

Responsible AI - Der Einsatz Künstlicher Intelligenz (KI) in der Öffentlichen Verwaltung durch erklärbare und faire KI- Methoden

Zukunftskongress-digital: Science Dialog 2020

Universität des Saarlandes und Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)

Oliver Gutermuth, Lea Mayer, Nijat Mehdiyev, Prof. Dr. Peter Fettke

Berlin, 16. Juni 2020

AUSGANGSLAGE

- 2018: Bundesweit **13.525** Prüferinnen und Prüfer aktiv
- Prüfung von **188.973** der **7.816.301** gemeldeten Betriebe
- Prüfungsquote von nur **2,4 %**
- Ergebnis der Prüfungen: Mehrergebnis i. H. v. **13,9 Mrd. €**
- **KI-Bedarf:** Fachkräftemangel, kognitive Belastung, Effektivität und Effizienz von Entscheidungen

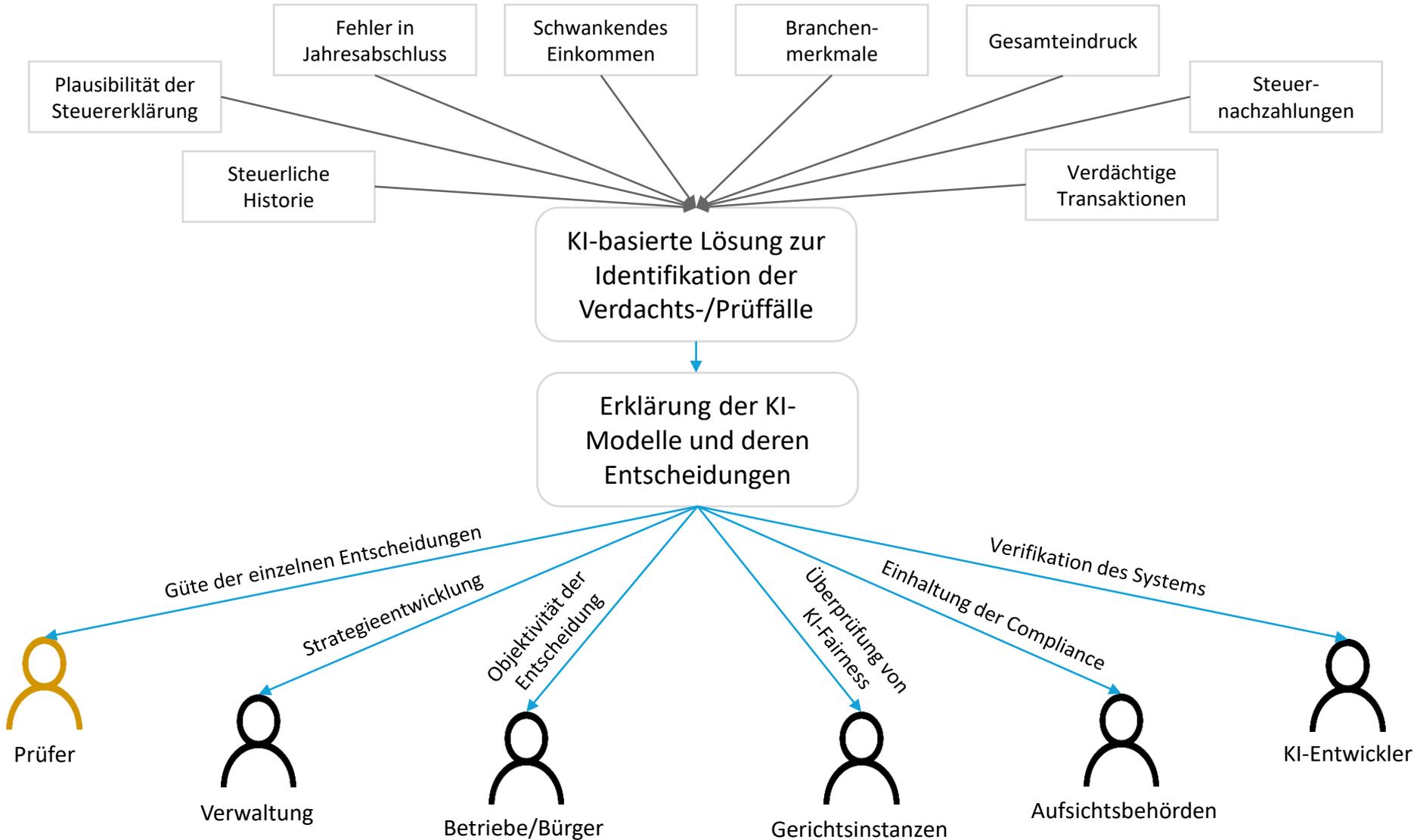
KI - HERAUSFORDERUNGEN

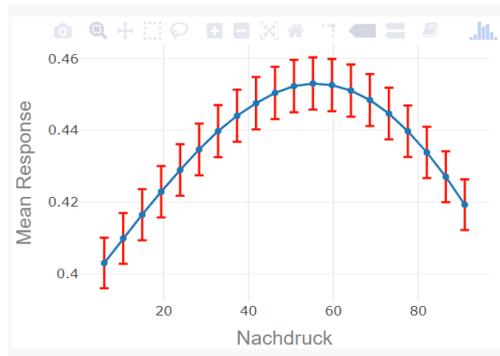
- Black-Box-Charakter
- Algorithmus-Aversion
- Akzeptanzbarrieren
- Bias und Ungerechtigkeit



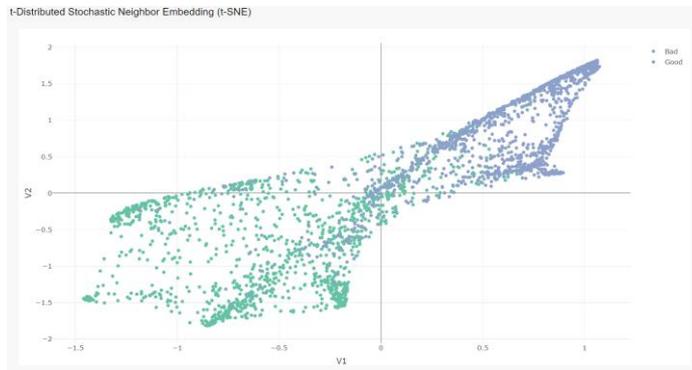
KI - NUTZEN

- **Effizienzsteigerung** durch
 - Automatisierung kognitiver Prozesse
 - Entscheidungsunterstützung bei kognitiven Prozessen
- **Innovation** in Form von
 - Neuen Geschäftsprozessen
 - Neuen Produkten und Dienstleistungen





