# **HEALTH RISK PREDICTION MODEL**

## **1. Introduction**

Obesity-related diseases, including diabetes, hypertension, and cardiovascular conditions, are major public health concerns. This project utilizes machine learning to predict an individual’s health risk based on key biometric parameters such as BMI, age, and blood pressure. The model provides personalized health recommendations to prevent obesity-related diseases.

## **2. Purpose**

The objective is to develop a classification model that categorizes individuals as either "At Risk" or "No Risk" based on their health data. This allows for early intervention and tailored lifestyle recommendations.

## **3. Methodology**

### **3.1 Data Collection & Processing**

* **Dataset**: A synthetic dataset of 100 samples was generated with the following features:
  + **BMI (Body Mass Index)**
  + **Age**
  + **Blood Pressure**
  + **Health Risk Label**: 1 (At Risk) or 0 (No Risk)
* **Data Splitting**:
  + Training Set: 80%
  + Testing Set: 20%
  + Stratified sampling was used to maintain class balance.

### **3.2 Model Selection & Training**

* **Algorithm Used**: **Random Forest Classifier**
* **Hyperparameters**:
  + Number of Estimators: 50
  + Max Depth: 5
  + Random State: 42

### **3.3 Model Evaluation**

* **Accuracy Score**: The model initially achieved **95% accuracy**, which is higher than the target (below 90%).
* **Confusion Matrix Analysis**:
  + **True Positives (TP)**: 6
  + **True Negatives (TN)**: 13
  + **False Positives (FP)**: 0
  + **False Negatives (FN)**: 1
* **Classification Report Metrics**:
  + **Precision**: 0.93 (for class 0), 1.00 (for class 1)
  + **Recall**: 1.00 (for class 0), 0.86 (for class 1)
  + **F1-Score**: 0.96 (for class 0), 0.92 (for class 1)
  + **Overall Accuracy**: **95%** (Needs reduction for better generalization)

### **3.4 Feature Importance**

A feature importance analysis showed that **BMI** had the most significant impact on risk classification, followed by blood pressure and age.

## **4. Results & Insights**

* The model demonstrated high accuracy but may be overfitting due to limited data.
* The confusion matrix shows that **only 1 false negative** occurred, meaning the model is slightly biased toward identifying individuals as "No Risk."
* Feature analysis suggests BMI is the most influential predictor, confirming its role in obesity-related diseases.

## **5. Conclusion & Future Work**

* The model effectively predicts health risks but needs further tuning for better generalization.
* Future improvements:
  + Expanding the dataset with real-world data.
  + Introducing lifestyle-based parameters (diet, activity level).
  + Optimizing hyperparameters further to prevent overfitting.
  + Deploying as a web-based tool for real-time health risk assessment.