

Shekhar Mittal

🏠 Palo Alto, CA | ✉️ shekhar.mittal@gmail.com | ☎️ 734.780.1120 | Visa Sponsorship NOT Required

EDUCATION

PHD, ECONOMICS

UCLA ANDERSON SCHOOL OF MANAGEMENT

June 2018 | Los Angeles, CA | 3.7

Thesis: Essays on Taxation in Emerging Economies [[pdf](#)]

MASTER OF PUBLIC POLICY, HARRIS SCHOOL

UNIVERSITY OF CHICAGO

June 2013 | Chicago, IL | 3.8

BACHELOR OF ENGINEERING, COMPUTERS

UNIVERSITY OF DELHI

June 2007 | Delhi, India | First Division

SKILLS

CAUSAL INFERENCE

A/B Testing (RCTs)
Non-Experimental Methods
Statistical Modelling
Policy Evaluation

MACHINE LEARNING

Feature Engineering
Real-world Model Evaluation
Sklarn, H2O, Turi/GraphLab

PROGRAMMING/TOOLS

Python, NumPy, Pandas, SciPy, SQL
GeoPandas, Matplotlib, Seaborn
PySpark, Stata
R, C, Hadoop
GitHub, Amazon Web Services

COURSERA CERTIFICATIONS

ML Foundations: A Case Study Approach
ML: Regression
ML: Classification
ML: Clustering & Retrieval

MEDIA

Indian Express, 2019. [[link](#)]
Financial Express, 2019. [[link](#)]
Times of India, 2019. [[link](#)]
IGC Blog, 2017. [[link](#)]

LINKS

in: [shekharmittal](#)

tw: [shekhar_m](#)

www: www.shekhar.me

g: [shekharmittal](#)

EXPERIENCE

UC BERKELEY | SCHOOL OF INFORMATION

Postdoctoral Scholar

Sept 2018 – Sept 2019 | Berkeley, CA

Led an analysis project that uses country level cellphone data to extract displacement signals and measure the impact of violence in Afghanistan

- **Data Preparation:** Used PySpark to clean, analyze and aggregate call detail records (~5TB level) from a cellphone provider in Afghanistan.
- **Data Summarization:** Constructed measures to define the level and duration of user level displacement due to occurrence of violent events.
- **Data Visualization:** Used GeoPandas, seaborn to build visualizations and statistical measures describing event level local population displacement.
- **Scale-up:** Used PySpark to scale the event level measures across thousands of violent events, and then carried out regression analysis to find factors affecting displacement.
- **Leadership:** Built and led a research team of interns and undergrads to maintain and scale the data science pipeline.

UCLA | ANDERSON SCHOOL OF MANAGEMENT

PhD Student

Aug 2013 – June 2018 | Los Angeles, CA

Ideated, built, piloted, and evaluated a machine learning pipeline to detect tax evading firms in a state-level tax system.

- **Data Gathering:** Convinced a state government in India to share sensitive tax returns data for capacity building and for research purposes. At that time no researcher had succeeded in this in India.
- **Data Preparation:** Transformed the 5 year universe of tax returns (MSSQL and ~1/2 TB) and prepared it for data cleaning and feature engineering using Stata.
- **Predictive Modeling:** Used Python, H2O, Graphlab, and Amazon AWS to build and pilot a random forest model that can detect the tax evading firms.
- **Model Evaluation:** Proposed and constructed real-world evaluation measures that are more relevant than the usual F1-score, accuracy etc.
- **Leadership:**
 - Built and managed a team of associates (~3) to enable local engagement.
 - Awarded multiple scientific grants (~\$210,000) to pilot and scale ML related tax capacity work, and to explore other government related data-driven projects.
- **Project Impact:**
 - Identified a prediction problem that is still leading to significant tax-evasion in multiple low-capacity countries.
 - Made presentations to multiple high level government officials and leaders (Chief Economic Advisor of India, Chief Ministers etc.).
 - Currently evaluating the efficacy of the ML model by building collaborations across multiple tax authorities in India.
 - Presented related economic & ML research in multiple prestigious global conferences. [[slides](#)]

Teaching Experience: Assisted in teaching the core economics course to 600 MBA students over 4 years.

PUBLICATIONS & WORKING PAPERS

- **Shekhar Mittal**, Ofir Reich and Aprajit Mahajan. *Who's Bogus? Learning to Identify Fraudulent Firms from Unbalanced and One-side Labelled Tax Returns Data*. In Proceedings of the 1st Annual ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018. [[pdf](#)]
- **Shekhar Mittal**, and Aprajit Mahajan. *VAT in Emerging Economies: Does Third Party Verification Matter?* WP. [[pdf](#)]
- Jan Luksic, and **Shekhar Mittal**. *Red Tape? The Revenue Impact of the VAT Filing Thresholds*. WP. [[pdf](#)]