

A close-up photograph of a woman with dark hair and glasses, looking directly at the camera with a slight smile. She is holding a white book cover in front of her face, which serves as the background for the title text.

Predicting NYT Best Sellers

Emily Ng

What does it mean to be a NYT Best Seller?

- Indicative of Quality
- Free Marketing
- Elevated Author Status
- Increased Future Book Deals
- Increased Chances of Movie Adaptations



▲ 14%

First Year
Sales

▲ 57%

First Time
Author Sales

Can you
optimize your
book for the
best chance to
get on the list?



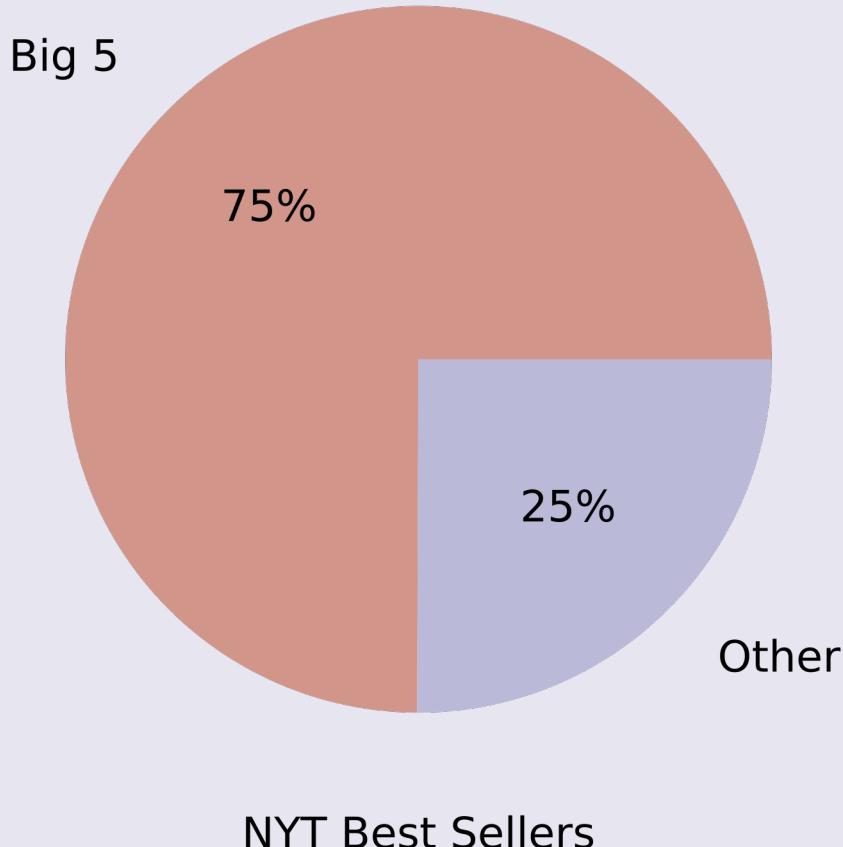
Data

