

Moderating comments for 'quality'

PROJECT FLETCHER

Taichi Hoshino

11 Nov 2016



The problem

There are great comments...



RW Remy Walls
Posted 17 days ago

Sir Ken Robinson presents a fair point when he states that school environments are training youth to think that there is no room for mistakes. This idea contradicts the belief that children should be able to express themselves and follow a path of creativity based on what they want to be. School systems are being hypocritical when they encourage children to do their best and be who they want to be, but overwhelm them with work, pressuring them to complete assignments and homework, consuming time and leaving little room for recreation. A very school in New York is experimenting with children, giving them less homework and more interactive assignments, allowing them to apply their creativity and extensive imagination. This research has shown that children excel more when they are able to use their imagination and it allows them to apply encouragement that they receive, opening up more opportunities for their future and making education easier.

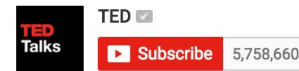
Robinson is able to fluently and eloquently deliver his speech with a balanced amount of comedy, gaining the audience's appeal in order to turn their attention to the issue of creativity. He is able to use this to show how schools and universities decide to steer away from recreational programs and more towards redundant and overwhelming studies, causing stress and depression. He also wisely alludes to Picasso, a creative man himself. He is able to apply the idea that "all children are born artists . . . [but] the problem is to remain [one]." This can take on many different interpretations, but he focuses on the idea that all children are naturally creative and must decide how to apply it to their lives. This effective rhetoric demonstrates Robinson's appeal, gaining approval from the audience and allowing him to establish a subtle credibility that will allow him to speak for many on the idea that the current education system is mistaken when they choose to turn away from the creative minds of our youth in contemporary society.

Reply Upvote

Report

And (hilariously) rubbish comments...

Do schools kill creativity? | Sir Ken Robinson



11,069,975 views

+ Add to Share More

104,530 1,412



SnowfurTheEevee 1 month ago

Short answer: yes.

Long answer: yeeeeeeeeeeesss.

Reply · 3



Popo Sandybanks 3 months ago

My kindergarten teacher crushed my dreams when she told me that I can't be a Jedi because they don't exist. Many tears were had that night.

Reply · 190



Riley Goss 2 months ago (edited)

I forgot 75% of what I learned in high school. The other 25% that I remember is useless.

