Karan Shukla

github: shuklak13 | medium: @shuklaswag | email: shuklak13@gmail.com | ph: 214-469-8677

Skills

- Programming: Python, Java, Scala, Git, Bash
- Data Engineering: Hadoop, Spark, SQL, Excel, Elasticsearch, Kafka, Docker
- Data Science: R, machine learning, statistical modeling, visualization

Education

- M.S., Computer Science [Fast Track] — University of Texas, Dallas — Spring 2018
  Data Science Specialization
- B.S., Computer Science [Honors] — University of Texas. Dallas — Spring 2017
  Minor in Statistics — Magna Cum Laude (GPA: 3.85) — National Merit — Phi Kappa Phi

Work Experience

- Strategy Consultant Intern at IBM Chief Analytics Office — Summer 2017
  - Extracted, cleaned, and transformed real-time text data with Elasticsearch, MongoDB, and Python for use in a text classification and event detection model
  - Led initiative to modernize a web application by…
    - containerizing it with Docker
    - deploying it to Kubernetes and Bluemix/Cloud Foundry clouds
    - creating a REST API for developer access, documented with Swagger
    - speeding up machine learning computations with Apache Spark
- Data Science Intern at IBM Extreme Blue — Summer 2016
  - Built real-time anomaly detection system using Spark, Kafka, and Scala using both supervised and unsupervised techniques
  - Used Scrum agile framework to bring our idea from conception to prototype to C-suite executive presentation in less than three months with a team of four interns
  - Recovered from disastrous mistakes using Git version control
- Data Science and Analytics Intern at Verizon — Summer 2015
  - Developed statistical programming approaches in R to solve a variety of consumer business problems, including…
    - allocating resources using regression techniques
    - identifying upselling opportunities via classification and segmentation
    - visualizing customer sentiment with ggplot2
  - Learned and implemented a variety of regression techniques, including…
    - linear and logistic regression
    - stepwise regression and LASSO for feature selection
    - time series models (autoregression)
  - Queried data with SQL and created visualizations in Excel

Graduate Coursework

- Machine Learning — Statistics for Data Science — Databases — Parallel Processing
  Cloud Computing — Distributed Computing — Big Data — Computer Networks

Independent Projects

- Tunify won Best Microsoft Hack — Aggregate + visualize geo-based Spotify playlists — HackTX 2017
- anonymizeR — Open-source R package for anonymizing data — Fall 15 – Spring 16
- SentweetSearch — Twitter sentiment analysis app built in Python — Fall 13 – Spring 14

Extracurriculars

- ACM UTD, Director of Public Relations — Summer – Fall 17
  - Led Slack community used by over a dozen engineering clubs and 500 students
  - Organized campus engineering events with student clubs and industry representatives
- ACM UTD, Co-Director of Projects — Spring 17
  - Coordinated with students, faculty, and company sponsors to host engineering talks