 

Fran Miguens KNN Project

* Description of dataset: The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.
* Question 1: Using KNN, can we predict account balance using age, job, education, and housing?
  + Accuracy was low because there are too many possibilities for account balance.
* Question 2: Using KNN, can we predict jobs using age, marital status, balance, education, and housing?
  + Accuracy continued to be low because there were too many jobs.
* Lets try adding more variables to predict job.
  + This doesn’t work because some variables are confounding variables.
* Question 3: Can we predict Marital Status using KNN and KFold?
  + Precision improved, but is still below the desirable place
  + Maybe this dataset was never meant for KNN or KFold
* Question 4: Using Logistic Regression can we predict a) Marital Status and b) If they will take the loan
  + We get our by far greatest results. This proves that the dataset is best suited for logistic regression.
* Question 5: Just to understand the dataset better, can we visualize it using graphs?
  + Allow me to present the graphs on the board :)