Generative Adversarial Networks for Imbalanced Learning in Customer Scoring

Johannes Haupt

Prof. Dr. Stefan Lessmann

GOR AR Analytics
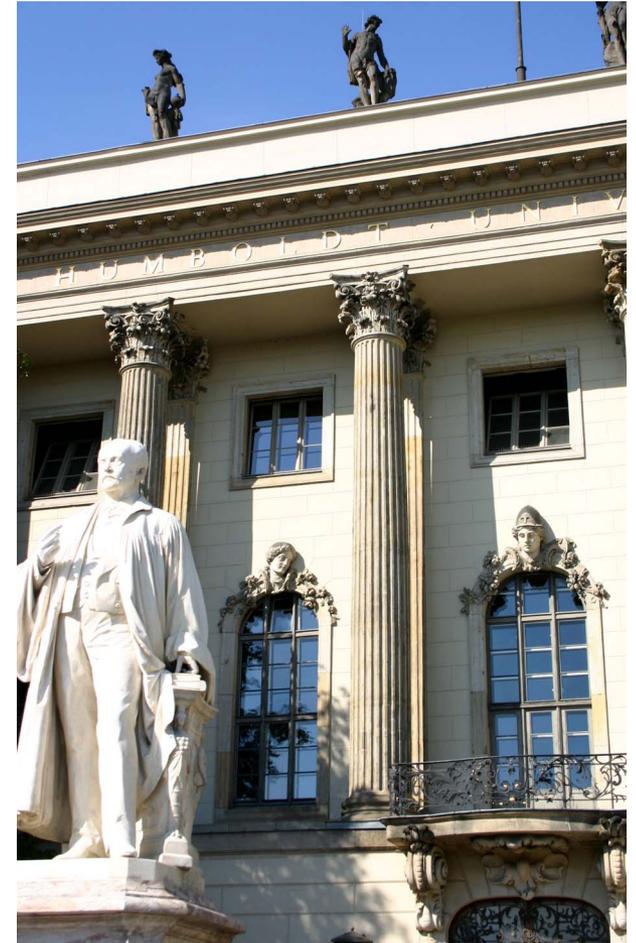
Agenda

Generative Adversarial Networks for Imbalanced Learning in Customer Scoring

- **Customer Scoring**

- **Classification of Imbalanced Data**

- **Generative Adversarial Neural Networks (GAN)**

- **GANs for Synthetic Data Generation**

Customer Scoring
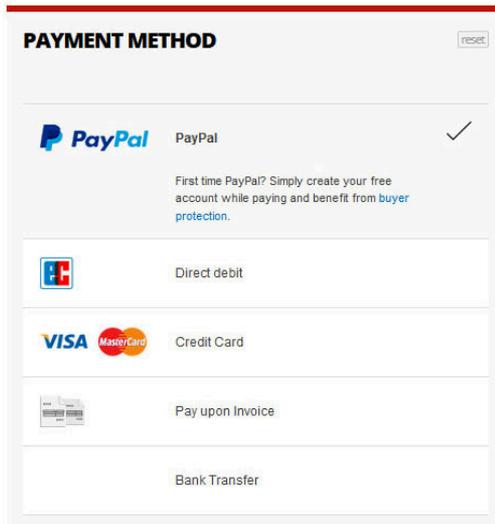
Payment Risk
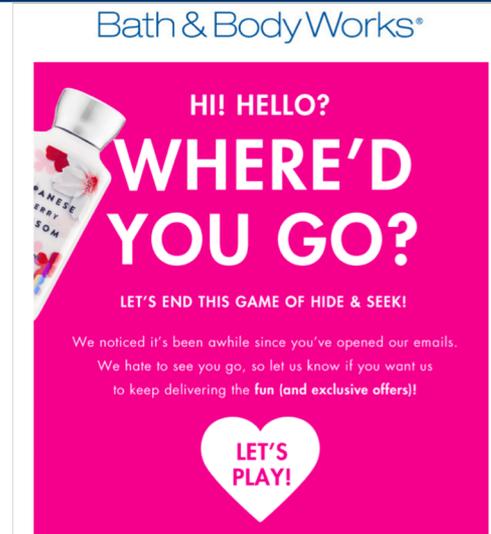
Churn Management

Product returns

Mail Targeting

Couponing
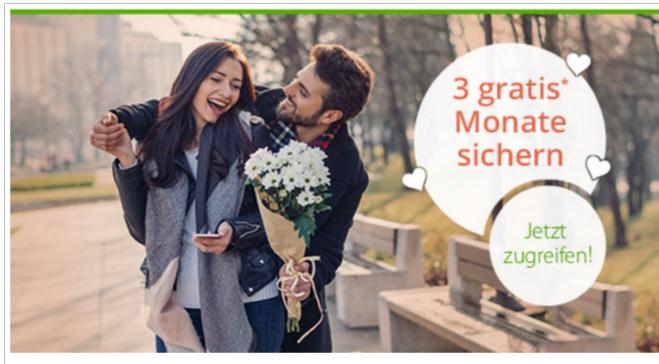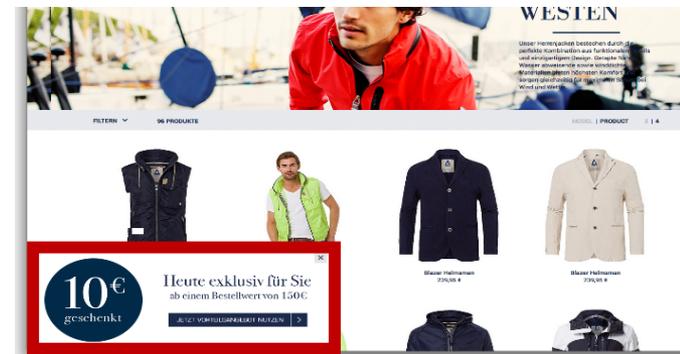
Classification of Imbalanced Data

