

Book Recommendation System

Group 5

Tejas Sharma
Jainav Sanghvi
Harsh Shah
Chaitanya Manas Cheedella

Presentation Overview

1. Novelty of the problem
2. Dataset creation
3. New ideas and Philosophy behind algorithm selection
4. Results and analysis (Algorithm + Data)
5. Future Scope

Novelty of the Problem

In today's fast-paced world, many people lead busy lives and often struggle to find time for leisure activities. Reading books, in particular, may not be a priority for some individuals who may choose to spend their limited free time on other activities, such as watching TV.

However, for those who are interested in reading but have difficulty finding a book that suits their preferences, our recommendation system aims to provide a solution. By offering personalized book recommendations based on individual tastes and interests, our system can help users discover books that they may not have found on their own.

Ultimately, our goal is to make reading more accessible and enjoyable for everyone, even those with busy lifestyles.

Novelty of the Problem

It is suited to run for new users (who have not read a book) as well as users which have read a book and want similar books.

To have a larger social impact, we have modelled our system in such a way that it can be extended, in the future, to recommending research papers/scientific books to those who may want to find similar content.

Dataset Creation

There are available datasets for research papers which can be used to curate recommendations, which means we would need to be web scraping our required information. Hence, we simulated the same for books as well.

Since we have decided to model our system to also be extended (in the future) to recommending research papers, we decided to see what are the common features that we can exploit.

The common things between them are -

- Abstract/Description
- Title/Book name
- KeyWords/Genres
- Number of Citations/Ratings

Dataset Creation

- Web scraping using Selenium and BeautifulSoup
- 50,000 books dataset (Comprehensive Literary Greats dataset)
- Columns used : 'id', 'title', 'series', 'author', 'ratings', 'description', 'genres', 'num_pages', 'first_published', 'awards', 'num_ratings', 'ratings By Stars'

(This dataset was chosen as it had many columns useful to us for the purpose of recommendations)

Workflow of the Recommendation System

The input to the system is as follows -

- 1) New reader - Genre + Reading capacity
- 2) Old reader - Previously liked books

Recommendation -

New reader : Top rated books based on data given

Old reader : Recommend books which similar to each given book. Also recommend books which the user may like as per the books given as input.

New Ideas

- 1) Extendable recommendation pipeline
- 2) Dataset curated to suit recommendation algorithms.
- 3) New ways to input to the transformers - adding more context to our inputs.
- 4) Multilingual transformers chosen.

Philosophy behind algorithm selection

As the dataset that we have created is mainly text based, we require the use of NLP techniques to make recommendations.

We started out by thinking of 2 approaches -

1. Create embeddings for the books by concatenating description, author and genres.
2. Create embeddings for the books using only description.
3. Create refined embeddings by adding contextual sentences to the description (eg. “The author of the book is :”).

The models that we built were then tested for both of these approaches.

Philosophy behind Algorithm Selection

To start out, we ran these models on word2vec using Skip-Gram and CBOW architecture.

Word2Vec gives an embedding for each word and to create an embedding for the description we have to take an average of the each word embeddings in the description.

As we have already seen in class, word2vec doesn't take the context into account and is expected to perform bad. (We will see in the analysis slides)

Philosophy of Algorithm Selection

We then moved on to the transformer model for getting embeddings. We used the sentence transformer library to use the pre-trained models.

Here are the list of pre-trained models available-

https://www.sbert.net/docs/pretrained_models.html

We decided to use paraphrase-multilingual-mpnet-base-v2. This model is suitable for large datasets containing different languages and is therefore suitable to our dataset.

The data is converted into a np.array and fed into the pre-trained model which encodes the input to give the sentence embeddings.

Analysis and Real-Time Inference

Skip Gram Word 2 Vec - The Martian

| title | description | | author | genres |
|-------------------------------------|---|---|------------------------------------|---|
| Reboot | five years ago wren connolly was shot three ti... | | Amy Tintera (Goodreads Author) | ['Young Adult', 'Dystopia', 'Science Fiction',... |
| The Watchers | the watchers is the first book in the watchers... | Lynnie Purcell (Goodreads Author), Holly Purce... | | ['Young Adult', 'Fantasy', 'Paranormal', 'Ange... |
| Midnighter and Apollo | midnighter and apollo are back with a vengeanc... | Steve Orlando (Writer), Fernando Blanco (Artist) | | ['Comics', 'Graphic Novels', 'LGBT', 'Dc Comic... |
| Times Square Red, Times Square Blue | if one street in america can claim to be the m... | | Samuel R. Delany | ['Nonfiction', 'Queer', 'History', 'New York',... |
| Strength of the Pack | a marine werewolf and his commander bring lege... | | Kendall McKenna (Goodreads Author) | ['M M Romance', 'Paranormal', 'Military Fictio... |
| The Devil's Son | chief of police brady simone is determined to ... | | Jennifer Loren (Goodreads Author) | ['Paranormal', 'Romance', 'Dark', 'Paranormal ... |

