

# Sumedh R. Sankhe

Boston, MA | [sumedh.sankhe@gmail.com](mailto:sumedh.sankhe@gmail.com) | 857-350-2859 | [linkedin.com/in/sankhe/](https://www.linkedin.com/in/sankhe/) | [sumedhsankhe.github.io/](https://sumedhsankhe.github.io/)

## Education

**Northeastern University, Khoury College of Computer Sciences, Boston, MA**

May 2020

Master of Science in Data Science

**Relevant Courses:** Algorithms, Machine Learning, Data Processing & Management, Natural Language Processing, Time Series Analysis

**Awards:** "The Best Real Impact Team" at **GE Aviation Hackathon**, Double Husky Scholarship

**Northeastern University, College of Engineering, Boston, MA**

May 2016

Master of Science in Industrial Engineering

**Relevant Courses:** Database Design & Management, Financial Management for Engineers, Lean Concepts & Applications, Probability & Statistics, Operations Research

## Technical Knowledge

Programming Languages	: R, Python, SQL, HTML, JavaScript, MATLAB
Softwares & Tools	: R-Studio, Jupyter Notebook, MySQL, PostgreSQL, Spark, Keras, TensorFlow, H2O.ai, Docker, Git, Tableau, PowerBI, R-Shiny, Flask, JIRA, Trello, Scrum
Visualization & Machine Learning & Statistical Packages	: ggplot2, shiny, tidyverse, data.table, caret, scikit-learn, matplotlib, NumPy, pandas, scikit-learn, pyspark, sparklyr, flask
Certification	: Six-Sigma Green Belt

## Professional Experience

**Northeastern University, Boston, MA**

Jan. 2020 – Aug. 2020

**Research Assistant**

- Created a computational biology web-application tying the MSstats ecosystem for conducting Proteomic Experiments
- Minimized software overhead by 100%, Incorporated capabilities to perform PCA to identify protein options in simulations
- Integrated simulation of data with varying parameters, applying various classification algorithms aiding identifying sample sizes
- Conceived self-regulated testing process prior to deploying to shiny-server, curtailing bugs by 25%

**Enel-X, Boston, MA**

Jul. 2019 – Dec. 2019

**Systems & Analytics Co-op**

- Constructed a R-Shiny web application (UTI) as component of data pipeline to inspect transformed interval data, UTI allowed fault identification and visualization in converted data minimizing process time by 15 days
- Improved data read/write speeds by 60%, diminished redundancy by 35%, managed storage model of converted utility data
- Devised a deterministic Nomination algorithm for customers enrolled in the TVA program leveraging past performance metrics and load data from Salesforce and DAG API's, shedding light on 5% reduction in performance year over year

**EnerNOC Inc., Boston, MA**

Jan. 2018 – Jul. 2018

**Systems & Analytics Co-op**

- Spearheaded development of Load Forecast & Reporting tool for the Polish Demand Response program facilitating Program Managers, supporting real time & on-demand Performance Calculation logic and pruning computation overhead
- Partnered with program managers to generalize & self-regulate a customer ranking algorithm across multiple programs with several measurable metrics, helping prioritize customers according to computed ranks and estimated earnings
- Designed and constructed a R-Shiny dashboard for easy report creation referencing Salesforce, multiple user inputs and pricing data from the Independent Electricity System Operator's (Ontario, Canada) website
- Saved 2-3-man hours daily, automated price crawling and populating records in Salesforce for the Energy Markets team

**Affect Mental Health, Harvard Innovation Labs, Boston, MA**

May 2017 – Aug. 2017

**Data Analyst**

- Reduced new signup response times to under 5 minutes leveraging Google Scripts to build an email response system
- Scraped Tweets and Reddit posts associated with key mental health tags utilizing BigQuery and Twitter API with R
- Analyzed extracted text to help marketing and fundraising teams to find influential personalities and subject matter experts
- Facilitated interviewers to build a template to extract key data points from interviews, curtailing data entry time by 25%

# Sumedh R. Sankhe

Boston, MA | [sumedh.sankhe@gmail.com](mailto:sumedh.sankhe@gmail.com) | 857-350-2859 | [linkedin.com/in/sankhe/](https://www.linkedin.com/in/sankhe/) | [sumedhsankhe.github.io/](https://sumedhsankhe.github.io/)

## Academic Projects

---

**Northeastern University, Boston, MA**

### **Baseline Load Calculator, Python**

Oct. 2019 – Nov. 2019

- Devised a baseline load calculator incorporating the New York Independent System Operator (NYISO) published guidelines
- The load calculator provided accurate baseline load values with the ability to include weather adjustment and incorporate previous dispatch days for use cases pertaining to Demand-Response events

### **Automated WebCrawler and Email System, R/Python/MySQL/Flask**

May 2019 – Jul. 2019

- Implemented an automated web-crawler in Python using BeautifulSoup, and storing the data in a MySQL database instance
- Setup an automated reporting and email system with R, updating a user of any new postings observed in the previous day, curtailing search times by 2 hours daily.

### **Opinion Generation using text summarization, Python, NLP**

Feb. 2019 – Apr. 2019

- Operationalized a n-gram based supervised phrase generation model which incorporated tunable parameters
- Flesch Readability Scores and Local PMI scores were calculated to measure and select appropriate candidate phrases
- Researched attention based seq2seq model and RNN models to improve summarization ability of the pipeline

### **Deep Embedded Clustering using Fashion-MNIST, Python/TensorFlow, Unsupervised Machine Learning**

Sept. 2018 – Dec. 2018

- Implemented multilayer autoencoders using TensorFlow architecture to train a deep learning model capable of clustering data, greedy layer wise training of the autoencoders was avoided by using batch normalization coupled with RELU Activation.
- Correlated findings with simpler classification algorithms like KNN, SVM and Decision Trees, found 15% increase in accuracy

### **Predicting House Prices in Boston, R/Tableau, Supervised Machine Learning**

Sept. 2017– Dec. 2017

- Implemented multiple machine learning algorithms to predict housing price in the Boston area, using the housing dataset
- Simple models like Linear Regression and MLR were tried out with some feature engineering to gain more insight in the data, more complex methods like xgboost and other tree-based methods were implemented to compare performance and accuracy

### **“Flashlight”, A Shiny app to visualize Boston Housing prices, Data Processing and Management**

Sep. 2017 – Dec. 2017

- Developed a R-Shiny app to visualize the finding off the housing prices, using the leaflet library to make the app interactive
- Using a SQL-Lite database on the backend to store and reference data