Mengmeng Wang

Data Scientist, Engineer, Researcher

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

Profile

Dedicated data scientist with a multidisciplinary background in electrical engineering, biomedical engineering, and computer science. Experienced in data cleaning, analysis, predictive modelling, and applying machine learning methods, including recent work with large language models. Skilled in Python, R, SQL and MATLAB, with experience working in Azure cloud environments and a strong interest in developing practical AI solutions, including building prototypes and exploring applications of Generative AI.

Skills

Programming Python (pandas, scikit-learn, Transformers, Hugging Face, Tensorflow), R, SQL

ML & AI Predictive modelling, Feature engineering, NLP, LLMs, Generative AI

Cloud & Platforms Azure, Cloud-hosted datasets and SQL, Git

Education

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

Master of Science (MSc) – Image Processing 2013 – 2014
University of Bristol

Bachelor of Engineering (BEng) – Telecommunications Engineering 2009 – 2013
Beijing University of Posts and Telecommunications

Experience

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

- Extracted, cleaned, and transformed large, complex health datasets (structured and unstructured) to enable advanced statistical and machine learning analyses.
- Designed and implemented end-to-end machine learning workflows, including NLP and large language models, covering data preparation, feature engineering, model training, and validation for predictive modelling and text analytics.
- Led 2 AI pilot projects to prove feasibility and evaluate effectiveness, developing prototypes that demonstrated practical applications and potential impact in mental health research and service delivery.
- Developed dashboards and visualisations to communicate analytical insights to technical and non-technical stakeholders.
- Led or contributed to 10+ analytical projects and authored/co-authored 20+ technical documents, research publications, and policy briefings supporting evidence-based decision-making.

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

- Delivered 100+ tutorials and practical Python workshops for undergraduate and graduate courses in Data Processing, Machine Learning, and Signals & Systems.
- Delivered hands-on instruction in data science and advanced ML topics, including data wrangling, formatting, visualisation, NLP, supervised and unsupervised learning, classification, regression, clustering, and neural networks.

Consulting Engineer May 2015 – Apr. 2017

China Academy of Information and Communications Technology

- Analysed domestic and international telecommunications and information technologies to inform strategic decisions.
- Produced 15+ consulting reports and data-driven deliverables for government and industry clients, providing actionable insights and recommendations.

Selected Projects

Predictive Modelling of Treatment Outcomes with Machine Learning and Explainable AI

- Built and optimised XGBoost classifiers on a large-scale clinical dataset to predict treatment outcomes.
- Implemented nested cross-validation for unbiased model evaluation, achieving an AUC of up to 0.78.
- Applied Shapley Additive Explanations (SHAP) to interpret prediction and quantify feature contributions.
- Performed feature engineering and preprocessing (scaling, one-hot encoding, up-sampling) to address heterogeneous clinical data.

Multimodal Clinical Outcome Prediction from Linked Health Data and Text

- Developed predictive models that integrate multimodal health data, combining structured electronic health records (EHRs) with text embeddings from clinical notes, using Azure cloud services.
- Performed feature engineering and preprocessing to improve data quality and model performance.
- Generated interpretable predictions from multiple machine learning algorithms to support clinical decision-making and outcome forecasting.

Topic Modelling & Clustering of Clinical and Survey Text

- Applied BERTopic and embedding-based techniques to extract latent themes from large-scale free-text, including clinician notes and open-ended survey responses.
- Clustered text documents into meaningful groups to support knowledge discovery.
- Delivered insights that supported health outcome research and survey-driven service improvement.

Multiview Clustering on Social Cognition Data

- Built a multiview clustering pipeline integrating multiple feature sets, with dimensionality reduction for joint and view-specific projections.
- Created pairwise visualisations to compare cluster structures and highlight inter-view patterns.

Financial Time Series Processing and Forecasting

- Developed a data processing pipeline for high-frequency Level 1 Limit Order Book data, including cleaning, outlier detection, and feature engineering.
- Applied statistical techniques (correlation, moving average, autoregression) to extract patterns and insights from financial time series.
- Implemented and evaluated multiple machine learning algorithms (decision trees, logistic regression, neural networks), comparing forecasting performance across models.

Volunteer Experience

Mentor, Girl Power in Engineering and IT Mar. 2021 – Dec. 2022

The University of Melbourne

- Mentored high school girls interested in science and engineering, fostering confidence and aspiration for tertiary education in STEM.

References

Available Upon Request