

**PEMBUATAN MODEL KLASIFIKASI DAN VISUALISASI
PADA TWEET AKUN TWITTER SUARA SURABAYA**

SKRIPSI

**Diajukan Untuk Memenuhi Sebagai Persyaratan
Dalam Memperoleh Gelar Sarjana Komputer
Program Studi Sistem Informasi**



Disusun Oleh :

ARILZA FARDLIANA
1635010010

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
SURABAYA**

2023

SKRIPSI

PEMBUATAN MODEL KLASIFIKASI DAN VISUALISASI PADA
TWEET AKUN TWITTER SUARA SURABAYA

Disusun Oleh:

ARILZA FARDLIANA

1635010010

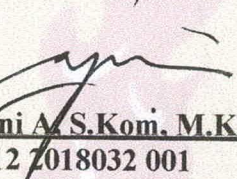
Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi
Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas
Pembangunan Nasional "Veteran" Jawa Timur
Pada Tanggal 29 Mei 2023

Pembimbing:

1.



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

2.

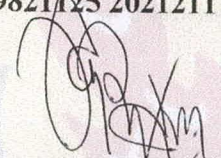

Amalia Anjani A. S.Kom. M.Kom.
NIP. 19920812 7018032 001

Tim Penguji:

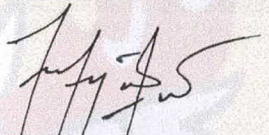
1.


Prisa Marga K. S.Kom. M.Cs.
NIP. 19821125 2021211 003

2.


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

3.


Seftin Fitri Ana Wati, S.Kom. M.Kom.
NPT. 212199 10 320267

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


Dr. Ir. Noxirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PENGESAHAN

**PEMBUATAN MODEL KLASIFIKASI DAN VISUALISASI PADA TWEET
AKUN TWITTER SUARA SURABAYA**

Disusun Oleh:


ARILZA FARDLIANA
1635010010

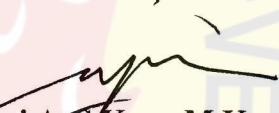
Telah disetujui mengikuti Ujian Negara Lisan Gelombang Mei
Periode 2023 pada Tanggal 29 Mei 2023

Menyetujui,

Dosen Pembimbing 1


Dosen Pembimbing 2


Dr. Eng Agussalim, S.Pd. M.T.
NIP. 19850811 201903 1 005


Amalia Anjani A. S.Kom. M.Kom.
NIP. 19920812/201803 2 001

Mengetahui,

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


Agung Brastama Putra, S.Kom. M.Kom
NIP. 19851124 2021211 003



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 : Arilza Fardliana
NPM : 1635010010
Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 29 Mei 2023 dengan judul:

**PEMBUATAN MODEL KLASIFIKASI DAN VISUALISASI PADA TWEET
AKUN TWITTER SUARA SURABAYA**

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

Surabaya, 20 Juli 2023

Dosen penguji yang memeriksa revisi:

1. **Prisa Marga K. S.Kom, M.Cs.**
NIP. 19821125 2021211 003

{  }

2. **Rizka Hadiwiyanti, S.Kom, M.Kom.**
NIP. 19860727 201803 2 001

{  }

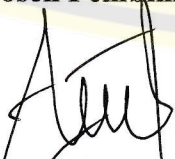
3. **Seftin Fitri Ana Wati, S.Kom, M.Kom.**
NIP. 212199 10 320267

{  }

Mengetahui,

Dosen Pembimbing 1

Dosen Pembimbing 2


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


Amalia Anjani A, S.Kom, M.Kom.
NIP. 19920812 201803 2 001



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

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Arilza Fardhiana

NPM : 1635010010

Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**PEMBUATAN MODEL KLASIFIKASI DAN VISUALISASI PADA TWEET
AKUN TWITTER SUARA SURABAYA**

