

## **PROPOSAL TUGAS UTS**

### **Optimasi Prediksi Niat Pembelian Pengunjung E-Commerce Menggunakan Random Forest dan Hyperparameter Tuning (Grid Search)**



**DISUSUN OLEH**

**Shanaya Balghis Riyona (241552010012)  
Thoriqurrahman Akrami (241552010014)**

**DOSEN PENGAMPU  
HENDRI KHARISMA, S.Kom., M.T.**

**Fakultas Sains dan Teknologi Jurusan Teknik Informatika  
Sekolah Tinggi Manajemen Informatika dan Komputer Tazkia. Raya Dramaga  
Blok Radar Baru No.8, RT.03/RW.03, Margajaya, Kec. Bogor Bar., Kota Bogor,  
Jawa Barat 16116, Indonesia**

## **Kata Pengantar**

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, taufik, dan karunia-Nya. Berkat pertolongan-Nya, penulis akhirnya dapat menyelesaikan penyusunan Laporan Pengajuan Judul Penelitian ini dengan baik dan lancar.

Laporan ini disusun sebagai salah satu tahap awal dalam proses penyusunan jurnal penelitian pada Program Studi Teknik Informatika. Melalui laporan ini, penulis berusaha memberikan gambaran umum mengenai topik penelitian yang diajukan, yang mencakup latar belakang, tujuan, manfaat, batasan masalah, serta metodologi penelitian yang akan digunakan.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna, baik dari segi isi maupun penyajiannya. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari dosen pembimbing maupun rekan-rekan agar laporan ini dapat menjadi lebih baik di kemudian hari.

Akhir kata, penulis berharap semoga laporan ini dapat memberikan manfaat, baik bagi penulis sendiri sebagai pengalaman akademik, maupun bagi pihak lain yang memiliki minat pada bidang penelitian ini. Semoga laporan ini juga dapat menjadi langkah awal yang baik menuju penyusunan penelitian yang sesungguhnya.

Bogor, 24 Oktober 2025

Penulis

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi yang semakin pesat telah membawa perubahan besar dalam berbagai aspek kehidupan, termasuk dalam bidang perdagangan. Saat ini, aktivitas jual beli tidak lagi terbatas pada transaksi langsung di toko fisik, melainkan telah bergeser ke arah perdagangan elektronik (e-commerce). Platform e-commerce memungkinkan pengguna untuk melakukan transaksi secara cepat, mudah, dan efisien hanya melalui perangkat digital seperti komputer maupun smartphone.

Peningkatan jumlah pengguna internet yang signifikan mendorong pertumbuhan pesat industri e-commerce, baik di dunia maupun di Indonesia. Namun, di balik tingginya jumlah pengunjung situs e-commerce, tidak semua pengunjung melakukan transaksi pembelian. Sebagian besar hanya melakukan penelusuran produk tanpa menyelesaikan proses pembayaran. Kondisi ini menimbulkan tantangan bagi pengelola e-commerce untuk memahami faktor-faktor yang memengaruhi niat pembelian (*purchasing intention*) dan bagaimana memprediksinya secara akurat.

Untuk menjawab tantangan tersebut, berbagai penelitian telah dilakukan dengan memanfaatkan Machine Learning guna memprediksi niat pembelian berdasarkan perilaku pengunjung situs. Salah satu dataset yang sering digunakan adalah Online Shoppers Purchasing Intention Dataset yang tersedia di UCI Machine Learning Repository. Dataset ini berisi data perilaku pengguna, seperti durasi kunjungan halaman, bounce rate, exit rate, serta atribut waktu dan jenis pengunjung.

Penelitian sebelumnya oleh Cemal Okan Sakar dkk. (2019) mengusulkan pendekatan berbasis Multilayer Perceptron (MLP) dan Long Short-Term Memory (LSTM) untuk memprediksi niat pembelian pengunjung secara real-time. Hasilnya menunjukkan bahwa pendekatan berbasis jaringan saraf mampu memberikan akurasi yang baik dalam klasifikasi niat pembelian. Meskipun demikian, penelitian tersebut belum mengeksplorasi secara mendalam potensi dari algoritma Random Forest, yang dikenal memiliki performa tinggi dalam menangani data kompleks serta mampu menghindari overfitting.

