

**ANALISIS SENTIMEN KOMENTAR INSTAGRAM PADA
PROGRAM KAMPUS MERDEKA DENGAN ALGORITMA NAIVE
BAYES DAN DECISION TREE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

BAYU WICAKSONO

20.11.3549

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS SAMIKOMYOGYAKARTA
YOGYAKARTA
2024**

**ANALISIS SENTIMEN KOMENTAR INSTAGRAM PADA
PROGRAM KAMPUS MERDEKA DENGAN ALGORITMA NAIVE
BAYES DAN DECISION TREE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

BAYU WICAKSONO

20.11.3549

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS SAMIKOMYOYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN KOMENTAR INSTAGRAM PADA PROGRAM
KAMPUS MERDEKA DENGAN ALGORITMA NAIVE BAYES DAN
DECISION TREE**

yang disusun dan diajukan oleh

Bayu wicaksono

20.11.3549

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Februari 2024

Dosen Pembimbing,



Nuri Cahyono, M.Kom

NIK. 190302278

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN KOMENTAR INSTAGRAM PADA PROGRAM
KAMPUS MERDEKA DENGAN ALGORITMA NAIVE BAYES DAN
DECISION TREE

yang disusun dan diajukan oleh

Bayu wicaksono

20.11.3549

Telah dipertahankan di depan Dewan Penguji
pada Tanggal 20 Februari 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Andi Sunyoto, M.Kom., Dr.
NIK. 190302052

Rumini, M.Kom
NIK. 190302246

Nuri Cahyono, M.Kom
NIK. 190302278



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Februari 2024

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Bayu wicaksono
NIM : 20.11.3549

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Komentar Instagram Pada Program Kampus Merdeka Dengan Algoritma Naive Bayes Dan Decision Tree

Dosen Pembimbing : Nuri Cahyono, M.Kom

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, 20 Februari 2024

Yang Menyatakan,



Bayu wicaksono

KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **“ANALISIS SENTIMEN KOMENTAR INSTAGRAM PADA PROGRAM KAMPUS MERDEKA DENGAN ALGORITMA NAIVE BAYES DAN DECISION TREE”**. Penyusunan skripsi ini ditunjukkan untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika.

Penulis menyadari bahwa terselesaikannya skripsi ini tidak terlepas dari bantuan berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih sebesar-besarnya kepada :

1. Bapak Nuri Cahyono, M.Kom, selaku Dosen Pembimbing Skripsi yang telah membantu, membimbing dan mengarahkan dalam penyusunan skripsi ini.
2. Kedua orang tua, serta keluarga yang telah memberikan dukungan secara moril dan materil dalam proses penyusunan skripsi ini.
3. Teman-teman mahasiswa yang telah mendukung pelaksanaan pengerjaan skripsi ini.

Penulis menyadari sepenuhnya dalam penyusunan skripsi ini masih banyak kekurangan, oleh karena itu penulis terus menunggu saran dan kritik yang membangun dan positif dari para pembaca dan pengguna skripsi ini. Semoga hasil penelitian ini dapat bermanfaat bagi pembaca dan pihak yang berkepentingan. Amin.

Yogyakarta, 20 Februari 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	21
2.2.1 <i>Text Mining</i>	21
2.2.2 Analisis Sentimen	21
2.2.3 Instagram	21
2.2.4 Program Kampus Merdeka	22
2.2.5 <i>Preprocessing</i>	22
2.2.6 VADER Sentimen	22
2.2.7 Google Collab	23
2.2.8 TF-IDF	23
2.2.9 Model Naïve Bayes Classifier	24

