Journal of Informatic Engineering (JOUTICA)
https://jurnalteknik.unisla.ac.id/index.php/informatika

Prediksi Pembelian Berdasarkan Click Through Rate Iklan Digital Menggunakan Algoritma Random Forest

Dewi Putriani[1], Sucipto[2*], Arie Nugroho[3]

[1]Universitas Nusantara PGRI Kediri,
Jl. Ahmad Dahlan No.76, Mojoroto, Kec. Mojoroto, Kota Kediri
dewiptr098@gmail.com

[2]Universitas Nusantara PGRI Kediri,
Jl. Ahmad Dahlan No.76, Mojoroto, Kec. Mojoroto, Kota Kediri
sucipto@unpkediri.ac.id
*Corresponding Author

[3]Universitas Nusantara PGRI Kediri,
Jl. Ahmad Dahlan No.76, Mojoroto, Kec. Mojoroto, Kota Kediri
arienugroho@unpkediri.ac.id

Abstrak:
Perkembangan teknologi digital telah mengubah strategi pemasaran, menjadikan iklan digital sebagai sarana utama untuk menjangkau konsumen secara lebih tepat sasaran. Namun, keberhasilan kampanye iklan tidak hanya bergantung pada tingkat klik (*Click Through Rate*/CTR), melainkan juga pada kemampuan sistem dalam mengidentifikasi pengguna yang berpotensi melakukan pembelian. Penelitian ini bertujuan untuk membangun model prediksi perilaku pembelian berdasarkan CTR dengan algoritma *Random forest* dan pendekatan CRISP-DM. Dataset yang digunakan berasal dari *Social Network Ads* dan terdiri dari 400 entri dengan atribut demografis seperti usia, jenis kelamin, dan estimasi gaji. Model dibangun dalam dua tahap, yaitu *baseline* dan hasil *tuning*. Evaluasi dilakukan menggunakan metrik klasifikasi, dan model hasil *tuning* berhasil mencapai akurasi sebesar 93%, *recall* 98%, dan *F1-score* 92%, menunjukkan performa yang unggul dalam mengenali kelas pembelian. Hasil ini menunjukkan bahwa *Random forest* dengan *tuning hyperparameter* dan *class weight* dapat menjadi solusi yang efektif dalam klasifikasi pengguna iklan digital dan mendukung pengambilan keputusan pemasaran yang lebih efisien.

Kata Kunci:
Iklan Digital, *Click Through Rate*, *Random forest*

Abstract:
The development of digital technology has changed marketing strategies, making digital advertising the main means to reach consumers in a more targeted manner. However, the success of an advertising campaign depends not only on the *Click Through Rate* (CTR), but also on the system's ability to identify users who have the potential to make purchases. This research aims to build a purchase behavior prediction model based on CTR with *Random forest* algorithm and CRISP-DM approach. The dataset used comes from *Social Network Ads* and consists of 400 entries with demographic attributes such as *Age*, *Gender*, and estimated salary. The model was built in two st*Age*s, namely *baseline* and *tuning* results. Evaluation was conducted using classification metrics, and the tuned model achieved 93% *accuracy*, 98% *recall*, and 92% *F1-score*, showing superior performance in recognizing the purchase class. These results show that *Random forest* with *hyperparameter* and *class weight tuning* can be an effective solution in digital advertising user classification and support more efficient marketing decision making.

Keywords:
Digital Advertising, *Click Through Rate*, *Random forest*

**Corresponding author: dewiptr098@gmail.com*

1. Pendahuluan

Dalam era saat ini, iklan digital telah menjadi komponen utama dalam strategi pemasaran banyak perusahaan. Iklan digital mampu menjangkau *audiens* secara luas dan tertarget, sehingga menjadi salah satu alat yang efektif untuk meningkatkan penjualan [1]. Namun, untuk mengoptimalkan efektivitasnya, diperlukan pemahaman yang lebih mendalam terhadap pola perilaku pelanggan [2]. Salah satu metrik yang sering digunakan dalam evaluasi performa iklan digital adalah *Click Through Rate* (CTR), yaitu persentase jumlah klik dibandingkan dengan jumlah tayangan iklan [3]. CTR memberikan gambaran awal tentang tingkat ketertarikan pengguna terhadap suatu iklan, namun metrik ini tidak selalu mencerminkan kemungkinan terjadinya pembelian [2]. Beberapa pengguna mungkin hanya mengklik iklan karena rasa penasaran tanpa niat untuk membeli. Oleh karena itu, diperlukan analisis lanjutan untuk memprediksi tindakan pengguna setelah melakukan klik. Salah satu pendekatan yang dapat digunakan adalah prediksi pembelian berdasarkan CTR, yang bertujuan untuk mengidentifikasi karakteristik pengguna yang tidak hanya tertarik tetapi juga memiliki potensi tinggi untuk melakukan transaksi [4].