Reply · 59

View all 4 replies

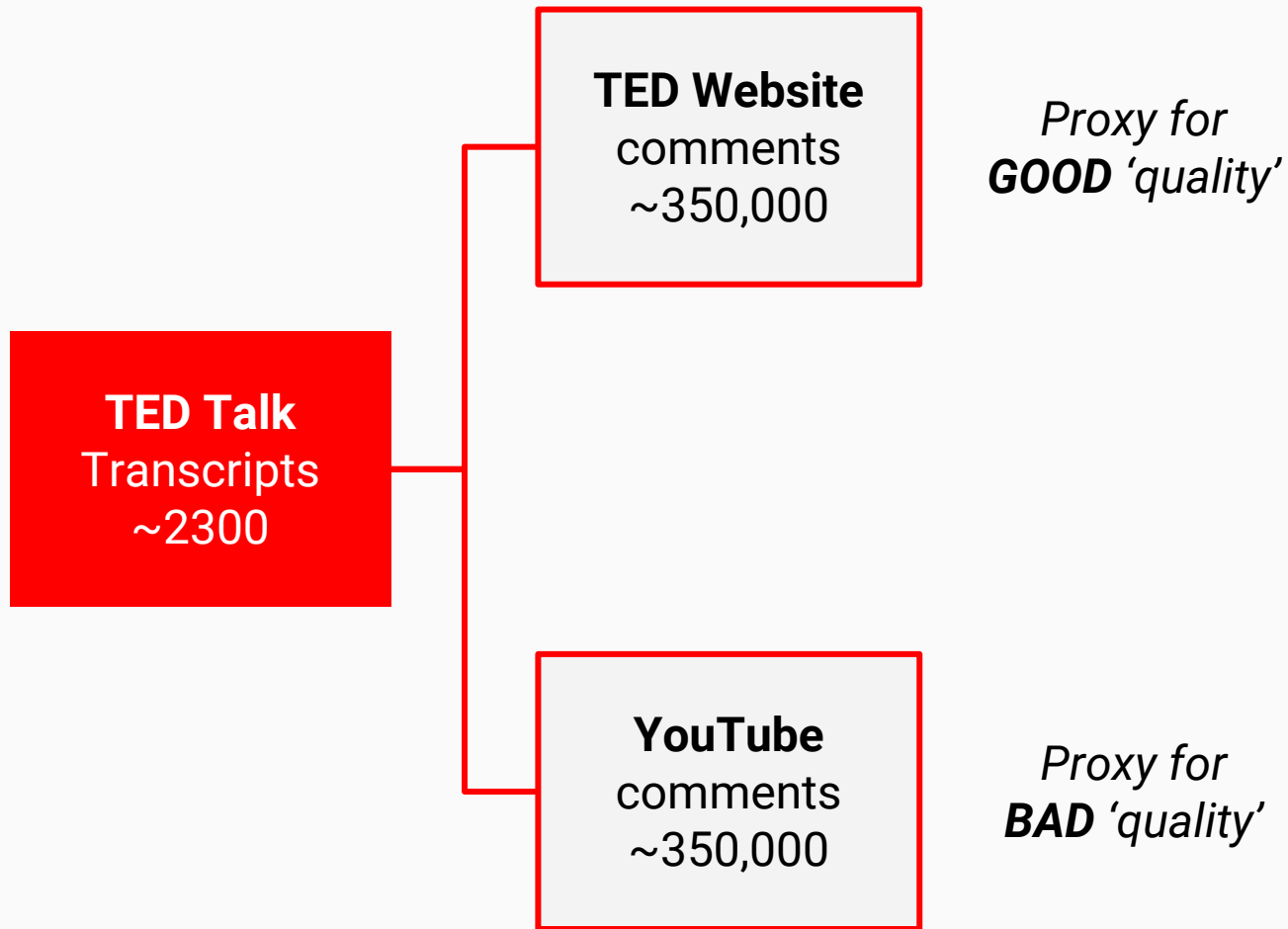
The question

Is there an effective way to automate the moderation of user submitted content to improve the '**quality**' of contributions?

Use cases

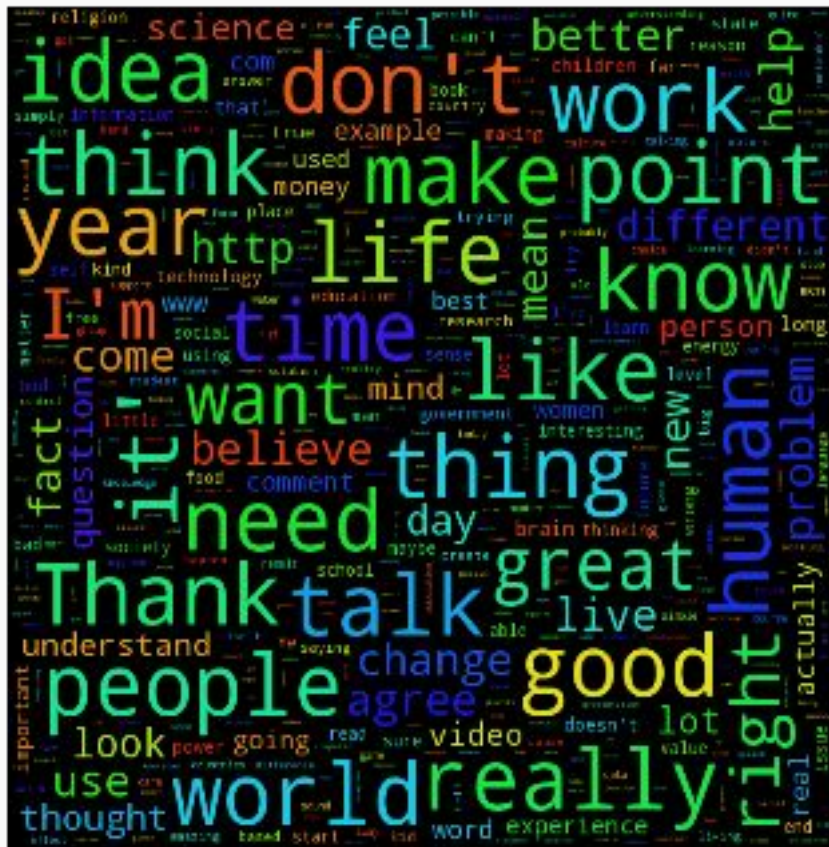
1. Contributors of high 'quality' content can be **identified** and encouraged to write more
2. Comments can be **ranked** based on 'quality' and can be made to be more visible to audiences
3. Provide live **feedback** to commenters (like a spell check) to guide people towards 'quality'

