

**ANALISIS SENTIMEN PADA TWITTER MENGENAI RESESI
DI INDONESIA MENGGUNAKAN ALGORITMA NAÏVE
BAYES**

SKRIPSI

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



Disusun oleh

SHINTA NABILLA EMI HAYU SETYAWAN

19.11.2821

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

YOGYAKARTA

2023

**ANALISIS SENTIMEN PADA TWITTER MENGENAI RESESI
DI INDONESIA MENGGUNAKAN ALGORITMA NAÏVE
BAYES**

SKRIPSI

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



Disusun oleh

SHINTA NABILLA EMI HAYU SETYAWAN

19.11.2821

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN PADA TWITTER MENGENAI RESESI DI
INDONESIA MENGGUNAKAN ALGORITMA NAÏVE BAYES**

yang disusun dan diajukan oleh

Shinta Nabilla Emi Hayu Setyawan

19.11.2821

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 04 Agustus 2023

Dosen Pembimbing



Theopilus Bayu Sasongko, S.Kom, M.Eng

NIK. 190302375

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PADA TWITTER MENGENAI RESESI DI
INDONESIA MENGGUNAKAN ALGORITMA NAÏVE BAYES**

yang disusun dan diajukan oleh

Shinta Nabilla Emi Hayu Setyawan

19.11.2821

Telah dipertahankan di depan Dewan Penguji
pada tanggal 04 Agustus 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Alfie Nur Rahmi, M. Kom
NIK. 190302240



Yoga Pristyanto, S. Kom., M.Eng
NIK. 190302412



Theopilus Bayu Sasongko, S. Kom, M. Eng
NIK. 190302375



Skripsi ini telah diterima sebagai salah satu persyaratan
Untuk memperoleh gelar Sarjana Komputer
Tanggal 04 Agustus 2023

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PENYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Shinta Nabilla Emi Hayu Setyawan
NIM : 19.11.2821

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Pada Twitter Mengenai Resesi Di Indonesia Menggunakan Algoritma Naïve Bayes

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom, M.Eng

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, 04 Agustus 2023

Yang Menyatakan,



Shinta Nabilla Emi Hayu Setyawan

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur atas kesempatan yang telah Allah SWT berikan dan dengan telah diselesaikannya skripsi ini. Saya mempersembahkan skripsi ini kepada :

1. Allah SWT yang telah memberikan segala kebaikan dan karunia-Nya.
2. Kedua orang tua dan adik saya tercinta yang telah memberi dukungan moral, mendoakan, dan selalu memperhatikan kesehatan saya di setiap kondisi.
3. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng selaku dosen pembimbing yang selalu membimbing, mengarahkan, dan bahkan memudahkan jalannya proses skripsi ini.
4. Seluruh teman-teman 19 S1 IF 04, yang telah berjuang bersama selama perkuliahan ini.
5. Teman-teman yang tidak bisa saya sebutkan satu persatu, yang selalu membantu memberikan solusi dalam proses pengerjaan skripsi dan selalu memberikan *support*.

KATA PENGANTAR

Puji dan syukur saya panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat dan karunianya sehingga saya dapat menyelesaikan usulan penelitian yang berjudul “Analisis Sentimen Pada Twitter Mengenai Resesi Di Indonesia Menggunakan Algoritma *Naïve Bayes*”. Skripsi ini membahas tentang hasil implementasi suatu metode *naïve bayes* dengan variasi *multinomial naïve bayes* pada sentimen mengenai resesi pada twitter. Kajian tersebut didasarkan pada banyaknya informasi mengenai resesi yang tersebar di sosial media khususnya twitter untuk mendapatkan informasi secara detail mengenai hal tersebut.

