RAHUL RAWAT

+91-9115676924 ◇ Zirakpur, Punjab
rahulrawat272chd@gmail.com ◇ LinkedIn ◇ Github ◇ Portfolio Website

PROFESSIONAL SUMMARY

Data Scientist & AI Engineer with 1.5+ years of experience delivering end-to-end ML and AI projects using deep learning, NLP, and AWS. Skilled at building scalable models and intelligent systems that turn data into real-world impact.

EXPERIENCE

Team Leader (Data Science) **Jan 2024 – Aug 2025**
Lambton College *Mississauga, ON*

- Designed a FinTech AI system to extract insights from PDF financial reports, building a **LangChain + FAISS**-based RAG pipeline powered by **LLMs** for real-time document summarization and intelligent query answering.
- Fine-tuned a **BERT classifier** for span vs. arithmetic query detection, achieving **96.59% accuracy**, and deployed it via **AWS SageMaker** to reduce inference time from **2 minutes to 2 seconds**.
- Built a structured **MySQL ETL pipeline** to clean and standardize 1,800+ layoff records using staging tables, conditional joins, and deduplication techniques as part of a separate data cleanup initiative.

GitHub: FinChatbot, MySQL ETL pipeline

Data Science Intern **Jul 2021 – Aug 2021**
Diginique TechLabs *Delhi, India*

- Developed and fine-tuned multiple **regression models** to predict continuous outcomes achieving an average **R^2 score of 0.90**, outperforming baseline benchmarks by 5%.
- Enhanced model robustness by optimizing feature selection, automating preprocessing pipelines, and integrating results into a scalable evaluation framework.

PROJECTS

FinTech Chatbot – AI-Powered Stock Forecasting Platform (Live Demo):

- Built a deep learning-based **stock price forecasting model** using historical market data and technical analysis, achieving $R^2 > 0.85$ on volatile tickers like TSLA.
- Developed a **LangGraph-powered multi-agent system** integrating **LLMs**, **RAG architecture**, and **intent classification** to route queries to **forecasting, financial Q&A, or document retrieval agents**.
- Implemented smart document retrieval using **LangChain**, **FAISS vector stores**, and **embedding-based chunking** to extract financial insights from uploaded PDFs.
- Deployed an interactive **Streamlit dashboard** for rapid prototyping; currently migrating to a fully **React-based UI** with a **FastAPI backend** for production scalability and improved UX.

CaptionAI – Image Captioning using Encoder-Attention-Decoder Architecture(Project Link):

- Designed a deep learning-based **image captioning system** using a modular **Encoder–Attention–Decoder** architecture in PyTorch, trained on the **Flickr8k dataset**.
- Used a **pre-trained ResNet-50 CNN** as the encoder to extract visual features, followed by an **LSTM decoder** with an **attention mechanism** to generate context-aware captions.
- **Implemented modular code design** separating feature extraction, sequence generation, and attention logic for improved readability, debugging, and scalability.
- Integrated with a **Streamlit-based UI** allowing users to upload images and receive generated captions in real time; future plans include BLEU score evaluation and transformer-based upgrades.

SKILLS

Data Science & AI: Python (Pandas, NumPy, Scikit-learn), **Deep Learning** (PyTorch, TensorFlow), **LLMs**, **NLP**, **Embeddings**, **RAG**, **SQL/MySQL**, **MongoDB**.

Frameworks & Tools: LangGraph, **LangChain**, **FastAPI**, **Docker**, **Git**, **Power BI**.

Platforms & Infra: AWS (S3, EC2, SageMaker), **Vector Databases** (FAISS, Chroma, Pinecone), **Streamlit**, **Jupyter Notebook**.

EDUCATION

Postgraduate Certificate in Big Data Analytics, **Lambton College** **Jan 2024 – Aug 2025**
Coursework: Big Data Tools, Neural Networks and Deep Learning, Big Data Algorithms and Statistics, Big Data Visualization, Natural Language Processing and Social Media Analytics, NoSQL (MongoDB, Neo4j, Redis).

College Projects:

- **Pneumonia Detection CNN** (Project Link):
  - Built a multi-model pipeline (Logistic Regression, Random Forest, CNN, and VGG16 with Transfer Learning) to classify X-ray images, achieving **91.67% accuracy**.
  - Deployed with a **Streamlit UI** and modular codebase.
- **Sentiment Analysis using NLP** (Project Link):
  - Preprocessed noisy text data using **regex** and vectorized using **BoW and TF-IDF**.
  - Compared **3 ML models** on binary sentiment classification, with thorough evaluation of performance metrics.

Bachelor of Science in Mathematics, **Panjab University** **Jul 2019 – Aug 2022**
Achieved 77% overall with 85% in mathematics coursework.

Relevant Subjects: Linear Algebra, Calculus I & II, Probability Theory, Numerical Analysis, Differential Equations.

Applied Work:

- Built regression and classification models **from scratch** using concepts from **Linear Algebra and Calculus**, implemented entirely in **NumPy**.

CERTIFICATIONS

- Machine Learning Specialization: Verify the certification here.
- TensorFlow Developer: Verify the certification here.