National College of Ireland

An Advanced Personalized Tweet Recommendation and Friend Suggestion System Using ChatGPT-3.5 Large Language Model, K-means Clustering, and Dynamic User Profiling

MSc Research Project
MSc. Data Analytics

Abhinandan Nahar
Student ID: X22202871

School of Computing
National College of Ireland

Supervisor: Prof. David Hamil

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Abhinandan Anil Nahar
Student ID:	X22202871
Programme:	MSc. Data Analytics **Year:** 2023-2024
Module:	MSc Research Project
Supervisor:	Prof. David Hamil
Submission Due Date:	12-08-2024
Project Title:	An Advanced Personalized Tweet Recommendation and Friend Suggestion System Using ChatGPT-3.5Large Language Model, K-means Clustering, and Dynamic User Profiling
Word Count:	6496 **Page Count** 22.

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Abhinandan Nahar

Date: 12-08-2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

AI Acknowledgement Supplement

[Insert Module Name]

[Insert Title of your assignment]

Your Name/Student Number	Course	Date
Abhinandan Nahar	MSc. Data Analytics	12-08-2024

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool

Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

[Insert Tool Name]	
[Insert Description of use]	
[Insert Sample prompt]	[Insert Sample response]

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Additional Evidence:

[Place evidence here]

Additional Evidence:

[Place evidence here]

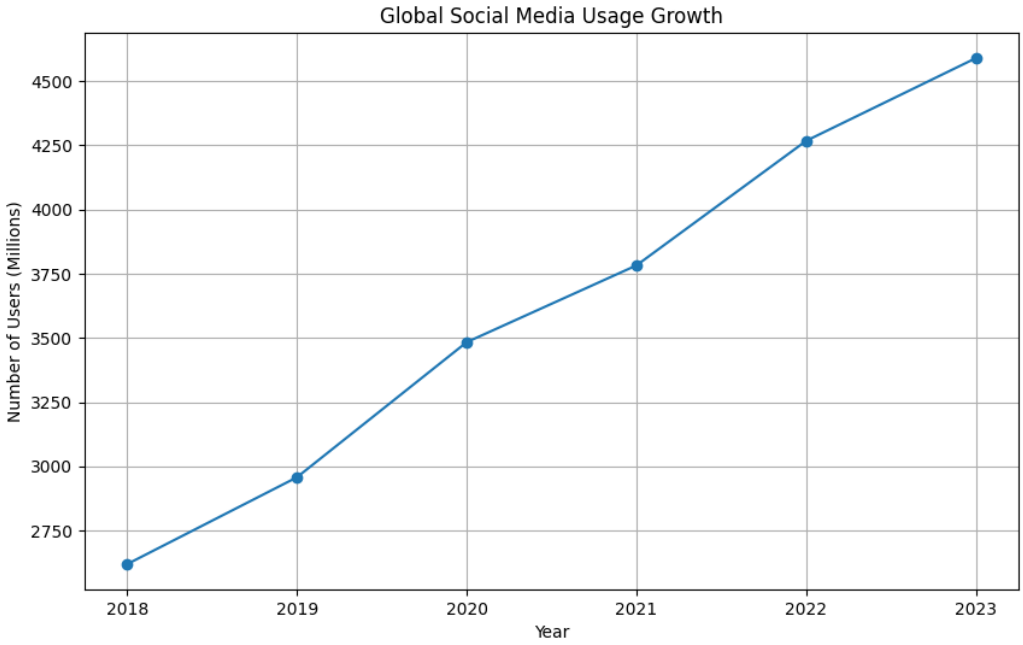
An Advanced Personalized Tweet Recommendation and Friend Suggestion System Using ChatGPT-3.5 Large Language Model, K-means Clustering, and Dynamic User Profiling

Abhinandan Anil Nahar

X22202871

Abstract

As has been exhibited from this research paper, there is an imperative need to improve the tweet recommendations and friend suggestions in social media sites such as the ‘X’. The system uses some of the following advanced technologies such GPT-3. 5 for interest scoring, K-means clustering for organizing the content and cosine similarity for friend suggestions. It also uses dynamic user interest profiling which captures and evolves as the amount of interestchanges and a new algorithm for recommendations. The paper also goes deep into the system specifications involving the technology architecture and the major algorithms. Performance findings reveal that the system successfully recommends the correct suggestions and friends. Thus, the existing challenges that are relevant to real-time processing of big data and the ethical issues remain unsolved, the study offers basic insights for further development of the recommendation systems of social media.



New developments in recent years in the fields of machine learning, natural language processing and artificial intelligence have made way for more effectiveness and efficiency of these recommendation systems (Pappalardo et al., 2024). Specifically, the use of 'large' generation models such as GPT-3. 5, for content analysis, holds interesting prospects to enhance the portrayal of tweets' context and specifics, which in turn may lead to more accurate interest identification and further suggestions.

1 Introduction

1.1 Background and Motivation

Social media today is a key aspect in the communication process as nearly all people share content in the social networks on a daily basis amounting to billions. However, out of all the said platforms, one that deserves its special attention is 'X', which offers real-time and brief content. Nevertheless, the total number of tweets produced results in difficulty for users to locate interesting and relevant content (Reuter et al., 2021). This information overload will make the user lose interest and stop using the mobile application.

Such systems have been developed as the solution to the mentioned problem, which can help to sort and recommend content that would be interesting to a specific user (P. S. et al., 2019). These systems consider the examinee's activity, the content attributes, and the social relations

to suppose what content could be of most interest to the examinee. Such systems can thereby complement the functions of the microblog heavily when it comes to the additional Tweets that would favor a specific user in the sphere of his or her interests as well as in the relation to the establishment of the connections.

