

# TEJESHWAR REDDY GANGIREDDY

+1 (716) -787-0430 | [tejeshwa@buffalo.edu](mailto:tejeshwa@buffalo.edu) | <https://github.com/teja145252> | <https://www.linkedin.com/in/teja-reddy55/>

## SUMMARY

Experienced Data Analyst with a robust background in data collection, data engineering, advanced data analysis, and business intelligence. Demonstrated expertise in designing and implementing ETL processes, transforming raw data into actionable insights, and developing scalable solutions for effective data management and business decision support. Proficient in advanced analytics, data visualization, and machine learning applications for comprehensive data-driven solutions.

## SKILLS & TOOLS

**Programming Languages:** | Python | Java | SQL | R | MATLAB | C | HTML5 | CSS3 |

**Databases:** | Postgres SQL | SQLite3 | MySQL | MongoDB | Cassandra |

**Machine learning:** Scikit-learn | TensorFlow | PyTorch | Reinforcement learning | OpenAI gym | TensorFlow RL | NLP | NLTK | spaCy | Hugging face transformers library | BERT | GPT 3.5 | Langchain | OpenCV | Pandas | VisiData | NumPy | Matplotlib | Seaborn | data manipulation | data visualization | Gen AI | prompt engineering | GAN's | Vector databases | VAE's |

**Tools/Technologies:** | Cloud services| Azure | AWS | snowflake | Version control |Git | GitHub | Spark | ETL Process | Apache Airflow | Docker | Databricks | Apache Kafka | CI/CD | Jenkins | Agile methodology| Microsoft BI stack (SSRS, SSIS, SSAS) | powerBI (DAX, slicers and filters, RLS, performance tuning) | Tableau |SharePoint dashboards | KPI Analysis | A/B testing | digital marketing channels (PPC , SEO, referral marketing)

## WORK EXPERIENCE

### RLHF Software Engineer

May 2024 – present

*Outlier AI, Inc, SF, CA.*

- Applied Reinforcement Learning with Human Feedback (**RLHF**) to enhance AI model performance by providing human feedback and ensuring accurate **API executions** based on user inputs. conducted rigorous monitoring and validation of model outputs, refining decision-making processes through iterative feedback loops, resulting in a **15% increase in model efficiency** and accuracy.
- Collaborated with cross-functional teams to align model outputs with business objectives and user requirements, developing and implementing training protocols to optimize model performance.
- Led the development of the **multitier deviation detection system**, Flamingo, designing initial prompts and evaluation methods to ensure accurate and relevant model responses. created techniques to identify coding deviations across multiple layers, managing multi-turn interactions and follow-up prompts to **refine model behavior and responses**.
- Conducted detailed assessments to establish task baselines, enhancing system robustness and error detection, streamlining workflow for deviation resolution, and **reducing resolution time by 20%**.

### Data Analyst intern

July 2023 – December 2023

*Eitacies Inc, Santa Clara, CA*

- Collected credit card datasets from various sources, gathering over 10,000 images mostly from online repositories, internal databases, to ensure a diverse and representative dataset for analysis. Conducted comprehensive **data profiling** using Python's **PIL library**, extracting metadata such as image dimensions and color histograms, and detecting inconsistencies with a 20% reduction in data inconsistencies. Led **data cleansing** initiatives, **standardizing** image formats to JPEG and resolutions to 300 DPI. Applied filters with **OpenCV** and PIL to enhance image quality, resulting in a 30% improvement in image clarity.
- Standardizing improved the efficiency of data processing tasks by 25%. Conducted **normalization** of data to improve consistency, including rescaling the pixel values by using **min max** normalization, which enhanced the reliability of data for analysis. Analyzed credit card image data to **identify trends and anomalies**, such as common design features and unusual patterns, providing actionable insights that guided further investigations.
- Developed a SaaS platform aimed at enhancing security and compliance in conference settings through cutting-edge computer vision and ML/AI techniques. **Annotated images** containing the region of interest using the **CVAT annotation tool** and built an object detection model using the pre-trained ultralytics YOLO object detection module. One of the core modules involved using object detection models to identify PCI DSS violations within conference recordings.
- Implemented real-time object detection capabilities. adjusted the model's parameters and training processes which have optimized performance by 35%. Integrated OCR technology, specifically **Tesseract OCR**, into our platform, I've enabled precise text detection and extraction from conference recordings. Also, used pre-trained BERT and GPT-3.5 base models for text classification.
- Implemented **Azure Stream Analytics** for streaming data ingest, facilitating real-time processing of incoming data streams. Leveraged Azure Machine Learning for **ML model orchestration**, ensuring efficient management and deployment of machine learning models. Models were deployed with a blue/green deployment strategy using **Azure App Service**, providing zero downtime and instant rollback capabilities for seamless updates and maintenance. Incorporated **Agile methodology** for iterative development and responsiveness to evolving requirements.
- Developed a detailed **data analysis** project **extracting** over 200 hours of audio recordings, **transforming** it by cleaning and transcribing them, and **loaded** it into a csv file then analyzed transcripts for harsh words, discriminatory language, or sexual bullying. Achieved an 85% accuracy rate in detecting inappropriate language using NLP techniques. Employed NLP techniques to categorize and flag inappropriate language, with an 85% accuracy rate in detecting and categorizing abusive or discriminatory behavior. Utilized VADER and Text Blob for **sentiment analysis** to identify negative sentiment indicative of inappropriate behavior. Achieved the said accuracy in detecting negative sentiments.

