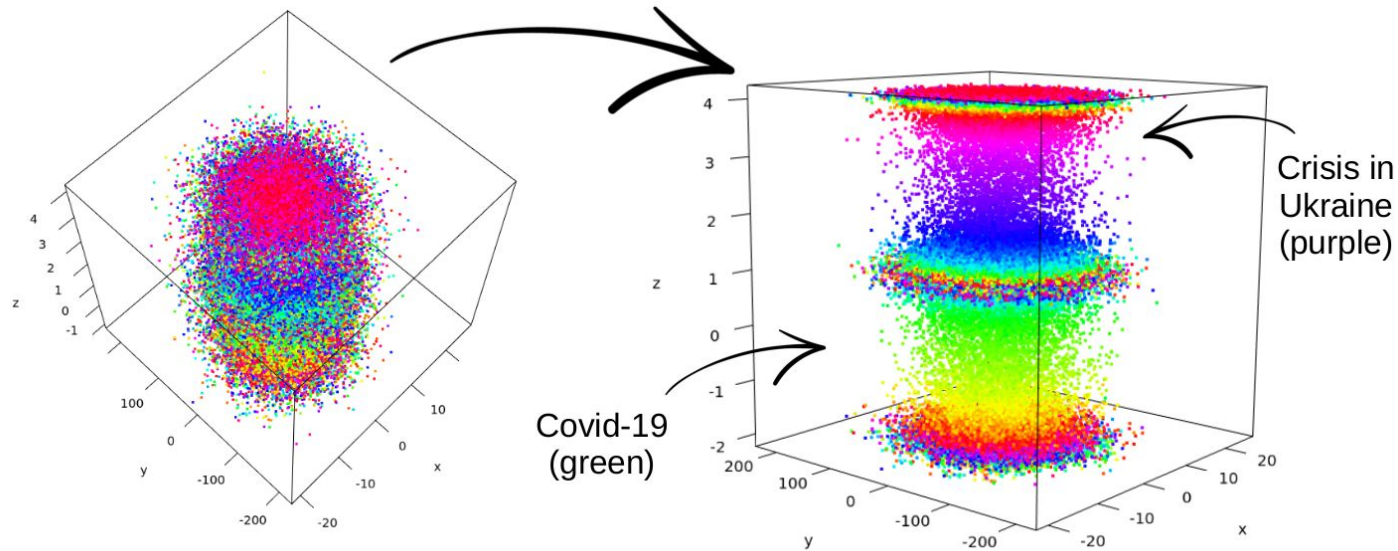


Methodische Grenzen Machine Learning basierter Social-Media-Screenings zur Einschätzung von Bedrohungslagen



MOTRA-K #2022

Prof. Dr. Dennis Klinkhammer

Agenda

- Social-Media-Screenings
- Machine Learning zur Klassifikation
- Zusätzliche Features mit NLP
- Funktionsselektion
- Empfehlungen
- Referenzen



Social-Media-Screenings

Social-Media-Screenings

- Explizite Erwähnung in der Cybersicherheitsagenda des Bundes
- Recherchertools für folgende Anwendungszwecke geeignet:
 - Kindesmissbrauch (BKA)
 - Extremismus (BfV)
 - etc.

Fortentwicklung der Cyberfähigkeiten des BfV und deren Nutzbarmachung im Verfassungsschutzverbund, insbesondere Modernisierung von

- Recherche-Tools zur Aufklärung von Extremismus in sozialen Medien sowie
- Datenhaltungs- und Analysesystemen bzw. -tools in der Aufklärung und Früherkennung staatlich gesteuerter Cyberangriffe

Social-Media-Screenings

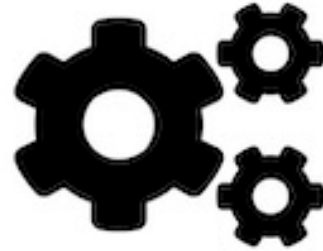
- Erfassung von Mustern im Prozess der (online) Radikalisierung
- Theoretischer Erkenntnisgewinn für Forschung und Praxis
- Identifikation von Hate-Speech, Fake-News und relevanten Akteuren

