EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

DSAR
DATA SCIENCE & ANALYTICS RESEARCH

Meeting at VSU, Voronezh

12/17/19, Vadim Borisov, Research Assistant / PhD student at Uni Tuebingen

Outline

- Who am I?
- The University of Tübingen, Germany / M
- Data Science and Analytics Research (DSAR) Group
- Explainable AI (XAI) - CancelOut
- Fairness in AI
- Deep Learning for Tabular Data

Who am I?

- **2013**: Got the master's degree from Voronezh State Technical University;
- **2014**: Moved to Bremen, Germany,
- **2015**: Did an internship in Kobe, Japan - Nanotubes and etc
- **2016**: Got my second master's degree from City Bremen University of Applied Science.
- **2016**: Fraunhofer Institute in Dresden, Germany
- **2017**: Research Assistant at Reutlingen University, Germany
- Since **2018**: I'm a Research Assistant at Tübingen University doing Machine Learning (ML).

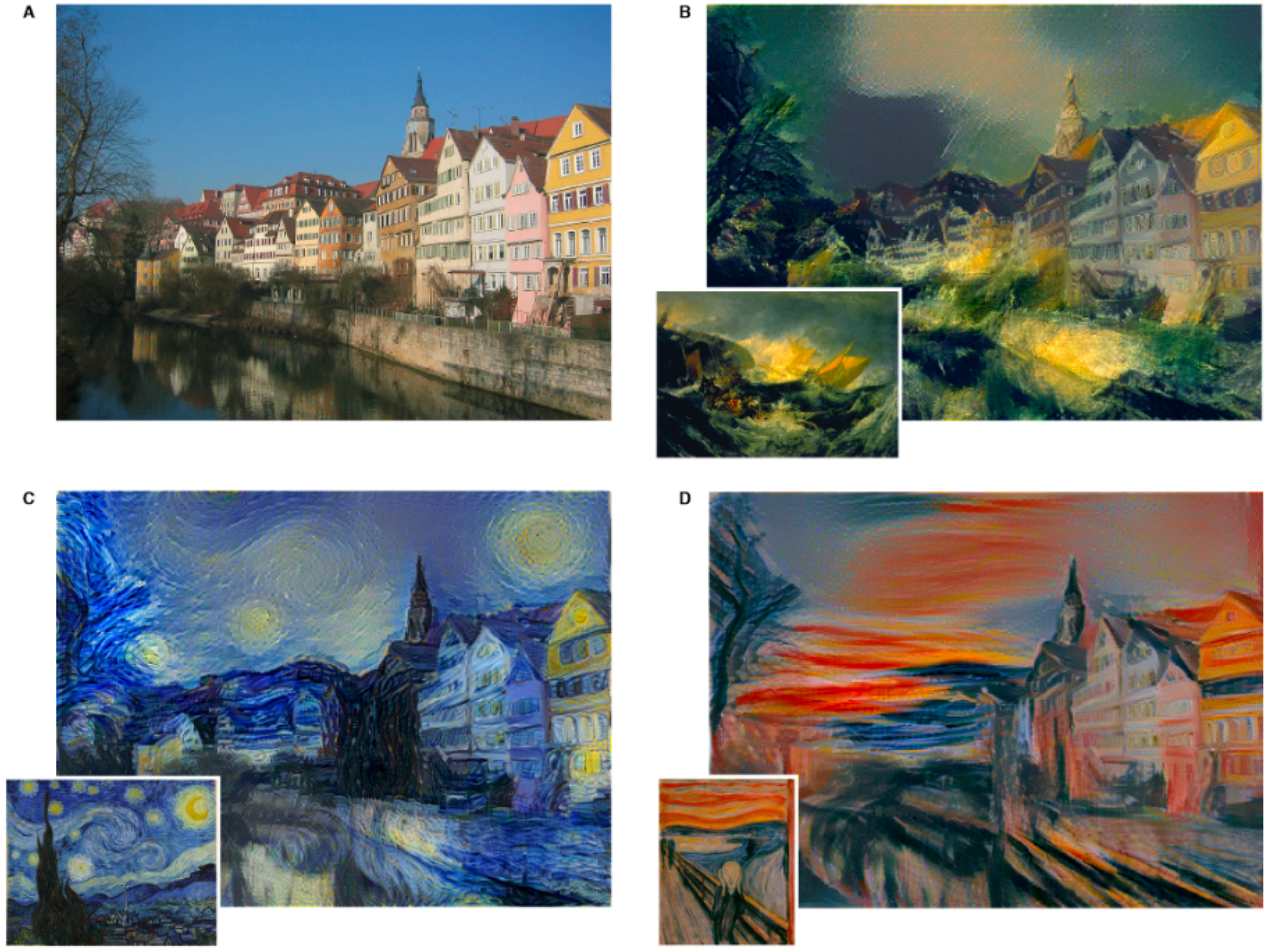
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Tübingen, Germany

