

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN WORK
FROM HOME MENGGUNAKAN RANDOM FOREST**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Audric Lysander

00000036711

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN WORK
FROM HOME MENGGUNAKAN RANDOM FOREST**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Audric Lysander

00000036711

UMMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Audric Lysander

Nomor Induk Mahasiswa : 00000036711

Program Studi : Informatika

Skripsi dengan judul:

Analisis Sentimen Masyarakat Terhadap Kebijakan Work From Home Menggunakan Random Forest

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tangerang, 6 Juni 2023



(Audric Lysander)

HALAMAN PENGESAHAN

Skripsi dengan judul

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBLIJAKAN WORK
FROM HOME MENGGUNAKAN RANDOM FOREST**

oleh

Nama : Audric Lysander
NIM : 00000036711
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

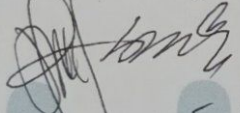
Telah diujikan pada hari Jumat, 16 Juni 2023

Pukul 13.00 s/d 15.00 dan dinyatakan

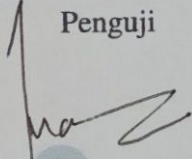
LULUS

Dengan susunan penguji sebagai berikut

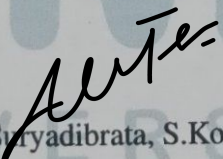
Ketua Sidang


(Adhi Kusnadi, S.T, M.S)
NIDN: 0303037304

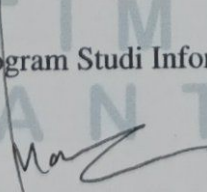
Penguji


(Marlinda Vasty Overbeek, S.Kom.,
M.Kom.)
NIDN: 0818038501

Pembimbing


(Alethea Suryadibrata, S.Kom., M.Eng.)
NIDN: 0322099201

Ketua Program Studi Informatika,


(Marlinda Vasty Overbeek, S.Kom., M.Kom.)
NIDN: 0818038501

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Audric Lysander
NIM : 00000036711
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN WORK
FROM HOME MENGGUNAKAN RANDOM FOREST**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 6 Juni 2023

Yang menyatakan



Audric Lysander

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Halaman Persembahan / Motto

”Believe in yourself and all that you are. Know that there is something inside you that is greater than any obstacle.”

Christian D. Larson



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis Sentimen Masyarakat Terhadap Kebijakan Work From Home Menggunakan Random Forest dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Alethea Suryadibrata, S.Kom., M.Eng., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.
6. Angeline Felicia, Cinthya, Grivia Trifosa Iskandar, dan Joel Alfa Nugraha beserta seluruh teman seperjuangan, yang telah memotivasi dan membantu penulis selama penyusunan laporan skripsi dan keseluruhan aktivitas perkuliahan.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 6 Juni 2023



Audric Lysander

ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN WORK FROM HOME MENGGUNAKAN RANDOM FOREST

Audric Lysander

ABSTRAK

COVID-19 merupakan virus penyebab pandemi yang telah mempengaruhi seluruh dunia. Banyak negara telah melakukan *lockdown* sesuai dengan arahan World Health Organization untuk mencegah terjadinya penularan virus COVID-19. Akibatnya banyak masyarakat yang tidak dapat keluar dari rumah yang menyebabkan sebagian besar bisnis beralih untuk menerapkan konsep *Work From Home*. Penerapan kebijakan ini mempengaruhi kinerja karyawan, salah satunya karena komunikasi yang sulit dilakukan. Salah satu media sosial yang dipakai oleh masyarakat untuk berkomunikasi dan menyebarkan informasi secara cepat untuk menangani masalah tersebut adalah Twitter. Oleh karena itu, penelitian ini dilakukan dengan tujuan untuk melakukan klasifikasi apakah kebijakan *Work From Home* yang diunggah melalui media sosial Twitter termasuk pada sentiment positif atau negatif. Model klasifikasi yang dipakai adalah algoritma *Random Forest* dengan bantuan *Lexicon Inset* untuk memberikan label, *back-translation* untuk memperbanya variasi kata, *TF-IDF Vectorizer* untuk memberikan bobot pada kata, dan menyeimbangkan data dengan berbagai metode. Dari penelitian ini didapatkan klasifikasi terbaik dilakukan menggunakan data yang melalui proses *back-translation*, dengan data yang telah diseimbangkan dengan metode *SMOTETomeks* dengan nilai *accuracy* sebesar 84%, *precision* sebesar 84%, dan *recall* sebesar 84%.

