TED Talks

Building an “awesome talk” detector

Overview/Agenda

The Data (30 seconds)

The Pipeline and Model (2.5 minute)

- Model Performance
- Pipeline/Model
- Model Features

Etc - Flask + TED Markov Chain (30 seconds)

The Data

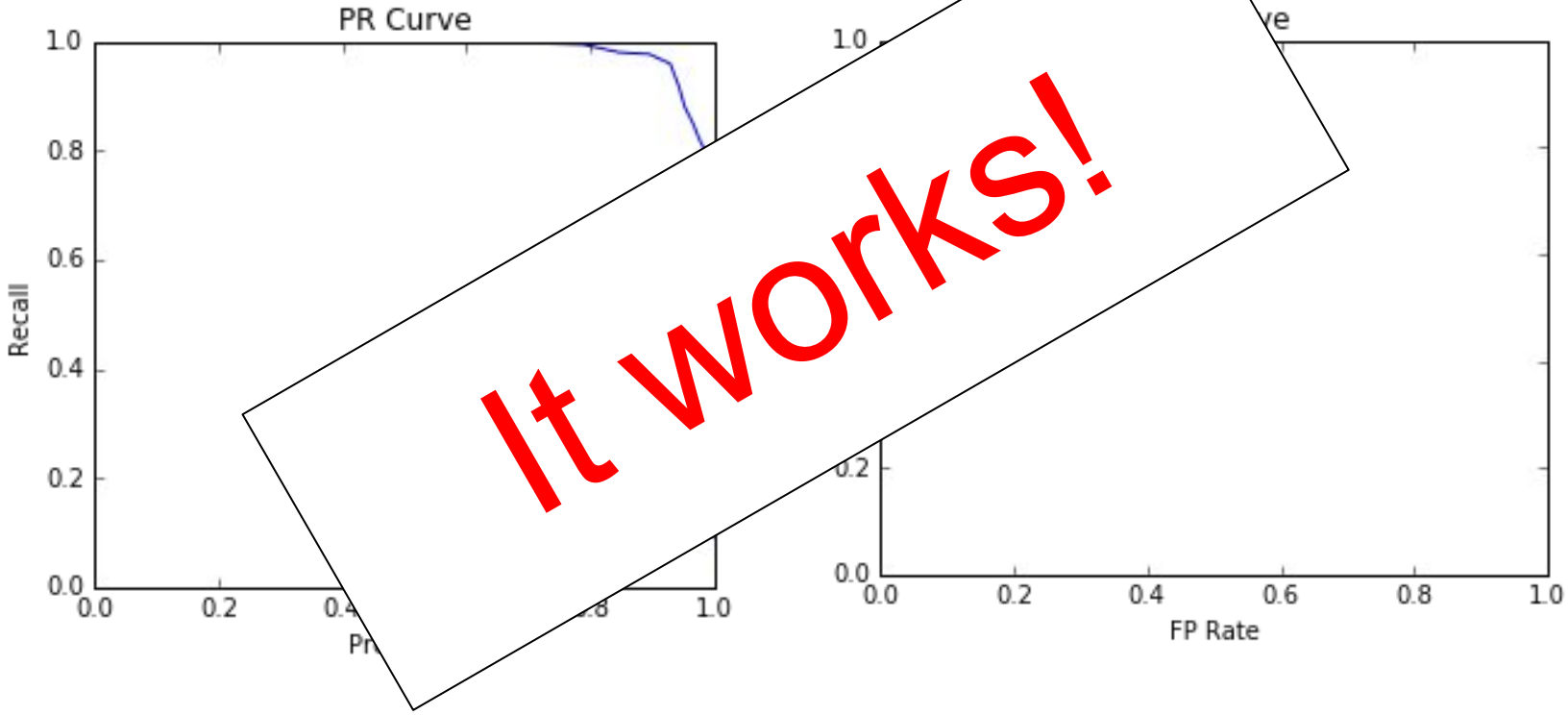
TED Talks

- 2153 Transcripts
  - 393 Topics
  - ~5-20 minutes long
  - Exciting!

Yale Open Lectures

- 1077 Transcripts
  - 23 Departments/Topics
  - ~60 minutes
  - Boring...

