

HALAMAN MOTTO

“Allah akan meninggikan orang-orang yang beriman diantaramu dan orang-orang yang diberi ilmu pengetahuan beberapa derajat.”

(Q.S. Al-Mujadilah: 11)

“Selesaikan apa yang sudah kamu mulai, lewati semua badai dan terjang semuanya tanpa mengubah tujuannya”



HALAMAN PERSEMBAHAN

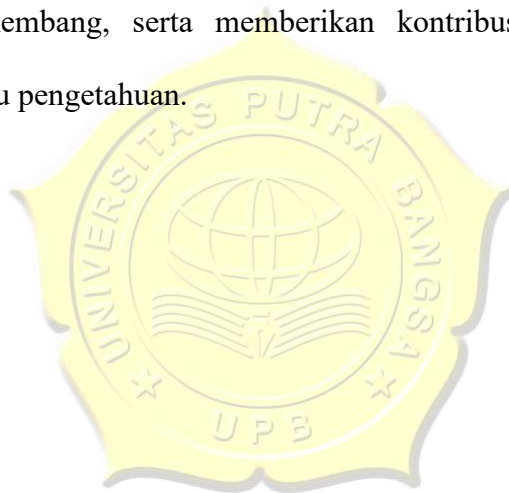
Dengan penuh rasa syukur kepada Allah SWT atas segala rahmat, nikmat, kesehatan, kekuatan, dan kemudahan yang telah diberikan sehingga skripsi ini dapat terselesaikan dengan baik. Karya sederhana ini penulis persembahkan kepada:

1. Allah SWT, Tuhan Yang Maha Esa, sebagai wujud rasa syukur atas segala karunia, petunjuk, dan pertolongan-Nya yang senantiasa menyertai setiap langkah penulis dalam menempuh pendidikan hingga menyelesaikan skripsi ini.
2. Kedua orang tua tercinta, bapak Amir Khasan dan Ibu Sri Sumarni yang menjadi sumber kekuatan, doa, kasih sayang, pengorbanan, serta dukungan tanpa batas. Dari kalian, penulis belajar arti kesabaran, ketulusan, dan kerja keras. Segala jeri payah yang kalian berikan menjadi sumber kekuatan terbesar dalam setiap langkah. Setiap doa yang bapak dan ibu panjatkan, setiap restu yang kalian berikan, dan setiap peluh yang kalian curahkan merupakan alasan penulis mampu bertahan dan menyelesaikan perjalanan panjang ini. Terima kasih atas segala perjuangan, keikhlasan, dan cinta yang telah diberikan. Semoga karya ini menjadi salah satu bentuk bakti dan kebanggaan bagi Bapak dan Ibu.
3. Kedua kakakku, Umi Ngafifatul Marghubah dan Anis Latifah, yang selalu memberikan doa, semangat, motivasi, serta dukungan selama proses perkuliahan hingga penyusunan skripsi ini. Terima kasih atas perhatian dan

dorongan yang diberikan sehingga penulis mampu menyelesaikan perjalanan ini.

4. Terakhir untuk penulis, terima kasih yang telah berjuang, bertahan, dan tidak menyerah dalam menghadapi berbagai tantangan selama proses perkuliahan dan penyusunan skripsi. Terima kasih telah terus melangkah meskipun dihadapkan pada berbagai kesulitan, keraguan, dan hambatan. Semoga pencapaian ini menjadi awal dari perjalanan yang lebih baik di masa depan.

Semoga skripsi ini dapat memberikan manfaat dan menjadi langkah awal untuk terus belajar, berkembang, serta memberikan kontribusi yang positif bagi masyarakat dan ilmu pengetahuan.



