Ishita Gopal, PhD

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

SUMMARY

Data scientist with expertise in machine learning, natural language processing (NLP), causal inference and statistics; skilled at handling structured and unstructured data (text, audio, networks); with a strong publication record.

Tools and Languages: Python (PyTorch, transformers, scikit-learn, Pandas), R, Git, SQL, AWS

DATA SCIENCE EXPERIENCE

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

- **Deep Learning for Attack Ad Detection:** Currently developing text and audio based deep learning models to identify negative advertising on YouTube (90% Accuracy). Investigating if audio features can exclusively classify ads, improving efficiency by eliminating the need for complex multilingual text analysis. [**Conference Presentation**]
- **Data Science Workshops**: Designed and led Python workshops on APIs, web scraping, machine learning (regression, classification, clustering), NLP and large language models (LLMs) for faculty and graduate students. **[Tutorials]**

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

- **COVID-19 Discourse Detection:** Built machine learning and statistical models to classify 1M+ tweets and quantify impact of health and policy changes on COVID-19 discourse, resulting in publication. Fine-tuned BERT LLM (85% F1 score). Improved performance by 10% over random forest and XGBoost. Used panel regressions to test hypotheses revealing significant content differences based on party affiliation and high correlation with state case counts. **[Paper]**
- **Text & Network models for Behavior Prediction:** Led a team of 6 to develop predictive models to identify who interacts with whom among policymakers using 300K+ observations from Twitter. Utilized topic models and lasso-assisted cosine similarity on 50K+ documents to quantify policy overlaps. Created visualizations that show clustering by party and state. Implemented permutation models on HPC, identifying gender and race as key predictors. **[Paper]**
- **Experimental Test of Peer Influence Diffusion:** Conducted a novel email experiment (1K subjects) that accounts for network effects, and quantified impact of peer effects on policy support diffusion. Tracked 6K+ policymakers, used backbone extraction to detect peers, employed zero-shot text classifiers on 90K bills to identify treatment policies and built regression models to analyze email response and click behavior. **[Working Paper]**
- **Experimental Evaluation of MTurk Samples and Misinformation Recall:** Ran 2 online experiments (9K participants) and assessed MTurk's recruitment limitations and the impact of Facebook comments on misinformation recall, resulting in publication. Demonstrated MTurk samples exhibit bias when treatment effects vary by age and digital literacy, providing critical guidance for academia and industry for analyzing social media behavior. **[Paper]**

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

- **Convolutional Neural Networks for Sensitive Data Detection:** Developed and deployed a credit card detection model for digital workspaces. Used 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, **Pennsylvania State University, USA** 2023
M.Sc. Economics, **University of Warwick, UK** 2015
B.A. (Hons) Economics, **Miranda House, India** 2014