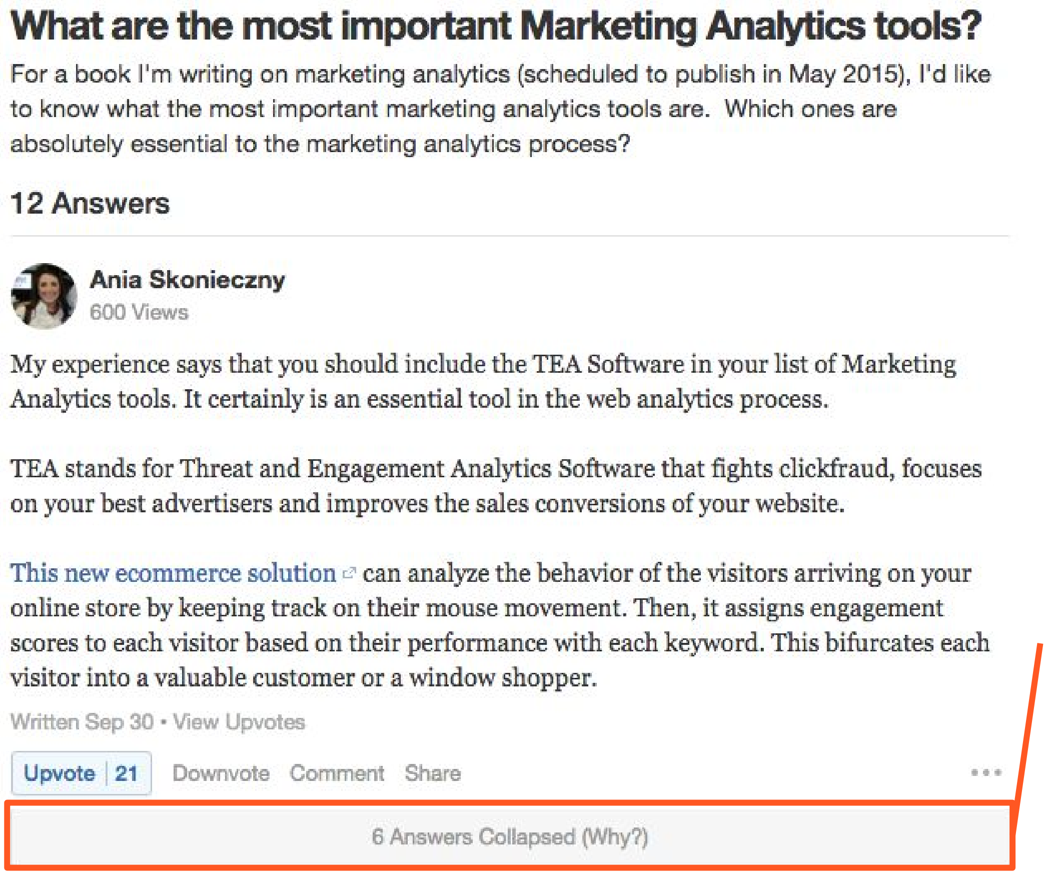
Predicting Collapsed Answers on Quora

Shannon Mitchell, Mason Montgomery, Akshat Verma, Mike Voltmer
SYS 6018 Final Project
10 Dec 2015

Problem Description

Quora



What are the most important Marketing Analytics tools?

For a book I'm writing on marketing analytics (scheduled to publish in May 2015), I'd like to know what the most important marketing analytics tools are. Which ones are absolutely essential to the marketing analytics process?

12 Answers

Ania Skonieczny
600 Views

My experience says that you should include the TEA Software in your list of Marketing Analytics tools. It certainly is an essential tool in the web analytics process.

TEA stands for Threat and Engagement Analytics Software that fights clickfraud, focuses on your best advertisers and improves the sales conversions of your website.

This new ecommerce solution can analyze the behavior of the visitors arriving on your online store by keeping track on their mouse movement. Then, it assigns engagement scores to each visitor based on their performance with each keyword. This bifurcates each visitor into a valuable customer or a window shopper.

Written Sep 30 • View Upvotes

6 Answers Collapsed (Why?)

Answers are manually reviewed and collapsed if insufficient

Q: What are the most important Marketing Analytics tools?

Collapsed Answer: Marketing is a group of people

Objectives

Objectives and Metrics

Objective: Predict whether an answer will be collapsed

Metrics:

$$\textbf{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}$$ (minimize false positives)

$$\textbf{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$ (minimize false negatives)

Q: When does something count as a musical instrument?

