Mengmeng Wang

Data Scientist, Engineer, Researcher

+61-466-989-626
mengmeng.wang.phd@gmail.com
/mengmwang
/mengmwang
/mengmwang.github.io

Profile

I am a dedicated data scientist and researcher with a multidisciplinary background in electrical engineering, biomedical engineering and computer science. I have developed data cleaning and analysis, predictive modelling, machine learning and large language models. I am experienced in Python, R, MATLAB, SQL, Excel, Power BI, Tableau, and various Python and R libraries.

Skills

Programming Languages Python, R, SQL, MATLAB

Technical Tools GitHub, Tableau, Power BI, Excel

Python Libraries Pandas, NumPy, Scikit-Learn, transformers, NLTK, SciPy, Matplotlib, JSON, RegEx

Professional Data Analysis, Technical Writing, Project Management, Teamwork, Communications

Education

Doctor of Philosophy – Statistical Signal Processing 2017 – 2023
The University of Melbourne

Master of Science – Image Processing 2013 – 2014
University of Bristol

Bachelor of Engineering – Telecommunications Engineering 2009 – 2013
Beijing University of Posts and Telecommunications

Experience

Data Scientist Feb. 2023 – Present
Centre for Youth Mental Health, Orygen

- Experienced in data extraction, wrangling, and statistical analysis using structured and non-structured data.
- Applied machine learning algorithms and large language models in data analysis and health outcomes prediction.
- Contributed to developing statistical analysis plans and writing technical reports, research papers and policy briefings.
- Designed and implemented end-to-end machine learning pipelines, from data preparation to model deployment and performance monitoring.
- Collaborated with cross-functional teams including clinicians, researchers, and policy makers to deliver data-driven solutions for mental health service improvements.
- Developed interactive dashboards and visualisations for non-technical stakeholders to support data-informed decision-making.

Data Processing & Machine Learning Tutor (Casual) Mar. 2019 – Nov. 2022
The University of Melbourne

- Delivered tutorials and practical workshops for three university-level courses: Data Processing, Machine Learning, and Signals & Systems.
- Taught key topics including data wrangling, format, visualisation, natural language processing, supervised and unsupervised learning, classification, regression, clustering, neural network.
- Demonstrated applied data science techniques using Python and Pandas, NumPy, Scikit-Learn libraries to teach modern data science and advanced machine learning concepts.

Selected Projects

Clinical NLP and Predictive Modelling in Medical Case Notes

- Applied advanced natural language processing (NLP) techniques and large language models (LLMs) to extract, structure, and utilise insights from unstructured clinical texts. This project includes text de-identification, topic clustering and outcome prediction.
- **Subproject 1 - Medical Case Note De-identification:** Developed an automated, large language models (LLMs) based de-identification pipeline to identify and identify and mask personally identifiable information (PII) from clinical notes. The solution integrates external data sources (eg. location-based information) and goes beyond generic NER by incorporating real-world domain-specific knowledge.
- **Subproject 2 - Topic Modelling & Clustering:** Implemented a BERTopic-based framework to extract latent themes and group similar clinical case notes. Identified key clinical topics through unsupervised clustering.
- **Subproject 3 - Outcome Prediction:** Designed and validated models using structured features and text embeddings to predict clinical outcomes.
- Tools used: Python, Hugging Face Transformers, BERTopic, scikit-learn.

EEG Data Analysis in Music Therapy

- Performed EEG Data Analysis and visualisations in music therapy research.
- EEG Data importing, cleaning, preprocessing, feature engineering, visualisations and statistical analysis.
- Collaborated with health professionals, doctors and music therapists, to investigate the impacts of music therapy on brain state and functional brain connectivity.
- Tools used: MATLAB and EEGLAB Toolbox.

Financial Timeseries Processing and Forecast

- Performed financial data analysis in Level 1 Limit Order Book data, across data processing pipeline.
- Data cleaning and preparation: outlier detection, data visualisation and feature engineering.
- Data analysis: correlation, moving-average, auto-regression analysis.
- Timeseries forecasting: auto-regression model and machine learning models (decision tree, logistic regression, neural networks).
- Tools used: Python, Jupyter Notebook, Pandas, NumPy, matplotlib, seaborn, statsmodels, sklearn, keras.

Customer Purchasing Behaviours Analysis

- Analysed customer transaction and purchase behaviour data to find patterns in customer behaviours.
- Perform data cleaning, preparation and visualisation to facilitate analysis.
- Draw insights on change in store layout in customer behaviours.
- Tools used: Python, Jupyter Notebook, Pandas, NumPy.

Selected Publications

- D. Baker, **M. Wang**, K. Filia, S.M. Teo, R. Morgan, M. Ziou, P. McGorry, V. Browne and C. Gao, "The changing impacts of social determinants on youth mental health in Australia." *The International journal of social psychiatry*, 71(1), 116–128, 2025.
- **M. Wang**, C. Davey and L. Johnston, "Correction of induced functional connectivity in filtered resting state fNIRS data," *The 27th Annual Meeting of the Organization for Human Brain Mapping (OHBM)*, 2021.
- **M. Wang**, L. Johnston and C. Davey, "Correction for time-varying signal power in fNIRS connectivity analyses," *Society of fNIRS Virtual Conference*, 2021.
- **M. Wang** and A. Seghouane, "Motion Artefact Removal in Functional Near-infrared Spectroscopy Signals Based on Robust Estimation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.
- **M. Wang**, F. Zhang and D. Agrafiotis, "A very low complexity reduced reference video quality metric based on spatio-temporal information selection," *IEEE International Conference on Image Processing*, 2015.

Volunteer Experience

Girl Power Mentor Mar. 2021 – Dec. 2022

The University of Melbourne

- Mentored Year 11/12 female students with interests in science and engineering.
- Coached girls to develop strength and confidence to pursue tertiary education in science and engineering.

References

Available Upon Request