

**ANALISIS SENTIMEN PENGGUNAAN GALON *BISPHENOL*
A MENGGUNAKAN ALGORITMA *SUPPORT VECTOR*
MACHINE MELALUI *CHI-SQUARE TEST***

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data**



Disusun Oleh:

CENDITYA AYU AURELIA

20083010023

**PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
SURABAYA
2024**

LEMBAR PENGESAHAN

**ANALISIS SENTIMEN PENGGUNAAN GALON BISPHENOL A
MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE
MELALUI CHI-SQUARE TEST**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data
pada Selasa, 14 Mei 2024

**Program Studi S-1 Sains Data
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya**

Oleh :

CENDITYA AYU AURELIA

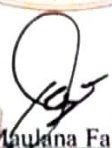
NPM. 20083010023


Disetujui oleh Tim Penguji Skripsi :

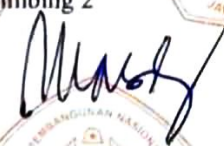
Penguji 1

Penguji 2


Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 21119950723270
Pembimbing 1


Tresna Maulana Fahrudin, S.ST., M.T.
NIP. 199305012022031007
Pembimbing 2


Trimono, S.Si., M.Si.
NIP. 199509082022031003


Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU
NIP. 197006192021211009


Fakultas Ilmu Komputer
Dekan,

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 196811261994032001

Mengetahui,
Program Studi Sains Data
Fakultas Ilmu Komputer
Koordinator,


Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU
NIP. 198012052005011002

Surabaya, Mei, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Cenditya Ayu Aurelia
NPM : 20083010023
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

**ANALISIS SENTIMEN PENGGUNAAN GALON *BISPHENOL A*
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*
MELALUI *CHI-SQUARE TEST***

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 29 Mei 2024

Hormat Saya



Cenditya Ayu Aurelia

NPM. 20083010023

ABSTRAK

ANALISIS SENTIMEN PENGGUNAAN GALON *BISPHENOL A* MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* MELALUI *CHI-SQUARE TEST*

Nama Mahasiswa / NPM : Cenditya Ayu Aurelia / 20083010023
Program Studi : Sains Data, FASILKOM, UPN Veteran Jatim
Dosen Pembimbing 1 : Trimono, S.Si., M.Si.
Dosen Pembimbing 2 : Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU

Abstrak

Air Minum Dalam Kemasan (AMDK) merupakan elemen utama yang penting untuk menjaga keseimbangan tubuh. Salah satu air minum dalam kemasan adalah galon yang umumnya menggunakan bahan *Polycarbonate* mengandung *Bisphenol A* (BPA) yang berpotensi memiliki dampak negatif pada kesehatan dan *Polietilen Tereftalat* tidak mengandung BPA yang berpotensi menyebabkan masalah lingkungan. Kondisi ini menimbulkan keresahan masyarakat di Twitter antara penggunaan galon BPA dengan galon non-BPA setelah menyebarnya pemberitaan terkait dampak yang disebabkan oleh galon BPA sehingga diperlukan analisis sentimen yang dapat mengkategorikan sentimen positif dan negatif. Penelitian ini dilakukan untuk menganalisis sentimen penggunaan galon BPA menggunakan seleksi fitur *Chi-Square* dan algoritma *Support Vector Machine* dengan *kernel Linear*, *Polynomial*, dan RBF. Tahapan penelitian meliputi pengumpulan data sebanyak 1257 *tweet*, *preprocessing* data, pelabelan data, pembobotan kata TF-IDF, seleksi fitur *Chi-Square*, *split dataset*, klasifikasi data, evaluasi model, dan analisis hasil. Tujuan dari penelitian ini untuk meningkatkan kesadaran masyarakat dalam memilih air minum kemasan galon yang aman dan ramah lingkungan. Penelitian ini menunjukkan hasil dari penerapan seleksi fitur *Chi-Square* dengan nilai taraf nyata 0.90 dapat meningkatkan akurasi hingga 94% pada *kernel Linear* dan *Polynomial*.

