Predicting the MALIGNANCY OF BREAST CANCER

**A Project Report for Long Term Industrial Training**

###### ***Submitted by***

DIGANTA DAS (CSE/2ND YEAR/GNIT)

BISHAL KR. SHAW (CSE/2ND YEAR/GNIT)

RAJVIR SINGH (CSE/2ND YEAR/GNIT)

***in partial fulfillment for the award of the degree of***

##### **B.Tech**

in

**Computer Science And Engineering**

Guru Nanak Institute Of Technology

****

At

**Ogma TechLab Pvt. Ltd.**



**Ogma TechLab Pvt. Ltd.**



**BONAFIDE CERTIFICATE**

Certified that this project work was carried out under my supervision

“Predicting the MALIGNANCY OF BREAST CANCER”

is the bonafide work of

***Name of the student: Diganta Das Signature:***

***Name of the student: Bishal Kumar Shaw Signature:***

***Name of the student: Rajvir Singh Signature:***

Name : Sourav Karmakar

**PROJECT MENTOR**

**SIGNATURE**

**Ogma TechLab Original Seal**

**Acknowledgement**

We take this opportunity to express my deep gratitude and sincerest thank to our project mentor, **SOURAV KARMAKAR** for giving most valuable suggestion, helpful guidance and encouragement in the execution of this project work.

We will like to give a special mention to my colleagues. Last but not the least we are grateful to all the faculty members of Ogma TechLab Pvt. Ltd. or their support.

CONTENTS

1. **Introduction**
   1. Problem Statement
   2. Objective
   3. Algorithms applied on it
2. **Method**

2.1 About the Breast Cancer dataset

2.2 Steps Followed

2.3 Accuracies obtained by applying algorithms

**3. Result OF Different MODELS**

3.1 Validating the Training Accuracy

3.2 Validating the Testing Accuracy

**4. Conclusion**

**5. Bibliography**

**Introduction**

**Problem Statement :---**

To build multiple Classification Algorithms and compare which one is performing better in the dataset of “MALIGNANCY OF BREAST CANCER”

**Objective :---**

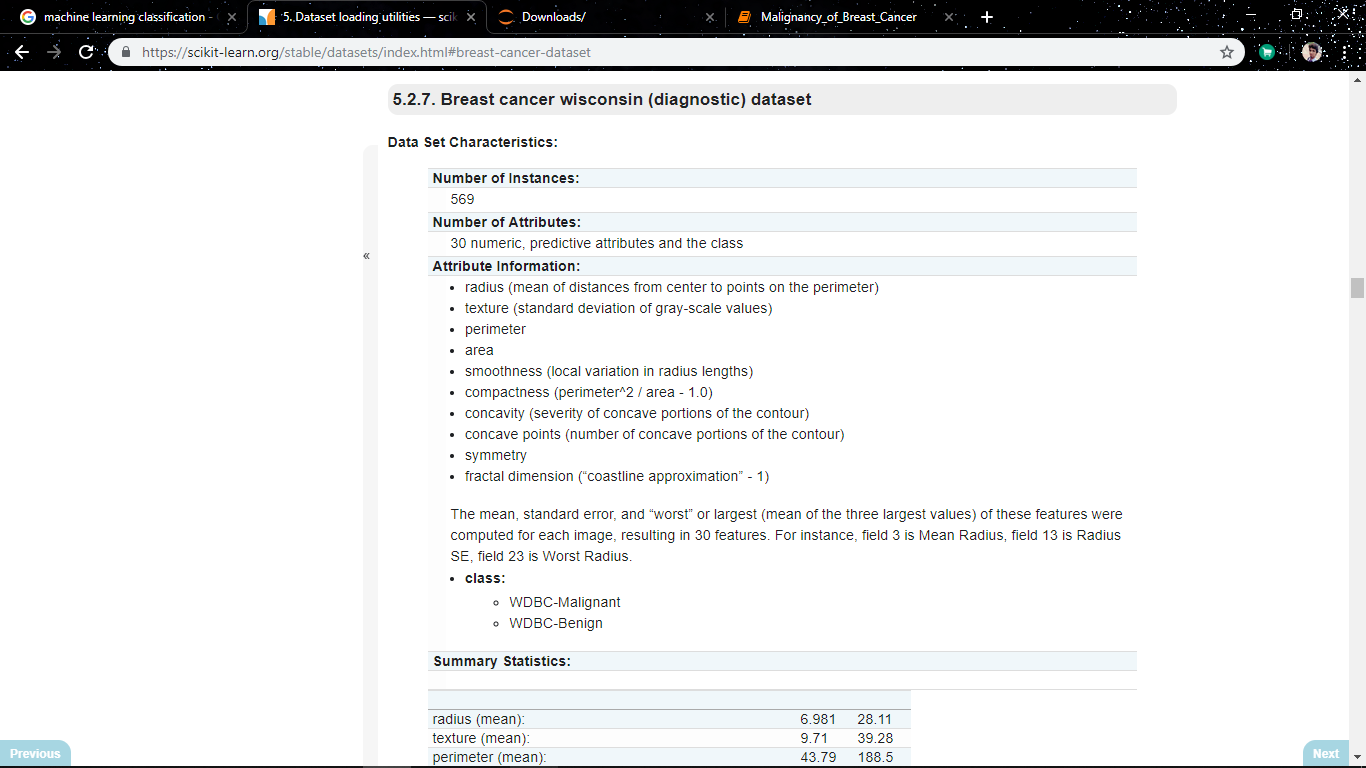
In this project, we were provided with a dataset of Breast Cancer and here we have to predict the MALIGNANCY OF BREAST CANCER by using different types of classification algorithms like decision tree, logistic regression, k-NN and SVM. We have applied this algorithms and made a comparative study of accuracies obtained from different classification algorithms applied on that dataset.

