



# Distinguishing Helpful User Reviews

An NLP classification analysis of book reviews by Andrew Dettor

# Why?

- Save time
  - Books especially
- Find what they want
- More informative



Suzanne rated it ★★★★★ · review of another edition  
Shelves: british-obsessions, books-i-own

Jul 02, 2007

"I'm going to keep going until I succeed — or die. Don't think I don't know how this might end. I've known it for years." — Harry Potter

Most seventeen-year olds don't view the possibility of an early death as being, well, possible. But then again, most seventeen-year olds haven't come face-to-face with death almost half a dozen times before their first kiss either.

In *Harry Potter and the Deathly Hallows*, the seventh and final installment of the ridiculously popular Harry Potter series, J.K. Rowl [...more](#)

1383 likes · [Like](#) · [see review](#)



Kassidy rated it ★★★★★ · review of another edition  
Shelves: favorites

Jun 21, 2014

It's hard for me to believe that I finished the Harry Potter series... I wish this series would go on forever!

I don't know how I've gone through my life without these books. I regret not reading it when I was younger, but I am so happy that I finally did, better late than never, right?

The world built in these books is so real and tangible, it is truly a masterpiece. The characters are engaging and brilliant, and I love how the plot is so interwoven and twisted within itself, especially in this I [...more](#)

1303 likes · [Like](#) · [see review](#)



Lily rated it ★★★★★

Oct 21, 2019

I can't believe its over.... I've FINALLY read the Harry Potter series. And WHAT an ending to the series this book was..

Video Review: [https://youtu.be/aeFPuw\\_Q3n4](https://youtu.be/aeFPuw_Q3n4) [...more](#)

699 likes · [Like](#) · [see review](#)



Jayson rated it ★★★★★ · review of another edition  
Shelves: 600-plus-pp, genre-juvenile-fantasy, genre-dystopian, read-in-2009, genre-young-adult, author-british, format-illustrated, subject-boarding-school

Aug 01, 2013

(A) 86% | Extraordinary

Notes: It ends too expository and convenient, but the story's compelling, and its new settings are a nice break from formula. [...more](#)

552 likes · [Like](#) · [see review](#)

# Data Collection

- Goodreads.com
- Lists
- Pages
- Books
- Reviews
- html parsing

goodreadsHomeMy BooksBrowseCommunitySearch booksSign InJoin

Listopia

Best Books of the 20th Century

The best books published during the 20th century (January 1st, 1901 through December 31st 2000)

See also

Most Rated Book By Year

Best Twentieth Century American Novels

Le Monde's 100 Books of the Century

Best Fantasy of the 20th Century

Best Forgotten Fantasy of the 20th Century

Best Romance Novels of the Twentieth Century

Best Science Fiction of the 20th Century

Best Forgotten Science Fiction of the 20th Century

Best Books By Millennium:

1000-1999

Best Books By Century:

21st, 20th, 19th, 18th, 17th, 16th, 15th, 14th, 13th, 12th, 11th, 10th, 9th, 8th, 7th, 6th, 5th, 4th

Best Books by Decade


2010-2019, 2000-2009, 1990-1999, 1980-1989, 1970-1979, 1960-1969, 1950-1959, 1940-1949, 1930-1939, 1920-1929, 1910-1919, 1900-1909, 1890-1899, 1880-1889, 1870-1879, 1860-1869, 1850-1859, 1840-1849, 1830-1839, 1820-1829, 1810-1819, 1800-1809

Lists for all books by Number of Ratings:

80,000 to 99,999 - 100,000 to 199,999 - 200,000 to 499,999 - 500,000 to 999,999 - 1,000,000 and more

All VotesAdd Books To This List


1



**To Kill a Mockingbird**  
by Harper Lee  
★★★★★ 4.27 avg rating — 5,039,586 ratings  
score: 978,933, and 9,911 people voted

Want to ReadRate this book★★★★★


2



**1984**  
by George Orwell  
★★★★★ 4.19 avg rating — 3,622,002 ratings  
score: 905,566, and 9,243 people voted

Want to ReadRate this book★★★★★


3



**Harry Potter and the Sorcerer's Stone (Harry Potter, #1)**  
by J.K. Rowling  
★★★★★ 4.67 avg rating — 8,059,885 ratings  
score: 712,208, and 7,287 people voted

Want to ReadRate this book★★★★★


4



**The Great Gatsby**  
by F. Scott Fitzgerald  
★★★★★ 3.93 avg rating — 4,230,159 ratings  
score: 701,535, and 7,199 people voted

Want to ReadRate this book★★★★★


5



**Animal Farm**  
by George Orwell  
★★★★★ 3.97 avg rating — 3,114,682 ratings  
score: 497,882, and 5,218 people voted

Want to ReadRate this book★★★★★

6




**The Hobbit, or There and Back Again**  
by J.R.R. Tolkien  
★★★★★ 4.52 avg rating — 2,114,682 ratings  
score: 497,882, and 5,218 people voted

Want to ReadRate this book★★★★★

Search listsSearch

RELATED NEWS



**A Debut Novelist's 2021 Reading in Review**

Mateo Askaripour is a Brooklyn-based writer whose first novel, *Black Buck*—which Colson Whitehead calls a "mesmerizing novel, executing a high..."

