VIEHGROUP

High Level Design (HLD)

Breast Cancer Prediction

Revision Number: 2.0

Last date of revision: 5 July 2022

Sarthak Wakchaure

VIEHGROUP

Document Version Control

Date Issued	Version	Description	Author
16/06/2022	1.0	Added Introduction and General description	Sarthak Wakchaure
20/06/2022	1.1	Added Design details and abstract	Sarthak Wakchaure
23/06/2022	1.2	Created and organized whole document	Sarthak Wakchaure
25/06/2022	1.3	Updated Design Details	Sarthak Wakchaure
28/06/2022	1.4	Checked and added Performance	Sarthak Wakchaure
1/07/2022	1.5	Added Conclusion and References	Sarthak Wakchaure
5/07/2022	2.0	Updated diagrams in Design Details	Sarthak Wakchaure

VIEHGROUP

Table of Content

Abstract

Women are seriously threatened by breast cancer with high morbidity and mortality. The lack of robust prognosis models results in difficulty for doctors to prepare a treatment plan that may prolong patient survival time. Hence, the requirement of time is to develop the technique which gives minimum error to increase accuracy. All experiments are executed within a simulation environment and conducted in JUPYTER platform. The aim of the project is to create an early finding for breast cancer diagnosis. It is a machine learning project that makes predictions for benign and malignant tumors in the diagnosis of breast cancer.

VIEHGROUP

Introduction

1. Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project

2. Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly technical terms which should be understandable to the administrators of the system.

3. Definition

The terms used in the projects are:

- Malignant & Benign- Types of breast cancer
- Breast Cancer
- Machine learning
- Feature selection
- Classification
- Prediction
- Logistic regression

VIEHGROUP

General Description

1. Product Perspective

The breast cancer prediction webapp uses a regression model based on machine learning to help identify if a patient has malignant or benign cancer. Knowing what form of breast cancer the patient has can help the doctor to treat the patient appropriately.

2. Problem Statement

To develop a machine learning-based method for predicting a patient's cancer type—malignant or benign.

3. Problem Solution

Develop the web application to predict whether patient has either malignant or benign cancer based on given inputs.

4. Further Improvement

It is necessary to conduct additional study in this area to improve the effectiveness of classification algorithms and enable them to make predictions on a wider range of factors. My goal is to obtain high accuracy by parametrizing our categorization systems. I'm researching a variety of datasets and how more machine learning techniques might be applied to define breast cancer. I want to increase accuracy while lowering mistake rates.

5. Data Required

Data consisting of texture mean, compactness mean, smoothness mean, radius mean, symmetry mean, fractal dimension mean, smoothness SE, symmetry SE, symmetry worst, and texture SE are required for training the model.

Data depends entirely on our problem definition.

6. Tools Used

- Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn are used to build the whole model.
- Visual Studio Code is used as IDE.
- For visualization of the plots, Matplotlib and Seaborn are used.

- Heroku is used for deployment of the model.
- Front end development is done using HTML & CSS.
- GitHub is used as version control system.

7. Constraints

The Breast cancer prediction website should be user friendly. Different inputs should be given for different patient.

8. Assumptions

The goal of the project is to classify the cancer is malignant or benign. Also, which features help predict breast cancer. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

