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 subsurface and production domain.

Proficient in natural language processing (NLP), GenAI LLM, subsurface relevant AI solutions. I am active in learning, and proactive in implementing 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. *Handwritten Detection and Recognition*
   - *Develop OCR (Optical character Recognition) algorithm for handwritten detection and sentence recognition from report, using both computer vision and deep learning techniques.*

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

3. *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.

4. *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.*

5. *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 – Sep 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.

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.

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

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 CO2 in next 7 or 14 days.
   - Created a predictive analytic dashboard using Power BI by ingesting the data using API from the Dataiku Server.

Geophysicist

Geosolutions, SLB *Kuala Lumpur, Malaysia*
Mar 2018 – May 2021

Seismic processing geophysicist, experience in narrow and wide azimuth survey at ultra-shallow, shallow, and deep water in different offshore locations such as offshore Malaysia, Myanmar, Australia, Western India, Gabon. Experience in seismic processing, mainly on seismic deblending, seismic denoise, demultiple, migration, velocity picking.

Automation work:

1. Linux Script for P190 & SEGD Extraction
2. Automated Parameter Analysis and Recommendation for Adaptive Deghosting

Intern Geoscientist

Beicip-Franlab Asia *Kuala Lumpur, Malaysia*
Jan 2016 – Aug 2016

1. Field Development Plan (FDP) for Angsi Field in Malay Basin.

2. GIS Operation Dashboard.
The purpose of the project is to overcome the communication barrier between offshore team and onshore officers by updating or generating the latest technical interpretation and economic analysis with a real time basis.
[Research Area: Gullfaks, Teapot Dome, North Sea]

Publications

1. Unsupervised Multitask Learning for Oil and Gas Language Models with Limited Resources, M. Marlot., D.N. Srivastava, F.K. Wong, M.X. Lee, ADIPEC, Abu Dhabi, UAE, October 2023. [Link]
2. Optimizing Performance in Big Data Handling for Enhanced Data Analytics, S. Atiq, M.X. Lee, EAGE Workshop on Data Science, 2023.
3. A survey of Natural Language Processing in Oil and Gas: Opportunities and Challenges, M. Marlot, M.X. Lee, EAGE Workshop on Data Science, 2023.
4. Unlocking Value from Text: Visualizing Insights with Natural Language Processing in Unstructured Oil and Gas Reports, M. Marlot, M.X. Lee, A. Irfan, P.K., Tellapaneni, L. Edwin, SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition, 2023. [Link]
5. Information Retrieval from Oil and Gas Unstructured Data with Contextualized Framework, M.X. Lee, M. Marlot, Third EAGE Digitalization Conference and Exhibition, Mar 2023. [Link]
6. Carbon Dioxide Emission Monitoring based on Prediction of Gas Fuel Rate using Machine Learning, M.X. Lee, EAGE Conference on Digital Innovation for a Sustainable Future, 2022. [Link]
7. Adaptive Deghosting Dashboard, M.X. Lee, A. Sazykin, SLB Technical Coordinators Meeting, 2021.
8. Imaging multi-order multiples - Shallow Water Case Study from Southeast Asia, B. Chowdhury, A. Sazykin, P. Kristiansen, S.Y. Lee, M.X. Lee, R. Alai, M. Shah, M. Nasrul, N. Nadzirah, SEG Kuwait "Seismic Multiples - The Challenges and the Way Forward" Workshop,2019.
9. Abstract of Optimum Notch Frequency Recovery using non-CMS approach. C.M. Lam, A. Verba, M.X. Lee, SLB Technical Coordinators Meeting, 2019.
10. Application of Simultaneous Inversion Characterizing Reservoir Properties in X Field, Sabah Basin, M.X.Lee, L.A. Luluan, IOP Conference Series: Earth and Environmental Science, Volume 88, 5th International Conferences on Geological, Geographical, Aerospace and Earth Sciences 2017 (5th AeroEarth 2017) 20–21 May 2017, Kuta, Bali, Indonesia. [Link]

Achievements

- 2nd runner up of EAGE Field Challenge 2017
  Represented Malaysia participating the EAGE Field Challenge 2017 organized by Total company at Paris, France with fully Integrated evaluation and field development project.
- AAPG L. Austin Weeks Recipient 2017
  Scholarship recipient for the 2017 American Association of Petroleum Geologists Foundation L. Austin Weeks Scholarship program
- Silver Award in Integrated Exploration and Production Opportunity Evaluation Project 2016
  New prospect finding for Bundi field integrating G&G knowledge. Performed reservoir probability and risk evaluation, and petroleum economic analysis.

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