MASOOD NAZARI

Senior Data Engineer | AI Specialist | FinTech & Regulatory Focus

Southampton, UK | 07938 935499 | Michaelnazary@gmail.com

LinkedIn   GitHub   Portfolio

PERSONAL PROFILE

Data Scientist & Analytics Engineer (MSc Distinction) with 5+ years of experience transforming complex datasets into strategic assets. Currently leading high-level data engineering initiatives at a major UK Clinical Trials Unit. Expert in deploying **Python/R pipelines**, **API-driven architectures**, and **Machine Learning models**. Bridging the gap between clinical operations and technical data science to deliver measurable efficiency gains (reducing manual workloads by **70%+**).

TECHNICAL SKILLS

> **Languages:** Python, R, SQL (PostgreSQL/T-SQL), Bash, SAS.
> **Engineering:** API Integration, ETL pipelines, Docker, Google Earth Engine, Azure/GCP.
> **Machine Learning:** Scikit-learn, XGBoost, Prophet, SHAP, LIME, Neural Networks (BERT, SpaCy).
> **BI & Visualisation:** Power BI, Tableau, R Shiny, Streamlit, Excel (Advanced).
> **Compliance:** Regulatory Compliance Frameworks (GDPR/GxP, Transferable to FCA/Basel III), Metadata Governance.

PROFESSIONAL EXPERIENCE

Clinical Data Coordinator
Southampton Clinical Trials Unit (SCTU) | Academic Research Centre

Sep 2024 – Present
Southampton, UK

Acting as the technical lead for data automation and reporting infrastructure across the Cancer Vaccine Launch Pad (CVLP).

> **Architected an Event-Driven Safety Validation Microservice (Python/FastAPI):** Designed a RESTful API ecosystem to automate Adverse Event (AE) reconciliation. Implemented strict schema validation that **reduced data ingestion errors to near-zero (<0.1%)** and flagged anomalies for immediate review.
> **Established a Real-Time Analytics Platform:** Engineered a containerised reporting layer (Streamlit/Docker) integrated with automated ETL pipelines. Reduced data latency from weekly to **near real-time**, enabling Trial Managers to make data-driven decisions 6x faster.
> **Governance-as-Code & Regulatory Compliance:** Implemented rigorous data lineage tracking within pipelines to satisfy **GDPR/GxP audits**. Created immutable audit logs that reduced QA retrieval time by **80%**, demonstrating a framework directly transferable to **FCA/Basel III standards**.
> **Technical Strategy & Infrastructure Audit:** Led a critical architectural review of 3rd-party integration tools. Identified OAUTH authentication vulnerabilities and authored a remediation roadmap that secured data access for 70+ external sites.

Data Scientist (Contract)
SC Solutions | Environmental Analytics

Jun 2024 – Sep 2024
Remote/Hybrid

> **Geospatial Pipeline Architecture:** Architected automated ETL pipelines using Google Earth Engine and Python APIs to ingest multi-terabyte satellite data. Reduced data preprocessing latency by **80%**, accelerating model iteration from days to hours.
> **ML Model Performance:** Developed the BioVision Analytics Hub (Streamlit/Plotly), utilising ensemble learning to reduce Root Mean Square Error (RMSE) by **15%** for real-time biomass prediction.
> **Commercial AI Implementation:** Scaled biomass prediction models across 10,000+ hectares, improving carbon stock estimation accuracy by **20%**. Integrated Explainable AI (SHAP) for audit-ready insights.

Senior Data Analyst & Accountant
Parsian E-Commerce Co. | Fintech

Oct 2019 – Sep 2023
Tehran

> **Fraud Detection System:** Engineered Machine Learning models for transaction anomaly detection, improving fraud capture rates by **26%** and reducing financial liability.
> **Financial Workflow Automation:** Designed SQL and Python-based reconciliation scripts to replace manual ledgers. Reduced manual handling time by **43%**, cutting the month-end closing cycle by **2 days**.
> **Strategic BI:** Built executive-level KPI dashboards to identify revenue bottlenecks. Increased decision-making efficiency by ∼27% by providing real-time visibility into cost drivers.

Sales Accountant
MaximTaxi | Private Transport

Jun 2018 – Sep 2019
Tehran

> **Data Integrity:** Leveraged SQL (PostgreSQL) to query large transactional datasets, identifying financial discrepancies and improving supplier payment accuracy by **23%**.
> **Reporting:** Developed automated queries for tax and compliance reporting, reducing reliance on manual Excel checks and minimising errors for regulatory audits.

