# Code Documentation

## Code Structure

.  
├── app.py # Streamlit frontend  
├── train\_models.py # ML model training and saving as .pkl  
├── data\_preprocessing.py # Data cleaning and feature engineering  
├── chronic\_kidney\_disease.csv # Dataset used  
├── \*.pkl # Saved models for use in UI

## Modules Used

* pandas, sklearn, xgboost – for data handling and ML
* pickle – to save/load trained models
* streamlit – UI for live input and prediction
* PySpark – for distributed processing and Spark DataFrames (optional)

## Execution Steps

1. Preprocess dataset using data\_preprocessing.py
2. Train models using train\_models.py
3. Launch frontend using:

streamlit run app.py

## Dependencies and Environment Setup

* Python 3.10+

pip install pandas scikit-learn xgboost streamlit

* Optional: Apache Spark, Java 11