Norman Khan PhD, MPhys (Hons)

8 Blisworth House, Whiston Road, London, E2 8SG

+44(0)7804 495292 — norman1000@gmail.com — github.com/nx1 — nx1.info/Norman-Khan

Experienced data scientist, researcher, and programmer specializing in Python pipeline development, big data processing, and statistical analysis. Currently as a postdoctoral researcher at IRAP (Toulouse), building ETL pipelines for reducing, processing, and analyzing large-scale archival astronomical data. Background spans academic research and industry data science, with expertise in machine learning, time-series forecasting, and algorithm optimization.

DATA SCIENCE / ML ENGINEER SKILLS

Programming Languages	- Python, C, Bash, HTML, Javascript, SQL
Python Modules	- Matplotlib, Pandas, Flask, pytest, SymPy, ctypes
Machine Learning	- scikit-learn, Keras (Tensorflow), PyStan, Prophet (Facebook), OpenCV
Development Tools	- Linux, Git, REST APIs, Docker, Amazon EC2/S3, and Microsoft Azure

WORK EXPERIENCE

Institute for Research in Astrophysics and Planetology (IRAP)   Toulouse, France - *2023*
- Led a collaborative research project with to develop a novel Python ETL pipeline and algorithm using Bayesian statistics to detect and classify rapid astrophysical transients in 25 years of archival X-ray data. github.com/nx1/EXOD2 — arxiv.org/abs/2503.14208
- Built a `Flask` dashboard to browse, visualize, and analyse results efficiently.
- Processed 87 billion data points, discovering 20,000 previously unknown X-ray phenomena.
- Our work has been published in the leading peer reviewed journal Astronomy and Astrophysics (A&A)

Shell - *Forecasting Solar and Wind Power*   London, UK - *2021 (3 months)*
- Developed, implemented and compared several regression machine learning models using `scikit-learn` to forecast the power output over a 48h window for a solar farm.
- Using historical trends and post-processed data obtained from numerical weather prediction models, I found that a random forest regressor produced the most accurate prediction ($R^2 > 0.95$).
- Experience with implementing a full end-to-end machine learning pipeline involving dimensionality reduction, feature extraction, hyperparameter tuning and model selection.

Senseye - *Characterising Anomalous Behaviour in Time Series*   Southampton, UK - *2019 (3 months)*
- Worked on predictive maintenance on industrial assets using predictive condition monitoring.
- Developed a predictive algorithm built on Facebook's `Prophet` that calculated the probability of a sensor measure having crossed a specified threshold over time.
- Built unit tests using `pytest` to ensure the delivery of robust production quality code.
- Developed algorithms that would identify domain specific features such as quasi-flat lines in time sensor data.
- Familiarity with Agile workflow using Jira, Atlassian in order to collaborate with other developers.

EDUCATION

University of Southampton - PhD Astrophysics: Accretion onto Compact Objects   *2018 - 2022*
- Studentship via STFC-funded DISCnet Centre for Doctoral Training, receiving interdisciplinary training in data-intensive science, big data handling, and data analytics: discnet.co.uk
- Thesis: Testing Precession of Super-Eddington Flows in Ultraluminous X-ray sources.
- Developed ETL pipelines using Python for the automated data analysis of astronomical sources.
- 5 semesters of Python teaching experience in undergraduate computer laboratory courses.

University of Southampton - MPhys Physics with Astronomy: First Class   *2014 - 2018*

PUBLICATIONS

The EXOD search for faint transients in XMM-Newton observations	*2025*
Long-term X-ray/UV variability in ULXs	*2023*
Testing Precession of Super-Eddington Flows in Ultraluminous X-ray sources	*2022*
The impact of precession on the observed population of ULXs	*2022*
Thermally driven winds in ULXs	*2022*
Predicting the self-lensing population in optical surveys	*2021*

TRAINING COURSES

DISCnet - Machine Learning Course	*2019*
DISCnet - Statistics and Data Analysis	*2019*
DISCnet - PyStan Probabilistic Programming Course	*2020*
DISCnet - HPC and Scalable Programming	*2019*
DISCnet - Introduction to Big Data	*2019*

CONFERENCE PRESENTATIONS

XMM-Newton Survey Legacy for Athena and Beyond	Toulouse, France - *2024*
New Results in X-ray Astronomy	Leicester, UK - *2022*
eROSITA Time Domain Astrophysics	Munich, Germany - *2020*

ADDITIONAL SKILLS & EXPERIENCE

Language Proficiency Fluent in French	
Data Science - Datadive: King's College London - Quantification of cardiovascular conditions	*2020*
Data Science - Datadive: Royal National Lifeboat Institution (RNLI) - Saving lives at sea	*2019*