NEW YORK UNIVERSITY

Tom Yang

yy2949@nyu.edu

New York City, NY

10003

SKILLS

- Professional in Python. Familiar with other languages/tools: bash, SQL, Hadoop, C, etc.
- Familiar with both Deep Learning Python modules (PyTorch, Tensorflow, etc.) and traditional machine learning algorithms (using sklearn)
- Have experience working on several NLP and CV projects
- Analyzing mathematical theory of ML/DL algorithms
- Professional in data-preprocessing jobs (data cleaning, data mining, web scraping, etc.)
- Good at Data Visualization (using Python/Excel)
- Strong learning ability: self-learned Hadoop online, learning C++ and AWS recently

EDUCATION AND TRAINING

Bachelor of Science: Data Science & Math

New York City, NY

- Dean's List Honoree Fall 2019 - Spring 2020
- Full GPA(4.0) in Freshman Year, 3.97 in Sophomore Year

WORKING EXPERIENCE

- Undergraduate Researcher (Data Science)

New York University

Starting From Feb 2022

- Work on using data science (causal inference, machine learning) to quantify traditionally hard problems in political science.
- Use deep active learning (semi-supervised learning) to do image classification for an image dataset with only limited labeled data.

- T.A. and Grader for Introduction to Machine Learning

New York University

Originally From Jan 2022 to May 2022 (now changing to Aug 2022 to Dec 2022 due to personal reasons)

- Teach the recitation, where students learn how to implement the theoretical algorithms covered in lecture into applicable codes with Python
- Design and grading homework that include both theory and coding.

OTHER RELATED EXPERIENCE

- Bud Challenge 2021 – Data Analysis track

Build a B2B, ML-based recommendation system for retailers. Got 3rd Place in the Final Round (More than 50 teams in the data track in total)

- NYU Datathon Spring 2021

Build an OCR model to read through a century of Ads from American Newspapers, and then use ML(NLP) to find some patterns of the Ads.

- Project: An analysis with ViBERT

Apply BERT in Computer Vision with different CV problems. Evaluate and compare different architectures and results. Work in-progress.

- Project: NLP Dataset Design

Current topic: Design, build, and evaluate a dataset of human-written codes for training. Aiming for using this dataset to build a model for coding plagiarism detection with GPT-3. Work in-progress.