

PREDICTIVE ANALYTICS MITTELS TEXT MINING IM FUSSBALL

DAVID JÄGERING

DAVID.JAEGERING@FOM-NET.DE

HYPOTHESEN

Stimmungsbeschreibung eines Artikels durch positiv und negativ klassifizierte Begriffe.

Positive Berichterstattung über eine Mannschaft in einer Kalenderwoche führt zu erfolgreicheren Ergebnissen.

Aus der Berichterstattung über negative Aspekte, wie Verletzungen oder Skandale, können Rückschlüsse auf den Ausgang einer Partie geschlossen werden.

Eine Kombination aus Data Mining und Text Mining führt zu präziseren Ergebnissen, als eine Einzelbetrachtung.

Durch historische Ergebnisse und der Sentimentanalyse lassen sich Trainerentlassungen vorhersagen.

TEXT CLUSTERING

30

RSS Feeds

8393

Artikel

LIGANEWS

TRANSFER

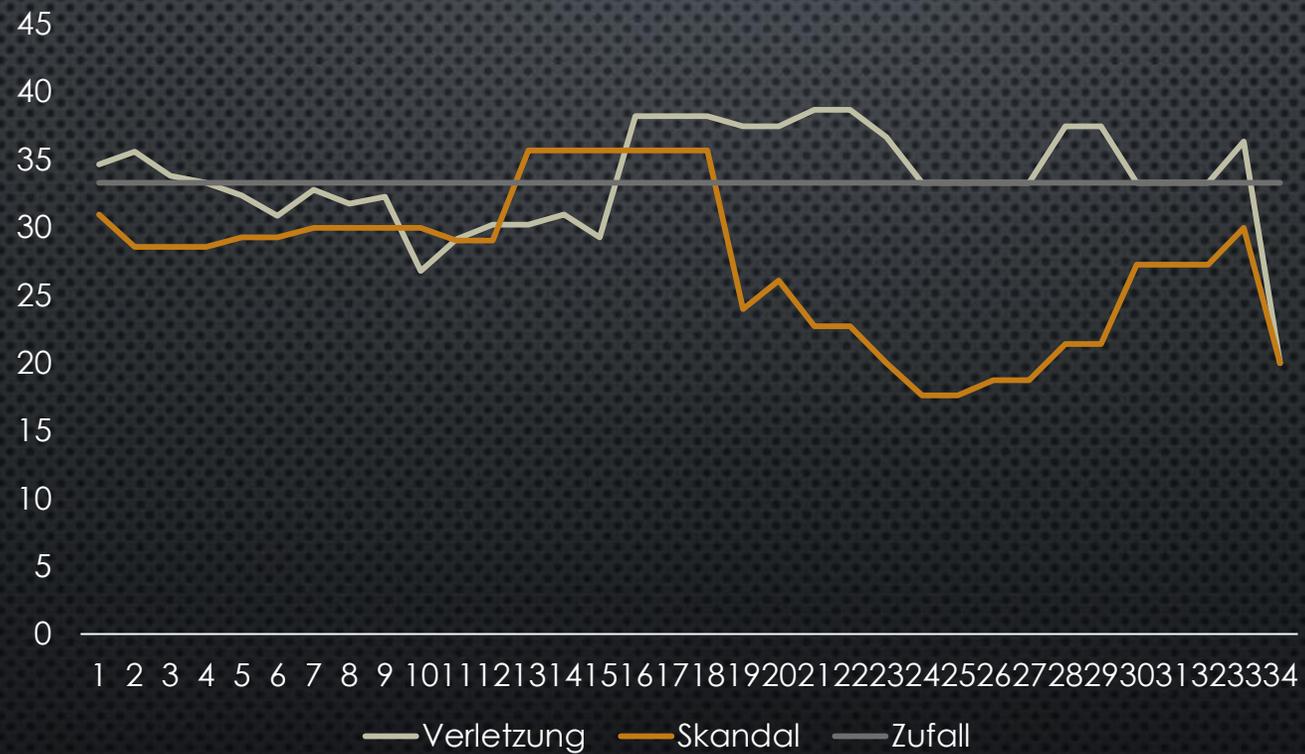
SKANDAL

VERLETZUNG

MANAGEMENT

REST

TEXT CLUSTER ALS INDIKATOR?





