Data Analytics: Exercise 1

1. Data mining is an incredibly popular field with deep origins. It is an incredibly trendy field because of the potential business value but should not be considered a trend as it will remain a fixture of computing and business analytics going forward. Data mining is essentially the process of discovering or extracting knowledge from large sets of data. Much of this is built on the application of various database and pattern detection tools but cannot be reduced to this alone. Pattern detection and statistics alone do not provide knowledge. Knowledge implies additional insight within a given context. Recognizing a pattern in data is important, but unless you can say something meaningful about that pattern in relation to the field or specific problem in question, we haven’t really learned anything about the data.

In this way, data mining essentially reduces to five major steps. First of all, we need to identify the source information and then select the individual data points to analyze. Once we have a feel for our data, we then need to use some statistics and models to extract the relevant information and then identify the key values from that extraction. We are not done yet, however, for data mining isn’t just extracting information, but knowledge. Our last step then is to look at the extracted information and perform some analysis so we can make an interesting or valuable assertion about the data set on the whole.

1. One instance of data mining that is incredibly beneficial is analytics in professional basketball. There is an incredible amount of data about players and games, including traditional stats like points scored, rebounds accumulated, etc. There is also increasingly new and specific data sets like player location and speed, in which cameras record the exact position of a player at any given time. Given that sports are by nature competitive, any insight into the game data can provide an edge to that team going forward. As a result, the more successful franchises are those with better data scientist and analysts. This requires highly trained and proficient data scientists, because statistics alone don’t provide the requisite knowledge. In order to make meaningful insights into the data, you also need to have an understanding for the game and players. I may be able to rely on a machine to give my shooting percentages at different locations on the court, but I need a trained human to be able to apply those statistics to real world situations and formulate plans or approaches to leverage that data to the teams advantage. In that way, data mining is an incredible asset given you have valuable data and skilled analysts.