**NEWS ARTICLE RECOMMENDATION**

PATNALA VENKATESH

venkatpatnala2005@gmail.com

***ABSTRACT- In an information-saturated world, news consumption varies widely based on individual interests, with some readers focusing on local updates, others on national issues, and still others on international affairs. This diversity in preferences underscores the need for a personalized approach to news delivery. To address this, we propose a News Article Recommendation System designed to filter and prioritize content based on user-specific interests, browsing patterns, and search history. By leveraging machine learning techniques, such as collaborative filtering, content-based filtering, and hybrid models, our recommendation system can dynamically adapt to each user's preferences, enhancing the relevance of the news articles they receive. This system aims to improve user engagement by reducing information overload and delivering a more efficient, targeted news experience.***

***Keywords -Recommendation system, News article recommendation, Content based filtering, Collaborative* *filtering, hybrid filtering***

**INTRODUCTION**

Staying updated with news is crucial for maintaining awareness of local and global developments, benefiting individuals across diverse fields. However, navigating the overwhelming volume of digital content to locate personally relevant articles presents a significant challenge. For instance, UPSC aspirants require news aligned with exam-relevant topics, while professionals in finance and business seek timely information on economic trends, market volatility, and sector-specific insights. This demand for precise content filtering highlights the importance of an **intelligent News Recommendation System** that dynamically tailors article suggestions based on user interests, search history, and browsing behavior.

In recent years, recommendation systems have gained prominence due to their effectiveness in addressing information overload, optimizing content relevance, and enhancing user engagement. By leveraging algorithms such as collaborative filtering, content-based filtering, and hybrid models, a robust News Recommendation System can provide users with a curated news feed that aligns closely with their specific needs. This study explores the application of these machine learning techniques to deliver personalized news recommendations, thereby improving accessibility, reducing cognitive load, and fostering a more efficient user experience in news consumption.

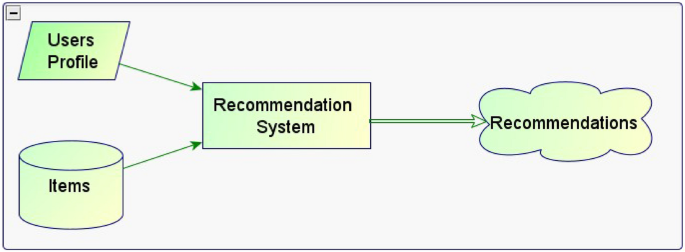
**RECOMMENDATION SYSTEM**

Recommender systems are sophisticated software tools and algorithms designed to suggest relevant items or content to users, based on their preferences, behaviors, and interactions with a platform. A recommender system, as a branch of machine learning, leverages data to guide users in finding products and content that align with their interests. Commonly used on websites and streaming platforms, these systems power “for you” or “you might also like” sections by analyzing user data to make tailored recommendations.

A core feature of any recommender system is its ability to predict user ratings for items before any direct feedback is provided. This prediction mechanism enhances the system’s effectiveness, helping users discover content that aligns with their preferences even before engaging with it directly. In a world overwhelmed by choice, recommender systems have become vital, helping users navigate vast amounts of content while increasing satisfaction and, in commercial contexts, driving sales.

Recommender systems function much like skilled salespeople, identifying what users might like based on their histories and preferences. They’re so ingrained in the digital landscape that many people use them without realizing it. By curating options in an otherwise overwhelming array of content, these systems enhance the user experience and often introduce users to items or content they might not have found on their own.

Some notable examples of recommender systems include Amazon’s product suggestions, personalized movie and TV recommendations on Netflix, suggested videos on YouTube, curated playlists on Spotify, and targeted content in the Facebook news feed and Google Ads.



There are three primary approaches to building a recommendation system:

1. Content-Based Filtering
2. Collaborative Filtering
3. Hybrid Approach

Content-Based Filtering

Content-based filtering is a recommendation system technique that suggests items to users based on their preferences and the similarity of items they have previously interacted with. This approach analyzes the characteristics of past interactions to recommend new, similar items that align with the user’s interests.

