**INDIVIDUAL ASSINGMENT  No. 2/3**                                   Due Date: March 8 | Weight: 10%

**Business Challenge:**
Use data provided in class to generate credit default forecasting model by applying the following steps*:
1.  Randomly divide data into training samples T(ols) and T(rt) and hold-out samples H(ols) and H(rt).
2.  Method 1: OLS
    a.  Used library algorithms to run OLS on T(ols) to generate and store the regression function.
    b.  Run regression function on H(ols) and generate MSE and $R^2$.
3.  Method 2: Regression Tree
    a.  Use library algorithms and set tuning parameters to tree depth 8, minimum units in each non-terminal node 20, and $R^2$ minimal improvement per split 1%.
    b.  Run algorithm on T(rt) to generate and store prediction algorithm.
    c.  Run prediction algorithm on H(rt) to generated MSE and $R^2$.
4.  Assess methods 1 and 2 based on MSE and $R^2$ and calculate value-add of best method.

**Deliverables:**
1.  **Executive Summary:** 10 Lines, non-technical, describing:
    a.  Business Understanding:  Target business value-add
    b.  Data Understanding:  Credentials, relevance, and top categories
    c.  Modeling:  Techniques, parameters, and performance metrics
    d.  Performance & Evaluation: Was value-add achieved, areas for further analytics

2.  **Technical Methodology:**

| STEPS | SPECS |
|---|---|
| 1. Data Management & Workflow | For each row, use technical language to provide step-by-step description of data & modeling decisions, why such decisions, whether performance was as desired, and areas for further analytics. Use business challenge above as model (Max > 20 words per line item). |
| 2. Training & Out-of-Sample Datasets | |
| 3. Model Parameters | |
| 4. Fit & Performance | |

3.  **Data & Code:** Submit files for following items (in Text, Excel, R, or Python):
    a.  Training & out-of-sample datasets
    b.  Algorithms
    c.  Interface for prediction & delivery

**Evaluation Criteria:**
1.  **Executive Summary:** Ability to explain and justify in non-technical language your outcome business value-add based on the process and methodology you chose for business understanding, data understanding, modeling, and performance & evaluation.
2.  **Technical Methodology:** Ability to use technical language to describe your process and methodology for above table items 1-4 and generate a step-by-step by step guide to duplicate your analytics using the datasets you provided as attachments.
3.  **Data & Code:** Usability of datasets and code as provided to duplicate analytics using the Technical Methodology you provided.

*Scoring Method*

| | |
|---|---|
| **Executive Summary | Clarity of:** | **3** |
| Data Understanding Description | 1 |
| Modeling Description | 1 |
| Performance & Evaluation Description | 1 |
| **Technical Methodology | Clarity of:** | **4** |
| Data Classification Description | 1 |
| Training & Out-of-Sample Datasets Description | 1 |
| Model Parameters Description | 1 |
| Fit & Performance Description | 1 |
| **Data & Code | Duplication:** | **3** |
| Training & Out-of-Sample Datasets Duplicability | 1 |
| Algorithms Duplicability | 1 |
| Interface for Prediction & Delivery Duplicability | 1 |
| Total (Max): | **10** |

*__Additional References:__ Sendhil Mullainathan, Machine Learning: An Applied Econometric Approach, Online Appendix, April 2017. In class material.*