

Kostas Papadopoulos

github.com/konpap94 • konpap94@bu.edu • 617-922-4803

ABOUT ME

I'm a Data Scientist and Software Engineer. I have work experience in data-driven consulting and clinical research. I developed a data visualization package for StratoDem and a cloud scanning tool with IBM Research. I have a passion for entrepreneurship and creativity. I enjoy learning skills that help me take ideas from conception to implementation.

EDUCATION

Boston University College of Arts and Sciences Expected May 2017
Bachelor of Science (double major in Computer Science and Neuroscience)

Graduate Coursework: Data Mining, Cloud Computing, Data Analytics

EXPERIENCE

Data Scientist Intern

StratoDem Analytics LLC, Boston MA
2016

June 2016 - August

- Worked in a team of analysts, using government data to build predictive models that provided profitable strategy insights to clients.
- Designed and co-developed "*Stratplotlib*", a data visualization python package built on top of folium and leaflet that improved visualization of geospatial data. <http://bit.ly/2fG0pnz>

Research Assistant

Department of Neuroscience at BIDMC, Boston MA

January 2015 - January 2016

- Analyzed and processed eye-tracking video data, identifying features that characterize the spatial attention of research subjects.
- Took initiative to write Python scripts to partially automate data processing which made the analysis of the data 50% faster.
- Created visualizations that yielded insights on the behavioral patterns of individuals.

PROJECTS

Duplicate document identification in big data: Implemented an efficient way to find duplicate book documents in big volumes of data with the use of locality sensitive hashing and sampling. <http://bit.ly/2n0ysKt>

Amazon natural language processing: Built a predictive model for user's star ratings by computing a TF-IDF vector for each user's review and finding its euclidean distance from each star's cluster.

StackOverflow Network Analysis: Visualized the interconnection of programming languages by constructing a network based on probabilistic Stack-Overflow tag matching
Visualization: <http://bit.ly/2lsC2N9>

Classifying the safety state of containers (Mentored by IBM):

Worked in a team of engineers to develop a tool that finds suspicious code in docker images. Intended for use in data centers, the tool works by mining security data from containers and then uses similarity matching to find malicious code snippets in the file binaries. <http://bit.ly/2fH0dFo>

Tripadvisor Hotel-Excellence: Web-scraped user hotel-reviews in order to build a model of factors that affect user ratings and perceived excellence of hotels. <http://bit.ly/2fG0nft>

SKILLS

Software and Programming Languages: Python (pandas, numpy, scikit-learn, scipy, networkX, matplotlib, seaborn, selenium etc.), SQL, Linux, LaTeX, Jupyter Notebook, Git

Statistical Methods: principal component analysis and dimensionality reduction

Machine Learning: clustering, feature extraction, text mining, natural language processing