New developments in recent years in the fields of machine learning, natural language processing and artificial intelligence have made way for more effectiveness and efficiency of these recommendation systems (Pappalardo et al., 2024). Specifically, the use of 'large' generation models such as GPT-3. 5, for content analysis, holds interesting prospects to enhance the portrayal of tweets' context and specifics, which in turn may lead to more accurate interest identification and further suggestions.

1.2 Problem Statement

Despite the progress in recommendation systems, several challenges persist in the context of social media platforms like 'X':

1. Preserving and recreating users' interest in a context where changes are likely to occur frequently.
2. In this way, the application should offer users fully personalized experience and at the same time avoid the formation of echo chambers, in which people will only see suggestions that coincide only with their own opinion.
3. The ability to analyze the incredible quantities of real-time data that are produced on the platform and on this basis timeously generate recommendations.
4. Coming up with algorithms of friends' suggestions that are not only based on the user's network but also give priority to the chances of friendship based on common activities and interests.

These challenges indicate the importance of the development of a new shallow model that incorporate the best of NLP, dynamic user profiling as well as a well-developed clustering algorithm to enhance the recommendation model that is tailored for social network sites.

1.3 Research Objectives

To address the identified problems, this research aims to achieve the following objectives:

Based on the given problem formulation, create a new approach to the recommendation of tweets using GPT-3.5 to content analysis and interest scoring for better results of interest mapping to the intending consumer.

1. Develop a dynamic user interest profiling system that is invoked periodically to allow for constant update in terms of user interest indicating the recommendations' timeliness.
2. Modify the system so that the friendship proposal works beyond a network and incorporates interest in each other like a combination of Askera and OkCupid.
3. Use and test an innovative K-means clustering method for the categorization of tweets to enhance the recommendation system's effectiveness and efficiency.
4. Evaluate the communication of the proposed system metrics on overall captures concerning users, correctness, and system results against the current baseline.

Thus, the following objectives are formulated for this research: When realizing these objectives, this work will yield the following possible outcomes which are expected to enhance the development of recommendation systems in social media applications such as ‘X’:

2 Related Work

2.1 Recommendation systems for social media

It can be said that recommendation systems within social medial platforms have become more complex in the recent past, in a bid to solve the problem of information overload as well as to increase the level of user interaction. Such systems tend to use various approaches to interpret the user’s behavior characteristics of the content; and relationships with other people to recommend appropriate contents (P. S. et al., 2019).

The main difficulty when approaching recommendation for social media for instance, is the volatility and rapidity of user interest. There is the scalable event-based clustering method proposed by Reuter et al. (2021) that intends to cluster related content in social media efficiently. This method highlights the need for real-time processing capacity when operating on big volumes of data shared on social networks.

Combining the approaches based on machine learning and deep learning has been a great improvement in social media recommendations. Aduwamai et al. (2024) provided a systematic literature review on applying these models for detecting depression on social media and exploring the direction for recommendation system not only enhancing the users’ experience but also enhancing their well-being.

Recent studies also examined the effects that recommenders with the help of AI have on people’s actions. Pappalardo et al. (2024) reviewed and discussed numerous methodologies and results and called for future research that includes the ethical perspectives and the possible ‘Application of AI Systems’. This highlights the need to set recommendations with unique and diverse features that are responsive to the user’s needs.

2.2 Interest-Based User Profiling

Interest-based user profiling is very important to have efficient recommendation systems in social networks. It is a response to the constant construction and evolution of the representations of users and their interests based on the activities and information available to them as well as their connections. This approach allows for better filtering and identification of the right information and recommendations to the users thus increasing the level of satisfaction.

L. R. D. and Pervin N. (2019) put forward an integrated method to generate the large-scale recommendation by including the prior social trust, social bias, and geo-spatial cluster. The article of them presents the problem of constructing users’ profiles from features which are not directly connected with observed content interaction. Thus, the integration of social as well as geographical factors proved the proficiency of researchers over generic recommendation systems with enhanced relevance.

Another challenge that arises in the process of creating user profiles is that the interests of users are time dependent. To overcome this problem, Parasuraman D. and Elumalai S. (2021) proposed a two-tier filtering approach; the first tier is mainly the collaborative filtering approach, while the second tier uses the content-based filtering approach. Their method is dynamic and evolves over time based on the users' changing preferences, resulting in better item recommendations. Based on these findings, a case for more dynamic user profiling methods is pointed out due to volatility of expressed interests on social media platforms.

Introduction of new techniques in AI and machine learning opens possibilities of better and more detailed profiling of the users. Based on the IoT and machine learning, Salina A. et al. (2022) introduced an architecture for recommender systems using social media content.

Their approach builds upon the huge amount of data produced by connected devices to enrich user profiles; this shows how the proposed ideas on integration of different types of data can be applied in the context of interest-based profiling.

Concerning the ethical aspect, the issue of detailed user profiling has also been discussed in the latest publications. Another survey on experience with AI-based recommenders and behavioral effects was carried out by Pappalardo et al. (2024) who underlined the importance of proper profiling that would be privacy-preserving and would not infringe upon users' agency. Thus, this research acknowledges the strategic role of personalization in user profiling techniques based on the concept of interests and the necessity of considering the ethical implications of such techniques as well.

2.3 Clustering in Social Media Analysis

Clustering methods are very important in social media analysis since it helps to group similar contents, users, or behaviors to enhance the recommendation systems, and organization of the content. They are effective especially when it comes to dealing with massive information flow that is characteristic of the social media.

