

Andy Fanthome

Data Scientist | Computer Science BSc

Phone: +447809277312 Address: Manchester, UK E-mail: fanthome.andrew@gmail.com

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

SUMMARY

A naturally curious and self-motivated data scientist with end-to-end data science experience including exploratory data analysis, model development, model deployment, and software engineering. Consistently worked with stakeholders company-wide to deliver high quality outcomes. Proficient in Python (pandas, sklearn, matplotlib, keras, scipy, seaborn), SQL, Azure, and AWS services.

RELEVANT WORK EXPERIENCE

ZALLY

January 2024 - Present | Manchester, UK

Data Scientist

- Increased machine learning model accuracy from 60% to over 90% by conducting over 200 recorded experiments with various setups, parameters, and models on datasets containing 450,000+ rows per user across more than 100 users.
- Developed an automated, end-to-end pipeline using **Python**, **CatBoost**, and **AWS S3** to label and filter irrelevant user data that made up 60% of all data.
- Built and deployed an interactive data science **dashboard** in **Python** hosted on **AWS EC2**, successfully delivering KPIs, visualizing data insights, and bridging the technical gap between the data science team and non-technical stakeholders.
- Created documentation and formatted code to **PEP8** standards for 54 data science modules, improving codebase maintainability.
- Improved scalability by using **Apache Airflow** to distribute tasks to run in parallel, decreasing total training time by 90%.
- Gained proficiency in key **AWS services** (**S3**, **SageMaker**, **DynamoDB**, **EC2**) and deployed machine learning models to production using these tools.
- Presented as a panellist at company events and engaged in data science forums, enhancing public speaking skills and professional development.
- Created and maintained unit tests for the data science codebase, always ensuring a minimum of 80% code coverage to enhance code reliability and maintainability.
- Branched, merged, reviewed, and deployed code using **Git** within **Azure DevOps**, supporting efficient team collaboration and code management.

NHS ALDER HEY INNOVATION

October 2021 - September 2022 | Liverpool, UK

Junior Data Scientist - Year in Industry University Placement

- Researched and developed a tool to flag patients at risk of readmission using **Random Forest** and **CatBoost** models to achieve an ROC of 0.85 that outperformed the results of other studies by 7.6%.
- Used **SQL** and **Python** to conduct data extraction, exploratory data analysis and data preprocessing on datasets of over 160,000 patient records from the last 5 years.
- Visualised model results in **Microsoft PowerBI** so that end users could view hundreds of patients' readmission risk scores weekly.
- Created an automated tool that evaluated models, producing rolling time-series graphs to highlight downward trends, saving over 1 hour per week previously spent on manual testing of each model.
- Followed Agile for collaborative, iterative development to deliver ahead of schedule over 2-week sprints and daily scrums.

THOMSON REUTERS

Summer 2018 | London, UK

Software Engineer - Summer Work Experience

- Automated a manual process in **Python** that split specified **Hadoop Hive** tables which saved 30 minutes of employee time daily and reduced human error.
- Designed a script using **RegEx** that compared **Kafka** topics in an output stream with a **YAML** file and automatically made missing topics in the output stream, saving employees over 2 hours per week.

EDUCATION

UNIVERSITY OF LIVERPOOL

September 2019 - July 2023 | Liverpool, UK

BSc (HONS) in Computer Science with a Year in Industry | Grade: First Class

Relevant Subjects: Big Data Analytics, Data Mining & Visualisation, Advanced Artificial Intelligence, Data Science, Software Engineering 1 & 2, Computer Based Trading, Data Structures & Algorithms

REIGATE COLLEGE

September 2016 – July 2018 | Reigate, UK

Mathematics A-Level: **B**, Computer Science A-Level: **B**, Chemistry A-Level: **B**, Media Studies AS-Level: **A**, EPQ: **A**

OTHER PROJECTS

FINAL YEAR PROJECT

- Built a chatbot web application deployed on **Google Cloud Run** using **HuggingFace** transformers, **Flask**, and **Python** – achieving BLEU and NIST scores of 0.342 and 1.046, respectively.
- Evaluated the web application on twenty users, showing 95% had an overall good experience.

KAGGLE COMPETITIONS / PERSONAL PROJECTS

July 2023 - Present

- Developed a binary classifier trained on 16,500 rows to flag people at risk of heart attack, achieving 92% Accuracy.
- Performed EDA of global covid-19 data, data cleaning/pre-processing, and conducted **ARIMA/SARIMAX time-series forecasting** of UK specific data.

NON-TECHNICAL EXPERIENCE

UNIVERSITY SOCIETY PRESIDENT

September 2021 – September 2022 | Liverpool, UK

- Organised venues for social events, recruited over 400 members, delegated tasks, completed risk assessments, ran a digital community on Instagram with over 650 followers, and encouraged harmony within the 3-person committee.

TECHNICAL SKILLS

Programming Languages: Python (Proficient), SQL (Proficient), HTML (Proficient), CSS (Intermediate), Java (Intermediate), R (Beginner)

Software: Jupyter Labs, AWS (S3, SageMaker, Lambda, DynamoDB, EC2), Apache Airflow, MySQL, SQL Server Management Studio, Power BI, Microsoft Excel, R Studio, Docker, Flask, PySpark, Keras, TensorFlow, Git, Azure, Jira, Notion, HuggingFace.

Technical: Classification, Statistical Analysis, Data Analysis, Data Cleaning, Feature Engineering, Feature Selection, Data Quality Checks, Experimentation, Automated Reporting, Model Tuning, Agile Methodologies.