

MURAPAKA HEMA KALYAN

Andhra Pradesh, India – 532001 | +91 8143994645 | mhemakalyan@gmail.com
[LinkedIn/in/hemakalyan](#) | [GitHub/KalyanM45](#) | [Medium/kalyan45](#)

PROFESSIONAL EXPERIENCE

ZoomRx Healthcare Technologies, Associate | Remote July 2024 – Present

- Extracted and analyzed structured/unstructured medical data from 20+ conferences, enabling real-time decision making using GCP-based pipelines.
- Optimized conference data processing scripts, improving computational efficiency by 30%.
- Developed prioritization algorithm, aligning research with client needs and reducing classification errors by 90%.
- Fine-tuned open-source LLMs (Llama-2) for entity extraction, improving accuracy to 95%.
- Employed OpenAI API with Multithreading to summarize the medical research and to extract medical entities.
- Analyzed user feedback on social data for client-specific drugs, increasing report accuracy by 30%

Hippo Cloud Technologies, Data Science Intern | Visakhapatnam Jun 2023 – Oct 2023

End to End Movie Recommendation Systems:

- Led a team of 7 interns, enhancing their skills in efficient modeling and effective data pre-processing techniques.
- Automated project lifecycles by creating data pipelines, reducing processing time by 50%.
- Conducted research and executed diverse end-to-end projects across different domains and use cases.

PROJECTS

OpenHealth: A Comprehensive AI Tool for HealthCare – [GitHub Link](#)

- Leading a team in developing an end-to-end SaaS application for remote healthcare with MLOps integration.
- Cleaned, normalized, and transformed 10K+ patient records, boosting model performance to 93%.
- Embedded Retrieval Augmented Generation Pipeline to communicate with Medical Reports.
- Deployed the model in AWS Elastic Beanstalk, Implementing CI/CD pipelines using AWS codepipeline.

Visa-Insight: End-to-End Visa Approval Prediction – [GitHub Link](#)

- Designed, scalable, modular pipelines using Gradient Boosting and Gridsearchcv, achieved 95% accuracy in visa prediction models
- Efficiently managed and optimized data retrieval from MongoDB server to ensure seamless integration.
- Used DVC for reproducible data versioning and ML pipeline tracking across experimentation stages.

TECHNICAL SKILLS

- **Programming:** Python, SQL, R
- **Machine Learning:** Scikit-learn, Langchain, OpenAI, XGBoost
- **Databases:** MySQL, MongoDB
- **Cloud Technologies:** AWS (EC2, S3, Elastic beanstalk), Google Cloud Platform (GCP), Azure Cloud
- **Libraries:** NumPy, Pandas, Scikit-learn, Matplotlib, Langchain
- **Big Data & MLOps:** DVC, MLFlow, Docker, Evidently, Jenkins, Kubernetes, Hadoop, PySpark, Airflow
- **Data Visualization Tools:** Microsoft Power BI, Tableau
- **Version Control:** Git, GitHub, GitLab
- **Office Tools:** Microsoft Excel

EDUCATION

- **B. Tech in Information Technology** | CGPA: 8.23, GMR Institute of Technology, Rajam 2020 – 2024
- **Intermediate** | CGPA: 9.4, Sri Chaitanya Junior College, Srikakulam 2018 – 2020
- **Class X** | CGPA: 9.7, Ravindra Bharathi School, Srikakulam 2017 – 2018

ACHIEVEMENTS

- Earned LinkedIn's Top Data Science Voice badge (Top 2%) for consistent thought leadership and community engagement.
- Recognized as the Best Data Science Content Writer in college for my open-source contributions.
- Delivered 10+ workshops and seminars on AIML topics, mentored 300+ students with a rating of 4.7/5.