**Predicting and visualizing breast cancer diagnosis**

**Background Information:**

Breast cancer is one of the common cancers occurring in women. In the United States breast cancer is the most diagnosed cancer after skin cancer. The tests taken to find how much the cancer is spread is costly and painful. An approach towards Data Mining helps us reduce the cost and easier to predict the severity of the cancer.

**Data collection:**

The data is collected from the website called UCI machine learning repository. The data set is “Diagnostic Wisconsin Breast Cancer Database”.

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)>

**The Problem:**

We are trying to find out if the cancer is Malignant or Benign using Data Mining techniques with the data provided.

**Possible solution:**

Visualization techniques.

Logistic regression to predict the cancer.

Correlation matrix between variables.