Bootcamp Info Sheet

Instructor

Name: *Elliot Stern*

Bio: *Elliot Stern 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.*

Bootcamp Details

Bootcamp Title: *Anomaly Detection*

Number of Days: *4*

Hours per Day: *3*

Type of Instruction: *Lecture with exercises and interactive knowledge checks*

Description: *Identifying the most impactful features for your model and detecting outlier data points are powerful machine learning techniques. This class will build upon foundational machine learning techniques to hone predictive skills and discover critical danger points in patterns. By the end of this course, students will be able to determine key features in models and identify anomalous data points.*

Target Audience: *Students who are comfortable using Python to manipulate data and creating basic visualizations, and have a foundation in classification techniques and clustering algorithms (KNN, K-means, logistic regression).*

Technologies: *Python & Anaconda*

- *python==3.11.4*
- *pandas==2.2.3*
- *matplotlib==3.9.2*
- *scikit-learn==1.5.2*
- *imblearn==0.0*
- *kneed==0.8.5*
- *statsmodels==0.14.4*

Prerequisites: *Students must be comfortable using Python to manipulate data and must know how to create basic visualizations. Additionally, students must have a foundation in classification techniques and clustering algorithms (KNN, K-means, logistic regression).*

Student References: *Class slides, class code, exercises, and exercises with answers.*

Bootcamp Syllabus

Day 1

- Introduction to anomaly detection
  - Definition of anomaly concepts and their use cases
  - Differentiating between types of anomalies
- DBSCAN
  - Concept of DBSCAN and its parameter estimation
  - Visualizing DBSCAN for an arbitrary distance
  - Optimize parameters of DBSCAN
  - Implement DBSCAN on time series data

Day 2

- Implementing anomaly detection
  - Implement DBSCAN on time series data
  - Examine why classification techniques are not useful for anomaly detection
  - Summary of different techniques for anomaly detection
  - Identifying SMOTE analysis and its implementation

Day 3

- LOF
  - Implementing Local Outlier Factor Algorithm
  - Optimizing LOF by tuning its hyperparameters
- Describe the isolation forest algorithm

Day 4

- Implement isolation forest to detect anomalies
- Preprocess time series data
- Review time series modeling basics
- Explain the concept of stationarity and differencing
- Explain how to measure linear relationships within time series