

Bootcamp Info Sheet

Instructor

Name: *Martin Laptev*

Bio: *Martin Laptev is a data scientist who enjoys using the Python and R programming languages and Command Line Tools to explore, analyze, visualize and present data. Martin believes that anyone, regardless of their background, can learn and benefit from technical skills. In addition to Data Society, Martin has taught Python Programming, Machine Learning, and Data Science courses for Foundation for the Advancement of Education in the Sciences (FAES), Software Carpentry, and General Assembly.*

Martin holds an MPH in Epidemiologic and Biostatistical Methods for Public Health and Clinical Research and a certificate in Data Science from Johns Hopkins University, a PhD in Tumor Biology from Georgetown University, an MS in Biotechnology from Jagiellonian University in Poland, and a BA in Biology from St. Mary's College of Maryland.



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*

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
- Describe an autoregressive (AR) model and moving average (MA) model