

KLASIFIKASI KESEHATAN MENTAL BERDASARKAN UNGGAHAN MEDIA SOSIAL MENGGUNAKAN TF-IDF DAN XGBOOST

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Yasmin Nuha Khoirunnisa

NIM: 09021282126044

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2025**

HALAMAN PENGESAHAN SKRIPSI

KLASIFIKASI KESEHATAN MENTAL BERDASARKAN UNGGAHAN MEDIA SOSIAL MENGGUNAKAN TF-IDF DAN XGBOOST

Sebagai salah satu syarat untuk penyelesaian studi di

Program Studi S1 Teknik Informatika

Oleh:

YASMIN NUHA KHOIRUNNISA

09021282126044

Pembimbing 1 : Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

Mengetahui,
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.
NIP 198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Jumat tanggal 26 Juni 2025 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Yasmin Nuha Khoirunnisa

NIM : 09021282126044

Judul : Klasifikasi Kesehatan Mental Berdasarkan Unggahan Media Sosial Menggunakan TF-IDF dan XGBoost

dan dinyatakan **LULUS**.

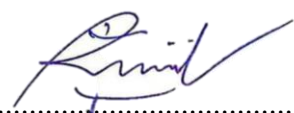
1. Ketua Penguji

Rifkie Primartha, S.T., M.T.
NIP 197706012009121004



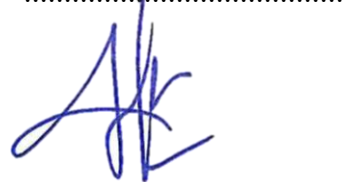
2. Penguji

Mastura Diana Marieska, S.T., M.T.
NIP 198603212018032001



3. Pembimbing

Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003



Mengetahui,

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.
NIP 198004182020121001



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Yasmin Nuha Khoirunnisa

NIM : 09021282126044

Program Studi : Teknik Informatika

Judul Skripsi : Klasifikasi Kesehatan Mental Berdasarkan Unggahan Media Sosial Menggunakan TF-IDF dan XGBoost

Hasil Pengecekan *Software iThenticate/Turnitin*: 8%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka saya bersedia menerima sanksi akademik Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 26 Juni 2025



Yasmin Nuha Khoirunnisa
NIM 09021282126044

MOTTO DAN PERSEMBAHAN

“The only way out is through.”

— Robert Frost

Kupersembahkan karya tulis ini kepada:

- Orang Tua
- Saudara
- Dosen Pembimbing
- Universitas Sriwijaya

ABSTRACT

Mental health is a critical issue in global public health, especially in the digital era where individuals often express their psychological conditions through social media posts. This study aims to develop a multi-class classification model to detect mental health conditions based on social media text using TF-IDF (term frequency–inverse document frequency) for feature extraction and XGBoost (extreme gradient boosting) as the classification algorithm. The dataset used consists of 53,043 English texts categorized into seven mental health classes: Normal, Depression, Suicidal, Anxiety, Stress, Bipolar, and Personality Disorder. The best-performing model, with hyperparameters set to learning rate = 0.2, max depth = 5, and $n_estimators = 1000$, achieved an accuracy of 78%.

Keywords: *XGBoost, TF-IDF, Multi-class Classification*

ABSTRAK

Kesehatan mental merupakan isu penting dalam kesehatan masyarakat global, terlebih di era digital di mana ekspresi kondisi psikologis individu dapat diamati melalui unggahan di media sosial. Penelitian ini bertujuan untuk mengembangkan model klasifikasi multi-kelas untuk mendeteksi kondisi kesehatan mental berdasarkan teks unggahan media sosial menggunakan teknik *term frequency-inverse document frequency* (TF-IDF) sebagai metode ekstraksi fitur dan algoritma *extreme gradient boosting* (XGBoost) sebagai metode klasifikasi. Dataset yang digunakan terdiri dari 53.043 teks berbahasa Inggris yang dikategorikan ke dalam tujuh kelas kesehatan mental: *Normal*, *Depression*, *Suicidal*, *Anxiety*, *Stress*, *Bipolar*, dan *Personality Disorder*. Hasil penelitian menunjukkan bahwa model terbaik dengan konfigurasi *hyperparameter learning rate* = 0,2, *max depth* = 5, dan *n_estimators* = 1000 mencapai akurasi sebesar 78%.

