Y2 ECLC20203 Elective: Artificial Intelligence and Machine Learning 22-23

**Breast Cancer Diagnostic Report**

Name: Harun Chowdhury

Student ID: 97146721

Date: 13/12/2022

Subject: Artificial Intelligent

**CONTENTS**

3

INTRO .....................................................................................................................................

3-4

ARTIFICAL INTEILIGENCE ..................................................................................................

LINEAR SEPRIBILTY ...........................................................................................................

4-5

ARCHITECTURE/MODEL ...................................................................................................

5-6

TRAINING & TESTING ........................................................................................................

6-7

CONCLUSION .......................................................................................................................

REFRENCES .......................................................................................................................

7

8

4

DATASET .................................................................................................................................

**Artificial Intelligence and Machine Learning**

Task

develop an application able of learning breast cancer data so that it can be used to create a breast cancer diagnostic tool for classification of malignant and benign biopsies.

Artificial Intelligence

Artificial intelligence is the skill to perform and mimic the way a human process information and task and replicate it through a computer system. This could be things such as visual perception, decision making or speech recognition, the purpose is to make day to day tasks easier.

When talking about AI there are 4 main types of AI used today, these consist of : theory of mind, reactive, self-aware and, limited memory.

Theory of mind AI is the able to make true decisions that are similar to humans, theory of mind would be able to comprehend and remember emotion and according to this, this will adjust its behaviour with the person communicating with it.

Reactive AI was a huge step in the ai world, this type of ai is unable to react with humans, unable to react to the world and lack imagination and will always have the same outcome in a situation it was previously been in. but these ai were such a huge step because it is the best for technology that don’t need external input such as self-driving cars or stopping spam emails from coming into our emails.

Self-aware AI is a type of AI that’s aware of their own emotion and emotions of others around it. This type AI is comparable to a human. Self-aware ai will also have needs, this type of AI hasn’t been developed yet but is a concept.

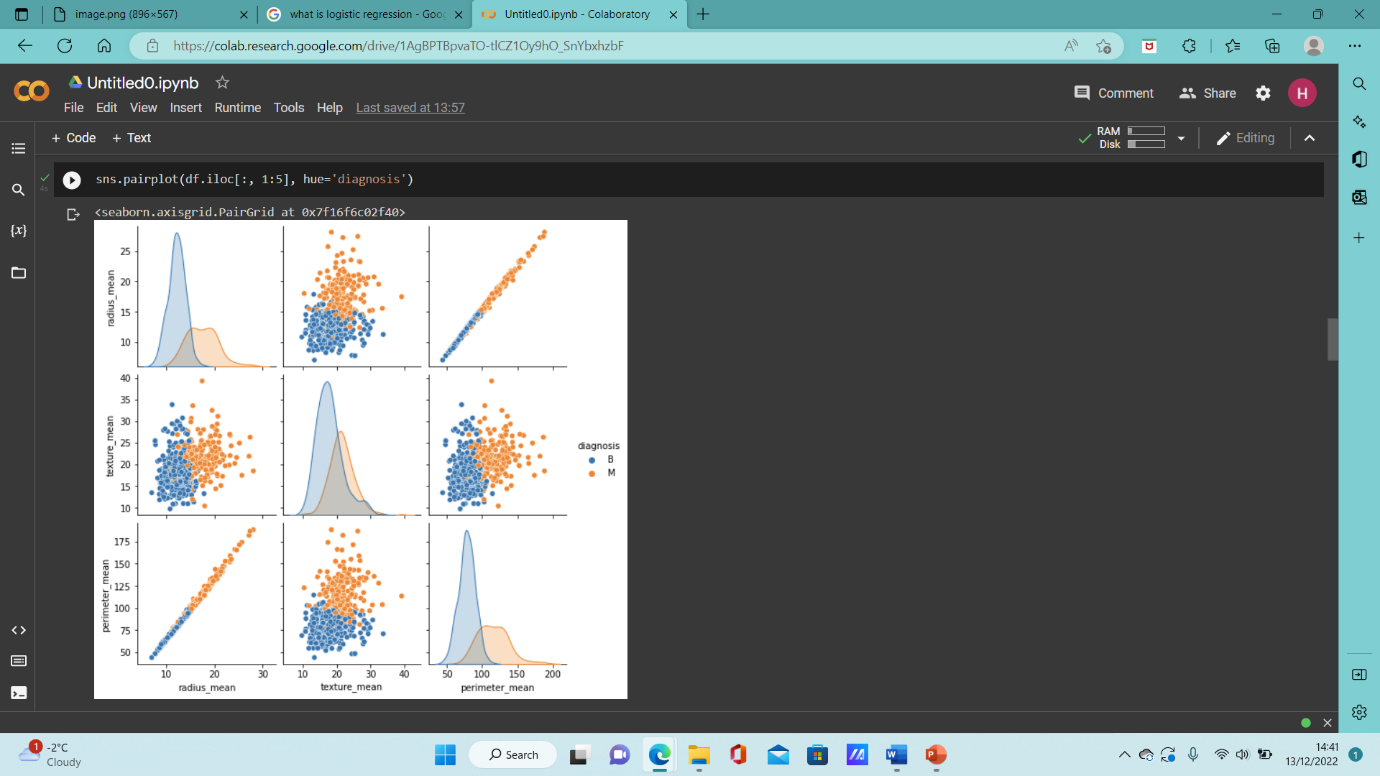
Limited memory ai is one of the most widely used AI used to this date. In limited memory AI it learns from the previous data and actions and constructs knowledge by monitoring activities or information. Limited memory uses observational, past data to make predictions and execute difficult tasks. This can be shown in self-driving cars, these monitor other cars speeds direction its traveling to help understand the road and make decisions depending on these factors. But suggesting by its name this AI type is limited because its not able to save data that it has monitored or previously used.