ABSTRAK

Program Makan Bergizi Gratis (MBG) merupakan salah satu program strategis pemerintah yang memunculkan beragam tanggapan masyarakat di media sosial, sehingga diperlukan analisis sentimen untuk memahami persepsi publik secara sistematis. Penelitian ini bertujuan menganalisis distribusi sentimen masyarakat terhadap Program MBG, mengevaluasi performa kombinasi metode IndoBERT–*Support Vector Machine* (SVM), serta menganalisis pengaruh distribusi data melalui penerapan *class weight*. *Class weight* merupakan teknik pembobotan kelas dengan memberikan bobot berbeda pada setiap kelas berdasarkan jumlah data, sehingga kelas minoritas memperoleh bobot lebih besar dibandingkan kelas mayoritas. Data penelitian berupa 3.590 komentar TikTok yang diperoleh melalui proses *scraping*, kemudian diproses melalui tahapan *pre-processing*, *hybrid labelling* (pelabelan otomatis, validasi berbasis *lexicon*, dan anotasi manual pada data ambigu), pembagian data latih dan data uji, ekstraksi fitur menggunakan *embedding* IndoBERT, standarisasi fitur, optimasi parameter menggunakan GridSearchCV, serta klasifikasi menggunakan SVM dengan empat skenario pengujian. Hasil penelitian menunjukkan bahwa distribusi sentimen didominasi oleh sentimen negatif sebanyak 2.106 komentar, diikuti sentimen positif sebanyak 745 komentar dan sentimen netral sebanyak 739 komentar. Performa terbaik diperoleh pada skenario dataset *non-hybrid* dengan penerapan *class weight*, yang menghasilkan akurasi 83,70%, *precision* 84,43%, *recall* 77,74%, *F1-score* 80,43%, dan MCC 71,11%. Selain itu, penerapan *class weight* terbukti meningkatkan performa model pada dataset yang masih mengalami ketidakseimbangan kelas, sedangkan *hybrid labelling* tidak meningkatkan performa klasifikasi, tetapi berhasil menghasilkan *ground truth* yang lebih berkualitas melalui validasi manual terhadap data ambigu. Temuan ini menunjukkan bahwa kombinasi *embedding* IndoBERT dan SVM efektif untuk analisis sentimen berbahasa Indonesia, sementara strategi penanganan distribusi data dan kualitas pelabelan perlu disesuaikan dengan karakteristik dataset agar menghasilkan model yang lebih andal.

Kata kunci: Analisis sentimen, IndoBERT, *Support Vector Machine* (SVM), Program Makan Bergizi Gratis, TikTok.

ABSTRACT

The Free Nutritious Meal Program (MBG) is one of the government's strategic programs that has elicited various responses from the public on social media, thus requiring sentiment analysis to systematically understand public perception. This study aims to analyze the distribution of public sentiment toward the MBG Program, evaluate the performance of the IndoBERT–Support Vector Machine (SVM) method combination, and analyze the impact of data distribution thru the application of class weight. Class weight is a class weighting technique that assigns different weights to each class based on the amount of data so that the minority class receives a higher weight compared to the majority class. The research data consists of 3,590 TikTok comments obtained thru a scraping process, which were then processed thru pre-processing stages, hybrid labeling (automatic labeling, lexicon-based validation, and manual annotation on ambiguous data), splitting into training and testing data, feature extraction using IndoBERT embedding, feature standardization, parameter optimization using GridSearchCV, and classification using SVM with four testing scenarios. The research results show that the sentiment distribution is dominated by negative sentiment with 2,106 comments, followed by positive sentiment with 745 comments and neutral sentiment with 739 comments. The best performance was achieved in the non-hybrid dataset scenario with the application of class weight, resulting in an accuracy of 83.70%, precision of 84.43%, recall of 77.74%, F1-score of 80.43%, and MCC of 71.11%. Additionally, the application of class weight proved to improve the model's performance on the dataset that still experienced class imbalance, while hybrid labeling did not enhance classification performance but successfully produced higher quality ground truth thru manual validation of ambiguous data. These findings indicate that the combination of IndoBERT embeddings and SVM is effective for Indonesian sentiment analysis, while strategies for handling data distribution and labeling quality need to be adjusted according to the dataset's characteristics to produce a more reliable m

Keywords: Sentiment analysis, IndoBERT, Support Vector Machine (SVM), Free Nutritious Meal Program, TikTok.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Publik Terhadap Program Makan Bergizi Gratis (MBG) pada Platform TikTok Menggunakan Metode IndoBERT-SVM (*Support Vector Machine*)” dengan baik. Shalawat serta salam senantiasa tercurah kepada Nabi Muhammad SAW yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang penuh dengan ilmu pengetahuan seperti saat ini.

