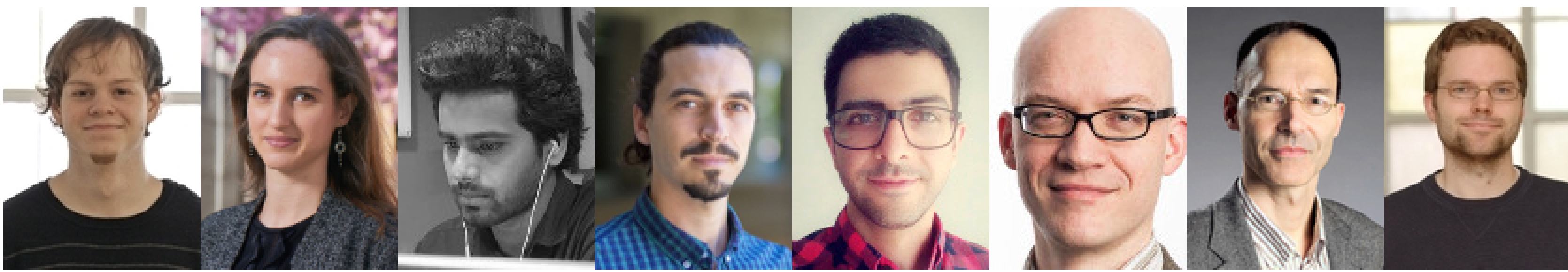
SemEval-2019 Task 4: Hyperpartisan News Detection



Johannes Kiesel[1] Maria Mestre[2] Rishabh Shukla[2] Emmanuel Vincent[2] Payam Adineh[1] David Corney Benno Stein[1] Martin Potthast[3]

Webis Bauhaus-Universität Weimar[1], Leipzig University[3]

FACTMATA[2]

@KieselJohannes

Task 4: Hyperpartisan News Detection Background

The left-right political spectrum is a system of classifying political positions, ideologies and parties. Left-wing politics and right-wing politics are often presented as opposed, although either may adopt stances from the other side. [Wikipedia]

A partisan is a politician who strongly supports their party's policies and is reluctant to compromise with political opponents.

The meaning of the term has changed dramatically: "partisan" now refers to an individual with a psychological identification with one of the major parties. [Wikipedia]

Task 4: Hyperpartisan News Detection Is it Fake News?

We see fake news as “disinformation displayed as news articles”

Task 4: Hyperpartisan News Detection Is it Fake News?

We see fake news as “disinformation displayed as news articles”

