Akshay Kulkarni

San Diego, CA
(857)-869-0944
akshkul02@gmail.com
linkedin.com/in/-akulkarni

Education

M.S. in Data Science - Northeastern University, Boston, MA — Sep 2018 to June 2020

Courses: *Algorithms, DBMS, Information Retrieval, Large Scale Parallel Processing, Un/Supervised ML*

B.E. [Hons] in E.C.E. - Birla Institute of Tech & Science, Dubai, U.A.E. — Aug 2012 to Aug 2016

Skills

Prog. Languages	: Python, R, SQL, Bash, Java, Scala, C++
Machine/Deep Learning	: SkLearn, PyTorch, TensorFlow/Keras, SpaCy, NumPy, OpenCV, NLTK
Distributed Frameworks	: Apache Spark, Beam, Hadoop MapReduce
Databases & Other Tools	: MySQL, Postgres, Mongo, Django, Flask, Pandas, Lucene, NetworkX
Web Development	: HTML5, CSS3, Bootstrap, VueJS, React, Bulma
VCS, DevOps & CI/CD	: GitHub, GitLab, GitHub-Actions, Docker

Experience

Research Data Scientist, *NCMIR, San Diego, CA* – Jul 2020 to Present

- Research & write performant code for high-throughput 2D/3D image processing & deep learning applications
- Maintaining CDeep3M codebase, a Deep Learning pipeline to perform segmentation for cellular ultrastructures
- Implement, bug fix & conduct performance evaluations on new deployed features, trained models & algorithms
- Integrating new functionality for CDeep3M-Preview, an online tool which allows end-users to rapidly test performance of any pre-trained neural network model shared & hosted on the CIL-CDeep3M Model Zoo

Machine Learning Research Assistant, *Vitek Lab, Boston, MA.* – Jan 2020 to Jul 2020

- Built open-source tools using statistical sampling & ML methods to automate Mass Spectrometry calibration
- Developed methods for prediction & inference of deviations in performance using simulation techniques with ML ensembles & isolation models to allow for pre-emptive corrections, saving time & costs.
- Implemented a web user-interface to improve demoing and usability as well as added support for on-the-fly data analysis, visualization and exports in a containerized manner for dev, test and prod environments

Graduate Teaching Assistant, *Northeastern University, Boston MA.* – Jan 2019 to Dec 2019

- Guided & mentored students during office hours to help reinforce & learn new concepts as the Head Teaching Assistant for courses in Database Design as well as Data Collection, Integration & Mining
- Managed & aided other TAs with student submission evaluations and conducting lab sessions & code reviews
- Supervised multiple groups during design & execution phases for progressive projects throughout the course

Data Analyst, *Predikly, Pune, MH* – Feb 2017 to May 2018

- Developed tools & dashboards to parse, analyze & visualize data consumed through 3rd party APIs and clients
- Implemented crawling/mining & tokenization in python on large corpora of texts for sentiment analysis as part of a News & Social Media Integration Platform

Projects

Data Mining & Clustering on COVID19 Twitter Networks – Feb 2020 to May 2020

[*Python, JavaScript, BigQuery, Beam, spaCy, GenSim, D3, VueJS, Surge.sh*] - *hashtag.surge.sh*

- Developed a tool to fetch/preprocess tweets & analyze them using NLP and graph-based clustering algorithms
- Deployed a web-app with Vue to display 3D visualizations of the networks with clustered groups for inference

Classification of Radiographs using Convolutional Neural Networks – Sep 2019 to Jan 2020

[*Python, PyTorch-FastAI, Keras, Google Compute Engine & Cloud Storage*]

- Classified high-res chest radiographs consisting of 14 possible diseases using various CNN architectures.
- Reduced computational overhead by using transfer learning with pre-trained ResNet50 & InceptionV3.
- Utilized techniques like one-cycle-policy & cyclic momentum to facilitate stable & faster convergence

Distributed Matrix Factorization for Recommender Systems – Oct 2019 to Dec 2019

[*Scala, Apache Spark, Breeze, AWS- EC2, S3, & Elastic MapReduce*]

- Deployed a distributed version of the A.L.S. algorithm with Spark & Breeze in Scala on multiple AWS clusters to perform factorization for Collaborative Filtering to generate latent representations & approximate missing data