Churn	Time Since Order	No. of orders	Last Order Value
0	5	4	25
0	1	2	15
0	4	2	30
0	10	1	51
0	7	4	20
0	6	3	49
0	3	1	15
1	2	3	65
1	1	5	53
1	3	2	57
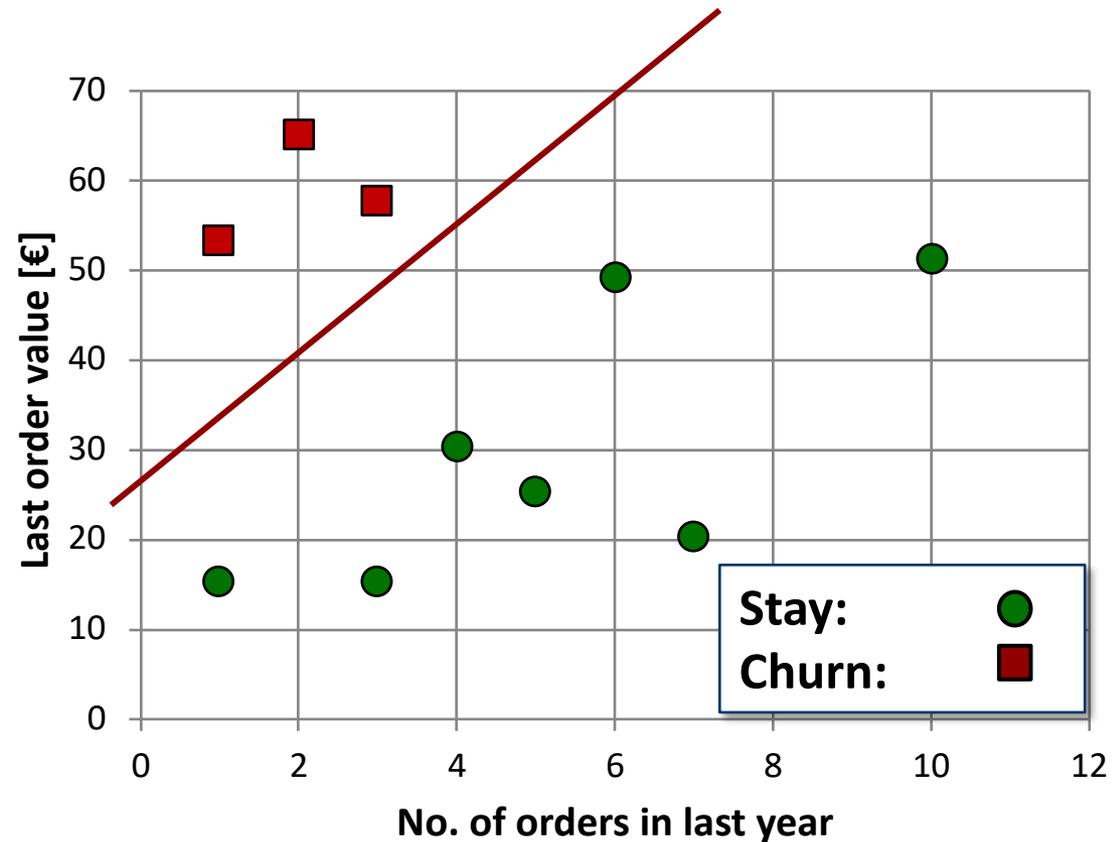
Classification of Imbalanced Data
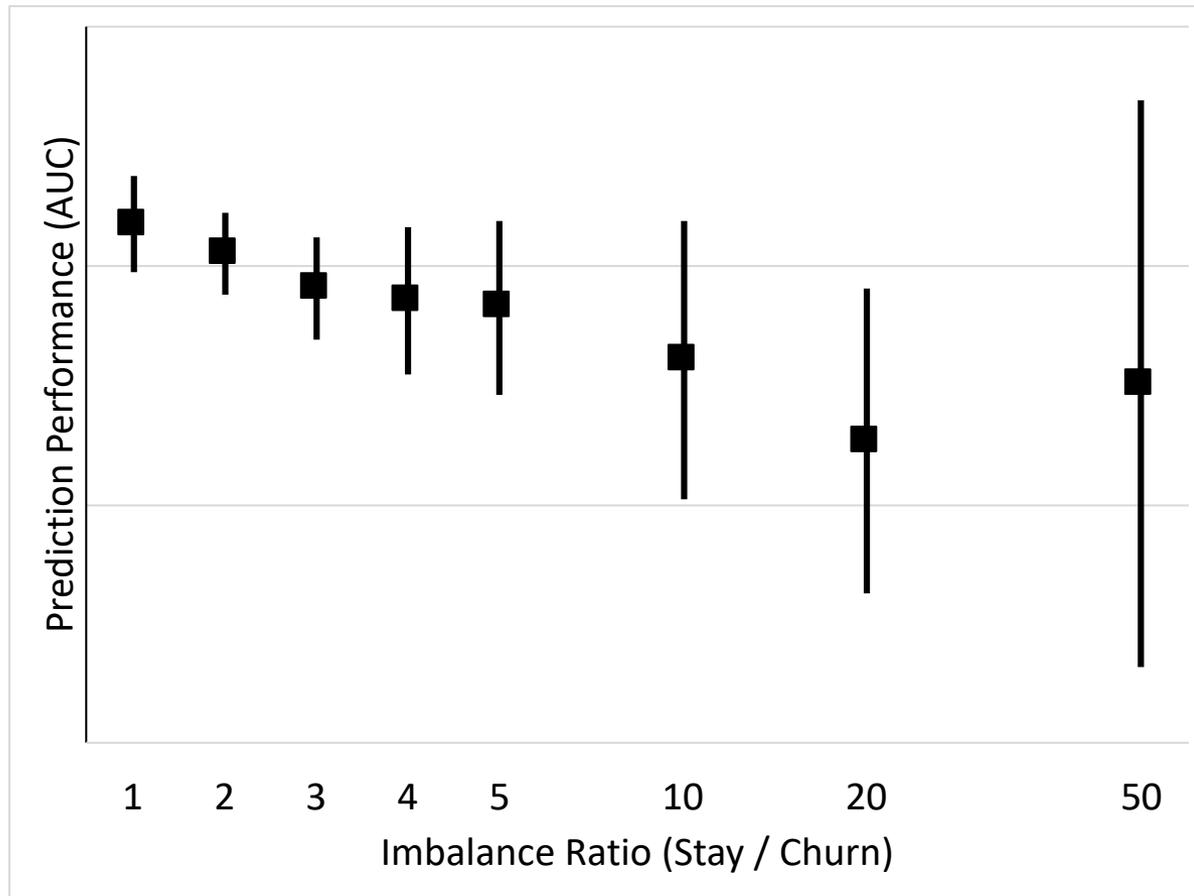
Synthetic Oversampling

