



LAPORAN SKRIPSI

**ANALISIS SENTIMEN TERHADAP KENAIKAN BAHAN
BAKAR MINYAK PADA *TWITTER* MENGGUNAKAN
ALGORITMA *DECISION TREE C4.5***

**RAYHAN PRASETYO
NIM. 201951213**

**DOSEN PEMBIMBING
Endang Supriyati, S. Kom., M. Kom.
Tri Listyorini, S. Kom., M. Kom.**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MURIA KUDUS
2023**

HALAMAN PERSETUJUAN

ANALISIS SENTIMEN TERHADAP KENAIKAN BAHAN BAKAR MINYAK PADA *TWITTER* MENGGUNAKAN ALGORITMA *DECISION TREE C4.5*

RAYHAN PRASETYO

NIM. 201951213

Kudus, 1 Januari 2023

Menyetujui,

Pembimbing Utama,



Endang Supriyati, S.Kom., M.Kom
NIDN. 0629077402

Pembimbing Pendamping,



Tri Listyorini, S.Kom., M.Kom
NIDN. 0616088502

HALAMAN PENGESAHAN

ANALISIS SENTIMEN TERHADAP KENAIKAN BAHAN BAKAR MINYAK PADA *TWITTER* MENGGUNAKAN ALGORITMA *DECISION TREE C4.5*

RAYHAN PRASETYO

NIM. 201951213

Kudus, 1 Februari 2023

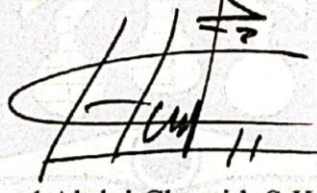
Menyetujui,

Ketua Penguji,



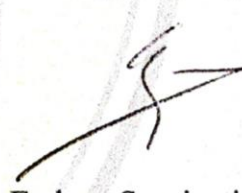
Evanita, S.Kom., M.Kom
NIDN. 0611088901

Anggota Penguji I,



Ahmad Abdul Chamid, S.Kom.,
M.Kom
NIDN. 0616109101

Anggota Penguji II,



Endang Supriyati, S.Kom.,
M.Kom
NIDN. 0629077402

Mengetahui

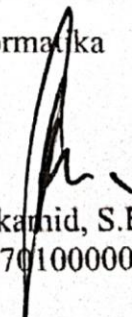
Dekan Fakultas Teknik




Moh. Djalil, ST., MT.
NIS. 0610701000001141

Ketua Program Studi Teknik

Informatika


Mukhamad Nurkani, S.Kom.M.Cs.
NIS. 0610701000001202

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini :

Nama : Rayhan Prasetyo
NIM : 201951213
Tempat & Tanggal Lahir : Jakarta, 16 September 2000
Judul Skripsi : Analisis Sentimen Terhadap Kenaikan Bahan Bakar Minyak Pada *Twitter* Menggunakan Algoritma *Decision Tree C4.5*

Menyatakan dengan sebenarnya bahwa penulisan Skripsi ini berdasarkan hasil penelitian, pemikiran dan pemaparan asli dari saya sendiri, baik untuk naskah laporan maupun kegiatan lain yang tercantum sebagai bagian dari Skripsi ini. Seluruh ide, pendapat, atau materi dari sumber lain telah dikutip dalam Skripsi dengan cara penulisan referensi yang sesuai.

Demikian pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar dan sanksi lain sesuai dengan peraturan yang berlaku di Universitas Muria Kudus.

Demikian pernyataan ini saya buat dalam keadaan sadar tanpa paksaan dari pihak manapun.

Kudus, 22 November 2022

Yang memberi pernyataan,



Rayhan Prasetyo
NIM. 201951213

KATA PENGANTAR

Puji Syukur kepada ALLAH S.W.T atas segala rahmatnya hingga dapat Menyusun tugas akhir ini dengan baik dan lancar.

Penyusunan Skripsi/Tugas Akhir ini ditujukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer .

