# KhaiSian Lim

Penang, Malaysia • +60 12 433 6851 • khaisianlim0571@gmail.com • https://github.com/Khai-Sian

Experienced Data Professional with a strong foundation in Machine Learning, Data Analytics, and Software Development. Proficient in designing and implementing data-driven solutions, building predictive models, and extracting actionable insights from complex datasets. Skilled in Python, SQL, and data visualisation tools, with experience collaborating cross-functionally to solve business problems. Academic projects and professional achievements demonstrate the ability to combine technical expertise with innovative problem-solving to drive impact across industries.

## WORK EXPERIENCE

**Coherent**
**Data Scientist** • Full-time Feb 2025 - May 2025

- Analysed global procurement data and supported the Finance Director and C-level stakeholders in identifying cost-saving opportunities from 160M to 220M; contributed to standardising savings calculations across international teams.
- Automated consolidation of global procurement data by building a Python-based validation tool, reducing processing time from 2 to 3 days to under 5 minutes and improving reporting accuracy.
- Developed a deep learning pipeline using **PyTorch** and **HuggingFace RoBERTa** for taxonomy classification; improved model accuracy from ~30% to over 60%.
- Created interactive **Power BI dashboards** to visualise global engineering software usage, enabling finance teams to identify and eliminate ~10–15% in unused subscriptions.
- Conducted feasibility analysis for local AI model hosting using GPUs; proposed infrastructure improvements to reduce dependency on external cloud inference.

**Intel**
**Software Developer (Graduate Trainee)** • Full-time Jan 2022 - Jun 2023

- Designed and implemented Python validators for XML-based product files, increasing code coverage and enhancing system reliability.
- Built Python-based validators for XML product files to ensure structural and schema compliance, improving validation consistency and reducing manual review time by ~40%.
- Collaborated cross-functionally to analyse requirements and design tailored solutions for file validation processes.
- Recognised by the product team for substantial contributions to a high-impact project on the Penang site.

**Flex**
**Software Developer** • Internship Jun 2021 - Aug 2021

- Optimised dashboard performance and usability through improved UI/UX design.
- Researched and integrated a speech recognition feature, reducing manual user interaction errors.

## EDUCATION

**Master in Data Science and Analytics (Distinction)**
**University of Leeds** • United Kingdom Sep 2023 - Dec 2024

- Average score: 74

**BSc in Computer Science**
**UOW Malaysia KDU Penang University College** • Penang — Jan 2019 - Dec 2021

- GPA: 3.91

**Certificate in SACE International**
**Disted College** • Penang — Jan 2018 - Dec 2018

- ATAR Score: **80.35** | Top 10 students graduated from the course

## PROJECTS

### Pokémon Card Detection and Identification (Personal Project)

- Trained a **YOLO object detection model** to locate Pokémon cards in diverse images.
- Used **FAISS** to identify the detected cards through vector similarity search.

### Wikipedia Article Classification (Master's Dissertation)

- Built a quality classification system for Wikipedia articles using text-based and structural features.
- Applied **Random Forest** and **K-Nearest Neighbors** as baselines; enhanced performance using **pre-trained BERT** and **LSTM** models.
- Conducted model evaluation using precision, recall, F1-score, and confusion matrices.

### Predictive Maintenance for Turbofan Engine (Bachelor's Dissertation)

- Developed a machine learning pipeline to predict the **Remaining Useful Life (RUL)** of turbofan engines using **NASA C-MAPSS** sensor data.
- Performed data preprocessing, feature selection, model training (e.g., XGBoost), and evaluation.
- Delivered actionable insights through visualisations to support predictive maintenance decision-making.

### IBM SpaceX Launch Prediction (Coursera Capstone)

- Completed end-to-end ML workflow including wrangling, analysis, and modeling for SpaceX rocket landings.
- Visualized results using **Plotly** and **Folium**; used logistic regression and decision trees.

### Customer Classification

- Applied classification and clustering algorithms to segment customers into "High-value" and "Low-value" groups.
- Leveraged **Random Forest**, **KMeans**, and feature engineering techniques.

## PUBLICATIONS

**Machine Learning Techniques for Predicting Remaining Useful Life (RUL) of Machinery for Sustainable Manufacturing Lines** — Jan 2023
Springer

https://link.springer.com/chapter/10.1007/978-3-031-36246-0_31

## SKILLS

- **Programming:** Python, R, SQL
- **Machine Learning:** HuggingFace, LangChain, PyTorch, Scikit-learn, TensorFlow

- **Data Visualization:** Matplotlib, Power BI, Seaborn, Tableau
- **Tools:** Microsoft Excel, MySQL, Visual Studio
- **In Progress:** Apache Spark, Cloud Platforms (AWS/GCP)