

Austin, Texas 78726
LinkedIn: [linkedin.com/in/rajatkabra](https://www.linkedin.com/in/rajatkabra)

Rajat Kabra
E-mail: rajatkabra19@gmail.com

Phone: +1 (669) 252-6058
GitHub: github.com/RajatKabra

Summary

Data science and machine learning professional with 6+ years of combined professional and research experience with the ability to work alongside diverse cross functional teams and take ownership.

Work Experience

SailPoint Technologies, Austin, TX (Senior Data Scientist) Aug 2018 to Present

- Successfully implemented certificate recommender system, reducing campaign completion time by more than 99%.
- Worked with engineers to draw data plans to reduce data consumption and storage requirements by over 90%.
- Proposed and productized a semi-supervised ML solution to reduce client onboarding time from months to minutes.
- Implemented visual interpretation layer in ML solutions to extract and interpret the reasoning behind all predictions.
- Partnered with business and product teams in stakeholder meetings to guide product decisions and resource allocation.
- Designed and analyzed A/B experiments to evaluate the impact of changes made in the product, design and demo.
- Worked with product managers to create ad-hoc data visualization and statistics reports for potential new clients.
- Implemented BERT and ELMo models for embedding and NER-POS methods to tag entities and identify linguistic features.
- Worked with engineering team to make pipeline changes for generating relevant features and transforming data.
- Reduced node distance computation time by 90% for large customers by replacing Jaccard algorithm with batched LSH.
- Owned and improved the core machine learning features of the acquired startups product and mentored 2 interns.

SAP Labs, Palo Alto, CA (Intern, Big Data Developer) Dec 2017 to May 2018

- Built pipeline to consume terabytes of data from sources like S3, HDFS, NoSQL database and automate metadata detection.
- Used Spark machine learning library for implementing pattern recognition algorithms to detect data patterns.
- Reduced data profiling and formatting time, and international data format conversion errors by over 80%.

San José State Tower Foundation, San Jose, CA (Research Assistant – Computer Vision) Jul 2017 to Dec 2017

- Worked on San Jose Smart city project to design a CNN powered auto-navigation drone to detect and clean graffiti.
- Implemented several models to detect the letters, style, background and artist signature from 15 graffiti classes.
- Performed data augmentation, used transfer learning and GPU training for implementation in Keras and Tensorflow.

Immaculate IT Solutions, India (Data Scientist) Jun 2015 to Jul 2016

- Made information retrieval system to extract useful business insights from data and define problems and their impact.
- Implemented end-to-end platform for performing user analytics using unsupervised machine learning.
- Designed product experiments to understand and analyze strong user engagement areas in the product for improvements.

Education

San José State University, San Jose, CA (GPA: 3.75) Aug 2016 to May 2018

- Master of Science, Computer Science

Rajiv Gandhi Technical University, India (GPA: 3.8) Sep 2011 to May 2015

- Bachelor of Engineering, Computer Science

Skills

Languages : Python, R, Java
Data : SQL, Spark, Dask, Hadoop, AWS, EMR, Cassandra, MongoDB
ML / DL / Stats : Regression, Hypothesis testing, Supervised, Unsupervised, Neural Networks, NLP
ML / DL Libraries : H2O, Spark ML, Keras, Sklearn, Tensorflow, SpaCy, NumPy, NLTK, Gensim, Pandas
Visualization : Matplotlib, LIME, Bokeh, GGplot2, Tableau

Patents

- Recommender system for certification campaigns & access requests using ensemble and boosted models. (Approved)
- Correlation Cube: Identity-Account Correlation using semi-supervised machine learning. (Filed)
- System and method for role validation in identity systems using machine learning and graph structures. (Filed)
- Machine Learning powered predictive platform for identity governance and administration. (Filed)

Personal Projects

Deep Learning Hub

- Deep learning hub for applications like image classification, image captioning, sequence analysis, and supervised learning.

Machine Learning Visualization

- User can upload data, train a model from 15 ML model options, tune the model and features, and visualize metrics.