Amogh Singh

Clemson, SC | vsamogh12@gmail.com | 864-324-3962 | amoghsingh | amoghsingh1297

Education

Clemson University | Clemson, SC **January 2021 – December 2022**
Master of Computer Science (Major: Data Science) | **GPA:** 3.5/4.0
Relevant Courses: Applied Data Science, Statistics, Data Analysis, Deep Reinforcement Learning, Data Structures, Cloud Computing

University of Mumbai | Mumbai, IN **August 2015 – May 2021**
Bachelor of Engineering in Computer Engineering
Relevant Courses: Machine Learning, Algorithms Analysis, Object-Oriented Programming, Artificial Intelligence, Software Engineering

Skills

Programming Languages: Python, SQL, Scala, Java, C++
Packages & Frameworks: Tensorflow, Keras, PyTorch, scikit-learn, OpenCV, Flask, NumPy, Pandas, Matplotlib
Data Analysis: Data Wrangling, EDA, Classification, Regression, Time Series Analysis
Databases & Software: MySQL, MongoDB, PostgreSQL, MariaDB, S3, Apache Spark, Apache Beam, MapReduce, Hive
Cloud & Applications: AWS, Google Cloud (BigQuery, Dataflow, AI Platform), Microsoft Azure, GitHub, Docker, Tableau

Experience

Graduate Teaching Assistant | Clemson University | Clemson, SC ***June 2021 – December 2021***

- Delivered concepts on designing, development, debugging and **problem-solving** using JAVA, & implementation of algorithm design on ARM architecture & provided constructive feedback with an analytical approach on App development and Unit testing
- Conducted weekly lab sessions on software fundamentals & mentored, graded, reviewed assignments for 100+ students
- Drafted 40+ assignments while working as a TA over 7 months and managed lab records, documented all activities, and reported back to management | **ARM assembly, Bash script, Git, Java, Linux, Shell Script**

Data Science Intern | Airrchip Inc | Mumbai, IN ***June 2020 – November 2020***

- Led the assignment to build an **ETL pipeline** using integrated data on **BigQuery** & ingest batches for the reporting dashboard
- Devised a web-scraper model for data extraction with BeautifulSoup to help a real-estate firm with price-assessment prototype
- Drafted weekly reports and monthly presentations to showcase progress and maintain efficient communication

Database Development & Analysis Intern | Furus Packaging Pvt Ltd | Mumbai, IN ***June 2018 – December 2018***

- Maintained organization's database using **SQL server**, scripted refined SQL queries to wrangle data and aggregate dozens of data extracts containing retail data information on Enterprise software
- Prepared multiple complex datasets using SQL joins, performed intricate tasks of writing triggers, user-defined functions, and procedures attaining a faster query processing of the system by 40%
- Designed attractive & intuitive Power BI dashboards, visualizations & reports based on the business needs of the client

Projects

Taxi Cab Fare Prediction using Supervised Learning *(Python, scikit-learn, NumPy, Pandas, Matplotlib)*

- Cleaned data (**7 million rows**) by imputing the missing values using the transformation pipeline feature of sklearn
- Applied multitude of approaches like linear regression, ridge regression, kNN regression resulting best RMSE of $3.99 and **improved it to $3.23** using RandomForest Regression with GridSearchCV for tuning the hyperparameters
- Utilized batch learning method for creating these models that potentially fits a robust pipeline

NBA Player Salary Prediction using Regression *(R, Dplyr, Shiny, tidyverse, Markdown, Docker)*

- Composed an analysis strategy to predict sportsperson's salary in R with a **90% accuracy** & leverage NBA data on scoring patterns to devise a prediction model by drawing meaningful insights from time-series graphs, scatterplots, and bar charts
- Explored the patterns in variables to obtain the best model by cross-validation relying on inter-variable relationships
- Generated **Docker** image and containers for the Shiny dashboard to deploy it as a web app on the Google Cloud service

Movie Recommender System *(Python, SciPy, Surprise, Netflix)*

- Retrieved data using API calls for a movie recommender & feature engineered for ratings from the user behavior
- Built user-item based collaborative filtering using kNN regression, singular value decomposition for matrix factorization with the best estimation model fitted with a recommendation accuracy of **80%** and **RMSE of 0.96** on a test-data of 0.2
- Estimated movie ratings & projected results against Netflix's recommender in heat maps, scatter plots, and box plots

MeTube System- Online multimedia sharing and streaming platform *(Bootstrap, JavaScript, PHP, SQL)*

- Developed a Full-Stack multimedia platform to host audio, image, and video media using the XAMPP stack
- Enabled plethora of features like subscriptions, video downloads, recommendations, messenger, and views tracking
- Implemented myriad RDBMS database design techniques by SQL & server-side scripting using PHP, hosted it on college Cloud

Co-Curricular Activities & Achievements

- Conducted research on development of Explainable AI in the healthcare field to scrutinize its bias & decision-making capability
- **Organized & managed IPython workshop** in collaboration with Professors & students, to teach scientific Python packages