

ARJUN KUMAR

arjunkumar@arizona.edu | [linkedIn - arjun-kumar-cse](#) | [github - Arjun-commit](#)

PROFESSIONAL SUMMARY

Data Scientist with expertise in Natural Language Processing, deep learning, and transformer architectures, skilled in conducting cutting-edge research and developing advanced machine learning models. Proficient in Python, R, PyTorch, and TensorFlow, with a demonstrated ability to leverage data for actionable insights and advanced research applications. Focused on language model innovations and skilled in publishing findings, I am eager to contribute to foundational advancements in LLMs and support their integration into production systems.

EDUCATION

University of Arizona, Tucson, AZ, USA

Master of Science in Data Science

4.0 GPA/4

August 2024 – Present

Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, MP, India

Bachelor of Technology in Computer Science and Engineering

3.5 GPA/4

August 2019 – May 2023

EXPERIENCE

Data Scientist Intern

Origins Tribe Pvt. Ltd.

August 2023 – January 2024

Internship

- Conducted NLP research and analysis to enhance user sentiment prediction from diverse online sources, supporting data-driven decisions for digital business optimization.
- Developed and deployed a sentiment analysis model using advanced NLP techniques and transformer-based models, reducing misclassification and improving prediction accuracy by 18%.
- Applied hypothesis testing and dataset refinement, resulting in a 50% improvement in model insights, contributing to increased engagement and customer satisfaction.

PROJECTS

GenAI Chat Agent | *Python, LangGraph, FastAPI*

November 2024 – January 2025

- GenAI Chatbot Agent: A chatbot is built using FastAPI, LangGraph, and LangChain, integrating LLMs like llama3-70b-8192 and mixtral-8x7b-32768 for dynamic model selection and interaction via a ReAct framework.
- Web Interface and Tools: The app includes a /chat API endpoint for processing messages with tools like TavilySearchResults and serves a web UI at /ui using Jinja2 templates.

Data Visualization and Analysis of Olympic data. | *Python, Scikit Learn*

August 2024 – December 2024

- Collaborated with a team to analyze and visualize patterns in an Olympic dataset, producing actionable insights that influenced sports strategy and training decisions. Utilized R and data visualization libraries (e.g., ggplot2, Shiny) to create compelling visualizations, effectively communicating findings to stakeholders.
- Implemented data cleaning and preprocessing techniques in R to ensure accuracy and reliability of visualized data.
- Created engaging presentations and documentation to showcase analytical results and foster discussions on data interpretation. Leveraged statistical analysis and P-testing to identify key trends and outliers within the Olympic dataset.

TECHNICAL SKILLS

Languages: Python, R, SQL, C/C++

Data Science and Machine Learning: Transformer architectures, Deep Learning, NLP, PyTorch, TensorFlow, Web scraping, Data wrangling, Feature engineering, Predictive modeling, Data visualization, ELT process, Statistics, A/B testing, P-testing, Experimental design, Machine learning algorithms (SVM, Random Forest, KNN, Decision Trees, Linear/Logistic Regression, K-means clustering)

Cloud Computing: IBM Cloud, AWS

Research & Development Skills: Hypothesis testing, Statistical analysis, Data exploration, Predictive modeling, Conference-ready research writing and publication

Data Visualization & Reporting: Power BI, Tableau, Microsoft Excel, R Markdown, Quarto, Django

Tools: Spark, Kubernetes, Docker, and Jenkins

Soft Skills: Effective Communication, Collaborative, Problem Solving, Critical Thinking, Adaptability in Dynamic Environments