IMDb

Spoiler Detection

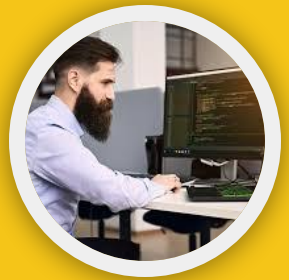
Kevin Chiv

Why Movie Spoiler Detection

- Users -

Less Spoilers

- Site Admin -

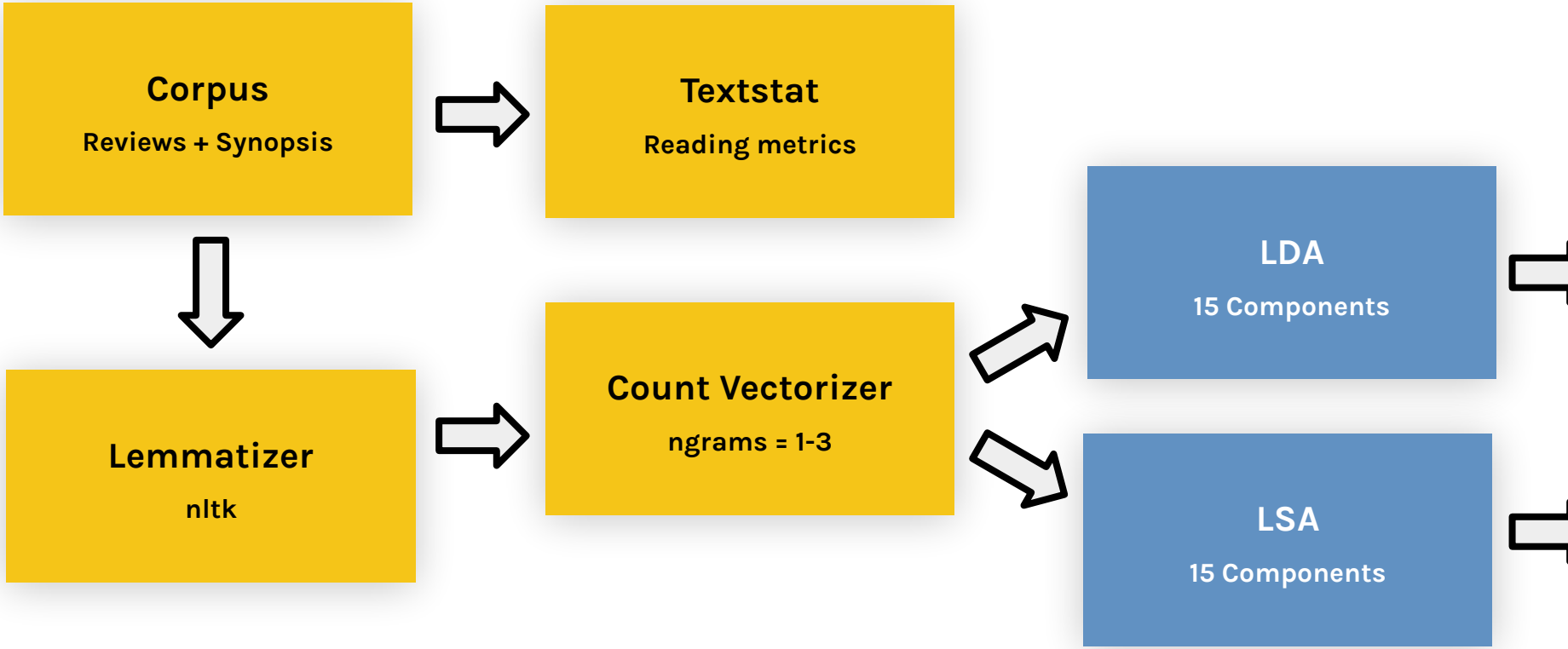


Less to Moderate

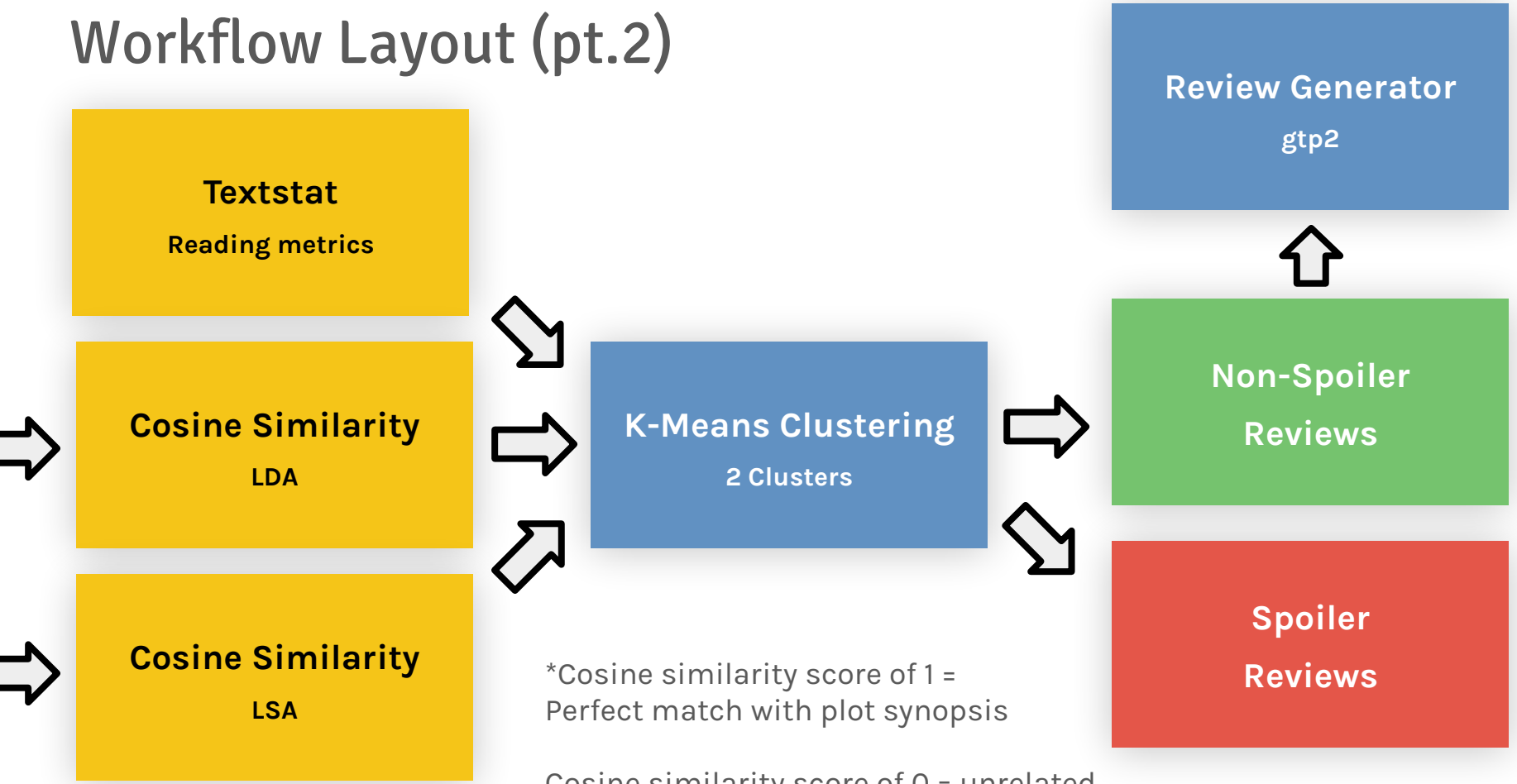
- Film Makers -

More Revenue

Workflow Layout (pt.1)



