AKSHAY KULKARNI

Boston, MA, US | **(857) - 869 - 0944**
kulkarni.akshay@northeastern.edu : **Email** | **Github :** github.com/Akshay-A-Kulkarni
akshaykulkarni.netlify.com : **Portfolio** | **Linkedin :** linkedin.com/in/-akshaykulkarni

EDUCATION

NORTHEASTERN UNIVERSITY — **Boston, MA**
Khoury College of Computer and Information Sciences — Sep 2018 - May 2020
M.S. in Data Science
Courses: Algorithms, DBMS, Information Retrieval, DMP, Large Scale Parallel Data Processing, Supervised & Unsupervised ML

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI – DUBAI CAMPUS — **Dubai, U.A.E.**
B.E. [with honors] in Electronics and Communications Engineering — Aug 2012 - Jul 2016

EXPERIENCE

Research Assistant – Northeastern College of Computer Science — **Boston, MA**
– Vitek Lab [MSstatsQC-ML] — Jan 2020 - Present

- Collaborating with an international group of researchers to explore and employ tree-based (Ensembles, XGBoost etc.) and Neural network machine learning models to detect sub-optimal operation and enhance longitudinal quality control in mass spectrometry-based proteomics.
- Implementing efficient parallelized models for the framework included in the open-source distribution of the tool as an R/Bioconductor package.
- Developing an interactive web interface for the MSstatsQC framework to enable users to upload and analyze optimal performance of MS runs.

Teaching Assistant – Northeastern University — **Boston, MA**
– CS3200 [Database Design] & DS4100 [Data Collection, Integration and Analysis] — Jan 2019 - Dec 2019

- Designed coursework and assignments. Evaluated submissions and mentored students in design and implementation of projects.
- Guided & tutored students in class or during office hours to help them reinforce learning concepts such as tidying, storing, analyzing data and employing Machine Learning techniques in R as well as designing databases in MySQL

Junior Analyst — **Pune, India**
– Predikly — Aug 2017 - May 2018

- Performed data cleaning, visualizations & contributed in development of dashboards & predictive analytics solutions with client-sourced data.
- Worked as part of the sentiment analysis team, assisting in mining and tokenization for large corpora of texts ranging from customer support incidents, product reviews, social media feeds, news articles, documents, and other sources.

Data/Sales Insights Intern — **Dubai, U.A.E.**
– Zio Technologies L.L.C — Aug 2015- Jan 2016

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

SKILLS

- **Languages:** Python, R, SQL, Java, Scala, Bash, MATLAB, Octave, HTML, CSS
- **Tools & Platforms:** IDEA & PyCharm, RStudio, Jupyter, GSuite, Git, Anaconda, Tableau, PowerBI, MongoDB, Docker, AWS, GCP
- **Libraries & Frameworks**: SkLearn, NumPy, Pandas, TensorFlow, Keras, PyTorch, Django, Flask, RShiny, Lucene, Spark, MapReduce, H2O

PROJECTS

Distributed Matrix Factorization for Recommender Systems [Python, Scala, Spark, AWS-EMR]

- Developed a scalable parallelized implementation to decompose a large ratings matrix into lower k-dimensional user & item latent factor matrices using ALS algorithm in order to optimize large scale computation of recommendations in explicit collaborative filtering.
- Deployed the algorithm on varying sizes of clusters on AWS-EMR to execute the factorization achieving near linear speedup and scaleup.

Classification of chest radiographs using Convolutional Neural Networks [Python, Google Cloud Platform]

- Built CNN architectures to classify chest x-rays from the NIH Chest X-ray Dataset containing 112,120 patient records (~50 GB pre-sampling) into one of the 14 possible conditions (Cardiomegaly, Effusion, Pneumothorax etc.)
- Reduced computational overhead by training on pre-trained CNN architectures such as ResNet50 & InceptionV3 in Pytorch-FastAI & Keras.
- Employed techniques such as one-cycle-policy and cyclic momentum to tune the network optimally and allow for significantly faster convergence while training models in deep learning cloud VM instances to achieve ~80% validation accuracy on binary classification across all labels.

Decision Support/Multi-Objective Optimization & Visualization Tool [Python, DASH, MYSQL, Heroku]

- Deployed a reusable database-agnostic tool with a React front-end and the core components developed in Python to help a user with taking an optimal decision in the presence of trade-offs between two or more conflicting objectives.
- Implemented functionality to display an interactive data table as well as calculate & highlight the Non-Dominated set of solutions to the user.
- Incorporated visualization methods to plot an interactive Pareto Frontier/Curve for the selected features or objectives for intuitive analysis.

Building a complete Search Engine/Information Retrieval Tool [Python, Apache Lucene]

- Implemented retrieval engine in Python and Apache Lucene, using several ranking algorithms such as BM25, QLM & Vector Space Model with pseudo-relevance feedback to rank crawled, parsed and cleaned documents
- Optimized the search engine by performing stopping, stemming and query expansion with Word2Vec-trained embeddings & generating summarization with Luhn's algorithm

Analytical Nature Quantification of occupations in O*NET Database - [R, SQLite, kernLab]

- Performed exploratory data analysis & implemented models like Linear Models, LDA, QDA & RandomForest on the O*NET database
- Analyzed the performance of Gaussian Process Classifier on the same multidimensional data pertaining to over 974 unique occupations.
- Achieved 85%+ on all models with ~96% accuracy on the Gaussian process model for quantifying the analytical nature of an occupation.