

Optimalisasi Klasifikasi Tipe Kepribadian Menggunakan Dataset MBTI dengan Metode SMOTE dan Perbandingan Dua Teknik Boosting

Diajukan Untuk Memenuhi

Persyaratan Guna Meraih Gelar Sarjana Strata 1

Teknik Informatika Universitas Muhammadiyah Malang



Elna Pocut Israeta

202010370311209

Bidang Minat

(Data Science)

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MALANG

LEMBAR PERSETUJUAN

**Optimalisasi Klasifikasi Tipe Kepribadian Menggunakan Dataset MBTI
dengan Metode SMOTE dan Perbandingan Dua Teknik Boosting**

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1

Informatika Universitas Muhammadiyah Malang

Menyetujui,

Malang, 4 Desember 2024

Dosen Pembimbing 1



Christian Sri Kusuma Aditya S.Kom.,

M.Kom

NIP. 180327021991PNS.

LEMBAR PENGESAHAN

**Optimalisasi Klasifikasi Tipe Kepribadian Menggunakan Dataset
MBTI dengan Metode SMOTE dan Perbandingan Dua Teknik
Boosting**

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

Elna Pocut Israeta
202010370311209

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian
pada tanggal 4 Desember 2024

Menyetujui,

Dosen Penguji 1



Ir. Galih Wasis Wicaksono S.kom.
M.Cs.

NIP. 10814100541PNS.

Dosen Penguji 2



Ir. Yufis Azhar S.Kom., M.Kom.
NIP. 10814100544PNS.



Mengetahui,
Ketua Jurusan Informatika



Ir. Galih Wasis Wicaksono S.kom. M.Cs.
NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : Elna Pocut Israeta

NIM : 202010370311209

FAK/JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“Optimalisasi Klasifikasi Tipe Kepribadian Menggunakan Dataset MBTI dengan Metode SMOTE dan Perbandingan Dua Teknik Boosting”** beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing



Christian Sri Kusuma Aditya S.Kom.,
M.Kom

Malang, 4 Desember 2024
Yang Membuat Pernyataan



ABSTRAK

Tipe kepribadian merupakan salah satu indikator penting yang dapat membantu memahami perilaku individu dalam berbagai konteks sosial dan profesional. Myers-Briggs Type Indicator (MBTI) adalah salah satu metode klasifikasi tipe kepribadian yang sering digunakan. Namun, dataset MBTI sering kali mengalami ketidakseimbangan kelas yang signifikan, yang dapat menurunkan performa model klasifikasi. Penelitian ini bertujuan untuk mengoptimalkan klasifikasi tipe kepribadian menggunakan metode SMOTE (Synthetic Minority Over-Sampling Technique) untuk menyeimbangkan data, serta membandingkan dua teknik boosting, yaitu XGBoost dan AdaBoost, guna meningkatkan performa model. Dataset yang digunakan dalam penelitian ini terdiri dari ribuan data kepribadian dengan distribusi yang tidak merata antar tipe. Setelah menerapkan SMOTE, dilakukan optimasi dan evaluasi model menggunakan metrik akurasi mikro dan makro. Hasil penelitian ini menunjukkan bahwa teknik XGBoost lebih unggul dalam menjaga keseimbangan performa dengan makro akurasi mencapai 95.24%, meskipun akurasi mikro sedikit menurun menjadi 61.96%. Sementara itu, AdaBoost menunjukkan penurunan yang lebih signifikan dalam akurasi mikro setelah penerapan SMOTE dengan nilai 44.32%, meskipun makro akurasi masih cukup tinggi pada angka 93.03%. Dengan demikian, XGBoost dipilih sebagai teknik yang lebih optimal dalam menangani ketidakseimbangan data pada klasifikasi tipe kepribadian.

Kata Kunci : MBTI, SMOTE, XGBoost, AdaBoost, Klasifikasi Kepribadian, Machine Learning

KATA PENGANTAR

Puji syukur kehadiran Allah SWT, karena berkat rahmat dan karunia-Nya sehingga penulis bisa menyelesaikan skripsi yang berjudul :

“Optimalisasi Klasifikasi Tipe Kepribadian Menggunakan Dataset MBTI dengan Metode SMOTE dan Perbandingan Dua Teknik Boosting”.