Collapsed Answer: … **1) You make music with it, and 2) Someone else asks you to show them how to do it** …

Related Work

Prior Work and Related Research

YAHOO!

Flags offensive content for human review

stackoverflow

Flags low quality content for human review

Q: Who was the worst woman in history

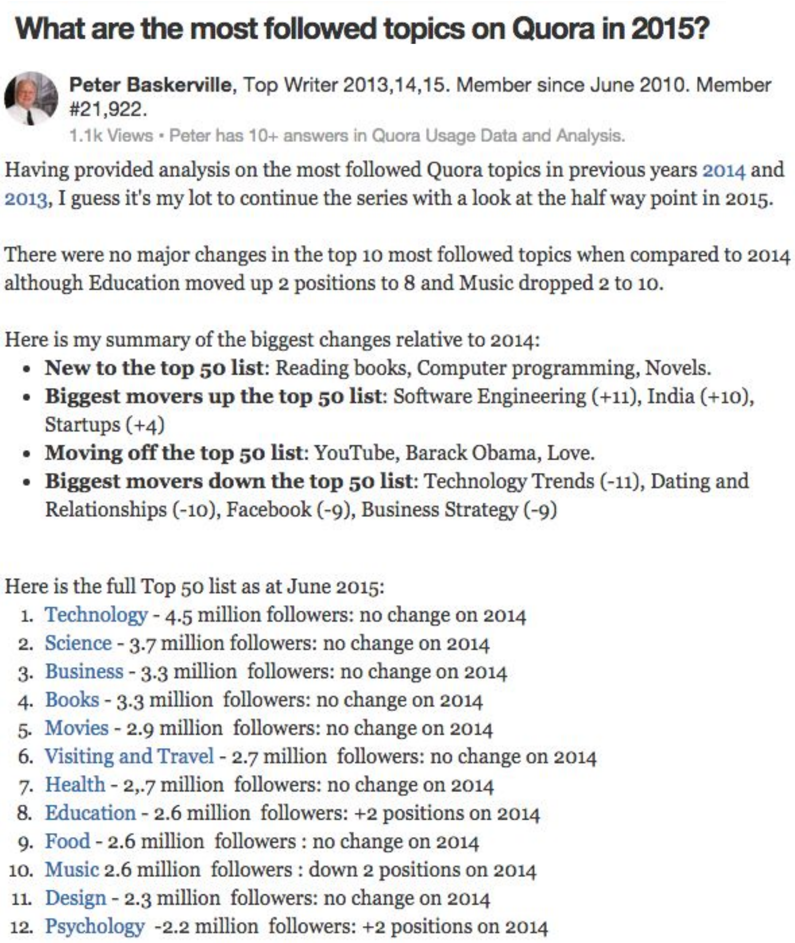
Collapsed Answer: Yoko Ono. She broke up the Beatles.

Technical Approach

Data

- Top 50 topics on Quora in 2015
- Dynamic content generation and user interaction → Selenium webdriver
- Scraped 4635 answers from 183 questions
- Features collected:

  Question Text, Answer Text,
  Upvotes, Views,
  Profile of Answerer,
  Accepted/Collapsed



What are the most followed topics on Quora in 2015?

Peter Baskerville, Top Writer 2013,14,15. Member since June 2010. Member #21,922.

1.1k Views • Peter has 10+ answers in Quora Usage Data and Analysis.

Having provided analysis on the most followed Quora topics in previous years 2014 and 2013, I guess it's my lot to continue the series with a look at the half way point in 2015.

There were no major changes in the top 10 most followed topics when compared to 2014 although Education moved up 2 positions to 8 and Music dropped 2 to 10.

Here is my summary of the biggest changes relative to 2014:

- **New to the top 50 list**: Reading books, Computer programming, Novels.
- **Biggest movers up the top 50 list**: Software Engineering (+11), India (+10), Startups (+4)
- **Moving off the top 50 list**: YouTube, Barack Obama, Love.
- **Biggest movers down the top 50 list**: Technology Trends (-11), Dating and Relationships (-10), Facebook (-9), Business Strategy (-9)

Here is the full Top 50 list as at June 2015:

1. Technology - 4.5 million followers: no change on 2014
2. Science - 3.7 million followers: no change on 2014
3. Business - 3.3 million followers: no change on 2014
4. Books - 3.3 million followers: no change on 2014
5. Movies - 2.9 million followers: no change on 2014
6. Visiting and Travel - 2.7 million followers: no change on 2014
7. Health - 2,.7 million followers: no change on 2014
8. Education - 2.6 million followers: +2 positions on 2014
9. Food - 2.6 million followers : no change on 2014
10. Music 2.6 million followers : down 2 positions on 2014
11. Design - 2.3 million followers: no change on 2014
12. Psychology -2.2 million followers: +2 positions on 2014

Q: How can I run faster than Usain Bolt?

