CLEF 2020 Thessaloniki

PAN

8th Author Profiling task at PAN

Profiling Fake News Spreaders on Twitter

PAN-AP-2020 CLEF 2020
Online, 22-25 September

Francisco Rangel
Symanto Research

Anastasia Giachanou
PRHLT Research Center
Universitat Politècnica de Valencia

Bilal Ghanem
Symanto Research

Paolo Rosso
PRHLT Research Center
Universitat Politècnica de Valencia

Introduction

Author profiling aims at identifying **personal traits** such as age, gender, personality traits, native language, language variety… from writings?

This is crucial for:

- Marketing.
- Security.
- Forensics.

Task goal

Given a Twitter feed, determine whether its author is **keen to spread fake news or not**.

Two languages:

English **Spanish**

Corpus

Methodology

1. Selection of fake news from Politifact and Snopes related sites (+ manual review).
2. Collection of tweets responding to the previous news:
    2.1. Manual inspection to ensure that the tweet refers to the news.
    2.2. Manual annotation of those tweets supporting vs. rejecting the news.
3. Timeline collection
    3.1. Manual review of the tweets to label the fake ones.
    3.2. Users with one of more fake tweets are keen to spread them. Otherwise, they are not.
    3.3. Removal of tweets referring explicitly to the fake news (to avoid bias).

	(EN) English			(ES) Spanish		
	Keen to spread fake news	**Not keen to spread fake news**	**Total**	**Keen to spread fake news**	**Not keen to spread fake news**	**Total**
Training	150	150	300	150	150	300
Test	100	100	200	100	100	200
Total	250	250	500	250	250	500

Evaluation measures

The **accuracy** is calculated per language and averaged:

$$ranking = \frac{acc_{en} + acc_{es}}{2}$$

Baselines

RANDOM	A baseline that randomly generates the predictions among the different classes
LSTM	An Long Short-Term Memory neural network that uses FastTex embeddings to represent texts.
CHAR N-GRAMS	With values for n from 2 to 6, with a SVM
WORD N-GRAMS	With values for n from 1 to 3, with a Neural Network
EIN	The Emotionally-Infused Neural (EIN) network with word embedding and emotional features as the input of an LSTM
Symanto (LDSE)	This method represents documents on the basis of the probability distribution of occurrence of their words in the different classes. The key concept of LDSE is a weight, representing the probability of a term to belong to one of the different categories: fake news spreaders / non-spreader. The distribution of weights for a given document should be closer to the weights of its corresponding category. LDSE takes advantage of the whole vocabulary

Participation

66 participants
33 working notes
22 countries

https://mapchart.net/world.html

Approaches

What kind of ...

Preprocessing

Features

Methods

... did the teams perform?

Approaches - Preprocessing

Twitter elements (RT, VIA, FAV)	Giglou; Hashemi; Pinnaparaju
Emojis and other non-alphanumeric chars	Buda; Pinnaparaju; Vogel; Giglou; Espinosa; Majumder; Lichouri; Shashirekha
Lemmatisation	Giglou; Hashemi; Lichouri; Shashirekha
Tokenisation	Vogel; Labadie; Fernández; Espinosa; Lichouri; Shashirekha; Baruah
Punctuation signs	Vogel; Koloski; Giglou; Espinosa; Hashemi; Lichouri; Shashirekha
Numbers	Pizarro; Vogel; Giglou; Espinosa; Hashemi; Shashirekha
Lowercase	Buda; Pizarro; Vogel; Pinnaparaju
Stopwords	Vogel; Koloski; Giglou; Espinosa; Hashemi; Lichouri; Shashirekha
Character flooding	Vogel; Labadie
Infrequent terms	Ikade
Short texts	Vogel