Read more...

Create New List  
All Lists  
Lists I Created  
Lists I've Voted On  
Lists I've Liked

Anyone can add books to this list.

MY VOTES

vote now

FRIENDS VOTES

get your friends to vote



## What makes reviews helpful?

- Reviewer expertise
- Timeliness of review
- Linguistic features
  - Subjectivity
  - Writing style

Mary

Follow

Add friend

More ▾

Sign in to Goodreads to learn more about Mary.

Sign In

140 ratings (3.41 avg)

119 reviews

Sources: [9, 17, 5, 13, 12]

# Defining helpfulness

- “X out of Y reviewers liked this review”
  - Proportion [14, 15, 5]
  - Not anymore?
- Classification
  - “Natural” cutoff [14]
  - Empirical [5]
- Regression [15, 5]
- Goodreads
  - discretization

★☆☆☆☆ Not happy, didn't perform

By [Amazon Customer](#) on August 2, 2016

Color: Blue | **Verified Purchase**

This product did not work as advertised. I'm disappointed with my purchase

► [Comment](#) | One person found this helpful. Was this review helpful to you?   [Report abuse](#)

★★★★★ Best purchase ever!

By [Amazon Customer](#) on July 27, 2016

Color: Blue

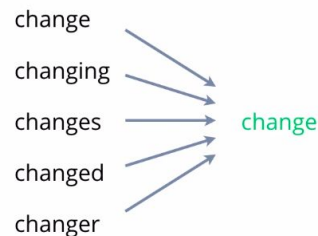
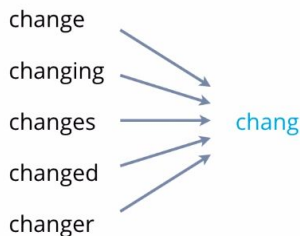
This is the best product ever, I highly recommend any one reading this review to purchase one immediately!

► [Comment](#) | One person found this helpful. Was this review helpful to you?   [Report abuse](#)

# Traditional NLP Methods

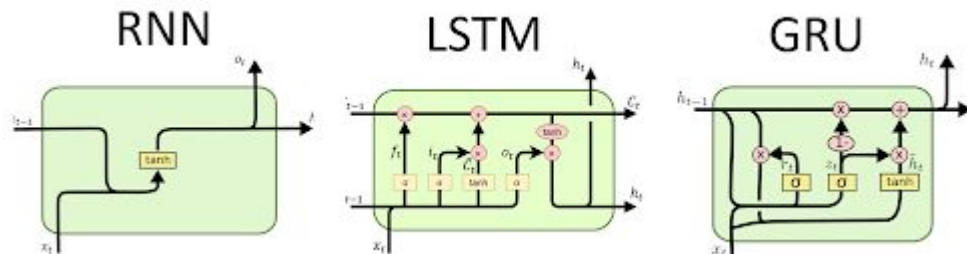
- Preprocessing
  - Lemmatization, stemming, word segmentation, part of speech tagging, tokenization, etc [4, 15, 2, 9, 1]
- Lexicon
  - LDA [6, 15]
- Involved [8]
- Expert knowledge [8]

## Stemming vs Lemmatization



# Using Deep Learning

- Feature engineering [8]
- Good performance [3, 2, 7, 11, 14]
  - Only text [14]
  - Other features [2]
- CNN, LSTM, GRU [7, 2, 14, 11]
- Poor interpretability



# Universal Language Model Fine-tuning [7]

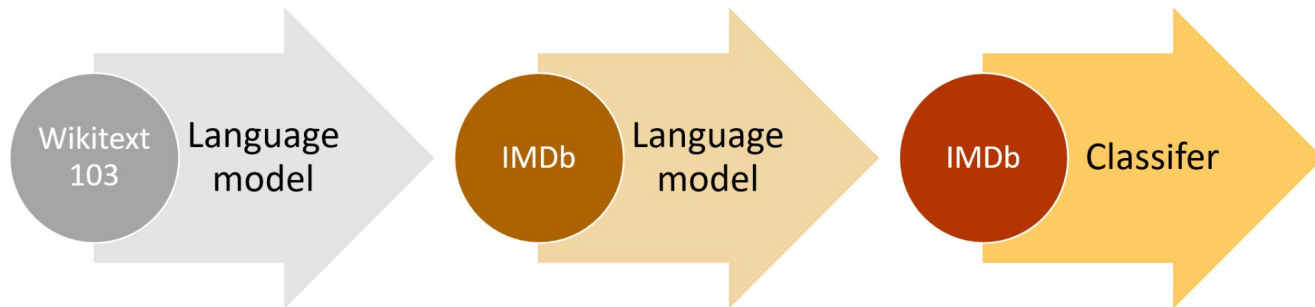
- Pretrained language model
  - Transfer learning
- Domain specific language model
- Classifier
- Better performance
- Less data





