Daily Transaction Data Analysis & Predictive Modeling

Project Overview

This project focuses on analyzing daily household transactions using data analytics and machine learning techniques. The goal is to derive insights from financial behavior, visualize transaction patterns, and build predictive models to assist in budgeting and financial decision-making.

Dataset Description

The dataset captures detailed information on daily transactions made by an individual. It includes the following fields:

- **Date**: Timestamp of the transaction
- **Mode**: Payment method (e.g., Cash, UPI, Card)
- **Category**: High-level classification (e.g., Food, Utilities)
- **Subcategory**: Specific transaction type (e.g., Groceries, Electricity)
- **Note**: Optional text description
- **Amount**: Monetary value of the transaction
- **Income/Expense**: Label indicating if the transaction is income or an expense
- **Currency**: Currency used (INR)

Data Preprocessing

- Converted Date to datetime format and extracted Month and DayOfWeek.
- Handled missing values in Subcategory and Note.
- Encoded categorical variables using Label Encoding.
- Scaled numerical features using StandardScaler.
- Verified that all records use the same currency.

Exploratory Data Analysis (EDA)

- **Income vs Expense**: Compared income and expense proportions and distributions.
- **Category & Subcategory Trends**: Identified where most money is spent.

- **Mode of Payment**: Analyzed preferences in payment methods.
- **Time Series Analysis**: Tracked monthly spending and earnings patterns.

Machine Learning Model

- **Model Used**: Random Forest Classifier
- **Target Variable**: Income/Expense
- **Train-Test Split**: 70-30 ratio
- **Evaluation Metrics**:
    - Classification Report
    - Confusion Matrix
    - ROC-AUC Score
    - Multi-class ROC Curve Visualization

Project Structure

- Daily Household Transactions.csv: Dataset
- transaction_analysis.py: Analysis & modeling script
- requirements.txt: Python dependencies
- README.md: Project documentation

How to Run the Project

1. **Install dependencies**:

pip install -r requirements.txt

2. **Run the Python script**

python transaction_analysis.py

Requirements

- Python 3.8+
- pandas

- numpy
- matplotlib
- seaborn
- scikit-learn

Conclusion

This project demonstrates how financial transaction data can be processed, analyzed, and used to train machine learning models. By analyzing spending habits, users can better manage their finances and make informed budgeting decisions. The Random Forest model effectively distinguishes between income and expense patterns and provides reliable performance insights using metrics like ROC-AUC.

Future Scope

- **Real-Time Budget Tracking**: Integrate with live transaction data sources to monitor expenses in real time.
- **Category Prediction**: Predict future transaction categories based on past habits and time of month.
- **Anomaly Detection**: Identify unusual or fraudulent transactions using advanced models.
- **Personalized Financial Insights**: Recommend savings strategies or alerts based on user spending behavior.
- **Integration into Mobile Apps**: Develop a finance management app using this model as the backend.