TEXT MINING MIT TWITTER – PREMIERE LEAGUE

20.593.365

Tweets

20

Vereine

52%

Baseline

42%

Text Mining

200

bewertete Spiele

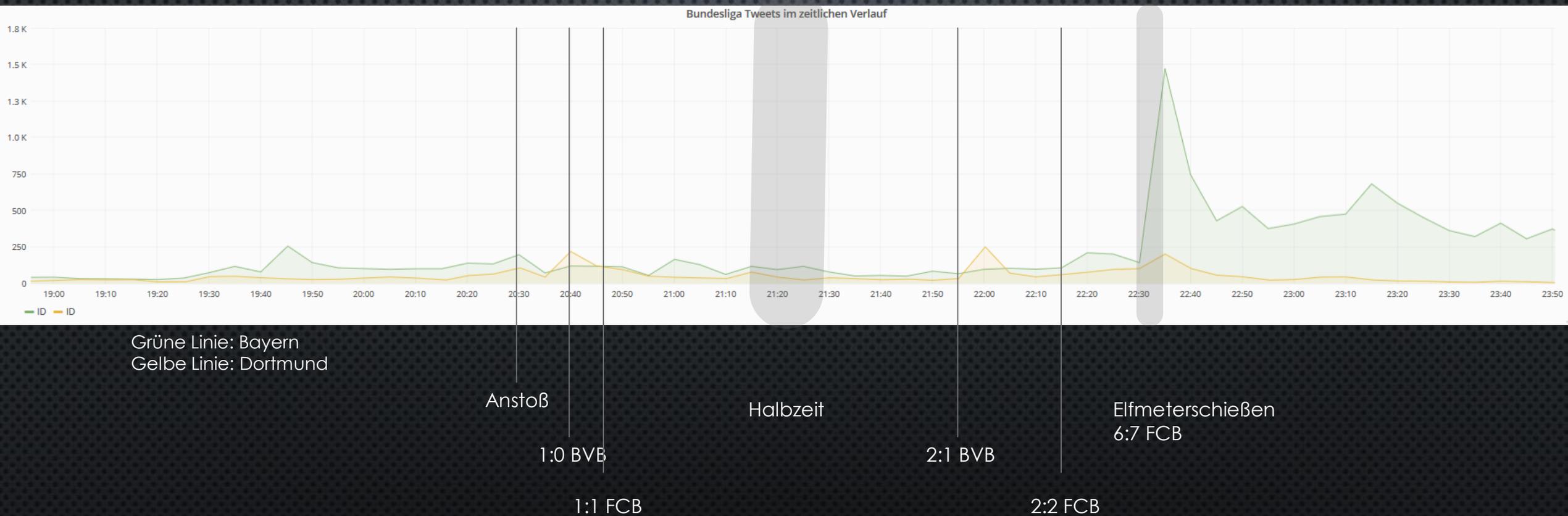
33

Tweets pro Minute

62%

Baseline & Text Mining

SUPERCUP FINALE AM 05.08.2017





Chelsea's Ghanaian defender Baba Rahman has joined Schalke 04 on loan. #cfc #schalke04
<https://t.co/4m5Cfb5wWG>

RT @amnestypress:
Qatar: Does Bayern Munich know about the conditions of migrant workers in its Doha winter training camp?
#FCBayern <https://t.co/4m5Cfb5wWG>

#HSV-Training ohne Hunt, Jung, Mavraj und Arp.
#Hamburg <https://t.co/9AoFRsVGwK>

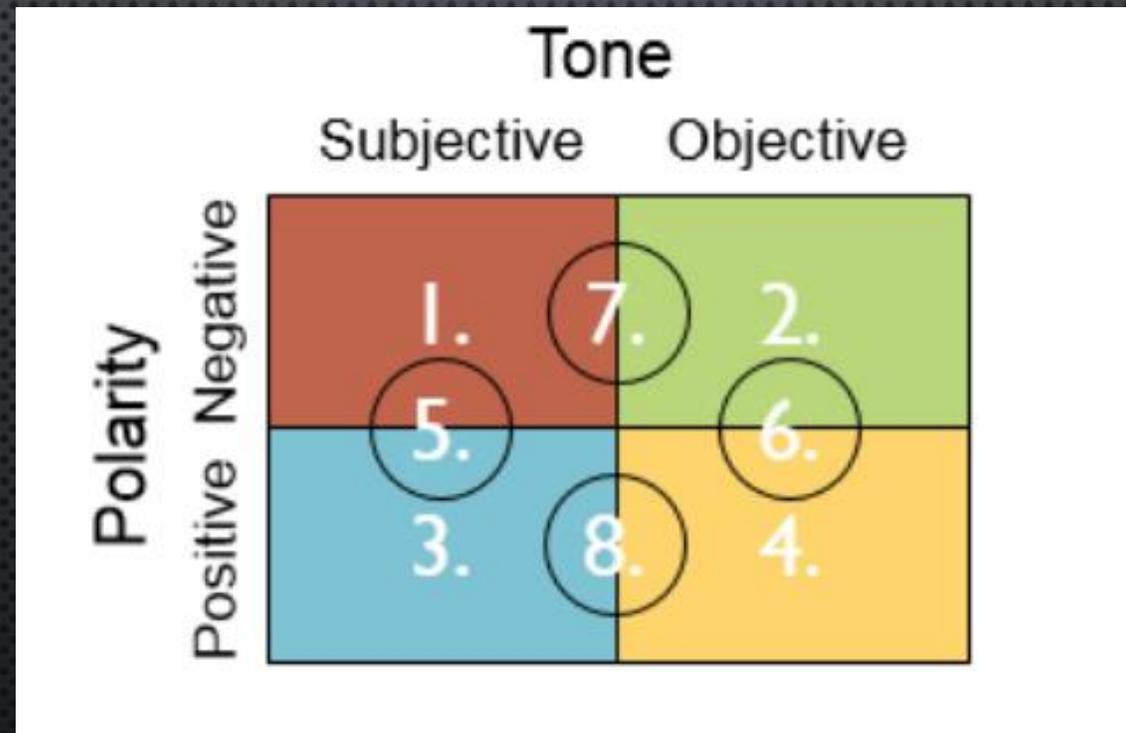
Sergio Gomez - Another Wunderkind for #BVB