Prediksi pembelian dilakukan dengan pendekatan data mining, yaitu proses penggalian informasi atau pola tersembunyi dari kumpulan data dalam jumlah besar [5]. Salah satu teknik dalam data mining yang relevan untuk tujuan tersebut adalah klasifikasi, yaitu proses pengelompokkan data ke dalam kategori tertentu berdasarkan fitur-fitur yang ada dalam data [6]. Proses data mining tersebut umumnya mengikuti tahapan dalam *Cross Industry Standard Process for Data Mining* (CRISP-DM), sebuah kerangka kerja standar industri yang mencakup enam fase utama: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment*. Pendekatan ini membantu peneliti dalam merancang solusi data mining secara sistematis agar hasilnya sesuai dengan kebutuhan bisnis yang ingin dicapai [7], [8]. Dalam konteks penelitian ini, CRISP-DM digunakan sebagai landasan metodologis untuk membangun sistem prediksi pembelian berdasarkan data iklan digital. Dalam hal ini, klasifikasi diterapkan untuk membedakan antara pengguna yang melakukan pembelian dan yang tidak setelah melakukan klik iklan, berdasarkan fitur-fitur tertentu yang tersedia dalam dataset.

Dalam penerapan strategi pemasaran iklan digital, diperlukan model prediksi yang memiliki nilai akurasi yang tinggi untuk membantu perusahaan dalam memfokuskan alokasi sumber daya pada kelompok pelanggan yang memiliki potensi konversi lebih besar. Upaya tersebut bertujuan untuk mengoptimalkan efektivitas anggaran iklan serta mendukung proses pengambilan keputusan strategis secara lebih efisien [4]. Penggunaan model prediktif memungkinkan perusahaan untuk mengidentifikasi pola perilaku pengguna yang tidak mudah diamati secara langsung, sehingga dapat memberikan wawasan baru dalam merumuskan pendekatan pemasaran yang lebih adaptif [9].

Dalam penelitian ini, algoritma *Random forest* digunakan untuk membangun model prediksi kemungkinan pembelian setelah pengguna melakukan klik pada iklan digital. Algoritma tersebut dipilih karena kemampuannya dalam menghasilkan prediksi yang stabil serta keandalannya dalam menangani variabilitas data [10]. *Random forest* bekerja dengan menggabungkan sejumlah pohon keputusan (*decision trees*) dan melakukan agregasi terhadap hasilnya, sehingga dapat meminimalkan risiko *overfitting* serta meningkatkan generalisasi model. Selain itu, *Random forest* memberikan hasil prediksi yang akurat dan stabil pada dataset berukuran kecil hingga sedang [11]. Selanjutnya, dalam *Random forest* terdapat fungsi parameter, salah satunya adalah *classweight*, yang digunakan untuk mengatasi permasalahan distribusi kelas yang tidak seimbang pada data [10]. Ketidakseimbangan data adalah ketimpangan antara jumlah pengguna yang melakukan pembelian dan yang tidak, dan hal tersebut

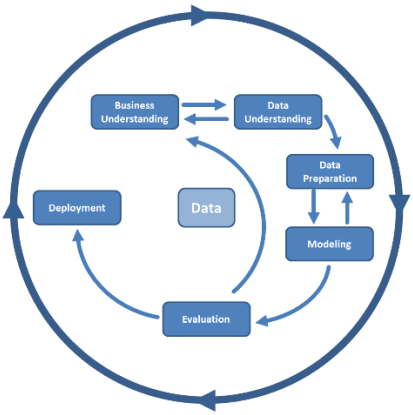
dapat menyebabkan bias dalam proses pembelajaran model [12]. Oleh karena itu, pengaturan bobot pada masing-masing kelas dilakukan agar model dapat memberikan perhatian lebih terhadap kelas minoritas. Sehingga diharapkan dapat meningkatkan akurasi prediksi serta menghasilkan output yang lebih adil dan representatif. Terdapat fungsi lebih lanjut, yaitu proses *tuning hyperparameter*, guna memperoleh performa model yang optimal. Proses tersebut mencakup penyesuaian parameter seperti jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), minimum sampel untuk pembelahan (*min_samples_split*), dan minimum sampel dalam daun (*min_samples_leaf*), serta strategi pembobotan kelas [13]. Dengan mengoptimalkan kombinasi parameter tersebut, diharapkan model yang dibangun dapat mencapai kinerja prediktif yang lebih baik dalam klasifikasi perilaku pembelian.

