

The image features a stack of approximately ten books on the left side, with varying thicknesses and some visible red and green spines. To the right, an open book lies flat on a dark wooden surface, showing two pages of text. The background is a solid, vibrant blue.

# New York Times Bestseller Analysis

Sameer Singh Rawat, LMU

# Overview

- Dataset - The data will comprise of the book titles and their respective summary. This will be generated by NYT api.
- Sentiment analysis will be done on the summary of each book.
- Topic modelling will be done on the summary but since summaries only describe in short detail, the descriptions will be fetched by google book api and topic modelling will be done on both.
- Recommendation - the columns such as authors, genres will be fetched through api and content based recommendation will be done based on the book user has already read.

Github link - [https://github.com/SameerR007/CSS\\_project](https://github.com/SameerR007/CSS_project)

# Dataset

1. Dataset will comprise of the books that has featured as NYT bestseller at the start of each month of 2023.
2. Dataset will contain the information about the books such as their author, genres, short description (referred as summary), detailed description (referred as description).
3. The difference between summary and description of random three books can be seen in the figure on the next slide.
4. Since summary seems to focus on solely the story of the book it is wise to perform sentiment analysis on them while description is more elaborative in size it seems better to do topic modelling on description.

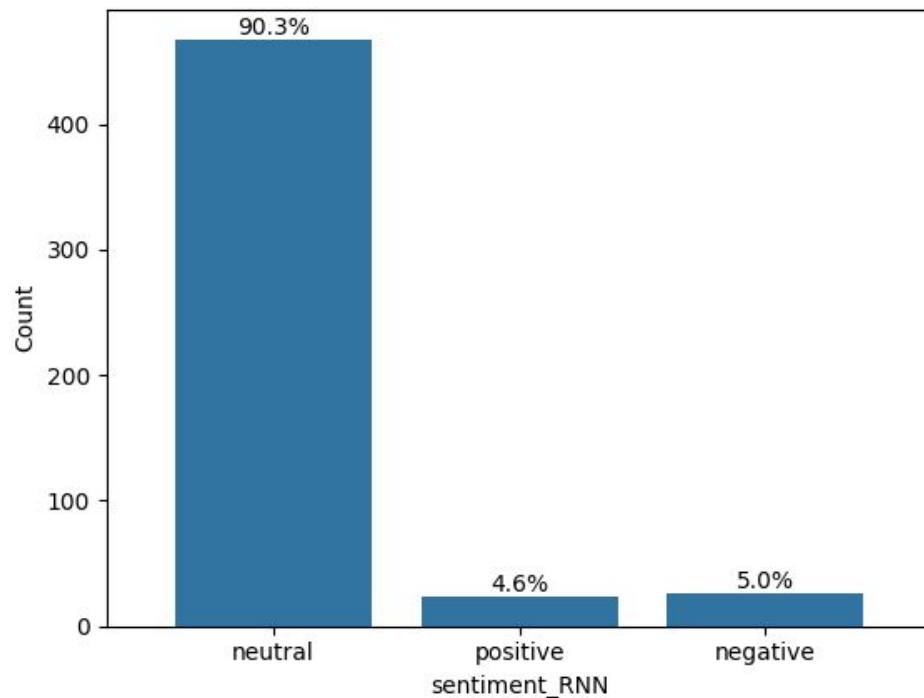
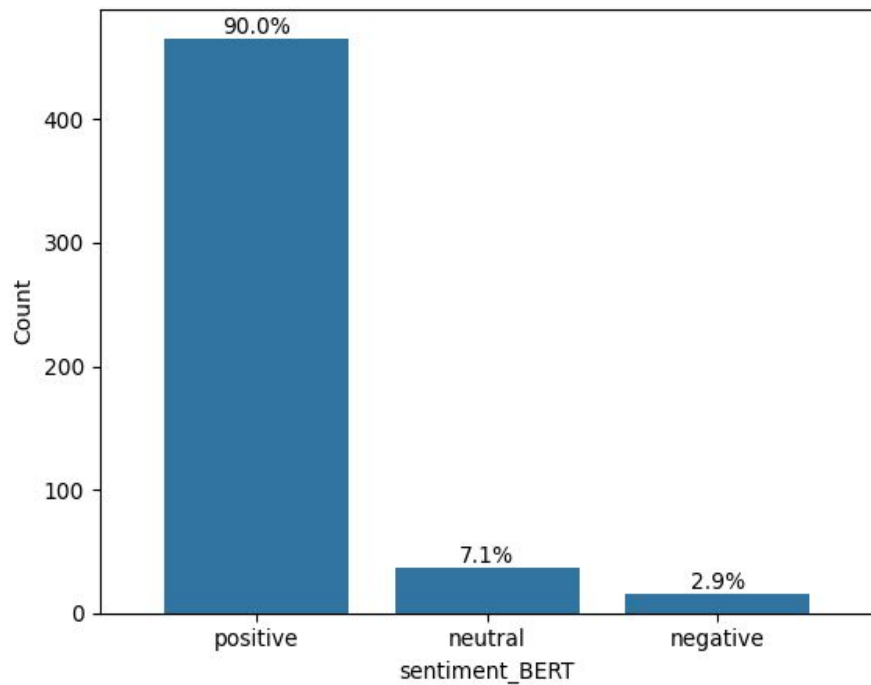
# Summary VS Description