Kata kunci: COVID-19, *Random Forest*, Sentimen, *Twitter*, *Work From Home*

U M M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Analysis of Public Sentiment on Work from Home Policy using Random Forest

Audric Lysander

ABSTRACT

COVID-19 is a virus that causes a pandemic that has affected the whole world. Many countries have implemented a lockdown in accordance with the directions of the World Health Organization to prevent transmission of the COVID-19 virus. As a result, many people cannot leave their homes, which causes most businesses to switch to implementing the Work From Home concept. Implementation of this policy affects employee performance, one of which is because communication. One of the social media used by the public to communicate and disseminate information quickly to deal with this problem is Twitter. Therefore, this research was conducted with the aim of classifying whether the Work From Home policy uploaded via social media Twitter includes positive or negative sentiment. The classification model used is the Random Forest algorithm with the help of Lexicon Inset to provide labels, back-translation to increase word variations, TF-IDF Vectorizer to add weight to words, and balance data with various methods. From this study, the best classification was obtained using data that went through a back-translation process, with data that had been balanced using the SMOTomeks methods with an accuracy of 84%, a precision of 84%, and recall of 84%.

Keywords: *COVID-19, Random Forest, Sentiment, Twitter, Work From Home*



DAFTAR ISI

| | |
|--|------|
| HALAMAN JUDUL | i |
| PERNYATAAN TIDAK MELAKUKAN PLAGIAT | ii |
| HALAMAN PERSETUJUAN | iii |
| HALAMAN PENGESAHAN | iii |
| HALAMAN PERSETUJUAN PUBLIKASI ILMIAH | iv |
| HALAMAN PERSEMBAHAN/MOTO | v |
| KATA PENGANTAR | vi |
| ABSTRAK | vii |
| ABSTRACT | viii |
| DAFTAR ISI | ix |
| DAFTAR GAMBAR | xi |
| DAFTAR TABEL | xii |
| DAFTAR KODE | xiii |
| DAFTAR LAMPIRAN | xiv |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang Masalah | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Permasalahan | 3 |
| 1.4 Tujuan Penelitian | 3 |
| 1.5 Manfaat Penelitian | 3 |
| 1.6 Sistematika Penulisan | 4 |
| BAB 2 LANDASAN TEORI | 5 |
| 2.1 Analisis Sentimen | 5 |
| 2.2 Twitter | 6 |
| 2.3 Text Pre-Processing | 6 |
| 2.4 Decision Tree | 7 |
| 2.5 Random Forest | 9 |
| 2.6 Term Frequency Inverse Document Frequency (TF-IDF) | 10 |
| 2.7 K-Fold Cross-Validation | 11 |
| 2.8 Confusion Matrix | 12 |
| BAB 3 METODOLOGI PENELITIAN | 14 |
| 3.1 Metodologi Penelitian | 14 |
| 3.1.1 Studi Literatur | 14 |
| 3.1.2 Pengumpulan Data | 15 |
| 3.1.3 Perancangan Sistem | 15 |
| 3.1.4 Pengujian dan Evaluasi | 15 |
| 3.1.5 Penulisan Laporan | 15 |
| 3.2 Perancangan Sistem | 15 |
| 3.2.1 Gambaran Umum Sistem | 16 |
| 3.2.2 Pengambilan Data | 17 |
| 3.2.3 Text Pre-Processing | 17 |
| 3.2.4 Data Labelling | 18 |
| 3.2.5 Apply TF-IDF | 19 |
| 3.2.6 Modelling | 20 |
| 3.2.7 Pengujian | 20 |
| 3.2.8 Evaluasi | 21 |
| BAB 4 HASIL DAN DISKUSI | 22 |
| 4.1 Spesifikasi Sistem | 22 |

