CSC 4792 – DATA MINING

PROGRAMMING EXERCISE 4//2019

You work as a business intelligence consultant at an insurance company where they sale various insurance products. The company is planning to run a direct market campaign to sell a pension product to its customer base. In preparation for this campaign, the company seeks a business intelligence solution that will help to predict which customers are most likely to respond to this marketing strategy. The expected solution will be used to target the marketing campaign only at those customers that are likely to purchase the pension product. The company has provided you with data from previous direct marketing campaigns that list customer information-specifically the annual salary and age of the customer – and whether the customer bought a product after they had been contacted via a direct marketing message. You are tasked to train the k-Nearest Neighbor model (with Euclidean distance) on this data and advise the marketing department whether they should contact a customer with the following profile: *Salary = 56,000* and *age=35*.

TASKS:

i. Plot the provided dataset to visualize its distribution
ii. Develop and train the k-Nearest Neighbor model on the dataset.
iii. For k=3, what advice do you give the marketing department for the customer with profile: *Salary = 56,000* and *age=35*. Should they contact the customer?
iv. Answer question *iii* for the Naïve Bayes model.

Link to the dataset:
https://drive.google.com/open?id=1D8E8JDoIzlT_NbaI5TDvRw3l4-S4G4K2