Generative Adversarial Networks
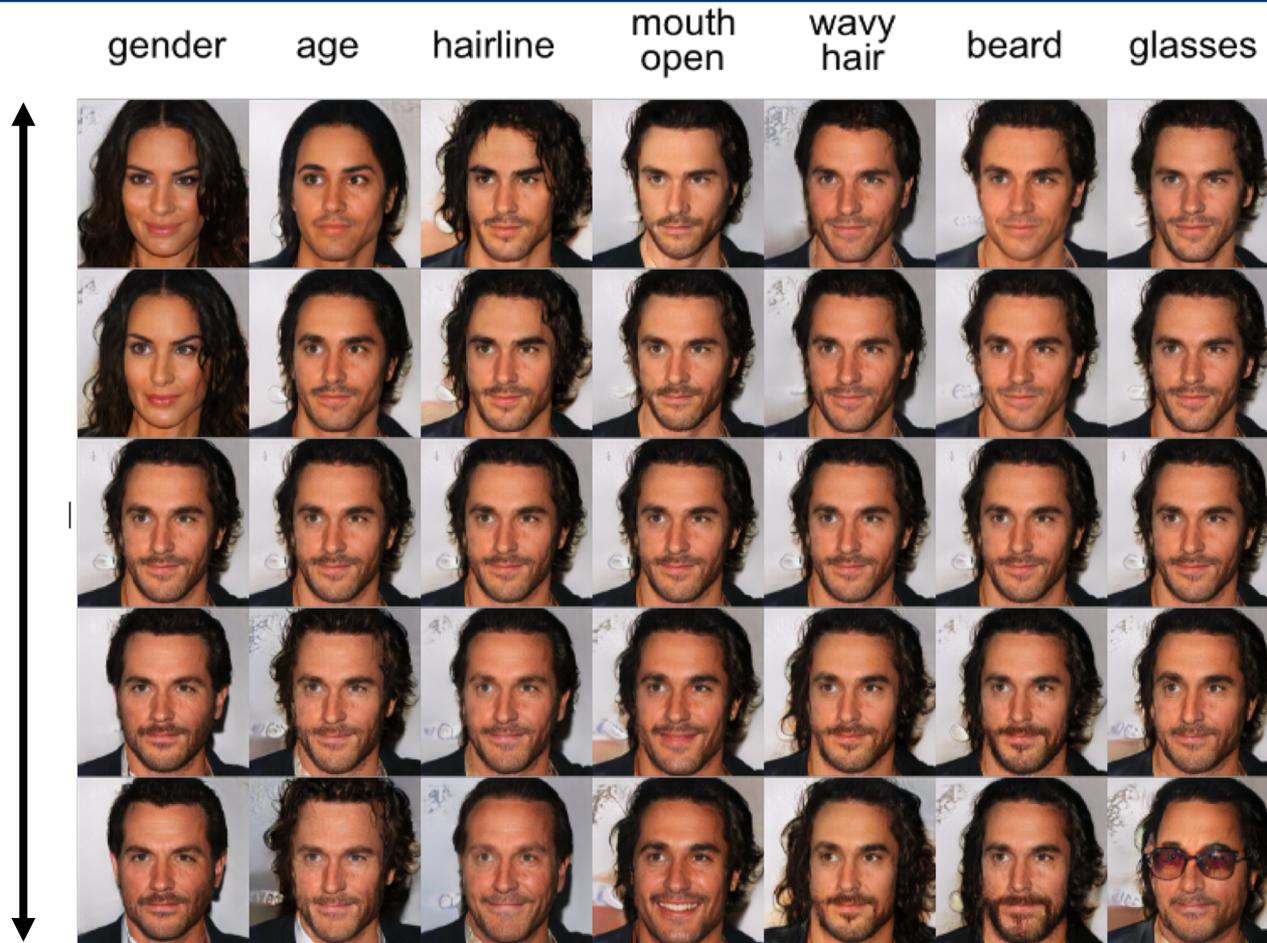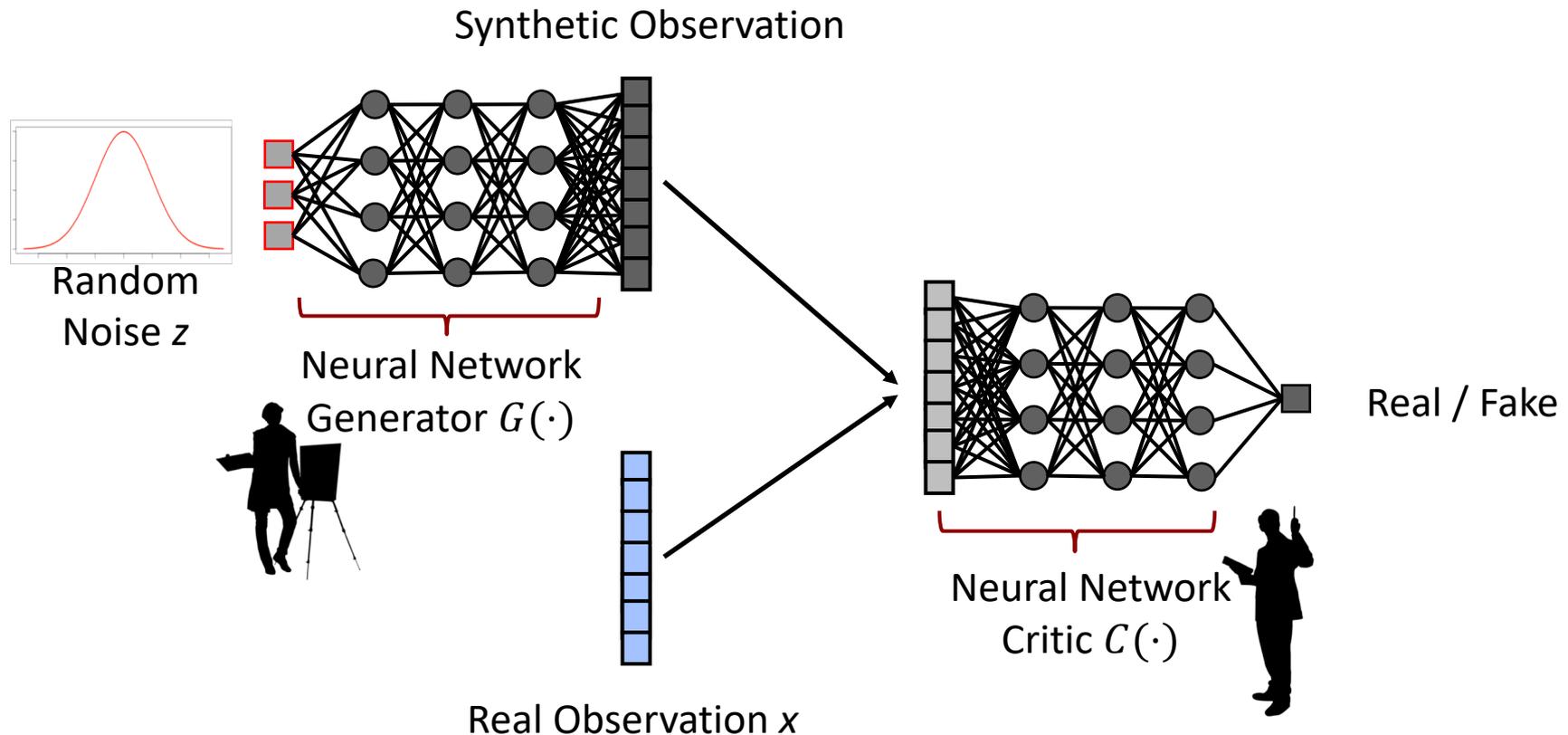
https://blog.insightdatascience.com/generating-custom-photo-realistic-faces-using-ai-d170b1b59255

