



# **Anomalieerkennung mit SAS Viya in den Eingangskanälen der HUK- COBURG**

# Vorstellung des Kontextes



**Wonach suchen Sie?**

**eVB-Nummer - Versicherungsbestätigung zur Zulassung/Ummeldung eines Fahrzeugs** ^

Möchten Sie ein neues Fahrzeug zulassen oder ein bereits versichertes Fahrzeug ummelden (z.B. Wohnungs- oder Halterwechsel)?

Gerne hilft Ihnen unser **eVB-Assistent**, damit Sie schnell die passende eVB-Nummer erhalten.

**Jetzt eVB-Assistent nutzen! Klicken Sie hier.** >

**eVB-Nummer - Wiederzulassung Ihres Fahrzeugs** v

**eVB-Nummer - Kurzzeitkennzeichen beantragen** v

**Kontaktformular**

**Ihre Nachricht**

**Thema**

Bitte wählen Sie v

**Versicherungsscheinnummer (optional)**

\_\_\_/\_\_\_\_\_-

z.B. 660/123456-A

**Nachricht** 1900

Bitte geben Sie Ihre Nachricht ein.

Ihre Angaben behandeln wir zur Bearbeitung Ihres Anliegens selbstverständlich nach Datenschutzgrundverordnung und Bundesdatenschutzgesetz.

[Hier finden Sie weitere Informationen zur Verwendung Ihrer personenbezogenen Daten und zu Ihren Rechten.](#)

gabe vornehmen.

KFZ-Versicherung

Welche Daten Ihres Kfz-Vertrags möchten Sie ändern?  
Unten sehen Sie eine Auswahl der häufigsten Themen, aber natürlich kann ich auch mehr.

Fahrer

Wie lange soll die Fahrerkreiserweiterung andauern?

Mehr als 21 Tage

Möchten Sie ihren Fahrerkreis dauerhaft erweitern?  
Gerne helfe ich Ihnen dabei. [Bitte klicken Sie hier.](#) Sie werden direkt zur Fahrerkreisänderung weitergeleitet.

👍 🗨

Nachricht schreiben >



# Vorstellung des Kontextes



# Dimensionen und Kriterien

## Anomalie

### Technisch/Formal

### Zeitlich

### Inhaltlich

### Beispiele

- › Unvollständige Anfragen, fehlende erwartete Eingaben, ungewöhnlich lange/kurze Texte, technisch fehlerhafte Parameter
- › Plötzliche Spitzen/Einbrüche einer Anfrageart; Häufigkeit deutlich höher/niedriger als üblich; Verschiebungen durch externe Ereignisse
- › Anfragen in untypischen Bereichen; ungewöhnliche Kombinationen von Informationen; Nichtbeachtung erwarteter Prozessmuster

### Indikatoren

- › Verletzung von Schema-/Validierungsregeln; Textlänge außerhalb üblicher Bandbreiten; unerwartete Wertebereiche
- › Abweichung von Basislinien und Saisonalität; signifikante Frequenzänderung gegenüber Erwartung
- › Thematische Unplausibilität; seltene Merkmalskombinationen; Kontext-Inkonsistenzen