The University of Tübingen

Founded in 1477.
Students ++28,515
Top 10 Universities in Germany.
An AI capital of Germany.

The University of Tübingen (Bethge Lab)



L. A. Gatys, A. S. Ecker, and M. Bethge Image Style Transfer Using Convolutional Neural Networks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016

http://bethgelab.org/research/machine_learning/style_transfer/

Max Planck Institute for Intelligent Systems, Tübingen

Bernhard Schölkopf,
Head of MPI Tübingen

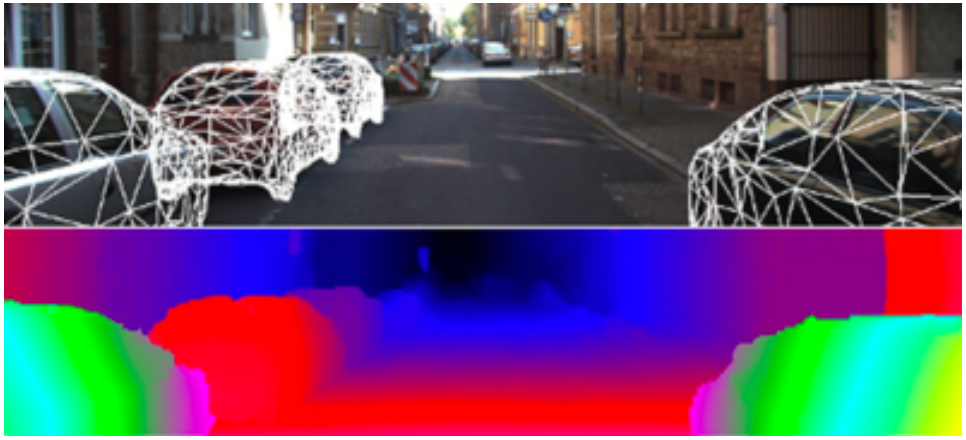
MAX-PLANCK-GESELLSCHAFT

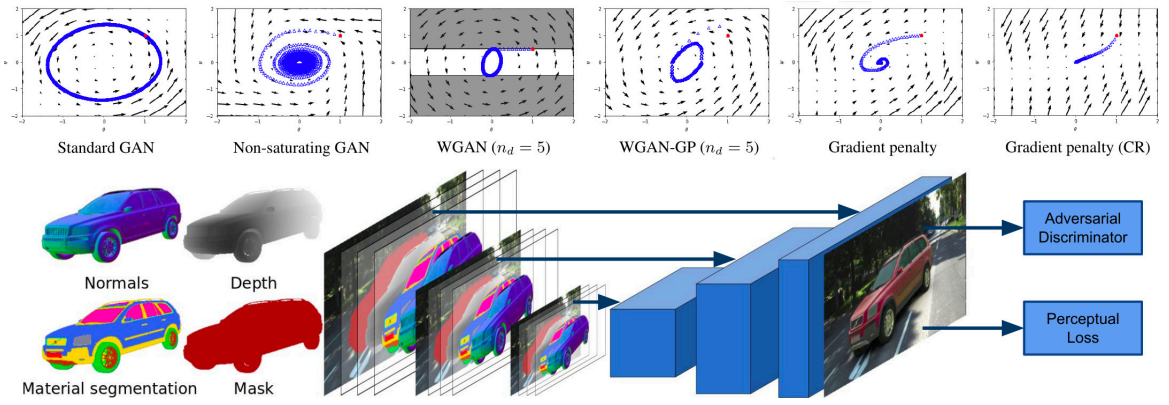
