



A Publisher Dilemma

Divya Uppal

Feb 23, 2022

The standard publishing workflow for journal manuscripts



Author selects
a journal



Author prepares
manuscript



Author submits
manuscript



Publisher places
manuscript into peer
review & revision



Publisher
accepts or
rejects
manuscript



Publisher offers
final revisions for
the manuscript



Publisher places
manuscript into
copyediting and
proofreading



Publisher publishes
the finalized work

Project Goal

Develop a model to predicts book rating depending on Author Past Performance.
Non Goal: Content of the manuscript

Target: Rating of the Book

Author Features:

• Average Rating	• Number of Followers
• Year of Birth	• Percentage Presence
• Country of Birth	• $(\text{Number_Rating}) * (\text{Number_Review})$
• Number of Rating	• Number of Review
• Distinct Work	

The Process

01

Gather Data

02

EDA

03

Feature Engg

04

Final Model

01

goodreads

**Meet your next
favorite book.**









Gather Data

The Data

Best Children's Books

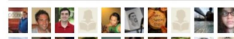
Cast your vote for the best children's books.

(If you love children's books, here's some groups you can check out: [Caldecott books](#), [Newbery books](#).)

	All Votes	Add Books To This List
1	 Where the Wild Things Are by Maurice Sendak ★★★★★ 4.22 avg rating — 906,753 ratings score: 114,640, and 1,173 people voted	Want to Read Rate this book ★★★★★
2	 The Very Hungry Caterpillar by Eric Carle ★★★★★ 4.30 avg rating — 447,153 ratings score: 108,872, and 1,118 people voted	Want to Read Rate this book ★★★★★
3	 The Giving Tree by Shel Silverstein ★★★★★ 4.37 avg rating — 1,004,542 ratings score: 106,239, and 1,090 people voted	Want to Read Rate this book ★★★★★
4	 Green Eggs and Ham by Dr. Seuss ★★★★★ 4.30 avg rating — 691,991 ratings score: 98,713, and 1,025 people voted	Want to Read Rate this book ★★★★★
5	 Goodnight Moon by Margaret Wise Brown ★★★★★ 4.29 avg rating — 338,036 ratings score: 79,793, and 826 people voted	Want to Read Rate this book ★★★★★
6	 Charlotte's Web by E.B. White ★★★★★ 4.18 avg rating — 1,600,438 ratings score: 74,344, and 784 people voted	Want to Read Rate this book ★★★★★



MAURICE SENDAK'S FOLLOWERS (1,756)



Maurice Sendak

Born	in Brooklyn, New York, The United States
Died	June 10, 1928
Genre	Arts & Photography, Children's Books, Science Fiction & Fantasy
Influences	William Blake, Antoine Watteau, Francisco Goya, Winsor McCay ...more

[edit data](#)

Maurice Bernard Sendak was an American writer and illustrator of children's literature who is best known for his book *Where the Wild Things Are*, published in 1963. An elementary school (from kindergarten to grade five) in North Hollywood, California is named in his honor.




Sendak was born in Brooklyn, New York, to Polish-Jewish immigrant parents, and decided to become an illustrator after viewing Walt Disney's film *Fantasia* at the age of twelve. His illustrations were first published in 1947 in a textbook titled *Atomics for the Millions* by Dr. Maxwell Leigh Eidinoff. He spent much of the 1950s working as an artist for children's books, before beginning to write his own stories.

