**Problem Statement – Insurance Sector: Claim Fraud Detection**

**Objective:**  
Develop a pipeline to detect potentially fraudulent insurance claims using historical claim and policyholder data.

**Key Tasks:**

1. **Data Preparation:**
   * Use or simulate data including claim amount, claim type, incident description, policy tenure, and previous claim history.
   * Clean and encode categorical data such as claim category and region.
2. **Feature Engineering:**
   * Extract features like claim-to-premium ratio, claim frequency, claim timing after policy start, etc.
3. **Model Building:**
   * Train a Random Forest or XGBoost model to predict claim fraud probability.
   * Use model explainability tools (e.g., SHAP) to identify top fraud indicators.
4. **Risk Scoring & Thresholding:**
   * Generate fraud risk scores and classify claims as Normal, Suspicious, or High Risk.
5. **Evaluation & Reporting:**
   * Use metrics (AUC, Precision, Recall) and visualize feature importance.
   * Produce a short summary report on fraud trends.

**Deliverable:**  
An end-to-end claim fraud detection pipeline that outputs risk scores and helps identify suspicious insurance claims.