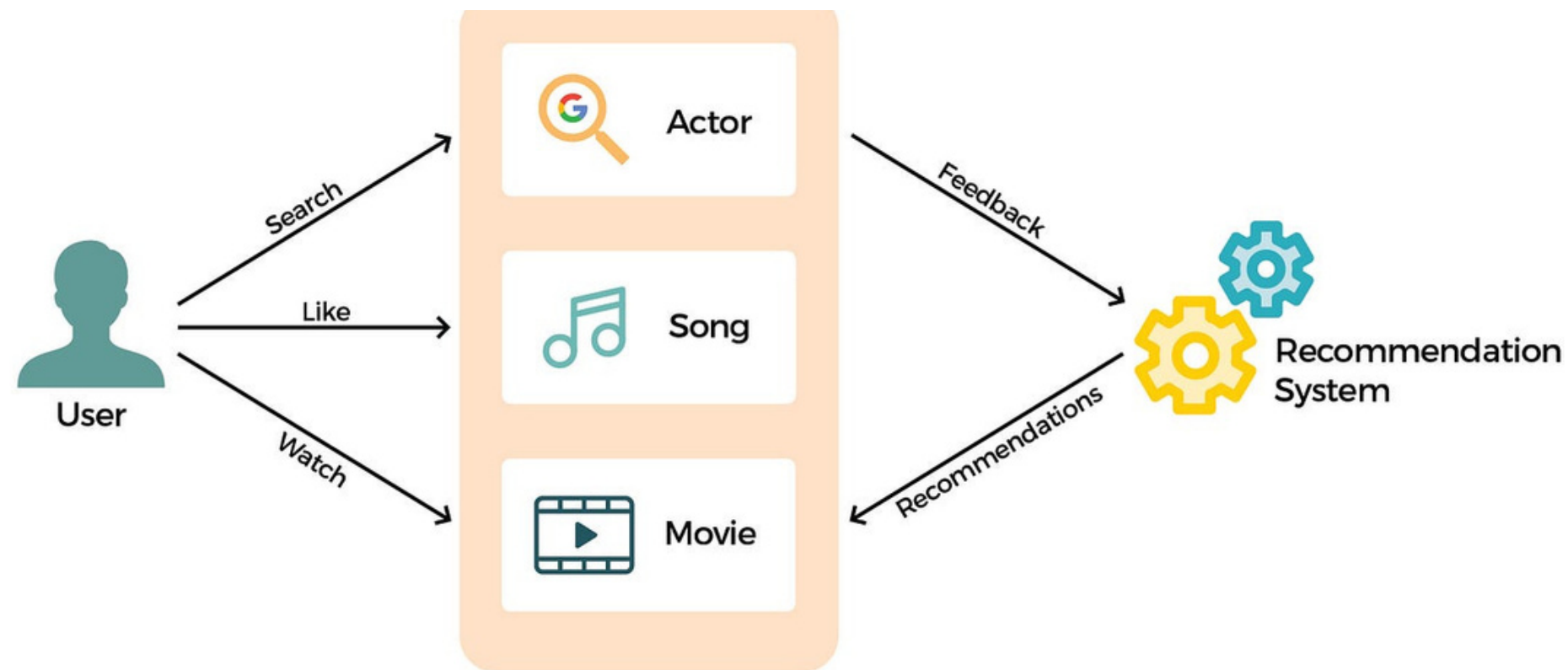


ADAPTIVE RECOMMENDER

Building a Recommendation System that can suggest personalized news articles or blog posts to users based on their reading history and interests, and that can be able to learn and adapt to the user's preferences over time and provide relevant and engaging content.



What is a recommendation system?

A recommendation system is a type of software that suggests items to users based on their interests, preferences, and past behaviors. It works by analyzing a large amount of data about **the User and the Items** being recommended, and then using algorithms to generate personalized suggestions.