Skripsi ini disusun sebagai salah satu syarat untuk meraih gelar Sarjana Sastra 1 (S1) pada Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang. Penulis menyadari bahwa masih terdapat banyak kekurangan dan jauh dari kata sempurna. Oleh karena itu, kritik dan saran sangat diperlukan guna perbaikan di masa yang akan datang.

Akhir kata, semoga skripsi ini dapat memberikan kontribusi positif, baik dalam pengembangan ilmu pengetahuan di bidang teknologi maupun dalam penelitian seputar klasifikasi kepribadian. Terlebih lagi, semoga skripsi ini bermanfaat bagi para pembaca dan pihak-pihak yang berkepentingan di masa mendatang.

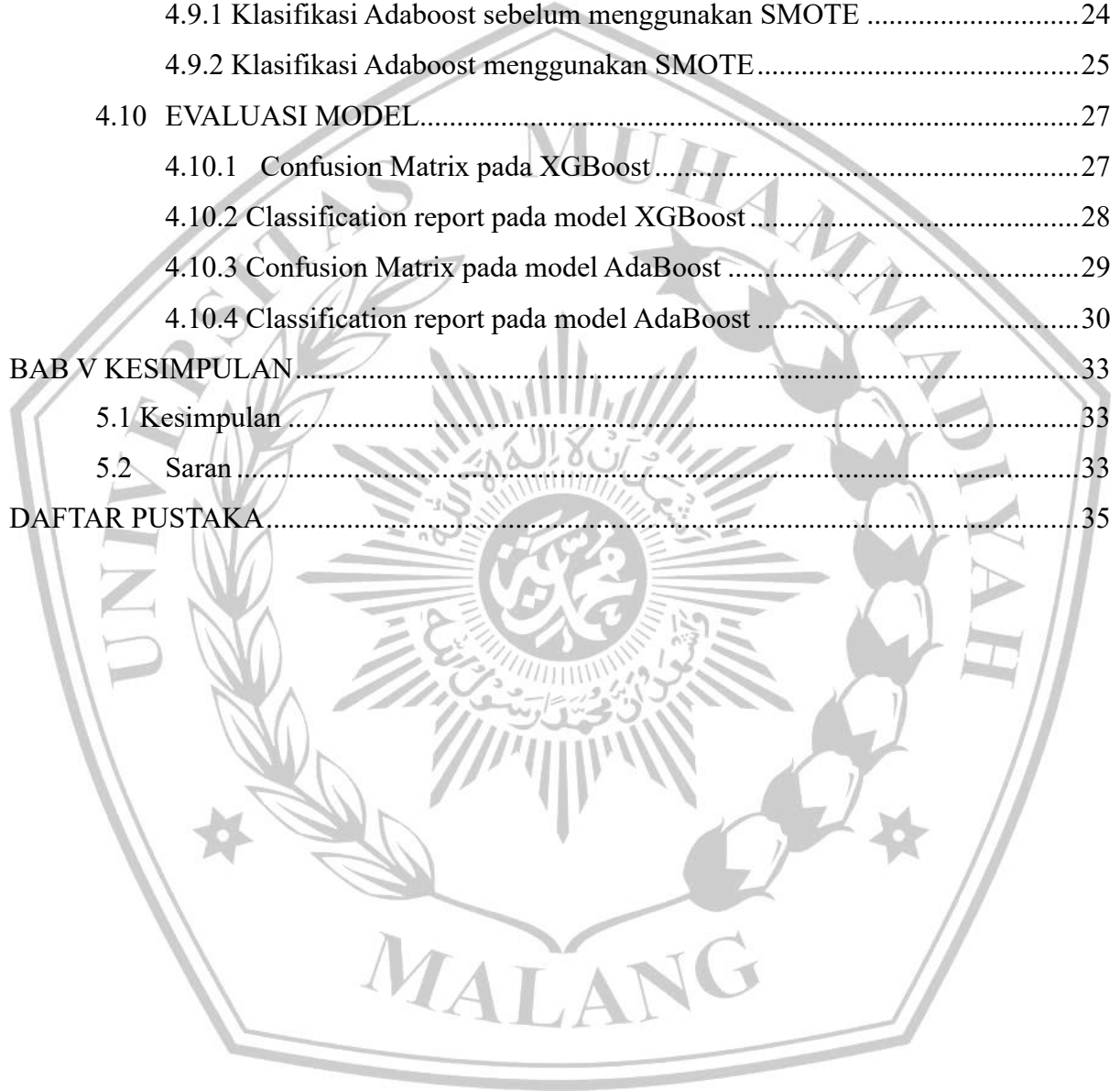
Malang, 02 November 2024

Penulis

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PERNYATAAN	ii
ABSTRAK.....	iii
ABSTRACT.....	iv
LEMBAR PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan	3
1.4 Batasan Masalah	3
1.5 Jadwal Kegiatan.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur.....	5
BAB III METODOLOGI PENELITIAN	8
3.1 Rancangan Penelitian	8
3.2 Myers-Briggs Type Indicator (MBTI).....	9
3.3 Eksplorasi Data.....	10
3.4 <i>Pre – Processing</i> Data	11
3.5 Pembagian Data.....	11
3.6 Penyeimbangan Data dengan SMOTE	11
3.7 Ekstraksi Fitur.....	12
3.8 Klasifikasi Hasil Ekstraksi tiap Model	12
3.9 Evaluasi Model.....	13
BAB IV HASIL DAN PEMBAHASAN	15
4.1 Library dan Package	15
4.2 Load Dataset	16
4.3 Label Encoding.....	17
4.4 Pre-Processing	17
4.5 TF-IDF.....	18

4.6	Splitting Data.....	19
4.7	SMOTE.....	19
4.8	Klasifikasi dengan XGBoost	21
4.8.1	Klasifikasi XGBoost sebelum menggunakan SMOTE	21
4.8.2	Klasifikasi XGBoost menggunakan SMOTE.....	22
