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 [3.75/4.0] 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

Programming Languages	:	Python, Java, R, SQL, Scala, JavaScript, Bash
ML/Deep Learning Tools	:	Scikit-Learn, PyTorch, TensorFlow/Keras, SpaCy, NumPy, OpenCV
Distributed Frameworks	:	Apache Spark, Beam, Hadoop MapReduce
Databases, ORMs & Other	:	MySQL, Postgres, MongoDB, Django, Flask, Lucene, HDF5
Web Development	:	HTML5, CSS3, Bootstrap, Node, VueJS, ReactJS, Bulma, DASH
Software, Tools & Cloud	:	Docker, Portainer, Jupyter, AWS, Google Cloud Platform

Experience

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

- Developing end-to-end ad-hoc ML solutions for dense segmentation, object detection, processing and analysis of high-throughput 2D & 3D complex imaging data with Python.
- Designing new features, implementing reusable toolkits & conducting performance evaluations for trained models & algorithms for CDeep3M2, a Deep Learning pipeline to perform segmentation tasks on various cellular ultrastructures
- Building APIs & backend data processing microservices for CIL-Preview, an online tool allowing end-users to rapidly test any pre-trained models shared & hosted on the CIL-CDeep3M Model Zoo. [cdeep3m.crbs.ucsd.edu/home/faq]

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

- Integrated features & built a containerized interface in R for an open-source tool 'MSstatsQC', providing real time monitoring, early detection & prevention of mass-spec instrumental issues. [http://msstats.org/msstatsqc/]
- Developed features to predict & interpret degradation in instrument performance using statistical analysis, data simulation & decision tree-based ensembles to allow for automatic instrument calibration & correction.
- Architected an un-supervised approach to detect & annotate unlabeled collected data and provide intuitive visualization for root-cause analysis of anomalous instrument behavior.

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

- Guided & mentored students during office hours/labs to help reinforce old & learn new concepts as the Head Teaching Assistant for courses in Database Design as well as Data Collection, Integration & Mining
- Conducted code reviews & supervised teams during progressive design & execution of projects in real applications.

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

- Designed tools & dashboards to analyze & visualize data sourced from clients or consumed via 3rd party APIs
- 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 Vue web-app to display 3D visualizations of computed clusters in the networks for interpretation

Convolutional Neural Networks for Diagnosis of Chest Radiographs – Sep 2019 to Jan 2020

[*Python, PyTorch-FastAI, Keras, Google Compute Engine & Cloud Storage*] - *github.com/Akshay-A-Kulkarni*

- Reduced computational overhead by using transfer learning with pre-trained ResNet50 & InceptionV3 models.
- 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*] - *github.com/Akshay-A-Kulkarni*

- 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