**Capstone Proposal**

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

Analyze the data of colleges and universities and classify them as either public or private. Based on this classification government funding can be given to schools who are public or scholarship to top students so as to improve public schools and bring them at par with private schools.

**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 Education department of the Government They may use the analysis to decide on which areas to increase the spending e.g. creating more board rooms for students as they may study in more concentrated way and provide more privacy. This may impact the graduation rates of students.

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

The data is taken from the ISLR package. The data frame used is the ‘College’ data frame. Fields to be studied are

* Private A factor with levels No and Yes indicating private or public university
* Apps Number of applications received
* Accept Number of applications accepted
* Enroll Number of new students enrolled
* Top10perc Pct. new students from top 10% of H.S. class
* Top25perc Pct. new students from top 25% of H.S. class
* F.Undergrad Number of fulltime undergraduates
* P.Undergrad Number of parttime undergraduates
* Outstate Out-of-state tuition
* Room.Board Room and board costs
* Books Estimated book costs
* Personal Estimated personal spending
* PhD Pct. of faculty with Ph.D.’s
* Terminal Pct. of faculty with terminal degree
* S.F.Ratio Student/faculty ratio
* perc.alumni Pct. alumni who donate
* Expend Instructional expenditure per student
* Grad.Rate Graduation rate

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

**Task 1:-**

Load the data and do some initial view of the data.

**Task 2:-**

Create some scatter plot and histogram to do some exploratory analysis of the data

**Task 3:-**

Create training and testing data set and create decision trees and do predictions. Check the analysis using confusion matrix

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

1. R Program.
2. Visualizations using ggplot.
3. Logistics Regression
4. Decision Three Model design & Prediction with test data and its accuracy.
5. Data Story on word document.