Bag of Words Word 2 Vec - The Martian

| title | description | | author | genres |
|-------------------------------------|---|---|------------------------------------|---|
| Reboot | five years ago wren connolly was shot three ti... | | Amy Tintera (Goodreads Author) | ['Young Adult', 'Dystopia', 'Science Fiction',... |
| The Watchers | the watchers is the first book in the watchers... | Lynnie Purcell (Goodreads Author), Holly Purce... | | ['Young Adult', 'Fantasy', 'Paranormal', 'Ange... |
| Midnighter and Apollo | midnighter and apollo are back with a vengeanc... | Steve Orlando (Writer), Fernando Blanco (Artist) | | ['Comics', 'Graphic Novels', 'LGBT', 'Dc Comic... |
| Times Square Red, Times Square Blue | if one street in america can claim to be the m... | | Samuel R. Delany | ['Nonfiction', 'Queer', 'History', 'New York',... |
| Strength of the Pack | a marine werewolf and his commander bring lege... | | Kendall McKenna (Goodreads Author) | ['M M Romance', 'Paranormal', 'Military Fictio... |

Transformer (Concatenated) - The Martian

| title | description | author | genres |
|---|---|--|---|
| The Prime of Miss Jean Brodie / The Girls of S... | the brevity of muriel sparks novels is equaled... | Muriel Spark | ['Fiction', 'Classics', 'Literature', 'Novels'... |
| The Monkey Wars | the controversy over the use of primates in re... | Deborah Blum | ['Science', 'Nonfiction', 'History', 'Biology'... |
| The Malloreon, Vol. 1: Guardians of the West /... | discover the magic of the malloreon—david eddi... | David Eddings | ['Fantasy', 'Fiction', 'High Fantasy', 'Epic F... |
| The Martian Chronicles | the strange and wonderful tale of man's experi... | Ray Bradbury | ['Science Fiction', 'Fiction', 'Classics', 'Sh... |
| I'll Never Let You Go | kevin is debonair well educated loyal and will... | Leonard Anderson Jr. | |
| Олень на заводской территории | олень на заводской территории это один из расс... | Kurt Vonnegut Jr., Курт Воннегут, Н. Пальцев (...) | |

Transformer (Description Only) - The Martian

| title | description | author | genres |
|--------------------------------|---|--------------------------------------|---|
| The Martian Chronicles | the strange and wonderful tale of man's experi... | Ray Bradbury | ['Science Fiction', 'Fiction', 'Classics', 'Sh... |
| Red Planet | jim marlow and his strangelooking martian frie... | Robert A. Heinlein | ['Science Fiction', 'Fiction', 'Young Adult', ... |
| Union | there are no holes in space none to hide in no... | Jolea M. Harrison (Goodreads Author) | |
| Planet of Adventure | stranded on the distant planet tschai young ad... | Jack Vance | ['Science Fiction', 'Fiction', 'Fantasy', 'Adv... |
| Deepsix | in the year 2204 tragedy and terror forced a s... | Jack McDevitt | ['Science Fiction', 'Fiction', 'Space Opera', ... |
| The Creatures That Time Forgot | from planet stories mad impossible world sunbl... | Ray Bradbury | ['Science Fiction', 'Short Stories', 'Science ... |

Transformer (Improvised) - The Martian

| | | | |
|---|---|---|---|
| The Creatures That Time Forgot | from planet stories mad impossible world sunbl... | Ray Bradbury | ['Science Fiction', 'Short Stories', 'Science ... |
| The Martian Chronicles | the strange and wonderful tale of man's experi... | Ray Bradbury | ['Science Fiction', 'Fiction', 'Classics', 'Sh... |
| Red Planet | jim marlow and his strangelooking martian frie... | Robert A. Heinlein | ['Science Fiction', 'Fiction', 'Young Adult', ... |
| Лунна дъга | роман по черному следу это первая часть фанта... | Sergey Pavlov, Сергей Павлов, Анастасия Цонева... | □ |
| Summary of The Martian: by Andy Weir Include... | summary of the martian by andy weir includes ... | NOT A BOOK | □ |
| Mars Trilogy | all three volumes of the worldwide bestselling... | Kim Stanley Robinson | ['Science Fiction', 'Fiction', 'Science Fictio... |
| 2001: A Space Odyssey | on the moon an enigma is uncovered\n\nso great... | Arthur C. Clarke | ['Science Fiction', 'Fiction', 'Classics', 'Sc... |
| Voyage | the space mission of a lifetime an epic saga o... | Stephen Baxter | ['Science Fiction', 'Fiction', 'Alternate Hist... |
| Green Mars | in the nebula award winning red mars kim stanl... | Kim Stanley Robinson | ['Science Fiction', 'Fiction', 'Hugo Awards', ... |

Why is our Improvised model better? Description + Refined representation of Author and Genre!

The Creatures that Time Forgot

From planet stories, mad impossible world sunblasted by day coldwacked by night and life condensed by radiation into eight days. Sim eyed the ship **if he only dared reach it and escape** - but it was more than half an hour distant, perhaps the limit of life itself. From the author of fahrenheit 45,1 **the martian chronicles** and the illustrated man originally published in the fall 1946 issue of planet stories it was later reprinted under the title frost and fire.

The Martian

Six days ago astronaut mark watney became one of the first people to walk on mars now he's sure he'll be the first person to die thereafter a dust storm nearly kills him and forces his crew to evacuate while thinking him dead mark finds himself stranded and completely alone with no way to even signal earth that he's alive—and even if he could get word out his supplies would be gone long before a rescue could arrive chances are though he won't have time to starve to death the damaged machinery unforgiving environment or plainold “human error” are much more likely to kill him first but mark isn't ready to give up yet drawing on his ingenuity his engineering skills — and a relentless dogged refusal to quit — he steadfastly confronts one seemingly insurmountable obstacle after the next will his resourcefulness be enough to overcome the impossible odds against him'

Why is our Improvised model better? Description + Refined representation of Author and Genre!

```
[28]: input_dataframe.iloc[36580]["genres"] #Creatures that Time Forgot
```

```
[28... "['Science Fiction', 'Short Stories', 'Science Fiction Fantasy', 'Classics', 'Fiction']"
```

```
[33]: input_dataframe.iloc[391]["genres"] #The Martian
```

```
[33... "['Science Fiction', 'Fiction', 'Audiobook', 'Adventure', 'Space', 'Adult', 'Science Fiction Fantasy', 'Thriller', 'Humor', 'Survival']"
```

```
[31]: input_dataframe.iloc[385]["genres"] #The Martian Chronicles
```

```
[31... "['Science Fiction', 'Fiction', 'Classics', 'Short Stories', 'Fantasy', 'Science Fiction Fantasy', 'American', 'Audiobook', 'Literature', 'Space']"
```

Bag of Words Word 2 Vec - Harry Potter

| title | description | author | genres |
|----------------------------------|---|-----------------------------------|---|
| The Adventures of Jewel Cardwell | what would you do if you knew you'd inherited ... | Fumi Hancock (Goodreads Author) | ['Young Adult'] |
| Thunderlight | vibrant scale swathed wings... abilities that ca... | Adrienne Woods (Goodreads Author) | ['Dragons', 'Fantasy', 'Young Adult', 'Romance...'] |
| The Immortal Lover | down through the ages of human history his nam... | A. White (Goodreads Author) | [] |
| Consequences | every action has consequenceswaking in an unfa... | Aleatha Romig (Goodreads Author) | ['Dark', 'Romance', 'Abuse', 'Contemporary', '...'] |
| Emily Post's Etiquette | features twenty new chapters that cover such a... | Peggy Post, Emily Post | ['Nonfiction', 'Reference', 'Self Help', 'Clas...'] |
| The Liberation of Gabriel King | gabriel king was a born chicken hes afraid of ... | K.L. Going (Goodreads Author) | ['Historical Fiction', 'Young Adult', 'Fiction...'] |

Transformer (Concatenation)- Harry Potter

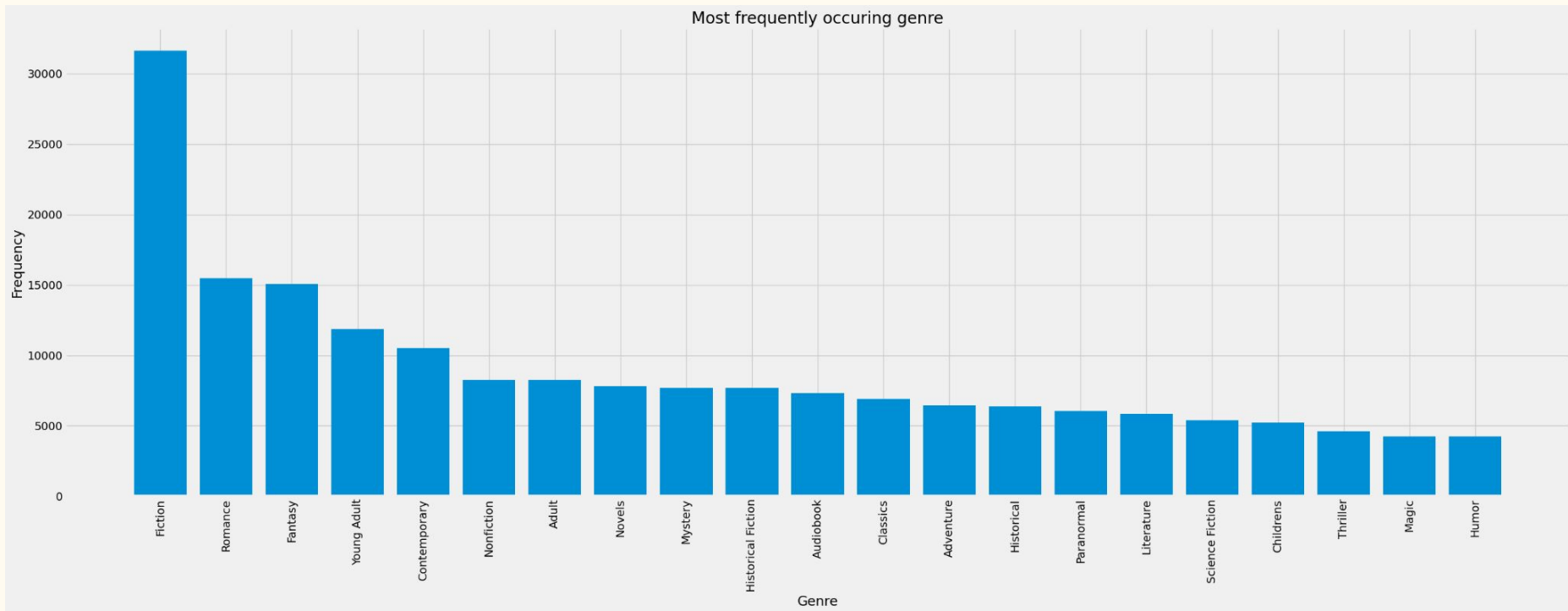
| | | | |
|---|---|---|---|
| Harry Potter and the Chamber of Secrets | ever since harry potter had come home for the ... | J.K. Rowling, Mary GrandPré (Illustrator) | ['Fantasy', 'Young Adult', 'Fiction', 'Magic',...] |
| Harry Potter and the Prisoner of Azkaban | harry potters third year at hogwarts is full o... | J.K. Rowling, Mary GrandPré (Illustrator) | ['Fantasy', 'Young Adult', 'Fiction', 'Magic',...] |
| Harry Potter Schoolbooks Box Set: Two Classic ... | there is an alternate cover edition for this i... | J.K. Rowling | ['Fantasy', 'Fiction', 'Young Adult', 'Childre...'] |
| The Harry Potter Collection 1-4 | the exciting tales of harry potter the young w... | J.K. Rowling, Mary GrandPré (Illustrator) | ['Fantasy', 'Young Adult', 'Fiction', 'Magic',...] |
| James Potter and the Hall of Elders' Crossing | what's it like to be the son of the most famou... | G. Norman Lippert (Goodreads Author) | ['Fantasy', 'Young Adult', 'Fiction', 'Magic',...] |
| James Potter and the Curse of the Gatekeeper | a summer of change brings james sirius potter ... | G. Norman Lippert (Goodreads Author) | ['Young Adult', 'Fiction', 'Magic', 'Fantasy',...] |

Transformer (Improvised) - “You may also like”

