

# AKSHAY KULKARNI

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

## SKILLS

**Languages:** Python, R, Java, SQL, Scala, Bash, HTML, CSS, JavaScript

**Databases:** MySQL/MariaDB, SQLite, MongoDB

**VCS & CI/CD:** Git, DVC, GitHub

**DataViz & Webapps:** Tableau, PowerBI, Looker, Plotly, DASH, Rshiny

**Libraries:** NumPy, Pandas, H2O, Scikit-Learn, OpenCV, NLTK, FastAI, TensorFlow, Keras, PyTorch, SpaCy, GenSim

**Frameworks:** Flask, Django, Vue.js, Apache Spark, Beam, Lucene, Docker, MapReduce

**Cloud Platforms:** Heroku, AWS, GCP, Render, Netlify

## EDUCATION

### Northeastern University

M.S. Data Science 3.8/4.0  
Boston, MA | Sep 2018 - May 2020

### Birla Institute of Technology & Science, Pilani – Dubai

B.E. [Hons] E.C.E.  
Dubai, U.A.E | Aug 2012 - Aug 2016

## INTERESTS

Fitness, Traveling, Anime, E-sports, Reddit, Photography, Binging Community/The Office/Parks&Rec

## NOTABLE PROJECTS

### Topic Modelling and Clustering on Complex COVID-19 Twitter Networks - [Python, JavaScript, BigQuery, Beam, spaCy, GenSim, Vue.js]

- Developed an end-to-end tool to allow users or organizations to extract & analyze information from tweets around areas of their interest.
- Constructed a streaming ingestion / ETL pipeline with beam to fetch, filter, process and store tweets from Twitter API to Google BigQuery.
- Designed network-graphs from streamed tweets to employ graph-based clustering as well as other unsupervised clustering algorithms.
- Implemented an NLP pipeline with pretrained language models to tidy, tokenize, lemmatize & handle specific parsing of social media text
- Employed methods like Latent Dirichlet Allocation & Named Entity Recognition to extract & visualize various topics surrounding COVID19.

### Distributed Matrix Factorization for Recommender Systems - [Scala, Apache Spark, AWS- EC2, S3, & Elastic MapReduce]

- Implemented a scalable parallelized version of the A.L.S. algorithm to process more than about 100 million user-item interaction ratings.
- Efficiently processed & generated approx. 8+ billion records which contained predictions of ratings for missing user-item interactions.
- Deployed the algorithm on varying sizes of MR and Spark clusters on AWS and displayed a near empirical linear speedup and scaleup.

### Classification of Radiographs using Convolutional Neural Networks - [Python, FastAI, Keras, Google Compute Engine & Cloud Storage]

- Classified 112,000 high resolution chest radiographs from the NIH Dataset into 10 possible conditions using various CNN architectures.
- Reduced computational overhead by training on pre-trained CNN architectures such as ResNet50 & InceptionV3 in Pytorch-FastAI & Keras.
- Achieved 85+% validation accuracy on binary classification while employing techniques such as one-cycle-policy & cyclic momentum to tune the network optimally and allow for significantly faster convergence while training the models in cloud VM instances.

## EXPERIENCE

### Northeastern University

#### DS Research Assistant – Olga Vitek Lab

Jan 2020 – Present  
Boston, MA.

- Developed an end-to-end open-source software to enhance individual & time-series Mass Spec workflows using simulation techniques with Tree-based & Neural network models.
- Created functionality for the open-source MSstatsQC package provided in R by leveraging Statistical Process Control & Machine Learning methods.
- Automated identification & inference of sub-optimal performance in workflows using parallelized Ensembles & XGBoost models as well as Unsupervised Isolation based models.
- Deployed a reactive web application to increase user interaction by enabling them to upload, test and visualize their data remotely for optimal analysis & maintenance of their workflows.

### Khoury College of Computer Sciences

#### Graduate Teaching Assistant –

Jan 2019 – Dec 2019  
Boston, MA.

- Guided and mentored 90+ students in class or individually during office hours to assist them in reinforcing and learning new concepts in Database Design & Data Collection and 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 of students during the design and implementation of their projects throughout the course.

### Predikly

#### Junior Data Analyst

Feb 2017 – May 2018  
Pune, India.

- Performed data cleaning, visualization & contributed in development of dashboards for predictive analytics solutions with client-data.
- Worked in the sentiment analysis team, performing mining & tokenization for large corpora of texts ranging from customer incidents, reviews, documents & other sources.

### Zio Technologies

#### Data/Sales Insights Intern

Aug 2015 – Feb 2016  
Dubai, U.A.E.

- Worked with the analytics team to analyze the sales of various AV components (Security, Digital signage) in parts of the Gulf region.
- Directly worked with the managers to strategize and explore new options and take decisions to help maximize profit from the analysis.

## EXTRA-CURRICULAR ACTIVITIES

- **Data Science Hub Founding Member**-Formed a student organization to provide academic & internship aid by holding outreach events, seminars & forums for all NEU-DS program students.
- **MSPCA-Angell Volunteer** - Attended workshops and volunteered during weekends and free time to help with cleaning kennels as well as feeding and walking the dogs.

