

## 入門：社会科学のための機械学習

日時：2022年3月3日（木）10:30～17:00

場所：オンライン開催

料金：一般 5,000 円、学生 2,500 円

講師：川田恵介（東京大学）

定員：35 名

### ■本コースの内容

機械学習、特に教師付き学習の社会科学への応用について基礎から学ぶコースです。教師付き学習は、データを“柔軟”に用いることで、よりデータ主導方の予測モデル構築を可能にします。学部/大学院レベルの計量経済学や政治学の教科書でもすでに紹介されており、標準的な分析ツールとなっております。

本コースでは、機械学習の基礎から社会科学の伝統的な関心である因果推論/比較研究への応用まで、高度な数学を用いずに紹介します。データを扱っている研究者・分析者のみならず、これからデータ分析をやりたい院生・学生の参加も歓迎いたします。

具体的内容としては、教師付き学習の各手法（LASSO 推定や Random Forest など）を紹介します。またセミパラメトリック推定法と併せて用いることで、平均効果の推定やその異質性探索に応用できることも論じます。

本年は、社会科学において特に有益だと考えられている、Stacking 法も紹介します。また近年に懸念が高まる機械学習の安直な応用による「誤発見リスクの爆発的増大」とその回避策について、重点的に解説します。

さらにR言語についての無料の統計クラウドサービスである R cloud と個票データを用いて、実際に自分の手で機械学習を用いた実証研究を体験してもらいます。

### ■次のような方におすすめです

- ・機械学習、とくに社会科学研究への応用に関心がある方
- ・「AI」を用いたデータの解析結果」を批判的に検討したい方

### ■注意事項

- ・SSJ データアーカイブのデータを利用するため、大学または公的研究機関所属の研究者・学生（学部生も可）、SSJ データアーカイブへデータを寄託されている民間研究機関の方のみを対象とします。
- ・R cloud への事前登録・必要パッケージのインストールを事前に済ませてください。登録・インストール方法を紹介した動画を事前に配布します。

### ■本コースの日程

「機械学習の外観」、「教師付き学習」、「コントロール変数法への応用」、「効果の異質性分析

への応用」

\* 進捗によって、内容が若干変わることがあります。