Penelitian sebelumnya yang menggunakan dataset *Social Network Ads* dari *Kaggle* telah menunjukkan potensi penggunaan algoritma *machine learning* dalam mengklasifikasikan perilaku pengguna terhadap iklan digital. Salah satu studi oleh Venkata tahun 2023 mengimplementasikan algoritma *Decision Tree* dan menghasilkan akurasi sebesar 82% dengan nilai presisi 80%. Hasil ini menunjukkan bahwa metode pohon keputusan mampu memberikan performa yang cukup baik dalam konteks prediksi pembelian berdasarkan klik iklan [14]. Namun, tingkat akurasi tersebut masih menyisakan ruang untuk peningkatan, khususnya dalam hal kestabilan prediksi dan kemampuan generalisasi terhadap data baru. Sementara itu, penelitian lain oleh Handayani tahun 2023 memanfaatkan algoritma *Light Gradient Boosting* (LGBM) dengan tambahan teknik *resampling* seperti *Random Over Sampling (ROS)*, *Random Under Sampling (RUS)*, *SMOTE*, dan *SMOTE-Tomek*. Dengan pendekatan tersebut, penelitian tersebut mampu mencapai akurasi sebesar 91,25%, presisi 84,38%, dan nilai *AUC* sebesar 0,92. Hasil tersebut memperlihatkan bahwa LGBM yang dikombinasikan dengan teknik penyeimbangan data mampu memberikan hasil yang kompetitif dalam klasifikasi pengguna iklan digital [15].

Meskipun demikian, pendekatan alternatif tetap diperlukan untuk mengevaluasi stabilitas performa model pada berbagai skenario dan ukuran dataset. Oleh karena itu, penelitian ini mengusulkan penggunaan algoritma *Random forest* yang dikenal memiliki ketahanan terhadap *overfitting*, kestabilan prediksi, serta fleksibilitas dalam menangani berbagai jenis dan ukuran data. Dengan dukungan metodologi CRISP-DM, algoritma ini diharapkan mampu menghasilkan model klasifikasi yang efektif dalam memprediksi kemungkinan konversi pengguna setelah mereka berinteraksi dengan iklan digital.

2. Metode

Gambaran tahapan penelitian yang dilakukan berdasarkan metode CRISP-DM dapat dilihat pada Gambar 1. Terlihat bahwa tahapan – tahapan utamanya adalah antara lain:



Gambar 1: Metode CRISP-DM

2.1. *Business Understanding*

Tahapan ini bertujuan untuk merumuskan tujuan bisnis, yaitu memprediksi kemungkinan pembelian berdasarkan metrik CTR untuk meningkatkan efektivitas kampanye iklan digital. Dengan memahami karakteristik pengguna yang berpotensi melakukan pembelian, perusahaan dapat mengalokasikan anggaran pemasaran secara lebih efisien.

2.2. *Data Understanding*

Data yang digunakan dalam penelitian ini berasal dari dataset *Social Network Ads* yang tersedia di platform *Kaggle*. Dataset tersebut terdiri dari 400 entri dan memuat atribut seperti usia, jenis kelamin, estimasi pendapatan, serta label pembelian. Distribusi data menunjukkan adanya ketidakseimbangan kelas, di mana jumlah pengguna yang tidak melakukan pembelian lebih banyak dibandingkan yang melakukan pembelian setelah mengklik iklan.

2.3. *Data Preparation*

Tahapan ini mencakup sejumlah proses pra-pemrosesan data, seperti encoding variabel kategorial, penskalaan fitur numerik agar memiliki skala yang sebanding, pembagian dataset menjadi data latih dan data uji, serta pemeriksaan dan penanganan potensi anomali atau inkonsistensi dalam label target.

2.4. *Modeling*

Tahapan ini merupakan proses pembangunan model klasifikasi menggunakan algoritma *Random forest*, yang merupakan metode ensemble berbasis decision tree. Untuk membangun *Random forest*, dilakukan proses sebagai berikut:

2.1.1. *Bootstrap* sampling adalah proses dibuat beberapa subset data latih secara acak menggunakan teknik *sampling with replacement*.

