Samar Dikshit
Boston, MA | 929-217-0015 | dikshit.s@northeastern.edu
samar14641.github.io | github.com/samar14641 | linkedin.com/in/samar-dikshit

EDUCATION

Northeastern University, Boston, MA — September 2019 – Present
Khoury College of Computer Sciences — Expected Graduation: December 2021
Candidate for Master of Science in Data Science, GPA: 3.78
Related courses: Data Management and Processing, Machine Learning, Information Retrieval, Algorithms

Manipal Institute of Technology, Manipal, India — July 2015 – July 2019
Department of Information and Communication Technology
Bachelor of Technology in Computer and Communication Engineering (Data Analytics minor)

TECHNICAL KNOWLEDGE

Programming Languages:	Python 3, R, C++, Cypher, SQL
Data Science Technologies:	Jupyter, Matplotlib, NetworkX, NLTK, NumPy, pandas, PyTorch, Rasa, scikit-learn, Seaborn, TextBlob, caret, tidyverse, MySQL, Neo4j, Elasticsearch, Git
Certifications:	Querying with Cypher in Neo4j 4.x, Data-driven Astronomy
Operating Systems:	Windows, Ubuntu

EXPERIENCE

Analytics Engineering Intern — January 2021 – August 2021
DTonomy Inc., Cambridge, MA

- Analyzed cyberattack data in Python based on the MITRE ATT&CK database and developed patterns to identify Defense Evasion and Lateral Movement attack patterns for the SOAR platform
- Created bots for Slack using Rasa 2 to use services like Google Analytics and AbuseIPDB
- Developed Node-RED automations to connect Elastic Security with the SOAR platform
- Analyzed user behavior and trends on the website to help increase the overall users and decrease bounce rate

Research Assistant — June 2020 – January 2021
Center for Complex Network Research, Northeastern University, Boston, MA

- Collected, processed, and explored data with over 1.5 million samples related to philanthropies, non-profits, and universities from the LittleSis API using Python
- Worked on matching the names of organizations and people across the new data and previous GuideStar data using TfidfVectorizer, CountVectorizer, and pairwise kernels, thereby expanding the previous network
- Created and analysed a new network with more than 97,000 relationships between sociopolitical entities to determine factors that influence grants and donations using NetworkX and pandas

Teaching Assistant — May 2020 – December 2020
Northeastern University, Boston, MA

- DS2000 Programming with Data, CS3000 Algorithms and Data

PROJECTS

Detecting Brain Tumours using Machine Learning — October 2020 – December 2020

- Trained a set of classifiers to detect a brain tumour when given an MRI scan in Python
- Used decision trees, adaptive boosting, and a convolutional neural network to obtain a peak sensitivity and accuracy of 98.27% and 99.17% with cross-validation, hyperparameter tuning, and feature selection

World War II Information Retrieval — February 2020 – April 2020

- Built a Python web scraper to collect over 90,000 documents related to World War II, processed the text using NLTK, and stored them in an Elasticsearch index
- Implemented PageRank and HITS to calculate relevance scores for the documents for a given set of queries

Similarities and Differences between News Sources in the United States — October 2019 – November 2019

- Scraped over 72,000 news articles from 8 news sites and used R to create visualizations illustrating bigrams, word associations, and sentiment to prove the existence of media bias depending on political lean
- Developed Fasttext and SVM models with a peak sensitivity and specificity of 92.86% and 92.54% respectively to classify articles by political lean based on their headlines

More projects can be found here.