

**ANALISIS SENTIMEN DETEKSI DEPRESI PADA
PENGGUNA TWITTER DENGAN MENGGUNAKAN
ALGORITMA RANDOM FOREST DAN EXTREME
GRADIENT BOOSTING (XGBOOST)**

SKRIPSI



Oleh :

SUSY RAHMAWATI

18081010048

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2023

**ANALISIS SENTIMEN DETEKSI DEPRESI PADA
PENGGUNA TWITTER DENGAN MENGGUNAKAN
ALGORITMA RANDOM FOREST DAN EXTREME
GRADIENT BOOSTING (XGBOOST)**

SKRIPSI

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam M
enempuh Gelar Sarjana Komputer
Program Studi Informatika



Oleh :

SUSY RAHMAWATI

18081010048

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2023

LEMBAR PENGESAHAN

**Judul : ANALISIS SENTIMEN DETEKSI DEPRESI PADA PENGGUNA TWITTER
DENGAN MENGGUNAKAN ALGORITMA RANDOM FOREST DAN EXTREME
GRADIENT BOOSTING (XGBOOST)**

Oleh : SUSY RAHMAWATI

NPM : 18081010048

Telah Diseminarkan Dalam Ujian Skripsi

Pada: Hari Kamis, Tanggal 26 Januari 2023

Mengetahui

Dosen Pembimbing

Dosen Penguji

1.



1.



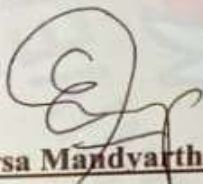
Eva Yulia Puspaningrum, S.Kom., M.Kom.

Fetty Tri Anggraeny, S.Kom., M.Kom.

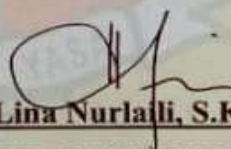
NIP 19890705 20212 002

NIP 19820211 2021212 005

2.



2.



Eka Prakarsa Mandvartha, S.T., M.Kom.

Afina Lina Nurlalli, S.Kom., M.Kom.

NIP 19880525 2018031 001

NIP 19931213 2022032010

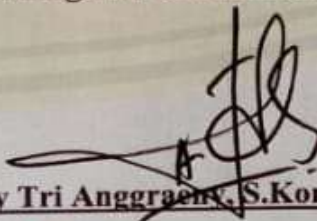
Menyetujui

Dekan

Koordinator

Fakultas Ilmu Komputer

Program Studi Informatika



Dr. Novirina Hendrasarie, ST., MT.

Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP 19681126 199403 2 001

NIP 19820211 2021212 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Informatika UPN "Veteran" Jawa Timur, yang bertandatangan di bawah ini :

Nama : SUSY RAHMAWATI

NPM : 18081010048

Menyatakan bahwa Judul Skripsi/Tugas Akhir yang Saya ajukan dan akan dikerjakan, yang berjudul :

***"ANALISIS SENTIMEN DETEKSI DEPRESI PADA PENGGUNA
TWITTER DENGAN MENGGUNAKAN ALGORITMA RANDOM FOREST
DAN EXTREME GRADIENT BOOSTING (XGBOOST)"***

Bukan merupakan plagiat dari Skripsi/Tugas Akhir/Penelitian orang lain dan juga bukan merupakan produk dan atau software yang saya beli dari pihak 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 gelas di UPN "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 24 Januari 2023

Hormat Saya,



SUSY RAHMAWATI

NPM: 18081010048

**ANALISIS SENTIMEN DETEKSI DEPRESI PADA PENGGUNA
TWITTER DENGAN MENGGUNAKAN ALGORITMA *RANDOM FOREST*
DAN *EXTREME GRADIENT BOOSTING (XGBOOST)***

Nama Mahasiswa : SUSY RAHMAWATI

NPM : 18081010048

Program Studi : Informatika

Dosen Pembimbing : Eva Yulia Puspaningrum, S.Kom., M.Kom.

Eka Prakarsa Mandyartha, S.T., M.Kom.

