Fraud Detection and Transaction Insights — Project Report

Project Title:

E-Commerce Transaction Fraud Detection Using Machine Learning

---

Objective:

To detect fraudulent transactions from an e-commerce dataset using exploratory data analysis and a supervised machine learning model (Decision Tree Classifier).

Tools & Technologies Used:

- **Programming Language:** Python
- **Libraries:** pandas, seaborn, matplotlib, scikit-learn
- **Dashboard:** Power BI

---

1. Data Cleaning & Preprocessing:

- Handled null values and date conversion using `pd.to_datetime()`.
- Removed invalid customer ages (less than or equal to 0).
- Extracted new features from date column:
    - `Transaction Hour`, `Day`, `Month`,

---

2. Exploratory Data Analysis (EDA):

- **Fraudulent transactions** are only around **5%** of total — indicating **class imbalance**.
- Fraud is more associated with:
    - **Debit card** and **Bank Transfer**
    - **Mobile devices**
    - Customers in **age group 25–41**
- Certain product categories like **toys & games** and **home and garden** have higher fraud cases.
- Fraud occurred more frequently during **specific hours** and **weekdays**.

---

3. Visual Insights:

- Countplots and bar graphs were created to show:
  - Payment methods vs fraud count
  - Fraud by age group, product category, device used
  - Time-based fraud trends (hour & date)
  - Top 10 locations with highest fraud cases

---

4. Machine Learning Modeling:

- Used **Decision Tree Classifier** from `sklearn`.
- Applied `class_weight='balanced'` to handle class imbalance.
- Evaluation metrics (after training):
  - Accuracy, Precision, Recall, F1-score (evaluated using confusion matrix)
- Top important features identified:
  - `Account Age Days`, `Transaction Amount`, `Customer Age`, `Transaction Hour`, `Payment Method_Credit Card`

---

5. Dashboard (Power BI):

- A separate interactive Power BI dashboard was created to visualize:
  - Total fraud count and amount
  - Fraud distribution by age group, location, and device
  - Hourly fraud trends and filter options

---

6. Key Insights:

- **Debit card** and **Bank Transfer** payments are more fraud-prone.
- **Mobile device usage** slightly increases fraud risk.
- Fraud mainly happens in customers aged **25–41**.
- Transaction timing and account age play a major role in fraud likelihood.

---

7. Conclusion:

This project successfully demonstrates how machine learning and data visualization can help detect and understand patterns in fraudulent transactions. It highlights key risk factors that businesses can monitor to reduce fraud.