BOSCH
Invented for life

amazon

Max Planck Institute for Intelligent Systems, Tübingen - Autonomous Driving

Andreas Geiger





Data Science and Analytics Research (DSAR) Group

Our Research

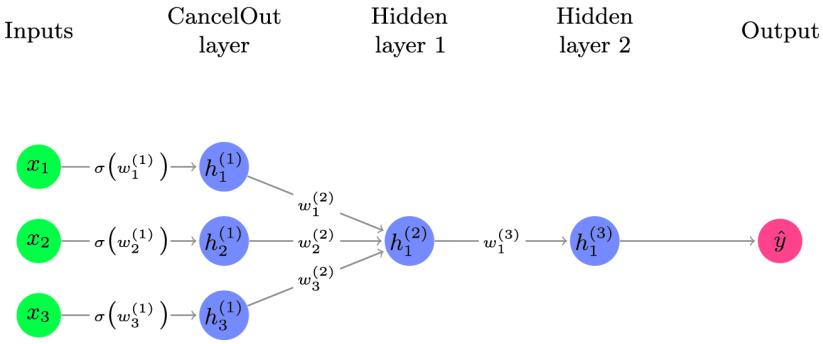
Online Feature Selection

pystreamfs

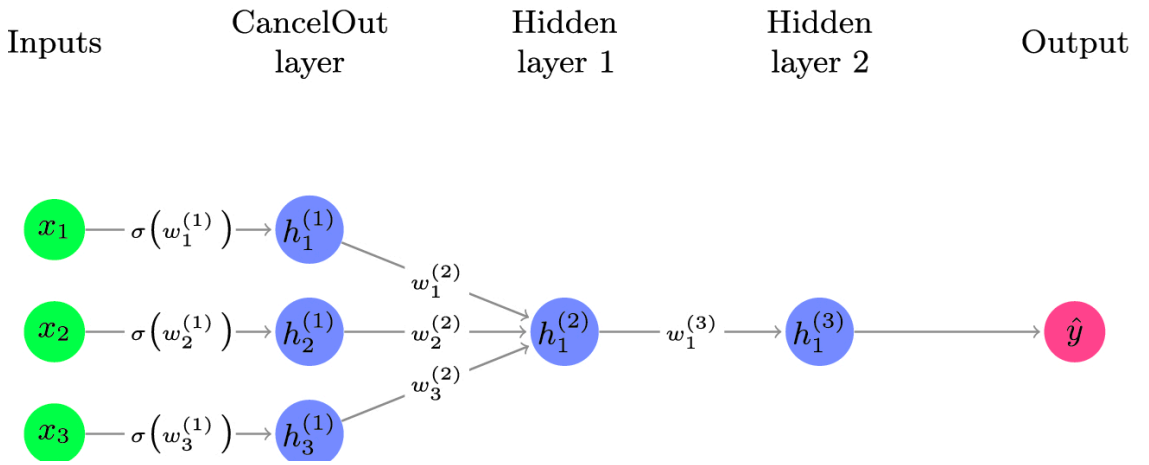
Gaussian Process Scaling using Deep Learning

Empowerment Through Attainable Counterfactuals (NeurIPS 2019 paper "Towards User Empowerment" M.Pawelczyk et. al.)

CancelOut: A Layer for Feature Selection in Deep Neural Networks (ICANN 2019 paper, V.Borisov et. al.)