Algoritma Random Forest bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dan menggabungkan hasil prediksinya untuk mendapatkan keputusan akhir yang lebih stabil dan akurat. Namun, performa optimal Random Forest sangat bergantung pada pengaturan parameter seperti jumlah pohon (*n\_estimators*), kedalaman maksimum pohon (*max\_depth*), dan jumlah fitur yang dipertimbangkan pada setiap pembagian (*max\_features*). Oleh karena itu, diperlukan proses Hyperparameter Tuning untuk menentukan kombinasi parameter terbaik. Salah satu metode yang umum digunakan adalah Grid Search, yaitu teknik pencarian sistematis terhadap seluruh kombinasi parameter untuk menemukan konfigurasi dengan performa tertinggi.

Berdasarkan hal tersebut, penelitian ini berfokus pada optimasi prediksi niat pembelian pengunjung e-commerce menggunakan algoritma Random Forest yang dioptimasi melalui metode Grid Search. Dengan pendekatan ini, diharapkan model yang dihasilkan mampu memberikan akurasi lebih tinggi dibandingkan model dengan parameter default, serta dapat membantu pengelola e-commerce dalam memahami perilaku pengguna dan meningkatkan tingkat konversi penjualan

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini dapat disusun sebagai berikut:

1. Bagaimana cara mengimplementasikan algoritma Random Forest untuk memprediksi niat pembelian pengunjung e-commerce berdasarkan Online Shoppers Purchasing Intention Dataset?
2. Bagaimana pengaruh Hyperparameter Tuning menggunakan Grid Search terhadap peningkatan performa model Random Forest dalam memprediksi niat pembelian pengunjung e-commerce?
3. Seberapa besar peningkatan akurasi, presisi, recall, dan F1-score yang diperoleh setelah dilakukan optimasi parameter dibandingkan dengan model tanpa tuning?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengimplementasikan algoritma Random Forest dalam memprediksi niat pembelian pengunjung e-commerce menggunakan data dari *Online Shoppers Purchasing Intention Dataset*.
2. Untuk melakukan proses optimasi hyperparameter pada algoritma Random Forest dengan metode Grid Search, guna memperoleh kombinasi parameter terbaik.
3. Untuk menganalisis dan membandingkan performa model sebelum dan sesudah dilakukan tuning menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.
4. Untuk menghasilkan model prediksi yang optimal yang dapat membantu pelaku bisnis e-commerce dalam memahami perilaku pengguna serta meningkatkan potensi konversi penjualan.

## 1.3 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis, yaitu sebagai berikut:

### 1. Manfaat Teoritis

- Memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang Machine

Learning, khususnya dalam penerapan algoritma Random Forest dan teknik Hyperparameter Tuning (Grid Search) untuk prediksi perilaku pengguna pada platform e-commerce.

- Menjadi referensi bagi penelitian selanjutnya yang ingin mengkaji lebih dalam tentang optimasi model klasifikasi pada dataset perilaku pengguna atau bidang serupa.
- Menunjukkan bagaimana metode tuning parameter dapat memengaruhi kinerja model secara signifikan dalam konteks prediksi niat pembelian.

## 2. Manfaat Praktis

- Membantu pelaku bisnis atau pengelola situs e-commerce dalam mengidentifikasi calon pembeli potensial berdasarkan perilaku kunjungan pengguna secara lebih akurat.
- Memberikan dasar pengambilan keputusan untuk strategi pemasaran yang lebih efektif, seperti personalisasi rekomendasi produk, promosi, atau strategi retargeting.
- Menyediakan model prediksi yang efisien dan dapat diimplementasikan pada sistem e-commerce nyata untuk meningkatkan tingkat konversi penjualan.

### 1.4 Batasan Penelitian

Agar penelitian ini lebih terarah dan fokus pada tujuan yang ingin dicapai, maka ruang lingkup penelitian dibatasi oleh beberapa hal sebagai berikut:

1. Dataset yang digunakan adalah *Online Shoppers Purchasing Intention Dataset* yang diperoleh dari UCI Machine Learning Repository, yang berisi data perilaku pengunjung situs e-commerce seperti durasi kunjungan, rasio keluar (*exit rate*), rasio pantulan (*bounce rate*), serta atribut waktu dan jenis pengunjung.
2. Penelitian ini hanya menggunakan algoritma Random Forest sebagai model utama untuk prediksi niat pembelian, tanpa membandingkannya dengan algoritma lain seperti SVM, MLP, atau LSTM.
3. Proses optimasi parameter terbatas pada penggunaan metode Grid Search untuk menentukan kombinasi hyperparameter terbaik, seperti *n\_estimators*, *max\_depth*, dan *max\_features*.
4. Evaluasi performa model dilakukan menggunakan metrik klasifikasi, yaitu akurasi, presisi, recall, dan F1-score, tanpa mempertimbangkan metrik lain seperti ROC-AUC atau waktu komputasi secara mendalam.
5. Penelitian ini tidak membahas aspek implementasi sistem real-time, melainkan berfokus pada tahap analisis dan optimasi model prediksi menggunakan data historis.

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Landasan Teori

##### 2.1.1 E-Commerce

E-commerce atau *electronic commerce* merupakan aktivitas jual beli barang dan jasa melalui media elektronik seperti internet. Perkembangan e-commerce sangat pesat seiring meningkatnya penggunaan teknologi informasi dan kemudahan akses internet. Platform seperti Tokopedia, Shopee, Lazada, dan Bukalapak telah menjadi sarana utama bagi masyarakat untuk bertransaksi secara digital.

Dalam konteks penelitian ini, e-commerce menjadi objek utama karena sistemnya menghasilkan berbagai data perilaku pengguna, seperti jumlah kunjungan, durasi interaksi, serta aktivitas klik, yang dapat dimanfaatkan untuk menganalisis **niat pembelian (purchasing intention)**.

##### 2.1.2 Niat Pembelian (Purchasing Intention)

Niat pembelian adalah keinginan atau kecenderungan konsumen untuk membeli suatu produk atau layanan setelah mengalami interaksi tertentu di situs e-commerce.

Menurut Ajzen (1991) dalam *Theory of Planned Behavior (TPB)*, niat adalah faktor utama yang mempengaruhi perilaku aktual seseorang. Dalam e-commerce, niat pembelian dapat dipengaruhi oleh berbagai faktor seperti:

- Kepercayaan (trust) terhadap situs atau penjual,
- Kemudahan penggunaan (usability) platform,
- Kepuasan pengalaman belanja (user experience), dan
- Harga serta promosi.

Analisis niat pembelian menjadi penting untuk membantu platform e-commerce mengoptimalkan strategi pemasaran dan retensi pelanggan.

##### 2.1.3 Data Mining dan Machine Learning

*Data mining* adalah proses menggali informasi berharga dari kumpulan data besar menggunakan teknik statistik, matematika, dan kecerdasan buatan. Salah satu cabangnya adalah *machine learning*, yaitu metode yang memungkinkan komputer mempelajari pola dari data untuk membuat prediksi.

Dalam penelitian ini, machine learning digunakan untuk membangun model prediksi terhadap niat pembelian berdasarkan perilaku pengguna di situs e-commerce.

### 2.1.4 Algoritma Random Forest

Random Forest adalah algoritma *supervised learning* yang bekerja dengan membangun banyak pohon keputusan (*decision trees*) dan menggabungkan hasilnya melalui proses voting atau averaging.

Kelebihan algoritma ini meliputi:

- Tahan terhadap *overfitting*
- Mampu menangani data berdimensi tinggi
- Menyediakan *feature importance* untuk mengetahui variabel paling berpengaruh

Dalam konteks prediksi niat pembelian, Random Forest digunakan karena kemampuannya menganalisis berbagai variabel perilaku pengguna secara simultan dan menghasilkan akurasi tinggi.

### 2.1.5 Hyperparameter Tuning dengan Grid Search

Grid Search merupakan metode sistematis untuk mencari kombinasi parameter terbaik pada model machine learning.

Prosesnya melibatkan:

1. Menentukan parameter dan rentang nilainya,
2. Melatih model dengan setiap kombinasi parameter,
3. Mengevaluasi performa menggunakan metrik seperti akurasi atau F1-score, dan
4. Memilih parameter dengan hasil terbaik.

Teknik ini meningkatkan performa model tanpa perlu mengubah struktur algoritma.

### 2.1.6 Analisis Penelitian Terkait

Penelitian yang menjadi dasar utama dari studi ini adalah karya Cemal Okan Sakar dan Yomi Kastro (2019) yang berjudul “*Real-time Prediction of Online Shoppers’ Purchasing Intention Using Multilayer Perceptron and LSTM Recurrent Neural Networks.*”

Penelitian tersebut berfokus pada penerapan algoritma Machine Learning berbasis Neural Network untuk memprediksi niat pembelian pengunjung situs e-commerce secara *real-time*.