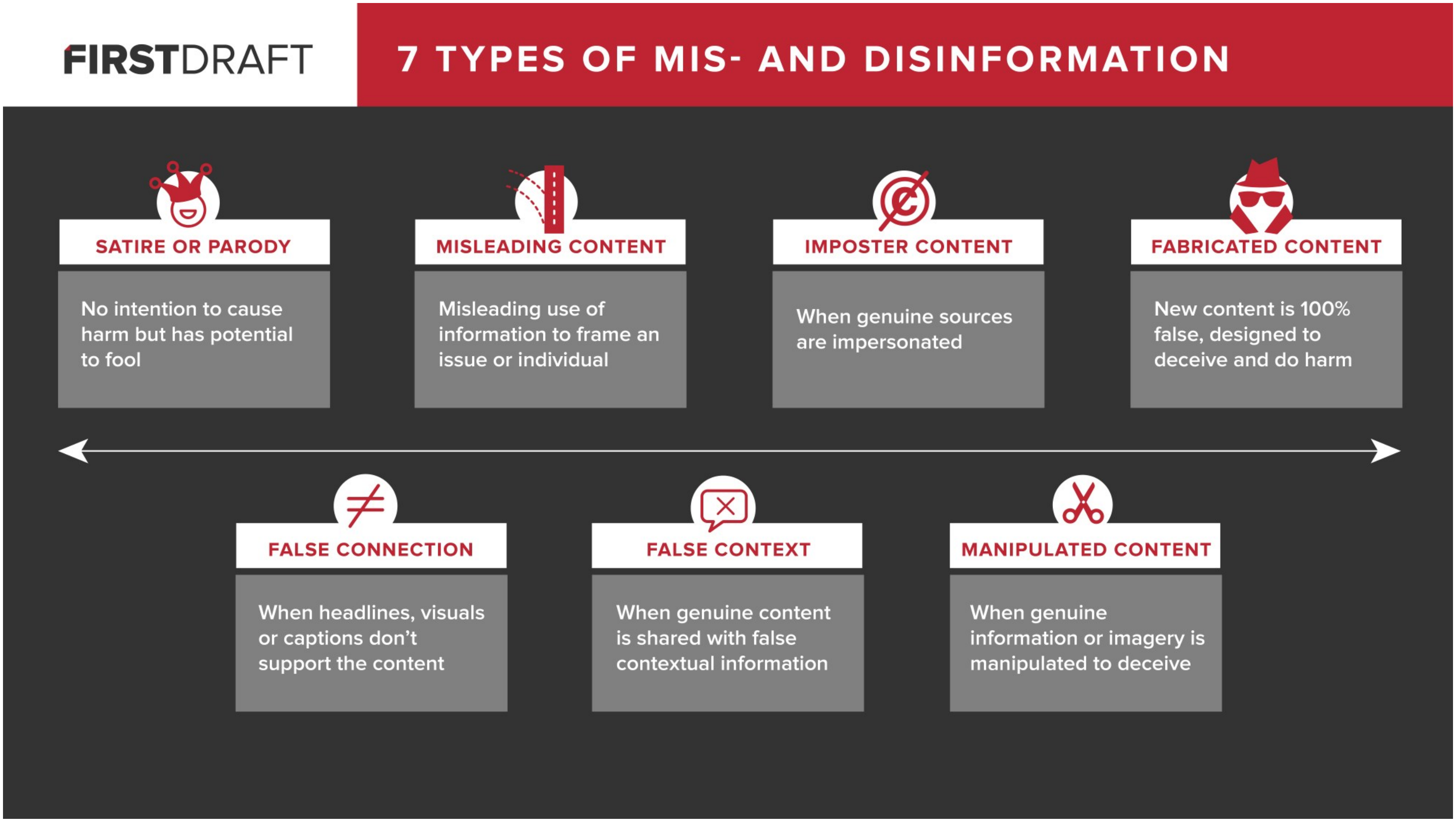


Image: Claire Wardle, First Draft

Task 4: Hyperpartisan News Detection   Is it Fake News?

We see fake news as "disinformation displayed as news articles"