Model Performance



Pipeline/Model

Munge

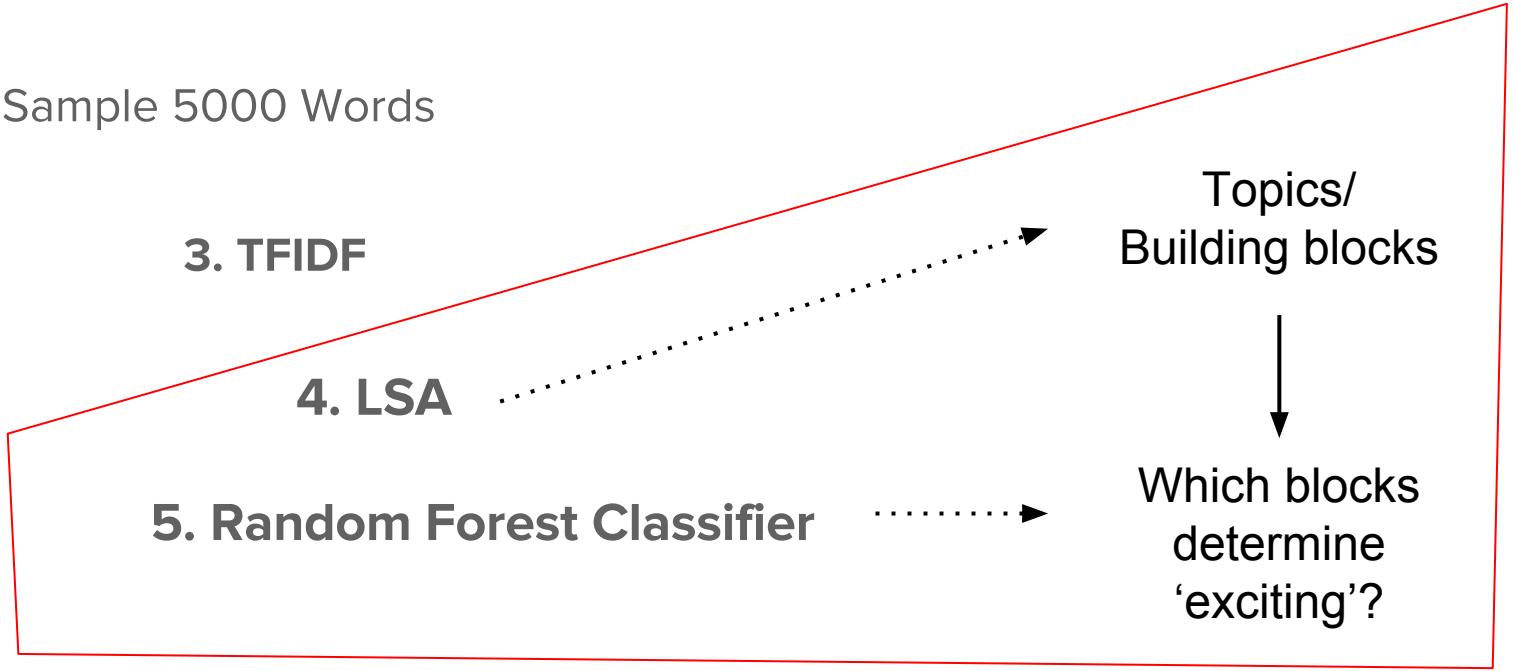
1. Scrub cheating words
2. Sample 5000 Words

Transform

3. TFIDF
4. LSA

Model

5. Random Forest Classifier



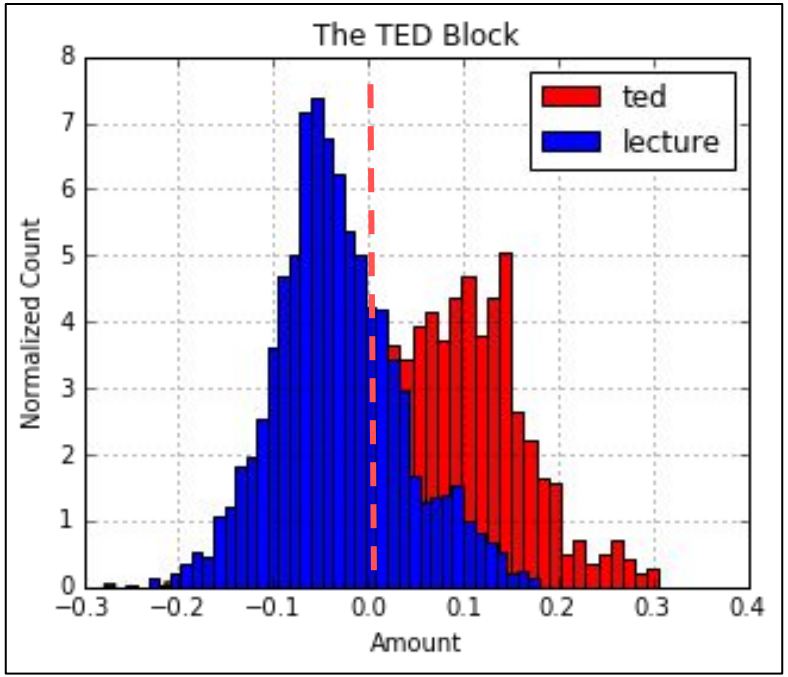
Building Block/Topics - The “TED” Block

LSA rank: #3

RFC rank: #3

Words:

Market, Energy, Price, Rate, Political, Hydrogen, Economy, Electron, Equilibrium, Democracy



Histogram of how much this block is used in each talk.

Flask App

TED Talks

OMG It's borrrrrringgg..

Paste Transcript Here:

Roberto Gonz\xe1lez Echevarr\xeda: I have been speaking in the last lectures about the ending of Part I of the Quixote, and today we finally get to the end of the novel. I remind you, again, that although you have both parts bound as one book \u2014 we all have it bound as one book \u2014 this is the end of the novel Cervantes set out to write, that he had no specific plan to write a second part, which would not be published anyway until a decade later, in 1615. It is very easy to make the mistake that what we\'re coming to now is a provisional ending, so avoid that mistake. It is very easy to make, because of the book being bound together, and also because of the very nature of the ending of Part I that is so complicated. It is also easy to fall in that mistake because you know that Part II is coming as part of your course, and that you are going to read it, but not the readers of the 1605 Quixote

SUBMIT

TED Talks

OMG It's like a TED talk! Oohh la laaa!

Paste Transcript Here:

Is there life beyond Earth in our solar system? Wow, what a powerful question. You know, as a scientist -- planetary scientist -- we really didn\'t take that very seriously until recently. Carl Sagan always said, "It takes extraordinary evidence for extraordinary claims." And the claims of having life beyond Earth need to be definitive, they need to be loud and they need to be everywhere for us to be able to believe it. So how do we make this journey? What we decided to do is first look for those ingredients for life. The ingredients of life are: liquid water -- we have to have a solvent, can\'t be ice, has to be liquid. We also have to have energy. We also have to have organic material -- things that make us up, but also things that we need to consume. So we have to have these elements in environments for long periods of time for us to be able to be confident that life, in that moment when

SUBMIT

TED Markov Chain Excerpt

"... And in fact, we need a big breakthrough here, something that's going to destroy civilization, I need to figure out how we can answer them. Let's go there and make a map of my city, New York City, showing inundation in red… "

Takeaways

- Built a model predicting talk awesomness
- Served it on a Flask app

Questions?

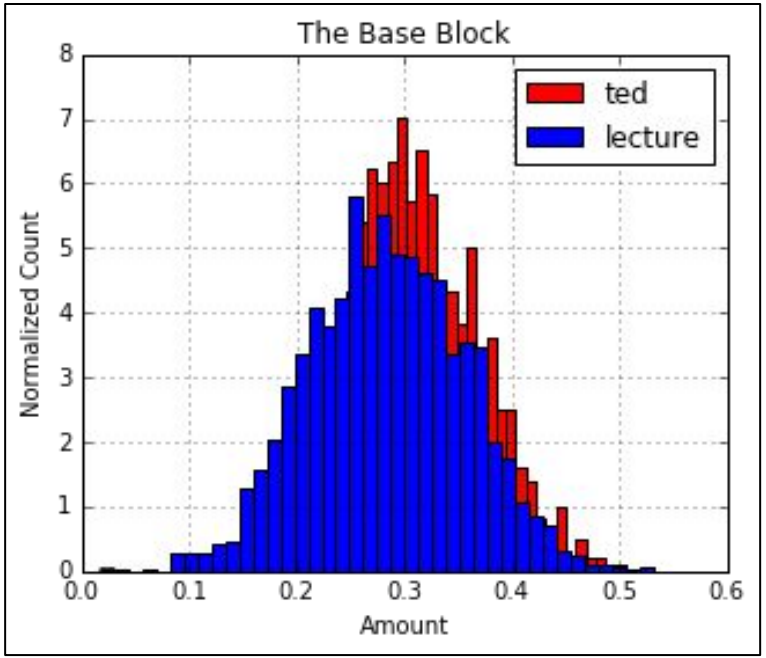
Building Block/Topics - The Base Block

LSA rank: #1

RFC rank: #20

Words:

People, Like, Just, Going, Know, Think, Really, Time, Things, Actually



Histogram of how much this block is used in each talk.