# Sarthak Kothari

www.sarthakkothari.me | Boston, MA, USA. | kothari.sar@husky.neu.edu | (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 Machine Learning, Unsupervised Machine Learning and Data Mining, Large-Scale Parallel Data Processing, Causality in Machine Learning, Introduction to Data Management and Processing, Algorithms

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

## TECHNICAL KNOWLEDGE

| | |
|---|---|
| **Development and 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. |

## WORK EXPERIENCE

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

- Developing high dimensional graph embedding methodologies using metapath2vec in python to quantify the success and impact in scientific careers of researchers.
- Developing models to use these embedding for predicting success, recommendation venues and identifying the author of the papers.

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

- Forecasting the demands for 550K+ Staples SKU's for the upcoming week by developing an ensemble of models which offered a 23% improvement to the existing model.
- Visualizing the delivery footprint for Staples operations to recognize key areas of interest using Python and Bokeh.
- Building data pipelines to process more than 200K+ weekly vendor records using Python and presented related visual stories to executive management.

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

- Migrated millions of records to Neo4j, a graph database, from MongoDB and Spark using Python in order to increase efficiency and data retrieval time for complex queries by 40%.
- Executed clustering analysis on the malware alert data harnessing the intelligence obtained by analyzing relationships between nodes in the graph database to explain how different malware files are associated with each other using Python.

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

- Modeled a logistic regression for lead scoring which would rate a lead's likelihood of becoming a customer for a client with 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

**Deconfounding Movie Revenue, Northeastern University, Boston, MA** Jul 2019 – Aug 2019

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

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

- Used different classes of Bandit Algorithms to study the exploration-exploitation trade-offs in each of them.
- Tested these systems on various datasets like Criteo, Coupon Purchase Prediction, etc. to understand the impact of the hyper parameters on reward and regret.

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

- Performed comparative study between Cannon's vs Simple Block Partitioning for Distributed Matrix Multiplication algorithm.
- Implemented the algorithms using MapReduce & benchmarked them on AWS EMR cluster using matrices of different sizes.

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

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