

---

Lord of the Rings 

VS



The Hobbits

---



# Agenda



01

## Introduction

About Tolkien  
Problem Statement

02

## Data & EDA

Data background  
Data cleaning  
Word cleaning  
Graphs

03

## Modeling

Naive bayes  
KNN  
Random Forest  
Logistics Regression  
Vectorizers

04

## Conclusion

And  
Recommendations





# 01

## Introduction



Spēāk FRIĒND  
and ENCĒR

# Introduction

## Tolkien

- CBE, FRSL, academic, genius
- Godfather of Fantasy writing
  - Created histories to characters and invented languages for the races
  - Started the practice of map drawing for fantasy land (Tolkien style)
  - Incorporation of Myth into Fantasy
  - Set the standard for how fantasy races are portrayed (Elves, Dwarves)



*J.R.R. Tolkien.*



# Introduction

Amazon bought the television rights for The Lord of the Rings for **US\$250 million in November 2017**, making a five-season production commitment worth at least US\$1 billion. This would make it the most expensive television series ever made.

## Problem Statement

There is an increase interest in Tolkien works recently and LOTR fans are revisiting topics on the two trilogy done by Peter Jackson: LOTR and The Hobbits.

The Rings of Power production team wants to help LOTR fans with differentiating the movies (LOTR and The Hobbits) in the online discussions. In that, LOTR fans will get the correct base knowledge required before the series is released.

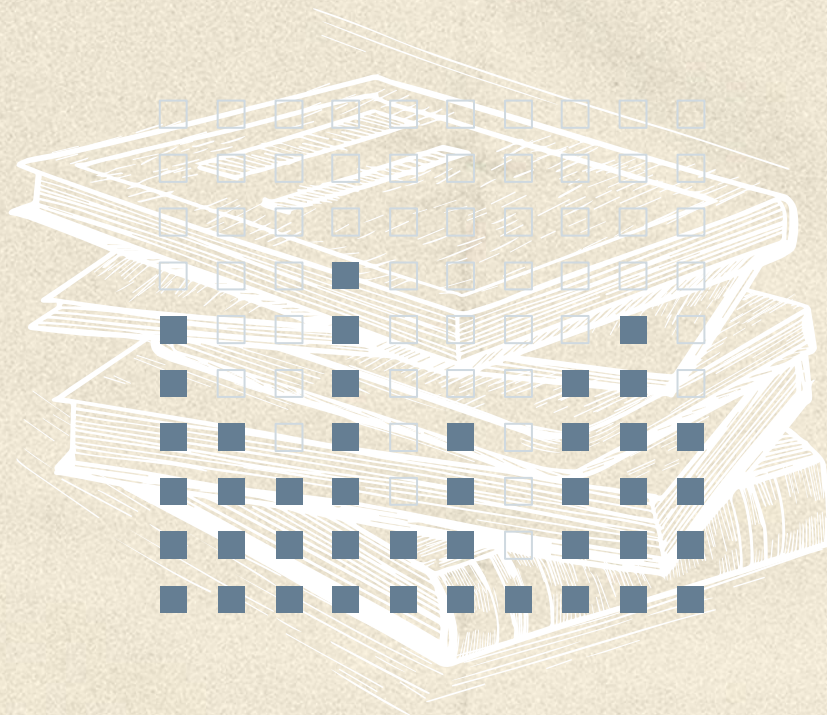


### Note:

Fans who only watched Percy Jackson movies are referred as LOTR fans  
While, hardcore Tolkien nerds who read the books are referred as Tolkien fans

