Aayush Vikram Shah

(206)-591-1763 | aayushah@uw.edu | in/aayush-shah19 | aayush1909.github.io | Seattle, WA

EDUCATION

University of Washington   **Seattle, WA**
Master of Science in Information Management (Specialization – Data Science) | GPA: 3.82|4   **Expected: Jun '21**

- **Coursework**: Foundations of Data Science, Machine Learning and Econometrics, Database Management Systems, Scaling and Applications in Data Science, Business Intelligence Systems, Basic Educational Statistics
- **Teaching Assistant**: Basics of Python Programming, Machine Learning Methods and AI, Introduction to Programming for Data Science, Foundations of Programming: Python, Python Programming Essentials

University of Mumbai   **Mumbai, India**
Bachelor of Technology in Computer Engineering | GPA: 3.7|4   **Aug '15 - Jun '19**

- **Coursework**: Data Structures, Big Data Analytics, Artificial Intelligence, Machine Learning, Image Processing, Data Warehousing and Mining, Neural Networks and Fuzzy Logic, Analysis of Algorithms, Object Oriented Programming, Applied Mathematics

SKILLS

Programming Languages: Python, R, Java
Tools & Tech Stack: AWS, SQL, Tableau, PyTorch, Tensorflow, Keras, Power BI, Hadoop, MS-Excel, Jira, Git

WORK EXPERIENCE

Software Engineering Intern **- TerraClear Inc**, Seattle, USA   **Jun '20 - Present**

- Interacting with cross-functional teams to gather requirements, analyzed business challenges to design the automation of the Ground Truth Data Pipeline process fed into the Neural Network Models.
- Targeting to reduce the manual efforts of the CV-AI team on data processing of RGB and Multispectral images by over **60%**

Data Analyst Intern **- Puratech Solutions**, Mumbai, India   **Jun '18 - Aug '18**

- Designed processes to collect consumer and demographic data of customers by scraping websites and storing them in databases
- Analyzed customer data using R to gain insights on customer behavior, increasing the customer retention rate by **9%**

Digital Analyst Intern **- Greymeter**, Mumbai, India   **Dec '15 - Jan '16**

- Re-engineered area-specific social media campaigns by analyzing student performance and collecting it in Excel
- Used Marketing Analytics, Google Analytics to increase the likes on social media platforms by **58%** in a month

PROJECTS

Data Scientist **- ML-Driven DNS Malware Detection System**, Seattle, USA (Github)   **Mar '20 - Jun '20**

- Developed and deployed RESTful APIs using AWS API Gateway to predict fully qualified domain names as DGA or Benign
- Implemented an XGBoost Classifier using AWS SageMaker on a custom **15 million** dataset uploaded to an S3 bucket
- Scaled the product to **1,000,000+** predictions per minute through AWS Lambda

Team Lead **- Grocery Retail Sales Analysis**, Seattle, USA (Github)   **May '20**

- Analyzed various store trends and created dashboards using Tableau to understand categories and items the store can focus on
- Predicted total revenue of a grocery store for the forthcoming months by applying FBProphet model on accrued time-series data
- Achieved the **2nd position** for the Best Machine Learning Model at the Premier University of Washington Hackathon

Data Scientist **- Fresh or Rotten?**, Seattle, USA (Github)   **Feb '20 - Mar '20**

- Mined reviews to classify them as fresh or rotten on the Rotten Tomatoes Movie Review data using Naïve Bayes algorithm
- Tuned the model's hyperparameters using K-Fold Cross Validation and used TF-IDF to attain an accuracy of **67.03%**

Data Analyst **- FIFA '20 Analysis**, Seattle, USA (Github)   **Jan '20**

- Analyzed the data to see the 20 most competitive clubs in the world and the best suited ones for young players to grow
- Applied Recursive Feature Elimination, Lasso and Ridge Regression to see the potential of players with an accuracy of **84%**
- Secured the **2nd runners-up position** at the Datathon held at the University of Washington

Data Analyst **– AccuWater**, Seattle, USA (Github)   **Oct '19**

- Built a prototype in 36 hours to determine the most prominent features causing algal growth in oceans using NASA's MODIS data
- Applied feature engineering and Random Forest to predict the algal growth with an accuracy of **90%** based on 11 features

Software Developer **- Stock Market Forecasting**, Mumbai, India (Github)   **Aug '18 - May '19**

- Designed a real-time investment portal on the Bombay Stock Exchange (BSE) data using Python
- Implemented a hybrid model of Neural Networks, Support Vector Regression and Sentiment Analysis with an accuracy of **95%**

Machine Learning Engineer **- Breaking CAPTCHA using CNN**, Mumbai, India (Github)   **Oct '17 - Nov '17**

- Tested the robustness of a CAPTCHA generator using CAPTCHA's drawn in a convoluted way
- Implemented CNN for alphabet/digit recognition which was able to break the CAPTCHA generator exposing security risks

Other Projects: App Rating Prediction (Python, Random Forest, Support Vector Machines, Linear Regression), Extracting main topics from a document (Python, Latent Dirichlet Allocation, NLP), Data Analysis on Titanic Data (R)