

## Rajat Kabra

Austin, Texas 78726

E-mail: [rajatkabra19@gmail.com](mailto:rajatkabra19@gmail.com)

Phone: +1 (669) 252-6058

LinkedIn: [linkedin.com/in/rajatkabra](https://www.linkedin.com/in/rajatkabra)

GitHub: [github.com/RajatKabra](https://github.com/RajatKabra)

### Summary

Data science and machine learning professional with 5+ years of combined professional and research experience with the ability to work alongside diverse cross functional teams.

### Work Experience

#### SailPoint Technologies, Austin, TX (Senior Data Scientist)

Aug 2018 to Present

- Successfully implemented recommender system, reducing certification campaign completion time by 99%.
- Lead engineering team 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.
- Successfully implemented visual interpretation layer in ML solutions to interpret the reasoning behind 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 NLP models to translate natural language into database queries using BERT model and POS methods.
- Implemented usage anomaly detection models using machine learning and probability distribution models.
- 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 pipelines to consume terabytes of data from S3, HDFS, NoSQL database sources and automate metadata detection.
- Used SparkML for implementing pattern recognition algorithms to detect data patterns, identifying data categories.
- Reduced data profiling and formatting time, and reduced 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

<b>Languages</b>	: Python, R, Java
<b>Data</b>	: SQL, Spark, Dask, Hadoop, AWS, EMR, Cassandra, MongoDB
<b>Machine Learning / Stats</b>	: Regression, Hypothesis testing, Supervised, Unsupervised, Neural Networks, NLP
<b>ML Libraries</b>	: H2O, Spark ML, Keras, Sklearn, Tensorflow, SpaCy, NumPy, NLTK, Gensim, Pandas
<b>Visualization</b>	: 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

- Visual interface 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 performance.