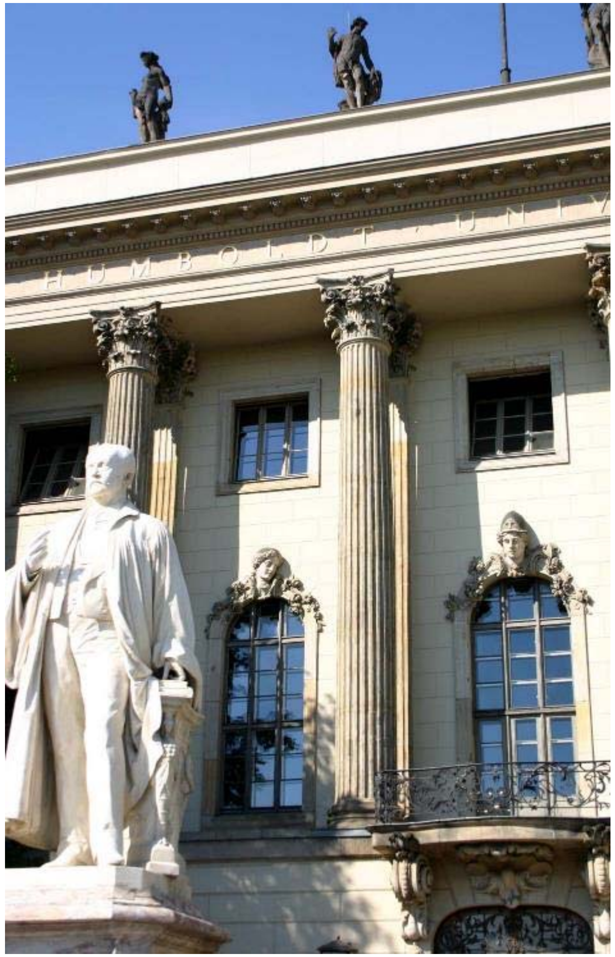
7th Annual SKBI Conference



E-Commerce Models and Causality: Novel Perspectives on Uplift Modeling

Stefan Lessmann

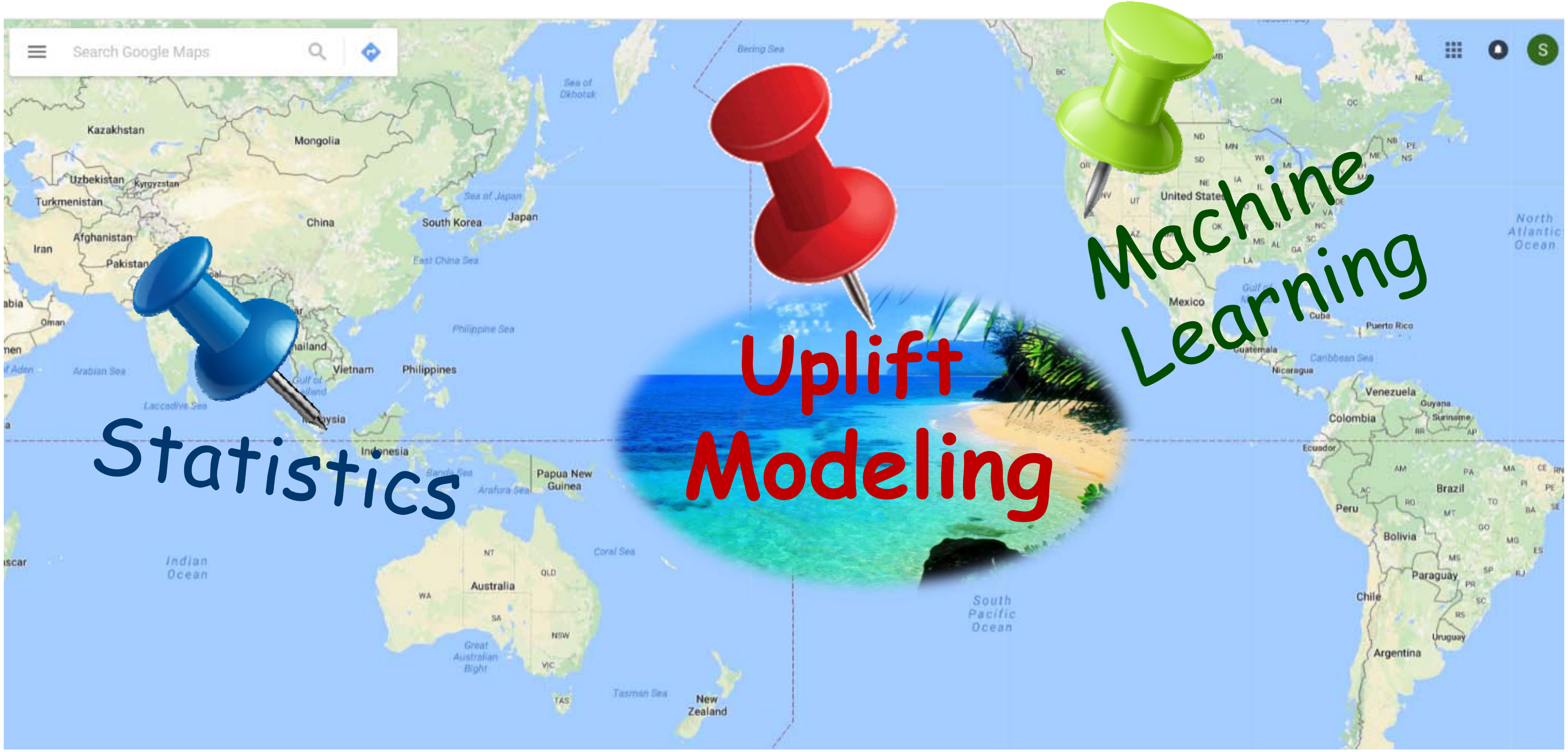
HU Berlin

Agenda

E-Commerce Models and Causality: Novel Perspectives on Uplift Modelling

