Medicos is a medical web application which does Breast Cancer Diagnosis by classifying the cells into Benign & Malignant based on the Input Parameters.

Medicos

Breast Cancer Diagnosis

Ravi Sharma

Index

Introduction

Medicos Web Application -

Medicos is a medical web application which does Breast Cancer Diagnosis by classifying the cells into Benign & Malignant based on the Input Parameters, The web application is created using "Streamlit" and python machine learning.

What is Streamlit?

Streamlit is an open-source Python library that allows you to create and share interactive web applications for data science and machine learning projects with minimal effort. It is designed to help data scientists and developers quickly build and deploy web apps using only a few lines of Python code.

How to run a streamlit file on the web browser?

Command to install "streamlit" on visual studio code (Run the visual studio code as administrator) -

1. Install the library in your visual studio code.

pip install streamlit

2. Import the library in your visual studio code.

import streamlit as st

3. Write basic streamlit code in the "main.py" file.

4. Open the terminal in your visual studio code

streamlit run main.py

What is classification?

Classification is a supervised machine learning method where the model tries to predict the correct label of a given input data. In classification, the model is fully trained using the training data, and then it is evaluated on test data before being used to perform prediction on new unseen data. Random Forest, decision trees and logistic regressions are some famous algorithms used.

Breast Cancer

What is breast Cancer?

Breast cancer is a type of cancer that forms in the cells of the breasts. It is the most common cancer diagnosed in women worldwide, though it can also occur in men. Breast cancer begins when cells in the breast grow out of control, forming a tumor that can often be seen on an X-ray or felt as a lump. If left untreated, these cancer cells can spread (metastasize) to other parts of the body.

The cells can be classified into two Categories "Benign" and "Malignant.

What are the Benign Cells?

"Benign" tumors or conditions are non-cancerous and do not invade nearby tissues or spread to other parts of the body. They typically grow slowly and are usually considered less harmful than malignant tumors. In the context of breast cancer, benign tumors are not cancerous and do not pose the health risks.

What are the Malignant Cells?

"Malignant" refers to a medical term used to describe cells or tumors that are cancerous. In the context of breast cancer or any other type of cancer, malignant cells are abnormal cells that have the potential to invade nearby tissues and spread to other parts of the body. A malignancy can be aggressive and requires medical intervention such as surgery, chemotherapy, radiation, or targeted therapy to manage and treat.

Why is there a need for Breast Cancer Classification?

Machine learning models can assist radiologists and pathologists in accurately diagnosing breast cancer, reducing human error and improving diagnostic precision. Early and accurate detection is critical for successful treatment outcomes.

Apart for the use in the medical sciences it also helps understand various students to understand how classification and supervised machine learning algorithms work, The Breast Cancer is perfect for classification purposes because there are exactly two target categories in the dataset that are Benign and Malignant hence making it very well suited for classification tasks.

Dataset Description

What is the source of the Dataset?

The dataset is loaded from the "Scikit-learn" Machine learning library. It was originally obtained from the University of Wisconsin Hospitals, Madison, from Dr. William H. Wolberg. The dataset is publicly available from the UCI Machine Learning Repository.

Dataset Details -

- Number of Instances: 569
- Number of Attributes: 30 numeric, predictive attributes and the class label.

Class Distribution -

- 212 Malignant (cancerous)
- 357 Benign (non-cancerous)

Target Variable -

- The target variable is the diagnosis of breast tissues, where:
- 0 denotes malignant
- 1 denotes benign

Various Features of the dataset –

They describe characteristics of the cell nuclei present in the image -

1. Radius (mean of distances from center to points on the perimeter)
2. Texture (standard deviation of gray-scale values)
3. Perimeter
4. Area
5. Smoothness (local variation in radius lengths)
6. Compactness (perimeter^2 / area - 1.0)
7. Concavity (severity of concave portions of the contour)
8. Concave points (number of concave portions of the contour)
9. Symmetry
10. Fractal dimension ("coastline approximation" - 1)

Medicos: Breast Cancer Diagnosis

Medicos is a medical web application which does Breast Cancer Diagnosis by classifying the cells into **Benign & Malignant** based on the Input Parameters!

What are Benign and Malignant cells?

Benign

"Benign" tumors or conditions are non-cancerous and do not invade nearby tissues or spread to other parts of the body. They typically grow slowly and are usually considered less harmful than malignant tumors. In the context of breast cancer, benign tumors are not cancerous and do not pose the health risks.

Specified Input Parameters

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	me
0	14.1273	19.2896	91.969	654.8891	0.0964	0.1043	

Prediction of Diagnosis

Predicted Diagnosis: There are signs of the presence of Malignant cells.

> There are symptoms of cancerous cells. If you also feel the following symptoms then seek immediate medical help!

1. Fatigue.
2. Shortness of breath.
3. Anemia.
4. Diarrhea.
5. Weight loss.
6. Drenching night sweats.
7. Abnormal lumps or bumps.

