

Data Science Project

Finish Off School

February 18, 2024

Abstract

Breast cancer remains a significant global health challenge, affecting millions of lives each year. In the pursuit of more effective diagnostic tools and personalized treatment strategies, the integration of machine learning techniques has emerged as a promising avenue. This project aims to harness the power of machine learning regression for breast cancer prediction, leveraging comprehensive datasets and cutting-edge algorithms to develop accurate models that can forecast tumor behavior and patient outcomes.

1 Project

The project is based on a UCI Machine Learning Repository, in one of [its challenge](#). It is based on a dataset of Breast Cancer patients with Malignant and Benign tumor. The goal is to perform regression to predict whether the given patient is having Malignant or Benign tumor based on the attributes in the given dataset.

The student will have to study the machine learning contents in [this repository](#) to learn the basic tools of data science.

Once this is done, the student will then work on the [project script](#) and follow the following steps:

1. Ensure the dataset is well loaded and that the feature selection is well made
2. Code a feature scaling
3. Split data for training and testing.
4. Try different regression methods and write a short report to comment the results