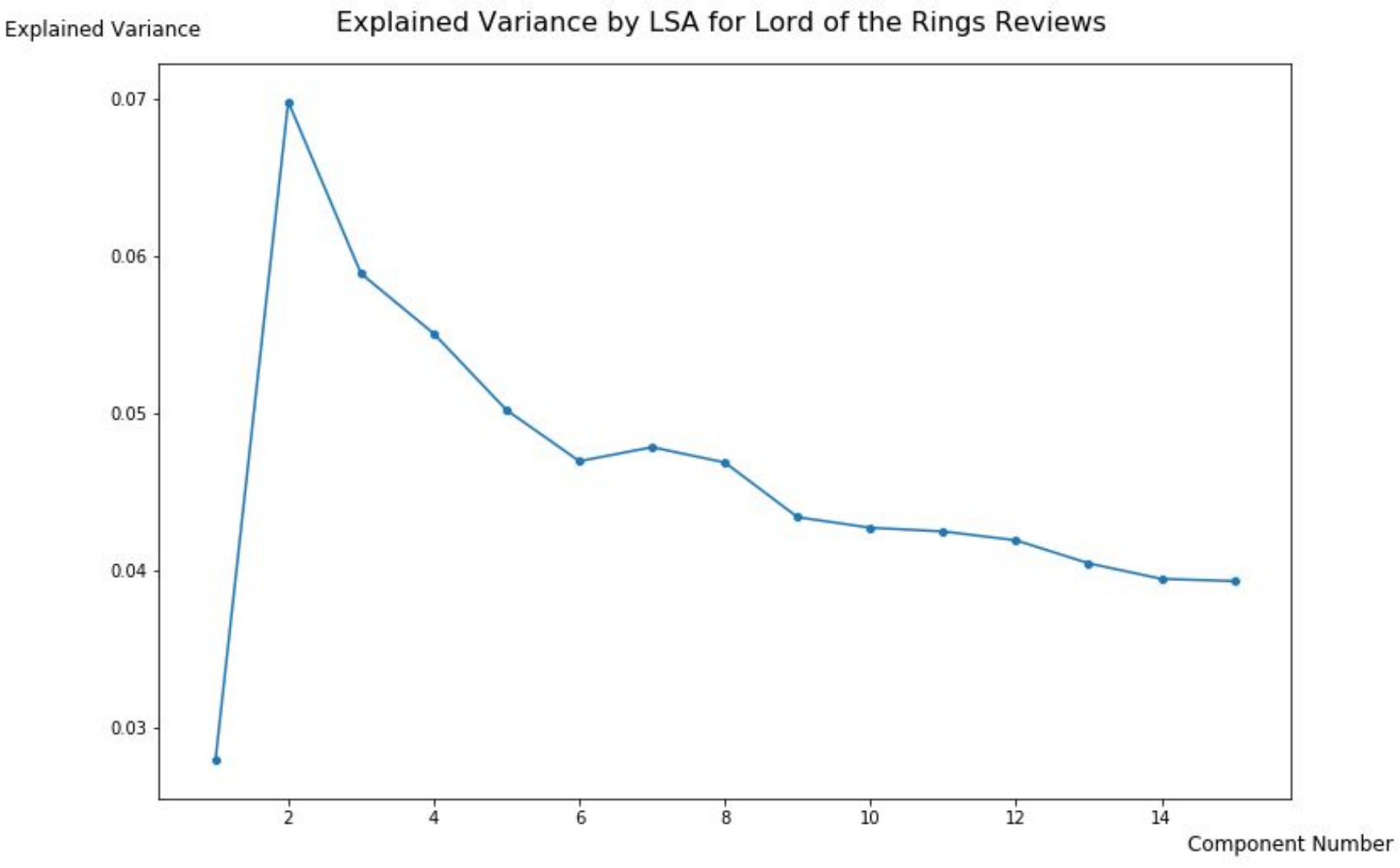
Workflow Layout (pt.2)