# Klassen von Anomalien

## Seltene Ereignisse

› Signifikante, aber selten auftretende Abweichungen von der Norm

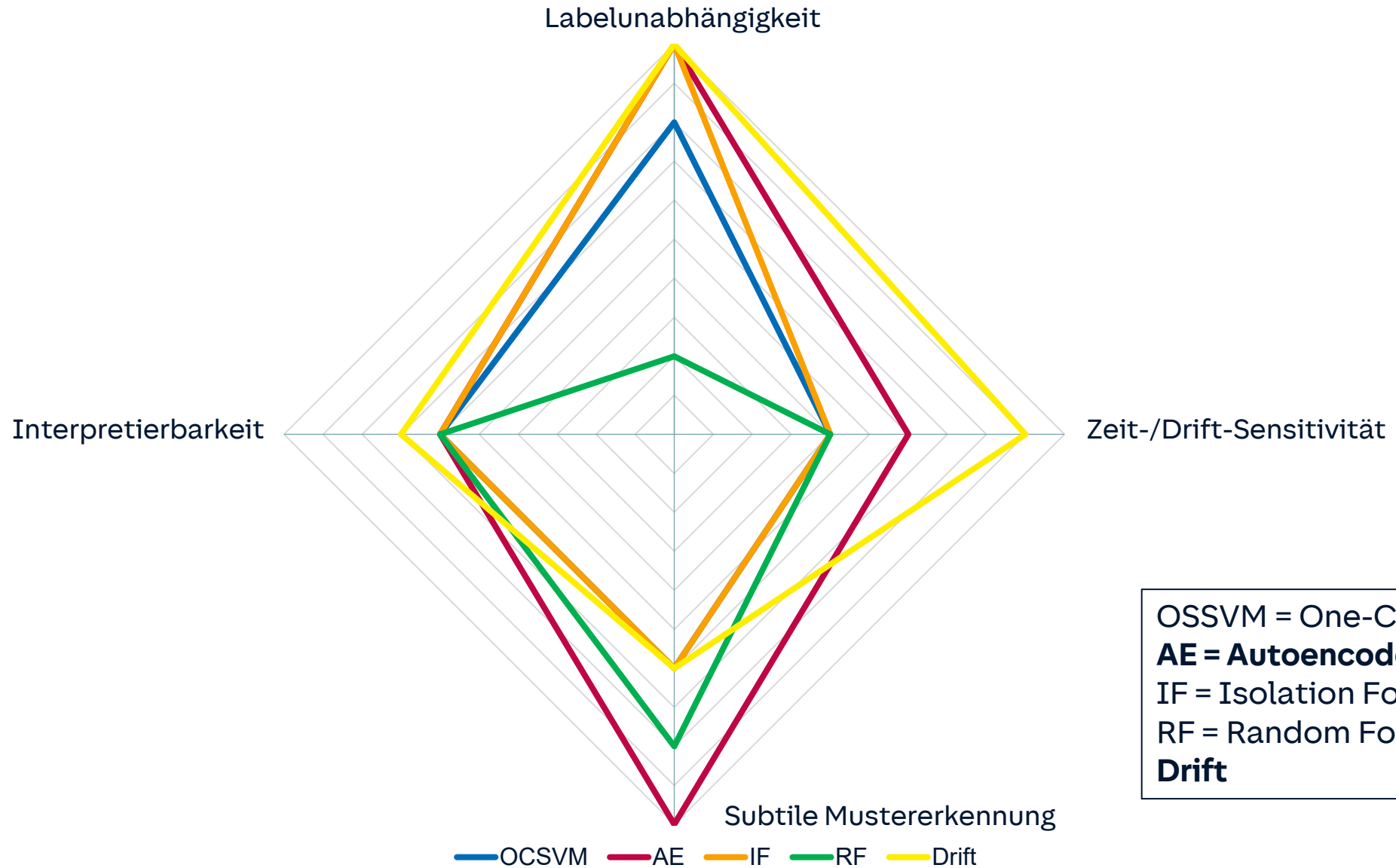
## Fehlerhafte Datenwerte

› Unplausible