

KLASIFIKASI AKUN BUZZER PADA TWITTER MENGUNAKAN ALGORITMA NAIVE BAYES

SKRIPSI

Diajukan untuk memenuhi persyaratan
dalam memperoleh gelar Sarjana
Komputer Program Studi Sistem
Informasi



Disusun Oleh:

CATUR ARPAL PERKASA

19082010121

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

SURABAYA

2023

KLASIFIKASI AKUN BUZZER PADA TWITTER MENGGUNAKAN
ALGORITMA NAIVE BAYES

Disusun Oleh:

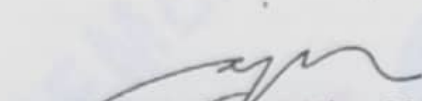
CATUR ARPAL PERKASA

19082010121

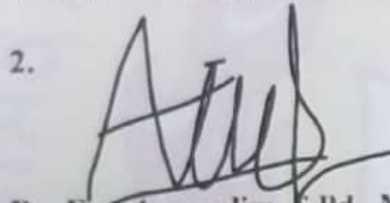
Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi
Program Studi Sistem Informasi Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Pada Tanggal 13 Januari 2023

Pembimbing:

1.



Amalia Anjani S., S.Kom, M.Kom.
NIP. 19920812 2018032 001

2.



Dr. Eng Agussalim, S.Pd., M.T.
NIP. 19850811 2019031 005

Tim Penguji:


1.


Prisa Marga K., S.Kom, MCs.
NIP. 19821125 2021211 003

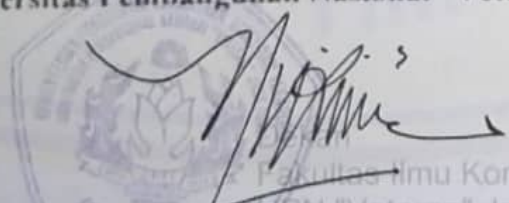
2.


Rizka Hadiwiyanti, S.Kom, M.Kom.
NIP. 19860727 2018032 001

3.


Dhian SMria V. K., S.Kom., M.Kom.
NPT. 201198 60 522249

Mengetahui,
Dekan Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur


Dr. Novrina Hendrasarie, S.T., M.T.
NIP. 19681126 199403 2 001



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
FAKULTAS ILMU KOMPUTER

KETERANGAN REVISI

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : Catur Arpal Perkasa
NPM : 19082010121
Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 13 Januari 2023 dengan judul:

KLASIFIKASI AKUN BUZZER PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES

Oleh karenanya mahasiswa tersebut diatas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diizinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, Januari 2023

Dosen penguji yang memeriksa revisi:

1. Prisa Marga K., S.Kom. MCs.
NIP. 19821125 2021211 003
2. Rizka Hadiwiyanti, S.Kom. M.Kom.
NIP. 19860727 2018032 001
3. Dhian Satria Y. K., S.Kom., M.Kom.
NPT. 201198 60 522249

{  }


{  }

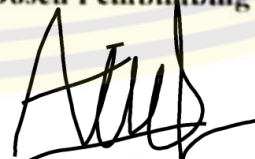
{  }
1902
15.01

Mengetahui,

Dosen Pembimbing 1

Dosen Pembimbing 2


Amalia Anjani A., S.Kom., M.Kom.
NIP. 19920812 2018032 001


Dr. Eng Agussalim, S.Pd., M.T.
NIP. 19850811 2019031 00

LEMBAR PENGESAHAN

KLASIFIKASI AKUN BUZZER PADA TWITTER MENGGUNAKAN
ALGORITMA NAIVE BAYES

Disusun Oleh:


CATUR ARPAL PERKASA
19082010121


Telah disetujui mengikuti Ujian Negara Lisan Gelombang Januari
Periode 2023 pada Tanggal 13 Januari 2023

Menyetujui,

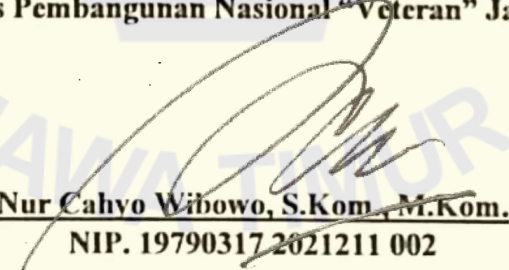
Dosen Pembimbing 1

Dosen Pembimbing 2


Amalia Anjani A., S.Kom, M.Kom.
NIP. 19920812 2018032 001


Dr. Eng Agussalim, S.Pd., M.T.
NIP. 19850811 2019031 005

Mengetahui,
Ketua Program Studi Sistem Informasi
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur


Nur Cahyo Wibowo, S.Kom, M.Kom.
NIP. 19790317 2021211 002



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
FAKULTAS ILMU KOMPUTER

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Catur Arpal Perkasa
NPM : 19082010121
Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**KLASIFIKASI AKUN BUZZER PADA TWITTER MENGGUNAKAN
ALGORITMA NAIVE BAYES**

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain dan juga bukan merupakan Produk / 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 terbukti benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari.

