

**ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA
ROKOK PADA MEDIA SOSIAL TWITTER MENGGUNAKAN
ALGORITME NAIVE BAYES**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

FIZAL FAOZI

16.11.0494

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA
ROKOK PADA MEDIA SOSIAL TWITTER MENGGUNAKAN
ALGORITME NAIVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

FIZAL FAOZI

16.11.0494

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA
ROKOK PADA MEDIA SOSIAL TWITTER MENGGUNAKAN
ALGORITME NAIVE BAYES**

yang disusun dan diajukan oleh

Fizal Faozi
16.11.0494

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 November 2022

Dosen Pembimbing,


Sumarni Adi, S.Kom., M.Cs.

NIK. 190302256

PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA ROKOK PADA
MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITME NAIVE
BAYES

yang disusun dan diajukan oleh

Fizal Faozi

16.11.0494

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 November 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Yuli Astuti, M.Kom.
NIK. 190302146

Anna Baita, M.Kom
NIK. 190302290

Windha Mega Pradnya D, M.Kom
NIK. 190302185



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 November 2022

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fizal Faozi
NIM : 16.11.0494

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Terhadap Kenaikan Harga Rokok Pada Media Sosial
Twitter Menggunakan Algoritme Naive Bayes**

Dosen Pembimbing : Sumarni Adi, S.Kom., M.Cs.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 25 November 2022

Yang Menyatakan,



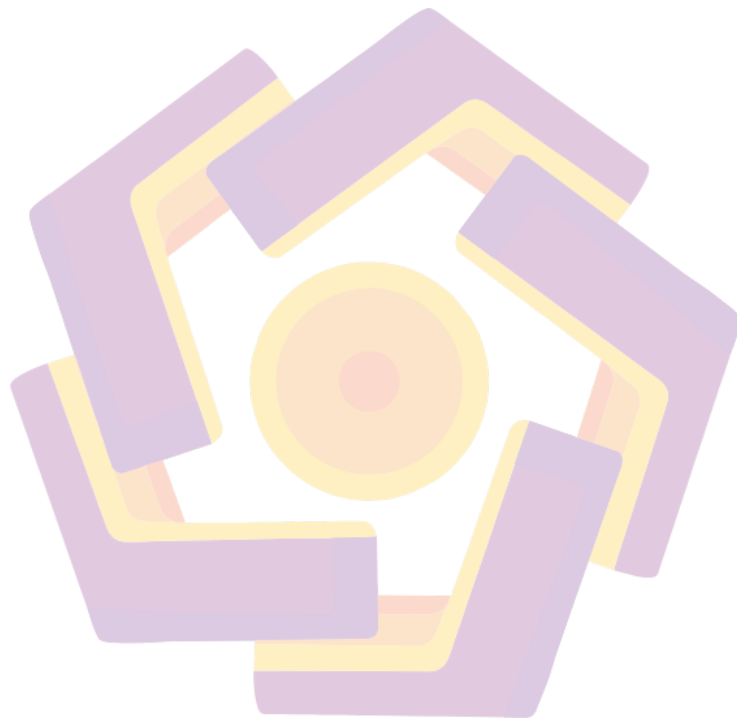
Fizal Faozi

PERSEMBAHAN

Segala puji saya panjatkan kehadiran Alloh SWT yang telah menentukan takdir yang luar biasa kepada saya, sehingga saya bisa menyelesaikan skripsi ini pada waktu yang tepat. Selain itu saya juga sangat berterimakasih kepada orang – orang yang telah membantu, mendoakan, dan mendukung saya baik secara langsung maupun tidak langsung. Dengan segala kerendahan hati dan juga kekurangan skripsi ini saya persembahkan kepada:

1. Kedua orang tua tercinta Ibu Muntaqiyah dan Bapak Mujahidin yang selalu mendoakan dan mensupport saya baik dalam bentuk finansial maupun dalam bentuk apapun. Terutama untuk ibu saya, engkau adalah ibu yang luar biasa bagi saya.
2. Guru – guru saya terutama Pengasuh PP. Al Falah Sumpiuh Banyumas KH. Nurul Huda Sekeluarga, Pengasuh PP. DNF Mojo Kediri Gus H. Achmad Ali Saudi Sekeluarga, dan Pengasuh PP. Al Muhdi Krapyak Sleman Yogyakarta KH. Agus Masruri Sekeluarga yang telah mendoakan dan mendidik ruhaniyah dan jasad saya agar senantiasa berada dijalan yang diridhoi Alloh SWT.
3. Ibu Sumarni Adi, S.Kom, M.Cs selaku dosen pembimbing yang selalu memberikan masukan serta bimbingan kepada saya dalam menyelesaikan skripsi ini.
4. Saudara dan kerabat saya terutama kakak saya Ghozin Khoironi dan adik saya Iin syarofah yang selalu memberi semangat dan nasehat pada saya.
5. Teman-teman kontrakan “RUKAN” Tito Harditya Warsito, Timur Dian Radha Sejati, Yusril, Dian Lesmana, M. Parman, Aji Syahroni, dan Ma’ruf Bagas Angghada atas dukungan, bantuan, dan masukan kepada saya dalam menyelesaikan skripsi ini.
6. Seluruh Teman-teman di PP Al Muhdi yang memberi dukungan dan saran kepada saya.

7. Teman-teman kelas 16.S1IF-08, yang selalu mengingatkan tentang menyelesaikan skripsi ini.
8. Serta semua pihak yang telah membantu serta mendukung saya yang tidak bisa saya sebutkan satu persatu.



KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah SWT atas berkat dan rahmatnya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul **“Analisis Sentimen Terhadap Kenaikan Harga Rokok Pada Media Sosial Twitter Menggunakan Algoritme Naive Bayes”** pada waktu yang tepat.

Penyusunan skripsi ini dilakukan untuk memenuhi salah satu persyaratan dalam menyelesaikan Program Strata-1 Informatika Universitas Amikom Yogyakarta. Dalam menempuh pendidikan Strata-1 Informatika sampai terselesaikannya penulisan skripsi ini banyak pihak yang telah memberi fasilitas, membantu, membina, dan membimbing penulisan, untuk itu khususnya kepada:

1. Bapak Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Hanif Al Fatta, S.Kom, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya Duhita, M.Kom., selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Ibu Sumarni Adi, S.kom, M.Cs., selaku dosen pembimbing yang selalu memberikan bimbingan dan masukan dalam menyelesaikan skripsi ini.
5. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah membekali penulis dengan beberapa disiplin ilmu yang berguna.
6. Kedua orang tua saya yang selalu mendoakan dan memberi dukungan moril maupun materil selama ini.
7. Seluruh teman – teman terutama kelas 16.S1IF-08, yang banyak berdiskusi dengan penulis selama masa studi.

Penulis menyadari dalam skripsi ini masih banyak kekurangan, oleh karena itu penulis akan menerima dengan senang hati apabila ada kritik dan saran yang membangun dari pihak lain. Penulis berharap skripsi ini dapat bermanfaat dan menambah pengetahuan bagi kita terutama tentang analisis sentimen.

Yogyakarta, 27 Februari 2022

Penulis

DAFTAR ISI

JUDUL	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI	iv
PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
INTISARI	xvi
ABSTRACT	xvii
BAB I	Error! Bookmark not defined.
PENDAHULUAN	Error! Bookmark not defined.
1.1 Latar Belakang Masalah.....	Error! Bookmark not defined.
1.2 Rumusan Masalah.....	Error! Bookmark not defined.
1.3 Batasan Masalah.....	Error! Bookmark not defined.
1.4 Maksud dan Tujuan Penelitian.....	Error! Bookmark not defined.
1.4.1 Maksud Penelitian.....	Error! Bookmark not defined.
1.4.2 Tujuan Penelitian.....	Error! Bookmark not defined.
1.5 Manfaat Penelitian.....	Error! Bookmark not defined.
1.6 Metode Penelitian.....	Error! Bookmark not defined.
1.6.1 Metode Pengumpulan Data.....	Error! Bookmark not defined.
1.6.2 Metode Analisis.....	Error! Bookmark not defined.