AGENDA

Find the Data 4

Choose Technique 5

Preprocessing 6

Model Building 7

Future Work 8

DATASET

	link	headline	category	short_description	authors	date
0	https://www.huffpost.com/entry/covid-boosters-...	Over 4 Million Americans Roll Up Sleeves For O...	U.S. NEWS	Health experts said it is too early to predict...	Carla K. Johnson, AP	2022-09-23
1	https://www.huffpost.com/entry/american-airlin...	American Airlines Flyer Charged, Banned For Li...	U.S. NEWS	He was subdued by passengers and crew when he ...	Mary Papenfuss	2022-09-23
2	https://www.huffpost.com/entry/funniest-tweets...	23 Of The Funniest Tweets About Cats And Dogs ...	COMEDY	"Until you have a dog you don't understand wha...	Elyse Wanshel	2022-09-23
3	https://www.huffpost.com/entry/funniest-parent...	The Funniest Tweets From Parents This Week (Se...	PARENTING	"Accidentally put grown-up toothpaste on my to...	Caroline Bologna	2022-09-23
4	https://www.huffpost.com/entry/amy-cooper-lose...	Woman Who Called Cops On Black Bird-Watcher Lo...	U.S. NEWS	Amy Cooper accused investment firm Franklin Te...	Nina Golgowski	2022-09-22

<https://www.kaggle.com/datasets/rmisra/news-category-dataset>

> This dataset contains around 210k news headlines from 2012 to 2022 from HuffPost.

> Columns:

category: category in which the article was published.

headline: the headline of the news article.

authors: list of authors who contributed to the article.

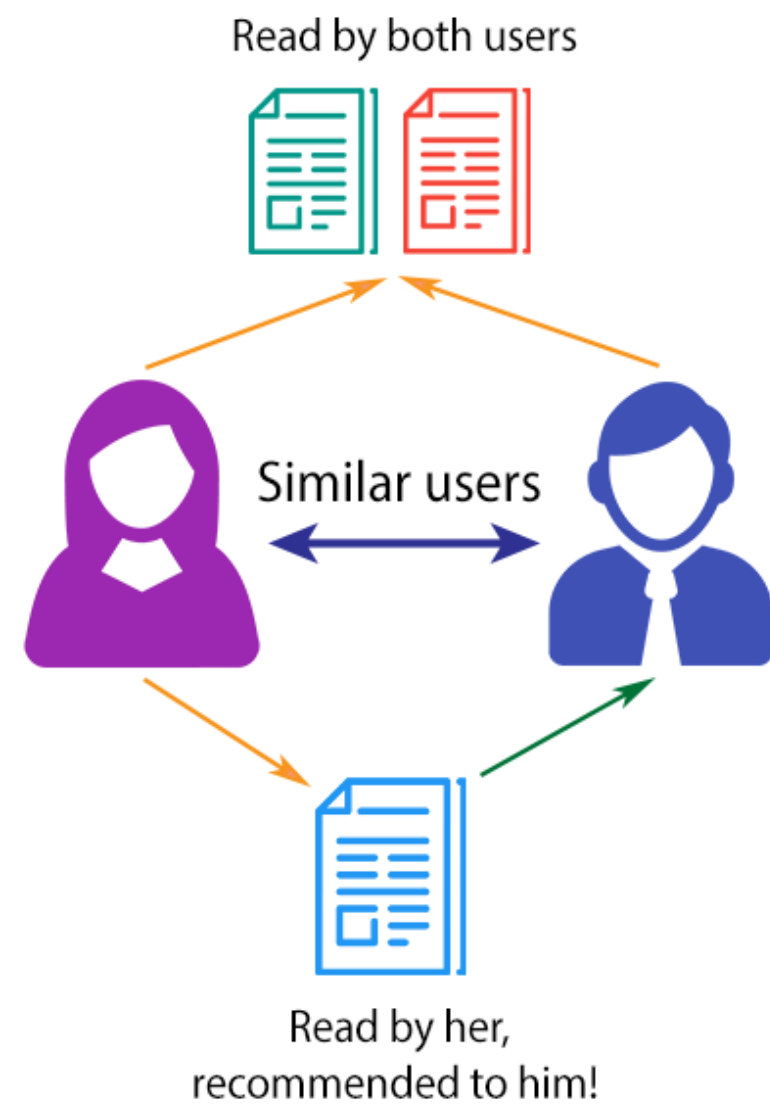
link: link to the original news article.

short_description: Abstract of the news article.