ABSTRAK

Depresi merupakan kondisi dimana seseorang merasa kecewa atau sedih ketika mengalami kegagalan, kehilangan, atau perubahan dan menjadi tidak wajar ketika mengalami ketidakmampuan untuk beradaptasi. Stres kronis dan peristiwa kehidupan yang penuh tekanan di awal kehidupan sangat kuat prediktor proksimal dari timbulnya depresi. Menurut WHO atau yang biasa kita sebut Organisasi Kesehatan Dunia, depresi adalah masalah paling umum keempat di bumi. Kurang lebih 20% perempuan dan 12% laki-laki mengalami depresi di beberapa titik dalam hidup mereka. Diperkirakan kurang lebih 300 juta orang, atau 4,4% populasi dunia, menderita depresi pada tahun 2015

Berdasarkan dari permasalahan yang ada, tugas akhir ini akan mengimplementasikan algoritma klasifikasi untuk mendeteksi depresi pada pengguna Twitter. Ia menggunakan algoritma *random forest* dan *extreme gradient boosting (XGBoost)*. Kedua algoritma ini diusulkan karena pendekatan dua arah dapat mengekstraksi informasi kontekstual lebih cepat sehingga tidak kehilangan makna dan konteks kalimat. Studi ini diharapkan dapat meningkatkan penyembuhan bagi pengguna media sosial Twitter yang depresi.

Berdasarkan hasil penelitian dapat diketahui bahwa Sistem Deteksi Depresi pada pengguna twitter dengan Bahasa Indonesia menggunakan algoritma Random Forest dan algoritma XGBoost. Dapat berhasil mengklasifikasikan data uji tweet berupa klasifikasi dengan keluaran prediksi depresi dan tidak depresi.

Algoritma *Random Forest* tanpa SMOTE memiliki performa yang lebih baik dengan nilai *accuracy* 87% daripada *Random Forest* menggunakan SMOTE memiliki nilai *accuracy* 81%. Algoritma *XGBoost* SMOTE memiliki performa yang lebih baik dengan nilai *accuracy* 99% daripada algoritma *XGBoost* tanpa SMOTE nilai *accuracy* 95%. Metode Oversampling menggunakan *Synthetic Minority Over Sampling Technique* (SMOTE) tidak dapat diterapkan dialgoritma *Random Forest* sebab beberapa memiliki akurasi cenderung menurun. Namun algoritma *XGBoost* memiliki akurasi yang cenderung meningkat setelah menggunakan SMOTE.

Kata Kunci: Depresi, Twitter, Algoritma *Random Forest*, Algoritma *XGBoost*, *Synthetic Minority Over Sampling Technique*.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat serta taufiq hidayahnya, sehingga penulis dapat menyelesaikan Tugas Akhir/Skripsi dengan judul “ANALISIS SENTIMEN DETEKSI DEPRESI PADA PENGGUNA *TWITTER* DENGAN MENGGUNAKAN ALGORITMA *RANDOM FOREST* DAN *EXTREME GRADIENT BOOSTING (XGBOOST)*” dengan lancar dan tanpa adanya hambatan apapun.

Selesainya laporan Tugas Akhir/Skripsi tidak terlepas dari beberapa pihak yang telah memberikan dukungan sehingga penulis dapat menyelesaikan laporan. Dengan hormat, penulis menyampaikan terima kasih yang sebesar-besarnya kepada beberapa pihak karena tanpa adanya dukungan dan bantuannya penulis tidak dapat menyelesaikan dengan lancar.

Dalam laporan Tugas Akhir/Skripsi penulis menyadari bahwa masih jauh dari sempurna, untuk itu penulis sangat mengharapkan saran, kritik dari semua pihak yang bersifat membangun untuk penyempurnaan laporan ini. Semoga laporan ini bermanfaat bagi kita semua. Aamin.

Surabaya, 24 Januari 2023

Susy Rahmawati

UCAPAN TERIMA KASIH

Segala puji hanyalah milik Allah, sumber segala hikmah dan ilmu pengetahuan karena berkat rahmat dan berkah-Nya penulis dapat menyusun dan menyelesaikan laporan Skripsi ini hingga selesai tepat pada waktunya. Ucapan terima kasih saya sampaikan ke berbagai pihak yang turut membantu memperlancar penyelesaian Skripsi ini, yaitu kepada