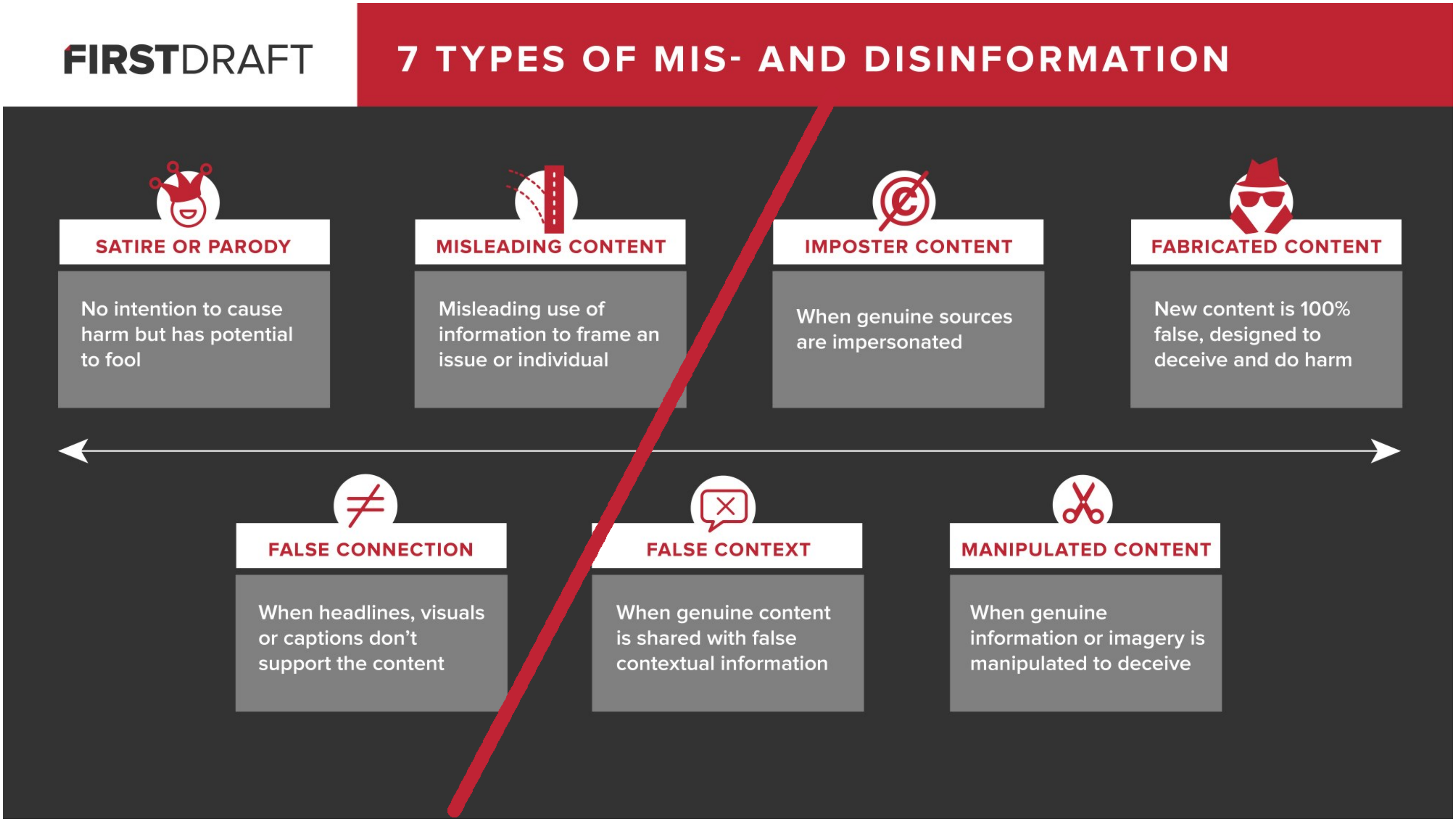


Image: Claire Wardle, First Draft

Task 4: Hyperpartisan News Detection Is it Fake News?

Motivations for mis- and disinformation:

FIRSTDRAFT **MISINFORMATION MATRIX**

	SATIRE OR PARODY	FALSE CONNECTION	MISLEADING CONTENT	FALSE CONTEXT	IMPOSTER CONTENT	MANIPULATED CONTENT	FABRICATED CONTENT
POOR JOURNALISM		✓	✓	✓			
TO PARODY	✓				✓		✓
TO PROVOKE OR TO 'PUNK'					✓	✓	✓
PASSION				✓			
PARTISANSHIP			✓	✓			
PROFIT		✓			✓		✓
POLITICAL INFLUENCE			✓	✓		✓	✓
PROPAGANDA			✓	✓	✓	✓	✓

Image: Claire Wardle, First Draft

Task 4: Hyperpartisan News Detection Is it Fake News?