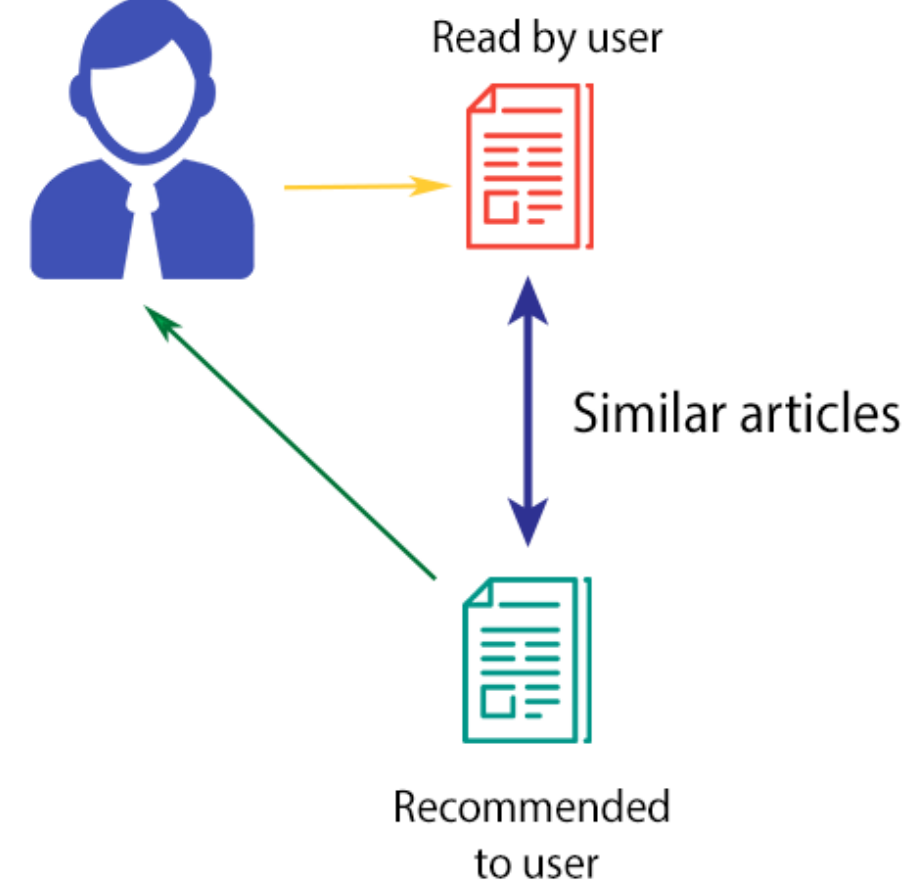
date: publication date of the article.

TECHNIQUE'S

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



Pre-processing of Data

df

	link	category	authors	date	text
0	https://www.huffpost.com/entry/covid-boosters-...	U.S. NEWS	Carla K. Johnson, AP	2022-09-23	Over 4 Million Americans Roll Up Sleeves For O...
1	https://www.huffpost.com/entry/american-airlin...	U.S. NEWS	Mary Papenfuss	2022-09-23	American Airlines Flyer Charged, Banned For Li...
2	https://www.huffpost.com/entry/funniest-tweets...	COMEDY	Elyse Wanshel	2022-09-23	23 Of The Funniest Tweets About Cats And Dogs ...
3	https://www.huffpost.com/entry/funniest-parent...	PARENTING	Caroline Bologna	2022-09-23	The Funniest Tweets From Parents This Week (Se...
4	https://www.huffpost.com/entry/amy-cooper-lose...	U.S. NEWS	Nina Golgowski	2022-09-22	Woman Who Called Cops On Black Bird-Watcher Lo...
...
209522	https://www.huffingtonpost.com/entry/rim-ceo-t...	TECH	Reuters, Reuters	2012-01-28	RIM CEO Thorsten Heins' 'Significant' Plans Fo...
209523	https://www.huffingtonpost.com/entry/maria-sha...	SPORTS	Unknown	2012-01-28	Maria Sharapova Stunned By Victoria Azarenka I...
209524	https://www.huffingtonpost.com/entry/super-bow...	SPORTS	Unknown	2012-01-28	Giants Over Patriots, Jets Over Colts Among M...
209525	https://www.huffingtonpost.com/entry/aldon-smi...	SPORTS	Unknown	2012-01-28	Aldon Smith Arrested: 49ers Linebacker Busted ...
209526	https://www.huffingtonpost.com/entry/dwight-ho...	SPORTS	Unknown	2012-01-28	Dwight Howard Rips Teammates After Magic Loss ...

