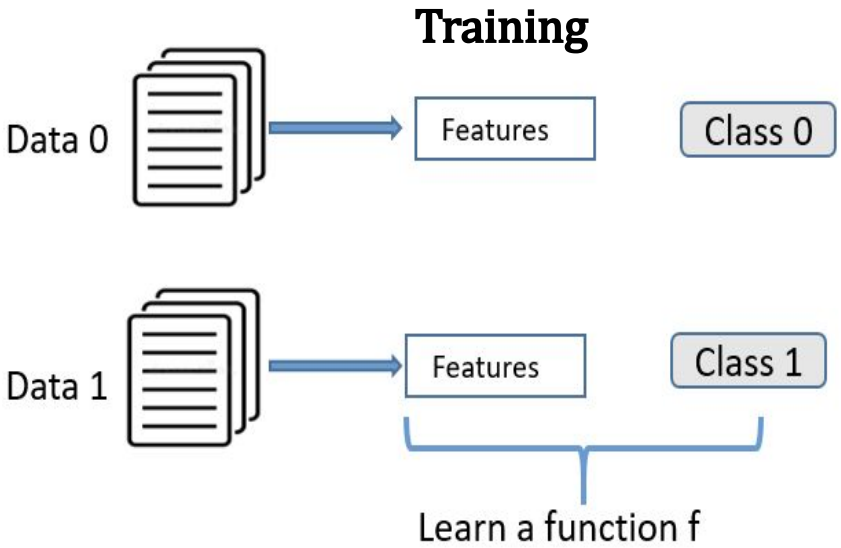
UniNE at PAN-CLEF 2020: Profiling Fake News Spreaders on Twitter

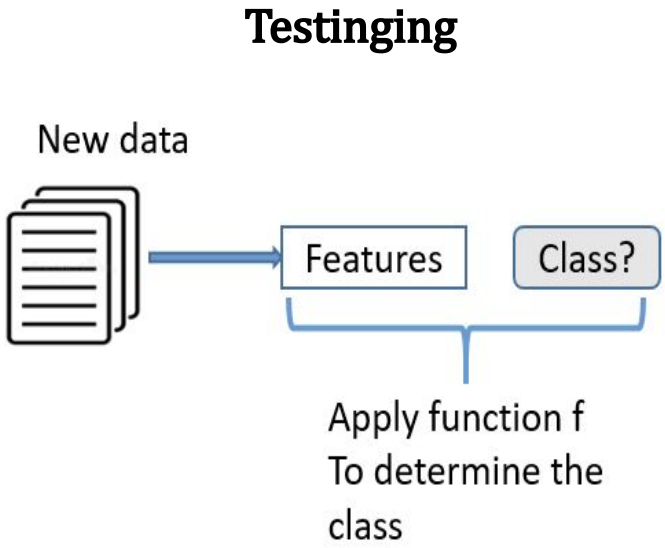
Catherine Ikae, Jacques Savoy
University of Neuchatel, Switzerland

The Task

Task: Given a Twitter feed, determine whether its author is keen to be a spreader of fake news.

