoBERTa untuk peningkatan lebih lanjut

No	Judul	Penulis	Tujuan	Algoritma	Hasil	Saran
3	Weibo Text Sentiment Analysis Based on BERT and Deep Learning	Hongchan Li et al.	Menganalisis sentimen teks Weibo selama pandemi COVID-19	BERT + BiLSTM + CNN + Attention Mechanism	Akurasi 92,7%	Menambahkan lebih banyak data dari berbagai periode untuk meningkatkan generalisasi model
4	BERT for Stock Market Sentiment Analysis	Matheus Gomes de Sousa et al.	Menganalisis sentimen berita keuangan untuk memprediksi pergerakan pasar saham	BERT	Akurasi 82,5%	Menggunakan dataset yang lebih luas dan menggabungkan data historis untuk meningkatkan akurasi prediksi pasar
5	An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian	Marco Pota et al.	Menganalisis sentimen tweet dalam bahasa Italia menggunakan BERT	BERT dengan teknik pra-pemrosesan	Akurasi 75%	Menggunakan pendekatan serupa untuk bahasa lain guna menguji efektivitas metode

BAB III

METODOLOGI

3.1 Analisis dan Justifikasi

Dalam sebuah penelitian, pemilihan topik merupakan salah satu aspek krusial yang menjadi fondasi dan menentukan kualitas dari hasil akhir, khususnya yang berfokus pada persepsi publik di platform media sosial. Untuk itu, penelitian ini tidak langsung menetapkan topik secara acak, melainkan memulainya dengan proses identifikasi dan analisis terhadap isu-isu kebijakan yang paling relevan dan ramai diperbincangkan oleh masyarakat di berbagai platform digital, terutama X dan YouTube. Proses seleksi ini divalidasi melalui analisis trend dan pencarian menggunakan *Google Trends*.

Justifikasi pemilihan topik ini didasarkan pada relevansi aktualnya di tengah masyarakat dan dinamika opini publik yang terbentuk di media sosial. *Google Trends* diharapkan dapat memberikan gambaran yang objektif dan aktual mengenai tingkat keterlibatan publik terhadap masing-masing topik, sekaligus memperkuat justifikasi pemilihannya dalam penelitian ini. Selain itu, setiap topik yang dipilih memiliki ciri khas dan sudut pandang yang berbeda, memberikan gambaran yang lebih luas dan beragam mengenai persepsi masyarakat.

Pendekatan analisis sentimen menggunakan model BERT (*Bidirectional Encoder Representations from Transformers)* dipilih karena kemampuan yang unggul dalam memahami konteks kalimat secara dua arah, sehingga lebih akurat dalam menangkap makna kata dalam konteks analisis sentimen. Berbagai penelitian telah menunjukkan bahwa BERT mampu mengungguli model klasik seperti SVM, *Naive Bayes*, bahkan LSTM dalam tugas-tugas NLP, termasuk klasifikasi teks dan analisis sentimen. Terlebih lagi, pengguna IndoBERT, yang secara khusus dilatih pada data berbahasa Indonesia, sangat efektif dalam memahami konteks lokal, termasuk penggunaan bahasa sehari-hari, *slang* serta struktur kalimat khas Indonesia, menjadikannya pilihan optimal untuk penelitian ini.

Untuk memahami prioritas kebijakan pemerintahan Presiden Prabowo Subianto secara lebih luas, peneliti terlebih dahulu menyusun daftar sepuluh kebijakan utama yang banyak diperbincangkan di ruang publik. Pemilihan kebijakan ini didasarkan pada data Google Trends, pemberitaan media daring, serta percakapan aktif di platform media sosial seperti X dan YouTube. Kesepuluh kebijakan ini dipilih karena memiliki intensitas diskursus publik yang tinggi dan berdampak langsung terhadap kehidupan masyarakat, baik dari sisi ekonomi, sosial, maupun keberlanjutan lingkungan. Berikut disajikan tabel pembanding yang merangkum tujuan utama serta permasalahan publik yang muncul dari masing-masing kebijakan tersebut:

Tabel 3.1 Dasar Permasalahan 10 Kebijakan

No	Kebijakan	Tujuan Utama	Permasalahan yang Muncul
1	Makan Bergizi Gratis	Meningkatkan gizi dan kesehatan anak-anak sekolah melalui penyediaan makan gratis	- Isu distribusi tidak merata - Kasus keracunan siswa - Kekhawatiran beban anggaran
2	Efisiensi Anggaran	Mengurangi pemborosan anggaran untuk efisiensi pemerintahan	- Pemotongan anggaran publik - Potensi penurunan layanan - Ketidaksiapan lembaga dalam efisiensi
3	Danantara	Optimalisasi dana kekayaan negara untuk mendorong ekonomi	- Potensi korupsi - Isu transparansi dan akuntabilitas - Rendahnya pemahaman publik
4	PPN 12%	Meningkatkan penerimaan negara untuk pembangunan dan pemulihan ekonomi	- Membebani masyarakat menengah ke bawah - Resistensi pelaku UMKM - Sosialisasi minim
5	Penghapusan Utang UMKM	Meringankan beban UMKM dan mendorong pertumbuhan bisnis kecil	- Kecemburuan sosial - Kekhawatiran moral hazard - Belum jelas kriteria UMKM yang dihapus hutangnya
6	Penurunan Harga Tiket Pesawat Domestik	Meningkatkan mobilitas masyarakat dan pariwisata domestik	- Realisasi tidak konsisten - Harga tetap tinggi saat musim ramai - Tidak semua rute terdampak
7	Bank Logam Mulia	Meningkatkan cadangan logam nasional dan kemudahan investasi logam mulia	- Belum jelas mekanisme dan regulasi - Kekhawatiran monopoli negara - Ketidaksiapan infrastruktur
8	Transisi Energi (Net Zero Emission 2050 dan B50)	Menuju energi bersih dan pengurangan emisi karbon	- Kesiapan industri rendah - Biaya implementasi tinggi - Ketimpangan antar wilayah
9	Peraturan Devisa Ekspor SDA	Menjaga devisa negara dan mendorong stabilitas ekonomi nasional	- Potensi menurunkan daya saing ekspor - Reaksi negatif dari pelaku ekspor besar

3.1.2 Perbandingan 10 Kebijakan Presiden Prabowo Berdasarkan *Google Trends*

Dalam rangka memahami sejauh mana perhatian publik terhadap kebijakan-kebijakan Presiden Prabowo Subianto, peneliti melakukan analisis menggunakan data dari Google *Trends*. Sebanyak sepuluh kebijakan dipilih berdasarkan intensitas pemberitaan, diskursus publik, serta potensi dampaknya terhadap masyarakat. Sepuluh kebijakan tersebut antara lain:

1. Program Makan Bergizi Gratis
2. Efisiensi Anggaran
3. Danantara
4. PPN 12%
5. Penghapusan Utang UMKM
6. Penurunan Harga Tiket Pesawat Domestik
7. Bank Logam Mulia
8. Transisi Energi (Net Zero Emission 2050 dan B50)
9. Peraturan Pemerintah tentang Devisa Ekspor SDA
10. Peningkatan Dana Desa dan BLT Desa

Analisis ini membandingkan popularitas masing-masing kebijakan dalam rentang waktu satu tahun terakhir. Data pencarian publik melalui Google menjadi indikator awal yang penting dalam mengukur tingkat perhatian masyarakat terhadap suatu isu, sekaligus mencerminkan dinamika opini yang berkembang di ruang digital. Melalui pendekatan ini, peneliti dapat mengidentifikasi kebijakan mana yang memperoleh *eksposur* tertinggi di kalangan masyarakat serta perubahan tren yang terjadi dari waktu ke waktu.

Hasil perbandingan ini diharapkan dapat memberikan konteks tambahan terhadap relevansi topik dalam studi analisis sentimen, sekaligus menjadi dasar dalam mengelompokkan dan menginterpretasikan respons publik secara lebih tajam pada tahap analisis lanjutan.

Visualisasi tren pencarian dari Google Trends pada dua gambar yang terpisah dibawah ini menggambarkan tingkat popularitas relatif dari sepuluh kebijakan

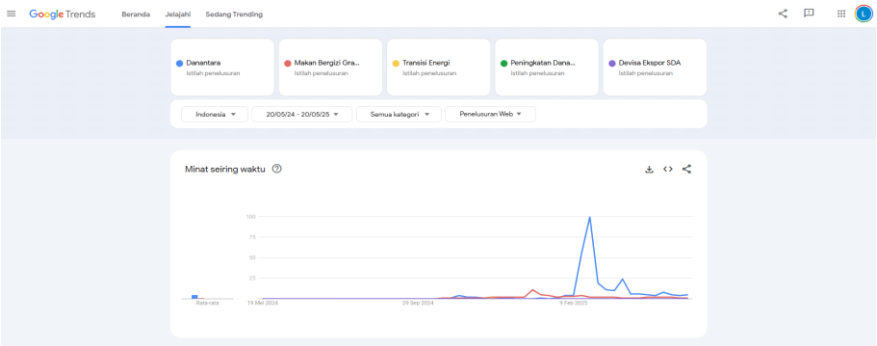
prioritas Presiden Prabowo Subianto di mata publik dalam kurun waktu satu tahun terakhir. Karena Google Trends hanya memungkinkan maksimal lima istilah perbandingan dalam satu waktu, maka analisis ini dilakukan dalam dua kelompok.

a. Gambar pertama menampilkan lima kebijakan:
   Danantara, Makan Bergizi Gratis, Transisi Energi, Peningkatan Dana Desa, dan Devisa Ekspor SDA.
   Dari grafik terlihat bahwa "Makan Bergizi Gratis" mendominasi pencarian dengan lonjakan tajam pada awal tahun 2025, menunjukkan perhatian masyarakat yang sangat tinggi terhadap program ini. Disusul oleh "Danantara", yang mulai menarik perhatian publik pada akhir Februari hingga Maret 2025. Tiga kebijakan lainnya menunjukkan tren pencarian yang lebih rendah namun tetap konsisten dalam kurun waktu tersebut.
b. Gambar kedua menampilkan kebijakan:
   Efisiensi Anggaran, PPN 12%, Bank Logam Mulia, Penurunan Harga Tiket, dan Penghapusan Utang UMKM.
   Dalam rafik ini, "PPN 12%" menunjukkan puncak pencarian yang signifikan, terutama saat wacana kebijakan ini mulai digaungkan pasca pelantikan. "Efisiensi Anggaran" dan "Penghapusan Utang UMKM" mengalami peningkatan minat dalam periode yang sama, meskipun dalam volume yang lebih moderat. Sementara itu, "Bank Logam Mulia" dan "Harga Tiket Pesawat" cenderung menunjukkan pola pencarian yang rendah.

Interpretasi awal dari data ini menunjukkan bahwa isu-isu yang berdampak langsung pada kehidupan masyarakat sehari-hari seperti kebutuhan pokok (makan bergizi), kenaikan pajak (PPN 12%), dan program strategis nasional seperti Danantara mendapatkan atensi yang jauh lebih besar. Sementara itu, isu teknis atau sektoral seperti devisa ekspor dan bank logam mulia cenderung kurang menarik perhatian publik secara luas di ranah digital. Analisis ini memperkuat pentingnya memahami persepsi publik sebagai bagian dari evaluasi dan pengembangan kebijakan. Temuan ini juga dapat menjadi acuan dalam menentukan fokus

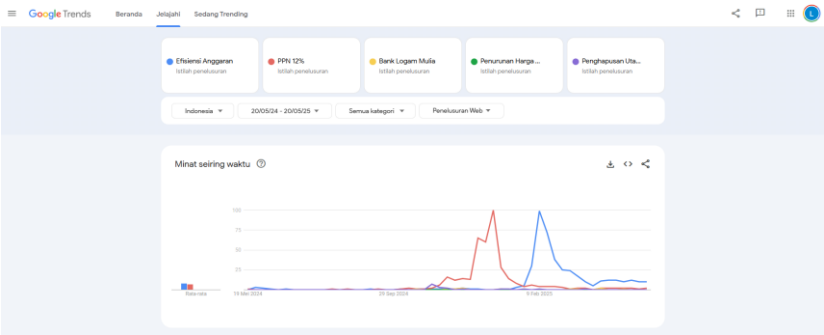
pendekatan analisis sentimen serta pemetaan opini masyarakat pada masing-masing kebijakan yang akan dibahas pada bab-bab berikutnya.

Pada Gambar 3.1 menunjukkan grafik lima kebijakan, yaitu Danantara, Makan Bergizi Gratis, Transisi Energi, Peningkatan Dana Desa, dan Devisa Ekspor SDA.



Gambar 3.1 Trend Google 5 Kebijakan (1)

Pada Gambar 3.2 menunjukkan grafik lima kebijakan, yaitu Efisiensi Anggaran, PPN 12%, Bank Logam Mulia, Penurunan Harga Tiket, dan Penghapusan Utang UMKM.



Gambar 3.2 *Trend* Google 5 Kebijakan (2)

Berdasarkan analisis dua grafik *Google Trends*, dapat disimpulkan bahwa empat kebijakan yang memiliki popularitas tertinggi secara keseluruhan adalah sebagai berikut:

1. Makan Bergizi Gratis

   Mencapai puncak tertinggi dalam grafik *Google Trends* (pada gambar pertama), dengan lonjakan pencarian yang signifikan pada awal tahun 2025. Hal ini mencerminkan perhatian publik yang sangat tinggi terhadap program

ini karena menyangkut kebutuhan dasar (makanan) dan menyasar langsung siswa sekolah di seluruh Indonesia.

2. PPN 12 %

   Menempati peringkat kedua dari segi volume pencarian (gambar kedua), terutama pada Maret 2025, saat kebijakan ini ramai dibahas di media sosial dan media daring. Topik ini sensitif karena terkait langsung dengan harga barang dan beban ekonomi masyarakat, sehingga memicu pertukaran pendapat yang luas.

3. Danantara

   Menunjukkan tren kenaikan tajam pada akhir Februari hingga Maret 2025, tepat setelah diumumkan secara resmi oleh pemerintah. Sebagai program strategis baru yang mengelola kekayaan negara (sovereign wealth fund), Danantara menarik perhatian karena kompleksitas dan potensi dampaknya terhadap ekonomi nasional.

4. Efisiensi Anggaran

   Meskipun tidak setinggi tiga kebijakan di atas, topik ini menempati urutan keempat tertinggi dalam grafik *Google Trends* (gambar kedua), terutama pasca pengumuman restrukturisasi belanja negara. Istilah ini banyak digunakan di media dan ruang diskusi kebijakan publik.

Keempat kebijakan ini dipilih sebagai fokus utama dalam penelitian karena puncak pencarian di *Google Trends* menunjukkan minat dan perhatian publik yang paling tinggi dibandingkan enam kebijakan lainnya. Selain itu, keempat kebijakan ini mewakili berbagai sektor utama, yaitu sektor sosial melalui program makan bergizi, sektor fiskal melalui kebijakan PPN, sektor ekonomi strategis melalui program Danantara, dan sektor tata kelola anggaran negara melalui kebijakan efisiensi anggaran. Volume pertukaran pendapat dan intensitas pemberitaan juga konsisten tinggi sepanjang Januari hingga Mei 2025. Respon masyarakat terhadap keempat kebijakan ini pun sangat beragam dan menunjukkan polaritas yang kuat, sehingga sangat ideal untuk dianalisis secara mendalam melalui pendekatan analisis sentimen.

Dengan dasar tersebut, peneliti menetapkan bahwa keempat kebijakan ini layak dianalisis secara lebih mendalam untuk memahami persepsi publik dan potensi keterlibatannya terhadap stabilitas serta penetapan kebijakan pemerintah baru di bawah Presiden Prabowo Subianto.

3.1.3 Program Makan Bergizi Gratis

Program makan bergizi gratis merupakan salah satu program prioritas yang diusung oleh Presiden terpilih Prabowo Subianto dalam visi-misi pemerintahannya. Program ini ditujukan untuk siswa sekolah dasar hingga menengah sebagai bagian dari upaya peningkatan kualitas sumber daya manusia, khususnya pada aspek gizi anak-anak. Selain itu, program ini juga diharapkan dapat memberikan dampak terhadap sektor ekonomi, pertanian lokal, dan ketahanan pangan nasional.

Topik ini dipilih karena menjadi salah satu isu yang paling banyak diperbincangkan oleh masyarakat setelah Pemilu 2024. Berdasarkan data dari *Google Trends*, pencarian dengan kata kunci *"makan bergizi gratis"* mengalami lonjakan signifikan sejak Januari 2025. Selain itu, topik ini juga menjadi trending di platform media sosial seperti X, dengan ribuan unggahan dan komentar yang menunjukkan berbagai respon dari masyarakat. Beberapa media nasional seperti Kompas, Detik, dan CNN Indonesia juga secara intensif memberitakan perkembangan isu ini, baik dari segi kesiapan anggaran, mekanisme pelaksanaan, maupun respon dari para ahli.

Program ini menimbulkan reaksi beragam di kalangan masyarakat. Sebagian mendukung penuh program ini karena dinilai berpihak pada rakyat kecil dan memberikan manfaat langsung kepada siswa dan keluarga kurang mampu. Namun, sebagian lainnya mempertanyakan efektivitas program, sumber pendanaan, dan potensi beban terhadap anggaran negara. Kondisi ini menunjukkan bahwa program makan bergizi gratis merupakan topik yang relevan untuk dianalisis guna melihat persepsi publik secara lebih mendalam.

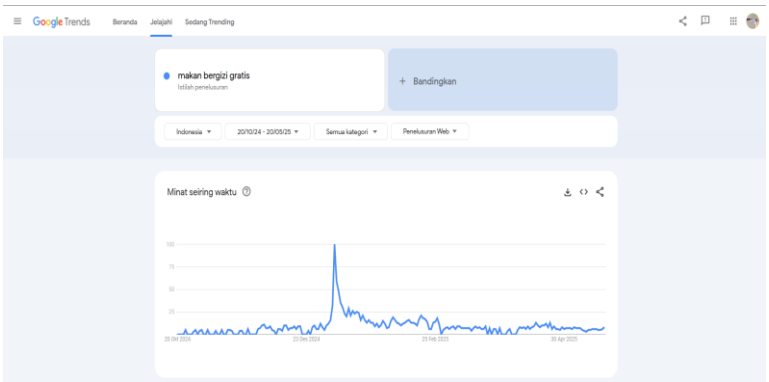
Banyaknya opini dan komentar dari masyarakat terhadap program ini menunjukkan bahwa topik ini sangat cocok untuk dianalisis menggunakan pendekatan analisis sentimen. Pendekatan ini dapat membantu mengelompokkan opini masyarakat ke

dalam kategori positif, netral, dan negatif, sehingga dapat diperoleh gambaran umum mengenai penerimaan atau penolakan publik terhadap program. Hasil analisis ini juga berpotensi memberikan *insight* bagi pembuat kebijakan dalam proses perumusan atau evaluasi program. Sub-bab berikut akan menyajikan data dan hasil pengamatan tersebut secara rinci topik makan bergizi gratis.

3.1.3.1 Data Google Trends Program Makan Bergizi Gratis

Gambar dibawah ini menunjukkan grafik tren pencarian Google dengan kata kunci "makan bergizi gratis" di Indonesia dalam rentang waktu 20 Oktober 2024 hingga 20 Mei 2025. Berdasarkan grafik tersebut, terlihat bahwa terdapat peningkatan yang sangat signifikan dalam volume pencarian pada akhir bulan Desember 2024. Puncak tertinggi terjadi tepat pada minggu keempat Desember.

Setelah mencapai puncaknya, tren pencarian mulai menurun namun tetap menunjukkan fluktuasi yang relatif stabil hingga Mei 2025. Pola ini mengindikasikan bahwa topik makan bergizi gratis sempat menjadi perhatian besar publik dan tetap relevan dalam perbincangan masyarakat selama beberapa bulan setelahnya. Temuan ini memperkuat validitas topik sebagai isu yang layak dianalisis lebih lanjut dalam penelitian ini, khususnya dalam konteks persepsi publik melalui media sosial.



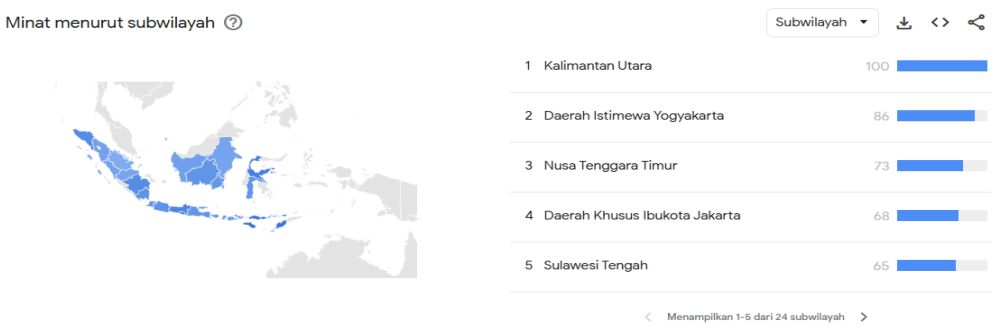
Gambar 3.3 *Trends* "Makan Bergizi Gratis"

Gambar dibawah ini menyajikan data distribusi minat pencarian terhadap kata kunci "makan bergizi gratis" berdasarkan wilayah geografis di Indonesia, seperti yang tercatat oleh *Google Trends* dalam periode 20 Oktober 2024 hingga 20 Mei 2025.

Hasilnya lima teratas menunjukkan bahwa Kalimantan Utara menjadi wilayah dengan tingkat minat tertinggi (skor 100), diikuti oleh Daerah Istimewa Yogyakarta (86), Nusa Tenggara Timur (73), DKI Jakarta (68), dan Sulawesi Tengah (65). Skor ini menunjukkan proporsi relatif dari pencarian di setiap wilayah, bukan jumlah absolut, dengan nilai 100 merepresentasikan wilayah dengan tingkat minat tertinggi.

Pola ini menggambarkan bahwa perhatian publik terhadap program makan bergizi gratis tersebar di berbagai wilayah Indonesia, tidak terbatas hanya di pusat pemerintahan seperti Jakarta. Dominasi Kalimantan Utara sebagai wilayah dengan minat tertinggi juga dapat mengindikasikan perhatian yang lebih besar dari daerah-daerah di luar pulau Jawa terhadap isu pemerataan akses makanan bergizi.

Temuan ini memperkuat argumentasi bahwa topik tersebut memiliki jangkauan dan daya tarik nasional, serta layak dianalisis lebih lanjut dalam konteks persepsi masyarakat secara luas.



Gambar 3.4 *Trend* "MBG" - 5 Sub Wilayah Teratas

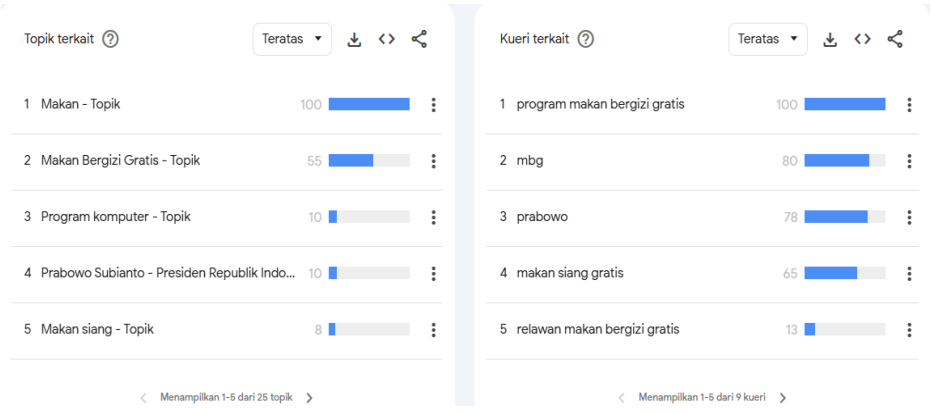
Gambar dibawah ini menampilkan topik-topik dan kueri pencarian yang paling sering dikaitkan dengan kata kunci *"*makan bergizi gratis*"* di *Google Trends* selama periode 20 Oktober 2024 hingga 20 Mei 2025. Data ini penting untuk memetakan konteks pencarian pengguna dan memahami keterkaitan isu secara lebih luas.

Topik yang paling banyak muncul adalah "Makan" dengan skor 100, menunjukkan tingginya minat masyarakat pada pencarian terkait konsumsi makanan secara umum. Di Posisi kedua ada "Makan Bergizi Gratis" dengan skor 55,

memperlihatkan perhatian khusus pada topik ini meskipun dalam cakupan sempit. Nama "Prabowo Subianto Presiden Republik Indonesia" juga muncul dalam daftar, dengan skor 10. Hal ini menunjukkan adanya keterkaitan antara topik makan bergizi gratis dengan figur Presiden Prabowo Subianto sebagai penggagasnya.

Sementara itu, kueri “program makan bergizi gratis” menjadi pencarian terbanyak dengan skor 100, menandakan bahwa masyarakat secara spesifik mencari informasi tentang program ini. Diikuti oleh istilah populer lain yang muncul adalah “mbg” dengan skor 80, singkatan informal dari makan bergizi, serta “prabowo” dengan skor 78, yang kembali menunjukkan kaitan kuat antara program dan tokoh politik tersebut. Kueri lain seperti “makan siang gratis” dan “relawan makan bergizi gratis” juga muncul, yang menunjukkan bahwa pencarian masyarakat mencakup pelaksanaan program dan keterlibatan aktor pelaksana di lapangan.

Temuan ini menunjukkan bahwa topik makan bergizi gratis bukan hanya berdiri sendiri, tetapi juga memiliki keterkaitan yang erat dengan aktor politik (Prabowo), bentuk kegiatan spesifik (makan siang), dan komunitas pendukung (relawan). Dengan demikian, data ini mendukung bahwa topik ini memiliki dimensi sosial-politik yang kompleks dan layak dianalisis dari sudut pandang persepsi publik.



Gambar 3.5 *Tren* “MBG” - Topik Teratas

3.1.4 Program PPN 12%

Program PPN 12% merupakan salah satu kebijakan fiskal yang dirancang oleh pemerintahan baru untuk meningkatkan penerimaan negara melalui pajak.

Kenaikan tarif Pajak Pertambahan Nilai (PPN) dari 11% menjadi 12% ini merupakan kelanjutan dari ketentuan Undang-Undang Harmonisasi Peraturan Perpajakan (UU HPP), yang telah dirancang sejak periode sebelumnya. Penerapan tarif baru ini dijadwalkan berlaku mulai 1 Januari 2025, dan menjadi bagian dari strategi jangka menengah dalam menjaga kestabilan APBN serta memperkuat kapasitas fiskal nasional.

Topik ini diangkat karena menjadi salah satu isu sentral yang menimbulkan perdebatan di masyarakat, baik dari kalangan pelaku usaha, akademisi, hingga masyarakat umum. Berdasarkan data *Google Trends*, pencarian kata kunci "PPN 12%" mengalami peningkatan tajam sejak diumumkan secara resmi dalam Rencana Kerja Pemerintah 2025. Selain itu, diskusi seputar kenaikan PPN juga menjadi perbincangan hangat di media sosial seperti X dan YouTube .

Reaksi publik terhadap kebijakan ini cukup beragam. Sebagian kalangan menilai bahwa kebijakan ini diperlukan untuk memperkuat ketahanan fiskal negara dan membiayai program prioritas nasional. Namun, tidak sedikit pula yang mengkhawatirkan dampak langsungnya terhadap daya beli masyarakat dan potensi kenaikan harga barang konsumsi. Pelaku usaha kecil dan menengah (UMKM) juga menyuarakan kekhawatiran bahwa kebijakan ini dapat menurunkan minat konsumsi masyarakat.

Kebijakan ini menjadi penting untuk dianalisis karena menunjukkan adanya polarisasi opini publik terhadap keputusan ekonomi negara. Analisis sentimen terhadap isu PPN 12% dapat memberikan gambaran yang lebih jelas mengenai penerimaan atau resistensi masyarakat terhadap kebijakan tersebut. Melalui pendekatan ini, opini publik dapat dipetakan ke dalam kategori positif, netral, dan negatif yang selanjutnya dapat menjadi acuan dalam penyusunan strategi komunikasi kebijakan publik dan mitigasi dampak sosial ekonomi.Sub-bab berikut akan memaparkan data dan pengamatan mengenai persepsi masyarakat terhadap kebijakan kenaikan PPN menjadi 12%.

3.1.4.1 Data *Google Trends* Program PPN 12%

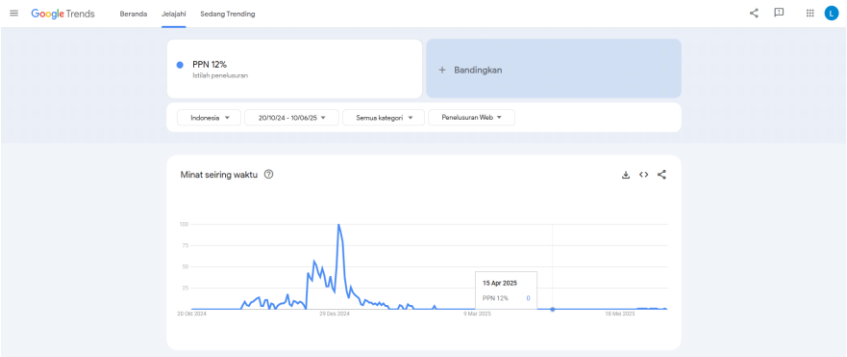
Untuk mengamati dinamika perhatian publik terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%, penelusuran data melalui platform *Google Trends* dilakukan dengan kata kunci "PPN 12%". Data yang ditampilkan mencakup kurun waktu Oktober 2024 hingga Januari 2025.

Berdasarkan grafik yang ditampilkan di bawah ini, terlihat adanya lonjakan signifikan dalam volume pencarian pada minggu pertama Januari 2025, bertepatan dengan banyaknya pemberitaan terkait rencana implementasi kenaikan tarif PPN yang akan berlaku mulai 1 Januari 2025, sesuai amanat Undang-Undang Harmonisasi Peraturan Perpajakan (UU HPP). Puncak perhatian masyarakat ditunjukkan dengan nilai pencarian mendekati angka 100 pada pertengahan Januari 2025, yang merupakan nilai tertinggi dalam skala *Google Trends*.

Setelah mencapai puncaknya, tren pencarian mulai mengalami penurunan secara bertahap pada Februari hingga Maret 2025, dan relatif stabil hingga Mei 2025. Penurunan ini dapat diinterpretasikan sebagai penurunan urgensi atau perhatian publik terhadap topik tersebut, seiring dengan meredanya pemberitaan dan masyarakat mulai beradaptasi terhadap kebijakan baru tersebut.

