Hazrul Akmal Hazarudin

https://hazrulakmal.github.io/

Email : hazrulakmal121@gmail.com
Mobile : +44-7570271356

Technical Experiences

- **Central Bank of Malaysia** — Kualua Lumpur, Malaysia
  Data Science Intern — *Feb. 2021 - Present*
  - **ML NLP Experiment**: Optimised Distil-BERT for multi-output financial news sentiment classification task with a limited dataset to achieve an F1-Score of 0.86 and a 50% reduction in inference time through Data Augmentation, Knowledge Distillation, and Transfer Learning.
  - **Text Document Extraction**: Conducted research on the usability of Open Source tools such as PyMuPDF, PDFMiner, Camelot and SOTA document processing model, LayoutLM to perform document extraction tasks.
  - **Engineering**: Developed a Python program to compute CoVaR to estimate systemic risk, enabling team analysts to outsource statistical CoVaR analysis to Python ecosystems.
  - **Statistical Modelling**: Analysed financial system systemic risk across 9 Malaysia's major banking providers using CoVaR method and presented the findings through interactive Power BI Dashboard.
- **LSESU Data Science Society** — London, UK
  Project Developer — *Sep. 2021 - Dec. 2021*
  - **Statistical Analysis**: Performed a market basket analysis on more than 1M retail transactions dataset to uncover associations between products that are purchased together and to identify patterns of co-occurrence.
  - **Data Pre-processing**: Pre-processed raw data into a structured format for analysis using Pandas. Applied an unsupervised machine learning clustering technique to segment different customers based on purchasing patterns
  - **Data Visualisation**: Designed clear and concise graphs using Seaborn, Plotly and Matplotlib to effectively showcase in-depth analysis for presentation.

Projects

- **Kaggle Feedback Prize 3 NLP Competition**: Built predictive models that automatically assess essay writing skills based on six different aspects. Submissions were ranked top 36% in leaderboard track and top 9% in efficiency track.
  - **Training Performance**: Optimised GPU usage through mixed-precision and gradient accumulation techniques to finetune 304M parameters DeBERTaV3-Large model.
  - **Architecture Tuning**: Engineered Attention and Mean Pooling heads and integrated layerwise learning rate decay (LLRD) into DeBERTa model to effectively improve embedding representations.
  - **ML Experiment Tracking**: Utilised Weight & Biases to track and visualize experiments and iterate quickly.
- **Malaysia Study Jam Mini Hackathon**: Customer segmentation prediction competition. Finished 1st place
  - **Hyperparameter Tuning**: Maximised various model performances via an automated tuning library, Optuna
  - **Predictive Modeling**: Applied feature engineering techniques such as Principal Component Analysis (PCA) and K-Means Clustering Algorithm to optimise data. Ensemble LightGBM, CatBoost, and Gradient Boosting models to achieve the highest mean F-score of 0.55.
- **Movie Database**: Created a website application database by using SQLite technology that allows users to create a login account, and give reviews, rates and comments on thousands of movies and TV shows.

Education

- **London School of Economics and Political Science** — London, UK
  Bachelor of Sciences in Mathematics, Statistics and Business; Expected - 1st Class — *Sep. 2020 – Present*
  - **Related Modules**: Machine Learning, Artificial Intelligence, Time Series & Forecasting, Financial Statistics, Bayesian Inference, Stochastic Processes, Algorithms & Data Structures, Databases and Principle of Finance
- **Kolej Tuanku Ja'afar** — Seremban, Malaysia
  *Advanced-Level; A*A*A*A*A* — *Aug. 2018 – June. 2020*
  - **Subjects**: Further Mathematics, Mathematics, Physics, Economics, Extended Qualification Project (EPQ)

Skills

- **Certificates**: **Natural Language Processing Specialization** by DeepLearning.AI, **Data Scientist in Python Path** by Dataquest.io, **Machine Learning for Time Series Data in Python** by DataCamp
- **Programming**: Python, R, SQL, Julia **Tools**: Git, Sklearn, PyTorch, XGBoost, Darts, Tidymodels, LaTeX