

MOTTO DAN PERSEMBAHAN

MOTTO

“Tugas kita bukanlah untuk berhasil. Tugas kita adalah untuk mencoba. Karena di dalam mencoba itulah kita menemukan kesempatan untuk berhasil.”

(Buya Hamka)

PERSEMBAHAN

Tugas akhir ini saya persembahkan kepada:

1. Keluarga kecil nan berarti besar untuk saya, Mamah Popon Nuraeni, Bapak Saripudin Hidayatulloh, Dede Wulansari yang tidak pernah berhenti memberikan dukungannya kepada penulis dan menjadi penenang penulis ketika penulis membutuhkan tempat yang tenang selama melaksanakan penelitian ini.
2. Seluruh dosen dan staff di Jurusan Informatika tercinta yang telah memberikan ilmu dan pengalamannya untuk penulis selama berkuliah di S1 Informatika Universitas Siliwangi.
3. Teman-teman seperjuangan satu organisasi di Himpunan Mahasiswa Informatika FT-Unsil, UKM LDK KISI, GenBI Tasikmalaya, Google Developer Student Clubs Indonesia yang telah berbagi ruang bersama penulis dalam berjuang mendapatkan pengalaman berorganisasi.
4. Semua pihak yang telah mendukung penulis dalam menyelesaikan penelitian dan laporan tugas akhir ini.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT., Karena berkat limpahan rahmat dan karunia-Nya penulis mampu menyelesaikan Tugas Akhir yang berjudul *“Implementasi Ensemble Machine Learning Classifier Dan Synthetic Minority Over-Sampling Technique Untuk Analisis Sentimen Terhadap Sustainable Development Goals Di Indonesia”* yang ditujukan sebagai salah satu syarat untuk memperoleh gelas Sarjana Komputer dari Program Studi Informatika Fakultas Teknik Universitas Siliwangi Tasikmalaya.

Proses penyusunan laporan Tugas Akhir ini tidak lepas dari bantuan berbagai pihak, baik dari segi moril maupun materil. Karena itulah, perkenankan penulis untuk mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Eng. Ir. H. Aripin selaku Dekan Fakultas Teknik, Universitas Siliwangi Tasikmalaya;
2. Bapak Ir. Rianto, S.T., M.T. selaku ketua Program Studi Informatika, Fakultas Teknik, Universitas Siliwangi Tasikmalaya;
3. Bapak Ir. Acep Irham Gufroni, S.Kom., M.Eng. sebagai Dosen Pembimbing I dan Ibu Irani Hoeronis, S.Si., M.T. berikut Ibu Ir. Rahmi Nur Shofa, S.T., M.T. sebagai Dosen Pembimbing II yang telah memberikan arahan dan bimbingannya selama penulis melaksanakan penelitian ini;
4. Bapak Dr. Aradea S.T., M.T. dan Bapak Husni Mubarak S.TP., M.T. selaku dosen Penguji yang telah menguji penelitian dari penulis serta memberikan arahan, kritik dan saran yang membangun sehingga peneliti mendapatkan hasil penelitian yang lebih baik.

5. Bapak Dr. Aradea S.T., M.T. selaku Wali Dosen serta seluruh jajaran dosen di Program Studi Informatika yang telah berjasa kepada penulis;
6. Semua pihak yang sudah membantu dalam penyelesaian Tugas Akhir ini.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna. Dengan dasar itu, penulis memohon kritik dan saran yang sifatnya membangun demi perbaikan di masa yang akan datang.

Penulis memohon maaf apabila terdapat banyak kekurangan dan kesalahan dalam pembuatan laporan Tugas Akhir ini. Semoga laporan ini dapat bermanfaat khususnya bagi penulis dan umumnya bagi yang membutuhkan.

Tasikmalaya, Desember 2022

Penulis

DAFTAR ISI

PENGESAHAN	i
PENGESAHAN PENGUJI.....	ii
LEMBAR PERNYATAAN KEASLIAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
MOTTO DAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	I-1
1.1. Latar Belakang.....	I-1
1.2. Rumusan Masalah.....	I-6
1.3. Tujuan Penelitian	I-6
1.4. Batasan Masalah	I-6
1.5. Manfaat Penelitian	I-7
BAB II TINJAUAN PUSTAKA.....	II-1
2.1. Analisis Sentimen	II-1
2.2. Twitter	II-1
2.3. Machine Learning.....	II-1
2.3.1. <i>Naïve Bayes</i>	II-2
2.3.2. <i>Support Vector Machine</i>	II-4

2.3.3. <i>K-Nearest Neighbor</i>	II-6
2.3.4. <i>Ensemble Machine Learning Classifier</i>	II-9
2.4. Imbalanced Dataset	II-11
2.4.1. <i>Synthetic Minority Over-sampling Technique</i>	II-11
2.5. Confussion Matrix	II-12
2.6. Penelitian Terdahulu	II-13
2.7. Penelitian Terdekat	II-18
2.8. Matriks Ruang Lingkup Penelitian	II-19
2.9. Keterbaruan Penelitian	II-20
BAB III METODOLOGI PENELITIAN	III-1
3.1. <i>Foundational Methodology for Data Science</i>	III-1
3.2. Tahapan Penelitian	III-2
BAB IV HASIL DAN PEMBAHASAN	IV-1
4.1. <i>Data Collection</i>	IV-1
4.2. <i>Data Understanding</i>	IV-4
4.3. <i>Data Preparation</i>	IV-5
4.3.1. <i>Cleansing</i>	IV-5
4.3.2. <i>Casefolding</i>	IV-7
4.3.3. Normalisasi	IV-8
4.3.4. <i>Tokenizing</i>	IV-9
4.3.5. <i>Stopword Removal</i>	IV-10
4.3.6. <i>Stemming</i>	IV-11
4.3.7. <i>Manual Labeling</i>	IV-13