Schlüsseltechnologie:
Machine Learning (ML)



Social-Media-Screenings

- Systematische Literaturrecherche
 - Berücksichtigung von 64 themenspezifischen Studien
 - Fokus auf Studien im Zeitraum von 2015 bis 2020
- (Einfaches) Machine Learning kommt häufiger zur Anwendung als klassische Verfahren aus der Statistik und Deep Learning Verfahren
 - Logistische Regression (LR)
 - Support Vector Machine (SVM)
 - Random Forest (RF)



Machine Learning zur Klassifikation

Machine Learning zur Klassifikation

- Grundzüge eines regressionsbasierten Klassifikationsverfahrens:

$$Y \sim X_1 + X_2 + X_3 + \dots + X_n$$

- Output (Y) und manuelle Klassifikationsvorgabe
- Input (X_i) als zu erlernende Eigenschaften (Features)

Analysing Social Media Network Data with R
Semi-Automated Screening of Users, Comments and Communication Patterns

a Working Paper and Tutorial by Dennis Klinkhammer

Abstract

Communication on social media platforms is not only culturally and politically relevant, it is also increasingly widespread across societies. Users not only communicate via social media platforms, but also search specifically for information, disseminate it or post information themselves. However, fake news, hate speech and even radicalizing elements are part of this modern form of communication: Sometimes with far-reaching effects on individuals and societies. A basic understanding of these mechanisms and communication patterns could help to counteract negative forms of communication, e.g. bullying among children or extreme political points of view. To this end, a method will be presented in order to break down the underlying communication patterns, to trace individual users and to inspect their comments and range on social media platforms; Or to contrast them later on via qualitative research. This approach can identify particularly active users with an accuracy of 100 percent, if the framing social networks as well as the topics are taken into account. However, methodological as well as counteracting approaches must be even more dynamic and flexible to ensure sensitivity and specificity regarding users who spread hate speech, fake news and radicalizing elements.

Machine Learning zur Klassifikation

- Dichotomisierung:
 - 0, wenn \leq Mittelwert und 1, wenn $>$ Mittelwert
 - Zweck: Varianzhomogenität der Features
- Beispiel für $Y = 1$ und X mit $(0, 1, 1)$ aufgrund erlernter Eigenschaften:

$$\mathbf{1} \sim 0.00 * \mathbf{X}_1 + 0.75 * \mathbf{X}_2 + 0.25 * \mathbf{X}_3$$

Machine Learning zur Klassifikation

- Dichotomisierung:
 - 0, wenn \leq Mittelwert und 1, wenn $>$ Mittelwert
 - Zweck: Varianzhomogenität der Features
- Beispiel für $Y = 1$ und X mit $(0, 1, 1)$ aufgrund erlernter Eigenschaften:

$$\mathbf{1} = 0.00 * \underline{\mathbf{0}} + 0.75 * \underline{\mathbf{1}} + 0.25 * \underline{\mathbf{1}}$$

Machine Learning zur Klassifikation

- Dichotomisierung:
 - 0, wenn \leq Mittelwert und 1, wenn $>$ Mittelwert
 - Zweck: Varianzhomogenität der Features
- Beispiel für $Y = 1$ und X mit (0, 1, 1) aufgrund erlernter Eigenschaften:



$$1 = 0.00 * \underline{0} + 0.75 * \underline{1} + 0.25 * \underline{1}$$

Mittelfristig keine stabilen Befunde!

Machine Learning zur Klassifikation

- Dichotomisierung:
 - 0, wenn \leq Mittelwert und 1, wenn $>$ Mittelwert
 - Zweck: Varianzhomogenität der Features
- Dichotomisierung über Normwerte:
 - Differenz zum Mittelwert in Relation zur Standardabweichung
 - z-Transformation mit 0, wenn $z \leq 1$ und 1, wenn $z > 1$