Motivations for mis- and disinformation: includes partisanship

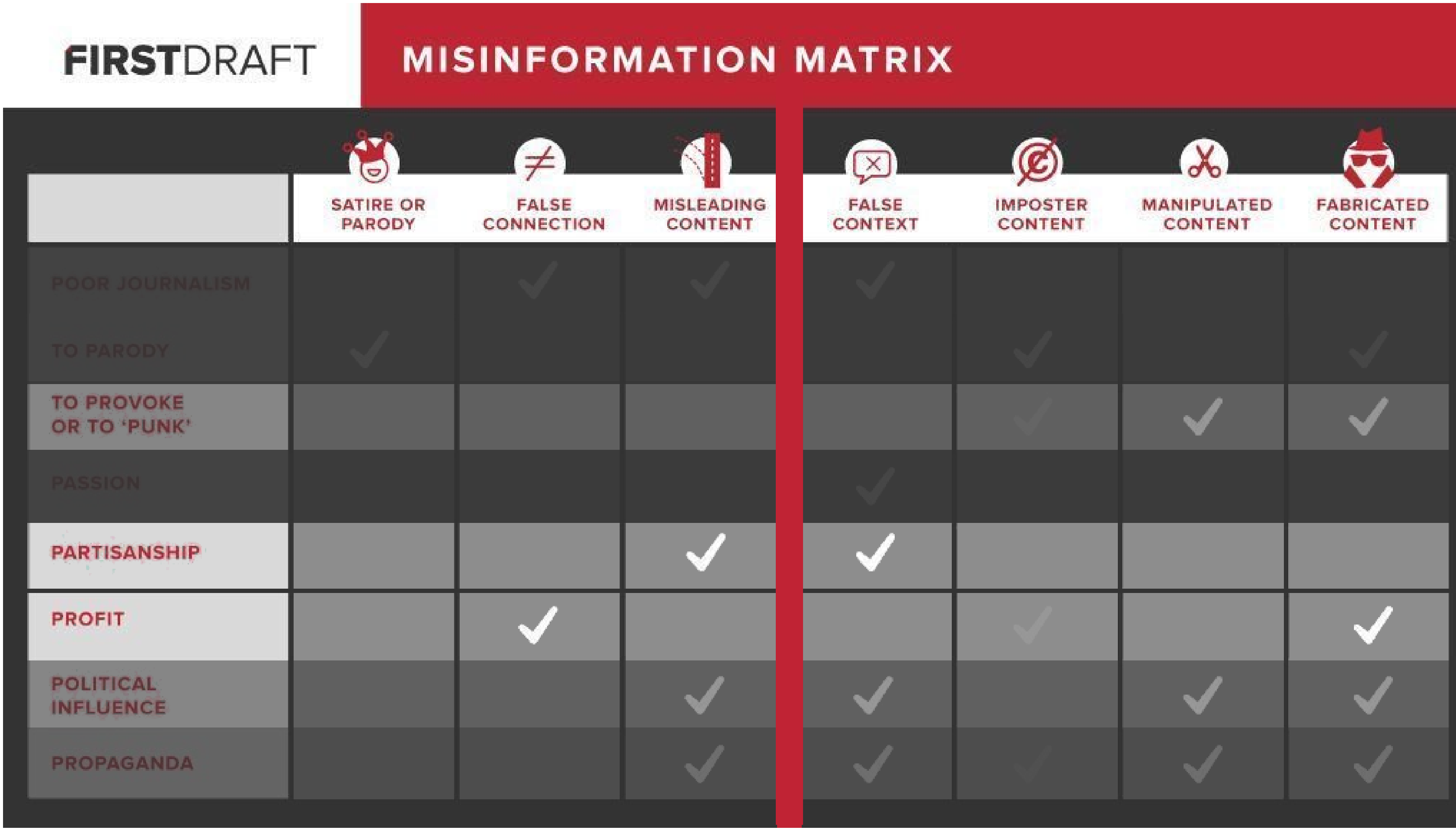
FIRSTDRAFT MISINFORMATION MATRIX

	SATIRE OR PARODY	FALSE CONNECTION	MISLEADING CONTENT	FALSE CONTEXT	IMPOSTER CONTENT	MANIPULATED CONTENT	FABRICATED CONTENT
POOR JOURNALISM		✓	✓	✓			
TO PARODY	✓				✓		✓
TO PROVOKE OR TO 'PUNK'					✓	✓	✓
PASSION				✓			
PARTISANSHIP			✓	✓			
PROFIT		✓			✓		✓
POLITICAL INFLUENCE			✓	✓		✓	✓
PROPAGANDA			✓	✓	✓	✓	✓

Image: Claire Wardle, First Draft

Task 4: Hyperpartisan News Detection   Is it Fake News?