Data ini menunjukkan bahwa kebijakan PPN 12% merupakan isu yang sempat mendominasi perhatian publik dalam waktu singkat, dan menjadi perbincangan hangat di berbagai media sosial serta platform pemberitaan digital. Lonjakan ini juga mengindikasikan bahwa publik memiliki kekhawatiran atau keingintahuan tinggi terhadap dampak langsung dari kebijakan ini terhadap harga barang dan jasa, terutama dalam konteks ekonomi pasca pandemi.

Oleh karena itu, data dari *Google Trends* ini menjadi indikator awal untuk mengukur sentimen dan persepsi publik terhadap isu kebijakan fiskal seperti PPN 12%, dan dapat digunakan sebagai salah satu acuan untuk melakukan analisis lanjutan melalui pendekatan analisis sentimen berbasis teks.



Gambar 3.6 *Trends* "PPN 12%"

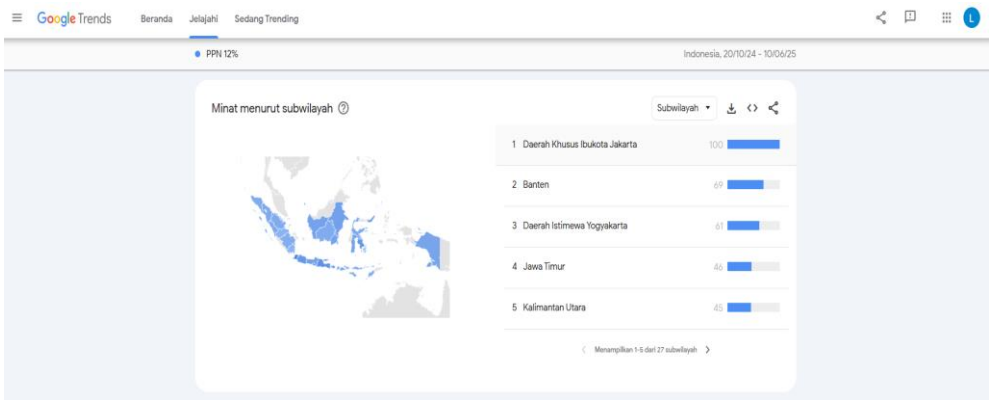
Selain tren waktu, *Google Trends* juga menunjukkan persebaran minat pencarian berdasarkan wilayah geografis. Berdasarkan data yang diambil dari periode 20 Oktober 2024 hingga 10 Juni 2025, terlihat bahwa minat terhadap topik "PPN 12%" paling tinggi berasal dari wilayah-wilayah dengan tingkat aktivitas ekonomi yang relatif padat dan tingkat akses internet yang tinggi.

Adapun lima besar wilayah dengan tingkat pencarian tertinggi adalah sebagai berikut:

1. Daerah Khusus Ibukota Jakarta menempati urutan pertama dengan skor indeks pencarian 100, yang menunjukkan minat paling tinggi secara relatif dibandingkan wilayah lain. Hal ini dapat dikaitkan dengan posisi Jakarta sebagai pusat pemerintahan dan ekonomi nasional, dimana kebijakan fiskal seperti kenaikan PPN langsung berdampak pada sektor perdagangan dan jasa.
2. Provinsi Banten berada di urutan kedua dengan nilai indeks 69. Banten merupakan wilayah penyangga DKI Jakarta dengan kawasan industri besar seperti di Tangerang dan Cilegon, sehingga respons masyarakat terhadap kebijakan PPN juga cukup tinggi.
3. Daerah Istimewa Yogyakarta menempati posisi ketiga dengan skor 61. Meskipun bukan pusat industri, tingginya minat ini bisa dikaitkan dengan kesadaran masyarakat yang tinggi terhadap isu-isu ekonomi nasional, serta aktivitas mahasiswa dan akademisi yang aktif dalam diskursus kebijakan publik.

4. Jawa Timur berada di posisi keempat dengan skor 46, mencerminkan tingginya perhatian masyarakat di wilayah ini yang memiliki pusat perdagangan besar seperti Surabaya dan Sidoarjo.
5. Kalimantan Utara berada di posisi kelima dengan skor 45, yang menunjukkan adanya perhatian cukup tinggi terhadap isu kenaikan PPN, meskipun secara populasi lebih kecil dibanding provinsi lain. Hal ini mungkin terkait dengan aktivitas ekspor-impor atau investasi yang berkembang di wilayah tersebut.

Grafik dibawah ini menunjukkan bahwa isu kenaikan PPN 12% tidak hanya menjadi perhatian di pusat pemerintahan, tetapi juga meresap ke berbagai daerah yang memiliki kepentingan ekonomi atau tingkat literasi publik yang tinggi. Distribusi minat pencarian ini penting sebagai indikator awal mengenai wilayah mana saja yang memiliki kesadaran atau kekhawatiran lebih besar terhadap perubahan kebijakan fiskal nasional.

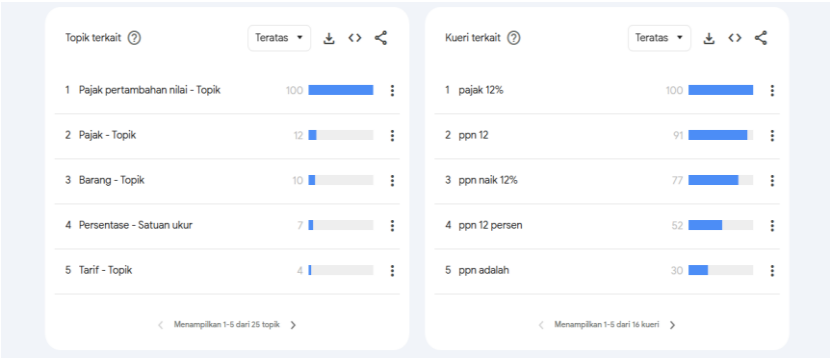


Gambar 3.7 *Trend* "*PPN 12%*" - 5 Wilayah Teratas

Berdasarkan grafik *Google Trends* dibawah ini, terdapat sejumlah topik dan kueri pencarian yang paling sering dikaitkan dengan istilah "PPN 12%" oleh masyarakat Indonesia. Pada kategori topik terkait, istilah *"*Pajak Pertambahan Nilai*"* muncul sebagai topik paling dominan dengan skor indeks maksimum, yaitu 100. Ini menunjukkan bahwa pembahasan mengenai PPN 12% secara langsung berhubungan erat dengan pemahaman masyarakat tentang jenis pajak tersebut secara umum. Diikuti oleh topik *"*Pajak*"*, *"*Barang*"*, serta istilah teknis seperti *"Persentase"* dan *"*Tarif*"*, menunjukkan bahwa pencarian seputar PPN tidak hanya

fokus pada kebijakan itu sendiri, tetapi juga pada aspek mekanisme dan teknis penerapannya.

Sementara itu, pada kategori kueri terkait, masyarakat paling banyak mencari dengan frasa langsung seperti *"pajak 12%"*, *"ppn 12"*, dan *"ppn naik 12%"*. Ini mencerminkan kepanikan atau kebutuhan masyarakat untuk segera memahami perubahan tarif yang akan diterapkan. Beberapa kueri lainnya, seperti *"ppn 12 persen"* dan *"ppn adalah"*, juga menunjukkan bahwa sebagian masyarakat mungkin masih mencari pengertian dasar dari PPN, termasuk bagaimana kenaikan ini mempengaruhi harga barang dan jasa sehari-hari. Keseluruhan data ini mengindikasikan bahwa topik PPN 12% menjadi perhatian luas baik dari sisi praktis maupun konseptual, serta mendorong masyarakat untuk mencari klarifikasi melalui mesin pencari terkait pengaruhnya terhadap kehidupan ekonomi sehari-hari.



Gambar 3.8 *Trends* "PPN 12%" - Topik Teratas

3.1.5 Program Danantara

Program danantara merupakan salah satu kebijakan strategis pemerintah dalam mendukung percepatan transformasi digital dan inklusi ekonomi di berbagai wilayah Indonesia, khususnya di daerah-daerah yang masih tertinggal secara infrastruktur teknologi dan keuangan. Program ini bertujuan untuk menjembatani kesenjangan akses terhadap layanan keuangan digital dengan cara memperkuat kolaborasi antara pemerintah, sektor swasta, serta masyarakat lokal.

Danantara berperan sebagai ekosistem digital yang mendorong penggunaan teknologi finansial untuk memperluas akses pembiayaan, meningkatkan literasi keuangan, serta memberdayakan pelaku usaha mikro, kecil, dan menengah

(UMKM) di daerah. Topik ini menjadi penting karena Danantara dianggap sebagai fondasi dari berbagai kebijakan digital keuangan negara di masa depan, termasuk sistem anggaran berbasis digital, distribusi dana desa, bantuan sosial, serta pengawasan belanja pemerintah. Sejak diperkenalkan secara resmi, inisiatif ini mendapat perhatian besar dari publik, terutama di kalangan birokrat, pelaku ekonomi digital, dan akademisi yang tertarik pada tata kelola keuangan yang adaptif terhadap era digital.

Berdasarkan data pencarian daring dan pemantauan media sosial, istilah "Danantara" mulai mengalami peningkatan popularitas sejak diumumkan dalam agenda prioritas nasional tahun 2025 dan diperkuat melalui pidato-pidato pejabat tinggi negara. Diskusi terkait program ini juga muncul di berbagai kanal media sosial seperti X, YouTube, dan forum-forum diskusi publik daring. Publik banyak menyoroti potensi manfaat Danantara dalam menciptakan efisiensi anggaran, mengurangi kebocoran dana, serta memperluas jangkauan layanan fiskal digital kepada masyarakat.

Respons publik terhadap program ini cukup positif, khususnya dari kalangan yang melihat pentingnya digitalisasi tata kelola pemerintahan. Namun, beberapa kritik juga muncul, terutama mengenai kesiapan infrastruktur digital di daerah terpencil, kekhawatiran terhadap keamanan data, dan potensi tumpang tindih dengan sistem yang sudah ada sebelumnya.

Analisis terhadap persepsi publik mengenai Danantara menjadi penting untuk mengidentifikasi sejauh mana kebijakan ini diterima oleh masyarakat, serta untuk memahami tantangan-tantangan yang harus diantisipasi dalam implementasinya. Sentimen publik terhadap program ini mencerminkan harapan sekaligus kekhawatiran atas arah transformasi digital sektor publik di Indonesia.

Sub-bab berikut akan menyajikan data dan pengamatan terkait opini masyarakat terhadap program Danantara yang diperoleh melalui berbagai kanal daring dan media sosial sebagai bagian dari pendekatan analisis sentimen.

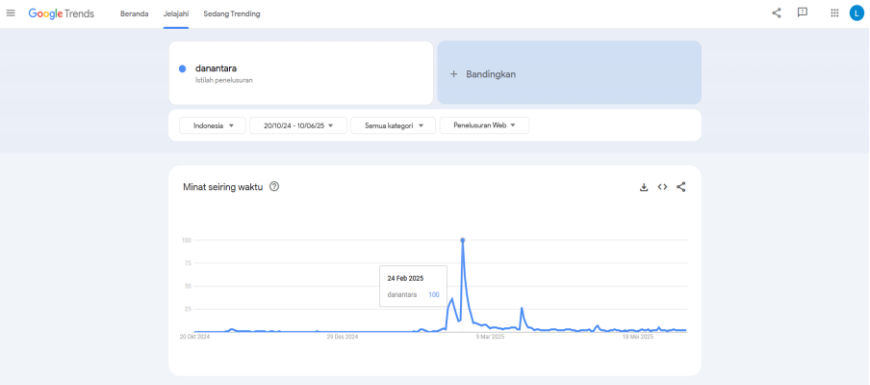
3.1.5.1 Data Google *Trends* Program Danantara

Untuk memahami seberapa besar perhatian publik terhadap Program Danantara, peneliti memanfaatkan data dari *Google Trends* sebagai salah satu indikator awal popularitas topik di ranah digital. *Google Trends* memberikan gambaran mengenai intensitas pencarian kata kunci "Danantara" di Indonesia selama periode waktu tertentu. Grafik pencarian ini mencerminkan minat masyarakat yang meningkat secara signifikan pada titik-titik waktu tertentu, khususnya saat program ini diumumkan secara resmi atau menjadi sorotan media nasional.

Berdasarkan data *Google Trends* yang ditampilkan di atas, puncak pencarian terhadap kata kunci "Danantara" terjadi pada tanggal 24 Februari 2025, yang bertepatan dengan momen pengumuman resmi dari pemerintah mengenai inisiasi program ini sebagai bagian dari agenda transformasi digital keuangan negara. Setelah puncak tersebut, frekuensi pencarian mengalami fluktuasi namun tetap menunjukkan adanya minat yang berkelanjutan hingga beberapa minggu berikutnya.

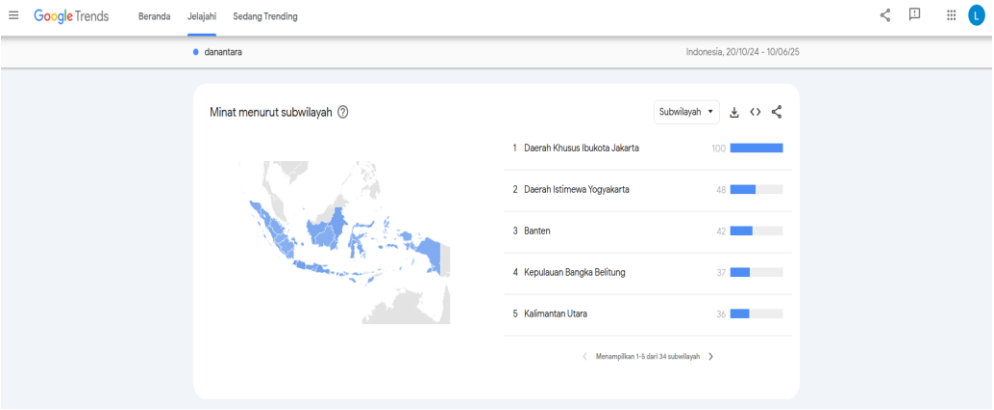
Tren ini menandakan bahwa peluncuran Program Danantara berhasil menarik perhatian masyarakat luas, baik dari kalangan pemerhati kebijakan, media, hingga masyarakat umum. Informasi ini menjadi dasar yang penting dalam mengkaji persepsi awal publik terhadap program tersebut sebelum dilakukan analisis lebih lanjut melalui media sosial dan kanal daring lainnya.

Dengan demikian, data *Google Trends* ini berfungsi sebagai bukti kuantitatif mengenai popularitas dan daya tarik awal Program Danantara di tengah masyarakat digital Indonesia.



Gambar 3.9*Trends* "Danantara"

Gambar di bawah ini menunjukkan peta intensitas pencarian kata kunci "Danantara" berdasarkan wilayah di Indonesia, sebagaimana direkam oleh Google Trends selama periode Oktober 2024 hingga Juni 2025. Berdasarkan visualisasi tersebut, terlihat bahwa minat tertinggi terhadap Program Danantara berasal dari wilayah Daerah Khusus Ibukota Jakarta, disusul oleh Daerah Istimewa Yogyakarta, Banten, Kepulauan Bangka Belitung, dan Kalimantan Utara. Hal ini menunjukkan bahwa perhatian publik terhadap program ini lebih terkonsentrasi di wilayah-wilayah yang memiliki akses informasi digital yang tinggi dan infrastruktur komunikasi yang lebih mapan. DKI Jakarta sebagai pusat pemerintahan dan ekonomi nasional menjadi wilayah dengan intensitas tertinggi, yang bisa jadi disebabkan oleh keterlibatan langsung instansi pemerintah dan media dalam penyebaran informasi terkait kebijakan ini. Temuan ini memperkuat asumsi bahwa persebaran informasi dan tingkat literasi digital mempengaruhi tingkat ketertarikan masyarakat terhadap kebijakan publik seperti Program Danantara. Data ini juga dapat dijadikan sebagai acuan dalam merancang strategi komunikasi publik yang lebih efektif dan inklusif, khususnya di daerah yang menunjukkan intensitas minat yang lebih rendah.



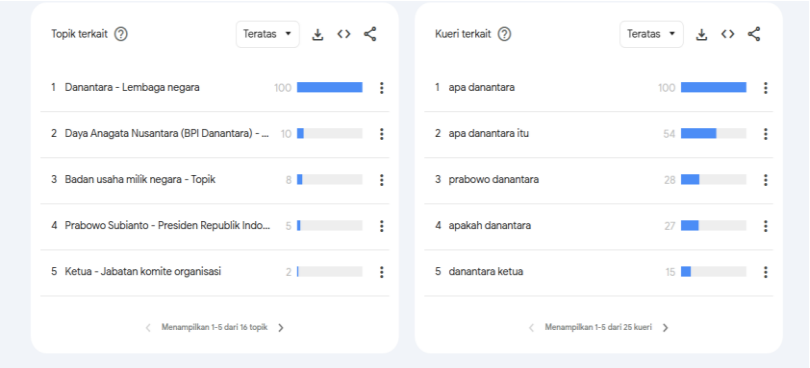
Gambar 3.10 *Trends* "Danantara" - 5 Sub Wilayah Teratas

Gambar dibawah ini menampilkan data topik dan kueri terkait terpopuler yang sering dicari masyarakat mengenai Program Danantara berdasarkan hasil pencarian Google Trends. Dalam kategori topik terkait, istilah "Danantara – Lembaga negara" menempati peringkat tertinggi, diikuti oleh "Daya Anagata Nusantara (BPI Danantara)", "Badan usaha milik negara", "Prabowo Subianto – Presiden Republik Indonesia", dan "Ketua – Jabatan komite organisasi". Hal ini mencerminkan bahwa pemikiran publik mengenai Danantara tidak hanya berfokus pada programnya semata, namun juga melibatkan institusi yang menaunginya, tokoh-tokoh terkait, serta struktur organisasi yang mungkin terlibat dalam kebijakan ini. Keterkaitan dengan nama Presiden Prabowo Subianto menunjukkan adanya persepsi bahwa kebijakan ini merupakan bagian dari agenda besar pemerintahan baru.

Di sisi lain, kueri terkait memperlihatkan tingginya rasa ingin tahu masyarakat terhadap definisi dan latar belakang program ini. Pencarian seperti "apa danantara", "apa danantara itu", "prabowo danantara", "apakah danantara", hingga "danantara ketua" mengindikasikan bahwa banyak masyarakat yang masih belum memahami secara jelas apa itu Danantara, siapa yang terlibat, dan bagaimana struktur kepemimpinannya. Ini menunjukkan adanya kesenjangan informasi publik terhadap program yang sedang ramai diperbincangkan tersebut.

Data ini mempertegas pentingnya penyampaian informasi yang jelas, terbuka, dan masif oleh pihak pemerintah atau lembaga terkait untuk menjawab pertanyaan-pertanyaan dasar masyarakat. Minimnya pemahaman publik bisa berdampak pada munculnya spekulasi dan misinformasi. Oleh karena itu, data kueri dan topik ini

sangat bermanfaat sebagai landasan dalam menyusun strategi komunikasi publik agar lebih tepat sasaran dan efektif.



Gambar 3.11 Trends "Danantara" - Topik Teratas

3.1.6 Program Efisiensi Anggaran

Program Efisiensi Anggaran merupakan salah satu inisiatif strategis pemerintah dalam merespons tuntutan reformasi birokrasi dan pengelolaan keuangan negara yang lebih optimal. Kebijakan ini menekankan pada pengurangan belanja yang tidak produktif, pengetatan alokasi anggaran pada pos-pos yang kurang berdampak langsung terhadap masyarakat, serta mendorong tata kelola fiskal yang transparan dan akuntabel. Tujuan utama dari program ini adalah untuk memperkuat kesehatan fiskal negara sekaligus memastikan bahwa setiap rupiah yang dibelanjakan memberikan manfaat maksimal bagi pembangunan nasional.

Topik ini dipilih karena menjadi salah satu kebijakan yang cukup menyita perhatian publik, terutama di tengah berbagai sorotan terhadap efisiensi penggunaan APBN. Berdasarkan penelusuran melalui Google Trends, pencarian terhadap kata kunci "efisiensi anggaran" meningkat secara signifikan pada periode awal tahun 2025, bertepatan dengan berbagai pernyataan resmi dari pemerintah terkait pemangkasan anggaran di sejumlah kementerian dan lembaga. Di samping itu, media sosial seperti X dan YouTube turut menjadi ruang diskusi aktif mengenai program ini, dengan munculnya beragam opini, kritik, dan dukungan dari masyarakat maupun pakar kebijakan publik.

Reaksi publik terhadap program efisiensi anggaran sangat beragam. Sebagian masyarakat menilai bahwa langkah ini merupakan keputusan tepat dalam

menghindari pemborosan dan meningkatkan efektivitas pembangunan nasional. Namun, tidak sedikit pula yang khawatir bahwa kebijakan ini justru berpotensi menghambat program sosial atau pelayanan publik yang penting, terutama jika dilakukan tanpa kajian mendalam dan mekanisme pengawasan yang memadai.

Dengan mencuatnya diskursus di berbagai kanal digital, program efisiensi anggaran menjadi topik yang sangat relevan untuk dianalisis melalui pendekatan analisis sentimen. Melalui pendekatan ini, opini publik dapat dikategorikan ke dalam sentimen positif, netral, maupun negatif, yang kemudian dapat dijadikan bahan evaluasi terhadap kebijakan yang diambil oleh pemerintah. Selain itu, hasil analisis ini juga diharapkan mampu memberikan masukan kepada pemangku kepentingan terkait strategi komunikasi publik serta pengelolaan risiko sosial dalam pelaksanaan program.

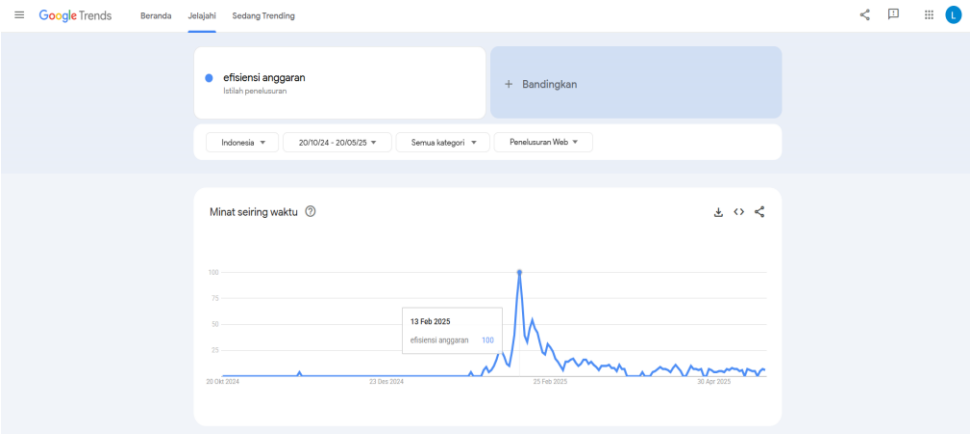
Sub-bab berikut akan menyajikan data-data dan pengamatan yang berkaitan dengan persepsi publik terhadap Program Efisiensi Anggaran, yang dikumpulkan melalui Google Trends, media sosial X, YouTube, serta pemberitaan media daring nasional dan internasional.

3.1.6.1 Data *Google Trends* Program Efisiensi Anggaran

Berdasarkan data dari *Google Trends*, topik "efisiensi anggaran" menunjukkan lonjakan pencarian yang signifikan pada pertengahan Februari 2025, khususnya pada tanggal 13 Februari 2025 yang menjadi puncak tertinggi minat pencarian masyarakat terhadap topik ini. Lonjakan ini berkaitan dengan pernyataan resmi pemerintah atau pengumuman kebijakan pemangkasan anggaran yang ramai diberitakan dan menjadi sorotan publik pada periode tersebut.

Setelah mencapai puncaknya, minat pencarian terhadap kata kunci ini mengalami tren penurunan secara bertahap, meskipun masih menunjukkan fluktuasi kecil hingga bulan Mei 2025. Tren ini menunjukkan bahwa wacana efisiensi anggaran cukup menyita perhatian publik dalam waktu tertentu dan menjadi topik yang banyak dibicarakan, terutama dalam konteks reformasi kebijakan fiskal nasional yang tengah dilakukan oleh pemerintah baru.

Data ini menjadi indikator penting bahwa kebijakan efisiensi anggaran berhasil membangkitkan diskursus di tengah masyarakat, sekaligus mencerminkan adanya kebutuhan publik untuk memahami lebih jauh dampak dan pelaksanaan dari program tersebut. Oleh karena itu, analisis sentimen terhadap persepsi publik pada momen tren pencarian ini menjadi penting untuk menggambarkan bagaimana kebijakan ini diterima, dipahami, dan dipersepsikan oleh berbagai kalangan.

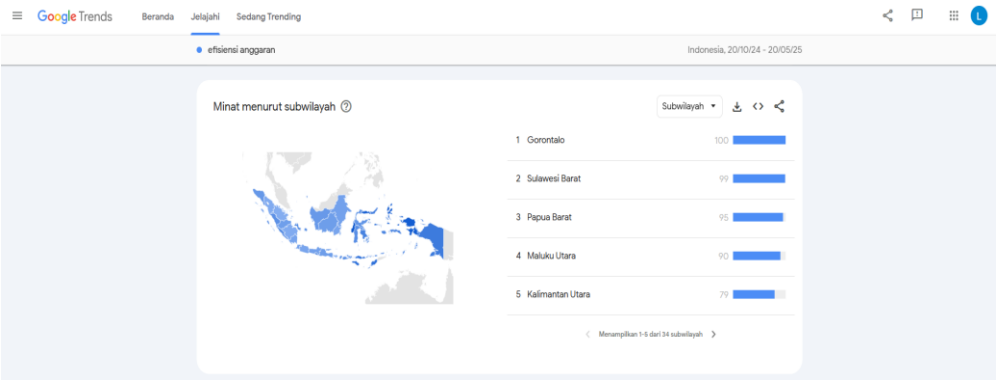


Gambar 3.12 *Trends* "Efisiensi Anggaran"

Selain menunjukkan lonjakan minat pencarian secara nasional, data Google Trends juga memetakan sebaran minat berdasarkan wilayah di Indonesia terhadap kata kunci "efisiensi anggaran". Berdasarkan grafik sebaran ini, provinsi Gorontalo tercatat memiliki tingkat minat tertinggi dengan skor pencarian 100, diikuti oleh Sulawesi Barat dan Papua Barat dengan skor yang hampir setara, masing-masing menunjukkan antusiasme regional yang cukup signifikan. Kemudian disusul oleh Maluku Utara dan Kalimantan Utara, yang juga menunjukkan ketertarikan terhadap topik ini.

Distribusi geografis ini mengindikasikan bahwa isu efisiensi anggaran tidak hanya menjadi perhatian di wilayah-wilayah pusat pemerintahan seperti Jawa, tetapi juga di daerah-daerah luar yang kemungkinan besar memiliki perhatian lebih terhadap kebijakan alokasi dan distribusi anggaran nasional. Besarnya minat dari wilayah-wilayah ini dapat mencerminkan harapan terhadap pemerataan kebijakan fisikal dan efisiensi pengeluaran negara yang berdampak langsung pada pembangunan daerah.

Temuan ini penting sebagai data pendukung dalam memahami konteks persebaran opini publik, sekaligus membantu mengidentifikasi wilayah-wilayah yang mungkin lebih sensitif terhadap kebijakan pemotongan atau penghematan anggaran. Selanjutnya, analisis lebih dalam dapat dilakukan untuk memahami persepsi masyarakat dari daerah-daerah ini terhadap program efisiensi anggaran yang diluncurkan oleh pemerintah pusat.

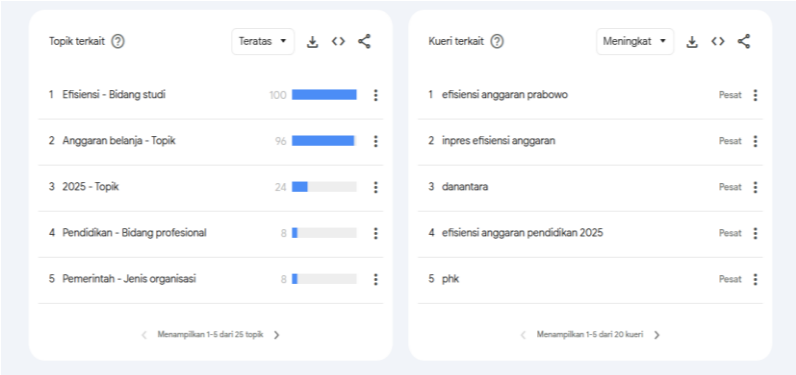


Gambar 3.13 *Trends* "Efisiensi Anggaran" - 5 Sub Wilayah Teratas

Berdasarkan data pencarian topik dan kueri terkait dalam Google Trends, terlihat bahwa istilah "efisiensi" mendominasi sebagai topik teratas dengan skor pencarian maksimum (100), diikuti oleh istilah lain yang berkaitan erat dengan konteks kebijakan anggaran seperti "anggaran belanja", "2025", serta "pendidikan" dan "pemerintah". Ini menandakan bahwa masyarakat tidak hanya menaruh perhatian terhadap konsep efisiensi itu sendiri, tetapi juga mengaitkannya secara langsung dengan konteks fiskal, pemerintahan, dan agenda kebijakan tahun 2025.

Sementara itu, dalam bagian *kueri terkait*, sejumlah frasa pencarian yang mengalami peningkatan signifikan antara lain adalah "efisiensi anggaran prabowo", "proses efisiensi anggaran", dan "efisiensi anggaran pendidikan 2025". Kueri seperti ini menunjukkan bahwa masyarakat secara aktif mencari informasi yang lebih spesifik mengenai rencana atau implementasi efisiensi anggaran yang dilakukan oleh pemerintahan Presiden Prabowo, baik secara umum maupun dalam konteks sektoral seperti pendidikan. Kueri "danantara" juga muncul sebagai bagian dari pencarian yang berkaitan, menandakan adanya keterkaitan persepsi publik antara lembaga pengelola investasi negara dan program efisiensi anggaran.

Data ini mencerminkan ketertarikan publik yang tinggi terhadap bagaimana kebijakan efisiensi anggaran akan berdampak pada sektor-sektor utama, serta menunjukkan bahwa program ini telah menjadi bagian dari diskursus nasional yang luas, mencakup aspek pemerintahan, pendidikan, dan pengelolaan belanja negara. Temuan ini menjadi dasar yang kuat dalam menganalisis persepsi publik terhadap kebijakan efisiensi anggaran menggunakan pendekatan analisis sentimen.



