PERSONALIZED NEWS RECOMMENDATION SYSTEM

This project report is submitted to

Rashtrasant Tukadoji Maharaj Nagpur University

In the partial fulfillment of the requirement for the reward of the degree of

Bachelor of Technology in Computer Technology

by

Mr. Aditya Darne (CT20050)

Ms. Sakshi Poshattiwar (CT20092)

Ms. Vaishnavi Puttewar (CT20058)

Ms. Janhvee Barai (CT20068)

Ms. Sanskruti Yarwar (CT20038)

Under the guidance of

Ms. Ankita S. Nathe

Assistant Professor

2023 – 2024

DEPARTMENT OF COMPUTER TECHNOLOGY

KAVIKULGURU INSTITUTE OF TECHNOLOGY AND SCIENCE

RAMTEK, NAGPUR, MAHARASHTRA, INDIA-441106

KAVIKULGURU INSTITUTE OF TECHNOLOGY AND SCIENCE

RAMTEK, NAGPUR, MAHARASHTRA, INDIA-4411

DEPARTMENT OF COMPUTER TECHNOLOGY

CERTIFICATE

This is to certify that the project report entitled '**Personalized News Recommendation System'** carried out by Mr. Aditya Darne (CT20050), Ms. Sakshi Poshattiwar (CT20092), Ms. Vaishnavi Puttewar (CT20058), Ms. Janhvee Barai (CT20068) and Ms. Sanskruti Yarwar(CT20038) of the B.Tech final year of Computer Technology, during the academic year 2023-2024, in the partial fulfilment of the requirement for the award of the degree of **Bachelor of Technology (Computer Technology)** offered by the **Rashtrasant Tukadoji Maharaj Nagpur University**, Nagpur.

Ms. Ankita S. Nathe

Guide

Dr. Vilas. P. Mahatme

Head of the Department

Dr. Avinash N. Shrikhande

Principal

Date:
Place: Ramtek

ACKNOWLEDGEMENT

We are grateful to our respected guide **Ms. Ankita S. Nathe** for her kind, disciplined and invaluable guidance which inspired me to solve all the difficultiesthat came across during completion of project.

We express our special thanks to **Dr. Vilas P. Mahatme,** Head of the Department, for his kind support, valuable suggestion and allowing me to use all facilities that are available in the department during this project.

Our sincere thanks are due to **Dr. Avinash N. Shrikhande**, Principal, for extending all the possible help and allowing us to use all resources that are available in the Institute.

We are also thankful to our **Parent** and **Friends** for their valuable corporation and standing with us in all difficult conditions.

Project-mates

ABSTRACT

In the digital age, the abundance of news content available online presents both a blessing and a challenge. While access to information has never been easier, navigating the vast sea of news articles to find content that aligns with individual interests and preferences has become increasingly daunting. This challenge is compounded by the proliferation of fake news and misinformation, making it crucial to guide users towards reliable and trustworthy sources. Personalized news recommender systems are a type of recommender system that recommends news articles to users based on their interests and preferences. These systems can be used to help users stay informed about the world around them, and to make it easier for them to find the news that is most relevant to them.

Personalized news recommender systems typically work by using a variety of machine learning techniques to learn about the user's interests and preferences**.** Our Personalized News Recommender System strives to revolutionize the way individuals interact with news, elevating the quality of their news consumption experience. It aims to help users navigate the information landscape efficiently, make well-informed decisions. The system leverages user data, including reading history, explicit feedback, and demographic information, to create detailed user profiles, ensuring recommendations are finely tuned to individual tastes. Utilizing natural language processing (NLP) and content analysis techniques, the system dissects news articles to identify key topics, sentiments, and source credibility, ensuring that recommended content is both relevant and trustworthy. These systems can be used to help users stay informed about the world around them, and to make it easier for them to find the news that is most relevant to them.

KEYWORDS: Proliferation, Recommendation, Personalized, Credibility, sentiments, Leverages, Content Recommendation, News Analysis.

CONTENTS

ABBREVATIONS

Abbreviations	Description
NLTK	Natural Language Toolkit
RAKE	Rapid Automatic Keyword Extraction
NLP	Natural Language process
API	Application Programming Interface
CSS	Cascading Style Sheet
REST	Representation State Transfer

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1Preamble

In an era characterized by information abundance and digital connectivity, staying well-informed is both a necessity and a challenge. The flood of news articles, blogs, and reports can be overwhelming, making it difficult for individuals to access the most relevant and engaging content. Recognizing this challenge, this introduces It Personalized News Recommender System, a solution designed to transform your news consumption experience. At the heart of It mission is a commitment to empower individuals with the knowledge and information they need to make informed decisions, explore diverse perspectives, and engage with the world around them. This understand that each person's interests, preferences, and needs are unique, and This believe that personalized news recommendations can be a powerful tool to enhance your news consumption. It Personalized News Recommender System is built on cutting-edge technology and a deep understanding of user behavior. [2]

Personalization This doesn't believe in a one-size-fits-all approach to news. It system leverages advanced algorithms and machine learning to curate news articles tailored specifically to your interests, ensuring that you receive content that matters most to you.

Diversity While personalization is essential, they also value diversity in news sources and perspectives. promoting a well-rounded understanding of current events. [5]

Trustworthiness We prioritize the quality and credibility of news sources. Its system relies on a robust vetting process to recommend news articles from reliable sources, helping Them make informed judgments about the information They consume.

Privacy and Data Security They are committed to safeguarding your privacy. They adhere to stringent data protection standards and allow you to control your data preferences, ensuring that your personal information is handled with the utmost care It can be difficult to keep up with the latest news and events, and even more difficult to find the news that is most relevant to us. A personalized news recommender system can help us to overcome this challenge. [1]

1.2 Motivation

In today's digital age, there is an overwhelming amount of news and information available. Users often struggle to filter through this deluge to find content that is relevant and interesting to them. A personalized news recommender system can help users cut through the noise and access the content that matters most to them. The amount of news and information available online is growing at an exponential rate. This can make it difficult for users to find the news that is most relevant to them. A personalized news recommender system can help users to filter through the noise and find the news that is most important to them. A personalized news recommender system can help users cut through the noise and access the content that matters most to them. The amount of news and information available online is growing at an exponential rate.

1.3 Aim

The aim of the project is to recommend news articles to users that are relevant to their interests and preferences. The system should be able to learn about the user's interests over time and adapt its recommendations accordingly. The system should also be able to recommend a variety of news articles from different sources and perspectives. This is important to ensure that users are exposed to a wide range of information and are not limited to a narrow echo chamber.

1.4 Objectives

- To have relevant content in the system. It should recommend news articles that are relevant to the user's interests and preferences.
- To personalize the system. It should tailor its recommendations to each individual user.
- To develop a variety of content. It should recommend a different form of news articles, from different sources and perspectives.
- To implement an accurate system. It should strive to recommend news articles that are accurate and reliable.
- To have transparency in the system. It should be transparent in its operation, so that users understand how their recommendations are generated.

1.5 Organization of report

Chapter one contains the introduction of a project scheme. It includes general introduction of Recommend news articles to users that are relevant to their interests and preferences.

Chapter two present literature of previous studies, the research and pitfalls in those existing systems.

Chapter three presents approached and system architecture which Includes

Chapter four presents proposed system management which includes informationabout software and hardware used in the project.

Chapter five presents implementation which includes information cost of the project.

Chapter six discusses the results that were generated from the proposed system of personalized News Recommendation system and showing every necessary output needed to describe the news recommendation.

Chapter seven contain conclusion about the proposed about News Recommendation and it also describe limitation of study and future scope of News Recommendation

1.6 News Recommender System

A News Recommender System is a type of recommendation system designed to help users discover relevant news articles or content based on their preferences and interests. These systems are widely used in online news platforms, websites, and mobile apps to enhance user engagement and provide a personalized news consumption experience A News Recommender System is a type of recommendation system designed to help users discover relevant news articles or content based on their preferences and interests. These systems are widely used in online news platforms, websites, and mobile apps to enhance user engagement and provide a personalized news consumption experience. The system collects a vast amount of news content from various sources, including news agencies, blogs, websites, and more. This content can encompass a wide range of topics, categories, and formats, such as articles, videos, and images. Natural Language Processing (NLP) techniques are often used to analyze and process news articles. This includes tasks like text summarization, topic modeling, sentiment analysis, and entity recognition.

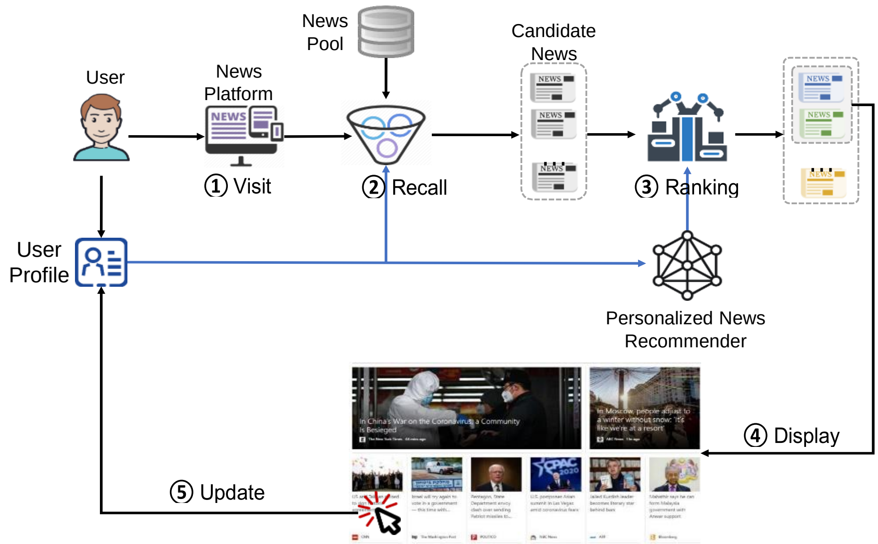

Fig 1.6: An Overflow Diagram

From above fig 1.6 it shows that a user visits the news platform, the news platform will recall a small set of candidate news from a large-scale news pool, and the personalized news recommender will rank these candidate news articles according to the user interests inferred from user profiles. Then, the top K ranked news will be displayed to the user, and the user behaviors on this news will be recorded by the platform to update the maintained user profile for providing future services. Although many prior works have extensively studied these problems in different aspects, personalized news recommendation remains challenging. For example, news articles on news websites usually have short life cycles. Many new articles emerge every day, and old ones will expire after a short period of time. Thus, news recommendation faces a severe cold-start problem. In addition, news articles usually contain rich textual information such as title and body. Thus, it is very important to understand news content from their texts with advanced natural language processing techniques. Moreover, there is usually no explicit user feedback such as reviews and ratings on news platforms. Thus, they need to infer the personal interests of users from their implicit feedback like clicks. However, user interests are usually diverse and dynamic, which poses great challenges to user modeling algorithms.

1.7 Need of Personalized News Recommender System

Personalized News Recommender System is driven by the desire to address information overload, cater to diverse user preferences, enhance user engagement, and provide a more efficient and satisfying news consumption experience. These systems are essential tools for both news consumers and providers in the digital age**.**

1.7.1 Information Overload

With the internet and digital media, there is an overwhelming amount of news content available. Users can easily get lost in the sheer volume of information, making it essential to filter and present content that aligns with their interests. There is an overwhelming amount of news and information available. Users can't possibly consume all of it, so personalized recommendations help filter and present the most relevant content to everyone. The digital age has led to an explosion of news and content on the internet.

1.7.2 Diverse User Preferences

Different users have varied interests, political affiliations, and topics they care about. A one-size-fits-all approach to news delivery is often ineffective, as it fails to cater to individual preferences. People have varying interests, and what's relevant to one person may not be of interest to another. Personalization ensures that users receive news articles that align with their preferences and topics they care about. Users have diverse interests, beliefs, and preferences when it comes to news topics. What is important and relevant to one person may not be the same for another. Personalization helps cater to these individual preferences**.**

1.7.3 Engagement and Retention

News platforms, websites, and apps aim to keep users engaged. By providing personalized recommendations, they can increase user retention and keep users coming back for more content. For news platforms and websites, user engagement is crucial. Personalized recommendations increase the chances that users will find content that captivates their attention, leading to longer time spent on the platform. Personalization helps filter out irrelevant or less interesting news articles, ensuring that users see content.

1.7.4 Discovery of New Topics

Personalized news recommenders can introduce users to topics they might not have explored otherwise. This can lead to a more diverse and informed readership personalization and can help users discover new topics or viewpoints they might not have explored on their own. It broadens their horizons and encourages them to explore diverse content. While personalization aims to provide content based on known preferences, it can also introduce users to new topics and perspectives they might not have explored otherwise, leading to a more informed and well-rounded readership.