oder inkonsistente Messungen/Eingaben

## Ungewöhnliche Muster

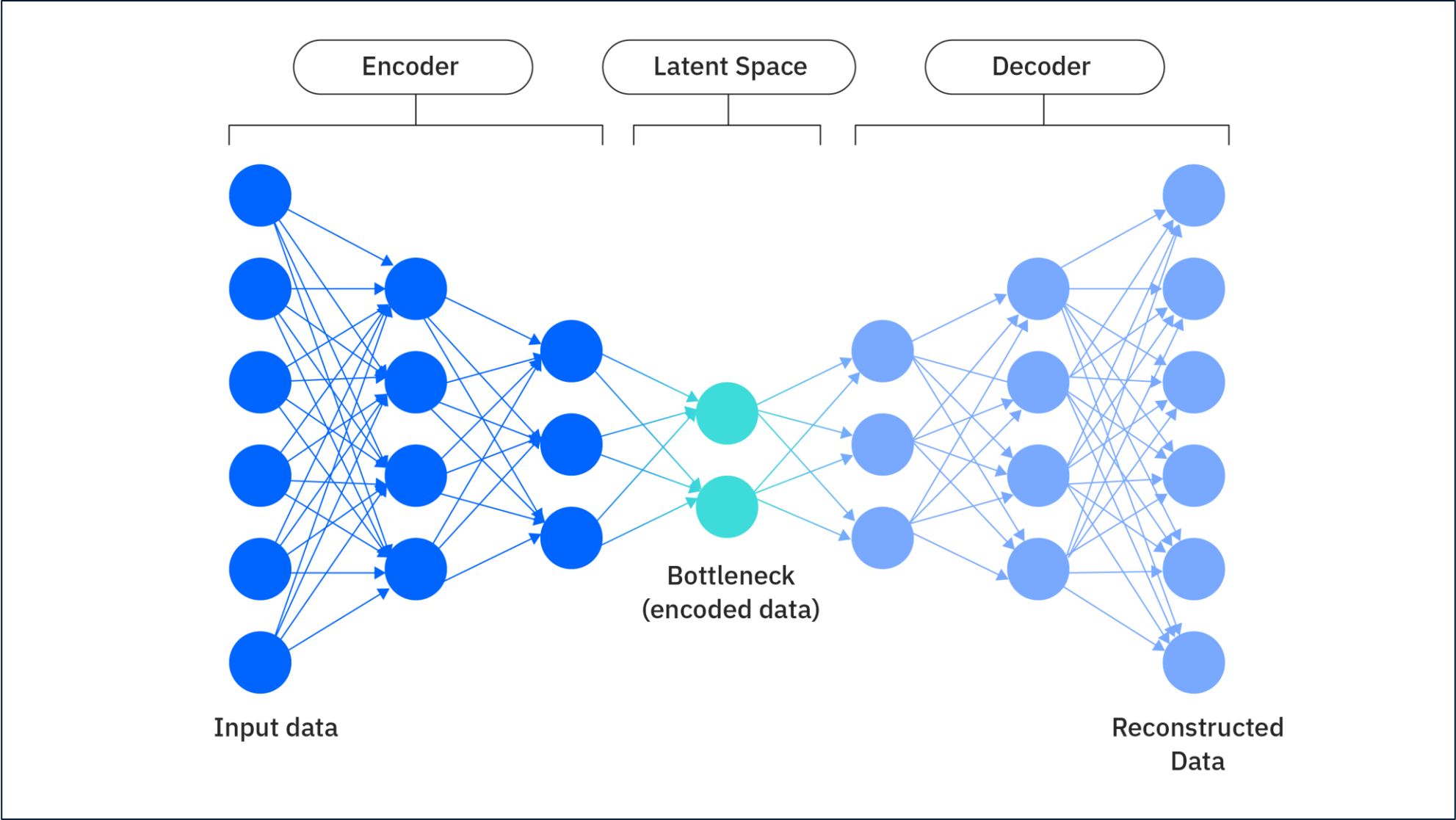
› Kombinationen und zeitliche Entwicklungen, die nicht der regulären Verteilung folgen

