

SANJITH KRISHNA VENKATESH KUMAR

West Lafayette, IN | 765-767-2708 | sanjithkrishnaus@gmail.com | <https://www.linkedin.com/in/sanjith-krishna> | <https://sanjith-v.github.io/portfolio/>

EDUCATION

PURDUE UNIVERSITY

Master of Science in Business Analytics and Information Management

Teaching Assistant for MGMT 58600 Python Programming, Merit Scholarship Recipient

West Lafayette, IN

Expected Aug 2025

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE (BITS), PILANI

Bachelor of Engineering in Chemical Engineering

Coursework: Probability and Statistics, Advanced Statistical Methods, Numerical Methods in Chemical Engineering

Hyderabad, India

May 2024

EXPERIENCE

PREDICTION GUARD

AI Engineering Intern

- Building a RAG-based AI compliance system for judicial hearings using AWS Bedrock, Claude, TensorFlow and a FAISS vector database to compare automatic speech recognition (ASR) transcripts with legal benchmarks using Cursor

Remote, USA

Jun 2025 – Present

KEARNEY STUDENT LAB (CAPSTONE PROJECT)

Data Science Consultant

- 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.

West Lafayette, IN

Jan 2025 – May 2025

FINSIRE

Data Scientist

- Developed and deployed a valuation model and an extensive database covering 16,000 cars and 7,000 motorbikes (2013-2023), automating collateral value assessment processes and enabled seamless 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.

Chennai, India

Sep 2023 - Aug 2024

PROJECTS

Spotify song recommender

- Clustered 1.2M Spotify tracks into listener-driven archetypes, revealing key shifts in music trends and guiding playlist curation.
- Built an Pytorch-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.

Mar 2025 - Jun 2025

GenAI Instagram Caption Generator

- Built a caption generator using Hugging Face (BLIP) and OpenAI GPT-4 Vision, achieving low-latency (<350ms) 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.

Feb 2025 - Apr 2025

Instacart Market Basket Analysis (SQL)

- 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.

Jan 2025 - Feb 2025

Data-Driven Capacitance Prediction Study ([Link to Paper](#)), BITS Pilani

- 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.

Sep 2022 - Dec 2023

Optimal Scheduling and Model-Based Control of Chemical Processes, National University of Singapore

- 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.

Jun 2023 - Dec 2023

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.