Penyusunan skripsi ini dilakukan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika Universitas Putra Bangsa. Selain itu, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Program Makan Bergizi Gratis (MBG) yang berkembang di media sosial TikTok menggunakan kombinasi metode IndoBERT sebagai ekstraksi fitur dan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi. Penelitian ini diharapkan dapat memberikan gambaran mengenai opini masyarakat terhadap program tersebut sekaligus menjadi referensi dalam pengembangan penelitian analisis sentimen di masa mendatang.

Dalam penyusunan skripsi ini, penulis menyadari bahwa keberhasilan penelitian tidak terlepas dari bantuan, bimbingan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Allah SWT atas segala rahmat, karunia, kesehatan, kekuatan, serta kemudahan yang diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua dan keluarga tercinta yang senantiasa memberikan doa, kasih sayang, dukungan, motivasi, dan pengorbanan yang tiada henti selama proses perkuliahan hingga penyusunan skripsi ini.
3. Bapak Awwaludin 'Abid, S.Kom., M.kom selaku Kepala Program Studi Informatika Universitas Putra Bangsa yang telah memberikan arahan dan dukungan selama masa perkuliahan.
4. Ibu Wahyuni Windasari, S.Si., M.Sc., selaku dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, masukan, serta motivasi selama proses penyusunan skripsi ini. Penulis juga menyampaikan terima kasih kepada seluruh dosen penguji yang telah memberikan kritik dan saran yang membangun demi penyempurnaan skripsi ini.
5. Seluruh dosen Program Studi Informatika Universitas Putra Bangsa yang telah memberikan ilmu pengetahuan, pengalaman, dan wawasan selama masa perkuliahan.
6. Teman-teman mahasiswa Program Studi Ilmu Komputer yang telah memberikan dukungan, bantuan, serta semangat selama proses penelitian dan penyusunan skripsi.

7. Sahabat penulis, Amira Nur Amalina, yang senantiasa memberikan dukungan, motivasi, perhatian, serta menjadi tempat berbagi cerita dan semangat selama proses perkuliahan hingga penyusunan skripsi ini.
8. Teman-teman magang yang hingga saat ini masih memberikan dukungan, bantuan, dan semangat kepada penulis dalam menyelesaikan skripsi ini.
9. Sahabat penulis, Cahya Maylinda dan Nurul Maftukhah, yang telah memberikan kontribusi, bantuan, dukungan, motivasi, serta semangat selama proses penyusunan skripsi ini. Terima kasih atas kebersamaan, perhatian, dan kesediaannya untuk selalu membantu, mendengarkan keluh kesah, serta memberikan hiburan dan dorongan ketika penulis menghadapi berbagai tantangan dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi perbaikan penelitian di masa yang akan datang. Penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca, menjadi referensi bagi penelitian selanjutnya, serta memberikan kontribusi dalam pengembangan ilmu pengetahuan khususnya pada bidang *Natural Language Processing* (NLP), analisis sentimen, dan data mining.

DAFTAR ISI

HALAMAN SAMPUL DEPAN	i
HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN UJIAN	iv
HALAMAN BEBAS PLAGIARISME	v
HALAMAN MOTTO	v
HALAMAN PERSEMBAHAN	vii
ABSTRAK	ix
<i>ABSTRACT</i>	x
KATA PENGANTAR.....	xi
DAFTAR ISI	xiv
DAFTAR TEBEL.....	xvii
DAFTAR GAMBAR	xviii
DAFTAR LAMPIRAN.....	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah.....	7
1.4 Tujuan Penelitian.....	8
1.5 Manfaat Penelitian	9
BAB II KAJIAN PUSTAKA	11
2.1 Tinjauan Teori	11
2.1.1 Analisis Sentimen	11
2.1.2 Media Sosial.....	12
2.1.3 Program Makan Bergizi Gratis (MBG).....	14
2.1.4 <i>Crawling Data</i>	15
2.1.5 <i>Preprocessing Data</i>	16
2.1.6 <i>Lexicon Based</i>	18
2.1.7 <i>Grid Search Cross-Validation</i>	19

