Lee Ming Xiang

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

I am a data scientist with petroleum geoscience background. I have six 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

Senior Domain Data Scientist

KL Innovation Factori, SLB

Jun 2021 – Present *Kuala Lumpur, Malaysia*

1. *Handwritten Detection and Recognition from Scanned Documents*

- Engineered a sophisticated system for detecting and recognizing handwritten text from scanned documents using deep learning methodologies.
- Designed and implemented a custom pipeline using a residual convolutional neural network (ResNet) architecture tailored for on-premises applications, addressing connectivity constraints in rig floor environments.

2. *Pre-stack Seismic Denoise using Deep Learning*

- Innovated a self-supervised learning model specialized in low-frequency noise attenuation, utilizing sequential inference networks for effective swell noise reduction.
- Enhanced traditional denoising techniques by integrating domain-specific knowledge and augmenting data, employing a residual attention mechanism within a denoising autoencoder framework.
- Achieved superior denoising performance, eliminating signal leakage and artifacts, providing more consistent and noise-free results compared to traditional methods.

3. *Operation Insights Retrieval using GenAI LLM*

- Developed a comprehensive pipeline for operational insights retrieval, incorporating prompt engineering, Retrieval Augmented Generation (RAG), and fine-tuning of large language models (LLMs), culminating in a user-centric web application deployment.
- Enhanced model performance by fine-tuning an open-source LLM, achieving a 94% accuracy rate—surpassing that of the pre-trained GPT-3.5 Turbo by 3%—through strategic hyperparameter tuning and rigorous post-processing safeguards.
- Conducted extensive research on the efficacy of different chunking strategies, data types, embedding techniques, and vector database models on the RAG framework, optimizing for specific use cases in geomechanics risk assessment for 1D modeling.

4. Real-time Acoustic Data Visualization and Analytics: Fiber Optics-to-Image Streaming Solution Analytics

- Developed a real-time data transmission pipeline converting high-frequency DAS signals to images, reducing data size by 71% while preserving resolution.
- Deployed an online web application for real-time image visualization.
- Implemented anomaly detection using STEGO, an unsupervised segmentation technique with self-supervised transformer and energy-based graph optimization.

5. Data Driven Optimization for Rock Physics Modelling Assisted by Machine Learning

- Innovated an optimization workflow combining data-driven and physics-based methods.
- Implemented global sensitivity analysis for parameter ranking and initialization in non-derivative optimization.
- Integrated rock physics constraints into function optimization.
- Developed RESTful APIs for PostgreSQL updates and deployed a SaaS web application with comprehensive visualizations.

6. *Oil and Gas Language Models- Unsupervised Multitask Learning*

- Mentoring to develop an unsupervised multitask learning fine-tuned GPT-2 medium oil and gas model with limited resources.
- Automated information extraction, and relationship extraction from Daily Production Report applying Natural Language Processing (NLP).

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**

7. *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 the correlation between the estimated petrophysical production rate and the actual production rate.

8. *Well Performance Analytic Dashboard*
   - Deployed 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, and data processing, to data analytics from Production Data Foundation, Dataiku to Spotfire.

9. *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 the 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, with experience in narrow and wide azimuth surveys at ultra-shallow, shallow, and deep water in different offshore basins. Experience in seismic processing, mainly on seismic deblending, seismic denoise, demultiple, migration, and velocity picking.
Automation work:

1. Conventional P190 information extraction & SEGD file list generation were done sequence by sequence has consumed significant production time. Developed an automated bash Linux script for automatic data extraction, saving production from 2 weeks to 20 minutes.
2. Automated Parameter Analysis and Recommendation for Adaptive Deghosting, saving testing time about 70%.

Publications

➢ *Data Science*

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]

➢ Geoscience

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