NEWS RECOMMENDATION SYSTEM

Presented By

Naitik Zaveri

Padmesh Naik

Vignesh Sundaram

Parth Dhruv

INDEX

INTRODUCTION

- In the digital age, people face an overwhelming amount of news and information, making it challenging to find articles that are of interest to them.
- This project proposes a news recommendation system that utilizes Cosine similarity and Content-based filtering to provide users with personalized news articles based on their interests.
- The system utilizes advanced techniques such as Word2Vec model to generate word embeddings and calculate article similarity.

Dataset

- The dataset includes 210,294 news articles in JSON format from diverse categories such as sports, politics, business, and entertainment.

- To offer tailored news articles based on user interests, the system utilizes properties such as category, headline, author, date, link, and brief description of the articles in the dataset.

Tools

Scikit-learn

Flask

Google Colab

Jupyter Notebook

APPROACH

TOKENIZATION

SENTENCE TRANSFORMER

TOKENIZATION

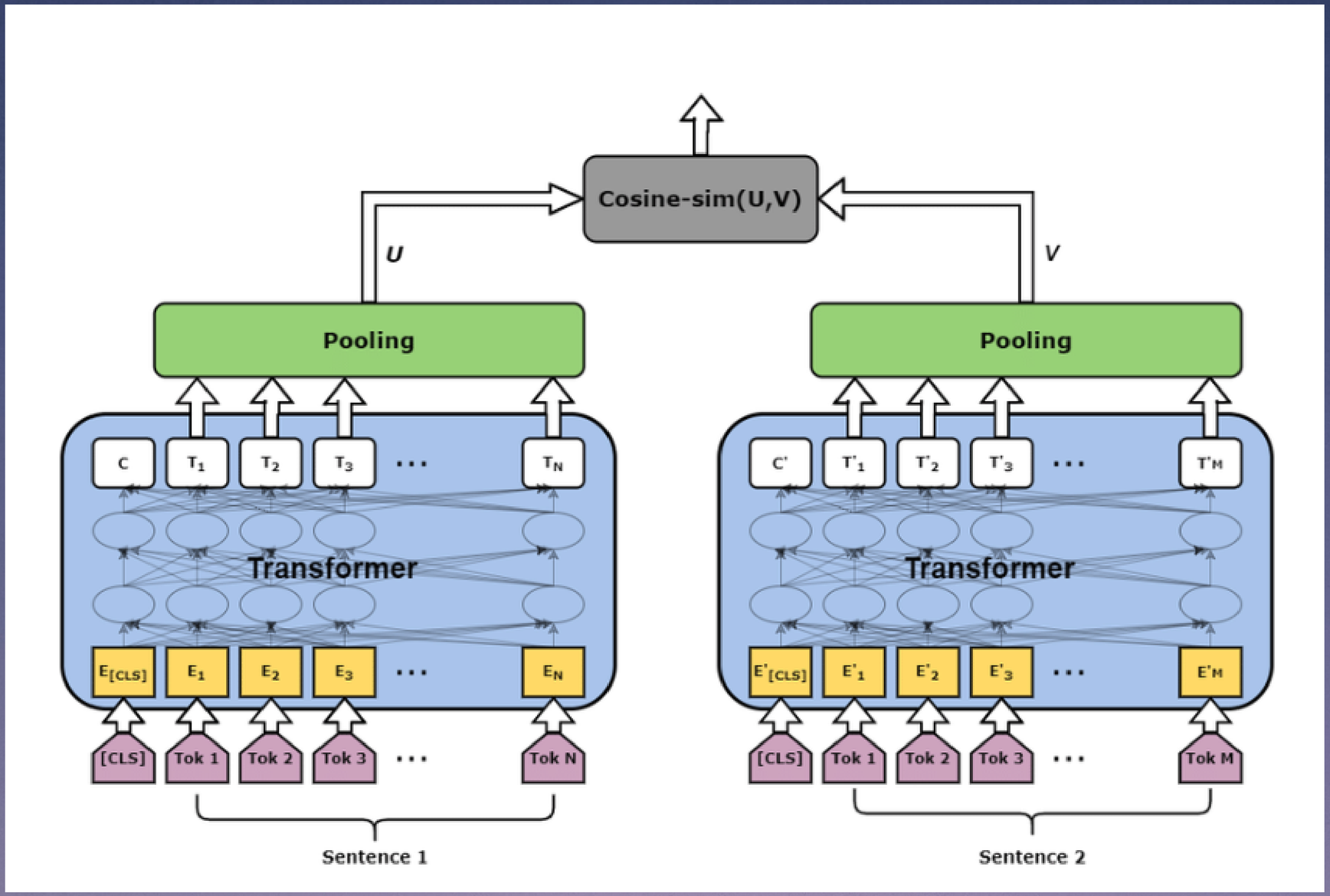
- Tokenization is an important NLP step as this is when the model understands the input that will be given to the model.
- Our model uses the Word Piece tokenizer in order to split the text into individual tokens.

