



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Nina Arndt, Melanie Budde, Ariana Sliwa

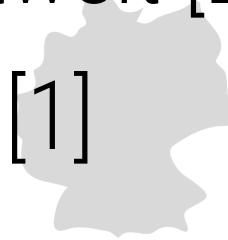
Analyse von News zur Suizidprävention

Abschlusspräsentation 19.03.2018



800.000 Suizide pro Jahr weltweit [2]

10.000 davon in Deutschland [1]



90% der Menschen, die durch Suizid versterben, litten an einer psychiatrischen Erkrankung, am häufigsten an einer Depression (> 50 %). [3]

Welchen Einfluss hat Medienberichterstattung?

Werther-Effekt:

Medienberichterstattung über einen Suizid kann weitere Suizide verursachen, die ohne nicht geschehen wären. [4]

Papageno-Effekt:

Berichte darüber, wie Menschen eine Krisensituation konstruktiv und ohne suizidales Verhalten bewältigen, haben zur Senkung der Suizidraten in der Zeit danach gesorgt. [5]

Idee im Rahmen des Big Data Projekts

- tägliche Analyse der aktuellen Medienberichterstattung auf das Stichwort "Suizid"
- Prüfung auf angebotene Hilfe in der Berichterstattung
- Sentimentanalyse der gesamten Medienberichterstattung des Tages
- Ausgabe als Excelliste mit Ampelwarnsystem

1 Woher die Daten kommen

- Recherche der RSS-Feeds der relevanten Medien
- Aufsetzen des Crawlers durch Max

Lessons Learned:

- Paywalls mögen uns nicht
- Ohne Fließtext bleibt wenigstens die “Description”

2 Aus Quellcode wird Fließtext: Preprocessing

```
filters = {
  "not_impl_filter": {},
  "newsfeed.zeit.de": {
    "standard_filter": {
      "entrypoints": [{"tag": "article"}],
      "delete": [
        {"tag": "script"},
        {"tag": "style"},
        {"tag": "figure"},
        {"tag": "input"},
        {"tag": "div", "attrs": [
          {"attr": "class", "values": ["metadata", "article-heading", "article-footer", "sharing-menu"]}
        ]}
      ]
    }
  }
},
```

[...]

3 Texte analysieren: Textmining

Pakete und Bibliotheken:

- nltk (z.B. tokenize, stopwords)
- pandas (Dataframes)
- matplotlib (Diagramme)

3 Suche nach relevanten Stichworten

```

#Search for keywords
def werthereffect(text):
    #returns werther=1 for the submitted text if the texts contains one or more of the given keywords
    #returns werther=0 otherwise
    keywords = ['selbstmord', 'suizid', 'leben genommen', 'sich umgebracht', 'freitod', 'selbsttoetung',
    werther = 0
    lowertext = text.lower()
    for k in keywords:
        if k in lowertext:
            werther = 1
    return werther

def papagenoeffect(text):
    #number of "Telefon-Seelsorge"
    keywords = ['0800-1110111', '0800-1110222', '0800/1110111', '0800/1110222', '0800/ 1110111'
    papageno = 0
    lowertext = text.lower()
    for k in keywords:
        if k in lowertext:
            papageno = 1
    return papageno

```


3 Sentimentanalyse mit dictionary

```
def count_sentiment(text):
    sentiment = 0
    raw_tokens = word_tokenize(text)
    tokens = [w for w in raw_tokens \
              if not w in customstopwords]
    for t in tokens:
        if t in filtered_neg_tokens:
            sentiment -= 1
        if t in filtered_pos_tokens:
            sentiment += 1
    if (len(tokens)>0):
        sentiment = normalize((sentiment/len(tokens))*100)
    return sentiment
```

3 Sentimentanalyse mit BayesClassifier

```
# create dictionary for Bayes Classifier
def create_word_features(words):
    useful_words = preprocessing(words)
    my_dict = dict([(word, True) for word in useful_words])
    return my_dict

train_set = neg_texts[:7] + pos_texts[:7]
test_set = neg_texts[7:] + pos_texts[7:]

def classification(text):
    classifier = NaiveBayesClassifier.train(train_set)
    words = create_word_features(text)
    return classifier.classify(words)
```

4 Was ist der mögliche Output?

	A	B	C	D	E	F	G	H	I	J
1	Suicide Prevention information based on media published on 2018-3-18-18-35									
2	Sentiment	49,8943								
3	Sentiment	0								
4										
5	At least one article was found that contains the word suicide:									
6	Medium	Title	Link	Sent1	Sent2	Papageno				
7	www.welt	WM 2018	https://w	50,1312	negative	0				
8	www.faz.r	Soros sch	http://ww	50,25	negative	0				
9	www.ndr.	Revision i	https://w	48,8602	negative	0				
10	www.ndr.	Elfjährige	https://w	49,8571	negative	0				
11	rtlnord.de	Schicken	http://rtl	50,3367	negative	0				
12	newsfeed	Fußball-W	http://ww	50	negative	0				
13										
14										
15										

