

NAUSHIL KHAJANCHI

California | 323-620-5784 | nkhajanc@usc.edu | linkedin.com/in/naushilkhajanchi

EDUCATION

University of Southern California	Los Angeles, California
Master Of Science in Computer Science (3.7/4.0)	August 2023-May 2025
<ul style="list-style-type: none">Coursework: Machine Learning in Data Science, Natural Language Processing, Data Structures & Algorithms, Deep Learning, Web Technology, Information Retrieval	
Indus University	India
Bachelor Of Technology in Computer Science and Technology (3.9/4.0)	August 2019-May 2023

EXPERIENCE

USC Ming Hsieh Institute	United States
Research Assistant (Machine Learning Data Science)	Dec 2024-Present
<ul style="list-style-type: none">Refined MATLAB scripts and Python algorithms, slashing runtime for large-scale trajectory analysis by 35% and expediting ensemble modeling within the ACCIDDA project's infectious disease simulations.Implemented data pipelines transforming Parquet files into structured datasets for infectious disease modeling, reducing processing time by 75% and enabling statistical analysis of multi-variant scenarios.Migrated workflows from Jupyter to R-Markdown, improving compatibility with LEMMA servers and boosting script efficiency.	
WorkUp	United States
Machine Learning & Data Engineer Intern	May 2024-July 2024
<ul style="list-style-type: none">Applied advanced NLP techniques (NER, semantic matching, keyword extraction) to analyze job descriptions and candidate profiles, building an automated matching engine that improved placement rates by 30%.Developed ML models to forecast hiring trends with 85% accuracy, enabling recruiters to time outreach based on market demand patterns and optimize targeting.Integrated data framework for recruitment analytics, boosting signal clarity by 40% and potentially saving \$10K/year.	
Prompt Softech	India
Data Scientist Intern	December 2022-May 2023
<ul style="list-style-type: none">Developed ML models to forecast business metrics using time-series and regression, improving planning accuracy by 25%.Engineered pipelines to clean and unify data from APIs and databases, ensuring analytics-ready datasets.Refined dashboards and visual reports to monitor model performance and KPIs, streamlining stakeholder decision-making.	
Ambica Trading Co.	India
Data Scientist Intern	November 2020-February 2021
<ul style="list-style-type: none">Applied statistical analysis to identify KPIs, optimizing inventory strategy to reduce holding costs by 15%, saving \$20,000/yearBuilt customer segmentation models using clustering techniques, enabling targeted marketing that increased repeat purchases by 22% and added \$10,000 in quarterly revenue.Designed Power BI dashboards analyzing stock levels, trends, and forecasts, leading to a 15% reduction in excess inventory.	

ACADEMIC PROJECTS

Citation Generation with LLaMA Fine-Tuning | Python, Tensorflow, HuggingFace

- Fine-tuned **LLaMA3 with LoRA techniques** on 10,000+ AAAI papers to improve citation accuracy by **20%**.
- Integrated NLP-based textual cue recognition to optimize citation placement and enhance academic writing assistance.
- Evaluated performance with **BLEU (0.82) & ROUGE-L (0.87)**, surpassing baseline models by **20% in citation fluency & precision**.

Twitter Sentiment Analysis Pipeline | Python, FastAPI, Docker, NLP

- Created a real-time data processing pipeline analyzing **10,000+ tweets daily**, integrating NLP-based sentiment classification.
- Built a RESTful API with FastAPI, improving query response time by **5x** and handling **100+ concurrent requests**.
- Designed an **interactive dashboard** using Dash & Plotly, enhancing sentiment visualization accuracy by **30%**.

Financial Market Analysis Platform | MEAN Stack, Time Series Analysis, GCP

- Devised a stock price prediction platform using **ARIMA, LSTM, and Random Forest** models with **82% directional accuracy**
- Engineered feature extraction from multiple data sources including price data, economic indicators, and sentiment signals
- Created interactive visualizations for correlation analysis, risk assessment, and performance attribution using HighCharts.

TECHNICAL SKILLS

Programming Languages: Python, C++, R, SQL, MATLAB, Javascript, NoSQL
Machine Learning & AI: OpenCV, LangChain, TensorFlow, PyTorch, Scikit-Learn, Hugging Face, NLTK
Data Engineering & Pipelines: ETL, Data Modeling, Apache Airflow, DBT, Workflow Orchestration
Cloud & DevOps: Azure, AWS, Kubernetes, GitHub, CI/CD Pipeline, PostgreSQL, MongoDB, Google Cloud Platform, Snowflake
Tools & Platforms: Git, JIRA, Asana, Confluence, Power BI, Tableau, Streamlit, Docker