

**PREDIKSI PEMBELIAN BERDASARKAN *CLICK THROUGH RATE*
IKLAN DIGITAL MENGGUNAKAN ALGORITMA
*RANDOM FOREST***

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH:

DEWI PUTRIANI

NPM: 2113030001

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI

2025

Skripsi oleh:

DEWI PUTRIANI

NPM: 2113030001

Judul:

**PREDIKSI PEMBELIAN BERDASARKAN *CLICK THROUGH RATE*
IKLAN DIGITAL MENGGUNAKAN ALGORITMA *RANDOM FOREST***

Telah disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi
FTIK UN PGRI Kediri

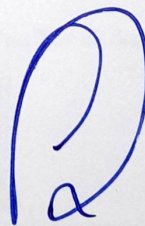
Tanggal: 1 Juli 2025

Pembimbing I



Sucipto, M.Kom
NIDN.0721029101

Pembimbing II



Arie Nugroho, S.Kom, M.M, M.Kom
NIDN.0712108103

Skripsi oleh:

DEWI PUTRIANI

NPM: 2113030001

Judul:

**PREDIKSI PEMBELIAN BERDASARKAN *CLICK THROUGH RATE*
IKLAN DIGITAL MENGGUNAKAN ALGORITMA *RANDOM FOREST***

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Prodi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 11 Juli 2025

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Pengujian:

1. Ketua : Sucipto, M.Kom.

[.....]

2. Penguji 1 : Aidina Ristyan, M.Kom.

[.....]

3. Penguji 2 : Arie Nugroho, S.Kom, M.M, M.Kom.

[.....]

Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer



Dr. SULISTIONO, M.Si.

NIDN/0007076801

PERNYATAAN

Yang bertanda tangan dibawah ini saya,

Nama	: Dewi Putriani
Jenis Kelamin	: Perempuan
Tempat/tgl. lahir	: Kediri/ 27 Juli 2002
NPM	: 2113030001
Fak/Jur./Prodi	: FTIK/ S1/ Sistem Informasi

menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 1 Juli 2025

Yang Menyatakan



DEWI PUTRIANI

NPM: 2113030001

MOTTO

"I screamed, cried, and almost gave up; I complained the whole way
— but I still crushed it."

Kupersembahkan karya ini buat:
Diri Sendiri dan Orang Tua

Abstrak

Dewi Putriani: Prediksi Pembelian Berdasarkan *Click Through Rate* Iklan Digital Menggunakan Algoritma *Random Forest*, Skripsi, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2025.

Kata Kunci: prediksi, *click through rate*, iklan digital, algoritma *Random Forest*.

Perkembangan pesat teknologi digital telah mengubah lanskap periklanan, terutama melalui digital advertising yang kini mendominasi berbagai platform daring. Meskipun metrik *click through rate* (CTR) kerap dijadikan indikator utama dalam mengukur efektivitas iklan digital, kenyataannya CTR yang tinggi tidak selalu berbanding lurus dengan keputusan pembelian. Hal ini memicu kebutuhan untuk membangun model prediksi yang tidak hanya mengandalkan klik semata, tetapi juga mampu memprediksi kemungkinan konversi secara lebih akurat.

Penelitian ini bertujuan untuk menghasilkan model prediksi pembelian berdasarkan CTR iklan digital dengan menggunakan algoritma *Random Forest*, yang dioptimasi melalui tuning hyperparameter dan penerapan *class weight* untuk menangani masalah ketidakseimbangan kelas. Metode yang digunakan dalam penelitian ini mengikuti pendekatan CRISP-DM dengan data *Social Network Ads* dari Kaggle.

Hasil penelitian menunjukkan bahwa model *Random Forest* yang dioptimasi mampu memberikan performa yang sangat baik dalam mengklasifikasikan keputusan pembelian pengguna. Model ini mencapai akurasi sebesar 93%, *recall* sebesar 0.98, dan *F1-score* sebesar 0.92. Performa ini menunjukkan peningkatan signifikan dibandingkan model-model sebelumnya, seperti *Decision Tree* (82%) dan *Light Gradient Boosting* (91%).

Temuan ini menegaskan bahwa kombinasi tuning hyperparameter dan penggunaan *class weight* menjadi strategi efektif dalam menangani ketidakseimbangan data dan meningkatkan performa model prediksi. Penelitian ini diharapkan dapat memberikan kontribusi praktis dalam strategi pemasaran digital berbasis data, serta menjadi acuan dalam pengembangan sistem rekomendasi yang lebih responsif terhadap kebutuhan pengguna.

