

PARIN SHAH

Los Angeles, CA | parin090199@gmail.com | (213) 285-8984 | [linkedin.com/in/parin-s-shah](https://www.linkedin.com/in/parin-s-shah) | parin09.github.io/portfolio/ | Tableau

EDUCATION

University of Southern California, Los Angeles Candidate for Master's in Computer Science (Data Science)	Aug 2021-May 2023 GPA: 4.0/4.0
Mukesh Patel School of Technology Management and Engineering, NMIMS, Mumbai Graduated as Bachelor of Computer Engineering with Distinction (Merit List)	Aug 2016-May 2020 83.5%

PROFESSIONAL EXPERIENCE

Unity Technologies, SF Data Engineer – Ads Team Cloud Cost Optimization -	Jan 2023-May 2023
---	--------------------------

- Saved **\$50K** (annualized) through optimizing several pipelines, dashboards, and datasets by refactoring the codebase.
- Extracted **asset inventory** data using GCloud API to keep track of resource labels over time by designing an **ETL** pipeline.
- Identified key areas for cost savings and optimizations by developing **GCP Cost Reporting** dashboards in Looker.
- Revamped and streamlined **LookML** codebase leading to increased modularity, reduced merge conflicts and over **4x** faster content validation times.

Unity Technologies, SF Data Engineer Intern – Ads Team Revenue Attribution -	May 2022-Aug 2022
--	--------------------------

- Discovered **1Mn+** new publishers using Unity to build games and advertising through Unity Ads by consolidating multiple datasets leading to an improved coverage over previous ETL from **2%** to **65%**.
- Identified a **63%** increase in revenue attributed to publishers using Unity to build games.
- Technologies used: SQL, Python, GCP, BigQuery, Looker, LookML, Apache Airflow, Github, Terraform, dbt.

Limechat, Mumbai Data Scientist	Jun 2020-Jun 2021
--	--------------------------

- Reduced manual labeling time by over **60%** by devising a tool to label new intents using **clustering** and **active learning**.
- Increased product sales by **28%** on average across companies by incorporating results of **A/B** test to determine optimal number of products to be displayed.
- Designed an ETL pipeline in python to process and upload over **50,000+** rows of semi-structured data per day from PostgreSQL server to product analytics tool (**Mixpanel**).
- Ingested data from disparate sources to Kafka and utilized **Kafka** streams for analytical processing and logging.
- Developed a custom analytics dashboard on Flask and VueJS displaying critical **KPIs** for each client with ability to download standard excel and pdf reports. Integrated caching algorithm using Redis leading to **90%** faster loading times.
- Lead entire DevOps initiatives by collaborating cross-functionally with various teams to design and incorporate infrastructure solutions that were used by **35** clients. Designed automated CI/CD pipelines on Gitlab for **100+** cloud servers.
- Setup Azure cluster to deploy horizontally scalable, clustered software using Kubernetes. Launched **50+** cronjobs to automate repetitive tasks.
- Technologies used: Python, SQL, Mixpanel, Flask, VueJS, Azure, Postgres, Redis, Docker, Kubernetes, Grafana, Kafka

Oracle Financial Services Software Limited, Mumbai Research Intern	May 2019-Jul 2019
---	--------------------------

- Interlinked **LDAP** server in Kubernetes for authorization purposes. Automated manual formation of access roles in banking systems helping **save 10+** hours of work per client.
- Incorporated **OpenID** connect to fetch relevant roles from server. Authorized multiple users by assigning permission to roles with the help of **Role Based Access Control (RBAC)**.

TECHNICAL SKILLS

- Programming Languages and Databases: Python, SQL, SAS, C++, Postgres, Redis, MySQL
- Python Libraries: Matplotlib, Plotly, Pandas, Numpy, Scipy, Tensorflow, Keras, Sci-Kit Learn
- Platforms and software: AWS, GCP, Azure, Apache Airflow, Tableau, Looker, Docker, Kubernetes, CI/CD, Gitlab, Jira, Notion
- Big Data: Apache Spark(PySpark), Hadoop, BigQuery, MapReduce, Kafka
- Web Technologies: Flask, Django, VueJS, ReactJS, D3.js

ACADEMIC PROJECTS AND PAPER

Immersive Visualization in Medical Imaging: Reports 3D (Paper) Python, Blender, Flask, VTK, Keras
--

- Constructed a web application to bridge the semantic gap between medical practitioners and laymen by leveraging the use of **Augmented Reality (AR)** as a graphically intensive solution.
- Devised novel algorithm for 3D **volumetric recreation** and **rendering** of organ. Generated 3D file to be viewed in AR.
- Achieved a **DICE** score of **87%** for segmenting HGGs (Brain Tumours) from MRI by applying U-Net CNN architecture.

Review of Credit Card Fraud Detection Techniques (Paper) Python, Numpy, Sci-Kit Learn
--

- Implemented **Synthetic Minority Oversampling Technique** to generate synthetic dataset to improve size of minority class.
- Employed **logistic regression** to classify transactions and achieved a F1 score of **94%**, Recall of **99%** and Precision of **90%**.
- Analyzed currently existing credit card fraud detection algorithms based on various parameters such as pre-processing, complexity, computation time, accuracy and listed the advantages and disadvantages with a suitable use-case.

Recommendation System Python, Spark, PySpark

- Designed a **hybrid recommender** system to predict the ratings by a user for a particular business on Yelp data.
- Achieved accuracy of **97.3%** by combining **item based collaborative filtering** and **model-based** recommendation systems.