| | | |
|-------|------------------------------------|----|
| 4.1.1 | Software | 22 |
| 4.1.2 | Hardware | 22 |
| 4.2 | Pengambilan Data | 22 |
| 4.3 | Perancangan Sistem | 24 |
| 4.3.1 | Text Pre-Processing | 24 |
| 4.3.2 | Split Data | 35 |
| 4.3.3 | TF-IDF | 35 |
| 4.3.4 | Handling Imbalanced Data | 36 |
| 4.3.5 | Modelling | 39 |
| 4.3.6 | Pengujian | 40 |
| 4.4 | Uji Coba | 41 |
| 4.5 | Evaluasi | 49 |
| BAB 5 | SIMPULAN DAN SARAN | 50 |
| 5.1 | Simpulan | 50 |
| 5.2 | Saran | 50 |
| | DAFTAR PUSTAKA | 52 |



DAFTAR GAMBAR

| | | |
|-------------|--|----|
| Gambar 2.1 | Struktur Dasar <i>Decision Tree</i> | 8 |
| Gambar 2.2 | Struktur <i>Random Forest</i> | 10 |
| Gambar 3.1 | Diagram alir metodologi penelitian | 14 |
| Gambar 3.2 | Diagram alir gambaran umum penelitian | 17 |
| Gambar 3.3 | Diagram alir <i>Text Pre-Processing</i> | 18 |
| Gambar 3.4 | Diagram alir <i>Data Labelling</i> | 19 |
| Gambar 3.5 | Diagram alir <i>apply TF-IDF</i> | 19 |
| Gambar 3.6 | Diagram alir <i>Modelling</i> | 20 |
| Gambar 3.7 | Diagram alir Pengujian | 20 |
| Gambar 4.1 | Hasil pengambilan data tweets menggunakan <i>library</i> <i>snsrape</i> | 23 |
| Gambar 4.2 | Hasil menghilangkan potongan teks | 25 |
| Gambar 4.3 | Hasil mengubah slang menjadi bahasa resmi | 26 |
| Gambar 4.4 | Hasil <i>back-translation</i> | 27 |
| Gambar 4.5 | Hasil <i>tokenization</i> | 28 |
| Gambar 4.6 | Hasil <i>stemming</i> | 29 |
| Gambar 4.7 | Hasil <i>data labelling</i> | 31 |
| Gambar 4.8 | Visualisasi persebaran label | 32 |
| Gambar 4.9 | Visualisasi <i>Bi-Gram</i> | 33 |
| Gambar 4.10 | Visualisasi <i>Tri-Gram</i> | 33 |
| Gambar 4.11 | Visualisasi <i>Word Cloud</i> | 34 |
| Gambar 4.12 | Hasil <i>encoding</i> | 35 |



DAFTAR TABEL

| | | |
|-----------|--|----|
| Tabel 2.1 | <i>Confusion Matrix</i> | 12 |
| Tabel 4.1 | Hasil <i>handling imbalanced data</i> menggunakan metode <i>hybrid</i> | 37 |
| Tabel 4.2 | Hasil <i>handling imbalanced data</i> menggunakan metode <i>oversampling</i> | 38 |
| Tabel 4.3 | Hasil <i>handling imbalanced data</i> menggunakan metode <i>undersampling</i> | 39 |
| Tabel 4.4 | Nilai <i>Precision, Recall, dan Accuracy</i> menggunakan data awal | 42 |
| Tabel 4.5 | Nilai <i>Precision, Recall, dan Accuracy</i> menggunakan Data <i>Hybrid</i> | 44 |
| Tabel 4.6 | Nilai <i>Precision, Recall, dan Accuracy</i> menggunakan Data <i>Oversampling</i> | 46 |
| Tabel 4.7 | Nilai <i>Precision, Recall, dan Accuracy</i> menggunakan Data <i>Undersampling</i> | 48 |



