JAMES LIANG

james.liang.cje@gmail.com | github.com/pooky1955 | linkedin.com/in/jamesyueliang

EDUCATION

DCS (Diploma of College Studies) in Honours Pure and Applied at Marianopolis College 2021 – 2023

Extracurriculars: Founder of the Marianopolis AI Club (recruited 50 active members; created a partnership with the AI Launch Lab; **hosted weekly workshops on Python**, machine learning, and deep learning)

DES (Secondary School Diploma) at Collège Jean-Eudes 2016 – 2021

LANGUAGES (ILR scale)

French (native)
English (full professional working proficiency)
Cantonese (professional working proficiency)
Mandarin (limited working proficiency)

TECHNICAL SKILLS

Machine Learning : Python, Scikit Learn, Tensorflow, Keras, OpenCV, NLTK , Pandas, Kaldi, R, Pytorch
Frontend Development : React, HTML, CSS, Javascript
Backend : Node.js, Flask, ASP.NET CORE, Go, Linux

EXPERIENCES

Project Developer | University of Toronto Machine Intelligence Student Team 2021

Significantly reduce image noise and data usage in video calls using deep learning, image to image translation of spectrograms and speech recognition [Website].

- Reviewed literature on audio resolution and used Wav2Vec to perform **real-time speech recognition**.
- Built **personalized speech synthesizer** using Tacotron and speaker embeddings generated from 15 second audio clip.
- Preprocessed data using **Pytorch** and audiomentations to acquire Mel Spectrograms and noisy versions of audio.
- Worked on using **MaskCycleGan-VC** to generate robust speech from noisy audio and **wav2vec** embeddings.
- Built a website using React and Flask where users can upload audio and download the denoised version.

Data Science Intern | Boehringer Ingelheim Canada 2021

Optimized marketing email schedule using historical data to increase click through rates and open rates.

- Preprocessed dataset using **R** scripts and **Pandas** from raw, untabulated data.
- Compared performance of various machine learning models such as **XGBoost**, **Random Forest**.
- Used state of the art **machine learning methods** (multitarget encoding, Linear Discriminant Analysis, SMOTE) to predict optimal send time.
- Predicted optimal hour range with a top 3 accuracy of 80% (**2.6 times higher** than baseline model).

Biostatistics and Data Science Summer Intern | Boehringer Ingelheim Canada 2020

- Created an **RShiny** app that allows users to download and filter clinical trials from clinicaltrials.gov.
- Created Python programs that use **NLP** and **Google's Tesseract** to extract specific features from scanned informed consents and categorize them according to their features.
- **Contributed to Synthea**, an open-source synthetic patient generator by generating localized (Canadian) demographics data (using Statistics Canada open datasets) [GitHub]

Winner at Pharmahacks 2020 [Devpost] | Hackathon (MILA, Montreal) 2019

- Experimented with different machine learning algorithms such as **XGBoost**, **Naive Bayes** combined with text preprocessing techniques such as **Word2vec**, **TF-IDF vectorizing**, and **count vectorizing**.
- **Deployed** ML algorithms to production using a **Flask REST API**.
- Created a custom trial success prediction by leveraging the successes of similar trials.

PROJECTS

- Created a **YouTube** channel with over 40 videos on machine learning, deep learning, and programming [YouTube].
- Created *Summarize And Ask*, a website that summarizes articles or texts using **MobileBERTSUM** and **BERT** to answer questions about the inputted text [GitHub].
- Created *iNews*, a project for the Montreal Regional Science Fair 2020 to guide the users into detecting fake and real news [GitHub] [*Science Fair canceled due to the 2020 pandemic*].
    - Created a **deep learning model** for stance detection achieving better performance than state of the art by applying **Word2vec embeddings** and **multi-channel convolutional architecture**. It was able to have a general accuracy of 85% compared to 82%.