RIJUL H. SHERATHIA

rsherathia@ucsd.edu | San Diego, California | +1 (619) 953-7403 | LinkedIn | GitHub

EDUCATION

UNIVERSITY OF CALIFORNIA, SAN DIEGO *September 2022 - December 2023 (Expected)*

Master of Science in Data Science, GPA: 3.9/4.0

- Coursework: Statistical Models, Machine Learning, Deep Learning, Scalable Data Systems, Big Data Analytics & Applications, Advanced Data Mining & Visualization, Data Ethics & Fairness, Advanced Optimization, Fraud Analytics, Computer Vision.

MIT WORLD PEACE UNIVERSITY, INDIA *July 2017 - June 2021*

Bachelor of Science in Computer Science and Engineering, GPA: 9.21/10

- Coursework: Database Management, Data Science, Data Warehousing, Data Structures, Python Programming, Object Oriented Programming, Design & Analysis of Algorithms, NLP, Advanced Java, Software Modelling and Design, Business Intelligence.

TECHNICAL SKILLS

Programming Languages	: Python(Pandas, NumPy, Sklearn, PyTorch), SQL, R, MATLAB, Java, XML, Shell Scripting
Databases	: Amazon Redshift (PostgreSQL), MySQL, NoSQL, MongoDB, Neo4j, Redis
ML/DL Application	: Predictive Modeling, Classification, Clustering, Statistical Analysis, NN Architectures
Data Science Skills	: PySpark, MapReduce, Hypothesis Testing, Tableau, Power BI, Advanced Excel
Cloud & Scheduler Platforms	: AWS (RDS, Redshift, S3, Lambda, EC2), Erwin Modelling, SnapLogic, GitHub, SVN

WORK EXPERIENCE

NATIONAL CENTER FOR MICROSCOPY AND IMAGING RESEARCH (NCMIR), UC SAN DIEGO, CA

Research AI/ML Scientist *June 2023 – Present*

- Implementing cutting-edge segmentation techniques (**U-Net Attention, Mask R-CNN, FPN**) on EM brain cells, achieving 76% accuracy.
- Exploring novel unsupervised learning approaches to uncover brain cell patterns, resulting in valuable insights for **neuroscientific research**.

INSTITUTE FOR NETWORK MEDICINE (iNetMed), UC SAN DIEGO, CA

Research Data Scientist *June 2023 – Present*

- Applied **EDA** to unveil genomics patterns, boosting gene regulatory network prediction **accuracy by 15%,** enabling targeted treatments**.**
- Engineering automated pipeline to **generate tailored plots** from RNA-Seq **genomics data**, resulting in effective and quicker insights.

ZS ASSOCIATES, PUNE, INDIA

Business Technology Solution Associate *July 2021 – July 2022*

- Engaged with BMS healthcare client to develop cutting-edge **ML techniques** & **ETL pipelines**, resulting in revenue savings of **$1 million**.
- Liaised with 10+ teams to aid in strategic solutions and performance, streamlining various business challenges using SQL, Python, and AWS Services, yielding **20% sales improvements.**
- Worked on the enhancement of **AWS Redshift** to cut space utilization and enhance performance thereby **reducing cost by 15%**.
- **Mentored 4 interns** in ETL job enhancements and familiarized them with orchestration and standard ETL procedures.
- Orchestrated 50+ analytical/ETL jobs on SnapLogic, with custom features like **task retry** and **anomaly detection** based on duration.

Business Technology Solution Associate – Intern *November 2020 – June 2021*

- Developed generic **Python script** automating file movement from Email to AWS S3, **reducing 20+ hours** of weekly manual effort.
- Constructed weekly operational **Tableau** dashboard analyzing meal expenses on Redshift views, **reducing expense overhead by 25%.**
- Deployed **15+ ETL pipelines** with Data Architecture engaging AWS Services, SnapLogic, Python, and SQL for various data sources.
- Acquired understanding of the business **healthcare domain** with the ability to incorporate and translate into technical design.

RELEVANT PROJECTS

Image-to-Text Generation using Generative Transformer *July – August 2023*

- Developed Generative Transformer model employing **Attention & Resnet-50**, training Flickr8k images to generate descriptive, accurate text, achieving outstanding results with BLEU-1 scores of **65.07%** and BLEU-4 scores of **38.79%**.

American Sign Language Detection and Recognition *May - June 2023*

- Performed data cleaning and preprocessing techniques (**skin masking and canny edge detection**) to enhance model performance.
- Implemented classification and CNN models, achieving accuracy of **97%**, and successfully integrated real-time prediction capabilities.

Credit Card Fraud Detection *May – June 2023*

- Conducted extensive **data cleaning, variable creation, and feature selection** techniques to train and explore various models.
- Successfully implemented selected model that resulted in savings of $21 million/year and achieved an FDR@3% of **56.98%** for the project.