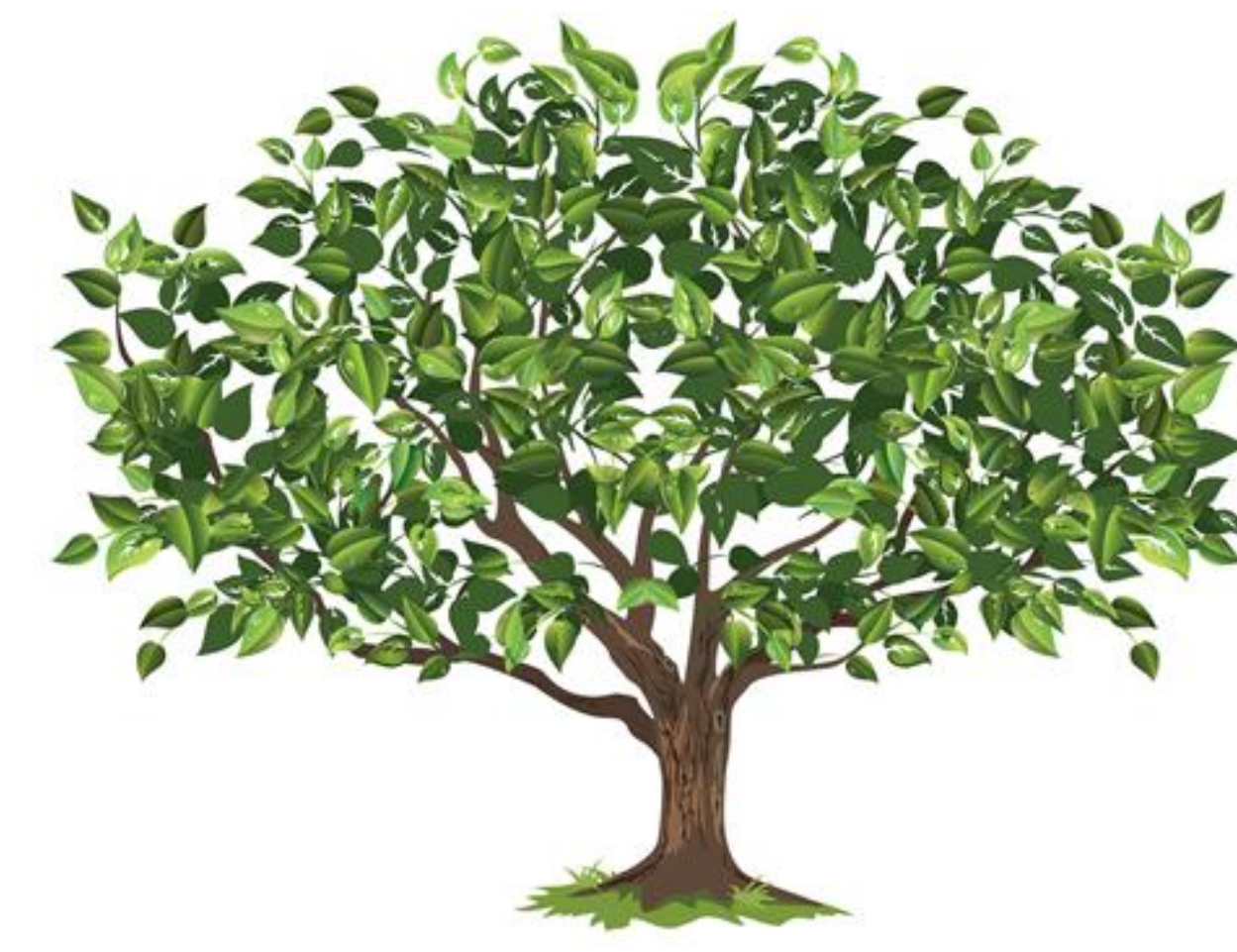


Tree Matching Solution for Self-Selection in Impact Surveys



Reema Gupta, Deepa Mani, Galit Shmueli

Sunil Mithas



Goal

Using survey data on thousands of people:

1. Assess the impact of a new technology
2. Rank employee groups by job satisfaction
3. Evaluate effect of experience on satisfaction

Challenges

- Self-selection (of various types)
- No control over study implementation (are samples really balanced?)