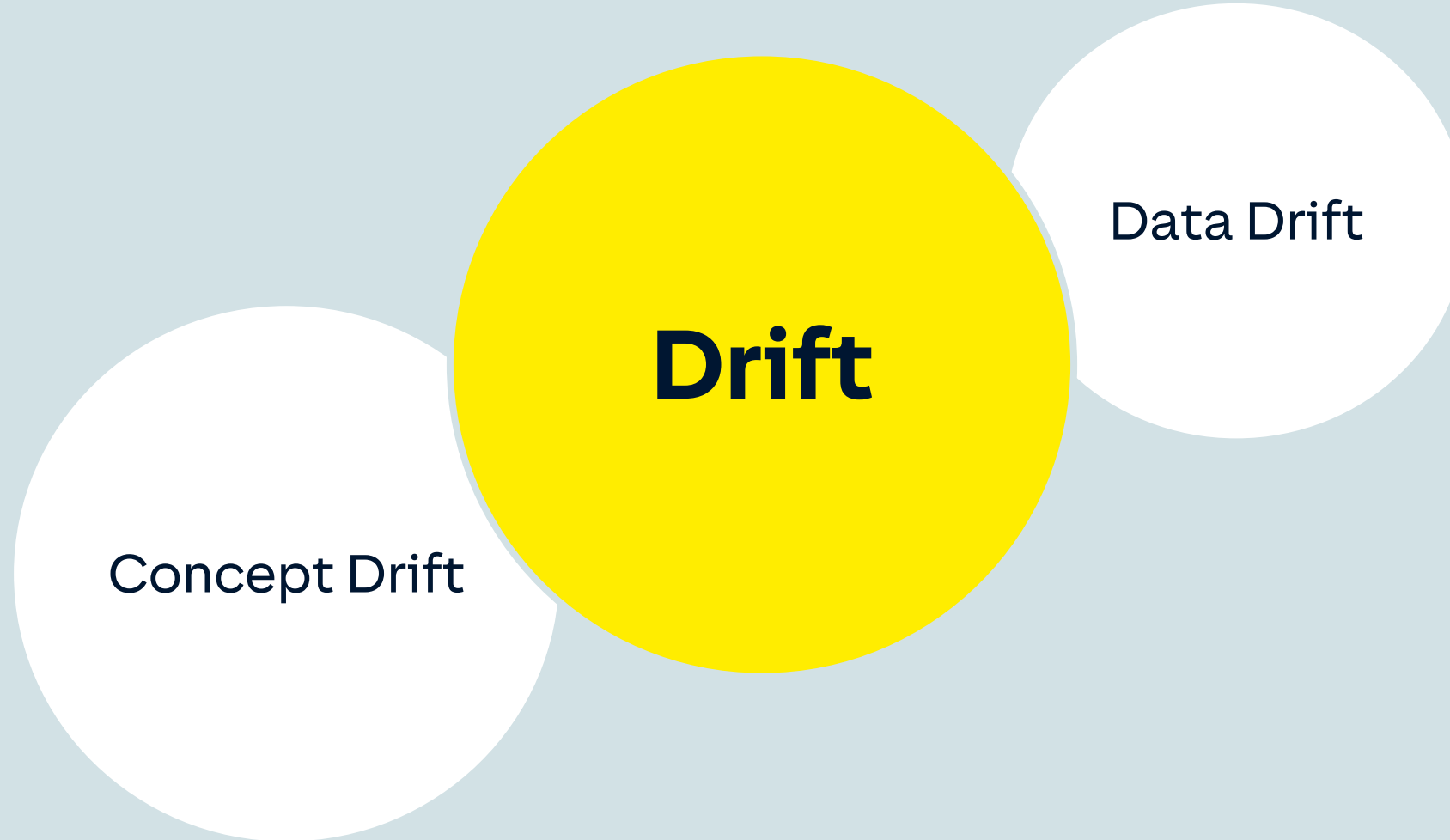
# Auswahl der Modelle



OSSVM = One-Class SVM  
**AE = Autoencoder**  
IF = Isolation Forest  
RF = Random Forest  
**Drift**

# Autoencoder







# Umsetzungsprozess



**1**

**SAS Data Integration Studio**

**2**

**SAS Studio - Datenvorverarbeitung**

**3**

**SAS Studio – Training Autoencoder**

**4**

**SAS Studio – Entwicklung Drifterkennung**

**5**

**SAS Visual Analytics - Reporting**

# Autoencoder im SAS Studio mit PROC NNET



## 1. Architekturwahl

- › Basierend auf Theorie und aktueller Fachliteratur
- › Mehrere Konfigurationen getestet

## 2. Hyperparameter-Tuning

- › Strukturierte Optimierung von Schlüsselparametern

Iterative  
Optimierung

A central diagram consisting of three concentric circles. The innermost circle is yellow and contains the text 'Iterative Optimierung'. The middle and outer circles are white with black outlines. Four curved arrows, one in each quadrant, point clockwise from the outer circle to the middle circle, and from the middle circle to the inner circle, indicating a continuous iterative process.