KATA PENGANTAR

Puji Syukur Kami panjatkan kehadirat Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan-Nya penyusunan skripsi ini dapat diselesaikan.

Skripsi dengan judul “Prediksi Pembelian Berdasarkan *Click Through Rate* Iklan Digital Menggunakan Algoritma *Random Forest*” ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Prodi Sistem Informasi FTIK UN PGRI Kediri. Pada kesempatan ini diucapkan terimakasih dan penghargaan yang setulus tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Dr. Sulistiono, M.Si. selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri.
3. Sucipto, M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
4. Sucipto, M.Kom. dan Arie Nugroho, M.Kom. selaku Dosen Pembimbing yang telah memberikan arahan, saran, dan motivasi dalam penyusunan skripsi ini.
5. Orang Tua dan Keluarga, yang selalu memberikan dukungan doa, dan semangat dalam menyelesaikan setiap tahapan pendidikan.
6. Teman – teman dan Rekan Mahasiswa, yang telah memberikan bantuan, kerja sama, serta dukungan selama proses penyusunan skripsi ini.

Disadari bahwa skripsi ini masih banyak kekurangan, maka diharapkan tegur sapa, kritik, dan saran – saran, dari berbagai pihak sangat diharapkan.

Akhirnya, disertai harapan semoga skripsi ini ada manfaatnya bagi kita semua, khususnya bagi dunia pendidikan, meskipun hanya ibarat setitik air bagi samudra luas.

Kediri, 1 Juli 2025

DEWI PUTRIANI
NPM: 2113030001

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN	iii
HALAMAN MOTTO	iv
ABSTRAK	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN.....	1
A. Latar Belakang dan Permasalahan	1
B. Batasan Masalah.....	5
C. Rumusan Masalah	6
D. Tujuan Penelitian.....	6
E. Manfaat Penelitian	6
BAB II KAJIAN PUSTAKA	7
A. Kajian Teori.....	7
1. Click Through Rate (CTR).....	7
2. Prediksi Keputusan Pembelian Konsumen	8
3. Data Mining	9
4. Metodologi CRISP-DM	11
5. Algoritma <i>Random Forest</i>	12
6. <i>Tuning Hyperparameter</i>	14
7. Penanganan Ketidakseimbangan Kelas.....	16
8. Evaluasi Model.....	17
B. Kajian Hasil Penelitian Terdahulu.....	18
C. Kerangka Berpikir.....	20

BAB III METODE PENELITIAN	22
A. <i>Business Understanding</i>	23
1. <i>Determine business objective</i>	23
2. <i>Assess situation</i>	23
3. <i>Determine data mining goals</i>	23
4. <i>Plan activities</i>	24
B. <i>Data Understanding</i>	24
1. <i>Kumpulkan data (Collect initial data)</i>	24
2. <i>Deskripsi Data</i>	25
3. <i>Eksplorasi Data</i>	25
4. <i>Verifikasi Kualitas Data</i>	25
C. <i>Data Preparation</i>	25
1. <i>Data Selection</i>	26
2. <i>Data Preprocessing</i>	26
3. <i>Data Transformation</i>	26
D. <i>Modeling</i>	27
1. <i>Select Modeling Technique</i>	27
2. <i>Generate Test Design</i>	28
3. <i>Build Model</i>	28
E. <i>Evaluation</i>	31
1. <i>Evaluate Results</i>	31
2. <i>Review Process</i>	32
F. <i>Deployment</i>	33
1. <i>Plan Deployment</i>	33
2. <i>Monitoring and Maintenance</i>	33
3. <i>Reporting</i>	33
BAB IV HASIL DAN PEMBAHASAN	34
A. <i>Hasil</i>	34
1. <i>Data Understanding</i>	34
2. <i>Data Preparation</i>	47
3. <i>Modeling</i>	53

4. <i>Evaluation</i>	63
5. <i>Deployment</i>	66
B. Pembahasan.....	66
BAB V PENUTUP	70
A. Kesimpulan	70
B. Saran.....	71
DAFTAR PUSTAKA	72
LAMPIRAN	76