Data sources



The differences are not immediately obvious

TED



YOUTUBE



Features and pipeline

Topic cluster

- NMF
- LSA (x 10)

Topic relevance

- Cosine similarity
- Levenshtein (token set ratio)

Comment characteristics

- Length
- # of punctuations
- First / Second person

Sentiment Analysis

- NLTK VADER
- Text Blob (polarity + subjectivity)

Troll factor

- Kaggle troll model
- Bad words list



**Random
Forest**

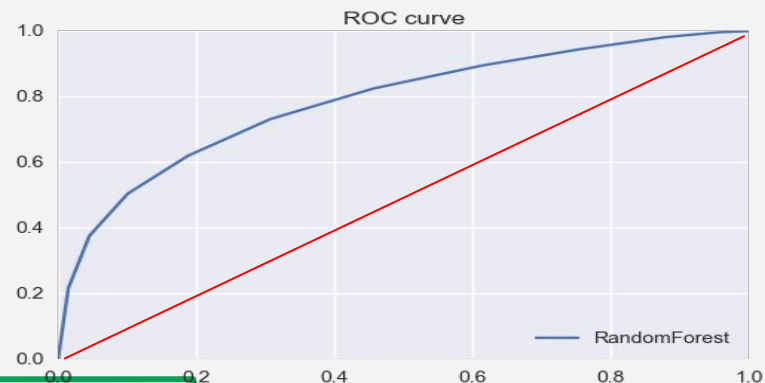
The result

Classification report

Model: RandomForest

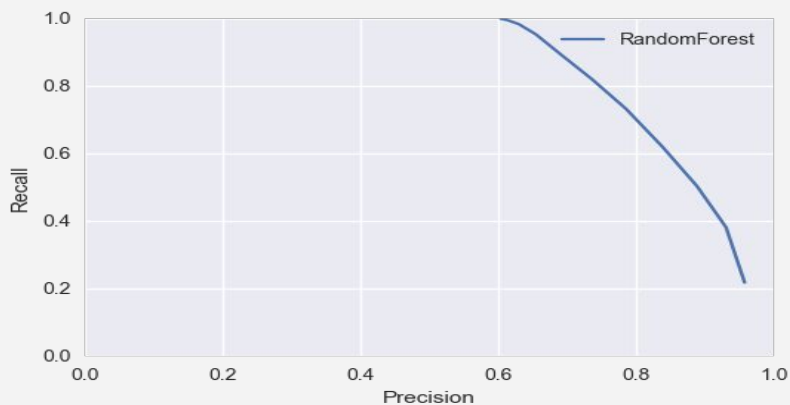
	precision	recall	f1-score
0	0.63	0.70	0.66
1	0.79	0.72	0.75
avg / total	0.72	0.72	0.72

ROC Curve



Encouraging!

