# q1

Your task in this exercise is to classify the fake and real news spread in Twitter by using Decision tree and Random forest model. we have two datasets consists of various features related to the fake and real news. (377 fake news and 365 real news). you should implement the following tasks: (to implement the following tasks, use the [scikit-learn library in python.])(https://scikit-learn.org/stable/index.html)

Also Write a concise report and explain your code and results.
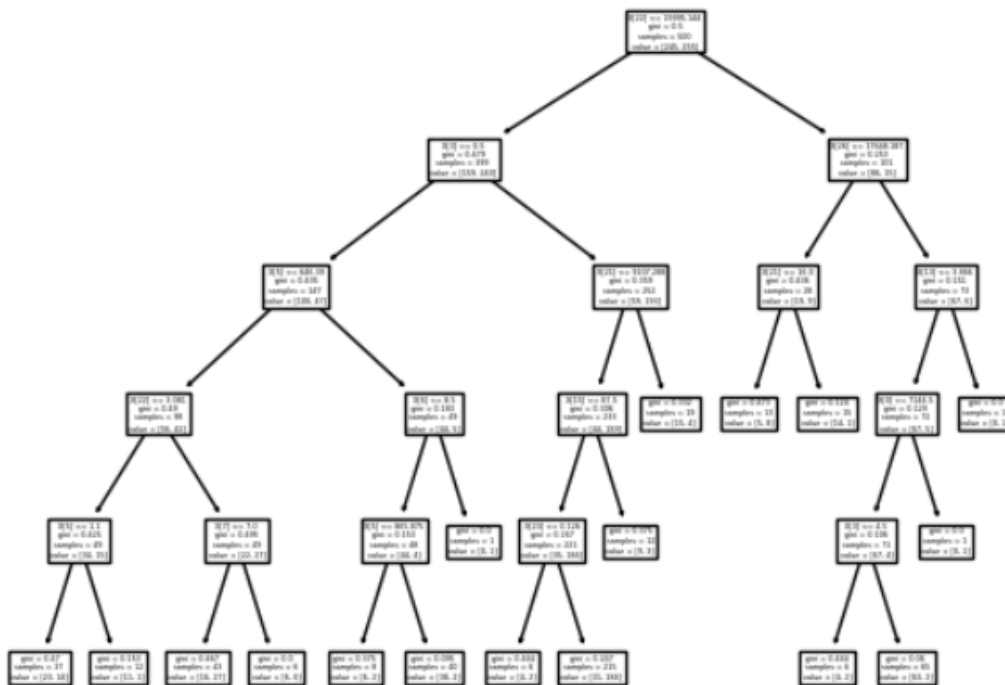
1-Read the two datasets and label fake news by 0 and label real news by 1. merge two datasets and shuffle data samples. Use 80 percent of data as training data and 20 percent of data as

test data.

2- In order to classify the news, use the Decision tree classifier with the Gini index method and compute confusion matrix, accuracy, precision, recall and F1 measures, finally as depicted figure, plot the constructed decision tree.

3- In order to classify the news, use the Decision tree classifier with the Information gain method and compute confusion matrix, accuracy, precision, recall and F1 measures, finally as depicted figure, plot the constructed decision tree.

4- In order to find the depth of the Decision tree in the range of [5,20] which achieves the best accuracy, implement 10-Fold Cross Validation method, and for each depth, compute the average accuracy and Determine which depth leads to the most accuracy. Finally compare the best achieved accuracy with the accuracy of two former tasks.

5- The Random Forest Algorithm combines the output of multiple (randomly created) Decision Trees to generate the final output. Write a document and explain how Random Forest algorithm works. (bonus)

6- Use the Random Forest algorithm and classify the news. Finally compare the accuracy of Random Forest model with the accuracy of three former tasks. (bonus)