Sokratis Papadopoulos

(646) 705-3295 | sokpapadop@gmail.com
sokratispapadopoulos.com | spapadopoulos | sokratis-papadopoulos

Skills

Programming: Python • Spark • SQL • R • MatLab
Tools: scikit-learn • pandas • numpy • scipy • statsmodels • gensim • pymc3 • nltk • beautifulSoup • Heroku
Machine learning: Supervised learning • Clustering • Natural language processing • Probabilistic programming
Statistical methods: Regression analysis • Time series analysis • Hypothesis testing • Panel data analysis • Causal inference • Bayesian methods • Spatial analysis
Visualization: Python (seaborn, matplotlib) • Tableau • Carto

Experience

Insight Data Science
New York, NY

Data Science Fellow — *Jan. 2019 - Present*

- Developed AirbnBoost, a web app that enables users to make faster and more informed Airbnb decisions.
- Built a machine learning-based pricing model, crawled open data to engineer location intelligence features, and characterized listings using topic modeling.
- Designed a flask app, deployed on Heroku cloud (bit.ly/2TFGqtw).

NYU Center for Urban Science + Progress
Brooklyn, NY

Graduate Research Assistant — *Sep. 2016 – Present*

- Analyzed spatio-temporal patterns in large-scale building energy data and built models to predict NYC energy consumption (~1M buildings) using Python and Apache Spark.
- Created a city-specific energy index for more than 7,500 NYC buildings in collaboration with the Mayor's Office using XGBoost and KMeans clustering. Findings used to support novel legislative frameworks. (bit.ly/2ByA6Nl)
- Applied hypothesis testing and Bayesian statistics for climate action impact evaluation.
- Contributed to grant proposals and presented policy recommendations to project stakeholders.

Masdar Institute
Abu Dhabi, UAE

Research Assistant — *Aug. 2013 – Sep. 2016*

- Used machine learning and agent-based modeling to quantify occupants' impact on building energy performance.
- Created an occupant behavior optimization framework based on machine learning and genetic algorithms. The tool speeds up the optimization process by 70% compared to brute-force simulation.
- Evaluated the accuracy and computational efficiency of time series methods and tree-based ensemble learning algorithms, such as Random Forest and Gradient Boosting, for electricity load forecasting.

Intelen Inc.
Athens, Greece

R&D Intern — *Jun. 2014 – Aug. 2014*

- Wrote scripts to automate the simulation of building occupant energy behavior using feedback from questionnaires.
- Created visualizations of household energy data using Python's seaborn and matplotlib to increase occupants' awareness of their consumption.

Education

New York University
New York, NY

PhD Civil & Urban Engineering (magna cum laude) — *Expected May 2019*

- Major: Data science
- Selected coursework: Machine learning, Big data management and analytics, Advanced topics in data science

Masdar Institute
Abu Dhabi, UAE

MS Engineering Systems & Management (magna cum laude) — *May 2015*

- Selected coursework: Applied statistics, Time series analysis, Systems optimization

Awards

Best Visualization 2017 United Nations, Data for Climate Action *(among 450 participating teams)*
Siemens Inc. Fellowship 2016-2018 Fully funded PhD research
Masdar Institute Scholarship 2013-2015 Fully funded MS studies