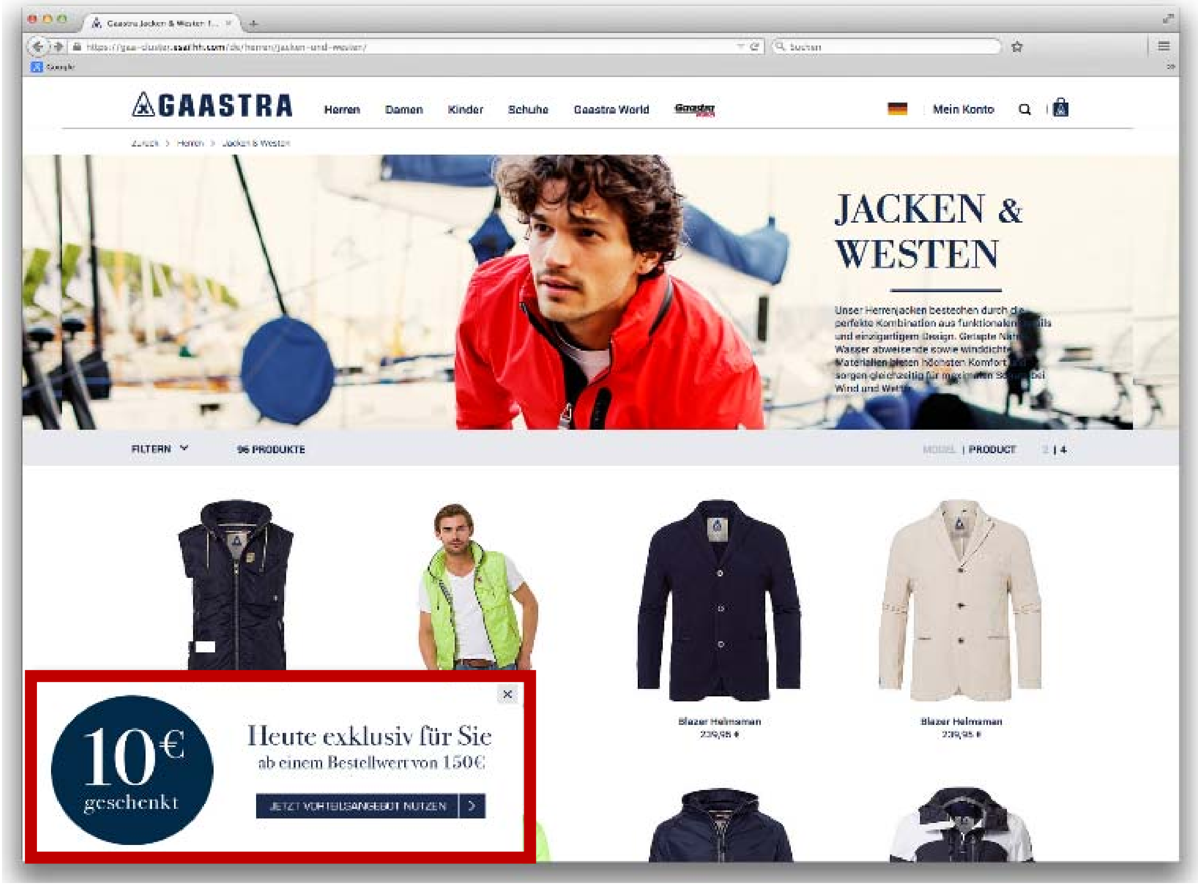
- **Uplift Modeling Fundamentals**
  - An informal and semi-formal introduction
  - Application context and prior work
- **Uplift Transformation**
  - Conversion modeling
  - Revenue uplift
- **E-Commerce Case Study**
  - Data and experiment design
  - Empirical results
- **Conclusions**