index	Summary	Description
0	In the sequel to "It Ends With Us," Lily deals with her jealous ex-husband as she reconnects with her first boyfriend.	PREVIOUS BOOK IN SERIES: IT ENDS WITH US, ISBN 9781501110368. Before 'It Ends with Us', it started with Atlas. Colleen Hoover tells fan favourite Atlas side of the story and shares what comes next in this long-anticipated sequel to the glorious and touching (USA TODAY) 'It Ends With Us'.
1	A battered wife raised in a violent home attempts to halt the cycle of abuse.	In this "brave and heartbreaking novel that digs its claws into you and doesn't let go, long after you've finished it" (Anna Todd, New York Times bestselling author) from the #1 New York Times bestselling author of All Your Perfects, a workaholic with a too-good-to-be-true romance can't stop thinking about her first love. Lily hasn't always had it easy, but that's never stopped her from working hard for the life she wants. She's come a long way from the small town where she grew up—she graduated from college, moved to Boston, and started her own business. And when she feels a spark with a gorgeous neurosurgeon named Ryle Kincaid, everything in Lily's life seems too good to be true. Ryle is assertive, stubborn, maybe even a little arrogant. He's also sensitive, brilliant, and has a total soft spot for Lily. And the way he looks in scrubs certainly doesn't hurt. Lily can't get him out of her head. But Ryle's complete aversion to relationships is disturbing. Even as Lily finds herself becoming the exception to his "no dating" rule, she can't help but wonder what made him that way in the first place. As questions about her new relationship overwhelm her, so do thoughts of Atlas Corrigan—her first love and a link to the past she left behind. He was her kindred spirit, her protector. When Atlas suddenly reappears, everything Lily has built with Ryle is threatened. An honest, evocative, and tender novel, It Ends with Us is "a glorious and touching read, a forever keeper. The kind of book that gets handed down" (USA TODAY).
2	A scientist and single mother living in California in the 1960s becomes a star on a TV cooking show.	From advice columnist Meredith Goldstein, a dazzling, romantic, and emotionally resonant YA debut about a teen science whiz in Cambridge, Massachusetts, who tries to crack the chemical equation for lasting love and instead wreaks havoc on herself and the boys in her life. For seventeen-year old Maya, the equation for happiness is simple: a dream internship at MIT + two new science nerd friends + a perfect boyfriend = one amazing summer. Then Whit dumps her out of the blue. Maya is miserable until she discovers that her scientist mother, before she died, was conducting research on manipulating pheromones to enhance human attraction. If Maya can finish her mother's work, maybe she can get Whit back. But when her experiment creates chaos in her love life, she realizes that maybe love and loss can't be understood using the scientific method. Can she learn to trust the unmeasurables of love and attraction instead?

