

**ANALISIS SENTIMEN PEMBELIAN BAHAN BAKAR
MINYAK PADA APLIKASI MyPertamina DENGAN METODE
NAIVE BAYES CLASSIFIER DAN *SYNTHETIC MINORITY
OVERSAMPLING TECHNIQUE***

SKRIPSI

**Untuk Memenuhi Persyaratan
Memperoleh Gelar Sarjana**

Oleh:

**I Komang Damai Armawan
19.51.0026**



PROGRAM STUDI S-1 SISTEM INFORMASI

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
PPKIA PRADNYA PARAMITA
MALANG
2022**

LEMBAR PERSETUJUAN

**ANALISIS SENTIMEN PEMBELIAN BAHAN BAKAR
MINYAK PADA APLIKASI MyPertamina DENGAN METODE
NAIVE BAYES CLASSIFIER DAN *SYNTHETIC MINORITY
OVERSAMPLING TECHNIQUE***

**Oleh:
I Komang Damai Armawan
19.51.0026**

**Telah disetujui oleh pembimbing untuk diujikan
Pada tanggal 9 Mei 2023**

Menyetujui,
Komisi Pembimbing

Pembimbing I

Pembimbing II

Dr. Mochamad Husni, S.Pd.I., MM., MMSI
NIK 01.09.10.036

Dr. Tubagus M. Akhriza., S.Si., MMSI., Ph.D
NIK 00.12.01.002

Mengetahui,
Ketua Program Studi S-1 Sistem Informasi

Dr. Dwi Safiroh Utsalina, S.Kom., MMSI
NIK 00.12.01.016

LEMBAR PENGESAHAN

**ANALISIS SENTIMEN PEMBELIAN BAHAN BAKAR
MINYAK PADA APLIKASI MyPertamina DENGAN METODE
NAIVE BAYES CLASSIFIER DAN *SYNTHETIC MINORITY
OVERSAMPLING TECHNIQUE***

**Skripsi oleh I Komang Damai Armawan
Telah dipertahankan di depan dewan penguji
Pada tanggal 19 Mei 2023**

Dewan Penguji,
Ketua

Dr. Rahayu Widayanti, S.E., M.M., MMSI
NIK 00.12.01.006

Anggota I

Eka Yuniar, S.Kom., MMSI
NIK 16.09.19.001

Anggota II

Dr. Mochamad Husni, S.Pd.I., MM., MMSI
NIK 01.09.10.036

Mengetahui,
Ketua Program Studi
S-1 Sistem Informasi

Mengesahkan,
Ketua STMIK PPKIA
Pradnya Paramita

Dr. Dwi Safiroh Utsalina, S.Kom., MMSI
NIK 00.12.01.016

Dr. Tubagus M. Akhriza., S.Si., MMSI., Ph.D
NIK 00.12.01.002

**PERNYATAAN
ORISINALITAS SKRIPSI**

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam Naskah SKRIPSI/TUGAS AKHIR ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu Perguruan Tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah SKRIPSI/TUGAS AKHIR ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang saya peroleh (SARJANA/AHLI MADYA) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No 20 Tahun 2003, pasal 25 ayat 2 dan pasal 70).

Malang, 7 Juli 2023

Mahasiswa,

Materai 10000

Nama : I Komang Damai Armawan

NIM : 19.51.0026

ABSTRAK

Armawan, I Komang Damai. 2023. *Analisis Sentimen Pembelian Bahan Bakar Minyak Pada Aplikasi MyPertamina Dengan Metode Naive Bayes Classifier Dan Synthetic Minority Oversampling Technique*. Program Studi Sistem Informasi STMIK PPKIA Pradnya Paramita. Pembimbing: (I) Dr. Tb. Mohammad Akhriza, S.SI, MMSI Ph.D, (II) Dr. Mochamad Husni, S.PdI., MM., MMSI