# 02

## Data & EDA





# Data Gathering



## Reddits

A network of communities where people can dive into their interests

lordoftherings

TheHobbit



## Web Scrapping

1800 submissions for each topic  
Using 'before' parameter: 19 July 2022

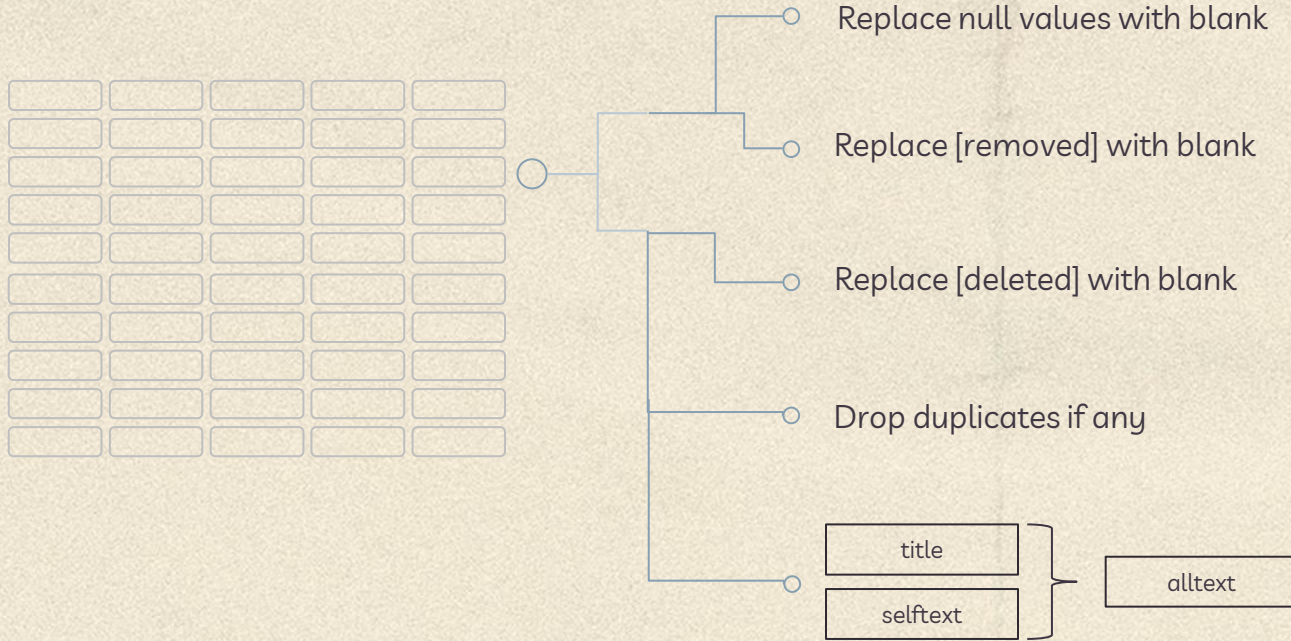
## Submissions



## Subreddits, Title, Selftext



# Data Cleaning





# Words Cleaning

Hi redditor ~~I am~~ currently running ~~a~~ LOTR Discussion forum ~~..~~. If ~~you are~~ interested ~~in such~~ topic ~~I will~~ recommend ~~you~~ go check out our ad below ! ? . .

