

# Shahaf Dan

(925) 791- 1544 ○ shahaf@bu.edu ○ [LinkedIn](#) ○ [GitHub](#) ○ [shahafdan.com](#)

Experienced and aspiring Data Scientist and an AI enthusiast. Interested in learning and growing in the “virtual” world of Data Science, Machine Learning, and Artificial Intelligence. Seeking **full-time opportunities starting May 2023**.

## EDUCATION

---

**Boston University** ○ BA Computer Science 2020 - 2023

- Python, Machine Learning, AI, Data Science, Statistics & Probability, Analysis of Algorithms, NLP, Software Engineering.
- Capstone Project: Used Python's SpaCy for NLP sentiment analysis over Twitter streaming data, fetched with Apache Spark.
- Activities: Alpha Epsilon Pi (President), Student Government (Treasure), TAMID Consulting Group (VP of Tech)

**Las Positas College** ○ AS Computer Science, AS Physics, AA Mathematics 2018 - 2020

- C++, PHP, Assembly, Data Structures, Computer Architecture, Databases, SQL, Calculus & Diff. Equations, Linear Algebra.
- Capstone Project: The Market Manager. Designed full stack software to automate the registration, volunteering, and inventory tracking of the market - a monthly grocery distribution event (SQL, PHP, JavaScript, HTML, CSS, AJAX, PDO)
- Activities: Alpha Gamma Sigma (President), Student Government (Events Director), CS Club (VP), Chess Club (President)
- **CS Dept. Scholar, Honors Certificate** (conducted 6 faculty-approved research projects), **Dean's List, Highest Honors**.

## CERTIFICATES

---

**MIT Professional Education** ○ **Applied Data Science Certificate** December 2022

- A 14-week accelerated bootcamp covering in depth Data Science topics, with engaging projects, live lectures, and quizzes.
- Regression/Classification, Statistical Modeling, Machine Learning, CNNs, RNNs, Pandas, Keras, TensorFlow, SciKit-Learn.
- Capstone Project: Developed a supervised Keras CNN model to classify cells' images as parasitized (Malaria) or uninfected.

**IBM (Coursera)** ○ **AI Engineering Certificate** November 2022

- A 6-course online program covering ML and AI Engineering topics with daily assignments, videos, and programming tasks.
- Neural Networks, Deep Learning, Image Recognition, Object Detection, Computer Vision, OpenCV, PyTorch, Watson Studio
- Capstone Project: Developed a supervised Object Detection model to detect presence of stop signs in images (OpenCV).

## WORK EXPERIENCE

---

**Hindsight Technology Solutions** ○ *Machine Learning Engineer Intern* January 2023 - Present

- Working on NLP (SpaCy) engine implementation on AWS cloud services for web-scraped text analysis.
- Created PoC for new recommender model and collaborated with engineers to design code for a reliable deployment pipeline.
- Boosted prediction speed for large datasets by developing a multi-layer clustering algorithm to group web-pages, which reduces prediction complexity from  $O(k)$  to  $O(\log k)$  while maintaining a high accuracy level of 91%.

**Boston University** ○ *Artificial Intelligence Research Assistant* January 2023 - Present

- Working with Professor Dokyun Lee towards a publication: AI models to operationalize creativity in AI generated artifacts.
- Working with Professor Eshed Bar-Ohn towards a publication: Educational AI Assistive tools for disabled students.
- Working with Professor Drori towards a publication: Analyze AI tools & ML models to detect AI generated math proofs.
- Developing TensorFlow and PyTorch ML models for various research projects, based on SOTA academic research products.

**AlgoTrade** ○ *Data & Machine Learning Engineer* June 2022 - October 2022

- Developed python scripts to generate statistical models on mass stock-option data collected over different strategies.
- Created PySpark ELTL pipeline from a streaming Interactive Brokers API to SQL database (Apache Spark, AWS RedShift)
- Designed a Deep Learning model to predict best option bundle purchase choices based on collected mass data.

**Spark!** ○ *Data Science Intern* January 2021 - May 2022

- Structured data ETL pipelines to map multi-sourced data (APIs, DBs) into an accessible and unified structure using Python.
- Designed Machine Learning models for classification/regression/clustering to predict desired insights from mass data.
- Worked on EDAs using python (Pandas, NumPy, Matplotlib) and generated user friendly business insights (Tableau).