

**ANALISIS SENTIMEN TERHADAP VAKSIN PADA MEDIA
SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE
BAYES DAN DECISION TREE**

SKRIPSI



disusun oleh

Arya Yoga Widyatama

18.11.2080

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**ANALISIS SENTIMEN TERHADAP VAKSIN PADA MEDIA
SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE
BAYES DAN DECISION TREE**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Arya Yoga Widyatama

18.11.2080

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP VAKSIN PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN
DECISION TREE**

yang disusun dan diajukan oleh

Arya Yoga Widyatama

18.11.2080

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 8 Juli 2022

Dosen Pembimbing,

Kusnawi, S.Kom. M.Eng

NIK. 190302112

PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN TERHADAP VAKSIN PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN DECISION TREE

yang dipersiapkan dan disusun oleh

Arya Yoga Widyatama

18.11.2080

telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Juli 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arief Setyanto, Dr.,S.Si, MT
NIK. 190302036

Donni Prabowo, M.Kom
NIK. 190302253

Kusnawi, S.Kom, M. Eng
NIK. 190302112

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Juli 2022

DEKAN FAKULTAS ILMU KOMPUTER

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

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 13 Agustus 2022



Arya Yoga Widyatama

NIM 18.11.2080

MOTTO

"Bermimpi adalah langkah pertama,
Dan bertindak adalah langkah selanjutnya"

"Percayalah bahwa hidup ini layak dijalani dan keyakinan kamu akan membantu
menciptakan fakta"

-William James-

"Jalani hidup sepenuhnya, dan fokus pada hal positif."

-Matt Cameron-

PERSEMBAHAN

Segala puji dan syukur kehadirat Allah SWT atas berkat, rahmat dan hidayah-Nya sehingga penulis mampu menyelesaikan skripsi dengan baik. Penulis juga mengucapkan terima kasih kepada pihak – pihak yang telah berkontribusi baik secara langsung maupun tidak langsung baik dalam penelitian maupun dalam penyusunan naskah. Skripsi ini saya persembahkan kepada :

1. Allah SWT untuk segala nikmat yang luar biasa, hambamu ini bersyukur atas nikmat yang telah diberikan.
2. Ibu dan bapak yang selalu memberikan kasih sayangnya secara tulus serta doa dan dukungannya selama menempuh perkuliahan ini dan terutama dalam penyusunan skripsi.
3. Sahabat dan rekan seperjuangan saya selama perkuliahan yaitu, Melano Habib, Yanuar Sadyatma, Dimas Priambodo, Nindra Reza, Wahid Rizka serta tidak lupa teman-teman lainnya dari Informatika 4.
4. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

KATA PENGANTAR

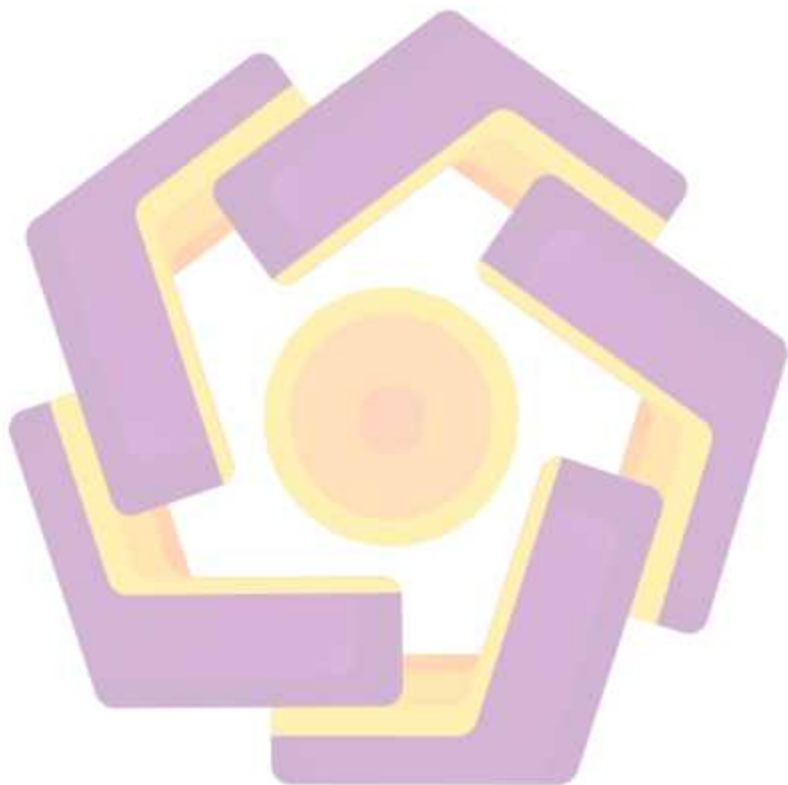
Segala puji dan syukur ke hadirat Tuhan Yang Maha Esa, karena dengan karunia dan rahmat-Nya, sehingga skripsi yang berjudul “Analisis Sentimen Terhadap Vaksin Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes Dan Decision Tree” dapat terselesaikan dengan baik.