Kata kunci: XGBoost, TF-IDF, Klasifikasi Multikelas

KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT atas berkat, rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Kesehatan Mental Berdasarkan Unggahan Media Sosial Menggunakan TF-IDF dan XGBoost” ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat lulus tingkat Strata-1 Program Studi Teknik Informatika pada Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis juga ingin mengucapkan banyak terima kasih kepada pihak-pihak yang telah memberikan dukungan, motivasi, bimbingan dan bantuan dalam menyelesaikan skripsi ini, yaitu:

1. Orang tua dan kakak yang telah memberikan doa, dukungan, dan restu kepada penulis selama ini.
2. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Alvi Syahrini Utami, M.Kom. selaku Dosen Pembimbing yang telah membimbing dan memberikan arahan kepada penulis dalam proses pengerjaan skripsi.
5. Bapak Rifkie Primartha, M.T. selaku Dosen Pembimbing Akademik.
6. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika yang telah memberikan ilmu kepada penulis selama masa perkuliahan.

7. Seluruh pihak yang telah membantu dan terlibat dalam proses pengerjaan skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kesalahan dan kekurangan yang disebabkan keterbatasan pengetahuan dan pengalaman penulis. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun agar dapat membantu penelitian selanjutnya. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi banyak orang.

Palembang, 26 Juni 2025

Penulis,

Yasmin Nuha Khoirunnisa
NIM 09021282126044

DAFTAR ISI

	Halaman
HALAMAN PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN KOMPREHENSIF	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	I-1
1.1. Pendahuluan	I-1
1.2. Latar Belakang Masalah.....	I-1
1.3. Rumusan Masalah	I-3
1.4. Tujuan Penelitian.....	I-3
1.5. Manfaat Penelitian	I-4
1.6. Batasan Masalah.....	I-4
1.7. Sistematika Penulisan.....	I-4
BAB I. PENDAHULUAN	I-5
BAB II. KAJIAN LITERATUR	I-5
BAB III. METODOLOGI PENELITIAN.....	I-5
BAB IV. PENGEMBANGAN PERANGKAT LUNAK.....	I-5
BAB V. HASIL DAN ANALISIS PENELITIAN.....	I-6
BAB VI. KESIMPULAN DAN SARAN	I-6
1.8. Kesimpulan	I-6
BAB II KAJIAN LITERATUR.....	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Kesehatan Mental.....	II-1
2.2.2 Text Classification	II-3

2.2.3	TF-IDF	II-4
2.2.4	XGBoost.....	II-5
2.2.5	<i>Text Pre-processing</i>	II-8
2.2.6	<i>Imbalanced Data</i>	II-11
2.2.7	<i>Confusion Matrix</i>	II-12
2.2.8	<i>Rational Unified Process</i>	II-14
2.3	Penelitian Lain yang Relevan.....	II-16
2.4	Kesimpulan	II-18
BAB III METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.2.1	Jenis Data	III-1
3.2.2	Sumber Data.....	III-1
3.3	Tahapan Penelitian	III-3
3.3.1	Menentukan Kerangka Kerja Penelitian	III-4
3.3.2	Menentukan Kriteria Pengujian	III-5
3.3.3	Menentukan Format Data Pengujian.....	III-6
3.3.4	Menentukan Alat Bantu Penelitian.....	III-6
3.3.5	Melakukan Pengujian Penelitian.....	III-7
3.3.6	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian.....	III-7
3.4	Metode Pengembangan Perangkat Lunak.....	III-8
3.4.1	Fase Insepsi	III-9
3.4.2	Fase Elaborasi	III-9
3.4.3	Fase Konstruksi.....	III-10
3.4.4	Fase Transisi.....	III-10
3.5	Kesimpulan	III-11
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem	IV-2
4.2.3	Analisis dan Perancangan	IV-3
4.3	Fase Elaborasi	IV-26