Gambar 3.14 *Trends* "Efisiensi Anggaran" Topik Teratas

3.1.7 Tagar Platform X Sebagai Indikator Kebijakan Presiden Prabowo

Dalam sebuah penelitian analisis sentimen, pemilihan studi kasus merupakan fondasi yang menentukan kualitas hasil akhir. Untuk itu, penelitian ini memerlukan bukti bahwa empat kebijakan Presiden Prabowo yang dipilih memang merupakan isu yang signifikan di mata masyarakat. Sub-bab ini akan menguraikan bagaimana fenomena tagar viral di Platform X digunakan sebagai tolok ukur utama untuk validasi topik tersebut.

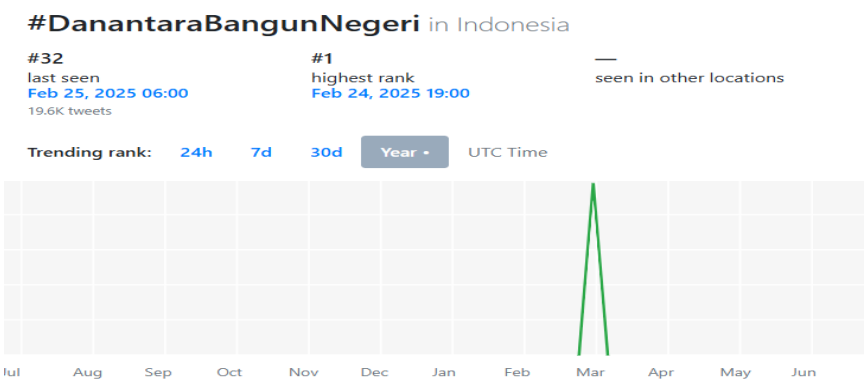
Platform X dapat diibaratkan sebagai cermin dari percakapan publik yang terjadi secara *real-time*. Ketika sebuah tagar kebijakan menjadi populer dan banyak digunakan, hal ini menunjukkan adanya "energi" publik yang besar terhadap isu tersebut. Energi inilah yang menjadi alasan kuat mengapa popularitas tagar sangat penting bagi penelitian ini. Setidaknya ada dua alasan. Pertama, viralnya sebuah tagar mengkonfirmasi adanya perdebatan publik yang beragam ada yang pro, kontra, dan netral. Kedua, dan yang terpenting untuk aspek teknis, popularitas ini menjamin pasokan data teks yang kaya dan melimpah. Ketersediaan data yang besar adalah syarat mutlak agar analisis sentimen menggunakan metode *deep learning*

dapat berjalan dengan baik. Maka dari itu, menjadikan tagar viral sebagai filter di tahap awal adalah sebuah langkah metodologis yang krusial untuk penelitian ini.

3.1.7.1 Kebijakan Danantara

Bukti relevansi publik untuk kebijakan Danantara dapat dilihat secara kuantitatif melalui analisis popularitas tagar di platform X. Salah satu tagar utama yang menjadi pusat diskusi untuk kebijakan ini adalah **#DanantaraBangunNegeri**.

Berdasarkan data tren yang terekam, tagar ini menunjukkan lonjakan popularitas yang sangat signifikan pada akhir Februari 2025. Seperti yang dapat diamati pada gambar dibawah ini, tagar **#DanantaraBangunNegeri** berhasil mencapai peringkat puncak sebagai **#1** ***trending topic*** di Indonesia pada tanggal 24 Februari 2025 pukul 19:00. Pada puncaknya, tagar ini telah digunakan dalam lebih dari 19.600 cuitan (*tweets*).



Gambar 3.15 *Trends* Tagar #DanantaraBangunNegeri di *Platform* X

Pencapaian peringkat #1 seperti yang ditunjukkan pada gambar di atas membuktikan bahwa pada waktu tersebut, diskursus mengenai kebijakan ini mampu mendominasi percakapan nasional di platform X. Lonjakan tajam pada grafik mengindikasikan adanya sebuah peristiwa pemicu yang berhasil menarik atensi publik secara masif. Volume cuitan yang mencapai puluhan ribu ini juga menandakan terbentuknya korpus data yang sangat besar dan kaya akan opini publik.

Bukti popularitas yang terukur ini menjadi justifikasi yang kuat mengapa kebijakan Danantara dipilih sebagai salah satu studi kasus dalam penelitian ini. Hal ini

memvalidasi bahwa terdapat cukup banyak data sentimen untuk diekstraksi dan dianalisis secara mendalam menggunakan model *deep learning* IndoBERT.

3.1.7.2 Kebijakan Makan Bergizi Gratis

Salah satu kebijakan yang paling banyak menyita perhatian publik dan menjadi fokus dalam penelitian ini adalah program "Makan Bergizi Gratis". Relevansi kebijakan ini di tengah masyarakat dapat dibuktikan secara kuantitatif melalui tingginya aktivitas percakapan di platform X, yang salah satunya terpusat pada tagar **#MakanBergiziBangunNegeri**.

Tagar ini digunakan secara luas untuk mensosialisasikan dan membangun narasi positif mengenai program tersebut, dengan mengaitkan gizi dengan pembangunan bangsa. Data tren menunjukkan bahwa tagar ini berhasil mendapatkan atensi publik yang luar biasa dalam periode yang sangat singkat. Seperti yang dapat dilihat pada Gambar dibawah ini, tagar **#MakanBergiziBangunNegeri** menduduki peringkat #1 *trending topic* di Indonesia pada tanggal 29 Desember 2024 pukul 08:00. Pada puncaknya, tagar ini telah digunakan dalam 15.800 cuitan (*tweets*).

Gambar 3.16 *Trends* Tagar #MakanBergiziBangunNegeri di *Platform* X

Pencapaian peringkat puncak sebagai topik paling populer se-Indonesia, meskipun dalam waktu singkat, menegaskan bahwa kebijakan Makan Bergizi Gratis mampu menjadi pusat percakapan nasional. Lonjakan tajam yang terlihat pada grafik mengindikasikan adanya sebuah peristiwa pemicu yang sangat efektif dalam menarik perhatian publik secara serentak.

Volume percakapan yang mencapai lebih dari 15.000 cuitan merupakan bukti konkrit terbentuknya kumpulan data opini publik yang kaya. Oleh karena itu, bukti popularitas yang terukur ini menjadi justifikasi yang sangat kuat untuk memilih kebijakan Makan Bergizi Gratis sebagai salah satu studi kasus utama dalam penelitian ini, sekaligus memastikan ketersediaan data yang memadai untuk dianalisis menggunakan model *deep learning*.

3.1.7.3 Kebijakan PPN 12%

Salah satu rencana kebijakan yang memicu reaksi publik paling kuat adalah kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%. Berbeda dengan kebijakan lain yang mungkin disambut dengan narasi beragam, diskursus mengenai kebijakan ini di platform X didominasi oleh sentimen kritis dan penolakan. Bukti paling nyata dari fenomena ini adalah viralnya tagar **#PajakMencekik.**

Nama tagar itu sendiri, yang memiliki konotasi negatif sangat kuat, menjadi indikator awal dari arah sentimen publik. Data tren menunjukkan bahwa tagar ini berhasil mendapatkan jangkauan yang sangat luas. Seperti yang dapat dilihat pada Gambar dibawah, tagar **#PajakMencekik** sukses menembus jajaran atas *trending topic* dengan menduduki peringkat #2 di Indonesia pada tanggal 20 Desember 2024 pukul 10:00. Popularitasnya didukung oleh volume percakapan yang sangat masif, yaitu mencapai 43.100 cuitan (*tweets*).

Gambar 3.17 *Trends* Tagar #PajakMencekik pada *Platform* X

Keberhasilan sebuah tagar dengan sentimen kritis menduduki posisi kedua terpopuler se-Indonesia adalah sebuah fenomena yang signifikan. Hal ini menunjukkan adanya gelombang penolakan publik yang terorganisir dan meluas.

Volume percakapan yang melebihi 43.000 cuitan juga menandakan terbentuknya korpus data sentimen negatif yang sangat besar dan kaya, menjadikannya sumber daya yang ideal untuk dianalisis.

Dengan demikian, bukti popularitas tagar kritis **#PajakMencekik** ini menjadi justifikasi utama mengapa kebijakan PPN 12% dipilih sebagai studi kasus. Adanya data sentimen negatif yang melimpah dan terukur ini sangat penting untuk menghasilkan analisis yang seimbang dan komprehensif mengenai persepsi publik terhadap kebijakan pemerintah dalam penelitian ini.

3.1.7.4 Kebijakan Efisiensi Anggaran

Kebijakan terakhir yang diidentifikasi memiliki relevansi publik yang tinggi adalah inisiatif "Efisiensi Anggaran". Wacana mengenai kebijakan ini bertujuan untuk membangun citra pemerintahan yang cermat dan bertanggung jawab dalam mengelola keuangan negara. Bukti kuat dari perhatian publik terhadap isu ini dapat dilihat pada tagar **#EfisiensiUntukBangsa**.

Tagar ini, yang secara langsung mengaitkan efisiensi dengan kepentingan bangsa, berhasil mendapatkan sambutan positif dan menjadi pusat diskusi. Data trend pada gambar dibawah menunjukkan performa tagar yang sangat impresif, dimana tagar tersebut berhasil menduduki peringkat #1 *trending topic* di Indonesia pada 09 Februari 2025 pukul 20.00. Pada saat yang sama, tagar ini juga mencapai puncak volume percakapan dengan total 10.500 cuitan (*tweets*).

Gambar 3.18 *Trends*Tagar #EfisiensiUntukBangsa di *Platform* X

Keberhasilan tagar **#EfisiensiUntukBangsa** menjadi topik terpopuler nasional menegaskan bahwa isu pengelolaan anggaran yang efisien merupakan topik yang

dianggap penting oleh publik. Lonjakan tajam pada grafik, seperti pada kasus-kasus sebelumnya, mengindikasikan adanya peristiwa pemicu yang efektif. Data ini sekali lagi memvalidasi pemilihan kebijakan ini sebagai studi kasus yang relevan dan memastikan ketersediaan data sentimen yang cukup untuk dianalisis.

Dari paparan bukti-bukti kuantitatif yang telah disajikan di atas, dapat ditarik kesimpulan yang kuat bahwa keempat kebijakan yang dipilih **Danantara, Makan Bergizi Gratis, PPN 12%, dan Efisiensi Anggaran** bukanlah topik yang dipilih secara acak. Masing-masing kebijakan telah terbukti mampu memicu diskursus publik yang masif di platform X, yang ditandai dengan viralnya tagar-tagar spesifik.

Keragaman sentimen juga terlihat dengan jelas. Ada kebijakan yang didominasi oleh kampanye dukungan positif (seperti **#EfisiensiUntukBangsa**), ada yang disambut dengan gelombang kritik dan penolakan yang kuat (seperti **#PajakMencekik**), dan ada pula yang memicu perdebatan pro-kontra yang seimbang. Fenomena ini menunjukkan bahwa keempat kebijakan tersebut merupakan isu yang dinamis dan "hidup" di tengah masyarakat, menjadikannya subjek yang sangat ideal untuk dianalisis persepsi publiknya.

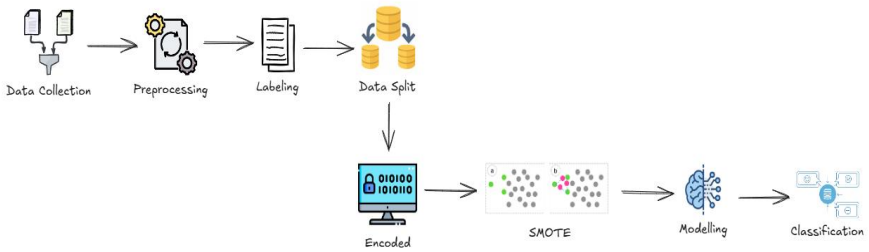
Dengan demikian, proses pemilihan topik yang didasarkan pada indikator terukur berupa popularitas tagar ini secara metodologis telah memvalidasi kelayakan penelitian ini. Ketersediaan data sentimen yang melimpah dan beragam dari keempat kebijakan ini menjadi landasan yang kokoh untuk melanjutkan penelitian ke tahap analisis sentimen mendalam.

3.2 Pendekatan Penelitian

Penelitian ini menganalisis sentimen publik terhadap kepemimpinan Presiden Prabowo Subianto, khususnya terkait program makan gizi gratis, efisiensi anggaran, danantara dan kebijakan PPN 12%. Analisis dilakukan melalui data dari pengguna X dan YouTube untuk menentukan apakah sentimen yang muncul bersifat positif, negatif, atau netral. Pendekatan penelitian analisis sentimen menggunakan model BERT (*Bidirectional Encoder Representations from Transformers)* dilakukan dengan memanfaatkan kemampuan BERT dalam memahami konteks dua arah pada teks.

Model ini terdiri dari dua tahap utama, yaitu *pre-training* dan *fine-tuning*, yang menggunakan BERT untuk melakukan klasifikasi sentimen secara akurat lalu untuk penelitian ini, Google Docs digunakan untuk menyusun laporan, Google Collaboratory digunakan untuk mengumpulkan dan memproses data dari X, sementara Apify digunakan untuk mengumpulkan data dari YouTube. Selain itu, Mendeley Desktop digunakan untuk mengelola dan menyisipkan referensi dalam laporan penelitian.

Dalam analisis sentimen, terdapat beberapa tahapan yang harus dilakukan. Rangkaian tahapan tersebut dapat dilihat pada gambar berikut.



Gambar 3.19 Tahapan analisis sentimen

3.2.1 Model IndoBERT

Penelitian ini mengadopsi model IndoBERT (*Indonesian Bidirectional Encoder*) sebagai fondasi utama dalam analisis sentimen. IndoBERT merupakan adaptasi dari arsitektur BERT yang telah dilatih secara khusus menggunakan korpus data berbahasa Indonesia yang besar dan beragam. Pemilihan IndoBERT didasarkan pada kemampuannya untuk memahami konteks bahasa Indonesia secara lebih mendalam dibandingkan model BERT versi umum yang dilatih dengan data berbahasa Inggris.

Kemampuan IndoBERT dalam menangkap nuansa bahasa lokal, termasuk penggunaan ragam bahasa tidak formal, slang, akronim, dan struktur kalimat khas Indonesia, sangat krusial dalam analisis sentimen terhadap komentar publik di platform media sosial seperti X dan YouTube. Data yang tidak terstruktur dan seringkali mengandung ekspresi informal memerlukan model yang mampu memproses konteks semantik secara akurat. IndoBERT, dengan arsitektur

transformer dua arahnya, memungkinkan pemahaman konteks kata secara utuh, baik dari kata-kata sebelumnya maupun sesudahnya dalam suatu kalimat.

Model ini terdiri dari dua tahap utama yaitu *pre-training* dan *fine tuning*. Pada tahap *pre-training*, IndoBERT dilatih pada data teks berbahasa Indonesia dalam skala besar untuk mempelajari representasi bahasa umum. Ini melibatkan tugas seperti *Masked Language Model* (memprediksi kata yang dihilangkan) dan *Next Sentence Prediction* (memprediksi apakah dua kalimat berhubungan). Setelah *pre-training*, model disesuaikan dengan tugas spesifik analisis sentimen dan klasifikasi topik yang ada dalam penelitian ini. Proses *fine-tuning* melibatkan penambahan lapisan *output* di atas model IndoBERT yang sudah terlatih, kemudian dilatih ulang dengan data berlabel dari komentar publik. Penggunaan IndoBERT diharapkan dapat menghasilkan klasifikasi sentimen (positif, negatif, netral) dan topik (MBG, PPN 12%, Efisiensi Anggaran, dan Danantara) yang lebih akurat dan relevan dengan konteks bahasa dan isu-isu di Indonesia.

3.3 Metodologi Pengumpulan Data

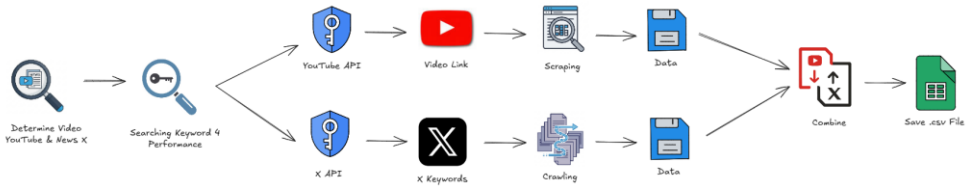
Berikut merupakan metodologi pengumpulan data dalam penelitian ini.

3.3.1 Studi Literatur

Studi Literatur dalam analisis sentimen mencakup berbagai pendekatan dan metode yang digunakan untuk memahami dan mengklasifikasi opini publik dari teks. Salah satu penelitian yang menunjukkan bahwa analisis sentimen dapat digunakan untuk mendapatkan informasi tentang tokoh politik, seperti calon gubernur, dengan tujuan menentukan opini di media sosial[1]. Selain itu, penelitian lain menelaah berbagai pendekatan analisis sentimen, seperti metode berbasis *machine learning* yang memerlukan data latih berlabel, dan metode berbasis leksikal (*Lexicon Based)* yang menggunakan kamus kata sentimen tanpa perlu pelatihan dataset [38]. Pendekatan lain yang sering digunakan adalah analisis Sentimen berbasis aspek, deteksi emosi, dan analisis sentimen tingkat lanjut, yang masing-masing bertujuan untuk mengevaluasi aspek tertentu, menentukan emosi dari pesan, dan memberikan penilaian spesifik yang sering digunakan dalam *e-commerce "Sentiment of Youtube Social Media Users Towards the Presidential Candidates of Ganjar Pranowo and Prabowo Subianto in the 2024".* Studi lain juga mengaplikasikan *BERT* untuk

analisis sentimen berbasis aspek (*aspect-based sentiment analysis),* yang memungkinkan analisis sentimen lebih detail pada aspek tertentu dalam teks. Studi-studi ini menunjukkan perkembangan signifikan dalam analisis sentimen publik pada berbagai topik menggunakan berbagai metode dan platform. Pada penelitian ini untuk pengumpulan data dilakukan dari berbagai sumber seperti, jurnal, buku, dan makalah yang relevan dengan analisis sentimen pada model *BERT* (*Bidirectional Encoder Representations*) pada kepemimpinan Prabowo Subianto menggunakan bantuan dari *platform* X dan Youtube.

Tahapan awal dalam melakukan analisis adalah pengumpulan data. Pengumpulan data dilakukan melalui dua cara pada dua *platform* yang berbeda, yaitu menggunakan metode *crawling* pada *platform* X dan *scraping* pada *platform* YouTube. Tahapan pengumpulan data dapat dilihat pada gambar berikut:



Gambar 3.20 Pengumpulan data X dan YouTube

3.3.2 Crawling Data

Crawling data merupakan proses pengumpulan data dengan dari sosial media X yang dilakukan secara otomatis untuk memperoleh data dalam jumlah besar. Dalam penelitian ini, proses *crawling* dilakukan untuk mengumpulkan opini publik terkait kepemimpinan Prabowo Subianto dengan fokus pada empat topik kebijakan utama, yaitu Makan Bergizi Gratis, PPN 12%, Efisiensi Anggaran, dan Danantara. Proses *Crawling* ini dijalankan menggunakan *Google Collaboratory* dengan bantuan pustaka python yaitu *tweety* yang berfungsi sebagai antarmuka ke Api resmi milik *platform* X

Langkah awal dalam proses ini dimulai dengan konfigurasi kredensial API yang terdiri atas API key, API *secret key, access token secret* yang diperoleh melalui akun pengembangan platform *X.* Setelah berhasil melakukan *autentikasi* dilakukan pencarian terhadap data menggunakan parameter kata kunci seperti “Prabowo”,

“Makan Gratis”, “ PPN 12%”, dan lainnya. Untuk mempersempit cakupan data, pencarian dibatasi pada periode tertentu serta difilter agar hanya mencakup *tweet* berbahasa Indonesia. Data yang berhasil dikumpulkan berupa teks utama (*tweet*), waktu publikasi dan atribut lain yang berkaitan. Seluruh hasil crawling kemudian disimpan dalam format .csv di google drive sehingga dapat digunakan langsung untuk tahap *preprocessing* dan pelabelan data.

Penggunaan Google Collab memberikan keuntungan dalam bentuk lingkungan komputasi awan yang terintegrasi dengan penyimpanan google drive, serta mendukung eksekusi kode *python* dengan pustaka terkait seperti *pandas* , *json* , dan *.csv* yang membantu dalam proses ekstraksi dan penyimpanan data.

Gambar 3.21 Tahapan Crawling Data

Melalui proses *crawling,* diperoleh sebanyak 3780 komentar publik yang relevan dengan masing-masing topik. Adapun jumlah data yang diperoleh dari setiap topik dapat dilihat pada tabel berikut.

Tabel 3.2 Jumlah Komentar Hasil Crawling Berdasar Topik

No	Topik	Jumlah Komentar
1.	Makan Bergizi Gratis	203
2.	PPN 12%	348
3.	Efisiensi Anggaran	1689
4.	Danantara	1540

Setiap data komentar dikumpulkan memiliki beberapa atribut atau kolom yang merepresentasikan informasi penting dari tweet tersebut. Kolom-kolom yang tersedia antara lain *conversation_id_str* sebagai ID dari *thread* percakapan, *create_at* sebagai waktu tweet dibuat, *favotite_count* yang menunjukkan jumlah like, *full_text* sebagai isi utama komentar, *id_str* sebagai ID unik tweet *image_url* untuk URL gambar jika ada, *in_replay_to_screen_name* sebagai *username* dari

tweet yang dibalas (jika merupakan balasan), serta *lang* yang menunjukkan bahasa yang digunakan dalam tweet tersebut.

Namun, karena fokus utama dari penelitian ini adalah menganalisis opini atau sentimen publik yang terkandung dalam isi komentar, maka kolom yang menjadi perhatian utama adalah kolom *full_text,* yang berisi teks utuh dari setiap komentar tweet yang dikumpulkan. Langkah ini bertujuan untuk mengurangi kompleksitas data, mempercepat proses pemrosesan, serta meningkatkan efisiensi dalam pembangunan model analisis sentimen yang akan dilakukan pada tahapan selanjutnya.

Untuk memberikan gambaran lebih lanjut mengenai bentuk data yang dianalisis berikut disajikan beberapa komentar hasil crawling:

Tabel 3.3 Hasil Crawling Berdasarkan Topik

No	Topik	Komentar (*Full_Text*)
1.	Makan Bergizi Gratis	tiap hari kenyang makan sabar https://t.co/xiZfskmwxv
		Spill dong tadi siang makan sama apa?! https://t.co/W0fiGI0i70
		Makan siang hari ini Hot Side Story Chicken by @hangry_id Jadi aku beli beberapa varian rasa 2 pcs Fried Chicken Hot Nashville Complete Set 1 pcs Fried Chicken Chicago BBQ Rice Set 5 pcs Fire Mala Chicken Wings Rasanya mantul banget apalagi Varian kalo yang Hot https://t.co/5gHu2zYEkX
		yuk utamain makan Real Food dibanding Ultra Processed Food https://t.co/kQ0fx36W2s
		Kamu butuh 5 hobi ini untuk menjadi versi terbaikmu: 1. Hobi yang bisa membuatmu fit dan sehat: Yoga Zumba Gym Makan sehat tidur teratur dll 2. Hobi yang bisa menghasilkan uang: Content creating bisnis high income skilss freelance dll 3. Hobi yang bisa membangun
2.	PPN 12%	PPN naik layanan dasar juga kena pajak. Tapi oligarki tetap dapat insentif. Jadi uang pajak ini untuk siapa sih sebenarnya? Biang keroknya ternyata adalah Undang-Undang (UU) Harmonisasi Peraturan Perpajakan (HPP). #RevisiUUHPP https://t.co/Qh1wrhgrUW
		Berikut penghitungan PPN untuk objek pajak yang menggunakan DPP Nilai Lain sesuai PMK 11/2025. https://t.co/M6pgkIN4fP
		Rakyat bingung blm ada kebijakan yg benar² pro rakyat. Kabinet gemoy PPN 12% carut-marutnya MBG dan kini malah sibuk dgn Revisi UU yg unfaedah & cuma menguntungkan elit/pejabat. Giliran didemo mereka tuding itu semua mahasewa. Waras? #IndonesiaGelap https://t.co/w7QHGVMErS
		Ya Allah capek banget loh ngadepin coretax ini. Min klo telat lapor spt masa ppn kena denda gk? @kring_pajak saya tiap hari mau lapor spt aja dipersulit sampe hari ini. Tadinya gk bisa di klik lapor dan bayar trus angka Lebih Bayar bln desember ilang dan skg error.

No	Topik	Komentar (*Full_Text*)
3.	Efisiensi Anggaran	Harusnya beban hidup gue juga kena efisiensi ini biar mendingan soalnya udah berlebihan
		Karena efisiensi anggaran beasiswa utk pendidikan dokter dokter gigi dan spesialis dari Kemenkes sementara dihentikan. Padahal Indonesia sudah punya masalah distribusi tenaga medis yg ga merata termasuk dokter spesialis. Situasi bisa memburuk khususnya di daerah terpencil yg https://t.co/dBHBUFTaPd
		Hari ini di aksi #IndonesiaGelap warga turun ke jalan atas begitu banyak rentetan kebijakan negara yang justru menyengsarakan warganya sendiri. Efisiensi anggaran 20 juta hektar hutan untuk pangan dan energi RUU Minerba yang dikebut semalam pendidikan dan kesehatan yang https://t.co/QvDXNHkbmT
		Ada yang udah packing? Pemerintah menjamin efisiensi anggaran tidak berdampak terhadap masyarakat secara langsung. Tapi ternyata banyak pekerja honorer diintai PHK kapan saja karena kantor mereka terdampak efisiensi anggaran. #kumparanplus #lipsus #efisiensianggaran https://t.co/lWipkAbaml
		Harusnya beban hidup gue juga kena efisiensi ini biar mendingan soalnya udah berlebihan
4.	Danantara	Speechless.... ▯ Dmn2 PHK.. Kalian dgn bangga bikin Danantara... Korupsi merajalela KPK tak berguna... Si Firli apa kabarnya.? Indonesia gelap n suram itu nyata!! https://t.co/7tJkqYz9Db
		Dana Anagata Nusantara - Danantara adalah Lembaga Pengelola Investasi atau Sovereign Wealth Fund (SWF) yang dibentuk pada 24 Februari 2025 nanti Danantara berarti energi atau kekuatan untuk tanah air Indonesia merupakan konsolidasi kekuatan BUMN Apa efek ke Pasar Modal? https://t.co/AtUWBKUpLB
		Danantara atau Badan Pengelola Investasi Daya Anagata Nusantara adalah lembaga investasi yang dibentuk oleh Presiden Prabowo Subianto untuk mengoptimalkan pengelolaan kekayaan negara melalui investasi strategis. Nama Danantara berasal dari kata Daya yang berarti energi atau https://t.co/t7VAqtXTtk
		@ndorokakung Danantara adalah peternak yg maksa ngumpulin ayam-ayam punya orang tp peternak ini tidak punya sistem yg bisa mengurus ayam dengan baik. Ayam-ayam yg dikumpulin berisiko mati dan bikin bangkrut peternakan beserta pemilik2 ayam.

Data komentar yang sudah dikumpulkan dan ditampilkan pada tabel diatas akan melalui tahapan pre-pemrosesan (*pre-processing*) untuk membersihkan dan menyiapkan data sebelum dilakukan analisis sentimen. Tahapan ini meliputi menyiapkan data sebelum dilakukan analisis sentimen. Tahapan ini meliputi penghilang URL, emoji, tanda baca yang tidak relevan, serta normalisasi teks agar hasil analisis menjadi lebih akurat dan efektif. Proses ini sangat penting untuk memastikan bahwa model yang dibangun dapat menangkap opini publik secara tepat dari data teks yang telah diperoleh.

3.3.3 Scraping Data

Scraping data merupakan teknik yang digunakan untuk mengekstrak data spesifik dari suatu halaman web. Dalam penelitian ini, teknik *scraping* digunakan untuk mengumpulkan komentar publik dari platform YouTube berkaitan dengan kepemimpinan Prabowo Subianto, khususnya yang membahas program makan siang bergizi gratis, kebijakan PPN 12%, efisiensi anggaran, dan danantara.

Alur pengumpulan data dimulai dengan penggunaan apify, sebuah platform otomatisasi web yang digunakan untuk menjalankan proses *scraping* secara efisien dan tanpa menggunakan kode secara manual. Apify terintegrasi dengan YouTube API, yang memungkinkan sistem mendapatkan akses resmi untuk mengekstrak komentar dari video YouTube berdasarkan video link yang telah ditentukan sebelumnya. Setelah link video ditentukan, sistem melakukan *scraping* terhadap komentar menggunakan tool bawaan apify, yang kemudian menyimpan data hasil ekstraksi dalam format .csv. Data inilah yang selanjutnya digunakan sebagai dasar untuk analisis sentimen dan topik dalam penelitian.

Gambar 3.22 Tahapan *Scraping* Data

Dari keempat program Prabowo yang dianalisis, terdapat total 7930 komentar. Jumlah komentar yang diperoleh dari masing-masing topik dapat dilihat pada tabel berikut.