Reuter et al. (2021) have developed a large-scale event-based clustering approach for social media data using record linkage methods. Their method solves the problem of how to cluster related contents in real-time, which is critical for contextual and timely recommendations as seen with 'X'. From this research, using more advanced clustering methodologies, there is significant possibility of enhancing the efficiency and the functionality of theisms of the analysis of systems on social media.

K-means clustering is still widely used and effective in the analysis of social media platforms. Multi et al. (2021) explored the application of text mining and more specifically, K-means clustering on the Twit collected data. Their work is diverse in terms of the exemplification of the utility of K-means in sorting the content of social media, which can improve the CBRS greatly.

Nigro L. and Fränti P. (2023) presented two medoid-based algorithms in the context of clustering sets which could be treating the sets as objects or considering the sets' contents. Although these algorithms have not been directly used in the context of social media in their work, they offer potential ideas for better clustering social media data especially in situations in which using the centroid-based method may not be effective.

As the clustering methods combined with other approaches in machine learning have demonstrated the future trends in the analysis of the social media. Seth A. and Zhang J. (2021)

presented a section on ‘‘personalized participatory media content recommendation using social network.’’ The paper reveals that their approach integrates clustering with network analysis to enhance content recommendation because clustering corresponds well with other kinds of network analysis, meaning that different clustering methods are best utilized when integrated with others.

Some of the current trends in social media are the ever-changing data type and user interactions thus clustering techniques must be modified to ensure they can accommodate the changes. The advancement of this study area in the future may include the creation of other durable and flexible clustering algorithms that support the processing of multimodal data (text, images, videos) in real-time to improve the efficiency and relevance of social media recommendation systems.

2.4 NLP in Tweet Analysis

NLP has now proved to be a crucial asset in the analysis of large volumes of textual data that are common on social media, especially ‘X’. The aspects of the text that are volume limitedto 280 characters, lack strict grammar and punctuation, and contain often overused, colloquial hashtags make it quite challenging and at the same time quite interesting to process with NLP tools and as a result these aspects have led to enhanced innovation in NLP particularly for the social media content.

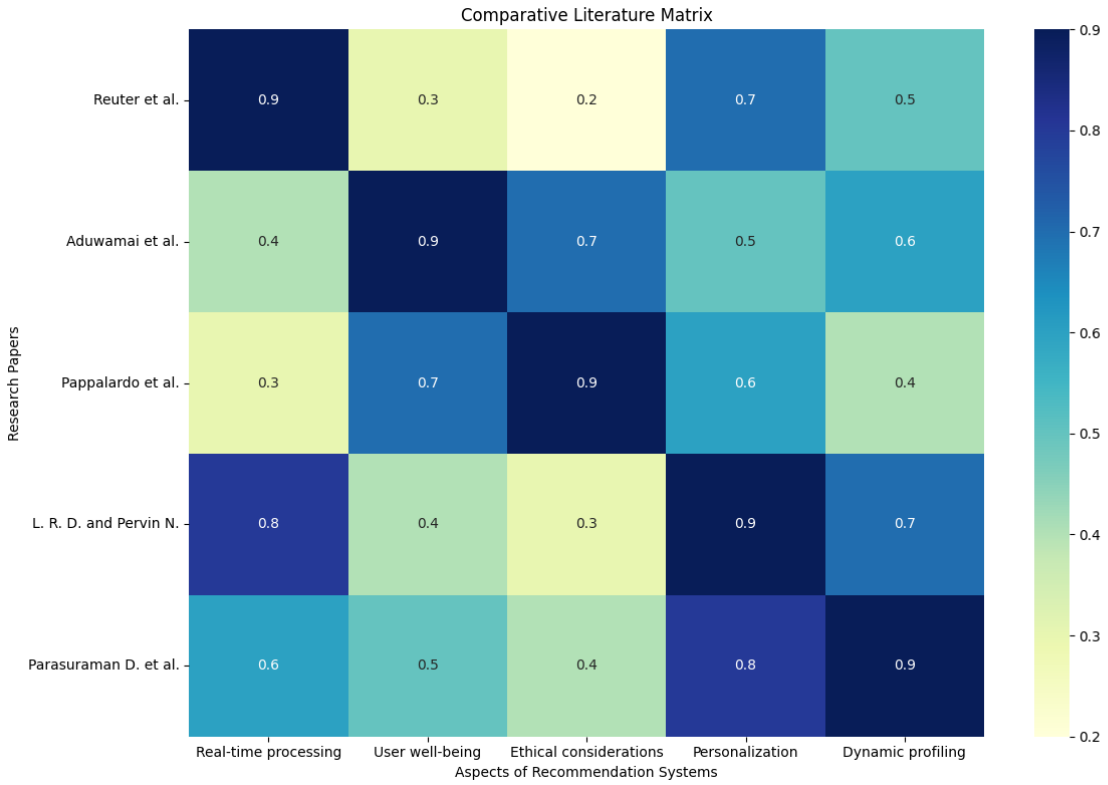
Hence, Claudiu et al. (2022) recently proposed significant work done in case of sentiment analysis of Romanian content on ‘X’ and the significance of using methods for NLP languages in this case. Their work presents the practical application of the necessity of developing NLP models that can work with multilingual and any-regardless-of-dialect social media textual data, which is essential for worldwide platforms such as ‘X’.