$$z = \frac{x - \bar{x}}{s_x}$$

Machine Learning zur Klassifikation

- Dichotomisierung:
 - 0, wenn \leq Mittelwert und 1, wenn $>$ Mittelwert
 - Zweck: Varianzhomogenität der Features
- Dichotomisierung über Normwerte:
 - Differenz zum Mittelwert in Relation zur Standardabweichung
 - z-Transformation mit 0, wenn $z \leq 1$ und 1, wenn $z > 1$



$$z = \frac{x - \bar{x}}{s_x}$$

Mittelfristig stabilere Befunde



Zusätzliche Features mit NLP

Zusätzliche Features mit NLP

- Neben quantitativen Features können auch qualitative Features im Machine Learning eingesetzt werden
- Natural Language Processing (NLP)
- Algorithmenbasierte Verarbeitung von Text- und Sprachdaten

Wörter -> Vektoren

Sentiment Analysis with R

Natural Language Processing for Semi-Automated Assessments of Qualitative Data

a Working Paper and Tutorial by Dennis Klinkhammer

Abstract

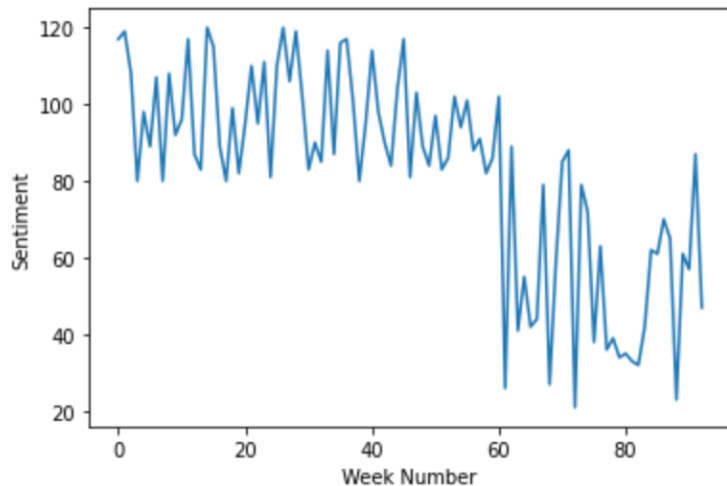
Sentiment analysis is a sub-discipline in the field of natural language processing and computational linguistics and can be used for automated or semi-automated analyses of text documents. One of the aims of these analyses is to recognize an expressed attitude as positive or negative as it can be contained in comments on social media platforms or political documents and speeches as well as fictional and nonfictional texts. Regarding analyses of comments on social media platforms, this is an extension of the previous tutorial on semi-automated screenings of social media network data. A longitudinal perspective regarding social media comments as well as cross-sectional perspectives regarding fictional and nonfictional texts, e.g. entire books and libraries, can lead to extensive text documents. Their analyses can be simplified and accelerated by using sentiment analysis with acceptable inter-rater reliability. Therefore, this tutorial introduces the basic functions for performing a sentiment analysis with R and explains how text documents can be analysed step by step - regardless of their underlying formatting. All prerequisites and steps are described in detail and associated codes are available on GitHub. A comparison of two political speeches illustrates a possible use case.

Zusätzliche Features mit NLP

- Klassische NLP Anwendungen:
 - Summarization -> Worum geht es in dem Text?
 - Classification -> Was für ein Text liegt vor?
 - Sentiment -> Welche Einstellungen beinhaltet der Text?
 - Generation -> Wie lässt sich ein Text sinnvoll erweitern?