2.1.2. Pembangunan pohon keputusan untuk setiap subset. Pada setiap *node*, terpilih fitur terbaik berdasarkan pengukuran ketidakmurnian (*impurity*) menggunakan *Gini Index* atau *Entropy*.

Gini Index dirumuskan pada Persamaan (1).

$$Gini(S) = 1 - \sum_{i=1}^{c} p_i^2 \quad (1)$$

Sedangkan *Entropy* dirumuskan pada Persamaan (2).

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \quad (2)$$

Untuk menentukan atribut mana yang paling optimal dalam membagi data, digunakan metrik *Information Gain*, yaitu selisih antara entropy sebelum dan sesudah pemisahan data berdasarkan suatu atribut. Rumus *Information Gain* dirumuskan pada Persamaan (3).

$$IG(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|}.Entropy(S_v) \quad (3)$$

Setelah semua pohon dibangun, masing-masing pohon menghasilkan prediksi terhadap kelas target. Pada tahap akhir, prediksi gabungan dari seluruh pohon diputuskan berdasarkan suara terbanyak melalui metode *Majority Voting*, yang dirumuskan pada Persamaan (4).

$$y = mode\{h_1(x), h_2(x), \dots, h_n(x)\} \quad (4)$$

Dengan pendekatan ini, *Random forest* mampu menghasilkan model prediktif yang stabil, akurat, dan lebih tahan terhadap *overfitting* dibandingkan dengan pohon keputusan tunggal. Selanjutnya, dilakukan proses *tuning hyperparameter* menggunakan teknik *Randomized Search* untuk menemukan kombinasi parameter terbaik seperti jumlah pohon (*n_estimators*), kedalaman pohon maksimum (*max_depth*), minimal sampel untuk split (*min_samples_split*), minimum sampel dalam daun (*min_samples_leaf*), dan parameter pembobotan kelas (*class_weight*).

2.5. Evaluation

Model yang telah dibangun dievaluasi menggunakan dua tahap, yaitu evaluasi performa pada data latih dengan teknik *cross-validation*, kemudian dievaluasi kembali pada data uji. Pengukuran performa dilakukan dengan metrik klasifikasi yaitu akurasi, presisi, *recall*, *F1-score*, dan *ROC-AUC*. Nilai hasil model dibandingkan dengan model *baseline* untuk mengukur adanya peningkatan performa.

2.6. *Deployment*

Tahap *deployment* dalam penelitian ini dilakukan secara konseptual sebagai simulasi integrasi model ke dalam sistem nyata. Meskipun belum diimplementasikan langsung, disusun rencana penerapan model ke dalam sistem rekomendasi atau dashboard analitik, termasuk strategi pemantauan performa jika dikembangkan lebih lanjut.

3. Hasil dan Pembahasan

3.1. Hasil

3.1.1. *Data Understanding*

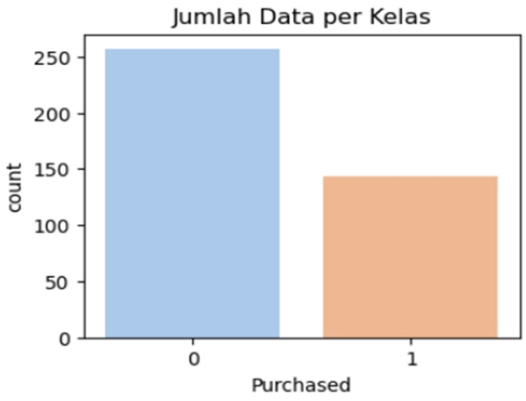
Dalam tahap ini dilakukan pemahaman terhadap data yang digunakan dalam penelitian. Dataset yang digunakan merupakan dataset publik dari *Kaggle* yang beralamat pada https://www.*Kaggle*.com/datasets/rakeshrau/social-network-ads. Berikut adalah deskripsi data dari dataset *Social Network Ads*.