CancelOut Layer for NNs



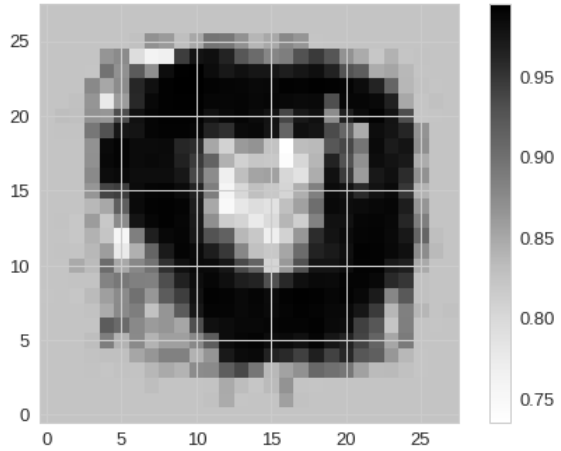
$$CancelOut(\boldsymbol{X}) = \boldsymbol{X} \odot g(W_{CO}) \quad (1)$$

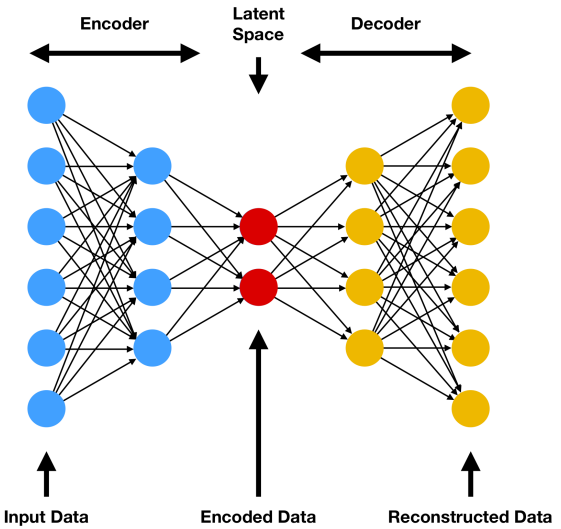
where $\odot$ indicates an element-wise multiplication, $\boldsymbol{X}$ is an input vector $\boldsymbol{X} \in \mathbb{R}_v^N$, W_{CO} is a weight vector $W_{CO} \in \mathbb{R}_v^N$, N_v is the feature size, and g is an activation function. Note, $g(x)$ denotes here element-wise application, e.g. $\boldsymbol{X} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, then $g(\boldsymbol{X}) = g\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = \left(\begin{bmatrix} g(a) \\ g(b) \\ g(c) \end{bmatrix}\right)$.

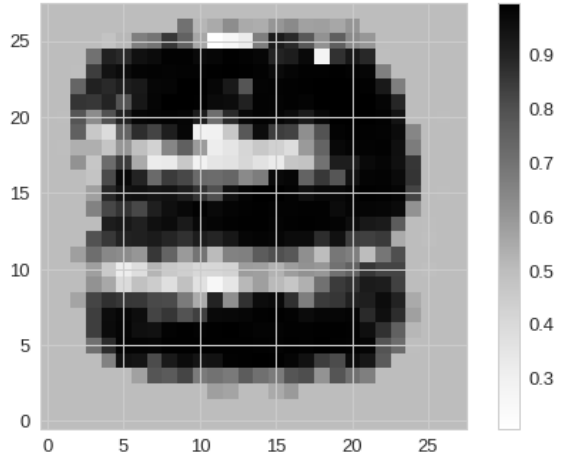
GitHub: https://github.com/unnir/CancelOut

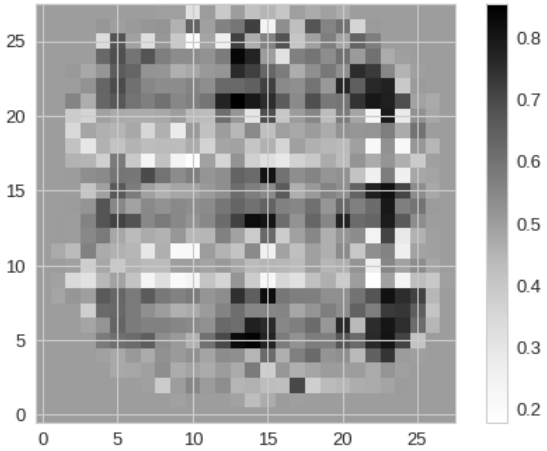
Paper: CancelOut: A Layer for Feature Selection in Deep Neural Networks (ICANN 2019 paper, V.Borisov et. al.)

