**Advanced Data Analysis MTH 9797 & STA 9797**

**Robert Kissell, PhD**

**Alternative Final Project Assignment**

Due: December 15 by 6pm

Submit via Blackboard

Sports Modeling Project

Develop Probability Prediction Model

Data: “Excel\_Alternative Final Project.xlsx”

**1: Game Results Data**

* College NCAA Football Scores for FBS teams.
* Tab: Game Results Data
* Includes the results of all games from 8/26/2017 – 12/2/2107
* All Teams have a unique ID from 1 to 131.
* Teams with ID=131 are Non-FBS teams and represent teams in lower divisions. All teams in a lower division are given ID=131.

**2: Develop a Prediction Model**

* *Sheet: Game Results Data*
* Develop a rating and prediction model for all teams.
* In class we have discussed three (3) different modeling approaches:
  + Logistic Regression using MLE
  + Logit Model based on normalized spreads (home pts minus away pts). Normalized spreads were computed by taking the z-score of the spread, and then computing the CDF of the z-score. This results in a normalized score between 0 and 1. Solve using the techniques for the Logit Model and computing the Wins Ratio.
  + Machine Learning with a Neural Network. We define an input matrix for all games with 1 if home team, -1 if away team, and the output variable Y is 1 if the home team won and 0 if the home team lost.

**3: Predict Game**

* *Sheet: Games to Predict*
* Rank all teams from best to worse.
* Predict Bowl Game Results
* Using your model or models, predict the winning team, the probability that the team will win, and the winning spread, e.g., the number of points the winning team is expected to win by.
* There are 39 games in total.
* There are 35 games being played at a neutral site and 4 games being played at one of the team’s home stadium.
* Please make appropriate adjustment when predicting the games. To make this easier, you may want to treat Team #1 as the home team and Team #2 as the away team.
* Provide all results in an Excel spreadsheet.

**4: Modeling Approach**

* Describe the modeling approach you used to develop the prediction model.
* How may models did you use.
* Did you encounter any obstacles when building the model?
* Did you encounter any data limitations while calculating team ratings?