DAFTAR TABEL

Tabel 4.1. Data Awal	34
Tabel 4.2. Deskripsi Dataset.....	35
Tabel 4.3. Nilai Kosong	44
Tabel 4.4. Inskonsisten Data	46
Tabel 4.5. Pemilihan Fitur	48
Tabel 4.6. Inkosistensi Label.....	48
Tabel 4.7. Hasil Penanganan	49
Tabel 4.8. Data Gender Sebelum Transformasi	50
Tabel 4.9. Data Gender Setelah Transformasi.....	50
Tabel 4.10. Fitur (X) yang Digunakan untuk Pemodelan	51
Tabel 4.11. Label (y) yang Digunakan untuk Pemodelan	51
Tabel 4.12. Data Fitur Numerik Sebelum Standarisasi	52
Tabel 4.13. Data Fitur Numerik Setelah Standarisasi	52
Tabel 4.14. Distribusi Data Latih dan Uji	53
Tabel 4.15. Data Sampel	56
Tabel 4.16. Hasil Bootstrap Sampling.....	56
Tabel 4.17. Hasil Cross-Validation.....	64
Tabel 4.18. Hasil Evaluasi.....	65
Tabel 4.19. Perbandingan Peforma Model.....	67
Tabel 4.20. Hasil Pengujian	68

DAFTAR GAMBAR

Gambar 2.1. Kerangka Berpikir	21
Gambar 3.1. CRISP-DM	22
Gambar 3.2. <i>Business Understanding</i>	23
Gambar 3.3. <i>Data Understanding</i>	24
Gambar 3.4. <i>Data Preparation</i>	26
Gambar 3.5. <i>Modeling</i>	27
Gambar 3.6. <i>Evaluation</i>	31
Gambar 3.7. <i>Deployment</i>	33
Gambar 4.1. Distribusi Usia	36
Gambar 4.2. Distribusi Gaji	37
Gambar 4.3. Distribusi <i>Gender</i>	38
Gambar 4.4. Distribusi <i>Purchased</i>	39
Gambar 4.5. Distribusi Usia berdasarkan Pembelian	40
Gambar 4.6. Distribusi Gaji berdasarkan Pembelian	41
Gambar 4.7. Distribusi Gender berdasarkan Pembelian	41
Gambar 4.8. Heatmap Korelasi Antar Fitur	42
Gambar 4.9. Duplikasi Data	44
Gambar 4.10. Boxplot Fitur Age	45
Gambar 4.11. Boxplot Fitur <i>EstimatedSalary</i>	45
Gambar 4.12. Pohon Keputusan Pertama	59
Gambar 4.13. Pohon Keputusan Kedua	60
Gambar 4.14. Pohon Keputusan Ketiga	60

DAFTAR LAMPIRAN

Lampiran 1 Dataset <i>Social Network Ads</i>	76
Lampiran 2 <i>Script Python</i>	77
Lampiran 3 Kartu Bimbingan Skripsi	78
Lampiran 4 Surat Keterangan Bebas Similarity.....	81
Lampiran 5 Bukti Halaman Awal Cek Similarity	82
Lampiran 6 Bukti Submit Artikel/LOA/Terbit.....	83

BAB I

PENDAHULUAN

A. Latar Belakang dan Permasalahan

Perkembangan teknologi digital telah membawa transformasi besar dalam dunia pemasaran, terutama dengan melalui periklanan *online* atau *digital advertising* yang saat ini menjadi salah satu sektor bisnis paling cepat berkembang dan paling menguntungkan (Handayani, 2023). Iklan digital saat ini telah menjangkau berbagai platform seperti media sosial, portal pencarian, *e-commerce*, aplikasi seluler, hingga video *online* dan permainan *online*. Hal tersebut menjadikan iklan digital dominan dalam industri periklanan (Yang et al., 2022). Menurut laporan Statista, pendapatan dari iklan digital di Amerika Serikat meningkat sebesar 12,2% pada tahun 2020 dibandingkan tahun sebelumnya, dan diperkirakan mencapai 989,82 miliar dolar pada tahun 2025 (Yang et al., 2022). Fenomena serupa juga terjadi di Indonesia dimana belanja iklan digital tumbuh secara signifikan, dengan mayoritas pengiklan kini beralih ke saluran digital seperti *Google*, *Instagram*, dan *TikTok* untuk menjangkau konsumen secara lebih tertarget.

Dalam persaingan dunia digital, perusahaan menginvestasikan dana yang sangat besar untuk strategi pemasaran, khususnya dalam bentuk iklan hal tersebut digunakan untuk mempromosikan produk dan jasa mereka (Zhou et al., 2023). Di Amerika Serikat, pengeluaran untuk iklan diperkirakan mencapai 345,3 miliar dolar pada tahun 2022 (Zhou et al., 2023). Oleh karena itu, penting bagi perusahaan untuk memastikan bahwa pengeluaran tersebut menghasilkan dampak yang signifikan terutama dalam hal konversi sehingga penyedia iklan perlu memahami dan menyesuaikan target audiens secara mendalam (Zhou et al., 2023). Salah satu metrik utama dalam mengukur efektivitas iklan digital adalah klik, yang merepresentasikan minat awal dari pengguna terhadap suatu merek (Yang et al., 2022; Zhou et al., 2023). Di Indonesia, metrik seperti *click through rate* (CTR) juga digunakan secara luas oleh pelaku bisnis digital, khususnya dalam kampanye penjualan berbasis *e-commerce* maupun iklan bersponsor.

