**Case Study: Predicting Loan Defaults at BankX**

**Background**

**BankX**, a mid-sized financial institution, wants to predict whether loan applicants are likely to default. By using machine learning, they aim to:

* Minimize financial risk
* Approve loans more efficiently
* Offer tailored financial advice to risky customers

The bank has historical data containing customer information such as:

* Age
* Income
* Credit score
* Loan amount
* Employment status
* Previous defaults
* Loan purpose
* Marital status

They decide to build a machine learning model using this data.

**Machine Learning Process**

* **Problem Type:** Classification (Predict Default: Yes or No)
* **Models Used:** Logistic Regression, Decision Tree, Random Forest
* **Best Model:** Random Forest (Accuracy: 92%)
* **Features Importance:** Credit Score, Income, Previous Defaults were top predictors.
* **Deployment:** Model is integrated into BankX's loan processing system.

### 1. What type of machine learning problem is this?

**Answer:**  
This is a **classification** problem because the output (default or no default) is categorical.

### 2. Which features were most important in predicting loan default?

**Answer:**  
The top three important features were **Credit Score**, **Income**, and **Previous Defaults**.

### 3. Why was Random Forest chosen over Logistic Regression and Decision Tree?

**Answer:**  
Random Forest was chosen because it achieved the highest accuracy (92%) and better handled the non-linear relationships and interactions between features.

### 4. What could the bank do if the model predicts a high risk of default for a customer?

**Answer:**  
The bank could:

* Offer smaller loans
* Increase interest rates
* Request collateral
* Offer financial counseling
* Decline the loan

### 5. How could BankX improve their model further?

**Answer:**  
They could:

* Collect more diverse and updated data
* Engineer new features (like debt-to-income ratio)
* Use more advanced models (like XGBoost or Neural Networks)
* Regularly retrain the model with new data
* Handle class imbalance better (if defaults are rare)

### 6. What are some ethical considerations when using this model?

**Answer:**

* **Fairness:** Ensure the model is not biased against certain groups (e.g., by gender or ethnicity).
* **Transparency:** Customers should understand why they were rejected.
* **Privacy:** Protect sensitive personal and financial data.
* **Accountability:** Decisions made by the model should be auditable.

### 7. What would happen if the model had a high false negative rate?

**Answer:**  
A high **false negative rate** means predicting a customer as "safe" when they actually default — leading to financial losses for the bank.