Collapsed Answer: Eat, drink sleep ..Keep doing this forever . But you cant run faster than Usain bolt

Feature Engineering

Answer Text

- Number of characters, words, sentences, spaces
- Percentage of uppercase letters
- Percentage of text abbreviations & stopwords
- Readability indices
- Entropy

Question Text & Question/Answer Comparisons

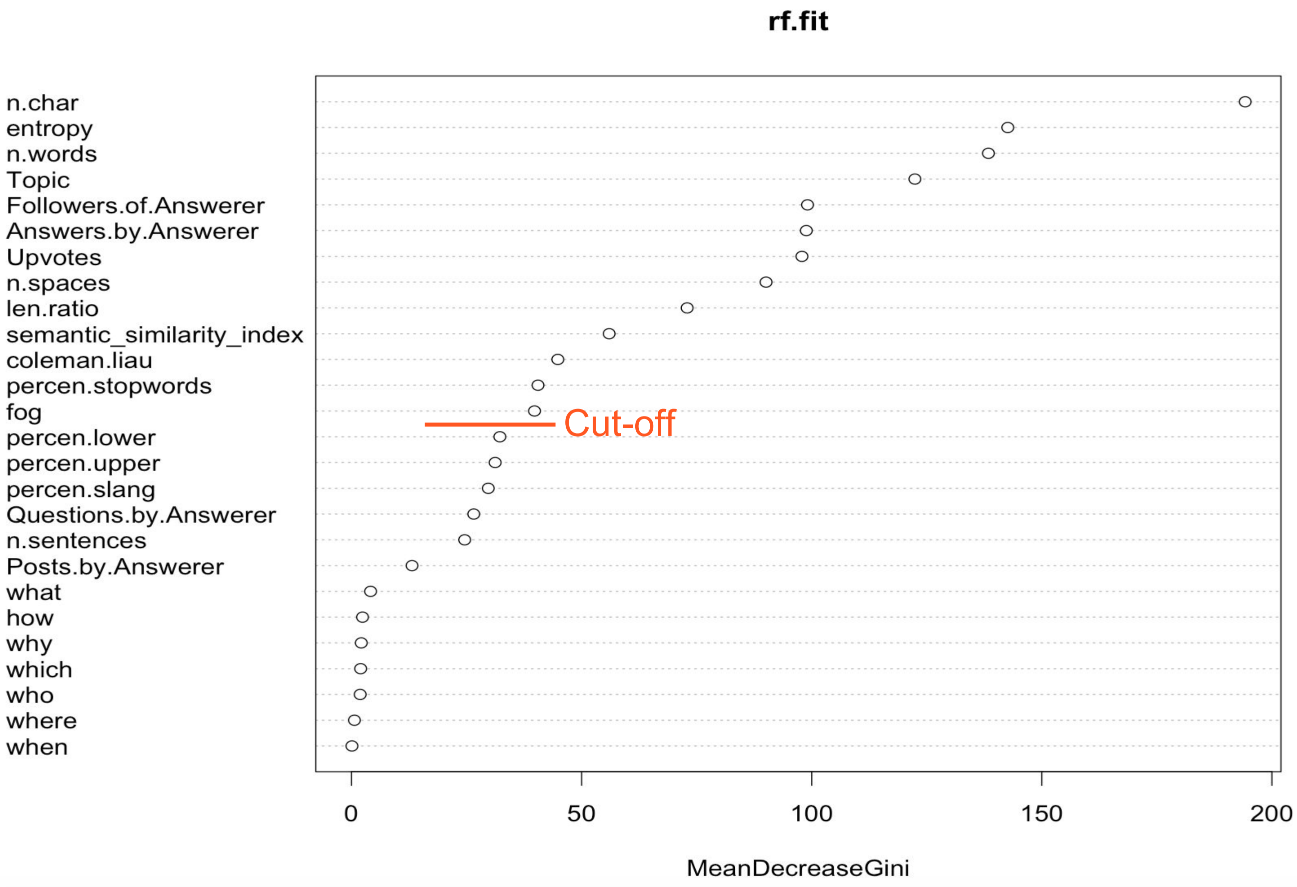
- Type of question
- Question/Answer length ratio
- Question/Answer semantic similarity index

Q: What is science?