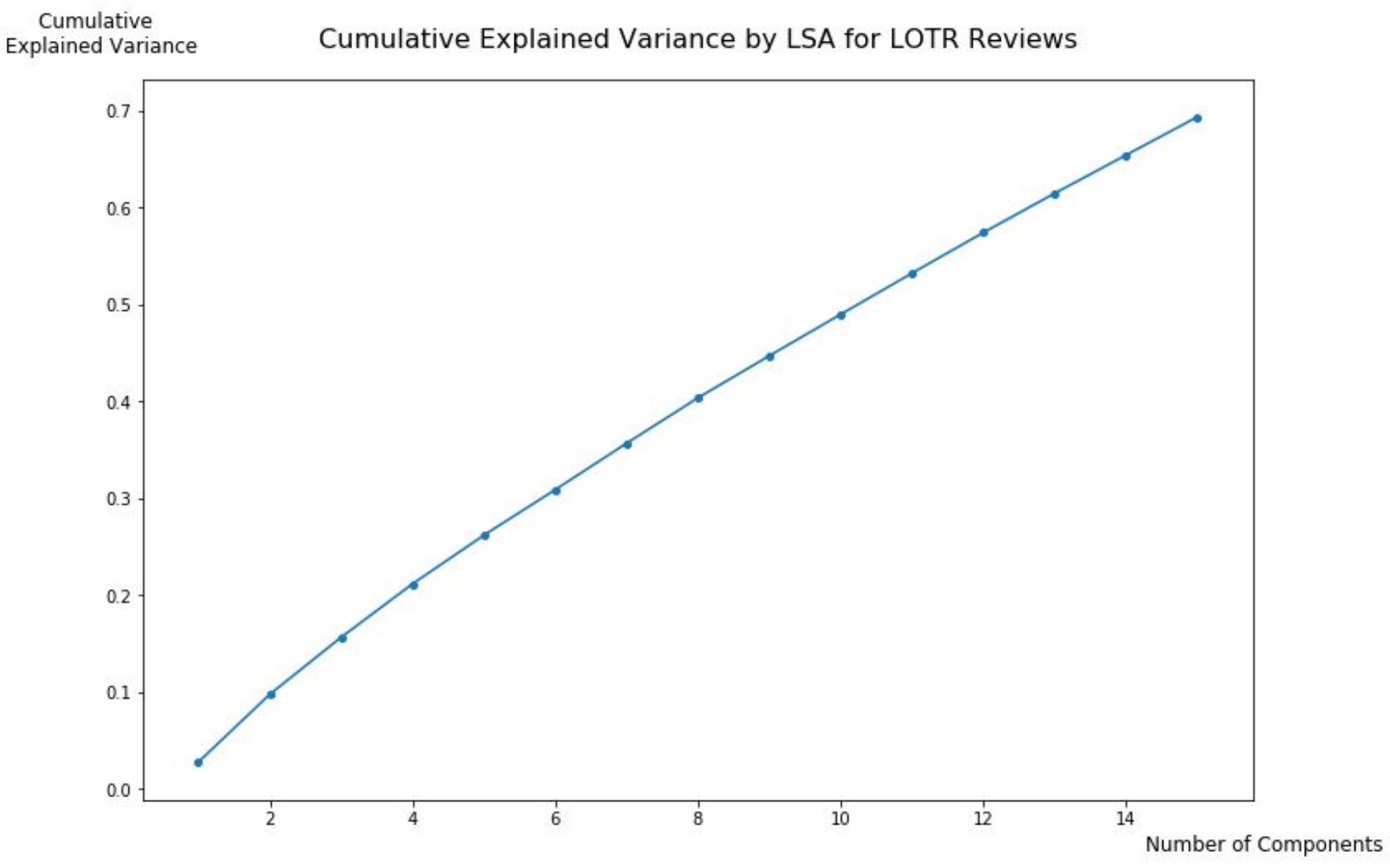


*Cosine similarity score of 1 =
Perfect match with plot synopsis

Cosine similarity score of 0 = unrelated

Let’s Look At An Example


Explained Variance by LSA for Lord of the Rings Reviews
Explained Variance
0.07
0.06
0.05
0.04
0.03
2
4
6
8
10
12
14
Component Number


Cumulative Explained Variance
Cumulative Explained Variance by LSA for LOTR Reviews
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2
4
6
8
10
12
14
Number of Components

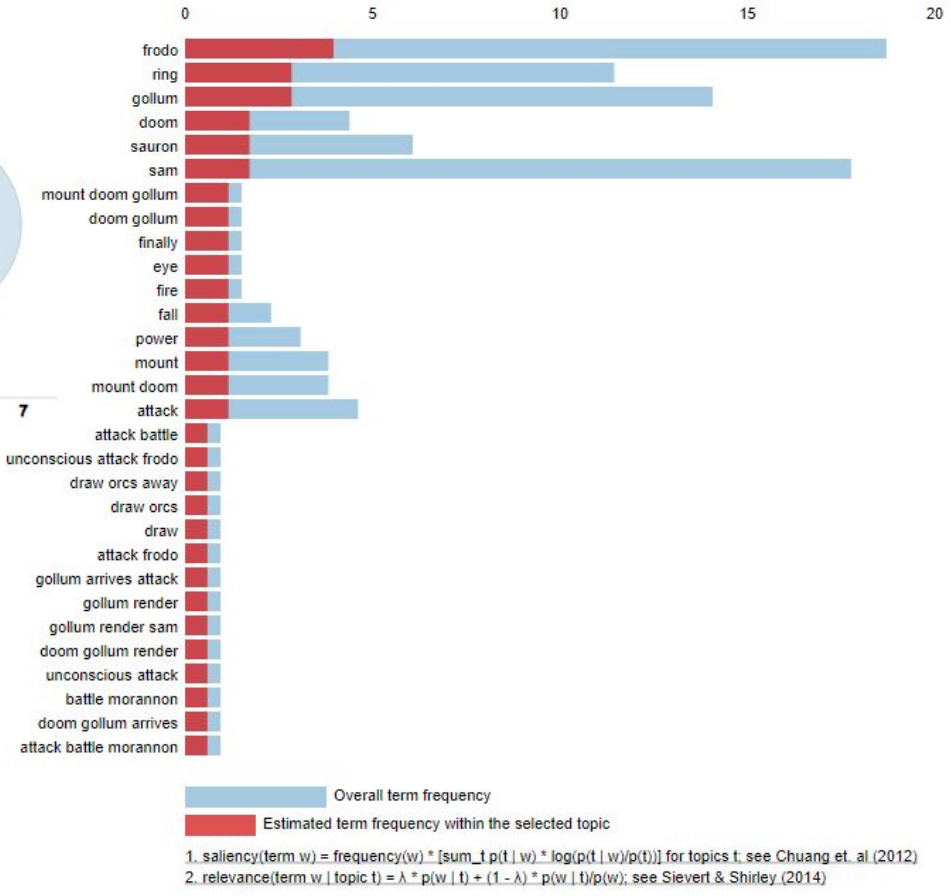
Selected Topic: 12 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric:(2)
λ = 1