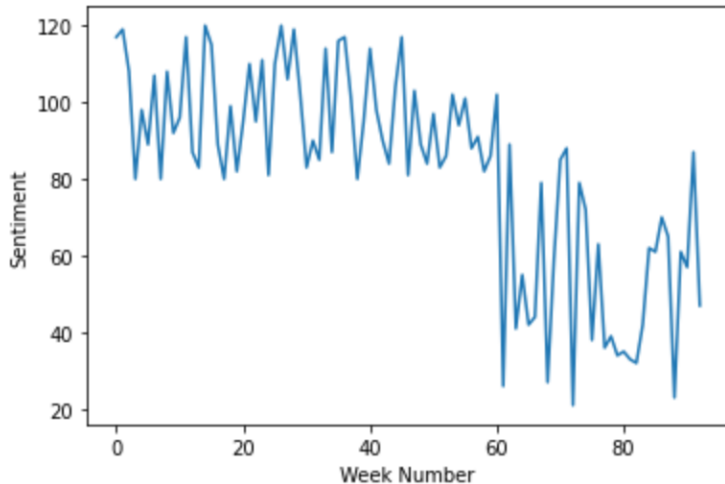
Zusätzliche Features mit NLP

- Oftmals werden Summarization, Classification und Sentiment in der Radikalisierungsforschung zur Anwendung gebracht
- Aber: Text- und Sprachdaten unterliegen einem zeitlichen Kontext



Zusätzliche Features mit NLP

- Oftmals werden Summarization, Classification und Sentiment in der Radikalisierungsforschung zur Anwendung gebracht
- Aber: Text- und Sprachdaten unterliegen einem zeitlichen Kontext

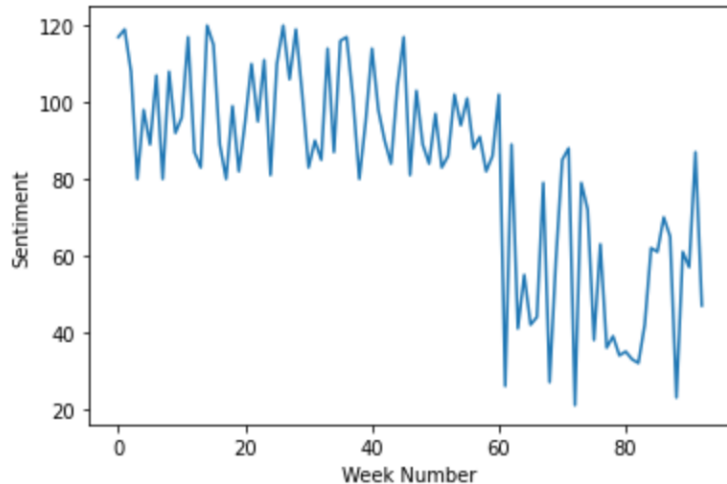


- Dadurch führen bspw. Mittelwerte der Sentiments bereits kurzfristig zu weniger stabilen Befunden



Zusätzliche Features mit NLP

- Oftmals werden Summarization, Classification und Sentiment in der Radikalisierungsforschung zur Anwendung gebracht
- Aber: Text- und Sprachdaten unterliegen einem zeitlichen Kontext



- Dadurch führen bspw. Mittelwerte der Sentiments bereits kurzfristig zu weniger stabilen Befunden
- Konditionale Mittelwerte führen hingegen zu stabileren Befunden





Zusätzliche Features mit NLP

- Viele Studien im Kontext der Radikalisierungsforschung weisen einen singulären Einsatz von NLP auf, bspw. Sentiments



Zusätzliche Features mit NLP

- Viele Studien im Kontext der Radikalisierungsforschung weisen einen singulären Einsatz von NLP auf, bspw. Sentiments 
- Dabei ermöglichen die Vektoren eine zielgerichtete Erweiterung des Klassifikationsverfahrens 

Vektoren als zusätzliche Features!

Zusätzliche Features mit NLP

- One Hot Encoding und Word Embedding ermöglichen Vektorisierung
- Beispiel für One Hot Encoding:
 - Marxism = [1, 0, 0, 0, 0, 0]
 - Fascism = [0, 1, 0, 0, 0, 0]
- Beispiel für Word Embedding:
 - Marxism (1) is (2) good (3) -> [1, 0, 0, 0, 1, 1]
 - Position + Häufigkeit (n)
 - [0, 0, 1, 0, 2, 1]
 - [0, 0, 0, 1, 3, 1]

Zusätzliche Features mit NLP

- Word Embedding berücksichtigt die Positionen und Häufigkeiten der Wörter in Text- und Sprachdaten
- Die Vektoren lassen sich zusätzlich um Sentiments und in Verbindung stehende Wörter erweitern