Uplift Modeling = Stats + ML + App Context



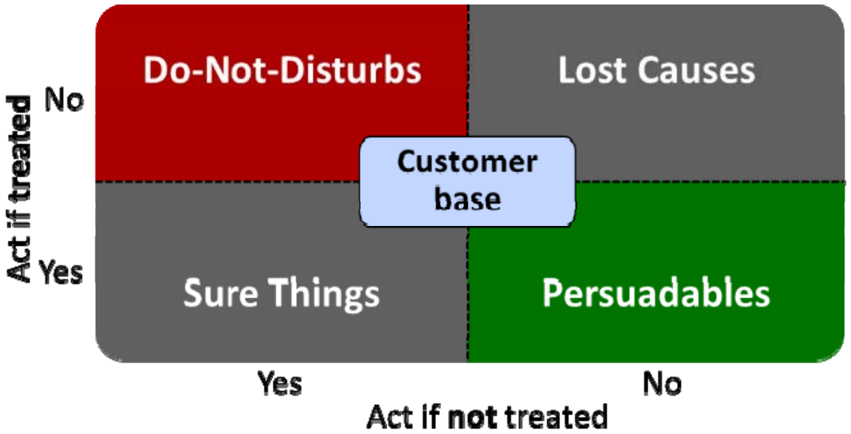
Application Context: E-Couponing

- **E-Shop Scenario**
  - Surfer visits shops
  - Clickstream data
    - Page visited
    - Time on page
    - Device
- **Decision(s)**
  - **Show incentive?**
  - Which incentive?
    - Percent or absolute discount
    - Discount size
    - Constraints



Aid decisions using via predictive model

Semi-Formal Introduction to Uplift Modeling



Observable outcome (e.g. from experiment)

Treated		Responded: Yes	Responded: No
	No	Control Responders	Control non-Responders
	Yes	Treated Responders	Treated non-Responders

Two-Model Approach

- M1 estimates response in treatment group
- M2 estimates response in control group
- Uplift score:

$$s(\boldsymbol{X}) = P(\text{act}|\boldsymbol{X}, T) - P(\text{act}|\boldsymbol{X}, C)$$

