

Rahul Vijay Dubey

rahuldubey.vjti@gmail.com

+1-858-260-9176

[LinkedIn](#)

MS in Computer Science, UC San Diego

2016-2018

B.Tech in Computer Engineering, VJTI, Mumbai, India

2010-2014

EXPERIENCE

Mindstrong Health

Senior Software Engineer, Machine Learning

Mountain View, CA

Oct'20 - present

• Machine Learning Infrastructure

[Python, Tensorflow, Databricks]

- Developing centralized feature-store infrastructure by leveraging Delta Lake on Databricks platform
- Developing wrappers around Databricks MLFlow to assist Data Scientists in modeling biomarkers, sleep, and hospitalization of patients

Yelp

Senior Machine Learning Engineer

San Francisco, CA

April'18 - Sept'20

• Ads Targeting System

[Spark, XGBoost, Java, Redshift Spectrum, Splunk]

- Designed the Machine Learning system for CTR prediction. Reduced log loss by over 10% and training cost by 15x thereby providing frequent retraining capabilities with consistent gains in metrics
- Developed feature generation pipeline with PySpark to interface with Java feature generation library to ensure consistent feature computation in online prediction service and offline training service

• Search Ranking System

[PySpark, XGBoost, Datalake]

- Developed the offline feature store for Search Ranking ML systems using Spark and Datalake (S3); reduced cost for feature-extraction cost by 10x and runtime by 70%, enabling large-scale feature engineering
- Implemented distributed learning-to-rank with XGBoost on Spark and achieved NDCG score of 0.77 on Restaurants ranking model

• User Modeling

[PySpark, MLlib]

- Developed end-to-end Machine Learning pipeline to predict gender, home-ownership, and auto-ownership attributes of a user; achieved AUCROC of above 75% for all 3 models
- Productionized gradient boosting model to predict user LTV in terms of ad-revenue based on user-activity; improved status-quo model by 25% for iOS and 70% for Android users
- Designed and implemented User-Business-Interaction feature store to power Recommendation models

• ML Mentorship & Consulting

[XGBoost, Spark]

- Mentored engineers on Search-Suggest team to create their first ML model to rank high-intent suggestions on iOS search for restaurant vertical queries which improved CTR and connection-rate metric by 12%
- Mentored Search-Intent team to create data-driven model to rank filters for movers-related filter queries and improved connection-rate metric by 40%

Samsung R&D Institute

Senior Software Engineer

Bangalore, India

July'14 - July'16

• Samsung-Fit Framework

[Java, Python]

- Developed a data-driven framework to increase user engagement towards fitness programs by applying Association Rule Mining over user's activity data
- Published the [research](#) in IEEE workshop on Health Informatics and Data Science, BIBM Conference, 2015

TECHNICAL SKILLS

• **Languages:** Python, Java

ML: XGBoost, LightGBM, MLlib, Tensorflow, Scikit-Learn

• **Technologies:** Spark, AWS EMR, Databricks

Databases: Redshift, PostgreSQL, MySQL, Deltalake