Classification “Y is a categorical column”

-K Nearest Neighbor

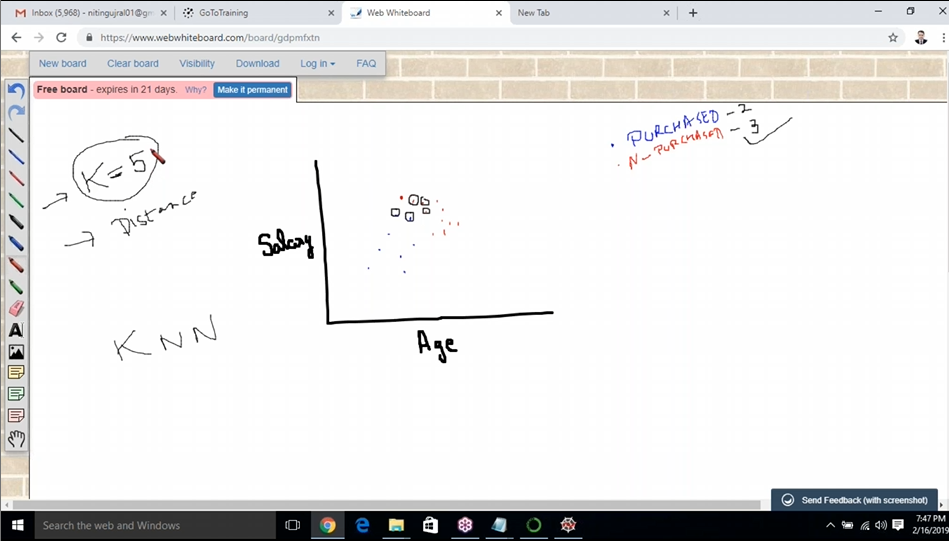
- Random Forest

Take out Garbage Data

Regression- Continuous Numerical Values

Classification categories(either string or number)

K nearest Neighbor Algorithm- Plots Salary against Age, while the dots are purchased vs not purchased, from that point, a distance is calculated to all other points. Then the 5 nearest points/neighbors. The greater category is the one predicted. If the points are equal, then it takes total distance from point, the less distance is the category it is classified as. K= amount of points you compare



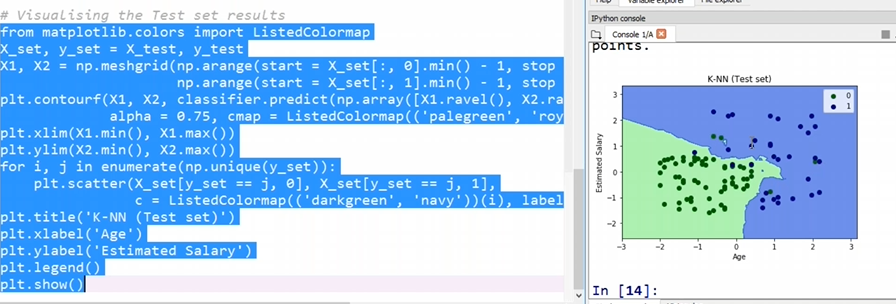
P=1 is Euclidian Distance

P=2 is Manhattan Distance

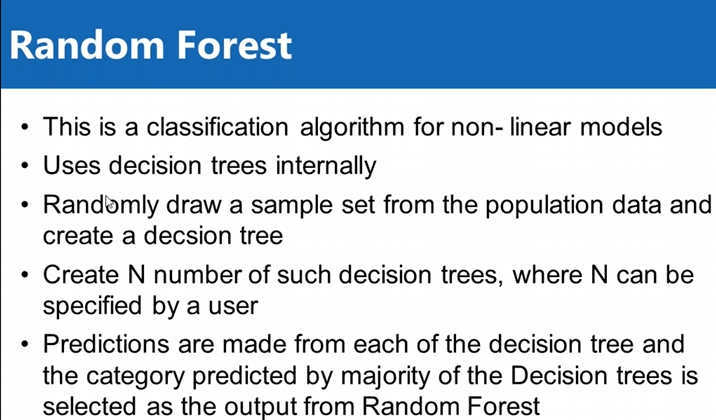
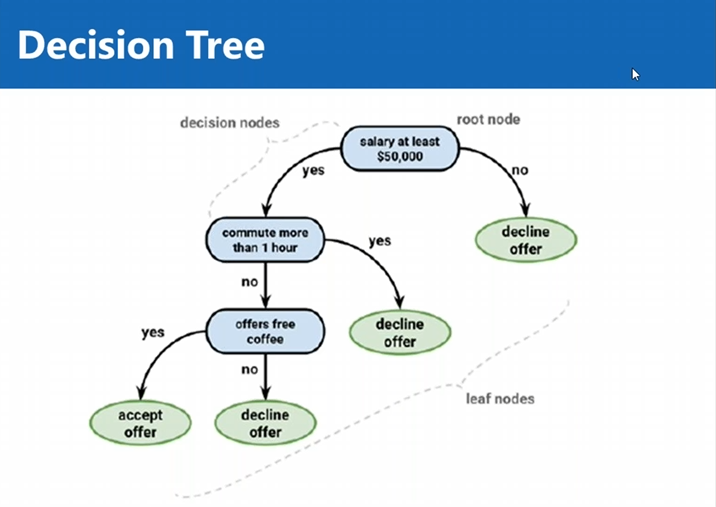
To check the accuracy of Classification- There is confusion matrix

Also ROC and AIC(for later)

(Learn Manhattan vs Euclidian)

  
There is also a Random Forest from classification,

A diagram of a tree

Description automatically generated with medium confidenceA diagram of a company's success

Description automatically generated

Maximize points of each category on each side of the line