Classic Solution: Propensity Scores

Measures of interest (Y)

Treatment (T) – with self selection

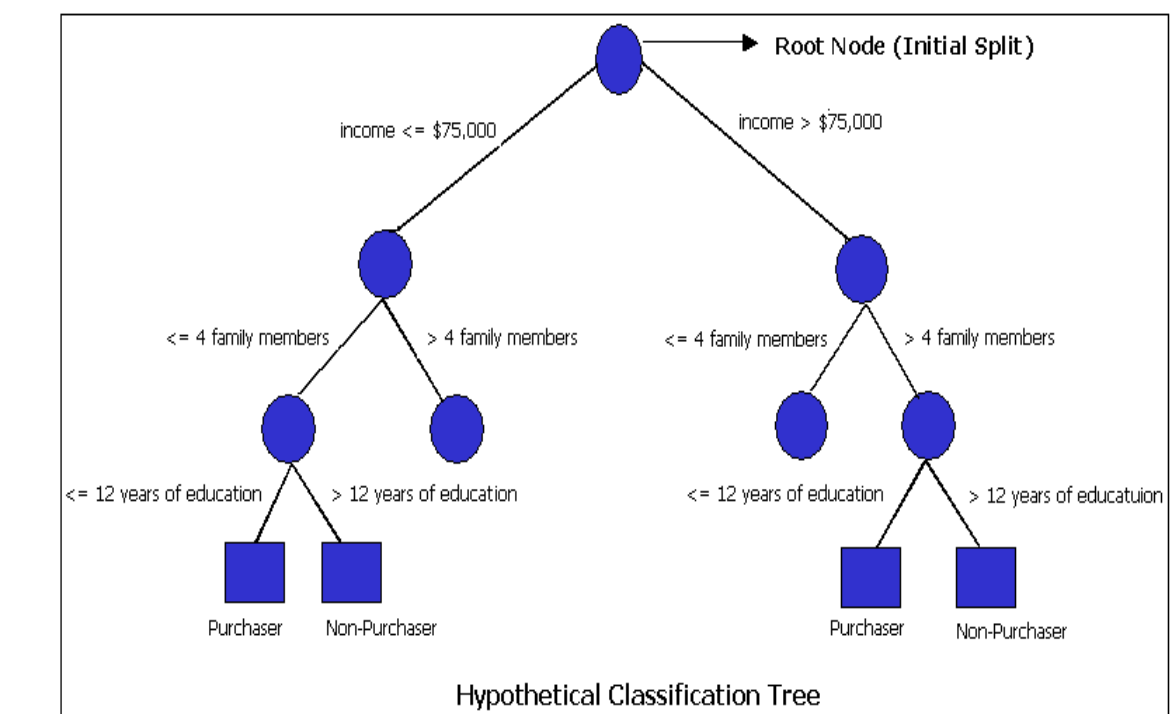
Pre-Treatment covariates (X)

$$\text{Step 1: match samples} \rightarrow \text{Get } PS = \text{logit}(T) = f(X) \rightarrow P(T=1|X)$$

Step 2: compare matched samples

$$Y = \beta_0 + \beta_1 T + \beta_2 X + \beta_3 PS + \epsilon$$

Proposed Solution: Classification & Regression Trees



People in same leaf node are similar in terms of their profile (X). Have same propensity score.