kaggle™
goodreads

The
New York
Times



Publisher



Penguin &
Random House



Harper
Collins



SIMON &
SCHUSTER

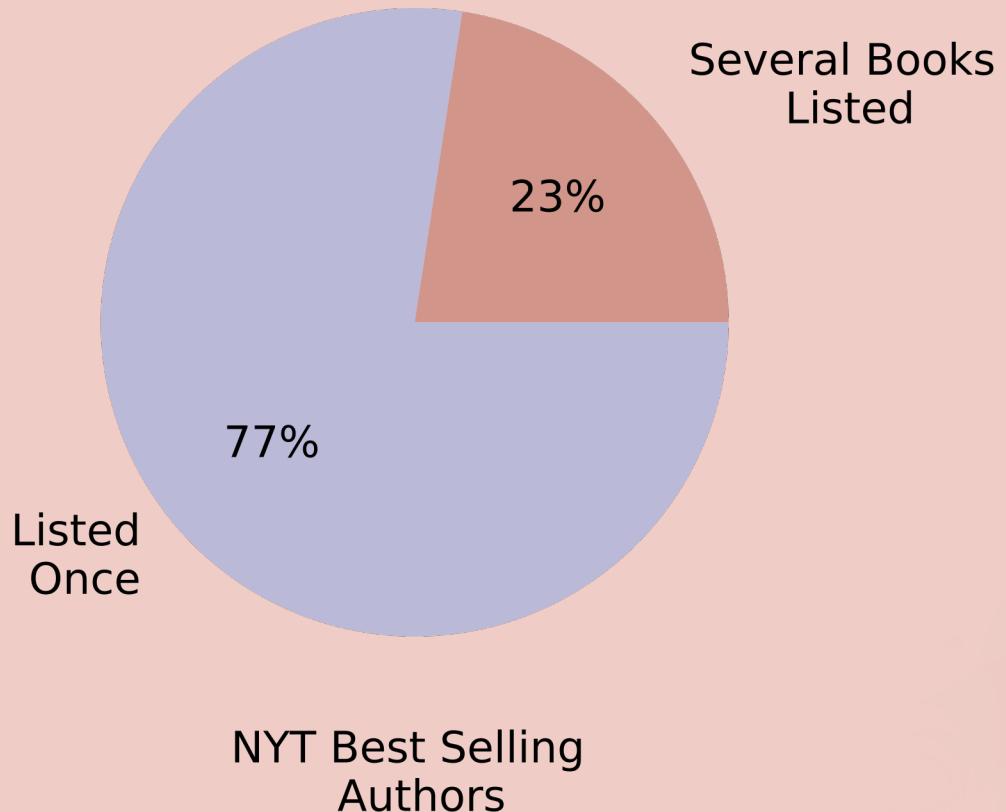


Hachette



MACMILLAN

Author



Genre

01 Fantasy



02 Fiction



03 Historical Fiction



04 Mystery



05 Nonfiction



06 Romance



07 Science Fiction



08 Young Adult



A close-up photograph of a person's hand resting on a stack of books. The hand is positioned vertically, with the fingers slightly spread. A small, intricate ring is visible on the middle finger. The background consists of numerous stacked books, creating a textured, layered effect.

