**MEGHNA REDDI**

973.906.9462 | [mreddi2001@gmail.com](mailto:mreddi2001@gmail.com) | [/in/meghnareddi](https://www.linkedin.com/in/meghnareddi/) | [Github](https://github.com/meghnareddi) | Jersey City, NJ

**EDUCATION**

New Jersey Institute of Technology, Newark, NJ **–** Master of Science (M.S.) – Data Science, Concentration in Statistics **–** GPA: 3.8

Mahatma Gandhi Institute of Technology, Hyderabad, IND – Bachelor of Technology (B.Tech.) – Computer Science – GPA: 3.5

**TECHNICAL SKILLS**

Programming Languages**:** Python | R | SQL | MySQL | MATLAB | C | C++ | HTML | JavaScript

ML Libraries: Scikit-learn | Keras | TensorFlow | XGBoost | Polars | NLTK | spacY | PyTorch | LangChain

**Data Visualization & BI:** MS Power BI | Tableau | RStudio | Matplotlib | Seaborn

**Cloud & Big Data Technologies:** AWS | Microsoft Fabric | Azure | Apache Spark | Hadoop | Databricks | BigQuery | Oozie | HBase

**Database & Automation Tools:** Oracle | Microsoft SQL Server | Docker | Git | Power Automate | JIRA | Confluence | Agile | Scrum

**Certifications:** [AWS Certified Cloud Practitioner](https://www.credly.com/badges/e23cfc99-837d-4ca6-9d3e-d52e807b0ed4/public_url) | [Generative AI with Large Language Models](https://www.coursera.org/account/accomplishments/certificate/Q8Z6ZIMPZ9JS) | [Google Data Analytics Professional](https://www.credly.com/badges/cbae6254-face-4e98-83a8-11d6f49a6f45)

**PROFESSIONAL EXPERIENCE**

**New Jersey Equity in Commercialization Collective |** Data Research Analyst– Newark, NJ **01/2024 – 12/2024**

* Enhanced data quality by reducing FPR by 60% in university name identification using regex and fuzzy matching in Python and processing 13GB of USPTO’s PatEx data stored in a **Microsoft Fabric’s Data Lakehouse**
* **Identified over 3000 male & female inventors using Fabric’s Python notebook based on inventor names across 8 NJ institutions**
* Increased inventor visibility by 20% by building a Power BI dashboard through Dataflow Gen2, analyzing gender demographics
* Orchestrated a scalable ETL pipeline in Data Factory, streamlining data collection, organization, and transformation for evolving datasets
* Enabled strategic decision-making by **using T-SQL within SQL Analytics Endpoint to query patent data**, generating actionable insights to stakeholders of NJECC, supporting women inventors
* Improved record match rate by 11% by automating REST API calls in Data Factory to validate genders using public sources like LinkedIn and university web pages, ensuring higher accuracy for analysis

**New Jersey Institute of Technology |** Teaching Assistant– Newark, NJ **01/2024 – 12/2024**

* Progressed assignment scores by 15% through 6 lab exercises in RStudio on Convolutional Neural Networks, Data Analysis and Large Language Models (LLMs), using real-world datasets to simulate industry challenges
* Boosted class average by 17% by delivering lessons in Python and MATLAB, providing feedback, and additional materials to solidify analytical concepts; Conducted a workshop on data visualization techniques and statistical modeling using Python and R, demonstrating best practices for data cleaning, wrangling, and interpreting complex datasets to derive actionable insights

**Zenoti India Pvt Ltd |** OperationsData Analyst – Hyderabad, IND **07/2022 – 07/2023**

* Improved customer training strategies by designing 10+ KPIs using DAX within Power Query to analyze undertrained customer accounts
* Increased biweekly course completions by 200% by delivering region-specific insights to managers, enabling targeted improvements
* Reduced churn in high-value key accounts by building an intuitive Power BI dashboard (Google Analytics connector) to track go-live activity, churn metrics, and operational KPIs
* Boosted quarterly revenue by $20,000 through data-driven retention strategies that prevented seasonal churn
* Automated weekly employee reports, saved 10+ hours by developing Power Automate flows, to extract, format using Python and HTML
* Bolstered customer onboarding rate by 30% by troubleshooting integration issues via JIRA, managing Northpass & Zenoti University ,and facilitating a seamless transition from legacy systems to Zenoti
* Collaborated with cross-functional Agile Scrum teams to refine user stories and align development with business goals, ensuring timely execution of testing and feature rollouts

**ACADEMIC PROJECTS**

**Bias Detection and Mitigation in LLM-Generated Text 10/2024 – 12/2024**

* Developed a robust bias detection framework for analyzing LLM outputs using the CrowS-Pairs dataset, identifying biases across demographic factors such as gender, race, age, and socioeconomic status
* Implemented bias identification pipelines leveraging AIF360 and FairSeq to evaluate pre-trained models, including GPT-2 and BERT
* Conducted multilingual bias analysis, with a focus on Indian languages, to explore cultural and linguistic disparities in AI-generated text
* Enhanced AI model trustworthiness and mitigated harmful stereotypes, contributing to improved NLP fairness and performance

**Training Optimization Algorithms in Neural Networks 02/2024 – 05/2024**

* Engineered a Python-based framework for layer-wise and end-to-end neural network training, implementing Coordinate Descent (CD) and Stochastic Gradient Descent (SGD) for optimization
* Achieved a test accuracy of 97.8% with CD and 97.06% with SGD after 15 epochs, demonstrating robustness in high-dimensional datasets
* **Analyzed parameter norms to uncover** local minima variations,improving understanding of model generalization, optimization behavior

**Time Series Forecasting on Divvy Bicycle Sharing System 10/2023 – 12/2023**

* **Developed an ETL pipeline** using **Spark jobs** to preprocess **10 years of ride-sharing data (25M+ records) from an S3 bucket**, enabling ride duration prediction and bike allocation optimization
* Performed exploratory data analysis (EDA) with Matplotlib, Seaborn, identifying trends in ride duration, user behavior, and station usage
* Trained and fine-tuned time series models (ARIMA, VAR, Prophet) reducing forecasting error by 20%, improving ride duration prediction accuracy for better resource management
* **Deployed models using Docker** and established **CI/CD pipelines** with GitHub Actions, ensuring **reproducibility, continuous integration, and automated deployment**

**AWS Image Recognition System 10/2024 – 12/2024**

* Optimized object detection and image captioning by implementing parallel processing across two EC2 instances, improving workload distribution and achieving over 90% detection accuracy
* Developed and deployed a scalable image recognition pipeline using AWS S3, SQS, integrating Amazon Rekognition to reduce image processing time by 2 seconds per image, enhancing efficiency and scalability

**Wine Quality Prediction using Spark and Docker Deployment on AWS 10/2024 – 12/2024**

* Deployed a 4-node EMR cluster on AWS to process wine quality data, enabling scalable model training and achieving an F1 score of 0.77
* Dockerized the application, reducing deployment time by 40%compared to traditional methods, ensuring faster, efficient model delivery
* Developed a distributed machine learning pipeline in Spark, implementing and evaluating Random Forest, Decision Tree, and Linear Regression to optimize training and model performance