AI Explainability Case Study: LIME-Based Interpretability in Heart Disease AI

1⃣⃣ Project Overview

📌 Objective: Use LIME (Local Interpretable Model-agnostic Explanations) to explain predictions of a Random Forest model for health risk classification.  
 📌 Problem: While the model was accurate, it lacked interpretability at the individual patient level.  
 📌 Goal: Provide clear, instance-specific explanations to aid decision-makers in clinical contexts.  
 📌 Tools & Libraries Used: Python, LIME, Scikit-learn

2⃣⃣ Dataset & Model Summary

* **Dataset**: Breast cancer dataset (used as a proxy for heart disease classification)
* **Features**: 30 clinical input variables (e.g., mean radius, texture)
* **Label**: Diagnosis (0 = malignant, 1 = benign)
* **Model Used**: Random Forest Classifier
* **Accuracy**: ~96% on test set

3⃣⃣ Explainability Technique

🔹 **Tool Used**: LIME (lime\_tabular)  
 🔹 **Explanation Type**: Local (Instance-specific rationale)

🔹 **Metrics Considered**:

* **Local Fidelity**: How well the local surrogate model approximates the global model
* **Human Interpretability**: Simplicity and understandability of output features

4⃣⃣ Key Python Snippets

explainer = lime.lime\_tabular.LimeTabularExplainer(

training\_data=X\_train\_scaled,

feature\_names=heart.feature\_names,

class\_names=heart.target\_names,

mode="classification"

)

exp = explainer.explain\_instance(X\_test\_scaled[0], model.predict\_proba, num\_features=5)

exp.show\_in\_notebook(show\_table=True, show\_all=False)

5⃣⃣ Insights Gained

📌 LIME clearly highlighted top 5 contributing features per instance.  
 📌 Clinicians can now review which factors led to a high-risk or low-risk diagnosis.

6⃣⃣ Challenges & Fixes

🔹 **Challenge**: Variability in LIME results due to sampling randomness.  
 🔹 **Fix**: Used fixed random state and averaged multiple runs to stabilize outputs.

🔹 **Challenge**: Feature names were lengthy and unclear in the output.  
 🔹 **Fix**: Applied preprocessing to rename and simplify for end-user clarity.

7⃣⃣ Key Learnings

📅 LIME bridges the gap between model predictions and clinical transparency.  
 📅 Simpler explanations improve real-world decision-making.  
 📅 SMF mission aligns with making AI decisions interpretable and ethical.

8⃣⃣ Next Steps  
 🚀 Compare LIME with SHAP and PDP on the same dataset.  
 🚀 Build multi-instance visual dashboards for clinicians.

📉 This case study is a key part of the SoulMindFusion Explainable AI Portfolio!