

# BLEND 360

## Ziyu (Zach) Xiong

### Data Science Associate

#### TOP 4 EXPERTISE

1. Machine Learning / Deep Learning
2. Computational Social Science
3. Time Series Analysis / Causal Inference
4. Computer Vision & NLP

#### TOP 4 SOFTWARE SKILLS

1. Python / PySpark / PyTorch
2. SQL
3. Tableau, R
4. Cloud Computing Platform (Databricks)

#### EDUCATION

- MS, Data Science, University of Rochester, Goergen Institute for Data Science, Rochester, New York
- BS, Computer Science, Oregon State University, Corvallis, Oregon

#### SUMMARY

Zach is a Data Science professional who has 1+ year academia research experience, including Computational Social Science, Natural Language Processing and Computer Vision, where two papers are published in 2021. Zach has a great interests on data science for social goods, especially hybrid data science with other industries.

#### SELECTED EXPERIENCE

##### Research Experiences

- **Social Media Research:** Streamed data from social media website (such as Twitter, Reddit). Cleaned and preprocessed large dataset (>100 GB). Applied advanced models and packages to Infer demographic features, such as Ethnicity, Age, and Gender. Trained a BERT model to predict sentiment. Applied and fine-tuned Latent Dirichlet Allocation (LDA) to extract topics in text-based dataset.
- **Homelessness Exploration:** Preprocessed dataset with external data sources, built data pipeline on Databricks platform. Performed state-level analysis to explore and visualize the homelessness trend across the states. Applied sharp-null test from causal inference and extracted the confounder effect of urbanization. Investigated policies impacted by different state governor election results using Regression Discontinuity Design.

##### Project Experiences

- **Multimodal Video Violence Detection:** Applied video backbone I3D to get the video representation for optic flow and RGB, performed VGGish on the mel-spectrum to get video embedding. Use Transformer based neural network to train the model with multiple instance learning. Conducted ablation studies on the contribution of different modalities.
- **RNA Structure Detection:** Preprocessed the dataset of RNA sequences over 60k samples with hundreds of features. Detect Pseudoknot structure using random forest and logistic regression with PCA. Achieved AUC ROC score of 70.2%.