

Analisis Sentimen Terhadap Objek Wisata di Kabupaten Ngada

Menggunakan Metode *XGBoost*

Laporan Tugas Akhir

Diajukan Untuk Memenuhi

Persyaratan Guna Meraih Gelar Sarjana

Informatika Universitas Muhammadiyah Malang



Sakinah Aulia Rahmah Putri Adnan

202010370311119

Bidang Minat

Data Science

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MALANG

2024

LEMBAR PERSETUJUAN

Analisis Sentimen Terhadap Objek Wisata di Kabupaten Ngada

Menggunakan Metode *XGBoost*

TUGAS AKHIR

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

Menyetujui,

Malang, 19 Juli 2024

Dosen Pembimbing 1



Dosen Pembimbing 2



Christian Sri Kusuma Aditya, S.Kom., M.Kom — Didih Rizki Chandranegara, S.Kom., M.Kom

NIP. 180327021991PNS.

NIP. 180302101992PNS.

LEMBAR PENGESAHAN

Analisis Sentimen Terhadap Objek Wisata di Kabupaten Ngada Menggunakan Metode XGBoost

TUGAS AKHIR

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

Disusun Oleh :

SAKINAH AULIA RAHMAH PUTRI ADNAN

202010370311119

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian
pada tanggal 19 Juli 2024

Menyetujui,

Dosen Penguji 1



Ir. Yufis Azhar S.Kom., M.Kom.

NIP. 10814100544PNS.

Dosen Penguji 2



Yuda Munarko S.Kom., M.Sc

NIP. 10806110443PNS.

Mengetahui,

Ketua Jurusan Informatika



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

NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : SAKINAH AULIA RAHMAH PUTRI ADNAN

NIM : 202010370311119

FAK./JUR. : TEKNIK/INFORMATIKA

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “**Analisis Sentimen Terhadap Objek Wisata di Kabupaten Ngada Menggunakan Metode XGBoost**” 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, 19 Juli 2024
Yang Membuat Pernyataan



Sakinah Aulia-Rahmah P. Adnan

ABSTRAK

Keindahan alam dan kekayaan budaya Kabupaten Ngada di Nusa Tenggara Timur menjadikannya destinasi wisata yang menarik, termasuk Kampung Adat Bena, Pemandian Air Panas Soa Mengeruda, dan Panorama Manu Lalu. Namun, pengelolaan pariwisata masih menghadapi berbagai tantangan, salah satunya adalah kurangnya optimalisasi pengembangan objek wisata. Penelitian ini bertujuan untuk memahami sentimen wisatawan melalui ulasan di *Google Maps* menggunakan metode *XGBoost*, sebuah algoritma *machine learning* yang terbukti efektif dalam klasifikasi teks. Data dikumpulkan menggunakan teknik *web scraping* menggunakan aplikasi *Octoparse* dan berhasil dihimpun sebanyak 1.160 ulasan untuk wisata Kampung Adat Bena, 584 ulasan untuk Pemandian Air Panas Soa Mengeruda, dan 394 ulasan untuk Panorama Manu Lalu. Data kemudian diberi label secara manual oleh lima penilai. Tahap *preprocessing data* mencakup *cleaning*, *case folding*, *stopword removal*, *tokenization*, dan *stemming*, dengan fitur teks dikonversi menggunakan TF-IDF. Model *XGBoost* dioptimalkan melalui *Grid Search* untuk menemukan kombinasi *hyperparameter* terbaik. Hasil penelitian menunjukkan bahwa algoritma *XGBoost* dengan *hyperparameter* optimal (*n_estimators*: 300, *learning_rate*: 0.3, *min_child_weight*: 1, *max_depth*: 3, *subsample*: 1.0) berhasil mengklasifikasikan sentimen dengan akurasi mencapai 84.6%, presisi 83.9%, *recall* 84.6%, dan *F1-score* 82.5%. Penggunaan *Grid Search* terbukti meningkatkan performa model secara signifikan dibandingkan tanpa optimasi *hyperparameter*. Metode *XGBoost* efektif dalam menganalisis sentimen ulasan pengunjung terhadap objek wisata di Kabupaten Ngada. Hasil analisis ini diharapkan dapat memberikan wawasan kepada pengelola dan pihak terkait dalam meningkatkan kualitas pengelolaan dan pengembangan pariwisata di wilayah Kabupaten Ngada.