## 4. Evaluierung

- › Vergleich der Modellvarianten
- › Auswahl der besten Konfiguration

## 3. Training

- › Auswahl geeigneter Verlustfunktionen und Optimierer zur Minimierung des Rekonstruktionsfehlers
- › Einsatz von Regularisierung

# Autoencoder im SAS Studio mit PROC NNET

## › SAS Help Center: Syntax: NNET Procedure

```

1  cas mycas;
2
3  data YWH24SAS.iris;
4      set sashelp.iris;
5  run;
6
7  proc nnet data=YWH24SAS.iris;
8      input SepalLength SepalWidth PetalLength PetalWidth;
9      target Species / level=nom;
10     train outmodel=YWH24OUT.nnetModel2 seed=1517878693;
11     autotune useparameters=custom
12         tuningparameters=(nhidden(LB=0 UB=2 INIT=0)
13                             nunits1(LB=1 UB=10 INIT=1)
14                             nunits2(LB=2 UB=15 INIT=2)
15                             regl1(LB=1e-03 UB=1e-02 INIT=1e-03)
16                             regl2(LB=1e-03 UB=1e-02 INIT=1e-03)
17                             );
18     optimization algorithm=LBFSG maxiter=100;
19 run;

```

The NNET Procedure

Model Information	
Modell	Neuronales Netz
Anzahl verwendete Beobachtungen	150
Anzahl eingelesene Beobachtungen	150
Ziel-/Response-Variablen	Species
Anzahl der Knoten	19
Anzahl Eingangsknoten	4
Anzahl Ausgangsknoten	3
Anzahl versteckte Knoten	12
Anzahl verborgene Schichten	2
Anzahl Gewichtsparameter	72
Anzahl Verzerrungsparameter	15
Architektur	MLP
Startwert für Ausgangsgewichtung	1517878693
Optimierungstechnik	LBFSG
Anzahl neuronale Netze	1
Zielwert	2,1371389237
Misclassification Rate for Validation	1.53E8

Iteration History

Iteration Number	Objective Function	Norm of Gradient	Loss	Validate Error	Step Size	Norm			Fit Error
						L1	L2	Maximum	
1	3.894868	0.215252	3.833981	0.940000	0	7.683624	1.080283	0.250000	0.940000
2	3.851448	0.325019	3.807222	0.940000	4.119746	5.552524	0.955563	0.313233	0.666667
3	3.773640	0.490541	3.729513	0.340000	1	5.529889	1.015490	0.332929	0.340000
4	3.767057	0.500286	3.720079	0.340000	1	5.895647	1.030099	0.332929	0.166667
5	3.521995	1.424922	3.430937	0.180000	1	11.36784	2.353511	1.235338	0.180000
6	3.434008	1.528495	3.340263	0.180000	1	11.72289	2.306250	1.085284	0.160000
7	3.390787	1.577735	3.295257	0.153333	0.001204	11.95300	2.308720	1.098828	0.153333
8	2.449462	3.116909	2.240689	0.153333	0.125000	26.10984	5.119803	2.544243	0.373333
9	2.358028	3.128688	1.911085	0.266667	1	55.84258	11.28140	5.316181	0.266667
10	1.923664	0.932357	1.463106	0.266667	0.500000	57.51258	11.81233	5.780525	0.180000
11	1.790185	0.887614	1.363790	0.160000	1	53.24889	10.92151	5.127238	0.160000
12	1.730442	0.843469	1.309802	0.160000	0.066561	52.52650	10.79618	5.063980	0.153333
13	1.730442	0.843469	1.309802	0.153333	1.62E-18	52.52650	10.79618	5.063980	0.153333

Optimierung hat den gewünschten Zielwert erfüllt.

