Application of Machine Learning Techniques for Fake News Classification

Kim Leone Souza da Silva[1], Paulo Canas Rodrigues[1] e Crysttian Arantes Paixão[1]
[1] University Federal of Bahia

Abbreviated abstract: Fake News consists of disseminating fake news in various social and digital media such as newspapers, television networks, and the internet. This causes serious problems, and its impacts can be felt in the real world. Thus, the present work aims to propose and evaluate strategies for processing and applying machine learning models, to improve the performance of classifiers in the problem of identifying Fake News in Brazilian news.

Related publications:

– MONTEIRO, Rafael A. et al, International Conference on Computational Processing of the Portuguese Language, (2018)

Problem, Data, Previous Works

Fake News consists of disseminating fake news in various social and digital media such as newspapers, television networks, and the internet. The main objective of this work is to propose and evaluate strategies for processing and applying machine learning models in order to improve the performance of classifiers in the problem of identifying Fake News.

The database used was Fake.br-Corpus and is available on Github (https://github.com/roneysco/Fake.br-Corpus)

•The database was built in 2018 by Monteiro;

•Contains 7,200 news from different segments, such as entertainment, politics, economy and religion;

•News was collected from January 2016 to January 2018

Methods

- Data processing

- Construction of the terms matrix;
  - Bag of words (binary, frequency, tf-idf);
  - N-grams (Unigrams, Bigrams, Trigrams);
  - There were 8 variables analyzed;
    - These variables were obtained by OpLexiconv3.0 from the lexiconPT package of the software R;

- Selection of variables (Stepwise Forward);

- Applying methods (k-NN, Naïve Bayes, SVM, Random Forest);

Results and Conclusions

The table show the results

- The best method was the binary unigram term matrix with the addition of the variable adj. The variable adj is the number of adjectives in the text.

The application of machine learning models achieved better performance when using a binary unigrams compared to other methods and the model with the highest performance was the SVM, with 96,81% accuracy.

Table 1: Performance of the models applied to each term matrix

	SVM		KNN		Naive-Bayes		Random Forest	
	ACC	F1-Score	ACC	F1-Score	ACC	F1-Score	ACC	F1-Score
Uni-binary	0,9644	0,9654	0,6519	0,4462	0,9431	0,9503	0,9537	0,9532
+cov	0,9681	0,9683	0,9398	0,9394	0,9495	0,9503	0,9593	0,9593
Uni-freq	0,95926	0,95861	0,7778	0,7273	0,9542	0,9517	0,9523	0,9520
+cov	-	-	0,9426	0,9425	-	-	0,9620	0,9627
Uni-tf-idf	0,9611	0,9610	0,7111	0,5948	0,9481	0,9459	0,9560	0,9560
+cov	0,9625	0,9628	0,9417	0,9416	0,9551	0,9552	0,9611	0,9618
Bi-binary	0,9343	0,9319	0,5511	0,1925	0,9097	0,9088	0,9181	0,9150
+cov	0,9528	0,9526	0,9380	0,9367	0,9204	0,9201	0,9556	0,9550
Bi-freq	0,9333	0,9310	0,6532	0,4922	0,8907	0,8845	0,9218	0,9202
+cov	0,9537	0,9536	0,9417	0,9413	0,9019	0,8980	0,9648	0,9635
Bi-tf-idf	0,937037	0,9342	0,6440	0,4341	0,8931	0,8858	0,9106	0,9042
+cov	0,9491	0,9473	0,9315	0,9304	0,9032	0,8957	0,9569	0,9578
Tri-binary	0,8486	0,8348	0,7014	0,5841	0,6583	0,4766	0,8356	0,8356
+cov	0,9292	0,9272	0,9407	0,9388	0,6750	0,5270	0,9495	0,9502
Tri-freq	0,8426	0,8281	0,7005	0,6187	0,6343	0,4317	0,8444	0,8392
+cov	0,9403	0,9357	0,9417	0,9401	0,6565	0,4638	0,9574	0,9562
Tri-tf-idf	0,8495	0,8296	0,6806	0,5751	0,5903	0,3249	0,8394	0,8321
+cov	0,9361	0,9358	0,9300	0,9285	0,6218	0,4080	0,9583	0,9582