Project: Introduction to Machine Learning Applications

First 3 Sections Due 04/08/2021 11:59 pm via LMS.
Final presentations in-class -- 04/26/2021; 04/29/2021; 05/03/2021
Final report due – 04/29/2021 11:59 pm via LMS

PROJECT OBJECTIVE
The goal is to develop an understanding of how individuals approach machine learning projects, seeing the entire process from exploratory data analysis to modeling and evaluation.

PROJECT SELECTION
Please look through the available Kaggle competitions at:
https://www.kaggle.com/competitions and select 1 project.

You should avoid image-based data or projects that only require visualizations. Ideally the project should be aligned with some type of an application related to business.

REPORT SECTIONS
1. **Executive Summary** (1 pages)
This should be a summary in your own words of the problem, data, and findings.

2. **Benchmarking of Other Solutions** (2 pages)
Identify 3 other Kaggle *solutions* completed by others. The solution should include a score on the Kaggle prediction task. You can find by selecting on the project and then clicking on the link to Kernels. Summarize the features, modeling approach, and performance in a table. Then do some further research to comment on the approach and try to characterize what make the kernel more or less successful than others.

3. **Data description and Initial Processing** (3 pages)
This section should include basic characterization of data. You should run and report basic statistics on the data and generate at least 3 visualizations. You can review other kernels to understand some different approaches to the data, but this section you are required to generate all analyses.

4. **Modeling** (3 pages)
Modeling should examine relevance of different independent variables (features) and different algorithms. You should examine at least 3 different models and be able to also explain the relevance of different independent variables.

5. **Appendix**
A GitHub repository of all modeling, well commented code.

NOTE: If you copy and paste from the Kaggle description that is plagiarism and you will be reported to the Associate Dean's office and receive a 0 on the project grade.

PROJECT EVALUATION
The description below describes an ideal project. Projects will be evaluated subjectively by the instructor according to this rubric.

- *Formatting (10 points).* The student presented the report in a format that indicated professionalism and care in the organization, writing, and presentation of the overall report.
- *Executive summary (20 points).* The student was able to present the results of modeling in a way that is rich and interesting as well. There is clear representation of key predictors and key algorithms used. There is a summary of the results and key findings.
- *Benchmarking of other solutions (30 points).* There is a clear insightful comparison of approaches, and he predictive characteristics of the different models are clearly compared in a table with appropriate conclusions. There are outside resources consulted in the description of specific algorithms if relevant.
- *Data description and initial processing (40 points).* The student was able to clearly present an overall picture of the data using techniques presented in the class. This includes basic structure field by field descriptions as well as visualization and basic statistics. Where necessary they have adequately used techniques for cleaning the data or generating new features.
- *Analysis of relevance of independent variables (25 points).* The student was able to clearly present justification of the value of different independent variables. Where possible, exploration of feature creation is provided.
- *Analysis of performance of different model types (25 points).* There are outside resources consulted in the description of specific algorithms if relevant. Outside sources give clarity and there is evidence of some model tuning.
- *Commented Code (20 points, as needed).* Clearly commented code has been provided in the assigned Jupyter notebook.

PROJECT SUBMISSION
- *The project report is to be submitted to the LMS.
- NOTE: If you copy and paste from the Kaggle description that is plagiarism and you will be reported to the Associate Dean's office and receive a 0 on the project grade.
- Final Code Submission must be done via GitHub.