Namun, dalam praktiknya keputusan pembelian tidak selalu sebanding dengan klik iklan (Handayani, 2023). Perusahaan harus mengetahui apakah iklan benar – benar dilihat dan dipertimbangkan oleh pengguna atau hanya diklik tanpa tindakan lebih lanjut (Handayani, 2023; Lakshmanarao et al., 2020). Oleh karena itu, selain menargetkan pengguna yang tepat penting pula untuk menyajikan iklan yang relevan dengan kebutuhan pengguna agar potensi konversi meningkat (Lakshmanarao et al., 2020). Iklan yang ditempatkan dengan strategi yang tepat tidak hanya meningkatkan kemungkinan klik tetapi juga berpengaruh terhadap keputusan pembelian (Lakshmanarao et al., 2020).

Dalam kondisi periklanan media sosial yang bersifat dinamis, memahami perilaku dan preferensi pengguna menjadi kunci dalam menyusun strategi kampanye yang efektif (I.V. Dwaraka Srihith et al., 2023). Tidak semua produk sesuai untuk semua kalangan sehingga faktor – faktor seperti usia dan penghasilan menjadi penting dalam menentukan segmen pelanggan yang paling tepat (I.V. Dwaraka Srihith et al., 2023a; Venkata et al., 2023a). Dengan demikian, kemampuan untuk mengklasifikasikan dan mengelompokkan audiens berdasarkan potensi pembelian sangat dibutuhkan untuk mengoptimalkan tingkat konversi iklan digital (I.V. Dwaraka Srihith et al., 2023a; Venkata et al., 2023a). Kemajuan dalam teknologi data juga telah memungkinkan perusahaan untuk melakukan analisis prediktif terhadap klik iklan dan penayangan melalui pendekatan berbasis data besar (*big data*) (Saraswathi et al., 2019). Berbagai model prediksi klik telah dikembangkan untuk berbagai format periklanan seperti iklan pencarian, iklan teks, hingga iklan video (Saraswathi et al., 2019). Seiring dengan berkembangnya jaringan periklanan digital analisis terhadap perilaku klik dan kemungkinan konversi menjadi semakin kompleks dan menuntut pendekatan algoritmik yang lebih akurat dan efisien (Saraswathi et al., 2019).

Seperti yang telah disebutkan sebelumnya semua klik pada iklan digital tidak selalu berarti pembelian. Meskipun perusahaan dapat memperoleh keuntungan finansial dari setiap klik iklan (Lakshmanarao et al., 2020), banyak ditemukan bahwa klik yang tinggi tidak serta merta diikuti oleh peningkatan konversi atau pembelian. Hal tersebut menimbulkan tantangan dalam mengevaluasi efektivitas kampanye iklan karena keberhasilan tidak cukup hanya dilihat dari rasio klik

tayang, tetapi juga dari sejauh mana klik tersebut menghasilkan tindakan nyata seperti pembelian (Lakshmanarao et al., 2020a; Yang et al., 2022). Ketidaksesuaian antara CTR dan konversi dapat terjadi karena berbagai alasan, termasuk kurangnya relevansi iklan terhadap kebutuhan pengguna atau pengalaman pengguna yang tidak memuaskan setelah mengklik iklan.

Dalam kondisi tersebut, penting bagi perusahaan untuk memahami lebih dalam perilaku pengguna termasuk preferensi dan potensi keputusan pembelian mereka (I.V. Dwaraka Srihith et al., 2023a; Venkata et al., 2023a). Iklan yang ditempatkan secara tepat dapat meningkatkan kemungkinan pembelian, terutama jika iklan tersebut sesuai dengan masalah atau kebutuhan pengguna (Lakshmanarao et al., 2020). Oleh karena itu, perusahaan perlu mempertimbangkan strategi yang lebih komprehensif dan tidak hanya mengandalkan CTR sebagai indikator utama, tetapi juga menggunakan pendekatan analitik berbasis data.