Within leaf node, compare measures (Y) between people with different T

Case 1: Impact Assessment of eGov Initiatives in India

Survey commissioned by the Govt of India in 2006

- > 9500 individuals who used passport services
 - Representative sample of 13 Passport Offices
- “Quasi-experimental, non-equivalent groups design”
- Equal number of offline and online users matched by geography and demographics

Current practice:

Estimate impact by comparing various performance outcomes across online/offline groups



Case 2: Job Satisfaction in IT Firms in India

Survey of employee job satisfaction in Indian IT firms

- Conducted in 2007 by market research firm
- >2500 Indian IT professionals, randomly selected from 31 firms
- Employee details: IT experience in current and prior firms, educational qualifications, age, gender
- Firm details such as firm size

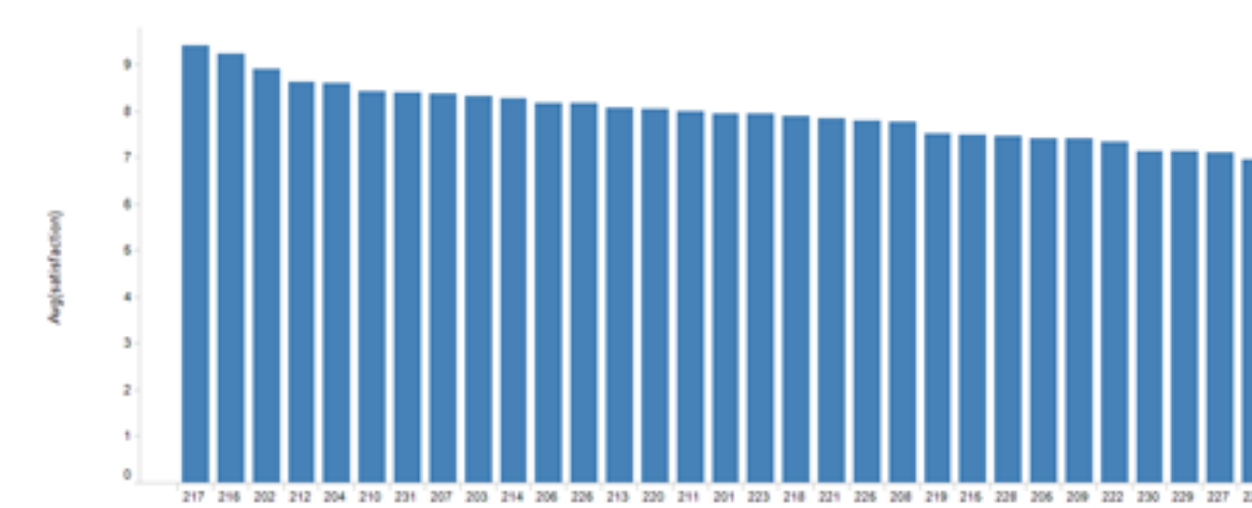
An Even Better Place To Work



Rank firms by satisfaction
Effect of experience on satisfaction

Rank firms by Satisfaction

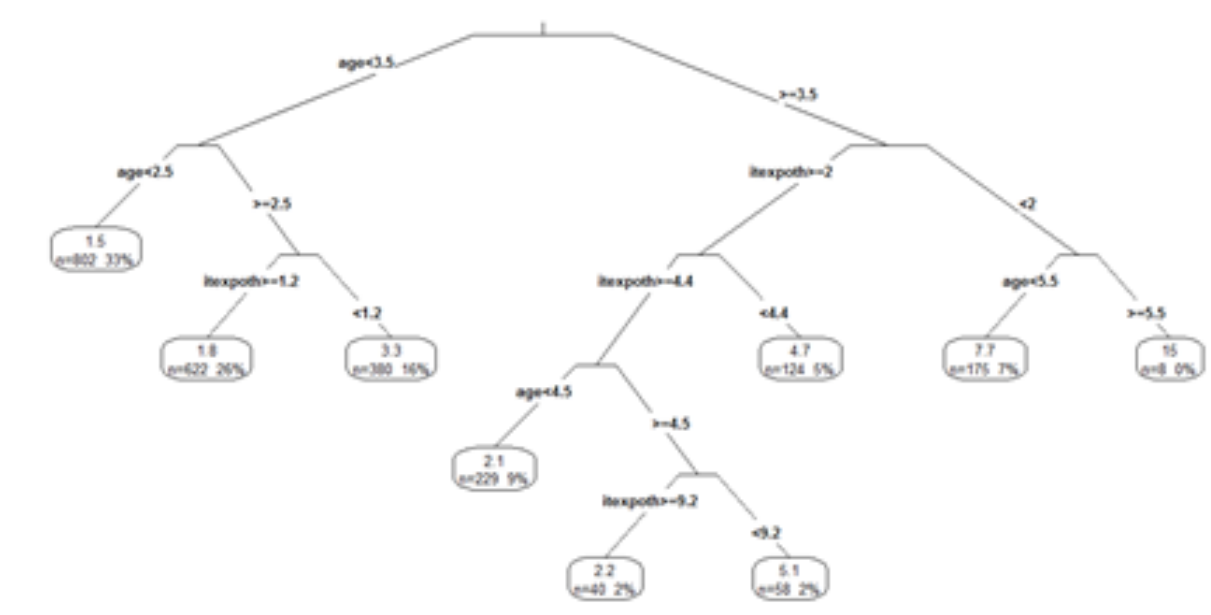
Average Satisfaction across 31 firms: Raw Data



Effect of Experience On Satisfaction

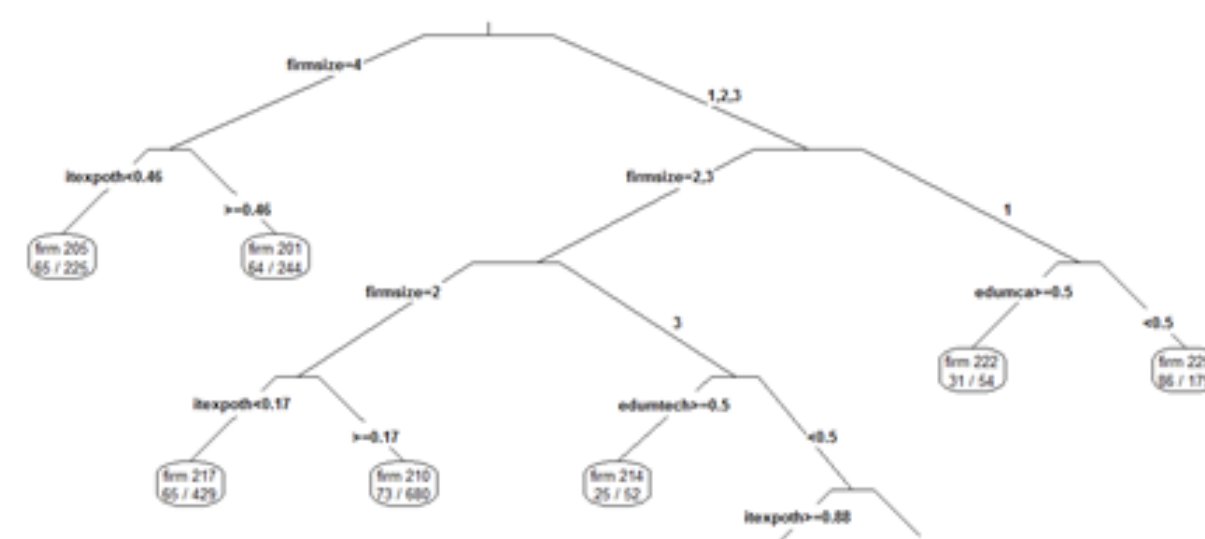
Self-selection variable: IT Experience in Current Firm

Employees with different demographics and previous experience might choose to stay longer/shorter at the current firm