- Languages: English and Spanish
- Genres: Twitter feeds



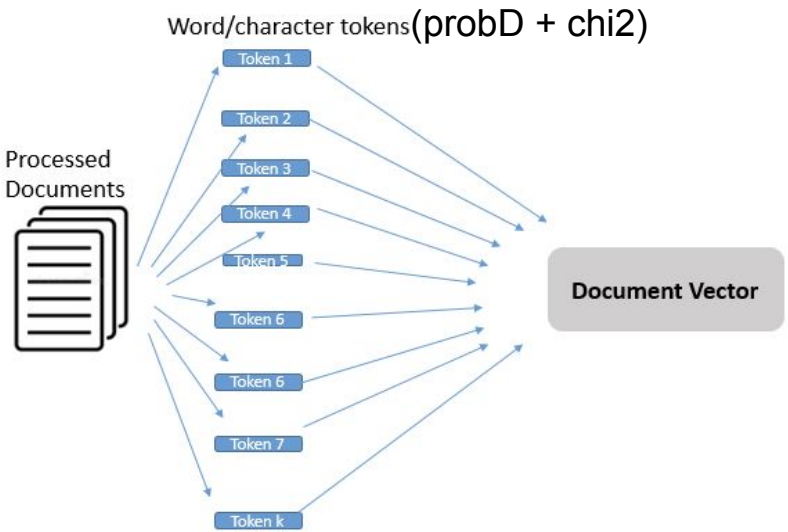


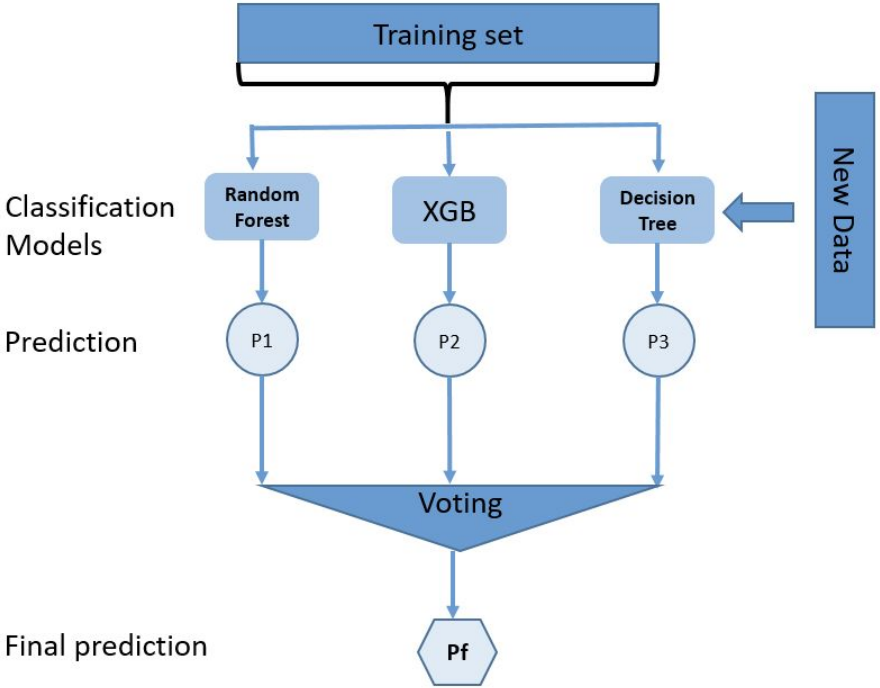
The Method

$$probD(t_i, c_0) = prob(t_i, c_0) - prob(t_i, c_1) = \frac{tf_{i0}}{n_0} - \frac{tf_{i1}}{n_1}$$

- unique vocabulary (VocUnC1, VocUnC2) belonging to the two categories (PtC1, PtC2) is determined by *probD*
- chi-square method was selected to reduce the feature space to a few hundred terms
- The documents belonging to two Categories are represented as vectors using a reduced set of features
- A classifier is implemented combining decision tree, random forest, and boosting as shown below

The Method


Word/character tokens(probD + chi2)
Processed Documents
Token 1
Token 2
Token 3
Token 4
Token 5
Token 6
Token 6
Token 7
Token k
Document Vector


Training set
Classification Models
Random Forest
XGB
Decision Tree
New Data
Prediction
P1
P2
P3
Voting
Final prediction
Pf

Evaluation with features ranked by chi2 values

	chi 100	chi 150	chi 200	chi 250	chi 300
Random Forest	0.77	0.76	0.77	0.71	0.79
Boosting	0.71	0.74	0.72	0.66	0.70
Decision Tree	0.68	0.72	0.63	0.64	0.63
Soft Voting	0.70	0.72	0.78	0.66	0.71
Majority Voting	0.72	0.81	0.78	0.72	0.75

TIRA results and conclusion

TIRA test set		
Fusion	English	Spanish
Soft Voting	0.675	0.700
Majority Voting	0.725	0.725

- Our feature selection technique is able to extract a reduced set of features upto 150.
- it is possible to identify those features more associated to normal tweets (e.g., I, this, film, review, episode, etc.) from tweets spreading fake news, (names of political leaders, says, post, president, she, he, democrat, etc)
- our analysis indicates that tweets containing fake news tend to include more references (URL) to other webpages than normal tweets, references used to support the misinformation or to justify some conspiracy theory. On the other hand, normal tweets present more retweets and hashtags.
- our attribution approach is based on a model combining three individual attributions computed by a decision tree, a boosting, and a random forest classifier