Partial Dependence Plots



t-SNE

Deep Case: Example Based Explanation

Select Case-ID for explanation: 6

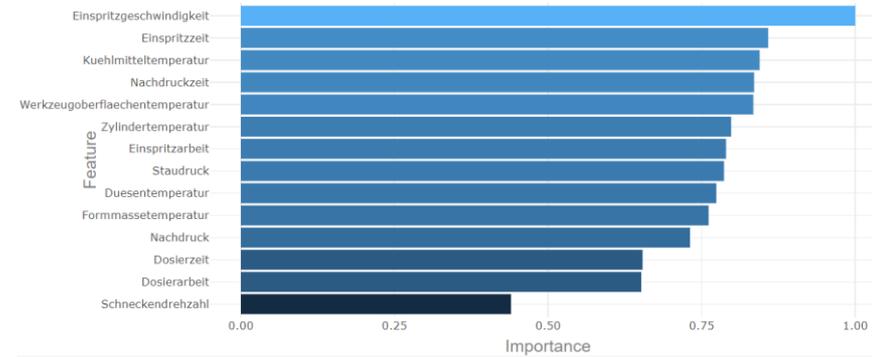
Case to Explain	Similar Cases	1	2	3	4	5	6	7	8	9	10
Nachdruck	11	200	1110	99	25	440	1000	1000	1740	150	150
Nachdruck	10	42	80	44	18	10	85	42	26	21	
Nachdruck	89	80	46	40	71	42	35	73	87	28	
Einspritzgeschwindigkeit	23	30	70	5	38	7	39	6	24	29	
Schneckenbreitheit	62	29	43	36	72	65	53	23	29	85	
Zylinderentemperatur	98	72	50	63	64	64	85	71	74	58	
Kuehlmittemperatur	43	26	26	34	49	42	30	30	31	33	
Staudruck	98	62	80	60	85	74	55	49	84	75	
Duesentemperatur	28	36	63	15	67	28	42	15	16	77	
Einspritzzeit	42	23	37	47	27	81	54	69	88	87	
Werkzeugoberflaechentemperatur	78	70	78	74	60	87	78	75	53	69	
Dosierzeit	102	76	61	21	69	79	50	32	39	54	
Einspritzzeit	79	98	88	47	65	71	88	80	44	74	
Dosierzeit	90	79	85	42	67	66	67	80	88	36	
Formmassestemperatur	14	-7	-3	31	26	20	14	37	64	14	
Predicted Value	Bad	Bad	Bad	Bad	Bad	Bad	Bad	Bad	Bad	Bad	
Bad	94.02%	Bad	Bad	Bad	Bad	Bad	Bad	Bad	Bad	Bad	
Good	5.98%	Bad	91.62%	95.77%	82.84%	93.64%	92.51%	94.85%	96.96%	95.34%	95.02%
Good		Good	3.38%	4.23%	7.16%	6.36%	7.49%	5.15%	3.14%	4.68%	4.98%

Local Accuracy of Global Model: 86 %

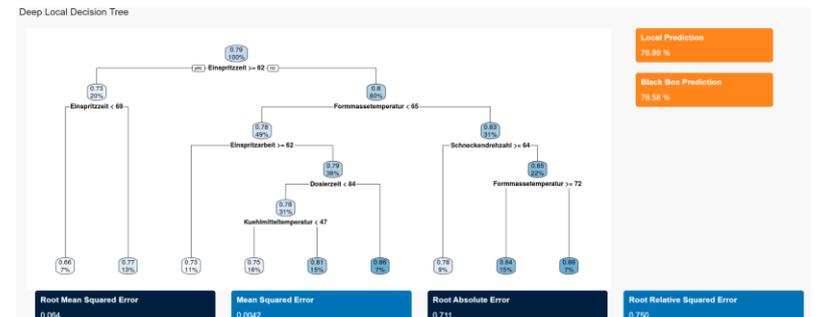
Fallbasierte und Kontrafaktische Erklärungen