2.2.10	<i>Decision Tree</i>	24
2.2.11	Synthetic Minority Over-sampling Technique (SMOTE)	25
2.2.12	Evaluasi Model	25
BAB III METODE PENELITIAN		27
3.1	Alur Penelitian	27
3.1.1	Pengambilan Data	27
3.1.2	<i>Preprocessing</i>	28
3.1.3	<i>Labeling</i>	29
3.1.4	Balancing Data	29
3.1.5	Pembobotan kata	30
3.1.6	<i>Splitting data</i>	30
3.1.7	Penerapan Complement Naïve Bayes dan Decision Tree	30
3.1.8	Evaluasi Hasil	31
3.2	Alat dan Bahan	31
3.2.1	Data Penelitian	31
3.2.2	Alat	32
BAB IV HASIL DAN PEMBAHASAN		33
4.1	<i>Scrapping Data</i>	33
4.2	<i>Preprocessing Data</i>	34
4.2.2	<i>Cleansing</i>	35
4.2.3	<i>Stopwords</i>	36
4.2.5	<i>Stemming Sastrawi</i>	38
4.2.6	Normalisasi Kata	39
4.3	<i>Labeling</i>	40
4.4	<i>Balancing Data</i>	41
4.5	Pembobotan Kata	42
4.6	<i>Splitting Data</i>	43
4.7	Penerapan Complement Naïve Bayes dan Decision Tree	44
4.7.1	<i>Complement Naïve Bayes</i>	44
4.7.2	<i>Decision Tree</i>	45
4.8	Evaluasi Hasil	46
BAB V PENUTUP		50
5.1	Kesimpulan	50
5.2	Saran	50
REFERENSI		52
LAMPIRAN		54

DAFTAR TABEL

Tabel 4. 2 Hasil <i>Case Folding</i>	34
Tabel 4. 3 Hasil <i>Cleansing</i>	35
Tabel 4. 4 Hasil <i>Stopwords</i>	36
Tabel 4. 5 Hasil <i>Tokenizing</i>	37
Tabel 4. 6 Hasil <i>Stemming</i> Sastrawi.....	38
Tabel 4. 7 Hasil <i>Normalisasi Kata</i>	39
Tabel 4. 8 Hasil <i>Labeling</i>	40
Tabel 4. 9 Frekuensi kemunculan kata.....	43
Tabel 4. 10 Skala pembagian data.....	43
Tabel 4. 11 <i>imbalance dataset CNB</i>	45
Tabel 4. 12 <i>balance dataset CNB</i>	45
Tabel 4. 13 <i>imbalance dataset Decision Tree</i>	46
Tabel 4. 14 <i>balance dataset Decision Tree</i>	46
Tabel 4. 15 Perbandingan Model <i>Naïve Bayes</i> dan <i>Decision Tree Confusion Matrix</i>	47

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	27
Gambar 3. 2 Alur Vader Sentimen	29
Gambar 4. 1 Hasil <i>Scrapping</i> data phantombuster	33
Gambar 4. 2 Visualiasi persebaran data	41
Gambar 4. 3 Visualisasi <i>Imbalance</i> data	42
Gambar 4. 4 Visualiasi setelah <i>balancing</i> data	42
Gambar 4. 5 <i>Confusion Matrix Decision Tree balance dataset</i>	48
Gambar 4. 6 <i>Confusion Matrix Complement Naïve Bayes balance dataset</i>	49