- Conducted in-depth data analysis on transcribed text to identify patterns of abusive behavior, creating detailed reports and providing recommendations for effective intervention strategies.

## EDUCATION

**Master's: Data science, University at Buffalo, The State University of New York,**

**February 2024**

**Course work:** Statistical data mining using R, Machine learning (supervised and unsupervised), Probability and Statistics, Data Models and Query Languages / CGPA: 3.5

## ACADEMIC PROJECTS

**Sleep Analysis Dashboard: Power Bi, DAX.**

**June 24 – July 24**

- The sleep analysis dashboard visualizes sleep health and its affecting factors. This dashboard was created using **Power BI**. I collected the sleep health and lifestyle dataset from Kaggle, cleaned and transformed it, **Reduced data processing time by 30%** through efficient cleaning and transformation techniques. including the division of blood pressure into separate systolic and diastolic columns.
- I **categorized** blood pressure levels into 5 distinct categories using DAX (SWITCH) based on the systolic and diastolic columns. I created 8 **DAX** measures to calculate various metrics for the dashboard.
- The dashboard visualizes general sleep patterns, sleep disorder analysis, health metrics comparison, the impact of occupation, gender and age distributions, and more. I configured **filters and slicers to enhance interactivity**, allowing users to explore different aspects of the data. Visualized key performance indicators (**KPI's**).
- **Row-level security** was not implemented, as it was out of scope for the project due to the open-source nature of the data.
- **Link to my dashboard:** [View My Power BI Dashboard](#)

**Inventory Management System: PostgreSQL, python**

**February 23 - May 23**

- Designed one of the best and most optimized inventory management systems. The major goal of the database design is to locate entities engaged in inventory management. then determined each entity's properties using these entities.
- The created database followed **BCNF**, also tested the database by updating and deleting data entities, used **Indexing** for Improved performance, and created several functions.
- Developed a user interface where users can give the SQL queries in the input box and fetch the results through a web page which is deployed on a local server. used **HTML, CSS, and ReactJS** to create user interfaces.

**Time series Analysis project (oil price dataset): R programming, ggplot2**

**February 23 - May 23**

- Preprocessed the dataset, imputed missing values using seadec imputation, and performed **augmented dickey fuller test** to check the non-stationarity of the dataset.
- Tested several models i.e., **ARIMA, SARIMA, Holt winters, ETS, STL, Prophet** etc. used AIC, BIC and RMSE values as metrics to evaluate the performance of models, also created several plots for analysis.
- Gained expertise in **identifying trends, seasonal patterns, cyclic patterns, and autocorrelation** in time series data.

**Recession Prediction: python (matplotlib, seaborn), R (ggplot2)**

**August 22 – December 22**

- Demonstrated a thorough comprehension of macroeconomic concepts leveraging time series data starting from 1960. observed correlations between features, including yield curve data. Utilized data of the past 58 years to provide predictions.
- Developed expertise in **model selection**, filtering features, and **hyperparameter tuning** to achieve a **predictive accuracy of 80%** for identifying economic recessions. Skilled in building predictive models for economic analysis.

**Customer Segmentation: python, SQL**

**August 22 – December 22**

- Led a customer segmentation analysis and determined distinct customer groups based on previous purchases and demographic data. **Improved model accuracy by 20%** over previous product-based segmentation methods.
- Developed a deep understanding of clustering and dimensionality reduction techniques, specifically deploying **Principal Component Analysis (PCA)**, to extract meaningful insights from customer data.
- Devised targeted marketing strategies resulted in a **12% increase** in overall customer engagement and an increase in sales.

## Website

<https://teja145252.github.io/>