

**ANALISIS SENTIMEN MASYARAKAT TERHADAP
PAGAR LAUT DI MEDIA SOSIAL TWITTER
MENGUNAKAN ALGORITMA XGBOOST, SVM,
DAN RANDOM FOREST DENGAN PENDEKATAN
SINGLE-LABEL DAN MULTI-LABEL**

Tugas Akhir

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat
Sarjana Komputer**



Dibuat Oleh:

MARTINUS BAGAS BILLY VALENTINO PUTRA

210711354

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2025**

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

ANALISIS SENTIMEN MASYARAKAT TERHADAP PAGAR LAUT DI MEDIA SOSIAL X DENGAN
PENDEKATAN SINGLE-LABEL DAN MULTI-LABEL

yang disusun oleh

Martinus Bagas Billy Valentino Putra

210711354

dinyatakan telah memenuhi syarat pada tanggal 28 Juli 2025

		Keterangan
Dosen Pembimbing 1	: Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D.	Telah Menyetujui
Dosen Pembimbing 2	: Aloysius Gonzaga Pradnya Sidhawara, S.T., M.Eng.	Telah Menyetujui
Tim Penguji		
Penguji 1	: Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D.	Telah Menyetujui
Penguji 2	: Prof. Dr. Ir. Alb. Joko Santoso, M.T.	Telah Menyetujui
Penguji 3	: Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D.	Telah Menyetujui

Yogyakarta, 28 Juli 2025

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Dr. Ir. Parama Kartika Dewa SP., S.T., M.T.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Martinus Bagas Billy Valentino Putra
NPM : 210711354
Program Studi : Informatika
Fakultas : Teknologi Industri
Judul Penelitian : Analisis Sentimen Masyarakat terhadap Pagar Laut
di Media Sosial Twitter Menggunakan Algoritma XGBoost, SVM, dan Random
Forest dengan Pendekatan Single-label dan Multi-label

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 24 Juli 2025

Yang menyatakan,

Martinus Bagas Billy Valentino Putra

210711354

HALAMAN PERSEMBAHAN

Karya tugas akhir ini saya persembahkan dengan penuh rasa syukur dan hormat kepada:

- Tuhan Yesus Kristus, sumber kekuatan, kebijaksanaan, dan kasih yang tiada henti, yang senantiasa membimbing dalam setiap proses kehidupan.
- Orang tua tercinta, yang selalu mendoakan, menyemangati, serta memberikan dukungan moral dan materi tanpa mengenal lelah.
- Dosen pembimbing dan seluruh dosen Program Studi Informatika Universitas Atma Jaya Yogyakarta, atas ilmu, arahan, dan dedikasi yang telah membentuk saya hingga titik ini.
- Teman-teman seperjuangan, yang telah menjadi bagian dari perjalanan ini dengan tawa, semangat, dan kebersamaan yang tak ternilai.
- Diriku sendiri, sebagai pengingat bahwa perjuangan, kerja keras, dan ketekunan adalah bekal berharga dalam menghadapi masa depan.

Semoga karya ini menjadi langkah kecil yang berarti untuk ilmu pengetahuan dan perubahan yang lebih baik bagi masyarakat.

KATA PENGANTAR

Puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan pembuatan tugas akhir “Analisis Sentimen Masyarakat terhadap Pagar Laut di Media Sosial Twitter Menggunakan Algoritma XGBoost, SVM, dan Random Forest dengan Pendekatan *Single-label* dan *Multi-label*” ini dengan baik.

Penulisan tugas akhir ini bertujuan untuk memenuhi salah satu syarat untuk mencapai derajat sarjana komputer dari Program Studi Informatika, Fakultas Teknologi Industri di Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan tugas akhir ini penulis telah mendapatkan bantuan, bimbingan, dan dorongan dari banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Yesus Kristus yang selalu membimbing dalam iman-Nya, memberikan berkat-Nya, dan menyertai penulis selalu.
2. Bapak Dr. Ir. Parama Kartika Dewa SP., S.T., M.T. IPU, selaku Dekan Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
3. Bapak Yohanes Sigit Purnomo Wuryo Putro, ST., M.Kom., Ph.D., selaku dosen pembimbing I yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
4. Bapak Aloysius Gonzaga Pradnya S., S.T., M.Eng., selaku dosen pembimbing II yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
5. Orang tua saya tercinta, yang telah menjadi sumber kekuatan dan inspirasi dalam hidup penulis. Terima kasih atas segala perjuangan, kasih sayang, dukungan moral maupun materiil, serta doa yang tak henti-hentinya dipanjatkan demi keberhasilan dan kelancaran studi penulis hingga tahap ini. Tanpa doa dan pengorbanan mereka, tugas akhir ini tidak mungkin terselesaikan dengan baik.
6. Teman-teman seperjuangan, baik dari lingkungan kampus maupun luar kampus, yang telah memberikan semangat, dukungan, dan kebersamaan

selama proses perkuliahan maupun dalam pengerjaan tugas akhir ini. Kebersamaan dan bantuan kalian sangat berarti bagi penulis.