Collapsed Answer: Solid is water

Feature Selection

Feature Importance given by Random Forest



Q: What is Physics?

Collapsed Answer: xkcd nails it.

Evaluation

Models

	Precision	Recall
Logistic Regression	0.75	0.57
Radial SVM	0.76	0.60
Random Forest	0.77	0.79

	Ground Truth	
Logistic	Accepted	Collapsed
Accepted	559	108
Collapsed	47	140

SVM	Accepted	Collapsed
Accepted	558	98
Collapsed	48	150

Random Forest	Accepted	Collapsed
Accepted	549	52
Collapsed	57	196

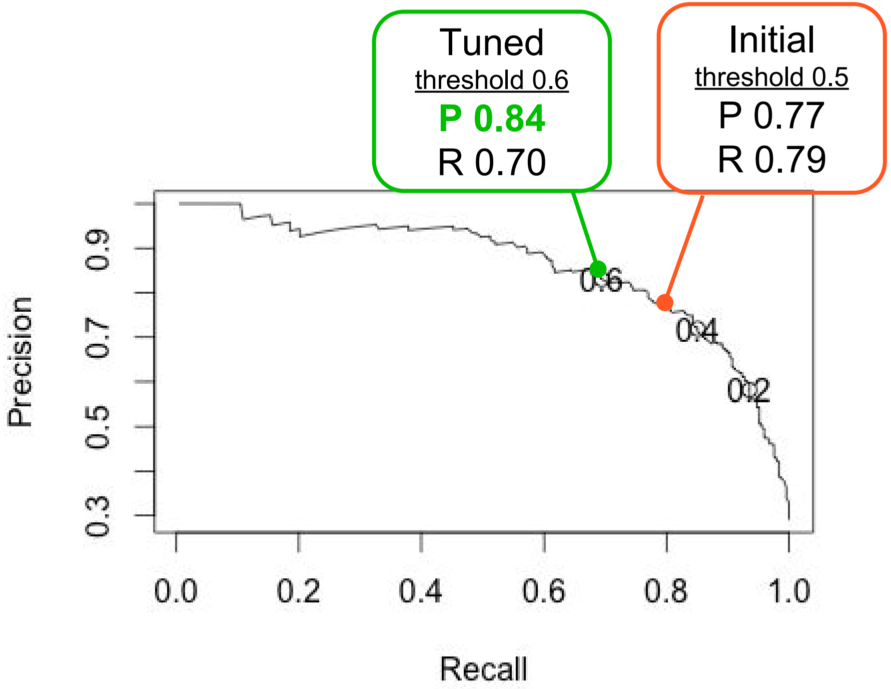
Q: What is mathematics?

Collapsed Answer: it is a subject where X is always missing and your duty is to find out Y… basically it is a subject which only teachers understand. :)”

Performance

- Tuned threshold to maximize precision with acceptable recall

- **Beats precision target of 0.66**



Tuned
threshold 0.6
P 0.84
R 0.70

Initial
threshold 0.5
P 0.77
R 0.79

Q: What are the most important lessons to learn about personal finance?

Collapsed Answer: Earn more than you spend

Limitations

- Does the model hold for less popular topics with fewer answers?

- Readability indices for short answers < 100 words aren’t reliable

  ex: “Alot” scores a 10th grade reading level

Q: How many genres of music are there? **Collapsed Answer: Alot**

Conclusions

Recommendations

- Implement model to flag answers for manual review

- If manual review confirms model success, implement tool to:
  - automatically moderate answers
  - warn users if answers will be collapsed before they are posted

Q: What are some of the best songs of all time? **Collapsed Answer: Definitely Bohemian Rhapsody from Queen**

Future Work

- Increase sample size and question/answer diversity
- Add features: part of speech counts, syntactic analysis
- Predict number of upvotes
- Apply additional model types: gradient boosting, neural network classifiers

Q: How did Rafael Nadal beat Novak Djokovic in the 2013 US Open Final?

Collapsed Answer: With his Tennis Racquet.

Questions?

Q: What are the best programming languages to learn today?
Collapsed Answer: Which is the best flower?

Q: Is there a liberal media bias?
Collapsed Answer: You betcha

Q: How do I understand physics easily?
Collapsed Answer: … 1. Drop an apple on your head 2. Try running on a slippery wet floor 3. Get down from a running bus …

Q: What is the difference between baking powder and baking soda?"
Collapsed Answer: The spelling.--- Edit ---Wow. Already flagged by someone for no reason. My answer is 100% correct.