

DSEG6111: Predictive Modeling Course Project Description

Dr. Kokkotos

Merrimack College – Summer 2022

In this project, you will have an opportunity to apply the algorithms and techniques you learned in class to some real-world problems.

- You need to consider three types of problems:
 - One problem with a quantitative response
 - One problem with a qualitative response
 - One problem specifically using principal components regression
- Find appropriate data related to the problems described above
Example: <http://archive.ics.uci.edu/ml/>
- Apply all the algorithms learned in class to the data
- Evaluate and compare the algorithms
- Submit a report, together with your data and code

Format-wise, the report should include the following sections:

- *Executive Summary*: 1-2 pages succinct summary of your key findings and recommendations, presented in the context of the goal of the project.
- *Data & Approach*: An overview of the data used in your analysis, any data re-engineering and related steps, a brief discussion of the overall approach, analytic goals, and data analytic techniques utilized.
- *Detailed Findings*: An in-depth description of data analytical findings; whenever possible/applicable shown graphically with written description.
- *Validity & Reliability Assessment*: A discussion of how the accuracy of analysis-driven recommendations should be tested in future implementation.
- *Appendix*: Any additional details such as parts of the output you consider worthwhile but not appropriate for inclusion in the body of the report. It is important to make sure, however, that only outcomes that are referenced elsewhere in your report and/or are necessary to understand the details of the work are included – inclusion of unrelated or unnecessary (to communicating or explaining of results) outcomes will adversely impact the final report grade.

You are to format the report as follows:

- *Length*: No more than 20 single-spaced pages (excluding any appendices).
- *Format*: MS Word; 12-point New Times Roman font.
- *Delivery*: Please submit an electronic copy of your report only
- *Content*: When including any charts/graphs or other parts of your statistical analyses-generated output, only those elements of the analytic output that are used and/or referenced in the result

discussion should be included; attaching unnecessary parts of the output will adversely impact the assignment grade as it will signal inattention to detail or a lack of full understanding of the analytic output. At the same time, not including evidence which is necessary to properly evaluate the analytic outcomes will also adversely impact the assignment grade. In short, it is critical to be thoughtful and purposeful.