**Harsha Teja**

**mlharshateja@gmail.com**

**Mobile: +91 7349781550**

**Professional Summary:**

With over 8 years of experience, I am a Data Scientist specializing in Data Science, Machine Learning, Deep Learning, and Python programming. I excel in developing analytical solutions and optimizing data processes, currently focusing on the AML RtB DBT migration project, where I handle data extraction, transformation, and migration from Oracle to Snowflake..

**Professional Experience:**

* **Currently serving as a Senior Technical Lead at HCLTech since April 2024,** working on the AML RtB DBT migration project for USAA, focusing on data migration from Oracle to Snowflake.
* **Previously worked as a Senior Professional I-Senior Data & AI Engineer at Capgemini from November 2022 to April 2024,** specializing in data science and deep learning projects.
* **Prior experience includes working as a Developer at KANINI Software Solutions from October 2021 to October 2022,** focusing on data science and machine learning projects.
* **Worked as a Developer at Tata Power Solar from July 2016 to October 2021,** where I contributed to data-driven initiatives and deep learning projects in the renewable energy sector.
* **Previously employed as a Senior Engineer at Badve Engineering from January 2013 to November 2015,** specializing in cost-cutting analysis and implementing PLC, robotic, and automation solutions.

**Academic Profile:**

* Completed M.Tech in Mechatronics from University College of Engineering, Jawaharlal Nehru Technological University (JNTU-H) in 2013 with a Percentage/DGPA of 79%.
* Obtained B.Tech in E.C.E from CIST, Kakinada, affiliated with Jawaharlal Nehru Technological University (JNTU-K) in 2011 with a Percentage/DGPA of 61.2%.

**Technical Skills:**

* **Machine Learning**: Machine Learning, Data Visualization, Predictive Modelling
* **Programming Languages**: Python ,PLC, **,** Basics of SQL.
* **Deep learning:-**  Deep Learning: CNN, Image Classification, TensorFlow, Keras
* **Natural Language Processing (NLP):** Text Classification, Sentiment Analysis, Named Entity Recognition, Word Embedding, Vectorizing
* **Reporting Tools:** MS Office (Word/Excel/Power Point), Plotly, Tableau
* **Others:**  Streamlit (Front-end Development)Flask, Statistics, PLC,HMI,Robotics.
* **Domain Knowledge:**  Automotive, Manufacturing,Robotics and Power
* **Soft Skills:**   Problem-Solving, Active Learning, Risk Analysis.

**Core Competencies:**

**Machine Learning**: Expertise in supervised learning, model optimization, and AML data transformation.

**Deep Learning:** Skilled in designing CNN architectures and using TensorFlow and Keras for advanced data analysis.

**Statistical Analysis & EDA:** Proficient in analyzing datasets, performing EDA, and feature engineering for model improvement.

**AWS Technologies:** Experience with SNS, SQS, Batch, DynamoDB, and Docker for efficient data processing and pipeline management.

**API Development**: Adept at creating APIs with AWS API-Gateway, Lambda, and DynamoDB for data integration.

**Jetson Projects:** Developed solutions on NVIDIA Jetson platforms for real-time data processing in automotive and robotics.

**POC Development:** Hands-on experience in executing POC projects and leveraging AI for AML compliance and data science.

**Project Experience:**

**Project #1:**

**Project : AML RtB DBT Migration for USAA**

**Technology Used- : Snowflake,DBT,MFP,Putty,Winscp,Datastage**

**Role** : **Developer**

**Date : April 2022 to still.**

**Client : USAA**

**Description:**

The project aims to migrate data from an Oracle database to Snowflake as part of the AML (Anti-Money Laundering) RtB (Run the Business) initiative for the USAA client. The goal is to efficiently manage and transform data, ensuring it aligns with AML compliance requirements.

**Responsibilities:**

**Data Extraction**: Utilized DataStage to extract data from Oracle databases.

**Intermediate Storage:** Temporarily stored extracted data in an intermediary database for pre-processing and validation.

**Data Transformation:** Employed DBT for data cleansing, aggregation, and enrichment to meet AML requirements.

**Final Data Load:** Loaded transformed data into Snowflake for scalable and efficient data handling.

Outcome: Enabled efficient AML operations and compliance checks with the successfully migrated and transformed data in Snowflake.