Surabaya, 19 Januari 2023

Hormat Saya,



Catur Arpal Perkasa
19082010121

Judul : **KLASIFIKASI AKUN *BUZZER* PADA TWITTER
MENGUNAKAN ALGORITMA NAIVE BAYES**

Pembimbing 1 : **Amalia Anjani, S.Kom, M.Kom.**

Pembimbing 2 : **Dr. Eng. Agussalim, MT.**

ABSTRAK

Strategi pemasaran dengan pendekatan eWOM saat ini menjadi tren baru dalam bidang pemasaran. Salah satunya dengan menggunakan konten fabrikasi yang disebarluaskan oleh *buzzer*. Konten fabrikasi merupakan informasi buatan yang dengan sengaja disebar oleh individu yang dikenal sebagai *buzzer*. Dalam penelitian ini, akan dikaji pola dari akun-akun yang terlibat dalam penyebarluasan kampanye IPO perusahaan *e-commerce* di Indonesia yang *trending* di Twitter.

Penelitian ini menggunakan 4 atribut untuk mengklasifikasikan akun *buzzer* yaitu jumlah *following*, jumlah *follower*, nilai sentimen dari *recent tweets*, dan umur akun. Data akan diproses dan dibersihkan sebelum melakukan analisis sentimen untuk memberikan bobot pada data. Kemudian data dilabeli sesuai dengan karakteristik yang telah ditentukan. Algoritma Naïve Bayes akan digunakan untuk melakukan klasifikasi akun *buzzer* dan *non-buzzer*.

Hasil penelitian yang berupa halaman pengklasifikasian *buzzer* dibuat dengan HTML dan CSS untuk *interface* dan Python Flask untuk *backend*-nya. Uji validasi dilakukan menggunakan dataset sebanyak 70 yang terbagi masing-masing 35 untuk *buzzer* dan *non-buzzer*. Hasilnya, pengklasifikasi *buzzer* salah melakukan prediksi kelas sebanyak 7 dari 70 data yang diujikan. Sehingga, pengklasifikasi *buzzer* ini memiliki nilai akurasi sebesar 0.90.

Kata kunci: *Buzzer*, *Gaussian Naïve Bayes*, Klasifikasi, Twitter

KATA PENGANTAR

Segala puji syukur penulis panjatkan atas kehadiran Allah SWT karena atas berkat dan hidayah dari Allah SWT yang membuat penulis dapat menyelesaikan skripsi yang berjudul "Klasifikasi Akun Buzzer Pada Twitter Menggunakan Algoritma Naive Bayes" sebagai syarat untuk menyelesaikan program studi Sistem Informasi di Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Dalam upaya menyelesaikan skripsi ini, tidak dapat dilepaskan dari dukungan yang diterima dari berbagai pihak, dan dalam kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua dan keluarga yang telah senantiasa mendoakan dan mendukung sehingga penulis terus termotivasi dalam menyelesaikan skripsi ini;
2. Bapak Nur Cahyo Wibowo, S.Kom, M.Kom. selaku Koordinator Program Studi Sistem Informasi UPN "Veteran" Jawa Timur yang turut membimbing serta memberi kesempatan untuk penulis, terutama dalam urusan administrasi;
3. Bapak Agung Brastama Putra, S.Kom, M.Kom selaku dosen wali yang selalu memberikan motivasi dan arahan dalam periode studi;
4. Bu Amalia Andjani, S.Kom, M.Kom dan Bapak Dr. Eng. Agussalim, M.T. selaku dosen pembimbing skripsi yang telah membimbing dengan sabar, memberikan ilmu berharga serta motivasi yang sangat bermanfaat sehingga penulis dapat menyelesaikan skripsi ini.

5. Seluruh bapak dan ibu dosen program studi Sistem Informasi UPN “Veteran” Jawa Timur yang telah memberikan banyak ilmu pengetahuan yang bermanfaat selama masa perkuliahan;
6. Terima kasih kepada Daffa, Tamariskom, Feby, Suhri, Kevin dan Sahabat Pena yang terlibat langsung dari masa penelitian akademis hingga penyusunan laporan.
7. Terima kasih kepada Sistem Informasi Angkatan 2019 (Solasifo) yang telah memberikan dukungan dan berjuang bersama mulai dari masa orientasi maba hingga saat ini;
8. Terima kasih kepada Google Developer Communities, yang telah mendorong dan memberi motivasi untuk penyelesaian penelitian ini.
9. Terima kasih kepada seluruh teman-teman yang selalu mendukung dan memberikan masukan dan semangat hingga saat ini.