Dataset yang digunakan adalah Online Shoppers Purchasing Intention Dataset dari UCI Machine Learning Repository, yang berisi data perilaku pengguna e-commerce selama satu tahun, termasuk aktivitas halaman yang dikunjungi, rasio pantulan (*bounce rate*), rasio keluar (*exit rate*), nilai halaman (*page value*), serta waktu dan jenis pengunjung (*visitor type*).

Dalam penelitian tersebut, penulis membandingkan dua model utama, yaitu Multilayer Perceptron (MLP) dan Long Short-Term Memory (LSTM), untuk mengklasifikasikan apakah seorang pengunjung memiliki niat membeli (*revenue = true*) atau tidak (*revenue = false*). Pendekatan ini dirancang untuk memproses pola perilaku pengguna secara berurutan sehingga dapat memberikan prediksi waktu nyata terhadap perilaku pembelian.

## Judul dan Peneliti

Komponen	Keterangan
Judul	Real-time Prediction of Online Shoppers' Purchasing Intention Using Multilayer Perceptron and LSTM Recurrent Neural Networks
Peneliti	Cemal Okan Sakar dan Yomi Kastro (2019)
Algoritma	Multilayer Perceptron (MLP) dan Long Short-Term Memory (LSTM)
Dataset	Online Shoppers Purchasing Intention Dataset (UCI Machine Learning Repository)
Akurasi	LSTM mencapai akurasi tertinggi sebesar $\pm 91\%$ , sedangkan MLP sedikit lebih rendah namun lebih efisien dalam pelatihan.

### 2.1.7 Metodologi Penelitian Sakar & Kastro (2019)

#### 1. Pengumpulan dan Pemrosesan Data

Dataset dikumpulkan dari log perilaku pengunjung situs e-commerce selama satu tahun. Data kemudian diproses dengan:

- Menangani nilai kosong (*missing values*).
- Mengubah variabel kategorikal menjadi numerik.
- Melakukan normalisasi untuk memastikan stabilitas model.

#### 2. Penerapan Model MLP dan LSTM

Dua model utama diterapkan:

- MLP digunakan untuk mempelajari hubungan antar fitur non-sekuensial (statik).
- LSTM digunakan untuk mempelajari pola perilaku pengguna secara berurutan (*sequential behavior*) yang menggambarkan perjalanan pengguna dari awal kunjungan hingga keputusan pembelian.

#### 3. Evaluasi Model



Evaluasi dilakukan menggunakan metrik Accuracy, Precision, dan Recall.

Model LSTM memberikan hasil terbaik dengan akurasi sekitar 91%, menunjukkan kemampuannya mengenali pola perilaku pembelian secara temporal.

#### 4. Analisis Hasil

Penulis menyimpulkan bahwa model berbasis jaringan saraf (terutama LSTM) efektif untuk mendeteksi niat pembelian, namun membutuhkan sumber daya komputasi tinggi dan tidak mudah diimplementasikan pada sistem e-commerce berskala besar.

##### 2.1.8 Analisis Relevansi dengan Penelitian Ini

Penelitian Sakar & Kastro (2019) menjadi acuan penting karena menggunakan dataset yang sama, yaitu *Online Shoppers Purchasing Intention Dataset*, dan berfokus pada prediksi niat pembelian pengguna. Namun, penelitian tersebut menggunakan model neural network (MLP dan LSTM) tanpa eksplorasi optimasi parameter secara sistematis atau pembahasan mendalam tentang algoritma berbasis *ensemble learning*.

Penelitian yang diusulkan ini berbeda dalam beberapa aspek penting:

1. Menggunakan algoritma Random Forest, yang termasuk *ensemble learning* berbasis pohon keputusan dan dikenal lebih interpretatif serta efisien dalam komputasi.
2. Melakukan Hyperparameter Tuning menggunakan Grid Search untuk menemukan konfigurasi parameter terbaik, sehingga model dapat mencapai performa optimal.
3. Menekankan pada analisis feature importance, untuk mengetahui atribut perilaku pengguna yang paling berpengaruh terhadap niat pembelian.