CancelOut Layer for NNs - Unsupervised Feature Selection









VAE (C-CHVAE) for Counterfactual Search

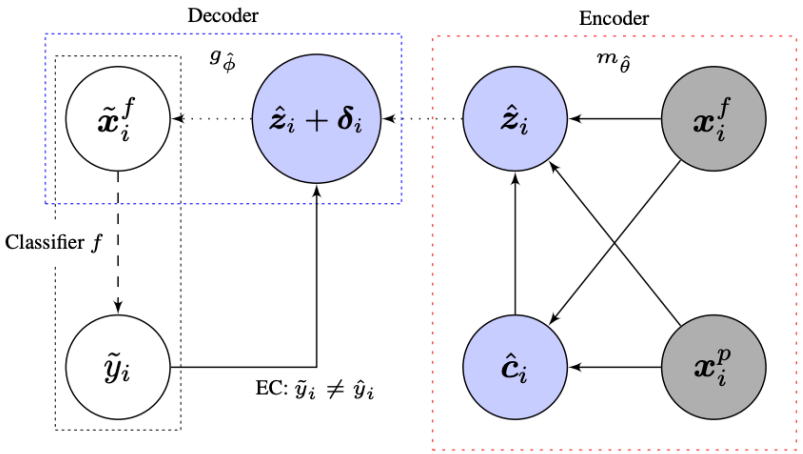
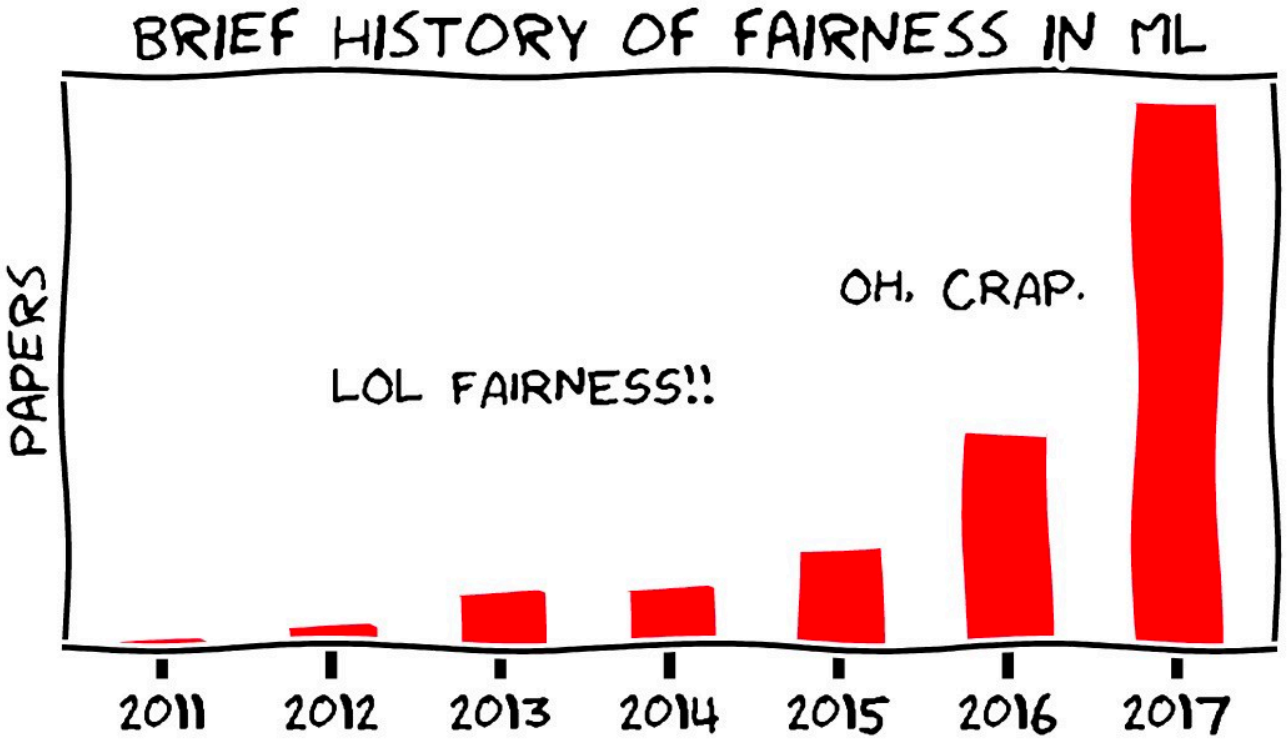


