POOJA PARAB

Bloomington, IN | pooparab@iu.edu | (812) 803-9788 | LinkedIn: linkedin.com/in/parab-pooja

EDUCATION

Indiana University, Bloomington, IN, United States **Aug 2022-Dec 2023**
Master of Science in Data Science
Relevant Coursework: Advance Database Concepts, Statistics, Data Mining, Algorithms, Machine learning.

D.J. Sanghvi College of Engineering, Mumbai, India **Jul 2015-May 2019**
Bachelor of Engineering in Electronics (Dept. Rank: 6) GPA: 8.9/10
Relevant Courses: Object-Oriented Programming, Structured Programming, Computer Networks, Digital Image Processing

TECHNICAL SKILLS & CERTIFICATIONS

- Programming Languages: Python (Pandas, TensorFlow, PyTorch, Scikit-learn, NLTK, Stanza), R (Shiny), Java
- Data Science Skills: Supervised learning (Regression, Classification), Unsupervised (Clustering, PCA), NLP
- Databases: MySQL, SQL Server, Oracle, PostgreSQL
- Data Visualization Tools: Tableau, Data-Studio
- Cloud Platforms: Amazon Web Services, Google Cloud Platform
- Other: Git, Docker, Kubernetes, Linux, Apache Airflow, Apache beam, Shell Scripting
- **AWS Certified Solutions Architect Associate** **Nov 2019**
- **Google Cloud Certified Associate Cloud Engineer** **Oct 2019**

EXPERIENCE

O'Neill School of Public and Environmental Affairs, Bloomington, IN
Research Assistant **Aug 2022-Present**

- Created an interactive web application for researchers using R (shiny) to extract data from complex Corelogic housing text datasets based on specified filters and queries, providing easy access to the required data subsets for individual research.
- Conducted exploratory data analysis and data visualization to assess the patterns in mortgage lending rates and home prices aiding researchers to deduce better conclusions.

Quantiphi Inc., Mumbai, India
Senior Data Engineer **Mar 2021-Jun 2022**

- Spearheaded a team of 5 to deliver 2 projects (AppFactory and Federated learning) and Proof of Concepts in the data engineering track and assisted Technical Architects to create a product backlog with Bayer Pharmaceutical as a Client.
- Introduced a platform to create Federated Learning experiments based on NVIDIA Clara's client-server approach which helped the client to train the model on local datasets without sharing sensitive medical data.
- Designed and Implemented a cloud-based architecture to create a secure MLOps platform (AppFactory) assisting clients in training and tracking their Machine Learning models with a configurable GPU, JupyterHub Server, and dependencies like TensorFlow, PyTorch, & NVIDIA CUDA.
- Accomplished cloud agnostic architecture by altering 27 Airflow DAGs, exclusively running on the AWS environment, to operate seamlessly on any other cloud with Roku as a client.
- Spearheaded the project to create the 5 tableau dashboards on medical image datasets, helping the client to select trials for model training by providing insights on the metadata and integrated with the web application.

Data Engineer **Jun 2019-Feb 2021**

- Constructed a data transfer pipeline for clinical trials using Dataflow to retrieve the data from SaMD, de-identify using a text detection model, and load data into Bigquery, reducing the processing time by almost 75%.
- Devised a data migration tool - 'Qinetic', using Dataproc, Dataflow, and Data Fusion to move TBs of data from OLTP sources (Oracle, SQL Server, and PostgreSQL) to Data warehouses (Redshift and BigQuery), for Quantiphi's internal use-cases.
- Deployed a web application on Elastic beanstalk, managed Mysql database, and configured billing alerts on AWS Cloudwatch for Simply Black Media as a client.

PROJECTS

Price Optimization for E-commerce Data **Oct 2022-Dec 2022**

- Obtained optimal price of product for boosting sales margin by plotting demand curve based on past transactions data using PCA, linear, ridge and lasso regression.

A Network Analysis of Game of Thrones **Jan 2022-Feb 2022**

- Performed network analysis on the datasets of five game of thrones books to obtain the well-known character using Pandas and Networkx libraries.
- Determined the famed character in the book by finding a correlation between three different measures i.e. degree of centrality, betweenness centrality, and PageRank.

Aspect Based Sentiment Analysis **Oct 2019-Dec 2019**

- Developed a flask API for extracting aspect (subject), opinion (adjective), and sentiment of the opinion from a long paragraph by implementing co-reference resolution using Neuralcoref and neural network pipeline for text analytics using stanza.
- Built a custom docker container, deployed it on AWS Sagemaker, and tested the robustness of the algorithm with the yelp dataset.

EXTRA-CURRICULAR & MENTORSHIP ACTIVITIES

- Conducted interviews and screening tests for hiring Data engineers in the organization.
- Mentored new joiners in the tableau and cloud certification drives arranged by the organization.