MADIHA (**Madi**) ANSARI

Data Science and Machine Learning Engineer

Mobile: 0423292380   **E-mail**: madihaalansari@gmail.com   **LinkedIn**: https://www.linkedin.com/in/madihaalansari/

Experienced data professional helping businesses to achieve data-driven solutions and implementing creative ways to cut costs by leveraging ML/AI-based technologies.

Core Competencies:

Project Management: End-to-end ML project planning, developing, testing, and monitoring, using frameworks like agile and CI/CD for automation and scalability.

Communication: Clear speaking and writing ability to communicate complex data concepts with management and cross-functional teams.

Collaboration: Ability to work closely with diverse key stakeholders to identify critical priorities through continuous feedback, ensuring on-time project delivery.

Technological Awareness: Always aware of the business needs. Identifying creative ways to implement task-specific technological solutions.

Leadership: Empowering others by encouraging and offering challenging goals with responsibility and appreciation.

Technical Competencies:

Programming Languages: Python, SQL, PowerShell/Bash, C++, VBA.

ML/AI Frameworks: Scikit-learn, Keras, Pytorch, TensorFlow, PySpark, OpenCV, BERT, VADER.

Cloud Technologies: AWS S3, AWS EC2, AWS RDS, Azure Analysis Services.

Data Visualisation Tools: Power BI, Python, and Tableau.

DevOps: CI/CD, APIs, Versioning, Logging, ML workflow, GitHub, Docker. Apache Spark.

Project Management: Knowledge of Agile, and Prince2 frameworks.

Professional Experience

---

1 - Ventura Motors, 'Business and Data Analyst' (Data Science and ML)   Sep 2021 – Present

Data Solution for ML Applications

Project: Electric Bus Performance and Battery Management System (Application)

Total Project worth $10 million.

- Implemented End-to-end, **Big data pipeline** in AWS for ingesting MQTT Broker data from CAN Bus.
- Applied advanced ML techniques like **Moving Averages, Principle Component Analysis,** etc. to preprocess, normalize, and transform noisy data for analytics.
- Analyzed critical vehicle metrics for **Predictive maintenance**.
- Collaborated with diverse stakeholders, including government, private, and research-based entities.
- *Improved data-driven decision-making in saving operational costs by 20%.*

MLOPS

Project: Intelligent Accidents Monitoring and Management System (Application) - IAMS.

Estimated annual accident damage $3 million.

- Developed and Deployed **End-to-end ML Application** on **AWS EC2** with bus accident data spanning over 20 years.
- Trained multiple supervised models like **SVM**, **XGBoost**, and **Random Forest** for predictive analytics.
- Trained and used **Silhouette Score** to evaluate performance for **DBSCAN** and **K-Means** clustering models to understand accident patterns.
- Used **GitHub** and **Docker** for CI/CD Implementation to streamline data integration and model monitoring.
- Collaborated with management and cross-functional teams.
- The System helped management mitigate road accident severity, thus by investing in customized driver training programs.
- *Helped reduce annual damage costs by 15% in better managing third-party accidents and insurance claims.*

* For this project I also incorporated time-series analytics from the ADAS system data in comparison to Non-ADAS vehicles.

Data Solution for ML/AI Applications

Project: Patronage Monitoring System for Public Buses in Southeast Victoria.

Business contracts worth 96% of revenue.

- Created **end-to-end data solution** for daily Passenger movement using AWS. I used the **GPR - Gaussian Process Regression** Model, a probabilistic approach to predicting passenger count.
- Preprocessed real-time data using **Scikit-learn.** Analytics based on time-series features like days of the week, peak-off peak, events, and season variations.
- The system helped in understanding diverse customer needs and route demands in the Eastern Region. Thus providing better service and saving operational costs for managing underutilized assets.

- *Comparison analysis found a 20% discrepancy in passenger counts with Myki ticketing system data*.

Text Analytics using AI modeling
Project: Customer Satisfaction Analysis
> Used NLTK Tokenization and Lemmatization techniques. Employed **VADER** for polarity scoring and **BERT** for part-of-speech tagging and named entity recognition(**NER**).

Image Processing using OpenCV
Project: Paperless Invoice data capture
- Automated text extraction from encrypted financial remittances in **PDF** files and other invoices in **image** formats**.**
- Utilized **OpenCV** and **PIL** to preprocess noise in images like **grayscale conversion**, **thresholding,** etc.
- Tuned the **Tesseract-OCR** engine to optimize performance for documents with complex layouts, like columns and tables.
- Finally, **Regex** pattern matching was used for the targeted extraction of invoice numbers, dates, payment amounts, and account details.
- *This helped in time and cost efficiency, by manually taking 10 mins to read per invoice a day, was cut down to 1 min per invoice with OCR enabled solution. It saved the company monthly $2,500 on 1000 docs per month.*

Performance Monitoring Analytics
Project: Vehicle Safety Inspection & Vehicle Off-Road Tracking System
- Developed an automated data pipeline for the **Power BI** application.
- Used various **Python** libraries to preprocess data for a traffic light predictive safety maintenance dashboard.
- *It Improved asset management significantly by achieving 100% safety compliance on 1000 vehicles thus saving $1,500 per vehicle if they default. Previously 5 vehicles were defaulting each month.*

Information Protection
Project: Risk Management for Data & Information Security.
> Worked on various data security frameworks like ISO 27001 & Essential 8 to define an Industry-specific Risk Register. Collaborated with government and non-government agencies in improving business private data security (resilience, response & recovery).