Demikian laporan tugas akhir ini dibuat, dan penulis mengucapkan terima kasih kepada semua pihak. Semoga laporan ini dapat bermanfaat bagi pembaca.

Yogyakarta, 24 Juli 2025

Martinus Bagas Billy Valentino Putra

210711354

DAFTAR ISI

LEMBAR PENGESAHAN	2
PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH.....	3
HALAMAN PERSEMBAHAN	4
KATA PENGANTAR.....	5
DAFTAR ISI.....	7
DAFTAR GAMBAR.....	10
DAFTAR TABEL	13
INTISARI	14
BAB I PENDAHULUAN.....	16
A. Latar Belakang	16
B. Rumusan Masalah	18
C. Batasan Masalah.....	19
D. Tujuan Penelitian.....	19
E. Metode Penelitian.....	21
1. Studi Literatur	22
2. Pengembangan Korpus.....	23
3. Pengembangan dan optimasi model.....	26
4. Evaluasi dan Implementasi	28
F. Sistematika Penulisan	30
BAB II TINJAUAN PUSTAKA.....	34
BAB III LANDASAN TEORI.....	41
BAB IV PEMBANGUNAN DATASET	51
A. Pengumpulan Dataset.....	51

B.	Pre-processing	53
C.	Pelabelan Data	70
BAB V PEMBANGUNAN MODEL		74
A.	Data Preparation	74
1.	Load Data	74
2.	Split Data + SMOTE	78
B.	Feature Extraction	80
1.	<i>Single-label</i>	81
2.	<i>Multi-label</i>	81
C.	Model Configuration	83
1.	Parameter	83
a)	<i>Single-label</i>	83
b)	<i>Multi-label</i>	84
2.	Pipeline	85
3.	Model Algorithm	89
a)	<i>Single-label</i>	89
b)	<i>Multi-label</i>	90
D.	Model Training	91
1.	<i>Single-label</i>	93
a)	Model Training <i>Single-label</i> tanpa SMOTE	93
b)	Model Training <i>Single-label</i> menggunakan SMOTE	94
2.	<i>Multi-label</i>	95
a)	Model Training <i>Multi-label</i>	95
G.	Model Evaluation	96
1.	<i>Single-label</i>	96

2. <i>Multi-label</i>	101
E. Penjelasan Model Terbaik.....	120
1. <i>Single-label</i>	120
2. <i>Multi-label</i>	120
BAB VI IMPLEMENTASI MODEL	122
A. Antarmuka <i>Streamlit</i>	122
1. <i>Single-label</i>	122
2. <i>Multi-label</i>	127
B. Evaluasi Hasil Antarmuka <i>Streamlit</i> (Best Model).....	130
BAB VII KESIMPULAN DAN SARAN.....	135
A. Kesimpulan	135
B. Saran.....	136
DAFTAR PUSTAKA.....	137

DAFTAR GAMBAR

Gambar 1. 1 Gambaran Keseluruhan dari studi penelitian	21
Gambar 1. 2 Operational Framework and Research Objectives	22
Gambar 1. 3 Diagram Alur Pengumpulan dan Anotasi Data.....	26
Gambar 1. 4 Tahapan Ekstraksi Fitur dan Proses Pelatihan Model dalam Klasifikasi Sentimen.....	28
Gambar 1. 6 Confusion Matrix [29].....	29
Gambar 1. 7 <i>Classification Report</i> [30]	30
Gambar 3. 1 Ilustrasi teknik <i>resampling</i> untuk mengatasi ketidakseimbangan kelas <i>oversampling</i> [31]	45
Gambar 3. 2 Ilustrasi perhitungan TF-IDF [32].....	46
Gambar 3. 3 Struktur umum <i>machine learning pipeline</i> menggunakan <i>scikit-learn</i> . [33]	49
Gambar 3. 4 Ilustrasi proses <i>Grid Search</i> [34]	49
Gambar 4. 1 Inisialisasi Autentikasi	51
Gambar 4. 2 Scrapping Tweet.....	52
Gambar 4. 3 Read hasil Scraping Tweet.....	53
Gambar 4. 4 Hasil Scraping Dataset	53
Gambar 4. 5 Kamus Normalisasi Data.....	55
Gambar 4. 6 Normalisasi Data dengan Lokasi dan Pengukuran.....	57
Gambar 4. 7 Fungsi Normalisasi dan <i>preserve</i> istilah khusus dalam Teks.....	58
Gambar 4. 8 Fungsi <i>normalize_abbreviations</i>	58
Gambar 4. 9 Fungsi Stemming.....	60
Gambar 4. 10 Fungsi Cleaning HTML Entities and Extra Whitespace in Text	61
Gambar 4. 11 Fungsi <i>Basic Text Cleaning</i>	62
Gambar 4. 12 Fungsi <i>Split_hashtag</i> dan <i>remove_stopwords</i>	63
Gambar 4. 13 Fungsi <i>clean_text</i>	64
Gambar 4. 14 Fungsi <i>preprocess_batch_texts</i> dan <i>get_text_statistics</i>	65
Gambar 4. 15 <i>Preprocessing process</i>	66
Gambar 4. 16 Kode print hasil	67
Gambar 4. 17 Save Data.....	70