Subsequent development in terms of language models has also been added to the stream analysis of tweets. Kumar (2024) discussed the opportunity of using conversational AI in decision-making process with focus on ChatGPT. This research signifies the potential of large language models in assessing the context and subjectivity of the tweets’ content, which might further enhance the facets of content classification and intent and interest identification of the users.

Subsequently, KOSAKA et al. (2024) focused on the joint learning and adaptation of the acoustic and language models for EQ-ASR in a style of affective tweets. Even though they were mainly concerned with speech, the ideas of how to adjust language models to the peculiarities of the social media content are applicable to the text-based tweets analysis, so there might be some ideas for the further integrating of the MHNLP to apply permanently to the SM.

In addition to text comprehension, the application of NLP in the ideology of the tweet proves to be more useful. Alipour and Esmaeilpour (2024) have worked on cryptocurrency return prediction using rule-based and machine learning tactics to analyze tweet sentiments. Their study supports the view of using more sophisticated NLP approaches for a more sophisticated content analysis across different fields such as finance and market analysis.

2.5 Comparative Literature Matrix



Paper	Year	Methodology	Relevance to Our Study
Reuter et al.	2021	Record linkage techniques	Highlights need for real-time processing of large data volumes
Aduwamai et al.	2024	Literature review	Explores AI applications for user well-being on social media
Pappalardo et al.	2024	Survey and analysis	Emphasizes need for diverse and responsive recommendations
L. R. D. and Pervin N.	2019	Integrated approach	Demonstrates proficiency over generic recommendation systems
Parasuraman D. et al.	2021	Collaborative and content-based filtering	Highlights importance of dynamic user profiling methods
Salina A. et al.	2022	IoT and machine learning integration	Shows potential for integrating diverse data types in profiling
Multi et al.	2021	K-means clustering	Demonstrates utility of K-means in sorting social media content
Nigro L. and Fränti P.	2023	Medoid-based clustering	Offers potential ideas for clustering social media data
Seth A. and Zhang J.	2021	Clustering and network analysis	Shows benefits of integrating clustering with other analysis methods

Claudiu et al.	2022	Sentiment analysis, NLP	Emphasizes need for language-specific NLP models
Kumar	2024	Conversational AI	Highlights potential of large language models in content assessment
KOSAKA et al.	2024	Speech recognition, NLP	Suggests ideas for adapting language models to social media content
Alipour and Esmaeilpour	2024	Sentiment analysis, ML	Demonstrates advanced NLP use for specialized content analysis

3 Research Methodology

3.1 Approach

Our approach starts with the system architecture that consists of several interconnected parts that are intended to recommend the most compelling tweets and friends' suggestions. They are data gatherer and data pre-processor, user interest identification module, tweet scorer module, recommendation module, cluster module, and the friend suggestion module. These components

collectively help in the handling of ‘X’ data, users’ behaviors analysis and generation of recommendations.

For data collection and pre-processing purposes, the basic data source or dataset that we employed for our study was downloaded from Kaggle which is an ‘X’ dataset. It is convenient to use this dataset to create a realistic scenario resembling a real ‘X’ system inthe analysis of this article. The steps in the pre-processing include data cleaning, tokenization, lower casing, removal of stop words and stemming/lemmatization. It is important to have the tweet data in a proper form for the subsequent analyses; the mentioned steps contribute to it.

Another function that is very important in our system is the User interest profiling. To indicate a broader coverage, a set of general interest categories that can be associated with typical ‘X’ing topics was defined and was used; this included technology, sports, politics, entertainments, food, health, finance, science, travelling, and education. These forms of interest categories form the foundation of the scoring system we have in place. For getting accurate interest scores of every created tweet, we used GPT-3. 5 model. This process involves designing a prompt that contains the text of the given tweet along with a request for scores for each of the interest categories followed by the sending of the prompt to the GPT-3. 5 API, converting the obtained JSON string into an array of scores for the corresponding categories and scaling the scores to the entire interval [0, 1].

The tweet recommendation algorithm is an equation that calculates the degree of relevance of a tweet to a user’s interests’ measurement. To obtain an overall relevancy value for a certain tweet, we calculate the product of the interest category coefficients and the corresponding value of the user’s interest profile for each coefficient and then sum up the products. Tweets are thereafter sorted in a descending order depending on the relevance scores obtained. As for the situation, personal user accounts’ information requires periodical updating, so, there is the concept of dynamic update. When the tweet is Liked or Retweeted by a user, their interest scores are adjusted with a positive coefficient while when they Unlike that specific content, the interest scores are adjusted with a negative coefficient.

The friend suggestion algorithm measures the extent to which two users are similar according to the interests they have. In other words, each user is represented by a vector in the space of interest categories and the cosine similarity is calculated between the users’ vectors. Friend’s suggestions are derived by computing the distance of the other users with a particular user, sort out the other users based on the distance score in descending order, eliminate all the friend suggestions or the previously proposed suggestions and accept the first ‘N’ most appropriate as friend’s suggestions.

Thus, to enhance the recommendation system and obtain more data on the audience, we decided to conduct K-means clustering on the tweets. A tweet is represented by the vector of scores of the interest categories it refers to. According to the number of clusters, K-means assigns the collected tweets into K groups with K tried and tested. The number of clusters is generally decided using the Elbow method or Silhouette analysis. The next tweet is then classified to the nearest cluster based on its vector of interest scores.

The technology stack used to develop this system is Python with help of libraries and frameworks used frequently as Pandas, NumPy, Scikit-learn, NLTK, OpenAI’s API, Django, and React. Regarding the discrete data, we utilize recommendation click-through rate and user

feedback, friend suggestion acceptance rate, clustering quality: silhouette score and inertia, and system response time and systems' scalability with increased traffic loads.