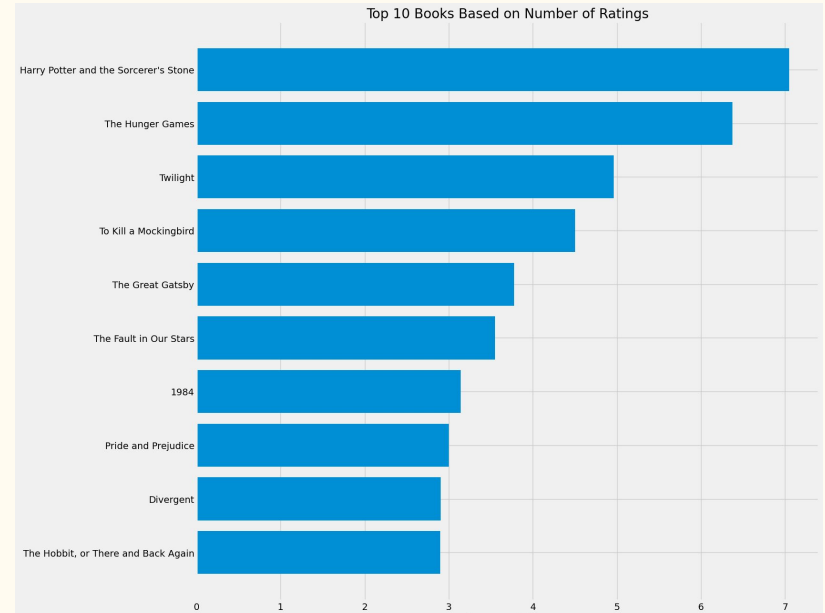
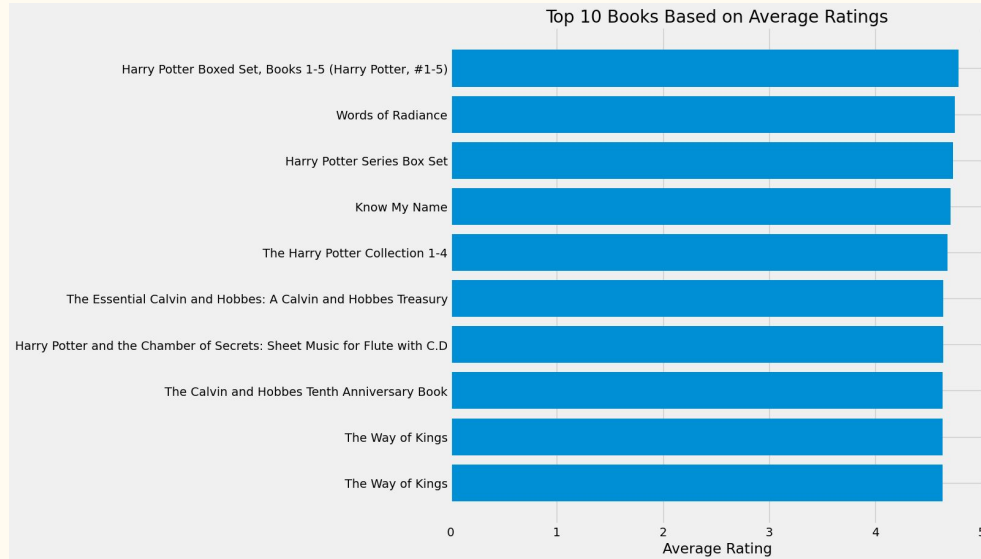
2 Methods - Average of Embeddings and Embedding of Combined Descriptions

| | title | description | author | genres | li |
|--|---|---|---|---|----|
| | The Martian Chronicles | the strange and wonderful tale of man's experi... | Ray Bradbury | ['Science Fiction', 'Fiction', 'Classics', 'Sh... | |
| | Red Planet | jim marlow and his strangelooking martian frie... | Robert A. Heinlein | ['Science Fiction', 'Fiction', 'Young Adult', ... | |
| | Union | there are no holes in space none to hide in no... | Jolea M. Harrison (Goodreads Author) | | [] |
| | Planet of Adventure | stranded on the distant planet tschai young ad... | Jack Vance | ['Science Fiction', 'Fiction', 'Fantasy', 'Adv... | |
| | Deepsix | in the year 2204 tragedy and terror forced a s... | Jack McDevitt | ['Science Fiction', 'Fiction', 'Space Opera', ... | |
| | The Creatures That Time Forgot | from planet stories mad impossible world sunbl... | Ray Bradbury | ['Science Fiction', 'Short Stories', 'Science ... | |
| | The Creatures That Time Forgot | from planet stories mad impossible world sunbl... | Ray Bradbury | ['Science Fiction', 'Short Stories', 'Science ... | |
| | Starship Titanic | sur la planète blérontin on sapprête à inaugur... | Terry Jones, Marie-Catherine Caillava (Transla... | ['Science Fiction', 'Fiction', 'Humor', 'Fanta... | |
| | Lost Moon: The Perilous Voyage of Apollo 13 | april 1970 the glory days of the apollo space ... | Jim Lovell, Jeffrey Kluger, Fred Sanders (Narr... | ['Nonfiction', 'History', 'Space', 'Science', ... | |
| | Doctor Who: Apollo 23 | for a few moments this afternoon it rained on ... | Justin Richards | ['Doctor Who', 'Science Fiction', 'Fiction', '... | |

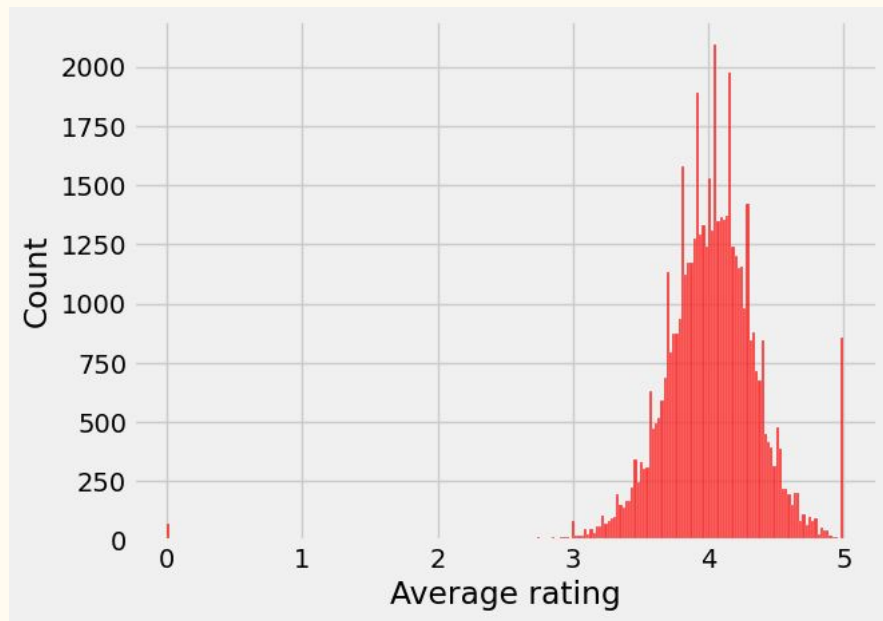
Dataset Analysis (EDA)