Approaches - Features

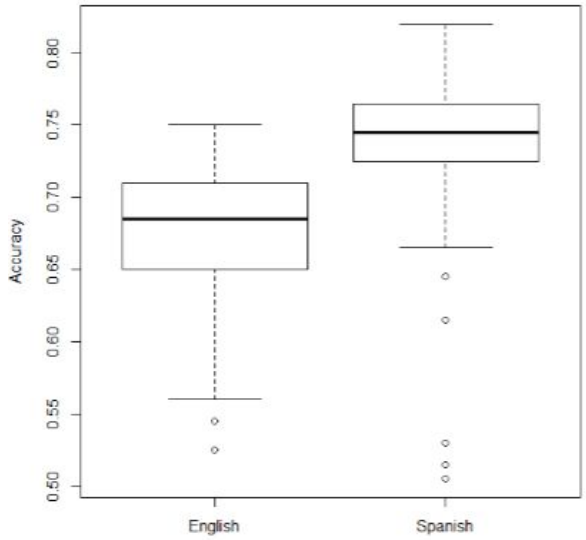
Stylistic features: - Number of occurrences - Verbs, adjs, pronouns - Number of hashtags, mentions, URLs... - Capital vs. lower letters - Punctuation marks - ...	Manna; Buda; Lichouri; Justin; Niven; Russo; Hörtenhuemer; Cardaioli; Spezanno; Ogaltsov; Labadie; Hashemi; Moreno-Sandoval;
N-gram models	Pizarro; Espinosa; Vogel; Koloski; López-Fernández; Vijayasaradhi; Buda; Lichouri; Justin; Hörtenhuemer; Spezanno; Aguirrezabal; Shashirekha; Babaei; Labadie; Hashemi;
Emotional and personality features	Justin; Niven; Russo; Hörtenhuemer; Espinosa; Cardaioli; Spezanno; Moreno-Sandoval;
Embeddings	Justin; Hörtenhuemer; Aguirrezabal; Ogaltsov; Shashirekha; Babaei; Labadie; Hashemi; Cilet; Majumder;
...BERT	Spezanno; Kaushik; Baruah; Chien;

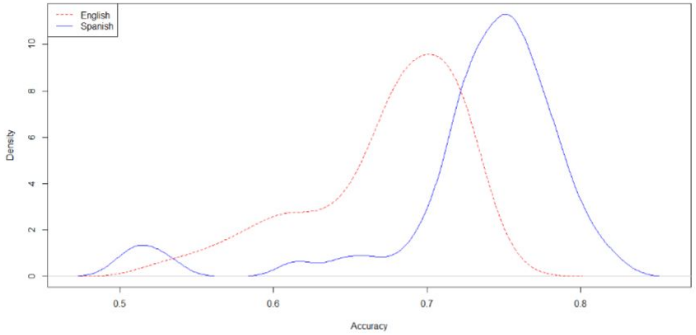
* 9 teams have used Symanto API to obtain psycholinguistic and/or emotional features

Approaches - Methods

SVM	Pizarro; Vogel; Koloski; Espinosa; Fernández; Hashemi; Lichouri; Aguirrezabal; Fersini
Logistic regression	Buda; Vogel; Koloski; Hörtennhuemer; Pinnaparaju; Aguirrezabal; Manna
Random Forest	Cardaioli; Espinosa; Hashemi; Aguirrezabal; Sandoval; Manna
Ensembles	Ikade; Shrestha; Shashirekha; Niven
Multilayer Perceptron	Aguerrizabal
NN with Dense Layer	Baruah
Fully-Connected NN	Giglou
CNN	Chilet
LSTM	Majumder; Labadie
bi-LSTM	Saeed
Ensemble (GRU + CNN)	Bakhteev

