Lian Chee Loong

ASSOCIATE DATA SCIENTIST · econometrics & finance major

+6016-2179937 | cheeloonglian@gmail.com | http://www.linkedin.com/in/chee-loong | https://cheeloong.github.io/

Work Exp.

Ernst & Young Advisory Services Sdn Bhd *KL, Malaysia*

ASSOCIATE DATA SCIENTIST – DATA AND ANALYTICS TEAM *Mar. 2018 - June. 2018*

JD: compiling business intelligence requirement for a major beverage company in Thailand and have co-written proposal to win a CoE project with a local newspaper company.

- Design dashboard storyline based on company organizational structure as well as business requirement
- Built interactive live sales dashboard with SAP Lumira Discovery 2.0
- Collaborate with business process improvement team to compile to-be business reporting requirement list
- Help with creating data warehouse architecture and BI implementation roadmap for client

Crave Asia Sdn Bhd *KL, Malaysia*

DIGITAL ANALYST – DIGITAL TEAM *Jan. 2016 – Apr. 2017*

JD: providing social media ad-hoc analysis and post-campaign digital report, distribution of media ad spent.

- Design weighted budget allocation on Facebook advertisement for multiple brands
- Analysed performance of FB ads and conduct advert optimization in terms of price and audience reach
- Integrating and evaluating advertising materials from different department
- Responsible to prepare and present the Post-Campaign Report to internal team/client.

Education

Monash University Clayton *Clayton, Australia*

B. COMMERCE IN ECONOMETRICS & FINANCE *Jul. 2013 – Mar. 2015*

- Weighted Average Mark 68.5%

Monash University Sunway *Sunway, Malaysia*

B. BUSINESS & COMMERCE IN ECONOMETRICS & FINANCE *Mar. 2012 – Jul. 2013*

- Weighted Average Mark 77%

Sunway University College *Sunway, Malaysia*

WACE, AUSTRALIAN MATRICULATION (AUSMAT) *Mar. 2011 – Dec. 2011*

- ATAR 77.25%

Technical

Fast.ai, Python, Rstudio, LaTeX, Markdown, Jupyter Notebook, NumPy, scitkit-learn, PyTorch, GCP, Google Colab, GitHub, HTML5, CSS3

17th June 2019

Projects

APPLICATIONS: VISION

Kaggle 2017: Planet Amazon Satellite Classifier with ResNet50 CNN

- Image classification modelling to label satellite image chips (Multi-labeled classification)
- Data Import using Kaggle API, pre-processing involving data augmentation and normalization, transfer learning using ImageNet, fine tuning model; freeze/unfreeze model training
- Attains 93.1% Validation F2 score (Late submission result: top 6% on Kaggle's private leaderboard)

Cambridge-driving Labeled Video Semantic Segmentation with ResNet34 encoder U-Net

- Image semantic segmentation to segment different objects on the snapshots taken from the video
- Apply weight decay, learning rate annealing, and transfer learning using ImageNet
- Attains 92.57% Validation Accuracy (November 2016 state-of-the-art was 91.5% Validation Accuracy)

Oxford Pets Dataset Image Restoration with GAN & Perceptual Loss CNN

- Feed low-res images as inputs and generate higher-res images as outputs
- Built pre-trained ResNet34 encoder U-Net, pre-trained Critics, Generative Adversarial Network (GAN), and perceptual loss CNN for super-resolution
- Model performance is evaluated on the picture quality generated by the model, validation accuracy is irrelevant.

APPLICATIONS: TEXT

IMDb Movie Reviews Sentiment Analysis

- Identifying viewers' sentiment about a movie (good/bad) using a subset of user reviews and movies IMDb dataset
- Text pre-processing involving tokenization and numericalization
- Pre-train source Language Model (ASGD Weight-Dropped LSTM)
- Built target Language Model (IMDb LM) and fine tuning after
- Attains 91.8% Validation Accuracy (Improvable with better GPU)

APPLICATIONS: TABULAR

Kaggle 2015: Rossmann dataset

- Forecast "Sales" for Germany's second largest drug store chain with given train set
- Importing original data & external datas uploaded by other Kagglers
- Feature Engineering / Data Cleaning: Joining different dataframes, string splitting on dataframes, extracting data fields, dealing with missing values, creating durations between data intervals
- Applying fast.ai preprocessors like `Categorify`, `FillMissing` and `Normalize`
- Applying Dropout Regularization (subclassing from PyTorch) and Batch Normalization
- Attains 12.04% Validation Loss (Late submission result: Top 27% on Kaggle's private leaderboard)

My projects are mostly done using deep learning models with fast.ai libraries and subclassing PyTorch, my go-to cloud server for GPU is Google Colab because it's free, I have also used GCP for a bit before Colab.

If you are interested to learn more about my work or myself, feel free to visit: https://cheeloong.github.io/