Kata kunci: Analisis Sentimen, *Google Maps*, *Grid Search*, Objek Wisata, *XGBoost*

ABSTRACT

The natural beauty and rich culture of Ngada Regency in East Nusa Tenggara make it an attractive tourist destination, including Bena Traditional Village, Soa Mengeruda Hot Springs, and Manu Lalu Panorama. However, tourism management still faces various challenges, one of which is the lack of optimization of tourist attraction development. This research aims to understand tourist sentiment through reviews on Google Maps using the XGBoost method, a machine learning algorithm proven effective in text classification. Data was collected using web scraping technique using Octoparse application and successfully collected as many as 1,160 reviews for Bena Traditional Village tourism, 584 reviews for Soa Mengeruda Hot Spring, and 394 reviews for Manu Lalu Panorama. The data was then manually labeled by five reviewers. The data preprocessing stage includes cleaning, case folding, stopword removal, tokenization, and stemming, with text features converted using TF-IDF. The XGBoost model was optimized through Grid Search to find the best hyperparameter combination. The results showed that the XGBoost algorithm with optimal hyperparameters ($n_estimators: 300$, $learning_rate: 0.3$, $min_child_weight: 1$, $max_depth: 3$, $subsample: 1.0$) successfully classified sentiments with 84.6% accuracy, 83.9% precision, 84.6% recall, and 82.5% F1-score. The use of Grid Search was shown to significantly improve model performance compared to no hyperparameter optimization. The XGBoost method is effective in analyzing the sentiment of visitors' reviews of tourist attractions in Ngada Regency. The results of this analysis are expected to provide insight to managers and related parties in improving the quality of tourism management and development in the Ngada Regency area.

Keywords: *Sentiment Analysis, Google Maps, Grid Search, Tourism Attractions, XGBoost*

LEMBAR PERSEMBAHAN

Segala puji bagi Allah Ta'ala atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir ini sebagai bagian dari perjalanan akademis menuju gelar sarjana di program studi Informatika, Universitas Muhammadiyah Malang. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Allah Ta'ala atas segala rahmat, hidayah, dan karunia-Nya yang tak terbatas, memberikan kekuatan dan ketenangan dalam setiap langkah perjalanan hidup dan studi penulis. Pencapaian ini tidak akan terwujud tanpa izin dan kehendak-Nya.
2. Kedua orang tua, Bapak Adnan Djata dan Ibu Ratna Dewi yang sangat penulis banggakan karena tiada henti melangitkan doa baiknya serta memberikan dukungan dan kasih sayang yang luar biasa sehingga penulis bisa sampai pada titik ini.
3. Saudari kandung penulis Kakak Hutami Rizky Putri Adnan yang selalu membantu dan mendoakan penulis tanpa henti dan juga kepada Adik Nafizah Naylal Husna Putri Adnan yang sangat penulis cintai karena selalu memberikan semangat dan dukungan selama pengerjaan tugas akhir ini.
4. Ibu Siti Sahara Bule, S.Pd sebagai guru sekaligus mentor yang senantiasa memberikan bimbingan, arahan, dan kontribusi yang sangat berharga serta memberikan pencerahan yang sangat berarti dalam perjalanan akademis penulis.
5. Bapak Christian Sri Kusuma Aditya, S.Kom., M.Kom, selaku Dosen Pembimbing I yang selalu sabar dalam membimbing penulis dan telah memberikan ilmu serta dukungannya selama proses pengerjaan tugas akhir ini.
6. Bapak Didih Rizki Chandranegara, S.Kom., M.Kom, selaku Dosen Pembimbing II yang juga telah sabar dalam membimbing dan banyak memberikan saran serta masukan untuk tugas akhir penulis.
7. Dosen wali penulis Bapak Aminudin, S.Kom., M.Cs. atas bimbingan dan dukungan yang telah beliau berikan selama masa studi penulis.
8. Dosen dan staf pengajar di program studi Informatika atas dedikasi, ilmu, dan bimbingannya sejak awal perkuliahan yang telah memberikan kontribusi besar terhadap pencapaian akademik penulis.

