# 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: 3.86**          **Expected: Jun 2021**

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

**University of Mumbai**                                                                                                        **Mumbai, India**

**Bachelor of Engineering (Computer Engineering) | CGPA: 8.86/10**                                              **Jun 2019**

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

## TECHNICAL SKILLS

**Programming languages:** Python, C, C++, Java, SQL, HTML, CSS, JavaScript, and PHP.

**Tools and services:** R, SAS, SSIS, SQL Server, Tableau, Power BI, AWS Lambda, AWS S3, AWS API Gateway, AWS SageMaker, MATLAB, WEKA, and Excel.

**Libraries:** Beautiful Soup, Caret, Dplyr, Ggplot2, Matplotlib, NLTK, Numpy, Pandas, PyTorch, Scikit-learn, Scipy, Seaborn, StatsModels, TensorFlow, and Tidyverse.

## WORK EXPERIENCE

**Amazon Web Services – Business Analyst Intern**                                                          **Jun 2020 – Sep 2020**
- Developing a centralized reporting solution for an internal Worldwide Public Sector team.
- Gathering requirements and data from various stakeholders and building sister dashboards using Tableau.
- Analyzing data relevant to headcount, time utilization, opportunities, and other business operations processes

**Symphony, India - Business Analyst Intern**                                                              **Dec 2017 - Mar 2018**
- Performed market research and analyzed survey data to reveal insights about user preferences.
- Developed a chatbot functionality using Firebase and DialogFlow that was integrated into the company application.
- Secured INR 100,000 in initial funding and won multiple startup competitions with the company.

**Central Drug Research Institute, India – Data Analyst Intern**                                           **Sep 2017 - Feb 2018**
- 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.
- Scraped news websites using Python to extract articles about AIDS in India to analyze public sentiment.

## PROJECTS

**DGA detection using XGBoost**                                                                             **Apr 2020 - Jun 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 Natural Language Processing techniques and bag-of-words approach to develop features for the model.
- Achieved an accuracy of about 63% on a dataset of 10,000 rows using smoothing and k-fold cross validation.

**Twitter Topic Modeling using LDA**                                                                                  **Jan 2020**
- Scraped Twitter data relevant to 'Data Science' and used Natural Language Processing techniques.
- Analyzed data to reveal insights about Twitter traffic over time and frequent terms used in discussing the topic.
- 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.
- Reached the final round of the hackathon which involved more than 165 participants.

**Recommendo - A Movie Recommendation System**                                                            **Sep 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 worked on the front-end of the web application.

**Other**: 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**