4.4. <i>Modeling</i>	IV-14
4.4.1. <i>Ensemble Machine Learning Classifier – Stacking</i>	IV-17
4.4.2. <i>Ensemble Machine Learning Classifier – Voting</i>	IV-19
4.4.3. Rangkuman Hasil Pemodelan	IV-22
4.5. <i>Evaluation</i>	IV-23
4.6. <i>Deployment</i>	IV-26
4.7. <i>Feedback</i>	IV-30
4.7.1. Feedback untuk Keseluruhan Proses Penelitian.....	IV-30
4.7.1. Feedback untuk Keseluruhan SDGs.....	IV-31
BAB V KESIMPULAN	V-1
5.1. Kesimpulan	V-1
5.2. Saran	V-2
DAFTAR PUSTAKA	
LAMPIRAN-LAMPIRAN	

DAFTAR TABEL

Tabel 2.1 Probabilitas Kosakata Positif	II-3
Tabel 2.1 Probabilitas Kosakata Positif	II-3
Tabel 2.3 <i>Contoh dokumen text</i>	II-7
Tabel 2.4 Hasil Pembobotan	II-8
Tabel 2.5 Kategori Imbalanced Dataset Penelitian Terdekat.....	II-11
Tabel 2.6 Confusion Matrix	II-12
Tabel 2.7 Penelitian Terdahulu	II-13
Tabel 2.8 Penelitian Terdekat	II-18
Tabel 3.1 Daftar Model Penelitian.....	III-5
Tabel 4.1 Keyword Pencarian Twit	IV-3
Tabel 4.2 Penjelasan Atribut Dataset	IV-4
Tabel 4.3 Hasil Proses Cleansing.....	IV-6
Tabel 4.4 Hasil Proses Casefolding	IV-7
Tabel 4.5 Hasil Proses Normalisasi	IV-8
Tabel 4.6 Hasil Proses Tokenizing.....	IV-8
Tabel 4.7 Hasil Proses Stopword Removal.....	IV-10
Tabel 4.8 Hasil Stemming.....	IV-11
Tabel 4.9 Manual Labeling	IV-13
Tabel 4.10 Daftar Model Penelitian.....	IV-14
Tabel 4.11 Parameter untuk setiap algoritma tunggal.....	IV-15
Tabel 4.12 Confusion Matrix	IV-23
Tabel 4.13 Evaluasi Model Terbaik.....	IV-25

DAFTAR GAMBAR

Gambar 1.1 Sustainable Development Goals.....	I-1
Gambar 1.2 Prinsip-prinsip Mendasar dari <i>SDGs</i>	I-3
Gambar 1.3 Peringkat Negara Berdasarkan Pengguna Twitter	I-5
Gambar 2.1 SVM berusaha menemukan hyperplane terbaik.....	II-5
Gambar 2.2 Ensemble Machine Learning Classifier – Stacking	II-9
Gambar 2.3 Ensemble Machine Learning Classifier – Voting	II-10
Gambar 3.1 Foundational Methodology for Data Science.....	III-1
Gambar 3.2 Skema Data Preparation	III-4
Gambar 4.1 Proses Data Collection	IV-1
Gambar 4.2 Proses Menyatukan Semua Hasil dari Data Collection.....	IV-2
Gambar 4.3 Subproses Restore <i>SDGs</i> 1-4	IV-2
Gambar 4.4 Subproses <i>SDGs</i> 1	IV-3
Gambar 4.5 Dataset hasil preprocessing	IV-12
Gambar 4.6 Ensemble Machine Learning Classifier – Stacking	IV-14
Gambar 4.7 Ensemble Machine Learning Classifier – Voting	IV-15
Gambar 4.8 Proses Modelling pada RapidMiner.....	IV-16
Gambar 4.9 Hasil Modelling dengan menggunakan Model A	IV-17
Gambar 4.10 Hasil Modelling dengan menggunakan Model B.....	IV-18
Gambar 4.11 Hasil Modelling dengan menggunakan Model C.....	IV-18
Gambar 4.12 Hasil Modelling dengan menggunakan Model D	IV-19
Gambar 4.13 Hasil Modelling dengan menggunakan Model E	IV-20
Gambar 4.14 Hasil Modelling dengan menggunakan Model F	IV-20
Gambar 4.15 Hasil Modelling dengan menggunakan Model G	IV-21

Gambar 4.16 Hasil Modelling dengan menggunakan Model H	IV-22
Gambar 4.17 Rangkuman Akurasi Model	IV-22
Gambar 4.18 Proses Pra-Deployment	IV-26
Gambar 4.19 Proses Deployment.....	IV-27
Gambar 4.20 Pie Chart Sentiment SDGs	IV-28
Gambar 4.21 Grafik Sentiment SDGs.....	IV-28
Gambar 4.22 Proses WordCloud	IV-31
Gambar 4.23 WordCloud SDGs Keseluruhan	IV-31
Gambar 4.24 WordCloud SDGs Berlabel Positif	IV-33
Gambar 4.25 WordCloud SDGs Berlabel Netral.....	IV-33
Gambar 4.26 WordCloud SDGs Berlabel Negatif.....	IV-33