# Lee Ming Xiang

✉ mingxiang1006@gmail.com   📞 (+60)12-9791319

I am a data scientist with petroleum geoscience background. I have five years of broad-based experience in building machine learning solutions in solving oil and gas industry challenges, specifically in production, petrophysics and geophysical domains.

Proficient in predictive modelling, data processing, data analytics, and Python scripting. I am active in learning, and proactive in trying innovative ideas for problem solving. LinkedIn profile

**SCAN ME**

## Core Skills

Machine Learning,
Deep Learning, Computer Vision, Natural Language Processing,
Data Visualization, Petroleum Geoscience

## Data Science Skills

Python, Pytorch, TensorFlow, Scikit-Learn, MLflow, Spotfire, Power BI, Dash, Azure, Structure Query Language (SQL), Oracle Database, Docker, Dataiku

## Geoscience Skills

Petrel, Omega, Vista, Techlog, Rock Physics, Seismic Processing, Seismic Interpretation, Static Modelling, Exploration & Production Cycle

## General Skills

Problem Solving, Teamwork, Adaptability, Organizing/ Planning, Decision Making, Proactive Learning, Fast Learner

## Working Experience

### Domain Data Scientist

*KL Innovation Factori, SLB*
*Oct 2023 – Present*                                              *Kuala Lumpur, Malaysia*

1. *Seismic Denoise using Self Supervised Learning*
   - Swell noise removal using self-supervised learning integrated with physics domain knowledge for better signal and artefact protection.

2. *AI Approach for Operation Insights Retrieval using GenAI LLM*
   - Comparing the state of art NLP model with GenAI LLM to explore the strength and limitation of GenAI LLM for operation insights retrieval.
   - Enhance the prompt and Retrieval Augmented Generation (RAG) workflow.
   - Fine tuning both local and on-cloud LLM.

3. *Real-time Acoustic Data Visualization and Analytics:  Fiber Optics-to-Image Streaming Solution Analytics*
   - *Data Processing and Transmission in real time, with conversion from signal to image*
   - *Real Time image processing and visualization*
   - *Anomaly detection using unsupervised learning.*

4. *Data Driven Optimization for Rock Physics Modelling Assisted by Machine Learning*
   - Develop innovative optimization workflow integrate data driven (data science, machine learning) and physics-based approach.
   - Deployed SaaS user centric web application for comprehensive visuals

### Data Analytics Engineer

*KL Innovation Factori, SLB*
*Jun 2021 – Oct 2023*                                              *Kuala Lumpur, Malaysia*

1. *Oil and Gas Language Models- Unsupervised Multitask Learning*
   - Explore the development of an oil and gas language model (LM) using an unsupervised multitask learning approach.
   - Publication

2. *Information Retrieval from Oil and Gas Unstructured Data*
   - Automated information extraction, and relationship extraction from Daily Production Report applying Natural Language Processing (NLP).
   - Mentoring in research for oil and gas language model and rapid risk identification from Daily Drilling Report.

3. *Pattern Recognition between Petrophysics and Production*
   - Cross domain machine learning prediction project in team to recognize the dominant factor in predicting the production potential.
   - Applied and compare various machine learning algorithms in predicting the hydrocarbon flag, perforation zone, permeability, and production rate.
   - Generated hypothesis testing to find correlation between estimated petrophysical production rate and the actual production rate.

4. *Well Performance Analytic Dashboard*
   - Deployed diagnostic Spotfire analytics dashboard to identify overperforming and underperforming well.
   - Positive feedback from stakeholders on the usability, and dashboard visualization.
   - Integrated workflow from retrieving data using API, data processing, to data analytics from Production Data Foundation, Dataiku to Spotfire.

