Neeha Rathna Janjanam

Data Science and NLP Enthusiast | MS in CS Candidate at NYU

neeharathnaj@gmail.com | (929)-689-4863 | https://www.linkedin.com/in/neeha-rathna-janjanam | https://github.com/nrjanjanam

EDUCATION

New York University, Master of Science, Computer Science (Graduating in Dec 2024) Sep 2023
Relevant Coursework:Machine Learning, NLP, Deep Learning, Information Visualization, Big Data, Database Systems
SRM University, B.Tech., Computer Science and Engineering May 2020
Relevant Coursework: Data Science and Big Data Analytics, Artificial Intelligence, Machine Learning, Object-Oriented Programming

TECHNICAL SKILLS

Programming Languages/ Constructs:	SQL, Python, R, C++, PySpark, SparkSQL Hadoop (Map Reduce)
Visualization Tools:	Microsoft PowerBI, Tableau (basics)
Databases:	Oracle, MySQL, MongoDB, SAP, Kafka, MS SQL Database
Tools:	AWS, Azure, Microsoft Office Suite, Google Suite, Git, Kubernetes, R Studio, Databricks
Others:	CI/CD Pipelines,Terraform, Snowflake, NLTK, Numpy, Pandas, ScikitLearn, Pytorch

RELEVANT PROFESSIONAL EXPERIENCE

Global Intern (Cloud - Data Intelligence), S&P Global Market Intelligence Jun 2024 - Present
- Developed an interactive **Power BI dashboard** to optimize **CI/CD pipelines**, supporting **Azure DevOps** to **GitHub** migration.
- Automated data extraction and transformation using **Azure DevOps REST APIs**, providing insights on pipeline performance.
- Strategized the **productionisation architecture** ensuring **data source decoupling** and **more affordable**, manageable hardware.
- **Documented** the project, including future recommendations, **earning recognition** for improving CI/CD process transparency.

Manager (Data Analyst and Engineer), Jio Platforms Limited Feb 2022 - Aug 2023
- Collected, transformed and analyzed **large-scale datasets (structured & unstructured)**; created dashboards used by **100K+** employees.
- Collaborated on **data analytics architecture**, contributing to its implementation using Agile practices, mentorship to juniors and interns
- Researched, designed, and implemented **cloud-based (Azure) data analytics** pipelines and a **real-time ETL ingestion system**.
- Conducted **data quality assessments**, identified anomalies, ensuring data integrity and reliability.
- Built a **Python multithreaded ingestion pipeline**, reducing ingestion time by **80%**.
- Managed detailed requirement documentation with **architecture diagrams** and technical specifications

Data Analytics GET Intern, Reliance Industries Limited Jun 2019 - Aug 2019
- Executed end-to-end data science project on **4GB** spool data, involving EDA and collaboration with domain experts.
- Developed advanced visualizations based on business parameters, enhancing data comprehension and decision-making
- Presented my findings in problem-solving sessions, generating actionable insights and fostering a dynamic team environment.

SELECTED PROJECTS

Enhanced Contract Clause Retrieval using Divide-and-Conquer Strategy with LLMs (Ongoing) Dec 2024
- Developed an automated pipeline leveraging **GPT 4o** to revolutionize legal contract analysis, focusing on clause extraction and summarization using **Azure AI Search** and **ADA-002** embeddings
- Implemented a **divide-and-conquer strategy**, chunking dense legal documents and processing them in parallel to address LLM context window limitations, , ensuring no clause mention is missed, achieving **>99% recall** in clause extraction using optimized document chunking techniques to handle documents of varying formats, reducing manual review time from hours to minutes.

Supervised Fine-Tuning for Mathematical Answer Verification Nov 2024
- Fine-tuned **LLama3-8B** using **LoRA** and SFT, optimizing rank and scaling parameters, achieving **82.4%** accuracy (**+26.4%** over baseline)
- Implemented **TRL**-based training pipeline, integrating custom prompt templates, sampling methods, and hyperparameter tuning; implemented comprehensive logging system for hyperparameter tracking/ reproducibility, checkpointing and followed best practices
- Reduced GPU memory usage by **50%** using **PEFT**; conducted experiments to optimize model training and inference performance

Assessing Robustness to Irrelevant Context: The Adversarial Context Task-Completion Method for Question-Answering Systems May 2024
- Developed Adversarial Context Task-completion (ACT) method and implemented **Retrieval-Augmented Generation (RAG)** frameworks to test **LLM (GPT 3.5)** robustness in question-answering with misleading context, using modified **SQuAD dataset**
- Investigated impact of **different prompting techniques** (Zero-Shot, Few-Shot, Chain-of-Thought) on model performance across relevant, irrelevant, and no context scenarios.
- Demonstrated significant performance reduction with irrelevant context (32.3% accuracy) compared to no context (42.5%) and relevant context **(81.2%**), highlighting importance of **high-quality retrieval in RAG-based frameworks**

Extractive Summarization and Named Entity Recognition based Tabulation of Scholastic Text Oct 2020
- Developed a system and wrote a paper for Automatic Summarization and Text-to-Table Conversion for student note-making
- Evaluated the performance of **7** summarization approaches using ROUGE metrics on datasets of **varying** lengths.
- Introduced novel Tabulation approach using Named Entity Recognition with **0.74** F1 score on an untapped historical dataset
- Innovatively introduced a Table-Slot comparison technique to assess the quality of generated tables.

SELECTED CERTIFICATIONS

- AWS Certified Cloud Practitioner (CLF-C02)
- Machine Learning with Python-From Linear Models to Deep Learning (92%) by MITx

HONORS AND ACHIEVEMENTS

- *NYU Graduate Merit Scholarship Recipient 2023* by Tandon School of Engineering, New York University
- *Best Paper Presentation 2020 & Research Grant* by Research Center, SRM Institute of Science and Technology