Ishita Gopal, PhD

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

WORK EXPERIENCE

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

- **Data Pipeline:** Developed Python and SQL pipelines that integrate web scraping, APIs, and transformer models to automate the search, retrieval, and transcription of YouTube political advertising data, supporting multiple projects.
- **Ad Classification & Strategy Prediction:** Trained large language models on ad text to classify 50K+ political ads and detect negative advertising on YouTube, achieving a 90% F1 score. Analyzed content variations and topic focus in election ads using structural topic models. Integrated findings into regression models to evaluate impact of ad spend and advertiser characteristics on content strategy and negativity, resulting in research outputs and working papers.
- **Python Workshops:** Led workshops on web scraping, APIs, machine learning, NLP (e.g. classification, topic models), and deep learning for text data (e.g. LLMs, transfer learning, zero-shot and few-shot learning). **[Workshops]**

Senior Researcher (Ph.D. Researcher), Center for Social Data Analytics **Aug 2018 – Dec 2023**
Pennsylvania State University (Select Published Projects)

- **Discourse Detection & Attention:** Developed ML and statistical models to classify 1M+ tweets to identify factors predicting attention to COVID-19 discourse and variations in issue focus across political groups. Fine-tuned a BERT-based LLM, achieving an 85% F1 score and a 10% improvement over Random Forest and XGBoost. **[Paper]**
- **Causal Test of Peer Influence Diffusion:** Led an experiment with 1K subjects to evaluate the impact of social connections on decision-making among U.S. politicians, controlling for network effects. Built a database of individual interactions and used network analysis to identify subject pairs for the study. Applied zero-shot models on 90K bills to identify treatment policies and ran regressions to analyze response rates and click behavior. **[Paper Accepted]**
- **Experimental Test of MTurk and Misinformation Recall on Facebook**: Conducted 2 online experiments with 9K participants to assess MTurk's recruitment limitations and the impact of Facebook comments on misinformation recall. Revealed biases in MTurk samples related to age and digital literacy, providing guidelines and best practices for deriving accurate insights from online samples. [**Paper**]
- **Text & Network Models for Behavior Prediction:** Led a cross-functional team of 6 and implemented network regressions on HPC to identify predictors of online engagement among U.S. politicians using 300K tweets. Measured peer similarities with cosine similarity on 50K bills. Developed visualizations to reveal interaction clusters. **[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.

Data Analyst **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

SKILLS

Tools and Languages: Python (Pandas, scikit-learn, PyTorch, transformers), SQL, R, Git, AWS
Skills: Machine Learning, Natural Language Processing, Causal Inference, Experiments, Statistics