I have chosen to use reactive ai for this task, this is because it reacts with the data given to it and gives out the same result as if it was either benign or malignant. Compare to the other AI types I deemed this to be the best suited because it only would have 2 answers and for all data will give either or results.

Linear Separability

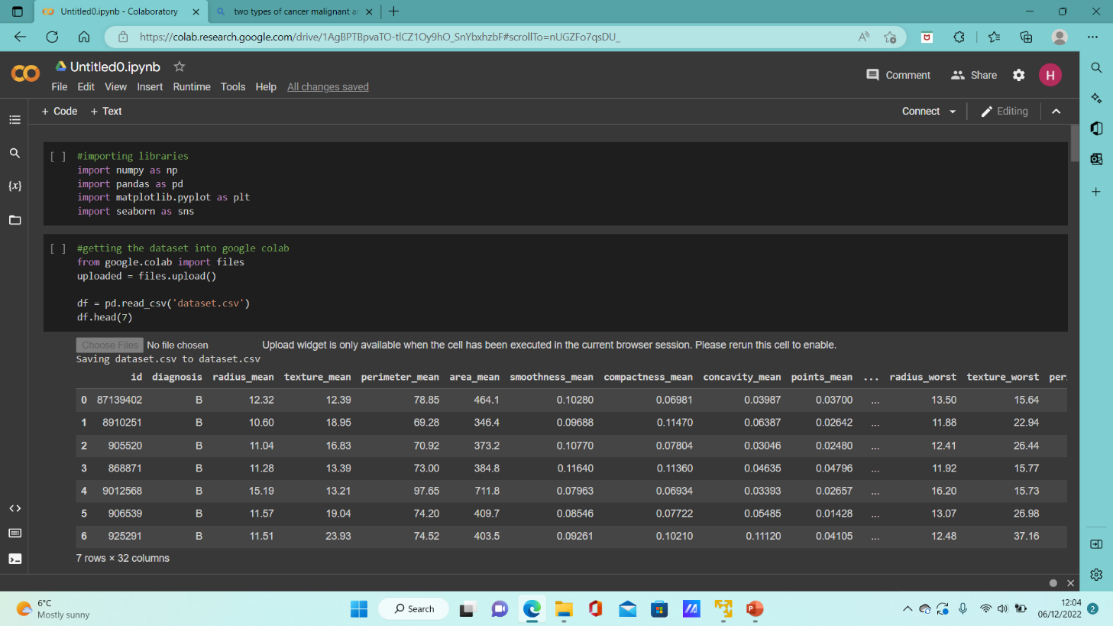
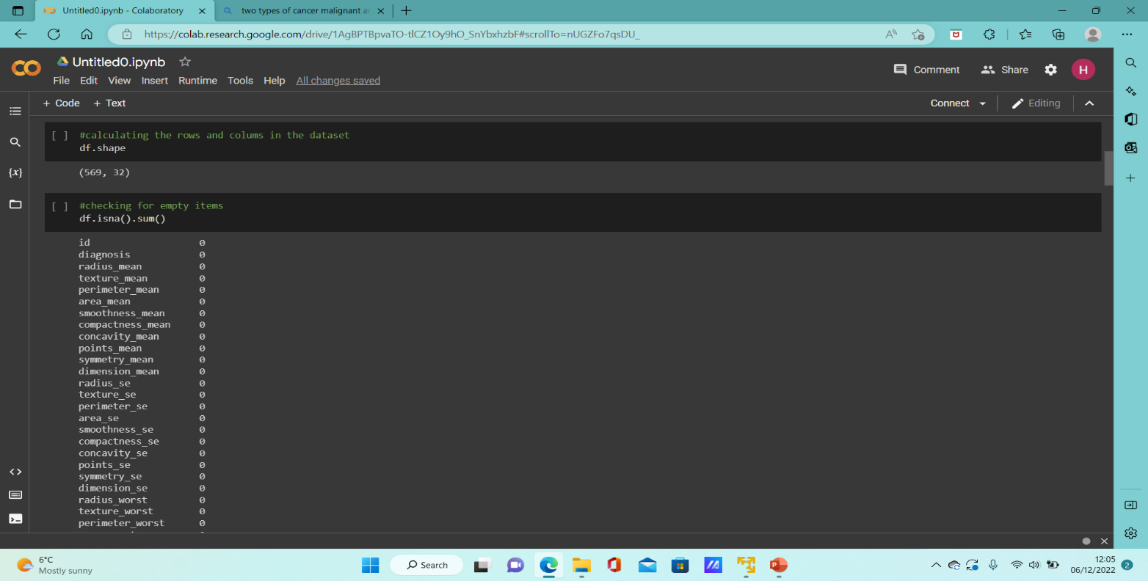
Linear separability implies that if there are two classes then there will be a point or line that splits the input features in such a way that all results of one class are in one space and the second class is in the other half, in terms of this task that will mean splitting benign and malignant.

This dataset I am currently working on is not linear separable because you can’t separate both results with a line. This is also show on the diagram below and you can clearly see you can’t separate the malignant vs benign.



The Database

The database I have used for this task I have got from the university, I didn’t need to clean it because all unnecessary data and values wasn’t in the code. To further check this I ran a code to check for any null values just after I imported the database int google collab.



Architecture/Models

There are numerus numbers of learning algorithms to choose from each having there own way of training and testing.

A decision tree algorithm is just like a flowchart, this is because all the nodes stands for the test on an input attribute and branches stands for the outcome of the test.