Generative Adversarial Networks

Synthetic Observation

Random Noise *z*

Neural Network Generator $G(\cdot)$

Real Observation *x*

Neural Network Critic $C(\cdot)$

Real / Fake

$$\min_{G} \max_{C} \mathop{\mathrm{E}}_{x \sim P_{real}} [\log C(x)] + \mathop{\mathrm{E}}_{z \sim P_z(z)} [\log\{1 - C(G(z))\}]$$

Generative Adversarial Networks

Synthetic Observation

Random
Noise z

Neural Network
Generator $G(\cdot)$

Real Observation x

Neural Network
Critic $C(\cdot)$

Real / Fake

$$\min_{G} \max_{C} \ \mathop{\mathrm{E}}_{x \sim P_{real}} [\log C(x)] + \mathop{\mathrm{E}}_{z \sim P_z(z)} [\log\{1 - C(G(z))\}]$$

Generative Adversarial Networks

Synthetic Observation

Random Noise z

Neural Network Generator $G(\cdot)$

Real Observation x

$$\frac{\partial Score}{\partial w_G}$$

Neural Network Critic $C(\cdot)$

Real / Fake Score
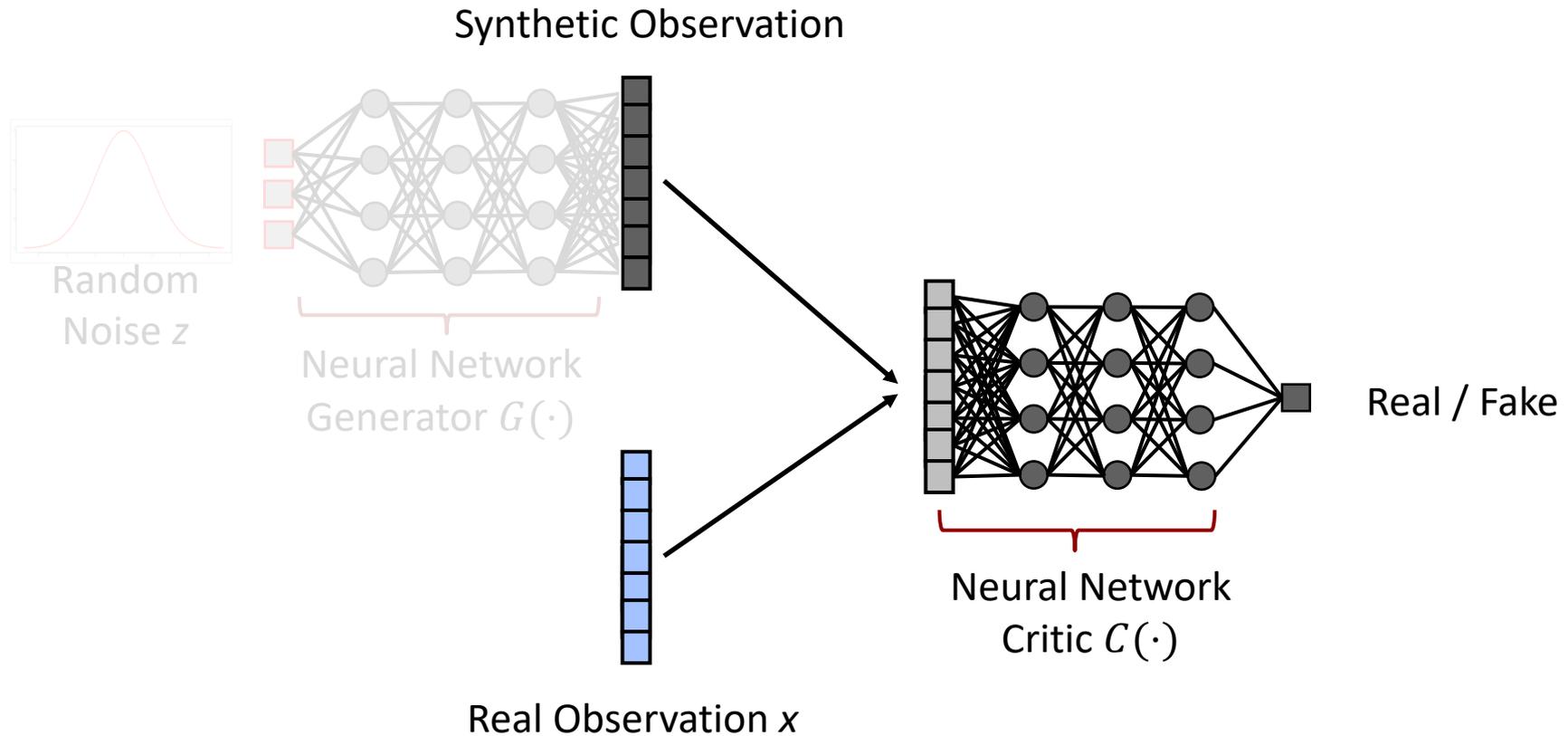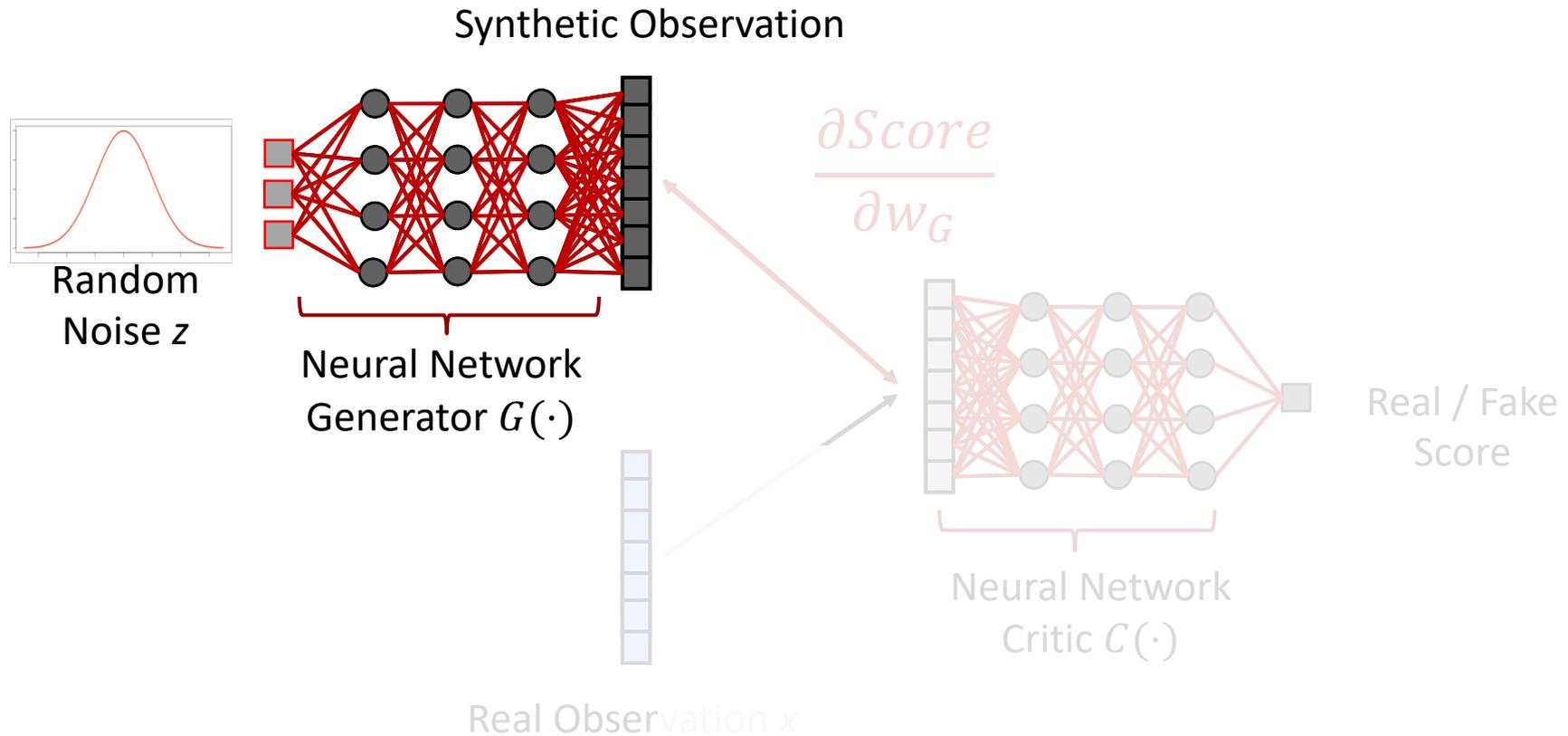
$$\min_{G} \max_{C} \operatorname*{E}_{x \sim P_{real}} [\log C(x)] + \operatorname*{E}_{z \sim P_z(z)} [\log\{1 - C(G(z))\}]$$

Conditional GAN

Group Information
(Stay / Churn)

Group Information
(Stay / Churn)

Real / Fake

Categorical GAN

Synthetic Observation

$$\frac{\exp(x_k)}{\sum_i \exp(x_i)}$$

0.1
0.7
0.2

ZIP code

1
0
0

Real Observation

How real?