Scout report by @ManuelVeth

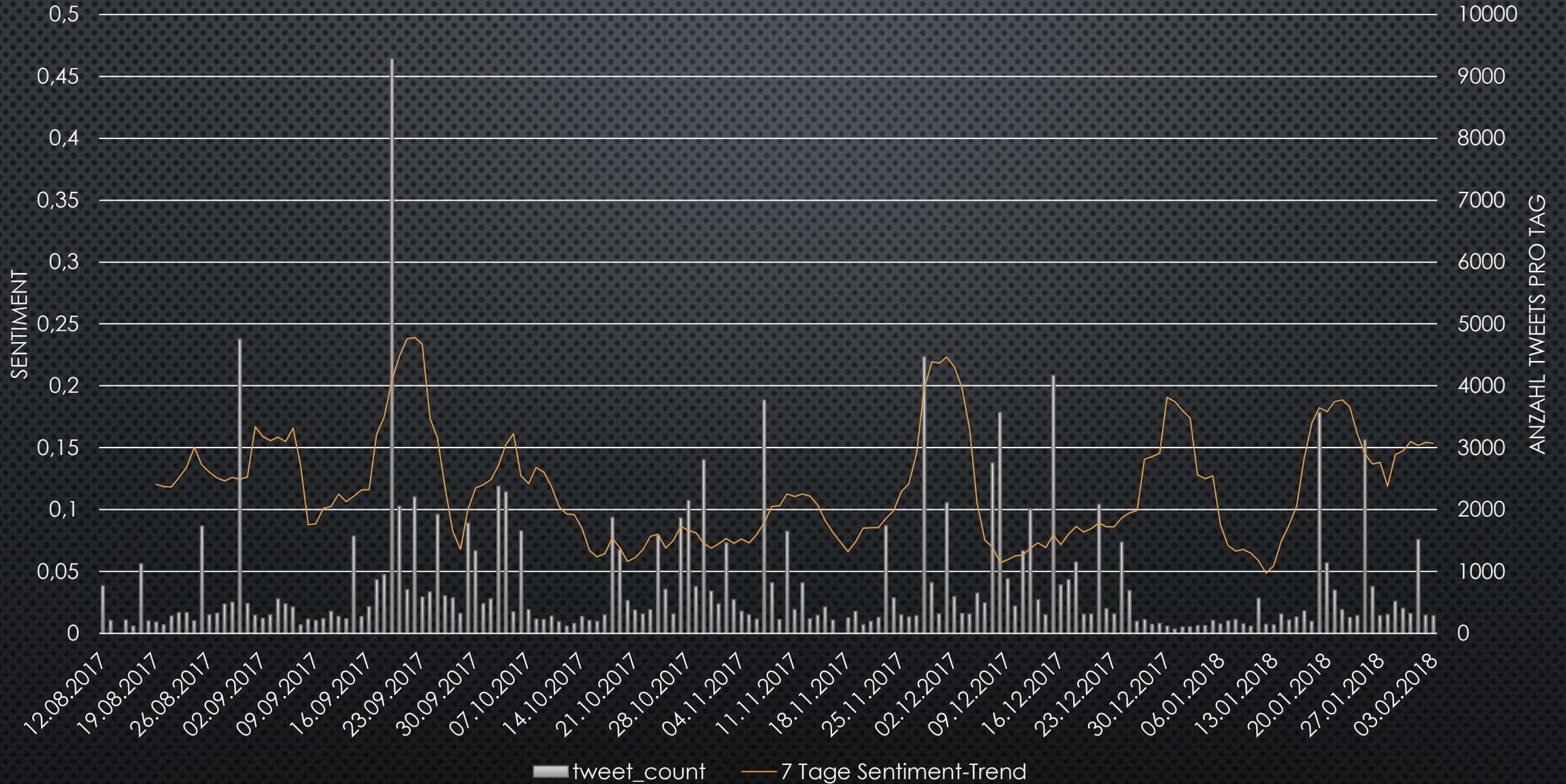
Read it right HERE ? <https://t.co/SeptgfV736>

Arsenal FC will pay #BVB a fixed transfer compensation in the amount of EUR 63.75 m.
(sixty three million seven hu... <https://t.co/JEiDBRFgGO>

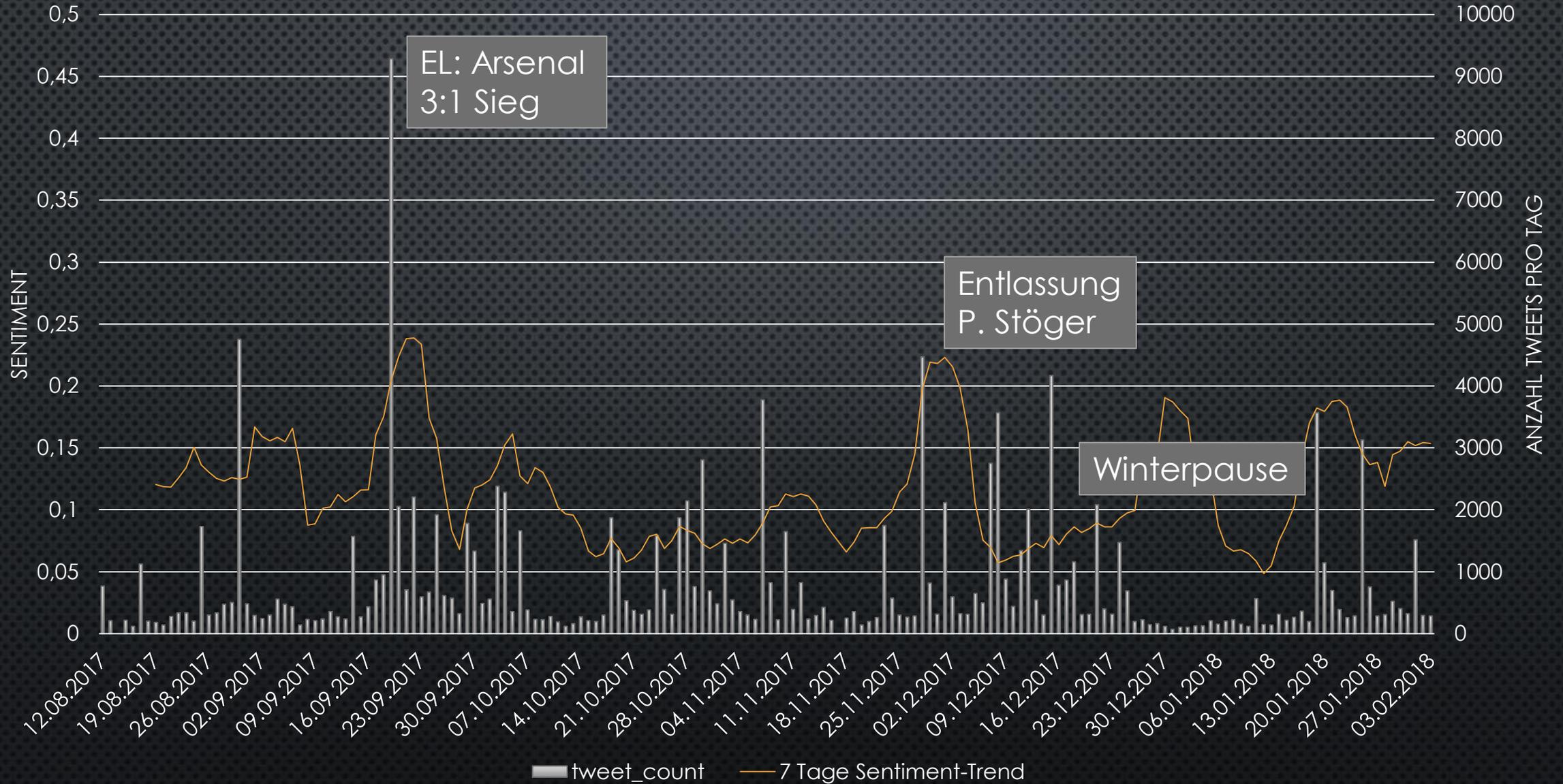
SENTIMENT MODELLE



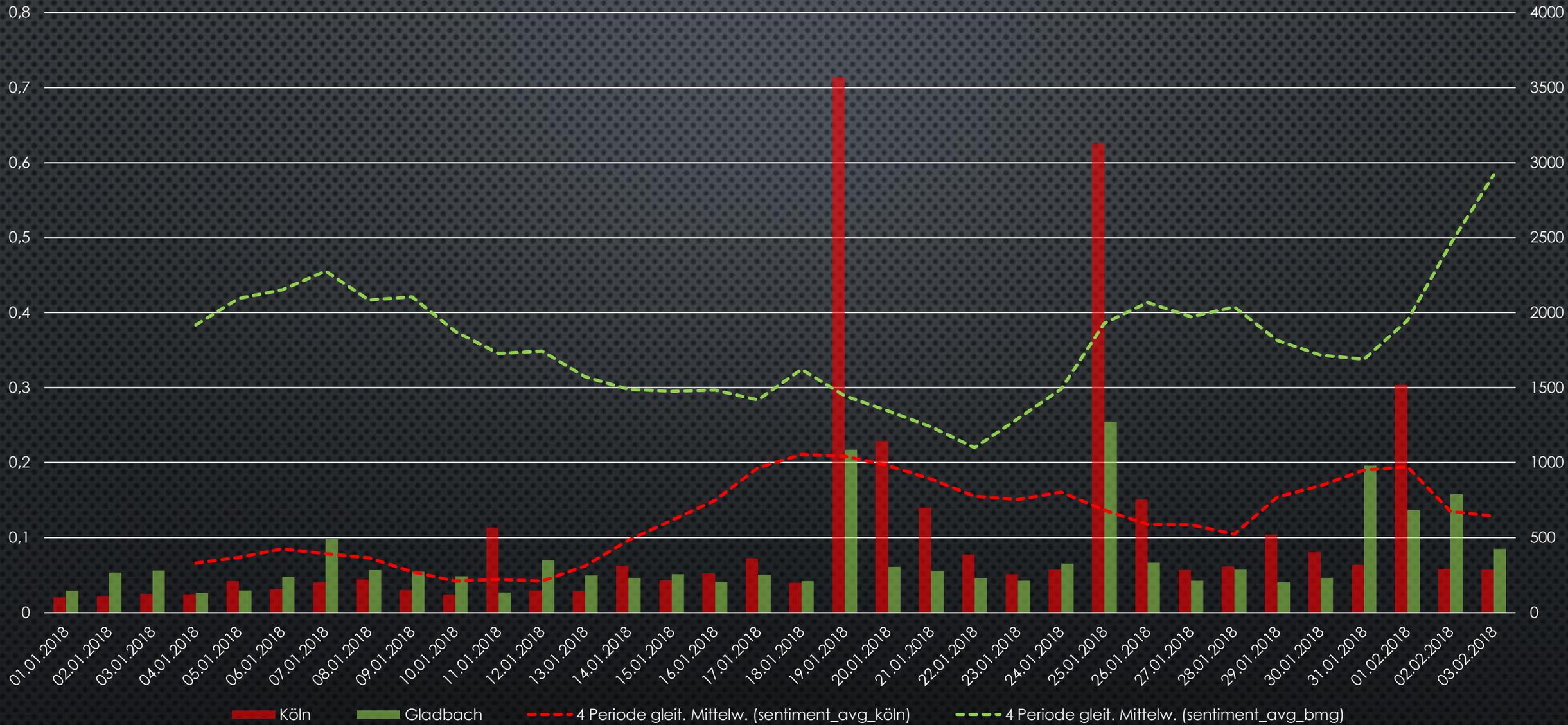
Sentiment Average - 1. FC Köln



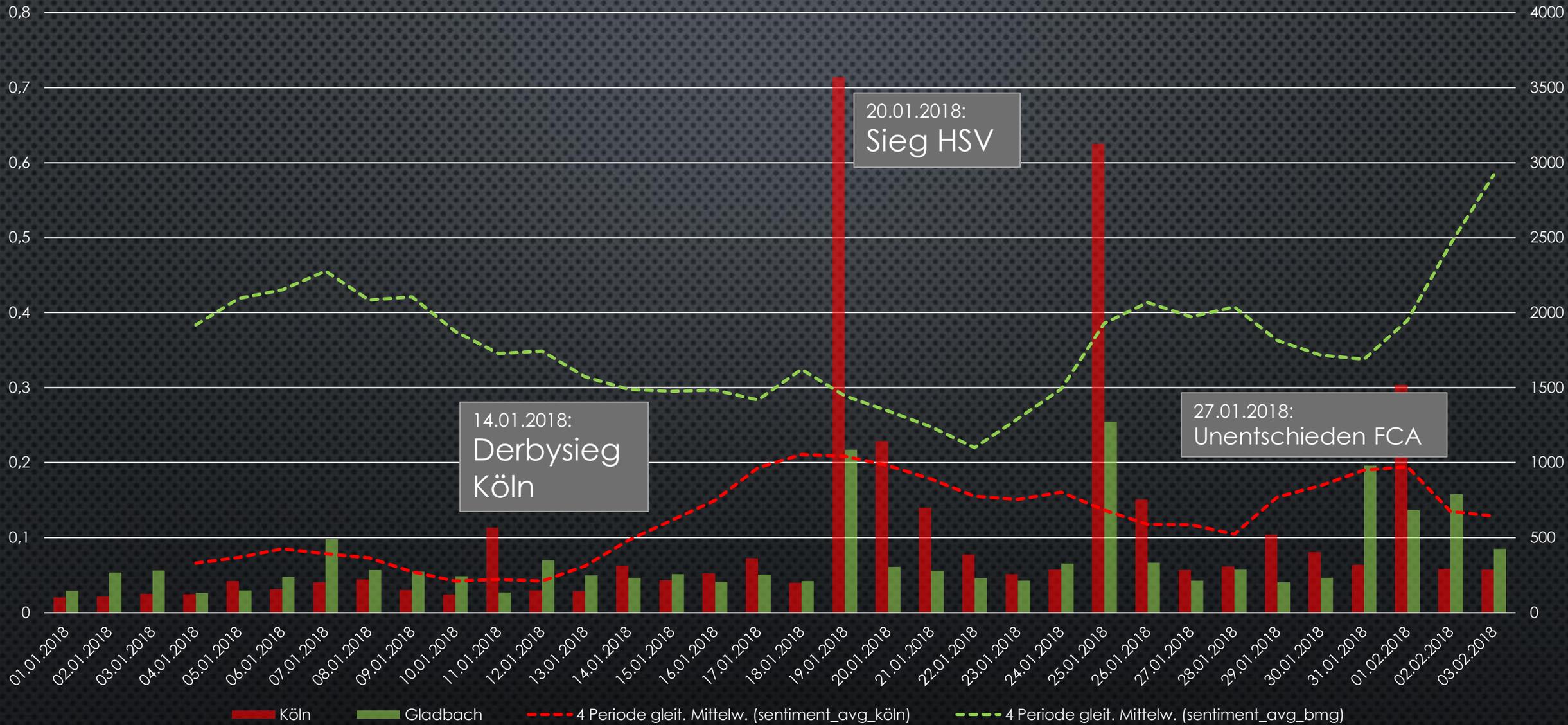
Sentiment Average - 1. FC Köln



Sentiment Average - 1. FC Köln // Borussia Mönchengladbach



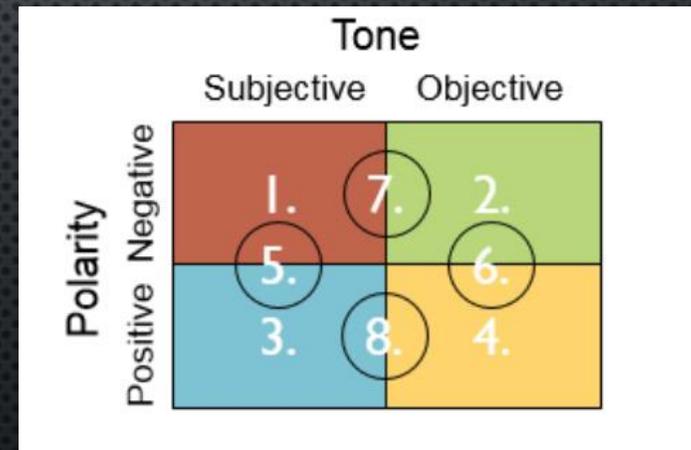
Sentiment Average - 1. FC Köln // Borussia Mönchengladbach



KOMBINATION DATA MINING UND TEXT MINING

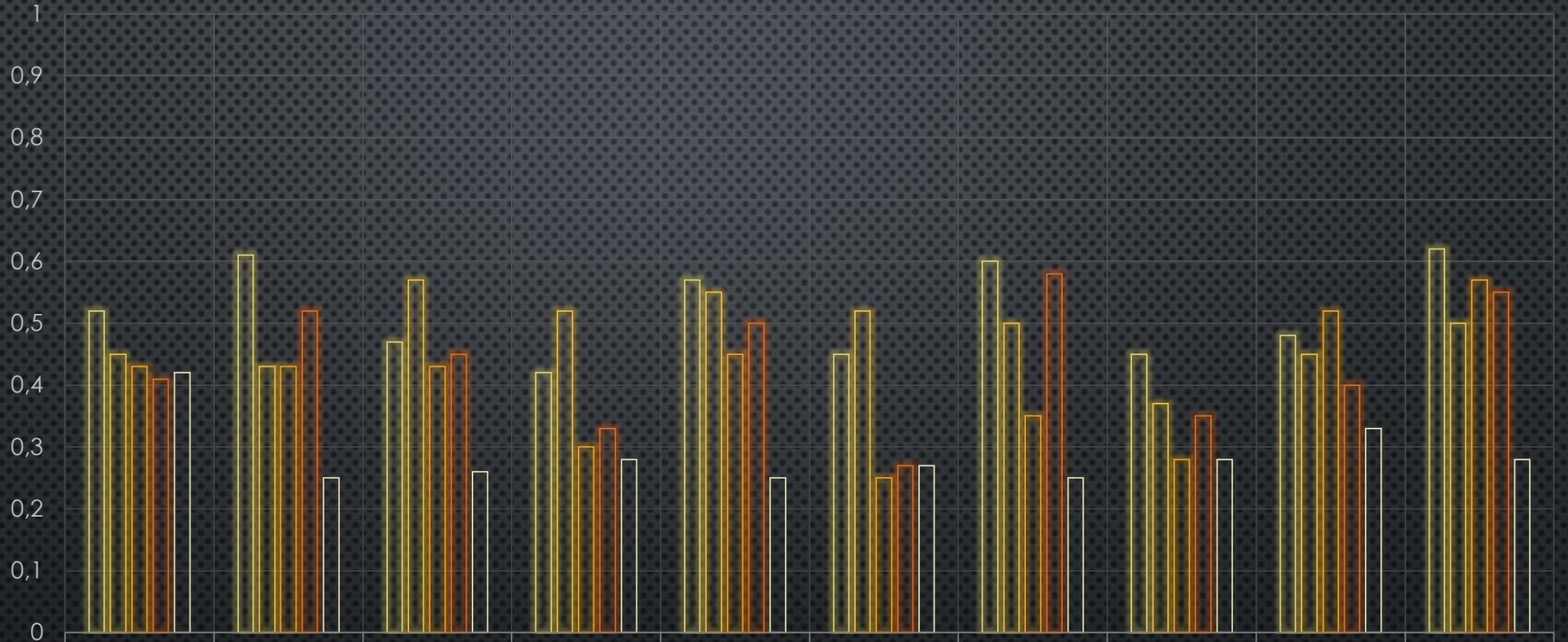
- BETEILIGTE MANNSCHAFTEN
- ENDERGEBNIS
- POISSON DISTRIBUTION
 - 10 JAHRE RÜCKWIRKEND
 - 7 SPIELE RÜCKWIRKEND
- DISTANZ ZWISCHEN DEN VEREINEN

- SENTIMENT MODELLE NACH SHUMAKER ET AL.



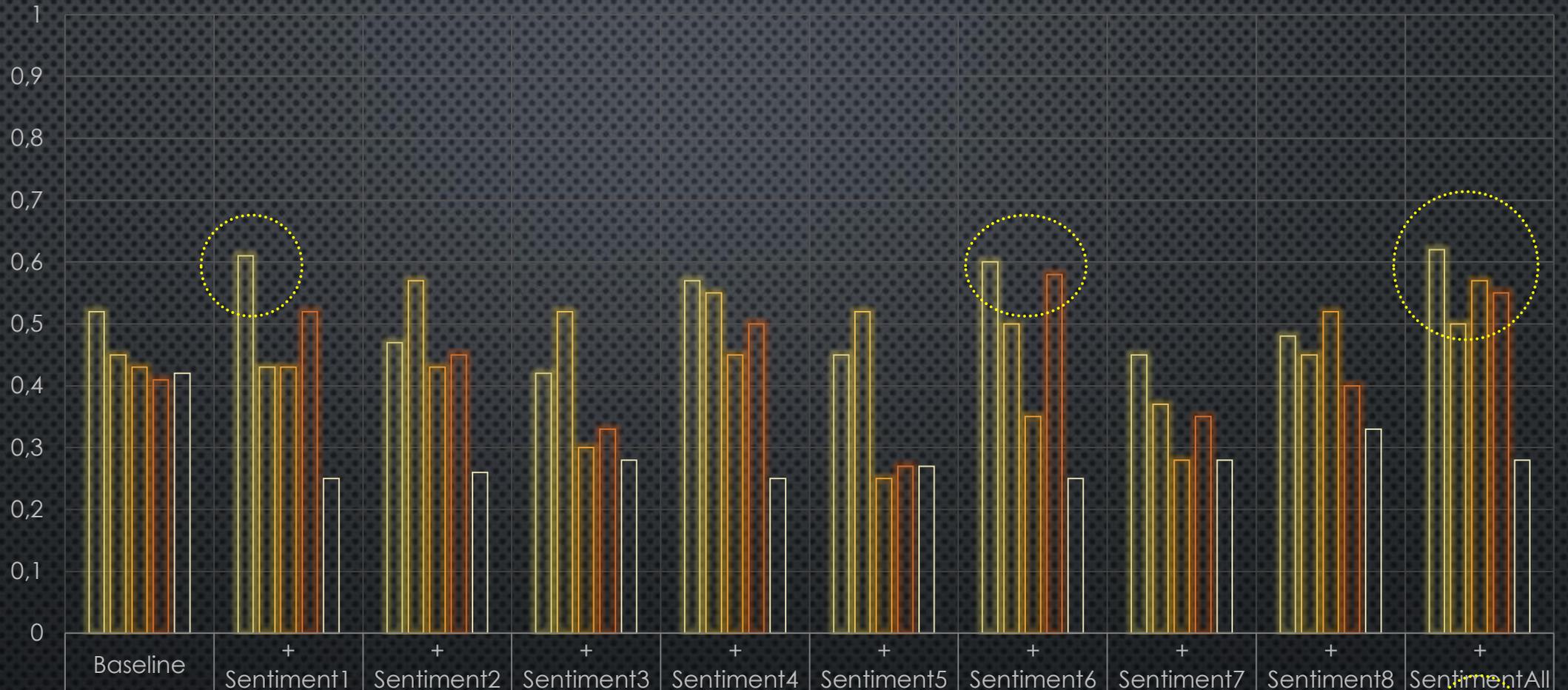
➔ BASELINE

Vergleich der Algorithmen



	Baseline	+ Sentiment1	+ Sentiment2	+ Sentiment3	+ Sentiment4	+ Sentiment5	+ Sentiment6	+ Sentiment7	+ Sentiment8	+ SentimentAll
Random Tree Forrest	0,52	0,61	0,47	0,42	0,57	0,45	0,6	0,45	0,48	0,62
Decision Tree	0,45	0,43	0,57	0,52	0,55	0,52	0,5	0,37	0,45	0,5
k-NN	0,43	0,43	0,43	0,3	0,45	0,25	0,35	0,28	0,52	0,57
Gaussian Process	0,41	0,52	0,45	0,33	0,5	0,27	0,58	0,35	0,4	0,55
Naives Bayes	0,42	0,25	0,26	0,28	0,25	0,27	0,25	0,28	0,33	0,28

Vergleich der Algorithmen



	Baseline	+ Sentiment1	+ Sentiment2	+ Sentiment3	+ Sentiment4	+ Sentiment5	+ Sentiment6	+ Sentiment7	+ Sentiment8	+ SentimentAll
Random Tree Forrest	0,52	0,61	0,47	0,42	0,57	0,45	0,6	0,45	0,48	0,62
Decision Tree	0,45	0,43	0,57	0,52	0,55	0,52	0,5	0,37	0,45	0,5
k-NN	0,43	0,43	0,43	0,3	0,45	0,25	0,35	0,28	0,52	0,57
Gaussian Process	0,41	0,52	0,45	0,33	0,5	0,27	0,58	0,35	0,4	0,55
Naives Bayes	0,42	0,25	0,26	0,28	0,25	0,27	0,25	0,28	0,33	0,28

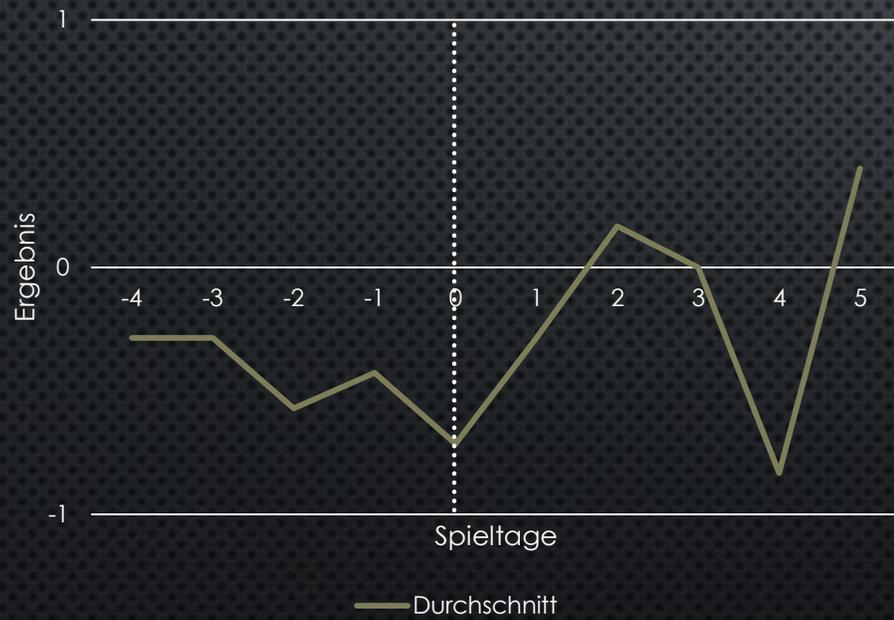
TRAINER ENTLASSUNGEN

Verein	Trainer	Entlassungsdatum
Wolfsburg	Jonker	18.09.2017
Bayern	Ancelotti	28.09.2017
Bremen	Nouri	30.10.2017
Köln	Stöger	03.12.2017
Hamburg	Gisdol	21.01.2018
Stuttgart	Wolf	28.01.2018