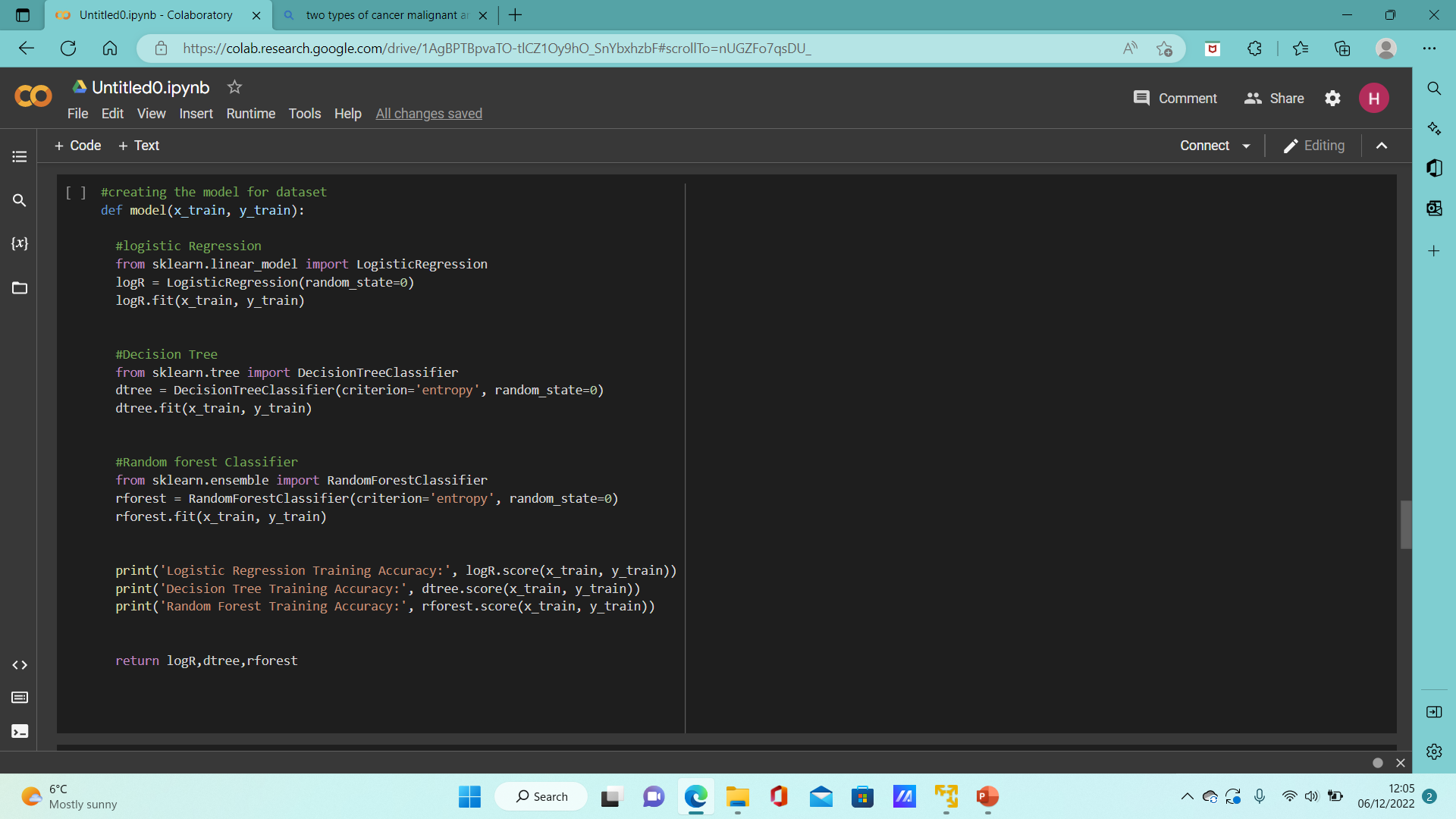
Random forest just like decision tree but is a collection of decision trees working together. The inputted dataset is split and entered into different decision trees. The average of all the results is considered and a final result is printed out. Random forests is more accurate compared to Decision tree algorithm and this is also showing through my testing on my project.

Logistic regression is an example of supervised learning, it is commonly used for calculating or predicting the probability if a yes or no question or event that occurs. This basically means any input you may us it will always have only 2 answers to output

I decided to do 3 models to test and train this dataset. I decided to do 3 to see which one was the most accurate within testing and training. The 3 models I chose were : decision tree, random forces, and logistic regression. I decide to do these specific ones because I wanted to compare classification and regression between these models because logistic is a regression and both decision and random forces can be used for classification.