Seiring dengan terus bertambahnya volume data yang dihasilkan dari aktivitas digital pengguna, tantangan utama dalam dunia pemasaran bukan hanya sekadar mengumpulkan data, tetapi juga bagaimana mengolahnya menjadi informasi yang bermakna untuk mendukung pengambilan keputusan strategis (Arie Nugroho et al., 2021). Dalam konteks ketidakseimbangan antara CTR dan konversi, dibutuhkan pendekatan analitik yang mampu memahami lebih dalam pola perilaku pengguna di balik interaksi mereka terhadap iklan. Salah satu pendekatan yang banyak digunakan adalah teknik data mining, yang bertujuan menggali pola tersembunyi dari data historis guna menghasilkan prediksi yang lebih akurat terkait kemungkinan pembelian (Arie Nugroho, Husin, et al., n.d.). Teknik ini umumnya melibatkan algoritma machine learning yang memungkinkan komputer belajar dari data dan membentuk model klasifikasi atau prediksi (Arie Nugroho, Husin, et al., n.d.). Dengan machine learning, data interaksi seperti klik iklan dapat dianalisis untuk memperkirakan seberapa besar kemungkinan pengguna akan melakukan pembelian, bukan hanya berhenti pada interaksi awal (Arie Nugroho et al., 2021; Arie Nugroho, Husin, et al., n.d.). Oleh karena itu, dalam penelitian ini digunakan algoritma *Random Forest* sebagai salah satu metode machine learning yang dikenal mampu menangani data tidak seimbang dan memberikan hasil prediksi yang stabil

serta akurat dalam konteks klasifikasi pembelian berdasarkan data iklan digital (Saraswathi et al., 2019).

Dalam penelitian ini, algoritma *Random Forest* dipilih untuk memprediksi kemungkinan setelah pengguna mengeklik iklan. *Random Forest* memiliki keunggulan dalam stabilitas prediksi dan bekerja dengan menggabungkan beberapa pohon keputusan sehingga mampu mengurangi risiko *overfitting* (Adriansyah et al., 2022; Arisusanto et al., 2023). Selain itu, algoritma ini memberikan hasil prediksi yang akurat dan stabil pada dataset berukuran kecil hingga sedang (Yoga Religia et al., 2021). Untuk meningkatkan performa model, penelitian ini menerapkan teknik *tuning hyperparameter* menggunakan metode *Random Search* yang efektif dalam menemukan kombinasi parameter terbaik secara efisien tanpa harus melakukan pencarian secara menyeluruh (Aribowo et al., 2024). Pendekatan tersebut membantu mengoptimalkan kinerja *Random Forest* sehingga dapat memberikan hasil prediksi mengeneralisasi lebih baik pada data baru. Penelitian ini juga menerapkan metode *class weight* untuk menangani masalah ketidakseimbangan data, dimana jumlah kelas antara pengguna yang melakukan pembelian dan yang tidak cenderung tidak seimbang (Adriansyah et al., 2022). Dengan pengaturan *class weight*, model *Random Forest* dapat memberikan bobot lebih tinggi pada kelas yang kurang terwakili sehingga meningkatkan performa prediksi secara keseluruhan dan menghasilkan prediksi yang lebih adil (Arisusanto et al., 2023).

Masalah ketidakseimbangan data merupakan tantangan umum dalam pemodelan prediksi pembelian berdasarkan klik iklan digital. Karena jumlah pengguna yang mengklik iklan namun tidak melakukan pembelian jauh lebih banyak daripada yang benar-benar membeli model cenderung bias terhadap kelas mayoritas, sehingga potensi mendeteksi pembeli yang sebenarnya sering terlewatkan (Handayani, 2023). Oleh karena itu, penting bagi model untuk mampu memperhatikan kelas minoritas agar prediksi menjadi lebih sensitif dan akurat terhadap perilaku pembelian yang sesungguhnya (Wardani et al., 2022). Penggunaan *class weight* menjadi solusi efektif karena memungkinkan model mempelajari pola dari kedua kelas dengan proporsi yang lebih seimbang tanpa harus melakukan perubahan signifikan pada data asli (Zhu et al., 2018).