Zusätzliche Features mit NLP

- Word Embedding berücksichtigt die Positionen und Häufigkeiten der Wörter in Text- und Sprachdaten
- Die Vektoren lassen sich zusätzlich um Sentiments und in Verbindung stehende Wörter erweitern

Standardfeatures



Zusätzliche Features mit NLP

- Word Embedding berücksichtigt die Positionen und Häufigkeiten der Wörter in Text- und Sprachdaten
- Die Vektoren lassen sich zusätzlich um Sentiments und in Verbindung stehende Wörter erweitern

Standardfeatures

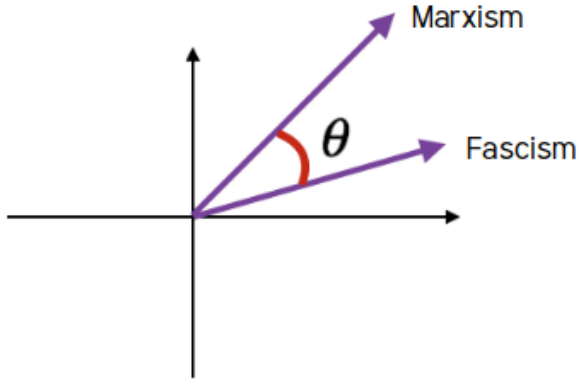


Bedarfsspezifische Features

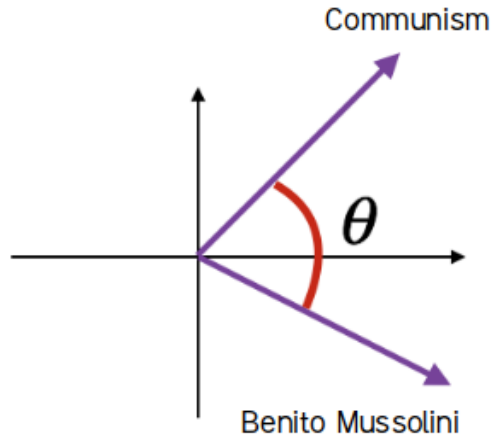


Zusätzliche Features mit NLP

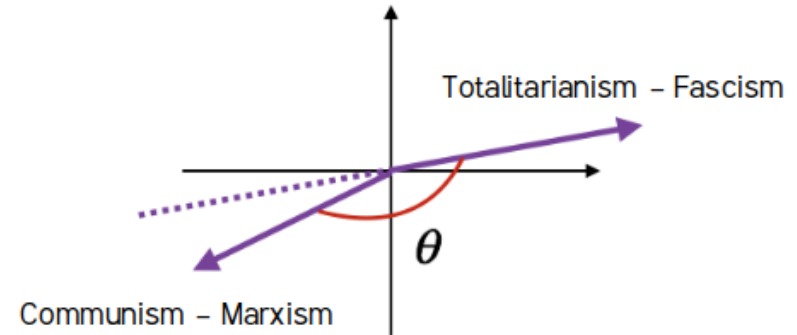
- Vektorbasierte Funktionsweise als bedarfsspezifisches Feature



θ is close to 0°
 $\cos(\theta) \approx 1$



θ is close to 90°
 $\cos(\theta) \approx 0$



θ is close to 180°
 $\cos(\theta) \approx -1$



Funktionsselektion

Funktionsselektion

- Gegeben seien folgende Anwendungsbeispiele:

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Funktionsselektion

- Gegeben seien folgende Anwendungsbeispiele:

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Funktionsselektion

- Gegeben seien folgende Anwendungsbeispiele:

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

Funktionsselektion

- Gegeben seien folgende Anwendungsbeispiele:

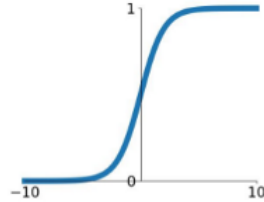
$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \quad ?$$

Funktionsselektion

- Machine Learning und Deep Learning Funktionen – eine Auswahl:

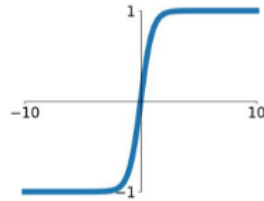
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