**Algorithms applied on it :---**

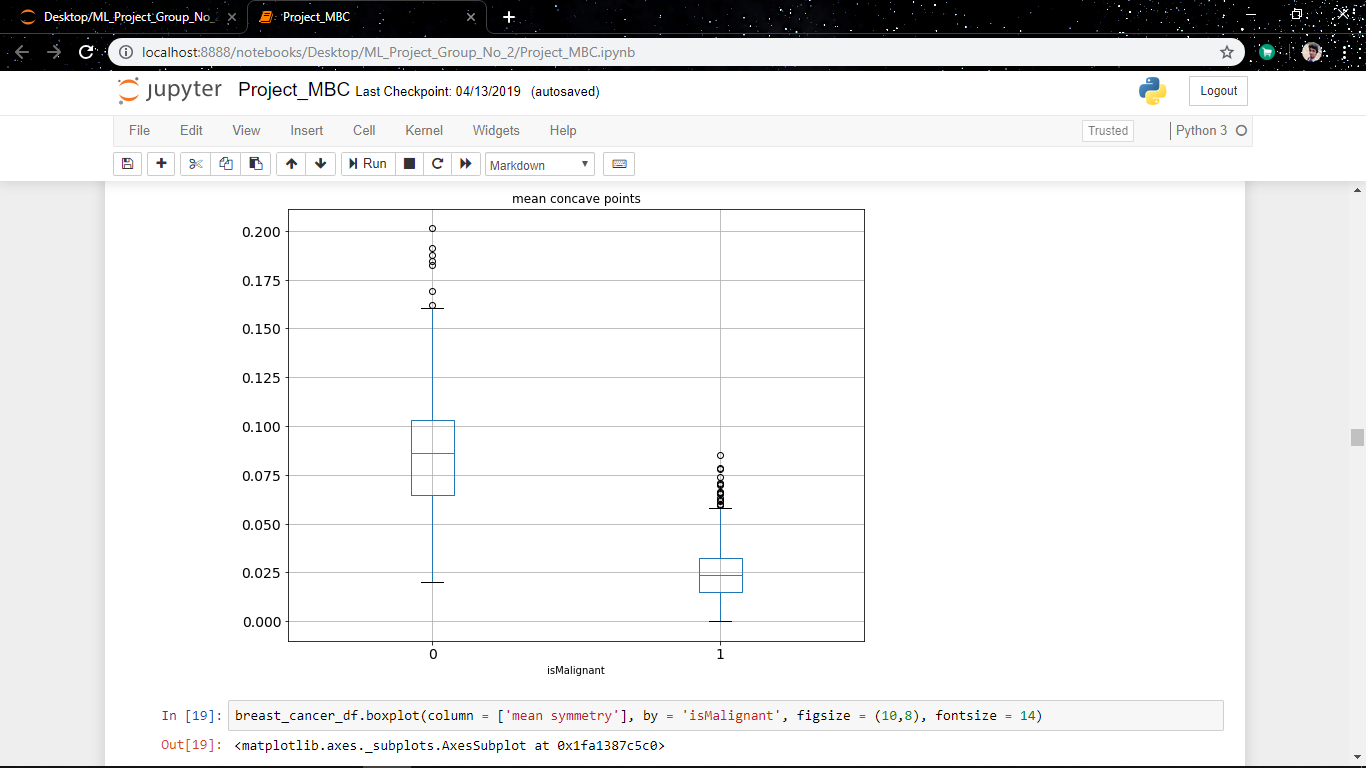
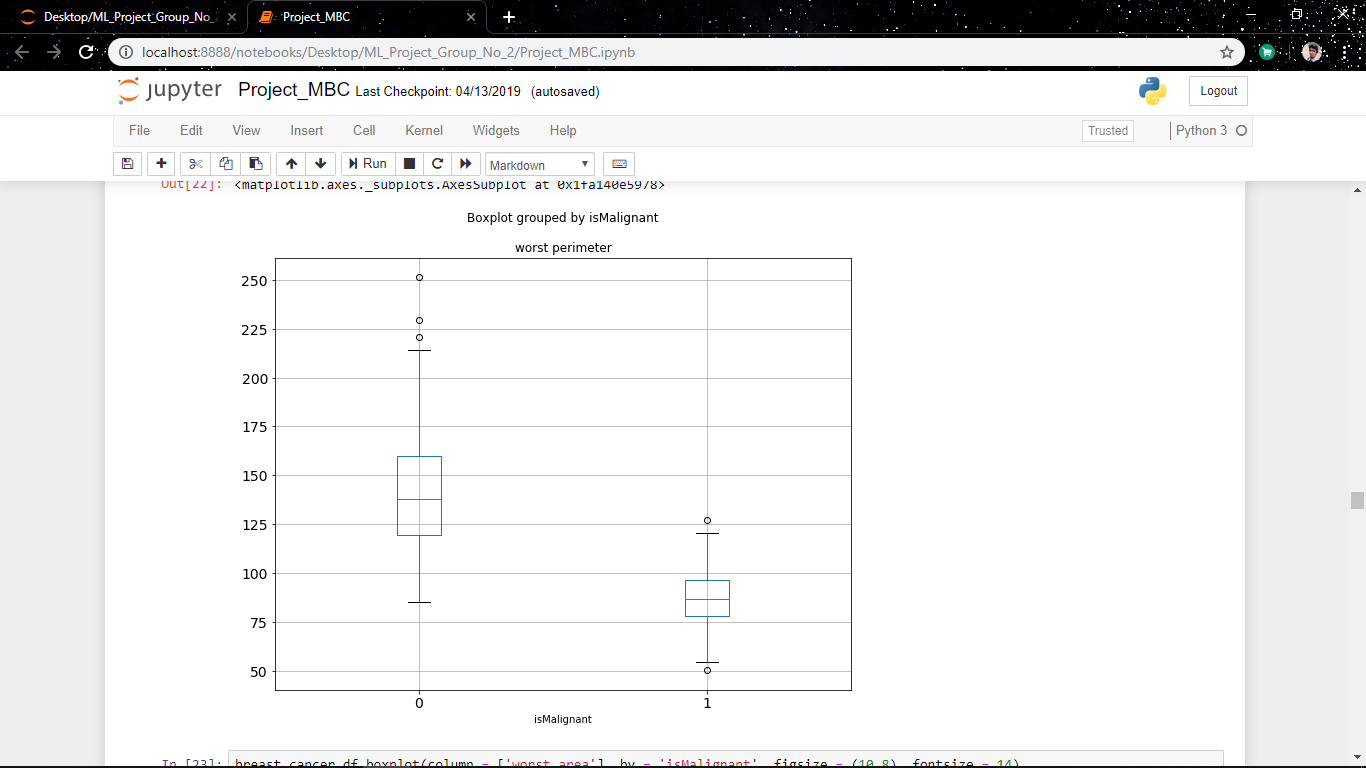
* Decision Tree
* Logistic Regression
* K-Nearest Neighbour (k-NN)
* Support Vector Machine (SVM)