4.9	Klasifikasi dengan Adaboost	24
4.9.1	Klasifikasi Adaboost sebelum menggunakan SMOTE	24
4.9.2	Klasifikasi Adaboost menggunakan SMOTE.....	25
4.10	EVALUASI MODEL.....	27
4.10.1	Confusion Matrix pada XGBoost.....	27
4.10.2	Classification report pada model XGBoost.....	28
4.10.3	Confusion Matrix pada model AdaBoost	29
4.10.4	Classification report pada model AdaBoost	30
BAB V KESIMPULAN.....		33
5.1	Kesimpulan	33
5.2	Saran	33
DAFTAR PUSTAKA.....		35



DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	8
Gambar 3. 2 Distribusi antar kelas	10
Gambar 4. 1 Library dan Package	16
Gambar 4. 2 NLTK.....	16
Gambar 4. 3 Load Dataset.....	16
Gambar 4. 4 Encoding.....	17
Gambar 4. 5 Lemmatizer.....	18
Gambar 4. 6 Preprocess Text.....	18
Gambar 4. 7 Cleaned Post.....	18
Gambar 4. 8 Ekstraksi Fitur TF-IDF	19
Gambar 4. 9 Splitting Data.....	19
Gambar 4. 10 Resampling SMOTE	20
Gambar 4. 11 Distribusi antar kelas setelah SMOTE.....	21
Gambar 4. 12 Klasifikasi tanpa menggunakan SMOTE	22
Gambar 4. 13 Hasil XGBoost non SMOTE	22
Gambar 4. 14 Klasifikasi dengan menggunakan SMOTE	23
Gambar 4. 15 Hasil XGBoost With SMOTE	24
Gambar 4. 16 Klasifikasi Adaboost tanpa SMOTE	25
Gambar 4. 17 Hasil AdaBoost Non SMOTE	25
Gambar 4. 18 Adaboost dengan SMOTE.....	26
Gambar 4. 19 Hasil AdaBoost With SMOTE.....	27
Gambar 4. 20 Confusion Matrix XGBoost	28
Gambar 4. 21 Classification Report XGBoost.....	29
Gambar 4. 22 Confusion Matrix AdaBoost.....	30
Gambar 4. 23 Classification Report AdaBoost	31

DAFTAR TABEL

Tabel 2. 1 Rincian Penelitian Terdahulu.....	5
Tabel 3. 1 Isi dataset MBTI.....	9
Tabel 3. 2 Pengujian Klasifikasi.....	13
Tabel 4. 1 Perbandingan Performa Model.....	31



