SANJITH KRISHNA VENKATESH KUMAR

Morgan Hill, CA | 765-767-2708 | sanjithkrishnaus@gmail.com | https://www.linkedin.com/in/sanjith-krishna | https://sanjith-v.github.io/portfolio/

EDUCATION

PURDUE UNIVERSITY — West Lafayette, IN
Master of Science in Business Analytics and Information Management — *Expected Aug 2025*
Teaching Assistant for MGMT 58600 Python Programming, Merit Scholarship Recipient

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE (BITS), PILANI — Hyderabad, India
Bachelor of Engineering in Chemical Engineering — *May 2024*
Coursework: Probability and Statistics, Advanced Statistical Methods, Numerical Methods in Chemical Engineering

EXPERIENCE

THERECORDXCHANGE — Remote,USA
AI Engineering Intern — *Jun 2025 – Present*
- Building a RAG-based AI compliance system for judicial hearings using AWS Bedrock, Claude, TensorFlow and a FAISS vector database to compare transcripts with legal benchmarks using Cursor.

KEARNEY STUDENT LAB (CAPSTONE PROJECT) — West Lafayette, IN
Data Science Consultant — *Jan 2025 – May 2025*
- Developed a predictive ensemble model (XGBoost + Random Forest) achieving 97.3% precision in identifying procurement transactions worth negotiating, enabling strategic vendor prioritization.
- Forecasted monthly procurement spend using Holt-Winters smoothing with a low Mean Absolute Percentage Error (MAPE) of just 8.46%, facilitating proactive financial planning for a global insurance client.
- Designed interactive dashboards and governance frameworks to centralize procurement analytics, driving actionable insights and enhancing decision-making transparency for the global insurance client.

FINSIRE — Chennai, India
Data Scientist — *Sep 2023 - Aug 2024*
- Developed and deployed a valuation model and an extensive database covering all vehicles sold in India (2013-2023), automating collateral value assessment processes and enabled instant collateral-based lending for a fintech platform.
- Built and optimized RestAPI + Flask endpoints to automate price updates for stocks and mutual funds, cutting data refresh time from 2 minutes to 2 seconds per call.
- Developed an NLP-driven bank statement analyzer using Word2Vec and BERT, reaching 65% transaction classification accuracy, approaching the 72% industry benchmark.

PROJECTS

Spotify song recommender — *Mar 2025 - Jun 2025*
- Clustered 1.2M Spotify tracks into listener-driven archetypes, revealing key shifts in music trends and guiding playlist curation.
- Built an XGBoost-based hit predictor (AUC=0.93), optimizing music marketing and discovery processes.
- Deployed a real-time Annoy + FastAPI recommender on Heroku, delivering scalable, instant song recommendations.

GenAI Instagram Caption Generator (Link to demo) — *Feb 2025 - Apr 2025*
- Built a caption generator using Hugging Face (BLIP) and OpenAI GPT-4 Vision, achieving low-latency responses.
- Deployed scalable FastAPI backend with Docker and Render/Heroku autoscaling, efficiently handling traffic surges.
- Optimized prompt engineering to improve caption relevance and quality, iteratively tuning model inputs based on user feedback.

Instacart Market Basket Analysis (SQL) — *Jan 2025 - Feb 2025*
- Engineered analytical SQL queries using advanced techniques (CTEs, window functions, aggregations) on 32M+ transaction records to quantify customer loyalty, churn risk, basket composition, and purchasing behaviors.
- Identified high-loyalty and churn-risk products, segmented customers by reorder habits, and derived actionable insights to inform targeted marketing strategies.

Data-Driven Capacitance Prediction Study (Link to Paper), *BITS Pilani* — *Sep 2022 - Dec 2023*
- Developed and fine-tuned machine learning models, including Bayesian Ridge Regression, K-Nearest Neighbors (R^2: 0.928, RMSE: 0.040), and Artificial Neural Networks (R^2: 0.893, RMSE: 0.049), to accurately predict specific capacitance in Ti3C2-based supercapacitors.

Optimal Scheduling and Model-Based Control of Chemical Processes, National University of Singapore — *Jun 2023 - Dec 2023*
- Designed a DLTI-based optimization model using Python and MATLAB, employing SLSQP algorithms to improve pulp plant operational efficiency during critical processes such as shutdowns and startups.

SKILLS

Programming & Databases: Python, SQL (MySQL, PostgreSQL), NOSQL (Firebase)
Python Libraries: Numpy, Pandas, scikit-learn, XGBoost, PyTorch, TensorFlow/Keras, BERT, Word2Vec, DistilBERT, Annoy, Pyomo, Hugging Face Transformers, OpenAI, Joblib, Selenium, Matplotlib, Seaborn.
Analytics & Experimentation: A/B Testing, Experimental Design, Causal Inference, Hypothesis Testing, Cohort & Retention Analysis
Cloud & Deployment: AWS (S3, EC2), GCP (BigQuery basics), Docker, Render, Heroku, Git
Visualization & Reporting: Tableau, Excel
Web & APIs: FastAPI, Flask, REST APIs, HTML/CSS/JS (basic front-end development)

LEADERSHIP AND ACHIEVEMENTS

Volunteered with the Digital Equity Foundation to bridge the digital divide, designing and delivering a tech curriculum to 1,500+ low-income students over 10 months; led a team of student volunteers at a partner school, ensuring all 98 students gained hands-on experience with G-suite tools.