Proses penyusunan hingga selesainya laporan skripsi ini tidak lepas dari bantuan, bimbingan dan dukungan dari berbagai pihak. Maka dari itu, saya tidak lupa mengucapkan terimakasih dan penghargaan kepada :

1. Bapak Prof. Dr. M. Suyanto , M.M, selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S. Kom,. M. Kom, selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Bapak Theopilus Bayu Sasongko, S. Kom,. M. Kom, selaku Dosen Pembimbing saya yang telah memberikan bimbingan dan *support* penuh dalam proses penulisan skripsi ini.
4. Bapak, Ibu, dan Adik saya tercinta yang telah memberikan *support*, motivasi dan doa terhadap saya.
5. Bapak dan Ibu Dosen serta staff Universitas AMIKOM Yogyakarta yang telah banyak memberikan ilmunya selama saya berkuliah.
6. Teman-teman jurusan Informatika Universitas AMIKOM Yogyakarta angkatan 19 yang telah berjuang bersama semasa perkuliahan.
7. Seluruh pihak yang tidak dapat disebutkan satu persatu yang telah banyak membantu dan mendukung sehingga skripsi ini dapat terselesaikan.

Yogyakarta, 28 Juli 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PENYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
INTISARI.....	xi
ABSTRACT.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur	6
2.2 Dasar Teori	11
2.2.1 <i>Data Mining</i>	11
2.2.2 <i>Sentiment Analysis</i>	12
2.2.3 Twitter	13
2.2.4 Pengertian Resesi	15
2.2.5 Python.....	15
2.2.6 <i>Naïve Bayes</i>	17
2.2.7 <i>Multinomial Naïve Bayes</i>	18
2.2.8 <i>Pre-processing</i>	19
2.2.9 TF-IDF (<i>Term Frequency – Inverse Document Frequency</i>).....	20

2.2.10	Pembagian Data.....	21
2.2.11	SMOTE (<i>Synthetic Minority Over-Sampling Technique</i>).....	22
2.2.12	Evaluation.....	22
BAB III METODOLOGI PENELITIAN.....		24
3.1	Alat dan Bahan Penelitian.....	24
3.1.1	Alat Penelitian.....	24
3.1.2	Bahan penelitian.....	24
3.2	Alur penelitian.....	26
3.3	Metode Pengolahan Data.....	26
3.3.1	<i>Crawling data</i>	27
3.3.2	<i>Preprocessing Data</i>	28
3.3.3	Pembobotan Kata.....	33
3.3.4	<i>Classification</i>	35
3.3.5	<i>Evaluation</i>	36
BAB IV HASIL DAN PEMBAHASAN.....		37
4.1	<i>Input Dataset</i>	37
4.2	<i>Preprocessing data</i>	39
4.2.1	<i>Cleaning Data</i>	39
4.2.2	<i>Case Folding</i>	44
4.2.3	<i>Tokenization</i>	45
4.2.4	<i>Stopword Removal</i>	45
4.2.5	<i>Stemming</i>	45
4.2.6	<i>Transformation</i>	46
4.3	<i>Pembobotan Kata</i>	46
4.4	<i>Classification</i>	56
4.5	<i>Evaluation</i>	57
BAB V PENUTUP.....		60
5.1	Kesimpulan.....	60
5.2	Saran.....	60
DAFTAR PUSTAKA.....		61

DAFTAR TABEL

Tabel 2.1 Studi literatur.....	8
Tabel 3.1 Perangkat Keras	24
Tabel 3.2 Perangkat Lunak	24
Tabel 3.3 Data awal.....	25
Tabel 4.1 Dataset.....	37
Tabel 4.2 data sebelum cleaning	40
Tabel 4.3 Dataset setelah labeling.....	41
Tabel 4.4 data setelah di cleaning	42
Tabel 4.5 Hasil case folding.....	44
Tabel 4.6 Hasil tokenization	45
Tabel 4.7 Hasil stopword removal	45
Tabel 4.8 Hasil stemming	46
Tabel 4.9 Hasil transformation.....	46
Tabel 4.10 Dokumen pembobotan kata	47
Tabel 4.11 Hasil tf.....	48
Tabel 4.12 TF(term frequency).....	50
Tabel 4.13 frekuensi kata untuk seluruh dokumen	52
Tabel 4.14 Hasil TF-IDF.....	54
Tabel 4.15 Classification report sebelum SMOTE	57
Tabel 4.16 Classification report sesudah SMOTE.....	58