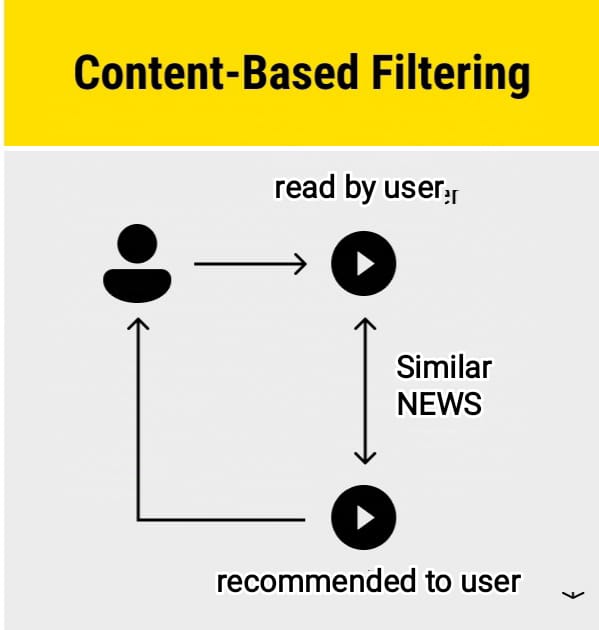


Fig-1 Content-Based Filtering.

The above Fig-1 shows content-based approach is widely applied across various platforms. For example, Netflix utilizes this method for movie recommendations, while YouTube leverages it for suggesting videos. In the context of news recommendation systems, this approach proves particularly effective, as it can personalize news suggestions based on a user’s searches, clicks, and browsing history. By analyzing these behaviors, content-based filtering can efficiently deliver news articles that closely match individual interests. The mathematics behind content-based recommendation systems relies on similarity measures to predict user preferences based on their past interactions. One key concept is **Term Frequency (TF)**, which calculates the frequency of each attribute within an item—for example, the number of times a specific word appears in an article. These TF values are used to create an attribute vector for each item, making it particularly effective for recommending news articles by identifying key terms and matching them to user interests and the popular common algorithm that measures the angle between two vectors, such as a user profile and an item profile. The cosine score ranges from -1 to 1, with higher scores indicating more similar items This is know as Cosine similarity.

Collaborative Filtering

Collaborative filtering is a recommender system that suggests items to users based on how other users with similar preferences have interacted with those items This approach identifies patterns in user-item interactions to provide recommendations based on shared interests within user groups, making it a robust method for personalized suggestions. Recommendation is based on similarities between use’s experience so it has better accuracy than content based recommendation technique and sometimes this helpful for good recommendations and user can feel good to use our website or application.

