

**Data Mining, CSCI 347, Fall 2019**  
**Essay Questions**

For the essay questions you may use your notes, the web and books. Do not discuss the problems with anyone aside from me. These questions do not have a single best answer. They will be graded on the logic, clarity and completeness of your answers. Ethical decision-making needs to be rationally addressed, using logical reasoning based on facts and commonly held values.

1. In one of two sentences, give the significant characteristics of each of the following learning algorithms. Include clustering, if your response to the second part of this question includes clustering. (5pts.)

Statistical learning (Naive Bayes)  
Decision trees  
Classification Rules  
Association Rules  
Linear regression  
Instance based learning

Presentations were made on the following special topics:

Stream learning, natural language processing, image mining, web mining (content, usage and structure), recommender systems, psycholinguistic data mining, mining in adversarial situations.

If web mining is seen as 3 separate topics, this gives 9 types of mining. For 4 of these topics, discuss the type of data mining algorithms discussed in class that are most likely to be helpful. In your discussion refer to characteristics that you have listed for the algorithm. (20 pts.)

2. In the book “Weapons of Math Destruction” Cathy O’Neil talks about mathematical models which are opaque, affect large numbers of people, and are potentially damaging, begin used to:
- Determine credit worthiness used to grant/deny loans
  - Screen applications and sort resumes, so human resource personnel don’t need to review all applications/resumes
  - Help policing, sentencing and setting parole, determining the likelihood of recidivism
  - Evaluate workers, teacher quality
  - Rand colleges
  - Target voters

In the Data Science Code of Professional Conduct, Rule 8 labeled “Data Science Evidence, Quality of Data and Quality of Evidence” states:

(h) A data scientist shall use reasonable diligence when designing, creating and implementing machine learning systems to avoid harm. ...

Discuss the obligations of a data scientist regarding dangerous algorithms, and what, if anything ought to be done. (10 pts.)