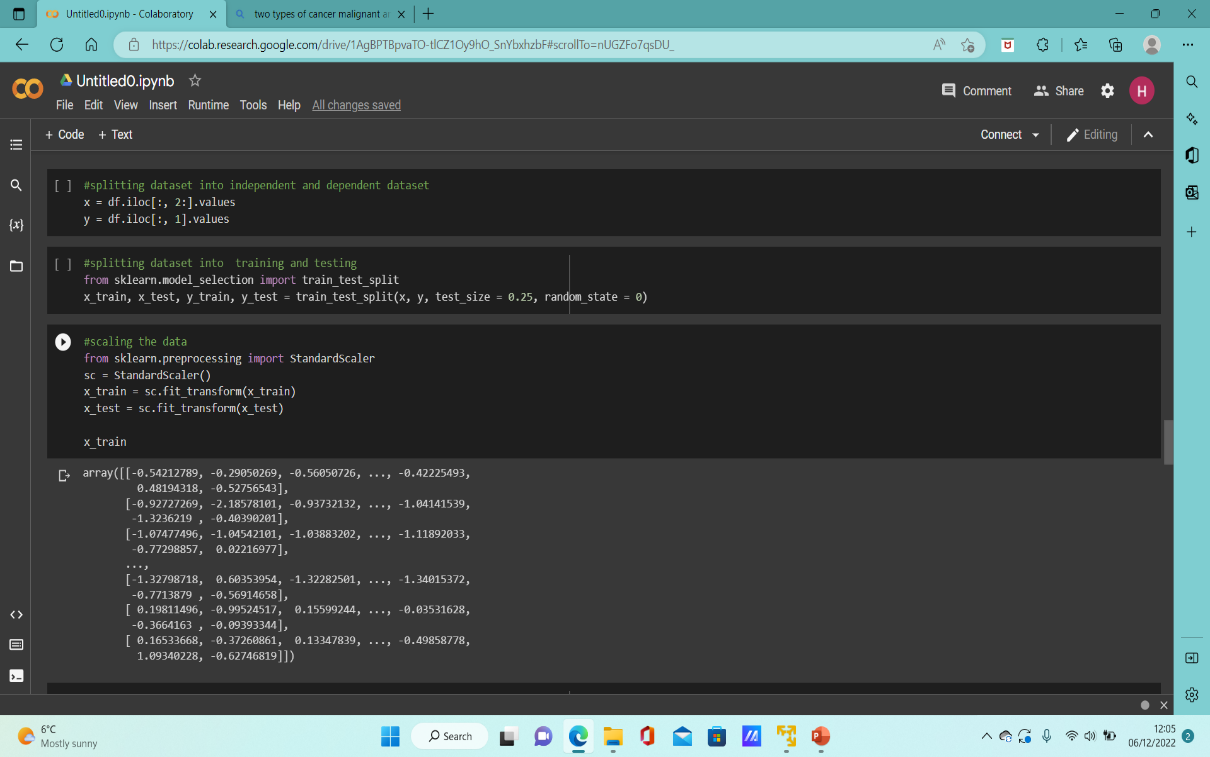
All these models are supervised learning models. In supervised learning, the system tries to study from the earlier examples provided. Compared to unsupervised which is a type of machine learning method used to find similarities in the data.

From this I can say these 3 models are the best for training and testing for breast cancer because you can compare which model is better, all are supervised learning so they already have a outcome for this instance either be malignant or benign, allowing all results to be more accurate than a unsupervised learning model.



Training and Testing

Before do the actual training and testing I had to change I the diagnostic result from the dataset from B & M to 1&0 so it this could work become the algorithms I chose mostly understand binary inputs and outputs. Then I split up the dataset into 2 sections, independent and dependent to make it easier for my code to read the different aspects and can split the data up for training and testing. Training I put 75% and 25% for testing.