**Method**

**About the Breast Cancer dataset: ---**

****

**Boxplot with different features against isMaligant**



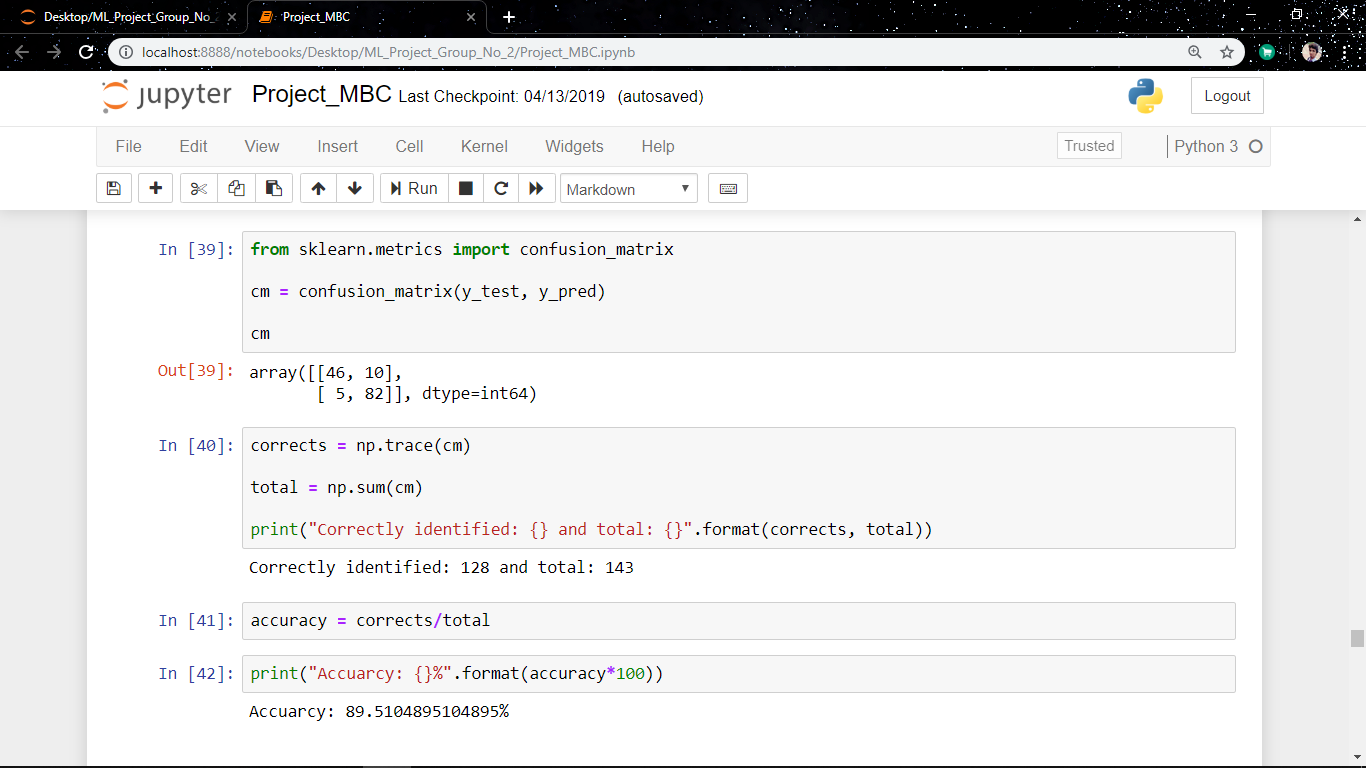
**Steps Followed: ---**

1. At first we analysed the data and explained it by boxplotting with isMaligant against all the features.
