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

Yelp San Francisco, CA

Sr. Machine Learning Engineer *April'18 - Present*

- **Ads Targeting System** *[Spark, XGBoost, Java, Redshift Spectrum, Splunk]*
  - Redesigned 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 Framework** *[PySpark, MLlib]*
  - Developed offline feature data store using Spark and reduced cost for feature-extraction cost by 10x
  - Implemented feature extractors in PySpark to optimize feature transformation for large-scale LTR models bringing down runtime by 70%
- **User Modeling** *[PySpark, MLlib]*
  - Developed end-to-end Machine Learning pipeline for user modeling tasks 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
- **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 Bangalore, India

Senior Software Engineer *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

Academic Projects

- **Quora duplicate question detection:** *[Keras, Tensorflow]*
  Developed Siamese network with bidirectional stacked LSTM to create sentence embeddings for Quora questions to detect duplicate question pairs on Quora; achieved F1 score of 71.35% using Keras and Tensorflow
- **Rating prediction on Amazon dataset:** *[NumPy, Scikit-Learn]*
  Built a collaborative filtering model on Amazon rating dataset of 40k users and 20k items to predict ratings. Achieved rank 20 among 328 teams in the Kaggle competition with MSE 1.12

Technical Skills

- **Languages**: Python, Java **Technologies**: Spark, EC2, EMR, Spectrum
- **ML**: XGBoost, LightGBM, MLlib, Keras, Tensorflow **Databases**: Redshift, PostgreSQL, MySQL