1.6.3	Metode Perancangan	Error! Bookmark not defined.
1.6.4	Metode Implementasi.....	Error! Bookmark not defined.
1.6.5	Metode Pengembangan	Error! Bookmark not defined.
1.6.6	Metode Testing	Error! Bookmark not defined.
1.7	Sistematika Penelitian	Error! Bookmark not defined.
BAB II		Error! Bookmark not defined.
LANDASAN TEORI		Error! Bookmark not defined.
2.1	Tinjauan Pustaka.....	Error! Bookmark not defined.
2.2	Pengertian Data Mining.....	Error! Bookmark not defined.
2.3	Pengertian Analisis Sentimen	Error! Bookmark not defined.
2.4	Text Mining	Error! Bookmark not defined.
2.5	Preprocessing	Error! Bookmark not defined.
2.6	<i>Term Weighting</i> TF-IDF.....	Error! Bookmark not defined.
2.6.1	<i>Term Frequency</i> (TF).....	Error! Bookmark not defined.
2.6.2	<i>Inverse Document Frequency</i> (IDF) ...	Error! Bookmark not defined.
2.7	<i>Naive Bayes Classifier</i>	22
2.8	<i>Confusion Matrix</i>	Error! Bookmark not defined.
2.9	<i>Flowchart</i>	Error! Bookmark not defined.
2.9.1	<i>Flowchart System</i> (Diagram Alir Sistem)	Error! Bookmark not defined.
2.9.2	<i>Flowchart Program</i> (Diagram Alir Program)	Error! Bookmark not defined.
2.10	Python.....	32
BAB III.....		34
ANALISIS DAN PERANCANGAN.....		34
3.1	Analisis Masalah.....	34

3.2	Solusi Yang Dipilih.....	35
3.3	Analisis Kebutuhan.....	36
3.3.1	Analisis Kebutuhan Fungsional.....	36
3.3.2	Analisis Kebutuhan Non-Fungsional.....	36
3.4	Perancangan Sistem	37
3.4.1	Gambaran Umum Sistem	37
3.4.2	Perancangan Pengumpulan Data Tweet.....	38
3.4.3	Pelabelan Data.....	39
3.4.4	Diagram Alur Preprocessing.....	40
3.4.5	Perancangan Proses Pembobotan TF-IDF.....	41
3.4.6	Diagram Alur <i>Naive Bayes Classifier</i>	41
3.5	Analisis Sentimen Klasifikasi Algoritme <i>Naive Bayes Classifier</i>	43
3.5.1	Pengumpulan Data.....	43
3.5.2	<i>Cleaning</i>	43
3.5.3	<i>Case Folding</i>	45
3.5.4	<i>Tokenization</i>	46
3.5.5	<i>Steaming</i>	47
3.5.6	<i>Filtering</i>	48
3.5.7	TF-IDF	50
3.5.8	<i>Naive Bayes Classifier</i>	62
3.5.9	Evaluasi Menggunakan <i>Confusion Matrix</i>	71
BAB IV		75
IMPLEMENTASI DAN PEMBAHASAN.....		75
4.1	Deskripsi Implementasi.....	75
4.2	Pengumpulan Data Tweet.....	75
4.3	Pelabelan Data.....	77

4.4	Preprocessing Data.....	78
4.4.1	<i>Cleaning Data</i>	78
4.4.2	<i>Case Folding</i>	80
4.4.3	<i>Tokenization</i>	81
4.4.4	<i>Steaming</i>	83
4.4.5	<i>Filtering</i>	84
4.5	Identifikasi Data.....	85
4.6	TF-IDF.....	85
4.7	<i>Splitting Data</i>	87
4.8	Klasifikasi <i>Naive Bayes</i>	87
4.9	Evaluasi.....	88
4.10	Analisis Data Keseluruhan.....	93
BAB V	96
KESIMPULAN	96
5.1	Kesimpulan.....	96
5.2	Saran.....	97
DAFTAR PUSTAKA	98