Dengan demikian, penelitian ini bertujuan untuk menghadirkan pendekatan yang lebih efisien, interpretatif, dan mudah diimplementasikan dalam sistem e-commerce konvensional, serta mengeksplorasi potensi peningkatan akurasi melalui optimasi parameter secara sistematis.

## **BAB 3**

### **METODE PENELITIAN**

#### **3.1 Jenis dan Pendekatan Penelitian**

Penelitian ini termasuk dalam jenis penelitian kuantitatif dengan pendekatan eksperimen komputasional (*computational experiment*). Fokus penelitian adalah menguji kinerja model *machine learning* dalam memprediksi niat pembelian pengunjung situs e-commerce menggunakan algoritma Random Forest yang dioptimasi dengan Grid Search.

Pendekatan kuantitatif digunakan karena seluruh proses analisis dilakukan berdasarkan data numerik dan hasil evaluasi model diukur menggunakan metrik kuantitatif seperti akurasi, presisi, recall, dan F1-score.

#### **3.2 Sumber Data**

Penelitian ini menggunakan dataset sekunder yang berasal dari repositori UCI Machine Learning Repository, yaitu:

Online Shoppers Purchasing Intention Dataset  
(<https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset>)

Dataset ini berisi 12.330 entri data dari pengunjung situs e-commerce selama satu tahun (2015–2016). Setiap entri merepresentasikan perilaku satu sesi kunjungan pengguna, dengan 18 fitur input dan 1 label output (Purchase/Non-Purchase).

Contoh beberapa fitur yang digunakan:

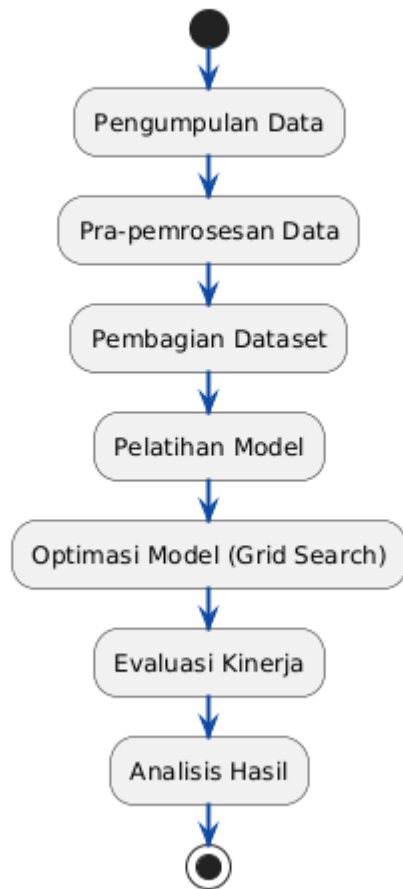
- Administrative, Informational, ProductRelated - jumlah halaman yang dikunjungi
- BounceRates, ExitRates, PageValues - metrik perilaku pengguna
- SpecialDay - kedekatan waktu dengan hari-hari khusus seperti Valentine's Day
- Month - bulan kunjungan
- VisitorType dan Weekend - tipe pengguna dan waktu kunjungan

Label (Revenue) bernilai TRUE jika pengguna melakukan pembelian, dan FALSE jika tidak.

#### **3.3 Tahapan Penelitian**

Secara umum, tahapan penelitian ini dapat digambarkan pada diagram berikut:

## Alur Proses Machine Learning



Penjelasan tiap tahap:

### 1. Pengumpulan Data

Data diunduh langsung dari situs resmi UCI Repository dan disimpan dalam format .csv. Dataset dipilih karena bersifat publik, lengkap, dan sudah banyak digunakan dalam penelitian terkait prediksi perilaku pengguna.

### 2. Pra-pemrosesan Data (Data Preprocessing)

Langkah ini bertujuan memastikan data bersih dan siap digunakan.

Tahapannya meliputi:

- Pengecekan nilai kosong (missing values) dan penanganannya (imputasi atau penghapusan).
- Normalisasi dan standarisasi untuk menyamakan skala fitur numerik.
- Encoding pada data kategorikal menggunakan *One-Hot Encoding*.
- Pembagian dataset menjadi data latih dan uji dengan perbandingan 80:20.

### 3. Pelatihan Model (Model Training)

Model utama yang digunakan adalah Random Forest Classifier, yang termasuk algoritma *ensemble learning* berbasis *bagging*.

