

Johnny Ma

phone: 914-434-4170. email: johnnyma@nyu.edu. address: 275 Park Ave, Brooklyn, NY. 11205.

Education

New York University, Center for Data Science

New York, NY

M.S. IN DATA SCIENCE, NLP TRACK

Sep. 2020 - Jun. 2022

- Current GPA: 4.0/4.0. Projects include building a Bandit recommendation system and VAEs for music generation.

The University of Chicago

Chicago, IL

B.A. IN ECONOMICS WITH HONORS, ART HISTORY MINOR

Sep. 2014 - Jun. 2018

- Cumulative GPA: 3.4/4.0. Member of Dean's List. Recipient of 2018 BFI Award for Outstanding Undergraduate Service.
- B.A. Thesis under Leo Burszty: Network Formation and Social Spillovers in Film Consumption on *Letterboxd.com*.
 - Using 10,000 scraped film diaries, found that site wide popularity drives viewing behavior more than friend network information.

Skills

Python Programming PyTorch, Sklearn, Librosa, FFmpeg, Transformers, Spacy, Fairseq, Selenium, Numpy, Pandas, Seaborn.

General Programming R (Web Scraping, Regression, Visualization), SQL, ArcGIS (geo-locating, ArcPy), Excel, VBA, Java, STATA, HTML.

Languages Fluent in English, 7 years of French knowledge, Mandarin spoken at home.

Experience

Computational Intelligence, Learning, Vision, and Robotics Lab (CILVR at NYU)

New York, NY

RESEARCH ASSISTANT UNDER HE HE AND SAM BOWMAN

Sep. 2020 -

- Generating minimal edit counterfactual and adversarial MNLI examples for model robustness via crowdsourcing.
- Extending Fairseq BART mask infilling for conditional generation of new MNLI hypotheses. Responsible for 100% of project Python code.

American International Group (AIG)

New York, NY

ACTUARIAL ANALYST ON COMMERCIAL CASUALTY TEAM UNDER ASHISH THAKUR AND WILLIAM ZHU

Jun. 2018 - Jun. 2019

- Used Python and R to analyze loss models for commercial auto trucking insurance group.
- Managed external team in data collection using SQL. Responsible for team finances and presentations to AIG executives.
- Trained insurance loss type classifier using natural language processing. Built Django and Excel back-end for internal usage.

Federal Reserve Bank of New York

New York, NY

SUMMER RESEARCH ANALYST ON MICROECONOMICS TEAM UNDER MAXIM PINKOVSKIY AND RAJASHIRI CHAKRABARTI

Jun. - Aug. 2017

- Used STATA, ArcGIS, and Python to clean, visualize, and run regressions on large-scale Equifax and administrative datasets.
- Worked closely with senior economists to identify models, drive project direction, and narrate complex regression results.
- Found negative impact of Great Recession on credit activity and positive impact of ACA rollout on educational attainment.

Projects

Predicting Screenplay Budget from Script Text

FOR DS-GA 1001: INTRODUCTION TO DATA SCIENCE

Oct. 2020 - Dec. 2020

- Extracted raw text and financial information of 1,000 scripts from IMSDB.com and BoxOfficeMojo.com.
- Built low-medium-high screenplay budget classifier using lexical features, named entity pipelines, and deep word embeddings.
- Obtained 73% F1 score using Logistic Regression, GBM, and BERT ensemble model. Awarded top 10 project in class of 140.

Stand-up Comedy: Quantifying Humor and Identity

FOR SUBMISSION TO JOURNAL OF CULTURAL ANALYTICS

Jun. 2019 - Jun. 2020

- Extracted 100 Netflix stand-up routine videos and subtitles, ran NN laughter detector, created textual and audio features.
- Created Logistic Regression classifier for laughter prediction, F1 score of 62% slightly improves on literature.
- Tagged identity jokes for textual causal matching regression. Found audiences laugh 35% longer at females telling female-based jokes.

Extracurricular

Center for Data Science Leadership Circle

New York, NY

GENERAL MEMBER

Sep. 2020 -

- Organizing social gatherings, moderating NYU alumni panels, and holding office hours for CDS students.

President of OEconomica (Undergraduate Economics Society)

Chicago, IL

PREVIOUSLY VICE PRESIDENT, RESEARCH DIRECTOR

Jun. 2015 - Jun. 2018

- Directed club activities by leading board, organizing faculty events, securing department and private funding, and expanding Annual Econometrics Game to national Ivy League circuit.
- Developed and taught research cohort program and R workshop curriculum. 60-80 active members.