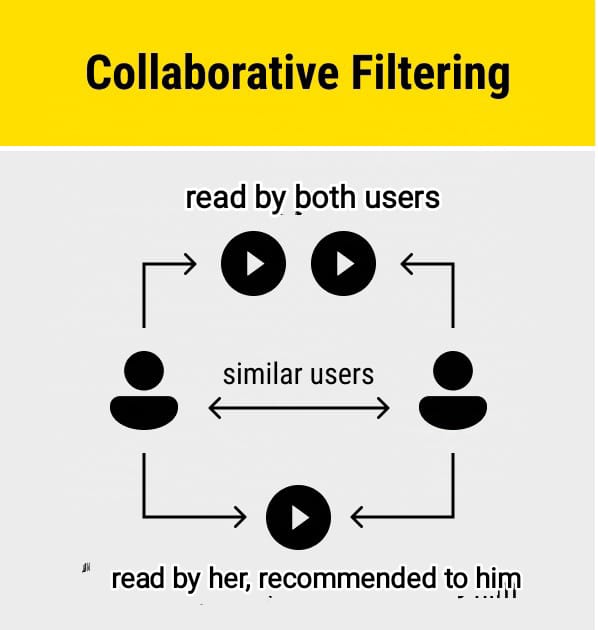
****

Fig-2 Collaborative Filtering

A clear example of collaborative filtering in action is the post and reel recommendations on platforms like Facebook and Instagram. Since friends often share at least some common interests, collaborative filtering leverages these shared preferences to personalize content. For instance, if your friend has liked or commented on a post, that post is more likely to appear in your feed or reel, enhancing your experience by showing content that aligns with your social circle’s interests. The above Fig-2shows you collaborative filtering. When combined with our news article recommendation system, this approach can lead to even more accurate and personalized news recommendations by factoring in both individual and social preferences. The mathematical foundation of collaborative filtering often involves **Matrix Factorization**, a popular technique for generating recommendations. One of the widely used methods in this approach is the Alternative Least Squares (ALS) model, which optimizes an objective function to minimize error. The ALS model employs a Root Mean Square Error (RMSE) loss function to measure the difference between predicted and actual ratings, iteratively adjusting until the best possible factorization is achieved. This allows for more accurate predictions of user preferences by breaking down complex user-item interactions into simpler, interpretable components.

Hybrid Approach

The hybrid approach integrates multiple recommendation techniques, such as combining collaborative and content-based methods, to enhance recommendation accuracy and relevance. By merging these techniques, hybrid systems leverage the strengths of each, yielding more effective and precise recommendations. This approach improves user satisfaction and is widely adopted for better predictive outcomes.

**Hybrid Techniques**: The hybrid approach can use various techniques to integrate these methods, such as:

* Weighted Blending
* Switching
* Feature Combination

**Pros of using Hybrid Approach:**

* Improved Accuracy
* Enhanced Personalization
* Better User Experience

The perfect example to understand you about hybrid approach is instagram. The below fig-3 shows you simple understand about hybrid approach how the techniques are integrated to get good recommendation.