Skripsi ini diajukan untuk memenuhi syarat akademik dalam menyelesaikan program Strata 1 Sarjana Ilmu Komputer di Universitas Amikom Yogyakarta. Selain itu, tujuan dari penulisan skripsi ini adalah untuk memberikan pengetahuan kepada pembaca mengenai Data Mining.

Proses penyelesaian skripsi ini tidak lepas dari bantuan, dukungan, kritik, dan saran dari berbagai pihak. Oleh karena itu penulis ingin menyampaikan terima kasih 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. Bapak Kusnawi, S.Kom, M.Eng. Selaku dosen pembimbing saya yang dengan sabar dan tulus membimbing serta memberikan ilmu kepada penulis.
4. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah banyak memberikan ilmunya selama kuliah.

5. Teman-teman dari kelas Informatika 4 yang telah menemani dan bersusah senang bersama selama perkuliahan.
6. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.



DAFTAR ISI

JUDUL.....	I
PERSETUJUAN.....	III
PENGESAHAN.....	IV
MOTTO.....	VI
PERSEMBAHAN.....	VII
KATA PENGANTAR.....	VIII
DAFTAR ISI.....	X
DAFTAR TABEL.....	XIII
DAFTAR GAMBAR.....	XV
INTISARI.....	XVI
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG MASALAH.....	1
1.2 RUMUSAN MASALAH.....	2
1.3 BATASAN MASALAH.....	3
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	3
1.5 MANFAAT PENELITIAN.....	4
1.6 METODE PENELITIAN.....	4
1.6.1 Metode Analisis.....	4
1.6.2 Metode Klasifikasi.....	5
1.6.3 Metode Evaluasi.....	5
1.7 SISTEMATIKA PENULISAN.....	6
BAB II LANDASAN TEORI.....	7
2.1 TINJAUAN PUSTAKA.....	7

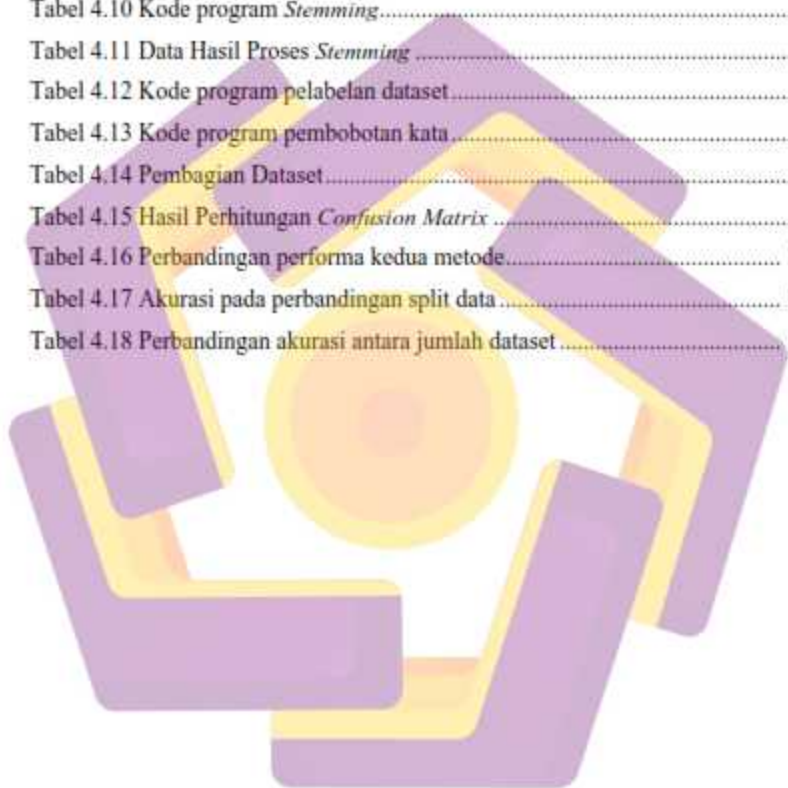
2.2	DATA MINING	11
2.3	ANALISIS SENTIMEN	11
2.4	KLASIFIKASI.....	12
2.5	TWITTER	12
2.6	VAKSIN	13
2.7	CRISP-DM.....	13
2.7.1	<i>Business Understanding</i>	14
2.7.2	<i>Data Understanding</i>	14
2.7.3	<i>Data Preparation</i>	14
2.7.4	<i>Modelling</i>	16
2.7.5	<i>Evaluation</i>	16
2.8	TF-IDF	17
2.9	NAIVE BAYES.....	18
2.9.1	<i>Multinomial Naïve Bayes</i>	19
2.10	DECISION TREE	21
2.11	CONFUSION MATRIX	22
2.12	PYTHON	23
BAB III METODE PENELITIAN.....		25
3.1	PEMAHAMAN BISNIS (<i>BUSINESS UNDERSTANDING</i>).....	25
3.2	PEMAHAMAN DATA (<i>DATA UNDERSTANDING</i>)	26
3.2.1	<i>Data Penelitian</i>	26
3.3	PERSIAPAN DATA (<i>DATA PREPARATION</i>).....	34
3.3.1	<i>Cleansing</i>	35
3.3.2	<i>Case Folding</i>	37
3.3.3	<i>Tokenizing</i>	39
3.3.4	<i>Filtering</i>	42
3.3.5	<i>Stemming</i>	44
3.3.6	<i>Pelabelan Data</i>	45
3.3.7	<i>Pembobotan Kata</i>	50
3.3.8	<i>Pembagian Data</i>	53