Many disadvantages

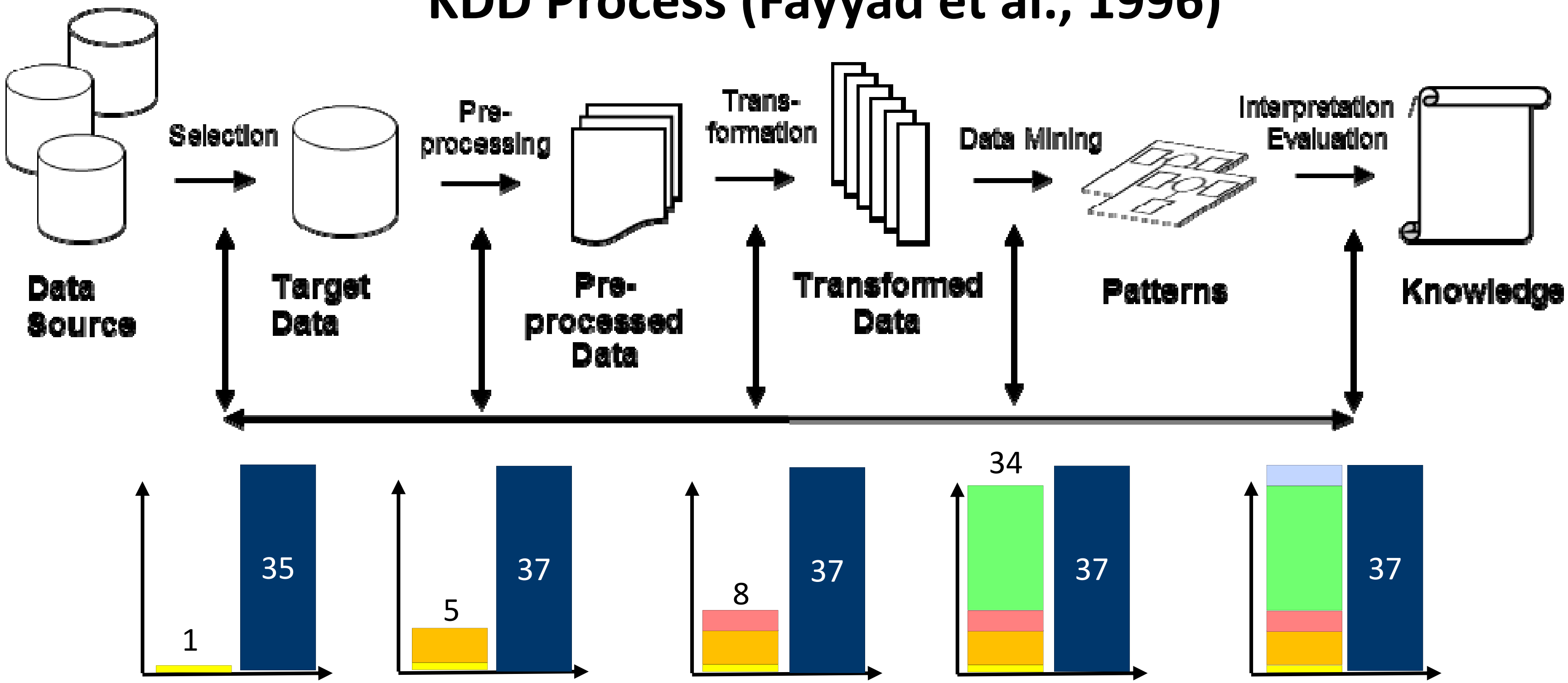
Desired info: who responds **because of** treatment

Act if targeted		Act if NOT targeted: Yes	Act if NOT targeted: No
	No	Do-Not-Disturbs	Lost Causes
	Yes	Sure Things	Ideal targets: Persuadables

[Radcliffe, 2007; Siegel 2011]

Prior Work on Uplift Modeling

KDD Process (Fayyad et al., 1996)

