Shikhar Bharat Shah

Seattle, WA |sbshah96@uw.edu| shikhar96.github.io | linkedin.com/in/sbshah96 | 206-226-9442

EDUCATION

University of Washington **Seattle, WA**

Master of Science in Information Management (Data Science specialization) | CGPA: 3.87/4 **Expected: June 2021**

Coursework: Data Science - Theoretical Foundations, Machine Learning & Econometrics, Data Science - Scaling, Applications, and Ethics, Business Intelligence Systems, Advanced Relational Database Management Systems, Basic Educational Statistics.

Teaching Assistant: Data Visualization for Data Scientists, Social Media Data Mining and Analysis.

University of Mumbai **Mumbai, India**

Bachelor of Engineering in Computer Engineering | CGPA: 8.86/10 **June 2019**

Coursework: Analysis of Algorithms, Applied Mathematics, Artificial Intelligence, Data Structures, Database Management Systems, Data Warehouse & Mining, Distributed Databases, Machine Learning, Operations Research, Soft Computing, Software Engineering.

TECHNICAL SKILLS

Programming languages: Python, SQL, C, C++, Java, HTML, CSS, JavaScript, and PHP.
Tools and services: R, SAS, SSIS, Tableau, Power BI, AWS Lambda, AWS S3, AWS API Gateway, AWS SageMaker, WEKA, and Excel.
Libraries: Beautiful Soup, Matplotlib, NLTK, Numpy, Pandas, PyTorch, Scikit-learn, Scipy, Seaborn, StatsModels, and TensorFlow.

WORK EXPERIENCE

PATH, Seattle WA – Data Science Intern **Sept 2020 – Present**

- Building web crawlers using Beautiful Soup and Selenium to extract news articles about vaccines in three African countries.
- Analyzing data using Python and Tableau to identify trends in news coverage and social media reach of articles.
- Performing sentiment analysis and topic modeling on textual data to support the vaccine acceptance study.

Amazon Web Services, Seattle WA - Business Analyst Intern **June 2020 – Sept 2020**

- Developed centralized reporting solutions for internal Worldwide Public Sector (WWPS) teams using Tableau dashboards.
- Gathered data from various stakeholders using SQL Server and built dashboards satisfying customer requirements.
- Analyzed business processes and suggested areas of improvement relevant to business operations for WWPS.

Central Drug Research Institute, India - Data Analyst Intern **Sept 2017 - Feb 2018**

- Scraped news websites using Python to extract articles about AIDS in India to analyze public sentiment.
- Collected epidemiological data about Parkinson's disease in Africa through a literature review.
- Analyzed disease prevalence and incidence rates and studied disease spread with factors like age, gender, and race.

RESEARCH

Microsoft AI for Accessibility – Student Researcher **Jan 2021 – Present**

- Building a web application that centralizes information about inclusive employers to support the differently abled.
- Mining social media data and building web crawlers to retrieve information about corporate culture and inclusive policies.
- Applying NLP techniques like sentiment analysis and topic modeling on textual data to reveal hiring insights.

UW HCDE Directed Research Group – Analysis Team Lead **Sept 2020 – Dec 2020**

- Performed data mining to extract news articles about QAnon and analyzed news coverage on QAnon misinformation beat.
- Analyzed Twitter data using PostgreSQL, Python, and Tableau to investigate data requests from journalists.
- Led a team in analyzing online misinformation relevant to voter fraud, QAnon, COVID-19, and other controversial topics.

PROJECTS

DGA detection using XGBoost **Apr 2020 - June 2020**

- Built an application classifying DGA and benign domain names using AWS Lambda, S3, API Gateway, and SageMaker.
- Generated a dataset of 5 million rows using more than 40 domain generation algorithms and other data sources.
- Used XGBoost for classification and hyperparameter tuning to achieve a testing accuracy of about 93%.

Movie Review Classification using Naïve Bayes **Feb 2020 - Mar 2020**

- Developed a Naïve Bayes classification model to predict if a movie is 'Fresh' or 'Rotten' based on movie reviews.
- Used NLP techniques, smoothing, and cross-validation to achieve an accuracy of 63% on a dataset of 10,000 rows.

Twitter Topic Modeling using LDA **Jan 2020**

- Scraped Twitter data relevant to 'Data Science' and used Natural Language Processing techniques.
- Used Latent Dirichlet Allocation (LDA) model to extract the most naturally discussed topics about the field.

Stockopedia - A Stock Recommendation System **Mar 2018**

- Built a system using Python recommending stocks to buy and sell to traders as part of the NSE FutureTech Hackathon.
- Analyzed financial data and calculated stocks' short-term prediction, risk diversification, and sentiment score.

Recommendo - A Movie Recommendation System **Sept 2017 - Oct 2017**

- Built a web application that recommends movies to a user based on similar movies and user ratings.
- Applied collaborative filtering and k-NN method in Python and built the system UI using HTML, CSS, and JavaScript.

Others: Amazon Product Review Clustering, Furniture Sales Forecasting using ARIMA, Sales Data Warehousing & Analysis, Grocery Store Sales Analysis and Forecasting, Customer Churn Prediction, Titanic Survival Prediction, Speech to ISL converter.

CERTIFICATIONS

AWS Certified Cloud Practitioner **Jul 2020**