Motivations for publishing hyperpartisan news are not just partisanship



	SATIRE OR PARODY	FALSE CONNECTION	MISLEADING CONTENT	FALSE CONTEXT	IMPOSTER CONTENT	MANIPULATED CONTENT	FABRICATED CONTENT
POOR JOURNALISM		✓	✓	✓			
TO PARODY	✓				✓		✓
TO PROVOKE OR TO 'PUNK'					✓	✓	✓
PASSION				✓			
PARTISANSHIP			✓	✓			
PROFIT		✓			✓		✓
POLITICAL INFLUENCE			✓	✓		✓	✓
PROPAGANDA			✓	✓	✓	✓	✓

Image: Claire Wardle, First Draft

Task 4: Hyperpartisan News Detection Data

Task: Given the text and markup of an online news article, decide whether the article is hyperpartisan or not.

- Dataset Annotated by Article: 1 273 articles.
- Manual annotation of each article by crowdworkers.
  - Articles from ∼500 US news publishers
  - Crowdworker reliability estimate by Beta reputation system (Ismail and Josang 2002)
  - Krippendorff's alpha of employed annotations (3 per article): 0.5
  - Public set: 645 articles; hidden test set: 628 articles, balanced
  - No publisher-overlap between sets

doi.org/10.5281/zenodo.1489920

Task 4: Hyperpartisan News Detection Data

doi.org/10.5281/zenodo.1489920

Task: Given the text and markup of an online news article, decide whether the article is hyperpartisan or not.

- Dataset Annotated by Article: 1 273 articles.
- Manual annotation of each article by crowdworkers.
  - Articles from ∼500 US news publishers
  - Crowdworker reliability estimate by Beta reputation system (Ismail and Josang 2002)
  - Krippendorff's alpha of employed annotations (3 per article): 0.5
  - Public set: 645 articles; hidden test set: 628 articles, balanced
  - No publisher-overlap between sets

- Dataset Annotated by Publisher: 754 000 articles.
- Manual annotation of each publisher by journalists.
  - Annotation of ∼400 US news publishers by BuzzFeed and Media Bias Fact Check
  - Crawling of articles feeds
  - Content wrappers were implemented for each publisher
  - Filtering to political news, English, at least 40 words, correct encoding
  - Public set: 750 000 articles, balanced; hidden test set: 4 000 articles, balanced
  - No publisher-overlap between sets

Task 4: Hyperpartisan News Detection Methods

Employed features

N-Grams character, word, part-of-speech
Embeddings BERT, Word2Vec, fastText, GloVe, ELMo, word clusters, sentences
Stylometry punctuation, structure, readability, lexicons, trigger words
Emotionality sentiment, emotion, subjectivity, polarity
Named entities nationalities, religious and political groups
Quotations count, discarded
Hyperlinks lists of hyperpartisan pages
Publication date year, month

Classifiers

Convolutional neural networks, Long short term memory, Support vector machines, Random Forest, Linear model, Naive Bayes, XGBOOST, Maximum Entropy, Rule-based, ULMFit