2. Then we applied all the classifier one by one on this dataset.
3. We calculated the accuracies for all the classification algorithms.
4. We compared the accuracies and tried to find out that which classification algorithms fits the best on this dataset.

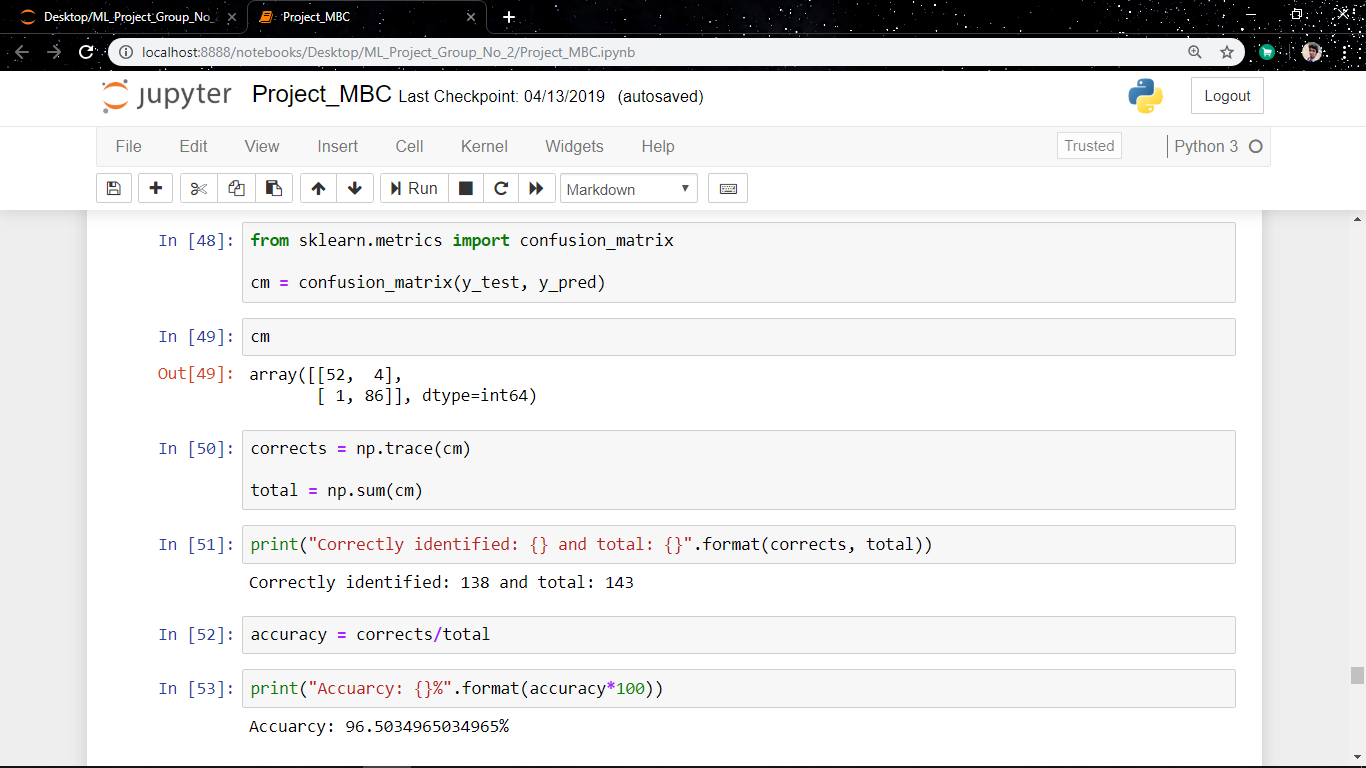
**Accuracies obtained by applying algorithms:----**