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, 20 Juli 2023

Hormat Saya,



1FAKX559438540
Arilza Fardhiana
NPM. 1635010010

Judul : **PEMBUATAN MODEL KLASIFIKASI DAN VISUALISASI PADA TWEET AKUN TWITTER SUARA SURABAYA**

Pembimbing 1 : **Dr.Eng Agussalim, S.Pd., M.T.**

Pembimbing 2 : **Amalia Anjani Arifiyanti, S.Kom, M.Kom.**

ABSTRAK

Sosial media berperan sebagai platform yang sangat efektif sebagai media penyebaran informasi, sehingga banyak digunakan oleh platform berita online salah satunya Suara Surabaya untuk memberikan informasi terkini, khususnya kejadian kemacetan, kecelakaan, dan tindak kriminal. Namun, dengan banyaknya kejadian yang terjadi di hari yang sama menjadikan informasi yang diberikan tidak tergambar dengan jelas, informasi yang disampaikan tidak terstruktur dan tertumpuk-tumpuk. Saat ini telah banyak metode yang digunakan untuk membuat model klasifikasi salah satunya Naïve Bayes Classifier. Pada penelitian ini kami memanfaatkan metode tersebut untuk pembuatan model klasifikasi tweet pada akun Twitter Suara Surabaya. Data tweet diklasifikasikan menjadi tiga kategori yaitu, kemacetan, kecelakaan, dan tindak kriminal. Tahapan membangun model klasifikasi ini adalah identifikasi masalah, studi literatur, pengumpulan data, ekstraksi lokasi menggunakan NER, pra-proses, pembuatan model, evaluasi model, dan visualisasi data. Hasil pembuatan model terbaik didapatkan dengan menggunakan Multinomial Naïve Bayes dengan perolehan F1 score sebesar 0.986 dengan melakukan perbandingan dataset 80:20 menggunakan metode hold out. Data tweet divisualkan dalam bentuk peta persebaran dan grafik dimana visualisasi data pada setiap kelas kejadian disajikan dalam bentuk peta persebaran sesuai dengan titik lokasi yang telah diketahui, dan juga dalam bentuk diagram untuk mengetahui jumlah kejadian pada setiap kategori.

Kata Kunci : Named Entity Recognizer, Multinomial Naïve Naves, Data Mining, Klasifikasi, Twitter

KATA PENGANTAR

Alhamdulillah Robbil Alaamiin... Dengan mengucapkan syukur kepada Allah SWT, yang telah melimpahkan berkah, rahmat serta hidayah-Nya, sehingga dapat menyelesaikan skripsi ini

Skripsi ini dibuat untuk memenuhi persyaratan dalam menempuh Program Studi S1/Sistem Informasi di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam penyusunan skripsi ini tidak terlepas dari berbagai pihak yang selalu mendukung dan selalu mendoakan demi kelancaran skripsi ini. Tidak lupa ucapan terima kasih disampaikan kepada :

1. Ibu Widah dan Bapak Rofian Rosad sebagai orangtua yang tidak ada henti-hentinya mendoakan dan memberikan semangat selama pembuatan laporan skripsi ini. Teruntuk adek saya Bagas dan Farel terimakasih karena telah menjadi warna dikehidupan saya
2. Bapak Dr.Eng Agussalim, S.Pd., M.T. Selaku dosen pembimbing 1 yang senantiasa memberikan bimbingan, dorongan dan arahan dalam penyelesaian skripsi.
3. Ibu Amalia Anjani Arifiyanti, S.Kom, M.Kom. Selaku dosen pembimbing 2 yang senantiasa memberikan bimbingan, dorongan dan arahan dalam penyelesaian skripsi.
4. Teman seperjuangan saya, Ana, Kikik, Lisa, Afifah, Rosa yang telah menemani saya dalam proses belajar selama perkuliahan ini, dan tidak

ada hentinya untuk memberi support agar saya segera menyelesaikan skripsi

5. Teruntuk sahabat saya Dina Novera, dan sepupu saya Laisa yang selalu mensupport saya dalam segala hal.
6. Teruntuk teman seperjuangan skripsi the last fighter, revi, efi, firda, linda, nafik, krisna, yeriko, mashita, beni yang banyak membantu saya dalam menyelesaikan administrasi dan perskripsian.
7. Mahasiswa Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur yang tiada hentinya memberikan dukungan dalam penyelesaian skripsi ini.
8. Seluruh teman-teman Sistem Informasi angkatan 2016 yang tetap kompak dan saling mendukung satu sama lain.
9. Dan semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu dalam penyusunan sampai terselesaikannya skripsi.

