

Data Scientist

**Data Scientist - UC Davis
San Francisco, CA**

Data Scientist, looking for opportunities to invest my skills in bringing actionable insights from the data

Authorized to work in United States for any employer

Work Experience

Data Scientist

UC Davis

September 2017 to Present

- Goal is to identify the subtypes in autism for the development of personalized and more effective therapies.
- We used hierarchical clustering methods to identify the clusters in the data based on some important features identified from random forest, further analysis to find the most significant behavioral groups is in progress.

Teaching Assistant

UC Davis

September 2017 to December 2018

- Worked as a teaching assistant in the school of business for the courses : Applied Business technology(using Python), Information systems(using PostgreSQL), Coding for Analytics(using Python).
- Worked as a teaching assistant in Graduate School of Nursing and Health professions for : Introduction to Computer science to Health professionals, Statistical Bio-Medical Computing.

Data Science Intern - Summer

GreyMatter Tech LLC

May 2018 to August 2018

- Worked as a Summer Intern on Natural Language Processing of voice capture info of Parkinson's patients.
- Evaluated the existing speech recognition API's for their efficiency and performance. Work selected for presentation at AMIA 2019 Informatics summit, among top 50 presentations in US.

Data Analytics Consultant

IQVIA Health

November 2015 to August 2017

- Worked as a functional lead developer in Scala using Spark. Built the data pipeline from RDBMS sources like Oracle, Netezza and migrated data into Hadoop (HDFS), then programmed custom algorithms in Scala and used Spark, Hive to do batch processing on the data. Successfully deployed product to production.
- Delivered business requirement to calculate distances using Geo-Coordinates of the practices and pharmacies and followed best practices of fine-tuning Spark to bring down the execution time of the pipeline by 70%.
- Identified the key points to make the product generic to deploy to production to multiple countries and developed modules to address the differences geographically, used Kibana to visualize the pipelines built.

Big Data Consultant

Capgemini

June 2014 to November 2015

- Worked as a developer using tools like Pig Latin and Hive for the analysis of the data from various clients like : AT&T, Sears, Centrica and wrote scripts to migrate the data from data warehouses to Hadoop data lake.
- Presented the results to clients which resulted in further 1M\$ deal and contract extension to identify the insights from data and present them using tableau dashboards.

Education

Master's in Data Analytics

University of San Francisco San Francisco, CA

August 2017 to December 2018

Bachelor's in Electronics and Communication Engineering

IIT Guwahati Guwahati, Assam

July 2010 to July 2014

Skills

Algorithms (1 year), Hadoop (3 years), Oracle (1 year), RDBMS (3 years), Scala (2 years), Python (3 years), Machine Learning (3 years), R (3 years), Spark (3 years), Big Data (3 years), NLP (Less than 1 year), Data Science (3 years), Business Intelligence (Less than 1 year), SQL (4 years)

Links

<https://www.linkedin.com/in/rohitchadaram>

<https://github.com/rohitchadaram/Project-Summary>

Certifications

Cloudera Hadoop Developer

August 2014 to Present

Given in Capgemini BigData Technologies Training Session.

Assessments

Indeed Assessments are tests that job seekers can use to demonstrate their knowledge and abilities.

Data Analysis — Expert

Feb 2019

View my full results at: https://share.indeedassessments.com/share_assignment/qw9jdgrsazwimv-4

Indeed Assessments provides skills tests that are not indicative of a license or certification, or continued development in any professional field.

Additional Information

RELEVANT TECHNICAL SKILLS

Machine learning algorithms and BigData tools: classification, regression, clustering, statistical analysis, docker and kubernetes, git, jira, confluence, neo4j, Cassandra, Hadoop Ecosystem (Cloudera and HortonWorks (Hive, PigLatin, Impala, Mapreduce)), Spark(Core, SQL, Streaming), RDBMS(Oracle, Netezza), Natural Language Processing.

Programming Languages: Scala, Python, Java, R, SQL

Packages and Visualization Tools: Python - Scikit, Pandas, Numpy, Matplotlib, Seaborn, R - caret, ggplot2, mice, Tableau, Kibana.