Semoga Allah SWT membalas kebaikan pihak-pihak diatas. Dalam pengerjaan skripsi ini penulis menyadari masih banyak sekali kekurangan. Semoga skripsi ini bermanfaat bagi semua pihak.

Surabaya, 9 Januari 2023

Penulis

DAFTAR ISI

ABSTRAK	i
DAFTAR ISI.....	iv
DAFTAR TABEL.....	vii
DAFTAR GAMBAR	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan.....	6
1.5 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	9
2.1 Dasar Teori	9
2.1.1 <i>Buzzer</i>	9
2.1.2 <i>Text Mining</i>	9
2.1.3 Klasifikasi	10
2.1.4 Analisis Sentimen	10
2.1.5 Twitter	12
2.1.6 Vader	13
2.1.7 Social Network Scrape.....	13
2.1.8 Python-Twitter	14
2.1.9 Pandas	15
2.1.10 Naïve Bayes	16
2.1.11 Flask	17
2.1.12 Sklearn.....	18
2.2 Penelitian Terdahulu.....	18
BAB III METODOLOGI PENELITIAN.....	23
3.1 Studi Literatur.....	23
3.2 Analisis Kebutuhan	24

3.2.1	Kebutuhan Data.....	24
3.2.2	Kebutuhan <i>Software</i> dan <i>Hardware</i>	24
3.3	Preparasi Data Klasifikasi	24
3.3.1.	<i>Scrape Username</i> dan <i>Tweet</i> Dengan SNScrape	25
3.3.2.	<i>Scrape</i> Informasi Akun	25
3.3.3.	Eksplorasi Dataset.....	26
3.3.4.	Visualisasi Distribusi Data.....	26
3.3.5.	Pelabelan Kelas	26
3.4	Perancangan Model Klasifikasi.....	28
3.4.1	Perancangan Model.....	28
3.4.2	Training Model Naïve Bayes	29
3.4.3	Pembobotan Klasifikasi	29
3.5	Pengklasifikasi <i>Buzzer</i>	29
BAB IV HASIL DAN PEMBAHASAN		31
4.1	Implementasi Kebutuhan.....	31
4.1.1	Kebutuhan Data.....	31
4.1.2	Kebutuhan <i>Hardware</i> dan <i>Software</i>	31
4.2	Preparasi Data Klasifikasi	32
4.2.1	<i>Scrape Username</i> dan <i>Tweet</i> Dengan SNScrape	32
4.2.2	<i>Scrape</i> Informasi Akun	34
4.2.3	Eksplorasi Dataset.....	38
4.2.4	Visualisasi Distribusi Data.....	40
4.2.5	Pelabelan Kelas	43
4.2.6	Penyeimbangan Kelas	45
4.3	Perancangan Model Klasifikasi.....	46
4.3.1	Perancangan Model.....	46
4.3.2	Training Model Naïve Bayes	48
4.3.3	Pembobotan Klasifikasi	49
4.4	Pengklasifikasi <i>Buzzer</i>	54
4.4.1	<i>Website</i> Python Flask	54
4.4.2	<i>Website Interface</i>	58
4.4.3	Uji Validasi	62
4.4.4	Hasil Uji Validasi.....	63

BAB V KESIMPULAN DAN SARAN.....	64
5.1 Kesimpulan.....	64
5.2 Saran.....	65
DAFTAR PUSTAKA	67

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	18
Tabel 3.2 Atribut Pelabelan Kelas	28
Tabel 4.3 Detail Jumlah Dataset	45

