Disease Prediction Dataset Overview

Dataset Files
- Training.csv
- Testing.csv

Training Dataset
- Shape: (4920, 133)
- 132 feature columns + 1 target column ('prognosis')
- Features: Symptoms (binary values: 0 = absent, 1 = present)
- Target: 41 unique diseases

Testing Dataset
- Similar structure to training set
- Used for evaluating the model

Missing Values
- 'Unnamed: 133' column: All values are NaN (dropped)

Example Features (Symptoms)
- itching
- skin_rash
- joint_pain
- fatigue
- nausea
- vomiting
- etc. (Total: 132 symptoms)

Target Variable
- prognosis: Disease name
- Encoded using LabelEncoder

Sample Diseases
- Fungal infection
- Allergy
- GERD
- Chronic cholestasis
- Drug Reaction
- And many more (41 total)

Usage
The dataset is used to train a machine learning model to predict diseases based on symptoms. Each row represents a unique symptom profile and its corresponding diagnosis.

Ideal for classification problems in healthcare using models like:
- Random Forest
- Decision Tree
- Naive Bayes
- SVM, etc.