Penelitian ini bertujuan untuk mengoptimalkan akurasi prediksi berdasarkan hasil-hasil penelitian sebelumnya yang menggunakan dataset *Social Network Ads* dari Kaggle (Rakesh Raushan, n.d.). Dalam studi-studi terdahulu, berbagai algoritma telah diaplikasikan untuk memprediksi kemungkinan pembelian berdasarkan interaksi pengguna dengan iklan. Salah satu penelitian menggunakan algoritma *Light Gradient Boosting* (LGBM) yang dikombinasikan dengan metode resampling seperti *Random Over Sampling* (ROS), *Random Under Sampling* (RUS), *Synthetic Minority Over-sampling Technique* (SMOTE), dan SMOTE-Tomek, berhasil mencapai akurasi sebesar 91.25%, presisi 84.38%, serta nilai *Area Under Curve* (AUC) sebesar 0.92, yang menunjukkan performa model yang cukup baik dalam menangani ketidakseimbangan data (Handayani, 2023). Sementara itu, penelitian lain yang menggunakan algoritma *Decision Tree* dengan dataset yang sama memperoleh akurasi yang lebih rendah, yaitu sekitar 82% dengan presisi 80% (Venkata et al., 2023). Perbedaan hasil ini mengindikasikan bahwa pemilihan algoritma dan teknik penanganan data sangat mempengaruhi kualitas prediksi yang dihasilkan. Oleh karena itu, pendekatan lain yang mengombinasikan algoritma *Random Forest* dengan teknik penanganan ketidakseimbangan data dan *tuning hyperparameter* diharapkan dapat menghasilkan model prediksi yang lebih stabil dan memiliki performa yang baik.

B. Batasan Masalah

Prediksi memiliki cakupan yang luas, maka dari itu agar penelitian lebih fokus terdapat beberapa batasan masalah yaitu antara lain.

1. Dataset yang digunakan adalah *Social Network Ads* dari Kaggle dengan jumlah 400 data.
2. Model yang digunakan adalah algoritma *Random Forest*.
3. Penanganan ketidakseimbangan kelas menggunakan pengaturan *class weight* pada *Random Forest*.
4. Pemilihan parameter menggunakan metode random search untuk optimasi parameter model.
5. Penelitian hanya sebatas pembentukan model algoritma tidak sampai tahap *deployment*.

C. Rumusan Masalah

Berdasarkan uraian latar belakang dan batasan masalah, terdapat beberapa pertanyaan dalam mengevaluasi efektivitas algoritma *Random Forest* saat memprediksi kemungkinan pembelian setelah klik iklan pada dataset *Social Network Ads*. Pertanyaan – pertanyaan tersebut adalah sebagai berikut:

1. Apakah model *Random Forest* mampu memberikan performa prediksi yang baik pada dataset dengan ketidakseimbangan kelas?
2. Apakah penggunaan metode *class weight* pada *Random Forest* dapat meningkatkan akurasi prediksi untuk kelas minoritas dalam dataset ini?
3. Berapa tingkat *accuracy*, *precision*, *recall*, *F1 score*, dan *AUC* yang dihasilkan oleh model setelah optimasi parameter menggunakan *random search*?

D. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan sebelumnya, maka tujuan dari penelitian ini adalah antara lain:

1. Menghasilkan model prediksi dengan akurasi yang lebih baik dengan menggunakan algoritma *Random Forest*.
2. Meningkatkan perhatian model terhadap kelas minoritas dengan menggunakan metode *class weight* pada algoritma *Random Forest*.

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat, antara lain:

1. Memberikan wawasan tentang efektivitas metode *class weight* dalam menangani kelas minoritas pada algoritma *Random Forest*.
2. Membantu bisnis atau praktisi pemasaran digital dengan meningkatkan akurasi penargetan iklan melalui analisis data yang lebih terstruktur.
3. Memungkinkan perusahaan untuk meningkatkan tingkat konversi dan efektivitas strategi pemasaran mereka.