# Autoencoder im SAS Studio mit PROC NNET



ROC Information																					
Variable	Event	Cutoff	TP	FP	FN	TN	FPR	FDR	TPR	FNR	TNR	ACC	KS	Youden Index	F1 Score	F0.5 Score	AUC	Gini	Gamma	Tau	Misclassification (Event)
P_Species	Virginica	0	50	100	0	0	1	0.666667	1	0	0	0.333333	0	0	0.500000	0.384615	0.980900	0.961800	0.967022	0.430336	0.666667
		0.010000	50	50	0	50	0.500000	0.500000	1	0	0.500000	0.666667	0	0.500000	0.666667	0.555556	0.980900	0.961800	0.967022	0.430336	0.333333
		0.020000	50	50	0	50	0.500000	0.500000	1	0	0.500000	0.666667	0	0.500000	0.666667	0.555556	0.980900	0.961800	0.967022	0.430336	0.333333

Fit Statistics					
Number of Observations	Squared Error			Mean Consequential Error	Multiclass Log Loss
	Divisor of Average	Average	Root Average		
150	150	0.066776	0.258411	0.153333	0.331140

Tuner Information	
Modelltyp	Neuronales Netz
Tuner Zielfunktion	Fehlklassifizierungsfehler Prozent
Suchmethode	GA
Größe Grundgesamtheit	10
Maximalanzahl Iterationen	5
Maximale Tuning-Dauer (in Sekunden)	36000
Validierungstyp	Single Partition
Validierung Partitionsanteil	0,30
Log-Level	2
Seed	1517878693
Anzahl der parallelen Bewertungen	2
Anzahl der Worker je Subsession	1

Tuner Results							
Default and Best Configurations							
Evaluation	Hidden Layers	Neurons in Hidden Layer 1	Neurons in Hidden Layer 2	L1 Regularization	L2 Regularization	Misclassification Error Percentage	Evaluation Time in Seconds
0	0	0	0	0.001000	0.001000	2.22	0.93
9	2	4	8	0.007743	0.001292	0.00	0.52
12	2	6	9	0.003162	0.003162	0.00	0.42
31	2	6	9	0.010000	0.003162	0.00	0.45
33	2	2	9	0.003162	0.003162	0.00	0.51
37	2	6	9	0.003162	0.010000	0.00	0.45
38	2	6	9	0.003162	0.001000	0.00	0.34
45	0	0	0	0.001486	0.001248	0.00	0.64
49	2	6	12	0.003162	0.003162	0.00	0.28
52	2	6	9	0.003162	0.005623	0.00	0.38
53	2	6	9	0.003162	0.001778	0.00	0.33

# Autoencoder im SAS Studio mit PROC NNET

Tuner History All Evaluated Configurations								
Evaluation	Iteration	Hidden Layers	Neurons in Hidden Layer 1	Neurons in Hidden Layer 2	L1 Regularization	L2 Regularization	Misclassification Error Percentage	Evaluation Time in Seconds
0	0	0	0	0	0.001000	0.001000	2.22	0.93
1	1	2	6	2	0.001292	0.005995	2.22	0.54
2	1	2	7	3	0.010000	0.002783	2.22	1.02
3	1	1	2	0	0.001000	0.003594	2.22	0.37

Best Configuration	
Evaluierung	9
Verborgene Schichten	2
Neuronen in verborgener Schicht 1	4
Neuronen in verborgener Schicht 2	8
L1 Regularisierung	0,00774269
L2 Regularisierung	0,00129157
Fehlklassifizierungsfehler Prozent	0,00

Tuner Summary	
Initiale Konfiguration Zielwert	2.2222
Beste Konfiguration Zielwert	0
Schlechteste Konfiguration Zielwert	28.8889
Initiale Konfiguration Evaluierungszeit in Sekunden	0.9308
Beste Konfiguration Evaluierungszeit in Sekunden	0.5242
Anzahl verbesserte Konfigurationen	1
Anzahl evaluierte Konfigurationen	84
Gesamt-Tuning-Dauer (in Sekunden)	19.0000
Paralleles Tuning Speedup	1.8067

Tuner Task Timing		
Task	Seconds	Percent
Modell Training	21.46	62.51
Modell Scoring	11.89	34.63
Summe Zielevaluierungen	33.37	97.21
Tuner	0.96	2.79
Gesamt-CPU-Zeit	34.33	100.00

Hyperparameter Importance	
Hyperparameter	Relative Importance
REGL2	1.0000
NUNITS1	0.7634
REGL1	0.7388
NHIDDEN	0.5097
NUNITS2	0.3553



# Drift-Erkennung im SAS Studio als Custom-Step



Detect Data Drift

Run Cancel Refresh

Definitions Output Options About

The Detect Data Drift steps can help to identify drift in the input data comparing a baseline dataset against a current one. For this comparison to work a SAS formatted date variable is required.