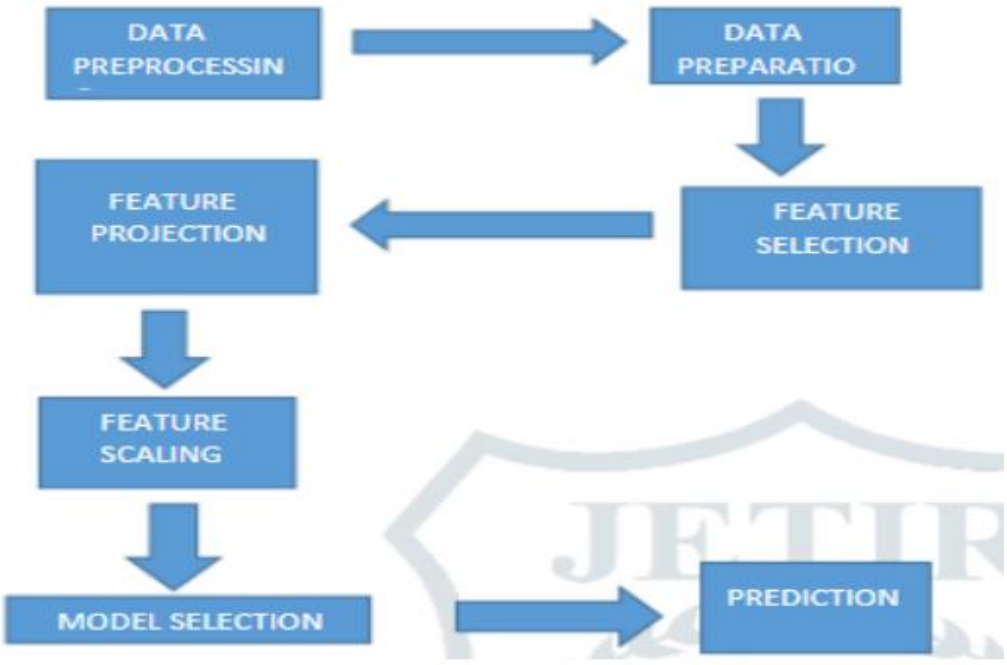
VIEHGROUP

Design Details

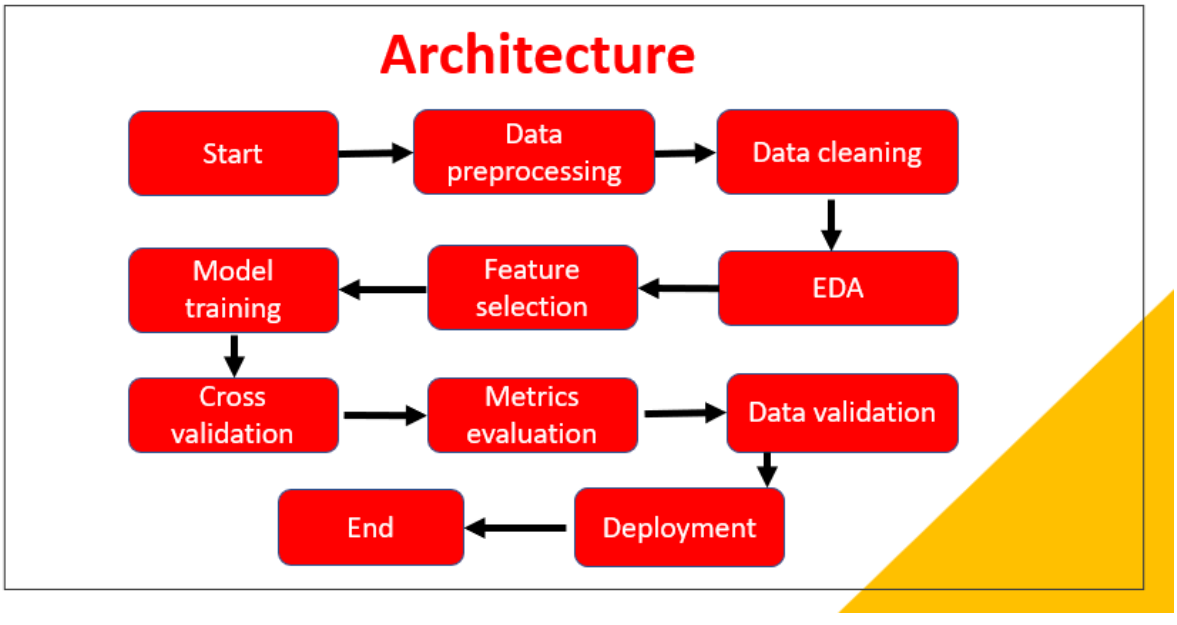
1. Process Workflow

For identifying the different types of anomalies, I will use a machine learning model. Below is the process flow diagram.

