



IMDB Movie Review Sentiment Classification

Ironhack DAFT
Nathan Fournillier
OCT 2020





01. Hypothesis

02. Steps to success

03. Obstacles

04. Solutions

05. Streamlit App



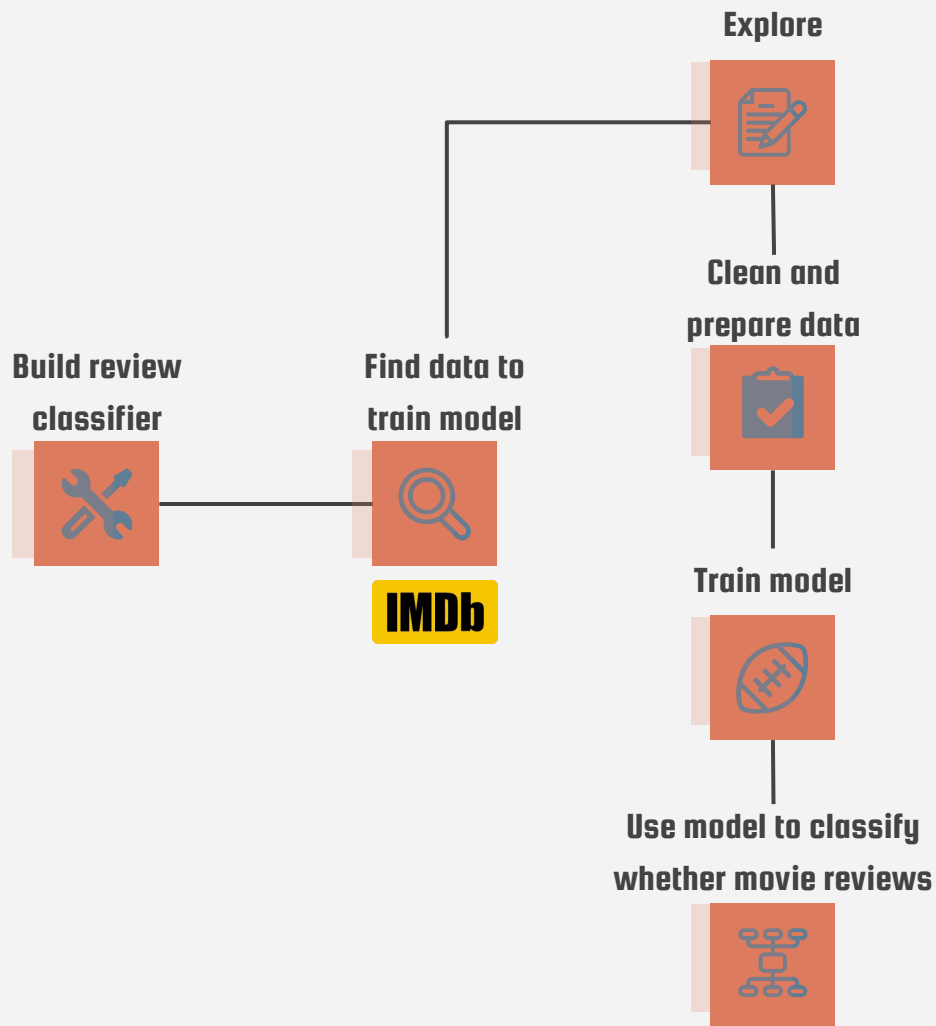


Are movie sequels ever
better than the original
movie?

Are remakes ever better
than the original?



