Data Mining Assignment 1

Identify a problem from your own experience that you think would be amenable to data mining. For that problem describe:

Problem: Fantasy League in Sports

Who should you have in your team? Which players are going to perform best for your team and allow you to beat the competition? The challenge at the start of the season is that there is very little if any data available to help you identify the winning players.

1. What the data is.

The data here is all about the player statistics and pre match analysis:

* First is where we give player stats of the last five matches in each format, they have played, for you to make that informed decision about which player to take.
* Another data comprises of the pre-game details about which players to pick, pitch and weather conditions.

2. What type of benefit you might hope to get from data mining.

Based on this, different models will be generated based on the different phases of the match, which are 2-Overs, 5-Overs, 8-Overs, 12-Overs, 16-Overs and 20-Overs. Based on this, a prediction on which team wins the ongoing match at that particular phase of the match is obtained.

3. What type of data mining (classification, clustering, etc.) you think would be relevant.

I think clustering would be more relevant because When there is little performance data available to train your model on, you have an advantage for unsupervised learning. In this type of machine learning problem, you can find similar players using some of their characteristics. This has been done using K-Means clustering. Ultimately this means you can get a better team more quickly at the start of the year, giving you an advantage.

4. Name one type of data mining that you think would not be relevant, and describe briefly why not.

Regression would not be relevant as it is numerical and usually used to measure future data and make certain predictions. Since, the data set consisting of many instances, In future it may improve the accuracy with more number of instances in the data set because with a larger number of instances the model will have the flexibility to deduce better rules and identify more patterns in the data set as compared to with a lesser number of instances.