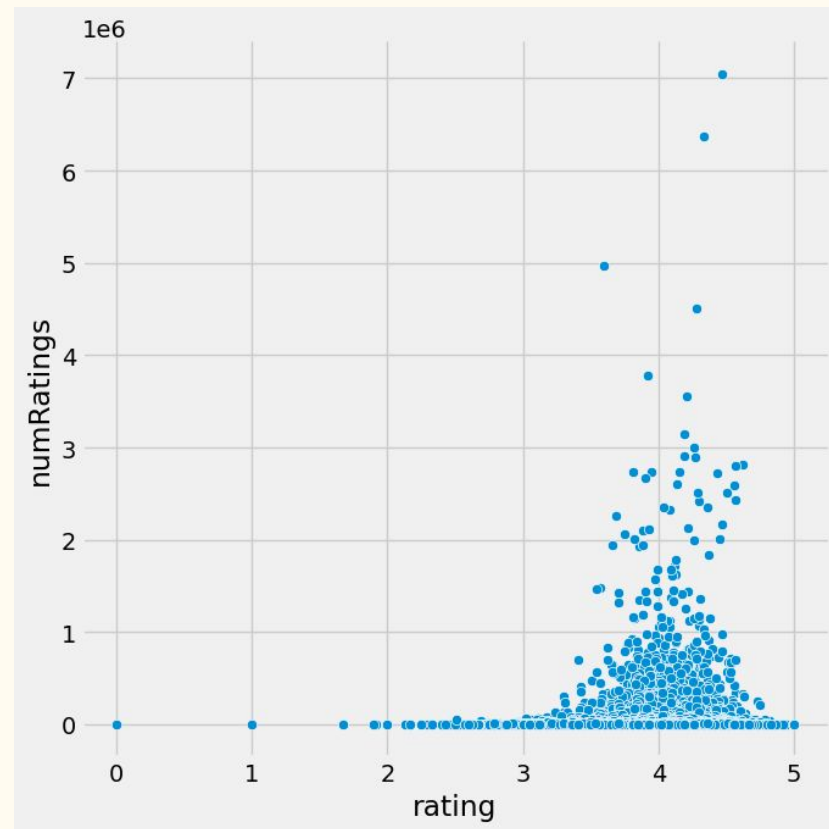
Top 10 books



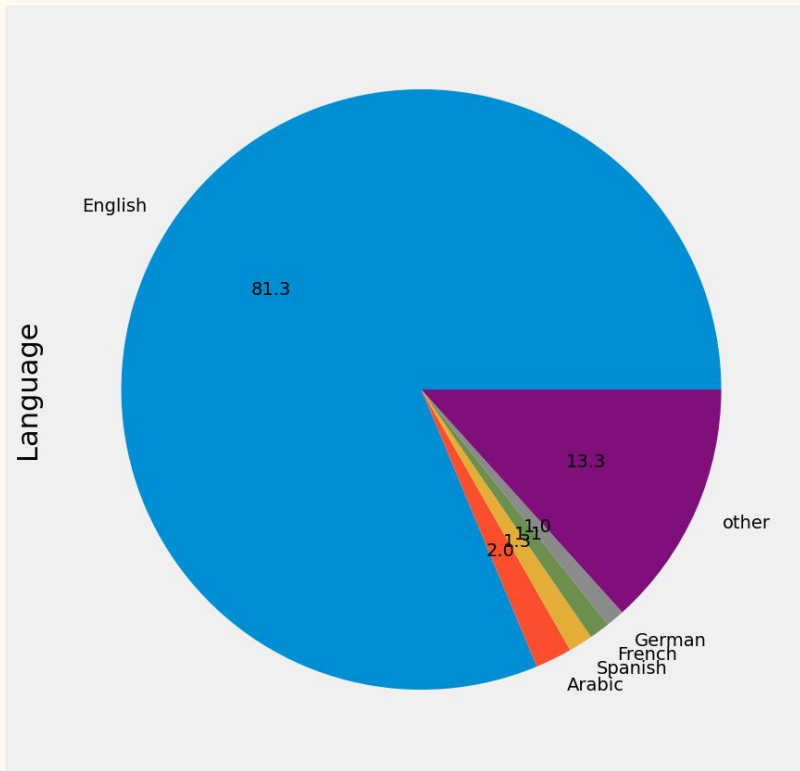
The average rating is calculated only for books which have more than 10k ratings.



