Summer School on Econometrics and Machine Learning
September 6 - 10, 2022

OVERVIEW OF THE SUMMER SCHOOL

This Summer School is an initiative under the FLAIR 4BD Project that seeks to **F**oster **L**earning, **A**nalysis, **I**nsights and **R**esearch **4 B**ig **D**ata. The initiative is supported by a grant from the Bill and Melinda Gates Foundation. The Summer School consists of a series of twenty two lectures which will provide a theoretical and empirical foundation in the field of Econometrics and Machine Learning.

RESOURCE PERSONS

Surabhi Singhal is a data science professional with six years of experience in developing business solutions for major healthcare, insurance and financial services organisations using a blend of econometrics, data science algorithms and statistics. She was chosen as a doctoral candidate at Stony Brook University, New York. She is currently working in data science division of a major financial services organisation. She is an alumnus and visiting faculty of Shiv Nadar University.

S Chandrasekhar is a faculty at the Indira Gandhi Institute of Development Research, Mumbai, India. He has been in the Institute for the last eighteen years. His research interests include understanding Income Dynamics in Rural & Urban India; Structural Transformation; Urbanization; Mobility & Labour Market Outcomes; Education Skills & Employability. He is an alumnus of The Pennsylvania State University and Delhi School of Economics.

Subrata Sarkar is a faculty at the Indira Gandhi Institute of Development Research, Mumbai, India. He has been in the Institute for the last thirty years. His specialisation is in econometrics and he works on applied problems on corporate governance, corporate finance and risk modelling. He is an alumnus of Presidency College Calcutta and University of Southern California.

SUMMER SCHOOL OUTLINE

Topic 1: What is Big in BIG Data

- Primer to BIG data
- Econometrics and ML: What They Do and Don't
- The Foot Soldiers of Econometrics
- The Missiles of ML
- Harnessing Images, Text and Voice
- Expectations from the Summer School

Topic 2: The Workhorse of Econometrics

- Ordinary Least Square Regression
- Probit and Logit: Modelling Binary Classification
- Cumulative Logit: Modelling Multiple Classification
- Multinomial Logit: Modelling Unordered Classification
- Datasets and Hands-on Session

Topic 3: Using Python for Statistics and Econometrics

- Introduction to the Python Language
- Introduction to Statistical Theorems
- OLS Model Assumptions
- Implications of Violations of OLS Assumptions using Simulated Data

Topic 4: Alternate Data

- Unavailability of survey or administrative data.
- What is alternate data?
- Questions which can be answered from alternate data?

Topic 5: Understanding the Labour Market from Job Postings

- Discuss of selected papers using data from online postings
- Understand the trends and patterns in skill assets sought by employers

Topic 6: PDF to Excel, Word Scrapping

- Hands-on Session
- Converting PDF to excel
- Basics of web scraping

Topic 7: APIs useful for Social Scientists

- Primer to BIG data
- How do we get started?

Topic 8: Introduction to Machine Learning Techniques

- Decision Tree
- Random Forest
- SVM
- XGBoost

Topic 9: Unstructured Data Evolution

- Data science things you can include in research
- Types of data sets to include in research
- Where can you use DS and where can you not

Topic 10: Advanced Visualisation

- Economic applications and recognition from representation
- Importance of visuals, advanced visualisation techniques
- Hands-on - Sankey diagram, Bubble Chart

Topic 11: NLP for Economic Problems

- Using NLP in Economic issues
- Text data analysis
- Different methods of vectorisation
- Live project: Twitter sentiments

Topic 12: AI for Economic Problems

- Where can AI be used in Economics
- What is Image Processing
- Live project – Covid identification using chest X-ray

Closing

What Lies Ahead?

The IGIDR - FLAIR 4BD team will be creating an Econometrics and Machine Learning Network (EMLN) soon. Details will be provided on the IGIDR website. Please make sure to update your Email ID with us.