0.0 0.2 0.4 0.6 0.8 1

Intertopic Distance Map (via multidimensional scaling)

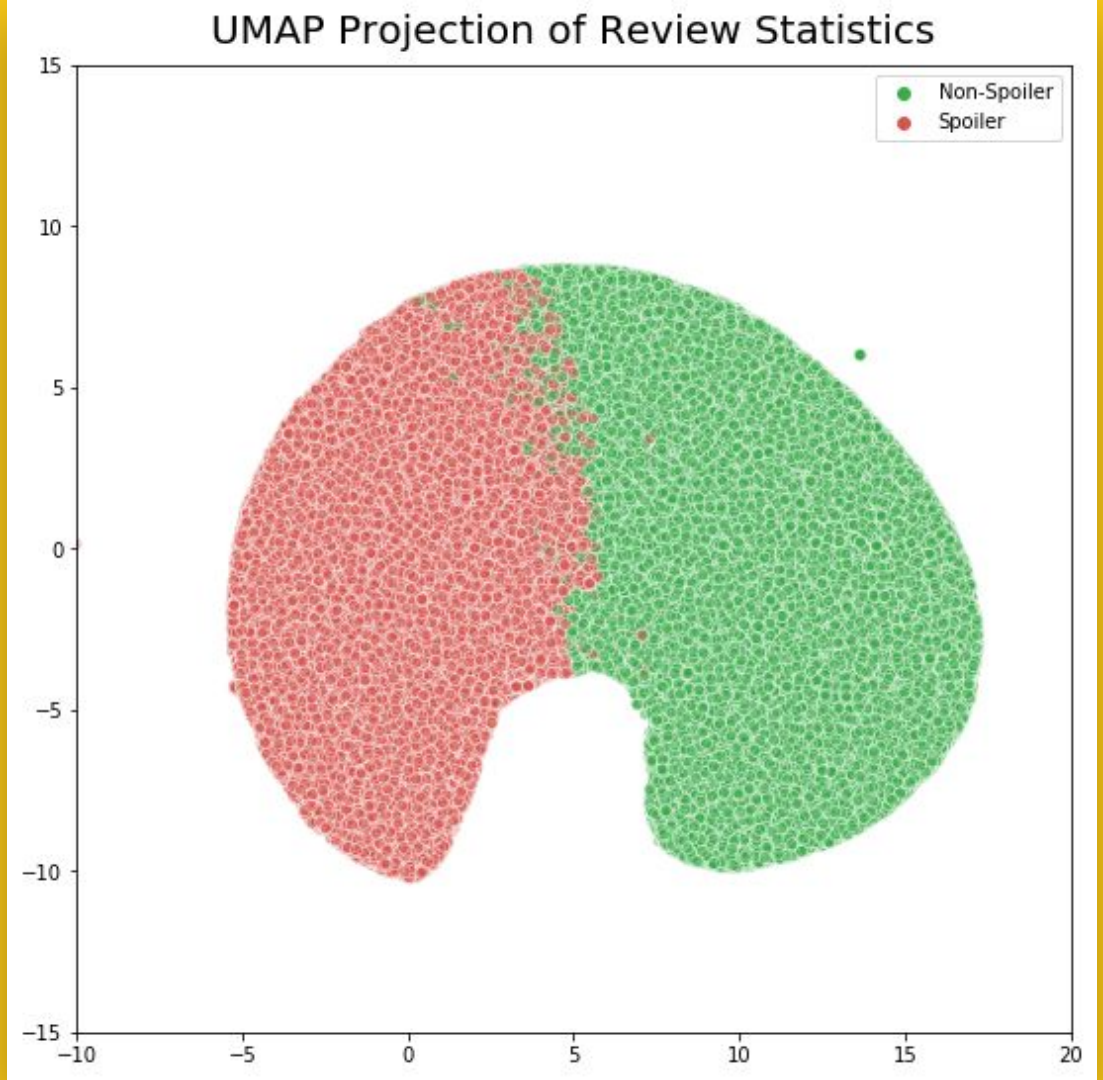
Top-30 Most Relevant Terms for Topic 12 (7.3% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Movie Review Statistics

