

# Rahul Vijay Dubey

LinkedIn

Email : rahuldubey.vjti@gmail.com

Mobile : +1-858-260-9176

## EXPERIENCE

---

### Yelp

Machine Learning Engineer

San Francisco, CA

April 2018 - Present

- **Ads Targeting**

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

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

- **User Modeling**

[Spark, 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
- Created a Spark ETL data publisher service to send these predictions to Online Datastore for real-time usage

- **User LTV Prediction**

[LightGBM, Pandas, Redshift]

- Designed data ETL pipeline and helped productionize gradient boosting model to predict user LTV in terms of ad-revenue based on first few days of activity on the mobile-app
- Provided error reduction of 25% for iOS users and 70% for Android users over status-quo model

### Trip Advisor

Machine Learning Intern

Needham, MA

June 2017 - Sept. 2017

- **Next-to-do Recommender System**

[Scikit-learn, Pandas, NumPy]

Built an LTR model to learn a point-wise scoring function to recommend "what to do next" using user's recent activity, location-based, and time-based features

- **Locally popular restaurants**

[Spark, Java, Hive]

Developed a probabilistic scoring system to compute probability of a restaurant being popular among locals of the city based on ratings data of the restaurants

## EDUCATION

---

### University of California, San Diego — La Jolla, CA

Master of Science in Computer Science

Sep. 2016 - Mar. 2018

3.74/4.0

### Veermata Jijabai Technological Institute — Mumbai, India

Bachelor of Technology in Computer Science

June. 2010 - June. 2014

8.8/10.0

## 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