Embedding Layer
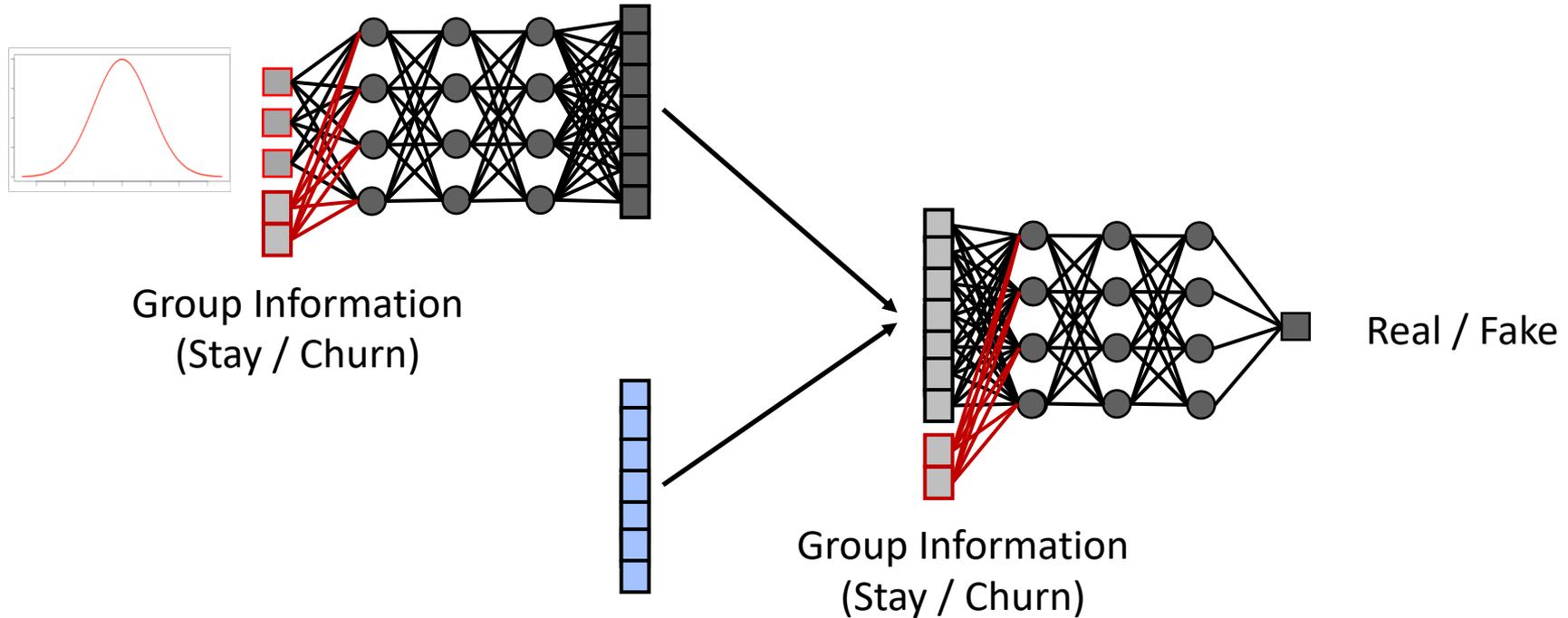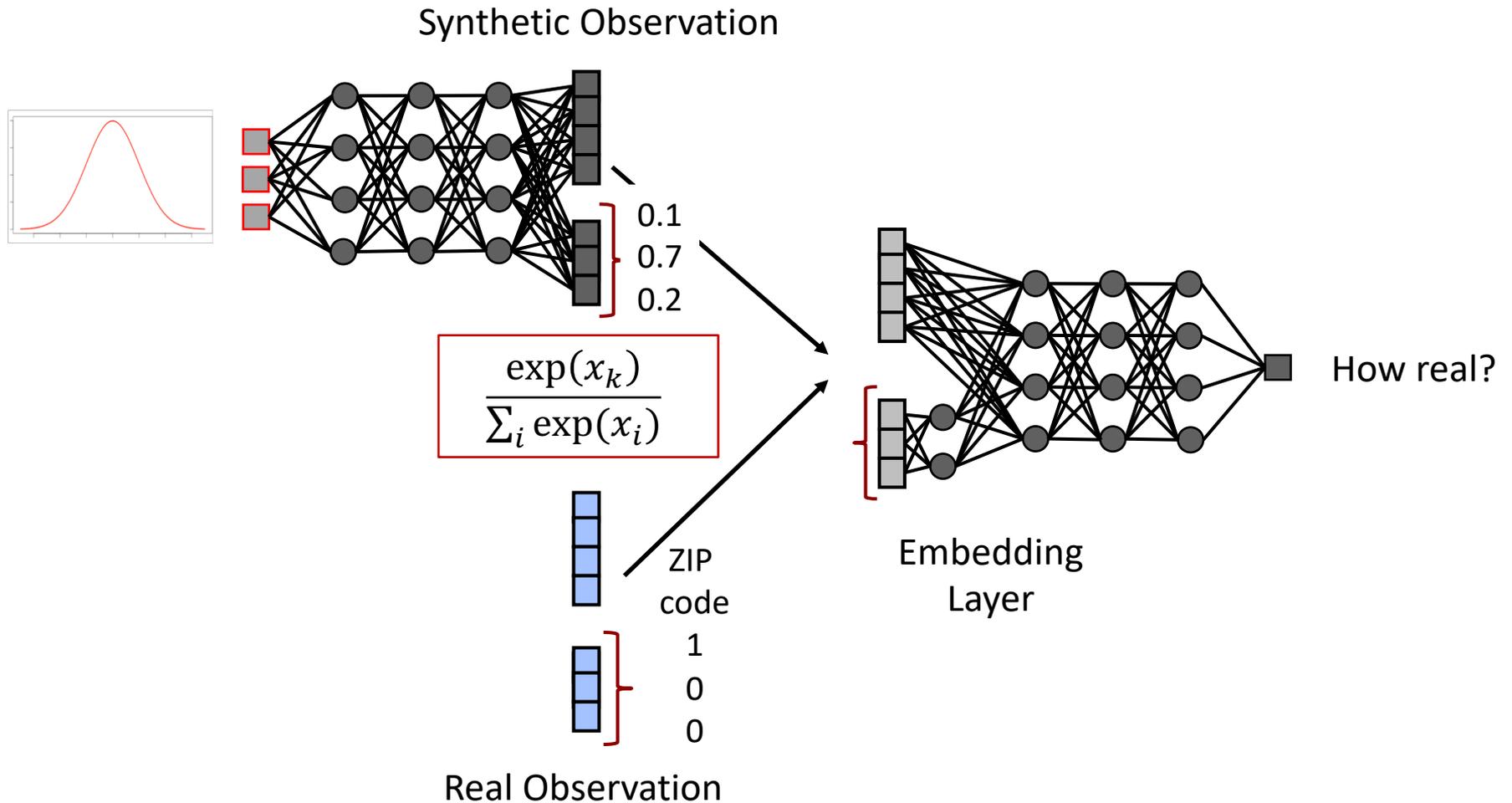
Empirical Evaluation