	Spoilers	Non-Spoilers
Cosine Similarity (LSA)	0.854	0.610
Cosine Similarity (LDA)	0.804	0.700
Cosine Similarity (NMF)	0.752	0.535
Word Difficulty Score	0.125	0.058
Count	221957	166083


UMAP Projection of Review Statistics
Non-Spoiler
Spoiler
15
10
5
0
−5
−10
−15
−10
−5
0
5
10
15
20

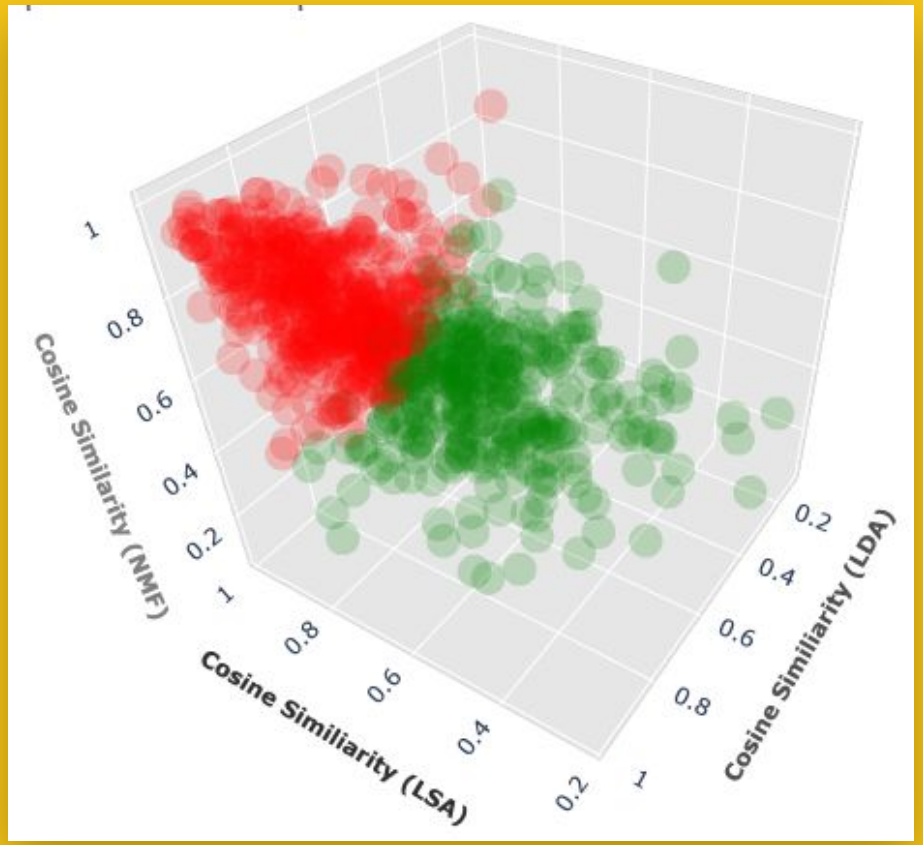
K-Means Clustering

Non-Spoiler Reviews

- Green
- Lower cosine similarity values

Spoiler Reviews

- Red
- About 57% of all reviews



Generated Non-Spoiler Review for LOTR

1000 iterations later . . . gpt2 presents* . . .

With an an average loss of 1.7 . . .

