**Chapter 3**

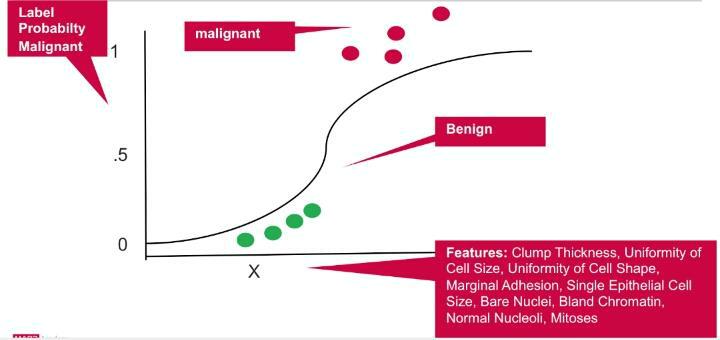
**ANALYSIS**

The Breast Cancer dataset is used for predicting the stage of the cancer in a person using Logistic Regression algorithm. The dataset consists of the information about the mean area, perimeter, radius, compactness, smoothness, etc. While working on this kind of data, which feature is important should be seen. The main aim is to make a model which can give us a correct prediction on the current stage of the cancer. Logistic Regression will be used for this dataset and checked if this gives us a good accuracy or not.

**3.1 Purpose**

The purpose of this project was to gain introductory exposure to programmatic data analysis concepts by analyzing the factors that determine the stage of the cancer based on the inputs given by the user. The project makes heavy use of NumPy, Pandas and Logistic Regression algorithm.

**3.2 Scope**



**Figure 3.1 Logistic Regression model**

This project has a scope that is allowed to maintain the datasets, it is used to train it, relevant and critical, for the application to run. It uses the Logistic Regression to predict the



Department of CSE, RNSIT 2018-2019 10