5. *CO2 Emission Monitoring based on Prediction of Gas Fuel Rate -Time Series Prediction*
   - Business impact award for 2021 SLB Asia Sustainability Hackathon.
   - Deployed Extra Tree algorithm in predicting the gas fuel rate to calculate the emitted $CO_2$ in next 7 or 14 days.
   - Created a predictive analytic dashboard using Power BI by ingesting the data using API from the Dataiku Server.
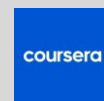
## Certifications

**Generative AI with Large Language Models**

**Build Basic GANs**

**Improving Deep Neural Networks: Hyperparameter Tuning**

**Neural Networks and Deep Learning**

**Dataiku Core Designer**

**Dataiku ML Practitioner**

**Dataiku Advanced Designer**
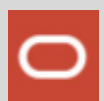
**Tibco Certified Associate Spotfire**

**Industrial Data Fundamentals**

**Data Fusion Fundamentals**

**OSDU Developer Training**

**Oracle Database Design & Programming with SQL**

**Azure AI & Data Fundamentals**

**Geosolutions Horizon Fixed Step Training Phase 1 ,2,3**

## Publications

- *SPE Publication:* [Oil and Gas Language Models- Unsupervised Multitask Learning](#)
- *EAGE Publication:*
  1. [Unlocking Value from Text Visualizing Insights with Natural Language Processing in Unstructured Oil and Gas Repots](#)
  2. [CO2 Emission Monitoring based on Prediction of Gas Fuel Rate -Time Series Prediction](#)
- *Research Gate Publication:* [Unlocking Value from Text Visualizing Insights with Natural Language Processing in Unstructured Oil and Gas Repots](#)

## Mentoring and Leadership

**International Petroleum Technology Conference 2025 Committee**          *Jun 2023 – Feb 2025*
Participate in organizing the Digital, Data Analytics, and Automation program, review submitted abstract for technical session.

**Technical Committee for SLB Machine Learning Innovative Competition**   *Jan 2023 - Jun 2023*
Review the data science challenges and DELFI technology stack used for the competition.

**Technical Committee & Mentor for APGCE GeoHackathon**          *July 2022 - Nov 2022*
Worked with Petronas management, geoscientists, and data scientist in developing oil and gas upstream data science challenges. Mentoring participants in applying data science to domain challenges.

**Technical Committee for SLB Beijing Geoscience Center**          *Dec 2021 - Jan 2022*
Introduced the hackathon challenge in forecasting the production decline curve using both production and formation data.

**Volunteer Speaker for Women Who Code Power BI Workshop**          *Jun 2019*

## Personal Projects

- Unsupervised Segmentation using Computer Vision
  https://github.com/mingxiang1006/Unsupervised_Seg
- Automatic Detection of Solar Roof Top using Computer Vision
  https://github.com/mingxiang1006/solar_ai
- Groove Defect Segmentation using Computer Vision
  https://www.kaggle.com/code/mingxiang1006/unet-seg
- Machine Learning with Optimized Parameters for Ecommerce Product Classification
  https://github.com/mingxiang1006/Ecommerce-Product-Classification/tree/main
- Future Sales Prediction
  https://github.com/mingxiang1006/Predict_Futre_Sale
- Telco Customer Churn Prediction
  https://github.com/mingxiang1006/Telco-Customer-Churn-Prediction
- Nasdaq Stock Portfolio Optimization
  https://github.com/mingxiang1006/Stock-Portfolio-Optimization

## Education

**Master of Data Science**
*Oct 2020 – Jun 2022*                    *University Malaya (UM), Kuala Lumpur*
Master Thesis: Generation of Carbon Dioxide Emission based on Prediction of Gas Fuel Rate using Machine Learning (Time Series Prediction)

**Bachelor of Technology (Hons) in Petroleum Geoscience**
*May 2012 – May 2017*                    *University Technology PETRONAS (UTP), Perak*
Majoring in Exploration Geophysics, Fundamental in Geology, Petrophysics, and GIS
Final Year Project: Application of Simultaneous Inversion in Sarawak Basin, Malaysia

**Student Exchange Program**
*Aug 2015 – Dec 2015*                    *Missouri University Science & Technology, United States*
Studied Petroleum Economics, Reservoir Characterization, General Psychology and Technical Communication