

Fake news detection for Romanian

Parfene Ioana

March 31, 2021

1 Introduction

1.1 Motivation

2 Method

2.1 Gathering Data

2.1.1 Web Sites

2.1.2 Crawling

2.1.3 Scraping

2.2 Preprocessing

2.2.1 Stop Words

2.2.2 Tokenization

2.2.3 Stemming

2.3 Data Models

2.3.1 Bag of Words

2.3.2 TF-IDF

2.4 Classifiers

2.4.1 Naive Bayes

Words		200				1000				All		
Model	BoW	BoW	T-I	T-I	BoW	BoW	T-I	T-I	BoW	BoW	T-I	T-I
Test	10%	20%	10%	20%	10%	20%	10%	20%	10%	20%	10%	20%
Result	82.5%	82.9%	82.6%	82.9%	90.3%	89.7%	87.4%	86.6%	93.3%	93.3%	73.0%	71.8%
Time	0.01s	0.01s	0.01s	0.01s	0.09s	0.08s	0.05s	0.06s	1.76s	1.67s	0.98s	1.08s

Figure 1: **Naive Bayes**, Accuracy Table, **Average** out of **100** runs

2.4.2 Passive Aggressive Classifier

Words		200				1000				All		
Model	BoW	BoW	T-I	T-I	BoW	BoW	T-I	T-I	BoW	BoW	T-I	T-I
Test	10%	20%	10%	20%	10%	20%	10%	20%	10%	20%	10%	20%
Result	83.0%	83.4%	84.3%	84.5%	93.1%	93.3%	93.5%	92.6%	94.3%	93.8%	94.2%	94.0%
Time	0.03s	0.03s	0.03s	0.03s	0.13s	0.13s	0.17s	0.15s	2.50s	2.3s	2.79s	2.63s

Figure 2: **Passive Aggressive Classifier**, Accuracy Table, **Average** out of **100** runs

3 Results

4 Conclusion

References