Tabel 1: Deskripsi Data

No	Variabel	Tipe Data	Deskripsi
1	*User ID*	Integer	Identifikasi unik setiap pengguna
2	*Gender*	Objek	Jenis kelamin pengguna
3	*Age*	Integer	Usia pengguna dalam tahun
4	*EstimatedSalary*	Integer	Estimasi gaji tahunan dalam mata uang tertentu
5	*Purchased*	Binominal (0/1)	Label target: 1 Pengguna melakukan pembelian 0 Pengguna tidak melakukan pembelian

Berdasarkan Tabel 1, Deskripsi Dataset, data terdiri atas variabel numerik seperti *Age* dan *EstimatedSalary*, serta variabel kategorial yaitu *Gender*. Sementara itu, variabel *Purchased* berperan sebagai label target dalam klasifikasi biner, di mana nilai 1 menunjukkan bahwa pengguna melakukan pembelian, dan nilai 0 menunjukkan tidak melakukan pembelian. Karena struktur datanya cukup sederhana, dataset ini cocok digunakan untuk mengidentifikasi pola perilaku pembelian pengguna berdasarkan informasi dasar seperti usia, jenis kelamin, dan estimasi pendapatan.

Selanjutnya dilakukan eksplorasi data untuk memahami karakteristik dari dataset secara lebih menyeluruh serta menilai pola distribusi data dan mendeteksi kemungkinan ketidakseimbangan kelas. Berdasarkan hasil eksplorasi, diketahui bahwa distribusi kelas pada variabel *Purchased* tidak seimbang. Berikut adalah visualisasi hasil eksplorasi dari distribusi kelas.



Gambar 2: Distribusi Pembelian

Berdasarkan Gambar 2, dapat dilihat bahwa jumlah antara pengguna yang tidak melakukan pembelian (kelas 0) lebih banyak yaitu berjumlah 257 dibandingkan dengan pengguna yang melakukan pembelian (kelas 1) hanya berjumlah 143. Kondisi tersebut perlu diperhatikan dalam proses pemodelan agar model tidak bias pada kelas mayoritas.

3.1.2. *Data Preparation*

Pada tahap *Data Preparation* dilakukan pengecekan data hilang (*missing value*) berdasarkan hasil tidak ditemukan nilai yang hilang sehingga tidak diperlukan tindakan lebih lanjut. Selanjutnya, dilakukan proses pengecekan nilai *outlier* berdasarkan hasil tidak ditemukan adanya *outlier* pada kedua fitur numerik maka tidak diperlukan proses penanganan data *outlier*. Selain memeriksa *outlier* tahap *Data Preparation* juga mencakup deteksi inkonsistensi label, yaitu kondisi ketika terdapat dua data dengan nilai fitur yang sama namun label targetnya berbeda. Dalam tahap tersebut ditemukan satu kasus inkonsistensi label yang ditunjukkan pada Tabel 2.

Tabel 2: Inkonsistensi Label

User ID	*Gender*	*Age*	*EstimatedSalary*	*Purchased*
15789109	*female*	32	117000	0
15694288	*female*	32	117000	1

Seperti yang ditunjukkan pada Tabel 2, terdapat dua entri data dengan kombinasi atribut yang identik (*Gender* = 0, *Age* = 32, *EstimatedSalary* = 117000), namun memiliki label *Purchased* yang berbeda, yaitu 0 dan 1. Ketidakkonsistenan ini dapat menyebabkan ambiguitas dalam proses pelatihan model, karena satu set fitur menghasilkan dua target klasifikasi yang bertentangan. Untuk menghindari potensi gangguan terhadap kinerja algoritma, salah satu entri dihapus, yakni entri dengan label *Purchased* = 0. Pemilihan untuk mempertahankan label 1 dilakukan karena data pembelian dinilai lebih

relevan untuk dianalisis dalam konteks prediksi konversi pada iklan digital, serta sebagai langkah untuk membantu menyeimbangkan distribusi kelas.

Lalu dilakukan transformasi fitur kategorial yaitu fitur *Gender* agar dapat digunakan dalam algoritma *machine learning* yang hanya menerima *input* numerik. Pada Tabel 3, merupakan hasil dari transformasi fitur kategorial.

Tabel 3: Transformasi Fitur Kategorial

Sebelum Transformasi	Sesudah Transformasi
male	1
male	1
female	0
female	0
male	1

Setelah proses transformasi variabel kategorial dilakukan, atribut *Gender* berhasil dikonversi ke dalam bentuk numerik. Nilai "*Male*" direpresentasikan sebagai 1 dan "*Female*" sebagai 0, sebagaimana ditunjukkan pada Tabel 3. Dengan konversi ini, variabel *Gender* telah siap digunakan sebagai fitur input dalam pemodelan menggunakan algoritma *Random forest*.