Tabel 3.4 Jumlah Komentar Hasil Scraping Berdasar Topik

No	Topik	Jumlah Komentar
1.	Makan Bergizi Gratis	2953
2.	PPN 12%	3248
3.	Efisiensi Anggaran	182
4.	Danantara	1547

Pada tahap *scraping* data, diperoleh data mentah yang memiliki kolom: nama pengguna, isi komentar, jumlah like, waktu komentar dipublikasikan, jumlah

balasan, dan tautan video. Karena penelitian ini hanya fokus pada isi komentar, maka kolom lain akan dihapus. Selanjutnya, dilakukan proses *preprocessing* untuk menyiapkan komentar agar dianalisis lebih lanjut. Untuk memberikan gambaran lebih lanjut mengenai bentuk data yang dianalisis berikut disajikan beberapa komentar hasil *scraping*:

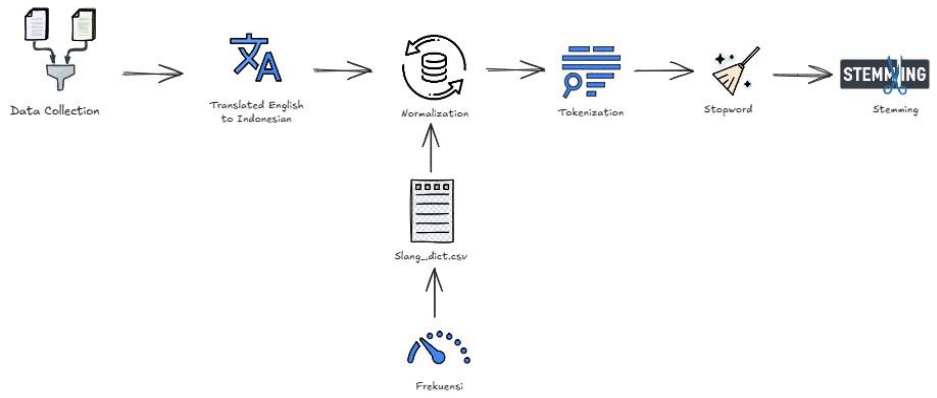
Tabel 3.5 Hasil *Scraping* Berdasarkan Topik

No	Topik	Komentar (*Full_Text*)
1.	Makan Bergizi Gratis	Yg lebih penting mah lapangan pekerjaan yg di banyakin..biar orang tuang bisa ngasih makan keluarganya..pemerintah gak perlu cape bikin program makan
		Saya setuju, yg gk setuju cuma orang yg udh kenyang makan dn tak peduli yg masih kurang gizi
		😱 😱 😱 ANAK SEKOLAH MINUM KOPI SAMA SUSU JAHE GRATIS,, BIAR BELAJARNYA KAGA NGANTUK 😱 😱 😱 😱
		Dgn sdm 58% adlh bukti bgsa in blm siap maju krn mentaly adlh mental pemalas d otaky isiny cma makan it knpa mereka g marh ktika d samakn binatg sma gibran.,
		kirain kita juga dapat makan gratis toh yang sd aja toh 😅 😅 mana udah berenti kerja makan gratis gausah kerja eh ternyata
2.	PPN 12%	PAJAK 12% Cara membunuh Rakyat sekarang udah binggung akan hidup gimana masa depan anak cucu kedepan. Rakyat diperas untuk menutupi APBN. Klu terus begini TUNGGU KEHANCURAN NKRI 😥
		Ngacooo korupsi mkin meraja lela eee pajak di naikin dengan dalih untuk kebaikan APBN ngacooo ngacooo
		Lebay orang2 nih, gaji minta naik tp PPN gag boleh naik? Ini negara nenek moyangmu... 😂
		Aneh yg dukung 12%..gmn kl sekalian 20% aja..sekaya apa kalian bs ngomong itu berdampak positif?terang2 harga bakal naik semua dan gaji ga bs mengikuti..gmn bs positif..aneh..
		hapuskan ppn 12% dan pph progresif krn memiskinkan rakyat
3.	Efisiensi Anggaran	hahaha beneran kena efisiensi dahlah gada ruang aman di negara ini. punten https://t.co/L2IOuVKvJE https://t.co/gZS5gK9EWX
		bangsattt memang kaum 58% gara2 kalian gw kerja bukannya cari nafkah malah nombok. submit jurnal scopus pake duit sendiri krn efisiensi ini. nelangsa 5 tahun gara2 presiden pilihan kaum 58%! tai lo semua.
		Efisiensi katanya Efisiensi ndasmu pea Noted .... PERUSAHAAN PENYELENGGARA RETRET 505 KEPALA DAERAH DI MAGELANG MILIK KADER GERINDRA https://t.co/IlHYojnoFP
		Ubur-ubur ikan lele Efisiensi anggaran nggak ngaruh lee!!
		Pendidikan dan Kesejahteraan Adalah Hak Bangsa Kebijakan efisiensi anggaran merupakan bagian dari Instruksi Presiden Nomor 1 Tahun 2025 bertujuan menghemat Rp.306 6 Triliun anggaran pada berbagai sektor.
4.	Danantara	danantara sangat bagus, 😊 asal uangnya tidak di korupsi pasti 7 atau 9 tahun lagi Indonesia akan seperti china dan Amerika,
		Sukses dan selalu beruntung program daya anagata nusantara (danantara) memberi dampak baik bagi bangsa dan negara

No	Topik	Komentar (*Full_Text*)
		Indonesia, sehat selalu dan panjang umur presiden RI ke-8 Prabowo Subianto dan wapres RI ke-14 Gibran Rakabuming.
		Apa sih kebijakan efisiensi ini sebenarnya? AUK AH GELAP. Kemana 300 triliun itu hasil efisiensi? Sebagian untuk Makan Bergizi Gratis selebihnya masuk Danantara // SELENGKAPNYA TONTON YOUTUBE ILC https://t.co/nLXoCej3Yd https://t.co/xiY3K2PyAT
		Danantara bakal jd lembaga super kebal kpk ppatk kejaksaan? Klo benar maka akan jd sumber masalah (korupsi) baru. Tambahan lg finalis terkorup jd pengawas. Mo dibikin jd apa negara ini wo?
		Kalo u cewek. Danantara itu ibu mertua u. Negara itu suami u. Jadi 80% penghasilan suami u dititipin ke ibu mertua dikelola sama beliau. Dengan keadaan suami u 11 bersaudara. Hehe.

3.4 Tahap Pengolahan Data

Untuk mempermudah pengolahan data, dilakukan tahapan *preprocessing*. Tahapan ini bertujuan agar proses analisis data dapat berjalan lebih mudah dan terstruktur. Tahapan *preprocessing* dapat dilihat pada gambar berikut



Gambar 3.23 Tahapan *Preprocessing*

3.4.1 Tahapan Translate

Dalam tahap *preprocessing*, proses *translate* digunakan untuk mengolah teks agar dapat diklasifikasikan sebagai positif, negatif, atau netral secara lebih akurat. Proses ini penting terutama ketika komentar mengandung campuran bahasa, seperti kombinasi bahasa Indonesia dan bahasa asing misalnya bahasa Inggris. Tahap ini bertujuan untuk menerjemahkan kata atau frasa asing ke dalam bahasa Indonesia agar seluruh teks dapat disajikan secara konsisten.Walaupun komentar yang dianalisis dalam penelitian ini sebagian besar menggunakan bahasa Indonesia, karena topiknya berkaitan dengan kinerja Presiden Prabowo Subianto, proses

translate tetap dilakukan untuk mengantisipasi keberadaan kata atau frasa asing guna menjaga konsistensi dan akurasi hasil analisis.

3.4.2 Case Folding

Pemrosesan teks yang esensial untuk menyamakan semua huruf dalam dokumen menjadi huruf kecil. Proses ini bertujuan untuk memastikan konsistensi data, sehingga kata yang sama, meskipun ditulis dengan kapitalisasi yang berbeda, diperlukan sebagai entitas tunggal. Dengan demikian, *case folding* mempermudah dan meningkatkan akurasi tahapan pengolahan teks selanjutnya.

3.4.3 Data Cleaning

Dalam analisis sentimen, terdapat tahapan *cleaning* adalah proses pembersihan teks dari elemen-elemen yang tidak relevan agar lebih mudah dianalisis oleh sistem. Data teks yang berasal dari komentar pada media sosial sering kali mengandung berbagai karakter atau format yang bisa mengganggu hasil analisis, seperti tanda baca, angka diubah menjadi huruf, simbol, hingga kata-kata yang tidak memiliki makna dalam konteks sentimen. Pada Tahapan *cleaning* dalam pemrosesan teks adalah proses membersihkan data komentar dari elemen- elemen yang tidak penting atau mengganggu analisis, seperti tanda baca, simbol, *lowercase,* tautan (*URL*), nama pengguna (*mentions),* angka, tagar (*hashtag*), spasi berlebih, dan karakter berulang. Tujuannya adalah menghilangkan noise agar data menjadi lebih bersih dan siap untuk diproses lebih lanjut oleh model. Berikut code dari yang digunakan untuk melakukan tahapan *cleaning.* Tahapan *cleaning* sangat penting karena teks mentah yang belum dibersihkan dapat menghasilkan analisis yang tidak akurat. Dengan membersihkan teks terlebih dahulu, sistem dapat lebih fokus pada kata-kata yang memiliki makna sentimen yang jelas.

3.4.4 Normalisasi

Tahapan normalisasi merupakan proses mengubah kata-kata yang tidak baku atau singkatan menjadi kata baku yang sesuai dengan kamus yang terdapat dalam file *slang_dict.csv.* Sebelum menyusun kamus tersebut, terlebih dahulu dilakukan perhitungan frekuensi kata pada data mentah untuk mengidentifikasi kata-kata tidak baku yang sering muncul dalam dataset. Berdasarkan hasil analisis frekuensi

tersebut, kemudian dibentuk kamus normalisasi yang memuat kata tidak baku dan bentuk bakunya. Kamus ini digunakan dalam proses normalisasi untuk menggantikan kata-kata tidak baku menjadi bentuk baku agar seluruh data teks menjadi baku dan siap untuk dianalisis.

Proses pembentukan kamus ini melibatkan identifikasi manual dan kurasi kata-kata yang sering muncul dalam komentar berbahasa Indonesia di media sosial. Ini termasuk singkatan (seperti "yg", "dgn","tp") akronim, serta variasi kata yang berasal dari imbuhan (*derivasi).* Misalnya, kata-kata seperti "membeli", " dibeli", "pembelian", atau "makanan", "memakan, " dimakan" akan dinormalisasikan ke bentuk dasarnya, yaitu "beli" dan "makan", utnuk memastikan konsisten reprsentasi makna. pendekatan ini esensial untuk mengatasi karakteristik bahasa tidak formal di platform media sosial, yang sering kali tidak mengikuti kaidah bahasa baku, sehingga membantu moodle untuk memahami inti makna kata secara lebih akurat dan mengurangi variasi yang tidak perlu dalam data.

Secara spesifik, proses pembentukan kamus ini melibatkan identifikasi manual dan kurasi kata-kata yang sering muncul dalam komentar berbahasa Indonesia di media sosial, termasuk ke singkatan(yg, dgn, tp), akronim, seperti variasi kata yang berasal dari imbuhan (*derivasi*). Misalnya, kata-kata seperti 'membeli', 'dibeli', 'pembelian' atau 'makanan', 'memakan', 'dimakan' akan dinormalisasikan ke bentuk dasarnya, yaitu 'beli' dan 'makan', untuk memastikan konsisten representasi makna. Pendekatan ini esensial untuk mengatasi karakteristik bahasa tidak formal di platform media sosial, yang sering kali tidak mengikuti kaidah bahasa baku, sehingga membantu model untuk memahami inti makna kata secara lebih akurat dan mengurangi variasi yang tidak perlu dalam data.

Sebagai contoh, Tabel 3.4. Contoh Normalisasi Kata Tidak baku dan Variasi Kata, dibawah ini menyajikan beberapa ilustrasi dari proses normalisasi yang diterapkan dalam penelitian ini

Tabel 3.6 Normalisasi Kata Tidak Baku

Kata Tidak Baku	Kata Baku
jgn	jangan
skrg	sekarang
udah	sudah
gak	enggak

Kata Tidak Baku	Kata Baku
capee	cape
pegangin	memegang
nyender	menyandar
bener	benar
manjain	memanjakan
bgt	banget
peleotin	meleot
bng	bang
ko	kok
gda	tidak ada
yahh	ya
termurahhhh	termurah
pengen	pengin
nyatakan	menyatakan
graaatiiiis	gratis
nyatakan	menyatakan
meriahh	meriah
cantikkkm	cantik
dlu	dulu
dr	dari
loe	kamu
yg	yang
slama	selama
spt	seperti
udh	sudah
ckup	cukup
kyk	kayak
msh	masih
mnding	mending
cm	cuma
klo	kalau
gw	saya
nih	ini
uu	undang-undang
gitu	begitu
gmn	gimana
tuh	itu
gua	saya
trus	terus
kini	saat ini

Kamus ini kemudian digunakan dalam fungsi *fix_slangwords* untuk menggantikan kata-kata tidak baku atau turunan menjadi bentuk baku agar seluruh data teks menjadi standar dan siap untuk dianalisis. Proses ii esensial untuk menjaga konsistensi data dan meningkatkan akurasi analisis sentimen, karena model tidak akan salah menafsirkan sentimen karena perbedaan ejaan atau singkatan. Selain normalisasi ini, tahapan *stemming* juga dilakukan(dengan *Sastrawi*) untuk menyederhanakan kata menjadi bentuk dasarnya (misalnya, ‘berlari, ‘pelari’, ‘berlarian’ menjadi ‘lari). Ini berfungsi sebagai langkah perlengkapan untuk

mengatasi variasi kata akibat imbuhan, dan penggunaan algoritma *Sastrawi* membantu memastikan akurasi *stemming* pada bahasa Indonesia.

3.4.5 Tokenizing

Tahapan *tokenizing* adalah proses yang dilakukan untuk memecah teks atau kalimat menjadi bagian-bagian yang lebih kecil, yang disebut token. Proses ini biasanya dilakukan dengan memisahkan kata-kata berdasarkan spasi atau tanda baca. Pada tahapan pemecahan kalimat menjadi kata-kata individu menggunakan metode seperti *split()* dalam Python, yang memisahkan kata berdasarkan spasi [25]. Tanpa *tokenizing,* analisis teks tidak dapat dilakukan secara efektif karena proses ini membantu dalam memahami struktur dan makna dari dataset ini.

3.4.6 Stopword

Tahapan *Stopword removal* adalah proses menghapus kata-kata yang dianggap tidak membawa makna penting dalam analisis teks, seperti kata sambung, kata depan, dan kata ganti, yang sering muncul dalam bahasa sehari-hari terutama pada saat ingin mengutarakan opini pada sosial media. Penghapusan *stopword* bertujuan untuk mengurangi kebisingan (noise) dalam data sehingga model fokus pada kata-kata yang lebih bermakna dan relevan dalam menentukan sentimen dan konteks kalimat. Dengan demikian, *stopword removal* dapat meningkatkan efisiensi proses dan kadang-kadang membantu meningkatkan akurasi model pada algoritma tradisional yang berkurang mampu menangani konteks panjang.

Namun, untuk model berbasis transformer seperti BERT, yang menggunakan mekanisme perhatian (*attention)* untuk memahami hubungan antar kata dalam kalimiat, penghapusan *stopwords* tidak selalu memberikan peningkatan akurasi yang signifikan. BERT secara otomatis mempelajari konteks kata secara menyeluruh, termasuk kata-kata fungsi, sehingga penghilang *stopwords* berlebihan berpotensi menghilangkan informasi kontekstual penting dan justru menurunkan performa model.

Oleh karena itu, dalam penelitian ini tahapan *stopword removal* dilakukan secara selektif dengan mempertimbangkan konteks model BERT. Penyesuaian ini

bertujuan untuk menjaga keseimbangan antara pengurangan noise dan pemeliharaan konteks penting dalam teks, sehingga diharapkan dapat membantu meningkatkan efisiensi pemrosesan tanpa mengorbankan kualitas analisis sentimen.

3.4.7 Stemming

Stemming adalah proses menyederhanakan atau mengubah suatu kata menjadi bentuk dasarnya (*stem*), yaitu bentuk kata sebelum mengalami penambahan imbuhan seperti awalan (*prefix*), akhiran (*suffix*), sisipan *(Infixes),* dan gabungan awalan dan akhiran *(Confixes)*. Teknik ini digunakan untuk mengurangi jumlah variasi kata yang memiliki arti serupa, sehingga semua varian kata dapat diproses sebagai satu entitas yang sama. Meskipun hasil stemming terkadang bukan merupakan kata baku dalam bahasa, teknik ini tetap efektif dalam menyatukan berbagai bentuk kata ke dalam satu bentuk dasar. Misalnya, kata "berlari," "pelari," dan "berlarian" akan di *stemming* menjadi bentuk dasar "lari". Contoh lainnya, kata "makanan," "memakan," dan "dimakan" akan dikembalikan menjadi "makan" [39].

Namun, terdapat risiko *overstemming*, yaitu kondisi ketika dua kata berbeda yang seharusnya tidak disatukan justru dipangkas menjadi bentuk dasar yang sama, sehingga dapat menurunkan akurasi analisis. Oleh karena itu, di gunakanlah algoritma *Sastrawi* yang dirancang khusus untuk Bahasa Indonesia agar hasil stemming tetap akurat dan relevan. Penyesuaian ini sangat penting, mengingat penelitian ini berfokus pada analisis komentar publik terhadap kinerja Presiden Prabowo Subianto, di mana mayoritas teks berasal dari komentar berbahasa Indonesia. Dengan demikian, penggunaan algoritma stemming yang sesuai akan membantu menjaga konteks dan makna asli dari komentar tersebut, sehingga hasil analisis sentimen menjadi lebih valid dan representatif.

3.5 Lexicon Based Labelling

Tahapan pelabelan data merupakan fase esensial dalam siklus hidup penelitian analisis sentimen berbasis *deep learning,* khususnya untuk mempersiapkan dataset yang akurat dan relevan bagi pelatihan model. Dalam studi ini, proses pemberian label sentimen pada data teks yang dikumpulkan dari platform X dan Youtube mengadopsi pendekatan *semi-otomatis* dengan memanfaatkan kamus berbasis

lexicon (*lexicon-based dictionary).* Pendekatan ini dipilih untuk menginisialisasi dan mempercepat proses anotasi sentimen pada volume komentar yang besar, mengklasifikasikannya ke dalam tiga kategori polaritas: positif, negatif, dan netral.

Implementasi metode *lexicon-based* dalam penelitian ini dilakukan dengan mengintegrasikan dua sumber utama. Pertama, digunakan kamus sentimen eksternal yang telah teruji dan banyak digunakan dalam penelitian analisis sentimen berbahasa Indonesia, seperti Indonesia Sentimen *Lexicon(InSet Lexicon)* Indonesia Sentimen Lexicon. Lexicon ini menyediakan daftar kata dengan bobot polaritas yang sudah ditentukan, memungkinkan klasifikasi awal secara otomatis. Kedua, untuk meningkatkan relevansi dan akurasi sebagai kamus lexicon yang dibangun secara khusus(*custom lexicon*) dari dataset komentar yang diperoleh dari penelitian ini sendiri. Pembentukan *custom lexicon* ini krusial untuk mengakomodasi karakteristik unik bahasa di media sosial, termasuk pengguna singkatan , akronim, bahasa gaul (*slang),* atau ekspresi kontekstual yang spesifik terkait opini publik mengenai kepemimpinan Prabowo Subianto dan kebijakan-kebijakan (Efisiensi Anggaran, PPN 12%, Danantara, dan Makan Bergizi Gratis). Proses ini memastikan bahwa lexicon yang digunakan dapat menangkap nuansa sentimen lokal yang mungkin tidak sepenuhnya tercakup oleh kamus umum.

Proses pembangunan *custom lexicon* ini melibatkan identifikasi kata-kata kunci dan frasa yang sering muncul dalam dataset yang telah melalui tahapan *preprocessing* (pembersihan, normalisasi, tokenisasi, *stopword removal, dan stemming).* kata-kata ini kemudian dianalisis secara manual oleh peneliti untuk menentukan polaritas sentimennya(positif, negatif, atau netral) berdasarkan konteks penggunaan dalam komentar. Apabila ditemukan kata atau frasa yang memiliki ambiguitas polaritas, yaitu dapat mengarah ke sentimen positif, maupun netral (atau bahkan negatif) tergantung konteksnya, penentuan label didasarkan pada konteks kalimat secara keseluruhan dan *intensitas* emosi yang tersirat. Sebagai contoh, kata “lumayan” bisa positif jika diartikan “cukup bagus”, tetapi bisa netral jika “ biasa saja”. Dalam kasus tersebut, peneliti akan merujuk pada keseluruhan kalimat dan frasa pendukung untuk menentukan polaritas dominan. Jika setelah analisis konteks tidak ada indikasi polaritas yang kuat atau emosi yang jelas, kata tersebut akan cenderung

dikategorikan sebagai netral. Pemilihan kata positif didasarkan ekspresi yang menunjukkan dukungan, persetujuan,apresiasi, atau harapan baik terhadap kebijakan atau kinerja Prabowo Subianto, seperti kata “bagus”, “bermanfaat”, atau frasa “ sangat membantu”. Sebaliknya, pemilihan kata negatif didasarkan pada ekspresi yang mencerminkan ketidakpuasan, kritik, penolakan, kekhawatiran atau sentimen tidak menyenangkan lainnya, contoh “boros”, “rugi”, atau frasa “membebani rakyat”. Setiap kata yang telah ditentukan polaritasnya kemudian ditambahkan ke dalam *custom lexicon* dengan bobot sentimen yang sesuai, yang nantinya akan berkontribusi pada perhitungan skor sentimen awal setiap komentar.

Setiap proses pelabelan awal secara *semi-otomatis* menggunakan lexicon tersebut, setiap label yang dihasilkan tidak serta merta langsung digunakan. Sebaliknya, label-label tersebut melalui tahapan validasi dan akurasi manual yang cermat oleh peneliti. Proses kurasi ini bertujuan untuk meminimalkan bias dan kesalahan yang mungkin timbul dari pelabelan otomatis, terutama pada komentar yang ambigu atau memiliki konteks kompleks. Pada tahap kurasi inilah, komentar dengan polaritas ambigu yang mungkin salah dikategorikan oleh sistem *lexicon-based* akan diperiksa ulang dan diperbaiki secara manual, memastikan bahwa label akhir mencerminkan interpretasi sentimen yang lebih akurat sesuai judgment peneliti. Selain itu, kurasi manual juga memastikan konsisten interpretasi sentimen di seluruh dataset, sehingga meningkatkan kualitas dataset secara keseluruhan dan menjadikan lebih representatif secara andal sebagai “*ground truth”* untuk pelatihan model . Dataset final yang telah divalidasi dan dilabeli secara akurat inilah yang selanjutnya akan menjadi fondasi bagi tahapan pemisahan data (*data splitting)* menjadi *training set* dan *testing set* untuk model *deep learning* IndoBERT. Penempatan tahapan pelabelan ini setelah *data collection* dan *preprocessing* namun sebelum *data splitting* sangat strategis, guna memastikan bahwa model dilatih dan di evaluasi menggunakan data yang telah dibersihkan dan dianotasi dengan benar,

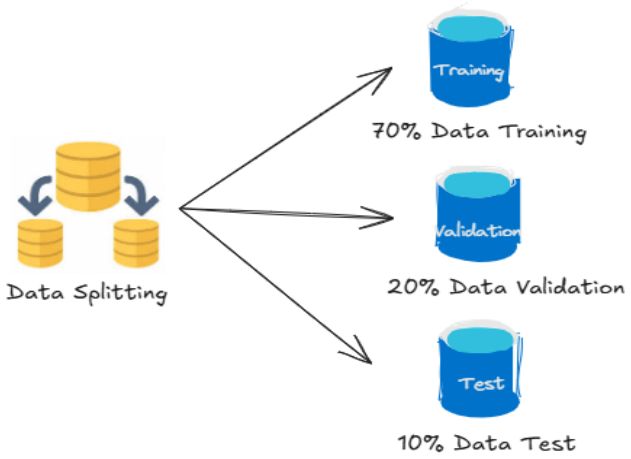
memaksimalkan potensi model dalam mempelajari pola sentimen secara optimal dan akurat pada data baru.

3.6 Tahapan Split

Dalam penelitian ini, analisis sentimen terhadap persepsi publik mengenai kepemimpinan Presiden Prabowo sentimen dilakukan dengan mengklasifikasikan komentar-komentar dari platform X dan Youtube menggunakan pendekatan *deep learning* berbasis model IndoBERT. IndoBERT merupakan model bahasa berbasis arsitektur *transformer* yang telah dilatih secara khusus memahami konteks dalam bahasa Indonesia, sehingga dinilai sesuai untuk digunakan dalam klasifikasi sentimen dan topik komentar publik.

Salah satu langkah krusial dalam proses pelatihan model adalah pembagian dataset menjadi tiga subset, yaitu data latih(*training set*), data validasi (*validation set*), dan data uji (*test set*). Dalam penelitian ini, dataset dibagi dengan proporsi 70%, untuk data latih, 20% untuk data validasi, dan 10% untuk data uji. Proporsi ini dirancang untuk memastikan bahwa model memperoleh data yang cukup untuk belajar pola-pola dari data (melalui training set), serta dapat dievaluasi secara teratur selama proses pelatihan(melalui validation set), dan akhirnya di uji performanya terhadap data yang benar-benar belum pernah dilihat sebelumnya (melalui test set).

Pembagian dilakukan menggunakan teknik *stratified splitting,* yaitu metode pembagian data yang mempertahankan proporsi distribusi label pada setiap subset. Teknik ini digunakan baik untuk label sentimen maupun topik agar setiap kelas terwakili secara seimbang dalam ketiga bagian data. Strategi ini penting untuk mencegah terjadinya bias pelatihan akibat ketidakseimbangan kelas, serta menjamin bahwa evaluasi model berlangsung secara adil terhadap seluruh label.



Gambar 3.24 Diagram Alur Split

Penggunaan data validasi sangat esensial dalam penelitian ini karena model dilatih menggunakan pendekatan *hyperparameter tuning,* yaitu proses pencarian kombinasi parameter pelatihan terbaik, seperti *learning rate,* jumlah epoch, dan *weight decay*. Selama proses pelatihan ,data validasi berfungsi sebagai pembelajaran langsung. Dengan demikian, model dapat menyesuaikan parameter yang digunakan dan menghindari *overfitting* terhadap data latih.

Sementara itu, data uji digunakan hanya satu kali di akhir proses untuk mengevaluasi performa akhir model. Evaluasi ini dilakukan untuk mengukur sejauh mana model yang telah di tuning dapat menggeneralisasikan hasilnya terhadap data baru yang tidak pernah dilibatkan dalam proses pelatihan maupun validasi, Dengan demikian, data uji memberikan gambaran objektif mengenai kemampuan model dalam mengklasifikasikan sentimen dan topik secara nyata.

3.7 Tahapan Encoded

Pada tahap ini, dilakukan konversi label yang masih berbentuk teks menjadi nilai numerik agar dapat digunakan dalam pelatihan model. Dua jenis label yang dikonversi adalah label sentimen dan label topik. Proses ini bertujuan agar model dapat mengenali setiap kelas secara terstruktur, karena algoritma *deep learning* seperti IndoBERT hanya dapat menerima input target dalam bentuk angka, bukan teks.

Untuk label sentimen, terhadap tiga kategori utama yang diklasifikasikan, yaitu "negatif", "netral", dan "positif". Masing-masing kategori tersebut diubah menjadi angka 0,1,dan 2. Konversi ini dilakukan dengan membuat pemetaan nilai secara manual berdasarkan daftar kategori yang tersedia. Sementara itu, untuk label topik, dilakukan pemetaan terhadap empat jenis topik utama yang menjadi fokus analisis dalam penelitian ini, yaitu "Danantara", "PPN 12%", "Makan Bergizi Gratis" , "Efisiensi Anggaran". Setiap topik diberikan representasi angka yang berbeda, yaitu dari angka 0 dan 3 secara berurutan. Hasil dari proses ini adalah dua kolom baru pada dataset yang merepresentasikan nilai label dalam bentuk numerik.

Agar sistem dapat digunakan kembali saat proses prediksi data baru (*inferensi*), struktur pemetaan label yang sudah dibuat disimpan dalam bentuk file khusus menggunakan metode serialisasi. Dengan cara ini ketika ada data komentar baru yang akan diklasifikasikan, sistem dapat menggunakan struktur pemetaan yang sama seperti saat pelatihan, sehingga menjamin konsistensi output.

3.8 Tahapan SMOTE

Dataset dalam penelitian ini dibagi menjadi tiga yaitu dengan proporsi 70% untuk data latih (*training set*), 20% untuk data validasi (*validation set*), dan 10% untuk data uji (*testing set*). Pembagian dilakukan menggunakan *stratifikasi* split untuk menjaga agar distribusi label sentimen dan topik tetap seimbang di setiap pembagian. Teknik *stratifikasi* ini penting untuk memastikan bahwa setiap subset tetap merepresentasikan proporsi kelas yang menjadi lebih representatif dan adil.

Hasil analisis awal terhadap distribusi label sentimen menunjukkan bahwa data memiliki ketidakseimbangan kelas, di mana kelas negatif memiliki jumlah yang dominan dibandingkan kelas lainnya. Ketidakseimbangan ini berpotensi menyebabkan model menjadi bias terhadap kelas mayoritas, serta mengabaikan representasi kelas minoritas dalam proses pembelajaran.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE), yaitu metode pembangkitan data sintetik untuk menyeimbangkan jumlah sampel antar kelas. Penerapan SMOTE

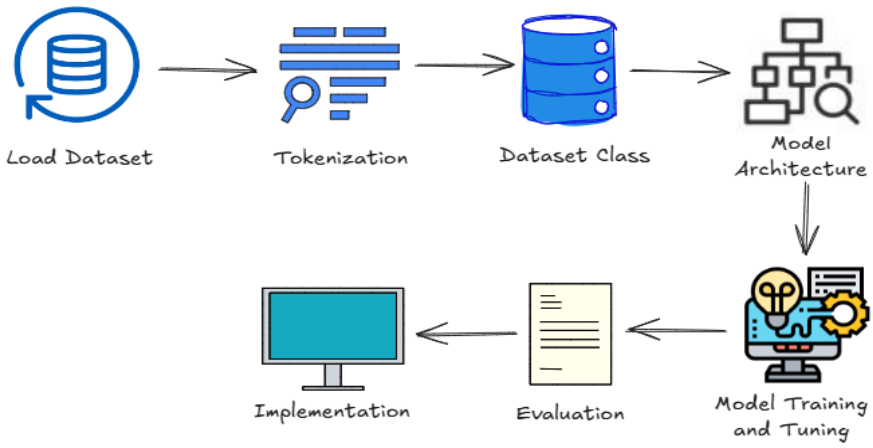
dilakukan secara khusus hanya pada data latih, yaitu 70% dari keseluruhan data. Proses oversampling ini tidak diterapkan pada data validasi maupun data uji, agar hasil evaluasi model tetap objektif dan tidak terpengaruh oleh data yang dihasilkan secara tidak alami(buatan). Dengan demikian, model dilatih menggunakan data latih yang telah seimbang secara kelas, sedangkan validasi dan pengujian dilakukan menggunakan data asli untuk memastikan performa model terhadap data dunia nyata.

3.9 Tahapan Modelling

Pada penelitian ini dilakukan tahapan pemodelan dengan menggunakan metode BERT (Bidirectional Encoder Representations from Transformers) untuk analisis sentimen. Pemilihan BERT didasarkan pada kemampuannya dalam memahami konteks kalimat secara dua arah, sehingga model ini lebih akurat dalam menangkap makna kata dalam konteks analisis sentimen. Berbagai penelitian telah menunjukkan bahwa BERT mampu mengungguli model klasik seperti SVM, Naive Bayes, bahkan LSTM dalam tugas-tugas NLP, termasuk klasifikasi teks dan analisis sentimen.