2.1.8	<i>Class Weight</i>	20
2.1.9	IndoBERT	22
2.1.10	<i>Support Vector Machine (SVM)</i>	24
2.1.11	<i>Confusion Matrix</i>	27
2.1.12	<i>Word Cloud</i>	30
2.1.13	<i>Google Colaboratory</i>	31
2.1.14	<i>Python</i>	32
2.2	Penelitian Terdahulu.....	32
2.3	Kerangka Pemikiran.....	36
2.4	Keaslian Penelitian.....	39
BAB III METODE PENELITIAN.....		41
3.1	Metode Penelitian.....	41
3.2	Tahapan Penelitian	41
3.2.1	Identifikasi Permasalahan	43
3.2.2	Perumusan Masalah	43
3.2.3	Studi Literatur	44
3.2.4	Pengumpulan Data	44
3.2.5	<i>Preprocessing Data</i>	46
3.2.6	<i>Hybrid Labeling</i>	48
3.2.7	Split Data.....	51
3.2.8	Ekstraksi Fitur	52
3.2.9	Standarisasi Fitur.....	54
3.2.10	<i>Hyperparameter Tuning</i>	56
3.2.11	<i>Class Weight</i>	57
3.2.12	Metode <i>Support Vector Machine (SVM)</i>	58
3.2.13	Evaluasi Model.....	59
3.2.14	Visualisasi Hasil	61
3.2.15	Hasil dan Kesimpulan	62
3.3	Metode Pengumpulan Data	62
3.4	Alat dan Bahan Penelitian.....	63
BAB IV HASIL DAN PEMBAHASAN.....		66

4.1	Pengumpulan Data	66
4.2	<i>Pre-processing</i> Data	67
4.2.1	<i>Cleaning</i>	68
4.2.2	<i>Handling Duplicate</i>	68
4.2.3	<i>Case Folding</i>	69
4.2.4	Normalisasi Slang	70
4.3	<i>Hybrid Labelling</i>	71
4.3.1	Pelabelan Otomatis (Model <i>Pre-trained</i>)	71
4.3.2	Validasi <i>Lexicon</i>	73
4.3.3	Anotasi Manual	76
4.3.4	<i>Intra-Annotator Consistency Check</i>	79
4.3.5	Data <i>Hybrid</i>	81
4.4	Split Data	82
4.5	Ekstraksi Fitur	83
4.6	Standarisasi Fitur	84
4.7	<i>Hyperparameter Tuning</i>	85
4.8	Penyeimbang Kelas (<i>Class Weight</i>)	86
4.9	Pemodelan <i>Support Vektor Machine</i> (SVM)	87
4.10	Evaluasi Model dan Analisis Hasil	90
4.10.1	Evaluasi skenario 1 (Dataset <i>Non-Hybrid</i> Tanpa <i>Class Weight</i>) ...	90
4.10.2	Evaluasi skenario 2 (Dataset <i>Non-Hybrid</i> dengan <i>Class Weight</i>) .	92
4.10.3	Evaluasi skenario 3 (Dataset <i>Hybrid</i> Tanpa <i>Class Weight</i>)	94
4.10.4	Evaluasi skenario 4 (Dataset dengan <i>Class Weight</i>)	95
4.10.5	Perbandingan Hasil Evaluasi Antar Skenario	97
4.11	Visualisasi	100
BAB V PENUTUP		105
5.1	Kesimpulan	105
5.2	Saran	106
DAFTAR PUSTAKA		108
LAMPIRAN		115

DAFTAR TEBEL

Tabel II - 1. <i>Confusion Matrix</i> Multikelas	28
Tabel II - 2. Penelitian Terdahulu	33
Tabel III - 1. Proses <i>Cleaning</i>	47
Tabel III - 2. Proses <i>Handling Duplicate</i>	47
Tabel III - 3. Proses <i>Case Folding</i>	48
Tabel III - 4. Proses Normalisasi	48
Tabel III - 5. <i>Inter-rater Reliability Cohen's Kappa</i>	50
Tabel III - 6. Kategori Performa	61
Tabel III - 7. Alat Penelitian	64
Tabel III - 8. <i>Library</i> Penelitian	64
Tabel III - 9. Bahan Penelitian	65
Tabel IV - 1. Hasil Filter Data	67
Tabel IV - 2. Hasil Jumlah <i>Handling Duplicate</i>	69
Tabel IV - 3. Aturan Pelabelan Hibrida	74
Tabel IV - 4. Pedoman Anotasi Sentimen	77
Tabel IV - 5. Distribusi Data Ambigu Annotasi Manual	78
Tabel IV - 6. Distribusi Split Dataset <i>Non-Hybrid</i> dan <i>Hybrid</i>	83
Tabel IV - 7. Ruang Pencarian <i>Hyperparameter</i>	85
Tabel IV - 8. Hasil Perhitungan Bobot Kelas Data <i>Non-Hybrid</i> dan <i>Hybrid</i>	87
Tabel IV - 9. Skenario Pemodelan	88
Tabel IV - 10. Distribusi Jumlah Data setiap Tahapan	88
Tabel IV - 11. Perbandingan <i>Classification Report</i> Antar Skenario	98
Tabel IV - 12. Distribusi Prediksi Setiap Skenario	103

