TEJESHWAR REDDY GANGIREDDY

**|**+1716-787-0430 **|** [tejeshwa@buffalo.edu](mailto:tejeshwa@buffalo.edu) **|** [https://www.linkedin.com/in/teja-reddy55**/**](https://www.linkedin.com/in/teja-reddy55/) **|** [https://github.com/teja145252](https://github.com/teja145252 )**|**

# SKILLS

| Python | Java | SQL | R | MATLAB | C | HTML5 | CSS3| Postgres SQL | SQlite3 | MySQL | MongoDB | Cassandra | Scikit-learn | TensorFlow | PyTorch | Reinforcement learning | OpenAI gym | TensorFlow RL | NLP | NLTK | spaCy | Hugging face transformers library | BERT | GPT 3.5 | Langchian | OpenCV | Pandas | VisiData | NumPy | Matplotlib | Seaborn | data manipulation | data visualization | Gen AI | prompt engineering | GAN’s | Vector databases | VAE’s | Azure | AWS | snowflake |Git | GitHub | Apache Spark | Hadoop | Hive | Impala | ETL Process | Apache Airflow | Docker | Databricks | Apache Kafka | CI/CD | Docker | Kubernetes| Jenkins | Agile methodology| Microsoft BI stack (SSRS, SSIS, SSAS) |MS Excel | Microsoft purview |Data Governance | powerBI (DAX, slicers and filters, RLS, performance tuning) | Tableau |Data modeling | Ad hoc analysis| KPI Analysis | A/B testing | digital marketing channels (PPC , SEO, referral marketing)

# EXPERIENCE

***OUTLIER AI INC***  *San Francisco, CA May 2024 – present*

## RLHF software engineer

* 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.
* Implemented a system that utilized **Natural Language Instruction** to guide ML models with specific questions and code. Assessed model performance based on the provided **Natural Language Explanations** and corrected the model's rationale as needed.
* 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%.**

***EITACIES INC***  *Santa Clara, CA Jul 2023 – Dec 2023*

## Data Science Intern

* 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. Managed and cleaned image metadata using **MS Excel**. 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.
* 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.
* Created and troubleshooted **SQL** queries for data extraction and analysis. Optimized complex SQL queries to improve performance and reduce execution time by 40%. Conducted data analysis to identify key metrics and trends, delivering actionable insights that informed business decisions.
* 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. **Utilized VADER** and **Text Blob** for **sentiment analysis** to identify negative sentiment indicative of inappropriate behavior. Achieved the said accuracy in detecting negative sentiments.

***SILICON WIRELESS SYSTEMS*** *Hyderabad, India January 2020 – August 2021*

## Data Science Associate

* Led the development of an advanced analytics platform that integrated data from IoT wireless machines, enabling real-time monitoring and predictive maintenance.
* Designed and implemented **data pipelines** to collect, clean, and preprocess large volumes of **sensor data** from IoT devices, ensuring high-quality inputs for analysis. Employed both supervised and unsupervised machine learning models using scikit-learn to analyze **IoT data**, including **anomaly detection**, clustering, and **predictive modeling**.
* Leveraged **deep learning frameworks** like **Keras and TensorFlow** to enhance model performance and tackle more complex data patterns.
* Utilized the trained models to generate actionable insights, such as early detection of equipment failures and optimization of operational efficiency. Continuously **refined models** and algorithms to improve accuracy and reliability, leading to a **20%** reduction in unplanned downtime and significant cost savings for the company.
* Worked closely with cross-functional teams, including engineering and operations, to **deploy models** into production and ensure seamless integration with existing systems.

# 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

# PROJECTS *For more information visit my* [*website*](https://teja145252.github.io/) *and* [*GitHub*](https://github.com/teja145252)

**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**](https://app.powerbi.com/view?r=eyJrIjoiZTg5NzQ5ZWQtM2U5NC00MjUzLWE0NTEtYWE2MzY3MDlmZjM1IiwidCI6Ijk2NDY0YThhLWY4ZWQtNDBiMS05OWUyLTVmNmI1MGEyMDI1MCIsImMiOjN9)

**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.

**Variant Call Format file parsing:** *python, SQL* **August 22 – December 22**

* Parsed a complex dataset with over 600,000 rows, containing predictor fields such as **FATHMM\_pred, LRT\_pred**, **MetaLR\_pred**, and others. Identified and corrected irregular data, ensuring data integrity.
* Optimized the data loading process, achieving a 20% faster load time compared to traditional methods. Designed a **normalized database** using refined data, creating 6 interconnected tables. Gained in-depth knowledge of **SQL** for data analysis and management.
* While the project focused on genetic data predictors, the experience with **large-scale data parsing**, optimization, and database management is directly applicable to analyzing bulk or single-cell **RNA-seq** data.

**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 customerengagementand an increase in sales.

# WEBSITE

*Please check my GitHub to view the source code of the website,* [*click here*](https://teja145252.github.io/) *to view my portfolio.*

# TRAINING & CERTIFICATION

Data Visualization with power BI – by great learning  
Oracle certified associate – java se8 programmer

Statistical Analysis – by Great learning