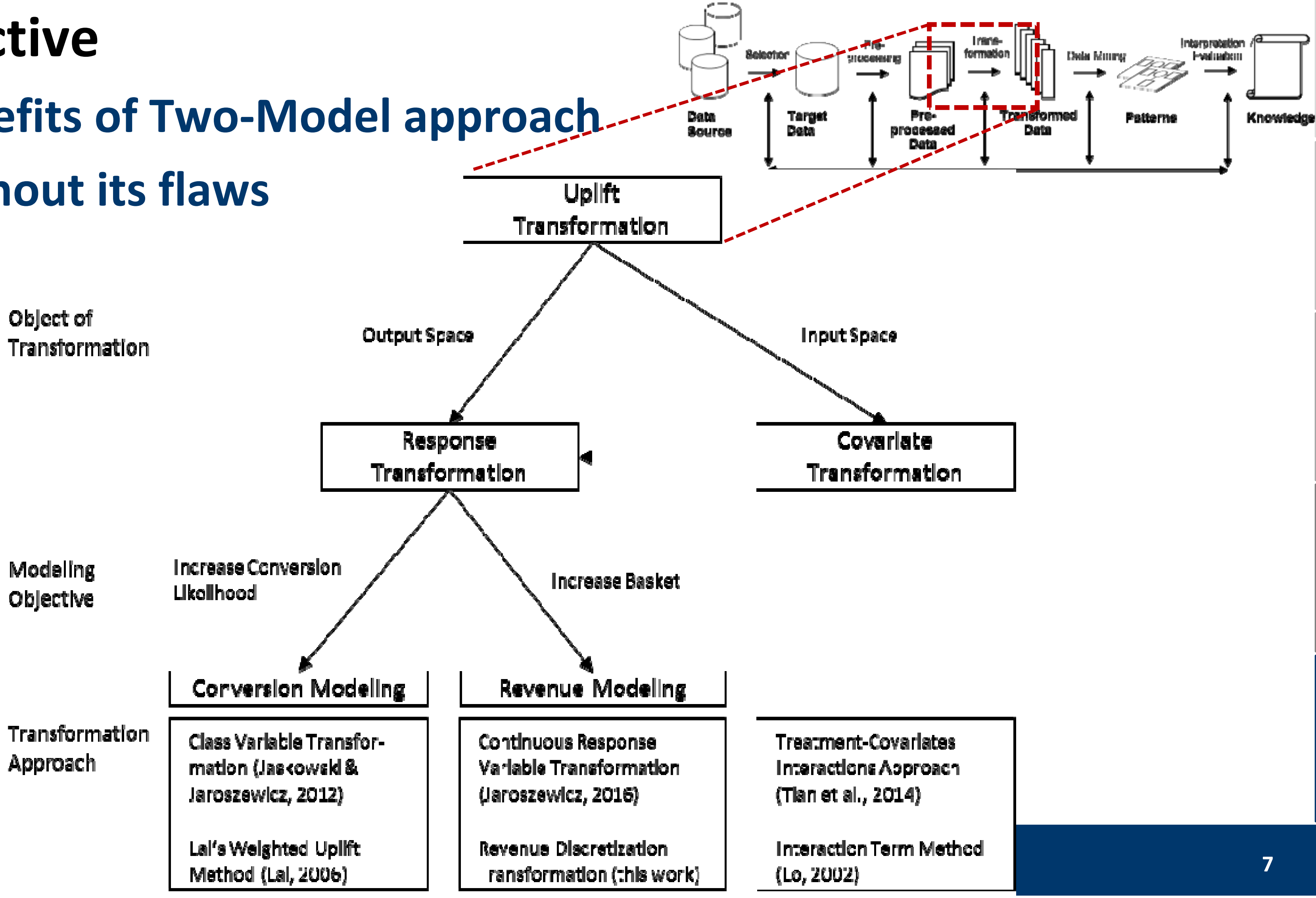


- Total of 37 studies
- About 65% published in 2012 or later
- Majority of studies develop novel algorithms

Uplift Transformation

Objective

- Benefits of Two-Model approach
- Without its flaws



Class Variable Transformation (cont.)