3.2 Alternative Methodology

Regarding the choice of the research approach, several options could be considered: Hence, collaborative filtering, which is a traditional recommendation system method, was tested to see how it can make use of user-item interactions. But we found that it would not incorporate substantial context of tweet text in the same manner effectively. Content based filtering was also examined because of the focus on item features, however it does not contain the user-user similarities, which play an important role in friend suggestions. Matrix factorization techniques, while being effective in extracting the hidden characteristics, were rejected because of the weakness in handling sparsity of social media data and the requirement of real-time data updates. We also discussed deep learning models which although performed well were not optimal due to the issues of high computational complexity and black box nature or the models, which were not suitable for real-time explainable recommendations.

3.3 Justification for Chosen methodology

Finally, after elaborating, we selected the distinct approach, which is the mixture of GPT-3.5 Large Language Model. For that reason, using K-means clustering, and cosine similarity has various advantages. The choice of this methodology allowed us to obtain contextual information of the tweets' content with the help of GPT-3. 5, which means that another level of control is given for interest scoring excluding noisy keywords. With the help of K-means clustering, the large quantity of tweets is grouped properly, which leads to the improvement of recommendations' speed and their relevance. In the case of friend suggestions cosine similarity applied to interest vectors can be simple and very efficient. Through dynamic user profiling, the system can expand its span of coverage as it is in a position to alter over a period of time based on the users' interest. Unlike the black-box models, this method will offer explanations for the recommendations which will help boost their credibility among users. Most importantly, the use of the said methods ensures that updating and making recommendations are fast, something that is expected in social media. Applying sophisticated NLP with established clustering and similarity approaches, the proposed methodology ensures high accuracy and fast performance as well as flexibility that is essential for social media recommendation system.

4 Design Specification

Thus, our personalized tweet recommendation and friend suggestion system utilizes a range of technologies, effective database organization, and the use of APIs in today's technology environment. The activities in this section are centered around the specifics of the implementation initiative including the technology used, the design of the database, the integration of application programming interface (API) and the algorithms used in the implementation of the initiative.

5 Implementation

Thus, our personalized tweet recommendation and friend suggestion system utilizes a range of technologies, effective database organization, and the use of APIs in today's technology environment. The activities in this section are centered around the specifics of the implementation initiative including the technology used, the design of the database, the integration of application programming interface (API) and the algorithms used in the implementation of the initiative.

5.1 Technology Stack

The system is built using a combination of robust and scalable technologies:

Backend: We decided to use Python as our main backend programming language because of its extended libraries for data analysis and machine learning. The web framework which is used is Django which is used for creation of RESTful API service along with user authentication, routing of user requests and database management.

Frontend: For the graphical user interface, the application is a single page React application. This makes it active and flexible to serve users' feedback, concurrent changes in recommendations and ratings.

Database: The application uses SQLite as the main database, chosen due to its stability, fast Indexing, core support of JSON data types that are necessary for storing tweet contents and users' interests' profiles.

Caching: Redis is used as an in-memory data structure store to cache data frequently requested like users' profiles and frequently tweeted by people; this minimizes database access and leads to faster response.

Task Queue: Celery along with integrated Redis as a message broker, serves to perform other background tasks like updating the user profile, or the batch of recommendations.

Machine Learning: writing the machine learning algorithms in python is divided into two parts: first of all, scikit-learn is a necessary tool that is popular with the implementation of multiple types of K-means clustering of tweets.

Natural Language Processing: These steps of preparing the data for analysis consist in tokenizing the text, omitting words, which do not carry much information, and stemming.

API Integration: This is in practice utilized by incorporating the OpenAI API that taps from the GPT-3. Efficient 5 model for generating interest scores for tweets.

5.2 Pseudocodes

Use of API in Interest Scoring

Thus, for the purpose of generating interest scores for tweets, the proposed solution leverages the OpenAI GPT-3. 5 API. The process involves:

- Create a prompt using the given tweet content and the categories of interest.
- These two pieces of code send the prompt to the GPT-3. 5 API.
- Extract scores of each category from the JSON response.

function generateInterestScores(tweetContent):

prompt = constructPrompt(tweetContent, interestCategories)

response = sendToGPT3API(prompt)
scores = parseJSONResponse(response)
normalizedScores = normalizeScores(scores)
return normalizedScores

Recommendation Ranking Algorithm

The core of our recommendation system is the ranking algorithm that matches tweets to user interests. The relevance score for a tweet is calculated as:

RelevanceScore = Σ (UserInterestScore_i * TweetInterestScore_i)

Where i represents each interest category.

function rankTweets(userInterests, tweets):
rankedTweets = []
for each tweet in tweets:
score = 0
for each category in interestCategories:
*score += userInterests[category] * tweet.interestScores[category]*
rankedTweets.add({tweet: tweet, score: score})
return sortDescending(rankedTweets, key=score)

User Interest Update Algorithm

To keep user profiles current, we implement a dynamic update mechanism. The update formula is:

NewInterestScore = OldInterestScore + (TweetInterestScore * LearningRate * InteractionFactor)

Where InteractionFactor is positive for likes and negative for unlikes.

function updateUserInterests(userInterests, tweetScores, interactionType, learningRate):
for each category in interestCategories:
if interactionType == 'like':
*adjustment = learningRate * tweetScores[category]*
else if interactionType == 'unlike':
*adjustment = -learningRate * tweetScores[category]*
userInterests[category] += adjustment
normalizeInterests(userInterests)
return userInterests