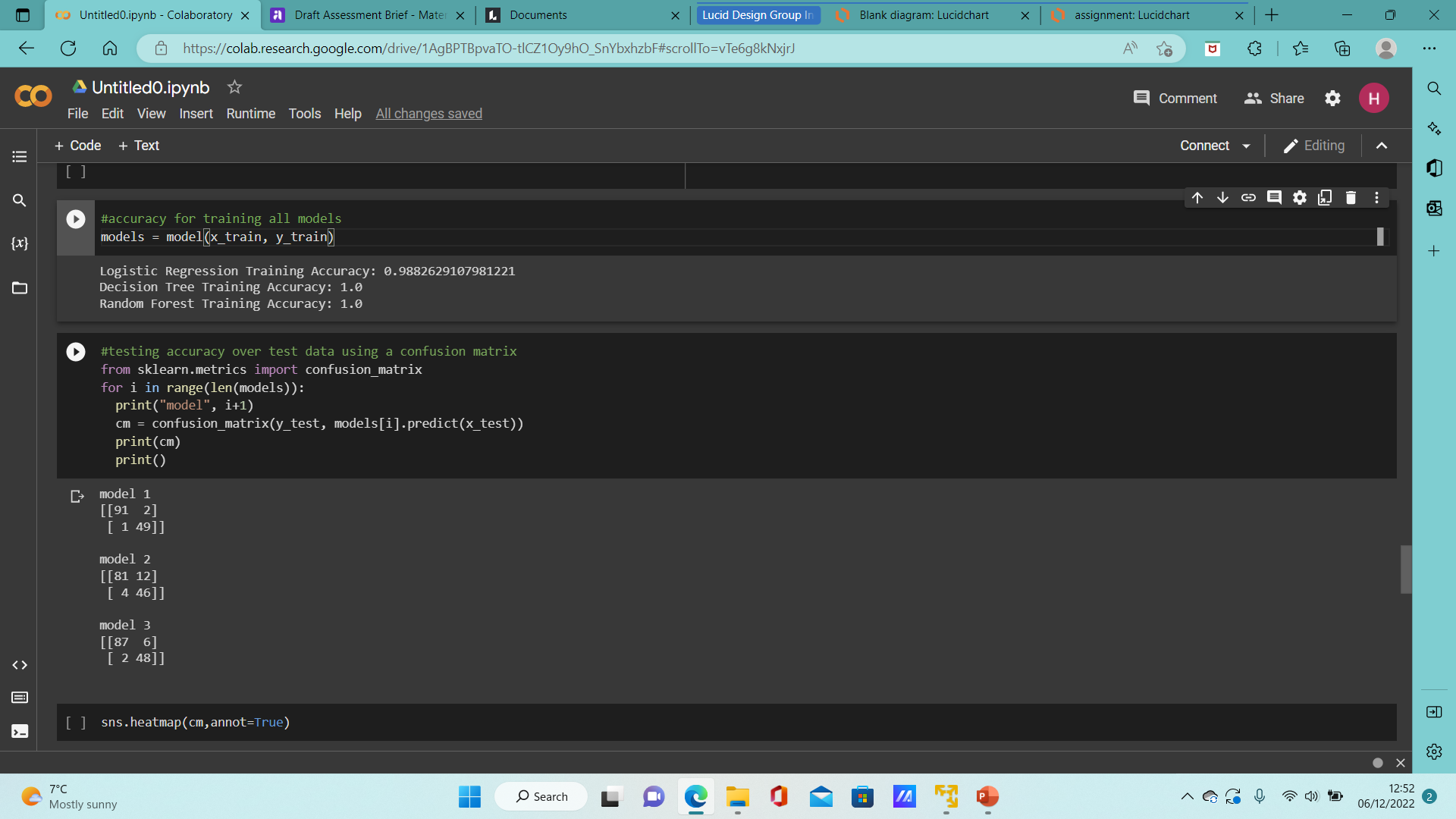
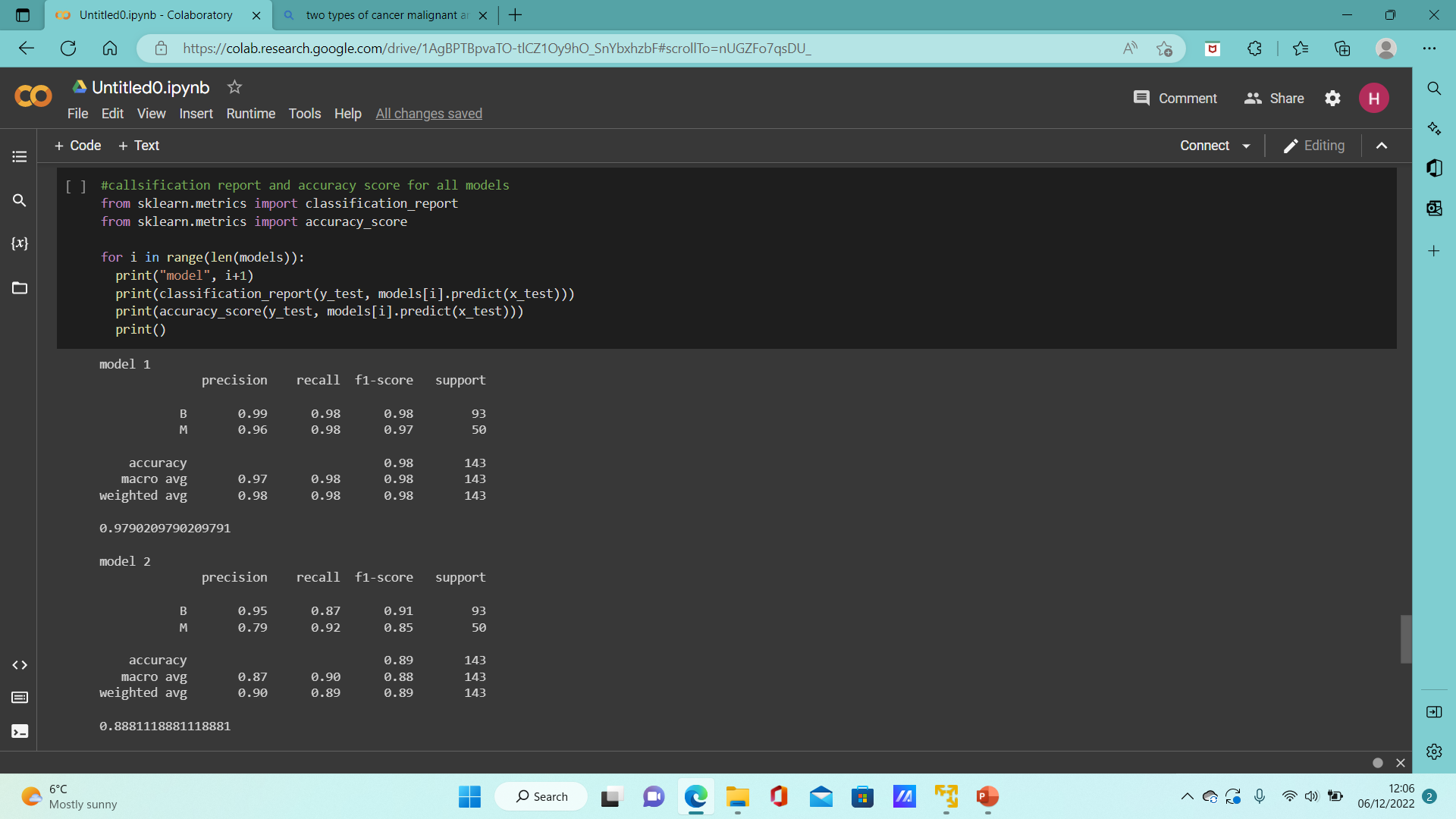
I first printed out the accuracy of training, and training onto of testing to see how all models preformed, this was showing in a confusion matrix, then printed out the accuracy of testing. From the results I could see training accuracy was much higher that testing.