Treated		Responded: Yes	Responded: No
	No	Control Responders	Control non-Responders
	Yes	Treated Responders	Treated non-Responders

$$s(\boldsymbol{X}_i) = P(Y_i = 1|\boldsymbol{X}, T) - P(Y_i = 1|\boldsymbol{X}, C)$$

Act if targeted		Act if NOT targeted: Yes	Act if NOT targeted: No
	No	Do-Not-Disturbs	Lost Causes
	Yes	Sure Things	Ideal targets: Persuadables

$$D = \{Y_i, \boldsymbol{X}_i, T_i\}_{i=1}^{n}$$

$$z_i = \begin{cases} 1 & if\ T_i = 1 \cap Y_i^c = 1 \\ 1 & if\ T_i = 0 \cap Y_i^c = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$s(\boldsymbol{X}_i) = P(z_i|\boldsymbol{X}_i)$$

Framework to model uplift using supervised classification.

[Jaskowski/Jaroszewicz, 2012; Lai, 2006]

Notation
- Customer index i
- Group membership $T_i \in \{0,1\}$
- Spending $Y_i \in \mathbb{R}$
- Conversion $Y_i^c \in \{0,1\}$
- Observations $\boldsymbol{X}_i \in \mathbb{R}^p$

Revenue Transformation (cont.)

Not all customers spend the same.

$$D = \{Y_i, \boldsymbol{X}_i, T_i\}_{i=1}^{n}$$

$$z_i^r = \begin{cases} +Y_i & \text{if } T_i = 1 \cap Y_i > 0 \\ -Y_i & \text{if } T_i = 0 \cap Y_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad z_i^r \in \mathbb{R}$$

$$z_i^b = \begin{cases} 0 & \text{if } z_i^r \in (-\infty, 0) \\ 1 & \text{if } z_i^r \in [0, +\infty) \end{cases}$$

$$z_i^b \in \{0,1\}$$

$$s^r(\boldsymbol{X}_i) = P(z_i^r | \boldsymbol{X}_i\ )$$

$$s^b(\boldsymbol{X}_i) = P(z_i^r | \boldsymbol{X}_i\ )$$

Framework to model revenue uplift using supervised regression or classification.

[Bodapati/Gupta, 2004; Jaroszewicz, 2016]

Notation
- Customer index i
- Group membership $T_i \in \{0,1\}$
- Spending $Y_i \in \mathbb{R}$
- Spending $Y_i \in \mathbb{R}$
- Conversion $Y_i^c \in \{0,1\}$
- Observations $\boldsymbol{X}_i \in \mathbb{R}^p$

