

KATTA SAI PRANAV REDDY

CONTACT

9347541040

kattapranavreddy@gmail.com

<https://www.linkedin.com/pranav-reddy-katta>

<https://github.com/ka1817>

EDUCATION

B-Tech - AI&ML
Anurag University, Hyderabad

CGPA - 8.13
2021-2025

Intermediate - MPC
Sri Chaitanya Junior College, Hyderabad

Percentage - 98
2019-2021

WORK EXPERIENCE

MACHINE LEARNING INTERN

iNeuron Intelligence Pvt. Ltd.

oct / 2024 - nov / 2024

- Performed customer segmentation using K-Means clustering on behavioral and demographic data to identify high-value customer groups, enabling data-driven marketing strategies.
- Built and deployed a classification model with Python and Scikit-learn to predict customer segments, improving targeting efficiency and personalization.

DATA SCIENCE INTERN

Unified Mentor Pvt. Ltd.

01-09-2024 to 01-10-2024

- Developed machine learning models in Python to predict employee attrition for Green Destinations, enabling proactive retention strategies.
- Presented data-driven insights to stakeholders through visualizations and actionable recommendations, aiding in attrition risk reduction

SKILLS

SQL Python HTML CSS Pandas Numpy

Matplotlib Seaborn Scikit-Learn Tensorflow

Keras NLP OpenCV Machine Learning C

CNN RNN Statistics Hugging Face

Langchain FastAPI Dockers

PROJECTS

1. AI Powered Product Intelligence Assistant for Big Basket

- Led end-to-end development of an AI-powered shopping assistant using Retrieval-Augmented Generation to enable context-aware product discovery through natural language and semantic search, delivering real-time, rating- and price-based recommendations.
- Built a scalable semantic search pipeline by embedding structured product data with thenlper/gte-small and indexing using FAISS; integrated CrossEncoder-based reranking to significantly improve retrieval precision for intent-driven queries within the retrieval framework.
- Implemented automated CI/CD pipelines using GitHub Actions and deployed the system on AWS EC2 via Docker, delivering a low-latency, production-grade architecture optimized for scalable, real-time inference and personalized product recommendations.

2. Telecom Customer Churn Prediction | End-to-End ML System with MLOps Principles

- Designed and implemented a scalable, production-ready ML system to predict telecom customer churn, integrating data preprocessing pipelines, model training with hyperparameter tuning, and real-time inference via a FastAPI backend.
- Containerized the application using Docker and orchestrated CI/CD workflows via GitHub Actions, enabling automated testing, model packaging, and seamless deployment to an AWS EC2 instance for scalable, production-grade service delivery.
- Achieved reproducible, maintainable MLOps through robust modularization, version control, and cloud deployment, empowering telecom providers with actionable churn insights and improving retention strategy responsiveness.

CERTIFICATIONS

- Certificate of Internship in Data Science - Unified Mentor
- Python for Data Science and Machine Learning
- Power BI
- Udemy - SQL
- Data Visualization with Python — Cognitive Class