MARK DUNBAR

SENIOR DATA SCIENTIST

+44 7913 716 076 | MARK-JAMES-DUNBAR@OUTLOOK.COM | GITHUB | LINKEDIN

SUMMARY

Data Scientist with 5+ years' experience delivering ML systems across government, telecoms, finance, and the public sector. Specialises in deep learning and ML engineering, signal processing, and time-series forecasting, with strong expertise in deploying production pipelines using TensorFlow, PyTorch, Kafka, and cloud platforms (AWS, GCP, Azure). Skilled at translating research into robust, operational tools and building scalable, containerised services (Docker, Kubernetes, Airflow). BSc in Physics and MSc Data Science.

PROFESSIONAL EXPERIENCE

DAINTTA | SENIOR DATA SCIENTIST

Mar 2024 – present

Productionized two underwater acoustic classification systems (TensorFlow/Keras, advanced digital signal processing and fourier analysis, Kubernetes) for field trials, designed to augment human analysis and decision-making with interpretable, explainable classifications, supplemented with outputs such as logits and latent layer saliency visualisations.

Led an extremely lightweight team to develop a computer vision model for underwater object detection alongside an analytical platform that enabled operators to manually label training data, validate predictions, and score imagery. Leveraged high volumes of open-source side-scan sonar data to achieve higher accuracy during live trials.

Provided forecasting subject matter expertise for a strategic demand analysis for HM Coast Guard, designing a data ingestion and preprocessing platform capable of handling multiple APIs and data formats. Worked with operators and human factors specialists to create a hierarchical weighted demand index and 5-year forecasting model (DuckDB, AWS, Prophet), directly informing national resource planning.

Expanded and mentored the data team, successfully hiring two professionals and performing line management duties. Hosted monthly Data COI meetings to foster cross-domain collaboration, knowledge sharing, and stronger stakeholder buy-in across the organisation.

CONVERA | DATA SCIENTIST

Jan 2023 – Mar 2024

Developed end-to-end machine learning projects including hierarchical revenue forecasting (Prophet) and churn prediction models, applying advanced time-series preprocessing and statistical analysis to large-scale transactional datasets.

Deployed and maintained ML pipelines on AWS SageMaker and Snowflake Snowpark, integrating with Redshift and Snowflake for scalable training, inference, and downstream analytics.

Engineered AWS migration strategy for ML infrastructure, working with solution architects to design reproducible environments and ensure smooth transition of existing models.

Optimised data pipelines and model outputs for intuitive, real-time executive dashboards (Tableau, Oracle BI), expediting the use and visibility of financial KPIs and enabling data-driven decision-making at the senior leadership team level.

SKY | DATA SCIENTIST

Aug 2021 – Jan 2023

Deployed real-time Kafka streaming pipelines to ingest and process telemetry from distributed edge devices, improving visibility into CDN health and enabling proactive monitoring of network issues.

Built network health monitoring models using streaming telemetry data, increasing operational insight for EU/US engineering teams, significantly reducing time-to-detection KPIs for critical incidents.

Analysed Sky Glass rollout data, identifying DHCP and signal quality issues, which informed proactive monitoring strategies and reduced customer-impacting incidents.

Developed performance monitoring dashboards for Sky Glass devices, enabling rapid root-cause analysis and cutting time-to-resolution for technical teams.

Optimised Kubernetes-based ETL pipelines on Google Cloud Platform, lowering operational costs and improving resource utilisation for large-scale data processing.

PROJECTS

ARRHYTHMIA CLASSIFICATION FOR ICD DEVICES

ML pipeline for ECG anomaly detection using feature extraction/selection, improving classification accuracy and demonstrating transfer of DS skills into healthcare. The codebase is featured in my Github. A brief 10-minute presentation I conducted on the project and its findings can be found on my Linkedin.

INTERACTIVE CNN MODEL & WEB APP WITH MLOPS

Developed PyTorch CNN webapp with containerised deployment, automated model versioning, and monitoring dashboard for real-time predictions. Showcases full MLOps lifecycle. Check out the code here, and the deployed application on huggingface.

RAG-LLM APPLICATION

Developed FastAPI-based document intelligence API (Ollama, ChromaDB) with PDF ingestion, semantic search, and per-user isolation. Demonstrated scalable retrieval-augmented generation. See the code for this project here.

EDUCATION

MSC DATA SCIENCE

Queen Mary University of London
Sep 2021 – Sep 2022

BSC PHYSICS

University of Glasgow
Sep 2015 – Sep 2019

CORE SKILLS

Languages & Frameworks: Python, R, SQL, TensorFlow, PyTorch, scikit-learn, Prophet, ARIMA, OpenCV
MLOps & Deployment: Docker, Podman, Kubernetes, Airflow, MLflow, AWS SageMaker, FastAPI
Cloud & Data Platforms: AWS, Azure, GCP, Snowflake/Snowpark, DuckDB, Redshift
Streaming & Data Engineering: Kafka, ETL pipelines (Airflow, Python), API ingestion
Specialisms: Signal Processing (Fourier Analysis, Digital Signal Filter Design), NLP (transformers, embeddings)