Kata kunci: *Analisis Sentimen, Galon BPA, Galon non-BPA, Chi-Square, Support Vector Machine.*

ABSTRACT

SENTIMENT ANALYSIS OF *BISPHENOL A* GALLON USAGE USING *SUPPORT VECTOR MACHINE* ALGORITHM THROUGH *CHI-SQUARE* *TEST*

Student Name / NPM : Cenditya Ayu Aurelia / 20083010023
Study Program : Sains Data, FASILKOM, UPN Veteran Jatim
Advisor 1 : Trimono, S.Si., M.Si.
Advisor 2 : Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU

Abstract

Bottled Drinking Water (AMDK) is the main element that is important to maintain the balance of the body. One of the bottled drinking water is gallons that generally use *Polycarbonate* material containing *Bisphenol A* (BPA) which has the potential to have a negative impact on health and *Polyethylene Terephthalate* does not contain BPA which has the potential to cause environmental problems. This condition causes public unrest on Twitter between the use of BPA gallons and non-BPA gallons after the spread of news related to the impact caused by BPA gallons so that sentiment analysis is needed that can categorize positive and negative sentiments. This research was conducted to analyze the sentiment of using BPA gallons using *Chi-Square* feature selection and Support Vector Machine algorithm with *Linear*, *Polynomial*, and *RBF kernel*. The research stages include data collection of 1257 tweets, data preprocessing, data labeling, TF-IDF word weighting, *Chi-Square* feature selection, split dataset, data classification, model evaluation, and result analysis. The purpose of this research is to increase public awareness in choosing safe and environmentally friendly gallon bottled drinking water. This research shows the results of the application of *Chi-Square* feature selection with a real level value of 0.90 can increase accuracy up to 94% on *Linear* and *Polynomial kernel*.

Keywords: *Sentiment Analysis, BPA Gallon, Non-BPA Gallon, Chi-Square, Support Vector Machine.*

KATA PENGANTAR

Puji dan syukur atas kehadiran Allah SWT. yang telah memberikan limpahan Rahmat dan nikmat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Penggunaan Galon *Bisphenol A* Menggunakan Algoritma *Support Vector Machine* Melalui *Chi-Square Test*”. Dalam proses penyusunan skripsi ini, penulis dengan rasa hormat mengucapkan terima kasih atas bantuan dan arahan kepada:

1. Kedua orang tua tercinta yang selalu memberikan dukungan penuh kasih sayang dan mendoakan yang terbaik atas setiap cita-cita penulis sehingga penulis termotivasi untuk terus berjuang dalam menyelesaikan skripsi ini.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Bapak Dr.Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Bapak Trimono, S.Si., M.Si. selaku Dosen Pembimbing 1 dan Bapak Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU selaku Dosen Pembimbing 2 yang senantiasa memberikan arahan, masukan, dan saran kepada penulis.
6. Ibu Kartika Maulida Hindrayani, S.Kom., M.Kom selaku Dosen Wali Penulis.
7. M. Andrew Arjunanda dan teman seperjuangan yang mendukung penulis.

Dalam laporan skripsi ini, penulis mohon maaf atas segala kekurangan dari materi maupun penulisan. Semoga laporan skripsi ini dapat memberikan kontribusi terhadap perkembangan ilmu pengetahuan terutama dalam bidang Sains Data.

Surabaya, 29 Mei 2024

Cenditya Ayu Aurelia

DAFTAR ISI

HALAMAN SAMPUL	ii
LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	8
2.1. Dasar Teori.....	8
2.1.1 Analisis Sentimen.....	8
2.1.2 Galon BPA	8
2.1.3 <i>Web Crawling</i>	9
2.1.4 Twitter.....	9
2.1.5 <i>Text Preprocessing</i>	10
2.1.6 Pembobotan Kata TF-IDF.....	11

2.1.7	Uji Dependensi <i>Chi-Square</i>	12
2.1.8	<i>Support Vector Machine</i>	12
2.1.9	<i>Confusion Matrix</i>	17
2.1.10	<i>Word Cloud</i>	19
2.2.	Penelitian Terdahulu	19
BAB III METODOLOGI PENELITIAN.....		26
3.1.	Variabel Penelitian dan Sumber Data	26
3.1.1	Variabel Penelitian	26
3.1.2	Sumber Data.....	26
3.2.	Langkah Analisis.....	28
3.2.1	Pengumpulan Data	28
3.2.2	<i>Preprocessing Data</i>	28
3.2.3	Pelabelan Data.....	30
3.2.4	Pembobotan Kata TF-IDF.....	31
3.2.5	Seleksi Fitur <i>Chi-Square</i>	32
3.2.6	<i>Split Dataset</i>	33
3.2.7	Klasifikasi Data.....	33
3.2.8	Evaluasi Model.....	35
3.2.9	Analisis Hasil	35
3.3.	Diagram Alir Penelitian	36
3.4.	Jadwal Penelitian.....	36
BAB IV HASIL DAN PEMBAHASAN		37
4.1.	Pengumpulan Data	37
4.2.	<i>Preprocessing Data</i>	40
4.2.1.	<i>Case Folding</i>	40

