**Capstone Proposal**

**1. What is the problem you want to solve?**

* My company uses Salesforce to manage its sales pipeline. An ongoing issue is knowing whether our pipeline is large enough, and generates enough “Award” wins to sustain future revenue growth targets.
* I want to use a machine learning categorization algorithm to forecast the win / lose outcomes (binary) of our sales pipeline leads.

**2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?**

The client is the CFO, and he cares about this problem because:

* Knowing the “Award” wins allows him to make sure that our sales pipeline can generate enough revenue to surpass future growth targets for his 3-year strategic plan.

Based on my analysis, what will the CFO do that they wouldn’t have otherwise?

* Supposing that my analysis forecasts wins more accurately than the current model, and it shows that we are not expected to reach growth targets, the CFO can make decisions that get the institute back on track.

**3. What data are you going to use for this? How will you acquire this data?**

The data comes from the Salesforce database of historical and current opportunities.

I will download a flat file (csv or excel) using Cognos.

**4. In brief, outline your approach to solving this problem (knowing that this might change later).**

* Obtain historical win / lose pipeline data for the past 5 years or more.
* Perform exploratory data analysis.
* Cross-validate the data by randomly splitting into training and test data.
* I will need to use a machine learning categorization method. My Initial thought is to start simple by use logistic regression, but other more sophisticated methods I have not studied yet, such as trees, random forests or gradient boosting could become the preferred methods once I study them.
* Train the algorithm for the Bid / No Bid (binary, 1 = Bid, 0 = No Bid) decision using historical data that we bid on.
* Test for Bid / No Bid decision using.
* Then train the algorithm on the Win / Lose (binary, 1 = Win, 0 = No Bid) decision.
* Test the algorithm on the Win / Lose decision.
* Compare my results versus the existing model that is being used.
* Write analysis document and slide deck to present the results.

**5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

1. Analysis document containing:
   * Code to clean up data.
   * Code to apply algorithm
2. Paper describing the steps taken, doing exploratory analysis, visualizing the data, and analyzing the effectiveness of the model.
3. Slide deck