SHARANG KULKARNI
202-817-9908 • sharang94@gmail.com • https://www.linkedin.com/in/sharangkulkarni/

EDUCATION

GEORGE WASHINGTON UNIVERSITY Washington, DC
Master of Science, Data Science (GPA: 3.77) Aug 2016 - May 2018
Relevant courses: Data Mining, Machine Learning I, Social Network Analytics, High Performance Computing, Supply Chain Analytics Machine Learning II, Advanced Statistics, Visualization of Complex Data

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA Surathkal, India
Bachelor of Technology, Electronics and Communication Engineering (GPA: 8.39/10) July 2012 - May 2016

SKILLS

Programming: R, Python, SAS, C, C++, Java, MATLAB **| Visualization frameworks**: Tableau, d3.js, ggplot, bokeh
Big Data: Hadoop, Spark **| Cloud Computing**: AWS, Azure, GCP **| Databases**: MySQL, MongoDB
Miscellaneous: MS Excel, Stata, Weka, ArcGIS, QGIS | **Other**: Bash, Git, CLI

Technical: Regression, Classification, Clustering, Deep Learning, Collaborative filtering, Time-series analysis, Association rules, Sentiment analysis, Topic modelling, Journey analytics and Validation methods

EXPERIENCE

HOPE ONE SOURCE Washington, DC
Data Scientist Nov 2016 - Present
- Heading the planning and construction of the data engineering framework for the Impact Analysis tool, using Python and AWS (DynamoDB, CloudWatch, EC2, S3 and Lambda). Inputs- SurveyMonkey (API) and internal data (POST API). Outputs- Attendance reports (Excel/PDF) and Tableau visualizations
- Developed data dashboard using d3.js to visualize demographic and geographic information of clients in DC, to help determine the most suitable location for providing a service
- Presented my work to James Jefferies, the 2018 IEEE President

GEORGE WASHINGTON UNIVERSITY Washington, DC
Data Science Analyst Sep 2017 – May 2018
- Contributed to an Insider Threat detection tool using AWS (S3 and EC2), Machine Learning and ElasticSearch
- I worked on building an ensemble of anomaly detection algorithms using multivariate inputs
- Tools/methodologies: AWS, Elasticsearch, Kibana, python, pandas, bokeh, scikit-learn

STATE OF DATA SCIENCE EDUCATION Washington, DC
Under Prof. Brian Wright (Data Science department, George Washington University) May 2017 – Sep 2017
- Performed data mining and analysis of course information across different MS in Data Science programs
- Created dynamic visualizations on Tableau, connected them to Google Sheet spreadsheets to update automatically

X-RAY SOURCES CLASSIFICATION Washington, DC
Under Prof. Oleg Kargaltsev (Physics department, George Washington University) Jan 2017 – May 2017
- Performed data cleaning, imputation and classification on datasets associated with energy emissions by x-ray sources

DATAKIND Washington, DC
Data Ambassador Oct 2016 – Dec 2016
- Built Random Forest model to model outcome desirability regarding children in foster care
- Helped develop app in bokeh to analyze the lifecycle of the children in the system
- Managed multiple subgroups of volunteers and eventually presented findings to a mixed audience of technical (data scientists) and non-technical stakeholders

STATUS PREDICTION FOR BIKESHARE SYSTEM Aug 2016 – Dec 2016

- Developed an automated data collection and analysis system to predict station statuses 20 minutes in advance, by using historical status and weather data (retrieved by web-scraping).
- Tools: Hadoop, R (Random Forest), SQL, Tableau

DEFAULT PREDICTION FOR LENDING CLUB Aug 2016 – Dec 2016

- Created a system to model and predict loan defaulters.
- Tools: Python (ensemble of Random Forest, Neural Network and Boosting methods) running in GCP

AUTOMATIC ESSAY EVALUATION Jan 2016 – May 2016

- Designed a tool to evaluate essays based on textual content, using Machine Learning techniques with a combination of statistical, grammatical and linguistic metrics derived from the text features
- Performance-wise, Random Forest outperformed Bayes and kNN after a parameter sweep

SEARCH ENGINE RESULTS RELEVANCE Aug 2015 – Dec 2015

- Contributed to the design of an algorithm to rate the effectiveness of a search engine using Machine Learning
- For feature extraction, we performed basic text cleaning followed by counting features in the search term
- The machine learning phase included SVD, SVM, Grid Search and Ensemble Average

DATAPHI LABS Bangalore, India

Data Scientist Intern June 2015 – July 2015

- Performed in-depth analysis of data pertaining to a fitness-tech platform, creating interactive dashboards containing temporal features and word clouds
- Queried data using SQL and then transformed text data to a structured format, to perform text analytics
- Developed programs to perform Journey Analytics and Habit Modelling using Machine Learning in Python

SONG GENRE CLASSIFICATION Jan 2015 – May 2015

- Designed an algorithm which classified the genre of 93% of test songs accurately
- Used signal processing techniques to extract MFCC's of each song, which act as explanatory features
- Implemented classification using K-Nearest Neighbors with KL divergence as the distance metric

SIMPLEWEALTH Bangalore, India

Social Media Analytics Intern Dec 2014 – Jan 2015

- Contributed to the project to identify influential users on Twitter and Facebook
- Used Python APIs to mine the data and stored the data in MongoDB after extracting relevant information
- The data was then used to populate Neo4j to identify second order influencers.
