Runshan Fu

runshan@cmu.edu
(+1) 412-652-8102
https://runshanfu.com

EDUCATION

Carnegie Mellon University, Pittsburgh, USA **2022**(expected)
Ph.D., Information Systems and Marketing
Committee: Kannan Srinivasan (Co-Chair), Param Vir Singh (Co-Chair), Yan Huang, Nitin Mehta

Shanghai University of Finance and Economics, Shanghai, China **2015**
Bachelor of Management, Information Management and Information Systems

University College London, London, UK **2014**
Visiting Student, Management Science and Innovation

RESEARCH INTEREST

Topics: Quantitative Marketing, Algorithmic Bias, Economics of AI, Fairness of ML, Crowd Lending
Methodologies: Structural Modeling, Analytical Modeling, Machine Learning

JOB MARKET PAPER

- **How Does Zestimate Affect Housing Market Outcomes Across Socio-economic Segments?**

PUBLICATION

- **"Un"Fair Machine Learning Algorithms** [SSRN]
  Runshan Fu, Manmohan Aseri, Param Vir Singh, Kannan Srinivasan
  Management Science, forthcoming
- **Crowds, Lending, Machine, and Bias** [SSRN]
  Runshan Fu, Yan Huang, Param Vir Singh
  Information Systems Research, 2021
- **AI and Algorithmic Bias: Source, Detection, Mitigation and Implications** [SSRN]
  Runshan Fu, Yan Huang, Param Vir Singh
  INFORMS Tutorials in Operations Research, 2020

WORKING PAPERS

- **Model Mis-specification and Algorithmic Bias** [arXiv]
  with Yangfan Liang and Peter Zhang

CONFERENCE PRESENTATIONS

"Un"Fair Machine Learning Algorithms
- INFORMS Marketing Science Conference 2019
- INFORMS Annual Meeting 2019
- Thirteenth Annual FTC Microeconomics Conference 2020

Crowd Bias and Machine Learning: Evidence from Crowd Lending
- Workshop on Information Systems and Economics 2018
- INFORMS Annual Meeting 2018
- INFORMS Marketing Science Conference 2019

When Algorithms Promote Inequality
- INFORMS Marketing Science Conference 2020
- CMU Symposium on AI and Social Good 2020

Teaching

Teaching Assistant

- Decision Analytics for Business and Policy (by Peter Zhang) — Spring 2020
- Digital Transformation (by Michael Smith) — Fall 2019, 2020
- Machine Learning for Problem Solving (by Leman Akoglu) — Spring 2017, 2018
- Business Intelligence & Data Mining (by Beibei Li) — Spring & Fall 2018
- Unstructured Data Analytics for Policy (by George Chen) — Spring 2018
- Unstructured Data Analytics (by George Chen) — Fall 2017
- Economic Analysis (by Karen Clay) — Fall 2017
- Economic Analysis (by Alessandro Acquisti) — Fall 2017
- Statistical Theory for Social and Policy Sciences (by Amelia Haviland) — Fall 2016

Service

Ad-hoc reviewer for: Mangement Science, Information Systems Research, Production and Operations Management, Conference on Information Systems and Technology (CIST), International Conference in Information Systems (ICIS).

Skills

Python, Ruby, Matlab, R, SQL, MongoDB

References

Kannan Srinivasan
H.J. Heinz II Professor of Management, Marketing and Business Technologies
Tepper School of Business
Carnegie Mellon University
kannans@cmu.edu

Param Vir Singh
Carnegie Bosch Professor of Business Technologies and Marketing
Tepper School of Business
Carnegie Mellon University
psidhu@cmu.edu

Yan Huang
Associate Professor of Business Technologies
Tepper School of Business
Carnegie Mellon University
yanhuang@cmu.edu

Nitin Mehta
Professor of Marketing
Rotman School of Management
University of Toronto
nitin.mehta@rotman.utoronto.ca

Abstracts of Selected Research Papers

How Does Zestimate Affect Housing Market Outcomes Across Socio-economic Segments?
(Job Market Paper)

We study the impact of Zillow's Zestimate on housing market outcomes and how the impact differs across socio-economic segments. Zestimate is produced by a machine learning algorithm using large amounts of data and aims to be an unbiased prediction of a home's market value at any time. Zestimate can potentially help market participants in the housing market as identifying the value of a home is a non-trivial task. However, inaccurate Zestimate could also lead to incorrect belief about property values and hinder the selling process. Meanwhile, Zestimate tends to be significantly more accurate for rich neighborhoods than poor neighborhoods, raising concerns that the benefits of Zestimate accrue largely to the rich, widening socio-economic inequality. Using data on Zestimate and housing sales in the United States, we show that Zestimate benefits the housing market as on average it increases both buyer welfare and seller welfare. Moreover, Zestimate actually reduces socio-economic inequality, as our results reveal that both rich and poor neighborhoods benefit from Zestimate but the poor neighborhoods benefit more.