4.3.1	Pemodelan Bisnis	IV-26
4.3.2	Kebutuhan Sistem	IV-28
4.3.3	Analisis dan Perancangan	IV-29
4.4	Fase Konstruksi	IV-36
4.4.1	Kebutuhan Sistem	IV-36
4.4.2	Implementasi	IV-37
4.5	Fase Transisi	IV-43
4.5.1	Pemodelan Bisnis	IV-43
4.5.2	Analisis dan Perancangan	IV-43
4.5.3	Implementasi	IV-43
4.1	Kesimpulan	IV-45
BAB V HASIL DAN ANALISIS PENELITIAN		V-1
5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2.2	Data Hasil Konfigurasi	V-4
5.2.3	Perbandingan Metode Penyeimbangan Data	V-8
5.3	Analisis Hasil Penelitian	V-9
5.3.1	Analisis Hasil Konfigurasi <i>Hyperparameter</i>	V-10
5.3.2	Analisis Metrik Evaluasi pada Model Terbaik	V-11
5.4	Kesimpulan	V-14
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1
6.3	Saran	VI-2
DAFTAR PUSTAKA		xv
LAMPIRAN		xvii

DAFTAR TABEL

Tabel II-1. Contoh Proses Lowercasing	II-9
Tabel II-2. Contoh Proses Cleaning	II-9
Tabel II-3. Contoh Proses Tokenizing.....	II-10
Tabel II-4. Contoh Proses Stemming	II-10
Tabel II-5. Tabel Confusion Matrix.....	II-13
Tabel III-1. Distribusi Data	III-2
Tabel III-2. Contoh Data	III-2
Tabel III-3. Format Tabel Pengujian Confusion Matrix.....	III-6
Tabel III-4. Konfigurasi Base Parameters.....	III-7
Tabel III-5. Konfigurasi Hyperparameter	III-8
Tabel III-6. Hasil Konfigurasi Setiap Model.....	III-8
Tabel III-7. Evaluasi Kinerja Model	III-8
Tabel IV-1. Kebutuhan Fungsional	IV-2
Tabel IV-2. Kebutuhan Nonfungsional	IV-2
Tabel IV-3. Tabel Distribusi Data.....	IV-3
Tabel IV-4. Tabel Pembagian Data.....	IV-4
Tabel IV-5. Tabel Distribusi Data Latih Sebelum dan Sesudah ROS	IV-4
Tabel IV-6. Contoh Data Unggahan Media Sosial	IV-5
Tabel IV-7. Hasil Proses Lowercasing	IV-6
Tabel IV-8. Hasil Proses Cleaning	IV-7
Tabel IV-9. Hasil Proses Tokenizing.....	IV-8
Tabel IV-10. Hasil Proses Stemming	IV-9
Tabel IV-11. Hasil Perhitungan TF dan IDF	IV-11
Tabel IV-12. Hasil Pembobotan TF-IDF	IV-13
Tabel IV- 13. Label One-Hot Encoding	IV-16
Tabel IV- 14. Nilai TF-IDF Kolom Pertama	IV-17
Tabel IV- 15. Nilai Gradient dan Hessian	IV-18
Tabel IV-16. Tabel Definisi Aktor	IV-21
Tabel IV-17. Tabel Definisi Use Case	IV-22
Tabel IV-18. Tabel Skenario Use Case Pelatihan Model	IV-22
Tabel IV-19. Tabel Skenario Use Case Klasifikasi Data.....	IV-24
Tabel IV-20. Tabel Skenario Use Case Evaluasi Performa Model.....	IV-25
Tabel IV-21. Implementasi Kelas pada Pelatihan Model	IV-37
Tabel IV-22. Implementasi Kelas pada Perangkat Lunak	IV-38
Tabel IV-23. Rencana Pengujian Use Case	IV-43
Tabel IV-24. Hasil Pengujian Use Case Pelatihan Model.....	IV-44
Tabel IV-25. Hasil Pengujian Use Case Klasifikasi Data	IV-44
Tabel IV-26. Hasil Pengujian Use Case Evaluasi Performa Model	IV-45
Tabel V-1. Konfigurasi Base Parameters	V-1
Tabel V-2. Konfigurasi Hyperparameter	V-4
Tabel V-3. Hasil Konfigurasi Setiap Model	V-5
Tabel V-4. Konfigurasi Hyperparameter Model Terbaik (Model 24)	V-6
Tabel V-5. Evaluasi Kinerja Model Terbaik (Model 24).....	V-8