Pada tahap ini dilakukan eksperimen awal tanpa optimasi untuk melihat performa dasar (baseline).

### 4. Optimasi Model (Hyperparameter Tuning)

Untuk meningkatkan akurasi, dilakukan optimasi hyperparameter menggunakan Grid Search, dengan parameter yang diuji antara lain:

- `n_estimators`: jumlah pohon (misal: 100, 200, 300)
- `max_depth`: kedalaman maksimum pohon
- `min_samples_split` dan `min_samples_leaf`: batas minimum sampel per node
- `criterion`: fungsi evaluasi seperti gini atau entropy

Proses Grid Search akan menguji setiap kombinasi parameter dan memilih hasil terbaik berdasarkan akurasi tertinggi pada validasi silang (*cross-validation*).

### 5. Evaluasi Model

Model yang telah dioptimasi dievaluasi menggunakan data uji dengan metrik:

- Accuracy : Persentase prediksi benar
- Precision : Ketepatan prediksi kelas positif
- Recall : Kemampuan mendeteksi kelas positif
- F1-score : Rata-rata harmonik antara precision dan recall
- Confusion Matrix : Tabel visualisasi hasil klasifikasi

### 6. Analisis Hasil

Langkah terakhir adalah menganalisis hasil model terbaik dan membandingkannya dengan penelitian sebelumnya (MLP dan LSTM). Selain itu, dilakukan analisis feature importance untuk mengetahui variabel perilaku pengguna mana yang paling berpengaruh terhadap niat pembelian.

## 3.4 Alat dan Bahasa Pemrograman

Penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan pustaka pendukung berikut:

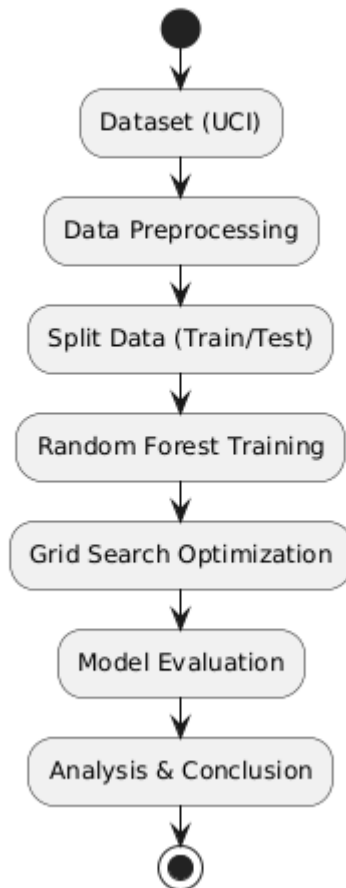
- Pandas, NumPy — untuk pengolahan data
- Scikit-learn — untuk implementasi Random Forest dan Grid Search
- Matplotlib, Seaborn — untuk visualisasi data dan hasil evaluasi

- Google Colab / Jupyter Notebook — sebagai lingkungan pemrograman

### 3.5 Kerangka Penelitian

Berikut diagram alur metodologi penelitian:

**Diagram Alur Metodologi Penelitian**



### 3.6 Output yang Diharapkan

Hasil akhir penelitian ini berupa:

1. Model Random Forest dengan parameter optimal yang mampu memprediksi niat pembelian secara akurat.
2. Analisis fitur yang berpengaruh terhadap keputusan pembelian.
3. Perbandingan performa dengan penelitian sebelumnya (MLP & LSTM).
4. Rekomendasi implementasi model dalam sistem e-commerce untuk meningkatkan konversi penjualan.

Dengan metodologi ini, diharapkan penelitian dapat menghasilkan model prediksi niat pembelian yang akurat, efisien, serta dapat memberikan wawasan yang berguna bagi pengelola platform e-commerce dalam memahami perilaku pelanggan.

## DAFTAR PUSTAKA

- [1] C. O. Sakar and Y. Kastro, "Online Shoppers Purchasing Intention Dataset," *UCI Machine Learning Repository*, 2019. [Online]. Available: <https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset>
- [2] C. O. Sakar and Y. Kastro, "Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks," *Neural Computing and Applications*, vol. 31, no. 10, pp. 6893–6908, 2019, doi: 10.1007/s00521-018-3523-0.
- [3] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001