Farrukh Nizam Arain

Phone: 07778 736005 | Location: Manchester, UK | Email: farrukh.arain@outlook.com
LinkedIn: linkedin.com/in/farrukh-nizam-arain | Portfolio: farrukh-nizam-arain.github.io/

PERSONAL STATEMENT

Senior Data Scientist with over 10 years of experience in delivering analytics solutions and business impact, now enhanced by a recent MSc in Data Science. Combines a strong track record in model development, data strategy, and stakeholder management (grounded in BSc Computer Science and MBA Finance) with freshly validated expertise in deep learning, NLP, and advanced machine learning techniques. Proven ability to lead projects from conception to deployment, translating complex analysis into commercial value. Seeking a senior or lead role to apply this updated, state-of-the-art technical skill set alongside deep industry experience to solve challenging business problems.

TECHNICAL SKILLS

Programming & Databases: Python, SQL, T-SQL, R, SQL Server, SSMS, MySQL, PostgreSQL
ML & Modelling: Scikit-learn, XGBoost, Random Forest, GBM, Logistic Regression, SVM, PCA, K-Means
Deep Learning & Vision: PyTorch, TensorFlow, CNN, ResNet, Transformers, PaddleOCR, YOLOv8
Data Engineering & ETL: Power Query, Power Pivot, EDA, Feature Engineering, Statistical Analysis
Python & Web Dev: NumPy, Pandas, Matplotlib, Seaborn, Plotly, Flask, Streamlit, Git
Explainable AI (XAI): SHAP, LIME, Grad-CAM, Model Interpretability
Visualization & BI: Power BI (DAX), Tableau, Excel (Advanced), Data Storytelling
Tools & Platforms: Jupyter, VS Code, Warehouse Management Systems (WMS)

WORK EXPERIENCE

Data Analyst | GXO Logistics, Inc. (Greene King Contract) | Manchester, UK | Jan 2026 – Present

- Develop and maintain the Demand Planning Dashboard using Power BI and Tableau, transforming supply chain data into daily/weekly insights for planners and senior management to optimise inventory and forecasting.
- Automate a high-volume suite of 15+ operational and KPI reports (e.g., OTIF, Supplier Performance) via SQL and advanced Excel, reducing manual effort, minimising errors, and ensuring timely delivery for strategic decision-making cycles.
- Perform root-cause analysis on supply chain performance metrics (stock-outs, order adherence), utilising statistical techniques to identify trends and present data-driven recommendations that inform process improvements.
- Manage end-to-end data integrity for critical planning systems by building and maintaining SQL/Access databases, ensuring reliable data flows for reporting and ad-hoc analytical requests across the department.

Administrative Assistant – Demand Planning | GXO Logistics, Inc. | Manchester, United Kingdom | Jun 2025 - Present

- Engineered automated ETL pipelines using Power Query and Power Pivot to consolidate raw data from multiple sources, forming the reliable foundation for all daily operational and KPI reporting.
- Built and maintained daily supply chain dashboards in Excel and Power BI (Intake, Stock Levels, Delivery Issues), providing the planning team with actionable insights into inventory and logistics bottlenecks.
- Ensured data integrity within the Warehouse Management System (WMS) by accurately managing purchase order closures, goods receipting, and complex datasets for reporting.
- Automated the generation and distribution of key daily reports (e.g., Unbooked POs, GNFR Demand Forecast), streamlining communication with internal teams and external suppliers and saving significant manual effort.

Research Project Assistant | Manchester Metropolitan University | Manchester, United Kingdom | Jul 2024 – Jun 2025

- Engineered a clinical AI platform for early lung cancer detection using miRNA biomarkers and Random Forest models, achieving 92.9% accuracy and 94.4% sensitivity to address the high false-positive rate of traditional screening methods.
- Optimised model efficiency and interpretability by implementing RFECV, reducing critical features by 82% (16 to 3 biomarkers) while maintaining a 91.4% cross-validation score, streamlining the path to clinical deployment.
- Developed a multi-modal food waste reduction platform, integrating a PyTorch ResNet34 image classifier (98% accuracy) and a PaddleOCR receipt parser (94.8% accuracy) with a PostgreSQL-backed recipe recommendation engine.
- Delivered scalable, end-to-end applications by deploying trained models via Streamlit and Flask APIs, featuring automated reporting and secure data handling to facilitate handover and stakeholder decision-making.
- Contributed to research papers and grant reporting for the UKRI Innovate UK programme, detailing methodological innovations and project outcomes that supported continued funding and commercial interest.