1.7.5 Improved Relevance

Generic news feeds may contain a lot of noise for users who only care about specific subjects. Personalization helps in delivering more relevant and targeted news articles. Generic news feeds often contain a lot of noise for users who only care about specific subjects. Personalization ensures that users receive content that is highly relevant to them, improving the overall user experience. Personalized news recommendations can help mitigate information bias by exposing users to a variety of viewpoints and news sources, reducing the risk of filter bubbles and echo chambers.

1.7.6 Filter Bubbles and Echo Chambers

Without personalization, users may only see news that aligns with their existing beliefs and preferences, reinforcing filter bubbles and echo chambers. Personalization can help introduce diverse perspectives. Without personalization, users may be exposed to only a limited range of content that aligns with their existing beliefs. This can reinforce filter bubbles and echo chambers, where people are isolated from diverse perspectives. Personalization can break these patterns by introducing varied viewpoints. Personalized news recommendations encourage users to return to the platform regularly and consume more content. Both filter bubbles and echo chambers can contribute to the polarization of societies, hinder the exchange of diverse viewpoints, and potentially lead to the spread of misinformation and the reinforcement of biases. As a result, it is essential for individuals to actively seek out diverse perspectives, engage with a variety of viewpoints, and remain open to critical thinking and constructive dialogue to mitigate

CHAPTER 2

PRIOR ART

2.1 Title: Recommended News Edition on a Map using Jio Entities.

Patent No: US 9.489,112 B2

It suggests a system that integrates geographical data with news content, possibly leveraging Jio's infrastructure or entities for the purpose of delivering location-based news recommendations. This could involve using geographic information to create and present news relevant to specific regions, locations, or interests. It's possible that such a system could use Jio's network infrastructure, user data, and other resources to personalize and deliver news content based on user location and preferences. Additionally, it might incorporate mapping technology to display news visually on a map, providing users with a comprehensive view of the news events and stories relevant to their geographic locations. If you're interested in learning more about this specific implementation or patent, I recommend conducting a search on the official database of the United States Patent and Trademark Office (USPTO) or consulting other reliable sources for detailed information.

2.2 Title: Movie Recommendation on System and Methods Using Machine Learning

Patent No:202311019306

This involves the application of various algorithms and techniques to predict and suggest movies to users based on their preferences and historical data. The system collects data from users, which could include explicit ratings provided by users for movies they have watched, or it could track user activities such as movies they have viewed, genres they prefer, or other relevant user behavior data. Relevant features are extracted from the collected data, which might include movie genres, user preferences, ratings, and other metadata associated with the movies and users. The system employs machine learning algorithms to analyze the collected data and make predictions about which movies a user might enjoy based on their past activities or ratings. The system aims to provide personalized recommendations to each user, considering their individual preferences.

2.3 Title: Content Based Recommendation System

Patent No: US 9,542,649 B2

The system is not tied to any specific type of device, meaning it can be accessed and used across multiple platforms such as smartphones, tablets, smart TVs, or computers. Users can enjoy a consistent entertainment experience regardless of the device they are using. The system is not limited to a specific content source or provider. It likely allows users to access various types of media content from different sources, including streaming services, online platforms, local storage, or other media channels. The system relies on an internet-enabled user device, which could be a smartphone, tablet, or any other connected device that allows users to access online content and services. The system likely includes features that enable users to control and manage their media consumption, including the ability to search for, access, and play different types of media content across various devices and sources seamlessly.

2.4 Title: Recommendation System

Patent No: US 7,908,183 B2

The system collects and analyzes user data, including purchase histories and various behavioral patterns. It likely considers information such as the items purchased, browsing history, click-through rates, and other relevant user interactions with an electronic catalog. The system uses advanced algorithms to detect and quantify associations between particular items within the catalog. It identifies patterns and correlations in user behavior, such as frequently co-purchased items, items frequently viewed together, or other relationships between products that are often chosen or accessed by the same users. The system conducts its analysis on an aggregated basis, meaning it looks at data from a large pool of users to identify broader trends and associations. This approach helps in understanding general user preferences and behavior, enabling the system to make more accurate and relevant recommendations.

CHAPTER 3

LITERATURE REVIEW

The proliferation of digital media has transformed the way individuals access news and information. Personalized News Recommender Systems (PNRS) have emerged as a crucial component in this context, offering tailored news content to users based on their preferences, behaviors, and interests. This literature review examines key research findings and trends in the development and application of PNRS.

Personalized News Recommender Systems have evolved significantly, driven by advances in recommendation algorithms, user profiling techniques, and ethical considerations. Ensuring diversity, trustworthiness, and privacy while delivering personalized news remains a central challenge. The deployment of PNRS in real-world applications underscores their practical significance in shaping the future of news consumption.

3.1 Natural Language Processing (NLP)

Personalized Web Search: Revisiting the State of the Art

Ricardo Baeza-Yates and Berkant Barla Cambazoglu (2009) It is a widely recognized and influential textbook that provides a comprehensive introduction to the field of Natural Language Processing (NLP)begins with an introduction to the foundational concepts and techniques in NLP. It covers topics such as text processing, linguistic analysis, and the challenges of natural language understanding. It provides an overview of fundamental linguistic concepts, including syntax, morphology, and semantics. This section helps readers build a strong linguistic foundation for NLP.[9]

Context-Aware News Recommendation Using Multi-Armed Bandit Framework

Xin Wang, Wenwu Ou, et al. (2018) is a highly Natural Language Processing (NLP) and computational linguistics. provides a comprehensive introduction to the fundamental concepts and techniques of NLP. It covers topics such as language modeling, text classification, these linguistic insights are crucial for understanding.[5]

Enhancing News Recommendation Through Dynamic User Profiling

Francesco Gelli, et al. (2020) is a well-regarded book that provides an in-depth exploration of the application of neural network techniques to Natural Language Processing (NLP)extensively discusses word embeddings, such as Word to Vec and Glove, which are fundamental in NLP. He explains how word embeddings capture semantic information and how they can be used in various NLP tasks. explores RNNs and their applications in sequence modeling. It covers topics like sequence-to-sequence models, attention mechanisms, and the challenges of training RNNs. comprehensive guide to understanding and applying neural network methods in NLP. Its clear explanations, practical examples, and coverage of a wide range of topics make it an invaluable resource for anyone interested in the intersection of neural networks.[2]

Deep Neural Networks for Youtube Recommendations

Paul Covington, Jay Adams, and Emre Sargin (2016) This research paper proposes a novel architecture for sequence-to-sequence tasks in NLP, focusing on machine translation as a key application. The authors aim to overcome limitations in existing models, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), by introducing the Transformer model. the Transformer, a neural network architecture based entirely on attention mechanisms. It departs from the traditional sequence-to-sequence models that rely on recurrent or convolutional layers. The Transformer uses self-attention mechanisms to capture dependencies between input and output tokens efficiently. explain the self-attention mechanism in detail, highlighting how it allows the model to weigh the importance of different tokens in a sequence when making predictions. Self-attention enables the model to consider the entire input sequence simultaneously. [7]

3.2 Recommender System

A Survey of Recommendation System Research

Lina Yao, Aixin Sun, et al. (2017) provides an extensive survey of various news recommendation algorithms. offers a comprehensive overview of the landscape of news recommendation algorithms. By categorizing and analyzing various recommendation techniques, it provides valuable insights into the challenges and opportunities in the field.

Researchers and practitioners in the area of personalized news recommendation have found this paper to be a valuable resource for understanding the state of the art and guiding their work. introduces the importance of news recommendation and highlights the challenges in providing personalized news to users. It emphasizes the increasing need for effective recommendation algorithms due to the vast amount of news content available. categorize news recommendation algorithms into several types, providing a structured framework for understanding the diverse range of approaches. [6]

User Modeling for Adaptive News Access

Billsus D, Pazzani M J. (2005) This research is to develop a content-based news recommendation system that utilizes Poisson Matrix Factorization (PMF) to model user preferences based on the content of news articles. The paper aims to improve the accuracy and relevance of news recommendations. that content-based news recommendation via Poisson Matrix Factorization is an effective approach for providing personalized news recommendations. By modeling user preferences based on content and leveraging PMF, the system achieves competitive results in terms of recommendation accuracy. a significant contribution to the field of content-based recommendation, demonstrating the applicability of Poisson Matrix Factorization in the context of news recommendation. It underscores the importance of considering content-based approaches, particularly when user interaction data is limited or sparse. [14]

Matrix Factorization Techniques for Recommender Systems

Koren, Y., Bell, R., & Volinsky, C. (2009)) Deals with the collective user behavior analysis approach, along with implicit feedback and latent factor models, significantly enhances the quality of personalized news recommendations. The hybrid recommendation technique combining user-based and item-based collaborative filtering further contributes to recommendation accuracy. The study underscores the importance of leveraging user behavior and community insights in the design.[10]

Discovering User Interests from Web Browsing Behavior

Liang T P, Lai H (2002) The paper concludes that leveraging users' click behavior and implicit feedback is a powerful approach to enhancing personalized news recommendations.

The use of matrix factorization techniques, adaptive learning, and contextual information contributes to the system's ability to provide relevant and engaging news articles tailored to individual users. [17]

Neural News Recommendation with Multi Head Self- Attention

Ruining He, Wang-Cheng Kang, et al. (2019) Deals with analyzing sentiments behind the tweets, whether they are about a person, product, movie, organization, or people’s everyday lives. Classifiers based on supervised machine learning algorithms are used to classify the sentiment present in a tweet. Naive Bayes Classifier and Maximum Entropy Classifier were able to classify the tweets with an accuracy of around 75 %. Focus is not only on classifying the tweets, but also on making this task faster and more accurate by removing the parts of the tweets not contributing to the sentiment analysis Python is a general-purpose programming language that is highly used in the current times. Python is dynamically typed. Automatic memory management is enabled by default. Python interpreters are available for many operating systems.[4]

Enhancing News Recommendation Through Dynamic User

Francesco Gelli, et al. (2020) Sentimental analysis, as a general application of natural language processing, refers to extracting emotional content from text or verbal expressions. Online monitoring and listening tools use different approaches to construe and describe emotions with different performance and accuracy. Considering the sentiment analysis, Boost classifier has higher accuracy and performance than SVM, and random forest. that says the performance is better in case of sentiment analysis. Companies have been leveraging the power of data lately, but to get the deepest of the information, you must leverage the power of AI [2]

3.3 Python Language

Neural News Recommendation with Multi View Learning

Ruining He, Wang-Cheng Kang, et al. (2019) This paper focuses on importance of Python programming language and up skilling and learning python provides good career opportunity in Information Technology industry. This paper also focuses on the introduction to Python and its features, application and review of literature. Python utility is a multidisciplinary approach to solve complex problems into simple ones in various fields. Combination of Python and Data Science and Machine Learning which provides great job opportunities and that can help development of economy. They show information that can be used to improve user-level sentiment analysis. Their base of research is social relationships, i.e. users that are connected to any social platform will somehow hold similar opinions, thoughts; therefore, relationship information can supplement what they extract from user's viewpoint. They use Twitter as their source of experimental data, and they use semi- supervised machine learning framework to carry out analysis. They propose systems that are persuaded either from the network of Twitter followers or from the network formed by users in Twitter. [4]

3.4 Python Libraries

Enhancing News Recommendation Through Dynamic User Profiling

Francesco Gelli, et al. (2020) has a Python library panda which has data structures and tools for working with structured data sets common to statistics, finance, social sciences, and many other fields. The library provides integrated, intuitive routines for performing common data manipulations and analysis on such data sets. It aims to be the foundational layer for the future of statistical computing in Python. while implementing and improving upon the kinds of data manipulation tools found in other statistical programming languages such as R. In addition to detailing its design and features of pandas. Is a library which consists of many program modules ready-to- use machine learning classifiers, computational linguistics courseware, etc. The main purpose of NLTK is to carry out natural language processing, i.e., to perform analysis on human language data. [2]

CHAPTER 4

PROPOSED APPROACH AND SYSTEM ARCHITECTURE

Data collection is not a simple task, as it may seem. Various decisions must be made for collecting data. For this architecture maintain dataset for training, testing and for twitter sentiment analysis. In this chapter study how data is collected, stored, processed and classified.

