

# VANESSA (ZIWEI) XU

1-929-988-3639

[vanessaziweixu@gmail.com](mailto:vanessaziweixu@gmail.com)

<https://www.linkedin.com/in/vanessa-ziwei-xu/>

<https://github.com/VavaXu>

## EDUCATION

**New York University, Center for Data Science**

**New York, NY**

*M.S. in Data Science*

*May 2023*

- **Relevant courses:** Probability & Statistics, Linear Algebra, Machine Learning, Big Data, Text as Data, Programming, Natural Language Processing, Capstone Project
- President – CDS Diversity Affinity Group; Mentor for Undergraduates - Data Science Club

**New York University, Steinhardt**

**New York, NY**

*B.Mus. in Music Business; Minor in Data Science, Performance Studies*

*May 2021*

- **Relevant courses:** Causal Inference, Database Design & Implementation
- Music Business Student Mentor

## SKILLS AND INTERESTS

- **Technical Skills:** Shell, Git, Python (PyTorch, NumPy, Pandas, Scikit-learn), R (quanteda), SQL (T-SQL, MySQL, SQLite, MySQL Workbench), NoSQL, MapReduce, Hadoop, Spark, Dask, Advanced Excel, Final Cut Pro
- **Languages:** Chinese (Fluent), French (Conversational)
- **Interests:** Singing (Final round auditions of Sing! China and performed at venues and theaters); Drama Writing, Director and Performer (LAMDA Gold Medal Distinction)

## PROJECTS

**Extracting causal political narratives from text – Deep-Learning based NLP Capstone Project**

*Sep 2022 - Dec 2022*

- Constructing machine learning pipeline for causal relation extraction from text as well as causal structure discovery in order to automatically extract narratives from a large corpus of 1 million political news articles
- Implementing text preprocessing methods, apply and test pipeline on data, draw informative conclusions on its performance, and suggest potential improvements to explore how politically aligned media outlets propagate different narratives about the same set of facts, how selective reporting is structured, and the issue of polarization
- Utilizing Deep-Learning based NLP methods and their Python implementations including BERT, GPT-3, as well as Deep Learning libraries including PyTorch or TensorFlow

**Using Human Rights Texts to Predict Country Income Category – NLP Project (R)**

*Mar 2022 - May 2022*

- Explored how effective human rights texts can be in predicting monetary values – countries' income levels
- Applied text analysis methods:
  - Text pre-processing: stemming, removing punctuation, stopwords, and numbers, converting to lowercase, replacing apostrophes with empty string
  - Tokenized, filtered, and made features into document term matrix in order to measure features by word counts
- Built Naïve Bayes Multi-class Classification model with Laplace smoothing in R to train 10,000+ human rights text data files
- Made country income classification predictions and achieved 82.46% accuracy; and generated a 10-page report

**Collaborative-filter Based Recommender System on MovieLens –Group Project (Python & Spark)**

*Mar 2022 - May 2022*

- Built Popularity Baseline Model and Spark's Alternating Least Squares Model (ALS) after preprocessing and partitioning to predict the top 100 movie recommendations for each user from the MovieLens dataset with 280k rows
- My contribution:
  - Built Single Machine Implementation–LightFM model, an implicit feedback recommender, and further improved model performance with an absolute increase of 1.4% in precision\_at\_k and improved model efficiency by achieving less runtime
  - Used hyperparameter tuning on both ALS and LightFM model to achieve the best possible model result
- Implemented fast search algorithm with Annoy and established a brute force algorithm to further decrease runtime

## WORK EXPERIENCE

**New York University, Center for Data Science**

**New York, NY**

*Graduate Teaching Assistant, Causal Inference*

*Jul 2022 - Aug 2022*

- Created course specific lesson plans and facilitated lab sessions in R language for Causal Inference
- Prepared weekly assignments, quizzes, exams and solutions for a class of 80 students
- Held Office Hours to help students work through complex questions and responded to inquiries and concerns
- Coordinated and released grading with the graders with careful consideration of confidential record

**New York University, Center for Data Science**

**New York, NY**

*Graduate Assistant*

*May 2022 - Aug 2022*

- Researched and generated detailed reports of alumni career placement stats with Python
- Reviewed and revised Career Development resources for undergraduate and graduate data science students
- Assisted with the development of internal undergraduate and graduate student portals, ensuring that website content was up-to-date and published content met guidelines
- Worked closely with the Academic & Student Affairs team on event logistics and student community building, including New Student Orientation and programming focused on diversity and inclusion