DAFTAR LAMPIRAN

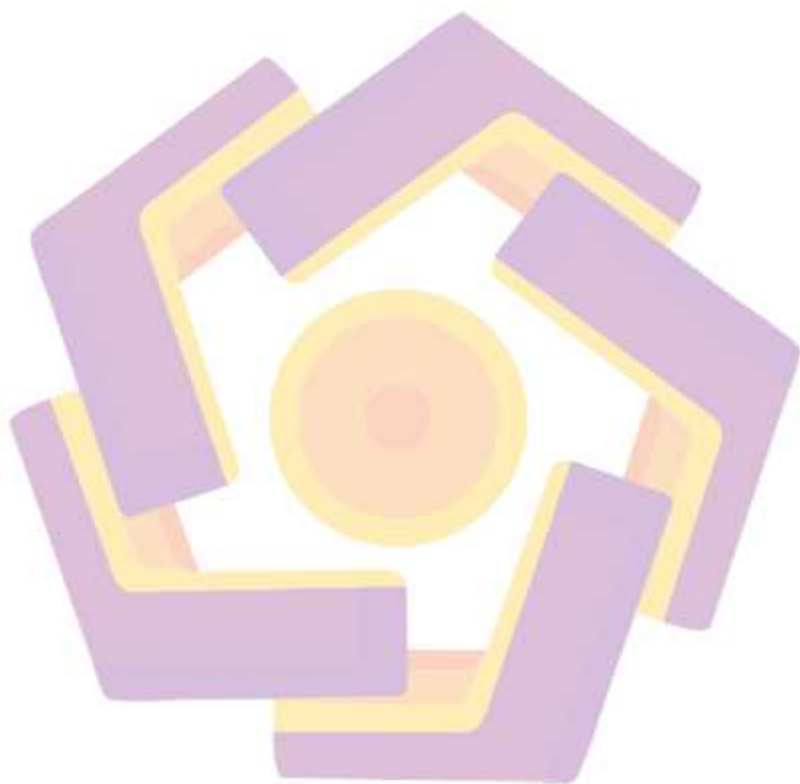


DAFTAR LAMBANG DAN SINGKATAN

SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
CNB	<i>Complement Naïve Bayes</i>
NLP	<i>Natural Language Processing</i>
TF	<i>Term Frequency</i>
IDF	<i>Inverse Document Frequency</i>
MNB	<i>Multinomial Naïve Bayes</i>



DAFTAR ISTILAH



INTISARI

Instagram merupakan salah satu sosial media yang digunakan merepresentasikan diri, berinteraksi, dan mencari informasi. Kita dapat mengambil sekumpulan informasi ke dalam bentuk dataset untuk diolah lebih lanjut. Berkaitan dengan hal itu, Program Kampus Merdeka sebagai objek analisis, mengingat Program Kampus Merdeka adalah program pemerintah yang saat ini sedang dijalankan oleh Kemendikbud. Pengambilan dataset yang didapat dari kumpulan komentar Instagram, tool yang digunakan adalah phantombuster. Menggunakan bahasa pemrograman python dengan tools Google Collab, dengan Algoritma Naïve Bayes Classifier dan Decision Tree untuk membuat model sentimen. Hasil scrapping mendapatkan 1764 data, dan sesudah dilakukan pre-processing menjadi 1694 data. Dari sentimen analisis yang telah dilakukan diperoleh hasil dari penerapan Algoritma Complement Naïve Bayes dan Decision Tree, sebelum dilakukan SMOTE over-sampling, perbandingan data positif dan negatif sebesar 35,06% banding 64,95%, dengan akurasi model Decision Tree 84% dengan skenario pembagian data 90:10 dan model Complement Naïve Bayes 81% pada skenario pembagian data 80:20. Setelah dilakukan balancing data menggunakan SMOTE over-sampling, akurasi pada model Decision Tree naik sebesar 1% dari 86% menjadi 85%, dengan skenario pembagian data 90:10, dan pada model Complement Naïve Bayes juga mengalami kenaikan sebesar 2%, dari 82% menjadi 83% dengan skenario pembagian data 80:20.

Kata kunci: *Sentimen Analisis, Naïve Bayes, Decision Tree, Machine Learning, Google Colab*

ABSTRACT

Instagram is a social media that is used to represent oneself, interact and search for information. We can take a collection of information into a dataset for further processing. In this regard, the Program Kampus Merdeka is the object of analysis, considering that the Program Kampus Merdeka is a government program which is currently being implemented by the Kemendikbud. Retrieving a dataset obtained from a collection of Instagram comments, the tool used is Phantombuster. Using the Python programming language with Google Colab tools, with the Naïve Bayes Classifier and Decision Tree algorithm to create a sentiment model. The scrapping results obtained 1764 data, and after pre-processing it became 1694 data. From the sentiment analysis that has been carried out, the results obtained from the application of the Complement Naïve Bayes Algorithm and Decision Tree, before SMOTE over-sampling are carried out, the comparison of positive and negative data is 35.06% versus 64.95%, with a Decision Tree model accuracy of 84% with the scenario data sharing 90:10 and the Complement Naïve Bayes model 81% in the 80:20 data sharing scenario. After balancing the data using SMOTE over-sampling, the accuracy of the Decision Tree model increased by 1% from 86% to 85%, with a 90:10 data sharing scenario, and the Complement Naïve Bayes model also increased by 2%, from 82% to 83% with an 80:20 data sharing scenario.

Keyword: *Sentimen Analisis, Naïve Bayes, Decision Tree, Machine Learning, Google Colab*