The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document

1 Architectural Components Overview

1.1 Data Source and Repository

1.1.1 Technology Choice

I have fetched my dataset from kaggle.

1.1.2 Justification

Kaggle is a huge repository of datasets and we can easily find a plethora of data out there. This is the reason I chose Kaggle as my data source.

1.2 Discovery and Exploration

1.2.1 Technology Choice

I have used Python for my Data Exploration tasks. The version of Python used is 3.7

1.2.2 Justification

Python provides a very good environment to work with Image data and a lot of support is provided by the Python community.

1.3 Actionable Insights

During the course of my data exploration, I found the slide images to be very interesting (and challenging). Slide images may need an experienced examiner to be classified.
Also, the dataset is highly imbalanced where positive cases(cancerous cells) are very less compared to negative cases(non-cancerous cells).

1.4 Applications / Data Products

1.4.1 Technology Choice

The model has been saved as a ".h5" file using Keras. This saved model can be used for web-application or a software that can be used in Pathology Labs for quick and better predictions of Breast Cancers with less amount of type 1 error.