Sameer Darekar
Data Scientist Intern - QxBranch

Fairfax, VA - Email me on Indeed: <u>indeed.com/r/Sameer-Darekar/6cf1481dd2bd8249</u>

Willing to relocate: Anywhere
Sponsorship required to work in the US

WORK EXPERIENCE

Data Scientist Intern
QxBranch - January 2017 to Present

Modelling Cyber Risk
• Working on creating a fully integrated solution that provides reinsurance analysts with a solution for assessing cyber risk of organizations to effectively write policies and build portfolios.
• Fuzzy joining various datasets based on company names and creating model for external breach.
• Creating models for key word extractions and text summarization for extraction of features from news articles and social media.
• Worked on temporal analysis by merging censys data with other sources and finding patterns before breaches.
Stochastic Model for Fantasy League Prediction
• Built stochastic model for predicting the number of points a player will score in a match of cricket using multiple number of simulations.
• Wrote script to gather all the data from ESPN by web scrapping and updating it every match.
• Generated equipoised matchups for maximizing profits in bets by expanding the elo ranking system in chess to cricket.

Telstra Network Disruption Classification - September 2017 to December 2017

• Predicted the severity of service disruptions on the Telstra network using Neural Network and Logistic Regression
• Used a dataset of features from their service logs and feature engineering to get a log loss score of 0.55
POS Tagging using Hidden Markov Model Fall 2017
• Implemented Parts of Speech Tagging using Hidden Markov Model(HMM) and higher order HMM using Viterbi Algorithm
• Got an accuracy of 94% on Brown Corpus at word level and 91% at sentence level.
Topic modelling using Semi Supervised and Unsupervised Techniques Fall 2016
• Used EM algorithm for identifying topics in text with a accuracy of more than 90% using a visibility factor of 0.5

Tetris game playing AI bot Fall 2016
• Created a bot to play Tetris game using two stage look ahead probabilistic search approach using python.
• Stood second in a class of 100 students considering of Ph.D. and masters' student in the tournament.

San Francisco Crime Classification - March 2016 to May 2016

• Finished in top 40% out of 2335 by using Gaussian Naïve Bayes Method (Python-Scikit Learn) and feature engineering.
• Found meaningful patterns, Spatial and Temporal Data plot to give insight of the crime in San Francisco using R(ggplot2), Tableau

SQL Developer

Accenture India Pvt. Ltd - August 2014 to November 2015

• Performed root cause analysis of the anomalous data generated due to faulty design of BFG (Big Friendly Giant) DB
• Worked on data cleansing with the help of Python and PL-SQL.

(Software Engineering Analyst) (SQL and Hive Developer)

Accenture India Pvt. Ltd - June 2012 to November 2015

SQL and Hive Developer

Accenture India Pvt. Ltd - April 2013 to July 2014

• Performed migration of Historical data of customers to HDFS with Sqoop and cleaned data with Hive for inconsistencies.
• Awarded with Accenture Stellar Award for generating reports using UDFs of potential money laundering customers.
• Completed Accenture's Banking Generalist Certification Program a fundamental banking domain certification program.

Database Tester

Accenture India Pvt. Ltd - August 2012 to April 2013

• Examined the backend flow of data from upstream and downstream applications while working in Orchestration Engine Team.
• Automated the manual task of validation of data with Python and VBA scripts.

EDUCATION

Masters in Data Science

Indiana University Bloomington
December 2017

Bachelor of Engineering in Information Technology

University of Pune
May 2012

SKILLS

APACHE HADOOP SQOOP (1 year), Hive (4 years), Python (2 years), SQL (4 years), Sqoop (1 year)

LINKS

http://darekarsam.strikingly.com

https://github.com/darekarsam

https://www.linkedin.com/in/darekarsam

ADDITIONAL INFORMATION

Technical Skills
Programming Languages: Python, R, Java, C#.net, C/C++, VBA Macros

Tools/Frameworks: Map Reduce, Pig, Hive, Flume, Sqoop, Shell Scripting, Tableau, Dataiku Data Science Studio, Tensorflow
Database/NOSQL Databases: PL-SQL, Mongo DB, HBase