Keith Ching

+61 415 376 087 | keithc07kh@gmail.com | Carlingford, NSW, 2118
e-Portfolio: https://kkh-portfolio.netlify.app/index.html

PROFESSIONAL SUMMARY

Data analyst and machine learning graduate with expertise in statistical modelling and predictive analytics. Proficient in exploratory data analysis, regression techniques, and hypothesis testing to extract insights from complex datasets. Experienced in building supervised learning models, feature engineering, and predictive modelling for classification tasks. Skilled in end-to-end workflows from data preprocessing to deployment, with strong capabilities in translating analytical findings into actionable recommendations through clear visualisations and stakeholder communication.

KEY PROJECTS

Identify Strengths and Weaknesses of AI and Human Reader on Breast Cancer Screening

Jul 2024 – Nov 2024 | *R, Power BI, SQL, Excel, Statistical Modelling*

- Executed exploratory data analysis (EDA) on government-sourced 600,000+ data to analyse comparative performance between AI and human readers. Employed multiple regression techniques, ultimately validating research findings through robust logistic regression analysis.
- Effectively communicated complex information through clear, structured presentations and written reports
- Conducted statistical hypothesis testing to quantify significance of performance differences, ensuring findings met clinical research standards for reproducibility and validity

Fannie Mae Mortgage Default Analysis Research Project

Nov 2024 | *R, Excel, Statistical Modelling*

- Conducted comprehensive statistical analysis on 420,000+ mortgage records using R to identify key default risk factors, revealing that loan modifications correlate with 98 times higher default rates and developing a predictive model
- Developed data-driven recommendations for lending practices by analysing credit scores, debt-to-income ratios, loan modifications, and more, providing actionable insights for minimising default risks in mortgage portfolios
- Achieved balanced predictive model performance with 74.4% specificity and 65.4% sensitivity through logistic regression, strategically optimising for minimising costly false negatives while maintaining robust identification of high-risk borrowers

LLM – SQL Query Generation with LoRA Fine-tuned CodeLlama

Nov 2025 | *Python, SQL, PyTorch, LoRA/QLoRA, Hugging Face Transformers, Git, VS Code*

- GitHub: https://github.com/kkh0714/SQL-Query-Generation-with-LoRA-Fine-tuned-CodeLlama

- Fine-tuned CodeLlama-7B-Instruct using LoRA (Low-Rank Adaptation) with 4-bit quantization on 78K SQL examples to generate queries from natural language questions
- Implemented parameter-efficient training achieving 87% SQL token accuracy and 57% exact match rate, demonstrating strong semantic understanding of SQL syntax and logic
- Developed comprehensive evaluation pipeline and inference system supporting interactive, single-query, and batch processing modes for production deployment

EDUCATION

Master of Data Science *Feb 2023 – Nov 2024*
Macquarie University – WAM: 89

Diploma of Information Technology (Higher Education) *Aug 2021 – Feb 2022*
Coder Academy

Bachelor of Commerce – Accounting & Finance *Mar 2014 – Jun 2017*
University of New South Wales (UNSW)

Extracurricular Activities

Global Leadership Program   *Feb 2023 – Nov 2024*

Actively participated and led discussions in six diverse thinktanks, spanning complex global issues like **Global Population Growth**, **Climate Change**, and addressing specific challenges such as **Taboo: health issues around the world**. This involvement honed critical analysis and the ability to synthesize multidisciplinary data. Furthermore, focused engagement in sessions on **The Science of Persuasion** and **High Impact Public Speaking** sharpened communication and influencing skills, enabling the effective delivery of complex, data-driven insights.

WORK EXPERIENCES

Data Science Intern   *Jul 2024 – Nov 2024*
Cancer Institute NSW - Sydney

- Assessed strengths and weaknesses of an AI breast screen mammography reader compared to radiologists
- Demonstrated the result using R and Power BI
- Hands-on experience in handling huge dataset
- Effectively communicated complex information through clear, structured presentations and written reports

Customer Service Agent   *Aug 2019 – Nov 2020*
Hamber Services Pty Ltd - Sydney

- Monitored and analysed customer transactions data
- Worked as a team to facilitate customer queries
- Effectively replied to client emails and phone calls with excellent written and verbal skills
- Prepared operation performance report
- Coached new colleagues and provided feedback

Family Business Member   *Jan 2018 – Aug 2019*
LEDeco - Sydney

- Launched official websites
- Analysed customers' online behaviour
- Participated online marketing courses held by Google about how to start a new business online and related marketing strategies
- Answered customer queries via email, Facebook, Instagram, and phone call

TECH SKILLS

- **Languages & Frameworks**: Python, R, SQL (PostgreSQL, MySQL, SQLite), JavaScript (ReactJS, React Native, TypeScript), Ruby (Ruby on Rails), HTML/CSS | PyTorch, TensorFlow, Hugging Face Transformers, LangChain
- **ML/AI**: Deep Learning (CV, NLP), Retrieval Augmented Generation (RAG), LLMs, LoRA/QLoRA, Transfer Learning
- **Tools & Platforms**: Git, VS Code, Jupyter Notebook, Power BI, AWS S3, Netlify, Heroku, Neo4j
- **OS**: MacOS, Windows

REFEREES

Chirag Mistry, *Data Scientist* - Cancer Institute NSW (contact details available upon request)

Costa Chan, *Manager* - Hamber Services Pty Ltd (contact details available upon request)