4 Was ist der mögliche Output?

	A	B	C	D	E	F	G	H	I	J
1	<u>Suicide Prevention information based on media published on 2018-3-18-18-39</u>									
2	Sentimen	69,0433								
3	Sentimen	62,5								
4										
5										
6										
7										
8										

5 Visualisierungen

14.03.2018

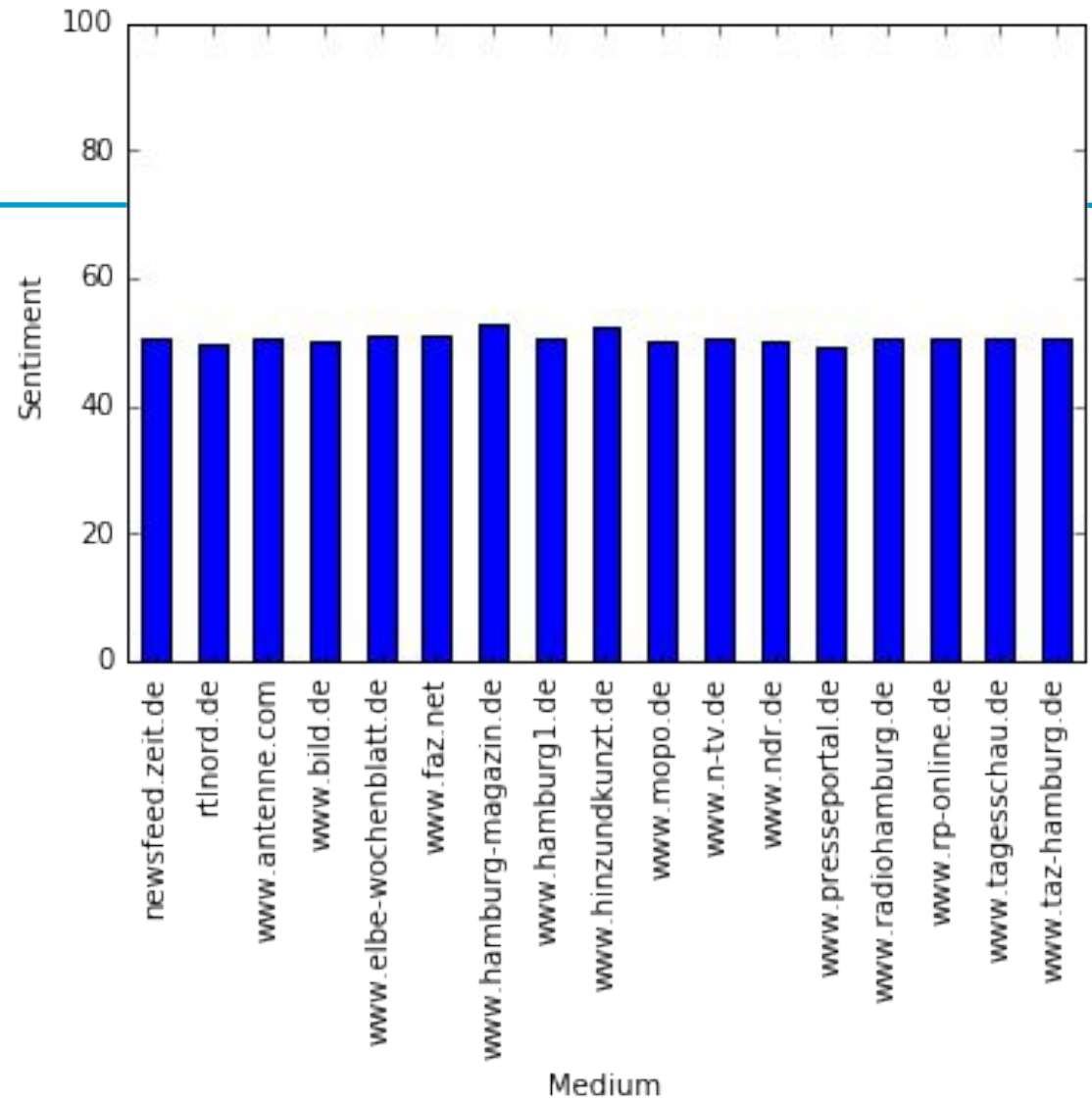
Rohdaten: 1.545 Einträge

Ausgabe:

Durchschnittlicher

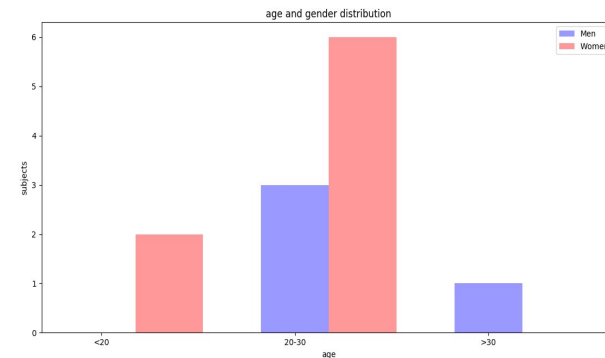
Sentiment nach Methode

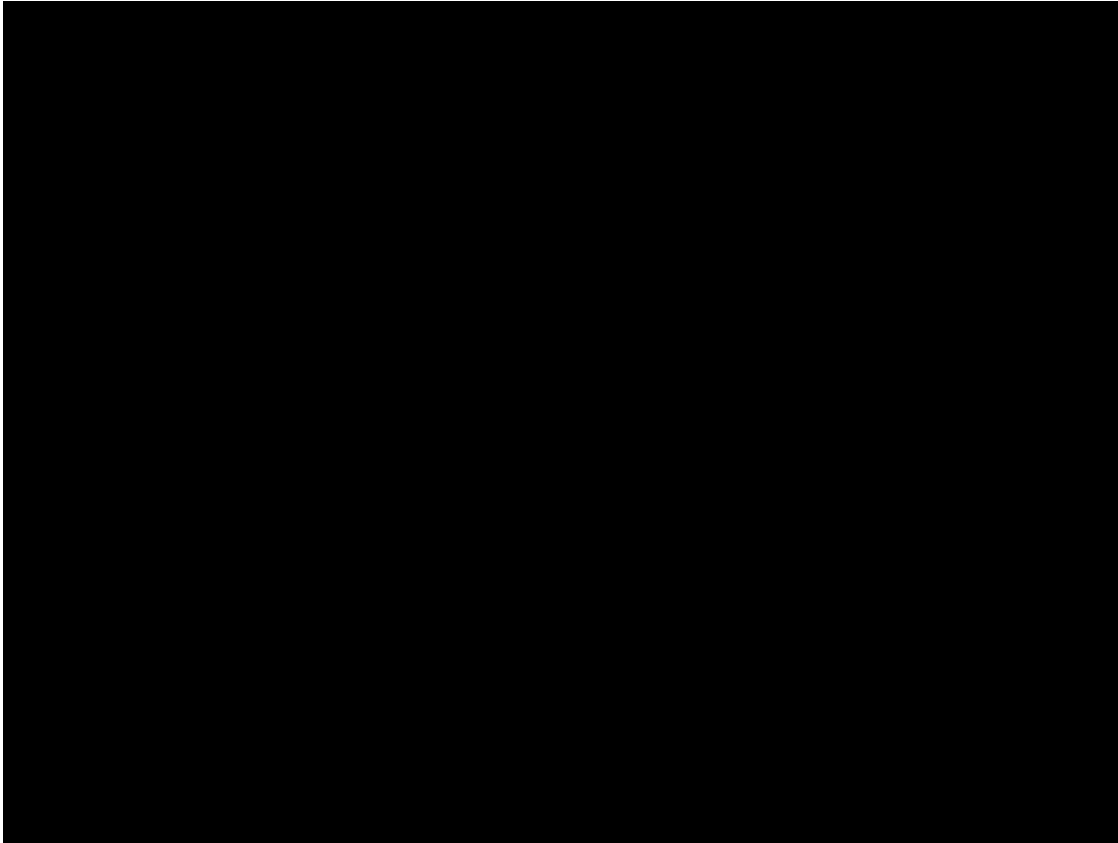
“Dictionary”



UKE // data-exploration

- Von Matlab nach Python
- Correlations-Analyse
- Visualisierung der Koordinaten





UKE // socket-implementation

- Problem: Datentyp 'cluster' von LabView zu Python
- Raw-Bytestream erfordert typecasting
 - fehleranfällig
 - schwer zu debuggen
- Lösungsansatz:
 - Server und Client kommunizieren über ZeroMQ / JSON

Literaturverzeichnis

[1] Statistisches Bundesamt. Todesursachen. URL

<https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/Todesursachen/Todesursachen.html>. Accessed: 14.03.2018

[2] WHO. Suicide Fact sheet, 2018. URL <http://www.who.int/mediacentre/factsheets/fs398/en/>. Accessed: 14.03.2018

[3] José Manoel Bertolote and Alexandra Fleischmann and Diego De Leo and Danuta Wasserman. Psychiatric Diagnoses and Suicide: Revisiting the Evidence. *Crisis*, 25(4):147–155, 2004. doi: 10.1027/0227-5910.25.4.147. URL <https://doi.org/10.1027/0227-5910.25.4.147>. PMID: 15580849.

[4] David P. Phillips. The Influence of Suggestion on Suicide: Substantive and Theoretical Implications of the Werther Effect. 39:340, 06 1974.

[5] Thomas Niederkrotenthaler and Martin Voracek and Arno Herberth and Benedikt Till and Markus Strauss and Elmar Etzersdorfer and Brigitte Eisenwort and Gernot Sonneck. Role of media reports in completed and prevented suicide: Werther v. Papageno effects. *British Journal of Psychiatry*, 197(03):234–243, sep 2010. doi:10.1192/bjp.bp.109.074633. URL <https://doi.org/10.1192/bjp.bp.109.074633>.