Shaun Yap

Contact: **shaunyap01@gmail.com** | (44) 7733933549 | linkedin.com/in/shaunyap0122/ | shaunyap01.github.io

CORE COMPETENCIES

I have extensively utilised R, Python and SQL for the past 5+ years, gaining deep familiarity with their capabilities and limitations, and effectively applying them to various data science and statistical projects.

Programming languages: R, Python, Julia, Tableau, Java, SQL, NoSQL, Node, React, CSS.

Software: Git, Github, Markdown, LaTeX, Microsoft Excel, Visual Studio Code.

Mathematics, Statistics, Programming, Statistical Analysis, Data Wrangling, Data Analytics, Data Preparation, Data Modeling, Database Management, Feature Engineering, Machine Learning, Deep Learning / Neural Networks, Artificial Intelligence, Large Language Models (LLMs), Model Validation, Research.

Intellectually curious; highly Resourceful; great Initiative; strong Leadership & Teamwork; effective Communication; Time management; Adaptability; Drive; Perseverance.

EDUCATION

PhD Researcher Digital Health Technologies (Visual System Vulnerability in Dementia) - University College London | UK — Sep 2025 – Present

4-year PhD **Research Studentship** fully funded by the EPSRC and The National Brain Appeal as part of the Tech4Health Doctoral Training Programme in Digital Health Technologies at University College London (UCL).

- Working with the UCL Queen Square Institute of Neurology and the UCL Hawkes Institute, under the supervision of Dr Keir Yong and Professor Andre Altmann.
- Research sits at the intersection of **neuroscience**, **data science**, and **digital health**, based across UCL Queen Square Institute of Neurology and the UCL Hawkes Institute.
- Research focuses on visual system vulnerability in dementia, combining **neuroscience**, **data science**, and **genomics** to improve detection and understanding of visual-led Alzheimer's disease.

MSc Data Science & Statistics (Distinction, 80% Avg.) - University of Exeter | UK — Sep 2023 – Nov 2024

- **Dissertation** "Evaluating Environment & Climate Truthfulness in Social Media using Deep Learning & Large Language Models" - received University **MSc Project Award (Best dissertation in cohort)**.
  - Focused on feature engineering, natural language processing (NLP), text vectorisation, traditional machine learning algorithms, deep learning architectures and large language models (LLMs).
- Applied **dimensionality reduction** on wine data; analysed factors influencing agricultural yield; **clustered** countries based on socio-economic and health indices using R/Python.
- Developed R functions for model **parameter optimisation**, applied **Bayesian statistical data modelling** techniques across various data sets, including binary and count data, and utilised Generalised Linear Models.
- Led group project to analyse tuberculosis (TB) risk in Brazil using **Generalised Additive Models** (GAMs); focused on quantifying **spatio-temporal risk** and recommending resource allocation for hospitals based on socio-economic covariates´ impact on TB rates.
- Developed proficiency in stochastic process properties and **Gaussian Processes** (1D and 2D GPs); elevated my ability to model and understand spatial data patterns.
- Applied **maximum likelihood** and **Bayesian approaches** to **Gaussian Processes**; demonstrated adaptability in model fitting and parameter estimation to uncover insights from spatial data.
- Designed and executed a **Dynamic Linear Model** incorporating linear trends and **seasonal** components, applied to monthly and quarterly datasets; demonstrated capability in handling and interpreting complex temporal data structures.
- Used **ARMA**, **ARIMA**, **SARIMA** & **SARIMAX** models for time series analysis; evidenced proficiency in model selection and validation; showcased ability to analyse temporal patterns and forecast future trends for global surface temperature.
- Detailed report on the impact of the 2011 Fukushima Disaster on plankton health, employed advanced **seasonality models** and **forecasting** techniques to analyse environmental data.
- Applied cutting-edge statistical learning techniques in Advanced Topics in Statistics projects: **Bayesian inference** for **Markov Chain Monte Carlo** (MCMC) with Gibbs Sampling in R and applying various classification methods in Python; optimising model hyperparameters for high accuracy in non-linearly separable data scenarios. Designed, built, and tested a **Java**-based **learning robot** utilising **Tremaux's algorithm** to find the quickest route to complete a maze.

BSc (Hons) Data Science - University of Warwick | UK — Sep 2020 – Jul 2023

- Used **R** extensively for university coursework. Used **Python** coupled with **Pandas**, **NumPy**, **Pytorch** for **Machine Learning** (ML).
- In **dissertation** project, contributed new methods, extensions & functionality (e.g. profile likelihood, regression modelling) to the **M-estimation** package using **Julia**, a high-level programming language which notably facilitates **automatic differentiation**. Leveraged **dual numbers** to support infinitesimals, enabling efficient calculation of complex gradient structures on large functions to optimise performance.
- **Exploratory Data Analyses** on (1) UK police stop-and-search dataset; (2) UN data on sustainable development goals. Very familiar with limitations of data, how data should be collected and cleaned.

- **Team project manager**; developed professional-level specification proposal to customer; created **React** web and mobile-accessible application with **Node.js** backend, **NoSQL** database for matching mentor and mentee in customer's company. Implemented **ML** in Python to leverage continual feedback ratings from user-set milestones for enhanced future matching. Facilitated team collaboration and version control using **Git** systems.
- Developed **Generalised Linear Models**, utilised **Sub-Classification methodologies** to (1) understand motivational factors involved in persuading rural people to source safer well water for drinking; (2) ascertain best way to treat cockroach infestations in apartment complexes.
- Developed and compared several **SVM machine-learning models** to read handwritten numbers. Used grid search method to ascertain model with best cross-validation error.
- Expertise in designing **efficient algorithms** and understanding computational problem complexity. Explored models of computation including **deterministic**, **non-deterministic**, and **randomised algorithms**, as well as **parallel** and **online computations**. Developed skills in classifying computational problems by hardness and devising strategies to address **intractable problems**.
- Proficiency in designing and analysing **efficient approximation** and **randomised algorithms** with provable performance guarantees for complex problems. Developed expertise in **LP relaxations**, **Chernoff bounds**, and the **probabilistic method**. Gained skills in **derandomisation** and **systematic algorithm design**, enhancing problem-solving capabilities through **mathematical proofs** and **advanced algorithmic strategies**.
- Tested sensitivity of **Expectation-Maximisation** algorithm through varied applications on large newsgroup datasets and critiqued clustering outcomes.
- Utilised **Relational Database Management Systems** to create a **PostgreSQL** GIG booking system to define and interrogate databases from procedural programs in **Java**. Client-specified constraints included: venues can be used by multiple gigs on the same day, need three-hour gap between gigs; no more than 20-minute interval in a gig line-up; number of standard tickets needed to be sold for promoters to be able to pay agreed fees.
- Designed **Java-based data structures and methods** for hosting restaurant information, customer reviews and customer favourites. Potential allergy issues are flagged on a customer-specific basis.

A-Levels - Garden International School, Malaysia
Jul 2018 – June 2020

- Mathematics – **A***
- Further Mathematics – **A***
- Chemistry – **A**
- Level 3 Extended Project – **A** "To What Extent Can Cryptocurrency Become Everyday-Use Currency?"

WORK EXPERIENCE

Data Scientist (R&D) – Bourn.ai | London, UK
Jul 2025 – Oct 2025

- AI fraud detection research and implementation.

Database Intern - EPIC Homes | Kuala Lumpur, Malaysia
Jul 2019 – Aug 2019

- Oversaw EPIC's database management operations, contributed to strategic planning efforts for the construction of green Epic Homes. EPIC is an award-winning social enterprise based in Malaysia.

Investment Banking Intern – AmBank | Kuala Lumpur, Malaysia
Jan 2019 – Jan 2019

- Internship with AmBank's Investment Banking Treasury Group based in Malaysia. Focused on credit rating analysis of specific quasi-government bonds.

CERTIFICATIONS

IBM Machine Learning Professional Certificate, 2024

- Developed practical skills in Supervised, Unsupervised, Deep Learning, and Reinforcement Learning, with a focus on Time Series and Survival Analysis through a 3-month, 6-course program covering the core algorithms and best practices in machine learning.

Deloitte Certificate of Completion (Technology Virtual Experience Program), 2023

- Completed tasks in Coding, Data Analysis, Data Modelling, Cyber Security, Forensic Technology.

J.P Morgan Certificate of Completion (Markets Virtual Experience Program), 2023

- Analysed customer, investment environment; developed investment proposal for customer; get buy-in from employer JPMorgan.

UNIVERSITY & PRE-UNIVERSITY AWARDS

- **MSc Project Award** (Awarded for excellence in dissertation research among MSc cohort) **– University of Exeter**, 2024
- UK Mathematics Trust, Senior Mathematical Challenge – Gold (**top 10%**), 2018 & 2019
- Duke of Edinburgh International Award – Silver, 2018
- Represented Federal Territory of Kuala Lumpur in Malaysian School Sports Golf Championship, 2016 & 2017
- Malaysian School Sports Golf WPKL - 1st Place, 2016; 2nd Place, 2017
- SEASAC Golf - 2nd Best Individual award; Team 2nd Place, 2018

FUNDRAISING / VOLUNTEERING EXPERIENCE

- Raised £22,146 in aid of Alzheimer's Research UK – Warwick, 2022
- Organised a 2-day magical event for Dania, a 4-year-old Make-A-Wish recipient battling leukaemia with a budget of MYR 5000 (~£950), 2018-2019
- Coached refugee children from Myanmar in Mathematics and English for three years, 2016-2019