

Sai Prasanna Kumar Kumaru

San Jose, CA, United States | +16692509619 | saiprasannakumar.kumaru@sjsu.edu

<https://www.linkedin.com/in/sai-prasanna-kumar-kumaru/> | <https://github.com/sqiprasanna> | <https://saiprasannakumar-sjsu.vercel.app/>

Data scientist with 3+ years of experience in building Big Data intensive AI products, leveraging Python, R, SQL, PyTorch, and LLMs. Skilled in Data Warehousing, Kubernetes, AWS, adept at handling large datasets, streamlining ETL processes and Business Intelligence.

EDUCATION

San Jose State University, San Jose, CA

Master of Science in Artificial Intelligence

Aug 2022 - May 2024

CGPA: 4.0/4.0

Indian Institute of Information Technology, Chittoor, India

Aug 2016 - May 2020

Bachelors in Technology in Computer Science specializing Artificial Intelligence

CGPA: 3.79/4.0

SKILLS

Languages & Databases: Python, R, C/C++, Java, Go, SQL, PostgreSQL, Matlab, SAS, NoSQL, MySQL, SaaS, DynamoDB.

Machine Learning Libraries: Keras, PyTorch, Pandas, PySpark, NLTK, Scikit-learn, LLM, TensorFlow, NumPy, Regression.

Cloud Tools: Docker, Kubernetes, Apache Spark, Snowflake, AWS (Glue, Lambda, Redshift, Sagemaker), Apache Airflow, MLFlow, GCP (BigQuery), Azure, Kafka, Hadoop, Looker, Hive, MapReduce, Atlassian.

Frameworks & others: Flask, PowerBI, Tableau, Redis, Statistics, Clustering, Data Mining, Agile, Deep Learning, experimental design, A/B Testing, IoT, Problem Solving, Google Cloud, SAP, LangChain, Computer Vision, OpenCV, DevOps, Terraform, GPU.

WORK EXPERIENCE

Data Scientist, BobiHealth, San Jose CA

Aug 2023 – May 2024

- Developed an end-to-end NLP-driven chatbot leveraging Hugging Face Transformers, **Langchain**, **GPT-4**, RAG, RLHF, ChromaDB, and **FastAPI** boosted user engagement and efficiency by **25%** with **90%** retrieval accuracy.
- Designed AI model for pregnancy risk analysis (**90%** accuracy) using **Random Forest** and collected data for analysis from public health datasets. Engineered data pipelines and provided predictive analytics with **Azure Databricks**.
- Orchestrated **Airflow** workflows for model training and deployment; optimized **CI/CD** pipelines and python data modeling to **AWS S3**, increasing agility by **15%**. Incorporated GDPR principles through FMEA ensuring compliance and data protection.
- Utilized **R** and **dplyr** for advanced data analytics methodologies to analyze pregnancy risk datasets, expertly writing **large-scale ad-hoc SQL** queries and creating interactive visualizations and **KPI metrics**.

Data Scientist, ZS Associates India Pvt Ltd, Bengaluru, India

Jul 2020 - Jul 2022

- Led redesign of customer engagement strategy, implementing sophisticated Machine Learning models like **LSTM** and **XGBoost** increasing sales forecasting and drove a **\$20M revenue** while improving lead qualification by **15%**.
- Developed a Writer lookalike model with **97%** accuracy by combining Positive-Unlabeled learning with **XGBoost**, containerized in **Docker** for efficient scaling, and orchestrated using **Kubernetes**.
- Re-engineered Omni-channel Resource **Optimization** algorithm using PuLP, Simplex, Gurobi, CPLEX for strategic call allocation to physicians, resulting in increased **operational efficiency** by **3x** and provided KT to client's team for deployment.
- Engineered a **CNN-LSTM** and **BERT**-based system for detecting fake news, enhanced with **Genetic Algorithm** for feature optimization, resulting in **94.5%** F1 score, extracted insights using Shapley values (**SHAP**) for **explainable AI**.
- Collaborated with cross functional teams to merge **ETL** pipelines with Machine Learning solutions, leveraging Excel and Matplotlib for data visualization to communicate complex data, insights and adherence to SLA with key stakeholders.

Data Scientist, Accenture, Bengaluru, India

Jan 2020 - Jun 2020

- Developed a **recommendation system** employing Neural Collaborative Filtering, Autoencoders, and SVD + Neural Network, achieving optimal performance with an **18% drop** in RMSE and MAE. Implemented through a robust web API and elevated cloud operations using **AWS** services like S3, and EC2.

Data Science Intern, Ericsson, Chennai, India

Jul 2019 – Dec 2019

- Developed a scalable Big Data image search engine, leveraging Java, **PySpark**, and Kafka for data processing, integrated **Elasticsearch** with OpenAI's CLIP model via **MLflow** for efficient image retrieval, product analytics.
- Enhanced real-time querying capabilities through a **FastAPI** server and developed a dynamic web app using **ReactJS**; handling over 10 million images and supported processing up to 1,000 queries per minute.

RESEARCH WORK AND PROJECTS

Knowledge graph relation extraction with Biomedical Text ([GitHub](#)) (*Research Publication – IEEE BigDataService*)

- Improved biomedical relation extraction by incorporating Flask, **Langchain**, ChatGPT API, and Weaviate with a **QLoRA**-fine-tuned LLMs - **LLaMA-2**, **Gemma-7B** models, attaining an F1 score exceeding BioBERT's benchmarks by **21%**. Paper published in **BigData Service 2024**, The 10th IEEE International Conference on Big Data Computing Services and ML Applications.

ACHIEVEMENTS AND CERTIFICATIONS

- Hackathon:** HireMeAI – AI interview bot with resume tailoring, UC Berkeley AI Hackathon using LLMs. (**Top 5%**). ([link](#))
 - Tech Stack** - React, Node, Flask, LangChain, OpenAI's API, Whisper, Anthropic Claude 3, LMNT, Python, LLMs, Groq.
- Work:** Research Assistant and Teaching Assistant for Machine Learning Course under Professor Kai Kai Lui at **SJSU**.
- Certifications:** Google Certified Associate Cloud Engineer Certification (2020). ([link](#)).