DAFTAR GAMBAR

Gambar II-1. Strategi Umum Klasifikasi Teks (K. Dalal & A. Zaveri, 2011)	II-4
Gambar II-2. Arsitektur RUP (Perwitasari et al., 2020).....	II-15
Gambar III-1. Tahapan Penelitian	III-3
Gambar III-2. Kerangka Kerja Penelitian	III-4
Gambar IV-1. Use Case Diagram Interaksi Peneliti dengan Sistem	IV-20
Gambar IV-2. Use Case Diagram Interaksi User dengan Sistem.....	IV-21
Gambar IV-3. Rancangan Antarmuka Halaman Klasifikasi Data.....	IV-27
Gambar IV-4. Rancangan Antarmuka Halaman Evaluasi Performa Model.....	IV-28
Gambar IV-5. Activity Diagram Pelatihan Model	IV-30
Gambar IV-6. Activity Diagram Klasifikasi Data.....	IV-31
Gambar IV-7. Activity Diagram Evaluasi Performa Model.....	IV-32
Gambar IV-8. Sequence Diagram Pelatihan Model	IV-33
Gambar IV-9. Sequence Diagram Klasifikasi	IV-34
Gambar IV-10. Sequence Diagram Performa Model	IV-35
Gambar IV-11. Class Diagram Pelatihan Model.....	IV-36
Gambar IV-12. Class Diagram Perangkat Lunak.....	IV-37
Gambar IV-13. Antarmuka Halaman Klasifikasi	IV-40
Gambar IV-14. Antarmuka Halaman Klasifikasi Setelah Proses Klasifikasi...	IV-41
Gambar IV-15. Antarmuka Halaman Performa Model	IV-42
Gambar V-1. Grafik Hasil Pelatihan Model Terbaik (Konfigurasi ke-24).....	V-7
Gambar V-2. Confusion Matrix Model Terbaik (Model 24).....	V-7
Gambar V-3. Perbandingan Akurasi Model	V-9

BAB I

PENDAHULUAN

1.1. Pendahuluan

Dalam bab ini diuraikan tentang latar belakang masalah penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, serta kesimpulan.

1.2. Latar Belakang Masalah

Kesehatan mental telah menjadi perhatian utama dalam bidang kesehatan masyarakat global. Berbagai jenis gangguan kesehatan mental, termasuk depresi, gangguan kecemasan, dan gangguan bipolar memberikan dampak negatif yang signifikan pada kesehatan fisik dan kesejahteraan individu. Kondisi ini semakin diperburuk oleh pandemi COVID-19 (Zhang et al., 2022).