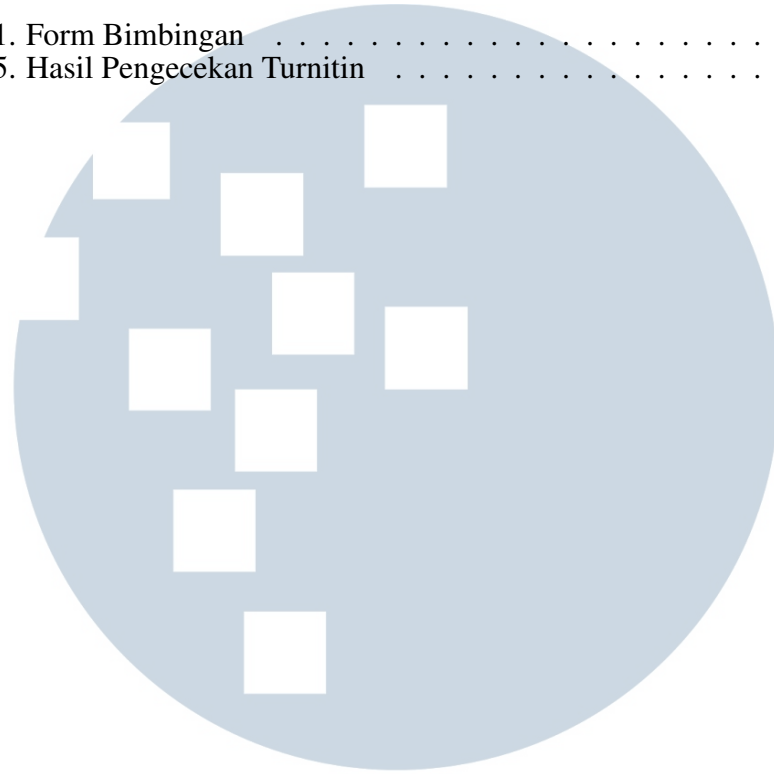
DAFTAR KODE

| | | |
|------|---|----|
| 4.1 | Potongan kode pengambilan data tweets | 23 |
| 4.2 | Potongan kode menghilangkan tweets spam | 24 |
| 4.3 | Potongan kode menghilangkan potongan teks | 25 |
| 4.4 | Potongan kode mengubah kata slang menjadi kata baku | 26 |
| 4.5 | Potongan kode <i>back-translation</i> | 27 |
| 4.6 | Potongan kode <i>tokenization</i> | 28 |
| 4.7 | Potongan kode <i>stemming</i> | 29 |
| 4.8 | Potongan kode <i>data labelling</i> | 30 |
| 4.9 | Potongan kode <i>encoding</i> | 34 |
| 4.10 | Potongan kode <i>split data</i> | 35 |
| 4.11 | Potongan kode TF-IDF | 36 |
| 4.12 | Potongan kode <i>handling imbalanced data</i> menggunakan metode <i>hybrid</i> | 36 |
| 4.13 | Potongan kode <i>handling imbalanced data</i> menggunakan metode <i>oversampling</i> | 37 |
| 4.14 | Potongan kode <i>handling imbalanced data</i> menggunakan metode <i>undersampling</i> | 38 |
| 4.15 | Potongan kode <i>hyperparameter tuning</i> | 39 |
| 4.16 | Potongan kode <i>modelling</i> menggunakan algoritma <i>Random Forest</i> | 40 |
| 4.17 | Potongan kode <i>cross validation</i> | 41 |



DAFTAR LAMPIRAN

| | |
|---|----|
| Lampiran 1. Form Bimbingan | 54 |
| Lampiran 5. Hasil Pengecekan Turnitin | 56 |



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA