

Text Mining & NLP mit Python und Power BI

Prof. Dr. Jens Albrecht, Prof. Dr. Roland Zimmermann

<https://github.com/jsalbr/tdwi-2023-text-mining>





Alle Daten zum mitmachen unter:
<https://github.com/jsalbr/tdwi-2023-text-mining>

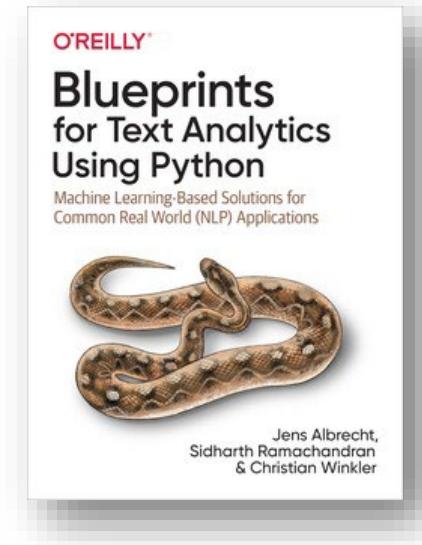
Zugang zu Google Colab:
<https://colab.research.google.com>

Power BI Desktop Download:
<https://powerbi.microsoft.com/de-de/downloads>



Prof. Dr. Jens Albrecht TH Nürnberg, Informatik

Data Warehousing, BI,
Data Science, NLP



Prof. Dr. Roland Zimmermann TH Nürnberg, BWL

BI, Information Design,
NLP, Process Mining



Herausforderungen mit Text

300 Kunden-
service-Anfragen
pro Tag

70.000 Verträge

1.000 Online-
Reviews zum
Produkt

30
Leserkommen-
tare pro Minute

100.000
dokumentierte
Change
Requests

... jede einzelne
muss manuell
kategorisiert
werden.

... ohne
automatisiertes
Ablagesystem.

... müssen für die
Einkaufs-
verhandlung alle
der Reihe nach
gelesen werden.

... in der eigenen
Online Community
und der
Community
Manager entdeckt
die dringendsten
eher per Zufall.

... und keinen
Überblick, wo die
Kostentreiber
stecken.

Text Analytics Methoden

Bereiche

1

Statistik

2

Unüberwachtes
Machine
Learning

3

Überwachtes
Machine
Learning

4

Semantische
Analysen

Methoden und Ziele
(Auszug)

VERTEILUNG

Überblick über Texte,
Datenqualität, Bias

KORRELATION

Zusammenhänge
Hypothesen

TOPIC MODELING

Identifikation von
Themen in Texten

CLUSTERING

Versteckte Muster in
Content-Archiv
analysieren

KLASSIFIKATION

Neue Texte in
bekannte Kategorien
einsortieren

REGRESSION

Trends in Texten
identifizieren

TRANSFORMERS &
LLMS

Semantik, Bedeutung
& Zusammenhänge
extrahieren

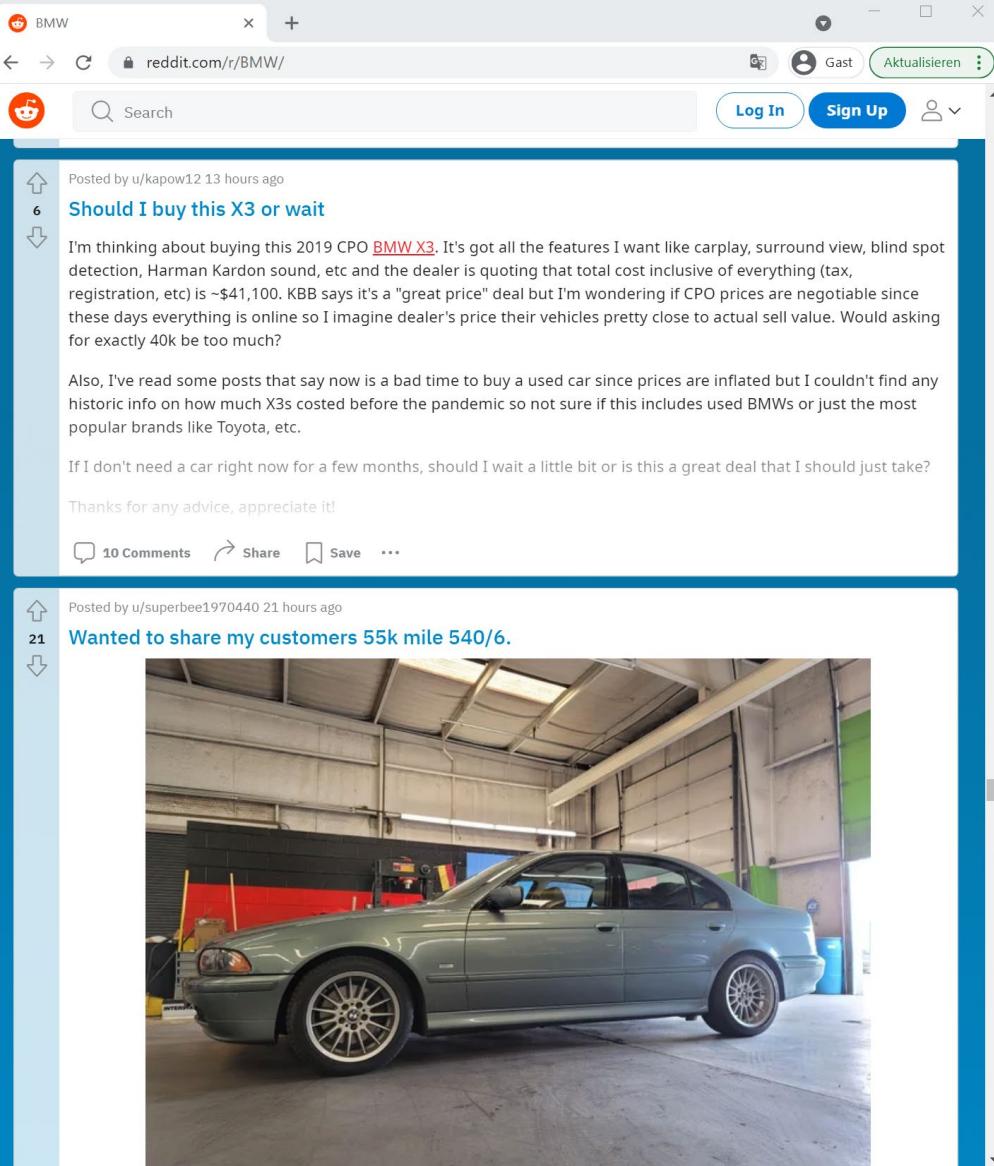
QUESTION
ANSWERING

Antworten
zu Fakten-Fragen
finden

Szenario



Übungsbeispiel: Forumsdaten aus Reddit



Reddit post by u/kapow12 13 hours ago: Should I buy this X3 or wait?

I'm thinking about buying this 2019 CPO [BMW X3](#). It's got all the features I want like carplay, surround view, blind spot detection, Harman Kardon sound, etc and the dealer is quoting that total cost inclusive of everything (tax, registration, etc) is ~\$41,100. KBB says it's a "great price" deal but I'm wondering if CPO prices are negotiable since these days everything is online so I imagine dealer's price their vehicles pretty close to actual sell value. Would asking for exactly 40k be too much?

Also, I've read some posts that say now is a bad time to buy a used car since prices are inflated but I couldn't find any historic info on how much X3s costed before the pandemic so not sure if this includes used BMWs or just the most popular brands like Toyota, etc.

If I don't need a car right now for a few months, should I wait a little bit or is this a great deal that I should just take?

Thanks for any advice, appreciate it!

10 Comments Share Save

Reddit post by u/superbee1970440 21 hours ago: Wanted to share my customers 55k mile 540/6.



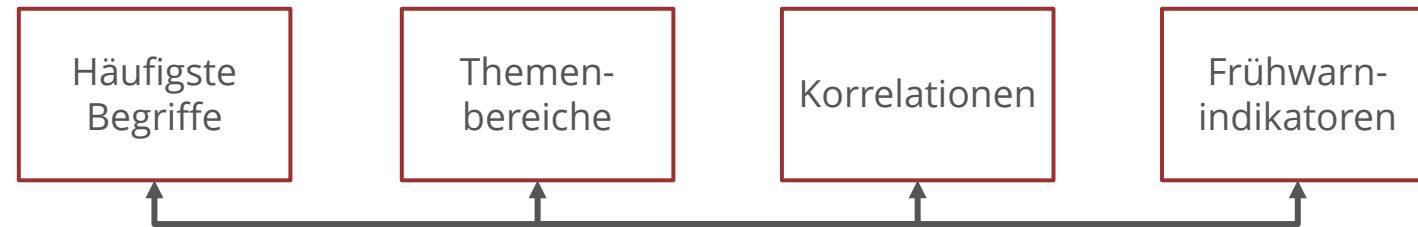
A photograph of a light blue BMW 5 Series sedan parked inside a garage. The car is positioned at an angle, showing its front-left side. The garage has a high ceiling with exposed beams and overhead lights.

Praktisches Anwendungsszenario: Störungsmeldungen in der Auto-Industrie nutzen

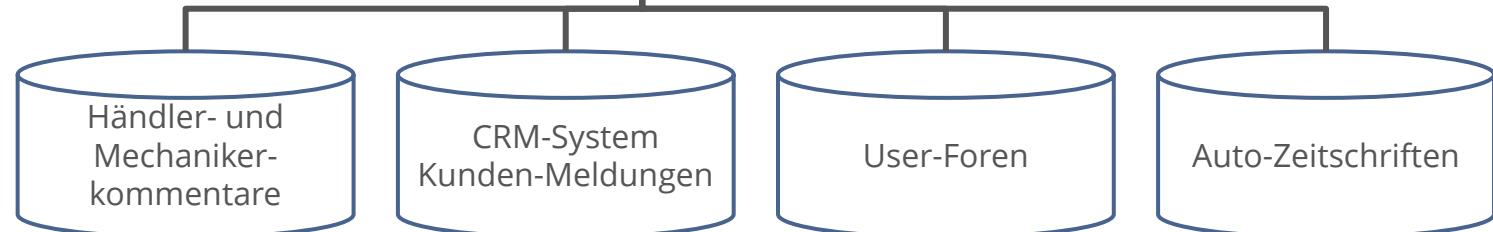
Aktionen



Analysen



Quelldaten



Explorative Datenanalyse mit PowerBI

Power BI als Explorationswerkzeug für Textdaten

- **Datenexploration mit BI-Tool leicht durch Fachanwender**
 - Business und Data Understanding durch Fachabteilung vorbereitbar
 - Arbeitsteilung mit Data-Scientists
- **Texte als Rohdaten sind anders als quantitative Tabellen zu behandeln:
Ziel = erste fachliche Einschätzungen zu Inhalten erzielen**
 - Darstellung von Texten in Tabellenform oft schlecht lesbar:
Thumbnails mit Preview und interaktiver Vergrößerung verbinden
Überblick und Detail für Rohtext-Visualisierungen
 - Rohtexte enthalten erkennbar viele unnütze Wörter:
Sind auch für erste Auszählungen (z.B. Wortwolken) zu ignorieren = Stopwortkonzept relevant
 - Interaktive Filteroptionen vermitteln rasch ersten Überblick über Inhalte der Texte
- Für spätere Phasen des Text-Mining-Prozesses: BI-Tool zur Kommunikation und Bereitstellung der – oftmals komplexen – Text-Mining-Ergebnisse weiternutzen!

Custom Visuals in Power BI für Exploration textueller Daten

Filters

Full Text Search  

Search...  

AND OR

Current Filter

Contains 'hydrogen' o... 

Show items when the value: contains hydrogen 

And Or 

Text Filter  

Search across your dataset right from the dashboard  

★★★★★

Word Cloud 

Create a fun visual from frequent text in your data  

★★★★★

Add

Tag Cloud - xViz 

Tag Cloud helps you get instant insight into the most prominent or prevalent terms in your data  

May require additional purchase  

★★★★★

Card Browser

Browse documents using double-sided cards, and click to view in place  

★★★★★

Strippets Browser

A quick way to view document contents  

★★★★★

Custom Visual von Prof. Albrecht

Spezielle Filter für Texte

Wortwolken

Textuelle Card-Darstellungen

Interdependente, interaktive Anwendung entsteht, die individuelle Sichten auf Rohtexte erlaubt

title  title 

Count of link by title

Hydrogen

Interaction zwischen den Visuals erlaubt die Exploration und auch detaillierte Untersuchung der Texte.

Filtern schränkt Wortwolke und Text ein

Click auf Wort in Wortwolke schränkt Texte und andere Wortwolke ein

Click auf Text schränkt Wortwolke auf einen Text ein

... bitte intensiv ausprobieren!

100MW alkaline electrolyser plant to be developed at Repsol industrial site

100MW green hydrogen electrolyser plant to be developed at Repsol industrial site. The H24All project has presented an application for European funding

Air Products and thyssenkrupp Sign Exclusive Strategic Cooperation Agreement for World's First Electrolysis Plants to Generate Green Hydrogen

LEHIGH VALLEY, PA Air Products (NYSE:APD), a world leader in industrial gases, and thyssenkrupp AG (ETR:THY) announced today that they have signed an exclusive strategic cooperation agreement to develop the world's first electrolysis plants to generate green hydrogen. The German Rostock-based engineering and cleantech company Apeox Energy (APEOX) will be the technology provider for the joint venture.

ARENA: \$70m for hydrogen electrolyser projects

ARENA: \$70m for hydrogen electrolyser projects. The Australian Renewable Energy Agency (ARENA) has today opened a \$70m funding round to help fast track the development of hydrogen electrolyser projects.

Anahí Kauai's™ Electrolysis System Starts World's™ Largest-Scale Hydrogen Supply Contract for the Fukushima Hydrogen Energy Research Field in Name

Anahí Kauai and its subsidiary, Anahí Kauai LLC, announced today that its electrolysis system has started supplying hydrogen to the Fukushima Hydrogen Energy Research Field in Name, Japan.

BioSolar announces green hydrogen electrolyzer technology

NewHydrogen Inc will focus

BioSolar to raise \$5m to accelerate development of electrolyzer technology

BioSolar to raise \$5m to

CFA Validates Elogenix™ Solid Oxide Electrolysis Technology

Europe's leading research

Clean Power Hydrogen delivers IrelandE™s first electrolyzer

Clean Power Hydrogen

Countries Roll Out Green Hydrogen Strategies, Electrolyzer Targets

Countries are increasingly

Visuals und einfaches ETL für Textdaten

Schnelle deskriptive quantitative Analysen

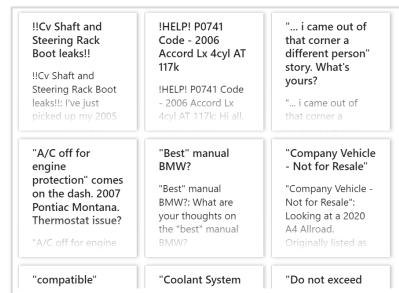
- Balkendiagramme für Mengen
 - Liniendiagramme für zeitliche Entwicklungen
 - Boxplots oder Violin-Plots für Verteilungen



Word Clouds und einfache Stopwörter

Textfilter

Thumbnails für Text-Details



Hier einen Filterbegriff eingeben (für Gesamttexte):

🔍📝

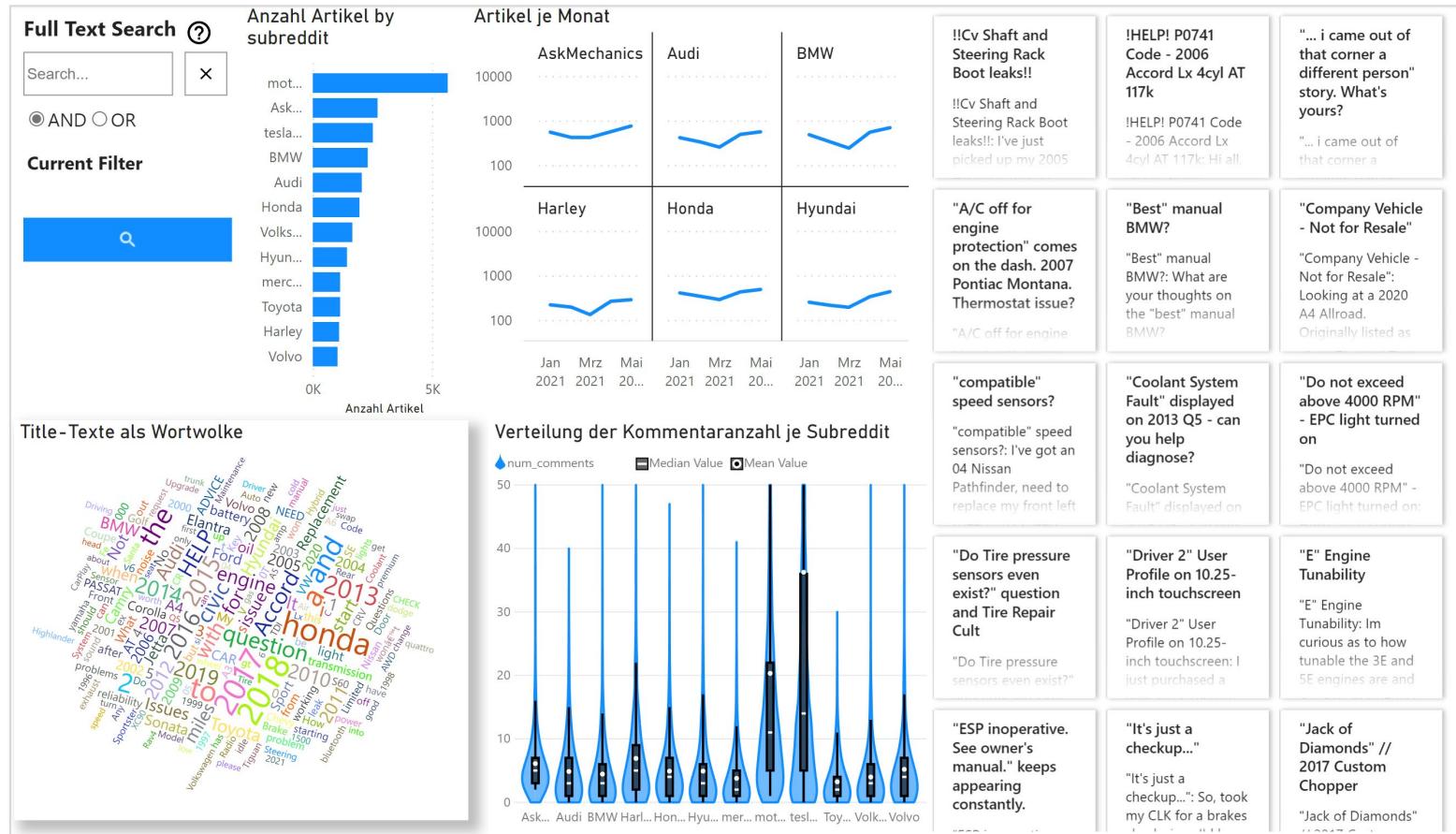
Full Text Search 



AND OR

Live-Demo: Übungs- und Lösungsdateien auf Github

<https://github.com/jsalbr/tdwi-2023-text-mining>



Datenaufbereitung

Textaufbereitung

```
text = "The BMW X5 and the Mercedes GLK are interesting cars. Does VW have models like these?"
```



Tokenization

The|BMW|X5|and|the|Mercedes|GLK|are|interesting|cars|.|Does|VW|have|models|like|these|?|



Part-of-Speech Tagging
& Lemmatization

```
['BMW', 'x5', 'Mercedes', 'GLK', 'car', 'VW', 'model']
```



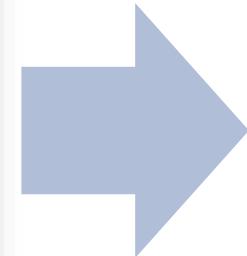
Named Entity Recognition

The **BMW ORG** **X5 PRODUCT** and the Mercedes GLK are interesting cars. Does **VW ORG** have models like these?

Klassifikation mit Python

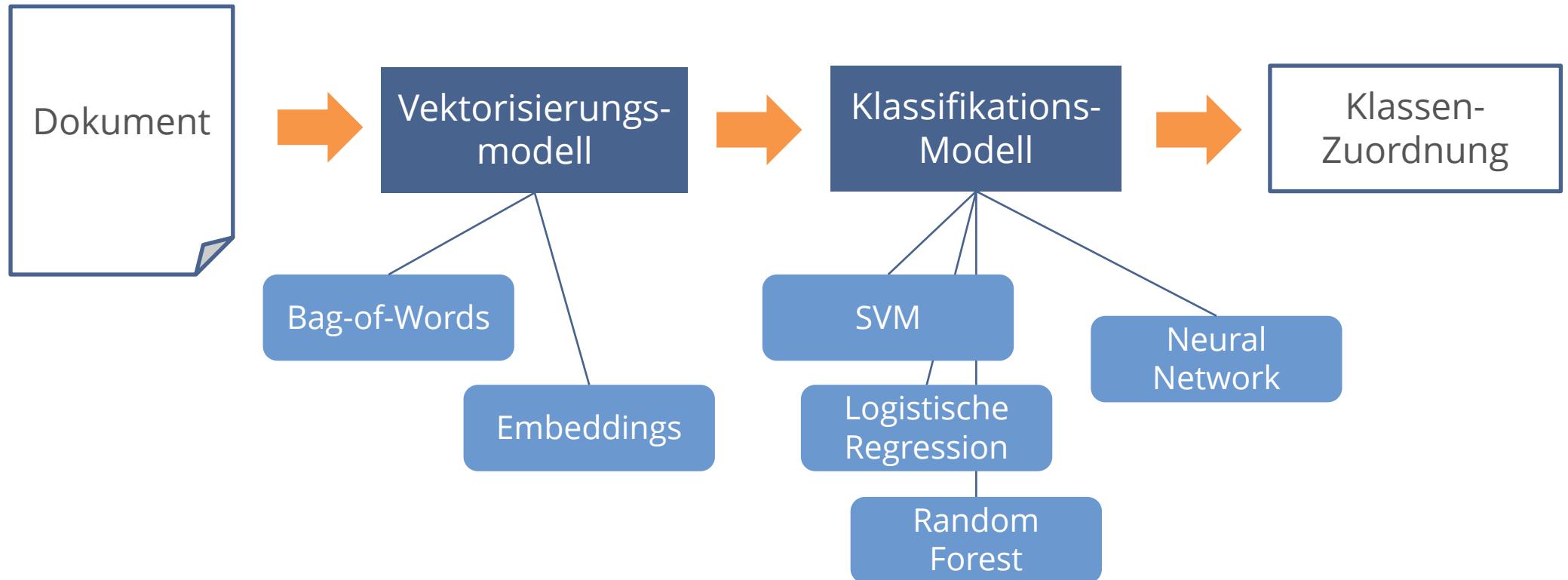
Klassifikationsaufgabe

	subreddit	title	text	lemmas	nouns
16571	Audi	Audi throwback 2013 A4 spy shots	Audi throwback 2013 A4 spy shots: Spy shots or 2013 A4 B8.5 facelift back in summer 2011 before reveal The 2012 Audi A4 is currently on sale in the U.S., so the freshened car is expected to make its debut late this year or early next before going on sale as a 2013 model. An all-new, lighter-weight B9 Audi A4 is due around 2015, so this new look will hold the A4 over until then. _URL_	audi throwback 2013 a4 spy shot spy shot 2013 a4 b8.5 facelift summer 2011 reveal 2012 audi a4 sale u.s. freshened car expect make debut late this year go sale 2013 model new light weight b9 audi a4 due new look hold a4 url _	audi throwback a4 spy shot spy shot a4 b8.5 summer 2011 audi a4 sale u.s. car debut late this year sale model weight b9 audi a4 look a4 url _
18064	motorcycles	Riding help	Riding help: Been riding for 10 years, I guess age caught me. I have a k6 750 love her since day one. Problem is my knees hurt when I ride, hell going to the store my knees hurt. Any suggestions?	riding help ride 10 year guess age catch have k6 750 love day one problem knee hurt ride hell go store knee hurt suggestion	riding help 10 year age k6 love day one problem knee hell store knee suggestion

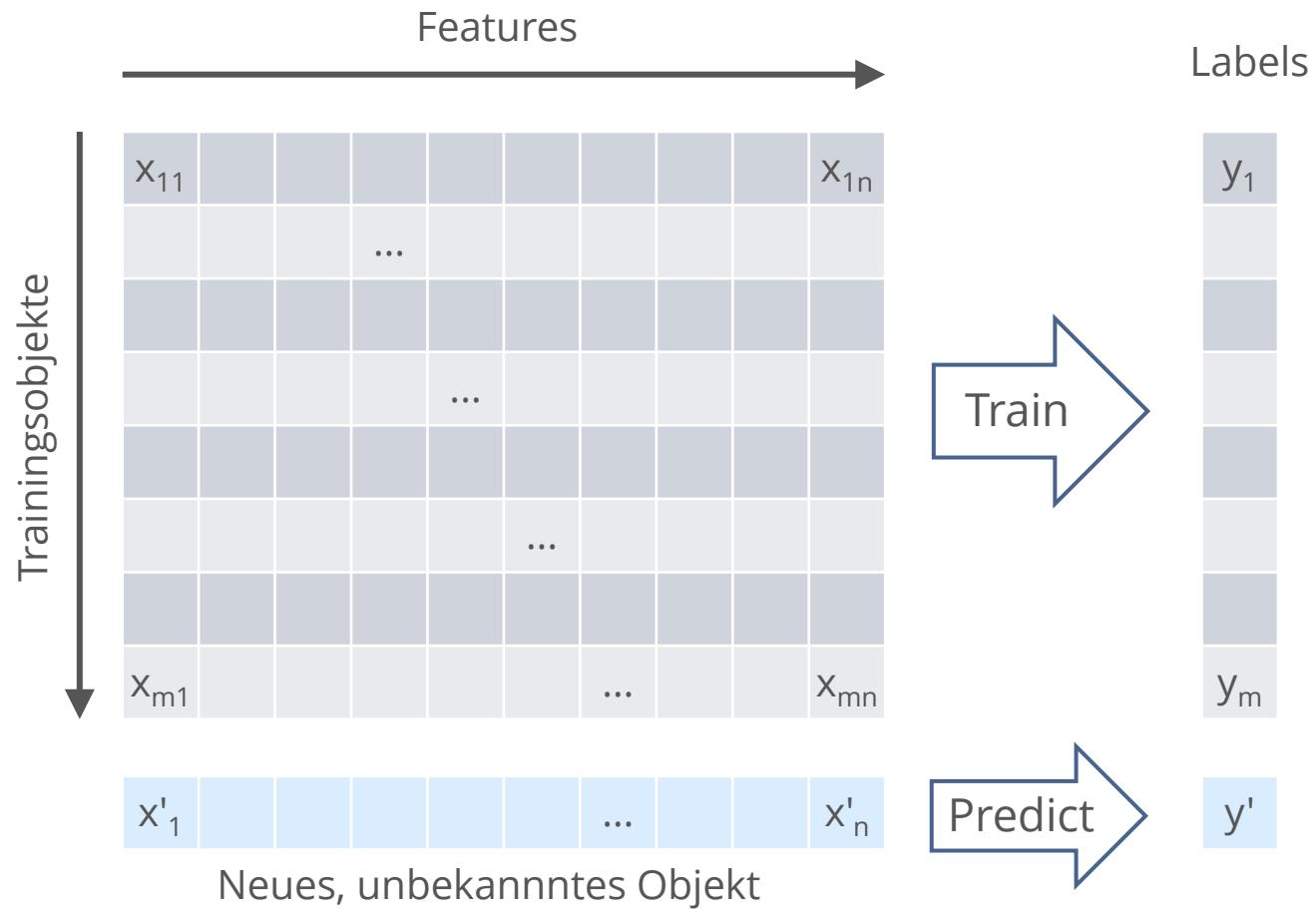


subreddit	
motorcycles	5654
AskMechanics	2713
teslamotors	2515
BMW	2303
Audi	2053
Honda	1957
Volkswagen	1657
Hyundai	1433
Toyota	1145
mercedes_benz	1145
Harley	1101
Volvo	1036

Text-Klassifikation



Supervised Learning: Trainings-Matrix X und Label-Vektor y



Bag-of-Words Vektorisierung (TF-IDF)

Dokumente

D_1 : „Pete likes London. Pete likes Paris.“

D_2 : „Pete does not like London.“

D_3 : „Pete likes London, but not Paris.“

	Pete	like	London	do	not	but	Paris
D_1	2	2	1				1
D_2	1	1	1	1	1		
D_3	1	1	1		1	1	1

Nur Worthäufigkeiten beachtet

- › Term Frequency (TF, TF-IDF)
- › Einfach, aber robust
- › Basis für viele Algorithmen
(Retrieval, Klassifikation, Topic Modeling)

Nachteile

- › Stark vereinfachendes Sprachmodell
- › Syntaktische und relationale Informationen gehen verloren

Verbesserung durch n-Gramme

Code-Beispiel

<https://github.com/jsalbr/tdwi-2023-text-mining>

Notebook "Classification"

Evaluation mit Power BI

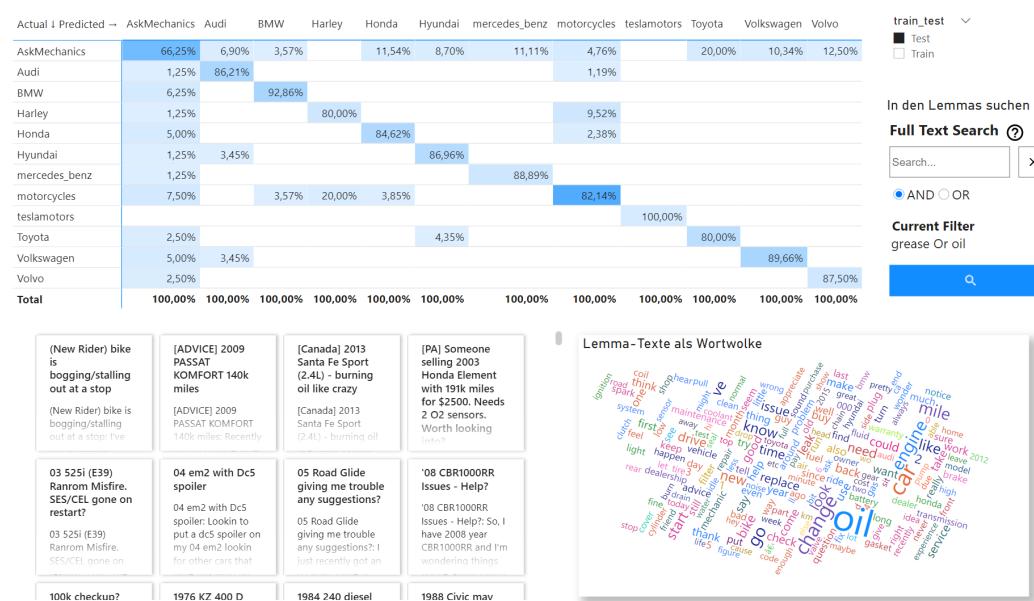
Confusion Matrix im Mehrklassen-Fall

		Actual								
		Predicted								
		True Negatives							False Positives	
1	0	12	0	0	0	0	0	1	0	0
0	1	0	10	0	0	0	0	0	0	2
0	0	0	0	10	0	0	0	7	0	0
0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	18	0	0	0	TN
0	0	0	0	0	0	0	15	0	0	0
0	0	0	0	0	0	0	0	17	0	0
		False Negatives							TP	FN
1	0	1	0	2	1	0	0	3	0	0
0	1	0	3	1	1	0	0	1	FP	TN
1	1	0	0	3	0	0	0	5	0	0

Text-Mining-Ergebnisse darstellen und inhaltlich prüfen

Aufbau einer interaktiven „Confusion Matrix“

- Matrix-Visual – erlaubt Filterung aus der Confusion-Matrix auf Texte und Wortwolken
- Anteil an Spalten (Prediction-Spalten)
→ Wieviel ist je Kategorie richtig vorhergesagt worden?
- Filter (Slicer) für Train vs. Test



Live-Demo: Übungs- und Lösungsdatei auf Github

<https://github.com/jsalbr/tdwi-2023-text-mining>

Actual ↓ Predicted →	AskMechanics	Audi	BMW	Harley	Honda	Hyundai	mercedes_benz	motorcycles	teslamotors	Toyota	Volkswagen	Volvo	
AskMechanics	66,25%	6,90%	3,57%		11,54%	8,70%		11,11%	4,76%		20,00%	10,34%	12,50%
Audi		1,25%	86,21%										1,19%
BMW	6,25%		92,86%										
Harley	1,25%			80,00%									9,52%
Honda		5,00%			84,62%								2,38%
Hyundai	1,25%	3,45%				86,96%							
mercedes_benz	1,25%						88,89%						
motorcycles	7,50%		3,57%	20,00%	3,85%			82,14%					
teslamotors									100,00%				
Toyota	2,50%					4,35%				80,00%			
Volkswagen		5,00%	3,45%								89,66%		
Volvo	2,50%											87,50%	
Total		100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	

train_test ▾
 Test
 Train

In den Lemmas suchen

Full Text Search ⓘ

X

AND OR

Current Filter
grease Or oil

🔍

(New Rider) bike is bogging/stalling out at a stop

[ADVICE] 2009 PASSAT KOMFORT 140k miles

[Canada] 2013 Santa Fe Sport (2.4L) - burning oil like crazy

[PA] Someone selling 2003 Honda Element with 191k miles for \$2500. Needs 2 O2 sensors. Worth looking into?

03 525i (E39) Ranrom Misfire. SES/CEL gone on restart?

04 em2 with Dc5 spoiler

05 Road Glide giving me trouble any suggestions?

'08 CBR1000RR Issues - Help?

100k checkup?

1976 KZ 400 D

1984 240 diesel

1988 Civic mav

Lemma-Texte als Wortwolke



Klassifikation mit semantischen Features

Semantik Lernen

Distributional Hypothesis (Firth, 1957)

"You shall know a word by the company it keeps."

Beispiel: Was ist "tezgüino"? Was ähnelt "tezgüino"?

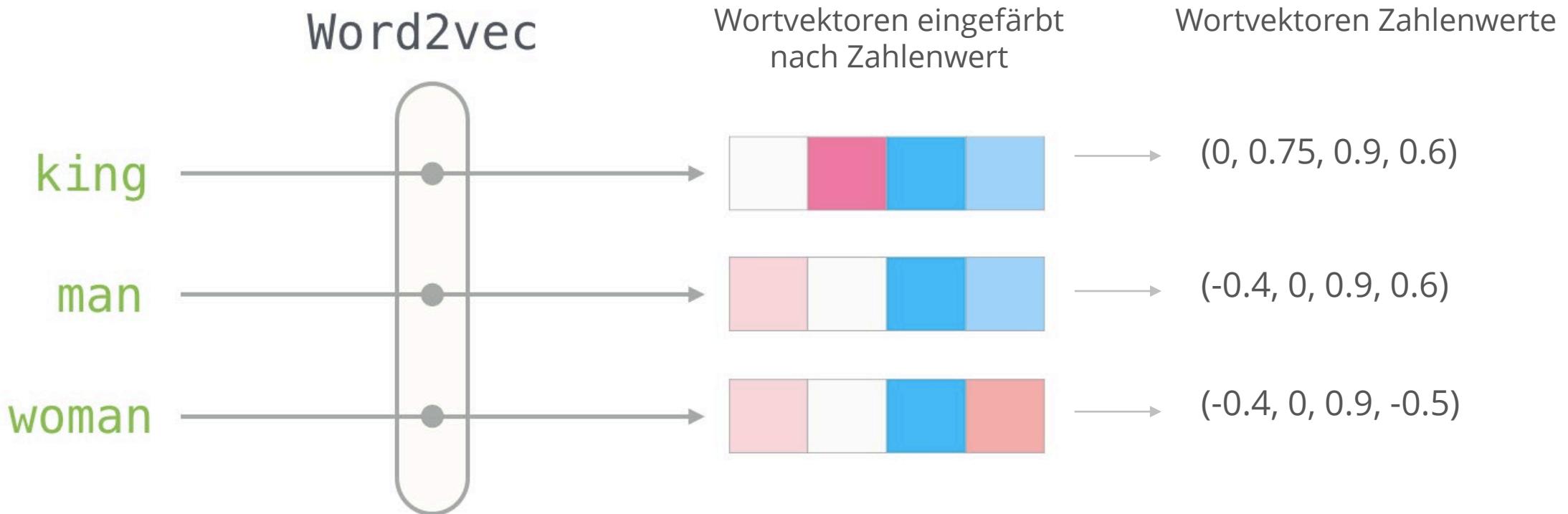
A bottle of ___ is on the table.

Everybody likes ___.

Don't have ___ before you drive.

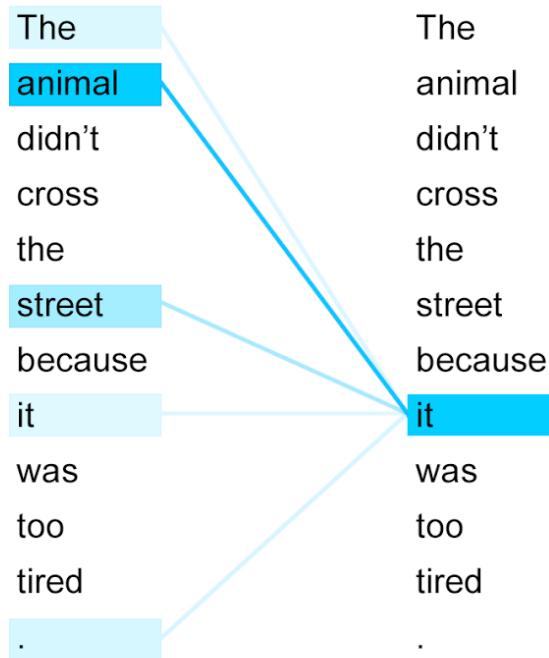
We make ___ out of corn.

Word Embeddings (Word2Vec, 2013)

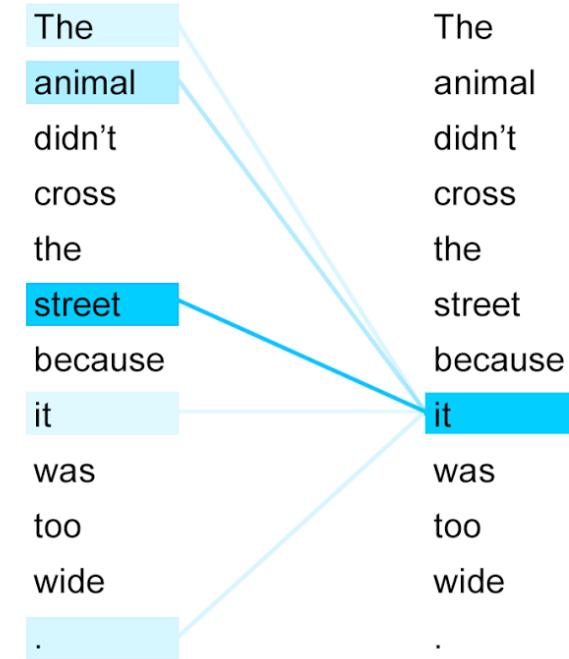


Attention und Transformer Modelle (Google 2017)

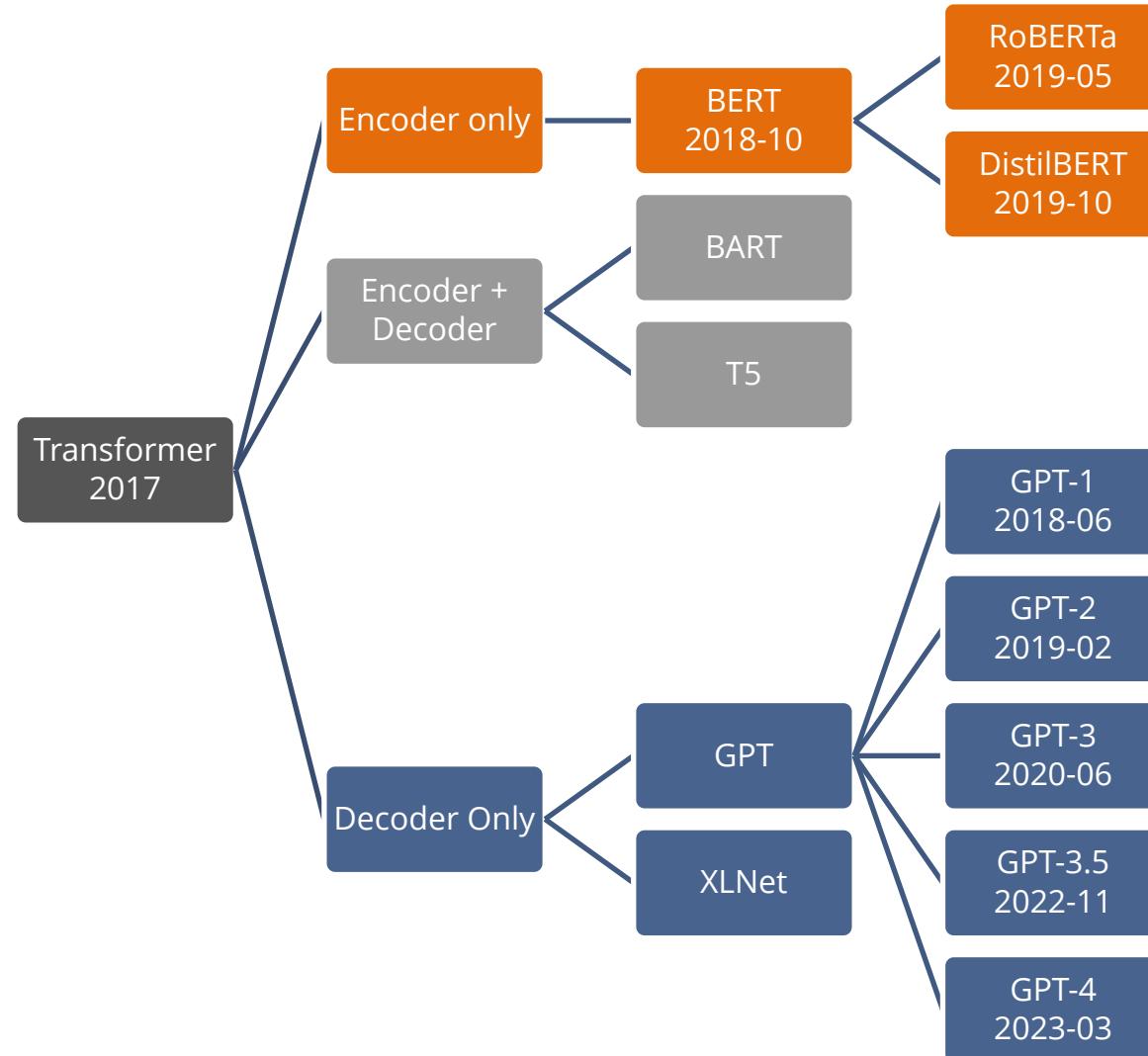
The animal didn't cross the street,
because it was too tired.



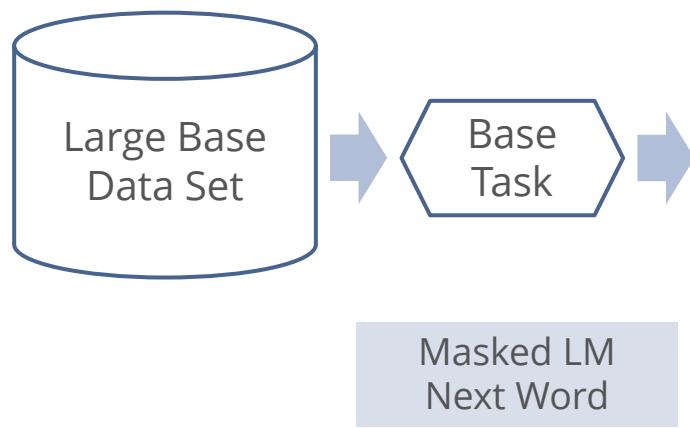
The animal didn't cross the street,
because it was too wide.



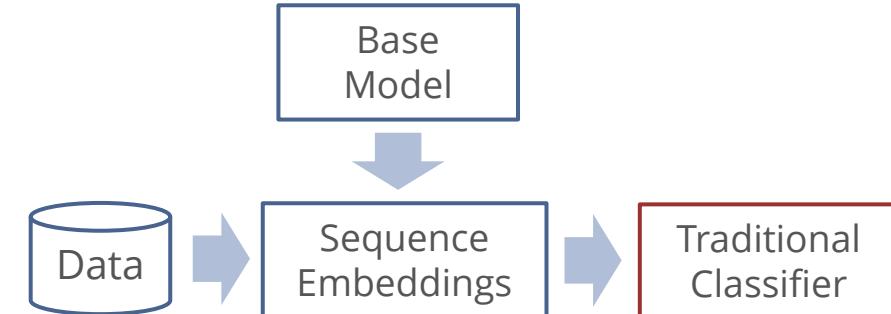
Transformer-Evolution



Klassifikation mit Transformer-Modellen



Variante 1: Nur Vektorisierung durch Basis-Modell



Variante 2: Fine-Tuning eines Base Models



Code-Beispiel

<https://github.com/jsalbr/tdwi-2023-text-mining>

Notebook "Classification BERT"

TF-IDF oder Embeddings?

TF-IDF + SVM

- Sehr schnelles Training, sehr schnelle Vorhersagen
- Funktioniert gut, wenn Klassen durch Schlüsselwörter gut separierbar
- Nicht gut, wenn Semantik der Texte relevant

Trainingszeit: 2 sek.

Accuracy: 86.4%

Vortrainierte Embeddings + SVM

- Training und Vorhersage mit dicht-besetzten Embedding-Vektoren deutlich langsamer als auf TF-IDF
- Ergebnisse häufig schlechter als mit TF-IDF

Embedding: 1 min.
Trainingszeit: 3 min.

Accuracy: 72.4%

Embeddings + Fine Tuning

- Training und Vorhersagen um mehrere Größenordnungen aufwändiger! (GPU erforderlich)
- Bessere Ergebnisse meist nur mit Fine-Tuning

Trainingszeit: 8 min.

Accuracy: 88.8%

Semantische Analysen mit Large Language Models

GPT



```
prompt = "Explain the term Business Intelligence with a joke in at most 100 words!"  
  
response = openai.Completion.create(  
    model="text-davinci-003",  
    prompt=prompt,  
    max_tokens=200  
)  
  
print(response['choices'][0]['text'])
```

- Q. How did the Business Intelligence expert collect data?
A. With his BI-noculars!

Code-Beispiel

<https://github.com/jsalbr/tdwi-2023-text-mining>

Notebook "Semantic_Analysis_LLM"

Sentiment Analyse mit GPT

Instruction: The text following the three dashes is about cars and their components.



Extract a list in json format where each entry includes the mentioned car component or part, the corresponding sentiment as "positive" or "negative", the respective car brand, the car model, and the year of making if specified.

Do not generate entries with neutral or no sentiment.

If a model name starts with a year, remove it from the model name and put it into a separate json field.

Each entry in the json list should have the fields "model", "vendor", "part", "sentiment", and "year".

I love my BMW i3 because it has a really strong engine for such a small car. But the battery could be stronger. We used to have a 2003 Audi A4 cabriolet which was very reliable but had problems with the sun roof.

```
[{"model": "i3", "vendor": "BMW", "part": "engine", "sentiment": "positive", "year": null},  
 {"model": "A4 cabriolet", "vendor": "Audi", "part": "sun roof", "sentiment": "negative",  
 "year": 2003}]
```

OpenAI Playground

Mode

Complete ▾

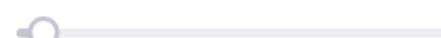
Model

text-davinci-003 ▾

Temperature 1



Maximum length 256



Verarbeitungskosten

	count	avg_words
subreddit		
BMW	500	146.92
Tesla	500	158.42
Volkswagen	500	143.12

ChatGPT models are optimized for dialogue. The performance of gpt-3.5-turbo is on par with Instruct Davinci.

[Learn more about ChatGPT ↗](#)

Model	Input	Output
4K context	\$0.0015 / 1K tokens	\$0.002 / 1K tokens
16K context	\$0.003 / 1K tokens	\$0.004 / 1K tokens

<https://openai.com/pricing>

→ Nur ca. 1\$ für 1.500 Text mit 150 Wörtern!

```
id = 'n8unhf'  
display_parts(id)
```

executed in 627ms, finished 11:22:58 2023-06-20

	id	vendor	model	year	part	sentiment
179	n8unhf	BMW	X3	2018	equalizer	negative
180	n8unhf	unknown	Volvo	None	sound system	positive

Anyone care to share their HK **equalizer** settings 2018 **X3**? I can't seem to get a great sound stage that isn't too tinny or hollow sounding. I'd like to see the settings you have configured if you've found something you're happy with. Although this is supposed to be a premium system, it in no way compares to my **Volvo**. I also wish **bmw** had thought about creating preset eq settings that are quick to change depending on what you are listening to or at least let customers create and save several of their favorite settings.

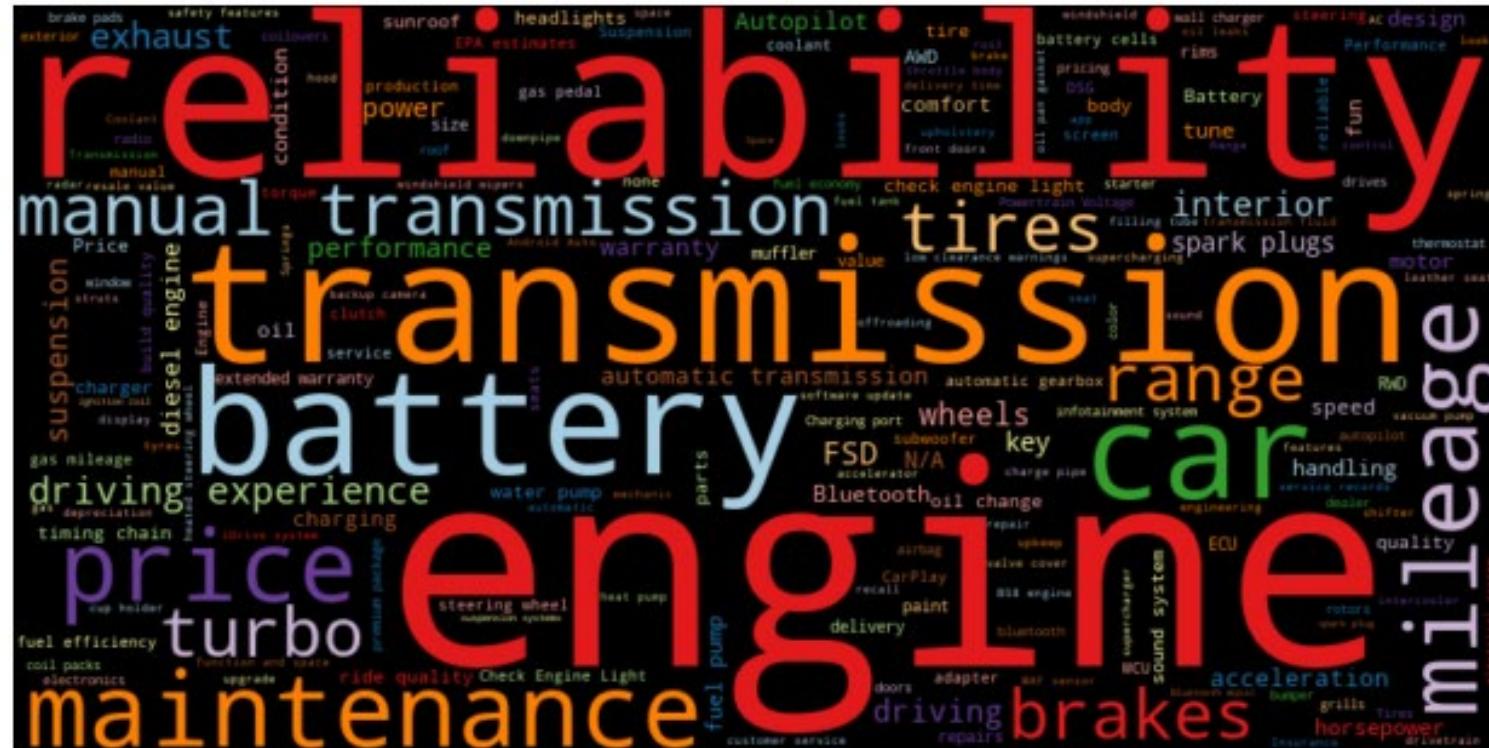
```
# Hallucination - this post says nothing about transmission or engine
# but only "from a quick inspection everything seemed good"
id = 'mxyt0d'
display_parts(id)
```

executed in 315ms, finished 12:50:41 2023-06-18

	id	vendor	model	year	part	sentiment
960	mxyt0d	BMW	745li	2003	center console	negative
961	mxyt0d	BMW	745li	2003	driver side door sill	negative
962	mxyt0d	BMW	745li	2003	engine	positive
963	mxyt0d	BMW	745li	2003	transmission	positive

2003 BMW 745li Question: A friend of mine bought a 2003 BMW 745li (I know...first thing I said was "may god help you..and instantly asked how much did you pay?") at 81,960 miles for \$3k. It runs good (for now...), lifted the car doesn't seem to be any noticeable problems underneath, but there are some cosmetic issues inside (some cracks in the center console but nothing too crazy, there some peeling by driver side door sill, other than that it is clean). Took it to a trusted mechanic and they said from a quick inspection everything seemed good. He kind of bought it on a whim, and didn't realize the insurance and maintenance cost in the long term. He's now is open to sell it. *I know it is hard to tell over a post but what do you think the car will go for in private?* CarMax is giving \$4800 through their online offer. I told him he should go with CarMax because of pretty large profit he's getting and easy sale. Trying to private sale on '03 7-series sounds like too much work. Mods: I am not trying to sell the car here, but if this is against rule 6, I apologize and I'll remove the post.

Genannte Komponenten



Erstellung einer Gruppierungstabelle mit ChatGPT

RZ

Please group all of the following words into 10 different technical groups, where each should be distinct compared to the others. Make sure that every part from the list is in one group. You are not allowed to skip any of the parts from the list. In case you can group a part into one of the 10 different technical groups you can create an 11th group "misc":

Group 1: Engine and Powertrain

- Engine
- Turbo
- Diesel engine
- Motor
- Spark plugs
- Acceleration
- Horsepower
- Tune
- Oil
- Gas pedal
- Torque
- Powertrain voltage
- Coilovers
- Coil packs
- Control
- Oil leaks

Group 2: Transmission

- Transmission
- Manual transmission
- Automatic transmission
- Gearbox
- DSG

Group 3: Battery and



RZ

Please create a table from this list with column headers "Group" and "Part"

Group | Part

Here's a table with the column headers "Group" and "Part":

Group	Part
Eng	Kopieren
Eng	Alles auswählen
Eng	Engine
Eng	Turbo

Here's a table with the column headers "Group" and "Part":

Group	Part
1	Engine
2	Engine
3	Diesel engine
4	Motor
5	Spark plugs
6	Acceleration
7	Horsepower
8	Tune
9	Oil
10	Gas pedal
11	Torque
12	Powertrain volt...
13	Coilovers
14	Coil packs
15	Control
16	Oil leaks
17	Transmission
18	Manual transm...
19	Automatic tran...
20	Gearbox
21	DSG
22	Battery
23	Battery cells
24	Charging

APPLIED STEPS

Source
Trimmed Text
Removed Duplicates

Minimale ETL-Schritte nötig und Verbindung der Tabellen

ChatGPT JSON - Power Query Editor

Datei **Home** **Transform** **Add Column** **View** **Tools** **Help**

Queries [4]

- Groups_manual_from_ChatGPT
- Reddit_Post_Gpt_Sentiment
- Parts_with_Count
- Reddit_raw_Posts

Table.FromRows(Json.Document)

	A ^b c Group	A ^b c Part
1	Engine	Engine
2	Engine	Turbo
3	Engine	Diesel engine
4	Engine	Motor
5	Engine	Spark plugs
6	Engine	Acceleration
7	Engine	Horsepower
8	Engine	Tune
9	Engine	Oil
10	Engine	Gas pedal
11	Engine	Torque
12	Engine	Powertrain voltage
13	Engine	Coilovers
14	Engine	Coil packs
15	Engine	Control

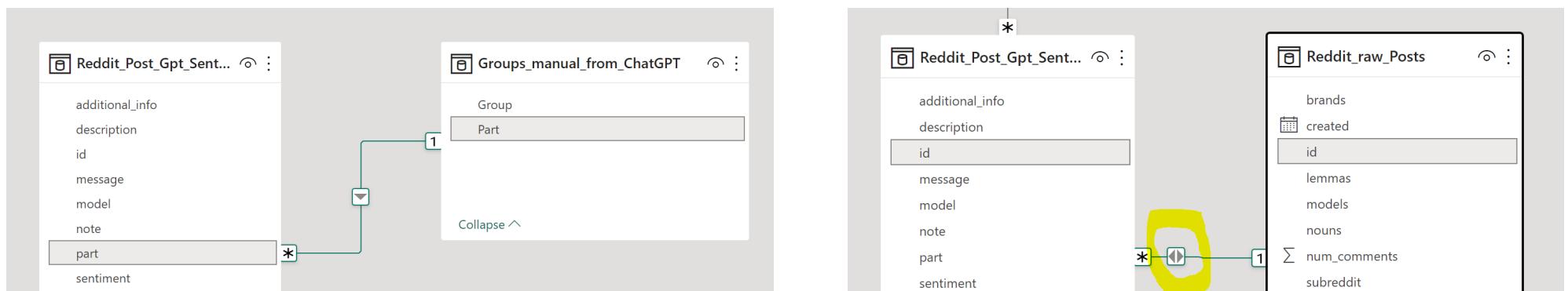
Query Settings

PROPERTIES

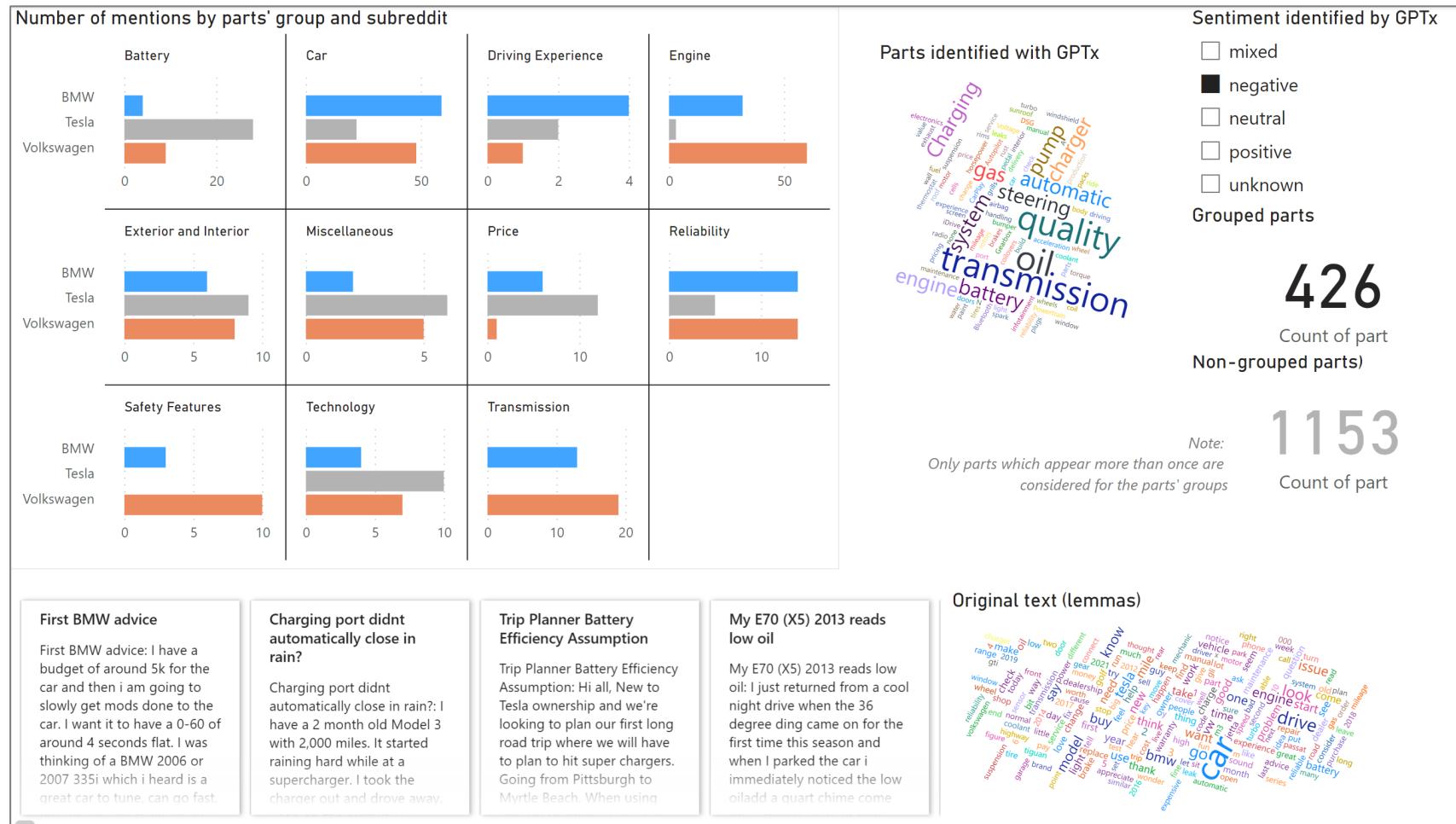
Name: Groups_manual_from_ChatGPT

APPLIED STEPS

- Source
- Trimmed Text
- Removed Duplicates



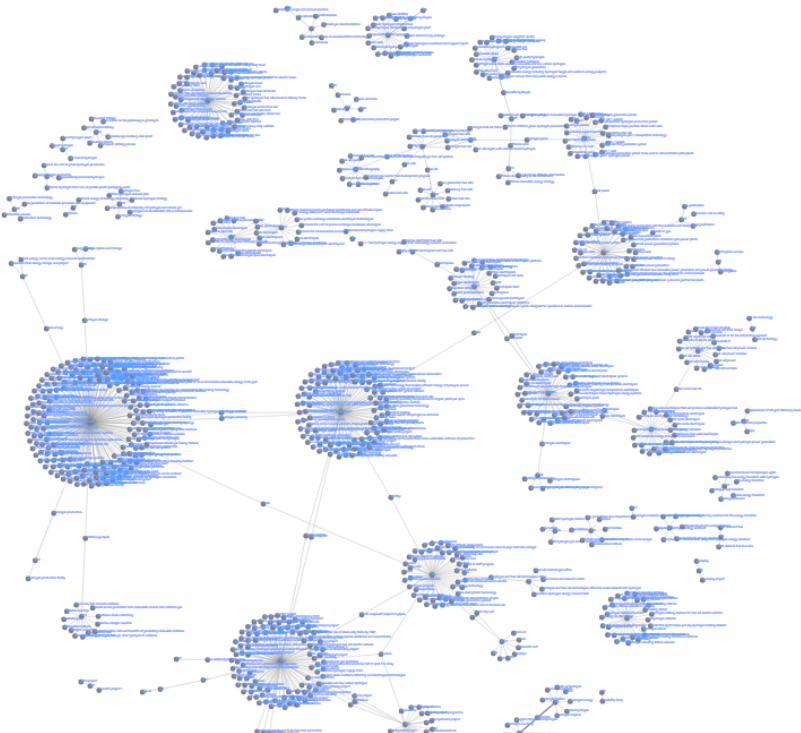
Teileanalyse auf Basis der Reddit-Daten



<https://github.com/jsalbr/tdwi-2023-text-mining/tree/main/PowerBI> → 4 Teileanalyse..

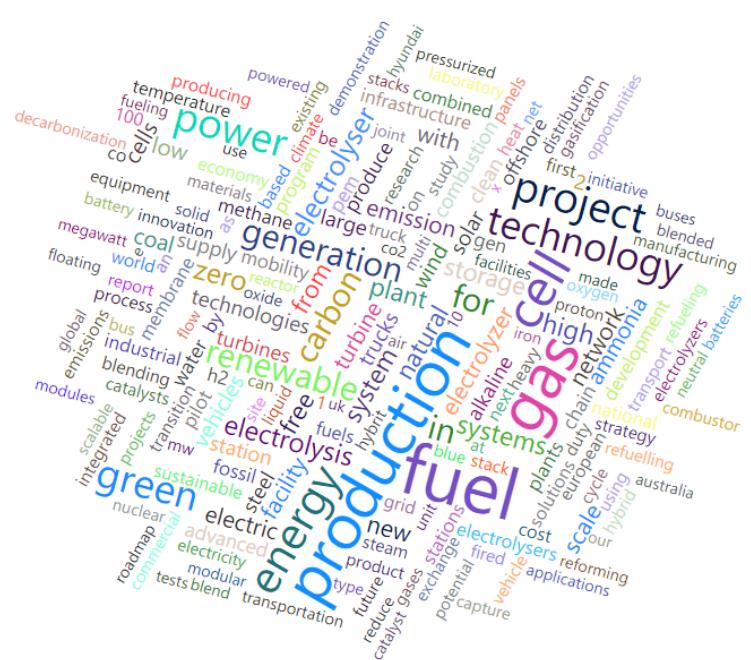
Vertiefende „Intelligence“ mit Power BI

Visualisierung von Topics und NER-Ergebnissen als Netzwerke



organisations	No. of Paragraphs
doe	80
itm power	80
air liquide	72
plug power	59
toyota	58
linde	56
hyundai	54
hydrogen council	44
nel	43
engie	42
siemens	41
uniper	41
arena	40
shell	39
Total	1845

What is the object of cooperation?

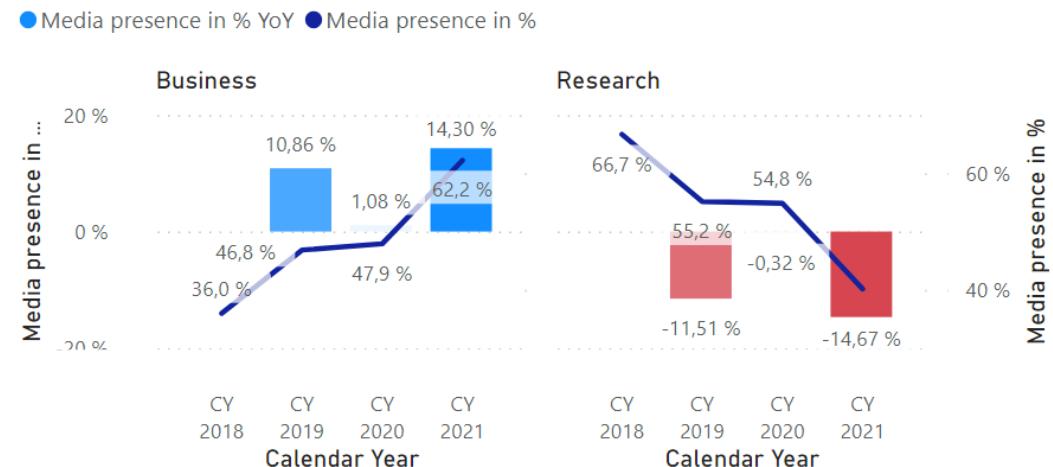


Veränderungen mit Time-Intelligence-Funktionen berechnen

Medienpräsenz und Änderung bei Topics

Calendar Year	CY 2021	No of paragraphs (coop)	Media presence in %	Media presence in % YoY
top_50_words				
production, hydrogen, gas, facility, ...	81	17,5 %	7,02 %	
hydrogen, ccus, boilers, bond, engi...	42	9,1 %	3,31 %	
zero, carbon, emission, project, free,...	37	8,0 %	1,71 %	
green, hydrogen, investment, blue, ...	28	6,0 %	2,51 %	
scale, large, project, production, exp...	23	5,0 %	-0,53 %	
hydrogen, nitrogen, polymers, socie...	22	4,8 %	-3,36 %	
natural, gas, network, hydrogen, ble...	18	3,9 %	-0,30 %	
Total	463	100,0 %	0,00 %	

Media presence in % YoY and Media presence in % by Calendar Year and Business or Research



Fazit



Prof. Dr. Jens Albrecht

TH Nürnberg, Informatik



Data Warehousing, BI,
Data Science, NLP
jens.albrecht@th-nuernberg.de
Für Beratungsprojekte:
jens.albrecht@data-knowhow.de

**Sprechen Sie uns an, für Ihre
akuten NLP-Fragen!**



Prof. Dr. Roland Zimmermann

TH Nürnberg, BWL



BI, Information Design,
NLP, Process Mining
roland.zimmermann@th-nuernberg.de
Für Beratungsprojekte:
zimmermann@architecting-analytics.com