1. Kedua Orang Tua, yang selalu memberikan dukungan dan pengertian kepada penulis sekaligus selalu mendoakan penulis dalam kondisi apapun untuk menyelesaikan laporan skripsi ini
2. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor UPN “Veteran” Jawa Timur.
3. Ibu Dr. Novirina Hendrasarie, ST., MT, selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom. M.Kom selaku koordinator program studi Informatika Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
5. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom. selaku pembimbing pertama penulis yang dengan sabar dan ikhlas meluangkan waktu, tenaga serta pikiran untuk membimbing dan mengarahkan penulis dalam penyelesaian proses skripsi ini.
6. Bapak Eka Prakarsa Mandyartha, S.T., M.Kom. selaku pembimbing kedua penulis yang telah memberikan waktu, tenaga serta dengan sabar untuk memberikan arahan, motivasi dan bimbingan kepada penulis sehingga penulis dapat menyelesaikan penyusunan skripsi ini.
7. Staff Dosen dan Tenaga Kependidikan Informatika UPN “Veteran” Jawa Timur yang telah memberikan berbagai ilmu selama mengikuti perkuliahan sampai akhir penulisan skripsi beserta pihak PIA yang

mempermudah penulis dari menemukan dosen pembimbing hingga proses ujian skripsi.

8. Seluruh teman-teman program studi Informatika angkatan 2018 terutama Nur Syiaul Husna, Miftahul Nuril Silviah dan lainnya yang tidak dapat disebutkan satu persatu yang telah membantu dan menjadi bagian pengukir cerita penulis selama masa perkuliahan.

Semoga Tuhan Yang Maha Esa memberikan balasan yang setimpal kepada semua pihak yang telah memberikan bantuan, bimbingan maupun motivasi yang bermanfaat bagi penulis. Penulis juga menyadari bahwa masih banyak kekurangan pada penulisan Laporan Skripsi ini. Oleh karena itu, saran dan kritik yang membangun akan penulis terima dengan senang hati dan penulis berharap semoga Laporan Skripsi ini dapat memberikan manfaat dalam perkembangan ilmu pengetahuan bagi pembaca.

DAFTAR ISI

| | |
|---|------|
| ABSTRAK..... | v |
| KATA PENGANTAR | vii |
| UCAPAN TERIMA KASIH..... | viii |
| DAFTAR ISI..... | x |
| DAFTAR GAMBAR | xii |
| DAFTAR TABEL..... | xiii |
| BAB I PENDAHULUAN..... | 1 |
| 1.1. Latar Belakang..... | 1 |
| 1.2. Rumusan Masalah | 3 |
| 1.3. Batasan Masalah..... | 3 |
| 1.4. Tujuan | 4 |
| 1.5. Manfaat | 4 |
| BAB II TINJAUAN PUSTAKA..... | 5 |
| 1.1. Penelitian terdahulu..... | 5 |
| 1.2. Dasar Teori..... | 6 |
| 1.2.1. Text Mining | 6 |
| 1.2.2. Analisis sentimen..... | 7 |
| 1.2.3. <i>Valence Aware Dictionary and sEntiment Reasoner (VADER)</i> | 8 |
| 1.2.4. <i>Synthetic Minority Over Sampling Technique (SMOTE)</i> | 9 |
| 1.2.5. Klasifikasi..... | 10 |
| 1.2.6. Evaluasi dan Model Validasi | 14 |
| 1.2.7. Model SEMMA | 15 |
| 1.2.8. Twitter..... | 16 |
| 1.2.9. Depresi..... | 16 |
| BAB III METODOLOGI..... | 19 |
| 3.1. Metode Penelitian..... | 19 |
| 5.1. <i>Sample</i> | 20 |
| 5.2. <i>Explore</i> | 20 |
| 5.3. <i>Modify</i> | 21 |
| 3.4.1. <i>Case Folding</i> | 21 |
| 3.4.2. <i>Cleaning</i> | 22 |
| 3.4.3. Stopword Removal | 22 |