SENTENCE TRANSFORMER

- The tokenized data for both the word that we want to search for and each and every data available for us in the dataset is given as an input to the Sentence Transformer model. The tokenized data is appended with CLS to begin with and a TOK to end the text with so that the Sentence Transformer model knows the start and the end.

- The Transformer model used here is BERT(Bidirectional Encoder Representations from Transformers). This encodes the data that we have as an input which is then given as an input to the Pooling layer.

- Since we are using two pooling layers, the output will be finding a cosine similarity between both the ground truth values and the text that we want.


Cosine-sim(U,V)
U
V
Pooling
Pooling
C
T1
T2
T3
TN
C'
T'1
T'2
T'3
T'M
Transformer
Transformer
E[CLS]
E1
E2
E3
EN
E'[CLS]
E'1
E'2
E'3
E'M
[CLS]
Tok 1
Tok 2
Tok 3
Tok N
[CLS]
Tok 1
Tok 2
Tok 3
Tok M
Sentence 1
Sentence 2

USER FEEDBACK

- Generate new recommendations, filter out irrelevant ones, and return improved results.
- Outcome: Refined, more relevant recommendations tailored to user needs.

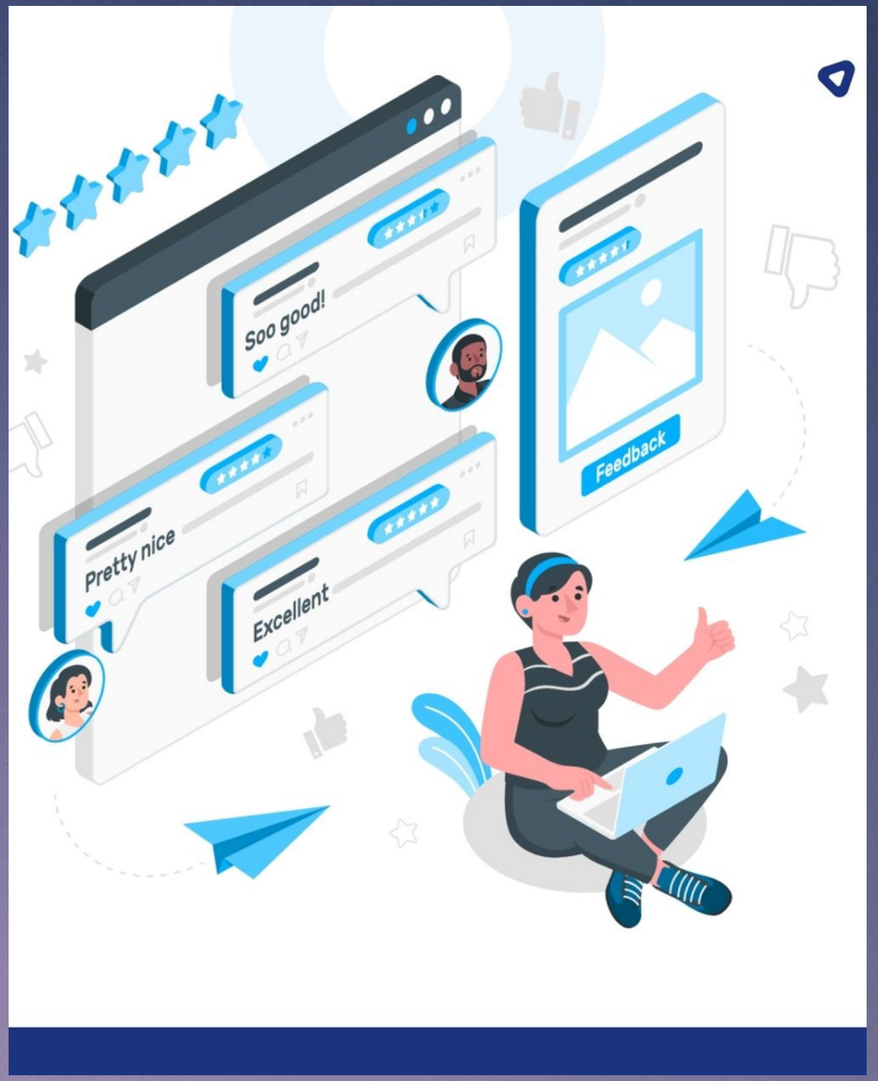
EVALUATION

Quantitative metric

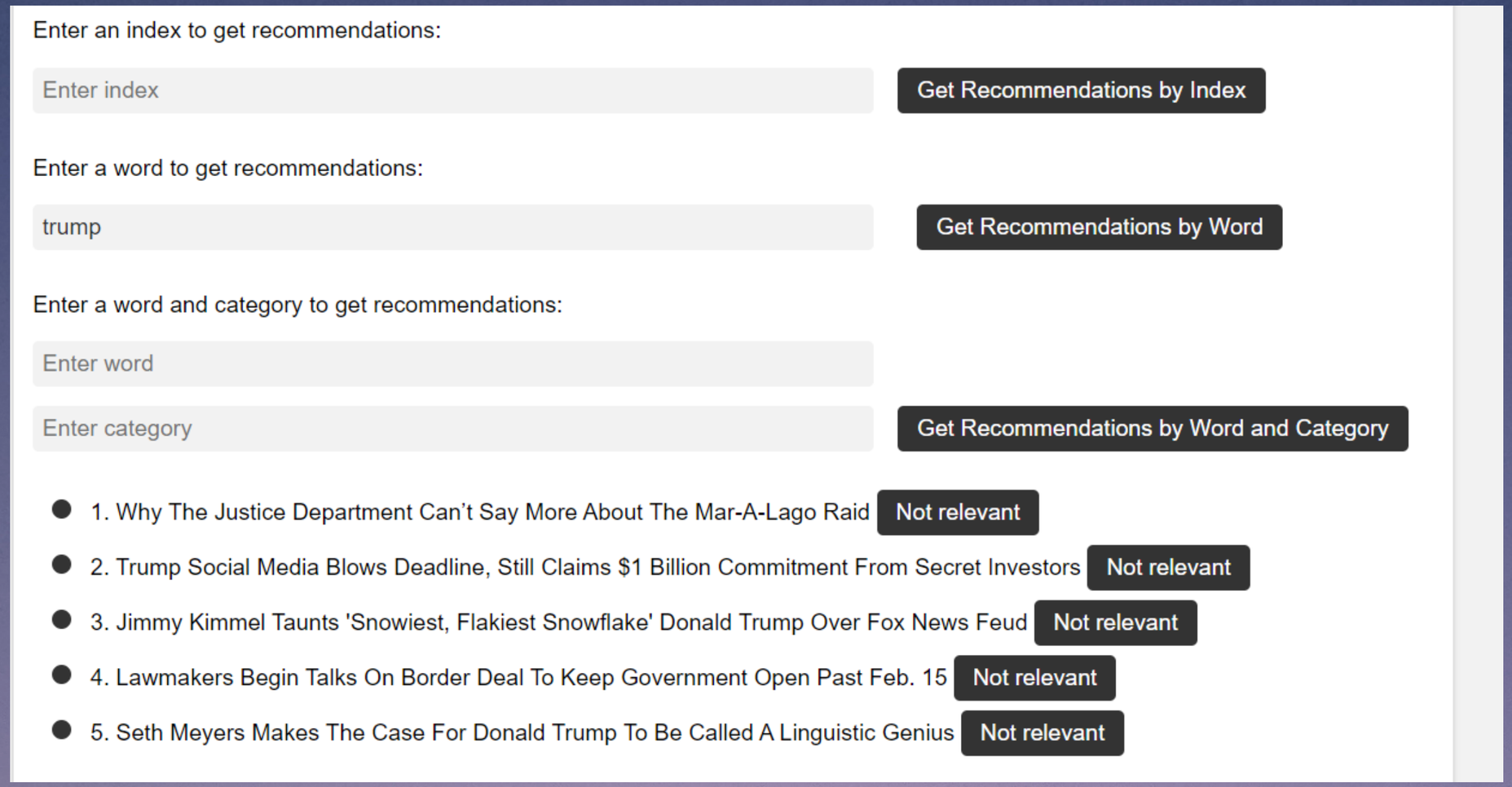
- Tested the recommendation system with a sample dataset or user queries.
- Collected user feedback on the relevance of the recommended articles (relevant or not relevant).
- Calculated the number of relevant articles out of the total recommendations provided.
- Accuracy of 80% was obtained.

Qualitative metric