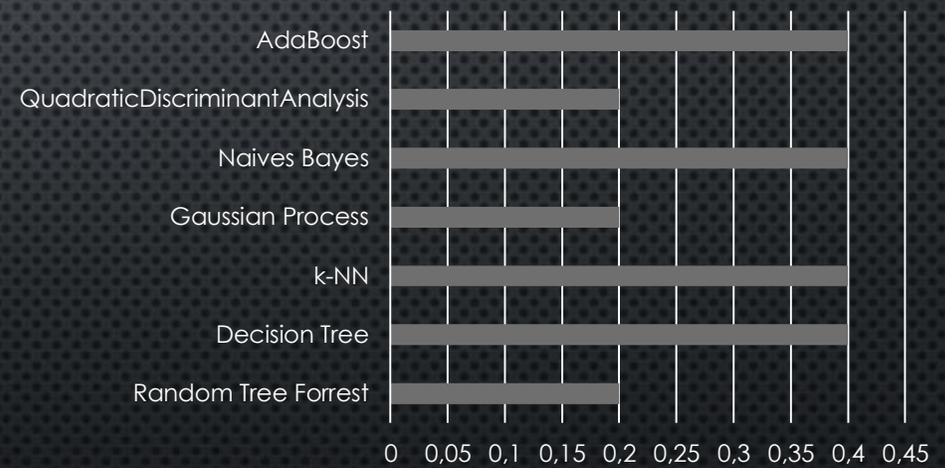
Verein	Trainer	Entlassungsdatum
Crystal Palace	de Boer	11.09.2017
Leicester	Shakespeare	17.10.2017
Everton	Koeman	23.10.2017
West Ham	Bilic	06.11.2017
West Bromwich	Pulis	20.11.2017
Swansea	Clement	20.12.2017
Stoke	Hughes	06.01.2018
Watford	Silva	21.01.2018

TRAINERENTLASSUNG

Ergebnisse vor/nach Trainerwechsel



Accuracy on Test Datasets



HYPOTHESEN



Stimmungsbeschreibung eines Artikels durch positiv und negativ klassifizierte Begriffe.



Positive Berichterstattung über eine Mannschaft in einer Kalenderwoche führt zu erfolgreicheren Ergebnissen.



Aus der Berichterstattung über negative Aspekte, wie Verletzungen oder Skandale, können Rückschlüsse auf den Ausgang einer Partie geschlossen werden.



Eine Kombination aus Data Mining und Text Mining führt zu präziseren Ergebnissen, als eine Einzelbetrachtung.



Durch historische Ergebnisse und der Sentimentanalyse lassen sich Trainerentlassungen vorhersagen.

FRAGEN ?

EXKURS - SUPERBOWL BEI TWITTER

1.388.237

Tweets

44

Sprachen

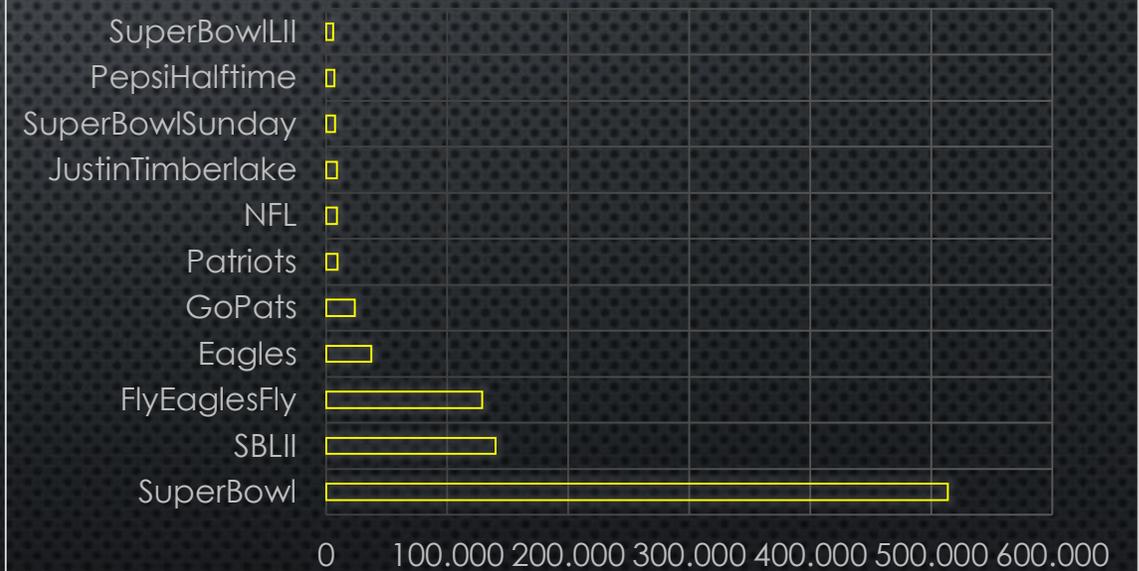
7

Stunden

52

Tweets pro Sekunde

Häufigkeit des Hashtags



27. SPIELTAG – PREMIERE LEAGUE

10.02.2018	tottenham	arsenal	H
10.02.2018	everton	crystalpalace	H
10.02.2018	stoke	brightonhoves	H
10.02.2018	westham	watfordfc	H
10.02.2018	mancity	leicester	H
11.02.2018	huddersfield	ournemouth	A
11.02.2018	newcastleud	manunited	A
11.02.2018	southampton	fcliverpool	A
12.02.2018	chelsea	westbrom	H