Selanjutnya dilakukan pembagian data menjadi dua bagian yaitu data latih dan data uji. Dalam penelitian ini digunakan rasio 70:30 yaitu data latih sebesar 70% dan data uji sebesar 30%. Serta dilakukan standarisasi fitur numerik agar menyelaraskan skala antar fitur, sehingga tidak ada satu fitur yang mendominasi proses pembelajaran model akibat perbedaan rentang nilai yang dimiliki.

3.1.3. *Modeling*

Setelah dilakukan analisis awal pada *Data Understanding* dan *Data Preparation* akan dilanjutkan penerapan model prediksi menggunakan algoritma *Random forest*. Algoritma *Random forest* bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dari sampel data yang berbeda melalui teknik *bootstrap* aggregation, kemudian menggabungkan hasil prediksi dari masing - masing pohon melalui proses voting mayoritas.

a. Model Awal

Model awal dibangun menggunakan parameter default dari algoritma *Random forest*. Tujuan dari pembuatan model *baseline* adalah untuk memberikan tolak ukur awal yang dapat dibandingkan dengan model hasil *tuning*. Pada tahap ini, model dilatih menggunakan data latih yang telah dipersiapkan tanpa melakukan penyesuaian parameter secara eksplisit. Model awal memberikan gambaran awal mengenai performa algoritma sebelum dilakukan optimasi lebih lanjut.

b. *Tuning Hyperparameter*

Untuk meningkatkan kinerja prediktif model, dilakukan *tuning* terhadap *hyperparameter Random forest*. Hasil *tuning hyperparameter* menggunakan random search, algoritma *Random forest* diatur dengan parameter optimal sebagai berikut:

1) *n_estimators* = 100
2) *max_depth* = 10
3) *min_samples_split* = 2
4) *min_saples_leaf* = 4
5) *bootstrap* = True
6) *class_weight* = 'balanced'

Parameter-parameter di atas digunakan dalam proses pembuatan model *Random forest*. "*n_estimators* = 100" berarti model terdiri dari 100 pohon keputusan. "*max_dept* = 10" membatasi kedalaman maksimum setiap pohon untuk mencegah *overfitting*. "*min_samples_split* = 2" menunjukkan jumlah minimum sampel yang diperlukan untuk membagi sebuah node, sementara "*min_samples_leaf*

= 4" memastikan setiap daun pada pohon memiliki setidaknya 4 sampel. Pengaturan "*bootstrap* = True" menunjukkan bahwa teknik *bootstrap* (pengambilan sampel dengan pengembalian) digunakan dalam pelatihan pohon. Terakhir, "*class_weight* = '*balanced*'" digunakan untuk menangani ketidakseimbangan kelas dengan secara otomatis menyesuaikan bobot kelas yang minoritas agar model tidak bias terhadap kelas mayoritas.

3.1.4. Evaluation

Setelah pemodelan dilakukan dengan dua pendekatan, yaitu model *baseline* dan model yang telah di*tuning*, langkah berikutnya adalah mengevaluasi performa masing-masing model. Evaluasi ini bertujuan untuk menilai kemampuan model dalam memprediksi kelas target secara akurat, serta memastikan bahwa model tidak hanya berkinerja baik pada data pelatihan, tetapi juga dapat menggeneralisasi dengan baik terhadap data uji yang belum pernah digunakan sebelumnya.

Evaluasi dilakukan dalam dua tahap, yaitu validasi silang pada data latih dan evaluasi akhir pada data uji. Tahap pertama menggunakan teknik *5-Fold Cross-Validation* untuk mengukur kestabilan model. Pada Tabel 4 menunjukkan hasil dari validasi silang.

Tabel 4: Hasil Validasi Silang

Iterasi	*Baseline*	*Tuning*
0	0.89	0.94
1	0.92	0.87
2	0.85	0.89
3	0.82	0.92
4	0.89	0.87
Rata – rata	0.87	0.90

Berdasarkan Tabel 4, terlihat bahwa model hasil *tuning* secara umum menunjukkan performa yang lebih baik dibanding *baseline*, terutama dalam hal konsistensi dan stabilitas akurasi. Tahap selanjutnya adalah evaluasi akhir menggunakan data uji yang tidak digunakan dalam proses pelatihan. Metrik yang digunakan meliputi akurasi, presisi, *recall*, *f1-score*, dan *ROC AUC*. Berikut adalah hasil dari evaluasi akhir.