Data Strategy for ESG
Project: Implementing Sustainability Standards Framework - Reporting for Corporate Carbon Footprint
> Developed strategic data required for tracking corporate emissions and reporting for compliance frameworks. Pioneered **Ventura's Corporate ESG report in 2023**.

2 - Miepol (Pty) Ltd, Data Scientist   Sep 2020 – Sep 2021

Data Science
Project: Bus Stop Management & Predictive Maintenance System
- Developed a real-time monitoring system that displayed **26,000 bus stops** audited daily across Victoria.
- Utilized time-series analysis and supervised ML models like **Regression**, and **Decision Trees** to create a predictive maintenance application in Power BI. The System triggered maintenance alerts based on set priorities.
- Monitoring **DDA Compliance**, **Risk Assessments**, and **Parts Warranty** on Bus Infrastructure.
- The project covered all aspects of bus infrastructure maintenance from financial to operational.
- *It saved the business more than $50,000 annually on purchasing complex data solutions and tailoring them to the stakeholder's needs.*

Data Engineering
Project: Wayfinding for Passengers
- Working with **GPS dataset** to identify Bus Stops that require Route Information to be updated.
- Automated ETL pipeline for Power BI Analytics.
- Helped the business save hours on manual data processing

3 - Teacher – Mathematics, STEM, and Programming in Python /C++   Aug 2011 - Oct 2019
STEM-focused education with project-based programming.

4 - LMKR, System Research Analyst   Jan 2005 – Nov 2007

Project: Analytics for Reservoir Performance for Oil and Gas.
- **Geospatial analysis:** Extracted relevant data from the **GIS** system to create Spatial distribution on well locations and production. Used Excel VBA to create functions to evaluate production rates.
- **Spatiotemporal analysis**: Utilised advanced **Excel** techniques for historical time-series analysis. Also created a correlation between various factors affecting production rates.

- **Seismic data analysis:** Extracted data from **Landmark Graphics** and used **MATLAB** to understand the effects of sub-surface changes on fluid flow.
- *It enhanced insights into oil and gas exploration activities.*

Project: Seismic Activity Monitoring and Anomaly Detection.
- Collected historical seismic data from repositories like seismic waves, magnitude, and timestamps(.csv).
- Applied **Fourier transformation** and **Wavelet analysis** to analyze the frequency and magnitude of waves using **MATLAB**.
- Based on historical data created inferences using basic **statistical analysis** for the likelihood of high and low-impact events. Compared, geological historical heat maps to identify earthquake trends surrounding a potential oil reservoir.
- Used **ARIMA** for time series data analysis to determine seasonal shifts and plotted the results in MATLAB.
- Established a threshold for seismic activity to trigger alerts/potential hazards and raise a flag for geologists and engineers.
- *Enhanced monitoring of seismic activity in oil and gas fields and improved understanding of potential risks for safety planning.*

5 - Petrosin Engineering, Database Coordinator   Sep 2003 – Dec 2004

Project: Data-driven Procurement Optimization for LPG Cylinder Supply
- Gathered inventory stocks, procurement costs, and client information. Tracked orders. Organized it into a structured database using **MS Access**. Created cross-application automation using **VBA**.
- Used advanced **Excel** to create analysis. Developed different procurement scenarios based on historical data. Did manual forecasting for future demand for cylinders and to understand client's needs.

Education and Certifications
—---

- **MLOps (AWS): Data Pipeline Automation & Optimization using Amazon Web Services.**

- **MLOps (Azure): Data Pipeline Automation & Optimization using Microsoft Azure Machine Learning.**

- **Apache Spark for Data Engineering and Machine Learning.**

- **Cyber Security Risk Management.**

- **Analyzing and Visualizing Data with Microsoft Power BI.**

Master of Data Science. University of San Diego, CA
Relevant Coursework: SQL for Data Science, Python for Data Science, Machine Learning, Probability and Statistics in Python, Big Data Analytics Using Spark, Data Protection & Security.

Bachelor of Computer Engineering (Hons). Comsats University
Relevant Coursework: Computer programming, Embedded Systems and Design, AI and Machine Learning, Computational Algorithms, Signals and Image Processing, Calculus and Discrete Mathematics, Linear Algebra, Systems Engineering, Database Design and Data Structures.

*References available on demand