**Result OF Different MODELS**

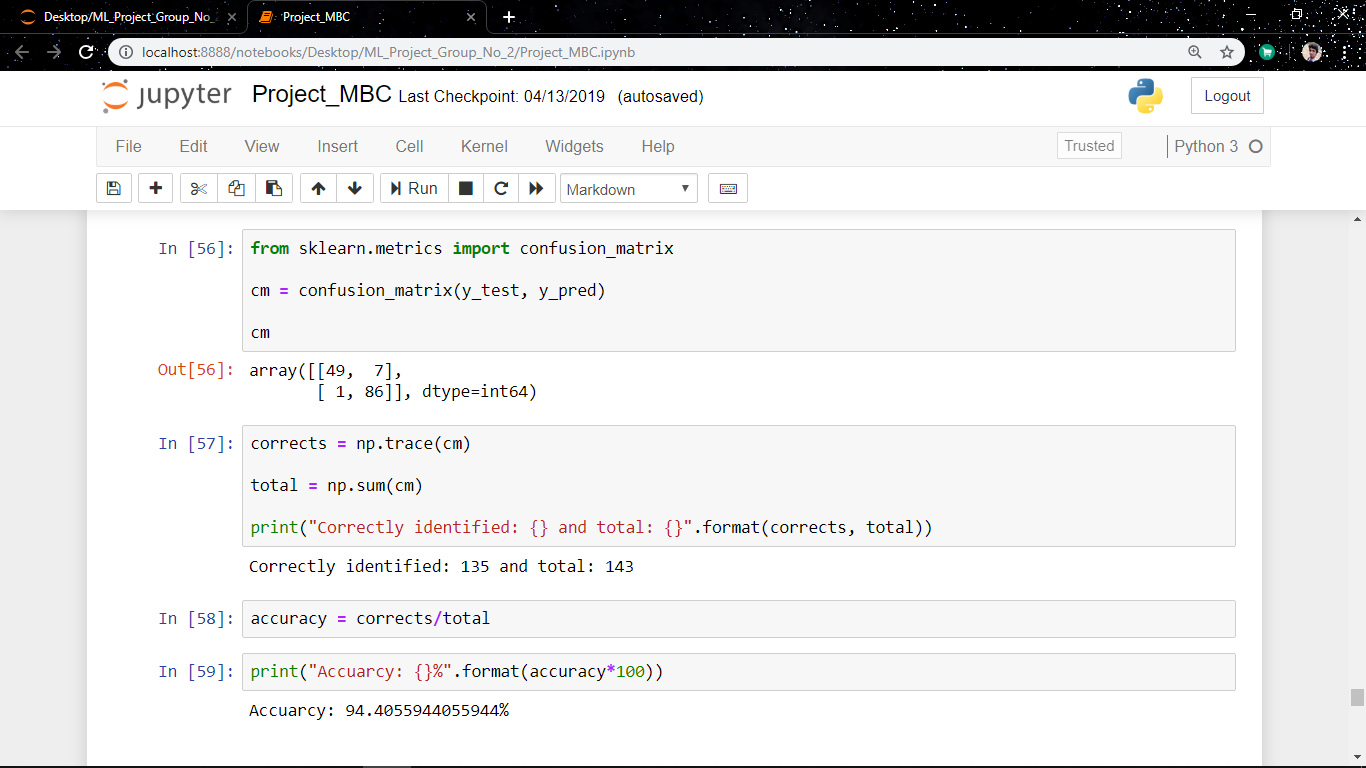
1. **Decision Tree:---**



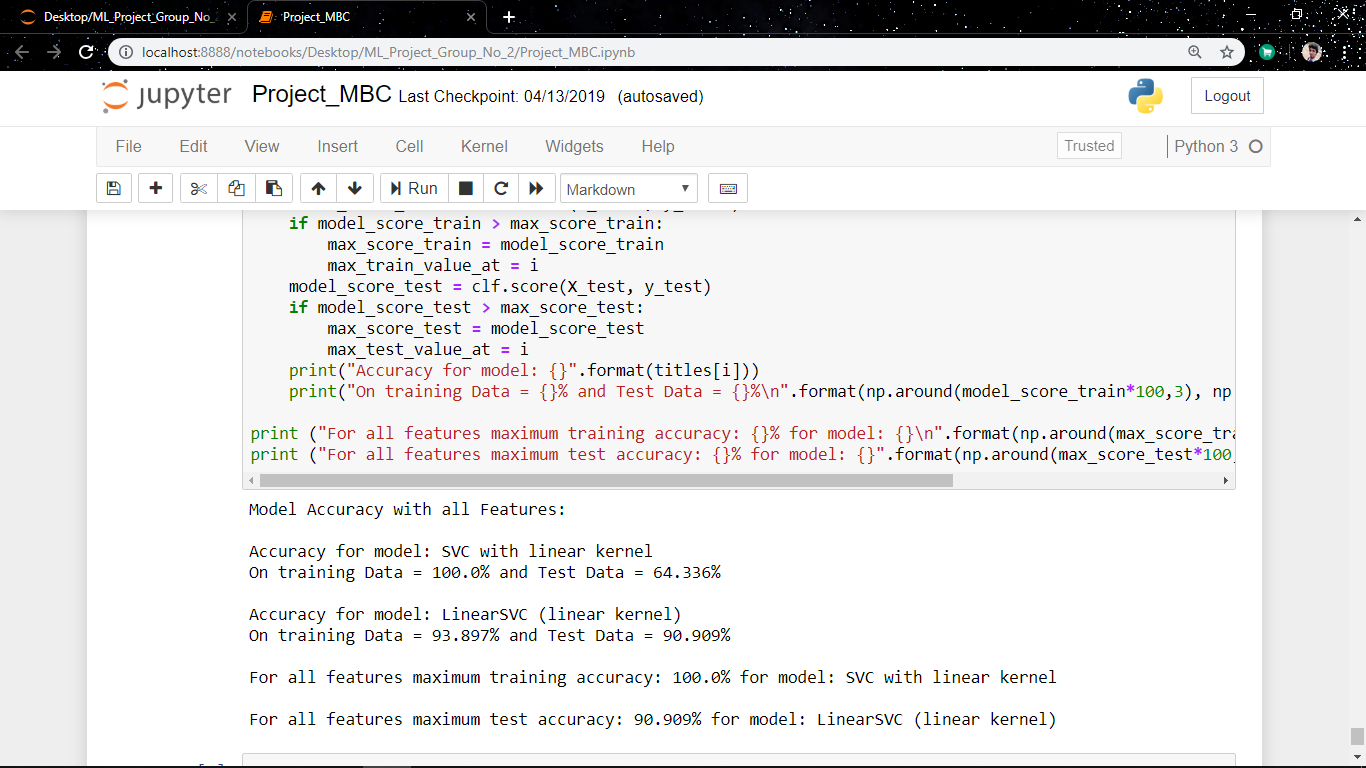
1. **Logistic Regression:---**



1. **K-NN Classification:---**



1. **SVM Classifier:---**



**Conclusion**

From this project we conclude that:-

* On this dataset we have differentiated the group of people coming under Malignant or Benign.
* In the given dataset, the Logistic Regression classifier works the best as the accuracy obtained is more than the other classifiers.
* Decision Tree as a classifier is showing the least accuracy then the other classifiers on this dataset.
* K-NN classifier and SVM classifier is also giving good accuracies on this dataset.
* We can also apply this classification algorithms on other datasets and obtain the best algorithm working on that dataset.

**Bibliography**

1. <https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html#sklearn.datasets.load_breast_cancer>
2. Machine Learning Modules.

………………..\*……………………