AKSHAY KULKARNI

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

EDUCATION

Northeastern University
Master of Science, Data Science
Boston, MA | Sep 2018 - Aug 2020

Birla Institute of Technology & Science, Pilani – Dubai
Bachelor of Engineering [Hons] ECE
Dubai, U.A.E | Aug 2012 - Aug 2016

SKILLS

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

Databases: MySQL/MariaDB, SQLite, MongoDB

VCS & CI/CD: Git, DVC, GitHub, GitLab, GitHub-Actions,

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

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

Frameworks & Platforms:Flask, Django, Spark, MapReduce, Apache Beam, Lucene, Docker, Heroku, AWS, GCP

EXPERIENCE

Northeastern University **Jan 2020 – Present**
Research Assistant *– Olga Vitek Lab* *Boston, MA.*

- Collaborating with an international group of researchers to enhance individual & time-series mass spectrometry proteomic workflows using statistical sampling, factorial design & simulation techniques with Tree-based as well as Neural network ML models.
- Created & Augmented a codebase employing statistical process control & ML methods to detect anomalies or classify sub-optimal runs of instruments.
- Implemented parallelized models (Ensembles/XGBoost) for the open-source distribution of MSstatsQC framework as an R/Bioconductor package.
- Deployed a web application to enable users to upload, test and visualize their data for optimal analysis & maintenance of their workflow.

Khoury College of Computer & Information Sciences **Jan 2019 – Dec 2019**
Teaching Assistant *– Database Design + Data Analysis* *Boston, MA.*

- Guided and mentored 90+ students in class or individually during office hours to assist them in reinforcing and learning new concepts.
- Assisted and managed other TAs in creating new coursework and assignments as well as evaluating student code submissions.
- Supervised teams of students during the design and implementation of their projects throughout the course.

Predikly **Feb 2017 - May 2018**
Junior Data Analyst *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 support incidents, product reviews, documents, & other sources.

Zio Technologies **Aug 2015 - Feb 2016**
Data/Sales Insights Intern *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.

NOTABLE PROJECTS

Topic Modelling and Clustering on Complex COVID-19 Twitter Networks - [*Python, JavaScript, BigQuery, Beam, spaCy, GenSim*]
- 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 and cyclic momentum to tune the network optimally and allow for significantly faster convergence while training the models in cloud VM instances.

Building a complete Search Engine/Information Retrieval Tool - [*Python, Apache Lucene, Docker*]
- Implemented retrieval engine with several ranking algorithms such as BM25, QLM & Vector-Space Model with pseudo-relevance feedback to rank crawled, parsed and cleaned documents. Benchmarked the performance of ranking algorithms to ascertain the most optimal one.
- Optimized the search engine by performing stopping, stemming and query expansion with Word2Vec-trained embeddings & generating summarization with Luhn's algorithm