Figure 1: **Counterfactual search**. The learned encoder, $m_{\hat{\theta}}$, maps heterogeneous protected and free features, $\boldsymbol{x}^p$ and $\boldsymbol{x}^f$, and latent mixture components, $\hat{\boldsymbol{c}}$, into a latent representation, $\hat{\boldsymbol{z}}$. The learned decoder, $g_{\hat{\phi}}$, reconstructs the free inputs $\boldsymbol{x}^f$ from the perturbed representation, providing a potential counterfactual, $\tilde{\boldsymbol{x}} = (\boldsymbol{x}^p, \tilde{\boldsymbol{x}}^f)$. The counterfactual acts like a typical observation from the data distribution. Next, we feed the potential counterfactual $\tilde{\boldsymbol{x}}$ to the classifier, f. We stop the search, if the EC condition is met.

$$\min_{\boldsymbol{\delta}} \sum_i \|\boldsymbol{x}_i^f - g_{\hat{\phi}}(\hat{\boldsymbol{z}}_i + \boldsymbol{\delta}_i; \boldsymbol{x}_i^p)\| \text{ s.t. } f(g_{\hat{\phi}}(\hat{\boldsymbol{z}}_i + \boldsymbol{\delta}_i; \boldsymbol{x}_i^p), \boldsymbol{x}_i^p) \neq f(\boldsymbol{x}_i^f, \boldsymbol{x}_i^p) \; \forall i.$$

Fairness in AI



Fairness in AI

- Criminal justice: recidivism algorithms (COMPAS)
  - Predicting if a defendant should receive bail
  - Unbalanced false positive rates: more likely to wrongly deny a black person bail

Table 1: ProPublica Analysis of COMPAS Algorithm

	White	Black
Wrongly Labeled High-Risk	23.5%	44.9%
Wrongly Labeled Low-Risk	47.7%	28.0%

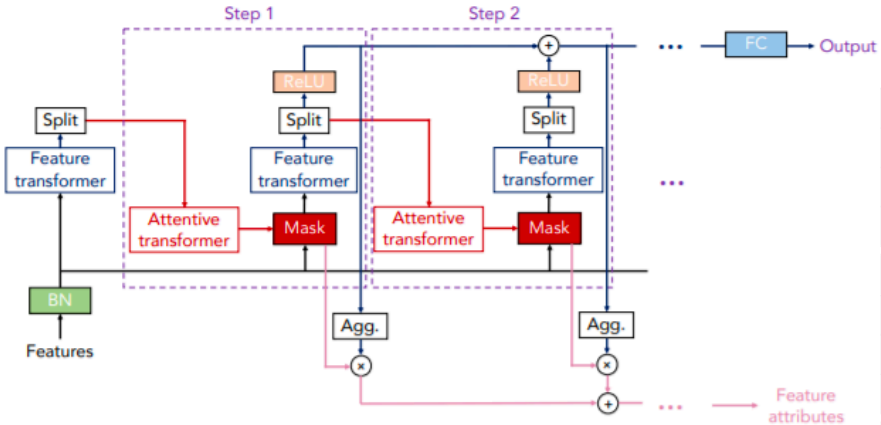
`https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing`

+++ What are we doing?

