IHRE DATEN. UNSERE EXPERTISE. INNOVATIVE LÖSUNGEN.

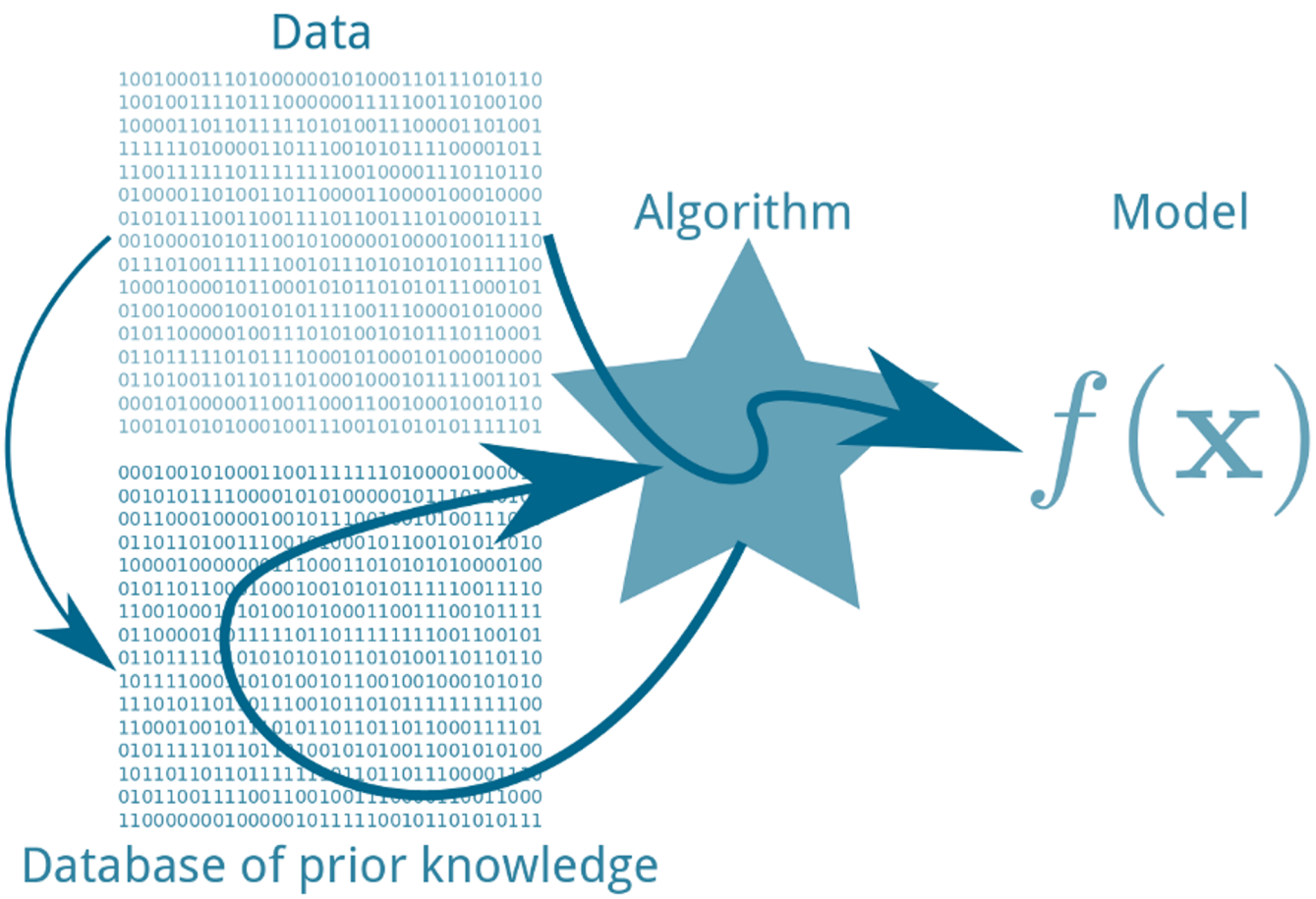
Simba n³ DATENLÖSUNGEN

TRANSFORMATION TIME: MEINE DATEN ALS ROHSTOFF

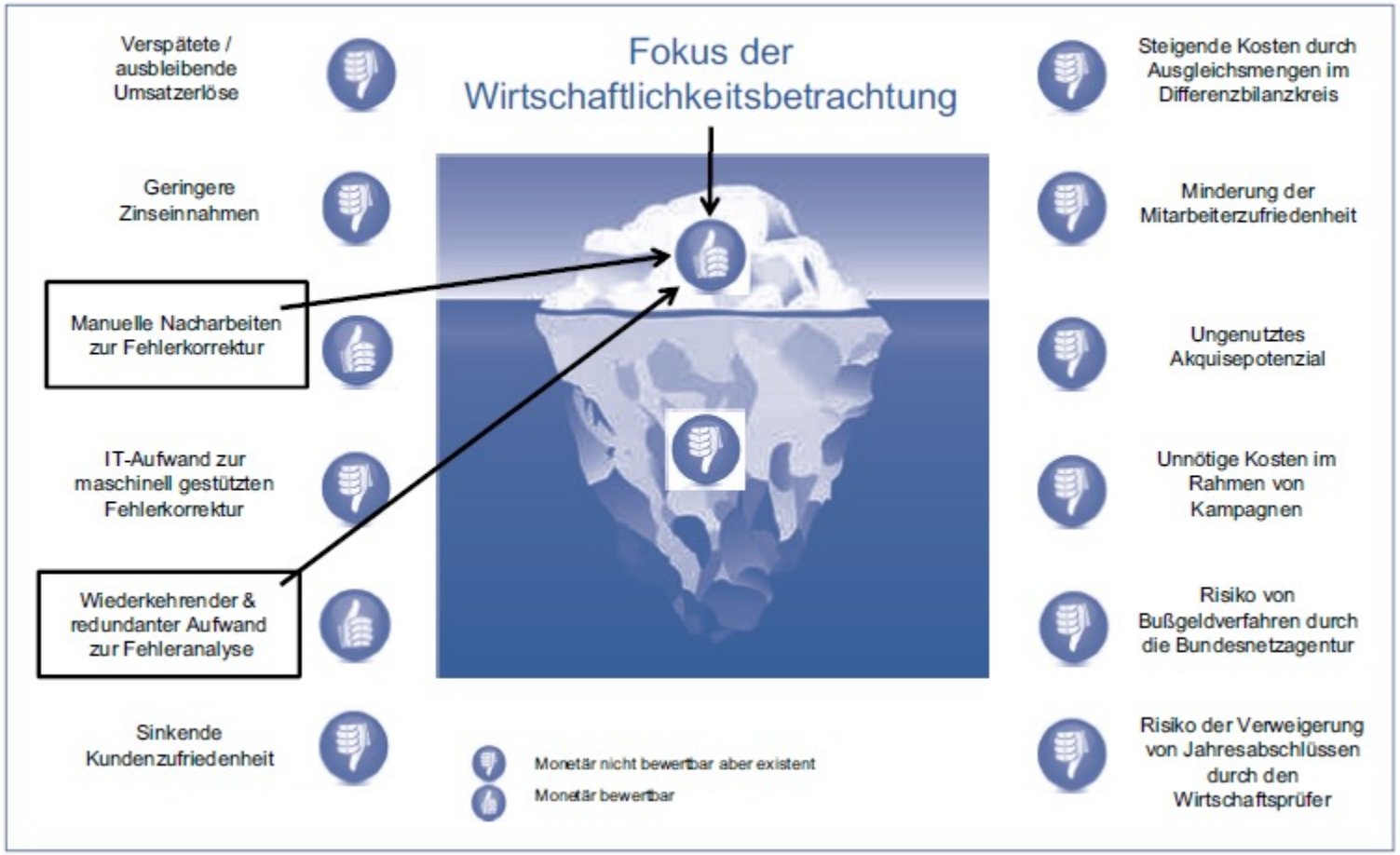
Simba n³ | Matthes Nagel

Südwestsachsen Digital | 05.07.2021

KI, AI, ML & Digitalisierung als Herausforderung



Mangelnde Datenqualität behindert die Digitalisierung



Auffinden unplausibler Datensätze: Kann man den Daten vertrauen?