9. Teman-teman dan semua pihak yang telah membantu penulis dalam menyelesaikan tugas akhir ini. Dukungan, semangat, dan bantuan kalian sangat berarti bagi penulis.

Malang, 19 Juli 2024

Sakinah Aulia Rahmah Putri Adnan



KATA PENGANTAR

Syukur Alhamdulillah penulis ucapkan kehadiran Allah Ta'ala yang telah melimpahkan rahmat dan karunia-Nya, sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul : **“Analisis Sentimen Terhadap Objek Wisata di Kabupaten Ngada Menggunakan Metode *XGBoost*”**. Shalawat serta salam senantiasa tercurahkan kepada kekasih Allah Ta'ala, panutan seluruh umat yakni baginda Rasulullah *Shallallahu 'Alaihi Wasallam* yang telah memberikan teladan akhlak dan budi pekerti bagi manusia.

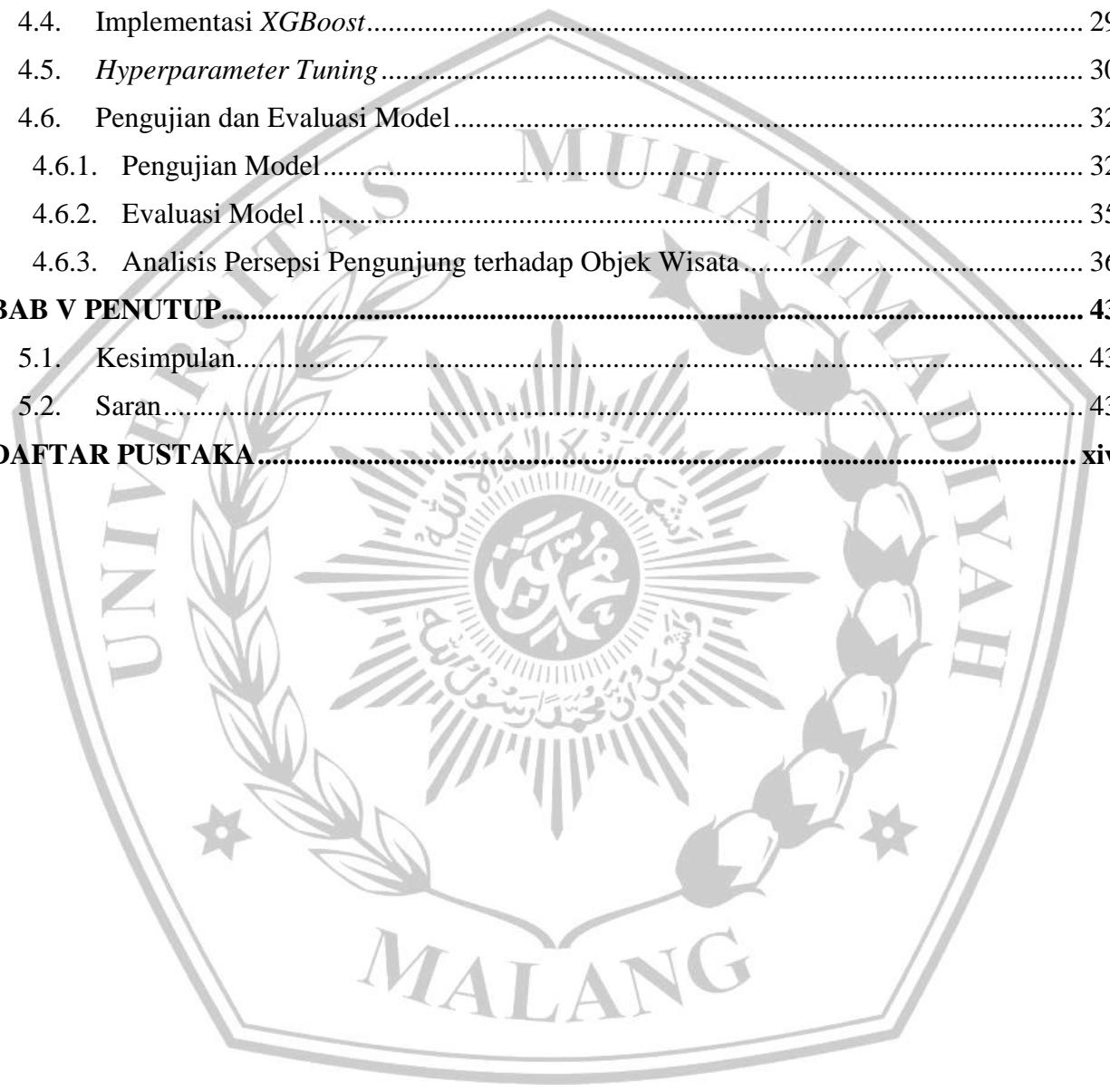
Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap objek wisata yang ada di Kabupaten Ngada dengan menggunakan metode *XGBoost*, sebuah algoritma *machine learning* yang memiliki kinerja tinggi dalam pengolahan data, yang diharapkan dapat memberikan gambaran yang lebih jelas mengenai persepsi dan pengalaman wisatawan. Yang kemudian hasilnya tidak hanya akan digunakan sebagai landasan evaluasi, tetapi juga sebagai pijakan strategis untuk merancang pengembangan sektor pariwisata di Kabupaten Ngada yang lebih adaptif dan berbasis data.