Kata Kunci: *Natural Language Processing, Klasifikasi, Naive Bayes Classifier (NBC), Synthetic Minority Oversampling Technique (SMOTE)*

Penerapan kebijakan aplikasi MyPertamina untuk pembelian BBM Bersubsidi mendapat respon yang beragam dari masyarakat, yang dituangkan kedalam media sosial. Respon masyarakat tersebut terbagi menjadi tanggapan netral, positif, dan juga negatif. Analisa manual bisa dilakukan untuk mengklasifikasikan semua tanggapan masyarakat ini kedalam ketiga kelas tersebut, namun akan memakan waktu yang panjang. Analisis sentimen dengan menggunakan metode *Naive Bayes Classifier* (NBC) bisa digunakan untuk klasifikasi tanggapan masyarakat mengenai penerapan aplikasi MyPertamina untuk pembelian BBM Bersubsidi dengan cepat dan akurat. Tujuan penelitian ini adalah untuk melakukan analisis sentimen terhadap penerapan aplikasi MyPertamina untuk pembelian BBM Subsidi pada masyarakat menggunakan Metode NBC dan mengimplementasikan *Synthetic Minority Oversampling Technique* (SMOTE). Pemisahan *dataset* dalam penelitian ini menggunakan tiga rasio yaitu 30% *testing set*, 40%, dan 50%. Hasil analisis sentimen dengan mengimplementasikan metode klasifikasi NBC dan SMOTE dengan rasio 30% *training set* menunjukkan hasil yang paling baik. Data awal sebanyak 972 data yang kemudian melalui proses *preprocessing* menjadi 712, kemudian algoritma SMOTE diimplementasikan untuk menyeimbangkan *training set*, menghasilkan 38% merupakan tanggapan netral, 35% positif, dan 27% negatif dengan akurasi 84%.

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Tuhan Yang Maha Esa atas segala rahmat dan hidayah-Nya peneliti dapat menyelesaikan skripsi dengan baik sebagai syarat untuk memperoleh gelar Sarjana Komputer..

Kami menyadari tanpa adanya dukungan dan kerja sama dari berbagai pihak, kegiatan skripsi ini tidak akan dapat berjalan baik. Untuk itu, kami ingin menyampaikan rasa terima kasih kepada:

1. Tuhan Yang Maha Esa yang telah memberikan petunjuk dan hidayah dalam pembuatan skripsi dan penyusunan laporan sehingga dapat berjalan dengan baik dari awal hingga akhir.
2. Kedua orang tua kami yang telah memberikan doa dan dukungannya.
3. Bapak Dr. Tubagus M. Akhriza, S.SI, MMSI, Ph.D, selaku Ketua STMIK PPKIA Pradnya Paramita.
4. Bapak Dr. Dwi Safiroh Utsalina, S.Kom., MMSI, selaku Ketua Program Studi S-1 Sistem Informasi.
5. Bapak Dr. Mochamad Husni, S.Pd.I., MM., MMSI, dan Bapak Dr. Tubagus M. Akhriza., S.Si., MMSI., Ph.D selaku dosen pembimbing skripsi.
6. Teman-teman dari Program Studi Sistem Informasi yang selalu memberikan semangat dan dukungan selama penyelesaian skripsi ini.
7. Dan seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan Skripsi dari awal hingga akhir yang tidak dapat kami sebutkan satu persatu.

Peneliti mengharapkan saran dan kritik dari berbagai pihak yang bersifat membangun demi penyempurnaan skripsi ini. Semoga skripsi ini berguna bagi pembaca secara umum dan peneliti secara khusus. Akhir kata, peneliti ucapkan banyak terima kasih.

