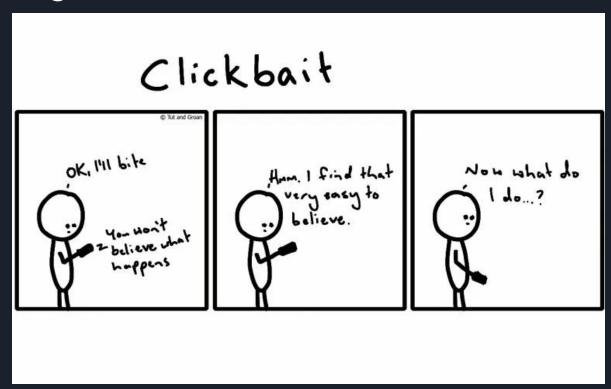
Text Classification -

or - Let's find patterns! Does hype and hyper trash leave a signature?

Laura Fedoruk - P.Eng. | M.Sc.

The problem: promoting media literacy vs. feeding the click-bait monster



HOW? Can we classify news vs. tabloids vs. political trash?

CUSTOM STOP WORDS.





UPDATED

Influx at the border needs not just compassion but immediate action: Scheer

Compassion must greet the asylum seekers coming across the Canada-U.S. border once they arrive, but more must be done to stop them from coming in the first place, Opposition leader Andrew Scheer says in proposing his own solutions to the problem.

Stephanie Levitz

Politics | 3 Hours Ago



CBC



AUGUST 22, 2017

UPDATE: Is illegal mosque complying with city's order to close its doors?

By David Menzies



N-GRAM REPETITION

YES WE CAN!!! JUST IN! 95% TRAINING SET ACCURACY!

SENTENCE LENGTH





It's Over

LIAR, LIAR? ELON MUSK DUMPS AMBER HEARD OVER FIBS

'He caught her in countless lies!'

What is the signature of journalism?

- Proper Noun (NNP), possessive ending (POS) "Laura's view"
- Cardinal Numbers (CN), Proper Noun (NNP) "300 Canadian Geese"
- Past Participle Verbs (VBN), Preposition (IN) "spoken of"
- Neutral headline sentiment
- Objective headlines

Tabloids!?

- !!!
- Relatively negative headline sentiment
- Short, multi-sentence headlines
- Preposition (IN), Personal Pronoun (PRP) "In her"
- Determiner (DT), Common Noun (NNS) "the book"

Opinion articles and rhetoric???

- ?
- Determiner DT, Proper Noun NNP "The Conservative Party"
- Wh-Determiner WDT "whatever"
- Wh-Pronoun WP "who"
- Adjective JJ, Common Noun NN, Coordinating Conjunction CC "good thought yet"

But where these articles are showing up is an ever evolving landscape...



Hey, America, Now You Have Our Worst People. You're Welcome.

Canada's ultraconservative Rebel Media may be falling apart, but America's alt-right movement could provide a warm welcome to its alumni.

Posted on August 19, 2017, at 5:48 a.m.



There's an app for that! ...

Finding	the Signature of Article Text
Enter exam	ple headline text
Enter exam	ple article body text

Media Literacy + Al Assistance

A few things to consider when evaluating the content you read online:

- Who is the originator of the content and when was the website created?
 Many exploitative news sites continually change domain names and have short histories.
- Are sources used and are they reputable?
 Tabloids and opinion based rhetoric tends to rely on unnamed sources with little backup or reference.
- Are opposing sources quoted when content is controversial?
 A tradition in journalism is to present opposing viewpoints when opinions or ideas are in question.
- Has the article been updated since it was released?
 Ethical journalism practice requires updating mistakes if they are discovered. A log of changes can give insight into the article and the standard of practice behind a publication.