Global ranking





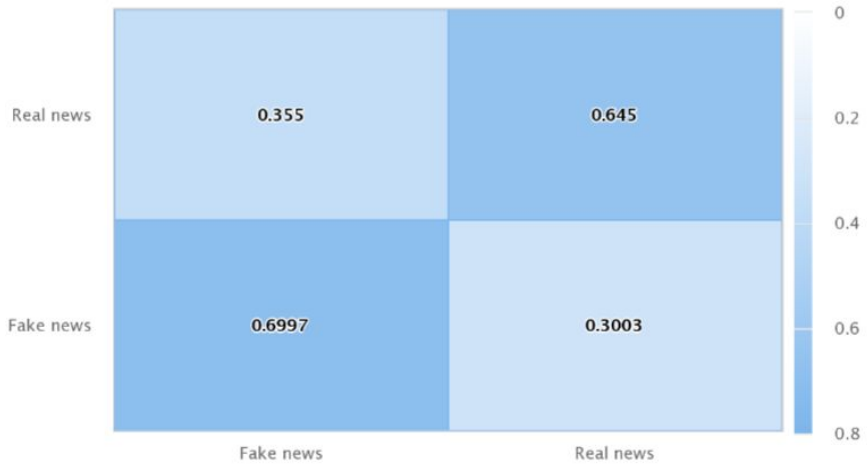
	PARTICIPANT	EN	ES	AVG
1	bolonyai20	**0.750**	0.805	0.7775
1	pizarro20	0.735	**0.820**	0.7775
	SYMANTO (LDSE)	*0.745*	*0.790*	*0.7675*
3	koloski20	0.715	0.795	0.7550
3	deborjavalero20	0.730	0.780	0.7550
3	vogel20	0.725	0.785	0.7550
6	higueraporras20	0.725	0.775	0.7500
6	tarela20	0.725	0.775	0.7500
8	babaei20	0.725	0.765	0.7450
9	staykovski20	0.705	0.775	0.7400
9	hashemi20	0.695	0.785	0.7400
11	estevecasademunt20	0.710	0.765	0.7375
12	castellanospellecer20	0.710	0.760	0.7350
	SVM + c nGrams	*0.680*	*0.790*	*0.7350*
13	shrestha20	0.710	0.755	0.7325
13	tommasel20	0.690	0.775	0.7325
15	johansson20	0.720	0.735	0.7275
15	murauer20	0.685	0.770	0.7275
17	espinosagonzales20	0.690	0.760	0.7250
17	ikae20	0.725	0.725	0.7250
19	morenosandoval20	0.715	0.730	0.7225
20	majumder20	0.640	0.800	0.7200
20	sanchezromero20	0.685	0.755	0.7200
22	lopezchilet20	0.680	0.755	0.7175
22	nadalalmela20	0.680	0.755	0.7175
22	carrodve20	0.710	0.725	0.7175
25	gil20	0.695	0.735	0.7150
26	elexpuruortiz20	0.680	0.745	0.7125
26	labadietamayo20	0.705	0.720	0.7125
28	grafiaperez20	0.675	0.745	0.7100
28	jilka20	0.665	0.755	0.7100
28	lopezfernandez20	0.685	0.735	0.7100
31	pinnaparaju20	0.715	0.700	0.7075
31	aguirrezabal20	0.690	0.725	0.7075
33	kengyi20	0.655	0.755	0.7050
33	gowda20	0.675	0.735	0.7050
33	jakers20	0.675	0.735	0.7050
33	cosin20	0.705	0.705	0.7050

	Participant	En	Es	Avg
37	navarromartinez20	0.660	0.745	0.7025
38	heilmann20	0.655	0.745	0.7000
39	cardaioli20	0.675	0.715	0.6950
39	females20	0.605	0.785	0.6950
39	kaushikamardas20	0.700	0.690	0.6950
	NN + w nGrams	*0.690*	*0.700*	*0.6950*
42	monteroceballos20	0.630	0.745	0.6875
43	ogaltsov20	0.695	0.665	0.6800
44	botticebria20	0.625	0.720	0.6725
45	lichouri20	0.585	0.760	0.6725
46	manna20	0.595	0.725	0.6600
47	fersini20	0.600	0.715	0.6575
48	jardon20	0.545	0.750	0.6475
	EIN	*0.640*	*0.640*	*0.6400*
49	shashirekha20	0.620	0.645	0.6325
50	datatontos20	0.725	0.530	0.6275
51	soleramo20	0.610	0.615	0.6125
	LSTM	*0.560*	*0.600*	*0.5800*
52	russo20	0.580	0.515	0.5475
53	igualadamoraga20	0.525	0.505	0.5150
	RANDOM	*0.510*	*0.500*	*0.5050*