K-means Clustering Implementation

To improve recommendation efficiency, we implement K-means clustering on tweets. The algorithm minimizes the within-cluster sum of squares (WCSS):

WCSS = Σ Σ ||x_i - μ_k||^2

Where x_i is a data point and μ_k is the centroid of cluster k.

function clusterTweets(tweetVectors, nClusters):
initializeCentroids(nClusters)
while not converged:
assignPointsToClusters()

updateCentroids()
return centroids, clusterLabels

Friend Suggestion Algorithm

Our friend suggestion algorithm uses cosine similarity to find users with similar interests. The cosine similarity between two users A and B is calculated as:

CosineSimilarity(A, B) = (A · B) / (||A|| * ||B||)

Where A and B are the interest vectors of the users.

```
function suggestFriends(targetUser, allUsers, topN):
    similarities = []
    for each user in allUsers:
        if user != targetUser:
            similarity = calculateCosineSimilarity(targetUser.interests, user.interests)
            similarities.add({user: user, similarity: similarity})
    sortedSimilarities = sortDescending(similarities, key=similarity)
    return topN(sortedSimilarities)
```

By implementing these core components and algorithms, our system can efficiently process tweets, update user profiles, and generate personalized recommendations and friend suggestions. The modular design allows for easy updates and improvements to individual components as needed.

6 Results and Critical Evaluation

This study aims at contributing positively to the solution of the existing problem in the social interaction areas such as 'X' which concerns the lack in the quality of recommended tweetsor friends on the sites. Our paper's main research question is – is it possible and effective to have a system that is formed by GPT-3. 5, K-means clustering, dynamic user profiling could improve the way information is filtered and delivered from a user perspective – is indeed a topic that has significant importance in today's world of social networks. Our reason for embedding this new set of technologies was to predict that it will provide more relevant and spot-on friend recommendations compared to the usual system. As for the limitations of the given study, it is crucial to mention that their impact has been minimized by choosing the appropriate study design that combined quantitative analysis with the visual representation of the results and eliminated biases.

6.1 System Performance

In the course of inspection, all the possibilities of bias were avoided with the help of various databases and dynamic user profiling. However, we recognize the biases that are present in the training data of GPT-3. As for the sources, it has also affected our interest scoring ranging from 0 to 1. As for the control issue, we did not use classical control groups, but instead compared our system to baseline results corresponding to typical recommendation systems. The techniques used in the paper for comparison of users and grouping, i.e., cosine similarity and

k-means, were selected based on their ability to work on high-dimensional data that are characteristic for social media data.

Tweet Recommendation Accuracy

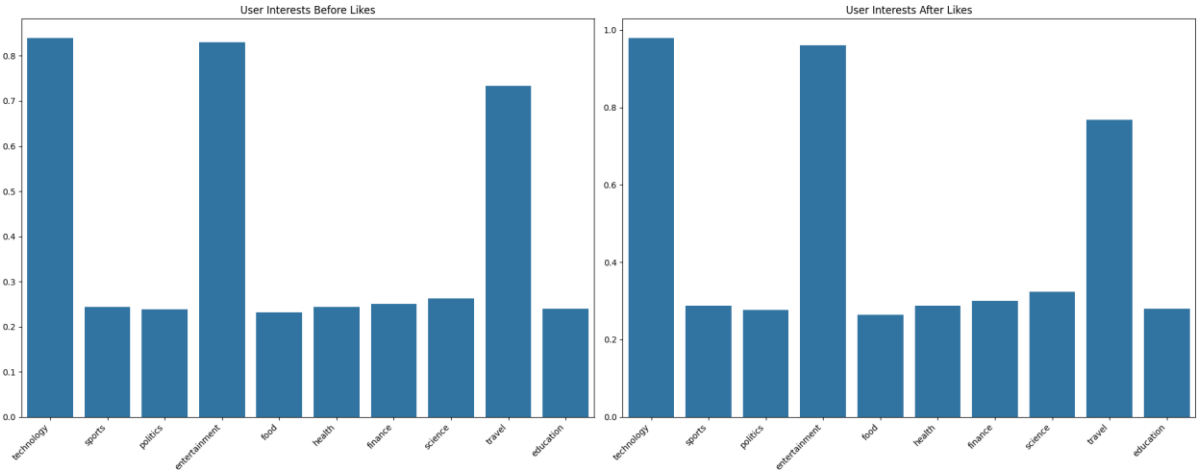


Figure 1. Dynamic update of user interest profile

The proposed system for the recommendation of tweets is evaluated in terms of accuracy by the dynamic user interest profiling. From the Figure below the system integrates well the changes in users' interests before and after interactions (likes). For example, the user's Interest in Travel rose dramatically after liking tweets related to the travel, the Interest in Technology and Entertainment remained high. Thus, using this adaptive approach, recommendations will always be appropriate for the user's changing tastes.