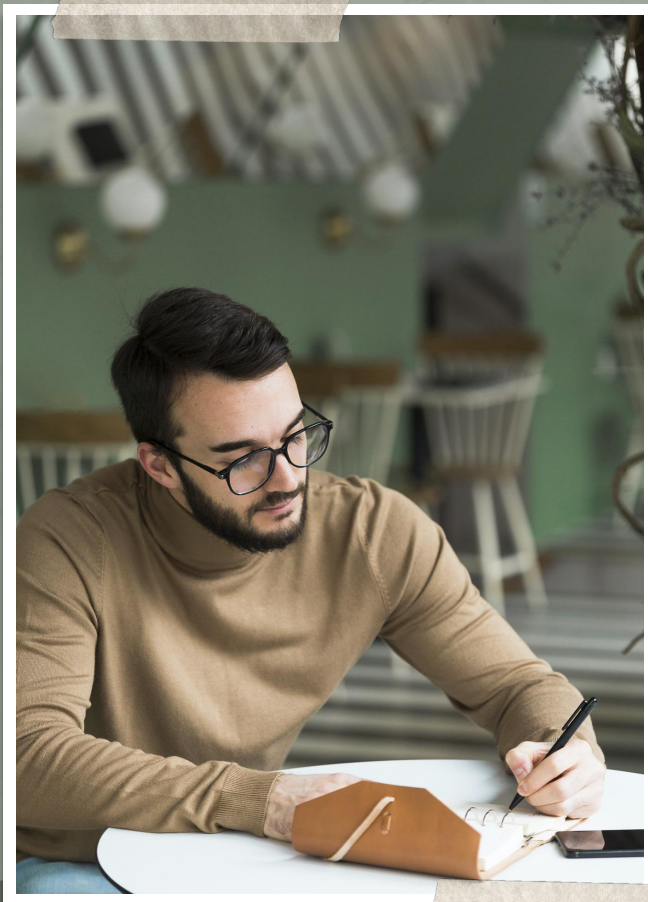
[...more](#)

MAURICE SENDAK'S BOOKS

[Combine Editions](#)

Average rating: 4.21 - 1,167,571 ratings - 22,374 reviews - 195 distinct works • Similar authors

 Where the Wild Things Are ★★★★★ 4.22 avg rating — 906,679 ratings — published 1963 — 142 editions	Want to Read Rate this book ★★★★★
 Chicken Soup with Rice: A Book of Months ★★★★★ 4.23 avg rating — 30,679 ratings — published 1962 — 33 editions	Want to Read Rate this book ★★★★★
 In the Night Kitchen ★★★★★	Want to Read