Pelaksanaan Skripsi ini tak lepas dari bantuan dan dukungan beberapa pihak, untuk itu penulis menyampaikan ucapan terima kasih kepada:

1. Moh. Dahlan, ST.,MT. selaku dekan Fakultas Teknik Universitas Muria kodus.
2. Endang Supriyati, M.Kom dan Tri Listyorini M.Kom selaku dosen pembimbing Skripsi yang dengan sabar memberi petunjuk, bimbingan, dan waktunya kepada penulis.
3. Kedua Orang tua dan kakak yang senantiasa memberikan memberikan semangat serta do'a untuk penulis.
4. Teman-teman seperjuangan yang telah membantu dalam penyusunan laporan skripsi ini.
5. Rekan-rekan yang telah memberikan semangat untuk penulis.

Penulis menyadari adanya kekurangan dan ketidaksempurnaan dalam penulisan skripsi ini, karena itu penulis menerima kritik, saran dan masukan dari pembaca sehingga penulis dapat lebih baik di masa yang akan datang. Akhirnya penulis berharap semoga bisa bermanfaat khususnya bagi penulis dan umumnya bagi para pembaca.

Kudus, November 2022

Penulis

**ANALISIS SENTIMEN TERHADAP KENAIKAN BAHAN BAKAR
MINYAK PADA *TWITTER* MENGGUNAKAN ALGORITMA *DECISION
TREE C4.5***

Nama mahasiswa : Rayhan Prasetyo
NIM : 201951213
Pembimbing :
1. Endang Supriyati, S.Kom, M.Kom
2. Tri Listyorini, S.Kom, M.Kom

RINGKASAN

Salah satu pengaruh dari kenaikan bahan bakar minyak di Indonesia yaitu masyarakat jadi berpikir dua kali dalam membeli bahan bakar minyak untuk sepeda motor terutama. Hal ini dapat dilihat dari cuitan di twitter tentang kenaikan bahan bakar minyak. Sosial media yang digunakan dalam penelitian ini yaitu *Twitter* Penelitian dilakukan untuk mengetahui sentimen masyarakat terhadap kenaikan bahan bakar minyak dengan menggunakan metode *Decision Tree C4.5*. Teknik klasifikasi di terapkan untuk mengklasifikasikan data *tweet*, Sebelum di lakukan klasifikasi, data *tweet* melalui *preprocessing* dan pembobotan *term frequency* dan *tf-idf*. Dari pembobotan ini menghasilkan data *training* dan *testing* yang akan dilakukan klasifikasi menggunakan pohon keputusan dengan Algoritma C4.5 untuk menentukan Sentimen Analisis. Hasil pengujian di peroleh dengan akurasi yang cukup tinggi yaitu 80%. *Analisis sentiment* masyarakat terhadap kenaikan bahan bakar minyak tersebut fokusnya ada pada tingkat polaritas respons atau pendapat ke dalam kategori positif dan negatif.

Kata kunci: Analisis Sentimen, Twitter, Decision Tree C4.5, preprocessing

SENTIMENT ANALYSIS OF THE INCREASE OF FUEL ON TWITTER USING THE DECISION TREE C4.5 ALGORITHM

Student Name : Rayhan Prasetyo
Student Identity Number : 201951213
Supervisor :
1. Endang Supriyati, S.Kom, M.Kom
2. Tri Listyorini, S.Kom, M.Kom

ABSTRACT

One of the effects of the increase in fuel oil in Indonesia is that people think twice about buying fuel oil for motorcycles, especially. This can be seen from the tweets on twitter about the increase in fuel oil. The social media used in this study is Twitter. The research was conducted to determine public sentiment towards the increase in fuel oil using the Decision Tree C4.5 method. Classification technique is applied to classify tweet data. Before classification, tweet data goes through preprocessing and weighting term frequency and tf-idf. From this weighting, it produces training and testing data which will be classified using a decision tree with the C4.5 Algorithm to determine Sentiment Analysis. The test results obtained with a fairly high accuracy of 80%. The analysis of public sentiment on the increase in fuel oil focuses on the level of polarity of responses or opinions into positive and negative categories.