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian.....	26
Gambar 3.2 Alur <i>Crawling Data</i>	27
Gambar 3.3 Alur <i>Cleaning Data</i>	28
Gambar 3.4 Alur <i>Case Folding</i>	29
Gambar 3.5 Alur <i>Tokenization</i>	30
Gambar 3.6 Alur <i>Stopword Removal</i>	31
Gambar 3.7 Alur <i>Stemming</i>	32
Gambar 3. 8 Alur <i>Text Preprocessing Pipeline</i>	33
Gambar 3. 9 Alur Pembobotan TF-IDF	34
Gambar 3. 10 <i>Classification Multinomial Naïve Bayes</i>	35
Gambar 4.1 Tahapan <i>Preprocessing data</i>	39
Gambar 4.2 Diagram Jumlah Analisis Sentimen.....	44
Gambar 4.3 Diagram setelah SMOTE	57
Gambar 4. 4 visualisasi <i>word cloud</i>	59

INTISARI

Resesi merupakan kondisi pada perekonomian negara yang sedang mengalami penurunan dimana itu sangat berdampak pada kehidupan sehari-hari masyarakat. Isu terkait resesi banyak di perbincangkan oleh masyarakat dengan opini yang berbeda-beda sehingga munculah suatu sentimen seperti negatif, positif, dan netral. Masyarakat kini banyak menggunakan media sosial untuk menyampaikan tanggapan mereka, salah satu media sosial yang sering digunakan yaitu Twitter. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen mengenai resesi di Indonesia pada Twitter menggunakan metode *Naïve Bayes* dengan variasi *Multinomial Naïve Bayes*. Data yang digunakan dalam penelitian ini diperoleh dengan cara mengumpulkan data dari Twitter dengan *keyword* resesi. Ada beberapa sentimen yang di analisis seperti positif, negatif, dan netral. Metode yang digunakan yaitu *Multinomial Naïve Bayes*, metode ini digunakan untuk menganalisis dan memprediksi sentimen yang ada pada tiap *tweet*. Ada beberapa tahapan yang dilakukan yaitu *crawling data*, *preprocessing data*, pembobotan kata, *classification*, dan *evaluation*. Pada tahap *preprocessing data* ada beberapa proses yang dilakukan yaitu *cleaning data*, *case folding*, *tokenization*, *stopword removal*, *stemming*, dan *transformation*. Pada tahap pembobotan kata dilakukan dengan TF-IDF (*Term Frequency- Inverse Document Frequency*). Dan pada tahap *classification* yaitu dengan menggunakan algoritma *multinomial naïve bayes*. Hasil pengujian dari dataset sebanyak 1006 data dan menggunakan *multinomial naïve bayes* sebagai klasifikasinya memperoleh hasil akurasi sebesar 74%, hasil tersebut merupakan hasil sebelum SMOTE (*Synthetic Minority Over-Sampling*). Setelah SMOTE hasil akurasinya sebesar 76%. Dan untuk hasil pengujian dengan variasi *7-fold cross validation* yaitu 73%.

Kata Kunci : *Naïve Bayes*, *Multinomial Naïve Bayes*, Twitter, Analisis Sentimen

ABSTRACT

A recession is a condition in a country's economy that is experiencing a decline where it greatly impacts people's daily lives. Issues related to the recession are widely discussed by the public with different opinions so that sentiments such as negative, positive, and neutral arise. People now use social media a lot to convey their responses, one of the social media that is often used is Twitter. Therefore, this study aims to analyze sentiment regarding the recession in Indonesia on Twitter using the Naïve Bayes method with a variation of Multinomial Naïve Bayes. The data used in this research was obtained by collecting data from Twitter with the recession keyword. There are several sentiments that are analyzed such as positive, negative, and neutral. The method used is Multinomial Naïve Bayes, this method is used to analyze and predict the sentiments that exist in each tweet. There are several stages carried out, namely crawling data, data preprocessing, word weighting, classification, and evaluation. At the data preprocessing stage, there are several processes carried out, namely data cleaning, case folding, tokenization, stopword removal, stemming, and transformation. The word weighting stage is carried out using TF-IDF (Term Frequency- Inverse Document Frequency). And in the classification stage, using the multinomial naïve Bayes algorithm. The test results from a dataset of 1006 data and using multinomial naïve Bayes as the classification obtain an accuracy of 74%, these results are the results before SMOTE (Synthetic Minority Over-Sampling). After SMOTE the accuracy is 76%. And for the test results with a 7-fold cross validation variation, namely 73%.

Keyword : *Naïve Bayes, Multinomial Naïve Bayes, Twitter, Sentiment Analysis*