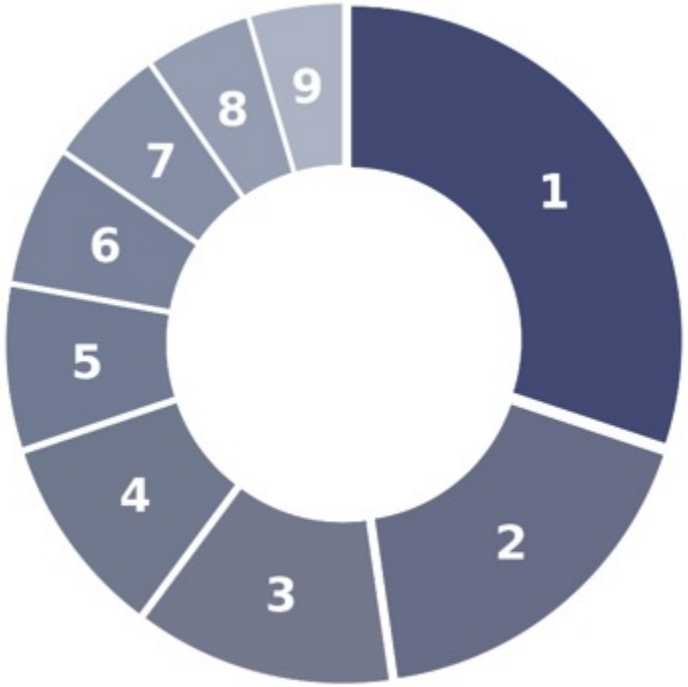
betroffene Datensätze	Ausreißer-stärke	Item1	Item2
82	326,6	Berufsgruppe=Rentner	Alter=40..50
82	284,9	Alter=30..40	Berufsgruppe=Rentner
164	171,5	Alter=10..20	Gemeinschaftskonto=ja
82	136,8	Alter=70..80	Berufsgruppe=Arbeiter
82	132	Familienstand=Kind	Alter=40..50
82	129,8	Familienstand=Kind	AnzahlAbbuchungen=100..200
82	122,2	Familienstand=verwitwet	Alter=30..40
246	121,6	AnzahlAbbuchungen=100..200	AnzahlZubuchungen=..5
164	109	AnzahlZubuchungen=30..40	AnzahlAbbuchungen=..10
82	98	Familienstand=Kind	Kontostand=20000..50000
82	96,6	Familienstand=Kind	Kontostand=10000..20000
328	87,8	Familienstand=Kind	Gemeinschaftskonto=ja
82	83,9	AnzahlAbbuchungen=50..70	AnzahlZubuchungen=..5
82	82	AnzahlZubuchungen=10..15	AnzahlAbbuchungen=300..500
82	79,7	Berufsgruppe=Techniker, Ingenieure	Alter=70..80
164	62,1	DauerKundenbezie-hung=25..29	Alter=20..30
164	57,6	Familienstand=Kind	Alter=30..40
164	52,6	Familienstand=Kind	Berufsgruppe=Angestellte
164	52,1	Alter=..10	Gemeinschaftskonto=ja
164	50,3	Alter=10..20	AnzahlAbbuchungen=200..300

betroffene Datensätze	Ausreißer-stärke	Item1	Item2
4	36,5	Verkauft=..05.2008	Emissionsklasse=Euro 4
4	36,1	Farbe=Schwarz	Lack=metallic
1	28,3	VK=6000..9000	EK=10500..13500
1	23,7	Modell=Caddy	Karosserie=Limousine
2	22	EZ=2008	Verkauft=..05.2008
1	20,3	Farbe=Weiß	Modell=Golf V
2	20,1	VK=9000..12000	EK=13500..16500
1	19,4	Durchschnittsverbrauch4.5..5.5	Verkauft=..05.2008
1	18,4	Verkauft=01.2009..03.2009	Zugang=01.2008..04.2008
2	18,1	Durchschnittsverbrauch=4.5..5.5}	Motor=Otto
1	16,7	Modell=Transporter T5	Motor=Otto
1	15,8	EK=22500..25500	Modell=Golf V
2	15,6	Antrieb=unbekannt	Emissionsklasse=Euro 4
1	15,5	EZ=2003	Modell=Golf V
1	15,3	Modell=Golf Plus	Lack=uni
1	15	Gewinn=..-4500	Motor=Otto
1	13,9	Gewinn=2500..3500	Modell=Golf V
1	12,7	Kosten=50..70	EK=7500..10500
1	12,5	Modell=Crafter	Antrieb=Frontantrieb
1	12,3	Verkauft=07.2009..09.2009	Zugang=01.2008..04.2008
1	12,1	EK=28500..	Motor=Otto
1	11,6	Art=Neufahrzeug	VK=9000..12000
1	11,4	Zugang=10.2007..01.2008	Verkauft=11.2008..01.2009
1	11,5	Art=Neufahrzeug	Zugang=04.2008..07.2008
2	11,5	EK=19500..22500	VK=12000..15000
2	11,3	Modell=Transporter T5	Karosserie=Limousine
1	11	EK=4500..7500	Modell=Passat Variant
1	10,5	VK=..6000	Farbe=Schwarz
2	10,4	EZ=..2001	Motor=Diesel
1	10	Modell=Transporter T5	Farbe=Schwarz