4.1Proposed Architecture

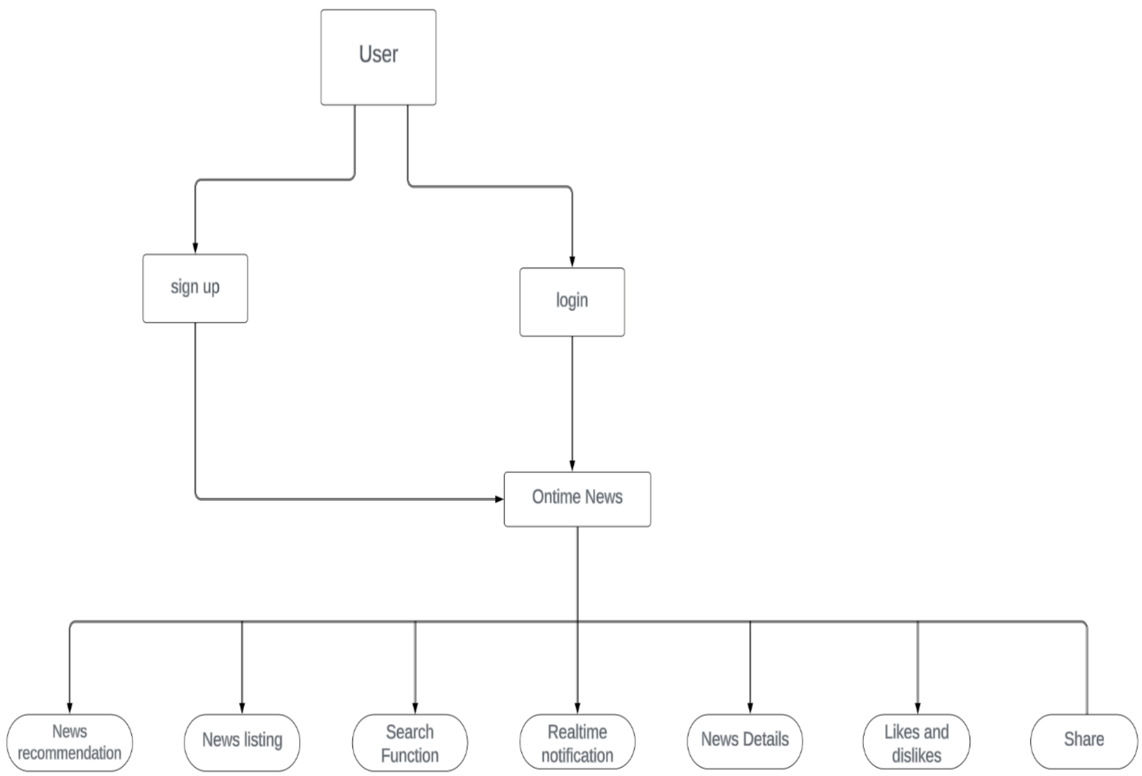

Fig 4.1: Overview of System

This architecture is designed to provide a comprehensive news platform that caters to both real-time and on-demand news consumption. It incorporates various functionalities such as user authentication, news search, news recommendation, and social media integration.

The architecture can be divided into the following components:

Frontend: This is the user interface of the news platform. It is responsible for displaying news articles, handling user interactions, and rendering news recommendations.

Backend: This component is responsible for processing user requests, managing data storage, and implementing the business logic of the news platform.

Database: This component stores all the news articles, user information, and other relevant data.

News API: This is an external service that provides real-time news updates. The news platform fetches news articles from this API and stores them in the database.

Recommendation Engine: This component analyzes user behavior and preferences to generate personalized news recommendations.

Social Media Integration: This component enables users to share news articles on their social media profiles.

User Authentication: This component manages user registration, login, and password recovery.

The architecture can be further divided into microservices for scalability and maintainability. Each microservice can be developed using a different programming language and framework, depending on the specific requirements of the functionality.

4.2 REST API (Representational State Transfer Application Programming Interface)

A REST API (Representational State Transfer Application Programming Interface) is an architectural style for designing networked applications. It specifies a set of constraints that must be adhered to, defining how data elements are presented to a client and how a client can request data from the server. The purpose of a REST API is to provide interoperability between computer systems on the internet. RESTful APIs are commonly used in web development, and they enable different systems to communicate with each other over the internet. They are known for their simplicity, scalability, and the ease with which they can be integrated into various programming languages and platforms. REST APIs are widely utilized for building web services, mobile applications, and other distributed systems. They have become standard for many web-based applications due to their flexibility and compatibility.

4.3 Data Collection

The data collection process for the "Personalized News Recommendations System" project involves the integration of a third-party news API, filtering and preprocessing of news articles using NLP techniques, and storage in a local MySQL database. This process ensures that the recommendation system can provide users with timely and relevant news recommendations tailored to their interests. Data maintenance and compliance with privacy regulations are integral parts of the data collection process, ensuring the system's ongoing functionality and user privacy protection.

4.3.1 API Integration

The project team integrates the third-party news API into the backend of the application. The API provides endpoints to retrieve news articles based on various parameters such as category, keyword, date, etc. Regular API calls are made to fetch news articles to maintain an updated dataset. Scheduled API requests ensure that the recommendation system always has access to the latest news content.

4.3.2 Data Filtering

The retrieved news articles may contain a wide range of information. To ensure relevance and quality. Articles are categorized into different topics (e.g., politics, sports, technology, entertainment), and only relevant categories are selected. Articles older than a specified threshold may be excluded to keep the data fresh. Depending on the target audience, articles in specific languages may be included or excluded.

4.3.3 Natural Language Processing (NLP)

After filtering, the news articles undergo NLP techniques to extract relevant features and information. Breaking down text into words or tokens. Eliminating common words (e.g., "the," "and") that do not contribute significantly to content understanding. Reducing words to their base form to improve text analysis. Identifying names, organizations, locations, etc.

4.4 Data Preprocessing

Data preprocessing is a crucial step in the "Personalized News Recommendations System" project. It involves cleaning, transforming, and structuring the raw news data to make it suitable for machine learning and recommendation algorithms. Techniques such as duplicate removal, missing data handling, text preprocessing, feature engineering, and data storage are applied to ensure the quality and relevance of the data. Preprocessed data is the foundation for accurate and personalized news recommendations, enhancing the overall user experience of the system. Data preprocessing is a critical step in the development of the "Personalized News Recommendations System." It involves cleaning, transforming, and structuring the raw news data obtained from the third-party API into a format suitable for machine learning and recommendation algorithms. This report outlines the data in the form preprocessing methods and techniques applied to prepare the data for analysis and recommendation generation. Data cleaning is the initial step in data preprocessing and focuses on identifying and rectifying issues in the raw data. In this project, data cleaning. Data cleaning is the initial step in data preprocessing and focuses on identifying and rectifying issues in the raw data. In this project, data cleaning.

Duplicate news articles may exist in the dataset due to multiple API calls or other reasons. These duplicates are identified and removed to avoid redundancy in the recommendation process. Some articles may have missing files or incomplete information. Missing data points are either imputed or the entire article is removed, depending on the extent of missing information. The primary content of news articles is textual data, and text preprocessing is essential for NLP tasks.

4.4.1 System Requirements

In this section, all the necessary details are described so that it is made easier to design the product and validate it according to requirements. Here, it is important to describe all inputs the software handle and all the outputs to better define interaction with other systems and facilitate integration. So, they have defined the hardware and software requirements required for successful working of this project.

A) Hardware Requirements

Hardware is the main part of this system. In this section, they will describe the hardware requirements for construction and deploying the system, including the personal computer, having the internet services.

Table 4.4.1 Hardware Requirements

Processor	Intel i3 (PC)
RAM	4GB minimum
Disc Space	25GB minimum

B) Software Requirements

In this section, they will describe the software requirements for constructing and deploying the system.

Table 4.4.2 Software Requirements

Operating System	Windows 8/9/10/11
Programming Language	Python
IDE	Python

4.5 Data Exploration

Data exploration is a crucial phase in the development of the "Personalized News Recommendations System." It involves analyzing and understanding the characteristics of the preprocessed news data before applying machine learning algorithms for recommendation. This report outlines the data exploration methods and findings to gain insights into the dataset. Basic statistical measures (mean, median, standard deviation) are calculated for numeric features like sentiment scores. Fig.4.5 shows a breakdown of the number of articles in each news category is examined to understand category distribution in the dataset. word frequency analysis is performed to identify the most common words and terms in the dataset. This can reveal topics that are frequently covered in the news articles. The modeling techniques are applied to discover latent topics within the news articles. This helps in understanding the thematic structure of the dataset. Bar charts or pie charts are created to visualize the distribution of news articles across different categories, providing a clear overview of the dataset's composition.

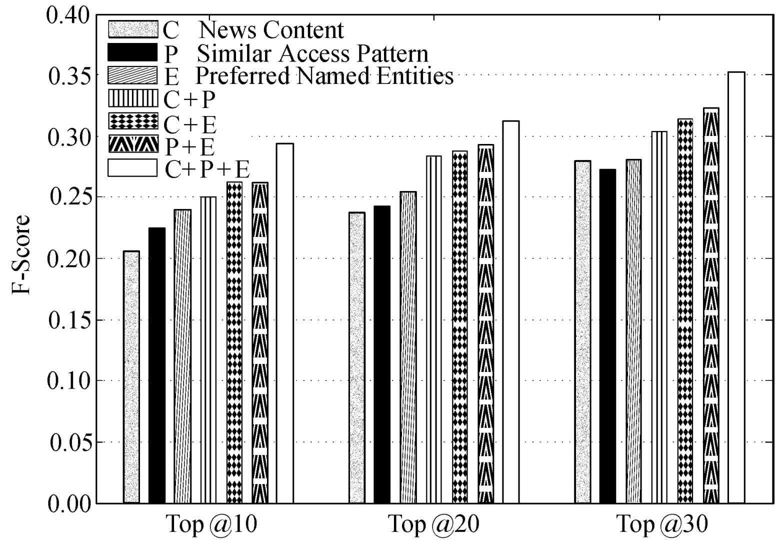

Fig 4.5 Frequencies of News Content

4.6 Data Labelling

Data labeling is a critical component in the development of the "Personalized News Recommendations System." The careful and systematic categorization of news articles into topics is fundamental to the accuracy and effectiveness of the recommendation system. The use of manual and automated labeling, clear guidelines, inter-annotator agreement, quality control, and ongoing maintenance ensured the creation of a high-quality labeled dataset, which serves as the foundation for training machine learning models for personalized news recommendations. The iterative nature of the labeling process and the feedback loop with annotators guarantee the continued success of the project. The primary objective of data labeling is to categorize news articles into specific topics or themes, such as politics, sports, technology, entertainment, and more. These labels will be used to create a structured and categorized dataset for training and improving machine learning models. Several methods were employed for data labeling, including both manual and automated approaches. team of human annotators was assigned to manually categorize news articles into predefined categories. Each article was assigned one or more category labels based on its content. A sample set of pre-existing labeled data was used to bootstrap the process. This dataset provided a foundation for building the labeling guidelines. A custom labeling interface was developed, allowing annotators to review and categorize articles efficiently. The interface enabled the selection of one or more categories from a predefined list. A quality control process was established to regularly review and audit the labeled data. Inconsistencies, errors, or mislabeling were identified and corrected. As the labeling process gained momentum, additional annotators were recruited to work on larger datasets. Automation techniques using pre-trained models were considered to augment the labeling process. Annotators continued to review and correct automated labels to ensure accuracy. Continuous communication and feedback with annotators were maintained.

4.6.1 Evaluation Metrices

There are many metrics to quantitatively evaluate the performance of news recommender systems.Most metrics aim to measure the recommendation performance in terms of the ranking relevance.For methods that regard the task of news recommendation as a classification problem, the Area Under Curve (AUC) score is a widely used metric, which is formulated as follows:

$$AUC = \frac{|\{(i\ ,\ j)|Rank(Pi) < Rank\ (nj))\}|}{NpNn}$$

Precision measures the proportion of recommended articles that are relevant to the user's preferences, while recall measures the proportion of relevant articles that are successfully recommended. High precision and recall values indicate that the system is providing accurate and comprehensive recommendations.

4.6.2 Learning Models

The machine learning model development process for the "Personalized News Recommendations System" is a critical component of the project. Collaborative filtering, content-based filtering, and hybrid models are explored and evaluated to provide accurate and personalized news recommendations. Model selection, validation, deployment, monitoring, and scalability considerations ensure that the system delivers high-quality recommendations, enhancing the user experience and engagement. The iterative nature of model development and the incorporation of user feedback guarantee the continuous improvement of the recommendation system.

4.6.3 Data preparation

The labeled news data is divided into training, validation, and test sets to facilitate model training and evaluation. Common splits might include 70% for training, 15% for validation, and 15% for testing.

The preprocessed news data, which includes vectorized text and category labels, is used for training the machine learning models. Text data has been transformed into numerical representations suitable for modeling.

4.6.4 Model Selection

Collaborative filtering techniques, such as user-based or item-based filtering, are considered for providing recommendations based on user behavior and preferences.

Content-based filtering models take into account the content of the news articles and user profiles to make recommendations. These models may involve natural language processing and feature engineering.

Hybrid models that combine collaborative filtering and content-based filtering are explored to provide more accurate and diverse recommendations.

4.6.5 Model Development

A collaborative filtering model is developed using techniques such as matrix factorization or singular value decomposition (SVD). It considers user-item interactions and generates recommendations based on user history and preferences. A content-based filtering model is built to analyze the content of news articles, extract relevant features, and match them with user profiles to provide recommendations. A hybrid model is created to combine the strengths of both collaborative and content-based filtering. This model aims to deliver recommendations that are both user-specific and content-relevant.

