Bootcamp Info Sheet

Instructor

Name: *Ryan Ellison*

Bootcamp Details

Bootcamp Title: *Classification in Python*

Number of Days: *4 days*

Hours per Day: *3 hours per day (11-2 eastern)*

Type of Instruction: *lecture with knowledge checks and hands on exercises*

Description: *Classification algorithms are powerful and intuitive data science tools that can predict behaviors and trends. This course teaches students how to implement algorithms like k-Nearest Neighbors, logistic regression and decision trees, as well as evaluate and interpret their results. By the end of the course, students will be able to build classification models to anticipate events and assess the accuracy of predictive algorithms.*

Target Audience: *Students who are comfortable using Python to manipulate data using numpy and pandas and must know how to create basic visualizations using matplotlib.*

Technologies: *Python & Anaconda. Packages that need to be installed:*

numpy==1.19.5

pandas==1.3.1

matplotlib==3.3.3

graphviz==0.19.1

seaborn==0.11.1

scikit-learn==0.24.1

scipy==1.4.1

wordcloud==1.8.1

Prerequisites: *Students must be comfortable using Python to manipulate data using numpy and pandas and must know how to create basic visualizations using matplotlib.*

Student References: *Class slides, class code, exercise files*

Bootcamp Syllabus

Day 1

- Introduction to Classification
- Building kNN models and performance metrics
    - Definition of kNN technique and application
    - Applying cross-validation to the kNN algorithm

Day 2

- Building kNN models and performance metrics
    - Applying cross-validation to the kNN algorithm
    - Identifying and evaluating performance metrics

- Logistic regression
    - Introduction to logistic regression and its relation to neural networks

Day 3

- Logistic regression
    - Assessing classifier performance (ROC curve, AUC, cutoff value selection)
    - Tuning & regularization of logistic regression
- Decision trees
    - Decision trees: Gini coefficient and information gain

Day 4

- Decision trees
    - Confusion matrices and misclassification rates
    - Evaluating the model for overfitting and underfitting