Deep Learning for Tabular Data

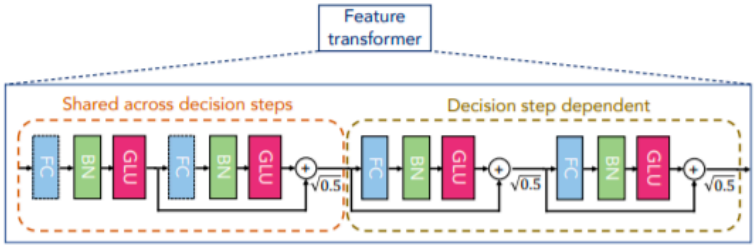
	A	B	C	D	E	F
1	Country	Salesperson	Order Date	OrderID	Units	Order Amoun
2	USA	Fuller	1/01/2011	10392	13	1,440.00
3	UK	Gloucester	2/01/2011	10397	17	716.72
4	UK	Bromley	2/01/2011	10771	18	344.00
5	USA	Finchley	3/01/2011	10393	16	2,556.95
6	USA	Finchley	3/01/2011	10394	10	442.00
7	UK	Gillingham	3/01/2011	10395	9	2,122.92
8	USA	Finchley	6/01/2011	10396	7	1,903.80
9	USA	Callahan	8/01/2011	10399	17	1,765.60
10	USA	Fuller	8/01/2011	10404	7	1,591.25
11	USA	Fuller	9/01/2011	10398	11	2,505.60
12	USA	Coghill	9/01/2011	10403	18	855.01
13	USA	Finchley	10/01/2011	10401	7	3,868.60
14	USA	Callahan	10/01/2011	10402	11	2,713.50
15	UK	Rayleigh	13/01/2011	10406	15	1,830.78
16	USA	Callahan	14/01/2011	10408	10	1,622.40
17	USA	Farnham	14/01/2011	10409	19	319.20
18	USA	Farnham	15/01/2011	10410	16	802.00

Deep Learning for Tabular Data

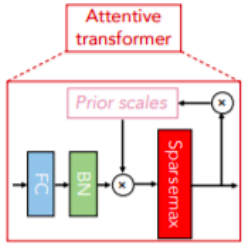


(a) TabNet architecture

Model	*Test accuracy (%)*
XGBoost*	89.34*
LightGBM*	89.28*
CatBoost*	85.14*
AutoML Tables (2 node hours)**	94.56**
AutoML Tables (5 node hours)**	94.95**
AutoML Tables (10 node hours)**	96.67**
AutoML Tables (30 node hours)**	96.93**
TabNet	**96.99**



(b) Feature transformer



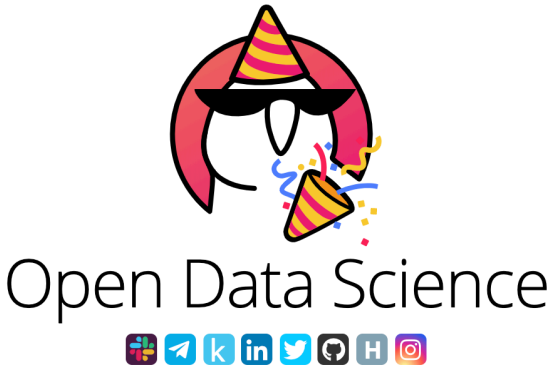
(c) Attentive transformer

TabNet from Google Brain.

https://arxiv.org/abs/1908.07442

My Tips for Mastering ML.

- Go to the lectures!
- Read papers if you want to understand the theory;
- Do Kaggle if you want to know the ML tools;
- ODS.ai;
- Collaboratel;
- И всегда спрашивайте вопросы!



Thank you.

Vadim Borisov
Sand 14 - C207
72074 Tübingen · Germany
Phone: +49 7071 29-00000
vadim.borisov@uni-tuebingen.de