**Homework #1**

1. Predictive analytics (PA) is a business intelligence technology that provides a predictive score for each customer, which is associated with the probability that determines a customer’s behavior. PA is a branch of data mining which focuses on the prediction of future probabilities and trends; transforming uncertainty into usable probabilities. The benefits of PA, include, but are not limited to: predicting a customer’s behavior towards a certain offer and performing the appropriate business tactic and determining consumer trends.

2. Supervised learning is associated with predictive models, where the analyst knows what the target attribute to be learned (an estimation or classification) from a set of input attributes. One example of supervised learning is a decision tree. Unsupervised learning is associated with descriptive models, where the analyst does not know what is to be learned from a set of input attributes but depends on correlations between instances to determine what is described within a collected piece of data. One example of unsupervised learning is clustering.

3. Classification supervised learning is associated with yes/no decisions or multi-value discrete decisions while estimation supervised learning is associated with a continuous value output which is monitored over time.

4. A few popular algorithms used for data mining/predictive analytics include: regression (linear, logistics, etc.) which measures correlations between variables, clustering which segments instances based on attribute values, ARIMA, simulations, naïve bayes, decision trees, and neural networks.

5. The CRISP-DM Model or Cross Industry Standard Process for Data Mining is an industry developed methodology standard for executing data mining and predictive analytics. The steps are as follows: 1. business understanding: how to define the problem for PA and identifying target variables 2. data understanding: getting the landscape of the data or interactions between business understanding and data understanding 3. data preparation: fix the problems in the data or create new variables from the core features 4. modeling: selecting modeling technique, generating test design, building the model and assessing the model 5. model evaluation: using test data to validate the model created using training data and 6. deployment: applying the model to real world applications.

6. Data is selected and split into training data (80%) and test data (20%) during model development and deployment. During model development, training data is selected, the proper modeling tools are applied, and a predictive model is created. During model verification, the test data is used to verify the predictive model created, the predictive model is further refined, and the model is verified. During model deployment, real world data is used, a predictive model is created accordingly, and the prediction response is monitored and the response is used to understand the effects of the prediction in a business setting.

7. Some data problems that can be expected when carrying out a PA project are: large amounts of missing data since bumps are harder to smooth over, sparse data, inaccurate data which associated with using the correct measuring metric to attain the proper impressions, nonstainary data overtime or values which are not constant, overfitting data which over optimizes the system unrealistically, and complexity in the model.

8. Web analytics is the measurement, collection, analysis, and reporting of internet data for the purpose of understanding and optimizing web usage; it is a continuous process. Web analytics can be used for measuring website traffic, business research, and marketing research.

9. A few Google Analytics generated metrics are:

pageview: the total number of views for a specific website page

time on site: the duration of time which the viewer was active on the website

unique visitors: tracking an individual users views/activities on the web

keyword report: the search engine is optimized through understanding past visitor and users selections of interest