

IDSN 599: Machine Intelligence

Final Project Part 3

Due: 12/6/2021 @ 11:59PM

Goal

For this last part of the final project, you are to train and test on your dataset.

Setup

- You get to create whatever Python scripts you need. You probably will need some custom transformer classes. They should go into their own Python script file.

```
'''  
Name  
IDSN 599, Fall 2021  
USC email  
Final Project Part 3  
'''
```

Requirements

A document that contains the additional sections to your document from part 2. Just add to your document that you submitted for part 2.

- ML Algorithms:** You are to train on at least 3 different ML algorithms. If you are doing a prediction problem, use the root-mean-squared error (RMSE) to determine your average error on the training. Choose the algorithm that has the smallest RMSE.
 - If you are doing a classification problem use the ROC graph to select your algorithm.
- Data Cleanup:** Discuss what you had to do to clean up your data. That may include deleting attributes/columns that have little to no correlation, filling in missing values, removing rows that have incomplete data that you cannot correct, etc. As part of this discussion, include the lines of code that you have in Python that performed this work. Explain why you did what you did. What was your rationale?
- Results:** Discuss how much of your data you used to train and how much to test. Did you have to separate your dataset into a test set and training set or was that already available to you. How well did you do on the test set? Tell what Python code produces your results.
- Analysis of results:** What did you learn from this dataset – whether predictor or classifier? Which methodology did you use to analyze the success of your model. Why did you choose that method? Depending on how much cleanup you had to do, did that bias your data or lead to a better result? If you had more time, could you have done other things to improve your result?

Deliverables

A compressed folder containing your document and all Python code you used to prepare your dataset for training of your ML algorithm.

Grading

| Item | Points |
|---------------------|--------|
| ML Algorithms | 10 |
| Data Cleanup | 10 |
| Results | 20 |
| Analysis of Results | 20 |
| Python Code | 20 |
| Total | 80 |