Dave Yeeles

dave@yeel.es 07896047892

PYTHON · LINUX · SCIKIT-LEARN · GIT · MATLAB · C · OPENSCAD · PANDAS · INDONESIAN (A2/B1)

Data scientist with 9 years experience building analysis pipelines and software tools for cutting-edge science. Expert in Bayesian inference, statistical modelling, and Python-based data engineering to deliver end-to-end solutions from prototype to deployment.

EMPLOYMENT

Scientific Software Engineer
Science and Technology Facilities Council
Oxfordshire, 2024-present

Built machine learning pipelines and computer vision tools for multimodal nanoscale bioimaging within a cross-disciplinary team. Led the full project lifecycle: developing statistical models, implementing ML algorithms that scale to millions of measurements, and deploying automated pipelines that run reliably in production to turn raw, messy data into actionable scientific insights.

Built a production ML pipeline that achieved world-first 5 nm resolution in single-molecule microscopy by resolving systematic biases in conventional techniques. Developed in Python and MATLAB using maximum-likelihood estimation and Bayesian inference (nested sampling) for parameter estimation, providing rigorous uncertainty estimates for downstream clustering stages.

Developed a statistical analysis framework in Python (NumPy, SciPy, pandas, scikit-learn), including custom multivariate kernel density estimation with adaptive bandwidth control. Performed model selection by benchmarking competing physical models, and implemented a novel approach for handling correlated circular and linear statistics using insights from differential geometry.

Scaled ML workflows to HPC infrastructure, orchestrating thousands of parallelised simulations for model validation. Containerised using Apptainer (Singularity) to ensure reproducibility across computing environments. Managed compute resource allocation and job scheduling with Slurm.

Designed and implemented a complete computer vision solution for automated multimodal image alignment. Pioneered the use of micromachined fiducial markers to bridge optical and electron microscopy, enabling the previously-impossible tracking and extraction of regions-of-interest across modalities. Led the full development cycle from component design in OpenSCAD, coordinating production with the fabrication team, to detection and analysis in Python with openCV.

Postdoctoral Research Fellow
Institute of Physics, Academia Sinica
Taipei, 2022-2024

Developed novel computational methods combining analytical modelling with numerical simulation to dramatically improve the efficiency of large-scale scientific computing. Modified production numerical relativity codebase and led major HPC simulation campaigns.

Increased computational efficiency of numerical black hole simulations by a factor of 10, enabling the simulation of systems that would otherwise have been unfeasible. Deployed parallelised C code on HPC cluster, again managing compute resources and workflow orchestration with Slurm.

Led a team of 10 to produce a major catalogue of black hole simulations. Required strategic resource planning to efficiently utilise the supercomputing budget, and an adaptable leadership style across skill levels from junior researchers to senior experts.

PERSONAL PROJECTS

Culinary Atlas of Indonesia
dave.yeel.es/culinary-atlas
2025

Exploring Indonesian cuisine through data science.
Scraped 50,000+ Indonesian recipes using Beautiful Soup and used unsupervised ML (along with UMAP dimensionality reduction, GMM clustering with BIC model selection, and iterative chi-square feature selection) to identify regional culinary families. Deployed a FastAPI backend and Leaflet.js frontend allowing users to explore geographic culinary patterns and discover recipe recommendations based on cosine similarity between dishes.

Proto-Language Reconstruction
[unpublished]
2025

Historical language reconstruction using machine learning.
Explored purely ML approaches to reconstruct proto-languages from modern descendents using PyTorch. Experimented with multiple architectures: GNN treating language families as graph structures, VAE for phonological feature encoding, attention mechanisms for sound correspondence. Tested by training on Romance cognates to 'reconstruct' Latin, but ultimately unsuccessful.

Diksonari blong Melanesia
dave.yeel.es/diksonari-blong-melanesia
2024

Quadrilingual dictionary of Melanesian creoles.
Collated and translated data from various sources into a single database. Automated data curation using Bash regex to rapidly generate a MySQL database powering an interactive online dictionary built using HTML and JavaScript.

Kamus Dwibahasa Ambon-Inggris
bahasanusantara.com/kamus-ambon
2023

Ambonese Malay – English dictionary.
As above, with additional data cleaning and orthography standardising steps. As well as the online tool, also uses Bash regex to rapidly generate a print volume typeset with a custom TeX class.

Digitising the Holle Lists
dave.yeel.es/holle
2023

Digitised 11 volumes of typewritten vocabularies of 300 indigenous Indonesian languages, collected over the past century, into a relational database using Tesseract OCR and Bash. The data is presented online as an interactive tool that displays all translations of a word simultaneously. The goal is to preserve endangered languages and widen digital access in rural communities.

EDUCATION

Royal Holloway, University of London
M.Sci. Physics, 2015

King's College London
M.Sc. Theoretical Physics
2016

Cardiff University
Ph.D. Astrophysics
2021