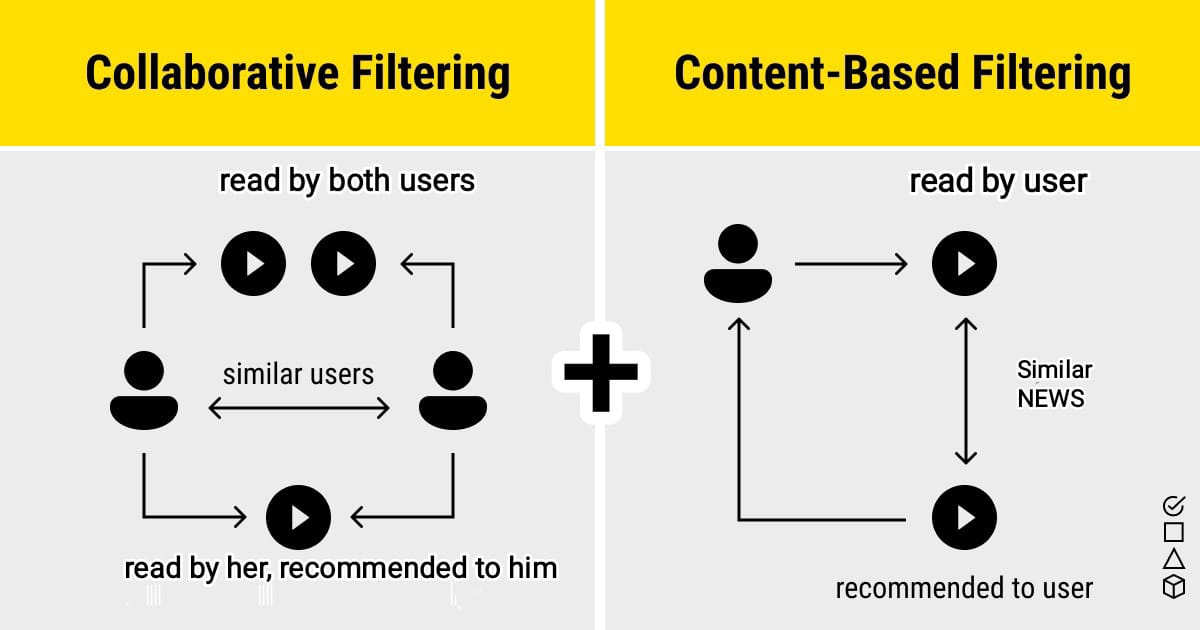
****

Fig-3 Hybrid Approach

**DATASET**

The dataset used for this project is the Microsoft News website Dataset, available from ka ggle. It contains articles across various news categories such as technology, sports, health, weather, travel and politics. The dataset contains a total of 43,460 rows and 8 columns. After removing null values, it has 36,422 rows while maintaining the same 8 columns. This cleaned dataset ensures more reliable analysis and modeling. This dataset’s diversity in topics and depth in each article makes it ideal for training a content-based recommendation system that can suggest relevant news based on textual similarity.

**CHALLENGES**

Locating a suitable dataset proved challenging due to the vast number of sources and datasets available online, requiring extensive reference-checking to ensure an up-to-date and relevant choice. Ultimately, I sourced the dataset from Kaggle. Identifying and removing null values presented another challenge; careful consideration was needed to avoid discarding essential information, as this could reduce model accuracy. Additionally, while cosine similarity is applied, it doesn’t require analyzing every row. Future work will involve improving data quality, including implementing grammar correction and minimizing redundancy**.**

**IMPLEMENTAIION AND RESULTS**

The code implements a content-based recommendation system using cosine similarity to suggest news articles based on textual content. It begins by importing essential libraries, loading a dataset from a CSV file, and examining the structure and contents of the data. Using `Tf-idfVectorizer`, it converts the text in the `lifestyle` column into numerical features that represent each article as a vector in a high-dimensional space. The TF-IDF transformation assigns weights to words based on their frequency across all articles, emphasizing terms that are more unique within each article. The cosine similarity function then calculates similarity scores between article pairs, allowing the system to quantify how closely related two articles are based on their content.

The core of the recommendation process lies in the `recommended articles` function, which, given an article title, identifies its most similar articles by ranking cosine similarity scores. This approach ensures that the recommendations are content-driven, relying purely on textual overlap and context rather than user behavior. When run on a sample title, the function outputs the top 5 related articles, making it valuable for suggesting similar news items in the lifestyle domain. This content-based filtering system is adaptable for any dataset where text-based similarity can drive recommendations.

**FUTURE SCOPE**

I will add a framework to this recommendation algorithm so users can interact easily. If we succeed, this news recommendation system can be used not only for article recommendations but also in police investigations. As crime rates are increasing daily, reviewing all old newspapers is becoming increasingly difficult. I will work on this aspect. So far, I have completed a paper focusing solely on content-based recommendations. In the future, I will conduct research on collaborative and hybrid approaches to enhance accuracy.

**CONCLUSION**

In summary, this research presents a content-based news recommendation system designed to improve user engagement and streamline access to pertinent information. By utilizing the content of articles, the system not only delivers tailored recommendations but also tackles the critical issue of analyzing historical news articles, especially given the rising crime rates. Although the current model showcases the effectiveness of content-based filtering, future studies should investigate collaborative and hybrid approaches to further enhance recommendation accuracy. By broadening the system's functionality, it could not only help users navigate news content more effectively but also aid law enforcement agencies in identifying trends and patterns in crime through the examination of historical data.

**REFERENCES**

* <https://content.iospress.com/articles/semantic-web/sw222991>
* <https://dl.acm.org/doi/abs/10.1145/3530257>
* <https://www.tandfonline.com/doi/full/10.1080/23808985.2022.2142149#abstract>
* <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8963698>
* <https://dl.acm.org/doi/abs/10.1145/358916.358995>
* <https://onlinelibrary.wiley.com/doi/full/10.1155/2020/8834908>
* <https://link.springer.com/article/10.1007/s11036-017-0842-9>

**github link:**

[**https://github.com/72venkatesh/News\_Recommendation**](https://github.com/72venkatesh/News_Recommendation)