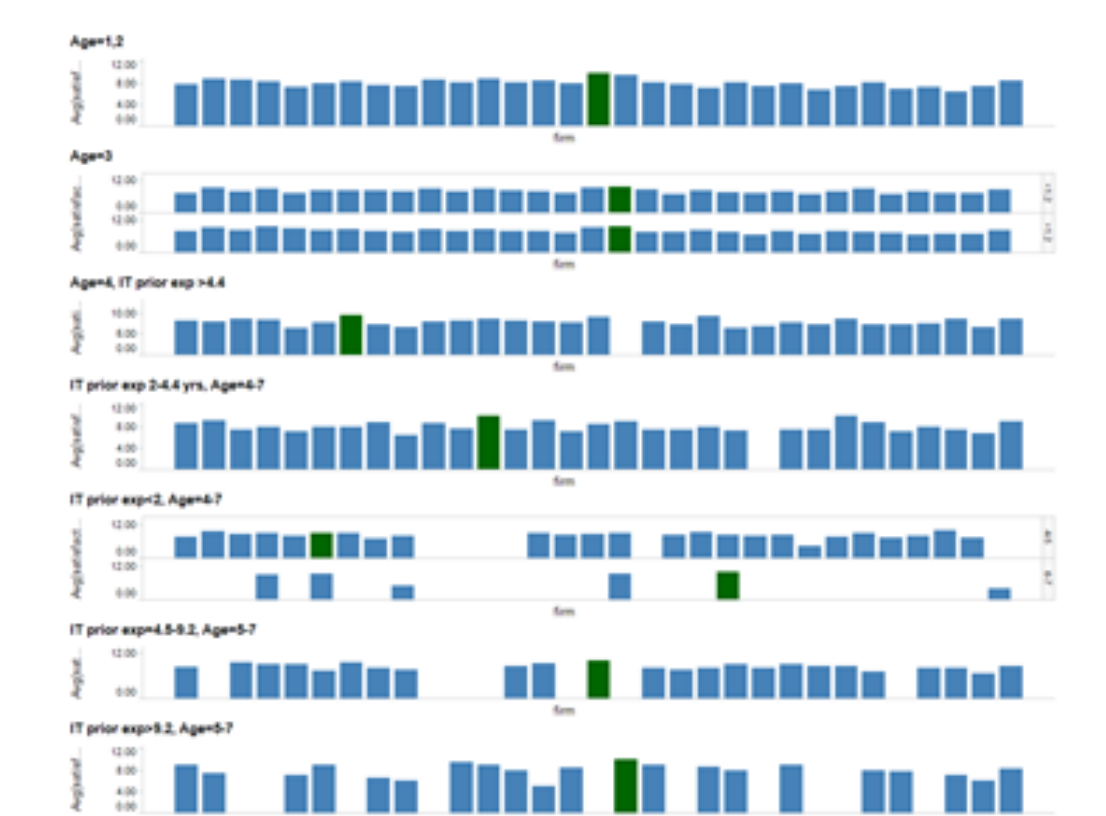


Self-selection variable: Firm (31 firms)