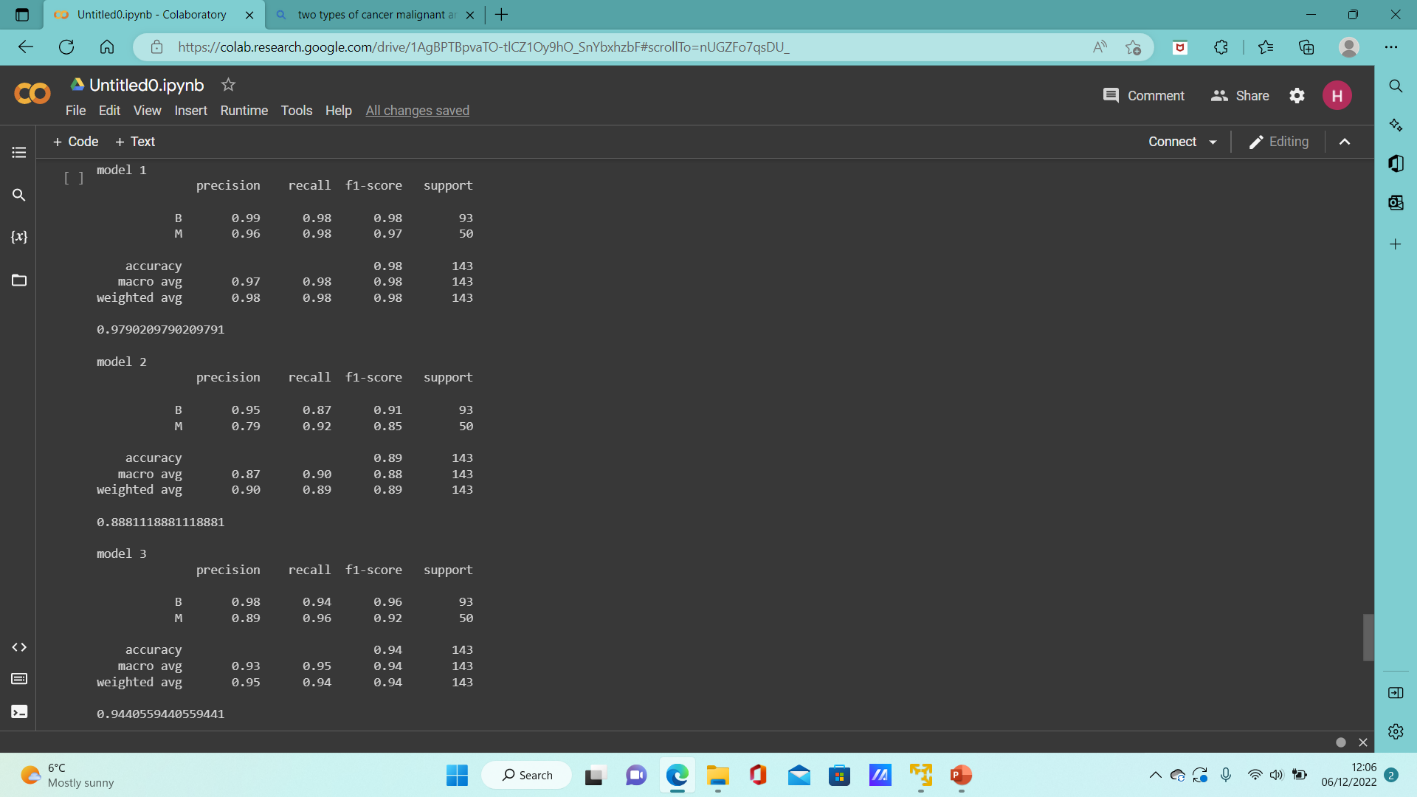
Top Left: Splitting data

Top Right: Accuracy on training and training onto of test

Bottom Left: Code for testing accuracy

Bottom right: Accuracy





Conclusion

This task was made to make detecting cancer easier and more accurate. This task was made for the medical word and was trained and tested to see which algorithm works the best and is the most accuracte with set given data, thorught this task I would say logistic regression is the best because the accuracy on training was above 90% and trainging ontop of testing gave a error of 2 which is the lowest between the three I have tested.

Furture plans for this task is to test bigger datasets and to see other spectrams that this could be used in and implamented for.

Reference

Marr, B. (2021) *Understanding the 4 types of Artificial Intelligence (AI)*, *LinkedIn*. Available at: https://www.linkedin.com/pulse/understanding-4-types-artificial-intelligence-ai-bernard-marr#:~:text=The%20most%20basic%20type%20of,conceive%20of%20past%20or%20future. (Accessed: December 13, 2022).

*linear separability* (no date) *Packt subscription*. Available at: https://subscription.packtpub.com/book/big-data-and-business-intelligence/9781788830577/2/ch02lvl1sec26/linear-separability (Accessed: December 13, 2022).

Tavasoli, S. (2022) *Top 10 machine learning algorithms you need to know in 2023: Simplilearn*, *Simplilearn.com*. Simplilearn. Available at: https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article (Accessed: December 13, 2022).