Precision Recall

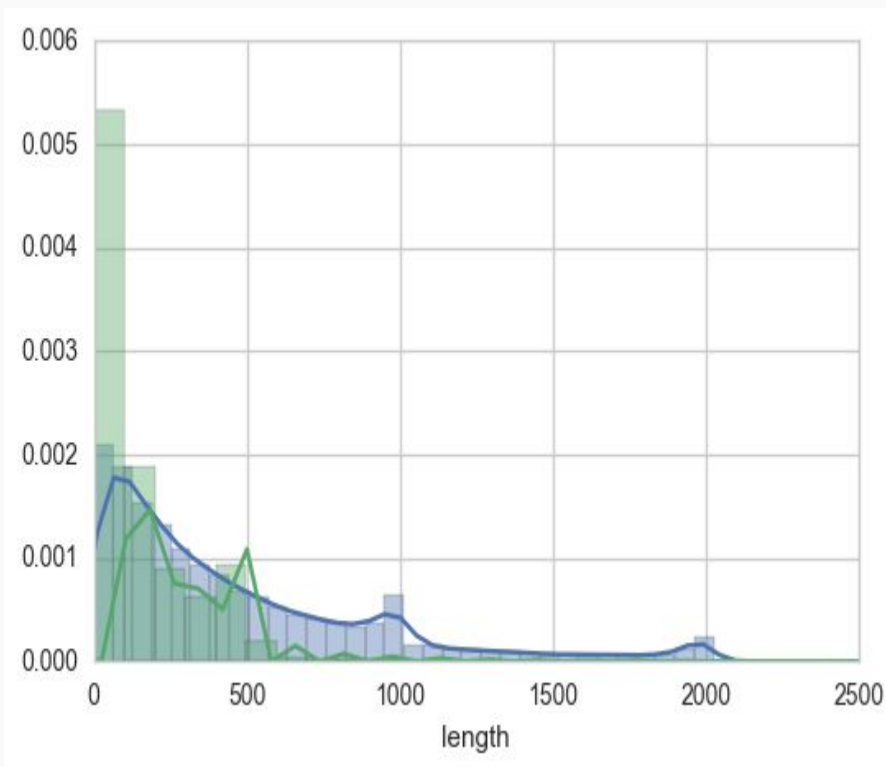


Feature Importance

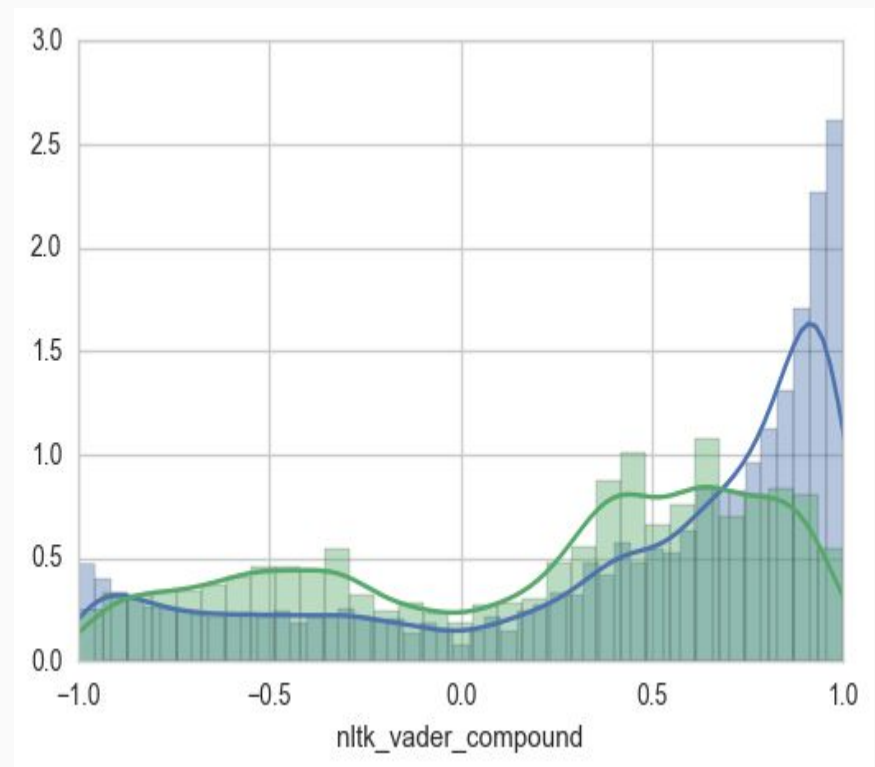
```
Model: RandomForest
('length', 0.17615958103603319)
('nltk_vader_compound', 0.078836964206574239)
(0, 0.065372154754284081)
(8, 0.051408859850673463)
('tb_polarity', 0.049437030260157971)
(6, 0.049049756883397397)
('lev_distance', 0.04550658777658168)
(2, 0.044254006995347761)
(9, 0.044134731991278403)
(1, 0.043478614756649103)
('cosine_sim', 0.043080432908143215)
```


The strongest predictors (examples)

Length of comment



NLTK VADER - sentiment



YouTube
TED

Next steps

1. Refine feature engineering and refine labelling
2. Refine supervised + unsupervised model performance
3. Build comment quality ranking algorithm for web applications