DAFTAR PUSTAKA

- Adiguno, S., Syahra, Y., & Yetri, M. (2022). *Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda*.
- Adriansyah, I., Mahendra, M. D., Rasywir, E., & Pratama, Y. (2022). Perbandingan Metode Random Forest Classifier dan SVM Pada Klasifikasi Kemampuan Level Beradaptasi Pembelajaran Jarak Jauh Siswa. *Bulletin of Informatics and Data Science*, 1(2). Retrieved from <https://ejurnal.pdsi.or.id/index.php/bids/index>
- Akbar, H., & Sanjaya, W. K. (2023). Kajian Performa Metode Class Weight Random Forest pada Klasifikasi Imbalance Data Kelas Curah Hujan. *Jurnal Sains, Nalar, Dan Aplikasi Teknologi Informasi*, 3(1). doi: 10.20885/snati.v3i1.30
- Alam, S., Dipa Makassar, U., & Perintis Kemerdekaan, J. K. (n.d.). *Analisis Prediksi Penjualan Kue menggunakan Metode Naive Bayes*.
- Anggoro, D. A., & Mukti, S. S. (2021). Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure. *International Journal of Intelligent Engineering and Systems*, 14(6), 198–207. doi: 10.22266/ijies2021.1231.19
- Aribowo, A. S., Cahyana, N. H., & Fauziah, Y. (2024). *Enhancing Semi-Supervised Sentiment Analysis Through Hyperparameter Tuning Within Iterations: A Comparative Study Using Grid Search and Random Search* (pp. 248–260). doi: 10.2991/978-94-6463-366-5_23
- Arisusanto, A., Suarna, N., & Dwilestari, G. (2023). Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma Random Forest Dengan Optimize Parameter Grid. *Jurnal Teknologi Ilmu Komputer*, 1(2), 43–47. doi: 10.56854/jtik.v1i2.51
- Citra Mawani, A., Li Hin, L., & Anubhakti, D. (2023). DETEKSI DINI GEJALA AWAL PENYAKIT DIABETES MENGGUNAKAN ALGORITMA RANDOM FOREST. In *Idealis: Indonesia Journal Information System* (Vol. 6, Issue 2). Retrieved from <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index> AjengCitraMawani|<http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>
- Feblian, D., & Daihani, D. U. (2017). IMPLEMENTASI MODEL CRISP-DM UNTUK MENENTUKAN SALES PIPELINE PADA PT X. *JURNAL TEKNIK INDUSTRI*, 6(1).doi: 10.25105/jti.v6i1.1526
- Firdaus, M. (2022). *PENENTUAN KLASIFIKASI DENGAN CRISP-DM DALAM MEMPREDIKSI KELULUSAN MAHASISWA PADA SUATU MATA KULIAH*.

- Fulazzaky, T., Saefuddin, A., & Soleh, A. M. (2024). Evaluating Ensemble Learning Techniques for Class Imbalance in Machine Learning: A Comparative Analysis of Balanced Random Forest, SMOTE-RF, SMOTEBoost, and RUSBoost *Scientific Journal of Informatics*, 11(4), 969–980 doi: 10.15294/sji.v11i4.15937
- Handayani, K. (2023a). PENERAPAN LIGHT GRADIENT BOOSTING DALAM PREDIKSI RASIO KLIK TAYANG. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1)
- Handayani, K. (2023b). PENERAPAN LIGHT GRADIENT BOOSTING DALAM PREDIKSI RASIO KLIK TAYANG. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1).
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 5, Issue 2). Retrieved from <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Istiana, N., & Mustafiril, A. (2023). Perbandingan Metode Klasifikasi pada Data dengan Imbalance Class dan Missing Value. *Jurnal Informatika*, 10(2), 101–108. doi: 10.31294/inf.v10i2.15540
- I.V. Dwaraka Srihith, T. Aditya Sai Srinivas, K. Owdharya, A. David Donald, & G. Thippana. (2023a). Clicks and Conversions: Unveiling the Power of Machine Learning in Social Media Ad Classification. *International Journal of Advanced Research in Science, Communication and Technology*, 1–6. doi: 10.48175/ijarsct-11401
- I.V. Dwaraka Srihith, T. Aditya Sai Srinivas, K. Owdharya, A. David Donald, & G. Thippana. (2023b). Clicks and Conversions: Unveiling the Power of Machine Learning in Social Media Ad Classification. *International Journal of Advanced Research in Science, Communication and Technology*, 1–6. doi: 10.48175/ijarsct-11401
- Lakshmanarao, A., Srisaila, A., & Kiran, T. S. R. (2020a). An Efficient Ad Click Prediction System using Machine Learning Techniques. *International Journal of Engineering and Advanced Technology*, 9(3), 1269–1272. doi: 10.35940/ijeat.C5518.029320
- Lakshmanarao, A., Srisaila, A., & Kiran, T. S. R. (2020b). An Efficient Ad Click Prediction System using Machine Learning Techniques. *International Journal of Engineering and Advanced Technology*, 9(3), 1269–1272. doi: 10.35940/ijeat.C5518.029320
- Nila, U., & Firliana, R. (2023). Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 175 Analisis Data Transaksi Penjualan Produk Pertanian Menggunakan Algoritma FP-Growth Kata Kunci-data transaksi, data mining, Algoritma FP-Growth. In *Agustus* (Vol. 7). Online.
- Nugroho, Adityo, Asror, I., & Firdaus, Y. (2023). *Klasifikasi Tingkat Kualitas*

Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest.

