Yahao (Frank) Yan

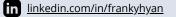github.com/frankyan1119   linkedin.com/in/frankyhyan   yahao_yan@outlook.com

Experience

Software Developer - NLP (Intern)

Deloitte - DSpace Lab

2019

- Developed an NLP model using Python to generate well-formed tax return reports from all sources of interview notes (audio, unformatted text, and hand writings) that saves weeks of manual work
- Created a data labeling UI using Vue and JavaScript to help users to label data more efficiently
- Built an API gateway for label collection to monitor label collection in real-time
- Integrated a weak supervised labeling system to the label collection process to minimize manual data labeling
- Improved frontend and backend functionalities according to user feedback on a weekly basis to deliver a seamless user experience
- Deployed the NLP application on AWS

Data Science Developer (Intern)

Bank of Montreal

2018

- Constructed a RNN/LSTM model using Keras to perform sentiment analysis on social media data for companies in the beauty industry
- Built regression models to predict market sales and revenue change and developed a statistical analysis notebook to investigate in correlations and trends
- Implemented a survey data analyzing API for data profiling, association rules, and cross-reference analysis
- Improved the runtime of aggregating two datasets containing daily/weekly data from 2016 to 2018 for multiple industries by over 90%

Technical Skills

Programming Languages

Python, JavaScript, C++, C, SQL, HTML/CSS

Frameworks and Tools

Django, Flask, Node, React, Vue, Heroku, Git, Docker

Databases

AWS S3, AWS DynamoDB, MongoDB, Neo4j, mySQL

Infrastructures

AWS EC2, AWS Lambda, AWS API Gateway, AWS SageMaker

Machine Learning

TensorFlow, Keras, NLP, RNN/LSTM, Timeseries

Education

University of Waterloo

BMath in Computational Mathematics (Co-op)

2015 - 2020

Projects

Weak Supervised Labeling System (NLP)

- Inspired by "Snorkel" (Stanford) to solve the problem of not having enough labeled data
- Built using python to automatically generate labels for large unlabeled text data sets
- The model generates on average 20% of data labels with 80% accuracy using only 10% labeled data

Quick, Draw!

- Built a drawing app using JavaScript and React that allows user to draw on a white board and provide ratings using an image recognition model
- The model is trained using the Google "Draw" data set

Game Development

- Built a multi-player game "Rock, paper, scissors" and single-player games "Tetris" and "Space shooter" using Python

Weather App

- Developed an asynchronized weather app that provides weather forecasts using JavaScript and Node.js

Interests

Data Science | Natural Language Processing