Bootcamp Info Sheet

Instructor

Name: *Elliot Stern*

Bio: *Elliot is a freelance Data Scientist with 7+ years of experience, including working with an Olympic team and creating an NBA salary model for Hazan Sports Management, an NBA sports agency. He co-founded a soccer computer vision company that secured $250,000 of funding. Elliot was a leader in creating computer vision products for the vertical jump and juggling a soccer ball. His sports algorithms have generated over $100,000 in profit and are more accurate than those of industry leaders such as ESPN and Yahoo. Elliot currently works as a data science freelancer taking on a myriad of data science challenges, including regression, computer vision, dashboard creation, and clustering problems, to name a few.*

Bootcamp Details

Bootcamp Title: *Classification using KNN, Logistic Regression, and Decision Trees 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:*

python==3.11.4

matplotlib==3.7.1

numpy==1.24.3

pandas==2.0.2

scikit-learn==1.2.2

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