Media sosial tidak hanya menjadi sarana interaksi, tetapi juga merupakan sumber data penting yang dapat digunakan untuk mengidentifikasi kondisi mental penggunanya. Berdasarkan data yang dipublikasikan dalam *Digital 2023: Global Overview Report*, tercatat lebih dari 4,76 miliar orang menggunakan media sosial pada tahun 2023, yang mencakup hampir 60% dari populasi dunia¹⁾. Dengan cakupan yang begitu luas, media sosial menjadi wadah bagi pengguna untuk mengekspresikan kondisi emosional dan psikologis mereka. Unggahan di media sosial dapat dianalisis menggunakan teknik *natural language processing* (NLP)

¹⁾ *Digital 2023: Global Overview Report*, We Are Social dan Meltwater, Januari 2023, halaman 10.

untuk mendeteksi tanda-tanda awal gangguan mental serta mendukung upaya pencegahan dan penanganan secara dini (Zhang et al., 2022).

Klasifikasi unggahan terkait kesehatan mental di media sosial memerlukan pendekatan yang efektif karena adanya keragaman gaya bahasa dan kompleksitas ekspresi emosional yang digunakan oleh pengguna. Menurut (Kowsari et al., 2019), salah satu teknik yang umum digunakan untuk ekstraksi fitur dalam klasifikasi teks adalah *term frequency-inverse document frequency* (TF-IDF). TF-IDF mengekstraksi informasi penting dari teks berdasarkan frekuensi kemunculan kata dalam dokumen dan distribusinya di seluruh kumpulan dokumen. Dengan memberikan bobot lebih pada istilah yang sering muncul dalam dokumen tetapi jarang ditemukan dalam keseluruhan koleksi dokumen, TF-IDF membantu meningkatkan akurasi klasifikasi teks dengan menekankan kata-kata yang paling relevan untuk tema atau kategori tertentu (Sazan et al., 2024).

Extreme gradient boosting (XGBoost) merupakan algoritma yang efektif dalam klasifikasi data. Algoritma ini dirancang untuk menangani data dalam berbagai skala dengan efisiensi tinggi, memanfaatkan komputasi terdistribusi dan paralel untuk mempercepat proses pembelajaran (*learning process*), sehingga hasil prediksi model dapat dioptimalkan (Haumahu et al., 2021). Penelitian oleh Hendrawan et al. (2022) menunjukkan bahwa kombinasi TF-IDF dan XGBoost memiliki akurasi 0,892, lebih tinggi dibandingkan dengan kombinasi TF-IDF dan Naïve Bayes yang memiliki akurasi 0,887.

Berdasarkan paparan di atas, penelitian ini akan mengembangkan sistem klasifikasi kondisi kesehatan mental berdasarkan unggahan di media sosial dengan

memanfaatkan TF-IDF sebagai metode ekstraksi fitur dan XGBoost sebagai algoritma klasifikasinya. Penelitian ini diharapkan dapat berkontribusi dalam pengembangan teknologi klasifikasi berbasis teks guna mendukung analisis isu kesehatan mental di era digital.

1.3. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang dan membangun perangkat lunak untuk mengklasifikasikan kondisi kesehatan mental berdasarkan teks unggahan di media sosial dengan menggunakan metode TF-IDF dan algoritma XGBoost?
2. Bagaimana kinerja model yang dibangun menggunakan TF-IDF dan XGBoost dalam mengklasifikasikan kondisi kesehatan mental?

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan perangkat lunak untuk mengklasifikasikan kondisi kesehatan mental berdasarkan teks unggahan di media sosial menggunakan metode TF-IDF dan algoritma XGBoost.
2. Mengetahui kinerja perangkat lunak yang dibangun menggunakan TF-IDF dan XGBoost dalam mengklasifikasikan kondisi kesehatan mental.

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

1. Menghasilkan perangkat lunak yang dapat dimanfaatkan sebagai alat bantu untuk deteksi awal yang lebih efektif dalam menangani permasalahan kesehatan mental di era digital.
2. Memberikan kontribusi dalam pengembangan metode klasifikasi teks dengan menggunakan TF-IDF dan algoritma XGBoost.
3. Menyediakan hasil penelitian yang dapat dijadikan sebagai referensi untuk penelitian-penelitian selanjutnya di bidang yang relevan.