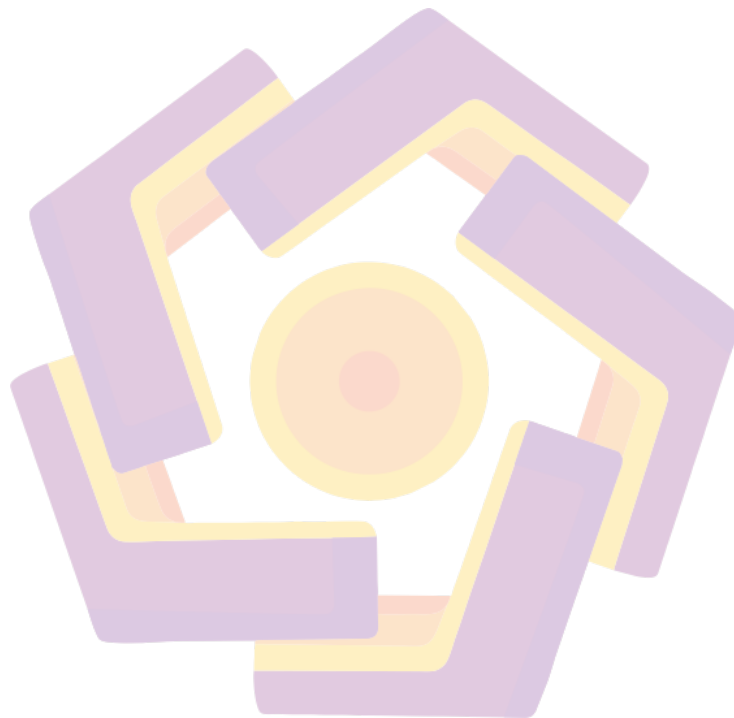
DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian.....	12
Tabel 2.2 <i>Confision Matrix</i> Model.....	26
Tabel 2.3 Daftar <i>Symbol Flowchart System</i>	29
Tabel 2.4 Daftar <i>Syymbol Flowchart Program</i>	31
Tabel 3.1 Data Tweet Sebelum <i>Cleaning</i> Daata	44
Tabel 3.2 Data Tweet Setelah Dilakukan <i>Cleaning Data</i>	44
Tabel 3.3 Data Tweet Setelah Dilakukan <i>Case Folding</i>	45
Tabel 3.4 Data Tweet Setelah Dilakukan <i>Tokenization</i>	46
Tabel 3.5 Data Tweet Setelah Dilakukan <i>Steaming</i>	47
Tabel 3.6 Data Tweet Setelah Dilakukan <i>Filtering</i>	49
Tabel 3.7 <i>Term Frequency</i> Positif	50
Tabel 3.8 <i>Term Frequency</i> Negatif.....	52
Tabel 3.9 <i>Data Document Frequency</i>	54
Tabel 3.10 Hasil Perhitungan IDF	56
Tabel 3.10 TF-IDF Positif.....	58
Tabel 3.11 TF-IDF Negatif	60
Tabel 3.12 Data Uji.....	70
Tabel 3.13 Hasil Pengujian Model Klasifikasi.....	71
Tabel 3.14 <i>Confusion Matrix</i>	72

Tabel 3.15 Nilai *Accuracy, Precision, Recall* 74

Tabel 4.1 Model *Confusion Matrix* 90

Tabel 4.2 *Confusion Matrix Model Naive Bayes* 90



DAFTAR GAMBAR

Gambar 3.1 Alur Proses Sistem	38
Gambar 3.2 Data Tweet	39
Gambar 3.3 Alur Prerocessing	40
Gambar 3.4 Alur Klasifikasi <i>Naive Bayes</i>	42
Gambar 4.1 <i>Sorce Code Crawling Data</i>	76
Gambar 4.2 Hasil <i>Crawling Data</i>	77
Gambar 4.3 Pelabelan Data	77
Gambar 4.4 Deklarasi <i>Library</i>	78
Gambar 4.5 <i>Sorce Code Input Data Dan Output nya</i>	79
Gambar 4.6 <i>Sorce Code Cleaning Data</i>	80
Gambar 4.7 <i>Output Cleaning Data</i>	80
Gambar 4.8 <i>Sorce Code Case Folding</i>	81
Gambar 4.9 <i>Output Case Folding</i>	81
Gambar 4.10 <i>Sorce Code Tokenization</i>	82
Gambar 4.11 <i>Output Tokenization</i>	82
Gambar 4.12 <i>Sorce Code Steaming</i>	83
Gambar 4.13 <i>Output Steaming</i>	83
Gambar 4.14 <i>Sorce Code Stopword</i>	84
Gambar 4.15 <i>Output Stopword</i>	84