Table in which to detect data drift:\*

WORK.HMEQ\_DATA\_DRIFT

Date column to use to create the two datasets:\*

captureDate

Type of interval used for comparison:

Day

Week

Semi-Month

Month

Quarter

Semi-Year

Year

Select all interval variables for tracking:

LOAN

VALUE

Select all nominal variables for tracking:

REASON

JOB

Results Output Data (9)

Distribution of LOAN

Curves

Normal(Mu=4837.1 Sigma=216.14)

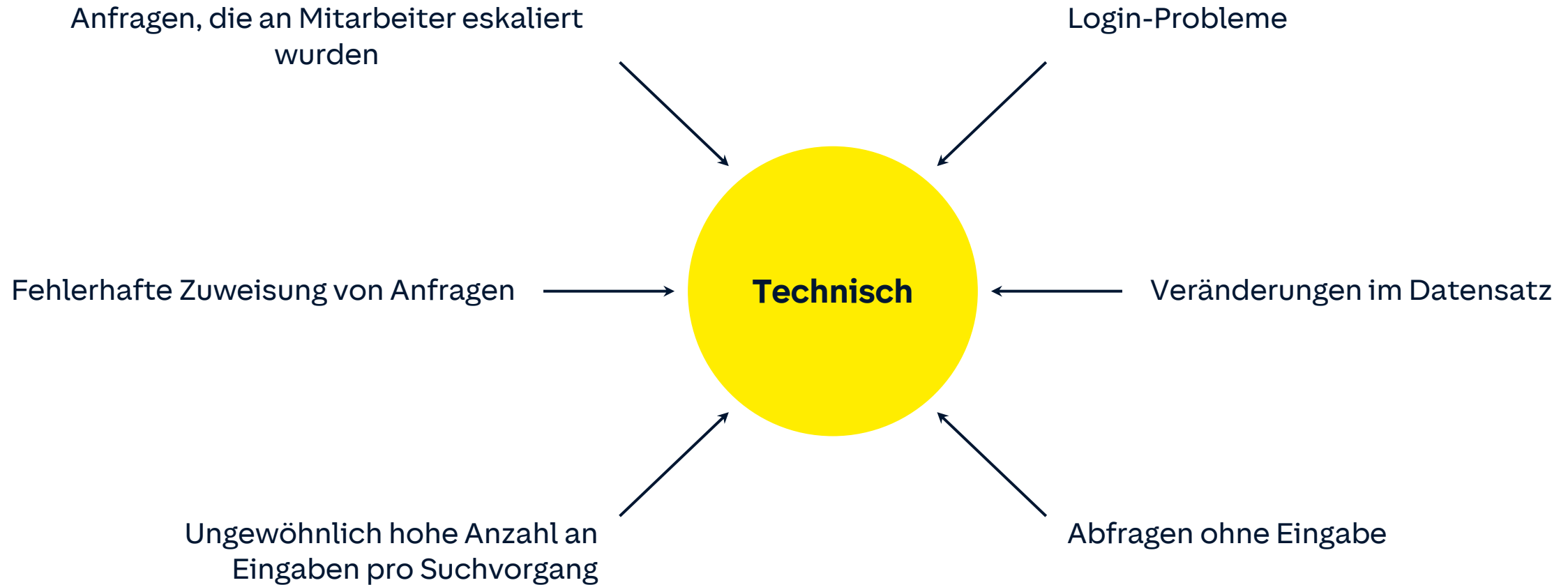
Kernel(c=0.79)

