# Predictive Modeling for Smartphone Purchase Behavior


## 💼 Prepared by:

**Name**: Mandar Kajbaje
**Domain**: Machine Learning – Classification
**Tools Used**: Python, Scikit-learn, Pandas, Matplotlib, Seaborn


## 📝 Abstract


This project focuses on building a predictive model to forecast whether a customer will purchase a smartphone based on their behavior and financial indicators. Using a synthetic dataset and machine learning models, we analyze key factors and apply classification techniques to predict purchase likelihood.


## 📌 Problem Statement

**Goal**: Predict if a customer will buy a smartphone (1) or not (0) using demographic and behavior-based features.
**Use Case**: Businesses can target marketing campaigns more effectively and increase conversion rates.

# 📂 Dataset Description

| Feature Name | Description |
|---|---|
| age | Age of the customer in years |
| income | Monthly income (₹) |
| time_on_website | Time spent browsing smartphone-related content (min) |
| previous_purchases | No. of past purchases (phones or accessories) |
| marketing_engaged | Whether the user engaged with ads (1 = yes, 0 = no) |
| search_frequency | How often the user searched phone terms |
| device_age | Age of current device (years) |
| will_purchase | Target variable (1 = yes, 0 = no) |

Dataset: **synthetic**, generated using Python.

# Step-by-Step Machine Learning Workflow

## Step 1: Import Libraries & Load Dataset

- Imported essential libraries: **pandas, numpy, matplotlib, seaborn**
- Loaded dataset into a DataFrame

## Step 2: Initial Exploration

- Checked data types and missing values
- Used **df.describe()** to understand feature distributions

## Step 3: Exploratory Data Analysis (EDA)

- Visualized age distribution, income spread, and class balance
- Heatmap used to explore feature correlations
- Found that features like **income, device_age**, and **marketing_engaged** have impact

## Step 4: Data Preprocessing

- Split data into features (X) and target (y)
- Performed 80-20 train-test split
- Scaled features using StandardScaler for normalization

## Step 5: Model Training

- Trained multiple models:
  **Logistic Regression**, **Random Forest**, and **SVM**
- Evaluated each using **accuracy_score** and **classification_report**

## Step 6: Model Evaluation

- Used confusion matrix and ROC curve
- Random Forest showed best performance (accuracy ~**75–80%**)

## Step 7: Hyperparameter Tuning

- Applied GridSearchCV on Random Forest
- Tuned **n_estimators** and **max_depth**
- Found optimal hyperparameters for better generalization

## Step 8: Model Interpretation

- Used SHAP to explain feature contributions
- Most important features:
  **income, time_on_website, device_age, marketing_engaged**

## 📊 Results Summary

| Model | Accuracy (Approx.) |
|---|---|
| Logistic Regression | 80.00% |
| Random Forest | 77.50% |
| Support Vector Machine (SVM) | 77.50% |

📌 Final selected model: **Logistic Regression** (Best Score: 78.50%)

# Conclusion

This project demonstrated how machine learning can predict smartphone purchasing behavior using behavioral and demographic data. Random Forest provided the best trade-off between performance and explainability. This can help businesses target high-potential customers effectively.