[Learn more about Malignant cells](#)
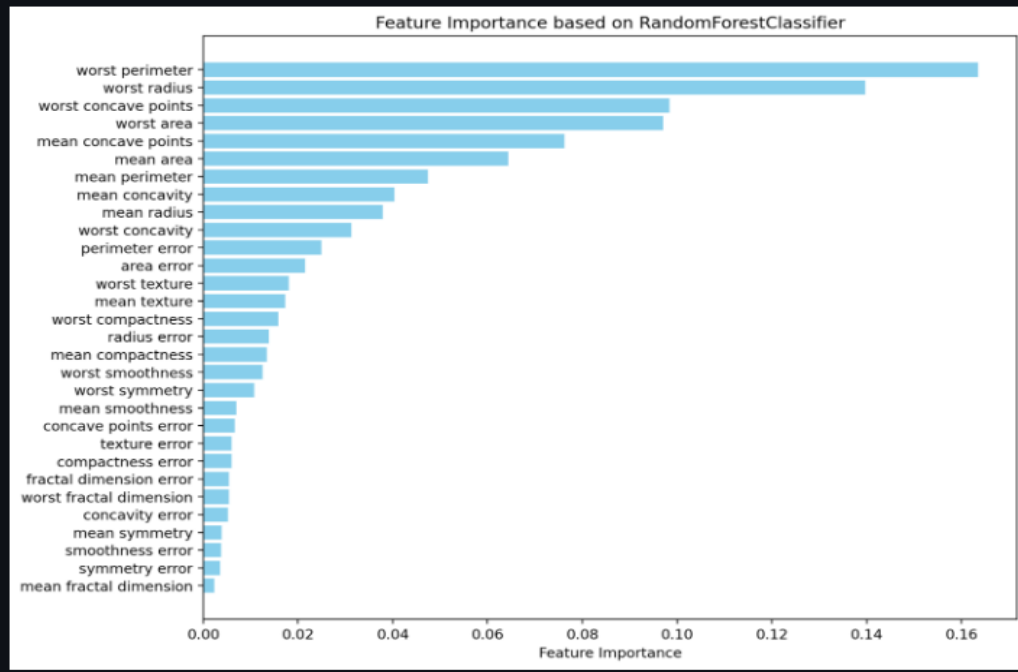
Specified Input Parameters 🔗

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	me
0	14.1273	19.2896	91.969	654.8891	0.0964	0.1043	

Prediction of Diagnosis

Predicted Diagnosis: There are signs of the presence of Malignant cells.
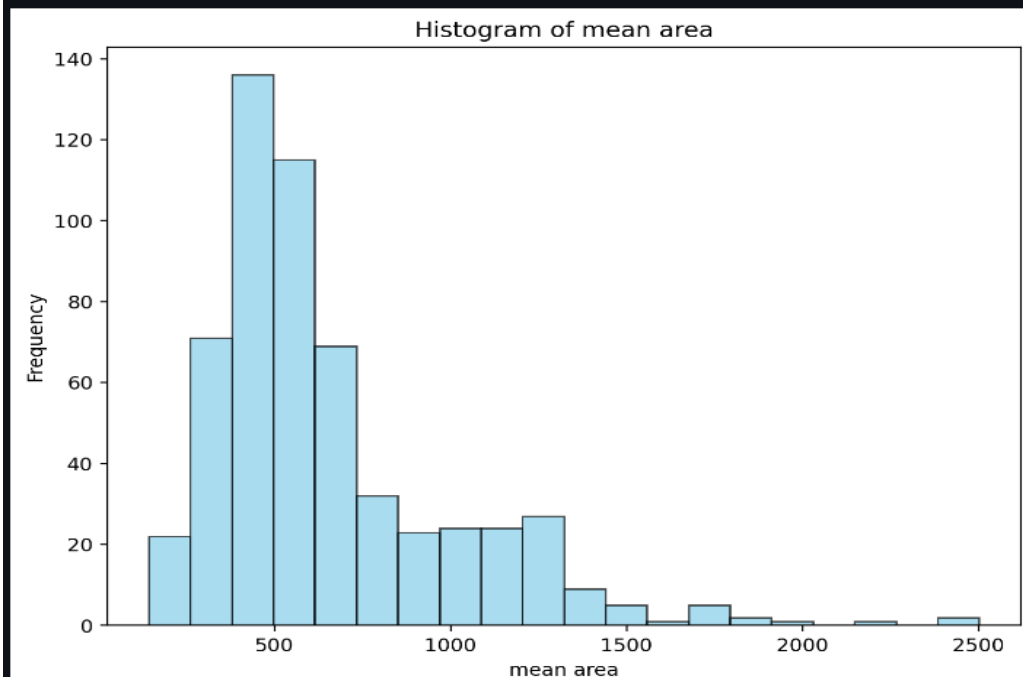
Plots

Feature Importance

Feature Importance based on RandomForestClassifier

Histograms of Selected Features

Select feature for histogram

mean area

Histogram of mean area

Scatter Plot 🔗

Select X-axis feature

mean texture	⌄

Select Y-axis feature

mean area	⌄

Scatter Plot: mean texture vs mean area