Ishita Gopal

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

SUMMARY

Data scientist with strong academic experience, adept at building hypothesis-driven solutions, leveraging diverse data types (text, networks, and audio), and using cutting-edge ML, statistics, and causal inference to solve real world problems.

Tools and Languages: Python, R, Git, SQL, AWS

DATA SCIENCE EXPERIENCE

Data Science Researcher ***Jan 2024 – Current***
Transdisciplinary Institute in Applied Data Sciences, Washington University in St Louis

- **Developing transformer-based deep learning models to detect attack political advertisements on YouTube (achieved 90% accuracy).** Investigating if audio features can exclusively be used for classifying ads, aiming to improve the efficiency of analysis by eliminating the need for complex (multi-lingual) text analysis.
- **Conducted multi-session data science workshops in Python for an interdisciplinary group of faculty and graduate students.** Developed comprehensive tutorials on web scraping, machine learning (classification, regression, clustering), NLP and LLMs. **[Tutorials]**

Doctoral Researcher ***Aug 2018 – Dec 2023***
Center for Social Data Analytics, Pennsylvania State University

- **Built machine learning classifiers and panel regression models to quantify how health and policy indicators predict policymakers' discussions of COVID-19 on Twitter.** Integrated government health and policy databases with Twitter data. Trained a large language model (BERT) to identify COVID-19 discourse in 1M+ tweets (F1 score of 85%). Enhanced F1 metric by ~ 10% compared to random forest and XGBoost algorithms. **[Published paper]**
- **Led a team of 6 and developed predictive models to analyze group dynamics and communication patterns of 4K+ policymakers on Twitter.** Collected ~300K observations of legislator interactions using Twitter's API. Created advanced visualizations, identifying clustering by party and state. Used permutation models on high- performance computing to show gender and race as significant predictors of cross-state interactions. [**Conditional accept paper**]
- **Conducted 2 online experiments (9K participants) to evaluate MTurk's recruitment limitations, providing crucial guidance for academia and industry.** Analyzed Facebook comments' impact on recalling misinformation, revealing bias from MTurk samples when treatment effects vary by age digital literacy. **[Published paper]**
- **Designed and implemented an email experiment (1K subjects) to test impact of peer effects on policy support diffusion, while accounting for network effects**. Tracked cosigning behavior of 6K+ US state legislators, identified peers using backbone extraction methods. Used zero-shot text classifiers (90K bills) to identify relevant treatment policies and built regression models to analyze email response and click behavior. **[Working Paper]**

Data Science Intern ***May - Aug 2022***
Aware HQ, Columbus

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

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

- Worked with government stakeholders to develop time series (ARIMA) models for electricity demand forecasting.
- Conducted scenario modeling to forecast impact of renewable uptake on coal capacity growth in India.

EDUCATION

Ph.D. Social Data Analytics & Political Science **Pennsylvania State University, USA** 2023
M.Sc. Economics, **University of Warwick, UK** 2015
B.A. (Hons) Economics, **Miranda House, India** 2014