Keputusan ini juga didukung oleh jurnal dari Google AI oleh Devlin et al. (2018), serta studi lanjutan seperti Sun et al. (2019) yang menunjukkan bahwa fine-tuning BERT menghasilkan performa yang sangat baik pada klasifikasi teks. Terlebih lagi, karena analisis ini difokuskan pada program kerja Presiden Prabowo Subianto yang menggunakan Bahasa Indonesia, maka digunakan IndoBERT yang secara khusus dilatih pada data berbahasa Indonesia. IndoBERT memiliki keunggulan dalam memahami konteks lokal, termasuk penggunaan bahasa sehari-hari, slang, serta

struktur kalimat khas Indonesia, sehingga lebih efektif dibandingkan BERT versi umum yang dilatih untuk Bahasa Inggris [21].



Gambar 3.25 Tahapan *Modelling*

3.9.1 Tahapan Persiapan Dataset

Tahap awal dalam *modelling* adalah dimulai dengan proses memuat dataset yang telah dipersiapkan sebelumnya. Dataset ini merupakan hasil dari proses *preprocessing* yang mencakup berbagai langkah pembersihan teks. Proses ini dilakukan untuk memastikan bahwa teks masukan bersih, konsisten, dan sesuai dengan format yang dibutuhkan oleh model. Dataset yang digunakan berisi komentar publik dalam bahasa Indonesia yang telah melalui tahapan preprocessing data. Data memiliki tiga kolom utama yaitu **final_text, label, dan topik.** Label sentimen dan topik kemudian dikodekan ke dalam bentuk numerik menggunakan *LabelEncoder* dari *scikit-learn* agar dapat digunakan oleh model klasifikasi.

3.9.2 Tahapan Tokenization dan Word Embedding dengan IndoBERT

Tokenisasi merupakan tahap awal dalam pemrosesan teks yang bertujuan untuk mengubah data teks mentah menjadi bentuk yang dapat dimengerti oleh model. Pada penelitian ini, proses tokenisasi dilakukan menggunakan tokenizer dari model IndoBERT (*indobenchmark/indobert-base-p1*). Tokenizer ini bertugas memecah teks komentar menjadi token-token individual, dan mengubahnya menjadi representasi numerik(*input_ids*) serta menghasilkan *attention_mask* untuk

menandai posisi token penting dan padding. Token ini telah dikodekan ini kemudian digunakan sebagai input utama ke dalam model *transformer.*

Setelah proses tokenisasi selesai, IndoBERT secara otomatis melakukan proses word embedding. Word embedding adalah proses konversi setiap token menjadi vektor berdimensi tetap mempresentasikan makna semantik dari token tersebut. Dalam model IndoBERT, embedding(representasi kata), *position embeddings* (menandai posisi token dalam urutan kalimat), *segment embeddings* (menandai hubungan antar segmen teks).

Proses embedding ini menghasilkan representasi vektorial yang kontekstual, artinya makna kata disesuaikan dengan konteks kalimatnya. Representasi inilah yang digunakan oleh lapisan-lapisan transformer selanjutnya dalam IndoBERT untuk mempelajari hubungan antar token dan menghasilkan output prediksi yang relevan. Dengan demikian, tahapan tokenisasi dan embedding menjadi pondasi penting dalam menghasilkan performa model klasifikasi yang baik.

3.9.3 Tahapan Arsitektur Model *Multi-Task* IndoBERT

Model yang diimplementasikan dalam penelitian ini merupakan hasil modifikasi dari arsitektur dasar IndoBERT, sebuah model bahasa berbasis *transformer* yang dilatih secara khusus untuk memahami teks dalam Bahasa Indonesia. Untuk mengakomodasi kebutuhan klasifikasi ganda yaitu analisis sentimen dan klasifikasi topik pada setiap komentar, model dikembangkan dalam bentuk *multi-task learning* di mana dua tugas klasifikasi dilakukan secara bersamaan dalam satu jaringan.

Secara arsitektural, model terdiri atas blok IndoBERT pretrained yang berfungsi sebagai *feature extraction* utama. Komponen ini menghasilkan representasi kontekstual dari input teks dalam bentuk *poller_output,* yaitu vektor yang merangkum informasi dari seluruh token dalam satu kalimat. Output tersebut kemudian diarahkan ke dua *fully connected layer (*linear layer) yang berfungsi sebagai *output head.* Output head pertama bertugas melakukan klasifikasi sentimen terhadap tiga kelas(negatif, positif, netral), sementara output head kedua digunakan

untuk klasifikasi topik komentar ke dalam empat kategori, yaitu Danantara, PPN 12%, Makan Bergizi Gratis, dan Efisiensi Anggaran.

Setiap output head menggunakan fungsi aktivasi *Softmax* untuk mengubah hasil output linear menjadi probabilitas dari masing-masing kelas. Model dilatih dengan menggunakan dua fungsi *loss* yang independen, yaitu *CrossEntropyLoss* untuk masing-masing tugas. Total *loss* yang digunakan dalam optimasi model dihitung sebagai penjumlahan dari kedua nilai loss tersebut. Pendekatan ini memungkinkan model untuk belajar dari dua target klasifikasi sekaligus dalam satu alur pelatihan terpadu, sehingga efisien secara komputasi dan mampu menangkap hubungan semantik yang mendalam antara teks dan target kelas.

Proses pelatihan menggunakan *optimizer* AdamW, yang secara khusus dirancang untuk stabilitas pada model berbasis *transformer,* serta didukung dengan pengaturan *hyperparameter* seperti *learning_rate, batch size, weight decay,* dan jumlah *epoch.* Model diimplementasikan menggunakan library *PyToch* dan *Hugging Face Transformers*, dengan dukungan GPU untuk mempercepat proses *training.* Setelah pelatihan selesa, model yang telah terbentuk tidak langsung digunakan bentuk utuh melainkan disimpan terlebih dahulu dalam format yang hanya memuat parameter-parameter hasil pelatihan dikenal sebagai (*state_dict*) tanpa harus mengulang proses pelatihan dari awal. Pada tahapan implementasi, model ini kemudian diaktifkan kembali dan digunakan untuk mengklasifikasi secara langsung terhadap input komentar yang diberikan oleh pengguna. Dengan pendekatan ini, sistem dapat bekerja secara efisien dan *responsif* karena model sudah dalam kondisi siap pakai.

3.9.4 Tahapan Pelatihan Model

Pada tahap ini, proses pelatihan model dilakukan untuk membentuk kemampuan model dalam mengenali dan mengklasifikasi komentar publik ke dalam kategori sentimen dan topik secara bersamaan. Model yang digunakan adalah IndoBERT, yang dimodifikasi menjadi arsitektur *multi-task* untuk menghasilkan dua output satu untuk klasifikasi sentimen (negatif, netral, positif), dan satu lagi untuk klasifikasi topik komentar terhadap program kerja Presiden Prabowo Subianto.

Proses pelatihan dilakukan dengan menggunakan algoritma optimasi *AdamW,* yang dikenal efisien dan stabil untuk model berbasis *transformer* seperti IndoBERT. Selama pelatihan, model dijalankan dalam beberapa *epoch*, yaitu putaran penuh di mana seluruh data latih digunakan untuk memperbarui bobot model. Pada setiap *epoch,* model menerima input berupa hasil tokenisasi (*input_ids* dan *attention_mask*) dan menghasilkan dua prediksi. Nilai prediksi ini kemudian dibandingkan dengan label sebenarnya menggunakan fungsi *CrossEntropyLoss* untuk masing-masing tugas. Nilai total *loss* dihitung dengan menggunakan kedua *loss* dari prediksi sentimen dan topik.

Setelah proses *forward pass,* model melakukan *backpropagation* untuk memperbarui bobot parameter berdasarkan nilai *loss* yang diperoleh. Proses ini terus berulang untuk setiap *batch* data hingga seluruh *epoch* selesai. Selain itu, setiap *epoch,* model langsung dievaluasi menggunakan data validasi. Evaluasi ini dilakukan untuk mengetahui apakah model mengalami peningkatan akurasi atau justru mengalami *overfitting.* Karena itulah, metrik seperti akurasi dan nilai *loss* dicatat pada setiap *epoch* untuk memantau performa model selama pelatihan.

Model yang menghasilkan kinerja terbaik berdasarkan evaluasi terhadap data validasi kemudian disimpan. Model ini akan digunakan kembali pada tahapan implementasi di Bab 4, dimana pengguna dapat menguji komentar secara langsung melalui antarmuka input. Dengan pendekatan ini, seluruh tahapan pelatihan tidak hanya fokus pada akurasi tinggi, tetapi juga menjamin bahwa model yang dikembangkan mampu diterapkan secara nyata pada skenario dunia nyata, sesuai dengan tujuan awal penelitian ini.

3.9.5 Tahapan *Hyperparameter Tuning*

Pada tahapan ini, dilakukan proses *fine-tuning* terhadap model IndoBERT dengan tujuan memperoleh konfigurasi *hyperparameter* yang memberikan performa paling optimal pada tugas klasifikasi sentimen dan topik. *Hyperparameter* merupakan parameter pelatihan yang nilainya ditentukan sebelum proses pelatihan di mulai, seperti *learning rate*, *batch size,* jumlah epoch dan *weight decay.* Penelitian ini

menggunakna pendekatan manual *grid search* untuk mengeksplorasi beberapa kombinasi nilai yang umum digunakan dalam penelitian model berbasis *transformer.*

Beberapa kombinasi yang diuji melibatkan nilai *learning rate* sebesar 1e-5, 2e-5, dan 3e-5, dengan *batch size* tetap sebesar 16, jumlah *epoch* maksimal 5, serta nilai *weight decay* 0,01. Evaluasi dilakukan berdasarkan akurasi validasi untuk dua keluaran model, yakni klasifikasi sentimen(dengan tiga label) dan klasifikasi topik(dengan empat label).

Hasil pelatihan menunjukkan bahwa seluruh konfigurasi memberikan *tren* penurunan *training loss* yang konsisten dan akurasi validasi yang stabil. Secara khusus, kombinasi *learning rate* 3e-5 memberikan hasil terbaik. Pada epoch ke 4 model mencatat *training loss* sebesar 0,9780, serta akurasi validasi sebesar 73,95% untuk klasifikasi sentimen dan 81,38% untuk klasifikasi topik. Jika di rata-rata, nilai akurasi validasi mencapai 77,66% yang menjadi nilai tinggi dibandingkan konfigurasi lainnya. Oleh karena itu, konfigurasi ini dipilih sebagai *parameter* final yang akan digunakan pada tahap implementasi dan pengujian model di bagian selanjutnya.

Pemilihan *epoch* terbaik ini tidak selalu diambil dari *epoch* terakhir, melainkan dari *epoch* yang menunjukkan performa validasi tertinggi selama proses pelatihan berlangsung. Strategi ini bertujuan untuk mencegah *overfitting* dan memastikan bahwa model yang di simpan merupakan model dengan generalisasi terbaik pada data validasi. Dengan demikian, proses tuning ini menjadi tahap penting dalam memastikan efektivitas penerapan model IndoBERT terhadap data komentar publik yang dianalisis dalam penelitian ini.

3.9.6 Evaluasi Model

Tahap evaluasi merupakan langkah akhir dalam proses pembangunan model BERT sebelum digunakan secara penuh. Pada tahap ini, performa model yang telah di *fine-tune* diuji menggunakan data uji (*testing set*) yang tidak digunakan saat pelatihan.

Tujuannya adalah untuk mengukur seberapa baik model dalam menggeneralisasi tugas klasifikasi sentimen dan topik pada data baru.

Tahapan evaluasi model meliputi penilaian seberapa baik model melakukan klasifikasi sentimen dengan menggunakan *confusion matrix* sebagai instrumen utama. Dari *confusion matrix,* dilakukan perhitungan metrik evaluasi seperti *accuracy, precision, recall,* dan *F1-Score* untuk mengukur performa model secara *kuantitatif.* Selain itu, visualisasi hasil evaluasi juga dilakukan untuk mempermudah validasi dan interpretasi performa model, misalnya dengan membuat tampilan website menggunakan framework tertentu seperti *Streamline*[37]. Dengan mengikuti tahapan ini, evaluasi model menjadi sistematis dan dapat memastikan model yang dihasilkan memiliki performa optimal sesuai tujuan aplikasi [48].

3.9.7 Tahapan Implementasi

Setelah proses *fine-tuning* selesai dilakukan pada model BERT, tahap selanjutnya dalam penelitian ini adalah implementasi sistem. Pada tahapan ini, model yang telah dilatih akan digunakan untuk melakukan klasifikasi secara langsung terhadap masukan dari pengguna. Ketika pengguna memasukkan sebuah kalimat atau komentar, sistem akan secara otomatis mendeteksi jenis sentimen dari kalimat tersebut, apakah termasuk positif, negatif, atau netral. Selain itu, model juga dirancang untuk mengidentifikasi topik atau program pemerintah yang dibicarakan dalam komentar tersebut. Program yang dimaksud mencakup beberapa isu kebijakan publik seperti “PPN 12%,” “Makan Bergizi Gratis,” “Efisiensi Anggaran,” dan “Danantara.” Dengan demikian, sistem tidak hanya melakukan analisis sentimen, tetapi juga analisis topik secara bersamaan (*multi-task classification*), yang memungkinkan pemetaan opini publik terhadap isu-isu tertentu secara lebih spesifik dan mendalam.

BAB IV

IMPLEMENTASI

Pada bab ini akan dijelaskan tahapan implementasi yang dilakukan berdasarkan hasil analisis dan perancangan pada bab sebelumnya. Bab ini juga membahas lingkungan implementasi, batasan implementasi, implementasi pengumpulan data, *preprocessing* data, penentuan sentimen, serta klasifikasi. Selain itu, bab ini menjelaskan proses pengolahan data dan pembangunan model untuk memprediksi sentimen suatu komentar terhadap keempat aspek kinerja Prabowo dengan menggunakan model BERT.

4.1 Lingkungan Implementasi

Dalam analisis kebutuhan yang mendukung penelitian analisis sentimen terhadap kepemimpinan Prabowo Subianto melalui platform X dan YouTube dengan menggunakan metode BERT (*Bidirectional Encoder Representations*), diperlukan sejumlah perangkat dan lingkungan pendukung untuk keperluan implementasi penelitian ini seperti perangkat keras (*hardware*) dan perangkat lunak (*Software*).

4.2 Pengolahan Data

Tahapan pengolahan data terdiri dari beberapa tahap, yaitu pengumpulan data, *preprocessing, labeling,* data split, encoded, smote, dan *modeling.* Untuk memberikan gambaran yang lebih jelas, subbab berikut akan menampilkan *pseudocode* dari masing-masing tahapan tersebut.

4.2.1 Implementasi Pengumpulan Data

Implementasi pengumpulan data dilakukan untuk memperoleh data yang digunakan dalam proses penelitian ini. Data yang dikumpulkan kemudian disimpan dalam format file CSV agar memudahkan proses pengolahan dan analisis selanjutnya. Pada penelitian ini, data diperoleh dari dua platform, yaitu X dan YouTube. Pengumpulan data dari platform X dilakukan dengan metode *crawling*, sedangkan dari Youtube menggunakan metode *scraping*. Berikut merupakan implementasi pengumpulan data pada platform X menggunakan metode *crawling.*

```
MULAI

  1. Hubungkan ke media penyimpanan Google Drive
  2. Tetapkan token autentikasi twitter
Jika token_autentikasi_twitter kosong maka tampilan "Peringatan
token tidak tersedia. Proses crawling tidak dapat dilakukan."
  Selain itu
  Tampilan Token twitter berhasil dimuat"
  3. Instalasi dependensi tweet-hervest dan Node. Js
     - Install node.js dan dependensinya
     - Install tweet-hervest melalui npm
  4. Daftar Topik kebijakan yang dianalisis
     Tetapkan daftar_topik =[
      {"nama" : "PPN12%", "Kata_kunci" : "Kata 1 atau Kata 2…"},
      {"nama": "MBG", "Kata_kunci" : "..."},
      {"nama" :"Danantara" ,"Kata_kunci" : "..."},
      {"nama" : "Efisiensi_Anggaran", "kata_kunci" : "..."}
      ]
  5. Simpan Hasil crawling menggunakan .csv

SELESAI
```

Berikut hasil data *crawling* komentar dari media sosial X.

1	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_scre	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username
2	1883164715654001085	Sat Jan 25 14:47:10 +0000 2025	4143	Di Hyderabad House New Delhi saya bersama Perdana Menteri @narendramodi menyaksikan penandatanganan sejumlah nota kesepahaman yang memperluas kemitraan strategis Indonesia dan India. Kerja sama ini mencakup berbagai bidang penting seperti ekonomi kesehatan keamanan maritim https://t.co/YwTjvWuE8	1883164715654	https://pbs.twimg.com/media/GU\	in	Indonesia	15	164	391	https://x.com/pra	40580714	prabowo
3	1885349408293949938	Fri Jan 31 15:28:21 +0000 2025	1038	Saya menerima kunjungan Menteri Pertahanan Prancis @SebLecornu dalam upaya memperkuat kerja sama pertahanan antara Indonesia dan Prancis. Bersama Menteri Pertahanan Sjafrie Sjamsoeddin saya menegaskan komitmen Indonesia untuk membangun industri pertahanan yang mandiri dan https://t.co/k3cVT1uVyY	1885349408293	https://pbs.twimg.com/media/Gipl	in	Indonesia	10	170	168	https://x.com/pra	40580714	prabowo
4	1882825683138807412	Fri Jan 24 16:19:58 +0000 2025	3920	Bapak Presiden @prabowo yth sebaiknya TNI dan POLRI menolak hadiah dari PT AGS utk membangun fasilitas di PIK-2 karena : 1) ini tmsk gratifikasi/sogokan 2) tanah tsb adalah tanah rakyat yg dibeli murah https://t.co/gZeCnQfiI5	1882825683138	https://pbs.twimg.com/media/GiE	in		68	351	1979	https://x.com/ms	1117990249806	msaid_didu
5				Sinergitas TNI-Polri menjadi pilar fundamental dalam mewujudkan misi Asta Cita yang diusung Presiden RI Jenderal TNI (Purn.) Prabowo Subianto. Adanya											

Gambar 4.1 Hasil *Crawling* Data

Sementara itu, pada pengumpulan data dari platform YouTube dalam penelitian ini dilakukan dengan metode *scraping* tanpa melalui penulisan kode secara manual. Proses dimulai dengan mendaftarkan API Key sebagai otorisasi akses ke data YouTube. Setelah itu, masukkan tautan video yang relevan ke dalam sistem. Langkah selanjutnya adalah memasukkan API Key yang telah didaftarkan sebelumnya untuk memungkinkan koneksi dengan server YouTube. Setelah koneksi berhasil, sistem akan melakukan *scraping* terhadap komentar pada video yang ditautkan. Data hasil *scraping* kemudian disimpan dalam format yang dapat

diolah lebih lanjut untuk keperluan analisis. Berikut hasil data *scraping* komentar dari media sosial YouTube.

1	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_scre	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username
2	1883164715654001085	Sat Jan 25 14:47:10 +0000 2025	4143	Di Hyderabad House New Delhi saya bersama Perdana Menteri @narendramodi menyaksikan penandatanganan sejumlah nota kesepahaman yang memperluas kemitraan strategis Indonesia dan India. Kerja sama ini mencakup berbagai bidang penting seperti ekonomi kesehatan keamanan maritim https://t.co/YwTjvWuE8	1883164715654	https://pbs.twimg.com/media/GU7	in	Indonesia	15	164	391	https://x.com/pra	40580714	prabowo
3	1885349408293949938	Fri Jan 31 15:28:21 +0000 2025	1036	Saya menerima kunjungan Menteri Pertahanan Prancis @SebLecornu dalam upaya memperkuat kerja sama pertahanan antara Indonesia dan Prancis. Bersama Menteri Pertahanan Sjafrie Sjamsoeddin saya menegaskan kombinasi Indonesia untuk membangun industri pertahanan yang mandiri dan https://t.co/k3dVT1uVyY	1885349408293	https://pbs.twimg.com/media/GIol	in	Indonesia	10	170	168	https://x.com/pra	40580714	prabowo
4	1882825683136807412	Fri Jan 24 16:19:58 +0000 2025	3920	Bapak Presiden @prabowo yth sebaiknya TNI dan POLRI menolak hadiah dari PT AGS utk membangun fasilitas di PIK-2 karena - 1) ini tmsk gratifikasi/sogokan 2) tanah tsb adalah tanah rakyat yg dibeli murah https://t.co/gZeChQfli5	1882825683136	https://pbs.twimg.com/media/GiE	in		68	351	1979	https://x.com/ms	1117990246806	msaid_didu
5				Sinergitas TNI-Polri menjadi pilar fundamental dalam mewujudkan misi Asta Cita yang diusung Presiden RI Jenderal TNI (Purn.) Prabowo Subianto. Adanya											

Gambar 4.2 Hasil *Scraping* Data

4.2.2 Implementasi Preprocessing

Implementasi *preprocessing* dilakukan untuk membersihkan dan menyiapkan data teks sebelum digunakan dalam pemodelan. Tahapan ini mencakup pembersihan karakter tidak relevan, normalisasi kata tidak baku, tokenisasi, penghapusan *stopword,* dan *stemming.* Pseudocode di bawah ini menggambarkan proses *preprocessing* yang diterapkan dalam penelitian ini.

```
MULAI

// import library yang diperlukan untuk NLP, pemrosesan data dan
normalisasi

BACA file "slang.dict.csv" dan ubah dalam bentuk dictionary:
slang_dict <- {slang : formal}

DEFINISIKAN fungsi angka_ke_kata(text):
Ganti semua angka dalam teks menjadi kata dalam Bahasa Indonesia

DEFINISIKAN fungsi cleaningText(text):
   -  Konversikan angka ke kata
   -  Hapus mention @
   -  Hapus hashtag
   -  Hapus retweet(RT)
```

```
   -  Hapus link
   -  Hapus karakter non-alfanumerik kecuali emoji
   -  Hapus newline dan tanda baca
DEFINISIKAN fungsi slangword(text)
   -  Pisah teks berdasar spasi (tokenisasi sederhana)

DEFINISIKAN fungsi filteringText(words):
   -  Buat daftar stopword Bahasa Indonesia dan Inggris
   -  Tambahkan daftar stopword tambahan
   -  Hapus semua token yang termasuk dalam stopwords
DEFINISIKAN fungsi stemmingText(token_list, original_text):
   -  Gunakan Sastrawi Stemmer
   -  Stem setiap token
DEFINISIKAN fungsi toSentence(tokens):
   -  Gabungkan daftar token menjadi satu kalimat
DEFINISIKAN fungsi full_normalization_pipeline(text)
   -  Simpan teks asli
   -  Bersihkan teks menggunakan cleaningText
   -  Normalisasi slang
   -  Tokenisasi teks
   -  Hapus stopwords
   -  Lakukan stemming
   -  Gabungkan hasil akhir
   -  Tabel Hasil akhir Cleaned, Tokenizer, Filtered, Stemmed,
      Final_Text
BACA dataset "Transelated(Kinerja).csv)
CEK apakah kolom "full_text" dan "translated_text" tersedia
JIKA TIDAK tersedia THEN
   Tampilkan pesan eror dan hentikan proses
Lakukan preporocessing pada setiap baris kolom " full_text":
   APPLY full_normalization_pipeline ke setiap baris
Gabungkan hasil preprocessing ke dataset asli
Simpan Hasil preprocessing ke file baru " Preprocessing_(Kinerja)"

SELESAI
```

Hasil dari *preprocessing* dapat dilihat pada Tabel 4.1

Tabel 4.1 Hasil Preprocessing

Komentar	Translate	Normalisasi	Tokenisasi	Stopword	Stemming	Hasil
Program *free lunch* ini sangat membantu anak-anak di daerah terpencil	Program makan siang gratis ini sangat membantu anak-anak di daerah terpencil	program makan siang gratis ini sangat membantu anak-anak di daerah terpencil	['program', 'makan', 'siang', 'gratis', 'ini', 'sangat', 'membantu', 'anak-anak', 'di', 'daerah', 'terpencil']	['program', 'makan', 'siang', 'gratis', 'sangat', 'membantu', 'anak-anak', 'daerah', 'terpencil']	['program', 'makan', 'siang', 'gratis', 'bantu', 'anak', 'daerah', 'terpencil']	program makan siang gratis bantu anak daerah terpencil
Kenaikan *tax* jadi 12% makin memberatkan #rakyat kecil.	Kenaikan pajak jadi dua belas persen makin memberatkan rakyat kecil	kenaikan pajak jadi dua belas persen makin memberatkan masyarakat kecil	['kenaikan', 'pajak', 'jadi', 'dua', 'belas', 'persen', 'makin', 'memberatkan', 'rakyat', 'kecil']	['kenaikan', 'pajak', 'dua', 'belas', 'persen', 'makin', 'memberatkan', 'rakyat', 'kecil']	['naik', 'pajak', 'dua', 'belas', 'persen', 'makin', 'berat', 'rakyat', 'kecil']	naik pajak dua belas persen makin berat rakyat kecil
Aplikasi danantara ini menurut saya sangat *user-friendly* dan praktis	Aplikasi Danantara ini menurut saya sangat mudah digunakan dan praktis	aplikasi danantara ini menurut saya sangat mudah digunakan dan praktis	['aplikasi', 'danantara', 'ini', 'menurut', 'saya', 'sangat', 'mudah', 'digunakan', 'dan', 'praktis']	['aplikasi', 'danantara', 'menurut', 'saya', 'sangat', 'mudah', 'digunakan', 'praktis']	['aplikasi', 'danantara', 'mudah', 'guna', 'praktis']	aplikasi danantara mudah guna praktis
Pemerintah perlu audit agar *budget* makin efisien dan tepat sasaran.	Pemerintah perlu audit agar anggaran makin efisien dan tepat sasaran	pemerintah perlu audit agar anggaran makin efisien dan tepat sasaran	['pemerintah', 'perlu', 'audit', 'agar', 'anggaran', 'makin', 'efisien', 'dan', 'tepat', 'sasaran']	['pemerintah', 'perlu', 'audit', 'anggaran', 'makin', 'efisien', 'tepat', 'sasaran']	['pemerintah', 'perlu', 'audit', 'anggar', 'efisien', 'tepat', 'sasar']	pemerintah perlu audit anggar efisien tepat sasar

4.2.3 Implementasi Labeling

Tahapan *labeling* dilakukan untuk memberikan label sentimen (positif, negatif, atau netral) berdasarkan skor hasil perhitungan bobot kata dari kamus sentimen. Setiap komentar diberi nilai sentimen dan kemudian diklasifikasikan sesuai nilainya. *Pseudocode* di bawah ini menggambarkan proses *labeling* yang digunakan dalam penelitian ini.

```
MULAI

// Unduh tokenizer NLTK
Download nltk tokenizer (“punk”)

//Baca lexicon berisi kata dan bobot sentimen
BACA File “lexicon_dengan_bobot(kinerja).csv”
KONVERSI menjadi dictionary: lex_dict <-{kata: bobot}

//Baca dataset hasil preprocessing komentar
BACA file”Processing(kinerja).csv)
PILIH kolom “final_text” saja
HAPUS data kosong (NaN) dari kolom tersebut
UBAH semua teks menjadi tipe string

//Definisikan fungsi proses_sentimen(text)
//Hitung skor sentimen berdasarkan jumlah total bobot kata-kata
dalam lexicon
Fungsi proses_sentimen(text)
IF text bertipe string THEN
    AMBIL semua kata yang ada didalam lex_dict
    JUMLAHKAN bobot berdasarkan kata-kata tersebut
    Kembalikan skor total
ELSE
    Kembalikan 0
//Hitung skor sentimen untuk setiap komentar
Terapkan proses_sentimen pada setiap baris “final_text”
Simpan hasil ke dalam kolom baru : “Sentimen score”
//Definisikan fungsi label_dari_skor(skor)
//Menentukan label berdasarkan nilai skor
Fungsi label_dari_skor(skor):
    IF skor>0 THEN return “positif”
    ELSE IF skor < 0 THEN return “negatif”
    ELSE return “netral”

//Terapkan pelabelan berdasarkan skor
Terapkan label_dari_sentimen_score pada kolom “sentimen_score”
simpan hasil ke kolom baru: “label”
```

```
//Simpan hasil akhir ke fie csv baru
Simpan dataframe final ke file "Labeling(Kinerja).csv"

SELESAI
```

Hasil dari *labeling* dapat dilihat pada Tabel 4.2

Tabel 4.2 Hasil Labeling

Komentar	Skor	Label	Topik
danantara ini inovatif sih semoga bisa bantu umkm ke level global	7	Positif	Danantara
kenaikan ppn bikin harga makin mahal rakyat makin susah	-3	Negatif	PPN 12%
pemerintah harus lebih transparan soal efisiensi anggaran	0	Netral	Efisiensi Anggaran
mbg ini cuma janji politik doang realisasinya minim banget	-5	Negatif	Makan Bergizi Gratis

4.2.4 Implementasi Data Splitting

Pada tahap ini, data yang telah diproses dan dilabeli dibagi menjadi tiga bagian, yaitu data latih 70%, data validasi 20%, dan data uji 10%. Pembagian dilakukan menggunakan fungsi train_test_split dari pustaka Scikit-learn untuk memastikan distribusi data tetap seimbang dan hasil evaluasi model lebih akurat. *Pseudocode* di bawah ini menjelaskan proses pembagian data.

```
MULAI

   1. Baca file dataset utama dari file CSV
   2. Acak urutan baris data agar tidak berurutan berdasarkan pola
      tertentu
   -  Gunakan seed  acak(random_state) untuk memastikan hasil
      acakan konsisten
   3. Hitung jumlah data untuk memastikan masing-masing subset:
   -  Simpan sebagai variabel total
   4. Hitung jumlah data untuk masing-masing subset
   -  Data pelatihan(train) = 70% dari total
   -  Data validasi (val) = 20% dari total
   -  Data pengujian (test) = sisa dari total 10%
   5. Bagi dataset sesuai jumlah yang telah dihitung
   -  Ambil baris awal sampai batas train_size = simpan sebagai
      data train
```

```
   -  Ambil baris setelah train hingga batas val_size berikutnya
      = simpan sebagai data val
   -  Ambil sisa baris terakhir = simpan sebagai data test.
   6. Simpan ketiga subset ke dalam file csv terpisah:
         -  train_70.csv
         -  val_20.csv
         -  test_10.csv

SELESAI
```

Hasil dari data split dapat dilihat pada tabel berikut

Tabel 4.3 Hasil Data Split

Jenis Data	Jumlah Data	Persentase
Data Latih (*Train*)	8197	70%
Data Validasi (*Val*)	2342	20%
Data Testing (*Test*)	1171	10%
Total	11710	100%

4.2.5 Implementasi Mapping

Pada tahap ini, label sentimen dan topik yang semula bentuk teks diubah ke dalam format numerik agar dapat digunakan dalam proses pemodelan. Proses ini dilakukan dengan membuat kamus untuk memetakan setiap label ke angka. Hasil encoding kemudian disimpan ke dalam kolom baru pada dataset. *Pseudocode* untuk proses ini ditampilkan di bawah.

```
MULAI


    Import pustaka
   -  Gunakan pustaka pandas untuk manipulasi data
    Load Csv File
   -  Load file 'train_70.csv" ke dalam variabel train_df
   -
    Preporocessing
   -  Hapus baris yang memiliki nilai kosong (NaN)pada kolom
      'final_text', 'label', atau 'topik'


    Define mapping dictionary:
```

```
   -  Buat  kamus  'label_map'  untuk  mengubah  label  sentimen
      (negatif/positif/netral) menjadi angka (0/1/2)
   -  Buat kamus 'topik_map' untuk mengubah topik menjadi angka
      sesuai urutan berikut:
         - Danantara = 0
         - PPN 12% = 1
         - Makan Bergizi Gratis = 2
         - Efisiensi Anggaran = 3

      Apply mapping
   -  Tambahkan  kolom  'label_sentimen'  ke  train_df  berdasarkan
      mapping dari kolom 'label'
   -  Tambahkan  kolom  'label_topik'  ke  train_df  berdasarkan
      mapping dari kolom 'topik'

Output
   -  Tampilan kombinasi unik dari label awal dan hasil mapping
      label_sentimen
   -  Tampilkan kombinasi unik dari topik awal dan hasil mapping
      label_topik

SELESAI
```

4.2.6 Implementasi SMOTE

Penerapan SMOTE dilakukan untuk mengatasi ketidakseimbangan jumlah data pada label sentimen. Teknik ini menambahkan data sintetis berdasarkan fitur teks yang telah direpresentasikan dalam bentuk vektor. Berikut merupakan *pseudocode* untuk implementasi SMOTE.

```
MULAI

    Import pustaka
   -  Gunakan pustaka pandas untuk manipulasi data
    Load Csv File
   -  Load file 'train_70.csv" ke dalam variabel train_df
   -
    Preporocessing
   -  Hapus  baris  yang  memiliki  nilai  kosong  (NaN)pada  kolom
      'final_text', 'label', atau 'topik'
```

```
    Define mapping dictionary:
   -  Buat  kamus  'label_map'  untuk  mengubah  label  sentimen
      (negatif/positif/netral) menjadi angka (0/1/2)
   -  Buat kamus 'topik_map' untuk mengubah topik menjadi angka
      sesuai urutan berikut:
         - Danantara = 0
         - PPN 12% = 1
         - Makan Bergizi Gratis = 2
         - Efisiensi Anggaran = 3

      Apply mapping
   -  Tambahkan kolom 'label_sentimen' ke train_df berdasarkan
      mapping dari kolom 'label'
   -  Tambahkan  kolom  'label_topik'  ke  train_df  berdasarkan
      mapping dari kolom 'topik'

Output
   -  Tampilan kombinasi unik dari label awal dan hasil mapping
      label_sentimen
   -  Tampilkan kombinasi unik dari topik awal dan hasil mapping
      label_topik

SELESAI
```

4.2.7 Implementasi Modeling

Tahap *modeling* merupakan inti dari proses analisis sentimen, dimana data yang telah diproses sebelumnya digunakan untuk melatih dan menguji model klasifikasi. Tahapan *modeling* mencakup pembuatan dataset, tokenisasi, model *definition,* model *training and tuning,* serta evaluasi terhadap kinerja model. Setiap tahap dijelaskan melalui *pseudocode* pada subbagian berikut.