Gambar 5. 2 Implementasi Split Data + SMOTE (optional)	78
Gambar 5. 3 Implementasi TF-IDF <i>Single-label</i>	81
Gambar 5. 4 Implementasi TfidfVectorizer <i>Multi-label</i>	81
Gambar 5. 5 Implementasi <i>Pipeline</i> Model <i>Klasifikasi</i> Teks	85
Gambar 5. 6 Implementasi Grid Search Cross Validation dalam Pipeline	87
Gambar 5. 7 Model Algoritma <i>Single-label</i>	89
Gambar 5. 8 Implementasi <i>Klasifikasi Multi-label</i> dengan Pendekatan Binary Relevance	90
Gambar 5. 9 Ringkasan Hasil Cross-Validation	91
Gambar 5. 10 Training Model <i>Multi-label</i>	95
Gambar 5. 11 Confusion Matrix XGBoost (a), Confusion Matrix SVM (b), Confusion Matrix RandomForest (c) <i>Single-label</i> tanpa SMOTE.....	98
Gambar 5. 12 Confusion Matrix XGBoost (a), Confusion Matrix SVM (b), Confusion Matrix RandomForest (c) <i>Single-label</i> menggunakan SMOTE.....	100
Gambar 5. 13 Hasil <i>Klasifikasi Multi-label</i> dengan XGBoost tanpa MLSMOTE: (a) Environment, (b) Economic, dan (c) Social.....	108
Gambar 5. 14 Hasil <i>Klasifikasi Multi-label</i> dengan SVM tanpa MLSMOTE: (a) Environment, (b) Economic, dan (c) Social.....	110
Gambar 5. 15 Hasil <i>Klasifikasi Multi-label</i> dengan Random Forest tanpa MLSMOTE: (a) Environment, (b) Economic, dan (c) Social.....	112
Gambar 5. 18 Hasil <i>Klasifikasi Multi-label</i> dengan Random Forest menggunakan MLSMOTE: (a) Environment, (b) Economic, dan (c) Social.....	118
Gambar 6. 1 Pengimporan Pustaka dan Pemuatan Model	122
Gambar 6. 2 Pengunggahan Data dan Pratinjau Awal	123
Gambar 6. 3 Proses <i>Preprocessing</i> dan Validasi	123
Gambar 6. 4 Implementasi <i>Preprocessing</i> Teks	124
Gambar 6. 5 Kamus Normalisasi	125
Gambar 6. 6 Visualisasi Data Hasil <i>Preprocessing</i>	126
Gambar 6. 7 Pemetaan Label dan Prediksi Model	127
Gambar 6. 8 Konfigurasi Awal Aplikasi <i>Streamlit</i>	127

Gambar 6. 9 Upload dan Pembacaan Dataset	128
Gambar 6. 10 Load Model dan Validasi Kolom	128
Gambar 6. 11 Transformasi Fitur dan Proses Prediksi	128
Gambar 6. 12 Mapping Label dan Konversi Hasil Prediksi	129
Gambar 6. 13 Penanganan Error dan Notifikasi	129
Gambar 6. 14 Tampilan Hasil Prediksi Sentimen.....	130
Gambar 6. 15 Hasil Prediksi <i>Single-label</i> dari <i>Best Model</i> (SVM)	131
Gambar 6. 16 Tampilan Hasil Prediksi Aspek dan Sentimen.....	132
Gambar 6. 17 Hasil Prediksi <i>Multi-label</i> dari <i>Best Model</i> (SVM).....	133

DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu	39
Tabel 3. 1 Sampel <i>Sentiment Analysis</i> [20].....	41
Tabel 3. 2 Sampel <i>aspect-based sentiment analysis (ABSA)</i> [12].....	42
Tabel 3. 3 Contoh Cleaning Text [20]	43
Tabel 3. 4 Conflusion Matrix untuk Klasifikasi <i>Multi-label</i>	50
Tabel 4. 1 Contoh Sampel Hasil Normalisasi	59
Tabel 4. 2 Contoh Implementasi Stemming.....	60
Tabel 4. 3 Hasil Text Preprocessing	67
Tabel 4. 4 Contoh Hasil Sentimen Umum	70
Tabel 4. 5 Hasil Labeling Manual Sentimen + Aspek	71
Tabel 5. 1 Load hasil Data set.....	75
Tabel 5. 3 Pembagian Dataset.....	79
Tabel 5. 4 Distribusi Label di Train Set.....	79
Tabel 5. 5 Distribusi Label Test Set.....	79
Tabel 5. 6 Distribusi Label <i>Train Set</i> setelah SMOTE (Optional).....	79
Tabel 5. 7 Parameter XGBoost, SVM, dan Random Forest <i>Single-label</i>	83
Tabel 5. 8 Parameter XGBoost, SVM, dan Random Forest <i>Multi-label</i>	84
Tabel 5. 9 Hasil Cross Model Training tanpa SMOTE.....	93
Tabel 5. 10 Hasil Cross Model Training menggunakan SMOTE.....	94
Tabel 5. 11 Rangkuman Matrix <i>Single-label</i>	96
Tabel 5. 12 Perbandingan Nilai <i>F1-Score</i> per Label dan Rata-Rata pada Klasifikasi <i>Multi-label</i> dengan dan tanpa ML-SMOTE.....	103
Tabel 5. 13 Rangkuman Matrix <i>Multi-label</i>	105
Tabel 5. 14 Sample Prediction	106

INTISARI

Analisis Sentimen Masyarakat terhadap Pagar Laut di Media Sosial Twitter Menggunakan Algoritma XGBoost

Intisari

Martinus Bagas Billy Valentino Putra

210711354

Isu pagar laut di pesisir Kabupaten Tangerang menjadi perhatian publik karena dampak lingkungan dan sosial yang ditimbulkannya serta minimnya transparansi dari pihak berwenang. Twitter sebagai media sosial banyak digunakan masyarakat untuk menyuarakan opini terhadap isu ini. Untuk memahami opini tersebut, analisis sentimen dengan pendekatan *machine learning* digunakan. Namun, pendekatan yang umum digunakan masih terbatas pada *single-label classification*, sementara isu pagar laut bersifat multidimensi dan lebih sesuai dianalisis menggunakan pendekatan *multi-label classification*. Penelitian ini bertujuan untuk membandingkan performa algoritma SVM, Random Forest, dan XGBoost dalam klasifikasi sentimen menggunakan pendekatan *single-label* dan *multi-label*, sekaligus mengisi celah penelitian melalui penerapan kedua pendekatan tersebut dengan dan tanpa penggunaan teknik SMOTE.

Penelitian ini dilakukan melalui empat tahap utama: studi literatur, pembangunan korpus data, pengembangan dan pelatihan model, serta evaluasi dan implementasi sistem. Data dikumpulkan dari Twitter dengan kata kunci terkait pagar laut, kemudian diproses melalui *cleaning*, *normalisasi*, *stemming*, dan pelabelan manual terhadap sentimen (positif, negatif, netral) dan aspek (lingkungan, sosial, ekonomi). Dataset dikonversi dalam format *single-label* dan *multi-label*, lalu fitur teks diekstraksi menggunakan TF-IDF. Tiga algoritma SVM, Random Forest, dan XGBoost dilatih dan dievaluasi menggunakan *classification report* dan

confusion matrix, serta diimplementasikan dalam antarmuka web menggunakan *Streamlit*.

Hasil penelitian menunjukkan bahwa pada pendekatan *single-label classification*, algoritma SVM tanpa SMOTE memberikan performa terbaik, sedangkan pada pendekatan *multi-label classification*, SVM tanpa SMOTE juga unggul dalam hal *F1-score*. Penelitian ini membuktikan bahwa pendekatan *multi-label* tanpa oversampling tetap dapat menghasilkan kinerja yang kompetitif, khususnya dalam konteks isu publik yang bersifat kompleks seperti pagar laut. Temuan ini diharapkan dapat mendukung pengambilan keputusan berbasis data dalam kebijakan publik dan analisis opini masyarakat di media sosial.

Kata Kunci: Analisis Sentimen, Pagar Laut, *Multi-label Classification*, *Machine learning*, Twitter

Dosen Pembimbing I :

Yohanes Sigit Purnomo Wuryo Putro, ST., M.Kom., Ph.D.

Dosen Pembimbing II : Aloysius Gonzaga Pradnya S., S.T., M.Eng.

Jadwal Sidang Tugas Akhir : 24 Juli 2025