4.6.6 Training and Validation

The models are trained on the training data and fine-tuned using the validation set. Hyperparameter tuning is performed to optimize model performance. Model performance is evaluated using relevant metrics, such as: Accuracy: Measuring how often the recommended articles match the user's actual preferences.

4.6.7 Model Selection and Deployment

Based on evaluation results, the most effective model or ensemble of models is selected for deployment. The model is integrated into the system's architecture for real-time recommendations. Continuous monitoring of model performance is essential. The model should adapt to changes in user behavior and evolving news content. Re-training and updates are scheduled to maintain accuracy.

4.6.8 A/B Testing

A/B testing is employed to compare the performance of different models or model versions in a real-world scenario. User engagement and satisfaction are monitored to make informed decisions about model updates. Considerations for scaling the recommendation system as the user base and news dataset grow are addressed. The architecture should handle increased traffic and data

From fig. 4.6 There are several key components in this framework. First, news modeling is the backbone of news recommendation, and a core problem is how to understand the content and characteristics of news. In addition, user modeling is required to understand the personal interest of users in news, and it is critical to accurately infer user interest from user profiles like behaviors. Based on the news and user representations built by the news and user models, the next step is ranking candidate news according to certain policies such as the relevance between news and user interest. Then, it is important to train the recommendation model with proper objectives to make high-quality news recommendations, and evaluating the ranking results given by the recommendation model is also a core problem in the development of personalized recommender systems. Besides, the datasets and benchmarks for news recommendation are also necessities in designing personalized news recommendation models. Moreover, beyond developing accurate models, improving the responsibility of intelligent systems has been a spotlight problem in recent years. How to develop responsible news recommender systems is a less studied but extremely important problem in personalized news recommendation

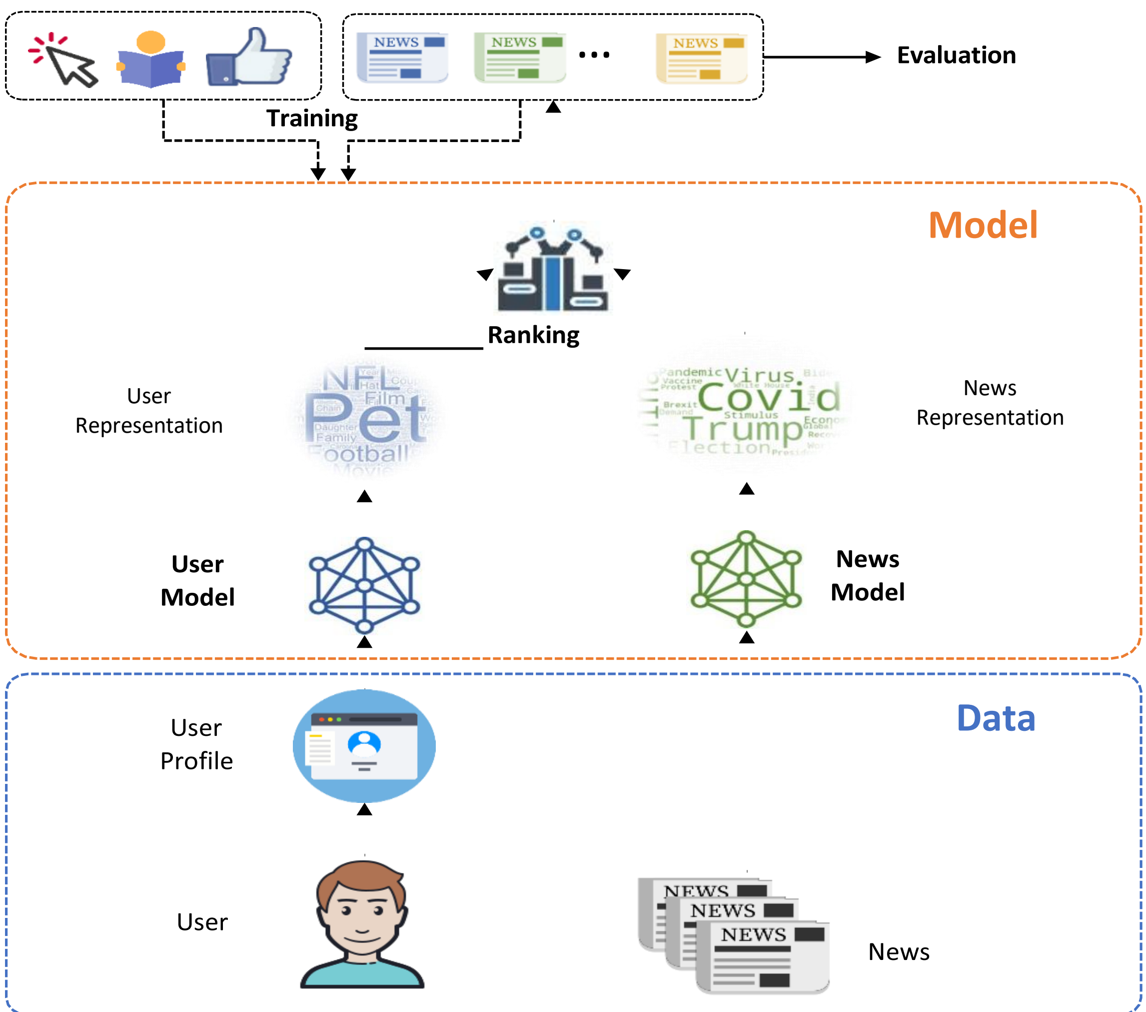

Fig 4.6: The Key Components in Developing Personalized News Recommendation Model

CHAPTER 5

TOOLS & TECHNOLOGIES

This section describes various technologies that are being used in the development ofthe proposed system. The functions and packages are explained along with features and components. Every module and its use in the proposed system has been highlighted.

5.1 Natural Language Process

Natural Language process is the intersection of applied science, Linguistics and Machine Learning that's attached the interaction between computers and humans in tongue. The fig 5.1 shows the venn diagram of NLP.

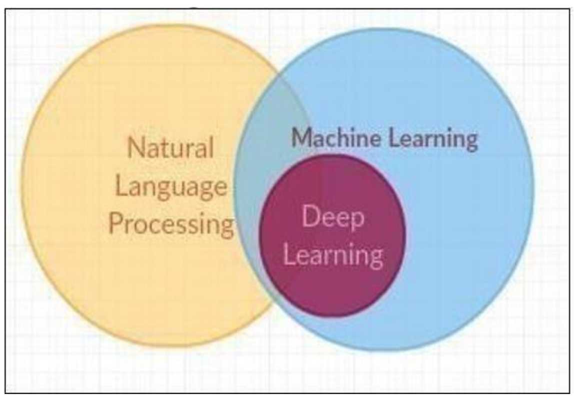

Fig 5.1 Venn diagram for NLP

NLP is far toward empowering PCs to grasp and deliver human idiom. Systems are used in separating of text, machine interpretation and Voice Agents like Alexa and Siri. Informatics is one amongst the fields that are profited from the advanced methodologies in Machine Adapting, significantly from Profound Learning methods. Regular idiom making ready methodology utilize the characteristic dialect tool cabinet for creating the principle organize in python tasks to figure with human dialect information. This is often easier to-use by giving the interfaces to a minimum of one than forty corpora and lexicon resources, for portrayal, for half passagessentences and to urge the words in its distinctive frame Marking, parsing, and gloss thinking for current reasoning quality to half up words, seeing the syntactical segments of these words, denoting the elemental subjects, doing this it serves to it machine by acknowledging the mostfactor to the substance.

5.2 Platform Used

Below are some platforms used for recommendation system

5.2.1 Windows 11

Windows ten is outlined because Microsoft works with the actual framework for PCs, tablets, and inserted gadgets etc. Microsoft discharged Windows ten is follow-up to Windows eight. It had been aforesaid on Gregorian calendar month that the window ten are invigorated rather than discharging it and framework as a successor. When the window ten is chosen or received it will be updated by inheritance squarely from window seven, eight or window ten. While not activity meddling and the framework design methodology. For maintenance shoppers run windows ten that helps in exchanging the applying on the past software package and setting to window 11. Shopper's pickup and fill or refresh window ten. With the assistance of window refresh partner window ten will be redesigned to physically begin associate degree overhaul for Windows. Windows ten is employed to focus on add capacities that through which IT offices allows to utilize mobile phones the board (MDM) programming to anchor and management gadgets helps in running operating framework. For given boarding programming as an example, Microsoft Framework Centre Arrangement Chief. Microsoft Windows ten is employed for varied validation advances, as an example, good cards and tokens. Further, Windows Hi has the biometric verification to Windows ten, wherever shoppers will sign on with a novel finger impression, or facial acknowledgment. The framework is employed to include virtualization-based security tools, as an example, Secluded Shopper Mode, Windows Safeguard Gizmo Watch and Windows shielded with Qualification monitor. Windows ten is a newer version for Bit Locker secret writing to substantiate databetween clients' gadgets, reposting instrumentation, messages and cloud administrations. Windows eight came up with the new plan and gave touch empowered motion driven UI like those on cell phones and tablets, however there wasn't abundant interpretation of well to customary workspace and digital computer PCs, significantly in business settings

In Windows ten, Microsoft ventured to deal with this issue and totally different behavior of Windows eight, as an example, associate degree absence of massive business neighborly highlights. The declaration of Windows ten in Gregorian calendar month 2014 from Microsoft was created and window business executive was made that point. There was the discharge from Microsoft to Windows ten by seeing the full population in Gregorian calendar month 2015. Then shoppers discovered that Windows ten is cordial than Windows eight because of it had been additional typical interface, that echoes the workspace partaking format of Windows seven. Windows ten consecrate Refresh that clad in August 2016, created some modifications to the assignment bar and start Menu. It to boot bestowed program augmentations in Edge and gave client's access to Cortana on the bolt screen. In Apr 2017, Microsoft discharged the Windows ten manufacturers Refresh that created Windows hello facial acknowledgment innovation faster and enabled shoppers to spare tabs in Microsoft Edge to examine later.

The Windows ten fall manufacturers refresh appeared in Gregorian calendar month2017, adding Windows Safeguard journey monitor to secure against zero-day assaults. There is fresh likewise enabled shoppers and IT to place applications running out of sight into vitality productive mode to safeguard battery life and enhance execution.

5.2.2 Front-end Development

HTML, CSS, Bootstrap

These fundamental web technologies form the core of the front-end development. HTML structures content, CSS styles it, and JavaScript adds interactivity and dynamic behavior.

Hypertext Markup Language (HTML)

Hypertext Markup Language is the core markup language used for structuring the content of a web page.

It defines the elements and their hierarchy on a web page, such as headings, paragraphs, lists, links, and images.HTML is responsible for creating the content structure and the basic layout of a web page HTML provides the structural foundation for web content. It uses a system of markup tags to define the structure and hierarchy of elements on a web page. Key features of HTML include:

Semantic Structure: HTML tags give meaning to content. Elements like `<header>`, `<nav>`, `<section>`, and `<footer` provide a clear structure to the page, making it more

accessible and improving search engine optimization (SEO). Content Presentation: HTML is responsible for presenting text, images, links, forms, and other content on a web page. Hypertext Links: HTML allows the creation of hyperlinks, enabling navigation between web pages and the dissemination of information.

Cascading Style Sheets (CSS)

CSS is used to control the presentation and styling of HTML elements. It defines how HTML elements should be displayed, specifying attributes like fonts, colors, margins, padding, and positioning It allows for the separation of content (HTML) from its appearance, making it easier to maintain and update the design of a website.

It defines attributes such as fonts, colors, margins, padding, and positioning to control the appearance of web elements.

It separates the content (HTML) from its presentation (CSS), making it easier to manage and update the design without altering the content.

It is used to create responsive web layouts that adapt to different screen sizes and devices, enhancing the user experience.

It supports animations and transitions for adding dynamic effects to web elements. It provides control over the layout of elements, including positioning them using techniques like flexbox and grid systems.

Bootstrap Framework

The Bootstrap framework is used for designing a responsive and visually appealing user interface. It simplifies the design process and ensures a consistent look and feel across various devices. Bootstrap is a popular front-end framework that builds upon HTML and CSS to streamline web development. Key features of Bootstrap include:

Responsive Design Bootstrap is built with responsive design in mind, making it easy to create web applications that look great on various devices, from desktops to mobile phones. Bootstrap provides a library of pre-designed components like navigation bars, buttons, modals, and forms, which can be easily customized to match the project's style.

Grid System Bootstrap includes a responsive grid system that simplifies layout design and column alignment, helping developers create flexible and consistent page structures.

Utility Classes: Bootstrap offers a set of utility classes that can be applied to HTML elements to control their appearance and behavior without writing custom CSS.