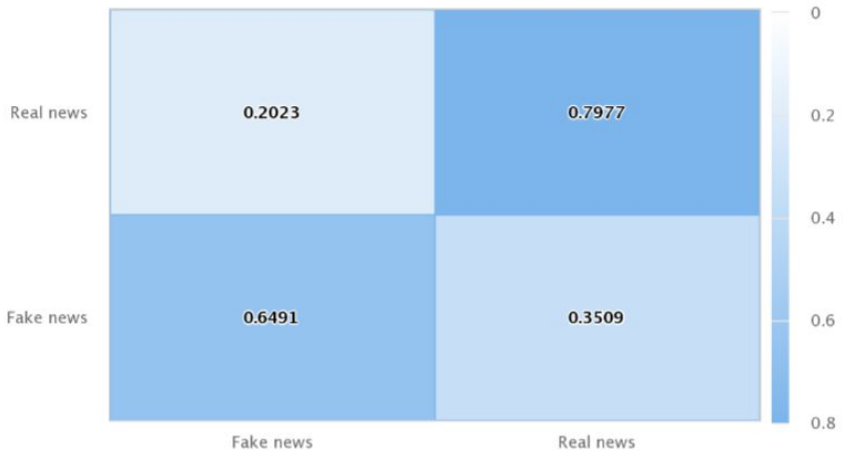
	Participant	En
54	hoertenhuemer20	0.725
55	duan20	0.720
55	andmangenix20	0.720
57	saeed20	0.700
58	baruah20	0.690
59	anthonio20	0.685
60	zhang20	0.670
61	espinosaruiz20	0.665
62	shen20	0.650
63	suareztrashorras20	0.640
64	niven20	0.610
65	margoes20	0.570
66	wu20	0.560

Confusion matrices

ENGLISH



SPANISH



Best results at PAN'20

Buda and Bolonyai
- n-Grams
- Stylistic features
- Logistic Regression ensemble

Pizarro
- word and char n-grams
- SVM

English		Spanish	
Buda and Bolonyai [9]	(0.750)	Pizarro [45]	(0.820)

