# RIJUL SHERATHIA

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

## EDUCATION

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

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

- Statistical Model, Machine and Deep Learning, Database Management, Fraud Analytics, Advanced Data Mining, Recommender System

**MIT WORLD PEACE UNIVERSITY, INDIA** **June 2021**

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

- Data Science, Data Structures, Business Analytics, OOP, Data Structures and algorithms, Machine Learning, AI, Software Engineering, Big Data Analytics, NLP, Operating Systems, Data Warehousing & Mining

## TECHNICAL SKILLS

| | |
|---|---|
| Programming Languages | : SQL, Python(Pandas, NumPy, Sklearn, PyTorch, Matplotlib), R, MATLAB, Java, Shell Scripting |
| Databases | : Amazon Redshift (PostgreSQL), MySQL, NoSQL, MongoDB, Neo4j, Redis, Cassandra |
| ML/DL Application | : Algorithms, Predictive Modeling, Classification, Clustering, CNNs, RNNs, LSTM, GANs, LLMs |
| Data Science Skills | : Statistical Analysis, MapReduce, A/B Testing, Hypothesis Testing, Spark, Kafka, Tableau, Power BI |
| Cloud & Scheduler Platforms | : AWS (RDS, Redshift, S3, Lambda, EC2, Glue, Athena), Erwin Modelling, SnapLogic, GitHub, SVN |

## WORK EXPERIENCE

**RESEARCH ML INTERN – NCMIR Lab, UC San Diego** **June 2023 - Present**

- Applying image segmentation (U-Net Attention, Mask R-CNN) on EM brain cells using TensorFlow, achieving 76% model accuracy.
- Implemented preprocessing techniques (Normalization, Noise Reduction), to improve the data quality, resulting in accurate models.
- Exploring unsupervised learning approaches to uncover brain cell patterns, resulting in insights for neuroscientific research.

**RESEARCH DATA SCIENCE INTERN – iNetMed Lab, UC San Diego** **June 2023 - Oct 2023**

- Utilized Python for exploratory data analysis (EDA) in genomics, unveiling patterns and enhancing gene network prediction by 15%.
- Engineered automated pipeline to process RNA-Seq data, generating customized plots, and facilitating informed decision-making.

**DATA ENGINEER / BUSINESS ANALYST – ZS Associates** **November 2020 - July 2022**

- Engaged with BMS healthcare client, leveraging advanced Python programming, SQL expertise, and data engineering skills to develop robust ETL data pipelines. Achieved significant cost savings of $1 million through streamlined processes.
- Revamped complex Python applications, optimizing intricate SQL queries, and implementing integration and QA/QC best practices. Elevated efficiency by 50% through data modeling, data warehousing, and ETL pipeline enhancements.
- Mentored a team of interns in ETL enhancements, data modeling, and orchestration, emphasizing Python and SQL skills.
- Orchestrated end-to-end CI/CD workflows, integrating Apache SVN and AWS services, facilitating rapid & reliable client delivery.
- Constructed an Interactive Tableau dashboard analyzing meal expenses on AWS Redshift to support data-driven decision-making, reducing expense overhead by 25%.

## RELEVANT PROJECTS

**Image-to-Text Generation using Transformer** **July 2023**

- Developed Generative Transformer model employing Multi-head Attention & Resnet-50 using PyTorch, training Flickr8k images to generate descriptive, accurate text, achieving outstanding results with BLEU-1 scores of 65.07% and BLEU-4 scores of 34.79%.

**YouTube Analysis Pipeline using AWS Services** **May 2023**

- Conducted YouTube analysis on AWS QuickSight by processing data with Python and publishing it to Kafka queue on AWS EC2
- Leveraged AWS Glue for ETL processing and AWS Athena for data querying and analysis, achieving response time under 2 sec.

**Credit Card Fraud Detection using Efficient Feature Engineering** **May 2023**

- Conducted data cleaning, variable creation, feature selection, and model exploration, reducing model training time by 40%.
- Successfully implemented Bagging and Boosting models that resulted in savings of $21 million and achieved an FDR@3% of 56.98%.

**Knowledge Graph Analysis for Company Domain Transition** **December 2023**

- Extracted JSON response from MongoDB's knowledge graph API, executing XML queries to obtain domain keywords.
- Utilized GraphQL(Neo4j) to identify competitors through data wrangling and NLP, combining patent keywords with API results.

## PUBLICATION / SEMINAR

**“Captcha for visually impaired people: A review” in IRJET** **May 2021**

- Proposed a new Captcha method for visually impaired people, performing the requested action by cursor tracking.
- Utilized NLP techniques to enable multiple languages audio conversion for communication across diverse linguistic backgrounds.

**“Threats of adversarial attacks on artificial intelligence” in MIT-WPU** **December 2020**

- Conducted seminar for 200+ students on adversarial attacks explaining the machine learning algorithms and techniques.
- Seminar also delivered research on preventive measures and defense against adversarial attacks on targeted predictive ML models.