Cross-Browser Compatibility: Bootstrap is tested and optimized for cross-browser compatibility, ensuring a consistent look and functionality across various browsers.

5.2.3 Backend Development Platforms

Python: Python is used for backend programming to handle data processing, API development, and integration with machine learning models. Python is a versatile and popular programming language used in various aspects of web development, including server-side programming, scripting, data processing, and automation. In the context of web development with HTML, CSS, and Bootstrap. Python offers several web frameworks, such as Django, Flask, and Pyramid, which enable server-side programming. These frameworks facilitate the development of dynamic web applications by handling HTTP requests, database interactions, and rendering web pages. Python is used to process and manipulate data in web applications. Libraries like Pandas are commonly employed to handle data, perform data analysis, and prepare data for presentation on the frontend. a popular language for integrating machine learning into web applications. Frameworks like Flask can be used to create APIs for serving machine learning models. Libraries like scikit-learn, TensorFlow, and Py Torch are used for model development and integration.

Flask: Flask, a micro web framework for Python, is employed for building the REST API that connects the frontend and backend components. It provides a lightweight and flexible solution for API development. Flask is a micro web framework for Python that is widely used for web development. It's a lightweight and flexible framework that's often preferred for smaller web applications, APIs, and prototypes. It is known for its minimalistic and lightweight design. It provides the core components needed for web development, but developers can choose to add additional libraries and extensions as needed, giving them the flexibility to customize their projects. It allows developers to define URL routes, associating them with specific functions that should be executed when a user accesses those routes. This makes it easy to create a structured and organized web application. It Flask is commonly used for creating RESTful APIs. It simplifies the process of building APIs that communicate with frontend applications or external services. It's often used in combination with libraries like Flask-RESTful to streamline API development.

5.2.4 Data Storage Platforms

MySQL Database

MySQL serves as the database management system for storing structured data, including preprocessed news articles, user interaction data, and other relevant information. MySQL plays a pivotal role as a relational database management system (RDBMS) that is used to store, manage, and retrieve structured data for web applications. It serves as a robust and efficient storage solution for web applications, housing structured data such as user profiles, content, configuration settings, and much more.

It follows the relational database model, allowing developers to create tables with relationships between them. This model is particularly useful for modeling complex data structures and maintaining data integrity. It uses SQL (Structured Query Language) for querying, inserting, updating, and deleting data. SQL is the standard language for interacting with relational databases**.**

5.2.5 Machine Learning Model Platforms

Python Machine Learning Libraries

Various Python libraries like scikit-learn, TensorFlow, or Porch are used for developing and deploying machine learning models. These libraries are essential for creating content-based and collaborative filtering models. Python Machine Learning Libraries used are:

Pandas-

Imagine Python as a toolbox full of tools that help you do different tasks. One of the tools in this toolbox is called "Pandas." Pandas is like a magical set of tools specifically designed to work with tables of information, kind of like how you use crayons to color pictures in a coloring book.

NLTK-

One of the tools in this box is called "NLTK," which stands for "Natural Language Toolkit." NLTK is like a special helper that makes it easier to work with words and sentences, just like how you use a magnifying glass to look closely at things.

NumPy-

Imagine you have a magical box called "NumPy" that helps you do math with numbers in Python. NumPy is like a super-powered calculator for your computer.

5.3 Modules Used

5.3.1 Pandas

Pandas is a Python library that is used for faster data analysis, data cleaning, and data preprocessing. Pandas are built on top of the numerical library of Python, called NumPy. Before it installs pandas, make sure it has NumPy installed in its system. If NumPy is not very familiar to it, then it needs to have a look at this article. Brush up it NumPy skills and then learn pandas.

Data Representation

Pandas provide extremely streamlined forms of data representation. This helps to analyze and understand data better. Simpler data representation facilitates better results for data science projects.

Less Writing and More Work Done

It is one of the best advantages of Pandas. What would have taken multiple lines in Python without any support libraries, can simply be achieved through 1-2 lines with the use of Pandas. Thus, using Pandas helps to shorten the procedure of handling data. With the time saved, it can focus more on data analysis algorithms.

Extensive Set of Features

Pandas are powerful. It provides it with a huge set of important commands and features which are used to easily analyze its data. can use Pandas to perform various tasks like filtering it data according to certain conditions

Wes McKinney, the creator of Pandas, made the Python library to mainly handle large datasets efficiently. Pandas help to save a lot of time by importing large amounts of data very fast.

Makes Data Flexible and Customizable

Pandas provide a huge feature set to apply on the data it has so that it can customize, edit and pivot it according to its own will and desire.

Made For Python

Python programming has become one of the most sought-after programming languages, with its extensive number of features and the sheer amount of productivity it provides.

5.3.2 NumPy

NumPy is a general-purpose array-processing package. It provides a high- performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. Many of its methods are mirrored by functions in the outermost NumPy namespace, allowing the programmer to code in whichever paradigm they prefer. This flexibility has allowed the NumPy array dialect and NumPy n-d array class to become the de-facto language of multi-dimensional data interchange used in python. NumPy can be used to perform a wide variety of mathematical operations on array. It adds powerful data structures to python that guarantee efficient calculations with arrays and matrices, and it supplies an enormous library of high- level mathematical functions that operate on these arrays and matrices.

It is a python library that provides a multidimensional array object, various derived objects and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, input & output, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more. It is a table of elements all of the same type, indexed by a tuple of non-negative integers.

Use Of NumPy

In Python have lists that serve the purpose of arrays, but it slows to process. NumPy aims to provide an array object that is up to 50x faster than traditional Python lists. The array object in NumPy is called Nd array, it provides a lot of supporting functions that make working with array very easy. Arrays are very frequently used in data science, where speed and resources are very important.

5.3.3 Scikit-learn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistency interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib. It is a simple and efficient tool for data mining and data analysis. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k- means etc.

This library works by splitting the dataset into two pieces: a training set and a testing set,

it also trains the model on training set and lastly test the model on the testing set and evaluate how well Its model did. As a high-level library, it defines a predictive data model in just a few lines of code, and then uses that model to fit your data. It is versatile and integrates well with other Python libraries. The main algorithms which are used in scikit-learn library are linear regression algorithm It can help your computer learn from examples. For instance, if you want your computer to recognize whether a picture
has a cat or a dog, scikit-learner can teach your computer by showing it lots of pictures and telling it which one's cats and which ones are dogs. Once your computer has learned. It can use that knowledge to make predictions. For example, if you show your computer a new picture, it can tell you whether it thinks there's a cat or a dog in it.

Scikit-learn can also help your computer sort things into categories. Imagine you have a bunch of different fruits, and you want your computer to group them by type (apples, bananas, oranges). Scikit-learn can do that for you.

t's like having a detective in your computer. Scikit-learn can find patterns in data, like figuring out what makes one thing different from another. For instance, it can help you understand what makes a fast car different from a slow one.

If you need your computer to make decisions, like whether to play outside based on the weather, scikit-learn can help by looking at past weather data and making a suggestion

5.3.4 Rapid Automatic Keyword Extraction (RAKE)

Stands for "Rapid Automatic Keyword Extraction." It's like a special tool that helps people find the most important words or phrases in a bunch of text, like a long article or a book. Imagine you have a big pile of books, and you want to quickly figure out what each book is mainly about without reading all of them. First, RAKE takes the text you give it (like a chapter from a book) and breaks it into smaller parts, like sentences or phrases.

Then, it looks at how often words appear in those smaller parts. Words that show up a lot are considered important.

RAKE looks for groups of words that appear together frequently. These groups are like clues to what the text is talking about. For example, if it sees "space exploration" or "rocket launch" mentioned a lot, it knows the text is probably about space.

RAKE ranks these important words and phrases by how often they appear and how they are connected to each other. The ones at the top are the most important ones.

Finally, RAKE gives you a list of these important words and phrases. It's like a summary of the main ideas in the text, so you can quickly understand what it's all about. It is designed to identify and extract important keywords or key phrases from a given text, such as a document, a webpage, or any other textual content. RAKE is particularly useful for text summarization, information retrieval, and content indexing.

It can be used to extract keywords and key phrases from news articles. These keywords can help in categorizing and tagging articles, making them more discoverable and searchable by identifying the most relevant keywords and key phrases in an article, RAKE can assist in generating content summaries or abstracts. These summaries can be used for quick previews, improving user engagement, and aiding in content recommendations. RAKE-extracted keywords can contribute to topic modeling. By identifying keywords in articles, the system can categorize articles into relevant topics or themes, enhancing the recommendation process.

5.3.5 Natural Language Toolkit (NLTK)

One of the tools in this box is called "NLTK" which stands for "Natural Language Toolkit." NLTK is like a special helper that makes it easier to work with words and sentences, just like how you use a magnifying glass to look closely at things. It can help you read and understand words. It knows about lots of different words and can tell you what they mean. It's like having a friendly dictionary on your computer. It can help you find specific words or count how many times a word appears. It's like the word detective! It can also ask NLTK to help you put words together to make sentences. It's like having a word puzzle solver. It can teach you interesting things about how languages work, like how sentences are structured and how words can change their form. It can use NLTK to make the computer talk to you or understand what you're saying. It's like having a conversation with your computer. NLTK is like a super-smart friend in your Python toolbox that helps you with everything related to words and sentences, just like how a magnifying glass helps you see things up close and better. The Natural Language Toolkit (NLTK) is a Python library for working with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources, such as WordNet. NLTK also includes a range of text processing libraries for tasks like tokenization, stemming, parsing, and more. It's a valuable tool for researchers and developers working on natural language processing and text analysis.

However, please note that my knowledge is based on information available up to September 2021, and there may have been developments or changes in the NLTK library since that time. If you're interested in the current state of NLTK, I recommend checking the project's official website or documentation for the most up-to-date information.

It's important to note that while NLTK is a useful library for text processing and analysis, building a comprehensive news recommendation system involves more than just NLTK. You would typically integrate NLTK with other libraries, data sources, and machine learning models to create a robust recommendation system.

Additionally, as of my last update in September 2021, NLTK is widely used in NLP tasks, but there are other powerful libraries like spacey, Genism, and machine learning frameworks like scikit-learn and TensorFlow that can be integrated into a news recommendation system for more advanced processing and recommendation capabilities.

Text Preprocessing NLTK provides a wide range of text preprocessing tools, such as tokenization, stemming, and lemmatization. These tools can be used to clean and normalize text data, making it more suitable for analysis.

Topic Modeling NLTK can be used in combination with other libraries like Genism or scikit-learn to perform topic modeling on news articles. Topic modeling can help identify the main themes or topics within news articles, which can be useful for content categorization and recommendation.

Sentiment Analysis NLTK offers sentiment analysis tools that can be used to determine the sentiment (positive, negative, neutral) expressed in news articles. Sentiment analysis can help personalize recommendations by considering a user's preferences for positive or negative news.

Named Entity Recognition (NER): NLTK includes NER capabilities that can extract entities like names of people, organizations, and locations from news articles. This information can be used to enhance recommendations, such as suggesting news articles related to specific entities.

Text Classification: NLTK can be used to build text classifiers. In the context of a news recommendation system, this could involve classifying news articles into categories or topics. This classification can help in recommending articles that match a user's interests.

5.4 Linear regression

Linear regression is a statistical technique commonly used in personalized news recommendation systems to predict user preferences based on various factors such as reading history, user demographics, and content features. By analyzing these factors and fitting a linear model, the system can estimate the relevance or interest level of news articles for each user. This helps in providing personalized recommendations that are more likely to be of interest to the user, improving the overall user experience.

a personalized news recommendation system, linear regression can be used to predict user ratings for news articles based on various features. These features can include user behavior, article content, and contextual information. The goal is to minimize the mean squared error between the actual user ratings and the predicted ratings. This approach is part of the broader field of recommendation algorithms, which aims to provide personalized suggestions by analyzing user behavior and interests.

The use of linear regression in personalized news recommendation systems is supported by research in the field of recommendation algorithms, where various machine learning techniques, including regression, are employed to provide personalized recommendations based on user behavior and interests

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. In the context of news recommendation systems, the dependent variable could represent user engagement metrics such as clicks, likes, or dwell time, while independent variables may include features like article topic, author, publication time, and user demographics.

CHAPTER 6

IMPLEMENTATION

Sentiment analysis steps are deeply intrinsic, comprising many different machine learning and NLP tasks and sub tasks. In this chapter, provide the implementation of sentiment analysis it represents a description of machine learning algorithm that is undertaken. It consists of various operations like data collection, data preprocessing, visualization of sentiment and accuracy.