02. STEPS TO SUCCESS



Exploring the data

```
example_text = df['reviews'][8]
example_text
```

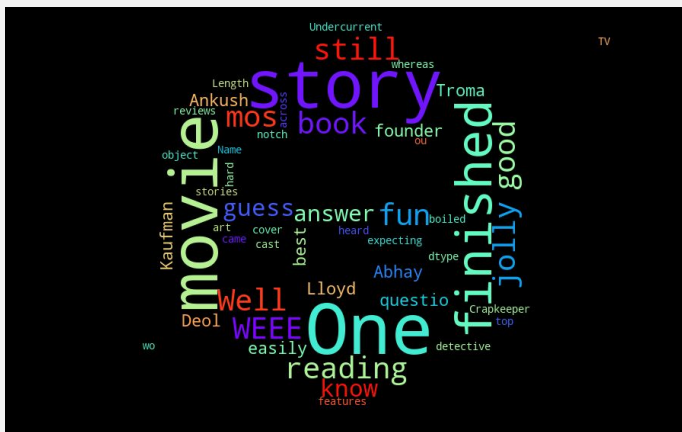
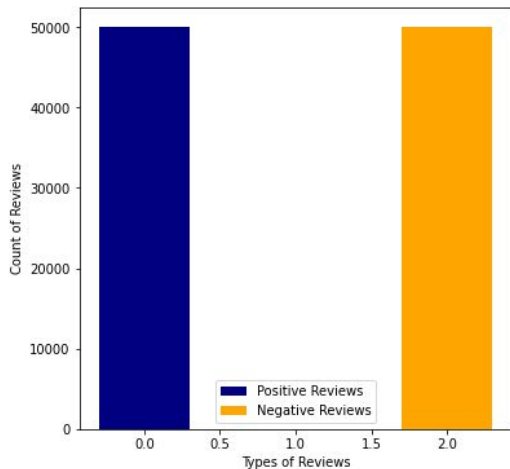
THE NIGHT LISTENER (2006) **1/2 Robin Williams, Toni Collette, Bobby Cannavale, Rory Culkin, Joe Morton, Sandra Oh, John Cullum, Lisa Emery, Becky Ann Baker. (Dir: Patrick Stettner).
The movie gives Williams a stand-out low-key performance.
What is it about celebrities and fans? What is the near paranoia one associates with the other and why is it almost the norm?
In the latest derange fan scenario, based on true events no less, Williams stars as a talk-radio personality named Gabriel No one, who reads stories he's penned over the airwaves and has accumulated an interesting fan in the form of a young boy named Pete Logand (Culkin) who has submitted a manuscript about the travails of his troubled youth to No one's editor Ashe (Morton) who gives it to No one to read for himself.
No one is naturally disturbed but ultimately intrigued about the nightmarish existence of Pete being abducted and sexually abused for years until he was finally rescued by a nurse named Donna (Collette giving an excellent performance) who has adopted the boy but her correspondence with No one reveals that Pete is dying from AIDS. Naturally No one wants to meet the fans but is suddenly in doubt to their possibly diverse ulterior motive so when the seed is planted by his estranged lover Jess (Cannavele) whose sudden departure from their New York City apartment has No one in an emotional tizzy that has only now grown into a tempest in a teacup when she comes back some investigating into and Pete's backgrounds discover some truths that he didn't anticipate.
Written by Patricia Piccoli and directed by Stettner, the film is a somewhat slow-paced drama with Terrence Anderson and the film's director Stettner and based on a story about a fan's hoax found out has some Hitchcockian moments though unorthodox like an old-fashioned pot-boiler done right.
Williams gives a stand-out, low-key performance as the conflicted good-hearted personality who genuinely wants to believe that his number one fan is in fact real and does love him (the one thing that has escaped his own reality) and has some unsettling dreadful moments with the creepy Collette whose one physical trait I will leave unmentioned but underlines the desperation of her character that can rattle you to the core.
However the film runs out of gas and eventually becomes a bit repetitive and predictable despite a finely directed piece of hoodwink and mystery by Stettner, it pays to listen to your inner voice: be careful of what you hope for.

Total Counts of both sets 1 25000

0 25000

Name: sentiment, dtype: int64

=====



'that movie was sick'



'no way in hell, I would go see that again'



' this movie really takes you places, first I was in the theatre, then I was at the box office asking for a refund'

NATURAL LANGUAGE PROCESSING

CLEANING



Tokenize sentences



Normalizing/Lemmatizing words



Removing punctuation, whitespaces and other noise from sentences



Part of speech tagging



Remove words that contain numbers

FEATURE ENGINEERING

```
>>> filtered_tokens[1].vector
array([ 1.8371646,  1.4529226, -1.6147211,  0.678362, -0.6594443,
        1.6417935,  0.5796405,  2.3821278, -0.13260496,  0.5750932,
        1.5654806, -0.6938064, -0.59607106, -1.5377437,  1.9425622,
       -2.4552505,  1.2321601,  1.0434952, -1.5102385, -0.5787632,
        0.12855647,  3.6501784,  2.6160972, -0.5710199, -1.5221789,
        0.00629176,  0.22760668, -1.922073, -1.6252862, -4.226225,
       -3.495663, -3.312053,  0.81387717, -0.00677544, -0.11603224,
        1.4620426,  3.0751472,  0.35958546, -0.22527039, -2.743926,
        1.269633,  4.606706,  0.34034157, -2.1272311,  1.2619178,
       -4.209798,  5.452852,  1.6940253, -2.5972986,  0.95049495,
       -1.910578, -2.374927, -1.4227567, -2.2528825, -1.799806,
        1.607501,  2.9914255,  2.8065152, -1.2510269, -0.54964066,
       -0.49980402, -1.3882618, -0.470479, -2.9670253,  1.7884955,
        4.5282774, -1.2602427, -0.14885521,  1.0419178, -0.08892632,
       -1.138275,  2.242618,  1.5077229, -1.5030195,  2.528098,
       -1.6761329,  0.16694719,  2.123961,  0.02546412,  0.38754445,
        0.8911977, -0.07678384, -2.0690763, -1.1211847,  1.4821006,
        1.1989193,  2.1933236,  0.5296372,  3.0646474, -1.7223308,
       -1.3634219, -0.47471118, -1.7648507,  3.565178, -2.394205,
       -1.3800384 ], dtype=float32)
```

Review

"Please give this one a miss
and the rest of the cast
rendered terrible
performances, the show is
flat.."



