Sarthak Kothari

www.sarthakkothari.me | Boston, MA, USA. | kothari.sar@husky.neu.edu | (617) 840-3450 | Availability: Jan 2019 – Aug 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, Information Retrieval.

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, SSIS, SSRS, Telerik, Excel, PowerBI, etc.
Python Packages & API's	: Numpy, Pandas, Scikit, SciPy, Tensorflow, BeautifulSoup etc.
Tools and frameworks	: Jupyter Notebook, RStudio, Eclipse, Git.
Web Technologies	: HTML, PHP.

WORK EXPERIENCE

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

- Generated insights for the housing services with the help of SQL to understand the demographics of on-campus student housing and prepare for house selection for upcoming semester by taking advantage of those insights.
- 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 – Technology Servicing

- 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

Landscape of Funding in Kickstarter Projects, Northeastern University, Boston, MA Jan 2018 – Apr 2018

- Performed exploratory data analysis (EDA) on the Kickstarter dataset from Kaggle using R in order to figure out the demographics and KPIs that prove vital in a Kickstarter campaigns success.
- Designed a statistical-model that would predict the likelihood of a Kickstarter campaign being successful.

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.
- Modeled a Logistic Regression model in Python using Numpy, Pandas and sklearn library of Python achieving 85% model accuracy using 5-fold cross validation scores.

Clustering of Data, Northeastern University, Boston, MA Jan 2018 – Feb 2018

- Conducted clustering based on unsupervised machine learning algorithm like K-Means, Gaussian Mixtures and DB-scan on multiple datasets having 10,000+ data points and 700+ features achieving accuracy of 75%.
- Executed and Implemented these algorithms using Numpy and Pandas library of Python.

Wikipedia Crawler, PageRank and Query Processing Models, Northeastern University, Boston, MA Sept 2017 – Dec 2017

- Implemented a wikipedia web crawler in Python to search Wikipedia pages for a keyword and its stem-words.
- Realized a PageRank algorithm in Python to rank the crawled webpage based on various properties of the pages.
- Devised indexer and query processing models to retrieve appropriate webpages based on query searched and performed query model evaluation to understand which query processing model retrieves the best result.