Task 4: Hyperpartisan News Detection

Results on dataset annotated by article

Team	Authors	Acc.	Prec.	Rec.	F1.
Bertha von Suttner	**Jiang et al.**	**0.822**	0.871	0.755	0.809
Vernon Fenwick	Srivastava et al.	0.820	0.815	0.828	**0.821**
Sally Smedley	Hanawa et al.	0.809	0.823	0.787	0.805
Tom Jumbo Grumbo	Yeh et al.	0.806	0.858	0.732	0.790
Dick Preston	Isbister and Johansson	0.803	0.793	0.818	0.806
Borat Sagdiyev	Palić et al.	0.791	**0.883**	0.672	0.763
Morbo	Isbister and Johansson	0.790	0.772	0.822	0.796
Howard Beale	Mutlu et al	0.783	0.837	0.704	0.765
Ned Leeds	Stevanoski and Gievska	0.775	0.865	0.653	0.744
Clint Buchanan	Drissi et al.	0.771	0.832	0.678	0.747
+ 32 more					

- 322 registrations
- 184 virtual machines assigned
- 42 software submissions from as many teams
- Ongoing submissions in TIRA

pan.webis.de/semeval19/
semeval19-web/leaderboard.html

Task 4: Hyperpartisan News Detection

Results on meta-learning dataset

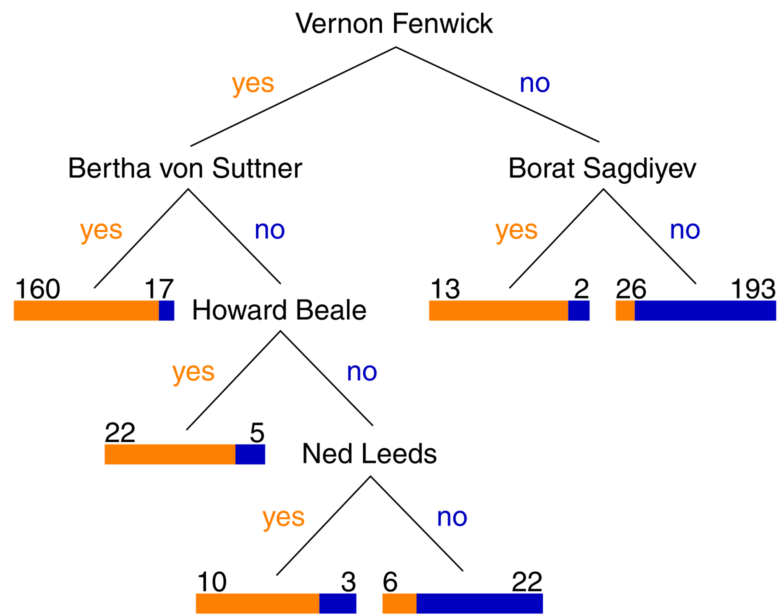
Team	Authors	Acc.	Prec.	Rec.	F1.
Bertha von Suttner alone	Jiang et al.	0.851	0.901	0.788	0.841

- Meta-learning dataset created from test dataset: 66% training, 33% test
- Accuracy higher (from 0.822)

Task 4: Hyperpartisan News Detection   Results on meta-learning dataset

Team	Authors	Acc.	Prec.	Rec.	F1.
Majority Vote	Kiesel et al.	0.885	0.892	0.875	0.883
J48-M10	Kiesel et al.	0.880	**0.916**	0.837	0.874
Bertha von Suttner alone	Jiang et al.	0.851	0.901	0.788	0.841