4.2.2.	<i>Cleaning</i>	41
4.2.3.	<i>Normalization</i>	42
4.2.4.	<i>Tokenizing</i>	43
4.2.5.	<i>Stopword Removal</i>	44
4.2.6.	<i>Stemming</i>	46
4.2.7.	<i>Handling Missing Values dan Remove Duplicate</i>	47
4.3.	Pelabelan Data.....	48
4.4.	Pembobotan Kata TF-IDF.....	50
4.5.	Seleksi Fitur <i>Chi-Square</i>	53
4.6.	<i>Split Dataset</i>	58
4.7.	Klasifikasi Data.....	59
4.8.	Evaluasi Model.....	66
4.8.1.	<i>Confusion Matrix dan Classification Report</i>	66
4.8.2.	Pengujian Dengan Data Baru	78
4.9.	Analisis Hasil	79
BAB V PENUTUP.....		82
DAFTAR PUSTAKA		84
LAMPIRAN		91
BIODATA PENULIS		100

DAFTAR GAMBAR

Gambar 2. 1. <i>Hyperplane</i> Terbaik Algoritma <i>Support Vector Machine</i>	13
Gambar 3. 2. Blok Diagram Tahapan <i>Preprocessing</i> Data.....	29
Gambar 3. 3. Blok Diagram Tahapan Pembobotan Kata TF-IDF	31
Gambar 3. 4. Blok Diagram Tahapan Seleksi Fitur <i>Chi-Square</i>	32
Gambar 3. 5. Diagram Algoritma <i>Support Vector Machine</i>	34
Gambar 3. 6. Blok Diagram Proses Analisis Penelitian.....	36
Gambar 4. 1. Hasil Proses <i>Case Folding</i>	41
Gambar 4. 2. Hasil Proses <i>Cleaning</i>	42
Gambar 4. 3. Hasil Proses <i>Normalization</i>	43
Gambar 4. 4. Hasil Proses <i>Tokenizing</i>	44
Gambar 4. 5. Hasil Proses <i>Stopword Removal</i>	45
Gambar 4. 6. Hasil Proses <i>Stemming</i>	47
Gambar 4. 7. Hasil Proses <i>Handling Missing Values</i>	47
Gambar 4. 8. Hasil Proses <i>Handling Remove Duplicate</i>	48
Gambar 4. 9. Visualisasi Distribusi Sentimen	49
Gambar 4. 10. Hasil Pelabelan Data Sentimen	49
Gambar 4. 11. Hasil Pembobotan Kata TF-IDF	51
Gambar 4. 12. Hasil <i>Degrees of Freedom</i> dan <i>Critical Value Chi-Square</i> 0.90... 55	
Gambar 4. 13. Hasil Seleksi Fitur Kata yang Terpilih.....	56
Gambar 4. 14. Hasil Seleksi Fitur Kata yang Terbuang	57
Gambar 4. 15. Hasil Seleksi Fitur <i>Chi-Square</i> Terpilih dari <i>Alpha</i> 0.90.....	57
Gambar 4. 16. Hasil Parameter Terbaik <i>Kernel Linear</i>	60
Gambar 4. 17. Hasil Parameter Terbaik <i>Kernel Polynomial</i>	60
Gambar 4. 18. Hasil Parameter Terbaik <i>Kernel RBF</i>	60