DAFTAR PUSTAKA

- [1] Nitesh V. Chawla (2003) SMOTEBoost: Improving Prediction of the Minority Class in Boosting. Diakses pada 8 Maret 2024 dari https://doi.org/10.1007/978-3-540-39804-2_12
- [2] Ryan, G., Katarina, P., & Suhartono, D. (2023). Mbti personality prediction using machine learning and smote for balancing data based on statement sentences. *Information*, 14(4), 217.
- [3] Amirhosseini, M. H., & Kazemian, H. (2020). Machine learning approach to personality type prediction based on the myers-briggs type indicator®. *Multimodal Technologies and Interaction*, 4(1), 9.
- [4] Rohith Muralidharan (2023) Myers-Briggs Personality Prediction. Diakses pada 8 Maret 2024 dari <https://doi.org/10.54105/ijdm.b1630.053123>
- [5] Ontoum, S., & Chan, J.H. (2022). Personality Type Based on Myers-Briggs Type Indicator with Text Posting Style by using Traditional and Deep Learning. *ArXiv*, abs/2201.08717.
- [6] Maulidah, M. (2023). KLASIFIKASI KEPERIBADIAN MENGGUNAKAN ALGORITMA MACHINE LEARNING. *Jurnal Informatika Dan Tekonologi Komputer (JITEK)*, 3(1), 66-73.
- [7] Pulungan, M. P., Purnomo, A., & Kurniasih, A. (2023). Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(7), 1493-1502.
- [8] Akbar, A. T., Husaini, R., Akbar, B. M., & Saifullah, S. (2020). A proposed method for handling an imbalance data in classification of blood type based on Myers-Briggs type indicator. *Jurnal Teknologi dan Sistem Komputer*, 8(4), 276-283.
- [9] Saputro, M. R., Mahdiyah, U., & Swanjaya, D. (2024). Perbandingan Metode Adaptive Boosting dan Extreme Gradient Boosting Untuk Prediksi Hasil Pertandingan Liga Spanyol. *Nusantara of Engineering (NOE)*, 7(1), 67-73.
- [10] Mendrofa, R. D., Siallagan, M. H., Amalia, J., & Pakpahan, D. P. (2023). Credit Risk Analysis With Extreme Gradient Boosting and Adaptive Boosting Algorithm. *Journal of Information System, Graphics, Hospitality and Technology*, 5(1), 1-7.

- [11]Lai, S. B. S., Shahri, N. H. N. B. M., Mohamad, M. B., Rahman, H. A. B. A., & Rambli, A. B. (2021). Comparing the performance of AdaBoost, XGBoost, and logistic regression for imbalanced data. *Mathematics and Statistics*, 9(3), 379-385.
- [12]Jian Yang (2022) A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm. Diakses pada 20 Maret 2024 dari <https://doi.org/10.3390/info13100475>.





UNIVERSITAS
MUHAMMADIYAH
MALANG



FAKULTAS TEKNIK

INFORMATIKA

informatika.umm.ac.id | informatika@umm.ac.id

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Elna Pocut Israeta
 NIM : 202010370311209
 Judul TA : Optimalisasi Klasifikasi Tipe Kepribadian Menggunakan Dataset MBTI dengan Metode SMOTE dan Perbandingan Dua Teknik Boosting

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	4%
2.	Bab 2 – Daftar Pustaka	25 %	16%
3.	Bab 3 – Analisis dan Perancangan	25 %	4%
4.	Bab 4 – Implementasi dan Pengujian	15 %	0%
5.	Bab 5 – Kesimpulan dan Saran	5 %	3%
6.	Makalah Tugas Akhir	20%	0%

*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)


 (.....)



Kampus I
 Jl. Bandung 1 Malang, Jawa Timur
 P. +62 341 551 253 (Hunting)
 F. +62 341 460 435

Kampus II
 Jl. Bendungan Sutami No 188 Malang, Jawa Timur
 P. +62 341 531 149 (Hunting)
 F. +62 341 582 060

Kampus III
 Jl. Raya Tlogomas No 246 Malang, Jawa Timur
 P. +62 341 464 219 (Hunting)
 F. +62 341 460 435
 E. webmaster@uman.ac.id