Tabel 5: Hasil Akhir Evaluasi

Metrik	*Baseline*	*Tuning*
Accuracy	0.90	0.93
Precision	0.87	0.86
Recall	0.87	0.98
F1-score	0.87	0.92
AUC Score	0.96	0.97

Berdasarkan Tabel 5 model hasil *tuning* menunjukkan peningkatan signifikan pada *recall* dan *f1-score*, hal tersebut menunjukkan kemampuan yang lebih baik dalam mengenali kelas positif (pembelian). Meskipun nilai presisi sedikit menurun, peningkatan pada metrik lainnya menunjukkan bahwa *tuning* berhasil meningkatkan performa model secara keseluruhan khususnya dalam konteks prediksi yang menekankan deteksi kelas minoritas.

3.1.5. *Deployment*

Tahap *deployment* dalam penelitian ini difokuskan pada pengembangan model prediksi tanpa dilanjutkan ke tahap implementasi dalam sistem atau aplikasi nyata. Oleh karena itu, pemanfaatan model hanya dijelaskan secara konseptual sebagai simulasi bagaimana model dapat diintegrasikan ke dalam sistem rekomendasi atau dashboard analitik. Selain itu, turut disadari pentingnya pemantauan performa model jika kelak diterapkan secara riil, termasuk kemungkinan retraining ketika terjadi perubahan pola data. Hasil penelitian ini disusun secara lengkap dalam bentuk laporan akademik yang

mencakup seluruh proses pengembangan model serta potensi penerapannya sebagai referensi pengembangan lanjutan.

3.2. Pembahasan

Setelah dilakukan evaluasi terhadap model prediksi yang dikembangkan dalam penelitian, langkah selanjutnya adalah membandingkan hasil performa model dengan penelitian terdahulu yang menggunakan algoritma berbeda dan dataset serupa. Tujuan dari perbandingan ini adalah untuk menilai keunggulan pendekatan yang digunakan, serta melihat posisi model dalam konteks penelitian sebelumnya. Perbandingan dilakukan berdasarkan nilai akurasi sebagai metrik utama yang ditampilkan dalam bentuk Tabel 6 yang merupakan hasil perbandingan nilai akurasi.

Tabel 6: Perbandingan Hasil Akurasi

No	Peneliti & Tahun	Algoritma	Dataset	Akurasi	Keterangan
1	Venkata (2023)	*Decision Tree*	*Social Network Ads*	82%	Tanpa optimasi, hanya *baseline* model
2	Kartika Handayani (2023)	*Light Gradient Boosting*	*Social Network Ads*	91%	*Resampling*
3	Penelitian ini	*Random forest*	*Social Network Ads*	93%	Model setelah *tuning hyperparameter* dan *classweight*

Berdasarkan Tabel 6, dapat dilihat bahwa model *Random forest* yang dikembangkan dalam penelitian ini berhasil memperoleh akurasi sebesar 93%, yang merupakan nilai tertinggi dibandingkan dua penelitian terdahulu dengan dataset yang sama, yaitu *Social Network Ads*. Penelitian oleh Venkata (2023) menggunakan algoritma *Decision Tree* tanpa optimasi lanjutan, hanya pada *baseline* model, dan memperoleh akurasi sebesar 82%. Sementara itu, Kartika Handayani (2023) menggunakan algoritma *Light Gradient Boosting* (LGBM) dengan pendekatan *resampling* untuk mengatasi ketidakseimbangan kelas, dan memperoleh akurasi sebesar 91%.

Hasil ini menunjukkan bahwa penerapan *Random forest* yang disertai dengan *tuning hyperparameter* dan penyesuaian *classweight* mampu meningkatkan performa model secara signifikan. Meskipun LGBM secara umum dikenal sebagai algoritma yang unggul dalam banyak kompetisi prediksi, hasil ini mengindikasikan bahwa dengan strategi pemodelan dan penanganan data yang tepat, algoritma *Random forest* tetap dapat menghasilkan kinerja yang kompetitif, bahkan lebih baik dalam konteks tertentu.