Gambar 4.16 <i>Sorce Code</i> identifikasi data.....	85
Gambar 4.17 <i>Sorce Code</i> TF-IDF	86
Gambar 4.18 <i>Output</i> TF – IDF.....	86
Gambar 4.19 <i>Sorce Code</i> <i>Spliting</i> data.....	87
Gambar 4.20 Deklarasi <i>Library</i>	87
Gambar 4.21 <i>Sorce Code</i> <i>Clasification Naive Bayes</i>	88
Gambar 4.22 <i>Accuracy Naive Bayes</i>	88
Gambar 4.23 <i>Sorce Code</i> <i>Evaluasi Confusio Matrix</i>	89
Gambar 4.24 <i>Output Confusion matrix</i>	89
Gambar 4.25 Data Keseluruhan Sebelum <i>Preprocessing</i>	93
Gambar 4.26 Data Keseluruhan Setelah <i>Preprocessing</i>	94
Gambar 4.27 Hasil TF-IDF.....	94
Gambar 4.28 Hasil Klasifikasi.....	95
Gambar 4.29 Hasil <i>Confusion Matrix</i> dan Nilai Akurasi.....	95

INTISARI

Media sosial adalah media *online* yang dimanfaatkan sebagai sarana komunikasi, berbagi informasi, menyampaikan opini secara *online* di internet salah satu *platformnya* adalah twitter. Banyak masyarakat yang cenderung lebih suka menyampaikan opininya di media sosial dibanding menyampaikannya secara langsung, terutama pada twitter. Banyak sekali *tweet* dari masyarakat yang berisikan sebuah opini terhadap suatu hal terutama sesuatu yang sedang *trending* contoh tentang kebijakan pemerintah menaikkan cukai hasil tembakau (CHT) sebesar 23% yang akan berlaku mulai tanggal 1 Januari 2020 dan menyebabkan naiknya harga jual eceran (HJE) sebesar 35%. Dengan adanya opini tersebut dapat dimanfaatkan sebagai bahan penelitian berupa analisis sentimen. Analisis sentimen merupakan proses penggunaan text *analytics* untuk mendapatkan berbagai sumber data di internet dari *platform* media sosial yang bertujuan mendapatkan opini dari pengguna platform tersebut. Oleh sebab itu, penulis tertarik melakukan sebuah penelitian yang berjudul Analisis Sentimen Terhadap Kenaikan Harga Rokok Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes Classifier. Dengan menggunakan algoritma naive bayes classifier sentimen masyarakat akan diklasifikasikan menjadi sentimen *positif* dan *negatif*. Dengan dilaksanakannya penelitian ini diharapkan bisa menghasilkan sebuah informasi yang akurat tentang respon masyarakat terkait kebijakan pemerintah menaikkan cukai hasil tembakau sehingga dapat dijadikan sebuah dasar atau pedoman bagi pemerintah untuk menyikapi respon dari masyarakat Indonesia.

Kata Kunci : Analisis Sentimen, Naive Bayes, Twitter, Pendapat

ABSTRACT

Social media is an online media that is used as a means of communication, sharing information, expressing opinions online on the internet, one of the platforms is Twitter. Many people tend to express their opinions on social media rather than expressing it directly, especially on Twitter. Lots of tweets from the public containing an opinion on something, especially something that is trending examples of government policies raising tobacco product excise (CHT) by 23% which will take effect on January 1, 2020 and cause an increase in retail selling prices (HJE) by 35 %. With this opinion can be used as research material in the form of sentiment analysis. Sentiment analysis is the process of using text analytics to get various sources of data on the internet from social media platforms that aim to get opinions from users of that platform. Therefore, the author is interested in conducting a study entitled Sentiment Analysis of Cigarette Price Increase on Twitter Social Media Using the Naive Bayes Classifier Algorithm. By using the Naive Bayes Classifier algorithm, community sentiments will be classified into positive and negative sentiments. The implementation of this research is expected to produce an accurate information about the public response related to government policies to raise the excise of tobacco products so that it can be used as a basis or guidelines for the government to respond to the response of the Indonesian people.

Keywords: *Sentiment Analysis, Naive Bayes, Twitter, opinion*