~~[https://discord.gg/6cf2QbUuMg](https://discord.gg/6cf2QbUuMg)~~

Hi redditor currently running LOTR Discussion forum if interested

topic recommend go check out our ad below

hi redditor currently running discussion forum if interested

topic recommend go check out our ad below

Remove HTML

Remove links

Extract only words using regex

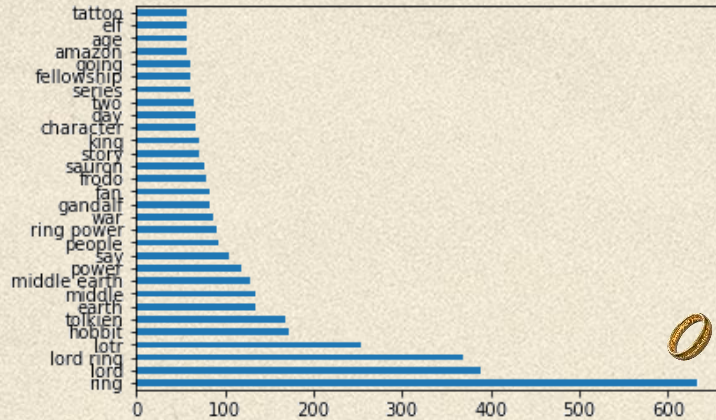
Remove stop words

Lowercase

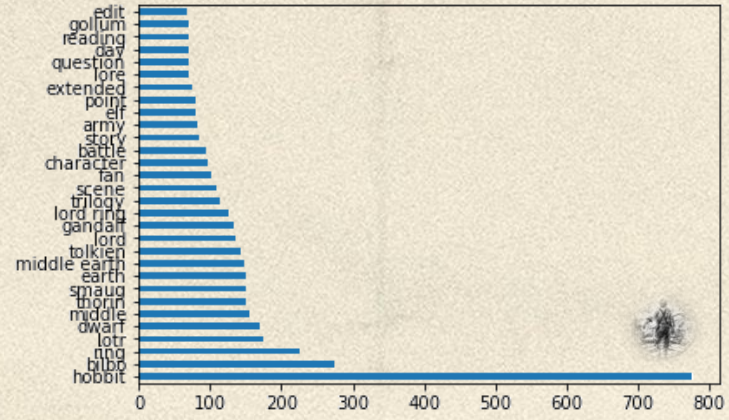
Lemmatize the words

# EDA

## LOTR



## Hobbits



Remove common words

ring

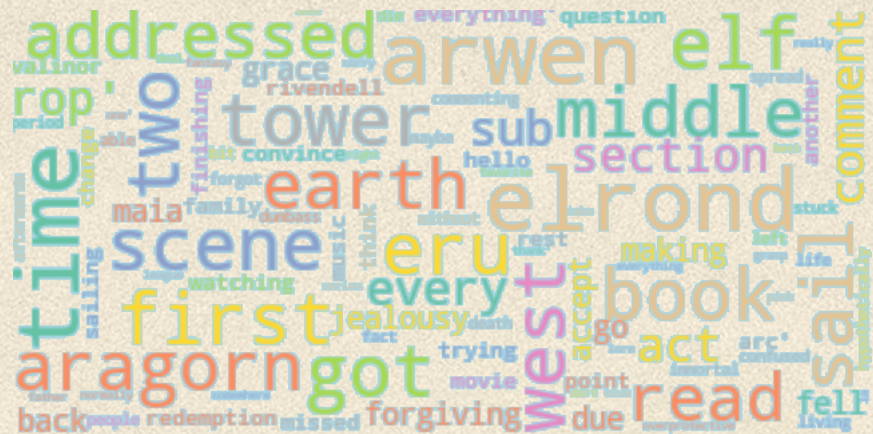
tolkien

middle earth

lord



# Word Cloud



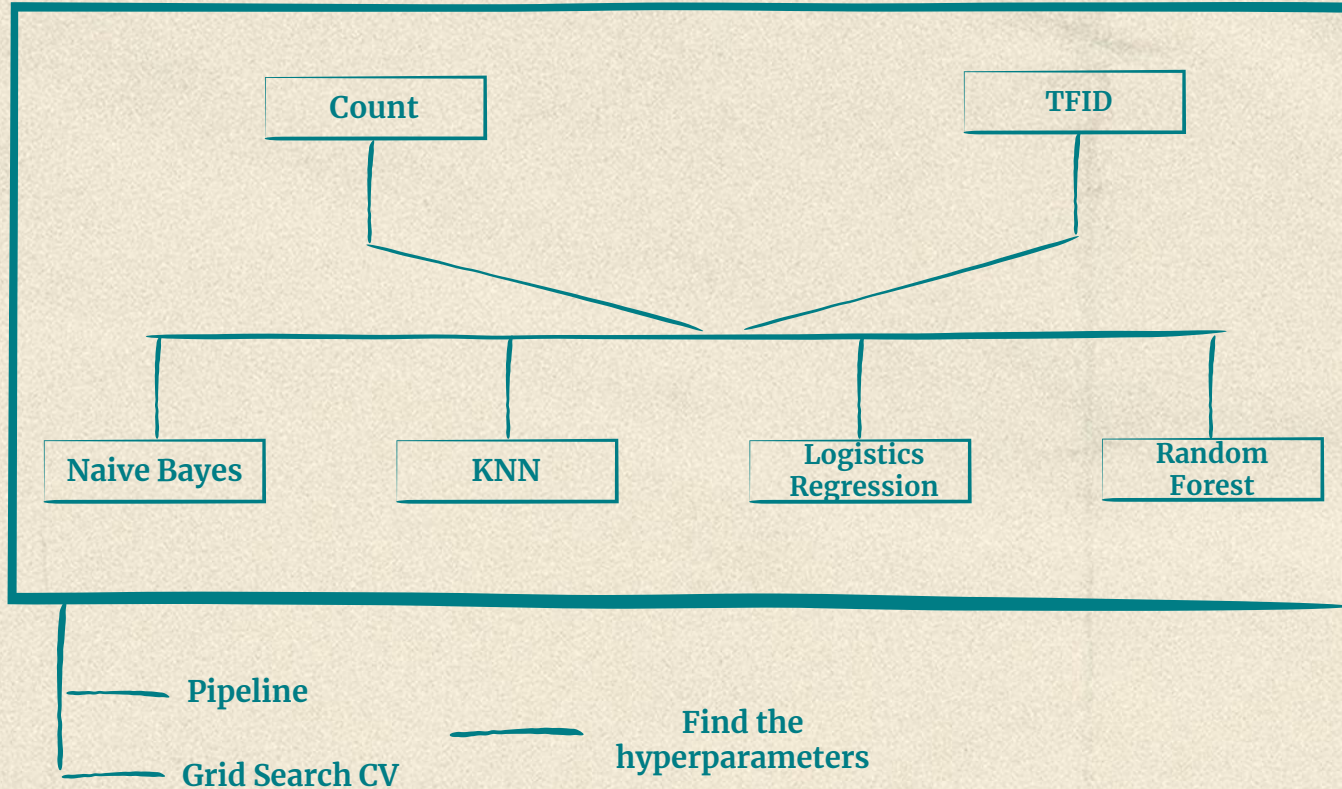
# 03

## Modeling



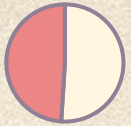


# Models & Vectorizers



# Results

|       |  | Model         | Best score | Train score | Test Score | Accuracy | Specificity | Recall | Precision | F1 Score |
|-------|--|---------------|------------|-------------|------------|----------|-------------|--------|-----------|----------|
| Count |  | Baseline      | 0.509      | 0.509       | 0.509      | 0.509    | 1.000       | 0.000  | 0.000     | 0.000    |
|       |  | Naive Bayes   | 0.701      | 0.798       | 0.703      | 0.703    | 0.758       | 0.647  | 0.721     | 0.682    |
|       |  | KNN           | 0.670      | 0.780       | 0.653      | 0.653    | 0.682       | 0.622  | 0.654     | 0.637    |
|       |  | Logistics Reg | 0.709      | 0.904       | 0.705      | 0.705    | 0.778       | 0.628  | 0.732     | 0.676    |
|       |  | Random Forest | 0.717      | 0.766       | 0.693      | 0.693    | 0.857       | 0.523  | 0.779     | 0.626    |
| TFID  |  | Naive Bayes   | 0.709      | 0.875       | 0.715      | 0.715    | 0.746       | 0.683  | 0.722     | 0.702    |
|       |  | KNN           | 0.585      | 0.730       | 0.607      | 0.607    | 0.685       | 0.526  | 0.617     | 0.568    |
|       |  | Logistics Reg | 0.723      | 0.868       | 0.714      | 0.714    | 0.787       | 0.637  | 0.743     | 0.686    |
|       |  | Random Forest | 0.714      | 0.769       | 0.693      | 0.693    | 0.857       | 0.523  | 0.779     | 0.626    |



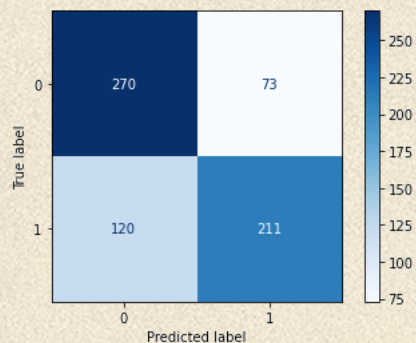
LOTR HOBB



# Results

| Model         | Best score | Train score | Test Score | Accuracy | Specificity | Recall | Precision | F1 Score |
