Krima Doshi

1216 E Vista Del Cerro Drive, Tempe, AZ 85281

+1 (646) 203-8862 | kjdoshi1@asu.edu | linkedin.com/in/krima-doshi | github.com/CodeHime

Technical Skills

Software/Programming: Python, R, Spark, Scala, Java, C, C++, Docker, AWS (lambda, ECS, EC2, EMR, CloudWatch, S3, SNS, SQS), Matlab, SciLab, Weka, Orange, Bash

Python Libraries: Pyspark, faiss, Pandas, Numpy, Scikit, Sklearn, Tensorflow, Keras, PyTorch, OpenCV, Pillow, matplotlib, seaborn

Databases: MySQL, AWS RDS, PostgreSQL, AWS Athena, Snowflake, MongoDB

Tools: Git, Basecamp, JIRA, Hubstaff, Grafana, Google DataStudio, Jupyter Notebook, MS Word, MS Excel, MS Powerpoint

Education

Masters, Computer Science [Big Data Systems] — **Aug 2021 - May 2023 (Expected)**

Arizona State University, Arizona, USA — *3.78/4*

Post Graduate Diploma, Data Science — **Nov 2020 - Aug 2021**

Gujarat Technological University, Gujarat, India — *9.5/10*

Bachelor of Technology, Information and Communications Technology — **July 2014 - May 2018**

Ahmedabad University, Gujarat, India — *3.38/4.33*

Relevant Coursework: Data Analytics and Visualization, Statistics and Exploratory Data Analysis, Fundamentals of Machine Learning, Deep Learning, Multimedia and Web Databases, Cloud Computing, Natural Language Processing

Experience

Data Engineer, IQM, Gujarat, India — July 2018 - Aug 2020

- Engineered multiple industry grade, scalable, well documented ETL pipelines for high-volume complex data and deployed the pipelines to production for downstream data science consumptions in AWS and Snowflake.
- Processed data in the terabyte scale from multiple data sources and trillions of data points using Spark and Pyspark.
- Analysed and optimized core queries in a database reducing the latency of responses from minutes to less than a second for real time consumption in SQL.
- Implemented built-in self-checks, auto-recovery, and alert support for the modules to make the ETL reliable and fail safe.
- Created dashboards to monitor ETL pipelines and analyse the distribution of incoming data in real time in Grafana.
- Composed visualizations to deliver the performance report of the campaign and voter targeting to the client.

Relevant Projects

Image Query System with Feedback from Scratch — Fall 2021

- Performed an exploratory analysis of the input image data and information contained by converting it into different vector models like hog, elbp, color moments, etc and how to select a distance measure based on the vector model.
- Developed coding algorithms for SVD, PCA, k-means, ASCOS similarity measure and personalized pagerank, SVM, Decision tree classifier, PPR classifier, LSH and VA-files from scratch.
- Compared the performance and drawbacks for each of the algorithms created and optimised the performance.
- Designed a feedback loop on SVM and decision trees for relevant and irrelevant images.

Sales analysis of motorcycle parts — Winter 2021

- Organised and grouped data into visualizations based on the insights obtained and enlisted all the business decisions they influence.

Sentiment analysis of tweets using BERT — Fall 2021

- Performed sentiment analysis by creating a Multilayer Perceptron with BERT (with PyTorch, BertTokenizer, Google Colab).
- Benchmarked the hyperparameters to achieve 82% accuracy on 80,000 tweets.

IBM SkillsBuild: Image captioning using Flickr8k Dataset — Summer 2021

- Generated image captions with a dataset of 8000 images using CNN, transformer encoder-decoder and positional embeddings in python.

Model for prediction of release year of song for YearPredictionMSD dataset — Spring 2021

- Developed and compared the performance of sklearn's linear regression and keras' neural networks for the prediction of the release year of a song from audio features (timbre attributes) in the YearPredictionMSD dataset.

Used Car Price Prediction — Spring 2021

- Performed data exploration, data cleaning, feature visualization and selection, and prediction using Linear Regression and Random Forest (sklearn).

Extra Curriculars:

Coordinator, Data Science Club, GTU 2021 | **Best Volunteer and Marketing Head,** IEEE AUSB, Ahmedabad University 2017