

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN ENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR ISTILAH	xiv
DAFTAR LAMPIRAN.....	xvii
ABSTRAK	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	7
C. Batasan Masalah	7
D. Tujuan Penelitian	8
E. Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA.....	10
A. Landasan Teori.....	10
B. Penelitian Sebelumnya.....	28
BAB III METODE PENELITIAN.....	36
A. Tempat dan Waktu Penelitian.....	36
B. Metode Pengumpulan Data.....	36
C. Alat dan Bahan Penelitian.....	38
D. Konsep Penelitian	40

BAB IV HASIL DAN PEMBAHASAN	50
A. Alur Penelitian	50
B. Hasil	71
BAB V PENUTUP.....	74
A. Kesimpulan	74
B. Saran	75
DAFTAR PUSTAKA	
LAMPIRAN	



DAFTAR TABEL

Tabel 2. 1. Perbedaan utama Klasifikasi dan Regresi	13
Tabel 2. 2. Perbedaan antara teknik Bagging dan Boosting	17
Tabel 2. 3. Definisi Komponen Confusion Matrix	23
Tabel 2. 4. Penelitian Sebelumnya	34
Tabel 3. 1. Atribut Dataset	37
Tabel 4. 1. Hasil Penggabungan Subset Dataset MIB	51
Tabel 4. 2. Fitur Input Setelah Preprocessing	52
Tabel 4. 3. Hasil Pembagian Dataset	53
Tabel 4. 4. Parameter Terbaik Hasil GridSearchCH.....	54
Tabel 4. 5. Hasil Rata-rata 5-Fold Cross Validation.....	56
Tabel 4. 6. Hasil Evaluasi Model pada Data Uji.....	56
Tabel 4. 7. Confusion Matrix Ketiga Model pada Data Uji.....	57
Tabel 4. 8. Perbandingan Training Score dan Testing Score	58
Tabel 4. 9. Perbandingan Performa Terbaik Ketiga Model	63
Tabel 4. 10. Komparasi Hasil Penelitian.....	70

DAFTAR GAMBAR

Gambar 3. 1. Alur Penelitian.....	40
Gambar 4. 1. Grafik ROC AUC.....	63
Gambar 4. 2. Data Akun Sintesis Uji.....	65
Gambar 4. 3. Probabilitas Prediksi Bot Random Forest	66
Gambar 4. 4. Probabilitas Prediksi Bot XGBoost.....	67
Gambar 4. 5. Probabilitas Prediksi Bot LightGBM	68
Gambar 4. 6. Heatmap Hasil Prediksi Ketiga Model.....	69



DAFTAR ISTILAH

1. X (Twitter): Platform media sosial mikroblogging yang menjadi objek utama penelitian, di mana pengguna berinteraksi melalui pesan teks singkat, gambar, dan video.
2. Machine Learning: Cabang dari kecerdasan buatan (AI) yang fokus pada pengembangan algoritma agar komputer dapat belajar dari data dan membuat prediksi tanpa instruksi pemrograman eksplisit.
3. Supervised Learning: Metode pembelajaran mesin di mana model dilatih menggunakan dataset yang sudah memiliki label jawaban (target) yang benar.
4. Random Forest: Algoritma pembelajaran mesin yang menggabungkan banyak pohon keputusan (decision trees) untuk meningkatkan akurasi dan mencegah model terlalu fokus pada data latih (overfitting).
5. XGBoost (eXtreme Gradient Boosting): Algoritma pengembangan dari gradient boosting yang dirancang untuk efisiensi tinggi, kecepatan komputasi, dan performa optimal dalam klasifikasi data.
6. LightGBM (Light Gradient Boosting Machine): Kerangka kerja gradient boosting yang menggunakan teknik pertumbuhan pohon berbasis daun (leaf-wise), sehingga lebih cepat dan menggunakan memori yang lebih sedikit.
7. Fake Account (Akun Palsu): Akun media sosial yang dibuat dengan identitas tidak asli, seringkali dioperasikan oleh Bot (otomatis) untuk menyebarkan informasi menyesatkan atau manipulasi opini.
8. Social Bot: Program otomatis yang mensimulasikan perilaku manusia di media sosial untuk melakukan tugas-tugas tertentu secara masif.
9. Feature Engineering: Proses pemilihan, manipulasi, dan transformasi data mentah menjadi fitur-fitur baru yang lebih informatif agar model pembelajaran mesin dapat bekerja lebih baik.
10. Metadata: Data yang memberikan informasi mengenai karakteristik akun, seperti jumlah pengikut, jumlah postingan, status verifikasi, dan tanggal pembuatan akun.