Lokale Feature Attribution basierte Erklärungen

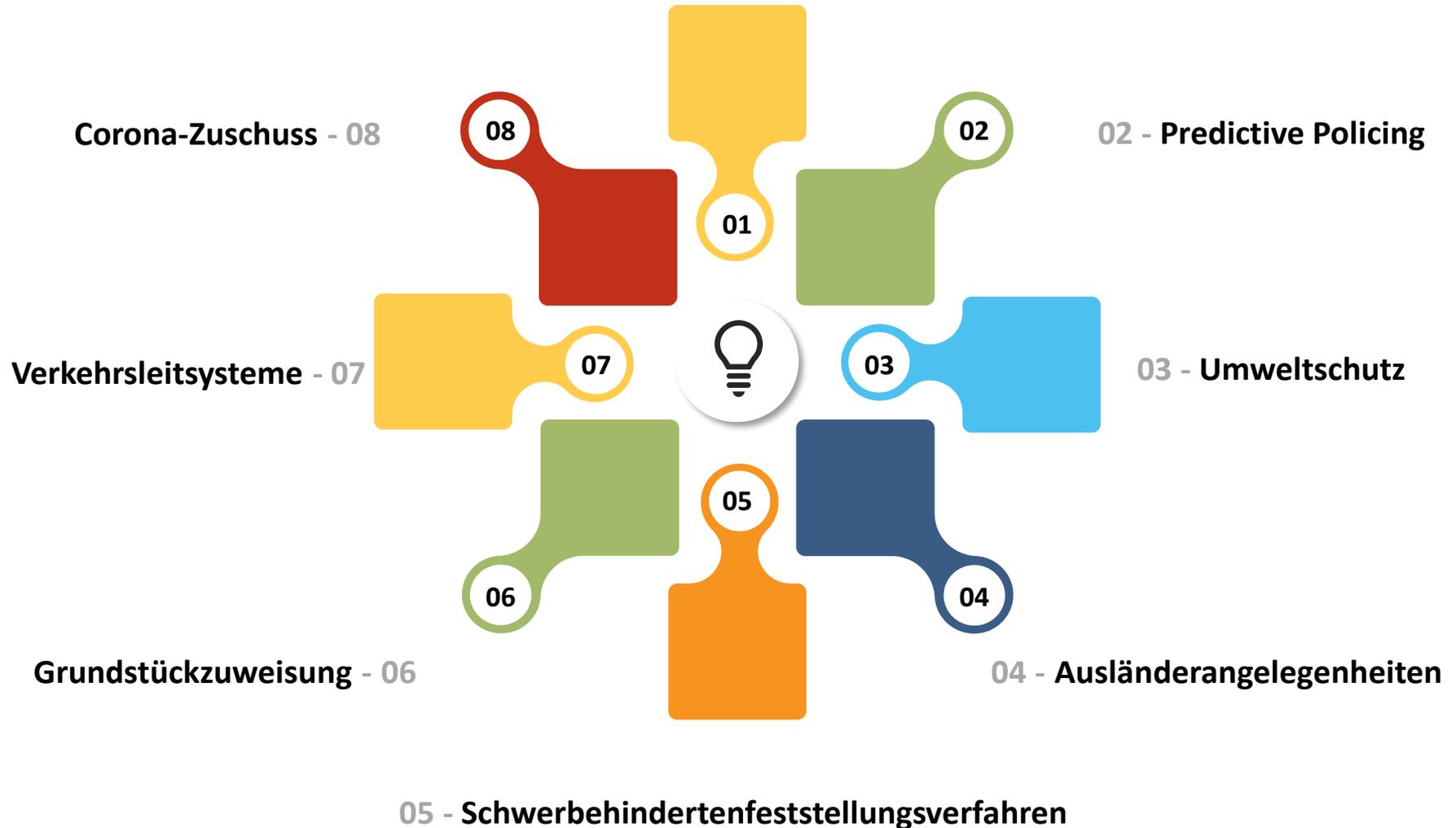


Feature Importance



Surrogate Baumbasierte Modelle

01 - Prüfung der Berechtigung auf Sozialhilfe



Forschungsziel: Entwicklung von XAI-Konzepten für die Öffentliche Verwaltung

- Verantwortung
- Transparenz
- (Soziale) Gerechtigkeit / Fairness

Etablierung von KI-Techniken in der Öffentlichen Verwaltung

- Stärkung von Akzeptanz und Vertrauen in die KI
- Anwendungsspezifische Erklärungsmethoden

Forschungsschwerpunkte

- Methoden zur Erklärung von KI (XAI)
- Schnittstellen zwischen KI und Verwaltungsabläufen
- Realisierung von Human-in-the-Loop-Konzepten



Vielen Dank für Ihre Aufmerksamkeit!

Kontakt

Oliver Gutermuth, Lea Mayer, Nijat Mehdiyev, Prof. Dr. Peter Fettke
Universität des Saarlandes und Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) und
Campus D 3 2
66123 Saarbrücken, Germany
[Oliver.Gutermuth, Lea.Mayer, Nijat.Mehdiyev, Peter.Fettke]@dfki.de
<http://www.dfki.de>