“Just got home from seeing "The Return of the King" at the midnight screening. The movie was great, I had expected a bit more but still felt that there was too much to cover. The action was fantastic, everything from the acting to the special effects. I would recommend this movie to anyone, especially those who haven't read the books…”

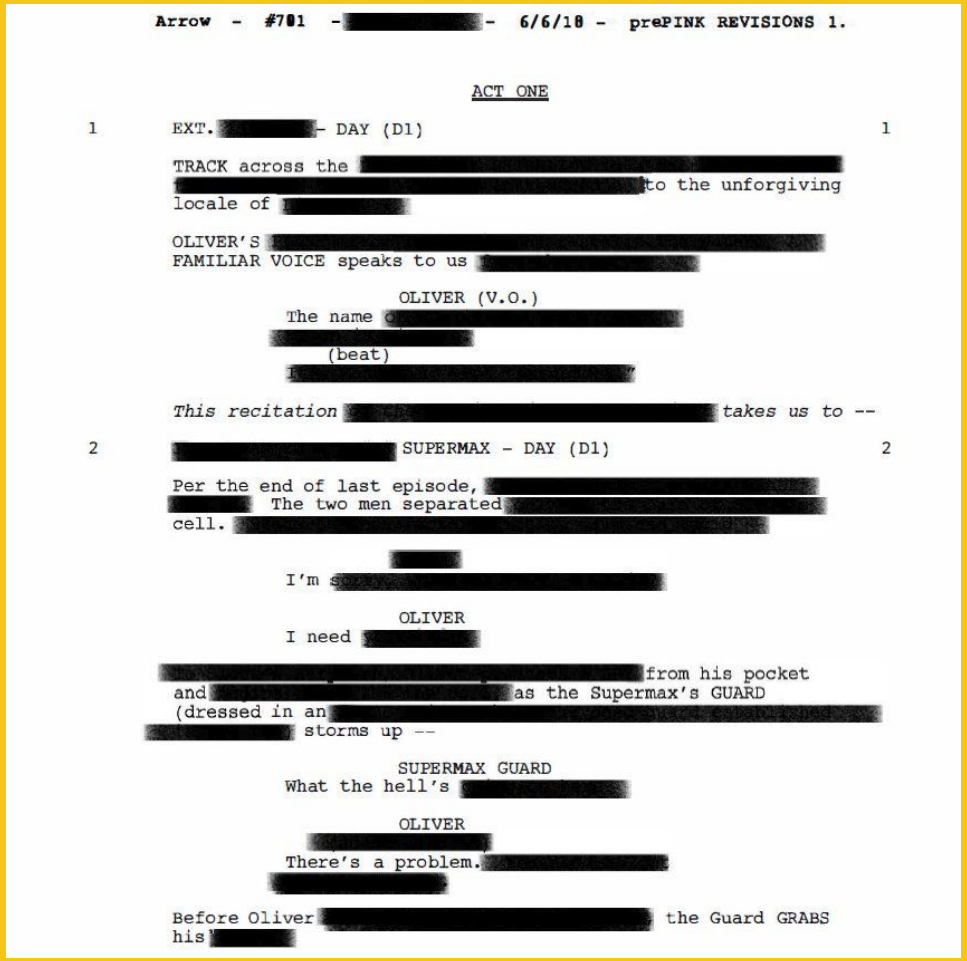
*Thanks to Seth Weiland

Recommendations

- Automatically flag text as spoilers with high cosine similarity against plot synopsis

- Hide spoiler reviews behind a “Click to See Spoiler” button

- Generate spoiler reviews to further improve the model



Arrow - #701 - [redacted] - 6/6/18 - prePINK REVISIONS 1.

ACT ONE

1 EXT. [redacted] - DAY (D1) 1

TRACK across the [redacted] to the unforgiving locale of [redacted]

OLIVER'S [redacted] FAMILIAR VOICE speaks to us [redacted]

OLIVER (V.O.)
The name [redacted]
(beat)
[redacted]

This recitation [redacted] takes us to --

2 [redacted] SUPERMAX - DAY (D1) 2

Per the end of last episode, [redacted] The two men separated [redacted] cell.

[redacted]
I'm [redacted]

OLIVER
I need [redacted]

[redacted] from his pocket and [redacted] as the Supermax's GUARD (dressed in an [redacted] storms up --

SUPERMAX GUARD
What the hell's [redacted]

OLIVER
[redacted]
There's a problem. [redacted]

Before Oliver [redacted] the Guard GRABS his [redacted]

~~The End~~

Thank You