Rahul Singh

+91 8210073691 srahulsingh7488@gmail.com LinkedIn GitHub Website

Education

Indian Institute of Technology Madras **CGPA: 8.11/10.0**
Bachelor of Science in Data Science Sep. 2022 – May 2026

DAV Public School, Khunti **Percentage: 96.4% (Class X) | 97% (Class XII)**
AISSE & AISSCE Apr. 2018 – May 2021

Experience

Sudha Gopalakrishnan Brain Centre, IIT Madras **Jul 2024 – August 2024**
Machine Learning Intern *Chennai, Tamil Nadu*

- Developed APIs for Brain Atlas database interaction and non-annotated image processing using Flask and Deep Learning, enabling seamless data retrieval and analysis.
- Built a Python-based system for SVG to GeoJSON conversion, verification, and visualization of brain regions, and conducted similarity analysis using ResNet-50 for advanced insights.

Projects

VisionRAD | *Python, Hugging Face, Unsloth, PyTorch, Transformers* | **GitHub** **Dec 2024 – Dec 2024**

- Developed a medical imaging AI system by fine-tuning Llama-3.2-11B vision and Qwen-VL-2B model with LoRA and 4-bit quantization, resulting in around 60% reduction in memory usage (from 30GB to 12GB) while maintaining 95% of the original model's accuracy in radiology image analysis.
- Leveraged Unsloth and Hugging Face libraries to optimize training on a 16GB T4 GPU, achieving 3x faster training speeds compared to traditional fine-tuning methods, despite memory constraints during model-saving.

Parallel Sorting Algorithms with CUDA | *Python, C, CUDA* | **GitHub** **Nov 2024 – Dec 2024**

- Implemented CUDA-accelerated merge and bitonic sort algorithms, resulting in 3.81ms sorting time for 50,000 elements with bitonic sort and 13.67ms with merge sort, showcasing efficient parallel processing on GPU hardware.
- Optimized GPU kernel performance using grid-stride loops and memory-efficient kernels, resulting in a 56x speedup compared to naive implementations and effective handling of non-power-of-two arrays through padding techniques.

GPT From Scratch | *Python, NumPy, PyTorch* | **GitHub** **Dec 2024 – Present**

- Developing a GPT model from scratch by implementing core components such as self-attention, multi-head attention, and transformer blocks, focusing on understanding the architecture and training dynamics of language models. Utilized Python and NumPy for efficient computation and model optimization.
- Built a custom autograd engine from scratch, enabling forward and backward propagation for neural networks, and applied it to binary classification tasks using a micrograd implementation, demonstrating foundational understanding of automatic differentiation and gradient-based optimization.

InfluConnect | *VueJS, Flask, SQLite, Redis, Celery* | **GitHub** **Sep 2024 – Nov 2024**

- Developed a full-stack responsive web app to connect sponsors and influencers, featuring an admin panel with CRUD operations for campaigns, ad requests, and profiles. Implemented JWT authentication and RBAC for secure access.
- Integrated Redis for caching and Celery for asynchronous tasks like email notifications, message reminders and generating monthly reports improving system responsiveness by 40% and reducing page load times by 30%.

InsightGenie | *Python, Pandas, Seaborn, Streamlit* | **GitHub** **Nov 2024 – Jan 2025**

- Developed a Python-based data analysis pipeline using Pandas, Seaborn, and Scikit-learn to automate dataset processing, visualization, and statistical analysis, including outlier detection, clustering, and trend analysis.
- Integrated LLM-powered narrative generation (GPT-4o mini) to create comprehensive, markdown-formatted reports with actionable insights, leveraging metadata from visualizations and statistical findings.

ScoreSage | *Python, Scikit-learn, GitHub Actions* | **GitHub** **Jan 2025 – Jan 2025**

- Developed end-to-end student performance prediction system using CatBoost and Flask, achieving 92% prediction accuracy through hyperparameter optimization and cross-validation, deployed to Azure App Service with CI/CD pipelines
- Engineered automated ML pipeline with feature engineering and model selection capabilities, reducing training time by 40% through parallel processing and GridSearchCV, resulting in 15% improvement in model performance metrics

Technical Skills

Languages: Python, C, Javascript, Java, Bash, SQL, HTML/CSS, Haskell
Frameworks: Scikit-learn, PyTorch, LangChain, HuggingFace VueJS, Flask, SQLite
Developer Tools: Git, GitHub, Linux, Arch, NeoVim, VS Code, PyCharm, IntelliJ, Eclipse
Libraries: pandas, NumPy, Matplotlib, Seaborn, SciPy, NLTK, XGBoost, Beautiful Soup

Achievements

- **Won 50,000 INR Cash Prize** in a GenAI Hackathon (sponsored by Lica.ai and Vidyo.ai) for developing an AI-powered Hiring Funnel that streamlined HR processes by automating candidate screening and interview scheduling.
- **Runner Up** at Hacksprint Hackathon (IIT Madras) for building EcoConnect, a web app promoting sustainability through community-driven initiatives like waste management, clean drives and eco-friendly business marketplace.
- **Individual Best Performer Award** at CrackTheCode Hackathon (sponsored by GeeksforGeeks and CodeChef) for creating a web app focused on Elderly Care, featuring health monitoring, medication reminders, and emergency alerts.
- **AIR 192** in Amazon ML Challenge among 5,000+ nationwide teams for developing an innovative machine learning solution using OCR and regex for text extraction from Amazon product images and NER for extracting key entities.