**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 import certain packages like pandas,numpy,seaborn,matplotlib,

Kneighborsclassifier, train\_test\_split, LogisticRegression and SVC functions into the notebook.

* Then the data is loaded into the jupyter notebook by using read\_csv function.
* After loading the data we need find the datatypes of data objects and need to convert them into the categorical data.
* As we are not getting the correct output from normal dataframe we need to convert the input to an array and then split
* After that we need to fit the data to kneighborsclassifier, LogisticRegression and SVC to predict the new values output.
* To find the accuracy of this methods we need to use the accuracy function.
* By comparing the accuracies of different methods with different test data size we can know the best method of approach.