Prabhu Subramanian

San Francisco, CA | (617)-982-4335 | subramanian.pr@northeastern.edu | github.com/prabhuSub

EDUCATION

MS in Information System, GPA: 3.8, **Northeastern University, Boston, MA** **Sep 2018 - Dec 2020**

Relevant Coursework: Data Management & Database Design, Designing Data Architecture & Business Intelligence, Big Data Architecture & Governance, Project Planning & Management

BE in Electronics & Telecommunication Engineering, **University of Pune, Pune, India** **June 2011 – May 2015**

TECHNICAL SKILLS

DBMS: MySQL, Microsoft SQL Server, PostgreSQL, Oracle, Snowflake, Aurora, Redshift, DynamoDB, Teradata
ETL & Tools: SSIS, SSAS, Alteryx, Talend, Athena, Putty, Docker, Kubernetes, Terraform, Jenkins, Atlantis
Programming languages: Python, SQL
Business Integration: Tableau, Microsoft Power BI, Excel, SAP, BI HANA
Cloud Engineering: Google Cloud Platform, Amazon Web Service, Microsoft Azure, MinIO, Grafana

PROFESSIONAL EXPERIENCE

Business Intelligence & Tools Analyst, Levi Strauss & Co., San Francisco, CA **Apr 2021 – current**

- Developing Python-based tools to automate reports on cloud integrating with SAP/BI HANA, SharePoint using shell scripting
- Developed and deployed Alteryx pipelines by integrating the tableau command-line interface and published extracts to Tableau dashboards to visualize the on-hand inventory
- Established ~90-100% automation for business users based on various complex manual file integrations

Graduate Programmer Analyst, SquarkAI, Boston, MA **Jan 2019 – Dec 2020**

- Improved the execution speed of Python scripts to load and predict machine learning models of H2O Java Virtual Machine by 50%, using parallel processing technique
- Developed scripts to connect MinIO framework with AWS S3 buckets to access the customer datasets to run the predictions and export the results back to AWS S3 bucket for user consumption
- Reduced the developer's testing time by 90% by designing automated testing scripts using Python and Git CLI

Data Engineer Intern, Peterbilt Motors Company (PACCAR), Denton, TX **Jan 2020 – Jun 2020**

- **Failure Mode Identification** *(Machine Learning, Neural Networks, Natural Language Processing, AWS)*
- Improved the speed of cleaning customer warranty data by 50% through preprocessing scripts
- Improved AWS pipeline for the internal team by establishing script logs and email notifications for pipeline completion
- Developed scripts to export CloudWatch metadata to visualize resource usage using d3 visualization
- **Platinum Score Card** *(Excel, ETL, Database, AWS)*
- Developed data warehouse pipeline from various data sources like Snowflake, Teradata, Excel, Microsoft SQL Server to calculate and publish Customer scores using Python Flask application
- Deployed AWS pipeline using Docker and Kubernetes leveraging AWS Services - SNS, SQS, SES, S3, EC2, Lambda

Software Engineer, Hewlett Packard Indigo, Pune, India **Nov 2015 – Jul 2018**

- Developed Enterprise Resource Planning dashboards for Supply Chain, Manufacturing, and Finance teams
- Improved application query optimization by 68% for the product report generation by improving database schema
- Customized and implemented supply-chain optimization tools like Kanban process using SQL functions, procedures, and triggers
- Maximized the operation of Customer Self Service application by 30% by leading the QA team for application query testing

PROJECTS

Gauguin Project - AutoViz (Visualization, Data Analytics)

- Led research group of 10 students by assisting and developing BOT using Google API & Selenium on cloud
- Helped achieve 70% accuracy with the Resnet50 model to identify and remove noise in the image datasets

H2O Hyperparameter Recommendation System (Text Analytics, Feature extraction, Feature Selection, Feature Engineering)

- Trained H2O AutoML on various datasets. Scrapped and stored the hyperparameters from the best model in the leaderboard
- Created ER model for the data scrapped. Built a pipeline around the ER model to push the data to Google Cloud SQL Storage

WORD2VEC – Model Interpretability (H2O, NLP, Visualization)

- Wrote a GitBook on model interpretability in collaboration with the professor to explain the model performance metrics
- Developed r2d3 visualization scripts by integrating TensorFlow & TensorBoard for visualizing results of BERT

MS Access to Snowflake (VBA, AWS, DevOps)

- Developed a VBA script to push the data from MS Access to Snowflake using Python ODBC
- Reduced server downtime by 90% using scheduled VBA script to perform a periodic transfer between the data sources.

Kaggle Competition - Kobe Bryant Shot Selection (H2O, AutoML, Kaggle, Visualization, XGBoost)

- Performed EDA on the data of Kobe's shot on the field and built a predictive model to predict successful shots given the position
- Decreased model Log Loss from 0.60 to 0.34 by hyperparameter tuning. Secured 1st rank on the leaderboard

AWARDS

- Honored with the best Information Systems Teaching Assistant award at Northeastern University
- Awarded company's leadership value holder certificate and badge

SERVICE AND LEADERSHIP

- President of AI club at Northeastern University 2019-2020 – NEU AI Skunkworks
- Represented Northeastern University as Graduate Student Ambassador for Information Systems