**Paper Title (Something Interesting)**

Naat Ambrosino, [nambrosino@bellarmine.edu](mailto:nambrosino@bellarmine.edu)

Jerrin Redmon, [jredmon3@bellarmine.edu](mailto:jredmon3@bellarmine.edu)

**ABSTRACT**

Up to 150 word summary of your project.

1. **INTRODUCTION**

Provide a one or two paragraph introduction to your project in which you describe the data set you are working with and the classification target (what are you trying to predict?) you will be pursuing.

1. **BACKGROUND**

In this section, provide some background for the problem for which the data were collected. For example, if you were using the mushroom data set, you write up some background on what a mushroom is, why the data were originally collected, what question(s) the authors were trying to answer, etc.

1. **EXPLORATORY ANALYSIS**

This section will be similar to your exploratory analysis project. First, provide a summary of the data set similar to your first exploratory analysis: *e.g. this data set contains 398 samples with 7 columns with various data types*. In this summary, provide the data types of your columns (in a table) and then rather than providing tabular statistics and plots for each variable, provide only statistics and plots that seem unusual. For example, if one or two variables have significant missing values or the distribution of the variable is skewed or looks unusual note that. Provide the unusual statistics or plots in this section. Provide any other appropriate plots (e.g. correlation matrix, heatmaps, bar charts, etc.) that you deem necessary.

**Table 1: Data Types**

|  |  |
| --- | --- |
| *Variable Name* | *Data Type* |
| V1 |  |
| V2 |  |
| V3 |  |

1. **METHODS**

In this section, describe how you prepared the data for your model and performed multiple experiments using different parameters for the model(s).

* 1. *Data Preparation*

Describe how you prepared the data for your model. For example, you might need to normalize the data, so variables with wider ranges of values don’t overshadow variables with smaller ranges. If you decide to drop variables from the model or create variables from existing columns, explain the process and the reasoning behind those decisions.

* 1. *Experimental Design*

You will run your model several times with different parameters to see what different results you get. In a table, describe your experimental parameters. Three or four experiments are sufficient. This is where you will describe how you divided your data into train, validate and test data sets. For example:

Table X: Experiment Parameters

|  |  |
| --- | --- |
| **Experiment Number** | **Parameters** |
| 1 | All four (4) raw features with 80/10/10 split for train, validate, and test |
| 2 | All four (4) normalized features with 80/10/10 split for train, validate, and test |
| 3 | All four (4) raw features with 70/15/15 split for train, validate, and test |
| 4 | All four (4) normalized features with 70/15/15 split for train, validate, and test |
| 5 | All four (4) features with a square root transform on displacement, horsepower, weight and acceleration with 70/15/15 split for train, validate, and test. |
| 6 | All four (4) features with a square root transform on displacement, horsepower and weight with a70/15/15 split for train, validate, and test. |
| 7 | All four (4) continuous features and the three (3) categorical features with a square root transform on displacement, horsepower, and weight with an 80/10/10 split for train, validate, test. |

* 1. *Tools Used*

Describe all of the software tools you used to perform your data preparation and model implementation. For example:

The following tools were used for this analysis: Python v3.5.2 running the Anaconda 4.3.22 environment for Apple Macintosh computer was used for all analysis and implementation. In addition to base Python, the following libraries were also used: Pandas 0.18.1, Numpy 1.11.3, Matplotlib 1.5.3, Seaborn 0.7.1, SKLearn 0.18.1, and Patsy 0.41. Provide a brief explanation of why you chose these tools.

1. **RESULTS**
   1. *Classification Measures*

Provide the classification measures for each experiment. For example, you could provide a contingency table for each model to measure how well it classifies data. You could also do an ROC curve (using SciKit Learn). You need to demonstrate how you are measuring the success/failure of the models.

* 1. *Discussion of Results*

Discuss which of your models provided the best classification (or some other outcome if not classification). Explain why you think your best model was the best and why your worst model was the worst.

* 1. *Comparison of Models*

Since you are running at least two machine learning classification models, compare the models and explicitly discuss which model was the best.

* 1. *Problems Encountered*

No project goes perfectly smooth. Discuss any problems you had with obtaining the data, preparing the data, implementing the model, or evaluating the model. **It would be highly unusual to indicate that you had no problems.**

**One of the column names being ‘class’ caused problems. I also had problems with sns.boxplot for the dfSick dataframe, with no such problem for any other dataframes. Also a problem that the confusion matrix function had where the y\_pred and y\_test should be flipped.**

* 1. *Limitations of Implementation*

Discuss the limitations of your model(s). Are there reasons your models might not be the best way to predict the target data? What other models might work better?

* 1. *Improvements/Future Work*

What would you like to do to improve your model in future work? Some items you might consider discussing are performing more experiments, using different models, adding or removing variables, finding a different data set, etc.

1. **CONCLUSION**

Finish up with a paragraph or two of summarizing your problem, the results and your conclusions (good model, bad model, needs more work, etc.).

**REFERENCES**

List any websites, books, articles, etc. that you found useful while you worked on this project. It is not necessary to cite the references in the paper unless you specifically mention it in the text.

*Division of Labor*

In this section after you references, provide a paragraph or two outlining what each team member did on this project. Please don’t tell you both did everything! Be honest about who did what.

**Other directions:**

1. 10-pt, Times New Roman, 1” margins all around (if you use this template you are already set).
2. **Ensure all tables and figures are numbered appropriately and referenced in the text.** See examples above and below.

|  |  |
| --- | --- |
| **Figure 1: Comparison of X/Y from dataset (single plot) (8 pt.)** | **Figure 2: (a) Function Output (b) A against B (multiple plots) (8 pt.)** |