# Sentiment Analysis

- Sentiment analysis will be done by state of art model BERT and also through a trained from scratch RNN model.
- RNN model is trained on the X(Twitter) dataset taken from kaggle. The dataset is divided into 80:20 train and test.
- This model gives a test set accuracy of about 80%.
- The predictions of sentiments of summary of total dataset with BERT model and RNN can be seen in the form of count plot in the next slide.
- RNN seems to struggle with classifying the sentiment of summaries in black and white, mostly outputting grey (neutral). While BERT mostly classifies the books to be of positive sentiment. (Figure on the next slide)

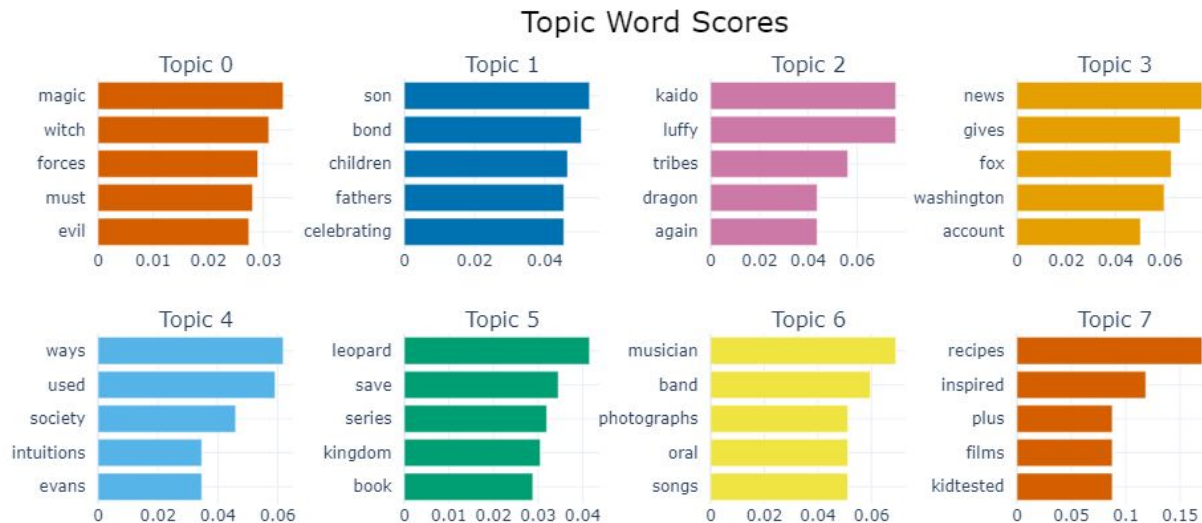
# BERT VS RNN



# Topic Modelling

- It will determine the topics around which people like to read stories.
- Again this will be done by state of art model BERTopic and a model implemented from scratch making use of Latent Dirichlet Allocation(LDA).
- Topics derived from BERTopic will be displayed on the recommended book as keywords.
- Topic modelling will also be done by LDA from scratch and the concept of how LDA is trained will be discussed (yet to be done)
- Discussion on both the approaches will be done.

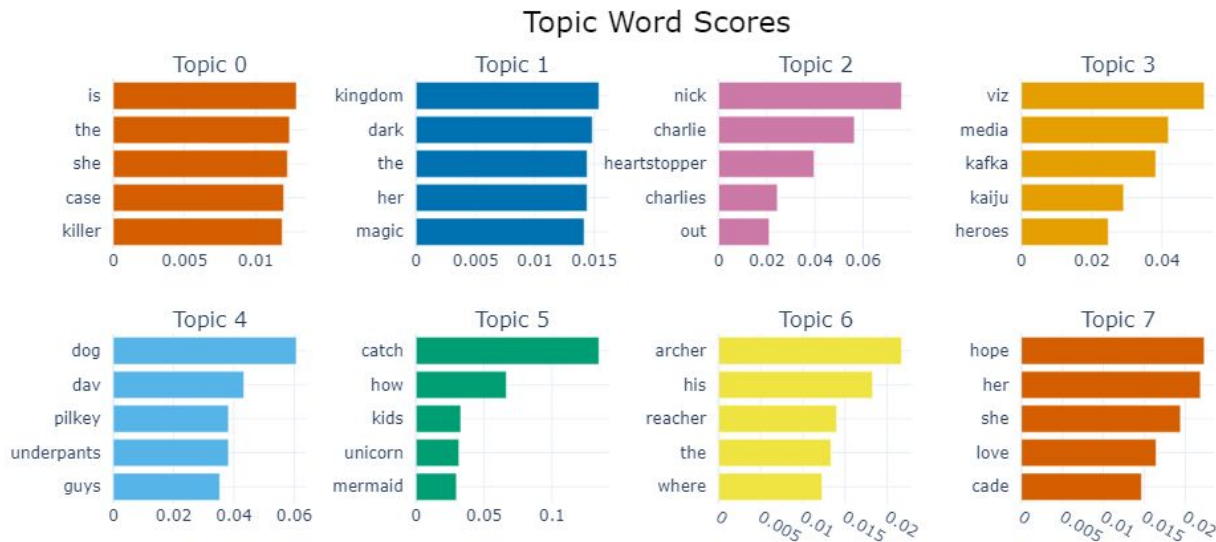
# BERTopic on summaries (short description)