Model

Strategy

Step 1

Address class imbalance



Class Imbalance

10%

Best Sellers

90%

Non Best
Sellers

Strategy

Step 1

Address class imbalance

Under-sampling with 9:1 ratio

Step 2

Compute TF-IDF on Words

Strategy

Step 1

Address class imbalance

Under-sampling with 9:1 ratio

Step 2

Compute TF-IDF on Words

TfidfVectorizer on Description

Step 3

Extract Top 25 Words

Lasso Logistic Regression on TF-IDF Features

Step 4

Compile Features

Define Feature Space

Step 5

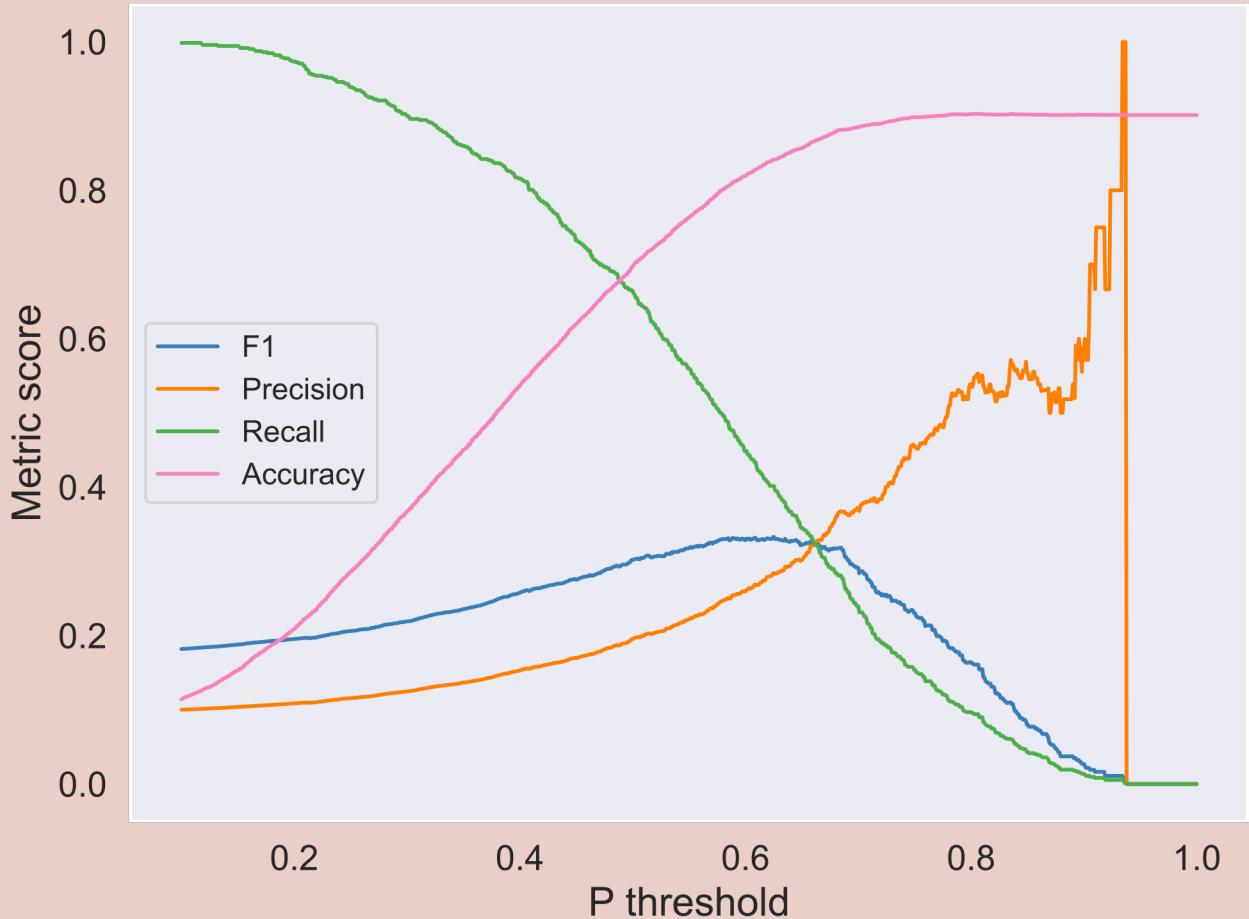
Fit Model

Random Forest Classifier

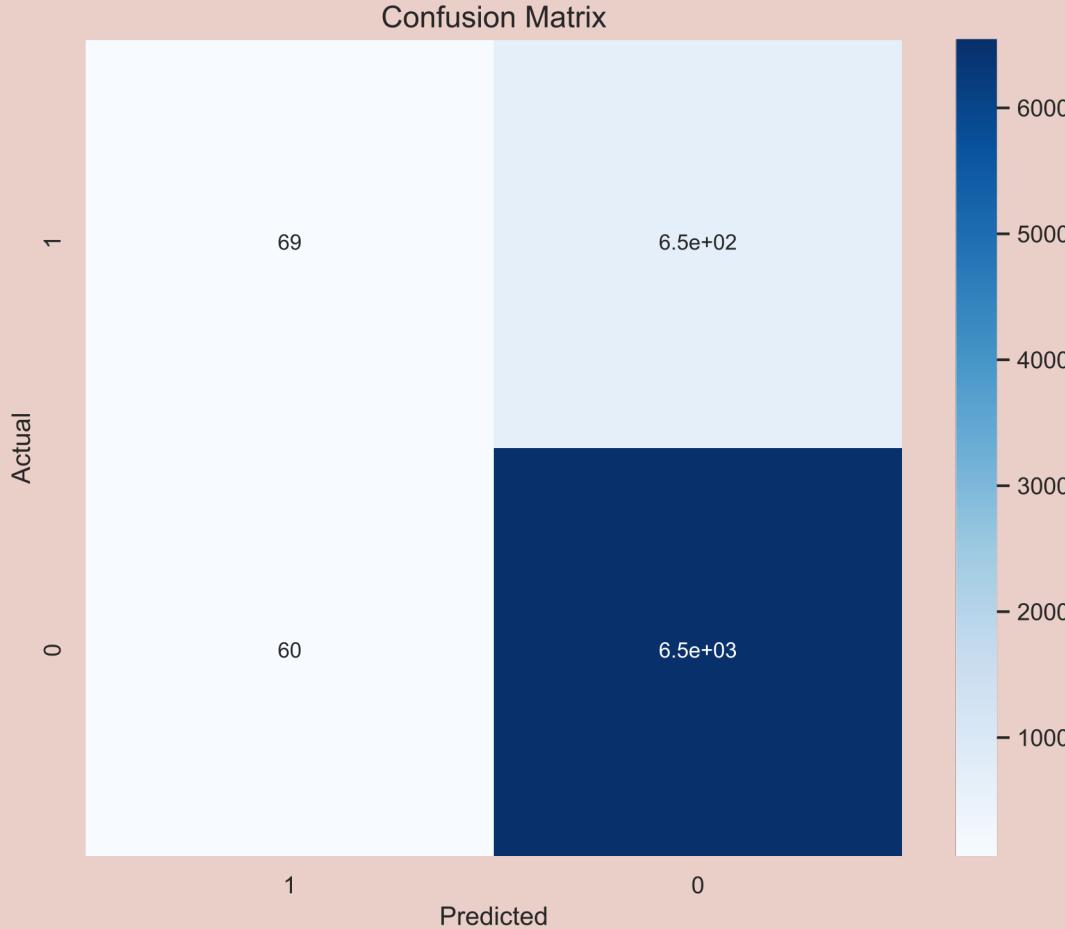
Results



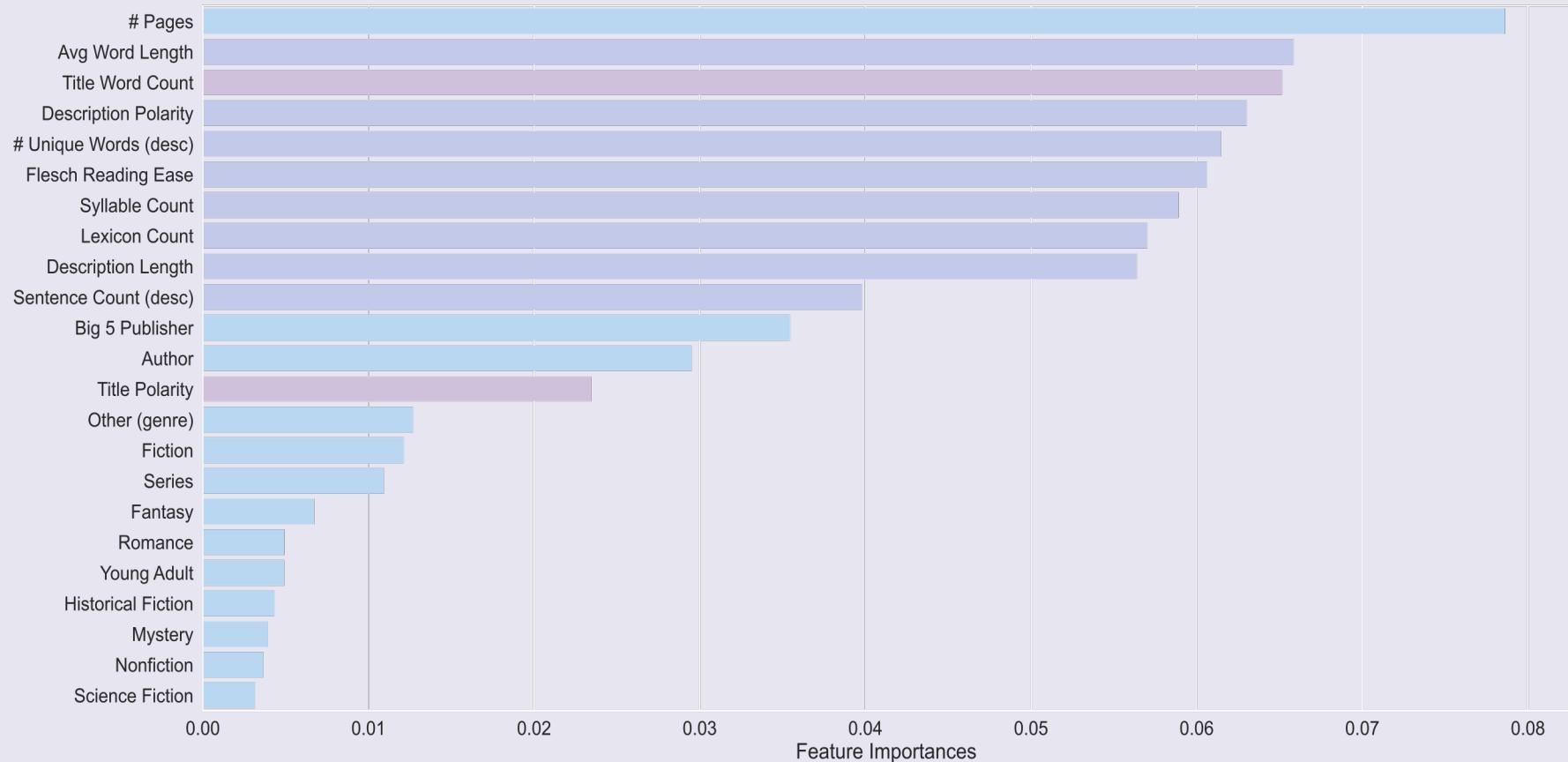
Metric Scores vs. Positive Class Decision Probability Threshold



Confusion Matrix at P = 0.8



Book Feature Importances





Underlying Themes

reader case york classic
father novel vampire volume come
bestselling family know woman
secret year want collection dan yang
past wife god daughter killer



New York Times Best Seller Predictor

Will your book become the next best seller?

Next Steps

01

Deeper look into how features influence outcome.

02

Expand model to include more book content. Investigate plot structure and literary analysis.