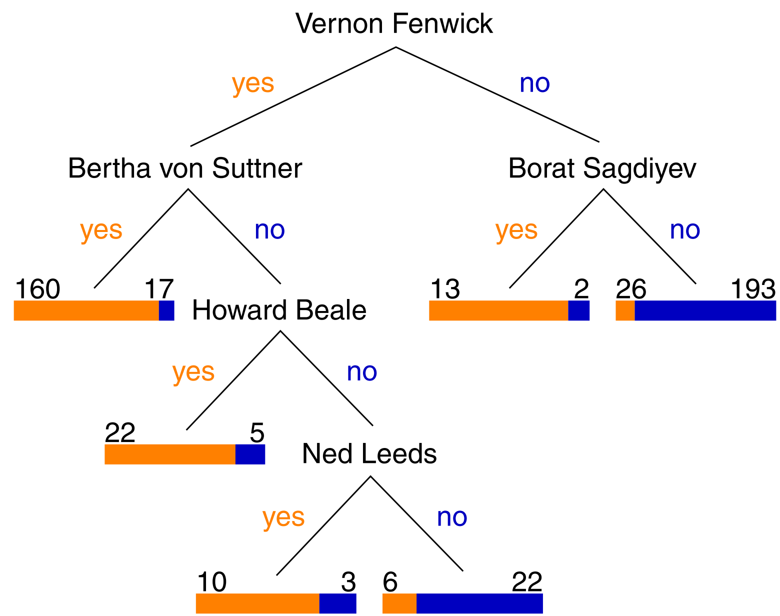
- Meta-learning dataset created from test dataset: 66% training, 33% test
- Accuracy higher (from 0.822)
- Baselines beat best single system



Task 4: Hyperpartisan News Detection Results on meta-learning dataset

Team	Authors	Acc.	Prec.	Rec.	F1.
Fernando Pessa	**Cruz et al.**	**0.899**	0.895	**0.904**	**0.900**
Spider Jerusalem	**Alabdulkarim and Alhindi**	**0.899**	0.903	0.894	0.899
Majority Vote	Kiesel et al.	0.885	0.892	0.875	0.883
J48-M10	Kiesel et al.	0.880	**0.916**	0.837	0.874
Bertha von Suttner alone	Jiang et al.	0.851	0.901	0.788	0.841

- Meta-learning dataset created from test dataset: 66% training, 33% test
- Accuracy higher (from 0.822)
- Baselines beat best single system
- Both participants beat the baselines
- They use a Random Forest and a weighted majority vote, respectively
- Ongoing submissions in TIRA



Task 4: Hyperpartisan News Detection

Results on dataset annotated by publisher

Team	Authors	Acc.	Prec.	Rec.	F1.
Tintin	**Bestgen**	**0.706**	0.742	0.632	0.683
Joseph Rouletabille	Moreno et al.	0.680	0.640	0.827	**0.721**
Brenda Starr	Papadopoulou et al.	0.664	0.627	0.807	0.706
Xenophilius Lovegood	Zehe et al.	0.663	0.632	0.781	0.699
Yeon Zi	Lee et al.	0.663	0.635	0.766	0.694
Miles Clarkson	Zhang et al.	0.652	0.612	0.832	0.705
Jack Ryder	Shaprin et al.	0.645	0.600	**0.869**	0.710
Bertha von Suttner	Jiang et al.	0.643	0.616	0.762	0.681
+ 16 more					
Robin Scherbatsky	Marx and Akut	0.524	**0.822**	0.062	0.116
+ 3 more					