Data Analyst | Feb 2010 – Sep 2023

Securebeans | Karachi, Pakistan

- Served as the primary data analyst for a cybersecurity consultancy, conducting stakeholder interviews and translating complex business challenges across diverse clients into data-driven strategies and process improvements.
- Engineered a time-series analysis model for a restaurant client (Pizza Palace) using T-SQL and SSMS, identifying underperforming menu items and seasonal trends to project a 7% increase in quarterly profitability.
- Built a sales-inventory model for Maven Toys using Power Pivot/Query, showing 50% of profits came from downtown stores and a $300K stock imbalance, prompting a strategic stock reallocation.
- Ran an end-to-end investment analysis pipeline, cleaning and analyzing 1,000+ "Unicorn" company datasets to pinpoint accelerated AI sector growth, directly shaping the firm's capital allocation roadmap.
- Led and mentored junior team members, fostering effective project contributions, while designing and presenting insights through interactive Tableau dashboards and reports to senior stakeholders to drive decision-making.

ACADEMIC AND TRAINING PROJECTS

Dynamic Airfare Prediction System | *MSc Dissertation* | 2024

- Engineered a Stacking Ensemble model on a dataset of 300,000+ flight entries to predict ticket prices with high accuracy (R^2: 0.994, MAPE: 10.85%).
- Conducted in-depth SHAP analysis to identify and explain key pricing drivers (e.g., advance purchase, airline), translating complex model logic into actionable business insights.
- Deployed the trained model as an interactive, end-to-end web application using Streamlit, demonstrating full-stack ML deployment capabilities.
- Tech: Python, Pandas, Scikit-learn, Stacking Ensemble, SHAP, Streamlit

Multi-Class Food Image Recognition System | *Deep Learning Module* | 2024

Manchester Metropolitan University | Manchester, United Kingdom

- Developed an 11-class image classifier using PyTorch, achieving 94.2% test accuracy on a dataset of 17,000+ images via transfer learning with pre-trained CNNs (ResNet-34, DenseNet121).
- Implemented advanced techniques including data augmentation for robustness and model explainability using SHAP and Grad-CAM to visually interpret prediction drivers.
- Containerised the final model and built an accessible Streamlit application for real-time image upload and classification.
- Tech: PyTorch, Torchvision, CNN, Transfer Learning, Grad-CAM, SHAP, Docker, Streamlit

AutoTrader's Car Price Predictor | *Machine Learning Module | 2023*

- Built a complete machine learning pipeline to estimate used car prices from 400,000+ advertisements, with the best ensemble model (Voting/Stacking Regressor) achieving an R^2 of 0.877.
- Applied SHAP analysis to determine and communicate the most significant price drivers (mileage, vehicle condition), providing transparent, explainable AI outputs.
- Operationalised the model by creating and deploying a functional Streamlit web application, showcasing the practical application of ML solutions.
- Tech: Python, Pandas, Scikit-learn, Ensemble Methods, SHAP, Streamlit

EDUCATION

MSc Data Science (Distinction) | Manchester Metropolitan University | 2024

- **Relevant Modules:** Principles of Data Science, Machine Learning Concepts, Advanced Machine Learning, Deep Learning, High Performance Computing and Big Data, Data Management and Governance, Computational Maths and Statistics, Algorithms and Data Structures.
- **Dissertation:** *Dynamic Airfare Prediction System* – Developed a predictive model for flight pricing using time-series analysis and machine learning techniques.

MBA (Finance) (Distinction) | Bahria University | 2010

- **Key Modules:** Strategic Management, International Business Analysis, Business Quantitative Techniques, Statistical Inference, Methods in Business Research.

BSc Computer Science (Software Engineering) (2:1) | Karachi Institute of Economics and Technology | 2007

- **Core Modules:** Data Structures, Database Management Systems, Artificial Intelligence and Expert Systems, Data Warehousing and Decision Support Systems, Elements of Statistics and Probability.

PROFESSIONAL CERTIFICATIONS

- **Google Data Analytics Professional Certificate** | Google | 2022
- **IBM Data Science Professional Certificate** | IBM | 2021
- **Python for Everybody Specialization** | University of Michigan (Coursera) | 2023
- **Tableau Certified (Desktop Fundamentals, Intermediate, Advanced & Prep Builder)** | Tableau | 2021

AWARDS AND ACHIEVEMENTS

- **Nominee, CDI Spring Awards –** 'Best AI Innovation in Healthcare' & 'Best Innovation in Operational Efficiency' | Manchester Metropolitan University | 2025
- **Finalist, Autotrader Data Science Hackathon** | Manchester Metropolitan University | 2024

ADDITIONAL INFORMATION

- **Languages:** English (Fluent), Urdu (Fluent), German (Basic)
- **Right to Work in the UK:** Graduate Route Visa valid until March 2027
- **Driving Licence:** Full clean UK driving licence
- **Interests:** Mentoring junior data analysts; hiking and landscape photography across UK national parks.

REFERENCES

References available upon request.