Sarthak Kothari

sarthakkothari.me | Boston, MA | kothari.sar@husky.neu.edu | +1 (617) 840-3450 | Availability: December 2019

EDUCATION

Northeastern University, Boston, MA Sept 2017 – Present
College of Computer and Information Science Expected Graduation – Dec 2019
Candidate of Master of Science in Data Science
Relevant Courses: Supervised & Unsupervised Machine Learning, Data Mining, Parallel Data Processing, Causality in Machine Learning

K. J. Somaiya College of Engineering, Mumbai, India Aug 2013 – Jun 2016
Secured a Bachelor of Engineering in Computer Science with *Distinction*

WORK EXPERIENCE

Center for Complex Network Research - Northeastern University, Boston MA Jun 2019 – Present
Data Science Research Assistant.

- Developed high dimensional embeddings for more than 3 million authors and 1 million publications to quantify the impact of author's presence in a paper from a network graph of scientific publications.
- Engineered metapath2vec approach in tensorflow for generating the high dimensional embeddings.

Staples Inc., Framingham, MA Jan 2019 – Aug 2019
Data Science Intern – Operations.

- Improved weekly demand forecasting by 23% for 550,000+ SKU's, by developing an ensemble of weak models.
- Visualizing delivery footprint for Staples operations to recognize key areas of interest with Python and D3.js.
- Built data pipelines to process more than 200,000+ weekly vendor invoice records through Python for reporting and visualizing shipment cost disparities.

Fidelis Cybersecurity, Bethesda, MD Jul 2018 – Aug 2018
Data Science Intern

- Increased data retrieval efficiency by 40% via migrating to Neo4j, a graph database, from MongoDB and Spark with Python.
- Executed clustering analysis on malware alert data by harnessing the intelligence obtained through analyzing relationships between nodes in Neo4j.

Hansa Cequity, Mumbai, India Aug 2016 – Jun 2017
Associate Analyst

- Modeled a logistic regression for lead scoring to rate a lead's likelihood of becoming a customer with an accuracy of 90%.
- Developed efficient data ingestion pipelines using SSIS for 100K+ records and build descriptive dashboards in SSRS enabling clients to evaluate product and campaign performance, boosting sales by 15%.

ACADEMIC PROJECTS

De-confounding Movie Revenue, Northeastern University, Boston, MA Jul 2019 – Aug 2019

- Implemented a Probabilistic-PCA model to produce latent confounders that drive movie's revenue using Pyro.
- Estimated causal effect of actor on movie's revenue by conditioning on actors with latent confounders in Pyro and PyTorch.

Bandit Algorithms in Advertising Recommendation, Northeastern University, Boston, MA May 2019 – Aug 2019

- Evaluated exploration-exploitation trade-offs using different classes of Bandit algorithms.
- Performed a comparative study between these algorithms by testing effects of hyperparameter on reward and regret for various Kaggle datasets, with UCB being the top performer.

Matrix Multiplication in Distributed Environment, Northeastern University, Boston, MA Nov 2018 – Dec 2018

- Executed a comparative study between Cannon's vs Simple Block Partitioning for Distributed Matrix Multiplication algorithm.
- Implemented the algorithms in MapReduce & benchmarked them on AWS EMR cluster on matrices of different sizes, with Canon's algorithm outperforming Simple Block Partitioning.

Predicting Student's Portuguese Grade, Northeastern University, Boston, MA Oct 2018 – Dec 2018

- Predicting student's final grade for Portuguese language course with demographic, socio-cultural and academic features.
- Building a complex-robust model which first predicts midterm performance & later predicts final grade using Python.

TECHNICAL KNOWLEDGE

Statistical Programming Languages	: Python, R, Java, D3.js, etc.
Machine Learning Skills	: Linear & Logistic regression, Clustering, PCA, etc.
Database tools & technologies	: Spark, Hadoop, Neo4j, MongoDB, SQL Server, MySQL, Postgres, Oracle
Data Mining, ETL and Visualization Tools	: Tableau, PowerBI, SSIS, SSRS, Telerik, Excel, etc.
Python Packages & API's	: Numpy, Pandas, Scikit, SciPy, Tensorflow, BeautifulSoup etc.
Tools, frameworks & Cloud Technologies	: AWS, Google Cloud Platform, Jupyter Notebook, RStudio, Eclipse, Git.