

HEMANTH RAMESH

+1 (857).390.5572 | ramesh.he@northeastern.edu | [LinkedIn](#) | [Github](#) | [Portfolio](#)

2+ years experienced, result-driven **Data Scientist** expertise in creating **end-to-end** data-intrinsic applications, addressing complex architecture & optimizing business problems. Proficient in predictive modeling, data analysis & scripting languages.

PROFESSIONAL EXPERIENCE

Data Scientist | Celstream Technologies, Bangalore, India

Jan - Aug 2022

Tech-Stack: *Python, ML algorithms, Recommendation Engine, AWS services, Docker, Flask, Data Analysis, Linux, Jira*

- Centralized ML and AI tasks across AWS environments using **Dataiku DSS**, boosted data pre-processing efficiency by **25%**
- Engineered **end-to-end ML pipelines** in DSS, that achieved digital transformation for **Enterprise AI** customers saving operational costs by 30% and cut manual labor by over **100 hours** monthly through strategic process optimization
- Tackled application performance, log management issues in production through adept Linux scripting, testing & debugging
- Designed dashboards with 10+ KPIs metrics, enhancing ML model performance tracking and data insights deliver
- Implemented collaborative filtering system, resulting in a **40% revenue increase** via personalized product suggestions

Data Intelligence Analyst | Lifesight, Bangalore, India

Sept - Dec 2021

Tech-Stack: *Python, Apache Spark, Airflow, Google Cloud Services, Jenkins, Snowflake*

- Developed data-driven attribution models & analyze mobility, audience & location data, resulting in **\$200K+** rise in revenue
- Automated ETL/ELT workflows using cloud data platforms, handling over **1M+** data rows, reducing manual effort by **75%**
- Reduced POC delivery time by over 60% through effective coordination between Sales/CSM and Data Engineering team

Data Scientist Intern | R3 Media Labs, Kanpur, India

Jan - Aug 2021

Tech-Stack: *Python, Regression models, MS SQL Server, MongoDB, Crazy Egg*

- Developed **XG Boost Model** for Customer Churn boosting customer retention of mooKIT Platform by **15%**
- Strategically analyzed 100GB of structured & unstructured click-stream data to identify user engagement strategies
- Enhanced platform revenue by **\$10,000** and user engagement by **12%** through strategic Python-based **A/B tests**

Graduate Student Mentor and Vice-Captain of Cricket Academy | Northeastern University

Present 2024

- Guided **30+** graduate students to academic success and led the team to victory in **Nationals** (Texas) and **Regionals** (MA)

KEY PROJECTS

LLM: Medical Chatbot | cTansformer, Langchain, PyPDF, Pinecone, sentence_transformer

Jan 2024

- Developed chatbot utilizing **LangChain** and **LLaMA 2** for intelligent content extraction from encyclopedia medical PDF book
- Employed Hugging Face 'all-MiniLM-L6-v2' and Pinecone's vector database for semantic text analysis and dynamic querying

NLP: Generating and scoring toxic comments | NLTK, Transformers

Aug 2023

- Orchestrated dual-model strategy, using **BERT** and **miniGPT**, achieving 85.44% accuracy and 55% reduction in model loss
- Conducted assessments of 14 unique models, achieving categorical cross-entropy loss values within 0.27 to 0.35

Machine Learning: End-to-End Regression model | Regressors, Flask, Azure, AWS, Docker, CI/CD, Git Actions

May 2023

- Engineered **production-grade ML model** employing column transfer pipeline, with 95.3% accuracy by R2 score comparison
- Launched application on AWS Beanstalk, Azure Web App, by **CI/CD** pipelines via **GitHub Actions** with container registries

MLOps: Credit Card Fault Detection | DVC, MLflow, Airflow, GCP

Nov 2023

- Executed real-time MLOps system, integrating **Airflow**-managed data pipelines with **DVC** and **GCP** storage
- Administered model tracking with **MLflow**, ELK for logging, Flask API for user interaction and automated retraining

Deep learning: DeepFake Detection | Pytorch, CNN, GANs

Apr 2023

- Leveraged MesoNet4 architecture using inverted residual blocks & **attention** mechanism on **150K** high-resolution images
- Fostered global unstructured pruning at 20% sparsity to enhance efficiency, achieving impressive **97%** accuracy rate

TECHNICAL SKILLS

Languages/Frameworks: Python (Pandas, NumPy, Scikit-learn, Scipy, Pytorch, TensorFlow, Keras, SpaCy, NLTK, Flask), SQL

Data Engineering: Apache Spark, MySQL, Hive, SQL Server, NoSQL, Snowflake, Airflow, Kafka, RESTful APIs

Technologies: ML Flow/DVC, AWS, GCP, Docker, Git, Tableau, Excel, Kubernetes, Linux, Power BI, JIRA, Warehousing

Modeling: Clustering, Classification, Regression, Recommendation, PCA, Time Series, Causal Analysis, Multi-Model, A/B tests

EDUCATION

Northeastern University (Boston, MA) | Master of Science in Data Science

April 2024

Courses: AI Metrics and Losses, Cloud architectures, NLP, Database Design, Data Mining, Neural Networks, ST: LLMs

PES University (Bangalore, India) | Bachelor of Engineering in Electronics and Communication