Saumil Shah

http://linkedin.com/in/saumil-shah-0257 | saumil0257@gmail.com |https://github.com/saumilsjsu | https://saumilsjsu.github.io/
+1(510) 402-9585 | San Jose, CA

Summary

- Industrial experience in creating automated python scripts for ETL (Extract-Transform-Load) processes.
- Proficient in Data Analysis and creating Visual Dashboards on SAS and Tableau.
- Experienced in Machine Learning, Statistical Analysis and Data Pre-Processing.
- Experienced in writing complex SQL queries and working with relational and NoSQL databases.
- Hands-on experience with Containerization technologies like Docker and Kubernetes and most of Amazon Web Services.

Education

Master of Science in Software Engineering, *San Jose State University, CA (GPA: 3.54)* 2019 – 2020 (Expected)
Relevant Coursework: Software Systems Engineering, Enterprise Software Platforms, Cloud Technologies, Data Mining, Large Scale Analytics, Machine Learning, Big Data, Human Computer Interaction.

Bachelor of Technology in Information Technology, *NMIMS, Mumbai* 2014 – 2018
Relevant Coursework: Data Structures, Designing Algorithms, Object-Oriented Programming, Operating Systems, Data warehousing, Data mining, Base SAS, Data visualization - SAS, Database management systems, Cloud Computing.

Skills and Tools

Programming Languages:	Python, Java, Golang, C++, PHP, CSS, HTML.
Databases:	Relational (MySQL, PostgreSQL) Non-relational (MongoDB, Cassandra, Redis).
Tools & Platforms:	Git, AWS, Docker, Kubernetes, SAS, MATLAB, Heroku, RabbitMQ, REST, Excel.
Data Science:	Tableau, Hadoop, Spark, NumPy, Pandas, Seaborn, Matplotlib, TensorFlow, Pytorch, Keras, Sklearn.
Web Frameworks:	Django, Flask

Project Work

Higgs Boson Classification |Python - Tableau| - *See Project* Oct 2019 – Dec 2019
- The aim of the project was to classify the events into "tau tau decay of a Higgs Boson" vs "background".
- Implemented Logistic Regression, Random Forest Classifier and XGBoost Classifier on the Higgs Boson dataset of 818238 instances and 33 features built from official ATLAS full-detector simulation by CERN.
- Carried out required data preprocessing steps and understood the complex data using data exploration on python and tableau.
- Performed a comparative analysis between the three mentioned classifiers.

Credit Card Fraud Detection (IEEE-CIS) |Python| - *See Project* Oct 2019 – Dec 2019
- Trained Naïve Bayes, Logistic Regression, Random Forest, SVM to detect fraudulent transactions.
- Dataset was provided by Vesta Corporation with 433 features and 1.9 GB of skewed and masked data.
- Enforced memory reduction and dimensionality reduction steps for optimal data preprocessing.
- Analyzed and constructed a comparative analysis between the classifiers based on five different metrics.

Online Food Ordering Application (Full Stack Project) |AWS - GO - NodeJS- Heroku| - *See Project* Feb 2019 – Apr 2019
- Created microservices for each module by building 6 Go APIs that were running parallelly.
- Dockerized the GO APIs and deployed them on AWS as EC2 instances over multiple regions.
- The architecture ensured fault tolerance and was scalable, where each service was running independently.
- Hosted frontend on Heroku.

Go Mini-Stack Project |AWS - GO - NodeJS| - *See Project* Feb 2019 – Mar 2019
- Configured Amazon RDS and Elasticache in a private network to connect to 2 GoAPIs under an internal load balancer.
- Deployed the dockerized APIs on AWS instances.
- Established Kong Gateway for URL cleaning in a public network.
- Connected the dockerized front-end NodeJS application connected to Kong for a GUI display.

Work Experience

Data Engineer Intern, *DeveloperDB*, Sunnyvale, CA July 2020 – Present
- Built data ETL pipeline to collect, prepare and make it suitable to analyze and process.
- Designed and developed automated scripts using Selenium to scrape technical candidates from search engines.
- Analyzed and created python scripts to parse and clean one of the biggest technical talent databases.
- Identified over 10K bad data merges, outdated data, wrong data and badly formatted data.
- Performed detailed testing of the python scripts and quality assurance of the output data achieving 97% accuracy.

Community Service

Volunteer, *Indian Development Foundation.*, Mumbai May 2015 – April 2016
- Represented IDF in their W2K mission for 200 hours.