3.4 PEMODELAN (<i>MODELLING</i>)	53
3.4.1 <i>Klasifikasi</i>	53
3.5 EVALUASI (<i>EVALUATION</i>)	66
3.5.1 <i>Confusion Matrix</i>	67
BAB IV HASIL DAN PEMBAHASAN	68
4.1 ANALISIS KEBUTUHAN.....	68
4.1.1 <i>Perangkat Keras dan Perangkat Lunak</i>	68
4.2 HASIL DAN PEMBAHASAN.....	68
4.2.1 <i>Pengumpulan Data</i>	68
4.2.2 <i>Preprocessing Data</i>	69
4.2.3 <i>Implementasi Pelabelan Data</i>	94
4.2.4 <i>Implementasi Pembobotan Kata</i>	96
4.2.5 <i>Hasil Pembagian Data</i>	97
4.2.6 <i>Hasil Pengujian Algoritma Naïve Bayes</i>	98
4.2.7 <i>Hasil Perbandingan Uji Algoritma</i>	100
4.2.8 <i>Hasil Uji Dengan Jumlah Dataset Berbeda</i>	101
BAB V PENUTUP	103
5.1 KESIMPULAN.....	103
5.2 SARAN.....	103
DAFTAR PUSTAKA	105

DAFTAR TABEL

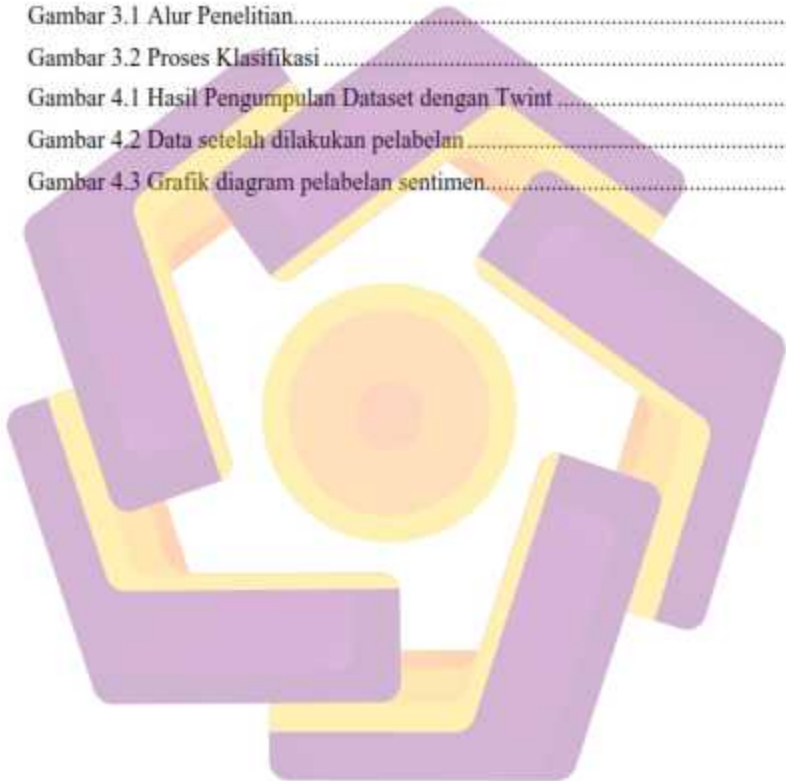
Tabel 2.1 Perbandingan hasil penelitian terdahulu	9
Tabel 3.1 Sampel Dataset Tweet.....	26
Tabel 3.2 Data hasil setelah proses <i>Cleansing</i>	35
Tabel 3.3 Data hasil setelah proses <i>Case Folding</i>	37
Tabel 3.4 Data hasil setelah proses <i>Tokenizing</i>	39
Tabel 3.5 Data setelah proses <i>Filtering</i>	42
Tabel 3.6 Data hasil setelah proses <i>Stemming</i>	44
Tabel 3.7 Sampel dataset setelah di labeling	46
Tabel 3.8 Sampel data perhitungan TF	50
Tabel 3.9 Sampel data perhitungan DF.....	51
Tabel 3.10 Sampel data perhitungan DF.....	51
Tabel 3.11 Sampel data perhitungan TF-IDF	52
Tabel 3.12 Sampel <i>data training</i>	54
Tabel 3.13 Hasil perhitungan <i>Prior Probabilities</i> setiap kelas	55
Tabel 3.14 Contoh term pada <i>data training</i>	56
Tabel 3.15 Model perhitungan klasifikasi	57
Tabel 3.16 Sampel <i>data testing</i>	58
Tabel 3.17 Contoh hasil perhitungan <i>Conditional Probabilitas</i>	59
