Breast Cancer Detection Project

Reem El-Habbaa

Table of contents

01 Introduction

This project is dedicated to the early detection of breast cancer through the application of machine learning techniques, Which focuses on constructing a predictive model that evaluates various indicators to accurately distinguish between:

- **Benign Tumors (Non-Cancerous)**
- **Malignant Tumors (cancerous)**

02 Dataset

For this machine learning project, we are utilizing a comprehensive breast cancer dataset sourced from **Kaggle**. This dataset is instrumental in training our model to identify patterns and make accurate predictions.

The dataset includes

- **Target: Diagnosis**
- **Variety of Features: Revealing generally the tumor size, shape, and cell characteristics**

Having Zero null or duplicate values

03 Preprocessing

Normalization

To ensure that our model treats all features equally, we've applied MinMaxScaler to normalize the dataset.

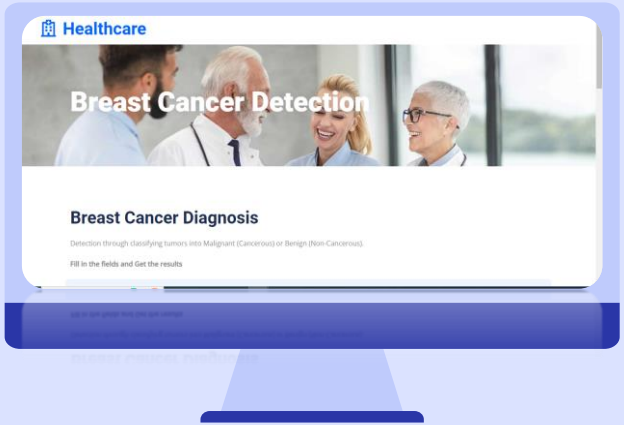
Encoding

Since our features are floating-point numbers, KBinsDiscretizer is used to encode them into discrete intervals.

04 Model

In this project, we've implemented **Logistic Regression** to predict breast cancer occurrences (Categorical target: Diagnosis), the result can be one of 2 classes: Benign or Malignant.
Having an **Accuracy score equals 97.37%** and performance as the following metrics:

Class	Benign	Malignant
Precision	99%	95%
Recall	97%	98%



Thank You

Reem El-Habbaa