Peneliti menyadari sepenuhnya bahwa dalam penyusunan tugas akhir ini masih terdapat kekurangan dan keterbatasan. Untuk itu peneliti mengharapkan kritik dan saran yang membangun dari berbagai pihak untuk penyempurnaan di masa mendatang. Akhir kata, peneliti berharap semoga tugas akhir ini dapat bermanfaat bagi semua pihak demi perkembangan dan kemajuan akademik.

DAFTAR ISI

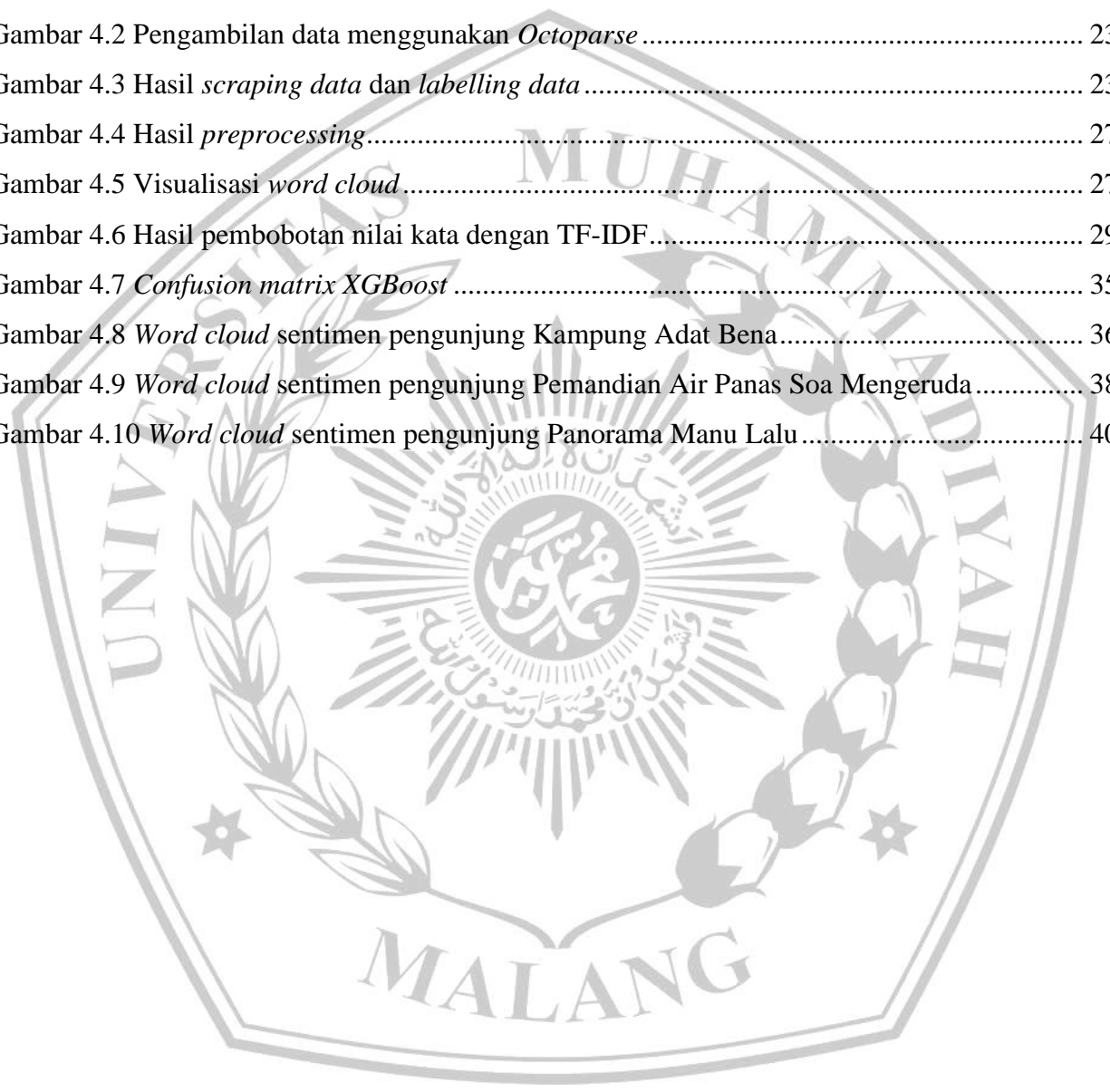
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERSEMBAHAN	vii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA	6
2.1. <i>State Of The Art (SOTA)</i>	6
2.2. <i>Machine Learning</i>	8
2.3. Analisis Sentimen.....	9
2.4. <i>Data Preprocessing</i>	9
2.5. <i>TF-IDF (Term Frequency – Inverse Document Frequency)</i>	10
2.6. <i>XGBoost (eXtreme Gradient Boosting)</i>	12
2.7. <i>Grid Search</i>	13
2.8. <i>Confusion Matrix</i>	14
BAB III METODOLOGI PENELITIAN	16
3.1. Pengumpulan dan Pengolahan Data.....	16
3.2. <i>Preprocessing</i>	17
3.3. <i>TF-IDF (Term Frequency — Inverse Document Frequency)</i>	19
3.4. <i>XGBoost</i>	19
3.5. <i>Hyperparameter Tuning</i>	19

3.6. Pengujian.....	21
BAB IV HASIL DAN PEMBAHASAN	22
4.1. Pengumpulan Dataset.....	22
4.2. <i>Preprocessing</i>	24
4.3. Hasil Pembobotan TF-IDF.....	28
4.4. Implementasi <i>XGBoost</i>	29
4.5. <i>Hyperparameter Tuning</i>	30
4.6. Pengujian dan Evaluasi Model.....	32
4.6.1. Pengujian Model.....	32
4.6.2. Evaluasi Model.....	35
4.6.3. Analisis Persepsi Pengunjung terhadap Objek Wisata.....	36
BAB V PENUTUP.....	43
5.1. Kesimpulan.....	43
5.2. Saran.....	43
DAFTAR PUSTAKA.....	xiv



