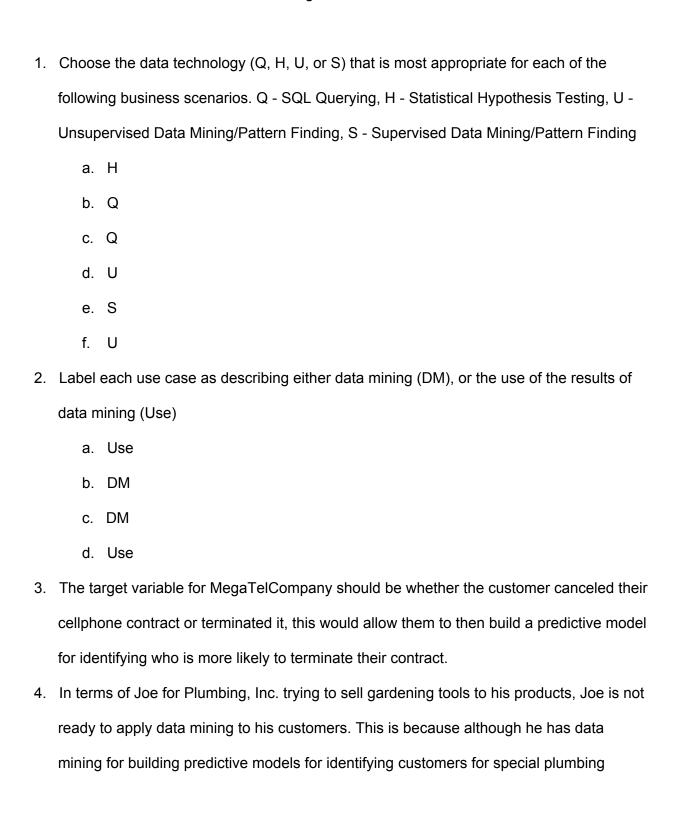
Assignment 1



Iheanyi Ekechukwu Data Mining September 9, 2014

offers, his data from his plumbing transactions are not good enough to accurately identify a subset of customers.

- 5. The Inductive Learning Hypothesis is given known information, at best we could create a hypothesis around this information we already know. An inductive bias is a different way of finding a solution to a problem, such as using a different predictive model. The difference between induction and deduction is that deduction uses already available information/evidence to support a specific conclusion. On the other hand, induction uses the supporting information/evidence to make a general statement, without necessarily making a firm conclusion.
- 6. Steps for Developing a Data Mining Task
 - a. Problem Statement
 - b. Research/background knowledge
 - c. Specify the task
 - d. Get all of the data
 - e. Understand your data
 - f. "Janitor" work data sanitizing
 - g. Linking/normalization/consolidation of your data
 - h. More "janitor" work
 - i. Dataset
 - j. Create target Function/machine learning model/training
 - k. Validation of data
 - I. Interpretation of data
 - m. Deployment
 - n. Monitoring

Iheanyi Ekechukwu Data Mining September 9, 2014

- 7. 3-5 risks or challenges that data mining/analytics applications may face in deployment
 - a. Overfitting data
 - b. Data generalization
 - c. Lack of optimal model to compare to
 - d. Having to rebuild models or build multiple models as more data becomes available and/or necessary