Ted (Tec Yan) Yap

New York, NY | 309.532.7995 | tyy231@nyu.edu

EDUCATION

NEW YORK UNIVERSITY
MS in Data Science
Expected May 2022 | New York, NY

ILLINOIS WESLEYAN UNIVERSITY
BS in Computer Science
BS in Mathematics
Minor in Finance
May 2020 | Bloomington, IL
Presidential Scholarship (full-ride)
Magna Cum Laude
Cum. GPA: 3.8 / 4.0

LINKS

Github:// **tedyap**
LinkedIn:// **ted-yap**

COURSEWORK

GRADUATE
Intro to Data Science
Probability & Statistics
(Passed Society of Actuaries Probability Exam)
Linear Algebra & Optimization
Machine Learning*
Natural Language Processing*
Big Data* *(* Spring 2021)*

UNDERGRADUATE
Advanced Accounting
Financial Statement Analysis
Artificial Intelligence
Functional Programming
Computer Architecture
Numerical Analysis
Calculus Series *(Teaching Asst 2x)*

SKILLS

PROGRAMMING
Python • R • SQL • C++ • Java
HTML • CSS • JavaScript • Assembly

TOOLS/FRAMEWORKS
AWS • Google Cloud Platform
PySpark • Apache Beam • Redis • Flask

DATA SCIENCE
PyTorch • Tensorflow • SciKit-Learn
SciPy • NLTK • Seaborn • Tableau
Scrapy • BeatifulSoup • Selenium

SOFTWARE ENGINEERING
Git • Agile • Docker

EXPERIENCE

MACHINE LEARNING ENGINEER INTERN at Insight Data Science
Jun 2020 - Aug 2020 | New York, NY

- 32 out of 1500 applicants chosen to be an AI Fellow Summer 2020.
- Built **SkinCredible**, an end-to-end machine learning pipeline that helps dermatologists analyze and monitor skin conditions over time.
- Trained an image classifier in **Tensorflow** that uses **CNN** and **LSTM** to discover time series patterns in 500GB of proprietary image dataset.
- Parallelized model training across multiple cloud GPUs to reduce training time.
- Leveraged **AWS Comprehend** to extract sentiments from dermatologists' notes and create proxy-labels for unlabelled dataset.
- Deployed model as a **REST API** hosted on **AWS EC2** in a Docker container.

DATA ENGINEERING INTERN at Advance Trading Inc
Jan 2020 – May 2020 | Bloomington, IL

- Developed modularized scripts for **ETL pipelines** with **Apache Beam** and **Google Cloud Dataflow** to migrate data from relational to **NoSQL** cloud databases.
- Built a database cache with **Redis** to reduce query processing time.
- Implemented a 'serverless' architecture with **Google Cloud Functions** to automatically pull data from **BigQuery** and generate daily margin reports, saving 5+ hours of analysts' time per day.

RESEARCH STUDENT at Illinois Wesleyan University
Jan 2020 – May 2020 | Bloomington, IL

- Conducted research on using Generative Adversarial Networks (GANs) to detect anomalies in text.
- Developed and published **ARAE-AnoGAN**, a mix model of autoencoders and GANs with word embeddings to detect short anomalous sentences in a document. Built in **Tensorflow** with custom loss functions.

OPERATIONS & TECHNOLOGY ANALYST INTERN at Citi
Jun 2019 – Aug 2019 | Kuala Lumpur, Malaysia

- Performed comprehensive exploratory analyses on credit card transactions using **Python** with **SQL** queries to uncover predictive signals of fraudulent accounts, in order to minimize financial loss.
- Automated fraud analyst scorecard generation using **VBA**, reducing human error and the time to generate feedback from one month to a day.

HIGHLIGHTED PROJECTS

NYC RESTAURANT INSPECTION | Data Science Class Project
Sep 2020 – Present | New York, NY

- Conduct feature engineering on restaurant inspection dataset to predict whether a restaurant will violate food safety regulations.
- Extract sentiments from Yelp reviews and test various classification models with **Spark MLlib** (Logistic Regression, SVM, Decision Tree) to optimize AUC.
- Process and join restaurant inspection dataset with Yelp reviews using **PySpark**.

ANSWER CORRECTNESS PREDICTION | Kaggle Competition
Aug 2020 – Present | New York, NY

- Train an attention based transformer neural network with **PyTorch** to predict the correctness of students' answers to multiple-choice questions.