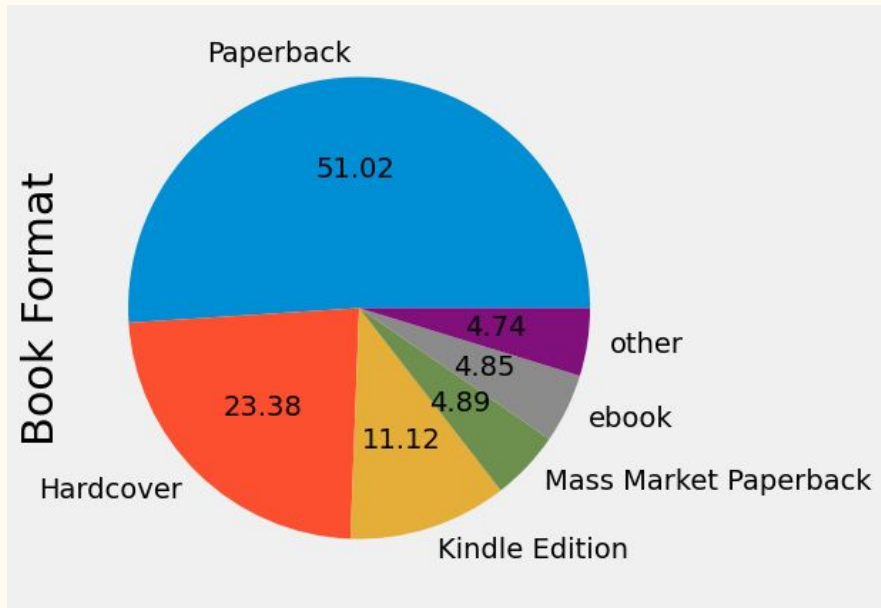
Counts per rating value



Rating v/s number of
rating for each data point

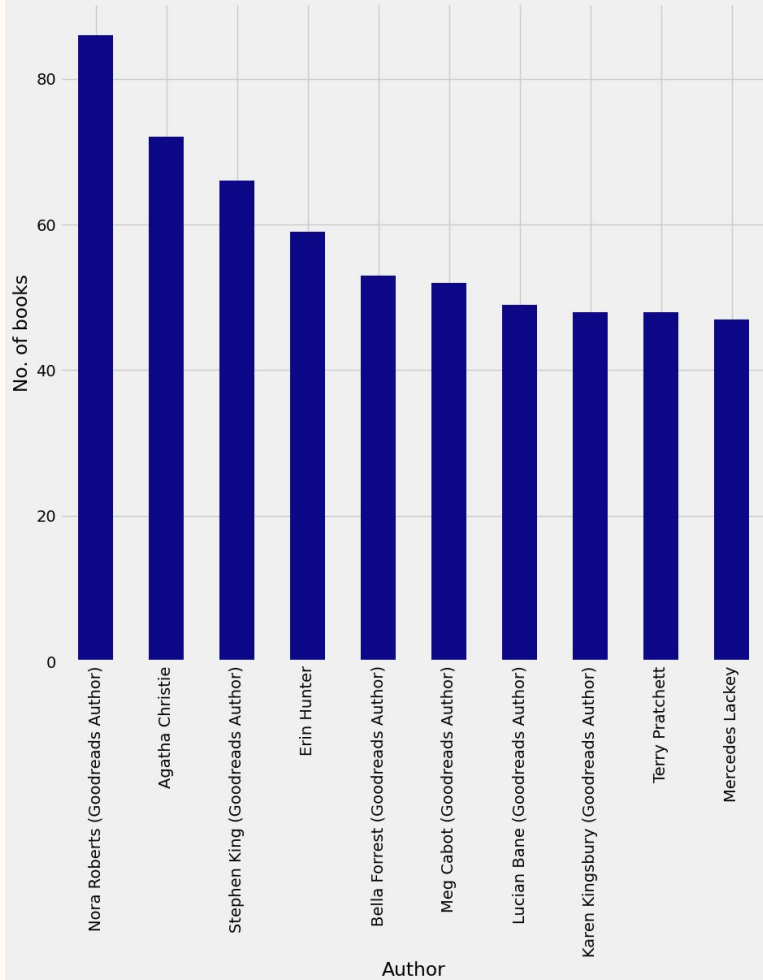


Book language distribution

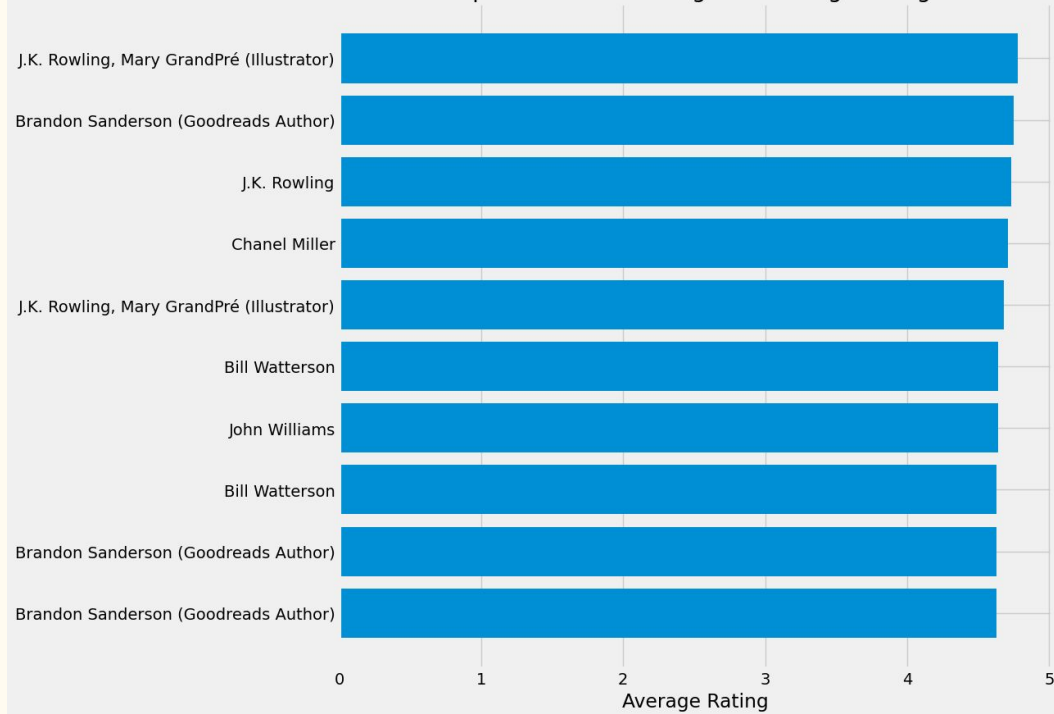


Book format distribution

Authors with most books



Top 10 Author with highest average ratings



Research Paper Recommendation System

- Dataset columns - 'Title', 'Abstract', 'Conclusion of the paper', 'Authors', 'Paper Citations', 'Keywords' (Web Scraped from IEEE Website)
- Model : Transformer (Improvised version)

Recommendation for : A 10-bit 50-MS/s SAR ADC With a Monotonic Capacitor Switching Procedure

| Title | Abstract | Conclusion of the paper | Authors | Paper Citations | Keywords |
|---|---|---|---|-----------------|---|
| Leakage current mechanisms and leakage reducti... | High leakage current in deep-submicrometer reg... | With the continuous scaling of CMOS devices, l... | K. Roy; S. Mukhopadhyay; H. Mahmoodi-Meimand | 1606 | , CMOS integrated circuits , leakage currents,... |
| The flipped voltage follower: a useful cell fo... | In this paper, a basic cell for low-power and/... | In this paper, a cell coined as FVF has been r... | R.G. Carvajal; J. Ramirez-Angulo; A.J. Lopez-M... | 470 | Voltage , Circuit synthesis , Consumer electro... |