11. Confusion Matrix: Tabel alat evaluasi untuk mengukur kinerja model klasifikasi dengan membandingkan nilai prediksi model dengan nilai aktual di lapangan.
12. Accuracy: Metrik untuk mengukur sejauh mana model benar dalam mengklasifikasikan akun asli dan palsu secara keseluruhan.
13. Precision: Tingkat keakuratan model dalam mengidentifikasi akun palsu di antara semua akun yang diprediksi sebagai palsu.
14. Recall: Kemampuan model dalam menemukan kembali seluruh akun palsu yang sebenarnya ada di dalam dataset.
15. F1-Score: Nilai rata-rata harmonik antara presisi dan recall yang digunakan untuk menunjukkan keseimbangan performa model.
16. Preprocessing: Tahap awal pengolahan data mentah, seperti pembersihan data dari nilai yang hilang (missing values) atau data duplikat, sebelum dimasukkan ke dalam model.
17. Dataset: Kumpulan data yang telah disusun secara sistematis untuk keperluan eksperimen dan analisis penelitian.
18. SMOTE (Synthetic Minority Over-sampling Technique): Metode penanganan ketidakseimbangan kelas (class imbalance) dengan membuat sampel sintetis baru pada kelas minoritas melalui proses interpolasi antar data yang sudah ada, tanpa membuang data asli.
19. Imbalanced Data: Kondisi di mana distribusi kelas pada dataset tidak seimbang, di mana satu kelas memiliki jumlah sampel yang jauh lebih banyak dibandingkan kelas lainnya sehingga berpotensi menyebabkan bias pada model.
20. GridSearchCV: Metode optimasi hyperparameter yang mencoba seluruh kombinasi parameter yang telah didefinisikan secara menyeluruh (exhaustive search) dan mengevaluasi setiap kombinasi menggunakan Cross Validation.
21. K-Fold Cross Validation: Teknik validasi model dengan membagi data latih menjadi K bagian (fold) yang digunakan secara bergantian sebagai data validasi untuk menghasilkan estimasi performa model yang lebih stabil dan andal.

22. ROC-AUC (Receiver Operating Characteristic — Area Under the Curve): Metrik evaluasi yang mengukur kemampuan model dalam membedakan kelas positif dan negatif secara keseluruhan, di mana nilai mendekati 1,0 menunjukkan kemampuan diskriminasi yang sangat baik.
23. Mutual Information: Ukuran statistik yang digunakan dalam seleksi fitur untuk mengukur tingkat ketergantungan antara setiap fitur dengan variabel target, sehingga fitur yang paling informatif dapat diprioritaskan.
24. Overfitting: Kondisi di mana model terlalu menyesuaikan diri dengan data latih sehingga performanya menurun secara signifikan ketika diuji pada data baru yang belum pernah dilihat sebelumnya.
25. MIB (Machine Intelligence and Bot detection) Dataset: Dataset publik yang dikembangkan oleh IIT-CNR, Italia, berisi data profil akun Twitter asli dan berbagai kategori akun bot yang telah melalui proses anotasi manual dan digunakan secara luas sebagai benchmark dalam penelitian deteksi akun palsu.

DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan Dosen Pembimbing 1

Lampiran 2. Kartu Bimbingan Dosen Pembimbing 2

