Title: Personal Financial Advisor

Team: Project Team 3

Members:

Shubham Talele - Business Analyst

Prathamesh Gavali - ML Engineer

Jai Juneja - Full-Stack Developer

Pankaj Chauhan - Data Analyst

Muskan Chauhan – Cloud Engineer

Shubham Aphashinkar - Data Engineer

Institution: CDAC PG-DBDA February 2025

Project Overview

Problem:

Users struggle to track savings and detect overspending from raw expense data.

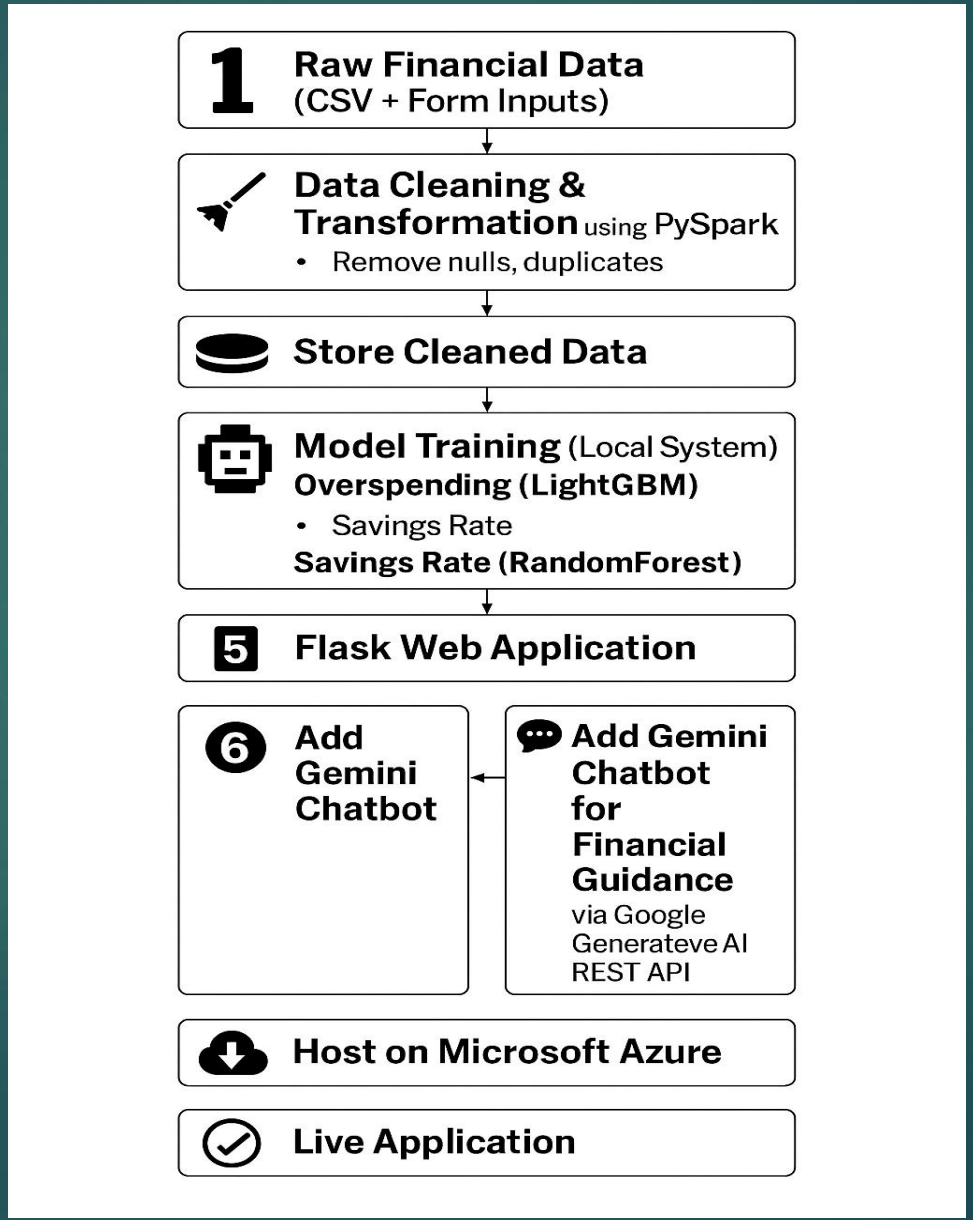
Objective: Build an intelligent, interactive platform to:

•Predict actual savings rate
•Flag overspending patterns
•Provide visual and chatbot-based guidance

Business Value & Impact :

- Helps users understand financial health in real-time
- Recommends actionable insights to meet savings goals
- Prevents over-budgeting and supports better planning
- Deployed on the cloud: accessible anywhere

Architecture Flow



Data Preprocessing(ETL)

- Dataset :
- Total Records are 275000
- Null values in Income [5454] , Age [5420] , random_note [137319] , temp_value [137366]
- So in this case we drop null records , we cannot impute that if we impute that it is meaning less
- For random_note and temp_value there is more than 50 % null records .
- So we drop these 2 columns & these two columns are meaning less for our needs
- And save cleaned file in csv mode
- Cleand data size now 255381

Exploratory Data Analysis & Insights

- Loaded dataset and checked shape, info, and summary stats
- Removed nulls and verified missing values
- Engineered new features like:
  - Spending_Rate, Actual_Savings_Rate, Savings_Gap
- Created Overspend_Flag for classification
- Encoded categorical columns: Occupation, City_Tier
- Visualized distributions: Income, Savings Rate, Spending Rate
- Analyzed overspending by Occupation and City Tier
- Plotted boxplot: Income vs Overspend
- Correlation heatmap for feature relationships
- Checked multicollinearity using VIF

ML Implementation

- *** Overspending Detection (Classification)**
- **Models Tried**: Logistic Regression, Random Forest, LightGBM
- **Best Model**: **LightGBM** (F1 = **0.59**, Recall = **0.99 , accuracy = 0.90**)
- **Why**: Best trade-off between accuracy and recall for minority class
- **Metrics Used**: F1-Score, Recall, Confusion Matrix
- *** Savings Rate Prediction (Regression)**
- **Models Tried**: Random Forest, Gradient Boosting, LightGBM
- **Best Model**: **Random Forest** (R^2 = **0.91**, MAE = **0.0218**)
- **Why**: Highest prediction accuracy with minimal error
- **Metrics Used**: RMSE, MAE, R^2 Score

Application Development

- **Frontend (HTML + CSS) :**
- Collects user inputs via interactive forms
- Displays savings insights, pie charts & feedback
- Integrated **Gemini Chatbot** for financial guidance
- **Flask Backend :**
- Handles input preprocessing & model loading
- Serves real-time predictions via Flask routes
- **Integration :**
- **HTML Form → Flask Route → ML Model → HTML Response**
- Bidirectional flow between chatbot and backend
- **User Experience :**
- Instant predictions with clear visual insights
- Natural interaction via chatbot for suggestions

Deployments on Azure

1. create resources => web app service => Runtime stack = python

   use free tier version

2. Deployement Center => Authenticated with Github => select User => Repository => branch => main

3. From overview we can able to access the hosted URL

4. With the help of hosted URL we are able to access the project

Results & Business Value

- **1. Key Findings**
- **Top overspending drivers**: Eating Out, Entertainment, Miscellaneous.
- Users with loans + high rent% show **consistently lower savings**.
- **2. Model Performance**
- **Overspending Detection**
- **Savings Rate Prediction**
- **3. Business Recommendations**
- Push *personalized nudges* to cut top 2 overspend categories.
- Partner features: insurance/loan refinance suggestions for high-loan users.

Results & Business Value

- **4. Future Enhancements (Easy & Practical)**
- **Auto email/SMS reminders** when user crosses a spend limit
- **One-click Excel/PDF export** of monthly report
- **Category-wise tips** (e.g., “Cut eating-out by 10% to hit your goal”)
- **User login + history tracking** (see past months’ trends)
- **Mobile-friendly UI / Android app later**

Hosted URL

- https://financial-eqhxdbgtf7eubagq.centralus-01.azurewebsites.net/