189814 rows × 5 columns

> Joined the 'Headline' column and 'Short Description' column and made one 'Text' column out of it.

> Replaced missing values in the 'authors' column with 'Unknown'.

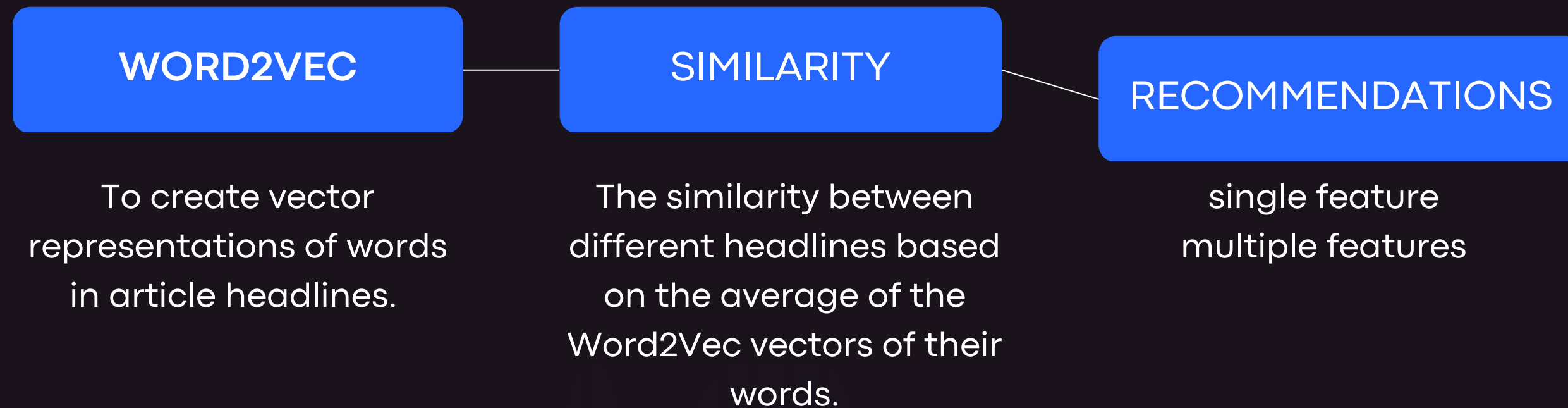
> Drop the rows with null values in the 'text' column.

> Removed Stopwords and did Lemmatization using nltk package.

> Vectorization of 'Text' column using Google's pre-trained Word2Vec model from gensim package.

Model Building

Based on content-based filtering technique.



FUTURE DIRECTIONS!



- > Building a simple UI/ Website for our content-based recommendation system.
- > Using this built website of ours, we will record the user interaction with our recommendation system and built up User data on the news articles like ratings, time spent on reading the article, search history etc.
- > Using this User data, we will then implement Collaborative Filtering on our recommendation system and get much Robust Recommendation System that will recommend articles not only based on new article features but also based on similar users.



Thank You!

MERIN ANTONY
SAKSHAM GAUTAM
SATYAM TOMAR
GAUTAM PRAKASH
LOMESH SONI

