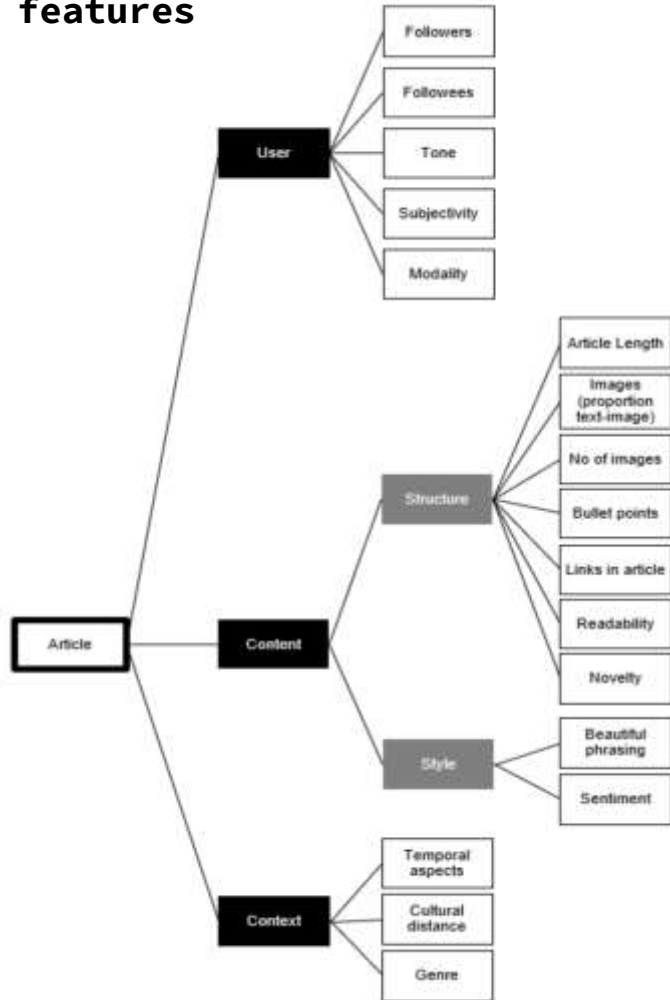


FEEDBACK MATTERS!  
PREDICTING THE  
APPRECIATION OF ONLINE  
ARTICLES A DATA-DRIVEN  
APPROACH

Catherine Sotirakou | PhD student | University  
of Athens - Faculty of Media & Communication  
Contact: @Cathrine\_Sot | cathrinesot@gmail.com

## Our features

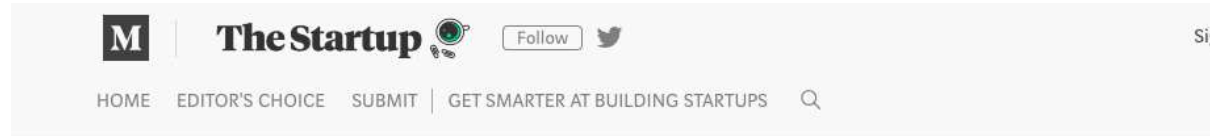


# PREDICTIVE MODEL

**Data:** We run experiments on a mixed-topic dataset of over 200K thousand articles published: Sept 2018-Sept 2019 on Medium.com and downloaded at the beginning of 2019. The articles had a large distribution of claps ranging from 84 to 157K claps

## DEPENDENT VARIABLE:

“Claps” (formerly called “Recommend”) on Medium represent whether readers liked the story or not and would recommend it to other users on the platform (a user can clap more than once). In the world of Medium, the success of an article is measured regarding claps count, which is the number of times the readers have been clapped.



Yesterday, Facebook lost around \$120 billion, and the internet cheered. That might sound like a nothing-burger like many of Facebook’s losses have been, but it wasn’t: It was nearly 20% of their total value (about as much as the entire worth of McDonald’s) and it disappeared in an afternoon, marking one of the biggest single-day collapses for a publicly traded company ever. After being marred by political, privacy, and even word-vomit scandals, Facebook’s invincibility finally fell away. And with their armor down, standing naked in front of the world, they finally learned the cost of all the problems they’ve caused.

But here’s an unpopular opinion: Months ago, after all of the Cambridge Analytica scandal had broken, and we already knew about their role in the election and fake news, Mark Zuckerberg told us all that he planned on



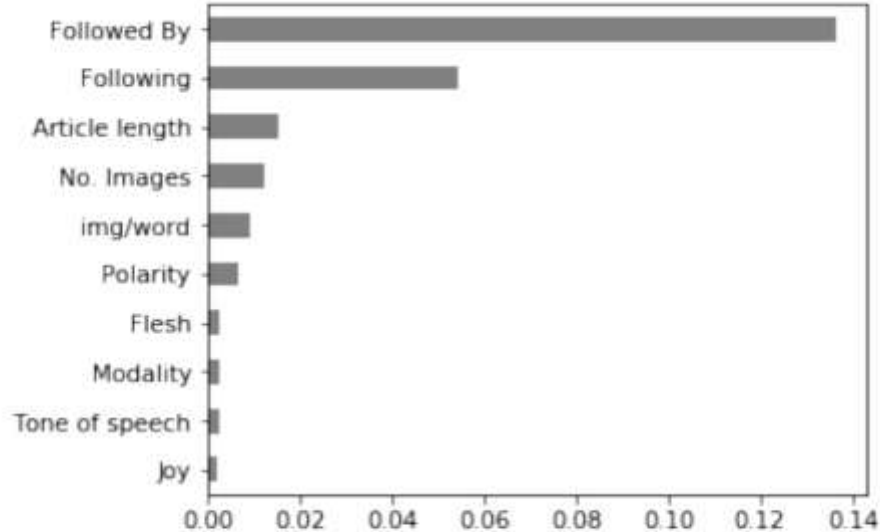
2

3

**Smart stories. New ideas. No ads. \$5/month.**

# METHOD - RANDOM FOREST

The table presents the ordered list of the importance of the variables of the selected categories, such as Followers, Article Length, Number of images etc., starting from the most important variables that affect the acceptance of an article.



The best score of all between the three different classifiers was generated by the XGBoost algorithm with F1-Score of 80%.

**Figure 2: Features importance score for high claps bucket**

THANK YOU !  
PLS GIVE ME SOME  
FEEDBACK!!!

Contact: @Cathrine\_Sot |  
cathrinesot@gmail.com |  
[aas2322@columbia.edu](mailto:aas2322@columbia.edu)