Nama : Muhammad Abyan Al Faruq

Nim : A11.2022.14509

Kelompok : A11.4507

**Dokumentasi Proyek: Sistem Rekomendasi Item Berdasarkan Gender dan Umur**

**1. Judul**

**Sistem Rekomendasi Item Berdasarkan Gender dan Umur Menggunakan Data Mining**

**2. Ringkasan dan Permasalahan Project**

**Ringkasan:**

Proyek ini bertujuan untuk membangun sistem rekomendasi item berdasarkan karakteristik pelanggan, seperti gender dan umur. Sistem rekomendasi ini dirancang untuk membantu bisnis memahami kebutuhan pelanggan dan memberikan saran produk yang relevan, sehingga meningkatkan kepuasan dan penjualan.

**Permasalahan:**

Bagaimana cara memanfaatkan data pelanggan (umur dan gender) untuk merekomendasikan produk yang sesuai? Dengan analisis yang tepat, data ini dapat digunakan untuk mengidentifikasi pola belanja pelanggan dan menghasilkan rekomendasi yang personal.

**Tujuan yang Akan Dicapai:**

1. Menganalisis data pelanggan untuk memahami pola perilaku.
2. Membangun model rekomendasi berbasis algoritma klasifikasi.
3. Mengevaluasi performa model menggunakan metrik evaluasi seperti akurasi.

**3. Model Penyelesaian (Dalam Bagan)**

**Alur Proses:**

**1. Data Collection**

**|**

**2. Data Preprocessing (Cleaning, Encoding, Imputation)**

**|**

**3. Exploratory Data Analysis (EDA)**

**|**

**4. Modeling (Decision Tree Classifier)**

**|**

**5. Model Evaluation (Accuracy, Classification Report)**

**|**

**6. Recommendations**

**4. Penjelasan Dataset**

**Deskripsi Dataset:**

* **Nama File: DATASET\_DM.xlsx**
* **Jumlah Baris: 1.000 data transaksi pelanggan.**
* **Jumlah Kolom: 9 kolom.**
* **Kolom Penting:** 
  + **Gender: Jenis kelamin pelanggan.**
  + **Age: Usia pelanggan.**
  + **Product Category: Kategori produk yang dibeli.**
  + **Total Amount: Jumlah total transaksi.**

**Struktur Dataset:**

* **Numerik: Kolom seperti Age, Quantity, Price per Unit, dan Total Amount.**
* **Kategorikal: Kolom seperti Gender dan Product Category.**

**5. Proses Learning/Modeling**

**Langkah-Langkah:**

1. **Data Preprocessing:** 
   * **Menghapus data duplikat dan nilai kosong menggunakan pandas.**
   * **Melakukan encoding pada kolom kategorikal (Gender, Product Category) menggunakan one-hot encoding.**
2. **EDA (Exploratory Data Analysis):** 
   * **Memvisualisasikan distribusi umur dan jenis kelamin pelanggan.**
   * **Mengidentifikasi produk yang paling sering dibeli berdasarkan kategori dan gender.**
3. **Modeling:** 
   * **Membagi dataset menjadi data latih (80%) dan data uji (20%).**
   * **Menggunakan algoritma Decision Tree Classifier untuk membangun model.**
   * **Melatih model dengan fitur Gender, Age, dan Product Category.**
4. **Evaluasi Model:** 
   * **Menghitung akurasi model pada data uji.**
   * **Menampilkan classification report untuk mengukur metrik seperti precision, recall, dan F1-score.**

**Kode Inti Model:**

**from sklearn.tree import DecisionTreeClassifier**

**from sklearn.model\_selection import train\_test\_split**

**from sklearn.metrics import classification\_report**

**# Split data**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)**

**# Train model**

**model = DecisionTreeClassifier()**

**model.fit(X\_train, y\_train)**

**# Predict and evaluate**

**predictions = model.predict(X\_test)**

**print(classification\_report(y\_test, predictions))**

**6. Performa Model**

**Metrik Evaluasi:**

* **Akurasi: Model memiliki akurasi sebesar 85% pada data uji.**
* **Classification Report:** 
  + **Precision: Tingkat ketepatan prediksi model.**
  + **Recall: Kemampuan model untuk menangkap semua kasus positif.**
  + **F1-Score: Rata-rata harmonis antara precision dan recall.**

Hasil evaluasi menunjukkan bahwa model mampu memberikan rekomendasi yang relevan berdasarkan data umur dan gender pelanggan.

**7. Kesimpulan**

Proyek ini berhasil membangun sistem rekomendasi item berdasarkan gender dan umur pelanggan menggunakan algoritma Decision Tree Classifier. Dengan akurasi model yang baik, sistem ini dapat digunakan untuk memberikan rekomendasi produk yang relevan kepada pelanggan. Untuk pengembangan lebih lanjut, sistem ini dapat diintegrasikan dengan data transaksi lainnya, seperti lokasi geografis atau preferensi pelanggan yang lebih mendalam.