Keywords : Analisis Sentiment, Twitter, Decision Tree C4.5, preprocessing

DAFTAR ISI

DAFTAR ISTILAH DAN SINGKATAN	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan.....	2
1.5. Sistematika Penulisan.....	2
BAB II TINJAUAN PUSTAKA.....	5
2.1 Analisa Sentimen.....	5
2.2 <i>Natural Language Processing</i>	5
2.3 <i>Text Mining</i>	5
2.4 <i>Scraping</i>	5
2.5 <i>Twitter</i>	6
2.6 <i>Preprocessing</i>	6
2.6.1 <i>Cleansing</i>	6
2.6.2 <i>Case Folding</i>	6
2.6.3 <i>Tokenizing</i>	7
2.6.4 <i>Stopword</i>	7
2.6.5 <i>Stemming</i>	7
2.7 Pembobotan kata	7
2.8 Metode <i>Decision Tree</i>	8
2.9 Diagram Alir	9
BAB III METODOLOGI.....	11
3.1 Kebutuhan Sistem.....	11
3.2 Metode Penelitian.....	11
3.3 Metode Pengumpulan Data	11
3.4 Perancangan Program Analisis.....	12
3.4.1 Tahap Pengumpulan Data	12
3.4.2 Tahap <i>Pre-Processing</i>	13
3.4.3 Klasifikasi <i>Decision Tree</i>	22
3.5 Skenario Sistem	24
BAB IV HASIL DAN PEMBAHASAN	27

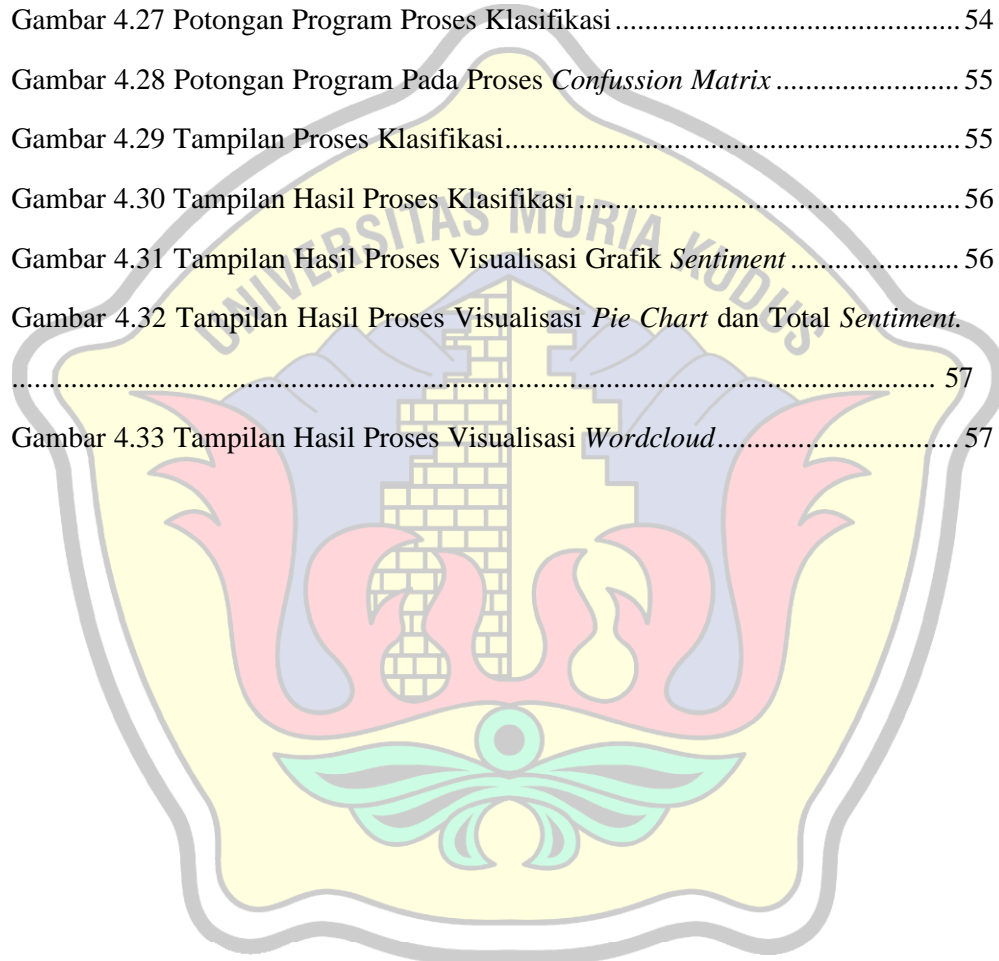
4.1 Hasil Pembahasan.....	27
4.1.1 Hasil <i>Scraping</i>	27
4.1.2 Hasil <i>Text Pre-processing</i>	30
4.1.3 Hasil Labeling.....	38
4.1.4 Hasil Klasifikasi.....	41
4.1.5 Hasil Visualisasi.....	44
BAB V PENUTUP	47
5.1 Kesimpulan.....	47
5.2 Saran.....	47
DAFTAR PUSTAKA	48
LAMPIRAN 1.....	49
BIODATA PENULIS	55