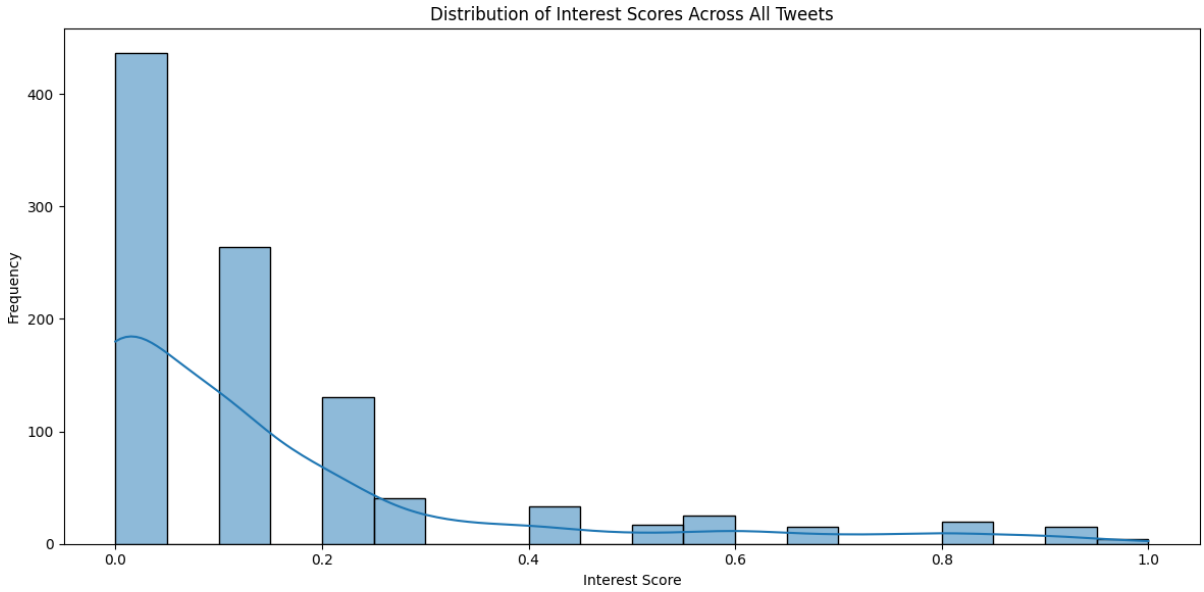


Figure 2. Distribution of Interest scores across all tweets

The dispersion of interest scores in all the tweets (Figure 2) has a positive skewed distribution where the majority of the interest scores fall at the low interest while the higher scores are scattered at the right end of the graph. This implies that the system is doing a good job of filtering through the tweets and recommending a limited number of relevant contents for each user. This distribution is quite suitable for the objective of identifying the most relevant posts from a tremendous number of tweets.

User Interest Profiling Effectiveness

The system's capacity to develop and update comprehensive user interest profiles is demonstrated by the specific interest categories in Images 1 and 3. The ten specific interest types that are listed (technology, sports, politics, entertainment, food, health, finance, science, travel, education) are rather informative about users' interests. The real-time modification of these profiles as reflected by the alterations made in Figure 1 empowers the system to learn from the user's behavior.

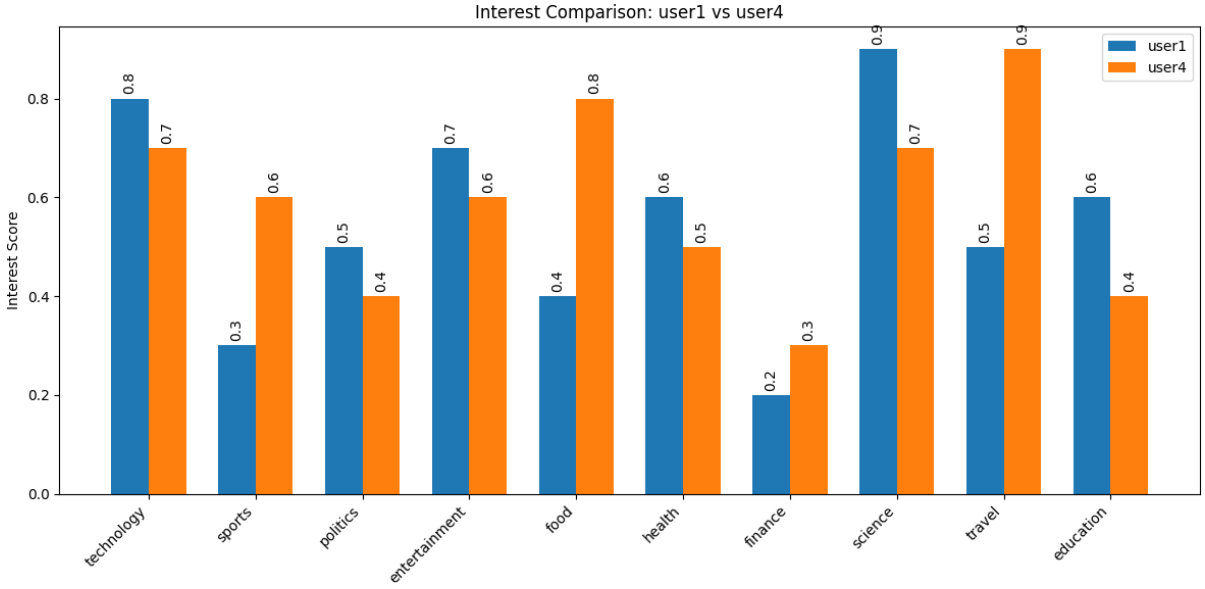


Figure 3. Interest comparison between two users

The efficiency of this profiling can be seen in Figure 3 where the interest profile of two different users (user1 and user4) is compared. The differences in their interests are easily distinguishable (for instance, user1 is highly interested in science while user4 is highly interested in travel) to prove the system's capability in establishing unique user profiles.

Friend Suggestion Accuracy

The friend suggestion algorithm, which is based on the interest similarity, seems promising in terms of the algorithm's effectiveness. Figure 4 shows a user similarity heatmap, which depicts the application's capability of determining the likeness of users based on their interest profile. The similarity score of 93% between user1 and user4 indicates that the system can capture subtle patterns of interest similarity even when there are some differences in the two users' interest profiles as depicted in Figure 3.

