Akshay Kulkarni

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

Github: github.com/Akshay-A-Kulkarni
Website: akshaykulkarni.netlify.com
LinkedIn: linkedin.com/in/-akshaykulkarni

EDUCATION

Northeastern University **Sep 2018 - May 2020**
Master of Science, Data Science Boston, MA.
Courses: Algorithms, DBMS, Information Retrieval, Large Scale Parallel Data Processing, Supervised & Unsupervised ML

Birla Institute of Technology & Science, Pilani – Dubai **Aug 2012 – Aug 2016**
B.E. [with honors] in Electronics and Communications Engineering Dubai, U.A.E.

SKILLS

Languages: Python, R, Java, SQL, Scala, Bash, HTML, CSS
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, Scikit-Learn, H2O, TensorFlow, Keras, PyTorch, FastAI, OpenCV, SpaCy, GenSim

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

EXPERIENCE

– **Northeastern University: Khoury College of Computer Science** **Jan 2020 – Present**
Research Assistant – Vitek Lab [MSstatsQC-ML] Boston, MA.
- Collaborating with an international group of researchers to enhance individual as well as longitudinal/time-series mass spectrometry proteomic workflows using statistical sampling, factorial design and simulation with Tree-based & Neural network ML models
- Developing tools using statistical process control & ML methods to detect anomalies or classify sub-optimal runs of instruments.
- Implementing parallelized models such as Ensembles & XGBoost within tools to be included in the open-source distribution of MSstatsQC framework as an R package.
- Building a web application to enable users to upload, test and visualize their data for optimal analysis & maintenance of their workflow.

– **Northeastern University** **Jan 2019 – Dec 2019**
Teaching Assistant – Database Design + Data Collection, Integration and Analysis Boston, MA.
- Guided or individually mentored 90+ students in class or during office hours to assist them in reinforcing or 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 and tokenization for large corpora of texts ranging from customer support incidents, product reviews, documents, and 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.

PROJECTS

Topic Modelling and Clustering on Complex COVID-19 Twitter Networks- [*Python, Google BigQuery, BEAM, spaCy, GenSim*]
- Implemented a streaming ingestion / ETL pipeline with beam to filter, process & store tweets from Twitter API to a BigQuery table.
- Generated network-graphs from streamed tweets to employ graph-based clustering as well as other unsupervised clustering algorithms.
- Created an NLP pipeline with pretrained language models to tokenize, lemmatize & handle specific cleaning of social media texts.
- Employed Latent Dirichlet Allocation & other topic modelling algorithms to extract and visualize various topics surrounding COVID-19.

Distributed Matrix Factorization for Recommender Systems - [*Scala, Apache Spark, AWS - EC2/Elastic MapReduce*]
- Implemented a scalable parallelized version of the A.L.S. algorithm to process roughly 100 million user-item interaction ratings.
- Optimally processed & generated approx. 8+ billion records containing predictions of missing user-item interaction ratings.
- Deployed the algorithm on varying sizes of clusters on AWS and achieved 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 radiographs from the NIH Chest X-ray Dataset into one of the 10 possible conditions such as Cardiomegaly, Effusion, Pneumothorax 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.

Decision Support/Multi-Objective Optimization & Visualization Tool - [*Python, DASH, SQLA-ORM, Heroku*]
- Deployed a database agnostic tool with React front-end and back-end developed in Python to allow users to perform decision support analysis in presence of trade-offs between two or more conflicting objectives.