# Language Modelling and Text Classification

- AWD\_LSTM
  - Wikipedia
- Training process
- Freezing/unfreezing





# Data Preprocessing

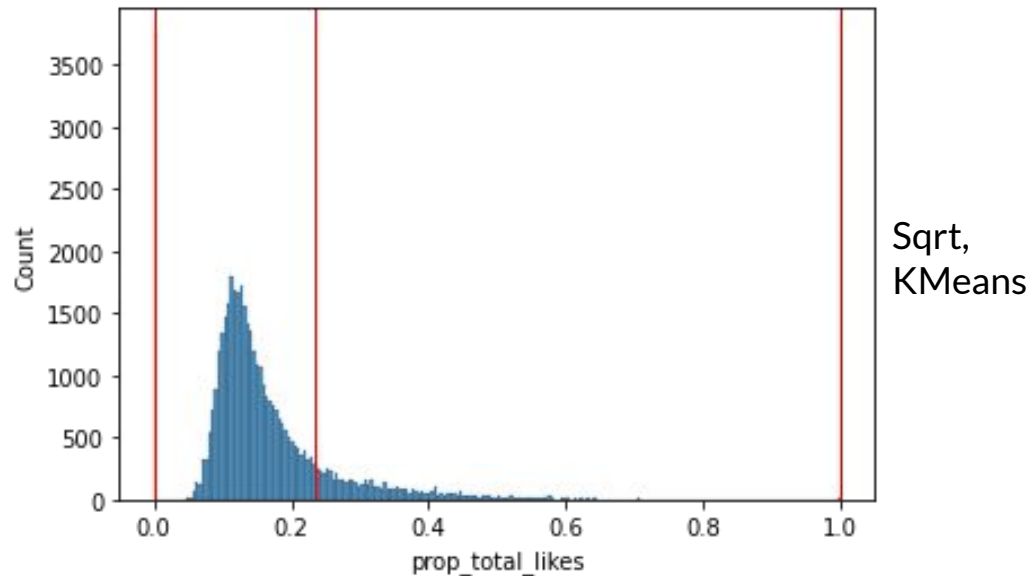
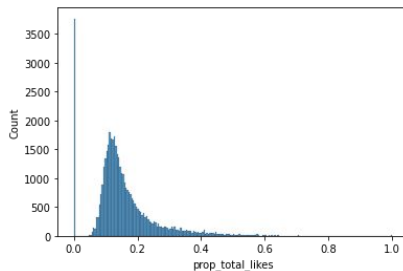
- Web scraping errors
- Not English
- Tokenization
- Numericalization
- Embedding matrix

xxbos i was so looking forward to getting my hands on the last of  
xxmaj the xxmaj dragon xxmaj knight 's and reading xxmaj adam 's  
and xxmaj margaret 's story . xxmaj as readers of the series will  
know , the series starts with xxmaj margaret 's death , but what  
xxmaj adam thought was the end of their love , was in fact , only the  
beginning . i am

```
defaults.text_spec_tok = [UNK, PAD, BOS, EOS, FLD, TK_REP, TK_WREP, TK_UP, TK_MAJ]
defaults.text_proc_rules = [fix_html, replace_rep, replace_wrep, spec_add_spaces, rm_useless_spaces,
                           replace_all_caps, replace_maj, lowercase]
defaults.text_postproc_rules = [replace_space]
```

# Different meanings of “helpful”

- Transformation
  - Square root
- Discretizations
  - 2 quartiles
  - 4 quartiles
  - KMeans clustering
  - Histogram binning





# Performance

- Language Model
  - 31% accuracy
- Classification
  - Poor performance

	0	1	micro-average
accuracy	0.573868	0.573868	0.573868
precision	0.570823	0.576018	0.573868
recall	0.487210	0.655392	0.573868
f1	0.525713	0.613146	0.573868

No sqrt,  
median

	0	1	2	3	micro-average
accuracy	0.722190	0.755995	0.386010	0.606738	0.617733
precision	0.121212	0.333333	0.245542	0.222982	0.235466
recall	0.016744	0.000480	0.686202	0.227760	0.235466
f1	0.029424	0.000958	0.361668	0.225346	0.235466

Sqrt, 4  
quartiles



# Conclusion

- Need more than just text
- Usefulness of ULMFiT
- CNN? GRU?
- Good language model
- Bad classifier



# Citations

- [1] to [21] are on the project report bibliography
- Each picture is cited in their respective slide's speaker notes