DAFTAR GAMBAR

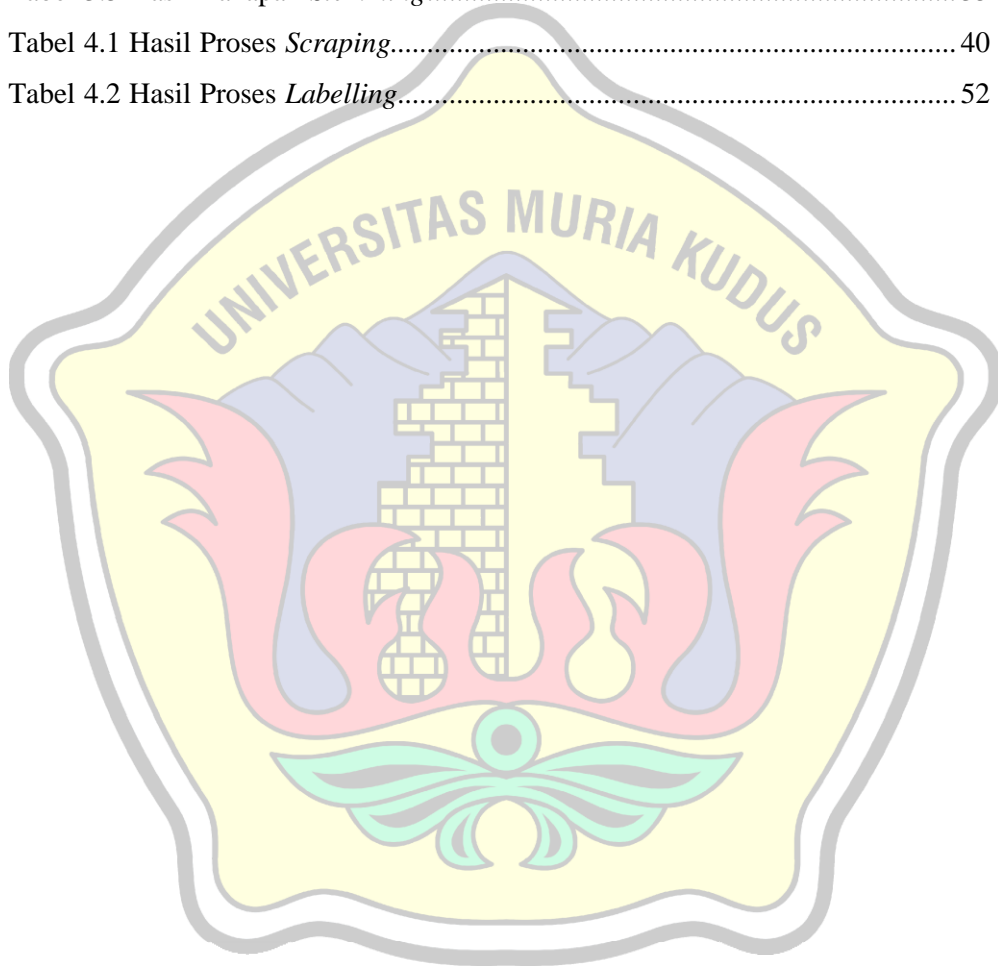
Gambar 2.1 Bagan <i>Decision Tree</i>	22
Gambar 3.1 Alur Program	25
Gambar 3.2 Tahapan Pengumpulan Data.....	26
Gambar 3.3 Tahapan <i>Pre-processing</i>	27
Gambar 3.4 Alur Klasifikasi <i>Decision Tree</i>	35
Gambar 3.5 <i>Confusion Matrix</i>	37
Gambar 3.6 Skenario Sistem	38
Gambar 4.1 API dan Token <i>Twitter</i>	39
Gambar 4.2 Program Tahapan <i>Scraping</i>	40
Gambar 4.3 Halaman Proses <i>Scraping</i>	41
Gambar 4.4 Tampilan Hasil <i>Scraping</i>	42
Gambar 4.5 Program Tahapan <i>Cleaning</i>	43
Gambar 4.6 Tampilan Tahapan Hasil <i>Cleaning</i>	43
Gambar 4.7 Potongan Program Tahapan <i>Case Folding</i>	44
Gambar 4.8 Tampilan Hasil Tahapan <i>Case Folding</i>	44
Gambar 4.9 Potongan Program Tahapan <i>Tokenizing</i>	45
Gambar 4.10 Potongan Program <i>Library nltk</i>	45
Gambar 4.11 Tampilan Hasil Tahapan <i>Tokenizing</i>	45
Gambar 4.12 Potongan Program Tahap <i>Stop Removal</i>	46
Gambar 4.13 Potongan Program Tahap Penambahan Variabel <i>Stopword</i>	46
Gambar 4.14 Tampilan Hasil Tahapan <i>Stop Removal</i>	47
Gambar 4.15 Potongan Program Tahapan <i>Stemming</i>	47
Gambar 4.16 Tampilan Hasil Tahapan <i>Stemming</i>	48
Gambar 4.17 Tampilan Hasil Tahapan <i>Pre-processing</i>	48
Gambar 4.18 Potongan Program Pada Proses <i>Pre-Processing</i>	49
Gambar 4.19 Tampilan Aplikasi Hasil <i>Pre-Processing</i>	49
Gambar 4.20 Potongan <i>Library</i> Pada <i>Python</i>	50

