Data Analytics and Applications

Week 6 Homework

Question 1 [5 points]
Research a predictive modeling technique or machine learning algorithm of your choice that we did **not** study in the Data Analytics and Application course.

Give a detailed description of the model/algorithm by addressing all of the following questions:

- is the model/algorithm designed to address supervised or unsupervised tasks (or both)?
- how is the model/algorithm fitted to the data (e.g., which loss function/criterion is used?)
- what are the parameters of the model? which, if any, are interpretable parameters? which, if any, are instead tuning parameters?
- what is the relationship between each tuning parameter and the tendency of the model/algorithm to overfit or underfit?
  (e.g. are large values of the parameters associated with overfitting or underfitting?)
- if the model is designed for supervised tasks, can it be used to meaningfully interpret the association between each predictor and the response variables? if so, how?

Question 2 [3 points]
Based on your own professional experience, describe a real industry problem in which the model/algorithm that you discussed in Question 1 can be used. In particular, address all of the following questions:

- who are the stakeholders in this problem?
- how does solving this problem by means of the chosen model/algorithm generate value for them?
- what kind of data needs to be available to effectively solve this problem using the model/algorithm of choice?

Question 3 [2 points]
In Questions 1 and 2, our reasoning was backwards: first we selected an algorithm, and then we thought of a problem that suits it. In this question, we will re-think about these two steps in a more appropriate order.

Given the problem of Question 2, and considering its context and the relevant stakeholders, what would be the most appropriate choice of model/algorithm to address it? Why? (This may or may not be the model/algorithm that you discussed in Question 1.)