NBER PhD Student Workshop in the Economics of Digitization
Applicant Information

- **Name:** Madhav Kumar
- **Affiliation:** MIT, Sloan School of Management
- **Year and Department:** 3rd year of Ph.D. in Quantitative Marketing
- **Focus:** Machine learning, online marketplaces, computational social science, data privacy, data sharing regulations
- **Advisors:** Prof. Sinan Aral, Prof. Catherine Tucker, Prof. Dean Eckles
- **Address:** E62-586, MIT Sloan School of Management, 30 Memorial Drive, Cambridge, MA - 02139, USA
- **Phone:** 617-971-6954
- **E-mail:** madhavk@mit.edu

NBER PhD Student Workshop in the Economics of Digitization
Madhav Kumar, Ph.D., MIT Sloan
Interest areas in digitization:

1. Data privacy, sharing, and regulation guidelines for the digital economy (with Prof. Catherine Tucker):

I am interested in exploring the role and value of data privacy, regulation, and sharing in an economy that is becoming increasingly dependent on personal data for its growth. In collaboration with Prof. Tucker, I plan to focus a part of my dissertation on understanding consumer data dynamics both from a positive as well as a normative perspective.

More specifically, we want to quantify the welfare effects and trade-offs involved in sharing personal data, especially in the financial domain. For example, having and sharing the history of financial transactions can help provide easier access to credit through online lending, at the cost of personal data being potentially shared to third-party vendors or being used without permission for other financial transactions such as insurance. On the other hand, unavailability of historical data may unfairly exclude consumers from the formal financial ecosystem, often leading them towards informal, high-risk sources. To operationalize this idea, we have partnered with a large Asian financial conglomerate to understand the interplay between data availability and credit availability in the scope of algorithmic lending.

Subsequently, we seek to formalize data sharing and usage guidelines while trying to maintain the balance between firm innovation and consumer protection. This involves developing a deep understanding of the current regulatory practices and juxtaposing them with projections of data consumption and exploitation by future AI-driven business strategies.

2. Applications of machine learning in e-commerce (with Prof. Aral and Prof. Eckles):

I also work on answering key marketing questions for online businesses using machine learning methods. For instance, I use deep learning to generate dense latent representations of products purchased and searched online. These representations convert high-dimensional discrete products into a continuous vectored representations of manageable size. We then use these representation to identify implicit relationships among products such as complementarity and substitutability and further generate product bundles that consumers would prefer. We test the efficacy of these bundles using a field experiment with a large online retailer.

The study serves two purposes - 1) it offers a novel way to design bundles in a large-scale cross-category retail setting, a previously unexplored context and 2) it allows us to test traditional theories on product bundling which still await empirical validation.

Madhav Kumar

madhavk@mit.edu • +1 617 971 6954

Education

Massachusetts Institute of Technology — Cambridge
Ph.D. in Business with focus on Machine Learning, NLP and Computational Social Science — *2016 –*

Indira Gandhi Institute of Development Research — Mumbai
M.Sc. Economics — *2009 – 2011*

Hindu College, University of Delhi — New Delhi
B.Sc. (Honors) Physics — *2005 – 2008*

PhD Coursework

Economics: Consumer Theory, Game Theory, Decision Theory, Contract Theory, Industrial Organization, Structural Modeling, Statistical Methods, Introductory and Advanced Econometrics, Consumer search, Network Analysis
Computer Science: Advanced Machine Learning, Advanced NLP, Science of Deep Learning
Business: Economics of IT and Digitization, Economics of Ideas, Innovation, & Entrepreneurship, Behavioral Marketing

Research

Dynamics of data privacy, regulation and consumer credit, with C. Tucker
Work in progress
Understanding the interplay between data availability and credit availability in the case of AI-driven lending to help develop data usage guidelines.

Generating optimal product bundles using deep learning, with S. Aral and D. Eckles
Work in progress
Identifying economic relationships using deep learning among a large assortment of retail products to generate consumer preferred bundles

Rapid spatial aggregation, with M. Loecher — Peer-reviewed
Communications in Computer and Information Science, Volume 499, Springer — *2015*
Exploited kd-trees as an efficient nearest neighbors search algorithm to map geo-tagged data points to irregular spatial polygons

Predicting usefulness of online reviews, with S. Upadhyay — Peer-reviewed
Proceedings of the 11^{th} Australasian Data Mining Conference, CRPIT — *2013*
Generated lexical feature sets from online reviews using NLP and combined them with business and reviewer characteristics to predict the number of useful votes for online reviews

Experience

Centre for Advanced Financial Research and Learning (CAFRAL), Reserve Bank of India — Mumbai
Research Associate — *Sep 2014 – Jun 2016*

- Assessed policy dynamics during unconventional monetary regimes using advanced computational methods
- Predicted multi-period corporate default using sequential Monte Carlo
- Forecasted investment horizons of foreign institutional investors to estimate tail dependence among flows

Fractal Analytics — New York/Mumbai
Data Scientist — *Jun 2011 – Mar 2014*

- Parsed unstructured adjuster notes using NLP and estimated severity of claims using stochastic gradient boosted trees for a Fortune 100 commercial insurance provider in the US
- Developed an integrated and automated framework to forecast claim amounts across business lines and geographies for the EMEA region of a large insurance company
- Predicted medication adherence for at-risk patients with chronic diseases for a Fortune 100 health insurer
- Designed predictive policing strategies using statistical models for a state police department in the US
- Optimized policy selection for audit for one of the largest workers' compensation organizations in the US

Software

RapidPolygonLookup, with M. Loecher — R package
Efficient nearest neighbors search for fast allocation of geo-tagged points to spatial polygons — *2014*