| | | |
|-----------------------------------|--|----|
| 3.4.4. | Stemming..... | 24 |
| 3.4.5. | <i>Valence Aware Dictionary and sEntiment Reasoner (VADER)</i> | 24 |
| 3.4.6. | <i>Synthetic Minority Over Sampling Technique (SMOTE)</i> | 25 |
| 3.4.7. | <i>Split train dan test data</i> | 26 |
| 3.4.8. | Ekstrasi Fitur..... | 26 |
| 3.5. | <i>Model</i> | 27 |
| 3.6. | <i>Assess</i> | 28 |
| BAB IV HASIL DAN PEMBAHASAN | | 29 |
| 4.1. | Lingkungan Penelitian | 29 |
| 4.1.1. | Spesifikasi perangkat keras..... | 29 |
| 4.1.2. | Spesifikasi perangkat Lunak..... | 29 |
| 4.2. | <i>Sample</i> | 30 |
| 4.3. | <i>Explore</i> | 32 |
| 4.4. | <i>Modify</i> | 33 |
| 4.4.1. | <i>Case Folding</i> | 34 |
| 4.4.2. | <i>Cleaning</i> | 34 |
| 4.4.3. | <i>Stopword removal</i> | 35 |
| 4.4.4. | <i>Stemming</i> | 37 |
| 4.4.5. | Pelabelan otomatis menggunakan vader..... | 39 |
| 4.4.6. | <i>Synthetic Minority Over Sampling Technique (SMOTE)</i> | 40 |
| 4.4.7. | <i>Split train dan test data</i> | 41 |
| 4.4.8. | Ekstraksi Fitur..... | 42 |
| 4.5. | <i>Model</i> | 43 |
| 4.6. | <i>Assess</i> | 44 |
| 4.6.1. | <i>K-Fold Cross Validation</i> | 44 |
| 4.6.1.1. | Algoritma Random Forest..... | 44 |
| 4.6.1.2. | Algoritma <i>XGBoost</i> | 46 |
| 4.6.2. | <i>Confussion Matrix</i> | 47 |
| 4.6.2.1. | Algoritma Random Forest..... | 48 |
| 4.6.2.2. | Algoritma <i>XGBoost</i> | 50 |
| BAB V PENUTUP..... | | 53 |
| DAFTAR PUSTAKA | | 54 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2. 1 Ilustrasi cara membuat titik SMOTE | 9 |
| Gambar 2. 2 <i>Bagging</i> | 11 |
| Gambar 2. 3 Boosting | 13 |
| Gambar 2. 4 Metodologi Semma | 15 |
| Gambar 3. 1 Model dari SEMMA | 19 |
| Gambar 3. 2 Alur pengumpulan data | 20 |
| Gambar 3. 3 <i>Flowchart Case Folding</i> | 21 |
| Gambar 3. 4 <i>Flowchart Cleaning</i> | 22 |
| Gambar 3. 5 <i>Flowchart Stopword removal</i> | 23 |
| Gambar 3. 6 <i>Flowchart Stemming</i> | 24 |
| Gambar 3. 7 <i>Flowchart Vader</i> | 25 |
| Gambar 3. 8 <i>Flowchart SMOTE</i> | 26 |
| Gambar 3. 9 <i>Flowchart TF-IDF</i> | 27 |
| Gambar 4. 1 API token Twitter Developer | 31 |
| Gambar 4. 2 Dataset..... | 33 |
| Gambar 4. 3 Daftar kamus stopwords removal dari sastrawi..... | 36 |
| Gambar 4. 4 Distribusi Data Label menggunakan VADER | 39 |
| Gambar 4. 5 Dataset (a) asli, (b) sesudah di SMOTE..... | 41 |
| Gambar 4. 6 hasil dari ekstraksi fitur menggunakan tfidf..... | 43 |
| Gambar 4. 7 <i>Confussion Matrix Random Forest</i> | 49 |
| Gambar 4. 8 <i>Confussion Matrix XGBoost</i> | 51 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 4. 1 Spesifikasi perangkat keras yang digunakan untuk penelitian..... | 29 |
| Tabel 4. 2 Spesifikasi perangkat lunak yang digunakan untuk penelitian | 29 |
| Tabel 4. 4 Fungsi <i>Case Folding</i> | 34 |
| Tabel 4. 5 Fungsi <i>Cleaning</i> | 34 |
| Tabel 4. 6 Fungsi <i>Stopword Removal</i> | 36 |
| Tabel 4. 7 Fungsi <i>Stemming</i> | 37 |
| Tabel 4. 9 Hasil Pelabelan..... | 39 |
| Tabel 4. 11 Perbandingan Dataset..... | 40 |
| Tabel 4. 16 Cross Validation algoritma Random forest Tanpa SMOTE | 45 |
| Tabel 4. 17 Cross Validation algoritma Random forest dengan SMOTE..... | 45 |
| Tabel 4. 18 Cross Validation algoritma XGBoost Tanpa SMOTE..... | 46 |
| Tabel 4. 19 Cross Validation algoritma XGBoost dengan SMOTE | 46 |
| Tabel 4. 20 Hasil pengujian algoritma Random forest | 48 |
| Tabel 4. 21 Hasil pengujian algoritma XGBoost..... | 50 |