Tabel 3.18 Hasil perhitungan pangkat nilai <i>Conditional Probabilitas</i>	59
Tabel 3.19 Sampel <i>data training</i>	62
Tabel 3.20 Nilai <i>data training</i>	62
Tabel 3.21 Perhitungan <i>Gain</i> untuk node akar.....	65
Tabel 3.22 Nilai <i>data training</i>	65
Tabel 3.23 Perhitungan <i>Gain</i> untuk simpul node	66
Tabel 3.24 Confusion Matrix	67
Tabel 4.1 Daftar perangkat keras dan lunak yang digunakan penelitian	68
Tabel 4.2 Kode program <i>Cleansing</i>	69
Tabel 4.3 Hasil Proses <i>Cleansing</i>	70
Tabel 4.4 Kode program <i>Case Folding</i>	75

Tabel 4.5 Data Hasil Proses <i>Case Folding</i>	75
Tabel 4.6 Kode program <i>Tokenizing</i>	79
Tabel 4.7 Data Hasil Proses <i>Tokenizing</i>	80
Tabel 4.8 Kode program <i>Filtering</i>	84
Tabel 4.9 Data Hasil Proses <i>Filtering</i>	85
Tabel 4.10 Kode program <i>Stemming</i>	90
Tabel 4.11 Data Hasil Proses <i>Stemming</i>	91
Tabel 4.12 Kode program pelabelan dataset.....	95
Tabel 4.13 Kode program pembobotan kata.....	97
Tabel 4.14 Pembagian Dataset.....	97
Tabel 4.15 Hasil Perhitungan <i>Confusion Matrix</i>	98
Tabel 4.16 Perbandingan performa kedua metode.....	100
Tabel 4.17 Akurasi pada perbandingan split data.....	101
Tabel 4.18 Perbandingan akurasi antara jumlah dataset.....	102



DAFTAR GAMBAR

Gambar 2.1 Icon Twitter.....	12
Gambar 2.2 Siklus tahapan <i>CRISP-DM</i>	13
Gambar 2.3 Tabel <i>Confusion Matrix</i>	22
Gambar 3.1 Alur Penelitian.....	25
Gambar 3.2 Proses Klasifikasi.....	54
Gambar 4.1 Hasil Pengumpulan Dataset dengan Twint.....	69
Gambar 4.2 Data setelah dilakukan pelabelan.....	96
Gambar 4.3 Grafik diagram pelabelan sentimen.....	96