DAFTAR GAMBAR

Gambar II - 1. Tren Pengguna Internet dan Media Sosial di Indonesia.....	13
Gambar II - 2. Perkembangan Pengguna Media Sosial di Indonesia.....	13
Gambar II - 3. Gambar Alur Kerangka Pemikiran.....	37
Gambar III - 1. Tahapan Penelitian	42
Gambar III - 2. Alur <i>Preprocessing</i> Data	46
Gambar III - 3. Tahapan <i>Embedding</i> IndoBERT.....	53
Gambar III - 4. TikTok <i>Comments Scraper</i> pada Apify.....	63
Gambar IV - 1. Hasil <i>Scraping</i> Apify	66
Gambar IV - 2. Hasil <i>Cleaning</i>	68
Gambar IV - 3. Hasil <i>Handling Duplicate</i>	69
Gambar IV - 4. Hasil <i>Case Folding</i>	70
Gambar IV - 5. Hasil Normalisasi Slang	71
Gambar IV - 6. Hasil Label Pelabelan <i>Non-Hybrid</i>	72
Gambar IV - 7. Distribusi Data <i>Non-Hybrid</i>	72
Gambar IV - 8. Distribusi <i>Confidence Score</i> dari Model.....	73
Gambar IV - 9. Status Data Berdasarkan <i>Threshold</i>	75
Gambar IV - 10. Hasil <i>Labeling Lexicon</i> dan Pelabelan Hibrida	75
Gambar IV - 11. Distribusi Label pada Data Ambigu.....	76
Gambar IV - 12. Hasil Pelabelan Manual Ketiga Anotator	78
Gambar IV - 13. <i>Confusion Matrix Intra-Annotator Consistency</i>	80
Gambar IV - 14. Hasil Dataset <i>Hybrid</i>	81
Gambar IV - 15. Distribusi Dataset <i>Hybrid</i>	82
Gambar IV - 16. Dimensi Fitur Hasil <i>Embedding</i>	84
Gambar IV - 17. Hasil Satandarisasi.....	84
Gambar IV - 18. Hasil Pencarian <i>Hyperparameter</i>	86
Gambar IV - 19. <i>Confusion Matrix</i> Skenario Pertama.....	90
Gambar IV - 20. <i>Classification Report</i> Skenario Pertama.....	91
Gambar IV - 21. <i>Confusion Matrix</i> Skenario Kedua	92
Gambar IV - 22. <i>Classification Report</i> Skenario Kedua.....	93
Gambar IV - 23. <i>Confusion Matrix</i> Skenario Ketiga	94

Gambar IV - 24. <i>Classification Report</i> Skenario Ketiga	95
Gambar IV - 25. <i>Confusion Matrix</i> Skenario Keempat	96
Gambar IV - 26. <i>Classification Report</i> Skenario Keempat.....	97
Gambar IV - 27. <i>Word Cloud</i> Sentimen Positif.....	100
Gambar IV - 28. <i>Word Cloud</i> Sentimen Netral	101
Gambar IV - 29. <i>Word Cloud</i> Sentimen Negatif	102



DAFTAR LAMPIRAN

Lampiran 1 Kode Program.....	116
Lampiran 2 Kamus Inset (<i>Indonesia Sentiment</i>).....	158
Lampiran 3 Link Dataset, Kamus InSet dan Hasil Analisis.....	159
Lampiran 4 Kartu Bimbingan	160
Lampiran 5 Kartu Sempro.....	161