- Nugroho, Arie.** (2022). Analisa Splitting Criteria Pada Decision Tree dan Random Forest untuk Klasifikasi Evaluasi Kendaraan. *JSITIK: Jurnal Sistem Informasi Dan Teknologi Informasi Komputer*, 1(1), 41–49. doi: 10.53624/jsitik.v1i1.154
- Nugroho, Arie, & Harini, D.** (2024). Teknik Random Forest untuk Meningkatkan Akurasi Data Tidak Seimbang. *JSITIK*, 2(2). doi: 10.53624/jsitik.v2i2.XX
- Nugroho, Arie, Husin, A., Provinsi, J., Hulu, T., & Hilir Indonesia, I.** (n.d.). SISTEMASI: Jurnal Sistem Informasi Analisis Performa Random Forest Menggunakan Normalisasi Atribut Performance Analysis of Random Forest Using Attribute Normalization. In Januari (Vol. 11, Issue 1). Retrieved from <http://sistemasi.ftik.unisi.ac.id>
- Nugroho, Arie, Soeleman, Ma., Anggi Pramunendar, R., & Nurhindarto, A.** (n.d.). PENINGKATAN PERFORMA ENSEMBLE LEARNING PADA SEGMENTASI SEMANTIK GAMBAR DENGAN TEKNIK OVERSAMPLING UNTUK CLASS IMBALANCE. doi: 10.25126/jtiik.2023106831
- Nugroho, Arie, Zaniul, A., & Fajar Guruh.** (2021). *Evaluation of Feature Selection Using Wrapper For Numeric Dataset With Random Forest Algorithm*. IEEE.
- Rakesh Raushan. (n.d.). *Social Network Ads*.
- Saputra, D. B., Atina, V., Nastiti, F. E., & Komputer, F. I. (2024). PENERAPAN MODEL CRISP-DM PADA PREDIKSI NASABAH KREDIT MENGGUNAKAN ALGORITMA RANDOM FOREST. In *Idealis: Indonesia Journal Information System* (Vol. 7, Issue 2). Retrieved from <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexDwiBagusSaputra>
<http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>
- Saraswathi, S., Krishnamurthy, V., Venkata Vara Prasad, D., Tarun, R. K., Abhinav, S., & Rushitaa, D. (2019a). Machine learning based ad-click prediction system. *International Journal of Engineering and Advanced Technology*, 8(6), 3646–3648. doi: 10.35940/ijeat.F9366.088619
- Saraswathi, S., Krishnamurthy, V., Venkata Vara Prasad, D., Tarun, R. K., Abhinav, S., & Rushitaa, D. (2019b). Machine learning based ad-click prediction system. *International Journal of Engineering and Advanced Technology*, 8(6), 3646–3648. doi: 10.35940/ijeat.F9366.088619
- Sucipto, S., Dwi Prasetya, D., & Widiyaningtyas, T.** (n.d.). *An Evaluation of the Impact of Dataset Size on Classification Performance in the Cognitive Bloom's Taxonomy*. Retrieved from <https://orcid.org/0000-0003-3412-002X>
- Venkata, I., & Srihith, D. (2023a). *Predicting Success: The Impact of Machine*

- Learning on Social Media Ad Classification*. doi: 10.5281/zenodo.8214337
- Venkata, I., & Srihith, D. (2023b). *Predicting Success: The Impact of Machine Learning on Social Media Ad Classification*. doi: 10.5281/zenodo.8214337
- Wang, Y., Ji, H., Yu, J., He, X., Li, Z., & Liu, T. (n.d.). *Click-Through Rate Prediction Based on User Behaviors and Feature Importance*. Retrieved from <https://github.com/jihuiqin2/sequence>
- Wardani, A. P. P., Adiwijaya, A., & Purbolaksono, M. D. (2022). Sentiment Analysis on Beauty Product Review Using Modified Balanced Random Forest Method and Chi-Square. *Journal of Information System Research (JOSH)*, 4(1), 1–7. doi: 10.47065/josh.v4i1.2047
- Wijiyanto, W., Pradana, A. I., Sopingi, S., & Atina, V. (2024). Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa. *Jurnal Algoritma*, 21(1). doi: 10.33364/algoritma/v.21-1.1618
- Yang, Y., & Zhai, P. (2022). Click-through rate prediction in online advertising: A literature review. *Information Processing and Management*, 59(2). doi: 10.1016/j.ipm.2021.102853
- Yoga Religia, Agung Nugroho, & Wahyu Hadikristanto. (2021). Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 187–192. doi: 10.29207/resti.v5i1.2813
- Zhou, X., & Richards, S. (2023). Predicting Click Behavior Based on Machine Learning Models. In *BCP Business & Management FIBA* (Vol. 2023).
- Zhu, M., Xia, J., Jin, X., Yan, M., Cai, G., Yan, J., & Ning, G. (2018). Class weights random forest algorithm for processing class imbalanced medical data. *IEEE Access*, 6, 4641–4652. doi: 10.1109/ACCESS.2018.2789428