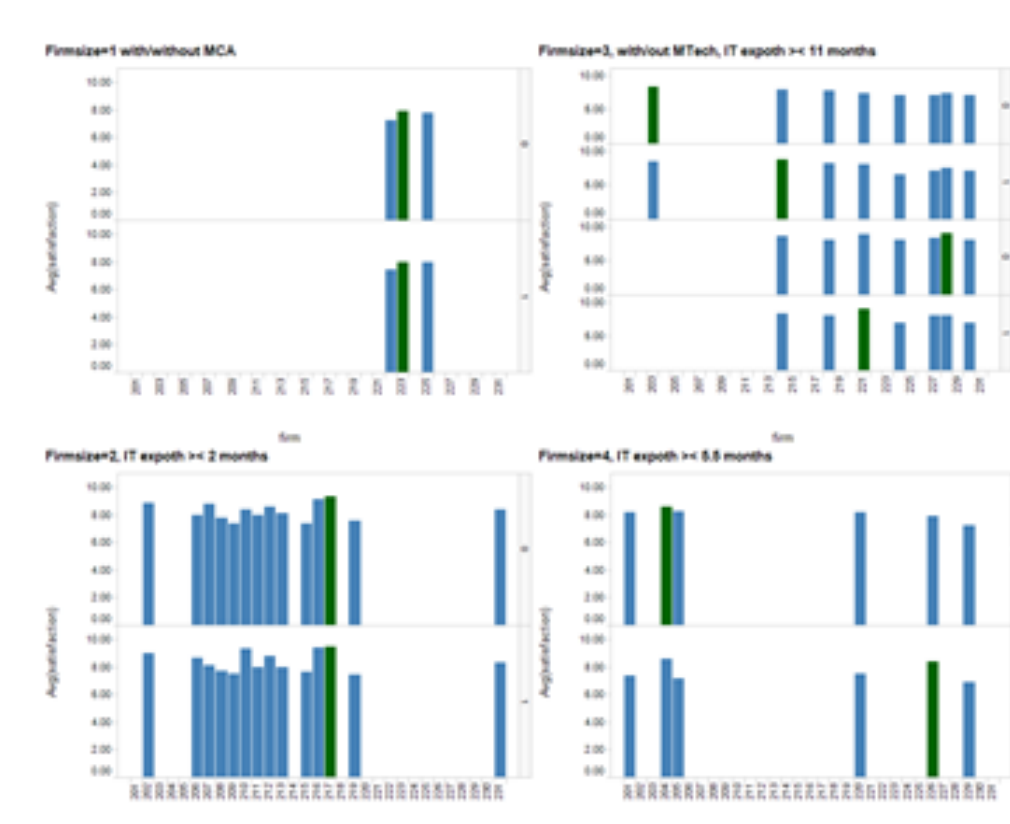
Employees with different demographics and previous experience might choose to join a certain firm



Average satisfaction across firms, after matching on pre-“current IT experience” choice variables



Average satisfaction across firms, after matching on pre-firm choice variables



Accounting for self-selection

Logistic

VARIABLE	ESTIMATE
Experience	0.125***
Awareness	1.462***
Clarity of Rules and Procedures	0.179***
Availability of Information	0.031
Clarity and Simplicity of Processes	0.061**
Convenience of Working Hours	0.043
Convenience of Location	0.076**
Design and Layout of Application Forms	0.123***

Classification Tree



People in each node are similar in terms of their profile (X). Have same propensity score.

Compare measures (Y) of online/offline within each leaf

Performance Metric	Unmatched sample			Propensity Matched sample		
	Offline	Online	Difference	Offline	Online	Difference
Number of steps	2.5	2.2	-0.30	2.1	2.2	0.11
Waiting time (min)	96.0	88.2	-7.80	105.5	88.2	-17.30
Total elapsed time in awaiting service (min)	43.0	43.0	0.00	44.4	41.1	-3.30
Proportion paying fines for crops (%)	42.4	48.3	5.70	48.4	48.3	-0.10
Proportion paying fines for payment office employees (%)	4.3	4.4	0.10	5.4	4.4	-1.00
Proportion using intermediary (%)	34.8	33.6	-1.20	34.3	33.6	-0.70
Satisfaction with overall QoS	3.5	3.6	0.10	3.5	3.4	-0.10
Overall government	3.9	3.5	-0.40	3.4	3.5	0.10
Overall trust in the government	3.2	3.0	-0.20	3.2	3.0	-0.20
Willingness to return system (%)	39.4	39.4	0.00	38.4	39.4	1.00

Advantages of Trees over Propensity Scores

1. Simpler to understand
2. Easier to present graphically
3. Clear information about which variables require matching
4. Easily generalized to multiple treatments
5. Easily generalized to continuous treatment

