REPLACE WITH COVER PAGE

Data Preparation 1

[10 points] Define and prepare your class variables. 1

[5 points] Describe the final dataset 1

Modeling and Evaluation 2

[10 points] Choose and explain your evaluation metrics 2

[10 points] Choose the method you will use for dividing your data into training and testing splits 2

[20 points] Create three different classification/regression models 2

[10 points] Analyze the results using your chosen method of evaluation. 2

[10 points] Discuss the advantages of each model for each classification task, if any. 2

[10 points] Which attributes from your analysis are most important? 2

Deployment 3

[5 points] How useful is your model for interested parties (i.e., the companies or organizations that might want to use it for prediction)? 3

# Data Preparation

## [10 points] Define and prepare your class variables.

Use proper variable representations (int, float, one-hot, etc.). Use pre-processing methods (as needed) for dimensionality reduction, scaling, etc. Remove variables that are not needed/useful for the analysis.

## [5 points] Describe the final dataset

Used for classification/regression (include a description of any newly formed variables you created).

# Modeling and Evaluation

## [10 points] Choose and explain your evaluation metrics

that you will use (i.e., accuracy, precision, recall, F-measure, or any metric we have discussed). Why are the measure(s) appropriate for analyzing the results of your modeling? Give a detailed explanation backing up any assertions.

## [10 points] Choose the method you will use for dividing your data into training and testing splits

(i.e., are you using Stratified 10-fold cross validation? Why?). Explain why your chosen method is appropriate or use more than one method as appropriate.

## [20 points] Create three different classification/regression models

(e.g., random forest, KNN, and SVM). Two modeling techniques must be new (but the third could be SVM or logistic regression). Adjust parameters as appropriate to increase generalization performance using your chosen metric.

## [10 points] Analyze the results using your chosen method of evaluation.

Use visualizations of the results to bolster the analysis. Explain any visuals and analyze why they are interesting to someone that might use this model.

## [10 points] Discuss the advantages of each model for each classification task, if any.

If there are not advantages, explain why. Is any model better than another? Is the difference significant with 95% confidence? Use proper statistical comparison methods.

## [10 points] Which attributes from your analysis are most important?

Use proper methods discussed in class to evaluate the importance of different attributes. Discuss the results and hypothesize about why certain attributes are more important than others for a given classification task.

# Deployment

## [5 points] How useful is your model for interested parties (i.e., the companies or organizations that might want to use it for prediction)?

How would you measure the model's value if it was used by these parties? How would your deploy your model for interested parties? What other data should be collected? How often would the model need to be updated, etc.