

|  |
| --- |
|  |
| Tumor Cancer Prdediction |
| AI Project |



|  |  |
| --- | --- |
|  | Decorative |
| Team Members |
| * Ali Sameh Saad * Abdalla Maged Mohamed * Yousef Ahmed Bekheit * Merna Tarek Rashed * Heba Hossam Mohamed * Samaa Salah eldeen Attia * Abdelmigeed Hatem Abdelmigeed |



Supervisors

* **Dr. Alaa Tarek**
* **Dr. Dina Elsayyad**



Introduction

This Project aims to identify using a set of data if the body cell is benign or malignant using Machine Learning models as decision tree, Random Forest , Naïve bayes , logistic regression and SVM using python language so we can detect then that that cell is a cancereous cell or not.



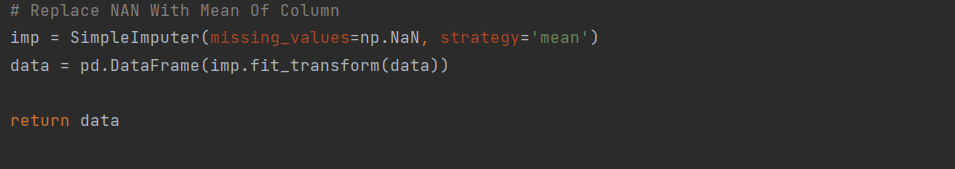
First…preprocessing:

Preprocessing isn’t just a main process inside this project, but It’s an essential step in any machine learning project in order to give better accuracy predictions as by removing useless data cells in the data set as duplicate, null values ,and outliers.

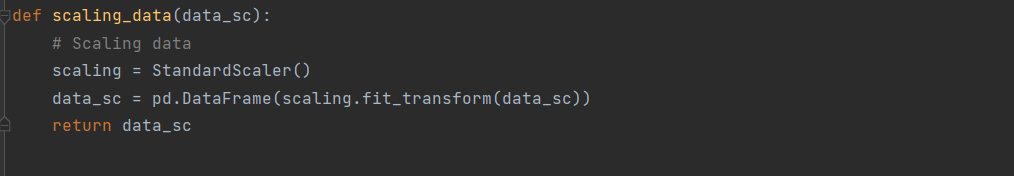
* **Data Cleaning**

Code cleaning started with changing string values inside the data set into zeroes and ones(m🡪1,B🡪0)using **label encoder** function, then unnecessary columns were removed using **drop** function, then dropping outliers using outliers and **IQR** statistics rules, then replacing null values with mean of coloumn.





* **Data Scaling**

****

Data scaling is done to decrease the big gap and difference between values of the data set so It was scaled using statistics and **Z-score** rules.



Second…. Data modelling

This project has Five models to get the best prediction of cancer cells accuracy using SVM, logistic regression, Linear Regression, Decision tree, and random forest.

* **Logistic regression**

Text

Description automatically generated

First, adding parameters to increase accuracy of predicted data, then using fit function to fit x\_train, Y-train data then detecting the accuracy of each train data and test data.

* **SVM**

Text

Description automatically generated

This model uses a line as a margin space for two dimensions while it uses a plane as a margin space for more than two dimesnions

* **Decision tree**

**Text

Description automatically generated**

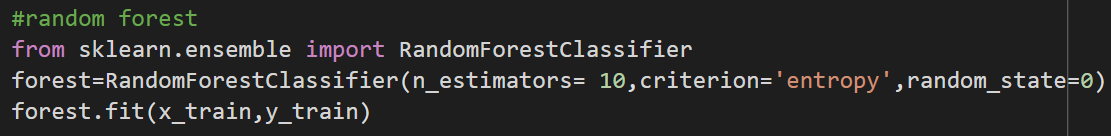
This model enables user to branch into smaller data sets using the entropy equation :

A picture containing text, clock, watch

Description automatically generated

Also, parameters used was random state was set it to “0” in order to stop random values to “0”, then fit the x\_train , and Y-train values.

* **Random Forest**

****

This model is considered a multitude of decision trees or average of each individual trees using the parameters : random\_state same as decision tree above, entropy equation

* **Naïve bayes**

Text

Description automatically generated

This model is mainly based on probability rules especially bayes rule to get the highest probability predictions.

P(x|y) = [P(X) (intersect.)P(Y)]/P(Y)