1.6. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. *Dataset* yang digunakan penelitian ini bersumber dari Kaggle.
2. *Dataset* yang digunakan penelitian ini menggunakan Bahasa Inggris.
3. Klasifikasi dibagi menjadi tujuh kategori kondisi kesehatan mental, yaitu *Normal, Depression, Suicidal, Anxiety, Stress, Bipolar, dan Personality Disorder*.

1.7. Sistematika Penulisan

Sistematika penulisan skripsi ini mengikuti pedoman penulisan proposal skripsi yang ditetapkan oleh Fakultas Ilmu Komputer Universitas Sriwijaya.

Setiap bab akan ditulis secara sistematis sebagai berikut:

BAB I. PENDAHULUAN

Bab ini menguraikan latar belakang, perumusan masalah, tujuan, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Bab ini membahas landasan teori yang digunakan dalam penelitian, meliputi kesehatan mental, klasifikasi teks, metode TF-IDF, algoritma XGBoost, *text pre-processing*, *imbalanced data*, *confusion matrix*, dan *rational unified process*. Bab ini juga mengulas penelitian-penelitian lain yang relevan.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas mengenai metodologi penelitian yang meliputi proses pengumpulan data, tahapan penelitian, serta metode pengembangan perangkat lunak menggunakan *rational unified process* (RUP). Tahapan penelitian mencakup penentuan kerangka kerja penelitian, kriteria pengujian, format data pengujian, alat bantu penelitian, pelaksanaan pengujian, analisis hasil pengujian, dan penyusunan kesimpulan.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas mengenai proses pengembangan perangkat lunak klasifikasi kesehatan mental berdasarkan unggahan media sosial menggunakan TF-IDF dan XGBoost. Pengembangan dilakukan menggunakan pendekatan RUP yang terdiri dari empat fase utama, yaitu insepasi, elaborasi, konstruksi, dan transisi.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini membahas hasil pengujian model klasifikasi yang dikembangkan menggunakan TF-IDF dan XGBoost, serta analisis performa model berdasarkan *hyperparameter* yang digunakan dan metrik evaluasi yang dihasilkan.

BAB VI. KESIMPULAN DAN SARAN

Bab ini memaparkan kesimpulan dari hasil penelitian dan saran untuk pengembangan penelitian selanjutnya.

1.8. Kesimpulan

Berdasarkan paparan di atas, penelitian akan dilakukan dengan metode yang telah ditetapkan, serta dengan sistematika penulisan yang baik dan benar.

