Helena H. Lee

linkedin.com/in/iamleehelenah | leehelenah@gmail.com | (708)-381-0187

◆ 3 years' work experience in the application of Predictive Modeling and Natural Language Processing (NLP)
◆ Live in San Francisco Bay Area (Sunnyvale); Authorized to work in the U.S. without restrictions

Professional Experiences

Oracle - Marketing Data Science team - Data Scientist (Python, PyTorch, SQL)   2020 May - Oct
Oracle's marketing data science team integrates digital marketing solutions that help B2B and B2C marketers
- Contract role focusing on analyzing marketing and sales data, including but not limited to: writing SQL queries, building data engineering pipelines, improving data quality, and training machine learning models
- Trained multi-label classification models on documents in PyTorch to identify the interests of potential customers. The training involves over 6 million textual paragraphs and more than 600 classes.
- Compared the pros and cons of different deep neural networks ranging from RNN, LSTM, Transformer, and BERT, thus reduced the model complexity by 27x and improved the training and inference speed

Living Analytics Research Centre - Research Engineer (Python, PyTorch, TensorFlow, ElasticSearch)   2018- 2020
LARC is the joint research centre of Carnegie Mellon University and Singapore Management University.
- Researched and implemented data science and statistical approaches to solve AI-related problems, predictive modeling, natural language processing, and natural language generation (NLG).
- Proficient in data exploration, data preparation, model development, and model deployment.
- As the first author, published 1 research paper at ACM DPH'19 and 1 demo paper at ACM WWW'20.
- Fine-tuned generative pre-trained language model GPT-2 on 1 million cooking recipe documents
- Deployed an interactive website which allows conditional recipe generation in real-time at RecipeGPT.org, receiving 1000+ visits across 18+ countries and 2 paper citations in 6 months after release
- Proposed and implemented the world's first model to identify recipes which are unhealthful for diabetics
- Compared different document classification models (NBBoW, Word2Vec, RNN) and achieved an F1 of 0.85

BrainCo - **Algorithm Team Intern** (Python, Scikit-learn)   2018
BrainCo is a brain-machine interface start-up founded by Harvard scientists, located in Somerville, MA
- Designed the proof-of-concept experiments to correlate people's attention level with brain wave signals
- Recruited 20+ volunteers across teams within the company to participate in electrophysiology (EEG) experiments

Acer - Research and Development Intern (Python, Tensorflow)   2017
Acer is one of the top personal computer vendors in the world, headquartered in Taiwan.
- Developed a computer vision model (MobileNet, a light-weighted convolutional neural network) using Tensorflow to assess the quality of diabetic retinopathy images with an accuracy of 93%

Education

National Chiao Tung University - M.S., Bioinformatics and Systems Biology   Taiwan 2018
Course Project: Attending Kaggle Corporación Favorita Grocery Sales Forecasting Challenge (Python, LightGBM)
- Conducted feature engineering and modeling on time-series sales data (Time, Store, Item, Quantity)
- Final world ranking: 58/1675 (Top 4%); Won the Silver medal

Skills

Languages: Python (Most of the time; 4+ years' experience), SQL, R, MATLAB, Spark
Expertise: Deep Learning (TensorFlow, PyTorch, Keras), Natural Language Processing (Gensim, StanfordNLP, SpaCy), Machine Learning (Scikit-learn), Cloud Computing (AWS), Version and environment control (Git, Docker)

Selected Publications

RecipeGPT: Generative Pre-training Based Cooking Recipe Generation and Evaluation System
Demo paper accepted by ACM WWW, 2020 [Paper] [Poster] [Source Code]
Estimating Glycemic Impact of Cooking Recipes via Online Crowdsourcing and Machine Learning
Short paper accepted by ACM DPH, 2019 [Paper]