

MEGHNA REDDI

973.906.9462 | mreddi2001@gmail.com | [/in/meghnareddi](https://in/meghnareddi) | [Github](#) | 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

Machine Learning: Linear & Logistic Regression | Decision Tree | SVM | PCA | Random Forest | KNN | K-Means | ARIMA | VAR | SARIMAX
Natural Language Processing & Deep Learning: NLTK | Spacy | Llama2 | BERT | GPT-3.5 | CNN | RNN | GAN | LSTM | Transformers
Data Analysis & Statistical Methods: Regression Analysis | Hypothesis Testing | A/B Testing | Time Series Analysis | Statistical Testing | Minitab | Predictive Modeling | Data Mining | Feature Engineering | Data Wrangling
Programming Languages: Python (Sklearn, Polars, TensorFlow, Keras, Pandas, Numpy, PyTorch, Seaborn) | R (ggplot) | SQL
AWS Services: EC2 | S3 | Sagemaker | Lambda | IAM | Quicksight | EMR | SQS | Lambda | Redshift | Rekognition
Azure Services: Azure AI/ML | AI Vision | AI Custom Vision | AI Language Service | Azure Data Lake Storage | Databricks | Fabric | Lakehouse | Data Factory | Synapse Analytics | Dataflow Gen2 | Data Pipelines | Notebooks | SQL Database
GCP Services: BigQuery | Cloud Storage
Data Visualization & Reporting: MS Power BI | Tableau | Looker | RShiny | Quicksight
Database & Automation: Oracle | MS SQL Server | MySQL | Docker | Git | Power Automate
Project Management: JIRA | Confluence | Agile | Scrum
Certifications: [AWS Certified Cloud Practitioner](#) | [Generative AI with Large Language Models](#) | [Google Data Analytics Professional](#)

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, processing USPTO's PatEx data (13GB) in 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 using Dataflow Gen2, analyzing gender demographics
 - Built a scalable ETL pipeline in Data Factory, streamlining data collection, organization, and transformation for evolving datasets
 - Leveraged SQL Analytics Endpoint to generate actionable insights, supporting NJECC stakeholders in decision-making
 - Improved record match rate by 11% by automating REST APIs, validating genders from LinkedIn and university web pages
- 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 CNNs, Data Analysis and Large Language Models (LLMs), using real-world datasets to simulate industry challenges
 - Conducted a data visualization & statistical modeling workshop using Python, teaching data wrangling and interpretation practices
- Zenoti India Pvt Ltd | Operations Data Analyst – Hyderabad, IND** **07/2022 – 07/2023**
- Improved customer training strategies by designing 10+ DAX KPIs in Power Query to analyze undertrained customer accounts
 - Increased biweekly course completions by 200% by delivering region-specific insights to PMs, enabling targeted improvements
 - Reduced churn by 12% in key accounts by developing a Power BI dashboard (Google Analytics connector) to monitor customer activity, churn rates, and revenue trends
 - Boosted quarterly revenue by \$20,000 via data-driven retention strategies preventing seasonal churn
 - Saved 10+ hours by developing Power Automate flows, to extract and format employee reports weekly using Python and HTML
 - Bolstered onboarding rate by 30% by troubleshooting integration issues via JIRA, managing Northpass & Zenoti University
 - Collaborated with cross-functional Agile Scrum teams to refine user stories, align development with business goals, ensuring timely execution of testing and feature rollouts
- Infor India Pvt Ltd | Data Solutions Consultant – Hyderabad, IND** **04/2022 – 06/2022**
- Reduced migration errors by 30% by optimizing SQL-based data migration processes with Postman and CRUD operations
 - Developed 10+ custom 4GL reports, integrating Pivot Tables and VLOOKUP in Excel to analyze supply chain data
 - Collaborated with R&D and customization teams to implement tailored data-driven solutions to meet client-specific requirements

ACADEMIC PROJECTS

- End-to-End Visual and Textual Processing & Support System using Azure AI** **01/2025 – 02/2025**
- Developed an AI-driven data processing system using Azure AI Vision to extract and search text from documents and images, improving data accessibility and metadata generation
 - Built an image classification model with Azure AI Custom Vision to categorize images, optimizing content management workflows
 - Integrated Azure AI Language Service for sentiment analysis, entity recognition, and key phrase extraction, enabling deeper insights from large datasets like customer reviews and feedback
 - Deployed a Question Answering system using Azure AI Language, reducing query response time by 35% through natural language processing and knowledge base integration
- Bias Detection and Mitigation in LLM-Generated Text** **10/2024 – 12/2024**
- Designed a bias detection pipeline using AIF360, FairSeq for fairness metrics, analyzing LLM outputs for biases across demographic factors
 - Evaluated Llama2, BERT, and other pre-trained models using CrowS-Pairs dataset, quantifying bias levels and fairness disparities
 - Conducted multilingual bias analysis, focusing on Indian languages to address cultural disparities in AI-generated text, improve inclusivity
 - 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

Predictive modeling & Time Series Forecasting on Divvy Bicycle Sharing System

10/2023 – 12/2023

- Developed an ETL pipeline using Spark jobs to preprocess 25M+ ride-sharing records from S3 bucket, to predict ride duration
- 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 resource management
- Deployed models using Docker and established CI/CD pipelines with GitHub Actions, ensuring reproducibility & continuous integration

AWS Image Recognition System

10/2024 – 12/2024

- Optimized object detection, image captioning by implementing parallel processing with 2 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 an ML pipeline in Spark, implementing Random Forest, Decision Tree, and Linear Regression to optimize model performance

Election Data Visualization and Analysis

10/2024 – 12/2024

- Built an interactive Tableau dashboard analyzing U.S. election data (5 GB) stored in AWS Redshift, uncovering voter turnout trends
- Designed complementary visualizations and filters in Tableau, allowing data exploration by state, gender, and year
- Tripled student engagement, attracting 500+ weekly dashboard views, increasing political awareness among students