Probable themes - fantasy(Topic 0), family(Topic 1), politics(Topic 3), art(Topic 6)



# BERTopic on descriptions



Probable themes - crime(Topic 0), fantasy(Topic 1),kafka(Topic 3),  
fantasy+children (Topic 6), positive sentiments such as love and hope(Topic 7)

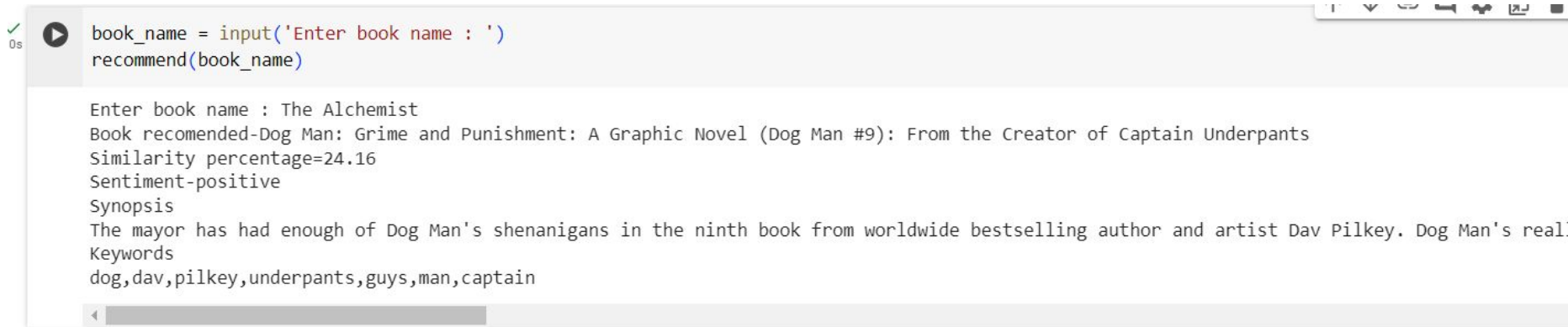
# Recommender system

In the final stage, we make a content based recommendation.

- Steps
  1. Make feature vectors combining the columns of genres, author, description for each book and treat it as a content.
  2. When a new book is entered the feature vector is formed for this new book by using google books api
  3. Finally the book that is most closest match to the entered book is recommended.

NOTE - CountVectorizer is used for encoding the feature vector. And cosine similarity for finding the closest match.

# Output demo



The screenshot shows a Jupyter Notebook interface. At the top left, there is a green checkmark and the text '0s'. To the right of this is a play button icon. The code cell contains the following Python code:

```
book_name = input('Enter book name : ')  
recommend(book_name)
```

The output of the code is displayed below the code cell:

```
Enter book name : The Alchemist  
Book recommended-Dog Man: Grime and Punishment: A Graphic Novel (Dog Man #9): From the Creator of Captain Underpants  
Similarity percentage=24.16  
Sentiment-positive  
Synopsis  
The mayor has had enough of Dog Man's shenanigans in the ninth book from worldwide bestselling author and artist Dav Pilkey. Dog Man's real  
Keywords  
dog,dav,pilkey,underpants,guys,man,captain
```

- Similarity percent as calculated by cosine similarity.
- Sentiment - as predicted by BERT
- Keywords - as suggested by BERTopic

Addition - the final model will be deployed in cloud using hugging face space and streamlit. Image of the recommended book will also be displayed.

Thank you!