Santosh Kanumuri

dkanumur@ttu.edu | 806-702-1121 | santoshkanumuri.tech
github.com/santoshkanumuri | linkedin.com/in/santoshkanumuri

Skills

Languages: Python, C, Java, JavaScript, TypeScript, NodeJS, PostgreSQL, MongoDB

Technologies & Tools: Git, Selenium, AWS(S3, EC2, RDS, Lambda, Fargate, ECS, LightSail), Heroku, Linux, ReactJS, NextJS, FastAPI, Django, Pinecone(Vector Database), Langchain, Pydantic, Ollama, OpenAI.

Education

Texas Tech University Aug 2023 - May 2025
Master of Science in Computer Science ***CGPA: 3.85/4.00***
Courses: Parallel Processing, Computer Security, Advanced Operating Systems, Neural Networks

Work Experience

Texas Tech University(Rawls College Of Business) May 2024 - Present

Graduate Assistant(Full Stack Developer)

- Engineered an art similarity search engine processing 35,000+ artworks with computer vision and vector embeddings (Pinecone), enabling price trend analysis with 92% accuracy and reducing search time by 70% while storing images in S3 and metadata in MongoDB.
- Built interactive art analytics platform (FastAPI backend) hosted on EC2 that performs real-time image comparison in under 3 seconds and generates market reports with 15+ data points when users upload images.
- Automated data collection from 3 auction houses using cloud-based cron jobs with email notifications, reducing manual processing by 80% and creating reliable ETL pipelines that process 1,000+ new artworks monthly.
- Developed an intelligent web search agent integrating Google, DuckDuckGo APIs with GPT for result summarization and citation generation, achieving 40% higher information accuracy compared to single-source searches, including a CSV mode that processes 500+ data points per hour and hosted on ECS using docker
- Created brand analysis system generating product similarity maps by applying NLP to 25,000+ Twitter posts, identifying 8 key market gaps and improving competitive analysis efficiency by 65%.
- Implemented personality analysis system using Google Gemini API to process 900+ LinkedIn profiles for talent analytics, reducing hiring mismatches by 35% and shortening recruitment cycles by 2 weeks.
- Consolidated fragmented financial data from Pitchbook into a centralized data warehouse, integrating 5+ data sources with 30,000+ records for comprehensive deals, investors, and company trend analysis, resulting in 45% faster market intelligence reports.
- Architected a local LLM(Llama3.2)powered innovation detection system that scraped and analyzed 1000+ company websites, classifying research domains with 88% accuracy and mapping innovation spectrums across industries, reducing manual classification time by 85% and enabling cross-sector innovation pattern identification.
- **Technical Skills:** Python, FastAPI, Flask, Streamlit, MongoDB, DynamoDB, AWS (S3, EC2, Lambda, Step Functions, LightSail), PyTorch, NLP, Computer Vision, Vector Databases, ETL

PricewaterhouseCoopers July 2022 – July 2023

Full Stack Developer

- Led development of CitiBank's internal document tracking system for loan documents as part of a 30-person team, reducing document retrieval time by 65% and improving loan processing efficiency by 40%.
- Engineered microservices architecture using Django backend with Amazon Aurora (PostgreSQL) database, processing 3,000+ document requests daily with 99.8% uptime.
- Developed responsive ReactJS front-end components with role-based access control for concurrent users, decreasing user training time by 3 weeks.
- Containerized application components with Docker for consistent deployment across development, testing, and production environments, reducing deployment errors by 85%.
- Built a centralized platform for task tracking, assignment details, and client feedback, streamlining project management.
- **Technologies:** Python, JavaScript, Django, ReactJS, REST API, PostgreSQL, Docker, AWS (EC2, S3, Lambda, Aurora), Git, Postman

Systema Automation Feb 2022 – Jun 2022

Python Developer Intern

- Designed and implemented a metrics visualization dashboard using Django and Chart.js, processing 5,000+ daily data points and reducing analysis time by 65%.
- Led a team of three developers using Agile methodologies, delivering all 12 sprint milestones on schedule with 94% test coverage.
- Integrated 5+ third-party REST APIs for real-time data collection, optimizing API calls to reduce latency by 70% and server load by 45%.
- Engineered responsive mobile-first UI with JavaScript and CSS Grid, increasing mobile user engagement by 40% and decreasing bounce rate by 25%.
- Implemented automated data validation processes that identified and corrected 98% of input errors before database insertion.
- **Technologies:** Python, Django, REST API, JavaScript, HTML5, CSS3, Chart.js, Git, PostgreSQL

Project Work

- **RaiderChat: AI-Powered Chatbot for Texas Tech** (Feb 2025): Developed an AI-powered chatbot using Retrieval-Augmented Generation (RAG) architecture to provide accurate answers based on Texas Tech college documents. Integrated a Large Language Model (LLM) with Chrome Vector Database for efficient document retrieval and utilized LangChain for seamless interaction, ensuring fast query responses and improved knowledge retrieval.
- **Alzheimer's MRI Classification with Vision-Language Models** (Nov 2024): Developed a model using vision-language models to match MRI images with text descriptions for diagnosing Alzheimer's stages: non-demented, very mild, mild, and moderate. Leveraged the Label Propagation for Zero-shot Classification with Vision-Language Models (ZLaP) technique.
- **YouTube Emotion Analyzer** (June 2024 – July 2024): Developed a Streamlit-based web application that searches YouTube for videos using keyword queries, extracts transcripts, and analyzes emotional trends within the content. Utilized the Gemini API for generating video summaries and extracting key insights. Implemented Natural Language Processing (NLP) techniques to classify emotions and visualize sentiment trends for various brand advertisements
- **CourseSync: Course Availability Tracker** (Sep 2024): Designed and developed a full-stack web application to monitor university course availability in real time. Built the frontend using React for a dynamic and responsive user experience, while the backend was developed with Flask to handle data processing and API interactions. Integrated web scraping techniques to fetch course availability data and implemented automated notifications for timely updates. Deployed the application on Heroku (backend) and Netlify (frontend) for seamless access.
- **ResultsRover – Telegram Bot for Automated Grade Retrieval** (Year): Developed a Python-powered Telegram bot that automates the extraction of student grades from Veltech University's results portal. Utilized Selenium for web scraping to fetch academic results, computed semester and cumulative GPAs, and provided detailed grade reports to students via email. Integrated with Telegram's API for a seamless user experience, allowing students to retrieve their results instantly. Hosted the bot on Heroku for reliable deployment and continuous availability.

Certifications

- **AWS** : Certified Cloud Architect.
- **MongoDB** : Certified Associate Developer
- **Google** : CyberSecurity Professional Certificate V2

Awards

- **Digital Acumen** – Awarded by PricewaterhouseCoopers India.