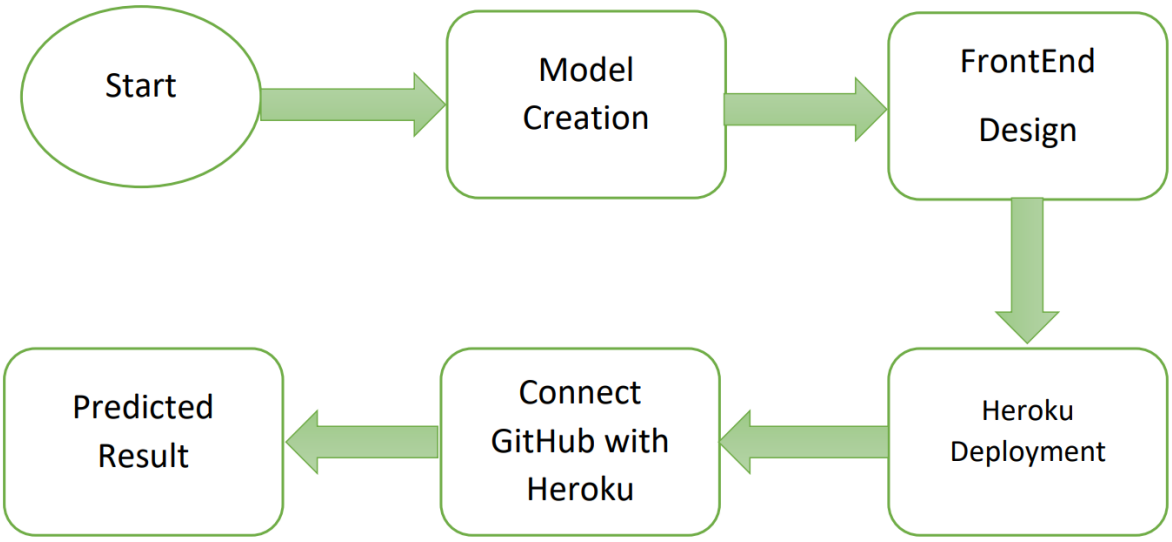
Diagram/Flowchart:



Model Training and Evaluation



Deployment Process



2. Error Handling

My attempt to link Heroku with GitHub first failed, but I was able to fix it. After connecting, I had a problem while showing the results on the website, which I fixed by using a youtube reference.

VIEHGROUP

Performance

1. Reusability

To avoid misleading the customer, breast cancer predictions should be as accurate as feasible. To forecast AQI, the best model available will be utilised. Our product adheres to reusability since we used HTML & CSS and because documentation is readily available.

2. Application compatibility

I adhere to Application compatibility because I am using Python for my project, which is cross-platform compatible.

3. Resource utilization

When we first started, we used high space to build the model. Our system just requires at least 2GB RAM and 1GB of storage once the model has been produced in order to operate the programme smoothly. System consumes less than 10% of its computing power whenever user attempts to forecast Breast cancer.

4. Deployment

The code is published on GitHub. The entire system is operational and is hosted on Heroku.

Conclusion

In this Python project, I developed graphs and results for the same breast cancer tumor predictor that we learnt to build using the Wisconsin dataset. A solid dataset has been shown to offer greater accuracy. The creation of prediction systems will result from the selection of appropriate algorithms with a strong home dataset. When a patient has been diagnosed with breast cancer, these systems can help determine the best course of therapy. Based on the stage of a patient's breast cancer, there are a variety of therapies available; data mining and machine learning may be a big help in selecting the course of therapy to be taken by extracting knowledge from such appropriate databases.

VIEHGROUP

References

1. https://ieeexplore.ieee.org/document/9544687
2. https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc20352470
3. https://www.geeksforgeeks.org/understanding-logistic-regression/
4. https://internship.viehgroup.com/courses/breast-cancer-prediction/
5. https://www.researchgate.net/publication/319688741_Breast_Cancer_Prediction_Using_Data_Mining_Method
6. https://seaborn.pydata.org/examples/regression_marginals.html
7. https://matplotlib.org/
8. https://pandas.pydata.org/docs/