SAURABH SHARMA

github.com/saurabhsharma2401 | +1 864 207 9214 | saurabs@g.clemson.edu

---

EDUCATION

Clemson University
Clemson, SC
Masters of Science in Computer Science| GPA 3.66/4
May 2024
Relevant Coursework: Cloud computing, Parallel architecture, Data structures, Algorithms, Applied Deep Learning

K J Somaiya Institute of Engineering and Information Technology| CGPA 7.69/10
Mumbai, India
Bachelor's of Engineering in Information Technology
September 2020

SKILLS AND INTERESTS

Key Technical Skills: Python, AWS, Snowflake, ETL, Data Warehousing, Big Data, C++, SQL, Keras, CNN, LSTM
Certifications: AWS SAA CO3 Solutions Architect Associate, COF-C01 SnowPro Core, Az-900 Azure Fundamentals
Software: Control-M, Toad (Oracle SQL), Service Now, Snowflake Word, Excel, AWS Cloud Services, Tableau
Interests: CI/CD, Distributed Systems, Cloud Computing, Data warehousing, Deep Learning, AI, NLP, DevOps, Jenkins, T-SQL

WORK EXPERIENCE

Graduate Teaching Assistant
Clemson, South Carolina
School of Computing, Clemson University
August 2023 – Present
- Leading discussions among multiple cohorts of 30 students each in recitation classes about latest ethical dilemmas in Computer Science and Tech society as a whole and graded their reflection based on multiple ethical frameworks

Graduate Research Assistant
Clemson, South Carolina
ECHO Lab, Clemson University
February 2023 – August 2023
- Worked on a project that involved Mental Workload Calculation with EEG and Eye-tracking Data
- Developed a Data pipeline that receives two real-time streams simultaneously and uses multithreading and multiprocessing to preprocess data in a 30-second window and calculate mental workload using eeglab MATLAB library
- Helped the Human Factors department collect mental workload data from participants performing mock Robot-Assisted Laparoscopic Surgery by using the above software. This data will help them personalize rest periods for efficient training

Senior Data Engineer
Mumbai, India
Larsen and Toubro Infotech
August 2020 – August 2022
- Created and automated data pipelines using Control-M for moving all of the client's(A fortune 500 company) data to Snowflake cloud warehouse
- Fixed broken pipelines and ensured the data was moved to Snowflake as scheduled, increasing system efficiency by 50%.
- Collaborated seamlessly with the Business Intelligence team to optimize data pipelines and enhance dashboard functionality, ensuring accurate and actionable insights for informed decision-making.
- Attained SnowPro Core certification, showcasing proficiency in Snowflake cloud warehouse and AWS integration. Utilized complex Snowflake SQL and harnessed advanced features like time travel on multi-TB tables to effectively address customer queries.

Deep Learning Intern
Mumbai, India
Innokrit Inventions, RiiDL
June 2019 – August 2019
- Collaborated and built a HandWritten text recognition model from scratch with various deployment-specific restrictions, primarily using Python and Keras API
- Researched and implemented the then state-of-the-art CRNN LSTM model, which helped the model retain the memory and gave 7% better accuracy than the existing CNN model
- Attained final test accuracy of 95.67%, which resulted in the model being used by several colleges at Mumbai University

---

SELECTED PROJECTS

Goodreads Data ETL Pipeline
November 2023
- Created an ETL pipeline that downloads live data from the Goodreads website locally, then in a parallelized two-stage process, cleans the data and pushes it to Amazon S3 buckets.
- Used Apache Spark, Airflow and Amazon Redshift to increase concurrency thus speeding up ingestion by 67%.

Precision Medicine Using Genomic Data and Diseases Prediction   **May 2020**

- Predicted the likelihood of developing breast cancer based on genomic profile
- Used high dimensional data to predict that early prognosis gives up to 40 years to make corrections in lifestyle
- Engaged in data preparation, feature selection, classification, and comparison. Compared the performance among different machine learning algorithms and got a final accuracy of 97% with the Decision Tree model

PUBLICATIONS

"[Machine Learning Approach for Predicting Breast Cancer using Genomic Data](#)" - Published at ICAST in May 2020.