Tentunya laporan skripsi ini masih banyak kekurangan, namun harapan kedepan semoga skripsi ini dapat ikut menunjang perkembangan ilmu pengetahuan, khususnya dalam bidang ilmu komputer.

Surabaya, Juli 2023

Penulis

DAFTAR ISI

| | |
|--------------------------------|-----|
| ABSTRAK | ii |
| KATA PENGANTAR | iii |
| DAFTAR ISI..... | v |
| DAFTAR GAMBAR | ix |
| DAFTAR TABEL..... | x |
| BAB I | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah | 4 |
| 1.3 Batasan Masalah..... | 4 |
| 1.4 Tujuan..... | 5 |
| 1.5 Manfaat..... | 5 |
| 1.6 Sistematika Penulisan..... | 6 |
| BAB II..... | 8 |
| 2.1 Penelitian Terdahulu..... | 8 |
| 2.2 Dasar Teori..... | 10 |
| 2.2.1 Twitter | 10 |
| 2.2.2 Suara Surabaya..... | 11 |
| 2.2.3 Visualisasi Data..... | 11 |
| 2.2.4 Open Street Map | 12 |
| 2.2.5 Text Mining..... | 12 |

| | | |
|--------------|---|----|
| 2.2.6 | Name Entity Recognition(NER) | 13 |
| 2.2.7 | Klasifikasi Teks..... | 14 |
| 2.2.8 | Pembobotan TF IDF..... | 14 |
| 2.2.9 | Naïve Bayes | 15 |
| 2.2.10 | Evaluasi Model..... | 18 |
| 2.2.11 | Python | 18 |
| 2.2.12 | PHP | 19 |
| BAB III..... | | 20 |
| 3.1 | Identifikasi Masalah | 21 |
| 3.2 | Tinjauan Pustaka | 21 |
| 3.3 | Analisis Kebutuhan Pmbangunan Model | 21 |
| 3.4 | Pengumpulan Data..... | 22 |
| 3.5 | Ekstraksi NER | 22 |
| 3.5.1 | Pelabelan | 25 |
| 3.5.2 | Cleansing..... | 25 |
| 3.5.3 | NER (Named Entity Recognizer)..... | 26 |
| 3.6 | Filter Data..... | 25 |
| 3.7 | Pra proses..... | 25 |
| 3.7.1 | Case Folding..... | 28 |
| 3.7.2 | Cleansing..... | 28 |
| 3.7.3 | Tokenisasi | 29 |

| | | |
|--------------|---|----|
| 3.7.4 | Stopword Removal..... | 29 |
| 3.7.5 | Stemming | 29 |
| 3.7.6 | TF-IDF | 30 |
| 3.8 | Pembagian Data Training dan Testing | 28 |
| 3.9 | Pembuatan Model Klasifikasi..... | 28 |
| 3.10 | Evaluasi Model..... | 30 |
| 3.11 | Visualisasi Data | 31 |
| BAB IV | | 32 |
| 4.1 | Identifikasi Masalah | 32 |
| 4.2 | Analisis Kebutuhan Pembangunan Model | 32 |
| | 4.2.1 Kebutuhan Hardware dan Software | 32 |
| 4.3 | Pengumpulan Data..... | 33 |
| 4.4 | Ekstraksi NER | 35 |
| | 4.4.1 Pelabelan | 35 |
| | 4.4.2 Cleansing..... | 36 |
| | 4.4.3 Named Entity Recognizer | 43 |
| 4.5 | Filter Data..... | 45 |
| 4.6 | Pra-proses | 45 |
| | 4.6.1 Case Folding | 46 |
| | 4.6.2 Cleansing..... | 47 |
| | 4.6.3 Tokenisasi | 51 |