TECHNICAL ENGINEERING & ARCHITECTURE

InsurePrice: Actuarial Risk Engine *| Open Source Development*

2024

- **Tech Stack:** Python, CatBoost, SHAP, Terraform (IaC), Azure Container Instances (ACI), FastAPI.
- **Cloud Architecture:** Deployed a scalable pricing engine on **Azure** using **Terraform** for Infrastructure-as-Code. Containerised the application with Docker to ensure reproducible deployments across dev/prod environments.
- **InsurTech Innovation:** Engineered a **Monte Carlo 'Digital Twin' simulator** to model driver risk across 10,000 commute scenarios. Developed a parametric 'IoT Maintenance Bond' prototype that triggers automatic payouts based on real-time vehicle telematics.
- **DevOps & Security:** Implemented a modular code structure with **Type Hinting** and automated testing. Designed the architecture to handle **Category-Level embeddings** natively, improving risk segmentation accuracy by 3.8% over baseline models.

Fraud Detection Dashboard API | *Open Source Development* 2024

- **Tech Stack:** Python, FastAPI (Async), NetworkX, Docker, Pydantic, Pandas.
- **System Architecture:** Engineered a high-concurrency **Async FastAPI backend** processing **5 million+ transactions** with sub-100ms latency. Implemented Pydantic schema validation to ensure strict type safety for financial data.
- **Risk Engine:** Deployed graph theory algorithms to detect **Money Mule accounts** and circular fraud rings. Integrated **Explainable AI (XAI)** to provide transparency on transaction rejections, ensuring compliance with banking regulatory standards.
- **Operational Intelligence:** Built a 5-tier dashboard hierarchy (Executive to Investigation) featuring a **Fraud Velocity Heatmap** and real-time smart alerts. Optimised memory usage for high-throughput data loading using efficient Pandas dtypes.

Bank Marketing Analytics Platform | *Open Source Development* 2024

- **Tech Stack:** Python, React (Frontend), Flask (REST API), Streamlit, Scikit-learn (SMOTE).
- **Architecture:** Designed a **dual-interface deployment strategy**: built a Streamlit app for rapid prototyping and a decoupled **Flask/React architecture** for production, serving real-time predictions via 7 REST endpoints.
- **FinTech Impact:** Engineered a 'Call Frequency' optimisation engine that reduced customer fatigue, projecting a **£170k annual ROI**. Integrated macroeconomic indicators (Euribor 3M, CPI) to model campaign performance under varying economic conditions.
- **Advanced Segmentation:** Applied **K-Means Clustering** and PCA dimensionality reduction to identify high-value customer personas. Implemented **SMOTE** to handle class imbalance, boosting minority class detection recall to 60%+.

NIHR Research Intelligence Dashboard | *Open Source Development* 2024

- **Tech Stack:** Python, Streamlit, Scikit-learn, Pandas, Plotly.
- **Strategic Impact:** Architected a portfolio intelligence platform managing **£171.6M in funding**. Delivered a verified **122:1 ROI** (£116k+ annual savings) by automating parliamentary reporting workflows.
- **Data Engineering:** Engineered a data quality pipeline using **Fuzzy Matching algorithms** to detect duplicate awards (5.6% rate) and enforce NIHR business rules, increasing dataset completeness score to 73.6%.
- **Custom Analytics Logic:** Developed proprietary algorithms for 'Market Opportunity Analysis' and 'Success Rate Benchmarking'. Created a parliamentary-ready UI with custom CSS/HTML components for high-stakes stakeholder presentations.

BioVision Analytics Hub | *Open Source Development* 2024

- **Tech Stack:** Python, LightGBM, XGBoost, Sentinel-1/2 APIs, Streamlit, Mapbox.
- **Big Data Engineering:** Architected an automated ETL pipeline to ingest and merge **multi-terabyte satellite imagery** (NASA GEDI, Sentinel SAR/Optical). Optimised data processing latency by implementing parallelised feature extraction for spectral indices.
- **Spatial Intelligence:** Developed an interactive geospatial dashboard with **K-Means Clustering** and Hotspot Analysis (Local Outlier Factor) to identify biomass patterns. Deployed ensemble ML models (LightGBM/RF) achieving **0.89 R^2** accuracy for forest ecosystem monitoring.

EDUCATION

MSc Data & Decision Analytics (Distinction) Sep 2023 – Sep 2024
University of Southampton, UK

- **Core Focus:** Advanced Machine Learning, Stochastic Processes, Game Theory, Operational Research.
- **Research:** Achieved Distinction for dissertation on "AI-driven geospatial forecasting" (deployed as the BioVision project).

BSc Materials Engineering (Distinction) Sep 2010 – Sep 2014
Semnan University

- **Key Achievement:** Ranked top 5% of cohort; specialized in computational modeling and simulation.