Prediction

This seems to be a
"Negative" review!

03. OBSTACLES



- ❑ Text files data frame issues
- ❑ Data Size & RAM
- ❑ Google collab crashing
- ❑ First model score 0.6
- ❑ Data cleaning
- ❑ Spacy Library import complications
- ❑ Unable to use trained model on a test text

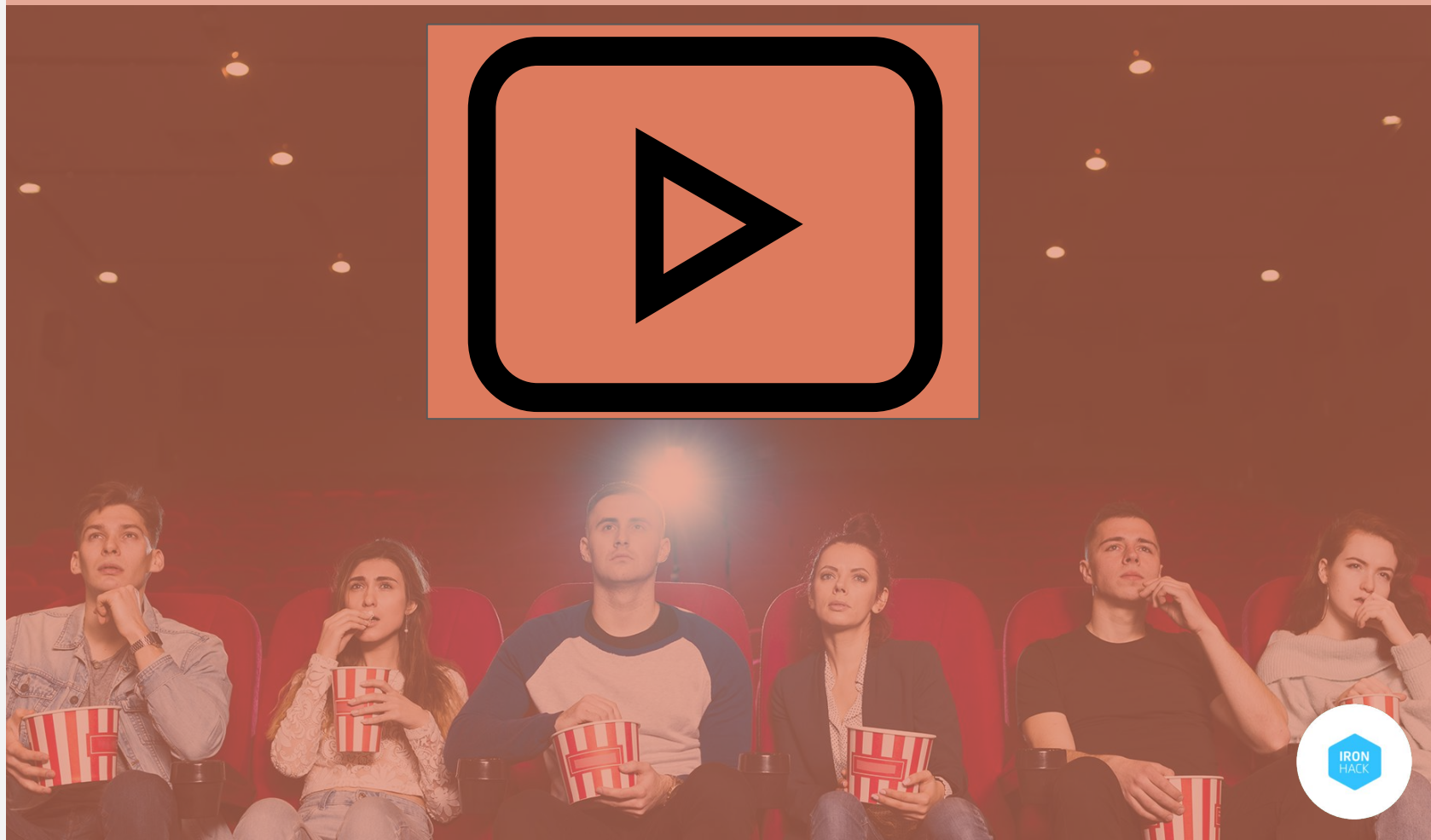




04. SOLUTIONS

- For overall flow of process I relied heavily on some notebooks, "Detecting Bad Customer Reviews with NLP" Johnathan Oheix
- Pickled dataframe often





A warm, intimate scene of a group of people dining together. A woman with blonde hair is the central figure, smiling and holding up a glass of white wine. To her left, another person's hand is visible holding a glass. In the foreground, a lit red candle in a holder and a bouquet of purple flowers are on the table. The background is softly blurred, showing a restaurant or bar setting with warm lighting.

THANK YOU

Any questions?