DAFTAR GAMBAR

Gambar 2.1 Diagram skema <i>XGBoost</i>	12
Gambar 3.1 Rancangan metode penelitian	16
Gambar 4.1 Aplikasi <i>Octoparse</i>	22
Gambar 4.2 Pengambilan data menggunakan <i>Octoparse</i>	23
Gambar 4.3 Hasil <i>scraping data</i> dan <i>labelling data</i>	23
Gambar 4.4 Hasil <i>preprocessing</i>	27
Gambar 4.5 Visualisasi <i>word cloud</i>	27
Gambar 4.6 Hasil pembobotan nilai kata dengan TF-IDF	29
Gambar 4.7 <i>Confusion matrix XGBoost</i>	35
Gambar 4.8 <i>Word cloud</i> sentimen pengunjung Kampung Adat Bena	36
Gambar 4.9 <i>Word cloud</i> sentimen pengunjung Pemandian Air Panas Soa Mengeruda	38
Gambar 4.10 <i>Word cloud</i> sentimen pengunjung Panorama Manu Lalu	40



DAFTAR TABEL

Tabel 2.1 <i>State Of The Art</i> (SOTA)	6
Tabel 2.2 Penjelasan teknik <i>Preprocessing</i>	10
Tabel 2.3 <i>Confusion matrix multiclass</i>	14
Tabel 3.1 Sampel data hasil <i>Scraping</i>	17
Tabel 3.2 Tahap <i>Preprocessing</i>	18
Tabel 4.1 Parameter <i>XGBoost</i>	31
Tabel 4.2 Parameter terbaik <i>XGBoost</i>	32
Tabel 4.3 Perbandingan hasil model <i>XGBoost</i>	34



DAFTAR PUSTAKA

- [1] G. B. Moggi, “Jumlah Kunjungan ke Kampung Adat Bena Meningkat, Mencapai 33.428 orang,” *Suluhdesa.com*, Jul. 07, 2023. Accessed: Nov. 29, 2023. [Online]. Available: <https://www.suluhdesa.com/nasional/5489389364/jumlah-kunjungan-ke-kampung-adat-bena-meningkat-mencapai-33428-orang>
- [2] A. Imron, “Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier,” Yogyakarta, 2020.
- [3] F. U. Haq and H. Rachmat, “Penggunaan Google Review sebagai Penilaian Kepuasan Pengunjung Dalam Pariwisata,” 2020. [Online]. Available: <https://www.researchgate.net/profile/Hardik>
- [4] G. K. Pati and E. Umar, “Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K-Nearest Neighbor,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, pp. 2309–2315, Oct. 2022, doi: 10.30865/mib.v6i4.4635.
- [5] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, “Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, pp. 171–180, Jan. 2022, doi: 10.35969/interkom.v16i4.192.
- [6] D. S. Utami and A. Erfina, “Analisis Sentimen Objek Wisata Bali Di Google Maps Menggunakan Algoritma Naive Bayes,” 2022.
- [7] A. Ikegami and I. D. M. B. A. Darmawan, “Analisis Sentimen dan Pemodelan Topik Ulasan Aplikasi Noice Menggunakan XGBoost dan LDA,” Bali, Nov. 2022.
- [8] I. P. A. P. Widiarta, R. Dwiyanaputra, and A. Aranta, “Analisis Sentimen Masyarakat Terhadap Kebijakan Penerapan PPKM di Media Sosial Twitter dengan Menggunakan Metode Xgboost.”

