Documentation: Cervical Cancer Risk Prediction

**Overview**

This code demonstrates a machine learning pipeline for predicting the risk of cervical cancer. It performs data preprocessing, model training, evaluation, and visualization of results.

**Installation**

Ensure you have Python installed on your system. You can install required libraries using pip:

pip install pandas numpy matplotlib seaborn scikit-learn

Code Description

**Importing Libraries**

pandas: Data manipulation and analysis.

numpy: Numerical computing library.

matplotlib: Data visualization library.

seaborn: Data visualization library based on matplotlib.

scikit-learn: Machine learning library for data preprocessing, modeling, and evaluation.

**Data Loading**

The code loads the cervical cancer risk dataset from the UCI Machine Learning Repository.

Preprocessing

Numeric and non-numeric columns are separated.

Missing values in numeric columns are filled with the mean and scaled using StandardScaler.

Missing values in non-numeric columns are filled with the most frequent value and encoded using OneHotEncoder.

**Model Training**

The data is split into training and testing sets using train\_test\_split.

The Random Forest Classifier is used as the classification model.

**Model Evaluation**

Accuracy, confusion matrix, and classification report are used to evaluate the model's performance.

Usage

Ensure Python and required libraries are installed.

Copy the code into a Python script or Jupyter Notebook.

**Run the code.**