Part 1: This article has 10 algorithms for data mining that are very significant. Each of these algorithms falls into a few topics consisting of clustering, association analysis, statistical learning, link mining, and classification. The 10 algorithms find various ways to “Mine” the information present in the data, whether it be on the World Wide Web or a simple data set. The authors have discussed pseudocode as well as impacts of the algorithms, they also talk about training data as well as how Google implemented its search. The main objectives of this article was to provide 10 reliable and efficient ways of data mining, but aside from that the authors tried to ways to actually use and implement the algorithms. The authors are also trying to “Inspire more researchers in data mining” to use the algorithms to further expand our knowledge about the impact these algorithms have.

Part 2: This article is very well written and was an enjoyable read. I have not read many scholarly articles so I did not absorb most of the information in my first read through of the article. I like the fact that they include pseudocode or in depth analysis for all of the algorithms they talk about. This made understanding these complex algorithms a bit easier to demystify. Something I did not like was the fact that they did not include any real world examples for the algorithm. Unlike Page Rank, which a real world example had been provided, k-means and other algorithms only provided explanations but did not provide applications where we would be able to use such algorithms. My favorite algorithm that I read about in this article was the Page Rank algorithm. It is very interesting to learn about something that I use in my day to day life. The explanation that the authors had made me actually understand the material through the first read, unlike most of the others. Another algorithm that I enjoyed learning about was Support Vector Machines. I like the fact that it is basic math that I learned in 7th grade that is being expanded upon. Projecting the data into a higher dimension just to get a clean cut of the data is very interesting to me, but very difficult to understand.

Part 3

1. Classification is the process of accurately predicting which class a set of data should be in. Classification does not necessarily need to use decision trees to help produce an answer. C4.5, Support Vector Machines, kNN, Naive Bayes, and CART all use Classification algorithms.
2. Clustering in data mining is when data is represented in groups that have similar attributes with others in the same group but vary with other groups. The clustering algorithms discussed in this article are k-means, Aproiri, and EM.
3. The algorithm that Google uses is Page Rank. Page Rank does use a directed graph where V is the set of all pages and E are the hyperlinks. Page Rank uses a directed graph because a website that references a different website might not reference back the original site.