Experimental Design

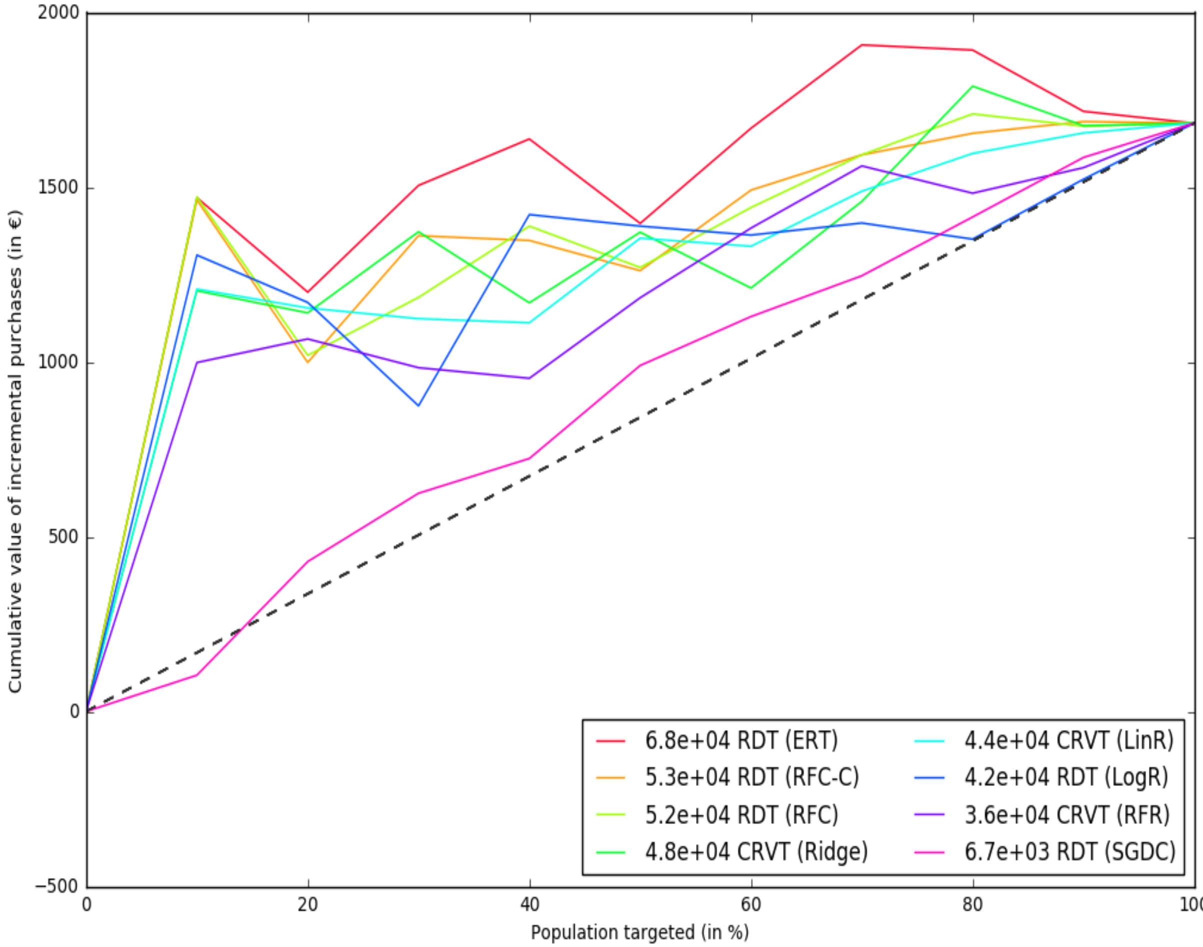
GROUP	OBSERVATIONS		CONVERSIONS		UPLIFT
Treatment	2,285,835	75 %	175,791	7.7 %	0.2 %
Control	766,155	25 %	57,285	7.5 %	
	3,051,990		**233,076**		

- **Data partitioning**
  - Training (40%)
  - Parameter tuning (30%)
  - Model comparison (30%)
- **Base learners**
  - Supervised classification & regression
  - Several meta-parameter settings
  - Logit, Lasso, SVM, kNN, RandomForest, GBM, Extremely Randomized Trees, …

VARIABLE	PAGE-TYPE
TimeToFirst	Cart
TimeSinceFirst	
TimeSinceOn	
	Sale
	Search
	Product
	Overview
ClicksPer	Product
ScrollHeightPer	Overview
TimeToBasketAdd	
TimeSinceLastConversion	
TimeSinceTabSwitch	
ViewCoutLastVisit	
TimeSinceLastVisit	
TimeSinceFirstVisit	
DurationLastVisit	
…	
62 Variables in total	

Empirical Results

- **Qini plot**
  - Target customers by uplift score
  - Incremental revenue per decile
- **CRVT**
  - Model z_i^r
  - Regression
- **RDT**
  - Model z_i^b
  - Classification
- **Selection of best base learners**



CRVT: Class response variable transformation (z_i^r)

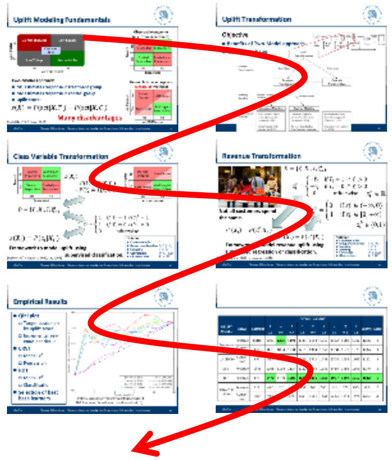
RDT: Response discretization transformation (z_i^b)

Empirical Results (cont.)