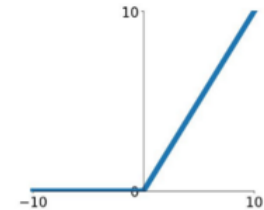
tanh

$$\tanh(x)$$



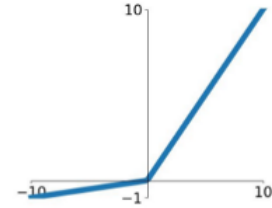
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

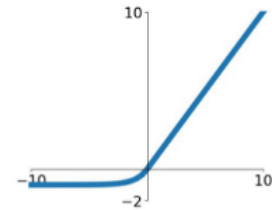


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Funktionsselektion

- Die Funktionen “lernen“ zugrundeliegende Muster unterschiedlich

- Sigmoid

```
binary_accuracy: 50.00%  
1/1 [=====] - 0s 31ms/step  
[[1.]  
 [1.]  
 [1.]  
 [1.]]
```

- tanh

```
binary_accuracy: 100.00%  
1/1 [=====] - 0s 49ms/step  
[[0.]  
 [1.]  
 [1.]  
 [0.]]
```

- ReLU

```
binary_accuracy: 50.00%  
1/1 [=====] - 0s 49ms/step  
[[0.]  
 [0.]  
 [0.]  
 [0.]]
```

Funktionsselektion

- Die Funktionen “lernen“ zugrundeliegende Muster unterschiedlich

- Sigmoid

```
binary_accuracy: 50.00%  
1/1 [=====] - 0s 31ms/step  
[[1.]  
 [1.]  
 [1.]  
 [1.]]
```



- tanh

```
binary_accuracy: 100.00%  
1/1 [=====] - 0s 49ms/step  
[[0.]  
 [1.]  
 [1.]  
 [0.]]
```

- ReLU

```
binary_accuracy: 50.00%  
1/1 [=====] - 0s 49ms/step  
[[0.]  
 [0.]  
 [0.]  
 [0.]]
```

Funktionsselektion

- Die Funktionen “lernen“ zugrundeliegende Muster unterschiedlich

- Sigmoid

```
binary_accuracy: 50.00%  
1/1 [=====] - 0s 31ms/step  
[[1.]  
 [1.]  
 [1.]  
 [1.]]
```



- tanh

```
binary_accuracy: 100.00%  
1/1 [=====] - 0s 49ms/step  
[[0.]  
 [1.]  
 [1.]  
 [0.]]
```



- ReLU

```
binary_accuracy: 50.00%  
1/1 [=====] - 0s 49ms/step  
[[0.]  
 [0.]  
 [0.]  
 [0.]]
```

Funktionsselektion

- Die Funktionen “lernen“ zugrundeliegende Muster unterschiedlich

- Sigmoid

```
binary_accuracy: 50.00%  
1/1 [=====] - 0s 31ms/step  
[[1.]  
 [1.]  
 [1.]  
 [1.]]
```



- tanh

```
binary_accuracy: 100.00%  
1/1 [=====] - 0s 49ms/step  
[[0.]  
 [1.]  
 [1.]  
 [0.]]
```



- ReLU

```
binary_accuracy: 50.00%  
1/1 [=====] - 0s 49ms/step  
[[0.]  
 [0.]  
 [0.]  
 [0.]]
```



Funktionsselektion

Warum verwenden 75% der Machine Learning bzw. Deep Learning basierten Social-Media-Screenings eine Sigmoid-Funktion?



