

# MADIHA (Madi) ANSARI

## Data Science and Machine Learning Engineer

Mobile: 0423292380

E-mail: [madihaalansari@gmail.com](mailto:madihaalansari@gmail.com)

LinkedIn: <https://www.linkedin.com/in/madihaalansari/>

Experienced data professional with a mathematics and engineering background, passionate about transforming raw data into actionable insights and driving strategic decision-making. Highly detail-oriented and adaptable with a growth mindset. Over a decade of multi-hat experience in data science, machine learning, and statistical analysis, coupled with strong data engineering skills.

### Core Expertise:

**Machine Learning & AI:** Developing, testing, and implementing machine learning models, including some experience with LLMs. AI-based Image processing and text handling.

**Data Engineering & Architecture:** Design and implementation of scalable data architectures, ETL processes, and database design.

**Statistical Analysis & Modelling:** Proficient in statistical concepts and techniques for experimental strategies and predictive modeling.

**Data Visualization & Reporting:** Creating dashboards and visualizations to provide actionable insights using Power BI and Tableau. Application Production for ML models.

**Stakeholder Communication:** Clear communication of complex data concepts and effective interaction with cross-functional teams.

### Technical Expertise:

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

**ML/AI Frameworks:** scikit-learn, Keras, Pytorch, TensorFlow, OpenCV, Hugging Face.

**Cloud Technologies:** AWS Cloud, Azure Analysis Services.

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

**Software/Data Engineering:** APIs, Versioning, Logging, ML workflow, GitHub, Docker. Apache Spark

**Database Management:** Proficient in managing large-scale datasets. on-prem and Cloud.

### Professional Experience

#### Ventura Motors, 'Business and Data Analyst' (Data Science and ML)

Sep 2021 – Present

- **Data Architecture and Design**

**Project: Electric Bus Performance and Battery Management System.**

**Trial worth \$10 million.**

Developed a robust data pipeline to efficiently capture and store high-frequency CAN bus data using **Amazon S3** and **AWS Glue**. The pipeline ingests critical metrics such as battery health, energy consumption, vehicle speed, and braking patterns. By seamlessly integrating this data with other relevant sources and applying sophisticated ML/AI techniques, I was able to preprocess, normalize, and transform the noisy data to uncover actionable insights. These insights enable predictive maintenance, optimize route efficiency, and enhance overall fleet management, leading to significantly improved data-driven decisions that drive operational cost efficiency.

**Stakeholders:** Government, private and local bodies, international, and research-based entities.

- **MLOPS Engineering**

**Project: Intelligent Accidents Monitoring and Management System - IAMS.**

**Estimated accident damage over \$3 million per year.**

Developed an ML-powered application called IAMS built on a large dataset. Deployed on **AWS cloud server instance (EC2)**. Implemented supervised (**SVM**, **XG Boost**, and **Random Forest**) and Unsupervised ML models (**DBSCAN**, **K-Means**). Evaluated performance based on relevant metrics. Used **GitHub** and **Docker** for continuous data integration, monitoring, and optimization (CI/CD).

**Team size: 4**

- **Data Architecture and Design for ML/AI Applications**

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

**Contracts worth 96% of business revenue.**

Developed end-to-end data solutions to understand patronage movement using geospatial and Myki data. Utilized unsupervised learning algorithms for public transport service improvements.

**Team size: 4**

**Stakeholders: Department of Transportation & Planning**

- **AI Modelling for Natural Language Processing**

**Project: Customer Satisfaction Analysis**

Used Python NLTK's Tokenization and Lemmatization techniques for text analytics. Employed VADER for polarity scoring and BERT for part-of-speech tagging and named entity recognition(NER).

- **AI Modelling for Optical Character Recognition**

**Project: Image Processing for Paperless Invoice data capture**

Extracting sensitive data from financial remittances in encrypted PDF files and invoices in various image formats. Worked with frameworks like OpenCV, Tesseract, and PIL.

- **Data Engineering and Analytics**

**Project: Vehicle Safety Inspection & Vehicle Off-Road Tracking System**

Modeled data flows and data structure for the Power BI app. Backend data processing using Python. Significantly boosted asset management by 100% compliance for about 1000 vehicles. It offers predictive safety maintenance insights across the business on real-time data.

**Collaborating team size: 4**

- **CyberSecurity - ISO 27001 Framework and Essentials 8**

**Project: Risk Management for Data & Information Security.**

Worked on security tools dataset to define trends utilizing ML practices for 'Threat Intelligence'. Collaborated with government and non-government agencies to improve public data security (resilience, response & recovery).

- **ESG Data Strategy and Engineering**

**Project: Sustainability Standards Framework - Reporting for Corporate Carbon Footprint**

Developed strategic data for tracking corporate emissions and managing reporting frameworks. Pioneered Ventura's Corporate ESG report 2023.

## **Miepol (Pty) Ltd, Data Scientist**

**Sep 2020 – Sep 2021**

Requirement analysis. Developed & deployed two apps called 'Traffic Management & Maintenance System' and 'Spatiotemporal Monitor for Wayfinding'. Implementing statistical analysis in Python, Working with streaming datasets, Data Quality and Integrity - ETL, Business Data Intelligence Strategies, Automation and Pipelines, Data Compliance & Risk Assessments.

## **Teacher – Mathematics, STEM, and Programming in Python /C++**

**Aug 2011 - Oct 2019**

STEM-focused education with project-based programming.

## **LMKR, Analytics Engineer**

**Jan 2005 – Nov 2007**

Geospatial analysis, Spatiotemporal analysis, and Seismic data analysis for oil and gas.

## **Petrosin Engineering, Systems Engineer**

**Sep 2003 – Dec 2004**

## **Education**

### **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, Pakistan**

Relevant Coursework: Computer programming in C++, 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.

## **Community Engagement and Volunteering**

- Glen Eira Strategic Transport Advisory Committee STAC (Active member).
- Conducted Workshops - Aimed at supporting govt's mission: 'Go Electric Plan' (Active member).
- City of Glen Eira Sustainability Champion Aug 23. (Project Name: 'Energy Smart')
- Training teachers for makeshift SOS schools in disaster-stricken areas (4 years)
- Served (SAYA) NFP organizations for orphaned and abused children (3+ years)

\*References available on demand