We build a structural model of a housing market where sellers and buyers face uncertainty about property values. Zestimate provides an unbiased signal of the property value. Our model captures two potentially countervailing effects of Zestimate. First, it reduces uncertainty in beliefs about property values. Second, it shifts the mean belief about property values towards the Zestimate. Since Zestimate predicts the property value with some error, the reported Zestimate could under or over estimate the property value. Hence, the mean belief about the property value may be shifted away from the true property value.

We estimate our model on a unique data set consisting of 3,724 properties listed in Pittsburgh between February and October 2019. The estimation results reveal that people in poor neighborhoods are more uncertain about property values initially before learning from Zestimate compared with those in the mid-range and rich neighborhoods. In a counterfactual analysis, we show that, on average, the introduction of Zestimate increases seller profit by 7.53%, buyer surplus by 4.42%, and total surplus by 6.16%. We also find that although Zestimate under-estimates the selling price of more than 40% of the properties in our sample, only 19.04% of the properties in the sample have lower seller profit with Zestimate than without. This suggests that having an undervalued Zestimate may still be better for the seller than not having a Zestimate due to the benefit of uncertainty reduction. In addition, Zestimate leads to the greatest total welfare increase in poor neighborhoods (7.09%) despite the fact that Zestimates are least accurate in these areas. One important reason that Zestimates are less accurate in poor neighborhoods is a lack of accurate data on home facts, which homeowners can voluntarily provide to Zillow to improve accuracy. In another counterfactual analysis, we increase Zestimate accuracy in poor neighborhoods to the same level as in other neighborhoods, and find that the positive impact of Zestimate on total surplus would further increase by 31.17%.

"Un"Fair Machine Learning Algorithms
(Forthcoming at *Management Science*)

Ensuring fairness in algorithmic decision-making is a crucial policy issue. Current legislation ensures fairness by barring algorithm designers from using demographic information in their decision-making. As a result, to be legally compliant, the algorithms need to ensure equal treatment. However, in many cases, ensuring equal treatment leads to disparate impact particularly when there are differences among groups based on demographic classes. In response, several "fair" machine learning (ML) algorithms that require impact parity (e.g., equal opportunity) at the cost of equal treatment have recently been proposed to adjust for the societal inequalities. Advocates of fair ML propose changing the law to allow the use of protected class-specific decision rules. We show that the proposed fair ML algorithms that require impact parity, while conceptually appealing, can make everyone worse off, including the very class they aim to protect. Compared to the current law, which requires treatment parity, the fair ML algorithms, which require impact parity, limit the benefits of a more accurate algorithm for a firm. As a result, profit maximizing firms could under-invest in *learning*, i.e., improving the accuracy of their machine learning algorithms. We show that the investment in learning decreases when misclassification is costly, which is exactly the case when greater accuracy is otherwise desired. Our paper highlights the importance of considering strategic behavior of stake holders when developing and evaluating fair ML algorithms. Overall, our results indicate that fair ML algorithms that require impact parity, if turned into law, may not be able to deliver some of the anticipated benefits.

Crowds, Lending, Machine, and Bias
(Published on *Information Systems Research*)

Big data and machine learning (ML) algorithms are key drivers of many fintech innovations. Although it may be obvious that replacing humans with machines would increase efficiency, it is not clear whether and how machines can improve human decisions. We answer this question in the context of crowd lending, in which decisions are traditionally made by a crowd of investors. Using data from Prosper.com, we show that a reasonably sophisticated ML algorithm predicts listing default probability more accurately than crowd investors. The improvement of the machine over the crowd predictions is more pronounced for highly risky listings. We then use the machine to make investment decisions and find that the machine improves upon investors' decisions and leads to greater welfare for both investors and borrowers simultaneously. When machine prediction is used to select loans, it leads to a higher rate of return for investors and more funding opportunities for borrowers with few alternative funding options. We also find suggestive evidence that the machine is biased in gender and race even when it does not use gender and race information as input. We propose a general and effective "debiasing" method that can be applied to any prediction-focused ML applications and demonstrate its use in our context. We show that the debiased ML algorithm, which suffers from lower prediction accuracy, still improves the crowd's investment decisions in our context. These results indicate that ML can help crowd-lending platforms better fulfill the promise of providing access to financial resources to otherwise underserved individuals and ensure fairness in the allocation of these resources.