Empfehlungen

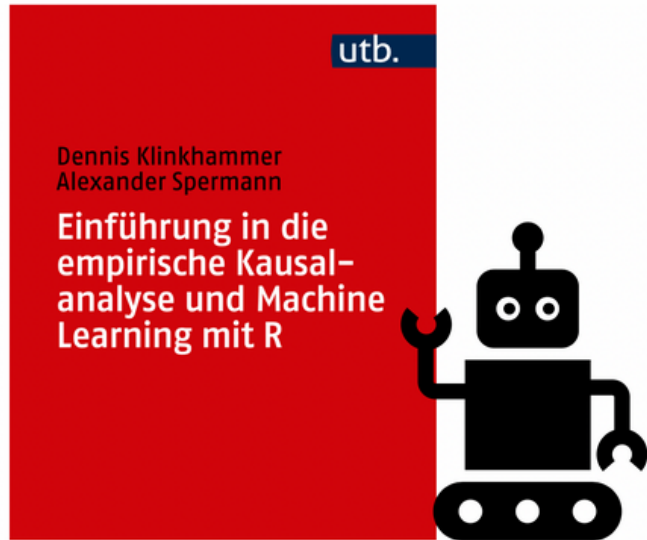
Empfehlungen

- Einhaltung der wissenschaftlichen Gütekriterien
 - Objektivität
 - Reliabilität
 - Validität
- Systematische Feature-Generierung
 - Überprüfung auf zusätzliche Informationen im Datenmaterial
 - Gleichzeitig gilt aber auch: KISS – Keep It Simple Stupid

Empfehlungen

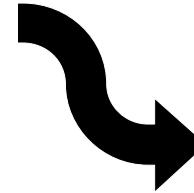
- (Geographische) Kontextgebundenheit
 - Transfer Learning funktioniert in der Regel nicht
 - Ähnliche Ideologien unterscheiden sich in ihrer Phänomenologie
- Methodische Grundlagen, insb. Mathematik und Statistik
 - Keine reine Anwendung von Tools und Algorithmen
 - Alternative: Spezifische Modellierung
 - Kontinuierliche Weiterbildung (!)

Vielen Dank für die Aufmerksamkeit



Statistical Thinking

Mehr Beispiele
zur Statistik mit
Python und R



Referenzen

- (1) Bundesministerium des Innern und für Heimat (2022): *Cybersicherheitsagenda des Bundesministeriums des Innern und für Heimat. Ziele und Maßnahmen der 20. Legislaturperiode*. Artikelnummer: BMI22011. Berlin.
- (2) Dragos, V., Kervarc, R., Bruyant, J.-P., & C. Moïse (2018): *Semantic approaches to analyse radicalised content*. Partnership against radicalization in cities network.
- (3) European Crime Prevention Network (2019): *European Crime Prevention Monitor. Radicalisation and violent extremism*. Brussels: European Crime Prevention Network.
- (4) Gaikwad, M., Ahirrao, S., Phansalkar, S. & K. Kotcha (2021): *Online Extremism Detection: A Systematic Literature Review with Emphasis on Datasets, Classification Techniques, Validation Methods and Tools*. IEEE Access.
- (5) Grogan, M. (2020): *NLP from a time series perspective. How time series analysis can complement NLP*. Towards Data Science.
- (6) Hamachers, A., Weber, K., Widmann, J. & S. Jarolimek (2020): *Extremistische Dynamiken im Social Web*. Frankfurt am Main: Verlag für Polizeiwissenschaft.

Referenzen

- (7) Iacus, S. M. & G. Porro (2022): *Using social networks to measure subjective well-being*. In: Significance. Volume 19. Issue 3.
- (8) Klinkhammer, D. (2020): *Analysing Social Media Network Data with R: Semi-Automated Screening of Users, Comments and Communication Patterns*. Cornell University (arXiv).
- (9) Klinkhammer, D. (2022): *Sentiment Analysis with R: Natural Language Processing for Semi-Automated Assessments of Qualitative Data*. Cornell University (arXiv).
- (10) Niederer, S. (2018): *Networked images: Visual methodologies for the digital age*. Inaugural lecture.
- (11) Ng, A. (2022): Deep Learning Specialization. DeepLearning.AI, Palo Alto, CA (USA).
- (12) Pereira-Kohatsu, J. & Quijano-Sanchez, L. & Liberatore, F. & M. Camacho-Collados (2019): *Detecting and Monitoring Hate Speech in Twitter*. Sensors (19).
- (13) Wienigk, R. & D. Klinkhammer (2021): *Online Aktivitäten der Identitären Bewegung auf Twitter. Warum Kontensperrungen die Anzahl an Hassnachrichten nicht reduzieren*. In: Forum Kriminalprävention (2/2021).