Gambar 4.21 Program Pada Pemberian Nilai Polaritas.....	51
Gambar 4.22 Proses Perhitungan <i>Labeling</i> Dengan Excel.....	51
Gambar 4.23 Tampilan Aplikasi Pada Proses <i>Labelling</i>	52
Gambar 4.24 Tampilan Hasil Dari Proses <i>Labelling</i>	53
Gambar 4.25 Potongan Program <i>Library Scikit Learn</i>	53
Gambar 4.26 Potongan Program Pembagian Dataset.....	54
Gambar 4.27 Potongan Program Proses Klasifikasi.....	54
Gambar 4.28 Potongan Program Pada Proses <i>Confussion Matrix</i>	55
Gambar 4.29 Tampilan Proses Klasifikasi.....	55
Gambar 4.30 Tampilan Hasil Proses Klasifikasi.....	56
Gambar 4.31 Tampilan Hasil Proses Visualisasi Grafik <i>Sentiment</i>	56
Gambar 4.32 Tampilan Hasil Proses Visualisasi <i>Pie Chart</i> dan Total <i>Sentiment</i>	57
Gambar 4.33 Tampilan Hasil Proses Visualisasi <i>Wordcloud</i>	57



DAFTAR TABEL

Tabel 2.1 Simbol Diagram Alir.....	23
Tabel 3.1 Tahapan <i>Cleaning</i>	27
Tabel 3.2 Hasil Tahapan <i>Case Folding</i>	29
Tabel 3.3 Hasil Tahapan <i>Tokenizing</i>	30
Tabel 3.4 Hasil Tahapan <i>Stop Removal</i>	34
Tabel 3.5 Hasil Tahapan <i>Stemming</i>	35
Tabel 4.1 Hasil Proses <i>Scraping</i>	40
Tabel 4.2 Hasil Proses <i>Labelling</i>	52



DAFTAR LAMPIRAN

Lampiran 1 60



DAFTAR ISTILAH DAN SINGKATAN

TF-IDF	: <i>Term Frequency Invers Document Frequency</i>
C4.5	: <i>Algoritma Data Mining</i>
Tree	: <i>DecisionTree Claasifier</i>
API	: <i>Application Programming Interface</i>
CSV	: <i>Comma Separated Values</i>
NLTK	: <i>Natural Language Toolkit</i>

