KUNAPULI VENKATA SAI KIRAN

Applied AI ML Lead, Vice President at JPMorgan Chase, Hyderabad, India

kirankunapuli@duck.com LinkedIn GitHub HuggingFace Kaggle **M** Medium

PROFILE

Data Science Lead with **13 years in IT** including **10+ years in data science** and **5 years in technical leadership**. Led a **team of 4 data scientists** growing ads revenue; built foundation models, ranking, recommendation systems and AI agents.

PROFESSIONAL EXPERIENCE

JPMorgan Chase **Jul 2024 – Present**

Applied AI ML Lead (Vice President) *Hyderabad, India*

- Developed the **Chase Customer Foundational Model** as part of an org-wide initiative: **38M-parameter transformer backbone** with masked and next-event pre-training. Built training data from **52B+ card, deposit, digital and offer events** within a base of **95M+ customers**; trained on **24 H100 GPUs** (Ray, 3 nodes) to generate **512-d** customer embeddings for downstream risk modeling, fraud detection and **category selection and ranking.**
- Re-architected and developed the **category selection and ranking model**; evaluated **8 challenger models**, including SAINT and FT-Transformer, against contextual bandit LinUCB and deployed distributed inference on **Ray Serve**.
- Migrated and deployed the **Account Origination Model** on Databricks/MLflow for **49 products**: monthly retraining, champion/challenger releases; moved 95M-customer batch scoring from Spark/EMR to Ray Data, **3.7x faster**.
- Developed the **Marketing Automation Agent**, a LangGraph/Bedrock audience-selection agent translating marketing goals into catalog-grounded features, SQL and modeling plans through **multi-query RAG and parallel retrieval**.
- Developed a **customer-level email opt-out risk model** using contact frequency, engagement and dormancy signals to predict next-week unsubscribes and inform audience prioritization and contact-policy decisions.
- Fine-tuned **SLMs to detect paycheck and rent transactions** using a **577K-pattern silver dataset** labeled with **Ministral 3 8B**; evaluated on human-labeled holdouts to support downstream insight generation.
- Created an **AutoResearch** agent for LLM-driven experiment loops and co-built **Casper**, a multi-agent VS Code extension for Jira, Confluence and Bitbucket; both used by **ML and engineering teams across CCB Digital**.

Tokopedia **Dec 2021 – Jul 2024**

Senior Lead Data Scientist *Hyderabad, India*

- Led **4 data scientists** delivering TopAds **ranking** and **keyword recommendation** from research to production.
- Developed the **Real-time Ads Ranking** system with a scalable inference pipeline serving **1K requests/s** at **sub-80 ms** latency, lifting CTR **3.25%** and ads revenue **1.57%** (1.6B IDR monthly).
- Optimized inference of an existing deep learning model with post-training quantization, increasing revenue by **1.43%** (1.4B IDR monthly) and saving up to **166M IDR** in inference costs monthly.

ADP **Nov 2019 – Dec 2021**

Data Scientist *Hyderabad, India*

- Built multilingual **Job** and **PayCode Classification** models for ADP DataCloud using Language-agnostic BERT (LaBSE) and multilingual Universal Sentence Encoder (mUSE), cutting time-to-market from **6 to 2 months**.
- Deployed models on AWS (Step Functions, ECR, SageMaker, API Gateway) as batch transform pipelines and real-time endpoints, scaling model serving capacity **3x** to handle growing inference demand.

DBS Bank **Oct 2017 – Oct 2019**

Data Scientist *Hyderabad, India*

- Built a LightGBM **cheque-risk classifier**, reducing false negatives **18%** through class-imbalance handling.
- Built **time-series** cash-demand forecasts for ATMs and CRS, guiding replenishment and reducing shortages by **15%**.

Tech Mahindra **Nov 2013 – Oct 2017**

Software Engineer *Hyderabad, India*

- Delivered **4 projects for General Electric** from data ingestion (APIs, R, Talend) to ETL and predictive analytics.
- Engineered cash-collection models (clustering, decision trees) and ARIMA payroll forecasting for General Electric.

TECHNICAL SKILLS

GenAI & Agents: LLMs, SLMs, Fine-tuning (LoRA, PEFT), Agentic RAG, Context Engineering, Multi-agent Systems, MCP

ML & DL: RecSys & Ranking, NLP, Forecasting, Embeddings, Foundation Models, Quantization, A/B Testing

Languages & Frameworks: Python, R, SQL, PyTorch, TensorFlow, Hugging Face, scikit-learn, LangGraph, LangChain

Cloud & MLOps: AWS (Bedrock, SageMaker), GCP (Vertex AI), Databricks (MLflow), PySpark, Ray (Ray Serve)

EDUCATION

GITAM University, Hyderabad **2009 – 2013**

Bachelor of Technology (B.Tech) in Electronics and Communication Engineering *CGPA 8.4*

AWARDS & COURSES

2nd place, Continued Pre-training of LLM on Legal Contracts, Summer School on NLP & LLM, IIIT Hyderabad Jul 2024

Super Techie Award, DBS Bank: Pride Connect Award for Consumer Banking Analytics Dec 2018