# SREE CHARAN REDDY KAILASAM

+14432859885 | sreechk1@umbc.edu | LinkedIn | GitHub


## SUMMARY


Fluent in the language of messy data, I specialize in transforming complex and unstructured information into intelligent systems that deliver measurable impact, not just insights. With experience fine-tuning Large Language Models (LLMs), developing predictive models, and deploying end-to-end data pipelines using Python, SQL, Spark, Airflow, and AWS, I build solutions that extract value from data and support scalable and automated decision-making across industries such as healthcare and finance.


## EDUCATION

**University of Maryland, Baltimore County, MD** Jan 2024 - PRESENT
Master of Science, Data Science (CGPA: 3.822/4)
Coursework**:** Natural Language Processing, Data Management, Platforms for Big Data Processing, Data Analysis and Machine Learning

**Vellore Institute of Technology, Vellore** Jul 2019- May 2023
Bachelor of Engineering, Computer Science and Engineering with specialization in Business systems (CGPA: 8.46/10)
Coursework: Artificial Intelligence, Data Science and Statistical Modelling, Probability and Statistics, Computer networks, Cloud Microservices and Applications, Advanced social, media and text analysis, Linear Algebra

## PROFESSIONAL EXPERIENCE

**Data Analyst I**

**Progment Software Technologies, Hyderabad, India** Nov 2022 - Dec 2023

- Engineered scalable ETL frameworks using **Python**, **SQL**, and **Apache Airflow** to extract data from APIs and relational databases, reducing manual processing by **40%** and enabling reusable data flows across departments.
- Designed and automated data ingestion pipelines with **AWS Lambda**, reducing data latency by **30%** and supporting real-time analytics through a cloud-based data lake architecture.
- Built interactive dashboards using **Power BI for executive reporting** and **Plotly Dash for custom analytics apps**, enabling real-time, audience-specific insights across financial data pipelines.
- Integrated **S3**, **Redshift**, and **Glue** for data lake architecture and near real-time analytics.
- Automated anomaly detection in financial data using **Time Series Forecasting (FBProphet, ARIMA)**.

**Machine Learning Research Assistant**

**Vellore Institute Of Technology, Vellore, India** Aug 2021- Oct 2022

- **Researched and developed multi-class image classification models** using **EfficientNet-V2**, focusing on optimizing performance for medical and satellite imaging datasets.
- Achieved a **12% increase in model accuracy** by implementing advanced data augmentation techniques, regularization strategies, and hyperparameter tuning.
- Managed large-scale data preprocessing using **PySpark**, and trained models with **W&B (Weights & Biases)** for experiment tracking.
- Used **Apache Kafka** for streaming ingest and integrated with **Delta Lake** to support scalable model training pipelines.

## PROJECTS

**AI-Powered Medical Assistant using NLP and Transformer Models**

- Developed a **production-ready AI healthcare assistant** that processed over **2,000 clinical transcripts**, using **ClinicalBERT** for symptom classification and **BART** for summarizing medical notes, enabling faster initial triage and response.
- Fine-tuned both models on **15,000+ samples** from **MTSamples** and **PubMedQA**, achieving **72% accuracy** in specialty prediction and improving model relevance by **25%** over baseline transformer models.
- Designed an optimized data preprocessing pipeline and deployed the end-to-end AI assistant via a **Flask web app**, achieving **30% noise reduction**, faster **LLM convergence**, and real-time symptom analysis with **<1.2s latency** using **transformer-based models**.

**Predicting Energy Consumption Patterns in Commercial Buildings Using Historical Energy Data**

- Built and deployed predictive models using **XGBoost** and **LightGBM** in **Python** to analyze historical energy usage and climate data across commercial buildings, contributing to a **10% reduction in operational costs** through optimized resource planning.
- Integrated data from external weather APIs and internal building management systems, and performed **feature engineering** to capture seasonal and occupancy-based consumption patterns.
- Enhanced model accuracy and robustness using **GridSearchCV**, **TimeSeriesSplit**, and **cross-validation**, resulting in a **15% improvement in prediction performance** over baseline models.

**Customer Segmentation using Apache Spark**

- Developed a customer segmentation pipeline on a dataset of **250,000+ records** using **Apache Spark** and **PySpark**, applying **RFM analysis** and **K-Means clustering** to identify and profile high-value customer segments.
- Improved cluster separation by **22%** using **Silhouette Score** optimization, enabling clear distinction between active, dormant, and high-potential users for targeted engagement.
- Implemented data preprocessing and feature scaling at scale using **Spark MLlib**, reducing end-to-end processing time by **40%** compared to traditional batch methods.
- Delivered actionable insights that supported personalized marketing strategies, resulting in a **15% increase in campaign efficiency** and informed the design of loyalty and retention programs.

## SKILLS

**Languages**: Python, SQL, R, Java, C++, Bash, JavaScript

**ML & AI**: LLMs, NLP (BERT, ClinicalBERT, BART), Transformers, TensorFlow, Keras, PyTorch, ARIMA, FBProphet, GridSearchCV, Cross-Validation

**Big Data**: Apache Spark, PySpark, Hadoop, Databricks, Kafka, Delta Lake

**LLM Tooling**: Hugging Face Transformers, LangChain, OpenAI API

**Pipelines & Cloud**: Apache Airflow, MLflow, AWS (Lambda, S3, Glue, Redshift), Azure, Docker, REST APIs

**Visualization**: Power BI, Tableau, Plotly Dash, Matplotlib, Seaborn

**Web & Tools**: Flask, Django, HTML, CSS, Bootstrap, Angular, Git, Jupyter, VS Code, Power Automate, Power Apps, Azure Data Studio, W&B

**Databases**: MySQL, PostgreSQL, MongoDB

**Soft Skills**: Communication, Analytical Thinking, Collaboration, Problem Solving

## CERTIFICATIONS

**Google:** Foundations of Data Science

**IBM**: Tools for Data Science, Python for Data Science, AI & Development

**Coursera:** Databases & SQL for Data Science with Python

**Databricks Academy:** Apache Spark with Databricks

**Udemy:** The Web Developer Bootcamp