Venkatalakshmi Kottapalli

Parsippany, NJ, 07054 | +12019544144 | venkatalakshmik23@gmail.com | LinkedIn | GitHub | Portfolio | Research

SUMMARY

Results-driven **AI/ML & Data Analyst** with **4+ years** of experience in **data analysis**, **data modeling**, **visualization**, and developing **AI/ML models**. Proven track record of delivering **40% ROI uplift** and transforming data into **actionable insights** with **95% accuracy**. Passionate about driving **business impact** through **data analytics**, **business intelligence**, **AI/ML**, and **data modeling**.

EDUCATION

Master of Science in **Artificial Intelligence**| Yeshiva University | New York, United States | GPA: 4.0/4.0 **Jan 2024 - Dec 2025**

Bachelor of Science in **Mathematics** | Adikavi Nannaya University | India| GPA: 3.8/4.0 (**University 1st Ranker**) **Jun 2015 - May 2018**

EXPERIENCE

AI/ML Engineer | Peblink, New York, USA **Sep 2025 - Present**

- Implemented AI-driven **optimization** models for **Goldman Sachs' $500M** Dallas campus initiative, identifying redundant management layers and improving operational efficiency.
- Built ensemble ranking system (XGBoost, Random Forest) to prioritize relocation candidates, reducing analysis time by **40%**.
- Developed **GPT-4** LangChain agents for relocation and retention scoring, automating 30+ workflows, reducing time by **25%**.
- Created interactive dashboards, heatmaps, and executive reports to visualize relocation scoring and cost-benefit analyses.

Machine Learning Co-op | ZSAnalytics, Boston, USA **May 2024 - Aug 2025**

- Developed **MLOps** pipelines on Azure ML and AKS, optimizing end-to-end model workflows for **95%** accuracy and **40%** ROI.
- Conducted **EDA** for data analysis and created 10+ features using feature engineering, boosting precision from **76%** to **88%**.
- Trained supervised ensemble machine learning models (XGBoost, LightGBM, Random Forest, SVM) with Optuna on 500K+ samples, achieving **88%** precision**, 0.91** AUC-ROC, **40%** accuracy improvement.
- Implemented **MLflow** and **CI/CD** pipelines with Azure DevOps and Git, automating and accelerating deployment by **30%**.
- Developed **RAG** system (LangChain, GPT-4, ChromaDB) indexing 100K+ descriptions with sentence-transformers, achieving key metrics Recall@K of **0.82**, **30%** engagement increase and reducing average query time from **6.2s** to **2.1s**.
- Presented analytical findings, results, and ROI improvements to stakeholders using dashboards, PowerPoint, visual reports.

Data Scientist | Cutso LLP, Hyderabad, India **Mar 2019 - Dec 2021**

- Performed **EDA** on large datasets (NumPy, Pandas, Matplotlib), executing quality checks improving model accuracy by **98%**.
- Conducted **statistical analysis** (Chi-Square, ANOVA) on 1M+ transactions, identifying 8 behavioral patterns for marketing.
- Implemented **clustering** (K-Means, DBSCAN, Hierarchical) into 5 groups (Silhouette: 0.68), improving retention by **18%**.
- Automated **KPIs dashboards** (Tableau), eliminating 50 hours monthly through real-time tracking of 30+ metrics.
- Delivered predictive analytics for inventory optimization, reducing stockouts by **25%**, improving profitability by **12%**.

RESEARCH & PROJECTS

Transformers (LLMs) Fine-tuning with DEKF | LLMs Optimization | Research **Oct 2025**

- Fine-tuned transformer-based **LLMs** using Decoupled Extended Kalman Filters (**DEKF**) for adaptive uncertainty estimation.
- Implemented in **JAX**, **NNX**, and **Dynamax** running on Google Cloud TPUs with XLA optimization, achieving **42%** lower compute, **38%** reduced memory, and **1.8×** faster convergence, cutting training time from **60 hrs** to **33 hrs** compared to LoRA.

AI Hackathon: Multi-modal Agent Conversational AI | Generative AI | GitHub **July 2025**

- Developed **GPT-4** powered multi-modal agent using **LangChain**, **LangGraph**, and hybrid **RAG** system (ChromaDB + SQLite)
- Handled 10K+ knowledge base documents, improving response accuracy by **35%** and securing **12th** place among **150+** teams.

Cardiomegaly Detection using Deep Learning| Computer Vision | Research Paper **May 2025**

- Developed **DeepCNN** model in **PyTorch** and **TensorFlow** for automated detection of cardiomegaly using chest X-ray images.
- Optimized with data augmentation, achieving **72%** accuracy, **30%** lower compute compared to **VGG16** about **100** epochs.

SKILLS

Programming Languages & Databases: Python, SQL, R, Cypher, PostgreSQL, ChromaDB, FAISS, Neo4j, SQL/NoSQL

AI/ML: Classification, Regression, Clustering, Time Series, CNN, RNN, GANs, Transformers, BERT, GPT, RAG, AI Agents, ETL/ELT

Frameworks: PyTorch, TensorFlow, Keras, Pandas, NumPy, Scikit-learn, Matplotlib, JAX, NLTK, LangChain, LangGraph, A/B Testing

MLOps & Cloud: MLflow, Docker, Kubernetes, Azure (ML Studio, AKS, ACR, DevOps), AWS (S3, RDS, IAM), GCP, CI/CD, Git

Data Visualization & Core: Power BI, Tableau, Advanced Excel, Matplotlib, Seaborn, OOP, Agile (Scrum), SDLC, REST APIs