Ishita Gopal

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

SUMMARY

Data scientist with 5 years of experience translating abstract questions into concrete data problems. Skilled at leveraging unstructured/structured data and utilizing advanced techniques in machine learning, statistics, and experimentation to extract insights and guide decision-making.

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, causal inference, time–series, panel-data models

DATA SCIENCE EXPERIENCE

Doctoral Research Scientist, Center for Social Data Analytics (C-SoDA), University Park
August 2018 – Current

- Led a team of 6 and implemented network simulation models on high performance computing to analyze complex behavioral patterns of 4K policymakers'. Used numeric/textual features and identified predictors of ties amongst individuals. Created advanced visualizations. Produced a working paper under review for publication.
- Developed machine learning and statistical models to quantify impact of health/policy indicators on policymakers' response to public health crisis. Leveraged deep learning model (BERT) to identify COVID-19 discourse in 1M+ tweets (F1 of 85%). Improved classification performance over random forest and XGBoost by ~10%. Built panel regression models for hypothesis testing. Published results in high-impact journal.
- Implemented and analyzed data from 2 online experiments (9,000 participants) to evaluate limitations of MTurk. Demonstrated MTurk should not be used when treatment effect heterogeneity is expected in age/digital literacy. Produced journal article that provides guidance for academia/industry.
- Designed causal inference study (1000 subjects) to estimate impact of peer effects on support for environmental policies. Used backbone extraction methods to account for underlying network effects. Implemented zero-shot text classifiers on 90K bills to identify treatment policies. Results provide behavioral insights about policymakers.
- Worked in cross-functional team to develop machine learning models aimed at detecting election integrity discourse. Evaluated performance of various language model architectures - BERT, RoBERTa, and XLNet. Implemented active learning approach to optimize model performance and decrease model training time.
- Created instructional materials (book, chapters, and tutorials) on using data science methods to analyze text and network data. Conducted coding workshops in Python/R to teach undergrads and peers.

Data Science Intern, Aware HQ, Columbus
May - August 2022

- Developed and deployed a credit card detection model with AWS SageMaker to flag sensitive data sharing in digital workspaces, to be used in Aware's product.
- 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, The Energy & Resources Institute, Delhi
August 2016 – August 2018

- Developed time series (ARIMA) models for electricity demand forecasting.
- Conducted 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, USA** Expected 2023
(Awards: Princeton University Dissertation Scholar, C-SoDA Predoctoral Fellow)
M.Sc. Economics, **University of Warwick, UK** 2015
B.A. (Hons) Economics, **Miranda House, India** 2014