**Money Maker or Money Loser?**

Braden Heuglin, [bheuglin2@bellarmine.edu](mailto:bheuglin2@bellarmine.edu)

**ABSTRACT**

I am applying a confusion matrix ML Classification model to a dataset on the stock price of AIRBNB over the last few years. My dataset contains the open and close price of the stock, as well as the low and high of each day. I will use the open, close, and low prices to predict what the high price will be. Once I predict the high price, I will display how accurate my prediction model is.

1. **INTRODUCTION**

One or two paragraph introduction to your project in which you briefly describe the data set you are working with and the ML **Classification** model you chose to apply to it.

As I mentioned in the abstract above, I chose to work with a dataset on the stock price of AIRBNB over that past few years.

1. **BACKGROUND**
   1. *Data Set Description*

Provide a more detailed description of the data set including a reference to where it can be found and why you chose to work with. An important feature of the data is who collected the data and why.

* 1. *Machine Learning Model*

Provide one or two paragraphs that describe the ML model (Logistic Regression) for your data set. Explain in narrative how the model works. You do not have to provide a large amount of mathematical, but you can if want.

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 appropriate plots (e.g. correlation matrix, heatmaps, bar charts, etc.) that you deem necessary.

**Table 1: Data Types**

|  |  |
| --- | --- |
| *Variable Name* | *Data Type* |
| Date | Categorical object |
| Open | Quantitative Int |
| High | Quantitative Int |
| Low | Quantitative Int |
| Close | Quantitative Int |
| Adjusted Close | Quantitative Int |
| Volume | Quantitative Int |

1. **METHODS**

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

* 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 |

* 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. Provide a brief explanation of why you chose these tools.

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

Provide the classification measures for each experiment using a confusion matrix and classification report.

* 1. *Discussion of Results*

Explain your model finding using a confusion matrix, the accuracy score and classification report.

* 1. *Problems Encountered*

No project goes perfectly smoothly. 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 not problems.**

* 1. *Limitations of Implementation*

Discuss the limitations of your model. Is there is reason it might not be the best way to model the data? What other models might work better?

* 1. *Improvements/Future Work*

What would you like to do to improve your model in future work? Do more experiments, use a different model, add/remove variables, find 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.

**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.)** |