DAFTAR PUSTAKA

- Forro, T., Volitaki, E., Malagon-vina, H., Klausberger, T., & Nevian, T. (2022). *Progress in Neurobiology Anxiety-related activity of ventral hippocampal interneurons*. 219(May). <https://doi.org/10.1016/j.pneurobio.2022.102368>
- Ghosal, S., & Jain, A. (2022). Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier. *Procedia Computer Science*, 218, 1631–1639. <https://doi.org/10.1016/j.procs.2023.01.141>
- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. 1–17.
- Haumahu, J. P., Permana, S. D. H., & Yaddarabullah, Y. (2021). Fake news classification for Indonesian news using Extreme Gradient Boosting (XGBoost). *IOP Conference Series: Materials Science and Engineering*, 1098(5), 052081. <https://doi.org/10.1088/1757-899x/1098/5/052081>
- Hendrawan, I. R., Utami, E., & Hartanto, A. D. (2022). Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text Classification. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 143–149. <https://doi.org/10.29408/edumatic.v6i1.5613>
- Işik, M., & Dağ, H. (2020). The impact of text preprocessing on the prediction of review ratings. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(3), 1405–1421. <https://doi.org/10.3906/elk-1907-46>
- Jang, B., Kim, M., Harerimana, G., Kang, S. U., & Kim, J. W. (2020). Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism. *Applied Sciences (Switzerland)*, 10(17). <https://doi.org/10.3390/app10175841>
- K. Dalal, M., & A. Zaveri, M. (2011). Automatic Text Classification: A Technical Review. *International Journal of Computer Applications*, 28(2), 37–40. <https://doi.org/10.5120/3358-4633>
- Kamal, M., Rehman khan, S. U., Hussain, S., Nasir, A., Aslam, K., Tariq, S., & Ullah, M. F. (2020). Predicting mental illness using social media posts and comments. *International Journal of Advanced Computer Science and Applications*, 11(12), 607–613. <https://doi.org/10.14569/IJACSA.2020.0111271>
- Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information (Switzerland)*, 10(4), 1–68. <https://doi.org/10.3390/info10040150>
- Liew, X. Y., Hameed, N., & Clos, J. (2021). An investigation of XGBoost-based algorithm for breast cancer classification. *Machine Learning with Applications*, 6(September), 100154. <https://doi.org/10.1016/j.mlwa.2021.100154>
- Merwin, R., Cuper, P., Lynch, T., & McMahon, K. (2020). Personality disorders. In *Functional Analysis in Clinical Treatment*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-805469-7.00019-X>
- Mutanov, G., Karyukin, V., & Mamykova, Z. (2021). *Multi-Class Sentiment*

- Analysis of Social Media Data with Machine.*
<https://doi.org/10.32604/cmc.2021.017827>
- Perwitasari, R., Afawani, R., & Anjarwani, S. E. (2020). Penerapan Metode Rational Unified Process (RUP) Dalam Pengembangan Sistem Informasi Medical Check Up Pada Citra Medical Centre. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTika)*, 2(1), 76–88. <https://doi.org/10.29303/jtika.v2i1.85>
- Qi, Z. (2020). The Text Classification of Theft Crime Based on TF-IDF and XGBoost Model. *Proceedings of 2020 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2020*, 1241–1246. <https://doi.org/10.1109/ICAICA50127.2020.9182555>
- Sazan, S. A., Miraz, M. H., & Muntasir Rahman, A. B. M. (2024). Enhancing Depressive Post Detection in Bangla: A Comparative Study of TF-IDF, BERT and FastText Embeddings. *Annals of Emerging Technologies in Computing*, 8(3), 34–49. <https://doi.org/10.33166/AETiC.2024.03.003>
- Siregar, P. S., Hatika, R. G., & Hayadi, B. H. (2023). Multiple Choice Question Difficulty Level Classification with Multi Class Confusion Matrix in the Online Question Bank of Education Gallery. *Journal of Applied Data Sciences*, 4(4), 392–406. <https://doi.org/10.47738/jads.v4i4.132>
- Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., & Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00349-y>
- World Health Organization. (2022a). *Mental Disorders*. <https://www.who.int/news-room/fact-sheets/detail/mental-disorders>
- World Health Organization. (2022b). *Mental health*. <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>
- World Health Organization. (2023a). *Depressive disorder (depression)*. <https://www.who.int/news-room/fact-sheets/detail/depression>
- World Health Organization. (2023b). *Stress*. <https://www.who.int/news-room/questions-and-answers/item/stress>
- World Health Organization. (2024). *Bipolar disorder*. <https://www.who.int/news-room/fact-sheets/detail/bipolar-disorder>
- World Health Organization. (2025). *Suicide*. <https://www.emro.who.int/health-topics/suicide/feed/atom.html>
- Wysocki, T. (2020). *Normality : a Two-Faced Concept*.
- Yang, Y., Khorshidi, H. A., & Aickelin, U. (2024). *A review on over-sampling techniques in classification of*. July. <https://doi.org/10.3389/fdgth.2024.1430245>
- Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: a narrative review. *Npj Digital Medicine*, 5(1), 1–13. <https://doi.org/10.1038/s41746-022-00589-7>