

# 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 Beam, Apache Spark,  
Hadoop MapReduce

### ML/DL Libraries: H2O, NLTK,

FastAI, SkLearn, TensorFlow,  
PyTorch, Keras, SpaCy

### Other Libraries: NumPy, BS4,

Pandas, OpenCV, Tidyverse,  
Lucene

### Web-Development: HTML5,

CSS3, Bootstrap, Bulma,  
VueJS, ReactJS

### Cloud Platforms: AWS, GCP,

Render, Netlify, Surge, Heroku

### Visualization Tools: Tableau,

PowerBI, ggplot2, Plotly, D3,  
DASH, Rshiny

### DevOps & CI/CD: GitHub,

GitLab, 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.

## WORK EXPERIENCE

### Vitek Lab – Northeastern University

Research Assistant - ML Modeling & Deployment

Jan 2020 – May 2020

Boston, MA

- Designed an end-to-end **open-source tool** to enhance individual & time-series Mass Spectrometry workflows using simulation techniques with Tree-based & Neural network models.
- Developed new functionality for the MSstatsQC Bioconductor **package/library** in R by leveraging Statistical Process Control & Machine Learning methods to process QC data.
- Automated **classification & inference** of sub-optimal performance in workflows using Tree Ensembles/XGBoost models in H2O as well as Unsupervised Isolation based models.
- Orchestrated the **containerization & deployment** of a web interface to improve scaling & user experience allowing remote processing & visualization of data for analysis & report generation.

### Khoury College of Computer Sciences – Northeastern University

Graduate Teaching Assistant

Jan 2019 – Dec 2019

Boston, MA

- Guided & mentored students in teams or individually during office hours to enable students in reinforcing & learning new concepts in Database Design & Data Collection & Mining.
- Assisted and managed other TAs in creating new coursework and assignments as well as evaluating/grading student submissions and conducting **code reviews**.
- Supervised teams during the design and implementation of their projects throughout the course.

### Predikly

Junior Data Analyst

Feb 2017 – May 2018

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 an automated **sentiment analysis** project.

### Zio Technologies

Data/Sales Insights Intern

Aug 2015 – Feb 2016

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.

## NOTABLE PROJECTS

### - Data Mining & Clustering on Complex COVID19 Twitter Networks

Feb 2020 – May 2020

[ Python, JavaScript, BigQuery, Beam, spaCy, GenSim, 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 social network analysis
- Deployed a SPA using **VueJS** to display **3D** visualizations & topic/groups surrounding COVID19.

### - 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**.