**CANCER**

**Data Description:**

The columns present in this dataset are:

* **Age:** This gives description a person belonging to which age.
* **Smoking:** This gives the description of a person is whether a smoker or not.
* **Yellow\_Fingers:** This gives the description of a person is suffering from yellow fingers or not.
* **Anxiety:** This gives the description of a person is suffering from anxiety or not.
* **Peer\_Pressure:** This gives the description of a person is suffering from peer pressure or not.
* **Chronic Disease:** This gives the description of a person is having any chronic disease or not.
* **Fatigue:** This gives the description of a person is suffering from fatigue or not.
* **Allergy:** This gives the description of a person is having allergy or not.
* **Wheezing:** This gives the description of a person is whether suffering from wheezing or not.
* **Alcohol Consuming:** This gives the description of a person is whether a alcoholic or not.
* **Coughing:** This gives the description of a person is whether suffering from coughing or not.
* **Shortness of Breath:** This gives the description of a person is whether suffering from shortness of breath or not.
* **Swallowing Difficulty**: This gives the description of a person is having any swallowing difficulty or not.
* **Chest Pain:** This gives the description of a person is whether suffering from chest pain or not.
* **Outcome:** This gives the description of a person is whether a cancer patient or not.

**Data Analysis:**

* First we need to load the packages like ‘caret’,’pROC’,’mlbench’,’e1071’ and ‘InformationValue’ to do the predictions in R.
* Then the data is loaded into the R file by using read.csv function.
* After loading the data we need to use the str function to know the characteristics of the data we loaded.
* If the output is of categorical type then we can convert the output into factor levels.
* By using sampling we need to split the data into training and test data.
* After that we need to fit the data to kneighborsclassifier, LogisticRegression and SVC to predict the new values output.
* While using knn we need use traincontrol to get more accurate results.
* On comparing jupyter with R it is easier to do predictions in R.