The training of dataset consists of the following steps:

6.1 Data Collections

The data collection process for the "Personalized News Recommendations System" project involves the integration of a third-party news API, filtering and preprocessing of news articles using NLP techniques, and storage in a local MySQL database. This process ensures that the recommendation system can provide users with timely and relevant news recommendations tailored to their interests. Data maintenance and compliance with privacy regulations are integral parts of the data collection process, ensuring the system's ongoing functionality and user privacy protection. The project team integrates the third-party news API into the backend of the application. The API provides endpoints to retrieve news articles based on various parameters such as category, keyword, date, etc. Regular API calls are made to fetch news articles to maintain an updated dataset. Scheduled API requests ensure that the recommendation system always has access to the latest news content. After filtering, the news articles undergo NLP techniques to extract relevant features and information. NLP processes include Breaking down text into words or tokens. Eliminating common words (e.g., "the," "and") that do not contribute significantly to content understanding. Reducing words to their base form to improve text analysis. Identifying names, organizations, locations, etc., in the articles. Determining the sentiment or tone of the articles (positive, negative, neutral). This process ensures that the recommendation system can provide users with timely and relevant news recommendations tailored to their interests. Data maintenance and compliance with privacy regulations are integral parts of the data collection process, ensuring the system's ongoing functionality and user privacy protection.

6.2 Data Preprocessing

Data preprocessing is a crucial step in the "Personalized News Recommendations System" project. It involves cleaning, transforming, and structuring the raw news data to make it suitable for machine learning and recommendation algorithms. Techniques such as duplicate removal, missing data handling, text preprocessing, feature engineering, and data storage are applied to ensure the quality and relevance of the data. Preprocessed data is the foundation for accurate and personalized news recommendations, enhancing the overall user experience of the system. Data preprocessing is a critical step in the development of the "Personalized News Recommendations System." It involves cleaning, transforming, and structuring the raw news data obtained from the third-party API into a format suitable for machine learning and recommendation algorithms. This report outlines the data preprocessing methods and techniques applied to prepare the data for analysis and recommendation generation. Data cleaning is the initial step in data preprocessing and focuses on identifying and rectifying issues in the raw data. In this project, data cleaning. Duplicate news articles may exist in the dataset due to multiple API calls or other reasons. These duplicates are identified and removed to avoid redundancy in the recommendation process. Some articles may have missing files or incomplete information. Missing data points are either imputed or the entire article is removed, depending on the extent of missing information. The primary content of news articles is textual data, and text preprocessing is essential for NLP tasks. The following text preprocessing techniques are applied. Text is split into individual words or tokens to facilitate further analysis. Tokenization is crucial for feature extraction and NLP tasks. Common words (stop words) that do not carry significant meaning are removed from the text. This step reduces noise in the data and improves analysis accuracy. Words are reduced to their base forms (lemmatization) or trimmed to their stems (stemming). This simplifies text analysis and ensures consistency in word representation. Named entities such as names, organizations, locations, etc., are identified and tagged in the text. It aids in identifying key information in articles. Sentiment analysis is performed to determine the sentiment or tone of each article (positive, negative, neutral). This information can be used to enhance recommendations. Data cleaning is the initial step in data preprocessing and focuses on identifying and rectifying issues in the raw data. In this project, data cleaning. Duplicate news articles may exist in the dataset due to multiple API calls or other reasons. These duplicates are identified and removed to avoid redundancy in the recommendation process.

6.3 Personalized Ranking

Many personalized news recommendation methods employ machine learning models for news modeling, user modeling and interest matching. How to train these models to make accurate recommendations is a critical problem. A few methods train their models by predicting the ratings on news given by users. For example, the system is trained by predicting unknown ratings in the user-news matrix. However, explicit feedback such as ratings is usually sparse on news platforms. Thus, most existing methods use implicit feedback like clicks to construct prediction targets for model training. For example, formulated the news click prediction problem as a binary classification task, and use cross entropy as the loss function for model training. It proposed to employ negative sampling techniques that combine each positive sample with several negative samples to construct labeled samples for model training. However, click feedback usually contains heavy noise and may not indicate user interest, which poses great challenges to learning accurate recommendation models.

From fig.6.3 An Example of User Modeling there are only a few methods that consider user feedback beyond click. For example, proposed to model click preference with click feedback and model reading satisfaction based on the personalized reading speed of users, and train the recommendation model to predict both clicks and user satisfaction. By optimizing objectives beyond news clicks, these methods make user engagement information and thereby can better understand user interest. In addition, these methods have the potential to recommend news articles that are not only clicked by users, but also indeed satisfy their information needs. Thus, designing engagement-aware training objectives is useful for news recommender systems to provide high-quality news suggestions.

Thus, most existing methods use implicit feedback like clicks to construct prediction targets for model training. For example, formulated the news click prediction problem as a binary classification task, and used cross entropy as the loss function for model training.. It proposed to employ negative sampling techniques that combine each positive sample with several negative samples to construct labeled samples for model training.

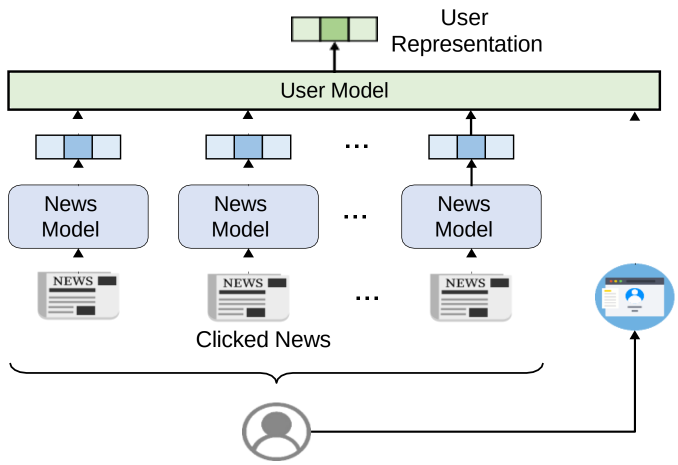

Fig 6.3: An Example of User Modeling

6.4 Responsible News Recommendation

Most endeavors on personalized news recommendation focus on improving the accuracy of recommendation results. In recent years, research on the responsibility of machine intelligent systems has gained high attention to help AI techniques better serve humans and avoid their risky and even harmful behaviors that can lead to negative societal impacts and unethical consequences. There are many aspects to improve the responsibility of personalized news recommender systems. For example, since many news recommendation methods are learned on private user data, it is important to protect user privacy in recommendation model training and online serving. Federated learning is a privacy-aware machine learning paradigm, which can empower the construction of privacy-preserving news recommender systems. Besides optimizing news recommendation accuracy, it is also important to promote the diversity of news recommendation results, which can satisfy users' needs on information variety and alleviate the filter bubble problem. Moreover, fairness is a critical aspect of responsible news recommendation, since the recommendation models learned on biased user data may inherit unwanted biases, which may lead to the prejudice of algorithms and further unfair recommendation results. To mitigate the unfairness issue of news recommendation methods, fairness-aware machine learning techniques can help build inclusive and fair algorithms to provide high-quality news recommendation services to different groups of users. These research fields on responsible news recommendation emerging in recent years have the potential to improve the quality

6.5 Electronic Feeds

The personalized news recommendation system (PNR) is a web application that requires the user to log in. The user logs in to the application and is required to select the news category of personal choice. Hence, after selecting the category, all live news was fetched from different electronic news portals through RSS feeds relating to this category. Users can manually input search keywords to fetch his/her news choices.

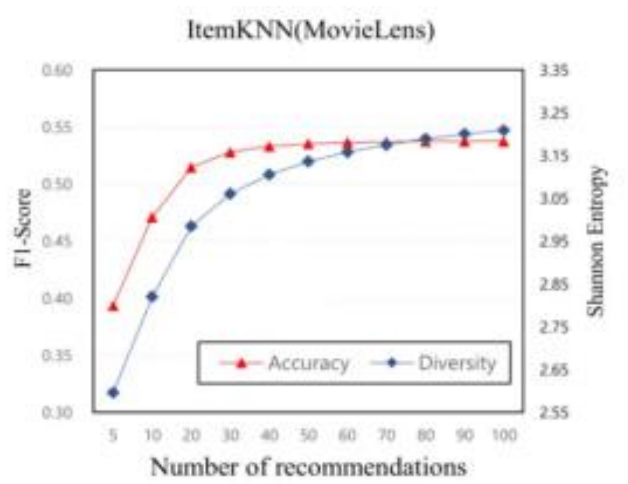

Fig: 1 Recommendation Working

To resolve existing problems, we propose the following: a web application that collects, parses, processes, annotates, and analyzes news from various RSS-fed channels which expresses either positive or negative information via crawling. The personalized news recommendation system (PNR) is a web application that requires the user to log in. The user logs in to the application and is required to select the news category of personal choice.

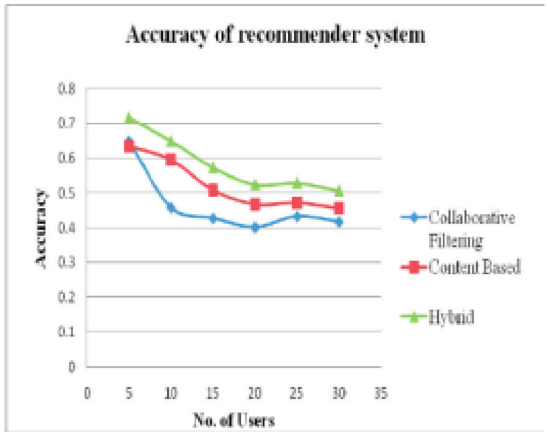

Fig: 2 Accuracy for Recommender

Hence, after selecting the category, all live news was fetched from different electronic news portals through RSS feeds relating to this category. Users can manually input search keywords to fetch his/her news choices. To resolve existing problems, we propose the following: a web application that collects, parses, processes, annotates, and analyzes news from various RSS-fed channels which expresses either positive or negative information via crawling.

6.6 Methodology

A. User Profiling

User profiling in personalized news recommendation systems involves the process of gathering and analyzing user data to understand their preferences, interests, and behavior. This methodology is crucial for delivering tailored news content to individual users based on their unique characteristics. There are several key steps involved in user profiling for personalized news recommendations. First, data collection is essential. This involves gathering various types of user data, such as browsing history, search queries, reading habits, and social media interactions. This data can be collected through user interactions with the news platform, cookies, user registrations, and other sources. Next, data preprocessing is performed to clean and organize the collected data. This may involve removing irrelevant information, handling missing data, and transforming the data into a suitable format for analysis. After preprocessing, the data is analyzed using techniques such as machine learning, natural language processing, and collaborative filtering to identify patterns and preferences. This analysis helps in creating user profiles based on factors like topics of interest, reading frequency, preferred sources, and engagement levels. Furthermore, the system utilizes these user profiles to generate personalized news recommendations. By matching user profiles with available news content, the system can deliver relevant and engaging articles to each user based on their specific interests and preferences. It's important to note that user profiling in personalized news recommendation systems must adhere to privacy regulations and ethical considerations. Users' consent and data protection should be prioritized throughout the entire process.

B. Content Analysis

Content analysis Used for personalized news recommendation systems involves the systematic examination and categorization of news articles, texts, and multimedia content to understand their attributes and relevance to individual users. This methodology is essential to effectively match content with user preferences and to provide personalized news recommendations. The Content analysis process through System in a personalized news recommendation system typically includes several key steps. First, the system collects a wide range of news articles, videos, and images from different sources. This content is then pre-processed to extract relevant features, such as keywords, topics, sentiments, and metadata. Natural language processing techniques can be used to analyze textual content, while image and video analysis methods can be used to understand visual content. Subsequently, the content was categorized and tagged based on different attributes such as topic, tone, relevance, and format. This categorization helps to organize the content and makes it easier to match user preferences. For example, news articles can be categorized into topics such as politics, technology, sports, and entertainment. Subsequently, machine learning and data mining algorithms are utilized to analyze the categorized content and identify patterns. These patterns help in understanding user preferences and behavior, which in turn informs the content recommendation process. In addition, the system incorporates user feedback and interaction data to refine the content analysis process continuously. By taking into account how users interact with recommended content, the system can adapt its analysis to better match individual preferences. Finally, the insights gained from content analysis can be used to generate personalized news recommendations for users. By exploiting the attributes and relevance of the analyzed content, the system can deliver tailored news experiences that align with each user's interests. It is important to note that ethical considerations should be taken into account when analyzing content in personalized news recommendation systems. Privacy, transparency, and fairness in content categorization and recommendation are essential for maintaining user trust and satisfaction. In conclusion, content analysis in personalized news recommendation systems involves the systematic examination, categorization, and utilization of news content attributes to deliver tailored recommendations. By understanding the nuances of news content and aligning it with user preferences, the system can improve the relevance and engagement of personalized news experiences for individual users.