| | | |
|----------------------|----------------------------------|----|
| 4.6.4 | Stopword Removal..... | 52 |
| 4.6.5 | Stemming | 53 |
| 4.6.6 | TF-IDF | 55 |
| 4.7 | Pembagian Data | 55 |
| 4.8 | Pembuatan Model Klasifikasi..... | 55 |
| 4.9 | Evaluasi Model Klasifikasi..... | 57 |
| 4.10 | Visualisasi Data | 59 |
| BAB V..... | | 64 |
| 5.1 | Kesimpulan..... | 64 |
| 5.2 | Saran..... | 64 |
| DAFTAR PUSTAKA | | 66 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 3.1 Diagram Alur Tahapan Penelitian..... | 22 |
| Gambar 3.2 Diagram Alur Tahapan Ekstraksi NER..... | 24 |
| Gambar 3.3 Diagram Alur Tahapan Pra-proses..... | 28 |
| Gambar 4.1 Source Code pengambilan data twitter..... | 33 |
| Gambar 4.2 Source Code Menghilangkan Pukul..... | 37 |
| Gambar 4.3 Source Code Menghilangkan Angka yang diikuti Titik..... | 38 |
| Gambar 4.4 Source Code Menghilangkan URI..... | 39 |
| Gambar 4.5 Source Code Menghilangkan Karakter | 41 |
| Gambar 4.6 Source Code Menghilangkan Spasi Berlebih..... | 42 |
| Gambar 4.7 Source Code Named Entity Recognizer..... | 44 |
| Gambar 4.8 Source Code Case Folding | 46 |
| Gambar 4.9 Source Code Menghilangkan Angka..... | 47 |
| Gambar 4.10 Source Code Menghilangkan Tanda Baca | 49 |
| Gambar 4.11 Source Code Menghilangkan Spasi Berlebih..... | 49 |
| Gambar 4.12 Source Code Tokenisasi | 51 |
| Gambar 4.13 Source Code Stopword Removal | 52 |
| Gambar 4.14 Source Code Stemming..... | 53 |
| Gambar 4.15 Source Code TF-IDF..... | 55 |
| Gambar 4.16 Source Code Pembagian Data | 55 |
| Gambar 4.17 Source Code Multinomial Naïve Bayes | 56 |
| Gambar 4.18 Source Code Bernoulli Naïve Bayes | 56 |
| Gambar 4.19 Source Code Gaussian Naïve Bayes | 57 |
| Gambar 4.20 Visualisasi Data..... | 60 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Penelitian Terdahulu | 8 |
| Tabel 2.2 Confusion Matrix | 20 |
| Tabel 3.1 Entitas spacy | 27 |
| Tabel 4.1 Pengumpulan Data | 34 |
| Tabel 4.2 Contoh Hasil Pelabelan Data | 35 |
| Tabel 4.3 Contoh Hasil Dari Menghilangkan Pukul..... | 37 |
| Tabel 4.4 Contoh Hasil Menghilangkan Angka yang diikuti Titik..... | 39 |
| Tabel 4.5 Contoh Hasil Menghilangkan URL | 40 |
| Tabel 4.6 Contoh Hasil Menghilangkan Karakter | 41 |
| Tabel 4.7 Contoh Hasil Menghilangkan Spasi Berlebih..... | 42 |
| Tabel 4.8 Contoh Hasil Ekstrak Lokasi | 44 |
| Tabel 4.9 Contoh Hasil Case Folding | 46 |
| Tabel 4.10 Contoh Hasil Menghilangkan Angka..... | 47 |
| Tabel 4.11 Contoh Hasil Menghilangkan Tanda Baca | 48 |
| Tabel 4.12 Contoh Hasil Menghilangkan Spasi Berlebih..... | 50 |
| Tabel 4.13 Contoh Hasil Tokenisasi | 51 |
| Tabel 4.14 Contoh Hasil Stopword Removal | 52 |
| Tabel 4.15 Contoh Hasil Stemming | 54 |
| Tabel 4.16 Hasil Skenario..... | 58 |
| Tabel 4.17 Confusion Matrix | 59 |