

Kiran Punna

Github.com | linkedin.com | kiranpunna58@gmail.com | +91 9381911235

Summary

AI/ML Engineer skilled in building end-to-end machine learning solutions, including OCR pipelines, real-time detection systems, and model deployment with FastAPI/Streamlit. Strong foundation in Python, deep learning, and data engineering, with hands-on experience developing production-ready AI projects and personalized recommendation systems.

Work Experience

Infosys Springboard — AIML Intern (Nov 2025 – Present)

Built an end-to-end medical risk classification and diet recommendation system using OCR-extracted clinical data, ensemble ML models (XGBoost, Random Forest), and rule-based medical labeling, achieving ~99% accuracy and preparing the pipeline for scalable MLOps deployment.

Projects

WildVision — Real-Time Animal Detection

GitHub

A real-time animal detection system identifying 95+ animal species with 97% accuracy using CNN models. Built with PyTorch and Flask, it streams live camera input with a responsive interface for identification and classification.

Tech Stack: Python, PyTorch, Flask, OpenCV, HTML, CSS, Bootstrap, NumPy, Pandas.

AI Vision Sentinel — Real-Time Face Recognition

GitHub

A face recognition system that detects faces in real time using PyTorch and OpenCV. Integrates web scraping with BeautifulSoup to retrieve information about the recognized person from Wikipedia, providing an intelligent data display.

Tech Stack: Python, PyTorch, OpenCV, BeautifulSoup, Flask, HTML, CSS, JavaScript.

Multi-Prediction ML System

GitHub

A machine learning system with prediction models for diabetes, book recommendations, and digit recognition. Built with Python and Streamlit for an interactive, user-friendly interface.

Tech Stack: Python, Pandas, NumPy, Scikit-learn, PyTorch, Streamlit, Matplotlib, Seaborn.

Real-Time Animal Prediction with YOLOv5 + SORT

GitHub

Real-time object detection and counting system for animals using YOLOv5 with SORT tracking. Built with OpenCV, achieving 94% accuracy for multiple species detection in live video streams.

Tech Stack: Python, YOLOv5, SORT Tracking Algorithm, OpenCV, PyTorch, NumPy, Pandas.

SatCap — Satellite Image Captioning

GitHub

A real-time satellite image captioning system that generates accurate, descriptive captions for images using pretrained CLIP and GPT-2 models. Built with FastAPI and Flask, it allows users to upload images and receive instant captions through a responsive web interface.

Tech Stack: Python, PyTorch, Transformers (CLIP & GPT-2), FastAPI, Flask, OpenCV, HTML, CSS, JavaScript, NumPy, Pandas.

DocQGen — Document-Based Exam Q&A Generator

GitHub

An AI-powered system that converts uploaded documents into exam-ready short-answer questions using RAG, ensuring zero hallucinations and strictly document-grounded outputs. Designed for exam preparation, question paper automation, and AI-assisted learning.

Tech Stack: Python, LangChain, NVIDIA LLaMA 3.1, Pinecone, FastAPI, HTML, CSS, JavaScript

Education

B.Tech (CSE-AI), Annamacharya Institute of Technology & Sciences

2023 – Present | CGPA: 9.3

Skills

Programming Languages: Python, Java, SQL, JavaScript

Databases: MongoDB, MySQL, PostgreSQL

Vector Databases: Pinecone, Chroma

Frameworks & Libraries: PyTorch, TensorFlow, Scikit-learn, NumPy, Pandas, FastAPI, Flask, React, Streamlit

Tools: Git, GitHub, Postman, Power BI

ML & Data Skills: Machine Learning, Deep Learning (CNNs, RNNs, LSTMs, GRUs, Transformers), NLP (NLTK, BeautifulSoup), Data Visualization, Data Preprocessing

Generative AI Applications: Large Language Models (LLMs), LangChain (LLM Orchestration), Hugging Face Transformers, Prompt Engineering, Retrieval-Augmented Generation (RAG)

Achievements

- **ML Mania — National Level Tech Fest:** Secured **2nd place** at Project Explore for the **NoteCraft** project.
- **Touchstone Elocution Competition:** Winner of college-level elocution and essay writing competitions.
- **Young Trucks Competition — Naukri:** Secured **98.5 percentile** at the national level.
- **ATF-ALGO Scholarship Program — Nationwide:** Achieved **Top 7.8%** in coding rounds and applications.