Shikhar Bharat Shah

Seattle, WA | sb.shah14@gmail.com | 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: 3.89/4 **June 2021**
Coursework: Data Science - Theoretical Foundations, Machine Learning & Econometrics, Data Science - Scaling, Applications, and Ethics, Business Intelligence Systems, Relational Database Management Systems, Machine Learning for Public Health Big Data.
Teaching Assistant: Data Visualization for Data Scientists, Social Media Data Analysis, Policy and Ethics in Information Management.

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, R, SQL, C, Java, HTML, CSS, and JavaScript.
Tools & Tech Stack: AWS, Hive, PostgreSQL, Spark, SSIS, SQLite, SQL Server, Power BI, Tableau, WEKA, Excel, Git.
Libraries: BeautifulSoup, Gensim, Keras, NLTK, Numpy, Pandas, PyTorch, Scikit-learn, Scipy, Seaborn, Selenium, TensorFlow, etc.

WORK EXPERIENCE

Microsoft – Capstone Student Researcher **Jan 2021 – Jun 2021**
- Developed a web application using Flask that centralizes information about inclusive employers to support the disabled.
- Mined social media data and built web crawlers to retrieve information about corporate culture and inclusive policies.

PATH, Seattle WA – Data Scientist Intern **Sept 2020 – Mar 2021**
- Developed web crawlers using Beautiful Soup and Selenium to extract news articles about vaccines in African countries.
- Analyzed data using Python and Tableau to identify trends in news coverage and social media reach of articles.
- Performed sentiment analysis and topic modeling on textual data to support the vaccine acceptance study.

Amazon Web Services (AWS), 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

Journalism Research through Collaboration DRG – 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.

Speech-to-ISL Translation System – Student Researcher, Author **Aug 2018 – Mar 2019**
Presented at ICCCIS 2021
- Built an application converting speech to Indian Sign Language using Android Studio, Google Speech API, Python, and Unity.
- Created a video motion database using Xbox Kinect and applied NLP techniques to match tokens with ISL motion gestures.

PROJECTS

DGA Domain Detection using XGBoost **Apr 2020 - June 2020**
- Developed 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.

Recommendo - A Movie Recommendation System **Sept 2017 - Oct 2017**
- Developed 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, Grocery Store Sales Analysis and Forecasting, Stock Recommendation System, etc.

CERTIFICATIONS

AWS Certified Cloud Practitioner **Jul 2020**
Tableau Data Scientist **May 2020**