# *Bootcamp Info Sheet*

## Instructor

**Name:** Joseph Issa

**Bio:** Dr. Joseph Issa joined Data Society as an adjunct professor for computer science in 2022. Over the years, the exposure to higher education heightened his interest in academic pursuits and influenced his efforts to assist others on their journeys. Education is a fantastic process of exploration, reflection, and confirmation. Dr Issa enjoys providing students with guidance through this process so they achieve their overall academic goals. His educational background includes a Ph.D. in computer engineering from Santa Clara University in December 2012. He also has a master's degree in Information and Data Science from UC Berkeley in 2022, a Master's degree in electrical engineering from San Jose State University in 1999, and a bachelor's degree in computer engineering from Georgia Institute of Technology in 1996. Dr Issa has both industry and academic experience in computer science and engineering. He has taught graduate and undergraduate computer and electrical engineering courses. He also worked at Intel Corporation in Santa Clara, California, for 16 years (1998-2014). Dr Issa has authored and co-authored over 30 journals and conference publications in computer science and engineering.

**Image:** *< Attach a headshot to accompany your bio >*

## Bootcamp Details

**Bootcamp Title:** Advanced classification

**Number of Days:** 4

**Hours per Day:** 3

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

**Description:** This course covers advanced topics in classification, including classification model parameter tuning and ensemble learning methods – specifically boosting. Students will learn how to make classification models perform better and how to choose the correct methods and parameters based on the problem at hand.

**Target Audience:** Students who are comfortable using Python to manipulate data and creating basic visualizations, and have a foundation in classification models and model accuracy measures.

**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 models and model accuracy measures.

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

# Bootcamp Syllabus

*Outline – Add or delete days/topics as needed.*

**Day 1**

- Random forests
  - Introduce random forest and discuss use cases
  - Summarize the concepts associated with random forest and bagging
  - Load dataset and implement random forest
  - Predict using random forest model and evaluate results

**Day 2**

- Random forests continued.
  - Introduce gradient boosting and how it compares to bagging
  - Discuss gradient tree boosting within scikit-learn

- Evaluate and optimize random forest model.
  - Evaluate base random forest model
  - Optimize random forest model using RandomizedCV method
  - Use performance metrics to compare optimized RF to base RF model

**Day 3**

- Gradient boosting
  - Implement base GBM model and asses its performance
  - Optimize GBM model using RandomizedCV method
  - Use performance metrics to compare all ensemble methods

- Support Vector machines
  - Hyperplane and classification using hyperplane

**Day 4**

- Support Vector machines continued.
  - maximal margin classifier and its pitfalls
  - support vectors and building a support vector classifier model
  - key differences between support vector classifier and support vector machine
  - optimizing the support vector machine model using grid search