

AKSHAY KULKARNI

📍 Boston, MA, USA. 📞 857-869-0944 ✉️ akshayavinashkulkarni@gmail.com



akshay-a-kulkarni.github.io

EDUCATION

Northeastern University

M.S. Data Science

Sep 2018 - May 2020

Boston, MA | **GPA - 3.8/4.0**

BITS, Pilani – Dubai

B.E. [Hons] E.C.E.

Aug 2012 - Aug 2016

Dubai, U.A.E | **CGPA 7.5/10**

SKILLS

Programming Languages:

Python, R, Java, Scala,
JavaScript, SQL, NoSQL

Databases: MySQL/MariaDB,

PostgreSQL, MongoDB

Distributed Frameworks:

Apache Spark, Apache
Beam, Hadoop MapReduce

ML/DL Libraries: H2O, NLTK,

FastAI, SkLearn, TensorFlow,
PyTorch, Keras, SpaCy

Other Libraries: NumPy,

BS4, Pandas, OpenCV,
Tidymverse, Lucene, GenSim

Web-Development: HTML5,

CSS3, Bootstrap, Bulma,
VueJS, ReactJS

Cloud Platforms: AWS-EC2,

AWS-EMR, GCP, Render,
Netlify, Surge, Heroku

Visualization Tools:

Tableau, PowerBI, ggplot2,
Plotly, D3, DASH, Rshiny

DevOps & CI/CD: Git, GitHub-

Actions, Docker

EXTRA-CURRICULARS

Data Science Hub Founding

Member - Helped form a student org/club to provide academic and internship aid by holding outreach events, seminars & forums for all DS program students at NEU.

EXPERIENCE

Vitek Lab – Northeastern University

Jan 2020 – May 2020

Research Assistant - Machine Learning & Web Application Development

Boston, MA

- Mass spectrometry instruments and proteomic workflows are difficult and expensive to maintain
- Increase in scale of samples, metrics & observations can introduce complexities in Quality control
- Built **open-source** tools with Statistical Process Control & ML to predict sub-optimal MS operation
- Implemented **detection & inference** of unforeseen instrument problems using **simulation** with Tree-based Ensembles & Unsupervised Isolation-based models to save valuable calibration time & costs
- Updated legacy functionality & contributed towards codebase of the BioC **library** MSstatsQC in **R**
- Developed a **web-UI** for easier processing, visualization & analysis of data with report generation
- Performed the **containerization** and **deployment** of the web application to improve UX & app scaling

Khoury College of Computer Sciences – Northeastern University

Jan 2019 – Dec 2019

Graduate Teaching Assistant

Boston, MA

- Head Teaching Assistant for **Database Design + Data Collection & Mining**
- Guided & mentored students during office hours to help reinforce & learn new concepts in courses
- Helped the professor in modifying coursework & creating new assignments with student feedback
- Managed and aided other TAs with student submission evaluations and conducting **code reviews**.
- Supervised teams during the design and implementation of their projects throughout the course

Predikly

Feb 2017 – May 2018

Junior Data Analyst

Pune, India

- Supported development of **dashboards** for predictive analytics solutions with client sourced data
- Performed **standardization & migration** of data with SQL scripts improving performance by 25%
- Worked on the development of a Cloud based Curated Business Intelligence Platform for US Financial Professionals to manage large security portfolios
- Implemented **mining & tokenization** functionality for large corpora of text ranging from user incidents, reviews, documents etc. as part of **sentiment analysis** toolkit project

Zio Technologies

Aug 2015 – Feb 2016

Data/Sales Insights Intern

Dubai, U.A.E.

- Directly worked with the managers to strategize and explore new options based on **A/B tests** and **KPIs** to take decisions to help maximize profit from the analysis
- Collaborated with the analytics team to analyze and target the sales of various AV components (Security, Digital signage) in parts of the Gulf region

PROJECTS

- Data Mining & Clustering on Complex COVID19 Twitter Networks

Feb 2020 - May 2020

[Python, JavaScript, BigQuery, Beam, spaCy, GenSim, ThreeJS, D3, VueJS]

hashtag.surge.sh

- Developed a comprehensive tool to extract tweets & generate **network-graphs** to mine & analyze the **communities** or clusters from any given seed keyword/area of interest
- Implemented an **ETL pipeline** to fetch, filter, process & store tweets from Twitter stream-API to Google BigQuery & employ NLP algorithms & **graph-based** clustering for network analysis
- Deployed a SPA using **VueJS** to display **3D** visualizations of topics/groups from mined COVID data

- Distributed Matrix Factorization for Recommender Systems

Sep 2019 - Dec 2019

[Scala, Apache Spark, Breeze, AWS- EC2, S3, & Elastic MapReduce]

- Generated **new recommendations** for users from existing 100 million+ video ratings by fully implementing the scalable A.L.S. algorithm in **Scala** with Spark & Breeze from scratch
- Efficiently processed 8+ billion records containing predictions of missing user-item ratings
- Deployed the algorithm on varying sizes of **Spark clusters** on AWS-EMR to ascertain viability of factorization for One Class Collaborative Filtering for most real-world data/production systems