Malang, 09 Januari 2023

I Komang Damai Armawan

DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS SKRIPSI	iv
ABSTRAK.....	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	5
1.3. Tujuan Penelitian	5
1.4. Batasan Masalah	5
1.5. Kontribusi Penelitian	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. <i>Artificial Intelligence</i>	7
2.2. <i>Data mining</i>	7
2.3. <i>Machine Learning</i>	13
2.4. <i>Web Scraping</i>	18
2.5. <i>Twitter</i>	19
2.6. <i>Term Frequency-Invers Document Frequency (TF-IDF)</i>	20
2.7. <i>Synthetics Minority Over-sampling Technique (SMOTE)</i>	22
2.8. <i>Naïve Bayes Classifier (NBC)</i>	24
BAB III METODOLOGI PENELITIAN	30
3.1. Analisis Permasalahan	30
3.2. Solusi yang Diusulkan	30

3.3.	Kerangka Kerja (<i>Framework</i>).....	31
3.3.1.	Kata Kunci	31
3.3.2.	Pengambilan Data	32
3.3.3.	<i>Preprocessing</i>	35
3.3.4.	<i>Splitting Dataset</i> Dan Pembobotan TF-IDF	39
3.3.5.	SMOTE	41
3.3.6.	NBC	41
3.4.	Rancangan Eksperimen.....	44
3.4.1.	Alat Pengujian.....	44
3.4.2.	Bahan Pengujian	45
BAB IV PENGUJIAN DAN HASIL		47
4.1.	Pembacaan <i>Dataset</i>	47
4.2.	<i>Preprocessing</i>	48
4.3.1.	Penghapusan Dupikasi	48
4.3.2.	Penghapusan Pengguna.....	49
4.3.3.	Pemotongan Tanggal	50
4.3.4.	Penghapusan Angka.....	50
4.3.5.	Penghapusan Tanda Baca	51
4.3.6.	<i>Stemming</i>	52
4.3.7.	Kamus Kata.....	53
4.3.8.	<i>Tokenizing</i>	54
4.3.9.	Pemberishan <i>Stopwords</i>	55
4.3.10.	Pemberian Label	59
4.3.	<i>Splitting Dataset</i> Dan Pembobotan TF-IDF	60
4.4.	SMOTE	63
4.5.	NBC	67
4.6.1.	Klasifikasi Pada <i>Testing Set</i> 30%	68
4.6.2.	Klasifikasi Pada <i>Testing Set</i> 40%	70
4.6.3.	Klasifikasi Pada <i>Testing Set</i> 50%	71
BAB V KESIMPULAN DAN SARAN.....		74
5.1.	Kesimpulan	74
5.2.	Saran	74

DAFTAR PUSTAKA	76
----------------------	----

DAFTAR TABEL

No	Nama Tabel	Halaman
2.1	Penelitian Yang Relevan.....	27
3.1	Kata Kunci dan Link.....	32
3.2	Contoh Data Yang Didapatkan.....	33
3.3	Proses Penghapusan Angka	36
3.4	Proses Penghapusan Tanda Baca.....	36
3.5	Proses Normalisasi.....	37
3.6	Proses Pembuatan Kamus Kata	37
3.7	Proses Tokenizing.....	38
3.8	Proses Pembersihan Stopword.....	38
3.9	Proses Labeling.....	39
3.10	Data Yang Akan Digunakan Dalam Perhitungan TF-IDF	40
3.11	Perhitungan TF-IDF.....	40
3.12	Proses Penyeimbangan SMOTE.....	41
3.13	Training Set Dan Testing Set Untuk Perhitungan NBC	42