4. Kesimpulan

Penelitian ini menunjukkan bahwa algoritma *Random forest* efektif digunakan dalam memprediksi perilaku pembelian pengguna berdasarkan data iklan digital. Model yang dibangun melalui pendekatan CRISP-DM dan disertai dengan proses *tuning hyperparameter* serta penyesuaian *classweight* berhasil mencapai performa tinggi, dengan akurasi sebesar 93%, *recall* 98%, *precision* 86%, dan *F1-score* 92%. Hasil ini menunjukkan bahwa model tidak hanya akurat secara umum, tetapi juga mampu mengenali kelas minoritas dengan sangat baik. Dibandingkan dengan penelitian terdahulu yang menggunakan algoritma *Decision Tree* dan *LightGBM*, model dalam penelitian ini menunjukkan peningkatan performa yang signifikan. Meskipun belum diimplementasikan secara langsung ke dalam sistem, hasil penelitian ini dapat menjadi dasar pengembangan sistem rekomendasi atau dashboard analitik untuk mendukung strategi pemasaran digital. Penelitian lanjutan dapat dilakukan dengan memperluas fitur data atau mengeksplorasi algoritma lain untuk meningkatkan generalisasi model.

Pustaka

[1] I.V. Dwaraka Srihith, T. Aditya Sai Srinivas, K. Owdharya, A. David Donald, and G. Thippana, "Clicks and Conversions: Unveiling the Power of Machine Learning in Social Media Ad Classification," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 1–6, May 2023, doi: 10.48175/ijarsct-11401.

[2] X. Zhou and S. Richards, "Predicting Click Behavior Based on Machine Learning Models," 2023.

[3] Y. Yang and P. Zhai, "Click-through rate prediction in online advertising: A literature review," *Inf Process Manag*, vol. 59, no. 2, Mar. 2022, doi: 10.1016/j.ipm.2021.102853.

[4] A. Lakshmanarao, A. Srisaila, and T. S. R. Kiran, "An Efficient Ad Click Prediction System using Machine Learning Techniques," *Int J Eng Adv Technol*, vol. 9, no. 3, pp. 1269–1272, Feb. 2020, doi: 10.35940/ijeat.C5518.029320.

[5] A. Nugroho, A. Husin, J. Provinsi, T. Hulu, and I. Hilir Indonesia, "SISTEMASI: Jurnal Sistem Informasi Analisis Performa *Random forest* Menggunakan Normalisasi Atribut Performance Analysis of *Random forest* Using Attribute Normalization." [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[6] R. Putra Primanda, A. Alwi, and D. Mustikasari, "url: http://studentjournal.umpo.ac.id/index.php/komputek DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri)," 2021. [Online]. Available: http://studentjournal.umpo.ac.id/index.php/komputek

[7] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," 2021. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[8] D. Feblian and D. U. Daihani, "IMPLEMENTASI MODEL CRISP-DM UNTUK MENENTUKAN SALES PIPELINE PADA PT X," *JURNAL TEKNIK INDUSTRI*, vol. 6, no. 1, Feb. 2017, doi: 10.25105/jti.v6i1.1526.

[9] S. Saraswathi, V. Krishnamurthy, D. Venkata Vara Prasad, R. K. Tarun, S. Abhinav, and D. Rushitaa, "Machine learning based ad-click prediction system," *Int J Eng Adv Technol*, vol. 8, no. 6, pp. 3646–3648, Aug. 2019, doi: 10.35940/ijeat.F9366.088619.

[10] A. Arisusanto, N. Suarna, and G. Dwilestari, "Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma *Random forest* Dengan Optimize Parameter Grid," *Jurnal Teknologi Ilmu Komputer*, vol. 1, no. 2, pp. 43–47, 2023, doi: 10.56854/jtik.v1i2.51.

[11] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi pada *Random forest* untuk Klasifikasi Data Bank Marketing," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 187–192, Feb. 2021, doi: 10.29207/resti.v5i1.2813.

[12] A. P. P. Wardani, A. Adiwijaya, and M. D. Purbolaksono, "Sentiment Analysis on Beauty Product Review Using Modified Balanced *Random forest* Method and Chi-Square," *Journal of Information System Research (JOSH)*, vol. 4, no. 1, pp. 1–7, Oct. 2022, doi: 10.47065/josh.v4i1.2047.

[13] D. A. Anggoro and S. S. Mukti, "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*, vol. 14, no. 6, pp. 198–207, Dec. 2021, doi: 10.22266/ijies2021.1231.19.

[14] I. Venkata and D. Srihith, "Predicting Success: The Impact of Machine Learning on Social Media Ad Classification," 2023, doi: 10.5281/zenodo.8214337.

[15] K. Handayani, "PENERAPAN LIGHT GRADIENT BOOSTING DALAM PREDIKSI RASIO KLIK TAYANG," 2023.