C. User-Content

In a personalized news recommendation system, a user content matching methodology involves aligning individual user preferences and characteristics with the relevant news content to deliver tailored recommendations. This methodology aims to ensure that the recommended content satisfies the specific interests and needs of each user. The process generally includes several key steps. First, the system gathers and analyses user data to build user profiles that include information such as reading habits, topics of interest, preferred sources, and interaction patterns. These data are then used to understand the unique preferences and behavior of each user. Subsequently, the system categorizes, and tags news content based on various attributes, such as topic, tone, relevance, and format. This categorization helps to organize the content and makes it easier to match user preferences. For example, news articles can be linked to topics such as politics, Science, Social Activities technology, sports, and entertainment. Subsequently, machine learning algorithms and data mining techniques were used to identify patterns and correlations between user profiles and categorized content. This analysis helps us to understand which types of content are most likely to resonate with specific user profiles. Real-time user interaction data and feedback are incorporated into the system to continuously refine the content-matching process. By considering how users interact with the recommended content, the system can adapt its matching methodology to better align with individual preferences. By the end, the insights gained from user content matching can be used to generate personalized news recommendations for users. By aligning the attributes and relevance of the analyzed content with the profile of each user, the system can provide tailored news experiences that are more likely to be relevant to individual users. It is important to note that ethical considerations should be taken into account when matching user content in personalized news recommendation systems. Privacy, transparency, and fairness in matching methodologies are essential to maintain user trust and satisfaction. In conclusion, user content matching methodology in personalized news recommendation systems involves aligning individual user preferences with relevant news content to deliver tailored recommendations. By understanding user-profiles and matching them with categorized content, the system can improve the relevance and engagement of personalized news experiences for individual users.

D. Feedback Loop

In the realm of personalized news recommendation systems, the feedback loop methodology involves continuous improvement in the algorithm of content recommendations based on user interactions and feedback. The whole purpose of this methodology is to enhance the accuracy and relevance of news recommendations for an extended period through the integration of user responses and preferences. The feedback loop methodology typically consists of several key steps. The system initially provides interesting news personalized news recommendations that align with user profiles and preferences. This is where the magic begins. Next, the users engage with the recommended content by clicking, reading, sharing, or even explicitly expressing their thoughts on articles. This is where the reading adventure takes off. Following user engagement, the system starts gathering and analyzing user feedback to gain insights into their interactions with the recommended content. This can include tracking various metrics, such as click-through rates, time spent on articles, sharing behavior, and explicit feedback signals. The data speaks volumes. Now, it's time to leverage machine learning algorithms and data analysis techniques to interpret user feedback and uncover patterns. This analysis helps us grasp which content types hold the greatest resonance with specific user profiles and preferences. It's like connecting the dots. In addition, the system incorporates valuable user feedback insights into the recommendation process and adjusts future recommendations accordingly. For instance, if a user consistently interacts positively with articles on a particular topic, the system will give high priority to recommending similar content in the future. It's all about personalization. The beauty of this methodology lies in its continuous nature. The system continuously adapts and learns from user behavior over time, leading to ever-improving accuracy and relevance in the recommendations. It's an ever-evolving journey. It's imperative to address privacy and transparency when implementing the feedback loop methodology in personalized news recommendation systems. Users must have control over their data and be fully informed of how their feedback contributes to the refinement of recommendations. Trust and transparency are key, the feedback loop methodology initialization in personalized news recommendation systems revolves around refining content recommendations based on user interactions and feedback. By continuously learning from user responses, the system enhances the relevance and engagement of personalized news experiences for each user. It's all about making the news truly personal.

E. Evaluation

The evaluation methodology for personalized news recommendation systems plays a crucial role in assessing the effectiveness and performance of recommendation algorithms in delivering relevant and captivating content to users. Typically, the evaluation process involves several key steps to measure the accuracy of the system, user satisfaction, and overall performance. Initially, the system gathers data on user interactions, including clicks, reading time, sharing behavior, and explicit feedback, regarding the recommended news content. This data will then be used to evaluate the ability of the system to deliver content based on user preferences and interests. Subsequently, various evaluation metrics were used to assess the performance of the recommendation algorithm. Common metrics include precision, recall, click-through rate, and user engagement. Precision measures the proportion of relevant items recommended to the user, whereas recall measures the proportion of recommended relevant items. The click-through rate indicates the percentage of users who click on a recommended item, providing insights into the effectiveness of capturing user interest through recommendations. In addition, qualitative insights into the relevance and quality of the recommended content can be obtained through user feedback surveys and user satisfaction ratings. This feedback is valuable for understanding user perception and satisfaction with personalized news recommendations. In addition, A/B testing and experimentation were conducted to compare the performance of different recommendation algorithms or features. The system can assess the impact of changes on user engagement and satisfaction by testing variations of recommendation systems with a subset of users. In addition, user retention and long-term engagement metrics are important to evaluate the impact of personalized news recommendations on user loyalty and continued usage. It should be noted that continuous evaluation and monitoring are essential for adapting to changing user preferences and behaviors. The evaluation methodology must be dynamic and iterative to ensure continuous improvement of the recommendation system. In conclusion, the evaluation methodology in personalized news recommendation systems involves assessing the system's accuracy, user satisfaction, and overall performance through metrics, user feedback, A/B testing, and long-term engagement analysis. By continuously evaluating and refining the recommendation algorithms, the system can enhance the relevance and effectiveness of personalized news.

CHAPTER 7

RESULTS AND DISCUSSION

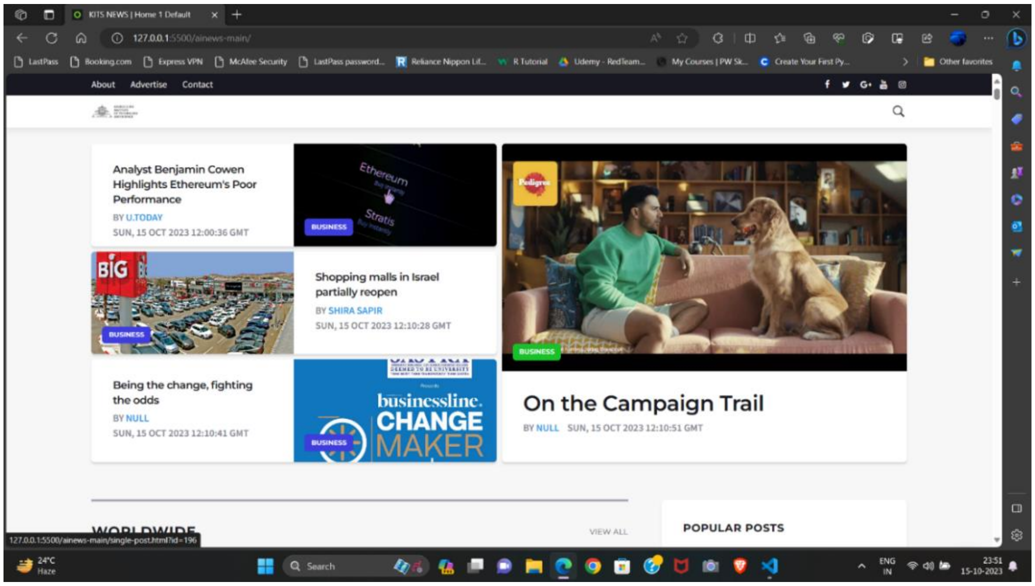

Fig 7.1 Home Page

The home page of It News Recommendation System serves as the gateway to a personalized and informed news consumption experience.

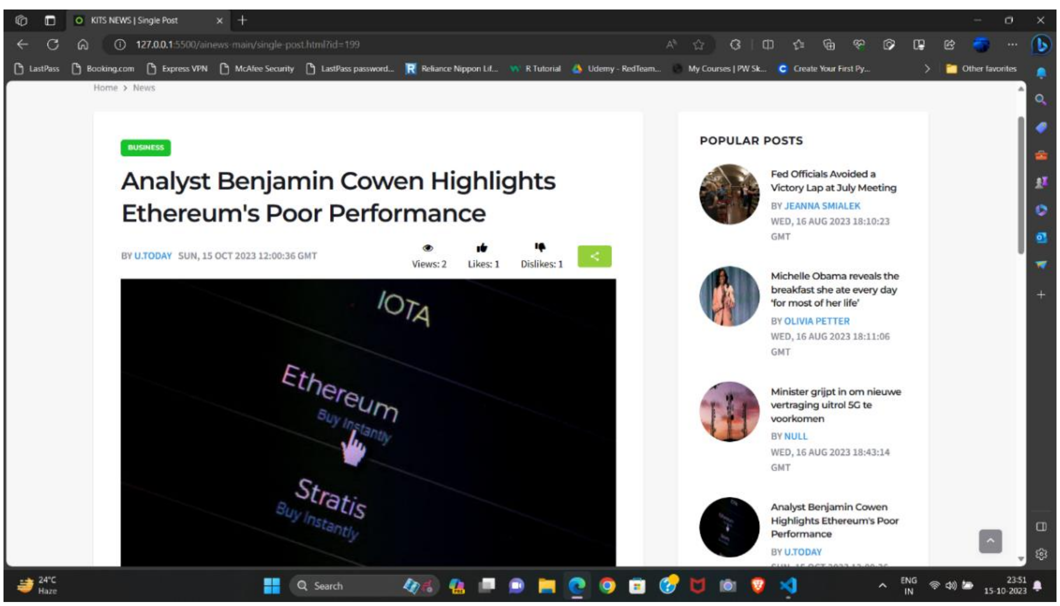

Fig 7.2 News Recommended

The presented Figure illustrates the user interface of our sophisticated News.

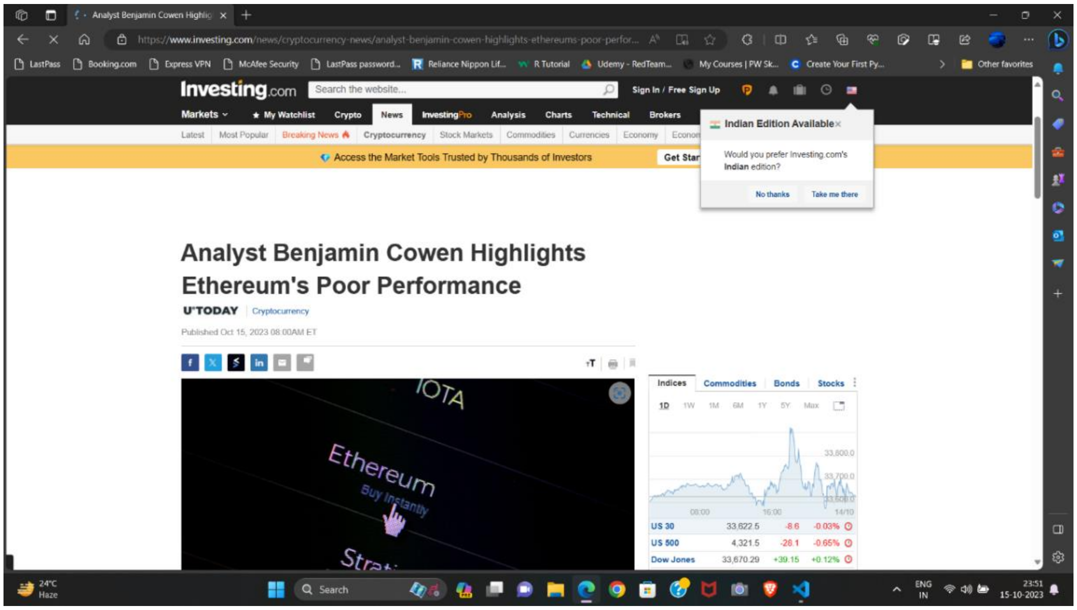

Fig 7.3 Redirected Page

Mention the title of the redirected page. This could be the headline of the recommended news article. Provide details about the news source. Include the name of the website or platform from which the news is being recommended.

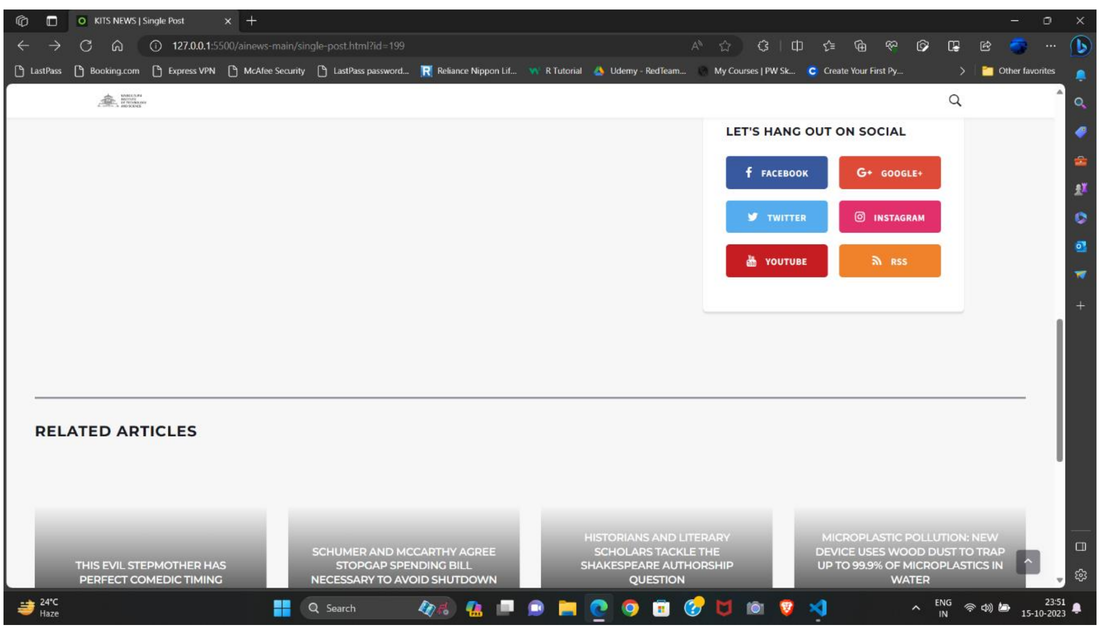

Fig 7.4 Social Handling and Recommended Articles

The social media and Recommended Article Page within It News Recommendation System serves as a dynamic and user-centric interface.