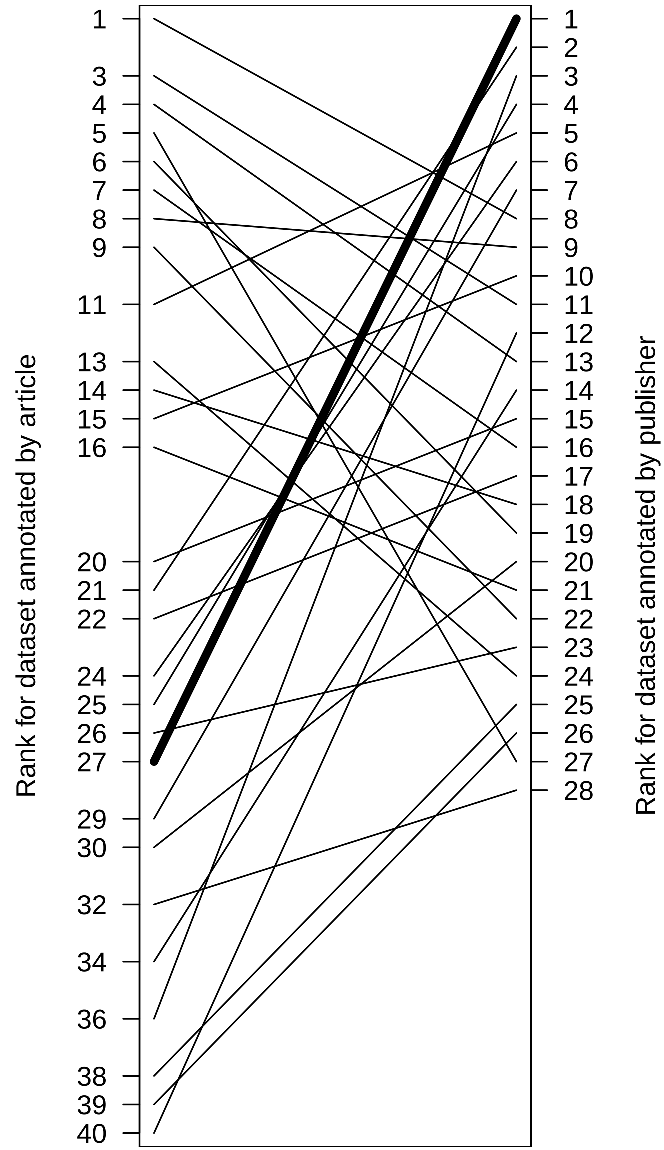
- 28 teams (of 42)
- Accuracy lower (from 0.822)
- Most teams focused on the other dataset
- Ranking very different
- Ongoing submissions in TIRA

pan.webis.de/semeval19/semeval19-web/leaderboard.html

Task 4: Hyperpartisan News Detection   Comparison of dataset rankings

Team	Authors
Tintin	**Bestgen**
Joseph Rouletabille	Moreno et al.
Brenda Starr	Papadopoulou et al.
Xenophilius Lovegood	Zehe et al.
Yeon Zi	Lee et al.
Miles Clarkson	Zhang et al.
Jack Ryder	Shaprin et al.
Bertha von Suttner	Jiang et al.
+ 16 more	
Robin Scherbatsky	Marx and Akut
+ 3 more	

- 28 teams (of 42)
- Accuracy lower (from 0.822)
- Most teams focused on the other dataset
- Ranking very different
- Ongoing submissions in TIRA



Task 4: Hyperpartisan News Detection Conclusion

- Two datasets, newest version downloaded ∼450 times
- Features reported to be especially efficient: embeddings, n-grams, sentiment
- Detailed analysis of hand-crafted features: Borat Sagdiyev
- So far, 10 teams released their code open source
- Very high accuracy: 0.8 to 0.9
- Submission still open!
- Challenge ahead: explainability

Task 4: Hyperpartisan News Detection Conclusion

- Two datasets, newest version downloaded ∼450 times
- Features reported to be especially efficient: embeddings, n-grams, sentiment
- Detailed analysis of hand-crafted features: Borat Sagdiyev
- So far, 10 teams released their code open source
- Very high accuracy: 0.8 to 0.9
- Submission still open!
- Challenge ahead: explainability

Task 4: Hyperpartisan News Detection Conclusion

- Two datasets, newest version downloaded $\sim$450 times
- Features reported to be especially efficient: embeddings, n-grams, sentiment
- Detailed analysis of hand-crafted features: Borat Sagdiyev
- So far, 10 teams released their code open source
- Very high accuracy: 0.8 to 0.9
- Submission still open!
- Challenge ahead: explainability

Task 4: Hyperpartisan News Detection Conclusion

- Two datasets, newest version downloaded $\sim$450 times

- Features reported to be especially efficient: embeddings, n-grams, sentiment
- Detailed analysis of hand-crafted features: Borat Sagdiyev
- So far, 10 teams released their code open source

- Very high accuracy: 0.8 to 0.9
- Submission still open!

- Challenge ahead: explainability