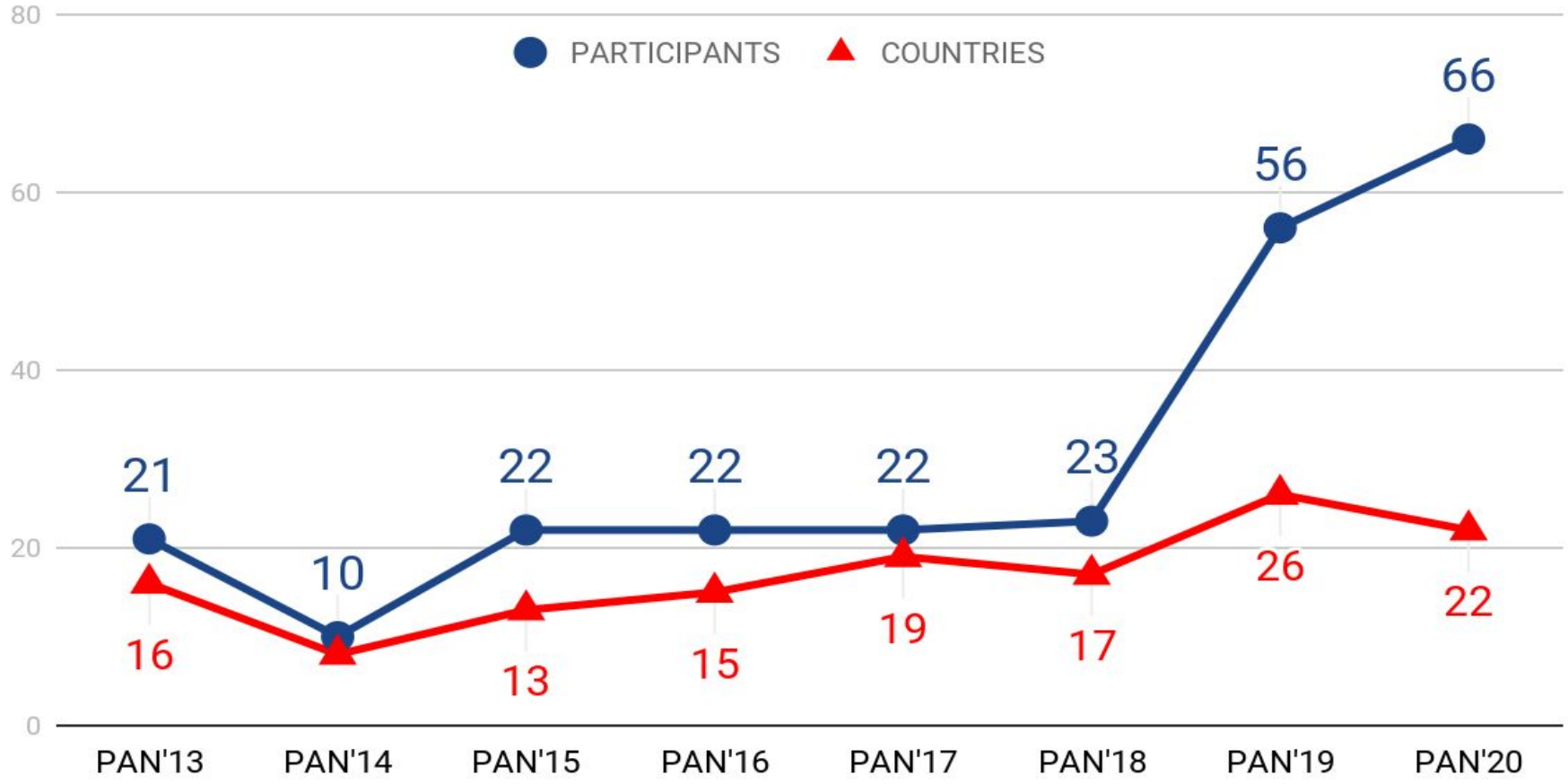
Conclusions

- Several approaches to tackle the task:
  - n-Grams + SVM prevailing.
- Best results in English:
  - Over 67% on average.
  - Best (75%): Buda and Bolonyai - n-Grams + Stylistic features + Logistic Regression ensemble
- Best results in Spanish:
  - Over 73% on average.
  - Best (82%): Pizarro - char & word n-Grams + SVM.
- Error analysis:
  - English:
    - False positives (real news spreaders as fake news spreaders): 35.50%
    - False negatives (fake news spreaders as real news spreaders): 30.03%
  - Spanish:
    - False positives (real news spreaders as fake news spreaders): 20.23%
    - False negatives (fake news spreaders as real news spreaders): 35.09%

Looking at the results, we can conclude:

- It is feasible to automatically identify Fake News Spreaders with high precision
  - ...even when only textual features are used.
- We have to bear in mind false positives since especially in English, they sum up to one-third of the total predictions, and misclassification might lead to ethical or legal implications.

Task Impact


80
60
40
20
0
PARTICIPANTS
COUNTRIES
21
10
22
22
22
23
56
66
16
13
15
19
17
26
22
PAN'13
PAN'14
PAN'15
PAN'16
PAN'17
PAN'18
PAN'19
PAN'20

Industry at PAN (Author Profiling)

Organisation

symanto
psychology ai

Sponsors

symanto
psychology ai

This year, the winners of the task are (ex aequo):

- Jakab Buda and Flora Bolonyai, Eötvös Loránd University, Hungary
- Juan Pizarro, Chile

2021 -> HATE speech spreadeRS

On behalf of the author profiling task organisers:

Thank you very much for participating
and hope to see you next year!!