**Project #2:**

**Project : NUANCE \_DASV**

**Technology Used- : Python, Pandas,Deeplearning ,Docker,AWS,CNN, ,ROS1, Jetson embedded systems**

**Role** : **Developer**

**Date : Nov 2022 to April-24.**

**Client : Capgemini-Germany(internal)**

**Description:**

NUANCE is a comprehensive data quality analysis solution based on machine learning techniques, designed to analyze the data generated by ADAS/AD (Advanced Driver Assistance Systems / Autonomous Driving) campaigns. The solution focus on detecting calibration misalignment of LiDAR and cameras. NUANCE is equipped with the ability to detect the presence of mud, dust, water and other contaminants on camera lenses that may compromise the quality of collected data. This feature is critical for ensuring that the data generated by ADAS/AD campaigns is of the highest possible quality, allowing researchers and engineers to make informed decisions about the performance of autonomous vehicles and related systems

**Responsibilities:**

* Designed AWS architecture for sensor miscalibration detection and data capture.
* Created Docker-based solutions for batch data processing, with results stored in DynamoDB.
* Implemented and fine-tuned CNN models for image analysis.
* Developed and tested APIs for front-end integration.
* Deployed ROS nodes on Jetson Xavier AGX for real-time processing and analysis.
* Engineered solutions to improve sensor alignment and blockage detection.
* Collaborated with robotics teams to integrate solutions into autonomous navigation systems.
* Conducted continuous testing and refinement in both simulated and real-world environments.

**Project #3:**

**Project : ML360**

**Technology Used- : Python, Pandas ,Numpy ,Matplotlib ,Skleran,NLP,CNN**

**Role** : **Developer**

**Date : Oct 2021 to Oct 2022**

**Client : Kanini(internal)**

**Description:**

ML360 is a web-based application that provides a comprehensive suite of data science tools for model building, including preprocessing, EDA, feature engineering, and model evaluation for regression and classification task

**Responsibilities:**

* Involved in requirement gathering and Architecture design of the project for implementation.
* Developed AutoML features for ease of use by non-data scientists and custom functionalities for data scientists.
* Built pipelines for machine learning processes, including preprocessing and model evaluation based on data size.
* Implemented text data vectorization techniques and parameter tuning for model optimization.
* Managed training and retraining of models as needed

**Project #4:**

**Project : AI-assisted Quality Assurance of Solar Panels.**

**Technology Used- : Python, Pandas ,Matplotlib ,CNN, Deep learning,Tensorflow,Opencv**

**Role** : **Developer**

**Date : Feb 2020 to Oct 2021.**

**Client : Tata Power(internal)**

**Description:**

Electroluminescence imaging is gaining traction in detecting PV module defects both during manufacturing and operation. Detect all defects irrespective of the individual defect criteria definition Type of defects covered: Crack: micro line and branch cracks Grid defect: Grid, dark black spot, chipped cell Short circuit, Working. With help AI system need to classify the good and Bad Modules/images.

**Responsibilities:**

* Involved in requirement gathering and Architecture design of the project for machine learning implementation.
* Applied computer vision techniques, CNN, TensorFlow, and OpenCV for fault identification in EL images.
* Artificial intelligence convolution neural network model to classify images of different defects.
* Image preprocessing
* Using Python Annotation of images by defect type.
* Building baseline models for the requirements with necessary data preparation.
* Parameter tuning process for optimal model hyper parameter.

**Project #5:**

**Project : Automation line - Equipment Failure Prediction**

**Technology Used- : Python, Pandas ,Numpy ,Matplotlib ,Random Forest Classifiers**

**Role** : **Developer**

**Date : Jan 2018 to Nov 2019.**

**Client :Tata Power**

**Description:**

In predictive maintenance, our goal was to prevent equipment failures by analyzing data collected from various machines, including Automatic machines, SPMs, and Robots. I gathered extensive data on machine failure, maintenance records, and other relevant information to develop predictive models. After testing multiple machine learning algorithms, including decision trees, I found that Random Forest emerged as the strongest technique due to its robustness. This approach helped us predict potential failures within specified time windows, mitigating production losses and reducing costs.

**Responsibilities:**

* Translate product requirements into analytical requirements/specification, design and develop required functionality.
* Involved in requirement gathering and Architecture design of the project for machine learning implementation.
* Participating in Data Preprocessing Techniques in order to make data useful for creating Machine Learning Models.
* Developed the model for machine failure predication for particular machines.
* Parameter tuning process for optimal model hyper parameters
* Run machine learning tests and experiments

**Project #6:**  :**Vacuum and Temperature Data analysis for Uptime improvement in Stringers Machines using IOT device.**

**Technology Used- : Python, Pandas ,Numpy ,Matplotlib ,Tableau 10**

**Role** : **Developer**

**Date : Dec 2016 to Nov 2017**

**Client :Tata Power**

**Description:**

The main aim of the project is improve the quality of the product at what interval vacuum and temperature is varying which leads to ,defect part is manufactured.. Installed the IOT device in machine to collect the data of temperature and vacuum in time interval. Extracted this data to the local host. Analyzed the data and in python , Prepared EDA

**Responsibilities**:

* Connecting the IOT sensor data to server.
* Performed Data Pre-processing using Python libraries like Pandas Techniques in order to make data useful
* Prepared exploratory data analysis in python.
* Finding the correlation between temperature ,Vacuum and time.
* Email notification to concern person when abnormal values observed.
* Perform statistical analysis and fine-tuning using test results.
* Prepared Dashboard in Tableau 10.
* Uptime of machine increased from 89 % to 95% and improved quality of product.