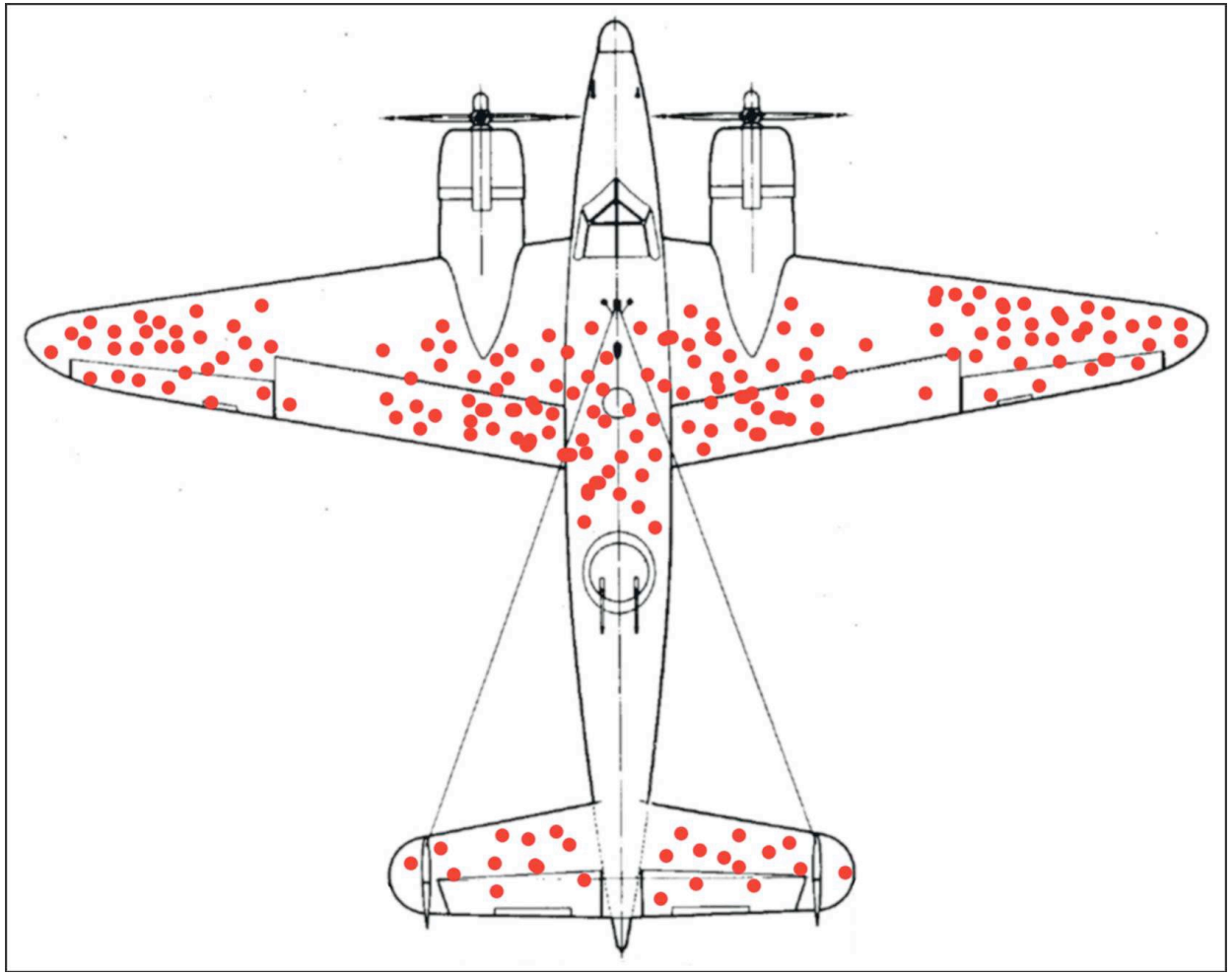
Betrugserkennung: Wie Algorithmen Steuersündern das Fürchten lehrt

Führende Ziffer	Wahrscheinlichkeit
1	30,1 %
2	17,6 %
3	12,5 %
4	9,7 %
5	7,9 %
6	6,7 %
7	5,8 %
8	5,1 %
9	4,6 %

https://de.wikipedia.org/wiki/Benfordsches_Gesetz



Alle Modelle sind falsch, einige sind nützlich



https://commons.wikimedia.org/w/index.php?curid=53081927, Abrufdatum 30.4.2018.

Simba n³ DATENLÖSUNGEN

VIELEN DANK!

Simba n³ GmbH | Dr.-Friedrichs-Straße 42, D-08606 Oelsnitz

Telefon | +49 37421 7224-0

E-Mail | info@nhochdrei.de

www.nhochdrei.de

Matthes Nagel

Managing Director

Simba n³ GmbH
Dr.-Friedrichs-Straße 42
D-08606 Oelsnitz

www.nhochdrei.de

Telefon | +49 37421 7224-103

E-Mail | matthes.nagel@nhochdrei.de