Gambar 4. 19. <i>Confusion Matrix Kernel Linear</i>	66
Gambar 4. 20. <i>Classification Report Kernel Linear</i>	67
Gambar 4. 21. <i>Confusion Matrix Kernel Polynomial</i>	68
Gambar 4. 22. <i>Classification Report Kernel Polynomial</i>	69
Gambar 4. 23. <i>Confusion Matrix Kernel RBF</i>	70
Gambar 4. 24. <i>Classification Report Kernel RBF</i>	70
Gambar 4. 25. <i>Confusion Matrix SVM Chi-Square 0.90 Kernel Linear</i>	72
Gambar 4. 26. <i>Classification Report SVM Chi-Square 0.90 Kernel Linear</i>	72
Gambar 4. 27. <i>Confusion Matrix SVM Chi-Square 0.90 Kernel Polynomial</i>	73
Gambar 4. 28. <i>Classification Report SVM Chi-Square 0.90 Kernel Polynomial</i>	74
Gambar 4. 29. <i>Confusion Matrix SVM Chi-Square 0.90 Kernel RBF</i>	75
Gambar 4. 30. <i>Classification Report SVM Chi-Square 0.90 Kernel RBF</i>	75
Gambar 4. 31. <i>Word Cloud Sentimen Positif</i>	80
Gambar 4. 32. <i>Word Cloud Sentimen Negatif</i>	81

DAFTAR TABEL

Tabel 2. 1. Tabel <i>Confusion Matrix</i>	18
Tabel 2. 2. Studi Literatur Penelitian Terdahulu	19
Tabel 3. 1 Hasil <i>Crawling</i> Data Twitter (1)	27
Tabel 3. 2 Hasil <i>Crawling</i> Data Twitter (2)	27
Tabel 3. 3. Jadwal Kegiatan	36
Tabel 4. 1. Algoritma <i>Crawling</i> Data Berdasarkan <i>Keyword</i>	37
Tabel 4. 2. Hasil Proses <i>Crawling</i> Data Berdasarkan <i>Keyword</i>	38
Tabel 4. 3. Hasil Proses <i>Crawling</i> Data Berdasarkan <i>Keyword</i> (2).....	38
Tabel 4. 4. Algoritma <i>Crawling</i> Data Berdasarkan Komentar.....	38
Tabel 4. 5. Hasil Proses <i>Crawling</i> Data Berdasarkan Komentar	39
Tabel 4. 6. Hasil Proses <i>Crawling</i> Data Berdasarkan Komentar (2)	39
Tabel 4. 7. Hasil Proses <i>Filtering</i> Data	40
Tabel 4. 8. Algoritma <i>Case Folding</i>	40
Tabel 4. 9. Algoritma <i>Cleaning</i>	41
Tabel 4. 10. Algoritma <i>Normalization</i>	42
Tabel 4. 11. Algoritma <i>Tokenizing</i>	43
Tabel 4. 12. Algoritma <i>Stopword Removal</i>	44
Tabel 4. 13. Algoritma <i>Stemming</i>	46
Tabel 4. 14. Algoritma Pelabelan Data	48
Tabel 4. 15. Algoritma Pembobotan Kata TF-IDF	50
Tabel 4. 16. Data Perhitungan Manual TF-IDF	51
Tabel 4. 17. Hasil Perhitungan Manual TF-IDF	52
Tabel 4. 18. Algoritma Seleksi Fitur <i>Chi-Square</i>	54
Tabel 4. 19. Hasil Seleksi Fitur <i>Chi-Square</i>	55

Tabel 4. 20. Data Perhitungan <i>Chi-Square</i>	57
Tabel 4. 21. Hasil Pembagian Data	59
Tabel 4. 22. Algoritma Penentuan Parameter <i>Kernel Linear, Polynomial, RBF</i> ..	59
Tabel 4. 23. Algoritma Mengambil Nilai <i>w</i> dan <i>b</i>	61
Tabel 4. 24. Algoritma Prediksi Sentimen Menggunakan Data Uji	62
Tabel 4. 25. Hasil Prediksi Data <i>Testing Kernel Linear</i>	63
Tabel 4. 26. Hasil Prediksi Data <i>Testing Kernel Polynomial</i>	64
Tabel 4. 27. Hasil Prediksi Data <i>Testing Kernel RBF</i>	65
Tabel 4. 28. Hasil Akurasi <i>Support Vector Machine</i>	76
Tabel 4. 29. Hasil Evaluasi Pengujian Data Baru	78

DAFTAR LAMPIRAN

Lampiran 1 91
Lampiran 2 96
Lampiran 3 98