Rajat Kabra

Austin, Texas 78726 E-mail: rajatkabra19@gmail.com Phone: +1 (669) 252-6058
LinkedIn: linkedin.com/in/rajatkabra GitHub: github.com/RajatKabra

Summary

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

Work Experience

SailPoint Technologies, Inc. (Data Scientist) Aug 2018 to Present
11120 Four Points Drive Suite 100, Austin, TX 78726

- Decision recommender for certification campaigns & access requests (Patent Approved)
  - Extracted over 30 meaningful features from raw data for training ensemble and boosted models with 96% ROC-AUC.
  - Implemented a visual dashboard to interpret and explain the reasoning behind predictions and feature importance.
- Correlation Cube: Identity-Account Correlation (Patent Filed)
  - Proposed, conceptualized a semi-supervised ML solution to reduce client onboarding time from months to minutes.
  - Single handedly implemented and productized the solution for scalability using Spark and H2O.
- Built Predictive Platform for role mining, access modeling, recommendation and anomaly detection. (Patent Filed)
- Successfully completed implementation for incremental training of ML models in production. (Patent Pending)
- Partnered with product managers in stakeholder meetings to guide product decisions and resource allocation.
- Partnered with QA teams on scripts to monitor ingested data quality and ML model performance in real-time.
- Removed performance bottleneck by replacing Jaccard algorithm by Spark LSH reducing computation time by 90%.
- Implemented time series forecasting models to predict spikes in client usage to boost their instance capacity.
- Used hypothesis testing to test the design interface that leads to faster user response and faster role creation.
- Implemented a usage anomaly detection solution using historic usage data and tree-based models.
- Used Word2Vec and TF-IDF for document clustering and community detection to identify anomalous user access.
- Used NER methods to tag entities and POS methods to identify linguistic features for machine translation tasks.
- Selected to present machine learning based solution at Navigate 2020, an international identity governance conference.

SAP America, Inc. (Intern, Big Data Developer) Dec 2017 to May 2018
3460 Hillview Avenue, Palo Alto, CA 94304

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

San José State University (Deep Learning Research Assistant – Computer Vision) Jul 2017 to Dec 2017
One Washington Square, San Jose, CA 95192

- Successfully implemented CNN model on an auto-navigation drone to detect and clean graffiti around the city.
- Implemented separate 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 (Data Scientist) Jun 2015 to Jul 2016
Mumbai, India

- Made information retrieval system to extract useful business insights from data and define problems and their impact.
- Mined user patterns to predict user behavior using machine learning and clustering and predict user interactions.
- Conducted product experiments to understand and analyze user engagement and interests, driving the roadmaps.

Education

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

- **Master of Science, Computer Science**

Rajiv Gandhi Proudyogiki Vishwavidyalaya, India (GPA: 3.8) Sep 2011 to May 2015

- **Bachelor of Engineering, Computer Science and Engineering**

Skills

Languages	:	Python, R, Java
Database/Big data tools	:	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 (learning)