

Dvija Muktesh Shah

Boston, MA - 02215 • +1 (508)-904-0452 • shah.dv@northeastern.edu • [LinkedIn](#) • [Github](#) • Available May 2022 - Dec. 2022

Education

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

Expected May 2023

Master of Science in Data Science

Relevant Coursework: Introduction to Data Management and Processing, Algorithms, Linear Algebra for Data Science, Supervised Machine Learning, Natural Language Processing.

Gujarat Technological University, L.J. Institute of Engineering and Technology, Ahmedabad, India

July 2015 - May 2019

Bachelor of Engineering in Computer Engineering

Relevant Coursework: DBMS, Calculus, Linear Algebra, Data Structures, Analysis and Design of Algorithms, Python, AI.

Work Experience

Acute Informatics Pvt Ltd, Ahmedabad, India

June 2019 - July 2021

Software Engineer

- Developed a product that enables customers to initiate loan applications across banks and receive the best possible offer. This product enables administrators to customize loan products with great flexibility to meet the needs of businesses. This product cuts the overall time it takes to obtain a loan by 80%.
- Designed a Web Scraper tool using Python's BeautifulSoup library to integrate with a third-party service for which Public API was not available. This entire manual process which used to take upto 7 days was cut down to some minutes.

Widhya, Remote

Jan. 2021

Data Science Intern

- Stock Price Prediction Model : Created a Stock Price Prediction model using linear regression to predict future stock prices at a particular given date, with an MSE score of 1.09.
- Covid-19 Analysis : Predicted the spread of Covid-19 using Quantitative modeling by modeling the exponential trend lines to forecast the rate of growth of the spread of disease.

The Sparks Foundation, Remote

Nov. 2020

Data Science and Business Analytics Intern

- Performed Prediction using Supervised Machine Learning by designing a Linear Regression Model using Python to predict the scores of students based on the number of hours they studied with an MAE of 4.97.

Projects

Spam or Ham SMS Classifier

Feb. 2021

Northeastern University, Boston, MA

- Constructed a Naive Bayes classifier for an imbalanced class SMS data set on which data upsampling and preprocessing was performed to forecast if a message is spam or not spam, and achieved an accuracy of 97.2%.
- Extracted the top 5 words with the highest probability of classifying an email as spam or ham.

Analyzing the Income Disparities between Foreign-Born and Native-Born U.S. workers

Sept. 2021 - Dec. 2021

Northeastern University, Boston, MA

- Extracted 2019 U.S. Census Public Use Microdata Area (PUMA) data using the Tidycensus package for a spatial analysis of state-wise immigrant distribution in the United States, and then used Exploratory Data Analysis (EDA), Hypothesis Testing, and Modelling with linear regression to weigh the features and yield the most significant elements out of approximately 1000 factors that influence income levels.
- Fostered Interactive map using Leaflet library that visualized the demographics of immigrants by PUMA.
- Based on findings, it was inferred that the earnings of native-born were comparatively higher than immigrant workers by 12%.

Analyzing Donald Trump's Tweets

Nov. 2021

Northeastern University, Boston, MA

- Converted a dataset of every single tweet from Donald Trump from 2015 to 2020 into a bag of words Tf-Idf (term frequency-inverse document frequency) for machine learning model building.
- Applied principal component analysis to determine which word has the largest effect on the number of retweets.

Criminal Face Identification System

July 2018 - July 2019

Gujarat Technological University, Ahmedabad, India

- Developed a Human Surveillance System that recognized faces in live video streams using OpenCV and Haar Cascade Classifier, and alerted authorized users in real-time if any of the faces matched criminal faces in the database.

Skills and Interests

Programming Language and Softwares: Python | R | SQL | Tableau | D3.js

Libraries / Packages: Numpy | Pandas | Scikit-learn | Matplotlib | NLTK | ggplot2 | tidyr | dplyr

Analytical Techniques: Regression Techniques | Classification Techniques | Statistical Analysis | Bayesian Analysis | Cross Validation