🧠 Rare Disease Diagnosis Assistant

A Machine Learning-based assistant to support doctors in diagnosing **rare diseases** based on patient-reported symptoms.

📌 Project Overview

Rare diseases often present with vague or overlapping symptoms, making accurate diagnosis a challenge for healthcare providers.

This project aims to **support medical professionals** by building a predictive model that uses **binary symptoms (0/1)** as input and suggests a possible disease class.

The system doesn't replace a doctor — it acts as a **decision-support tool**, highlighting conditions with low confidence and encouraging human verification when needed.

🛠️ Technologies Used

- **Python**
- **Pandas / NumPy / Matplotlib / Seaborn**
- **Scikit-learn**: `RandomForestClassifier`, `train_test_split`, `classification_report`
- **Machine Learning**: Supervised classification
- **Performance Metrics**: Accuracy, Precision, Recall, F1-score
- **Explainability Logic**: Warnings for diseases with low data representation

📂 Dataset

- Input: Binary symptom indicators (0 = no, 1 = yes)
- Output: Disease label (rare or common)
- Format: CSV with features representing symptoms and a target column `diseases`

- `Data set link: https://www.kaggle.com/datasets/dhivyeshrk/diseases-and-symptoms-dataset`

Note: The dataset was cleaned and verified to only contain binary values in features.

🔍 Model Workflow

1. **Data Cleaning & Preprocessing**
2. **Train/Test Split**
3. **Model Training** using `RandomForestClassifier`
4. **Prediction on unseen data**
5. **Overfitting Check** via comparison of training/testing accuracy
6. **Evaluation Metrics**
7. **Custom Logic for Confidence Messages**:
    a. If **F1-score < 0.6** or sample count is low → model warns the doctor
    b. If metrics are reliable → model confirms prediction reliability

💬 Doctor Guidance Messages

The model prints messages like:

- ✅ *"You can rely on this result. The model performed well and was trained on enough data."*
- ⚠️ *"Doctor, please verify this result by yourself, due to low sample count."*

🚀 Future Enhancements

- Add **Neural Network** alternative model
- Integrate with medical APIs for validation
- Improve explainability via SHAP / LIME
- UI for non-technical medical users

Disclaimer

It must **not** be used for real-life diagnosis without medical supervision.

Always consult a licensed physician for actual medical decision-making.

Author

Basma Sameh

AI & Data Science Student

GitHub: [basmasameh84]