4.2.7.1 Implementasi Load Dataset

Pada tahapan ini, proses implementasi dimulai dengan menggunakan dataset yang telah melalui tahapan pra proses dan pembagian sebelumnya. Dataset yang digunakan terbagi menjadi dua bagian utama, yaitu pelatihan (*train*) dan data *val* (*validation*). Data pelatihan yang digunakan merupakan hasil dari proses

penyeimbangan distribusi kelas menggunakan teknik SMOTE, sehingga jumlah data untuk setiap kelas sentimen menjadi seimbang. Sementara itu, data validasi berfungsi untuk mengevaluasi kinerja model selama proses pelatihan berlangsung, guna menghindari terjadinya *overfitting.*

Kedua dataset tersebut disimpan dalam format .csv dan dimuat ke dalam lingkungan kerja menggunakan pustaka pandas. Setelah dimuat, data ini akan digunakan pada tahap-tahap selanjutnya, yaitu tokenisasi teks dan pelatihan model klasifikasi berbasis IndoBERT.

```
Mulai

   1. Siapkan pustaka untuk memproses data
         -  Gunakan pustaka pandas untuk membaca file csv
         -  Gunakan pustaka transformers jika diperlukan untuk
            tahapan tokenisasi selanjutnya
         -
SELESAI
```

4.2.7.2 Implementasi Tokenization

Tokenizer dari model IndoBERT dimuat menggunakan *Auto Tokenizer* dari *Hugging Face Transformers.* Tokenizer ini berfungsi untuk mengubah data teks menjadi format token yang dapat diproses oleh model BERT. Berikut merupakan *pseudocode* untuk implementasi tokenizer.

```
MULAI

// Tentukan nama model pre-trained yang akan digunakan
model_name "indobenchmark/indobert-base-p1"

// Load tokenizer berdasarkan model yang telah ditentukan
tokenizer AutoTokenizer.from_pretrained(model_name)

SELESAI
```

4.2.7.3 Implementasi Dataset Class

Sebelum data dapat digunakan dalam pelatihan model IndoBERT, perlu dilakukan proses *transformasi* dari bentuk data frame ke dalam format yang dapat dipahami

oleh model *deep learning.* Oleh karena itu, dibuatlah sebuah kelas *IndoBERTMultiDataset,* yang merupakan turunan dari *Dataset* milik *PyToch.* Kelas ini bertanggung jawab untuk menyiapkan input teks beserta label-nya menjadi tensor-tensor yang dibuat oleh model IndoBERT.

Pada metode ***__init__*,** objek dari dataset dibuat dengan menerima data dalam bentuk *DataFrame* (yang telah berisi kolom *final_text, label_sentimen* dan *label_topik*), serta *tokenizer* dari model IndoBERT. Data komentar disimpan dalam bentuk list pada atribut *self.texts,* sementara label sentimen dan label topik juga disimpan dalam bentuk list masing-masing.

Proses tokenisasi dilakukan pada setiap item ketika dipanggil melalui fungsi __getitem__. Di sini, komentar (teks) diubah menjadi representasi numerik menggunakan tokenizer IndoBERT, dengan panjang maksimum token (*max_len*) yang telah ditentukan, yaitu 128 token. Tokenisasi dilakukan dengan padding agar agar panjang input seragam, serta *truncation* agar tidak melebihi batas panjang maksimal model.

Selain input teks yang telah di tokenisasi (*input_ids*) dan *attention_mask,* fungsi ini juga mengembalikan dua buah label dalam bentuk tensor *PyTorch.* label_sentimen dan label_topik. Keduanya disiapkan sebagai label klasifikasi untuk dua tugas berbeda yang dilakukan secara bersamaan oleh model (*multi-task classification*).Dengan adanya kelas dataset ini, proses pelatihan model menjadi lebih efisien dan terstruktur, karena data yang diberikan ke model sudah terformat sesuai dengan kebutuhan arsitektur IndoBERT.

```
MULAI

Class IndoBERTMultiDataset inherits from Dataset:

    Function_init_(df, tokenizer, max_len=128):
   -  Simpan daftar teks dari kolom 'final_text'
   -  Simpan daftar label sentimen dari kolom 'label_sentimen'
   -  Simpan daftar label topik dari kolom 'label_topik'
   -  Simpan tokenizer dan nilai panjang maksimal tokenisasi
```

```
   Function __getitem__(index):
   -  Tokenisasi teks ke-index dengan padding dan turncation
   -  Kembalikan dictionary berisi:
      - input_ids: tensor token input
      - attention_mask: tensor attention mask
      - label_sentimen: tensor label sentimen (tipe long)
      - label_topik: tensor label topik (tipe long)

   Function _len_():
   -  Kembalikan jumlah total data (jumlah teks)

SELESAI
```

4.2.7.4 Implementasi Pelatihan dan Tuning

Proses pelatihan model dilakukan dengan memanfaatkan arsitektur IndoBERT yang dimodifikasi untuk tugas *multi-klasifikasi,* yaitu klasifikasi sentimen dan topik secara bersamaan. Model dilatih menggunakan algoritma optimasi *AdamW,* yang dikenal memiliki stabilitas dalam proses pembuatan bobot dan secara umum sangat sesuai untuk model *transformer* seperti IndoBERT.

Setiap epoch dalam proses pelatihan terdiri dari *forward pass* untuk menghasilkan prediksi dari data masukan, dilanjutkan dengan perhitungan total loss sebagai gabungan *loss* sentimen dan *loss* topik. Nilai total *loss* inilah yang digunakan dalam proses *backpropagation* untuk memperbarui parameter model. Evaluasi terhadap data validasi dilakukan pada akhir setiap *epoch* untuk memantau performa model serta menghindari *overfitting.*

Dalam pelatihan ini juga diterapkan strategi *early stopping* , yaitu menghentikan proses pelatihan secara otomatis apabila akurasi validasi tidak menunjukkan perbaikan dalam beberapa *epoch* berturut-turut. Strategi ini diterapkan guna memastikan bahwa model tidak dilatih secara berlebihan (*overfitting*) dan tetap efisien dalam penggunaan sumber daya komputasi. *Early stopping* diimplementasikan dengan *patience* sejumlah 5 epoch, artinya pelatihan akan berhenti jika tidak ada peningkatan akurasi selama 5 epoch berturut-turut.

Untuk memperoleh konfigurasi pelatihan terbaik, dilakukan proses *hyperparameter tuning* menggunakan metode *grid search* manual. Dalam proses ini, model dilatih dan di evaluasi berbagai kombinasi parameter pelatihan, antara lain

1. *Learning rate :* 1e-5 , 2e-5, dan 3e-5
2. *Batch size* : 16
3. *Epoch:* : 5
4. *Weight decay :* 0.01

Setiap kombinasi diuji dan dibandingkan berdasarkan akurasi validasi dari kedua tugas (*sentimen* dan topik). Hasil dari tuning menunjukkan bahwa model mencapai akurasi validasi terbaik pada kombinasi learning 3e- 5, batch size 16, dan *epoch* 5 untuk kedua tugas klasifikasi.

```
MULAI

// Inisialisasi device (GPU jika tersedia)

// Definisikan grid hyperparameter:
   -  learning_rate (lr)
   -  batch_size
   -  epochs
   -  weight_decay

// Siapkan Dataset:
   -  train_dataset = IndoBERTMultiDataset(train_df, tokenizer)
   -  val_dataset = IndoBERTMultiDataset(val_df, tokenizer)

// Inisialisasi list kosong untuk menyimpan hasil grid search

// FOR setiap kombinasi hyperparameter dalam param_grid:
   -  Inisialisasi model IndoBERTMultiOutput
   -  Buat DataLoader untuk data latih dan validasi
   -  Inisialisasi optimizer dan fungsi loss

   -  Set nilai awal:
      - best_val_acc = 0
      - best_model_state = None
```

```
      - patiance_counter = 0

      TRY:
          FOR setiap epoch:
               -  Set model ke training mode
               -  total_loss = 0

               FOR setiap batch dalam train_loader:
         -  Ambil  input_ids,  attention_mask,  label_sentimen,
            label_topik
         -  Forward  pass  ke  model  dapatkan  out_sentimen  &
            out_topik
         -  Hitung loss total = loss_sentimen + loss_topik
         -  Backpropagation dan update bobot
         -  Akumulasi total_loss

   -  Hitung rata-rata training loss dan cetak hasilnya

   -  Set model ke eval mode
   -  Inisialisasi  list  untuk  menyimpan  prediksi  dan  label
      sebenarnya

   FOR setiap batch dalam val_loader:
         -  Lakukan prediksi out_sent dan out_topik tanpa gradien
         -  Simpan prediksi dan label sebenarnya ke list

   -  Hitung akurasi:
      - acc_sent = akurasi prediksi sentimen
      - acc_topik = akurasi prediksi topik
      - avg_val_acc = rata-rata kedua akurasi

   -  IF avg_val_acc lebih baik dari best_val_acc:
      - Simpan model terbaik
      - Reset patience_counter
EXPECT jika pelatihan dihentikan manual:
   -  Cetak pesan "Training interrupted manually"

- Simpan model terbaik atau model terakhir
- Simpan tokenizer
```

```
- Tambahkan hasil eksperimen ke list 'results':
    - parameter, akurasi, path model

SELESAI
```

Hasil proses pelatihan dan tuning model direkap dalam sebuah tabel yang berisi kombinasi *hyperparameter* (*learning rate, batch size,epoch, weight decay*) dan hasil akurasi validasi untuk masing-masing tugas klasifikasi (*sentimen* dan topik). Tabel bertujuan untuk memudahkan pembaca melihat performa masing-masing konfigurasi dan menjadi dasar dalam menentukan parameter terbaik. Pencatatan hasil dilakukan secara langsung dari output pelatihan, kemudian disusun ke dalam bentuk tabel yaitu tabel hasil pelatihan dan tuning model.

Tabel 4.4 Hasil Pelatihan dan Tuning Model

Learning Rate	***Epoch* Terbaik**	***Val Accuracy Sentimen***	***Val Accuracy* Topik**	***Avg Val Accuracy***
1e-5	3	0.7114	0.8100	0.7607
2e-5	3	0.7323	0.8130	0.7727
3e-5	4	0.7395	0.8138	0.7766

Berdasarkan hasil pada Tabel 4.4, model dengan *learning rate* sebesar 3e-5, *batch size* 16, dan *weight decay* 0.01 menunjukkan rata-rata akurasi validasi tertinggi dalam klasifikasi sentimen dan topik. Nilai tersebut diperoleh pada *epoch* terbaik, yaitu *epoch* ke-4 perhitungan manual yang dapat dilakukan seperti berikut:

1. Akurasi sentimen dari *epoch* terbaik 73,95%
2. Akurasi topik dari *epoch* terbaik 81,38%
3. Rata-rata akurasi $\frac{73,95 + 81,38}{2}$ = 77,66%

Pemilihan nilai dari *epoch* terbaik didasarkan pada prinsip *early stopping* yaitu menghentikans pelatihan saat performa validasi sudah tidak lagi meningkatkan secara signifikan. Pendekatan ini umum digunakan menghindari *overfitting* dan memastikan model mempertahankan generalisasi terbaik terhadap data baru. Dengan demikian, nilai akurasi yang digunakan dalam tabel merupakan representasi paling optimal dari masing-masing model selama proses pelatihan berlangsung.Dari seluruh pencobaan kombinasi parameter, konfigurasi dengan *learning rate* 3e-5 inilah yang memberikan performa validasi terbaik secara

konsisten. Oleh karena itu, konfigurasi ini dipilih sebagai model akhir yang digunakan dalam penelitian ini.

4.2.7.5 Implementasi Evaluation

Setelah model selesai dilatih dan konfigurasi terbaik diperoleh dari hasil tuning, langkah selanjutnya adalah mengevaluasi kinerja model menggunakan data uji. Data uji ini sengaja dipisahkan sejak awal dan tidak dilibatkan dalam proses pelatihan atau validasi, agar hasil evaluasi benar-benar mencerminkan kemampuan model ketika dihadapkan pada data baru.

Evaluasi dilakukan untuk dua tugas klasifikasi sekaligus yaitu, klasifikasi sentimen(positif, negatif, netral) dan klasifikasi topik (Danantara, PPN 12%, Makan Bergizi Gratis, dan Efisiensi Anggaran. Pada tahap ini, komentar-komentar dari data uji diberikan sebagai input ke model, kemudian hasil prediksi model dibandingkan dengan label asli yang dimiliki oleh data tersebut.

Beberapa metrik yang digunakan dalam evaluasi adalah akurasi, presisi, *recall,* dan *f1-score.* Metrik ini membantu menunjukkan seberapa tepat model dalam mengenali kelas yang benar dan seberapa konsisten performanya. Selain itu, juga dibuat *confusion matrix* untuk melihat secara langsung bagaimana distribusi prediksi model terhadap masing-masing kelas, termasuk di mana saja model sering salah dan benar.

Hasil evaluasi yang diperoleh menunjukkan bahwa model mampu bekerja dengan baik dan cukup stabil, karena akurasi pada data validasi. Artinya, model tidak hanya baik saat dilatih, tetapi juga bisa digunakan untuk memprediksi data baru dengan hasil yang tetap akurat.

```
MULAI

// Import library:
   -  matplotlib & seaborn untuk visualisasi
   -  sklearn.metrics untuk evaluasi klasifikasi

// Hitung Confusion Matrix dan classification Report:
```

```
   -  Untuk label_sentimen:
      - cm_sent = confusionmatrix(true_sent, pred_sent)
      - report_sent = classification_report(true_sent, pred_sent,
      target_names=sentimen_labels)

   -  Untuk label_topik:
   -  cm_topik = confusion_matrix(true_topik, pred_topik)
   -  report_topik      =      classification_report(true_topik,
      pred_topik, target_names=topik_labels)

// Cetak hasil classification report:
   -  Tampilkan report_sent
   -  Tampilkan report_topik

// Definisikan fungsi plot_confusion_matrix(cm, labels, title):
   -  Buat plot heatmap ukuran 6x5
   -  Tampilkan   confusion   matrix   dengan   anotasi   angka
      (annot=True)
   -  Tambahkan judul, label sumbu X dan Y

// Visualisasikan Confusion Matrix:
   -  Plot cm_sent dengan label: ["Negatif", "Netral", "Positif"]
Plot  cm_topik  dengan  label:  ["Danantara",  "PPN  12%",  "Makan
Bergizi Gratis", "Efisiensi Anggaran"]

SELESAI
```

4.2.8 Implementasi Akhir

Tahap implementasi akhir merupakan puncak dari seluruh rangkaian penelitian, dimana model klasifikasi sentimen dan topik yang telah dilatih, di-*tuning*, dan dievaluasi kemudian diintegrasikan ke dalam sebuah sistem yang fungsional. Proses dimulai ketika pengguna memasukkan sebuah kalimat atau komentar melalui *output* yang telah disediakan. Teks tersebut akan melewati seluruh alur *preprocessing* yang sama persis seperti yang diterapkan pada data latih untuk memastikan konsistensi. Setelah teks bersih, model IndoBERT yang telah disimpan akan dimuat dan digunakan untuk melakukan prediksi. Sistem kemudian akan menampilkan tiga hasil yaitu label sentimen (positif, negatif, atau netral), label

topik kebijakan yang paling relevan, serta distribusi persentase yang menunjukkan seberapa relevan komentar tersebut terhadap masing-masing dari empat kebijakan tersebut.

```
MULAI

// Import library:
   -  PyTorch, Transformers, Pandas, IPyWidgets, IPython display

// Definisikan kelas model IndoBERTMultiOutput:
   -  Inisialisasi:
      - Load IndoBERT pre-trained
      - Tambahkan dropout
      - Tambahkan dua classifier (sentimen dan topik)
   -  Forward pass:
      - Masukkan input_ids dan attention_mask ke BERT
      - Ambil pooled_output dan apply dropout
      - Keluarkan dua output: out_sentimen dan out_topik

// Load model dan tokenizer
   -  Deteksi device (GPU/CPU)
   -  Load tokenizer dari direktori lokal
   -  Load model IndoBERTMultiOutput dan bobot pretrained
   -  Set model ke eval mode
   -  Definisikan label-label untuk sentimen dan topik

// Inisialisasi widget penampung hasil (output_container)

// Definisikan fungsi analisis_komentar(komentar):
   -  Tokenisasi komentar (max_length=128)
   -  Lakukan prediksi tanpa gradien:
      - Ambil output sentimen dan topik dari model
      - Hitung label prediksi dengan argmax
   -  Buat DataFrame untuk menampilkan distribusi probabilitas
   -  Buat widget output:
      - Tampilkan komentar, hasil prediksi, dan tabel distribusi
   -  Tambahkan hasil baru ke bagian atas output_container

// Buat input dan tombol:
```

```
   -  Textarea untuk komentar
   -  Button "Analisis Komentar"

// Definisikan event handler tombol:
   -  Ambil isi textarea
   -  Jika tidak kosong:
      - Jalankan analisis_komentar()
      - Kosongkan textarea

// Buat dan tampilkan form antarmuka:
   -  Judul + textarea + tombol
   -  Tampilkan form dan hasil analisis

SELESAI
```

BAB V

HASIL

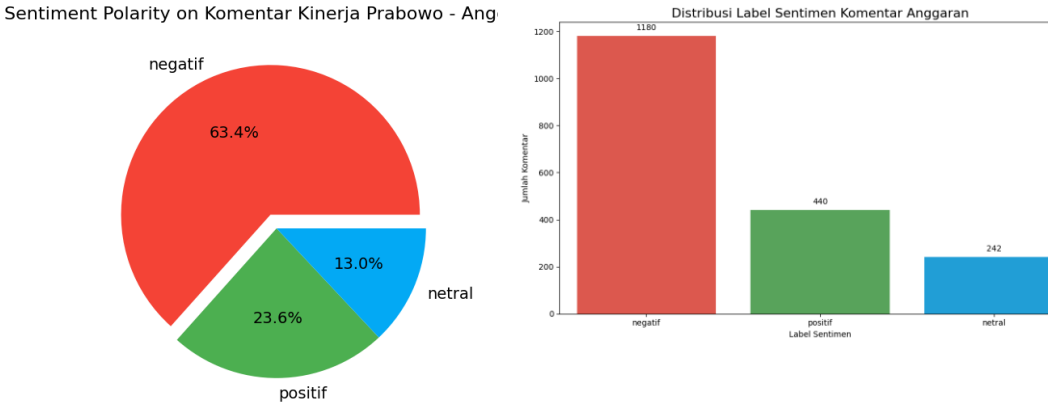
5.1 Hasil Analisis Sentimen Setiap Program

Pada bagian ini, akan dipaparkan hasil analisis sentimen secara rinci untuk masing-masing dari empat kebijakan utama yang menjadi fokus penelitian. Hasil yang disajikan disini diperoleh dari proses pelabelan awal menggunakan metode *Lexicon Based*, yang bertujuan untuk memetakan distribusi opini publik berdasarkan data komentar yang telah dikumpulkan dari platform X dan YouTube.

Untuk mempermudah pemahaman, hasil analisis untuk setiap program akan divisualisasikan dalam bentuk diagram lingkaran (*pie chart*) yang menunjukkan komposisi persentase sentimen, serta diagram batang (*bar chart*) yang menggambarkan jumlah total komentar untuk setiap kategori (positif, negatif, dan netral). Selanjutnya, akan disajikan pula evaluasi performa model IndoBERT yang digunakan, yang diukur melalui *Confusion Matrix* serta metrik kuantitatif seperti *accuracy*, *precision*, *recall*, dan *F1-Score*. Terakhir, akan ditampilkan hasil dari implementasi akhir sistem sebagai demonstrasi dari kemampuan model dalam mengklasifikasikan data baru.

5.1.1 Program Anggaran

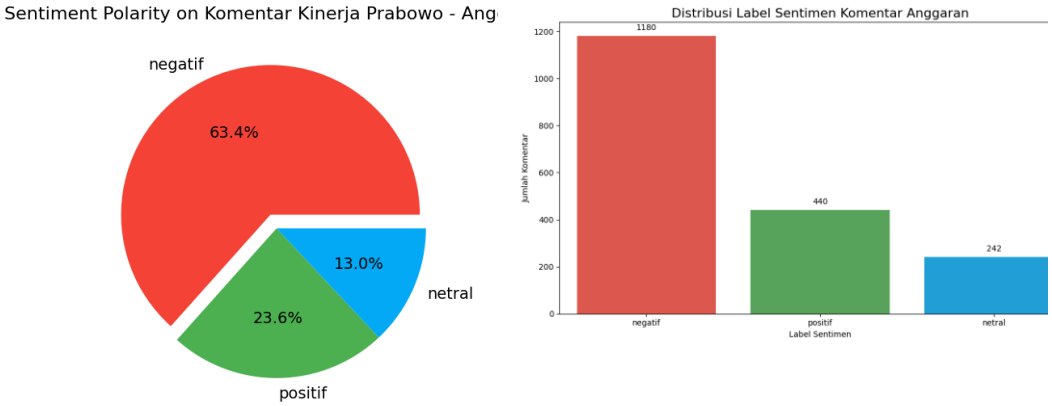
Hasil analisis sentimen terkait program efisiensi anggaran menunjukkan adanya distribusi sentimen yang beragam. Berdasarkan penelitian ini yang menggunakan metode *Lexicon Based*, sentimen negatif terhadap program efisiensi anggaran mencapai sekitar 63.4%, sentimen positif terhadap program efisiensi anggaran mencapai sekitar 23.6%, sentimen netral terhadap program efisiensi anggaran mencapai 13.0%. Diagram di bawah ini menunjukkan bahwa mayoritas sentimen terhadap program ini adalah negatif.



Gambar 5.1 Diagram Pada Program Anggaran

5.1.2 Program Danantara

Hasil analisis sentimen terkait program danantara menunjukkan adanya distribusi sentimen yang beragam. Berdasarkan penelitian ini yang menggunakan metode *Lexicon Based*, sentimen negatif terhadap program danantara mencapai sekitar 50.9%, sentimen positif terhadap program danantara mencapai sekitar 29.8%, sentimen netral terhadap program danantara mencapai 19.3%. Diagram di bawah ini menunjukkan bahwa mayoritas sentimen terhadap program ini adalah negatif.

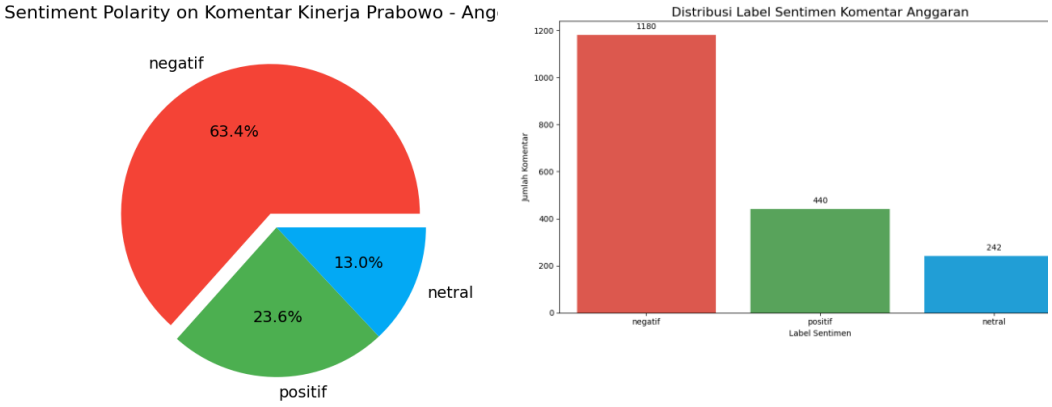


Gambar 5.2 Diagram Pada Program Danantara

5.1.3 Program PPN 12%

Hasil analisis sentimen terkait program PPN 12% menunjukkan adanya distribusi sentimen yang beragam. Berdasarkan penelitian ini yang menggunakan metode *Lexicon Based*, sentimen negatif terhadap program PPN 12% mencapai sekitar 58.3%, sentimen positif terhadap program PPN 12% mencapai sekitar 21.3%,

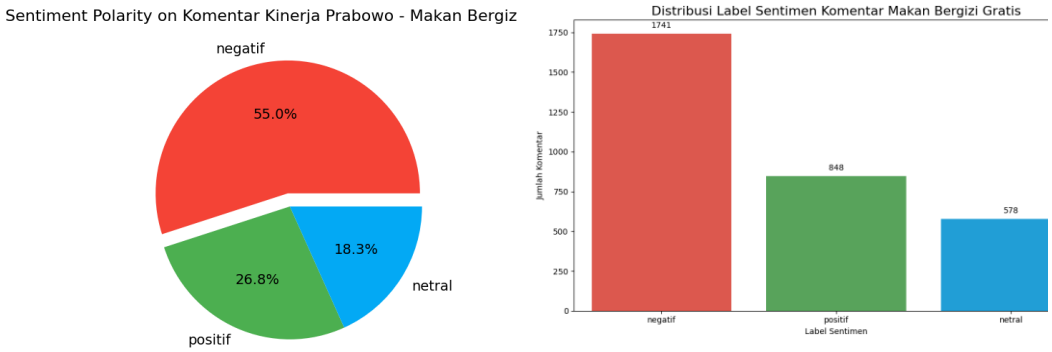
sentimen netral terhadap program PPN 12% mencapai 20.4%. Diagram di bawah ini menunjukkan bahwa mayoritas sentimen terhadap program ini adalah negatif.



Gambar 5.3 Diagram Pada Program PPN 12%

5.1.4 Program Makanan Bergizi Gratis

Hasil analisis sentimen terkait program makan bergizi gratis menunjukkan adanya distribusi sentimen yang beragam. Berdasarkan penelitian ini yang menggunakan metode *Lexicon Based*, sentimen negatif terhadap program makan bergizi gratis mencapai sekitar 55.0%, sentimen positif terhadap program makan bergizi gratis mencapai sekitar 26.8%, sentimen netral terhadap program makan bergizi gratis mencapai 18.3%. Diagram di bawah ini menunjukkan bahwa mayoritas sentimen terhadap program ini adalah negatif.