INTISARI

Pandemi Covid-19 telah melanda Indonesia sudah hampir 2 tahun lamanya. Virus ini sangat berdampak sekali di berbagai sector bidang. Salah satu cara pemerintah untuk menanggulangi penyebaran virus agar tidak meluas yaitu dengan mengadakan program vaksinasi Covid-19 yang masih berlangsung hingga saat ini. Vaksin merupakan sebuah produk biologi yang dapat menimbulkan sistem kekebalan terhadap penyakit tertentu dan berfungsi untuk menstimulasi sistem imun tubuh. Dengan diadakannya program ini, banyak juga masyarakat yang setuju dan kurang setuju karena menganggap vaksin akan memberikan efek samping. Adapun tujuan dari penelitian ini yaitu untuk menganalisis sentimen masyarakat terhadap vaksin menggunakan data dari media sosial twitter. Dengan melakukan analisis sentimen, pemerintah atau masyarakat dapat melihat seberapa besar pengaruh vaksin yang dihasilkan untuk mengatasi penyebaran Covid-19 di Indonesia.

Dataset yang digunakan diambil menggunakan library *Twint* yaitu sebanyak 3000 tweet. Berdasarkan data yang diambil dari twitter akan diklasifikasikan menjadi 3 kelas, yaitu positif, netral dan negatif. Penelitian ini menggunakan metode pendekatan *CRISP-DM* (Cross-Industry Standard Process for Data Mining) untuk tahapan-tahapan analisis dan algoritma yang digunakan pada penelitian ini untuk klasifikasi adalah *Naïve Bayes* dan *Decision Tree*. Untuk mengukur performa dari kedua algoritma tersebut, peneliti menggunakan metode *Confusion Matrix*. Hasil pengukuran akhir berupa nilai akurasi, presisi dan recall.

Hasil penelitian yang telah dilakukan, algoritma *Naïve Bayes* dengan perbandingan dataset 80:20 hanya menghasilkan nilai akurasi sebesar 78%, nilai presisi sebesar 78%, nilai recall sebesar 80%. Sedangkan *Decision Tree* mendapatkan hasil akurasi sebesar 79%, presisi 80%, dan recall 80%. Hasil tersebut didapatkan berdasarkan perhitungan *Confusion Matrix*. Maka dapat disimpulkan bahwa pada penelitian ini algoritma *Decision Tree* mempunyai performa lebih baik. Berdasarkan penelitian ini, opini masyarakat terhadap vaksin cenderung mengarah ke sentimen netral.

Kata Kunci : *Vaksin, Decision Tree, Naïve Bayes, CRISP-DM, Twint, Confusion Matrix.*

ABSTRACT

The Covid-19 pandemic has hit Indonesia for almost 2 years. This virus is very impactful in various sectors of the field. One of the government's ways to prevent the spread of the virus from spreading is to hold a Covid-19 vaccination program, which is still ongoing today. Vaccines are biological products that can cause the immune system to fight certain diseases and function to stimulate the body's immune system. With the holding of this program, many people also agree and disagree because they think vaccines will have side effects. The purpose of this study is to analyze public sentiment towards vaccines using data from the social media twitter. By conducting a sentiment analysis, the government or the public can see how much influence the vaccine produced has to overcome the spread of Covid-19 in Indonesia.

The dataset used is taken using the Twint library, which is 3000 tweets. Based on the data taken from Twitter, it will be classified into 3 classes, namely positive, neutral and negative. This study uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) approach for the analysis stages and the algorithms used in this study for classification are Nave Bayes and Decision Tree. To measure the performance of the two algorithms, the researcher used the Confusion Matrix method. The final measurement results are in the form of accuracy, precision and recall values.

The results of the research that have been carried out, the Naïve Bayes algorithm with a dataset comparison of 80:20 only produces an accuracy value of 78%, a precision value of 78%, and a recall value of 80%. Meanwhile, Decision Tree gets 79% accuracy, 80% precision, and 80% recall. These results are obtained based on the calculation of the Confusion Matrix. So it can be concluded that in this study the Decision Tree algorithm has better performance. Based on this research, public opinion on vaccines tends to lead to a neutral sentiment.

Keywords : *Vaccine, Decision Tree, Naïve Bayes, CRISP-DM, Twint, Confusion Matrix.*