| A Systematic Review of Compressive Sensing: Co... | Compressive Sensing (CS) is a new sensing moda... | Introduction of CS has revolutionized many are... | Meenu Rani; S. B. Dhok; R. B. Deshmukh | 270 | Sensors , Transforms , Mathematical model , Sp... |
| A low-power low-noise CMOS amplifier for neuro... | There is a need among scientists and clinician... | An 80-μ W fully integrated CMOS biosignal ampl... | R.R. Harrison; C. Charles | 1283 | Low-noise amplifiers , Low-frequency noise , P... |
| SegNet: A Deep Convolutional Encoder-Decoder A... | We present a novel and practical deep fully co... | We presented SegNet, a deep convolutional netw... | Vijay Badrinarayanan; Alex Kendall; Roberto Ci... | 8360 | Decoding , Neural networks , Training , Comput... |

Future Scope

- 1) Further change how we input information to the transformer - Ordering of the different aspects, identifying themes, etc.
- 2) Use more metrics and data - Append more columns to our description (eg. num pages, awards, etc).
- 3) Fine Tune the model - Take the pretrained model and train it exclusively on the dataset to make it more relevant to book/paper recommendations.
- 4) Incorporate user feedback - Take into account the tastes of multiple users to get Hits and Misses. This will improve our model judgement metrics. We can also implement a collaborative filtering aspect.
- 5) Different models - To make our research paper recommendation system, we can use a better transformer like SciBERT.