Jacky Angara

7 Ghim Moh Road, Singapore, +65 82429244, jackyangara@gmail.com, jackyangara.github.io

Education

The Hong Kong Polytechnic University   Hong Kong
Bachelor of Science (Hons.) Computing   Aug '15 - Jun '19

Professional Experience

Shopee   Singapore
Associate Data Scientist   Sep '19 – Present

- Fine-tuned a pre-trained model (BERT) to perform sentiment analysis on noisy structured data. Created a preprocessing pipeline to clean the noisy data. Prepared a text labelling guideline for local teams to label the data accordingly.

- Performed data clustering on noisy text data to assist a slot-filling task. Analysed the multi-dimensional data to find the optimum amount of clusters using mathematical methods. Implemented active learning to overcome the problem faced by both unsupervised and supervised learning..

- Experimented on different word/sentence embeddings (e.g. BERT embedding, fastText embedding, word2vec, GloVe, TF-IDF) to compare the weaknesses and strengths in different NLP tasks.

- Researched on different python web frameworks to find one that suits the application. Developed and deployed APIs using falcon web framework which consistently performs really well on high load.

- Crawled data from online sources using multiprocessing on multiple proxies while maintaining a low number of HTTP requests to prevent web traffic disruption.

- Performed unit testing (Python's unittest and PyTest) and load testing (JMeter) on APIs before entering UAT and Live deployment.

- Application of various machine learning algorithms, statistical modelling and evaluation metrics using scikit-learn.

Data Science Intern   Jun '19 – Aug '19

- Researched on slot-filling techniques and developed a machine learning application by tweaking Stanford's Universal Dependency Parser.

- Developed internal preprocessing modules in a chatbot.

- Experimented different neural networks on entity tagging to capture user intention on chat data.

HKT   Hong Kong
Software Developer Intern   Feb '18 – Jul '18

- Developed a dashboard web application with ASP.NET MVC and ASP.NET WebAPI as a CRM software. Deployed the web application to IIS Server.

- Researched on machine learning use cases in telecommunication companies. Introduced to Speech-to-Text Conversion by Google and discussed with the manager about my findings.

- Created an application to convert .wav file into binary data, which enables voice data storage. Learned different attributes in a .wav file (encoding, sample rates) and compression techniques (lossless and lossy compression).

TEDx HKPolyU   Hong Kong
Head of IT   Aug '17 – Jun '19

- Designed and developed landing pages for TEDx events. Integrated the website with Facebook Analytics for marketing and lead conversion.

- Developed a mobile application for video streaming and blog sharing services (https://github.com/jackyangara/tedxhkpolyu)

- Managed team members while ensuring both individual and team growth

Projects

Financial News Analytics	Aug '18 – Apr '19

- Developed a hybrid sentiment analyzer using Naïve Bayes classifier and recurrent neural network
- Crawled news data from different sources as training data
- Incorporated technical analysis into the prediction model to penalize fluctuations
- Nominated for "Best Final Year Project" award

Avocado Price Analysis	Aug '18 – Nov '18

- Link: https://github.com/jackyangara/avocado
- Discovered strong correlations between data using multivariate data analysis
- Offered a "Research Assistant" role on the topic of natural language processing

Coauthorship Network Analysis	Aug '18 – Nov '18

- Link: https://github.com/jackyangara/coauthorship
- Learned relationships between different nodes and how to assess different relationships mathematically in a network

Skills

Programming Languages:	Python, JavaScript
Frameworks and Libraries:	PyTorch, Selenium, Django, Falcon, networkx, Matplotlib, Seaborn, Flutter, Node.js, React.js, Express.js, Mongoose, d3.js
Mathematical Tools:	Scikit-learn, NumPy, pandas, SciPy
Technology:	MongoDB, Docker, MySQL, Nginx, Gunicorn, JMeter, Git, Shell Script
Machine Learning:	BERT, Transformer, SVM, Seq2Seq, CNN, RNN, Decision Tree, Naïve Bayes, k-Means, DBSCAN
Languages:	Bahasa Indonesia (Fluent), English (Fluent)