Gambar 5.4 Diagram Pada Program Makan Bergizi Gratis

5.1.5 Keempat Kinerja

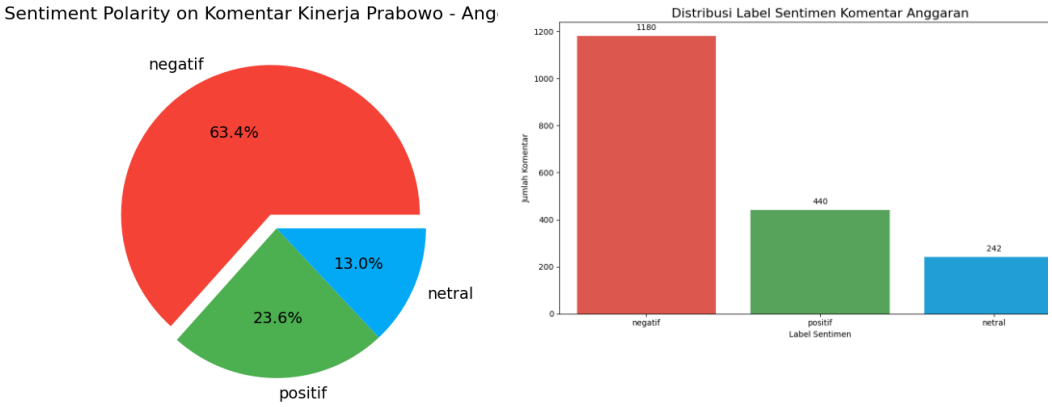
Pertama, dari sisi kinerja model, model *multi-task* IndoBERT yang dikembangkan menunjukkan performa yang solid dan efektif. Model terbukti lebih andal dalam melakukan klasifikasi topik dengan akurasi keseluruhan mencapai 82%, dibandingkan dengan tugas klasifikasi sentimen yang mencapai akurasi 74%.

Kekuatan utama model adalah kemampuannya dalam mengidentifikasi topik "Efisiensi Anggaran" (*F1-Score* 0.93) dan sentimen "Negatif" (*F1-Score* 0.85). Sementara itu, tantangan terbesar model terletak pada klasifikasi sentimen "Netral" (*F1-Score* 0.51), yang menunjukkan adanya kesulitan dalam menangani komentar yang bersifat ambigu.

Kedua, dari sisi distribusi data uji, analisis terhadap 1.171 komentar yang dievaluasi menunjukkan bahwa wacana publik mengenai empat kebijakan ini didominasi oleh sentimen negatif. Terdapat 659 sampel komentar negatif dalam data uji, jumlah ini secara signifikan lebih banyak dibandingkan dengan 297 sampel positif dan 215 sampel netral. Dari sisi topik, diskusi paling banyak terjadi seputar kebijakan PPN 12% (354 sampel) dan Makan Bergizi Gratis (324 sampel), yang mengindikasikan bahwa isu-isu yang berdampak langsung pada ekonomi dan kesejahteraan masyarakat menjadi perhatian utama dalam diskursus publik.

Secara ringkas, temuan utama penelitian ini menunjukkan dua hal: (1) Model IndoBERT yang dibangun mampu memetakan dinamika opini publik dengan andal, dan (2) Diskursus publik di media sosial terhadap empat kebijakan awal pemerintahan cenderung didominasi oleh kritik dan sentimen negatif, terutama pada isu-isu yang menyentuh langsung kehidupan sehari-hari masyarakat.

Hasil analisis sentimen terkait keempat program Prabowo menunjukkan adanya distribusi sentimen yang beragam. Berdasarkan penelitian ini yang menggunakan metode *Lexicon Based*, sentimen negatif terhadap keempat program mencapai sekitar 56.3%, sentimen positif terhadap keempat program mencapai sekitar 25.4%, sentimen netral terhadap keempat program mencapai 18.4%. Hasil diagram dibawah ini memperlihatkan bahwa penerimaan publik terhadap program kerja Prabowo cenderung positif, namun tetap diperhatikan adanya suara-suara kritis sebagai bagian dari evaluasi terhadap implementasi program kedepannya. Berikut adalah hasil klasifikasi sentimen terhadap komentar publik mengenai keempat program ini.



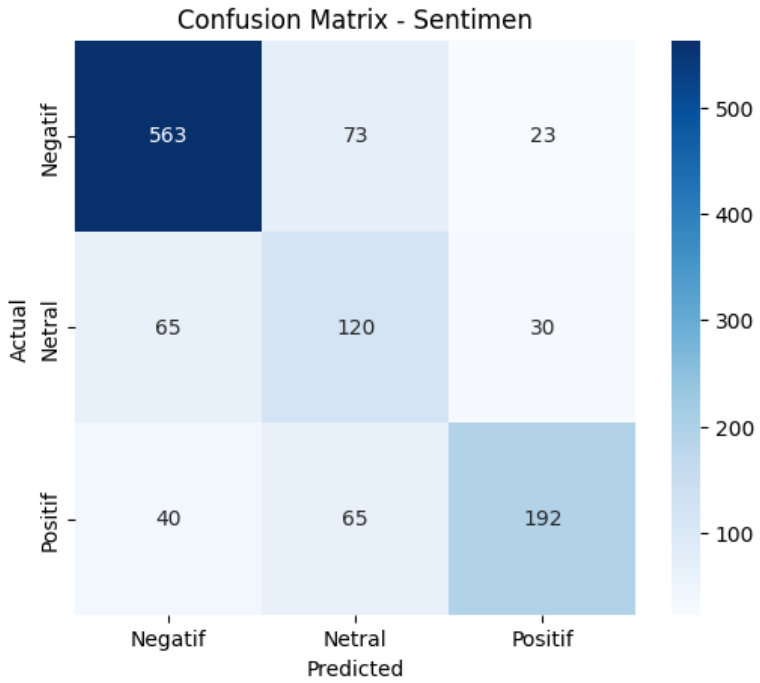
Gambar 5.5 Diagram Pada Keempat Program

5.2 Confusion Matrix

Confusion Matrix adalah sebuah tabel yang digunakan untuk mengevaluasi performa sistem klasifikasi dengan membandingkan hasil prediksi model dengan kelas sebenarnya. Tabel ini terdiri dari baris (kelas aktual) dan kolom (kelas prediksi), yang menunjukkan jumlah prediksi benar maupun salah untuk masing-masing kelas. Model tidak selalu membuat prediksi yang tepat, karena adanya komentar ambigu, pendek, atau pelabelan yang kurang konsisten. Untuk mengatasi hal ini, dilakukan evaluasi dan perbaikan pada data latih serta pelabelan agar model lebih mudah membedakan sentimen.

5.2.1 Confusion Matrix Sentiment

Confusion matrix pada Gambar dibawah ini menyajikan visualisasi yang rinci mengenai kinerja model IndoBERT dalam tugas klasifikasi sentimen pada data *test*. Matriks ini membandingkan label sentimen yang sebenarnya (*Actual*) dengan label yang diprediksi oleh model (*Predicted*), sehingga dapat melihat secara spesifik dimana model berhasil dan dimana model mengalami kebingungan. Sumbu vertikal merepresentasikan kelas asli dari data, sementara sumbu horizontal merepresentasikan kelas hasil prediksi model.



Gambar 5.6 Confusion Matrix Sentiment

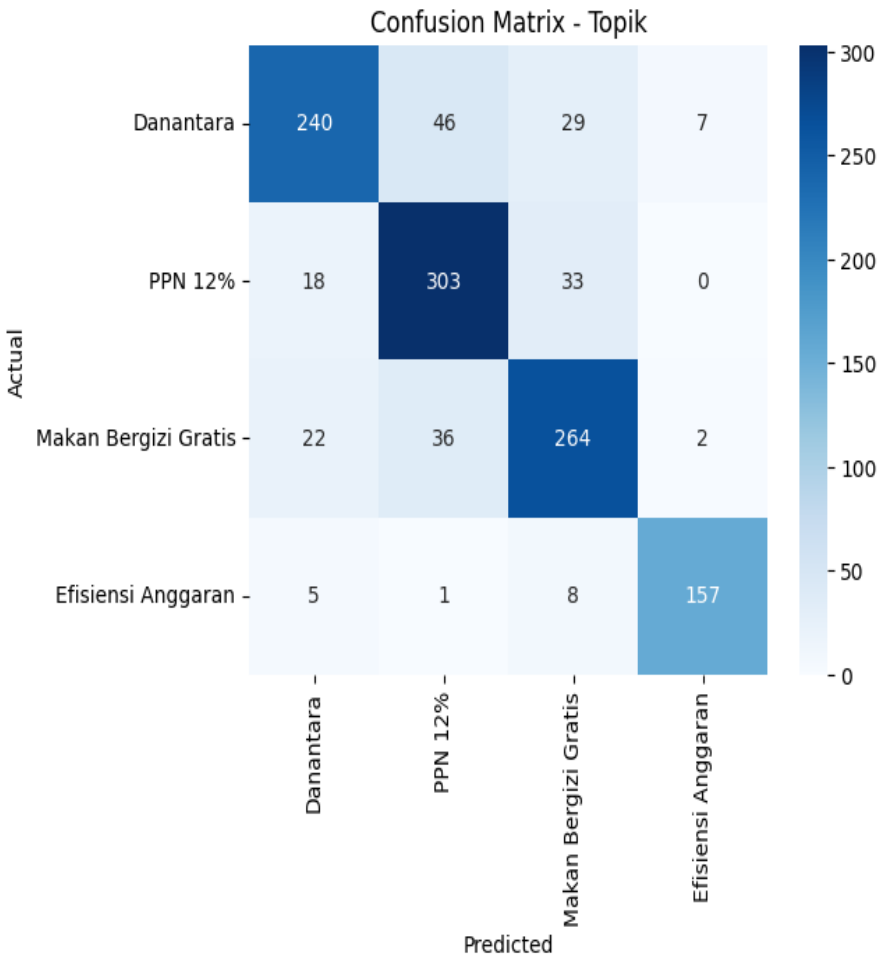
Berdasarkan matriks di atas, terlihat bahwa kemampuan model bervariasi untuk setiap kelas. Performa terbaik ditunjukkan dalam mengenali sentimen negatif, di mana 563 dari 659 komentar berhasil diklasifikasikan dengan benar. Meskipun demikian, model menghadapi tantangan signifikan pada kelas lainnya. Tantangan terbesar terletak pada identifikasi sentimen netral. Dari total 215 komentar netral, hanya 120 yang diprediksi dengan benar, sementara sebagian besar sisanya keliru diklasifikasikan sebagai negatif (65 komentar) dan positif (30 komentar). Kebingungan serupa juga terjadi pada kelas positif, di mana sebanyak 65 komentar yang seharusnya positif, justru salah diidentifikasi sebagai netral.

Secara keseluruhan, analisis *confusion matrix* ini menyimpulkan bahwa model sangat andal dalam mendeteksi opini negatif yang cenderung memiliki kosakata yang lugas dan bernuansa emosional yang kuat. Namun, model masih memiliki kelemahan dalam membedakan antara komentar positif dan terutama komentar netral. Hal ini kemungkinan besar disebabkan oleh ambigunya bahasa dalam komentar netral yang seringkali tidak mengandung penanda emosi yang kuat. Selain itu, komentar netral sering bersifat informatif, datar, atau menggunakan kata-kata yang muncul dalam konteks positif maupun negatif, sehingga menyulitkan model

untuk menangkap intensi emosionalnya. Komentar pendek atau tidak lengkap juga menjadi tantangan, karena memberikan konteks yang terbatas bagi model dalam melakukan prediksi. Kemungkinan lain berasal dari data pelatihan yang belum sepenuhnya mencakup keragaman gaya bahasa atau ekspresi netral yang lebih halus, sehingga menyebabkan model cenderung bias terhadap kelas sentimen yang lebih eksplisit.

5.2.2 Confusion Matrix Topic

Gambar dibawah ini menyajikan *confusion matrix* yang menggambarkan performa model dalam tugas klasifikasi topik kebijakan. Matriks ini secara visual menampilkan kemampuan model untuk membedakan komentar yang membahas "Danantara", "PPN 12%", "Makan Bergizi Gratis", dan "Efisiensi Anggaran".



Gambar 5.7 Confusion Matrix Topic

Secara keseluruhan, matriks ini menunjukkan bahwa model memiliki performa yang sangat baik dan andal dalam membedakan keempat topik kebijakan. Kinerja paling menonjol terlihat pada topik Efisiensi Anggaran, di mana model berhasil memprediksi dengan benar 157 dari 1171 komentar, dengan tingkat kesalahan klarifikasi yang sangat minimal. Performa baik ini disebabkan oleh penggunaan kosakata yang sangat spesifik pada topik ini (seperti 'hemat', 'pangkas'), sehingga mudah dibedakan dari topik lainnya.

Performa yang kuat juga ditunjukkan pada topik PPN 12% dan Makan Bergizi Gratis, dengan masing-masing 303 dan 264 prediksi benar. Meskipun demikian, matriks ini juga mengungkap area di mana model mengalami kebingungan dalam membedakan topik. Salah satu contoh yang mencolok adalah tumpang tindih prediksi antara topik Danantara dan 36 komentar Makan Bergizi Gratis secara keliru diklasifikasikan sebagai PPN 12%. Hal ini mengindikasikan bahwa ketiga topik tersebut kemungkinan memiliki kemiripan konteks wacana, terutama dalam hal terminologi ekonomi, kebijakan pemerintah, dan pembangunan. Kosakata yang bersifat umum seperti "program", "anggaran", atau "subsidi" mungkin muncul di ketiga topik tersebut, sehingga menyulitkan model untuk menangkap perbedaan spesifik antar isu tersebut. Selain itu, beberapa komentar yang membahas lebih dari satu topik secara bersamaan turut menjadi tantangan bagi model, karena pendekatan klasifikasi yang digunakan bersifat satu label (*single label*), dimana model hanya diizinkan memilih satu topik utama. alam konteks tersebut, komentar multi-topik cenderung menyebabkan ambiguitas dalam prediksi. Meski demikian, secara umum hasil ini tetap menunjukkan bahwa model klasifikasi topik yang dikembangkan sudah cukup efektif dan akurat dalam mengenali serta membedakan isu-isu utama dalam kebijakan publik.

5.3 Evaluasi

Berdasarkan hasil tahapan dari klasifikasi, dilakukan evaluasi dalam setiap skenario. Berikut merupakan rumus yang digunakan untuk menghitung *accuracy, precision, recall, f1-score*.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$F1\text{-}Score = 2x\frac{Precision\ x\ Recall}{Precision + Recall}$$

$$Macro\ Avg = \frac{1}{N}\sum_{i=1}^{N} Precision_i$$

$$Weighted\ Avg = \frac{\Sigma_{i=1}^{N}(F1_i \times Support_i)}{\Sigma_{i=1}^{N} Support_i}$$

Berikut merupakan perhitungan manual confusion matrix 3x3 untuk sentimen dan 4x4 untuk topik.

1. Confusion Matrix 3x3 Sentimen

   $$Accuracy = \frac{563 + 120 + 192}{1171} = \frac{875}{1171} = 0.75$$

   Untuk *precision, recall,* F1, mengambil salah satu contoh sentimen yaitu sentimen dari positif

   - *True Positif* (TP) = 192
   - *False Positif* (FP) = 30 (netral ke positif) + 23 (negatif ke positif)= 53
   - *False Negatif* (FN) = 65 (positif ke netral) + 40 (positif ke negatif) = 105

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} = \frac{192}{192 + 53} = \frac{192}{245} = 0.78$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} = \frac{192}{192 + 105} = \frac{192}{297} = 0.65$$

$$F1\text{-}Score = 2x\frac{Precision\ x\ Recall}{Precision + Recall} = 2x\frac{0.78\ x\ 0.65}{0.78 + 0.65} = 0.71$$

$$Macro\ Avg = \frac{1}{N}\sum_{i=1}^{N} Precision_i = \frac{0.84 + 0.47 + 0.78}{3} = 0.70$$

$$Weighted\ \ F1 = \frac{\Sigma_{i=1}^{N}(F1_i \times Support_i)}{\Sigma_{i=1}^{N} Support_i} = \frac{(0.84\times659)+(0.47\times215)+(0.78\times297)}{1171} = \frac{553.56+101.05+231.27}{1171} = \frac{886.27}{1171} = 0.75$$

Hasil evaluasi dari setiap kategori ditunjukkan pada Tabel 5.1.

Tabel 5.1 Classification Report - Sentimen

Kategori	Precision	Recall	F1-Score	Support
Negatif	0.84	0.85	0.85	659
Netral	0.47	0.56	0.51	215
Positif	0.78	0.65	0.71	297
—	—	—	—	—
			0.75	1171
Macro Avg	0.70	0.69	0.69	1171
Weighted Avg	0.76	0.75	0.75	1171

Pada tabel klasifikasi sentimen di atas, secara keseluruhan, model menunjukkan kinerja yang baik dengan nilai *Weighted Avg F1-Score* sebesar 0.75. Namun, jika dianalisis lebih dalam per kategori, terlihat adanya performa yang bervariasi. Model menunjukkan performa terbaik dalam mengenali sentimen Negatif, dengan *F1-Score* yang sangat kuat mencapai 0.85. Tingginya nilai ini kemungkinan besar disebabkan oleh jumlah data (*support*) pada kelas Negatif yang paling dominan (659 sampel) serta penggunaan kosakata kritis yang cenderung lebih jelas.

Sebaliknya, tantangan terbesar model terletak pada klasifikasi sentimen netral, yang hanya mencapai *F1-Score* sebesar 0.51. Meskipun ada sedikit peningkatan dari hasil sebelumnya, angka ini tetap menunjukkan bahwa model seringkali kesulitan dan keliru dalam mengidentifikasi komentar yang bersifat netral. Untuk sentimen positif, model menunjukkan kinerja yang solid dengan *F1-Score* 0.71, yang lebih baik dari netral namun belum sebaik kelas negatif.

Dari hasil ini, dapat disimpulkan bahwa model yang dikembangkan sangat andal dalam mendeteksi opini negatif, namun masih memerlukan perbaikan dalam membedakan nuansa antara komentar positif dan terutama opini netral. Kelemahan pada kelas netral ini kemungkinan besar disebabkan oleh karakteristik komentar

yang seringkali ambigu, tidak memiliki kata kunci emosional yang kuat, dan jumlah datanya yang lebih sedikit dibandingkan kelas negatif.

2. Confusion Matrix 4x4 Topik

$$Accuracy = \frac{240+303+264+157}{1171} = \frac{964}{1171} = 0.82$$

Untuk *precision, recall,* F1, mengambil salah satu contoh sentimen yaitu sentimen dari makan bergizi gratis

- *True Positif* (TP) = 264
- *False Positif* (FP) = 29 (Danantara ke MBG) + 33 (PPN ke MBG) + 8 (Efisiensi ke MBG) = 70
- *False Negatif* (FN) = 22 (MBG ke Danantara) + 36 (MBG ke PPN) + 2 (MBG ke Efisiensi) = 60

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} = \frac{264}{264 + 70} = \frac{264}{334} = 0.79$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} = \frac{264}{264 + 60} = \frac{264}{324} = 0.81$$

$$F1\text{-}Score = 2x\frac{Precision\ x\ Recall}{Precision + Recall} = 2x\frac{0.79\ x\ 0.81}{0.79 + 0.81} = 0.80$$

$$Macro\ Avg = \frac{1}{N}\sum_{i=1}^{N} Precision_i = \frac{0.84 + 0.78 + 0.79 + 0.95}{4} = 0.84$$

$$Weighted\ F1 = \frac{\Sigma_{i=1}^{N}(F1_i \times Support_i)}{\Sigma_{i=1}^{N} Support_i} = \frac{(0.79\times322)+(0.82\times354)+(0.80\times324)+(0.93\times171)}{1171} = \frac{254{,}38+290.28+259.2+159.03}{1171} = \frac{962.89}{1171} = 0.82$$

Hasil evaluasi dari setiap topik ditunjukkan pada Tabel 5.2.

Tabel 5.2 Classification Report - Topik

Kategori	Precision	Recall	F1-Score	Support
Danantara	0.84	0.75	0.79	322
PPN 12%	0.78	0.86	0.82	354
Makan Bergizi Gratis	0.79	0.81	0.80	324

Kategori	Precision	Recall	F1-Score	Support
Efisiensi Anggaran	0.95	0.92	0.93	171
—	—	—	—	—
Accuracy			0.82	1171
Macro Avg	0.84	0.83	0.84	1171
Weighted Avg	0.83	0.82	0.82	1171

Pada tabel klasifikasi topik di atas, model menunjukkan performa yang sangat baik secara keseluruhan, dengan tingkat akurasi mencapai 82%. Kinerja ini secara signifikan lebih tinggi dibandingkan dengan tugas klasifikasi sentimen, yang mengindikasikan bahwa model lebih mudah mengenali topik kebijakan daripada nuansa emosi.

Performa yang paling menonjol ditunjukkan pada topik Efisiensi Anggaran, yang mencapai *F1-Score* luar biasa sebesar 0.93. Performa tinggi ini disebabkan oleh penggunaan kosakata yang sangat spesifik dan unik terkait topik ini (misalnya: 'hemat', 'pangkas', 'birokrasi') sehingga mudah dibedakan dari topik lain.

Topik-topik lainnya seperti PPN 12% dengan *F1-Score* 0.82, Makan Bergizi Gratis dengan *F1-Score* 0.80, dan Danantara dengan *F1-Score* 0.79 juga berhasil diklasifikasikan dengan tingkat akurasi yang baik dan konsisten. Angka-angka ini menunjukkan bahwa model secara umum mampu dalam membedakan pembahasan mengenai keempat kebijakan tersebut. Kesimpulan yang dapat diambil, model klasifikasi topik ini terbukti sangat efektif dan dapat diandalkan, terutama dalam mengidentifikasi topik 'Efisiensi Anggaran'. Hal ini menunjukkan bahwa perbedaan kosakata antara topik kebijakan cukup jelas bagi model untuk dipelajari.

5.4 Perbandingan Model Tanpa SMOTE dan Dengan SMOTE

Sebagai bagian dari evaluasi performa, dilakukan perbandingan antara model yang dilatih tanpa teknik penyeimbangan data dan model yang dilatih dengan SMOTE. Perbandingan ini dilakukan untuk menilai pengaruh teknik penyeimbangan data terhadap performa klasifikasi. Fokus utamanya adalah untuk mengamati perbedaan hasil prediksi, khususnya dalam menangani kelas-kelas minoritas pada tugas klasifikasi sentimen dan topik. Perbedaan performa dari kedua pendekatan ini dapat

terlihat dari hasil evaluasi yang mencerminkan dampak penerapan SMOTE terhadap akurasi dan kemampuan generalisasi model.

5.4.1 Latar Belakang dan Tujuan Perbandingan

Data dibagi dengan rasio 80% untuk data latih dari 20% untuk data uji. Pada skenario ini, meskipun tidak dilakukan proses berdasarkan kata-kata yang paling sering muncul di dalam dataset itu sendiri. *Lexicon* tersebut bersifat terbatas namun cukup relevan terhadap konteks data yang dikumpulkan, sehingga pelabelan awal cenderung menghasilkan distribusi label yang relatif seimbang. Dengan seimbangnya distribusi sentimen positif, negatif dan netral hasil dari proses pelabelan tersebut, tidak ada indikasi awal yang menunjukkan perlu penyeimbangan data menggunakan metode seperti SMOTE.

Namun, setelah dievaluasi secara lebih dalam, pendekatan *lexicon* yang hanya didasarkan pada kata-kata dalam dataset sendiri ditemukan kurang *efektif* dalam menangkap sentimen secara utuh. *Lexicon* terbatas tersebut memiliki keterbatasan dalam mengenali sinonim, kata-kata bermakna emosional yang jarang muncul, serta konteks linguistik yang lebih luas. Untuk memperluas cakupan dan memperbaiki akurasi pelabelan, maka digunakanlah *InSet Lexicon* sebuah kamus sentimen berbahasa Indonesia yang telah banyak digunakan dalam penelitian sebelumya. Kamus ini mencakup ribuan kata positif, negatif, dan netral yang telah ditentukan secara akademik.

Setelah proses pelabelan diperbarui dengan bantuan *InSet Lexicon*, ditemukan bahwa distribusi sentimen tidak lagi seimbang. Sebagian besar komentar terklasifikasi ke dalam kelas negatif, sementara kelas positif dan netral menjadi minoritas. Kondisi ini menyebabkan ketimpangan label yang cukup *signifikan,* sehingga berpotensi memunculkan bias dalam pelatihan model. Jika dibiarkan, model cenderung terlalu fokus pada kelas mayoritas dan mengabaikan pola dari kelas-kelas lain yang lebih jarang muncul. Oleh karena itu, pada tahapan ini mulai diterapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan jumlah data dari setiap kelas.

Selain SMOTE, dilakukan pula tuning *hyperparameter* dengan mencoba beberapa konfigurasi pelatihan. Parameter yang diuji adalah *learning rate* dengan tiga variasi yaitu 1e-5, 2e-5, dan 3e-5. Penyesuaian ini bertujuan untuk menemukan pengaturan yang paling stabil dan optimal dalam proses pembelajaran model. Data kemudian dibagi menjadi 70% untuk pelatihan 20% untuk validasi, dan 10% untuk pengujian agar hasil evaluasi dapat mencerminkan kemampuan generalisasi model secara menyeluruh.

5.4.2 Eksperimen Tanpa SMOTE

Eksperimen pertama dilakukan menggunakan data hasil pelabelan awal dengan *lexicon custom.* Data pelatihan dan pengujian dibagi dengan rasio 80:20. Model dilatih selama tiga epoch, dengan evaluasi menggunakan nilai *loss* dan *perplexity* (PPL) untuk mengukur performa prediksi dan pemahaman konteks. Meskipun sempat terjadi sedikit penurunan *loss* pada epoch pertama, performa validasi tidak menunjukkan peningkatan *signifikan* dan *stagnan* pada epoch selanjutnya. Hal ini menyebabkan aktivasi *early stopping.*

Tabel 5.3 Hasil Pelatihan Model Tanpa SMOTE

Epoch	*Train Loss*	*Train PPL*	*Val Loss*	*Val PPL*	Keterangan
1	8.3534	4244.64	8.0720	3203.42	Performa awal sedikit membaik
2	8.1007	3296.71	8.0870	3251.82	Tidak ada peningkatan
3	8.0724	3204.68	8.0764	3217.55	Pelatihan dihentikan otomatis

Nilai *loss* dan *perplexity* yang tidak menunjukkan bahwa model kesulitan memahami pola data, dan akurasi *validasi* tidak dicapai secara optimal.

5.4.3 Eksperimen Dengan SMOTE dan Tuning Hyperparameter

Setelah pelabelan diperbarui menggunakan *InSet Lexicon,* distribusi label dalam data menunjukkan ketimpangan yang cukup besar, di mana kelas negatif mendominasi dan kelas positif serta netral menjadi minoritas. Ketimpangan ini dapat menggunakan proses pelatihan model karena dapat menyebabkan model bias terhadap kelas mayoritas. Untuk mengatasi hal tersebut, diterapkan metode *Synthetic Minority Over- Sampling Technique* (SMOTE) pada data latih. SMOTE bekerja dengan menggunakan sampel sintetis dari kelas minoritas berdasarkan

kedekatan spasial pada ruang vektor, sehingga menghasilkan distribusi label yang lebih seimbang.

Setelah proses SMOTE, data dibagi menjadi tiga bagian: 70% data pelatihan, 20% data validasi, dan 10% data pengujian. Proporsi ini dipilih karena secara umum pembagian 70-20-10 dianggap sebagai kompromi yang baik antara kebutuhan pelatihan model, evaluasi performa pada data baru (validasi), pengujian akhir secara objektif. Data pelatihan digunakan untuk melatih bobot model, data validasi digunakan untuk mengevaluasi performa antar *epoch* dan mengaktifkan mekanisme *early stopping,* sedangkan data pengujian digunakan pada tahap akhir untuk mengukur performa generalisasi model.

Pada tahap ini juga dilakukan eksperimen *tuning hyperparameter,* yaitu proses penyesuaian nilai parameter pelatihan model untuk mendapatkan hasil yang paling optimal. Salah satu *hyperparameter* sangat berpengaruh terhadap proses pelatihan adalah *learning rate. Learning rate* menentukan seberapa besar bobot model diperbarui setiap kali terjadi propagasi mundur (*backpropagation*). Nilai *Learning rate* yang terlalu besar dapat menyebabkan model melompat-lompat di sekitar nilai minimum dan tidak stabil, sedangkan nilai yang terlalu kecil dapat membuat proses pelatihan lambat dan berisiko terjebak pada local minimum.

Dalam eksperimen ini, digunakan tiga nilai *learning rate* yang berbeda, yaitu 1e-5, 2e-5, dan 3e-5. Pemilihan rentang nilai ini didasarkan pada rekomendasi umum dan praktek terbaik dalam *fine-tuning* model BERT variannya, termasuk IndoBERT. Seperti yang dijelaskan dalam diskusi komunitas. *Data Science Stack Exchange* , kisaran *learning rate* antara 1e-5, 2e-5, sampai hingga 5e-5 merupakan nilai yang paling umum digunakan dalam *fine-tuning* model IndoBERT. Nilai 2e-5 disebut sebagai *default* karena stabil dalam banyak kasus, sementara 1e-5 lebih cocok untuk pelatihan dengan dataset yang lebih besar atau ketika mengutamakan stabilitas, dan 3e-5 dapat mempercepat konvergensi pelatihan dengan tetap berapa dalam batas aman.

Selain, *learning rate* parameter *batch size* dan *epoch* juga ditentukan. *Batch size* yang digunakan adalah 16, yang merupakan pilihan umum untuk pelatihan model BERT di lingkungan terbatas GPU. Nilai ini cukup kecil untuk menghindari masalah memori, namun masih cukup besar untuk memberikan *gradien* yang representatif. Model dilatih selama maksimal lima *epoch* untuk masing-masing kombinasi *hyperparameter.* Mekanisme *early stopping* diterapkan apabila tidak ada peningkatan akurasi validasi dalam dua *epoch* berturut-turut, untuk mencegah *overfitting* dan efisiensi waktu pelatihan.

Hasil pelatihan menunjukkan bahwa nilai *train loss* menurun secara konsisten, dan akurasi validasi pada klasifikasi sentimen meningkat. Berikut adalah ringkasan hasil pelatihan berdasarkan kombinasi *learning rate*.

Tabel 5.4 Hasil dengan SMOTE dan Tuning Hyperparameter

Learning Rate	**Epoch Terbaik**	***Train Loss***	***Val Accuracy***	**Keterangan**
1e-5	3	1.1053	0.7114	Peningkatan Signifikan
2e-5	3	1.0591	0.7323	Lebih stabil, hasil lebih merata
3e-5	4	0.9780	0.7395	Performa terbaik secara keseluruhan

Dari tabel tersebut, terlihat bahwa *learning rate* 3e-5, menghasilkan hasil terbaik dengan akurasi validasi klasifikasi sentimen 73,95% pada epoch keempat. Hal ini menunjukkan bahwa pemilihan *learning rate* yang tepat, dikombinasikan dengan strategi penyeimbangan data menggunakan SMOTE dan pengaturan *batch size* serta *epoch* yang sesuai, mampu memberikan kontribusi yang nyata dalam peningkatan performa model IndoBERT untuk tugas klasifikasi sentimen.

