Ishita Gopal

Tel: +1 (925) 394-8793 Mail: ishitagopal@gmail.com Web: ishitagopal.github.io

Data scientist with 6 years of experience using machine learning and statistics to derive insights from real world data. Highly collaborative, skilled at research & communicating technical topics.

SKILLS

Tools and Languages: Python, R, SQL, Git, AWS, Azure
Data Science & Statistics: Machine learning, natural language processing (NLP), deep learning, network analysis, inferential statistics, hypothesis testing, experimental design, time–series, panel-data models

WORK EXPERIENCE

Data Science Intern, Aware HQ **Summer 2022**

- Developed and deployed a credit card detection model with AWS SageMaker to flag sensitive data sharing in digital workspaces. Used deep learning (EfficientNets) for transfer learning on hand-labeled data, utilized data augmentation to reduce overfitting, resulting in a 90% accuracy rate.

Doctoral Research Scientist, Center for Social Data Analytics **Aug 2018 - Current**
Produced 3 publications, 4 working papers, presented at 8 conferences.

Predictive Modeling, NLP and Network Analysis with Social Media Data:

- Used high-performance computing and simulations to analyze tie formation patterns in Twitter networks of 4K US legislators. Showed that gender and race are reliable predictors of inter-state ties. Employed penalized regressions and cosine similarity on 10K bill texts to generate robust measures of policy similarity.
- Developed an NLP pipeline and fine-tuned BERT to identify COVID-19 discussions in 1M+ tweets by US legislators, achieving an 85% F1 score that outperformed Random Forest/XGBoost by 8%. Used results in hierarchical models and showed that health/policy indicators strongly correlated with pandemic discussion.
- Constructing text classifiers and utilizing topic models to identify protest discussions in 3M+ posts extracted from group chats on Telegram. Employing panel regressions to establish the association between online protest discourse and subsequent on-ground repression.

Causal Inference for Understanding Behavior:

- Designed a field experiment (1000 subjects) to test how peer influence impacts US legislators' willingness to support green policies. Designed information-based treatments, accounted for latent network effects, and used backbone extraction methods to identify peers.
- Executed 2 online survey experiments (9000 participants) and showed that in-group commentary on Facebook increases acceptance of misinformation. Analyzed results using linear regression and GLMs.

Economist, The Energy & Resources Institute **Aug 2016 – Aug 2018**

- Developed time series (ARIMA) models for electricity demand forecasting.
- Used scenario modeling to forecast the impact of renewable uptake on coal capacity growth in India. Results provided policy assessment support to the Indian Government.

EDUCATION

Ph.D. Social Data Analytics & Political Science, *Pennsylvania State University* Expected 2023
M.Sc. Economics, *University of Warwick* 2015
BA (Hons) Economics, *Miranda House* 2014