TOMMY NEELD

Lead Data Scientist

tommy.neeld@gmail.com 07812846259 London, UK LinkedIn GitHub

Technical Skills

Data Science: Python (scikit-learn, Pandas, Numpy, PySpark, Kedro), MLOps (SageMaker, MLFlow), NLP (PyTorch, TensorFlow, Sentence Transformers, LangChain, SetFit), Forecasting (ARIMA, Prophet, RNN, DeepAR)

Engineering: Python (Django, FastAPI, Scrapy), Ionic, Angular, C, Cloud (AWS, Azure), Git, Docker, CICD (Azure, GitLab, Terraform), DB (PostgreSQL, MongoDB), VectorDB (Weaviate), Kubernetes (EKS, Helm), Data Pipelines (Serverless - Step Functions, Lambda, EventBridge, Fargate; Other - Kedro, Airflow)

Work Experience

London Stock Exchange Group

Lead Data Scientist, Innovation Labs — London, UK (March 2021 - Present)

I lead a team of data scientists within LSEG's 'Innovation Lab' function. I work across the organisation on a project-by-project basis. Within a multi-disciplinary consultancy style team, we take ideas from exploration through to implementation. I've worked with the FTSE team developing indices, our Post Trade team predicting repo failure rates, the Yield Book team developing fixed income analytics, and Sustainable Finance teams improving ESG scoring methods.

Relevant projects:

- **Market microstructure modelling** - Developed market impact and volume forecasting models (across several venues) to help equity traders understand interday market conditions using our in-house Execution Management System. This involved testing Deep Recurrent Neural Network architectures, including Amazons' DeepAR, alongside auto regressive and moving average benchmarks. Using custom features, I developed a neural network-based market impact model that consistently outperformed traditional parametric techniques (e.g., Kissell I-Star). I deployed the model and supporting infrastructure on AWS - utilising Amazon SageMaker, DynamoDB, Lambda and API Gateway - and built a frontend to showcase using Plotly Dash. The model is trained on approximately one year of rolling level 1 tick data queried from GCP's BigQuery data warehouse.

- **Thematic Index generation** - Collaborated with FTSE researchers to develop methods of generating Thematic Indexes based on any input theme description. Tested various data sources including company fundamentals, annual reports, company descriptions, business classifications etc. to develop an approach. Researched a range of NLP modelling techniques, including lexical search (BM25), along-side more advanced large language models such as BART zero-shot and pre-trained deep learning models like Google's USE. Deployed the best-performing model, namely a fine-tuned sentence transformer trained in a contrastive Siamese manner. Built an end-to-end solution which included a Weaviate vector search engine deployed on a Kubernetes cluster with a simple font-end for low-latency index generation. Wrote a research paper and currently supporting architects to productionalise.

- **Fixed Income analytics (ongoing)** - Supervising a member of my team to develop methods of estimating the potential premium associated with green bonds i.e., 'greenium'. We are developing this analytic using techniques such as propensity score matching. I am reviewing code and proposed methods. Once built, users will be able to better compare ESG bonds on primary and secondary markets.

EY

Lead Data Scientist, Emerging Technology — London, UK (April 2019 - March 2021)

Working included EY's Emerging Technology team, I focussed on building tools to help clients. As the founding data scientist, I helped establish the function and, in-time, led a team. My primary focus was on building a singular product from the ground up; this was a complex web-scraping regulatory intelligence application which I built working directly with clients - a tool that remains a significant revenue stream for the team.

Modules of the product:

- A Django application with asynchronous task execution using Celery and Redis deployed on Azure App Services with PostgreSQL Hyperscale.
- Scheduled scraping of white-listed sites and APIs using Scrapy and an NLP pipeline for scraped content to identify topics, clean and classify text.
- Client-specific relevancy modelling that uses human-in-the-loop and active-learning to order news by importance.
- Required user auth, account management, UI, CI/CD and model training pipelines.

Amey Strategic Consulting

Senior Consultant / Data Scientist — London, UK (April 2018 - April 2019)

Operating as a senior consultant within a data science centric offering, I would work to identify potential projects, develop proposals, and then work on client-site.

Relevant projects:

- **IoT asset management application** - An application that consumes readings from thousands of IoT sensors spread across physical assets such as bridges; with an architecture which utilises Kafka, Cassandra and Flask. I developed time-series forecasting models to predict behaviour under adverse weather conditions and cluster-based anomaly detection algorithms to alert Engineers to unusual events.

- **Big data text classification** – Working on forum data, I developed models to determine country of origin using NLP techniques. This work required use of John Snow Labs Spark NLP to manage large data volumes.

IBM

Technology Consultant, Graduate Scheme — London, UK (Sept 2011 - Sept 2013)

Joined the graduate scheme at IBM as a technology consultant. Primarily operated as a business analyst and developer working with SAP ERP solutions on client site.

Education

UCL, PhD (2014–2018)

- Cross-disciplinary PhD in Energy Demand Studies (Physics and Computer Science).
- Developed a new technology to improve gas combustion efficiencies using Machine Learning.
- Initiated two lucrative collaborations with Imperial College and Bosch - consulted for Bosch for a period.

Bath University, Physics Masters (2006–2010)

- First Class Honours specializing in computational Physics - courses in C and MatLab.
- Created a Monte Carlo simulation (in C) simulating a complex nano-fabrication phenomenon.

Personal Projects

FlockFinder

CTO and Co-Founder — Kent, UK (2020 - 2022)

During lockdown I built a product that provides farmers with an application to help them meet the burdensome administrative duties associated with livestock management.

- I scoped the original use-case and built the application from the ground up; a cross-platform Ionic app with a Django and PostgreSQL backend and various 3rd party integrations including Stripe, Firebase and the GoogleMaps API.
- The business is generating revenue and has been recognised for innovating in the agricultural sector; the application has won awards and been featured in the press.