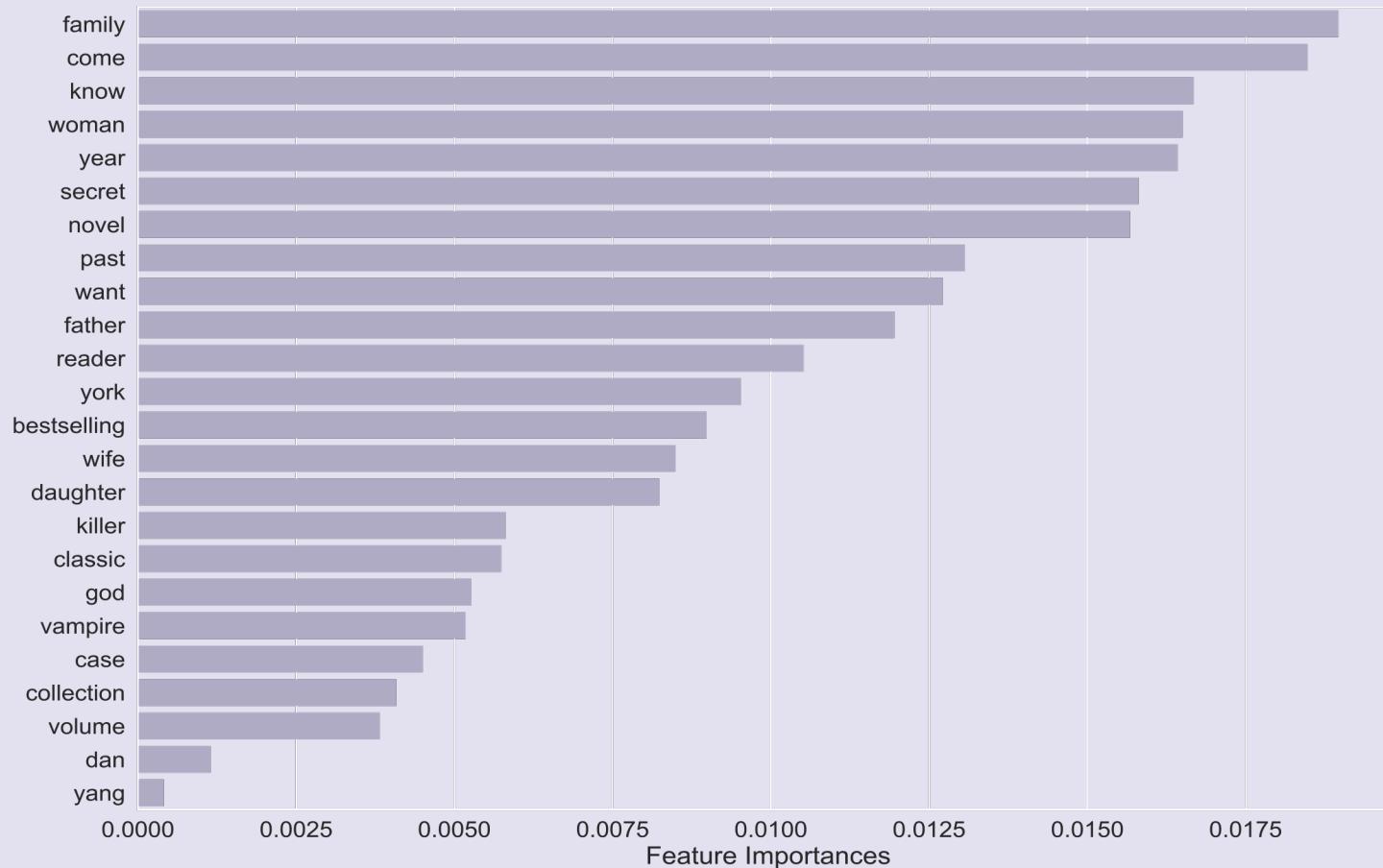
CREDITS: This presentation template was created by Slidesgo,
including icons by Flaticon, and infographics & images by Freepik.

Please keep this slide for attribution.

Thanks!

Does anyone have any questions?

Word Importances



TF-IDF

TF-IDF : Term Frequency – Inverse Document Frequency

Appendix: TF-IDF

- Woman reading a book on picnic blanket
- Crop student reading book in park
- Girl reading a book face up
- Young woman reading book on rest
- Dad celebrates fathers day with daughters
- Smiling women reading books in tent
- Person choosing books at the bookstore
- Multiracial teen friends reading book near fence
- Stack of vintage hardcover books
- Top view of some books

If you want to apply the same filter as the pictures already included in the template, follow these instructions.

In Google Slides / PowerPoint:

1. select the image, click Format options, open the Recolor drop-down menu and choose Grayscale
2. Create a new shape with the same shape as the photo to which you want to apply the filter.
3. Select the color for the filter (#ce6753) and reduce the opacity until you're happy with the result.
4. If you want to modify the photo, move the filter to a side first.