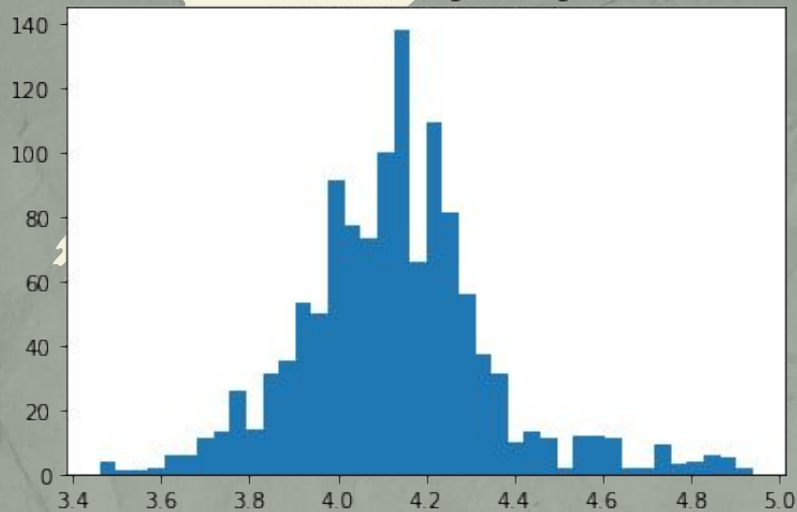


02

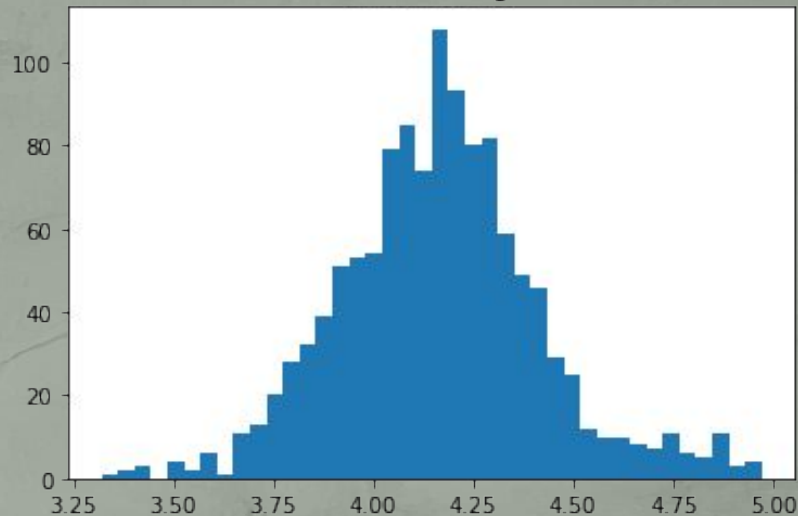
EDA

Data Distribution

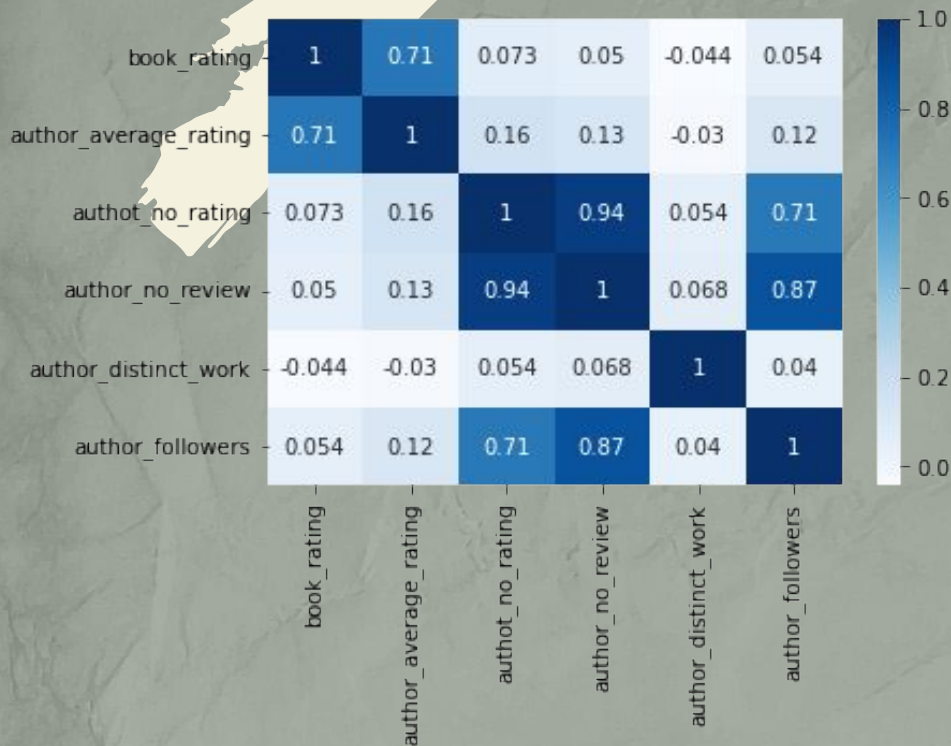
Author Average Rating



Book Rating



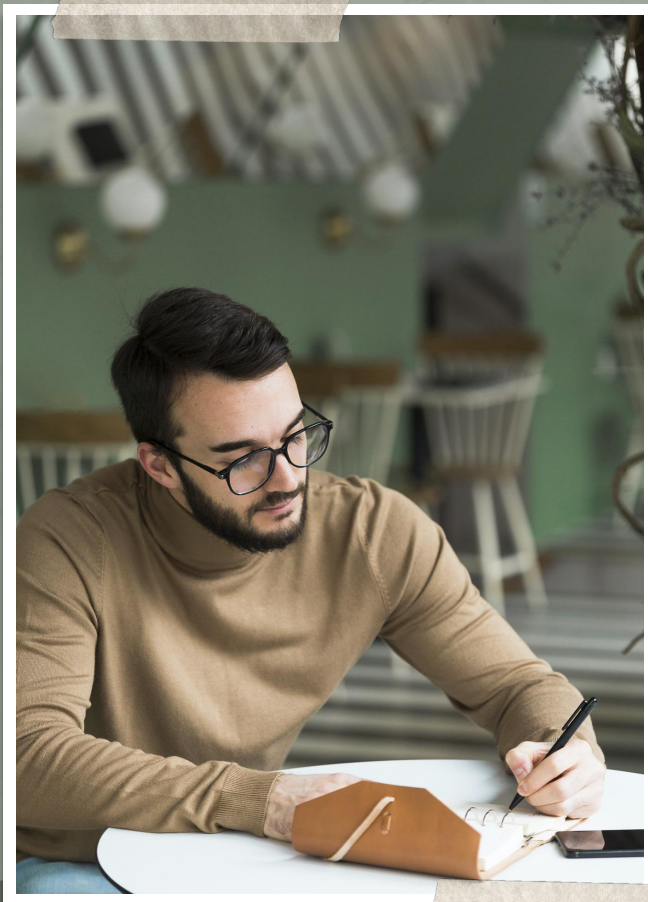
Correlation Matrix



Strongest Feature correlation*:

- Author Average rating: 0.71
- Author No of Rating: 0.073
- Author No of Followers: .054

*Only top correlated features are displayed

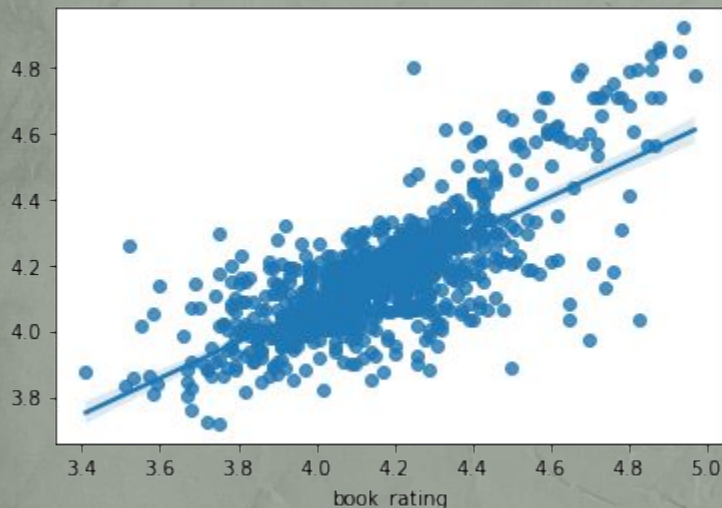


03

Feature Engineering

Initial Model

- Feature Engineering
- High multicollinearity in data
- Many insignificant features
- Author Distinct work is negatively correlated to book rating



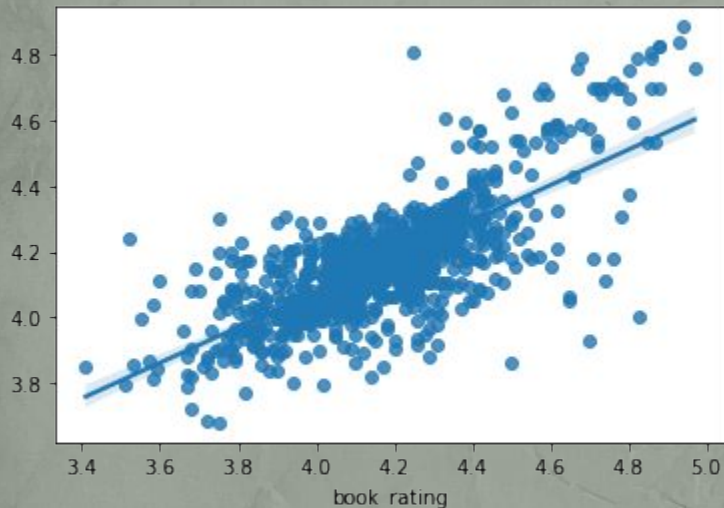
R^2 .4855

MAE = .1156

Final Model

Features:

- Square Author Average Rating
- Number of Rating
- Number of Reviews
- Interaction between Rating & Review



