ROHAN SADALE

1700 N 1st St, Apt 320,   **Phone**: +1(612)404-4593   **Email**: sadal005@umn.edu
San Jose CA 95112   **LinkedIn**: linkedin.com/in/rohansadale

Education

University of Minnesota, Minneapolis, MN, USA   September 2015 - June 2017
Master of Science, Data Science   GPA: 3.793/4.000

College of Engineering (COEP), Pune, MH, India   July 2008 - July 2012
Bachelor of Technology, Computer Engineering   GPA: 7.99/10.00

Technical Skills

Languages: Python,Go, Shell, Ansible and Javascript
Databases: MongoDB, MySQL and Oracle 11g
Softwares/Utilities: Hadoop, Spark, Tableau, scikit-learn and Tensorflow

Industry Experience

- **Veritas Technologies LLC**, Mountain View, CA, USA
  Data Science Intern   June 2016 - September 2016
  - Designed a data collection pipeline to extract appliance health *(performance)* information from MongoDB and unstructured archived logs files *(on AWS S3)*; *intelligently scrapped ~300GB of archived data*
  - Performed Exploratory Data Analytics on appliance data to understand failure correlations and trends/patterns in their performance over time
  - Communicated and presented findings, orally & visually in a way that can be easily understood by hardware design teams and business counterparts *(used Tableau dashboards for creating visualizations)*

- **Barclays**, Pune, MH, India
  Technology Analyst   July 2012 - July 2015
  - Designed a streamlined Database-ETL system for generation of financial reports for APAC regulatory bodies
  - Implemented underlying Database architecture, data models and methods for data extraction from heterogeneous trading databases; developed data transformation and statistical analysis packages using Python
  - Maintained and enhanced batch framework using Perl and Shell scripting; delivered Functional specifications, Scope documents, Interface documents and data mapping documents
  - Practiced agile methodology for project deliveries; configured JIRA for the implementation

Academic/Personal Projects

- **[Kaggle] Animal Shelter Outcomes:** Built an ensemble model using Random Forest, Logistic Regression and Xgboost to predict the outcome of animals as they leave Animal Center; performed exploratory data analysis to get insights which could help shelters focus their energy on specific animals who need a little extra help in finding new home. *(I was ranked in top 8% (from ~1600 participants).*

- **Clustering and Classification of emails:** Implemented Ridge Regression classifier, kNN classifier, Centroid-based classifier and Spherical k-means for classifying and clustering emails into categories.

- **Web Scraping & Data Analysis:** Developed a scrapper to extract restaurant information from Yelp. It used a novel technique to locate regions precisely using Google Maps API and extracted information from those regions through Yelp. Later, performed Exploratory data analysis to get insights and Sentiment analysis on the user reviews to suggest quality improvements for restaurant owners.

- **A Scalable Movie Recommendation System:** Built a movie recommendation system using Big Data Tools. The project involved multiple steps such as extracting data from MySQL and Twitter (scraping), data cleansing using Apache Hive and building a scalable item-based recommender using Apache Mahout.

- **MapReduce like Compute Framework:** Developed a MapReduce framework for performing computations in distributed fashion; implemented techniques like proactive fault tolerance, fault detection and recovery to guarantee correct results even in the presence of node failures.

- **Distributed File System:** Implemented a Distributed File System in which clients shared files with each other. Multiple clients could read/write files at the same time and the files would be replicated (distributed) across several file servers. Consistency across these files was achieved using Quorum based techniques.