6/7/10

Course ID: COMP 578

Semester: Summer 2010

Name: Alvin Little

<div align="center">Assignment #1: Exercises</div>

1.7: 2)

Suppose that you are employed as a data mining consultant for an Internet search engine company. Describe how data mining can help the company by giving specific examples of how techniques, such as clustering, classification, association rule mining, and anomaly detection can be applied.

Clustering: For an internet search engine company, clustering could be used to group related items together. Once you know what types of things people are commonly searching for together, you can use the information to recommend links based on what they might likely be also interested in. An example would be amazon.com recommendations of purchases based on what other products similar users have purchased.

Classification: When classification is used, it is often used to predict something based on features in a trained data set. Once the classification criteria is determined with a known data set, you can use those rules to predict attributes for new data you may want to analyze for pre-processing or quick analysis. One use that might be useful for a search engine company is sorting user types based on search terms and figuring out what class the user likely falls into, which you can then use to target advertisements for that user. For example, if a user is searching for clothing or perfume, you might use classification to guess they're a girl, and then target other advertisements aimed at girls.

Association Rule Mining: By using the data and mining it for features, you can often find rules that seem to associate attributes with other attributes strongly and interesting patterns emerge. An online search company might find that search terms follow each other in order or certain terms lead to certain links being clicked over others. Also you could have a user search for one thing and find an association to a type of media content like video, audio, or books. You could search the name of a TV show as an example and the user might end up clicking almost exclusively on a site that streams episodes of that show on youtube.com or hulu.com.

Anomaly Detection: When analyzing data, sometimes you want to find any objects that fall outside of the normal data set. An example commonly used is intrusion detection or fraud detection in

online networking and banking system respectively. As far as search engines go, if you were going to track potential terrorist activities via search results, you could use anomaly detection to find odd search terms that stick out from the typical daily collection of search terms people use on the web.