- Collected user feedback through surveys to understand how satisfied they are with the recommended articles.
- Asked them to rate the relevance, usefulness, or overall satisfaction on a Likert scale (1 to 5).
- Average rating score of 4.2 was obtained.



OUTPUT/DEMO


Enter an index to get recommendations:
Enter index
Get Recommendations by Index
Enter a word to get recommendations:
trump
Get Recommendations by Word
Enter a word and category to get recommendations:
Enter word
Enter category
Get Recommendations by Word and Category
1. Why The Justice Department Can’t Say More About The Mar-A-Lago Raid
Not relevant
2. Trump Social Media Blows Deadline, Still Claims $1 Billion Commitment From Secret Investors
Not relevant
3. Jimmy Kimmel Taunts 'Snowiest, Flakiest Snowflake' Donald Trump Over Fox News Feud
Not relevant
4. Lawmakers Begin Talks On Border Deal To Keep Government Open Past Feb. 15
Not relevant
5. Seth Meyers Makes The Case For Donald Trump To Be Called A Linguistic Genius
Not relevant

CONCLUSION

- The news recommendation system combined cosine-similarity and content-based filtering techniques

- **By providing relevant news articles that were aligned with user interests, the system aimed to enhance user engagement and retention.**

- **user feedback was gathered on the usefulness and relevance of the recommended articles, which assisted in identifying areas for improvement and improving the recommendations**

FUTURE WORK

- Improving feature selection
- Using Explainable recommendations to understand why the recommendations were made
- Incorporating user feedback

THANK YOU!