

Julia Lee

Data Scientist

Singapore

- Email me on Indeed: <http://www.indeed.com/r/Julia-Lee/6fd9f823ead19f1a>

- Strong foundation in data manipulation, statistical analysis, programming and modelling
- Hands-on work experience in Data analysis, visualisation and Machine learning projects
- Ability to translate data insights into business decisions and tech solutions
- Ability to think critically, commercially, and strong problem-solving skills
- Strong interpersonal skills and extensive experience of key stakeholder engagement

Proficient and hands-on experience in using SQL, R, Python, Machine Learning libraries, Big data framework, visualization tools:

- Numpy, Pandas, Scikit-learn, Tensorflow, Keras, OpenCV, XGBoost, Spacy, NLTK;
- Big data framework Hadoop/Spark PySpark;
- AWS Cloud services: EC2, S3, EMR, Lambda function, Alexa, Textract;
- Microsoft Azure: App Service, Function Service, Cognitive Services;
- Visualization Tools: Tableau, Power BI, Gephi

Work Experience

Data Scientist

Advice RegTech (FinTech Start-up) - Sydney NSW

December 2018 to Present

- Analysed financial advice documents to find out patterns, summarise auditing result, design report template and dashboard based on findings. (Tableau/PowerBI/Python)
- Performed unstructured data extraction, text data cleaning and analysis: bank statement analysis, application form verification, text classification, named entity recognition, signature recognition
- Solved business problems with self-trained models and APIs: LSTM model, NLTK, Spacy, AWS Textract, Azure OCR etc.
- Deployed models to Cloud using Microsoft Azure Machine Learning WebService, AppService and Function Service.

Product Manager

Sudao Information Technology - Yaoshibang (E-commerce Marketplace)

January 2015 to March 2018

- Led product ideation, product planning and established new strategy and products of B2B/B2C business e-commerce marketplace in retail medicine industry and released Mobile Application, Supplier Management System, Sales Data Management System etc.(from version 1.0).
- Conducted data analysis based on user transaction data and behaviour log to optimise user experience, recommending algorithm and target promotion.
- Designed and optimised online purchasing process and logic based on payment data analysis as well as integrating online payment and loan services provided by the third parties (Alipay/Wechat Pay/Yibao/Yilian).

- Coordinated cross-functional team in designing, engineering, testing and operation and responsible for product development to product launch as well as supporting product operation and monitoring after launch.

Product Manager

China Telecom Bestpay (Online Payment)

May 2014 to December 2014

- Conducted new function planning, data analysis, product specification documentation, prototype design and development monitoring for user experience improvement on online payment and transfer.

- Improved interaction process by function levelling based on the frequency of use.

Associate Consultant

SAMSUNG SDS (Global IT System Consulting Service)

May 2011 to September 2013

- Participated in industry enterprise system optimization project (PLM consulting)to assist professional consultants for gathering requirements, performing data analysis and documenting process flows.

- Participated in pre-sales activities, supported proposal preparation and writing proposal documents, prepared presentation materials, and assisted in conducting workshops and seminars.

- Interacted with the key account, managed relationships with strategic partners.

Education

Master of Data Science in Data Science

University of Sydney - Sydney NSW

March 2018 to July 2019

Bachelor of Science in Mechanics

Sun Yat-sen University - Guangzhou Canton

September 2007 to June 2011

Civil and Environmental Engineering

Dankook University - Seoul

September 2009 to December 2009

Stanford University Online Learning

Skills

Microsoft Office (10+ years), Python Programming (2 years), R Programming (2 years), Sql (3 years), Machine Learning (2 years), Data Science (2 years), Natural Language Processing (NLP) (1 year), Deep Learning (1 year)

Links

<https://github.com/Sweetflowerjulia>

<https://www.kaggle.com/juliaflower>

<https://www.linkedin.com/in/julia-lee-1080887b/>

Additional Information

DATA SCIENCE PROJECTS: PROJECTS:

NLP with Deep Learning:

- Text Summarization: Generating Amazon online review summaries with Transformer model and pre-trained BERT model by fine-tuning.
- Topic Classification: Using Keras library to build LSTM model with GloVe word embeddings.
- Question Answering Model: Extracted features and trained seq2seq with attention model with WikiQA Corpus dataset.
- Chatbot: Built a simple seq2seq model with Microsoft BotBuilder Personality Chat Datasets.

Big data analysis with Spark MapReduce framework on AWS

- Found out top trending videos by analysing viewers, countries, genres and number of clicks.
- Analysed positive and negative reviews of products by summarising ratings, users, review sentences on AWS EMR cluster with Spark

Elo Merchant Category Recommendation Challenge in Kaggle

- Conducted feature engineering and built prediction model with LightGBM, XGboost to predict loyalty score of each customer.

Visual analytics

- Visualised global terrorist data(1970-2017) by combining different type of visualisation using Tableau, Gephi and Python

Image classification

- 1) From scratch: data pre-processing, classification model building and validation using python.
- 2) Using packages: TensorFlow CNN; Scikit-learn: KNN, RandomForest, SVM

Mislabeled data classification (programming language: R)

- Built regression model to predict mislabeled biomedical data provided by FDA competition.