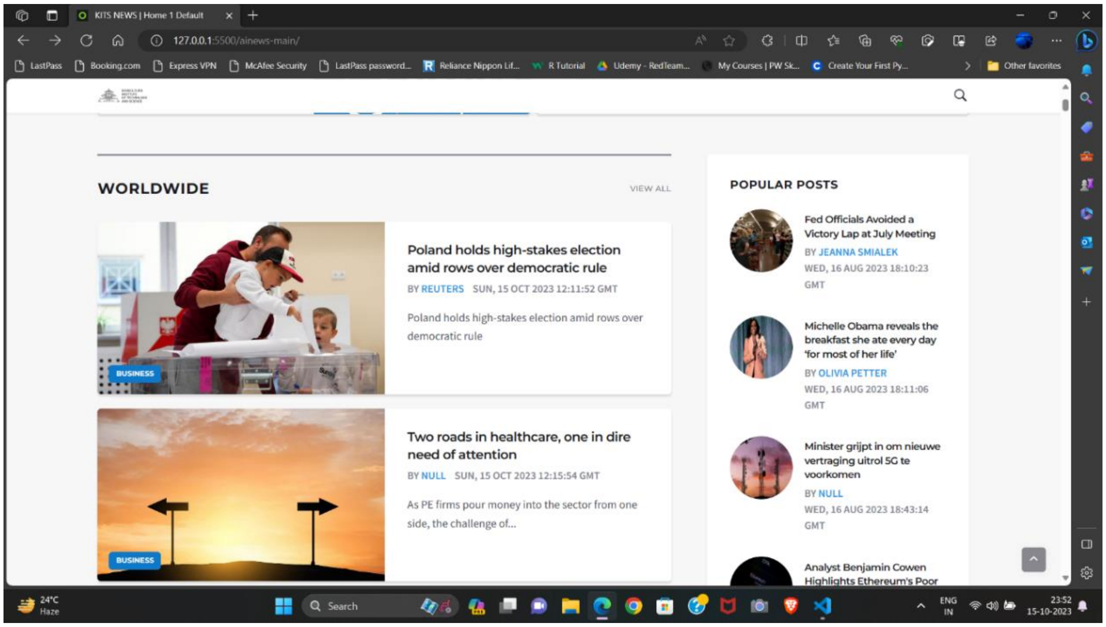

Fig 7.5 Worldwide News Section

The Worldwide News section serves as the global pulse of It News Recommendation System, encapsulating a diverse range of news stories from around the globe.

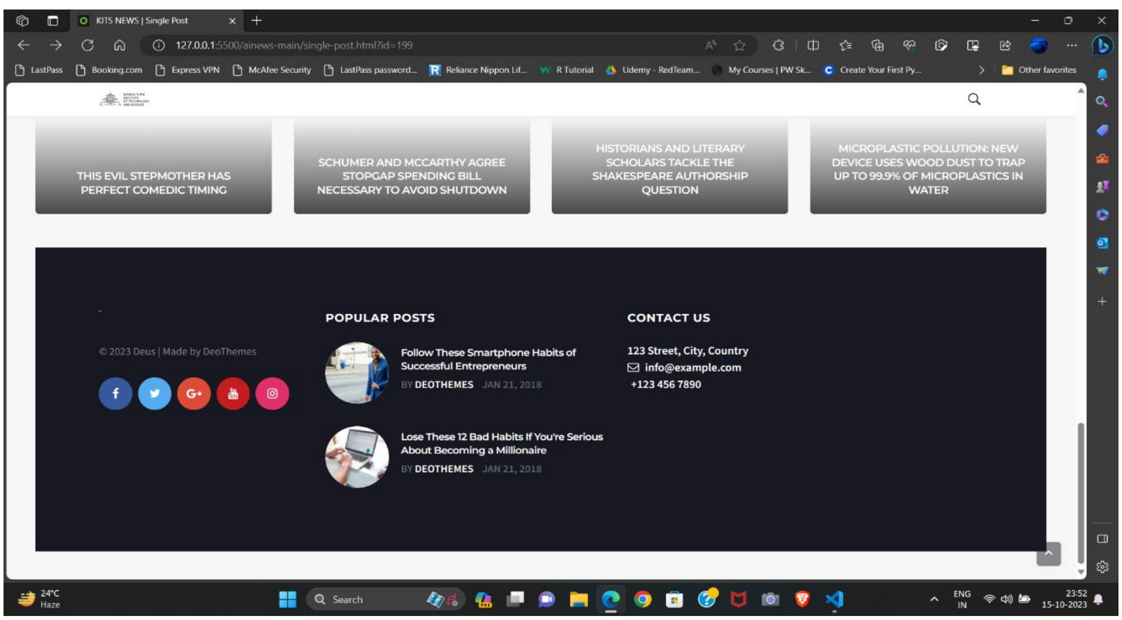

Fig 7.6 Contact Us

The "Contact Us" section serves as a vital component within the It News Recommendation System, providing users with a direct and efficient channel for communication.

CHAPTER 8

CONCLUSION

It brings together a diverse set of tools and technologies to create a sophisticated platform for news recommendation. By combining data collection, preprocessing, analysis, and machine learning, the system aims to provide users with personalized and relevant news content. The integration of web development technologies for the frontend and backend, along with advanced data analysis techniques, positions the project to deliver a seamless and tailored news consumption experience to users. The utilization of NLTK and other Python libraries enhances the system's capabilities in text analysis and content recommendation. Overall, this project represents a significant step towards enhancing the way users consume and engage with news content in the digital age.

8.1 Limitations of study

The quality and reliability of news data obtained from third-party APIs may vary. Inaccurate or biased news articles could affect the recommendations, potentially leading to issues with the system's accuracy and fairness. The system may inadvertently introduce bias in recommendations based on the data it uses or the algorithms it employs.

The project relies on external news APIs for data. Any interruptions or changes in these APIs' availability could impact the system's data collection and recommendation processes. Depending on the sources integrated, the system's recommendations may be limited in terms of content diversity, potentially favoring certain publishers or topics.

New users or users with limited interaction history may receive less accurate recommendations due to the "cold start" problem, where the system lacks sufficient data to personalize recommendations effectively. Ensuring fairness in recommendations is a complex and ongoing challenge. If user feedback is limited, it may hinder the system's ability to adapt and improve over time. As the user base grows, scalability challenges may arise, especially in terms of data storage and processing.

8.2 Future Scope of Work

Some of future scopes that can be included in personalized news recommendation system such as collaborative filtering, deep learning models, or reinforcement learning, to improve the accuracy of recommendations and adapt to evolving user preferences. Expand the system to support multimedia content, including images and videos. Incorporating these media types can provide a more comprehensive news consumption experience. Work on integrating real-time news updates from sources and incorporate mechanisms to notify users of breaking news or important updates. Extend the system to support multiple languages, broadening its user base and catering to diverse linguistic preferences. Develop algorithms or mechanisms to assess and filter content based on quality and reliability to ensure that recommendations are based on credible sources. Implement a feedback loop for users to rate and provide feedback on recommended articles, allowing the system to continually improve its recommendations. Create user segments or clusters based on behavior and preferences to provide more targeted recommendations for specific user groups. Continuously monitor and address potential bias in recommendations to ensure fairness and ethical use of AI in content curation. As technology progresses, the field of recommendation algorithms continues to evolve. Future research could focus on developing and implementing advanced algorithms, such as deep learning models, reinforcement learning, or hybrid approaches. The aim is to enhance the accuracy and efficiency of personalized news recommendations by leveraging cutting-edge machine learning techniques. The future of personalized news recommendation systems involves a more seamless integration of user feedback. This could include exploring innovative ways to capture explicit feedback from users, such as ratings or comments. This may include websites, mobile applications, and social media platforms. Ensuring a consistent and personalized experience across various platforms presents both technical and user experience challenges. This includes dynamic personalization strategies that adapt in real-time to users' changing preferences and behaviors. Machine learning models can continuously update user profiles based on real-time interactions, ensuring that recommendations stay relevant and timely.

References

1. Lichun Sun, et al. (2021) "Personalized News Recommendation Based on Multi-channel Attention Network". In Proc. Transactions on Knowledge and Data Engineering, 17(6), 734-749.
2. Francesco Gelli, et al. (2020) "Enhancing news recommendation through dynamic user profiling". In Proc. of the 10th international conference on the World Wide Web (WWW '01), 285-295.
3. Quoc-Tuan Truong, et al. (2019) "Time-aware personalized news recommendation with long- and short-term user modeling". In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 230-237.
4. Ruining He, Wang-Cheng Kang, et al. (2019)"Neural News Recommendation with Multi-Head Self-Attention". In Proc. of the third ACM international conference on Web search and data mining, 81-90
5. Xin Wang, Wenwu Ou, et al. (2018)"Context-aware news recommendation using multi-armed bandit framework". In Proc. Introduction to recommender systems handbook. In Recommender Systems Handbook (pp. 1-35). Springer.
6. Lina Yao, Aixin Sun, et al. (2017) "A Survey of Recommendation System Research". Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37.
7. Paul Covington, Jay Adams, and Emre Sargin (2016) "Deep Neural Networks for YouTube Recommendations". In Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 627-636.
8. Liu J, Dolan P, Pedersen E R. (2010) “Personalized News Recommen- Dation Based on Click Behavior”. In Proc. the 14th International Conference on Intelligent User Interfaces, Hong Kong, China, Feb. 7-10.31-40.
9. Ricardo Baeza-Yates and Berkant Barla Cambazoglu (2009) "Personalized Web Search: Revisiting the State of the Art". Recommended system application developments: A survey. Decision Support Systems, 54(1), 233-243
10. Koren, Y., Bell, R., & Volinsky, C. (2009) "Matrix Factorization Techniques for Recommender Systems". Novelty and diversity in recommender systems. ACM Computing Surveys (CSUR), 47(1), 1-41.

11. Yifan Hu, Yehuda Koren, and Chris Volinsky (2008) "Collaborative Filtering for Implicit Feedback Datasets". Effects of transparency on trust in and acceptance of a content-based art recommender. User Modeling and User-Adapted Interaction, 19(4), 367-399.
12. Kim H R, Chan P K. (2008), “Learning Implicit User Interest HierarChy for Context in Personalization”. Applied Intelligence ,28(2): 153-166.
13. Burke R. Hybrid, (2005) “Systems for Personalized Recommendations”. In Proc. Workshop on Intelligent Techniques for Web Personalization, Acapulco, Mexico,.133-152.
14. Billsus D, Pazzani M J. (2005) “User Modeling for Adaptive News Access”. User Modeling and User-Adapted Interaction
15. Carreira R, Crato J M, Gon¸calves D, Jorge J A. (2004) “Evaluating Adaptive User Profiles for News Classification.” In Proc. the 9th International Conference on Intelligent User Interfaces, Funchal, Brtngal 206-212.
16. Blei D M, Ng A Y, Jordan MI (2003). “Latent Dirichlet Allocation”. The Journal of Machine Learning Research, 3: 993- 1022.
17. Liang T P, Lai H (2002). “Discovering User Interests from Web Browsing Behavior”: An application to internet news services. In Proc. HICSS, Hawaii, USA, Jan. 7-10,
18. Tan A H, Teo C, (2002). “Learning User Profiles for Personalized In- Formation Dissemination”. In Proc. IEEE International Joint Conference on Computational Intelligence, Horolulu, USA, May 12-17,183-188.
19. Jurafsky D, Martin J H, Kehler A, (2000) Vander Linden K, Ward N. “Speech and Language Processing”. Prentice Hall
20. Hofmann T. (1999) “Probabilistic Latent Semantic Indexing”. In Proc. the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berke- ley, USA, Aug. 15-19, 50-57.