DAFTAR GAMBAR

No	Nama Gambar	Halaman
2.1	Tahap-Tahap Data Mining.....	9
3.1	Kerangka Kerja Penelitian.....	31
3.2	Tampilan Apify Console.....	33
4.1	Pseudocode Pembacaan Dataset.....	47
4.2	Hasil Pembacaan Dataset.....	47
4.3	Pseudocode Proses Penghapusan Duplikasi	48
4.4	Chart Hasil Proses Penghapusan Duplikasi.....	48
4.5	Pseudocode Proses Seleksi Pengguna	49
4.6	Chart Hasil Penghapusan Pengguna	49
4.7	Pseudocode Pemotongan Tanggal	50
4.8	Chart Hasil Pemotongan Tanggal.....	50
4.9	Pseudocode Proses Penghapusan Angka Dari Data	51
4.10	Tampilan Hasil Penghapusan Angka Dari Data	51
4.11	Pseudocode Proses Penghapusan Tanda Baca.....	52
4.12	Tampilan Hasil Penghapusan Tanda Baca	52
4.13	Pseudocode Proses Stemming	53
4.14	Tampilan Hasil Stemming	53
4.15	Pseudocode Proses Pembuatan Dan Pengunduhan Berkas kamus_kata.xlsx	54
4.16	Pseudocode Proses Pengunduhan kamus_kurangdari_10.xlsx.....	54
4.17	Pseudocode Proses Tokenizing.....	55
4.18	Tampilan Hasil Tokenizing	55
4.19	Tampilan Hasil Analisa Berkas kamus_kata.xlsx	56
4.20	Tampilan Hasil Analisa Berkas kamus_kurangdari_10.xlsx.....	57
4.21	Pseudocode Proses Penghapusan Kata-Kata Penting Pada Stoplist.....	58
4.22	Pseudocode Proses Kombinasi Stoplist.....	58
4.23	Pseudocode Proses Pembersihan Stopword	58
4.24	Tampilan Hasil Pembersihan Stopword	59
4.25	Pseudocode Proses Pelabelan	59
4.26	Chart Anggota Tiap Kelas	60

4.27	Tampilan Hasil Pelabelan	60
4.28	Pseudocode Proses Pemisahan Dataset	61
4.29	Chart Perbandingan Training Set Dan Testing Set Rasio 30%	61
4.30	Chart Perbandingan Training Set Dan Testing Set Rasio 40%	61
4.31	Chart Perbandingan Training Set Dan Testing Set Rasio 50%	62
4.32	Pseudocode Pembobotan Kata Di Training Set Dan Testing Set.....	62
4.33	Pseudocode Proses Pembuatan Dataframe Tf-Idf	63
4.34	Tampilan Dataframe Tf-Idf Training Set	63
4.35	Tampilan Dataframe Tf-Idf Testing Set	63
4.36	Chart Perbandingan Kelas Di Training Set Sebelum Oversampling Rasio 30%.....	64
4.37	Chart Perbandingan Kelas Di Training Set Sebelum Oversampling Rasio 40%.....	64
4.38	Chart Perbandingan Kelas Di Training Set Sebelum Oversampling Rasio 50%.....	65
4.39	Pseudocode Proses Oversampling	65
4.40	Chart Perbandingan Kelas Di Training Set Setelah Oversampling Rasio 30%	66
4.41	Chart Perbandingan Kelas Di Training Set Setelah Oversampling Rasio 40%	66
4.42	Chart Perbandingan Kelas Di Training Set Setelah Oversampling Rasio 50%	67
4.43	Pseudocode Pembuatan Model Klasifikasi.....	67
4.44	Pseudocode Prediksi Kelas Pada Testing Set	68
4.45	Chart Hasil Klasifikasi NBC Tanpa Implementasi SMOTE Rasio 30%.....	69
4.46	Chart Hasil Klasifikasi NBC Dengan Implementasi SMOTE Rasio 30% ..	69
4.47	Chart Hasil Klasifikasi NBC Tanpa Implementasi SMOTE Rasio 40%.....	70
4.48	Chart Hasil Klasifikasi NBC Dengan Implementasi SMOTE Rasio 40% ..	71
4.49	Chart Hasil Klasifikasi NBC Tanpa Implementasi SMOTE Rasio 50%.....	72
4.50	Chart Hasil Klasifikasi NBC Dengan Implementasi SMOTE Rasio 50% ..	72

DAFTAR LAMPIRAN

No	Nama Lampiran	Halaman
7.1	RIWAYAT HIDUP	82
7.2	BERITA ACARA BIMBINGAN.....	83