5.4.4 Analisis

Eksperimen yang dilakukan pada dua skenario pelatihan model IndoBERT tanpa SMOTE dan dengan SMOTE serta *tuning hyperparameter* menunjukkan perbedaan performa yang cukup signifikan, baik dari sisi stabilitas proses pelatihan maupun akurasi validasi. Perbandingan ini bertujuan untuk menganalisis secara menyeluruh bagaimana dampak penyeimbangan data dan konfigurasi pelatihan terhadap kemampuan model dalam mengklasifikasi sentimen komentar masyarakat secara akurasi dan adil.

Pada skenario tanpa SMOTE, data pelatih diambil dari hasil pelabelan menggunakan *lexicon* terbatas yang dibangun berdasarkan dataset itu sendiri. Meskipun distribusi label tampak seimbang secara *kuantitatif,* hasil pelatihan menunjukkan bahwa model mengalami kesulitan dalam memahami konteks data. Hal ini ditandai dengan nilai *train loss* dan *perplexity*(PPL) yang tetap tinggi selama proses pelatihan, serta tidak adanya peningkatan berarti pada performa validasi. Pelatihan bahkan terhenti lebih awal (*early stopping*) pada *epoch* ketiga karena tidak ditemukan perbaikan performa dalam dua *epoch* berturut-turut. Ini menunjukkan bahwa kualitas pelabelan dan representasi data memiliki pengaruh besar terhadap efektivitas model.

Sebaiknya, pada skenario kedua, data dilabeli ulang menggunakan *InSet Lexicon,* sebuah kamus sentimen berbahasa Indonesia yang lebih luas cakupannya. Hal ini membuat hasil pelabelan menjadi lebih representatif, namun juga mengungkapkan adanya ketimpangan distribusi label, khususnya dominasi sentimen negatif. Ketimpangan ini diatasi dengan menerapkan SMOTE (*Synthetic Minority Over-Sampling Technique*) pada data pelatihan, yang menghasilkan distribusi kelas yang lebih seimbang. Untuk mendapatkan konfigurasi pelatihan yang optimal, dilakukan *tuning* terhadap *hyperparameter* model, khususnya pada nilai *learning rate.*

Learning rate adalah salah satu *hyperparameter* paling penting dalam proses pembelajaran model, karena menentukan seberapa besar bobot model diperbarui dalam setiap langkah pelatihan. Pada eksperimen ini, diuji tiga nilai *learning rate* yaitu 1e-5, 2e-5, dan 3e-5. Notasi 1e-5 merupakan notasi ilmiah yang berarti 0.00001 demikian pula 2e-5 yang berarti 0.00002 dan 3e-5 yang berarti 0.00003.

Pemilihan nilai ini didasarkan pada praktek umum dan rekomendasi dari komunitas ilmiah dalam *fine-tuning* model BERT. Menurut diskusi teknis di *Data Science Stack Exchange,* kisaran nilai *learning rate 1e-5* hingga *5e-5* merupakan rentang yang umum digunakan dalam *fine-tuning* model *transformer* seperti IndoBERT. Nilai 1e-5 cenderung lebih konservatif dan stabil, cocok untuk dataset besar atau model yang mudah *overfitting.* Nilai 2e-5 dianggap sebagai default yang seimbang

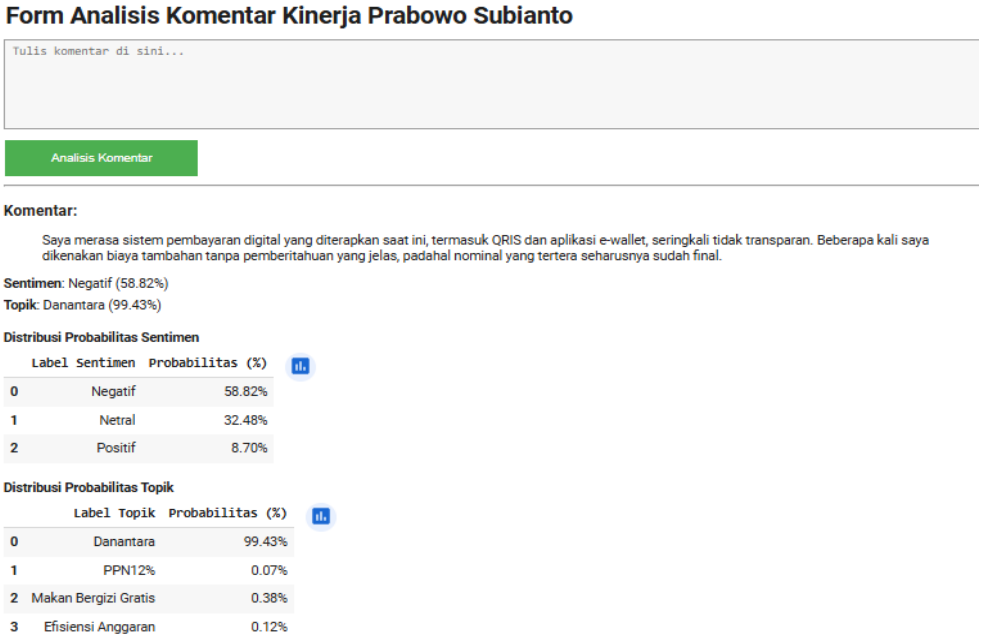
antara stabilitas dan kecepatan konvergensi. Sedangkan 3e-5 relatif lebih agresif dan cocok untuk mempercepat pelatihan pada dataset yang tidak terlalu besar, namun tetap dalam batas yang aman secara numerik.

Secara keseluruhan, eksperimen ini memperlihatkan bahwa pendekatan pelatihan model IndoBERT yang menggunakan pelabelan berbasis *InSet Lexicon,* penyeimbangan data dengan SMOTE, serta *tuning hyperparameter* secara sistematis, terbukti lebih efektif dibandingkan pendekatan awal tanpa penyeimbangan data. Model hasil pendekatan kedua tidak hanya lebih akurat dalam mengenali sentimen, tetapi juga lebih adil dalam menangani ketimpangan kelas dan variasi gaya bahasa dalam data komentar media sosial. Dengan strategi pelatihan yang tepat, model IndoBERT mampu mengklasifikasi opini publik secara lebih *representatif* dan dapat diandalkan untuk mendukung analisis kebijakan berbasis persepsi masyarakat.

5.5 Implementasi

Form analisis komentar ini dibuat agar pengguna bisa langsung mencoba model yang telah dibangun dengan cara memasukkan komentar ke dalam kolom teks yang tersedia. Komentar yang dimasukkan akan diproses oleh model IndoBERT yang telah dilatih untuk mengklasifikasikan dua hal, yaitu sentimen (negatif,positif,atau netral) dan topik pembahasan (seperti Makan Bergizi Gratis, PPN 12%, dan sebagainya).

Setelah total “Analisis Komentar” ditekan, sistem akan menampilkan hasil prediksi sentimen dan topik beserta distribusi *probabilitasnya* dalam bentuk tabel. Tampilan ini dibuat menggunakan *ipywidgets* di *Google Collab* yang memungkinkan antarmuka *interaktif* tanpa perlu membangun aplikasi *web* tambahan. Penggunaan *ipywidgets* memudahkan proses uji coba dan demonstrasi sistem secara langsung kepada pengguna atau penguji, sehingga hasil analisis dapat langsung terlihat dalam satu tampilan.



Gambar 5.8 Implementasi *Form* Analisis

5.6 Pengujian Hipotesis Berdasarkan Hasil Analisis Sentimen

Berdasarkan hasil analisis sentimen, dilakukan pengujian untuk menentukan apakah hipotesis awal penelitian sesuai dengan temuan data. Hipotesis yang menyatakan akan adanya resistensi publik terhadap kenaikan PPN menjadi 12% berhasil dikonfirmasi. Temuan ini didukung oleh tingginya sentimen negatif (58,3%) yang menunjukkan bahwa kekhawatiran masyarakat atas beban ekonomi memang terbukti, sesuai dengan perkiraan awal. Sama halnya, hipotesis untuk Program Makan Bergizi Gratis (MBG) yang memperkirakan adanya pandangan yang terbelah juga dikonfirmasi. Data menunjukkan adanya sentimen positif yang cukup tinggi (26,8%) meskipun sentimen negatif tetap mendominasi (55,0%), yang berarti publik mendukung gagasannya namun mengkritik keras pelaksanaannya.

Di sisi lain, hipotesis untuk kebijakan Efisiensi Anggaran ditolak. Asumsi awal bahwa publik akan memberikan apresiasi tidak terbukti, karena data didominasi oleh sentimen negatif yang sangat kuat sebesar 63,4%. Hal ini menunjukkan kekhawatiran publik jauh lebih besar daripada dukungan yang diharapkan. Hipotesis untuk program Danantara juga ditolak. Data tidak menunjukkan adanya perdebatan yang seimbang antara harapan dan kekhawatiran. Tingginya angka sentimen netral (19,3%) justru menandakan adanya kekurangan pemahaman di

tengah masyarakat mengenai program baru ini, bukan perdebatan pro dan kontra yang tajam.

BAB VI

PEMBAHASAN

6.1 Analisis Hasil Sentimen pada Setiap Program

Berdasarkan hasil pengujian model IndoBERT terhadap data uji, analisis sentimen yang lebih mendalam dapat dilakukan untuk memahami bagaimana persepsi publik terbentuk terhadap masing-masing dari empat program kerja Presiden Prabowo. Alih-alih hanya melihat distribusi awal, pembahasan ini akan menginterpretasikan temuan dari kinerja model untuk menjawab rumusan masalah penelitian.

Secara keseluruhan, data uji menunjukkan bahwa diskursus publik mengenai keempat kebijakan ini cenderung didominasi oleh sentimen negatif. Hal ini terlihat dari komposisi data uji di mana komentar negatif (659 sampel) jauh melampaui komentar positif (297 sampel) dan netral (215 sampel), seperti yang dirangkum pada Tabel 5.1. Namun, jika dianalisis lebih dalam, setiap kebijakan memicu pola sentimen yang unik.

Pada Program Makan Bergizi Gratis (MBG), meskipun terdapat banyak kritik terkait implementasi, hipotesis awal mengenai adanya apresiasi publik terkonfirmasi. Topik ini menjadi satu-satunya yang memunculkan sentimen positif dalam jumlah yang signifikan. Banyaknya komentar positif yang berhasil diidentifikasi model menunjukkan bahwa ide dasar dari program ini diterima baik oleh masyarakat karena menyentuh kebutuhan esensial. Kritik yang muncul (sentimen negatif) umumnya tidak menolak program itu sendiri, melainkan menyoroti aspek pelaksanaan di lapangan, seperti keamanan pangan dan tantangan distribusi.

Sebaliknya, pada kebijakan fiskal seperti Kenaikan PPN 12% dan Efisiensi Anggaran, sentimen negatif yang lugas menjadi temuan utama, sesuai dengan hipotesis awal. Model IndoBERT menunjukkan performa yang sangat kuat dalam mengenali sentimen negatif dan topik "Efisiensi Anggaran". Ini mengindikasikan bahwa ketidaksetujuan dan kekhawatiran publik diekspresikan menggunakan

kosakata yang jelas, tajam, dan mudah dikenali oleh model, seperti "beban", "mahal", "pangkas", dan "korupsi".

Hal yang paling menarik ditemukan pada kebijakan Danantara. Hipotesis awal yang menduga adanya harapan (positif) dan kekhawatiran (negatif) ternyata kurang tepat. Hasil evaluasi model yang menunjukkan kesulitan dalam mengklasifikasikan sentimen pada topik ini, ditambah dengan banyaknya data yang akhirnya jatuh pada kategori netral, justru memperkuat dugaan adanya kesenjangan pemahaman publik. Banyaknya komentar yang ambigu atau hanya bertanya menunjukkan bahwa masyarakat cenderung mengambil sikap menunggu karena kurangnya informasi yang jelas mengenai lembaga investasi baru ini.

Dari pembahasan ini, dapat disimpulkan bahwa persepsi publik tidaklah utuh. Model IndoBERT berhasil menangkap dinamika ini: sentimen positif kuat pada kebijakan populis (MBG), sentimen negatif yang jelas pada kebijakan yang dianggap membebani (PPN & Efisiensi Anggaran), serta sentimen ambigu dan netral pada kebijakan baru yang kompleks dan kurang tersosialisasi (Danantara).

6.2 Keefektifan Model BERT dalam Analisis Sentimen

Penggunaan model BERT dalam analisis sentimen tidak hanya memberikan sejumlah keunggulan, tetapi juga memiliki beberapa keterbatasan. Secara umum, BERT mampu memahami konteks bahasa Indonesia dengan baik, termasuk dalam mengenali kata-kata tidak baku dan bahasa slang, karena telah melalui proses pra-pelatihan menggunakan korpus berbahasa Indonesia (seperti IndoBERT). Salah satu keunggulan utama BERT adalah kemampuannya membaca kalimat dari dua arah (*bidirectional*), yang memungkinkan model ini menghasilkan prediksi sentimen yang lebih akurat.

Namun demikian, model BERT juga memiliki beberapa kelemahan. Salah satunya adalah kurang optimal dalam menganalisis teks yang sangat pendek, seperti komentar di YouTube yang hanya terdiri dari beberapa kata atau didominasi oleh emoji. Selain itu, model ini juga rentan mengalami *overfitting*, terutama ketika digunakan pada dataset yang jumlahnya terbatas. Oleh karena itu, meskipun BERT

tergolong unggul dalam memahami konteks, penerapannya tetap perlu disesuaikan dengan karakteristik data yang digunakan.

6.3 Rekomendasi untuk Penelitian Lanjutan

Penelitian selanjutnya disarankan untuk mengembangkan cakupan dataset dengan menambahkan data dari berbagai platform media sosial lainnya, seperti TikTok, Facebook, dan Instagram, agar analisis mencakup representasi opini publik yang lebih luas. Proses pelabelan data juga sebaiknya dilakukan secara manual oleh ahli, guna meminimalkan bias yang dapat timbul dari pendekatan *lexicon-based*.

Selain itu, model analisis dapat disempurnakan dengan menggabungkan BERT dengan metode lain, seperti CNN, yang berpotensi lebih efektif dalam mengenali ekspresi sentimen melalui emoji atau simbol visual lainnya. Dari sisi implementasi kebijakan, pemerintah disarankan untuk memprioritaskan program-program yang memperoleh respons positif dari masyarakat, seperti program Makan Bergizi Gratis (MBG), serta melakukan evaluasi lebih mendalam terhadap kebijakan yang menuai kontroversi, seperti kenaikan tarif PPN sebesar 12%. Upaya membangun kanal komunikasi dua arah melalui media sosial juga perlu dilakukan, agar pemerintah dapat merespons kekhawatiran publik secara *real-time.*

6.4 Tantangan dan Solusi

Dalam proses analisis sentimen, terdapat beberapa tantangan yang perlu dihadapi, antara lain banyaknya komentar spam serta penggunaan bahasa gaul yang bervariasi, yang menyulitkan tahapan *preprocessing* data. Untuk mengatasi hal tersebut, solusi yang diterapkan adalah melakukan pembersihan data secara lebih ketat, termasuk menggunakan kamus slang yang terus diperbarui agar mampu mengenali dan menyesuaikan berbagai variasi kata tidak baku yang muncul dalam komentar-komentar pengguna.

BAB VII

KESIMPULAN DAN SARAN

Pada bab ini akan dijelaskan mengenai kesimpulan yang diperoleh dari hasil pengamatan dan pembahasan Tugas Akhir dan saran untuk penelitian selanjutnya.

7.1 Kesimpulan

Berdasarkan hasil analisis sentimen menggunakan model IndoBERT terhadap komentar publik di platform X dan YouTube, penelitian ini berhasil menjawab rumusan masalah dan menguji hipotesis yang telah diajukan. Temuan utama menunjukkan bahwa persepsi masyarakat terhadap empat kebijakan awal Presiden Prabowo Subianto didominasi oleh sentimen negatif. Dari keseluruhan komentar yang dianalisis, sebanyak 56,3% menunjukkan kecenderungan negatif terhadap keempat kebijakan tersebut.

Program Efisiensi Anggaran mendapatkan persentase komentar negatif tertinggi sebesar 63,4% diikuti oleh kebijakan PPN 12% sebesar 58,3%, Program Makan Bergizi Gratis sebesar 55,0%, dan Danantara sebesar 50,9%. Fakta ini menunjukkan bahwa bahkan kebijakan yang awalnya diasumsikan akan diterima secara positif, seperti program bantuan sosial Makan Bergizi Gratis, tetap menuai reaksi skeptis dari publik. Hal ini menandakan bahwa persepsi masyarakat sangat dipengaruhi oleh kekhawatiran terhadap pelaksanaan teknis, beban fiskal, serta kepercayaan terhadap lembaga pelaksana.

Dari sisi teknis, model IndoBERT yang dikembangkan dalam arsitektur *multi-task learning* terbukti menjadi alat yang efektif dan andal untuk menganalisis teks politik berbahasa Indonesia. Model menunjukkan kinerja sangat baik dalam klasifikasi topik dengan akurasi 81,38% dan kinerja yang cukup baik dalam klasifikasi sentimen dengan akurasi 73,95%. Meskipun andal dalam mendeteksi sentimen negatif, model teridentifikasi memiliki tantangan dalam membedakan komentar netral yang ambigu, yang menjadi area penting untuk perbaikan di masa mendatang. Secara keseluruhan, penelitian ini tidak hanya berhasil memetakan persepsi publik secara kuantitatif, tetapi juga mengonfirmasi bahwa analisis sentimen dalam

penelitian merupakan instrumen yang berharga untuk mengukur *respons* masyarakat terhadap kebijakan pemerintah di era digital.

7.2 Saran

Berdasarkan hasil yang telah diperoleh, pendekatan analisis sentimen dengan menggunakan model IndoBERT terbukti memberikan performa yang baik dalam memetakan persepsi publik terhadap kebijakan pemerintahan. Pemilihan model ini sesuai secara ilmiah, karena IndoBERT telah dirancang khusus untuk memahami struktur dan konteks bahasa Indonesia, serta mampu melakukan klasifikasi multitugas yang mendalam. Meskipun demikian, hasil penelitian juga menunjukkan bahwa masih terdapat tantangan dalam menangani komentar yang sangat pendek, bermakna ganda, atau disampaikan secara tidak eksplisit, yang berpotensi mempengaruhi akurasi klasifikasi.

Untuk itu, penelitian lanjutan disarankan mengkesplorasi pengguna dataset yang lebih luas, tidak hanya dari sisi jumlah, tetapi juga dari keberagaman gaya bahasa dan sumber media sosial. Komentar dari *platform* seperti TikTok, *Facebook,* atau Instagram dapat memberikan warna baru dalam ekspresi publik yang selama ini belum terwakili, sehingga hasil analisis menjadi *komprehensif.* Semakin beragam data yang digunakan, semakin tinggi peluang model dalam memahami nuansa opini yang kompleks dan dinamis.

Selain itu, meskipun pendekatan IndoBERT berbasis *deep learning* telah memberikan hasil yang menjanjikan, pendekatan lain yang lebih ringan atau berbasis metode tradisional juga tetap relevan untuk mempertimbangkan, terutama jika tujuan penelitian diarahkan pada efisien komputasi atau keterbatasan sumber daya. Pendekatan semacam ini, apabila dikembangkan dengan cermat dan dipadukan dengan strategi representasi data yang baik, dapat menjadi alternatif yang tidak kalah kompetitif, khususnya untuk kasus dengan volume data terbatas atau kebutuhan implementasi cepat.

Kualitas pelabelan data juga menjadi hal penting yang perlu diperhatikan. Penggunaan metode otomatis seperti pendekatan leksikon memang memberikan

efisiensi pada tahap awal, namun validasi manual dan semi-manual tetap diperlukan, terutama untuk komentar-komentar yang sulit diklasifikasikan secara jelas. Validasi ini berperan besar dalam meningkatkan ketepatan pelabelan, yang pada akhirnya akan berdampak langsung terhadap performa model secara keseluruhan.

Secara praktis, hasil dari penelitian ini dapat menjadi dasar yang kuat bagi instansi pemerintah atau pemangku kebijakan dalam memahami opini publik secara lebih sistematis. Kebijakan yang memunculkan respons negatif dapat ditindaklanjuti dengan pendekatan komunikasi yang lebih terbuka dan menjelaskan secara transparan alasan di balik kebijakan tersebut. Sebaliknya, kebijakan yang mendapatkan tanggapan positif dimaksimalkan sebagai model sosial dalam membangun kepercayaan publik. Selain itu, sistem analisis seperti yang dikemangkan dalam penelitian ini dapat diadaptasi lebih lanjut sebagai alat bantu untuk pemantauan opini publik secara *real-time* dan berbasis data.

DAFTAR PUSTAKA

[1] F. D. Pratama and U. Lampung, "Sentiment of Youtube Social Media Users Towards the Presidential Candidates of Ganjar Pranowo and Prabowo Subianto in the 2024 Election," vol. 9, no. April, pp. 19–36, 2024, doi: 10.15294/ipsr.v9i1.1910.

[2] A. N. Muhammad, S. Bukhori, and P. Pandunata, "Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes-Support Vector Machine (NBSVM) Classifier," *Proc. - 2019 Int. Conf. Comput. Sci. Inf. Technol. Electr. Eng. ICOMITEE 2019*, vol. 1, pp. 199–205, 2019, doi: 10.1109/ICOMITEE.2019.8920923.

[3] S. Pokhrel, "No TitleΕΛΕΝΗ," *Αγαη*, vol. 15, no. 1, pp. 37–48, 2024.

[4] A. Qodir and M. Rusdin, *Political discourse analysis of joko widodo's political dynasty issues in jakarta post and tempo reporting thesis*. 2024.

[5] Muhammad Devit and Nanang Ridwan, "Dampak Framing Hasil Survey Pemilihan Presiden 2024 Mempengaruhi Persepsi Publik," *J. Kaji. Ilm.*, vol. 23, no. 3, pp. 233–242, 2023, doi: 10.31599/jki.v23i3.2805.

[6] J. Chen and S. Xia, "Are Online Users Influenced by What Other Users Say? Meta-Analyzing the Cognitive, Emotional, and Behavioral Impact of Online Comment Valence," *Cyberpsychology*, vol. 18, no. 5, 2024, doi: 10.5817/CP2024-5-8.

[7] S. S. Sohail, M. M. Khan, and M. A. Alam, "An Analysis of Twitter Users From The Perspective of Their Behavior, Language, Region and Development Indices -- A Study of 80 Million Tweets," 2021, [Online]. Available: http://arxiv.org/abs/2105.10245

[8] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.

[9] Y. Gunawan, I. Purnama, and R. Rohani, "Sentiment Analysis of Twitter towards the 2024 Indonesian Presidential Candidates Using the Naïve Bayes Algorithms," *Int. J. Sci. Technol. Manag.*, vol. 5, no. 4, pp. 953–961, 2024, doi: 10.46729/ijstm.v5i4.1154.

[10] A. R. Shaikh, H. Alhoori, and M. Sun, “YouTube and science: models for research impact,” *Scientometrics*, vol. 128, no. 2, pp. 933–955, 2023, doi: 10.1007/s11192-022-04574-5.

[11] N. M. Samsudin, C. F. B. Mohd Foozy, N. Alias, P. Shamala, N. F. Othman, and W. I. S. Wan Din, “Youtube spam detection framework using naïve bayes and logistic regression,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 14, no. 3, pp. 1508–1517, 2019, doi: 10.11591/ijeecs.v14.i3.pp1508-1517.

[12] U. K. Immanuel, “Sentiment Analysis of Public Opinions Regarding " Ideas of Presidential Candidates " in YouTube Video Comments with Robustly Optimized BERT Pretraining Approach,” vol. 8, no. 1, pp. 12–28, 2024.

[13] Harman Akbar Tullah *et al.*, “Sentiment Analysis Of Indonesian Civil Servan Candidates 2023 Twitter Network With Naive Bayes Algorithm Method,” *Inspir. J. Teknol. Inf. dan Komun.*, vol. 13, no. 2, pp. 49–63, 2023, doi: 10.35585/inspir.v13i2.66.

[14] M. Ahmad, S. Aftab, M. S. Bashir, and N. Hameed, “Sentiment analysis using SVM: A systematic literature review,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, pp. 182–188, 2018, doi: 10.14569/IJACSA.2018.090226.

[15] M. Al-Khazaleh, M. Alian, M. Biltawi, and B. Al-Hazaimeh, “Sentiment Analysis for People’s Opinions about COVID-19 Using LSTM and CNN Models,” *Int. J. online Biomed. Eng.*, vol. 19, no. 1, pp. 135–154, 2023, doi: 10.3991/ijoe.v19i01.35645.

[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[17] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, 2018, doi: 10.1002/widm.1253.

[18] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, “Research on the application of deep learning-based BERT model in sentiment analysis,” *Appl. Comput. Eng.*, vol. 71, no. 1, pp. 14–20, 2024, doi: 10.54254/2755-2721/71/2024ma.

[19] I. G. and Y. B. and A. Courville, “Deep learning 简介 一 、

什么是 Deep Learning ?," *Nature*, vol. 29, no. 7553, pp. 1–73, 2016, [Online]. Available: http://deeplearning.net/

[20] A. R. Abas, I. Elhenawy, M. Zidan, and M. Othman, "Bert-cnn: A deep learning model for detecting emotions from text," *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 2943–2961, 2022, doi: 10.32604/cmc.2022.021671.

[21] P. Sayarizki and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," *J. Comput.*, vol. 9, no. 2, pp. 61–72, 2024, doi: 10.34818/indojc.2024.9.2.934.

[22] A. Ahmad Aliero, B. Sulaimon Adebayo, H. Olanrewaju Aliyu, A. Gogo Tafida, B. Umar Kangiwa, and N. Muhammad Dankolo, "Systematic Review on Text Normalization Techniques and its Approach to Non-Standard Words," *Int. J. Comput. Appl.*, vol. 185, no. 33, pp. 975–8887, 2023.

[23] B. Sengupta, E. Vasquez, and Y. Qian, "Deep Tensor Encoding," 2017, [Online]. Available: http://arxiv.org/abs/1703.06324

[24] R. Feldman, "Techniques and applications for sentiment analysis: The main applications and challenges of one of the hottest research areas in computer science," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, 2013, doi: 10.1145/2436256.2436274.

[25] B. Ramadhani and R. R. Suryono, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 714, 2024, doi: 10.30865/mib.v8i2.7458.

[26] F. Yunina, "Analisis Efektivitas Dan Efisiensi Pelaksanaan Anggaran Belanja Badan Perencanaan Pembangunan Daerah (Bappeda) Aceh Tengah," *J. Akunt. Muhammadiyah*, vol. 9, no. 1, pp. 74–81, 2019, doi: 10.37598/jam.v9i1.647.

[27] N. R. Oktaviani and V. C. Mulyandani, "Analisis Efektivitas Dan Efisiensi Anggaran Belanja Langsung Perwakilan Badan Kependudukan Dan Keluarga Berencana Nasional Provinsi Jawa Barat Tahun Anggaran 2016-2020," *Indones. Account. Lit. J.*, vol. 2, no. 2, pp. 484–493, 2022, doi: 10.35313/ialj.v2i2.3219.

[28] I. M. Putri, "Kenaikan Ppn 12% Dan Dampaknya Terhadap Eknomi," *J. Ilm. Manajemen, Ekon. Akunt.*, vol. 8, no. 2, pp. 934–944, 2024, doi:

10.31955/mea.v8i2.4077.

[29] S. K. Indihome, M. Bagas, D. Putra, and E. Setiawan, "METODE LEXICON BASED UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KINERJA ISP," vol. 8, no. 6, pp. 12079–12087, 2024.

[30] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," *Proc. 7th Int. Conf. Lang. Resour. Eval. Lr. 2010*, pp. 1320–1326, 2010, doi: 10.17148/ijarcce.2016.51274.

[31] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, no. December, pp. 391–394, 2017, doi: 10.1109/IALP.2017.8300625.

[32] J. Jones, W. Jiang, N. Synovic, G. K. Thiruvathukal, and J. C. Davis, "What do we know about Hugging Face? A systematic literature review and quantitative validation of qualitative claims," 2024, doi: 10.1145/3674805.3686665.

[33] L. A. Kumar and D. K. Renuka, "State-of-the-Art Natural Language Processing," *Deep Learn. Approach Nat. Lang. Process. Speech, Comput. Vis.*, pp. 49–75, 2023, doi: 10.1201/9781003348689-3.

[34] W. Jiang *et al.*, "An Empirical Study of Pre-Trained Model Reuse in the Hugging Face Deep Learning Model Registry," *Proc. - Int. Conf. Softw. Eng.*, pp. 2463–2475, 2023, doi: 10.1109/ICSE48619.2023.00206.

[35] P. Deshpande and S. Jahirabadkar, "Comparative Analysis of Lexical Chain, Bidirectional Encoder Representations from Transformers (BERT) and Graph based Approaches for Condensation of Low Resource Language Documents," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 16s, pp. 08–15, 2024.

[36] K. H. Sasongko and A. M. Hilda, "Analisis Sentimen Terhadap Film ' Dirty Vote ' Pada Media Sosial X dan Youtube dengan Algoritma Naive Bayes dan SVM," vol. 6, no. 1, pp. 73–85, 2024, doi: 10.47065/josyc.v6i1.6150.

[37] Anugerah Simanjuntak *et al.*, "Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 1, pp. 60–67, 2024, doi: 10.22146/jnteti.v13i1.8532.

[38] Y. Nooryuda Prasetya and D. Winarso, "Penerapan Lexicon Based Untuk

Analisis Sentimen Pada TwiterTerhadap Isu Covid-19," *J. Fasilkom*, vol. 11, no. 2, pp. 97–103, 2021, [Online]. Available: https://ejurnal.umri.ac.id/index.php/JIK/article/view/2772/1566

[39] A. S. dan N. Surojudin, "Analisis Dan Perbandingan Stemming Algoritma Porter Dengan Algoritma Ahmad Yusoff Sembok Dalam Dokumen Teks Bahasa Indonesia," *Pros. Semin. SeNTIK*, vol. 4, no. 1, pp. 347–357, 2020, [Online]. Available: https://ejournal.jak-stik.ac.id/index.php/sentik/article/view/3304

[40] *Pencarian peraturan*. (n.d.). Database Peraturan | JDIH BPK. https://peraturan.bpk.go.id/Search?keywords=efisiensi+anggaran&tentang=&nomor=