DAFTAR GAMBAR

Gambar 2.1 Tahapan Dalam Analisis Sentimen	11
Gambar 3.1 Alur Metode Penelitian	23
Gambar 3. 2 Alur Perancangan Model Klasifikasi	28
Gambar 4.1 Baris Perintah Untuk <i>Scrape</i> Data	33
Gambar 4.2 <i>Source Code</i> Untuk <i>Parse Username</i>	33
Gambar 4.3 Contoh Hasil <i>Scrape</i>	33
Gambar 4.4 <i>Source Code</i> Mengekspor ke Format CSV	34
Gambar 4.5 <i>Source Code Scrape</i> Hingga 10 <i>Tweet</i> Terakhir dari <i>Username</i>	34
Gambar 4.6 Contoh Data Hasil dari Proses <i>Scrape</i>	35
Gambar 4.7 <i>Source Code</i> Penimbangan Sentimen	35
Gambar 4.8 Contoh Data Setelah Penimbangan	36
Gambar 4.9 <i>Source Code</i> untuk <i>parse compound</i> serta mencari nilai <i>mean</i> dan median.....	36
Gambar 4.10 Contoh Data Nilai <i>Mean</i> dan Median dari Nilai <i>Compound</i>	37
Gambar 4.11 <i>Source Code</i> Mendapatkan Informasi <i>account_age</i> , <i>following</i> , dan <i>followers</i>	38
Gambar 4.12 Contoh Data yang Telah Digabungkan dengan Informasi Akun	38
Gambar 4.13 <i>Source Code</i> Untuk Mengkalkulasi Nilai Sentimen	39
Gambar 4.14 Contoh Hasil Data	39
Gambar 4.15 <i>Source Code</i> Menampilkan Visualisasi Distribusi.....	41
Gambar 4.16 Visualisasi Distribusi Nilai <i>Compound</i> Pada <i>Tweet</i> Yang Berkaitan Dengan Topik IPO <i>E-commerce</i>	41
Gambar 4.17 Nilai <i>Mean</i> Berdasarkan Topik	41
Gambar 4.18 Visualisasi Distribusi Nilai <i>Compound</i> Dari <i>Tweet</i> Milik Akun yang Ikut Meramaikan Topik IPO <i>E-commerce</i>	42
Gambar 4.19 <i>Source Code</i> Fungsi Untuk Melakukan Pelabelan.....	43
Gambar 4.20 <i>Source Code</i> Fungsi Untuk Melakukan Pelabelan.....	44
Gambar 4.21 Contoh Hasil Data Setelah Pelabelan.....	44
Gambar 4.22 <i>Source Code</i> Untuk Penyeimbangan Kelas dan Jumlah Kelas Setelah Diseimbangkan.....	45
Gambar 4.23 <i>Import Library</i> dan Dataset	46
Gambar 4.24 Persiapan Dataset dan Contohnya	47
Gambar 4.25 Pembagian Dataset Sebelum Training dan <i>Initialize</i> Gaussian Naïve Bayes	47
Gambar 4.26 <i>Source Code</i> untuk <i>Fit</i> Model	48
Gambar 4.27 <i>Source Code</i> untuk Melakukan Prediksi Pada Data.....	48
Gambar 4.28 Grafik <i>Confusion Matrix</i> dari Skenario 1.....	50
Gambar 4.29 Grafik <i>ROC Curve</i> dari Skenario 1	50
Gambar 4.30 Grafik <i>Precision-Recall Curve</i> dari Skenario 1	51

Gambar 4.31 Grafik <i>Confusion Matrix</i> dari Skenario 2.....	52
Gambar 4.32 Grafik <i>ROC Curve</i> dari Skenario 2.....	52
Gambar 4.33 Grafik <i>Precision-Recall Curve</i> dari Skenario 2.....	53
Gambar 4.34 Model Disimpan Dengan Format <i>.pkl</i>	54
Gambar 4.35 Import <i>Library Flask</i> dan <i>Pickle</i>	55
Gambar 4.36 Membuat <i>Instance Flask</i> dan <i>Load Model</i> Pada <i>Flask</i>	55
Gambar 4.37 Membuat <i>Instance Twitter</i> dengan <i>Personal Credential</i>	55
Gambar 4.38 <i>Function</i> Untuk Mendapatkan <i>Following, Followers</i> , dan Umur Akun.....	56
Gambar 4.39 <i>Function</i> Untuk Mendapatkan <i>Mean</i> dari Nilai <i>Compound</i>	56
Gambar 4.40 <i>Function</i> Yang Akan di- <i>Export</i> Sebagai Modul.....	57
Gambar 4.41 Contoh Hasil <i>Scrape</i> dari <i>Function</i>	57
Gambar 4.42 <i>Source Code</i> Untuk <i>Route</i> ke <i>index.html</i>	58
Gambar 4.43 <i>Source Code</i> <i>index.html</i>	59
Gambar 4.44 <i>Source Code</i> <i>style.css</i>	60
Gambar 4.45 <i>Interface</i> Halaman Pengklasifikasian Akun <i>Buzzer</i>	60
Gambar 4.46 <i>Interface</i> Hasil Pengklasifikasian Akun <i>Buzzer</i>	61
Gambar 4.47 Data <i>Buzzer</i> Untuk Uji Validasi.....	62
Gambar 4.48 Data <i>Non-Buzzer</i> Untuk Uji Validasi.....	63
Gambar 4.49 Contoh Hasil Uji Validasi.....	63