UPLIFT MODEL	GOAL	LEARNER	DECILE (WEIGHT) 1 (0.9)	2 (0.8)	3 (0.7)	4 (0.6)	5 (0.5)	6 (0.4)	7 (0.3)	8 (0.2)	9 (0.1)	WAVG	RANK
Response	Revenue	Ridge	1191	**1205**	1191	1141	1153	1213	1483	1683	1647	1236	4
Response	Conver-sion	ERT	881	1034	1231	1261	1384	1656	1708	1673	1579	1244	3
Lai (2006)	Conver-sion	RFC	893	1034	1140	1258	1311	1439	1508	1639	1701	1192	5
CRVT	Revenue	LinR	1208	1154	1124	1112	1354	1331	1488	1596	1655	1246	2
RDT	Revenue	ERT	**1470**	1199	**1505**	**1638**	**1396**	**1668**	**1907**	**1892**	**1717**	**1512**	**1**
Tian et al. (2014)	Revenue	RFR	422	570	749	864	448	521	535	626	895	597	7
Tian et al. (2014)	Conver-sion	GBC	775	843	905	1031	1281	1399	1438	1427	1579	1044	6

Conclusions

- **Summary**
  - Uplift modeling and uplift transformation
  - Conversion versus revenue uplift
  - Response variable transformations for revenue
  - Promising results in e-marketing case study
- **Closing thoughts**
  - Smart decisions require the right model
  - The right model is not necessarily complex
  - Smart decisions require the right objective
  - Decision maker model mismatch



Comments, Questions, Critic

Stefan Lessmann

School of Business and Economics
Humboldt-University of Berlin
Unter den Linden 6, D-10099 Berlin, Germany

Tel.   +49.30.2093.5742
Fax.   +49.30.2093.5741

stefan.lessmann@hu-berlin.de
http://bit.ly/hu-wi

References

- Bodapati, A., & Gupta, S. (2004). A direct approach to predicting discretized response in target marketing. *Journal of Marketing Research, 41*, 73-85.
- Jaskowski, M., & Jaroszewicz, S. (2012). Uplift Modeling for Clinical Trial Data. In N. Elhadad & M. Hauskrecht (Eds.), *ICML 2012 Workshop on Clinical Data Analysis*. Edinburgh, Scotland.
- Jaroszewicz, S. (2016). First ideas on revenue uplift modeling. Personal Communication. October 15, 2016.
- Lai, Y.-T., Wang, K., Ling, D., Shi, H., & Zhang, J. (2006). Direct Marketing When There Are Voluntary Buyers. In *Proceedings of the 6th International Conference on Data Mining (ICDM)* (pp. 922-927). Hong Kong, China: IEEE Computer Society.
- Lo, V. S. Y. (2002). The true lift model: A novel data mining approach to response modeling in database marketing. *ACM SIGKDD Explorations Newsletter, 4*, 78-86.
- Radcliffe, N. J. (2007). Generating Incremental Sales: Maximizing the Incremental Impact of Cross-Selling, Up-selling and Deep-Selling Through Uplift Modelling. In. Edinburgh: Stochastic Solutions Ltd.
- Rzepakowski, P., & Jaroszewicz, S. (2012). Uplift modeling in direct marketing. *Journal of Telecommunications and Information Technology, 2*, 43-50.
- Siegel, E. (2011). Uplift Modeling: Predictive Analytics Can't Optimize Marketing Decisions Without It. In: Prediction Impact white paper sponsored by Pitney Bowes software.
- Tian, L., Alizadeh, A. A., Gentles, A. J., & Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association, 109*, 1517-1532.