Mid-Term Presentation

Sameer Singh Rawat
sameer.rawat@campus.lmu.de

December 22, 2023

Newyork Times Bestseller Analysis

Outline

Data Collection

1. Collecting the data of newyork times bestseller books through NYT api.

2. The data comprises of book titles, book authors, genres, book summary.

Outline

Sentiment Analysis

1. To analyze the sentiment of each of the NYT bestseller's summary, and see the stories of which sentiment are preferred by majority of audience in general.

2. BERT or hugging face models can be used for the purpose of sentiment analysis

Topic Modelling

1. To analyze around which topics most of the NYT bestseller stories revolve. It can give us an idea about the type of books people are currently liking to read.

2. BERTopic or Latent Dirichlet Allocation(LDA) can be used for this purpose.

Outline

Recommender System

1. In the end we would make an interactive interface for the user to input the book they have already read and give the most similar book from the books that has featured in Newyork Times Bestsellers.

2. We will make use of count vectorizer and cosine similarity to get the most similar NYT bestseller book to the inputed book. And also display its synopsis and the sentiment of the synopsis.

3. Finally we will deploy it in cloud for public use.

Thank you!