| Objective	Industry	Source	Obs.	Numeric Var.	Categorical Var. (Levels)	P(+1)	Minority Cases
Profitability	E-Commerce	DMC05	50,000	5	27 (651)	0.06	3000
Response	E-Commerce	DMC10	32,428	20	18 (209)	0.19	6161
Response	Finance	UCI-Coil00	9,822	59	26 (128)	0.06	589

Within 5-fold CV evaluation:
5-fold CV for Sampler + Model tuning
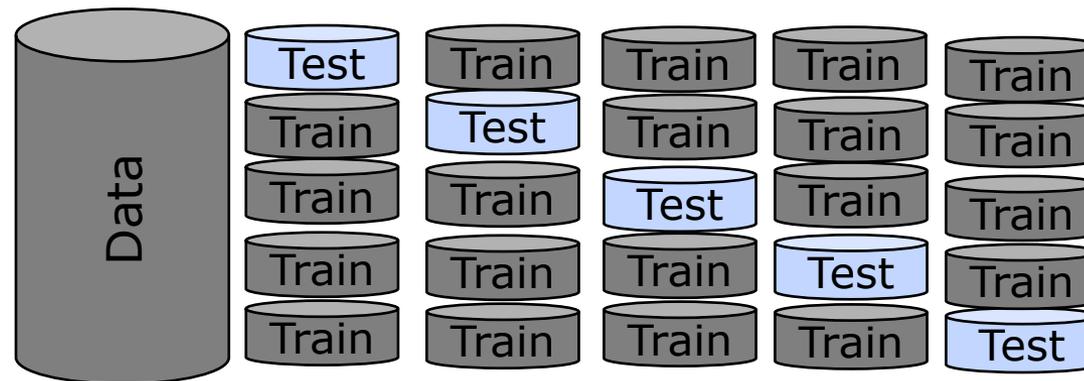
Empirical Evaluation
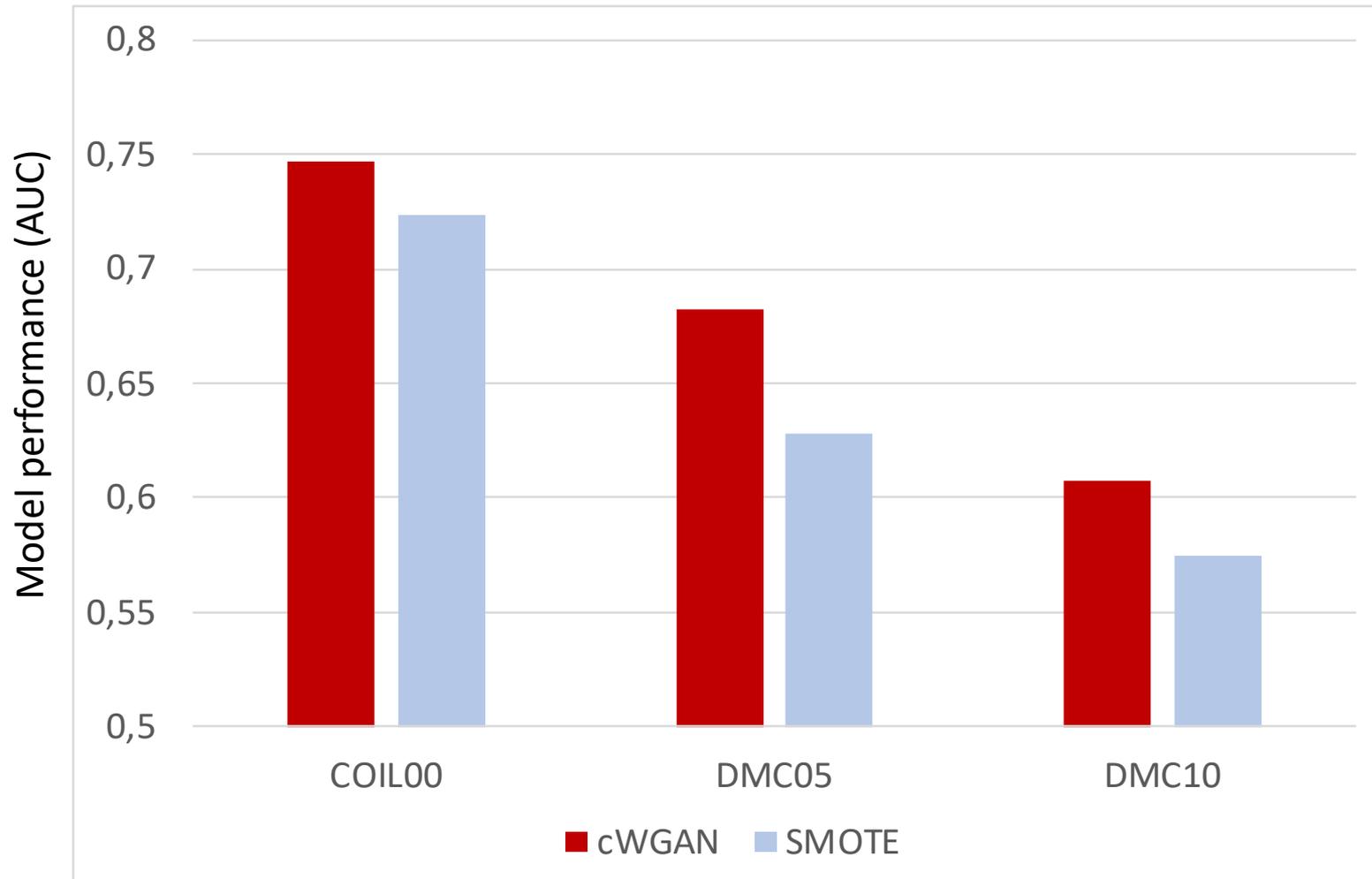
Empirical Evaluation
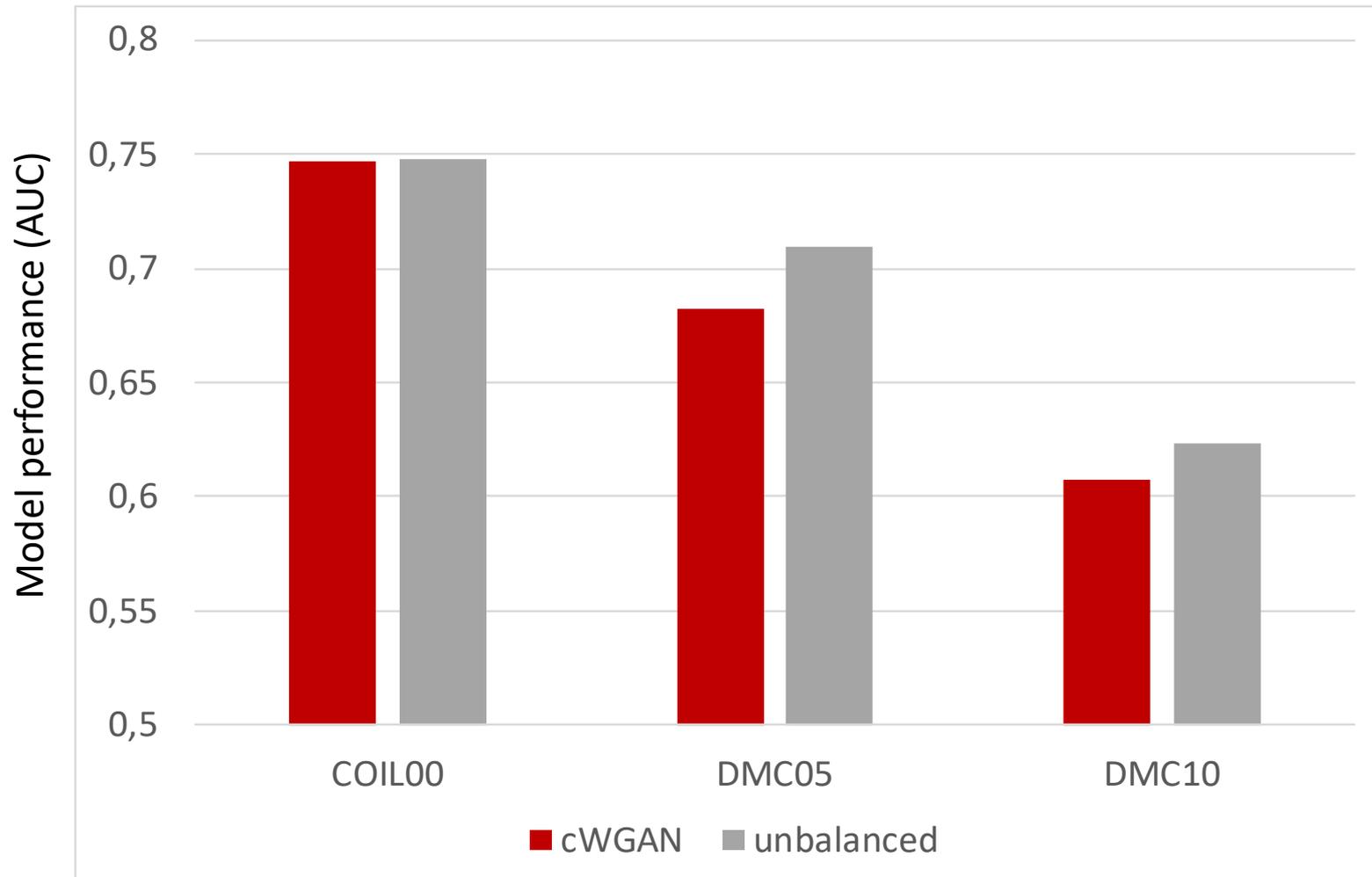
Takeaways

- **Generative Adversarial Networks (GAN) are powerful tools to generate synthetic data**

- **GANs are applicable to business data**

- **GANs show potential to improve existing scoring models**

Use Cases

Also use GANs to

- **Correct biased experiment samples**
- **Reduce bias in automatic decision making**
- **Create anonymized artificial data to share**
- **Multiply data that is costly to acquire**
- **Detect outliers and fraud**

Thank You For Your Attention

Prof. Dr. Stefan Lessmann

Johannes Haupt

Chair of Information Systems

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