

NATHANIEL FORDE | RESUME

Experienced Data Scientist / Statistician

📍 Dublin, Ireland
✉️ nathanielf.github.io

📞 +353 86 052 5426
✉️ nathaniel.forde@gmail.com

Short Profile

Expert Statistician/Data Scientist with 10+ years' experience delivering ML products - ranging across probabilistic modelling to causal inference in high-growth tech and regulated industries. Skilled in leading teams, scaling data platforms, and translating R & D into business impact. Open-source contributor (PyMC, Bambi, CausalPy, PyMC-Marketing) and experienced in experimentation strategy, MLOps, and customer-facing analytics.

Core Skills

- ML/Stats: Probabilistic AI, Causal inference, Bayesian methods, Time-series, Survival analysis, Factor analysis, Deep Learning, Survey Analysis, Marketing Analytics
- MLOps & Cloud: AWS (SageMaker, Lambda, ECS, S3), ML pipelines, CI/CD, Docker, Spark, Statsig, Snowflake-ML
- Data Tools: scikit-learn, PyMC, Pandas, dbt, Pytorch, Bambi, CausalPy, statsmodels
- Leadership: Team growth (0→9), cross-functional influence, experimentation strategy, mentoring

Experience

Staff Data Scientist - Personio (HR SAAS)- Dublin

April-2024-Present

- * Scaled the data science org from 0→9 scientists, establishing hiring processes, best practices, and technical roadmaps.
- * Built and Designed Proactive Insights and Engagement Survey products, driving product stickiness and upsell opportunities.
- * Designed and rolled out in-house experimentation framework; A/B testing adoption increased 3x, and test velocity improved 40% YoY after Statsig integration.
- * Led long-term forecasting and MMM models (revenue/costs) that improved planning accuracy by 15%, directly impacting company OKRs.
- * Mentored junior DSs and set standards for documentation, modelling, and productionization of ML workflows. Adopted Manager role during Paternity cover period.

Senior Data Scientist - Personio (HR SAAS)- Dublin

Nov 2021-April 2024

- * Built Personio's first end-to-end ML deployment pipeline (Python/AWS), now serving 10+ live models.
- * Developed survival models (time-to-failure) used for reliability prioritization; reduced error by 5%.
- * Performed root-cause analytics for UX latency issues, resulting in 35% improvement in customer-facing performance metrics.
- * Led experimentation with LLMs and NLP pipelines for customer feedback analysis, enabling automated diagnostics at scale.

Data Scientist (R & D) - Cartrawler (E-commerce)- Dublin

Nov 2019-2021

- * Designed dynamic pricing and demand forecasting models improving margin 4% and reducing loss ratios 8% YoY.
- * Built customer segmentation based on price sensitivity that fed into real-time pricing strategies.
- * Developed interlocking claim-propensity and loss-ratio forecasting models for insurance products.
- * Published work on Discrete Choice modelling for consumer preference
- * Tech stack: Python, PyMC, scikit-learn, Spark, in-house ML pipeline infrastructure.

Lead Insights Analyst - Paddy Power/Betfair (Gaming)- Dublin

Feb 2018-2019

- * Created NLP models predicting sentiment on customer chat data;
- Built propensity models for Responsible Gambling interventions, influencing C.5M in budget allocation.
- Partnered with industry think tank on predictive models for problem gambling; findings shaped regulatory strategy.
- Team Lead and Supervisor to Junior Analysts; Responsibilities included annual planning and performance reviews

Software Developer - Marsh & McLennan - Innovation Centre (Insurance)- Dublin

Mar 2017-2018

- * Developed web-apps (Go/Postgres/React) to deliver MCMC risk prediction models (earthquakes, floods, fires) to underwriters.

Data Analyst - Marsh & McLennan - Innovation Centre (Insurance)- Dublin

Mar 2014-2017

- Built statistical metrics and dashboards improving data quality 15% across 20+ global business units.
- Developed CRUD app (Shiny) for customer QA and spearheaded the company's first data quality framework.

Selected Publications and Open Source Work

Bayesian Vector AutoRegressive models of Irish GDP - PyMC- Open Source

Dec 2022

- Developed hierarchical Bayesian VAR & forecasting models; https://www.pymc.io/projects/examples/en/latest/time_series/bayesian_var_model.html

Justifying Instruments in IV Designs - CausalPy- Open Source

Jun 2024

- Core Docs on IV methods; https://causalpy.readthedocs.io/en/latest/notebooks/iv_weak_instruments.html

Education

Msc Logic - University of Amsterdam: Institute of Logic, Language and Computation-Amsterdam, Netherlands

2011-13

- Research thesis: models of causal inference and dependency relations.

MPhil (Research) Philosophy - Trinity College Dublin- Dublin, Ireland

2009-11

- Modal logics & identity; awarded IRCHSS Scholarship.

BA International - UCD/Sorbonne Université- Dublin Ireland & Paris, France

2005-09

-

Leadership and Community

PyCon & PyData Speaker - PyCon Ireland & PyData Berlin- various

2023-Present

- Frequent mentor and speaker at PyCon Events, Bayesian meetup groups, and internal DS guilds

Hiring Strategy - Personio- Dublin, Ireland

2021-Present

- Co-led Personio's data science hiring strategy: designed technical interview frameworks, improving candidate signal and DEI balance.

Open Source Contributor - PyMC eco-system- online

2021-Present

- Contributor to open-source probabilistic programming and causal inference libraries with 100k+ total users.