The UNIVARIATE Procedure

Fitted Normal Distribution for LOAN

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	4837.143
Std Dev	Sigma	216.1397

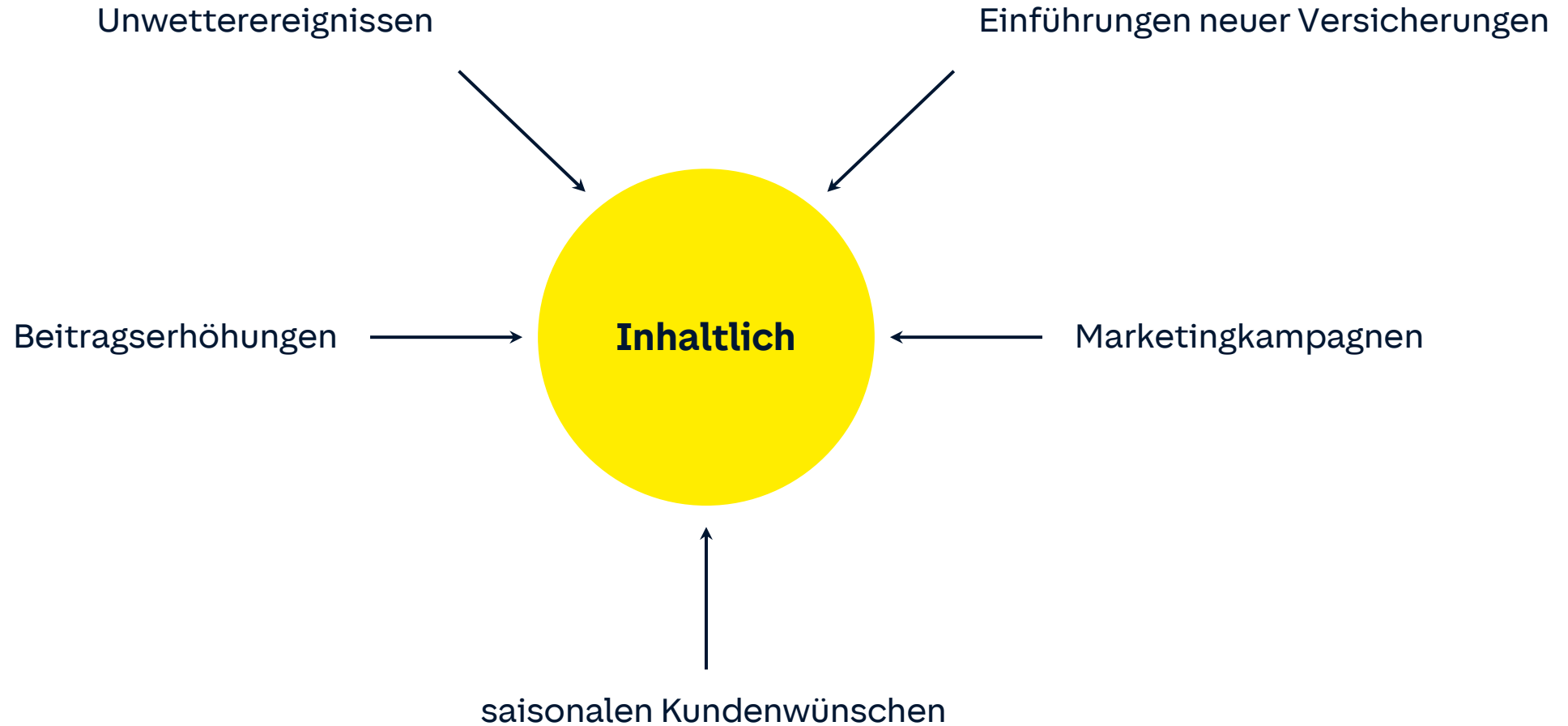
Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic	p Value	
Kolmogorov-Smirnov	D 0.28870527	Pr > D	<0.010
Cramer-von Mises	W-Sn 1.37545302	Pr > W-Sn	<0.005



# Ergebnisse



Erkennung der Auswirkungen von ....



**Vielen  
Dank**