- [9] I. Cholissodin, Sutrisno, A. A. Soebroto, U. Hasanah, and Y. I. Febiola, *AI, Machine Learning & Deep Learning (Teori & Implementasi) "From Basic Science to High Scientific Solution for Any Problem,"* 1.01. Malang, 2020.
- [10] D. A. Ramadhan and E. B. Setiawan, "Analisis Sentimen Program Acara di SCTV Pada Twitter Menggunakan Metode Naive Bayes dan Vector Machine," *e-Proceeding of Engineering*, vol. 6, no. 2, pp. 9736–9743, Aug. 2019.
- [11] F. S. Pamungkas and I. Kharisudin, "Analisis Sentimen dengan SVM, Naive Bayes dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 628–634, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [12] I. Febriansyah, M. Fikry, and Yusra, "Analisis Sentiment di Twitter terhadap Anies Baswedan sebagai Bakal Calon Presiden 2024 Menggunakan Metode K-Nearest Neighbor," *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 3, pp. 1061–1070, Jul. 2023, doi: 10.33379/gtech.v7i4.2723.
- [13] M. Furqan, S. Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 51–60, Feb. 2022, doi: 10.33633/tc.v21i1.5446.
- [14] F. D. U. Arif, "Perbandingan Kinerja Algoritma Random Forest, XGBoost dan Lightgbm dalam Klasifikasi Emosi Komentar Reddit," Universitas Islam Negeri Syarif Hidayatullah, Jakarta, 2023.
- [15] H. H. Sinaga and S. Agustian, "Perbandingan Metode Decision Tree dan XGBoost untuk Klasifikasi Sentimen Vaksin Covid-19 di Twitter," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 3, pp. 107–114, Dec. 2022, doi: 10.25077/teknosi.v8i3.2022.107-114.
- [16] L. Wilianto, T. H. Pudjiantoro, and F. R. Umbara, "Analisis Sentimen Terhadap Tempat Wisata dari Komentar Pengunjung dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Barat," *Prosiding SNATIF*, pp. 439–448, 2017.

- [17] A. M. AL-Shatnwai and M. Faris, “Predicting Customer Retention using XGBoost and Balancing Methods,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, 2020, doi: 10.14569/IJACSA.2020.0110785.
- [18] A. Samih, A. Ghadi, and A. Fennan, “Enhanced sentiment analysis based on improved word embeddings and XGboost,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 2, pp. 1827–1836, 2023.
- [19] G. A. Shadfila, “Implementasi Metode Extreme Gradient Boosting (XGBOOST) untuk Klasifikasi pada Data Bioinformatika (Studi Kasus : Penyakit Ebola, GSE122692),” 2020, Accessed: May 24, 2024. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/29276>
- [20] P. Liashchynskiy and P. Liashchynskiy, “Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS,” pp. 1–11, Dec. 2019.
- [21] B. Si, Z. Ni, J. Xu, Y. Li, and F. Liu, “Interactive Effects Of Hyperparameter Optimization Techniques and Data Characteristics on The Performance Of Machine Learning Algorithms For Building Energy Metamodeling,” *Case Studies in Thermal Engineering*, vol. 55, pp. 104–124, Mar. 2024, doi: 10.1016/J.CSITE.2024.104124.
- [22] R. Chairunisa, Adiwijaya, and W. Astuti, “Perbandingan CART dan Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 5, pp. 805–812, 2020.
- [23] A. Sumiah and N. Mirantika, “Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan,” *BUFFER INFORMATIKA*, vol. 6, no. 1, pp. 1–10, Apr. 2020.
- [24] Saifullah, “Deteksi Kelayakan Fisik Air Untuk Konsumsi Menggunakan Naive Bayes Classifier,” Universitas Islam Negeri Maulana Malik Ibrahim, Malang, 2019.
- [25] R. Guo, Z. Zhao, T. Wang, G. Liu, J. Zhao, and D. Gao, “Degradation State Recognition of Piston Pump Based on ICEEMDAN and XGBoost,” *Applied Sciences*, vol. 10, no. 18, p. 6593, Sep. 2020, doi: 10.3390/app10186593.

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Sakinah Aulia Rahmah Putri Adnan

NIM : 202010370311119

Judul TA : Analisis Sentimen Terhadap Objek Wisata di Kabupaten Ngada Menggunakan Metode XGBoost

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	6%
2.	Bab 2 – Daftar Pustaka	25 %	8%
3.	Bab 3 – Analisis dan Perancangan	25 %	7%
4.	Bab 4 – Implementasi dan Pengujian	15 %	8%
5.	Bab 5 – Kesimpulan dan Saran	5 %	4%
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)


(.....)