

# HEMANTH RAMESH

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

Result-driven **Data Scientist** with expertise in creating data-intensive applications, addressing complex architectural and scalability challenges in diverse sectors. Proficient in predictive modeling, data processing, data mining algorithms, and scripting languages.

## PROFESSIONAL EXPERIENCE

**Data Scientist** | Celstream Technologies, Bangalore, India

Jan - Aug 2022

- Scalably deployed Dataiku DSS across **3 AWS** environments, empowering ML/AI **teams** and boosting task efficiency
- Implemented collaborative filtering system, resulting in a **40% revenue increase** via personalized product suggestions
- Optimized **Machine Learning workflows** DSS with **Python** and **Docker**, reducing manual work by over **100 hours** monthly
- Improved backend NLP algorithms for contract analysis and invoice matching, achieving a **9%** increase in operational efficiency
- Enhanced AWS EC2 performance by **15%** and decreased manual effort by **40%** using targeted Linux scripting
- Designed dashboards with **10+ KPIs**, delivering vital data insights to engineers for enhanced decision-making

**Data Intelligence Analyst Intern** | Lifesight, Bangalore, India

Sept - Dec 2021

- Collaborated with **7+ engineers** on PySpark and GCP to build attribution models for mobility, audience, and location data
- Analyzed **1TB** of customer data in predictive study, leading to **24%** revenue uplift through data-driven insight and optimization
- Optimized cloud **ETL/ELT** pipelines for **100M+** data rows, using **SQL** and **Snowflake** for processing and analysis
- Reduced POC delivery time by over **60%** through effective coordination between Sales/CSM and Data Engineering teams

## RESEARCH AND PROJECTS

**LLM - Medical Chatbot using LLaMA 2** | cTansformer, Langchain, PyPDF, Pinecone, Sentence\_transformer [\[Link\]](#)

Jan 2024

- Spearheaded the development of chatbot using **LangChain** and **LLaMA 2**, extracting content from encyclopedia PDF book
- Employed **Hugging Face's** "all-MiniLM-L6-v2" model to embed text chunks, creating semantic representations for text analysis
- Managed **Pinecone's** vector database for semantic indexing and searches, utilized RetrievalQA chain for dynamic query

**NLP - Generating and scoring toxic comments, NER Task** | NLTK, Transformers [\[Link\]](#)

Oct 2023

- Orchestrated dual-model strategy, using **BERT** and **miniGPT**, achieving 85.44% accuracy and 55% reduction in model loss
- Implemented robust k-fold cross-validation (**k=14**) methodology and pioneered **NER system** using transfer learning [\[Link\]](#)
- Conducted assessments of **14 unique models**, achieving precise categorical cross-entropy loss values within 0.27 to 0.35

**Machine Learning End-to-End Student Performance Regression model** | Regressors, Azure, AWS, Dockers [\[Link\]](#)

Sept 2023

- Engineered **production-grade ML model** employing column transfer pipeline, with 95.3% accuracy by R2 score comparison
- Developed and implemented a **Flask** framework, encapsulating it into a **Docker image** for deployment
- Launched application on AWS Beanstalk and Azure Web App, by **CI/CD** pipelines via **GitHub Actions** with container registries

**MLOps - Credit Card Fault Detection** | MLOps, DVC, MLflow, Airflow, GCP [\[Link\]](#)

Sept 2023

- Executed real-time MLOps system, integrating Airflow-managed data pipelines with **DVC** and **GCP** storage
- Maximized CI/CD workflows using GitHub and conducted rigorous testing for Apache **Airflow DAGs**
- Administered model tracking with **MLflow**, ELK for logging, Flask API for user interaction and automated retraining

**Deep learning Research Assistant - DeepFake Detection : Northeastern University** | Pytorch, CNN, GANs [\[Link\]](#)

Jul 2023

- Generated **deep learning model** using the MesoNet-4 architecture to identify deepfake images on custom dataset
- Employed inverted residual blocks and attention mechanism on 150K high-resolution images generated by **StyleGAN**
- Fostered global unstructured pruning at 20% sparsity to enhance efficiency, achieving impressive 97% accuracy rate

## TECHNICAL SKILLS

**Languages:** Python (Pandas, NumPy, Scikit-learn, OpenCV, Pytorch, TensorFlow, Keras, Transformers, NLTK, Flask), R, Matlab

**Data Engineering:** ETL, Alteryx, Apache Spark, MySQL, NoSQL, Snowflake, Airflow, Kafka, REST APIs

**Tools & Technologies:** MLOps, Azure, AWS, GCP, Docker, Terraform, Git, Tableau, Excel, Kubernetes, Agile Methodology

**Modeling:** Clustering, Classification, Regression, Recommendation, PCA, Time Series, A/B Testing, NLP, Causal Analysis, LLMs

## EDUCATION and EXTRAS

**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

**Graduate Student Mentor and Vice-Captain of Cricket Academy**

Mar – Dec 2023

- Supervised **30+** graduate students to academic and professional success via workshops, mentorship, and sharp advice
- Led the team to **Nationals victory** (Texas) and **Regionals** (Northeast), inspiring leadership and fostering winning collaboration

**PES University** (Bangalore, India) | Bachelors in Electronics and Communication

Aug 2022