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, 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, Matlab, C, etc.
Machine Learning Skills	: Linear & Logistic regression, Clustering, PCA, etc.
Database tools & technologies	: Neo4j, MongoDB, Spark, Hadoop, 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 and frameworks	: Jupyter Notebook, RStudio, Eclipse, Git.

WORK EXPERIENCE

Staples Inc., Framingham, MA Jan 2019 – *Present*
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.

Student Housing Services, Northeastern University, Boston, MA Feb 2018 – Jun 2018
Graduate Assistant – Database and Operations

- Leveraged historical data using SQL to predict the student housing behavior for the upcoming semester.
- Visualized descriptive reports on Telerik reflecting the current housing trends for students living in university housing.

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 a completely automated descriptive dashboards in SSRS enabling clients to evaluate product and campaign performance and boosting sales by 15%.
- Restructured and reprogrammed SSIS data integration pipelines to efficiently ingest 100k+ of records from various sources into the system on daily basis.

ACADEMIC PROJECTS

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 which then becomes a feature to predict the final grade using Python.

Twitter Follower 2-Hop Neighbors, Northeastern University, Boston, MA Sept 2018

- Implemented replicated and reduce side joins in MapReduce as well as Spark to find 2-hop followers for the user.
- Executed the algorithms for 11M+ users in the Twitter Edges dataset using AWS EMR cluster.

Exploring and Modelling on Medical Appointment No-Show Dataset, Northeastern University, Boston, MA Mar 2018

- Excavated insights on medical appointment data-set to identify patients having high risk of not showing up to an appointment and identified KPIs that are specific to patients who do not show up for an appointment.
- Successfully designed and developed a classification model to predict no-shows in Python using scikit-learn and numpy.