$$R^2 = .4914$$

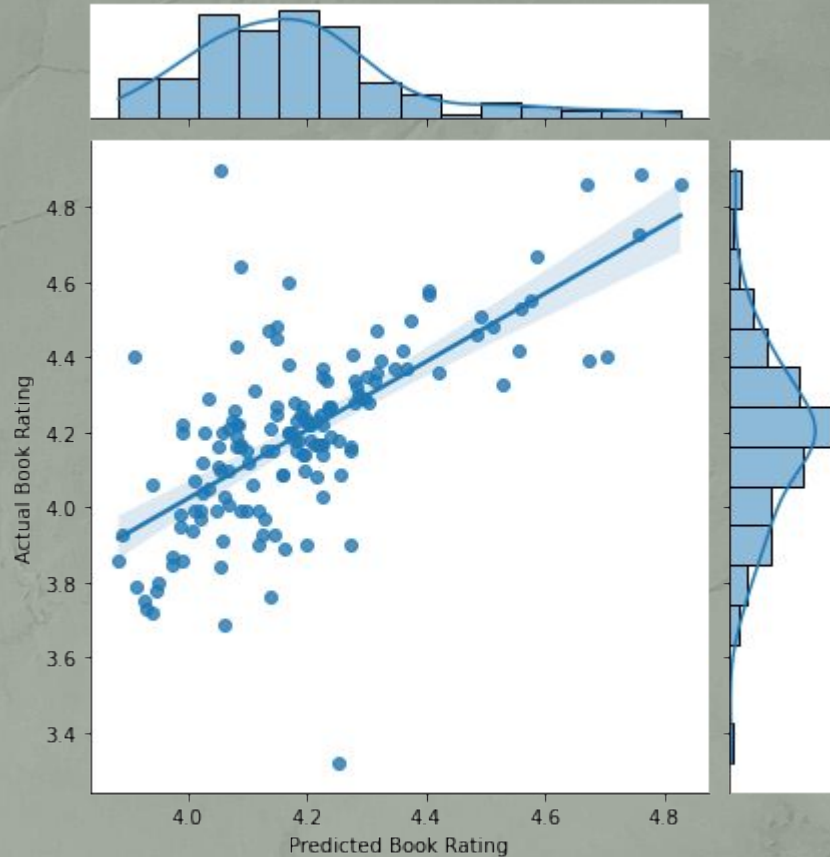
$$MAE = .1130$$

04

Final Model

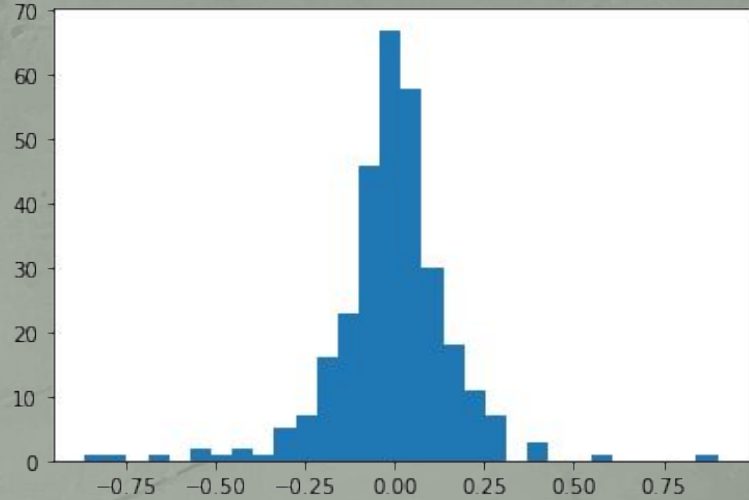
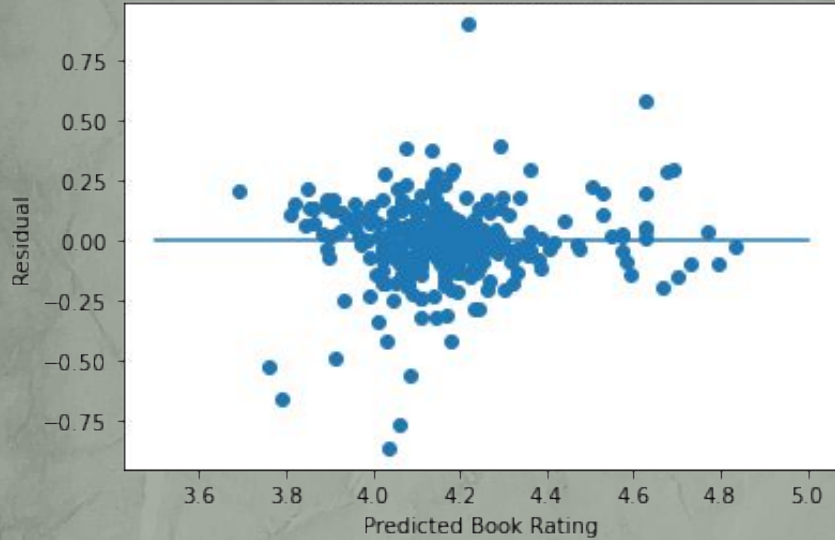


Model Performance



Residual Plot

Residuals vs. Predictions



Performance

Author Rating	Author No Rating	Author No Reviews	Book Rating	Predicted Rating
4.3	556616	37639	4.3	4.294747
4.14	1344	565	4.6	4.179611
4.04	20271	1416	3.99	4.098574
4.33	10310268	402766	4.35	4.235018
3.93	881114	37679	3.98	3.985006

Future Work

- Acquire Author Reviews from different websites
- Use NLP to measure sentiment in Reviews
- Analyse text available in Manuscript
- Incorporate Sales Data

Thank You.