|---------------|------------|-------------|------------|----------|-------------|--------|-----------|----------|
| Logistics Reg | 0.723      | 0.868       | 0.714      | 0.714    | 0.787       | 0.637  | 0.743     | 0.686    |

Confusion matrix



Useful words for Hobbits

|               | coef       |
|---------------|------------|
| <b>hobbit</b> | 222.263938 |
| <b>bilbo</b>  | 21.084557  |
| <b>gollum</b> | 14.443830  |
| <b>smaug</b>  | 9.261529   |
| <b>dwarf</b>  | 6.732842   |
| <b>thorin</b> | 4.819926   |
| <b>edit</b>   | 3.476050   |
| <b>bombur</b> | 3.453213   |

Useful words for LOTR

|                   | coef     |
|-------------------|----------|
| <b>ring</b>       | 0.111069 |
| <b>lord</b>       | 0.155078 |
| <b>frodo</b>      | 0.175312 |
| <b>aragorn</b>    | 0.249595 |
| <b>lord ring</b>  | 0.268696 |
| <b>mordor</b>     | 0.270472 |
| <b>fellowship</b> | 0.276850 |
| <b>power</b>      | 0.288995 |
| <b>eye</b>        | 0.294030 |
| <b>galadriel</b>  | 0.312845 |

# 04

## Conclusion & Recommendation





# Conclusions

- Fans should use the TFID Vectorizer, Logistic Regression model to differentiate if the post is related to LOTR or The Hobbits.
- The main characters in the movie are important to differentiate the posts.  
Dwarf characters such as Thorin, Bombur are prominent in The Hobbits movie as well as in the online discussions.  
While, LOTR discussions are mainly about the foes (black riders, eye, Morgoth).
- Characters such as Bilbo, Gandalf are in existence in both movies. Hence, it is best for the production team to state that clearly during fans meet and greet session / conference to clear the confusion.



# Recommendations

- Production team to work with YouTubers who specialize in recaps / Tolkien works.  
In order to provide bite sized videos on the background for LOTR fans before the upcoming series.



- Also, recommend the production team to work with streaming platform (mainly Amazon Prime) to stream LOTR and the Hobbits trilogy.



- Future work will recommend for the production team to use other online discussion forums to gather data.



And of course...  
the best recommendation will be...

**WATCH LOTR & HOBBITS**



# Thanks



“All that is gold does not glitter,  
Not all those who wander are lost;

...

...

Renewed shall be blade that was broken,  
The crownless again shall be king.”



– J.R.R Tolkien

---