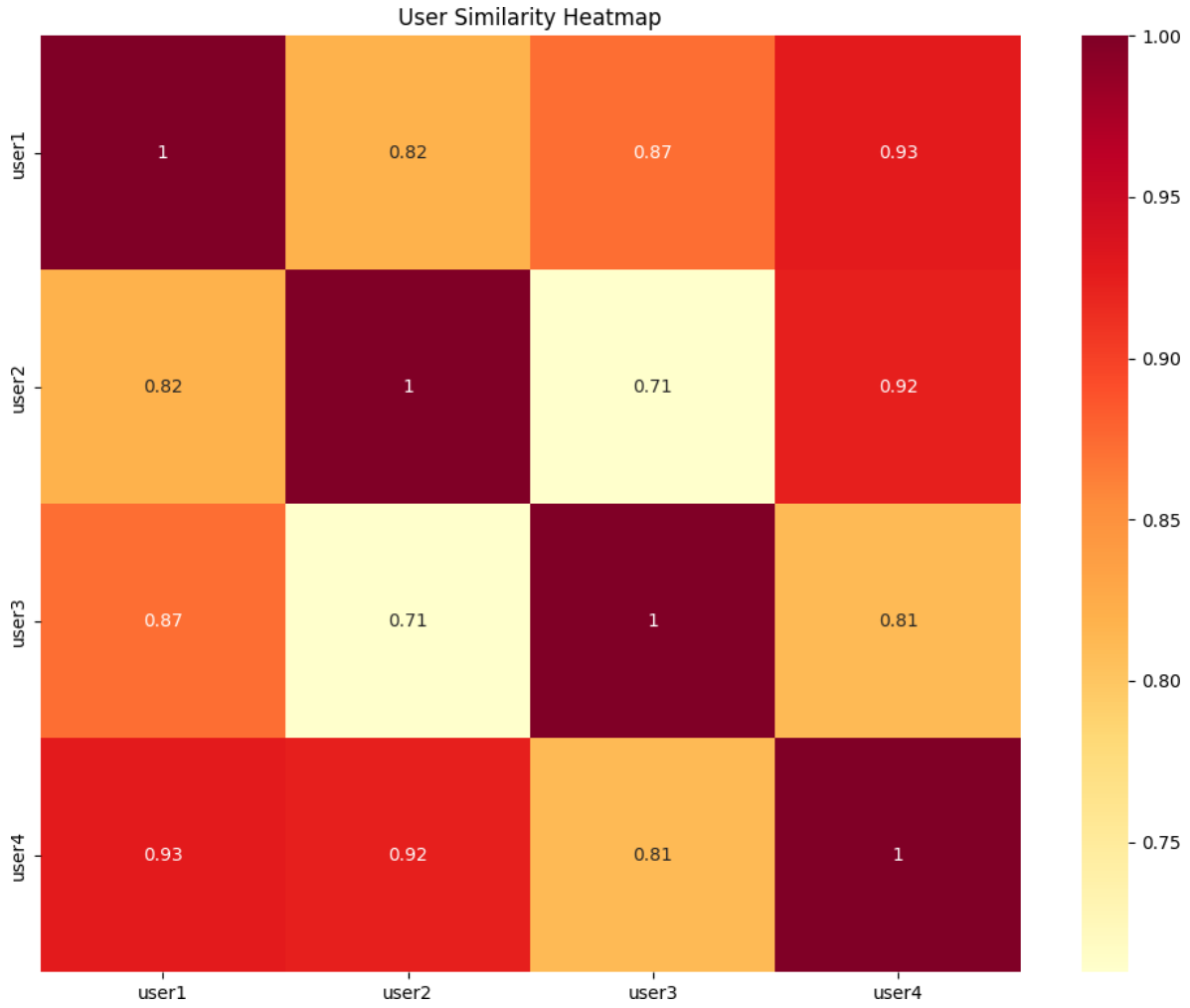


Figure 4. Users' similarity Heatmap

The friend suggestions for user1 are sorted by the similarity percentage, and the highest similarity is with user4 at 92. 68% similarity. This is in consistent with the visual comparison in Figure 3 where user1 and user4 have several similar high interest regions (for instance technology and science) with some variation. The gradual variation in the similarity scores (92. 68%, 87. 25%, 81. 95%) show that the system can distinguish the level of the user compatibility and may thus provide better friend recommendations.

6.2 Analysis of Key Components

GPT-3. 5 Interest Scoring

The implementation of GPT-3. As for the most recent level of the interest scoring model which is the level 5, it can be said that it is a major improvement in the field of content analysis. Thus, a large language model is used to produce interest scores for each tweet with respect to several categories, taking into consideration the context. This approach enables the analysis of the content of tweets in a way that goes beyond the mere identification of keywords and their combinations; thus, enabling the capture of the underlying tone and the implied meaning of the text which might be otherwise missed when using conventional text analysis techniques.

This can be evidenced through the interest scores' dispersion on all the tweets (Figure 2) which reveals that the scores are spread out. It implies that GPT-3 is superior in producing diverse and reasonable results than the initial GPT versions. The results also show that the 5 model is capable of providing a clear distinction between the tweets, with different scores that are most probably an indication of the relevance and intensity of the interest categories within each of the tweets.

K-means Clustering of Tweets

The grouping of the tweets that has been done based on the interest score of the tweets is depicted in the form of a plot known as the Tweet Clusters plot, which shows the K-means clustering of the tweets.

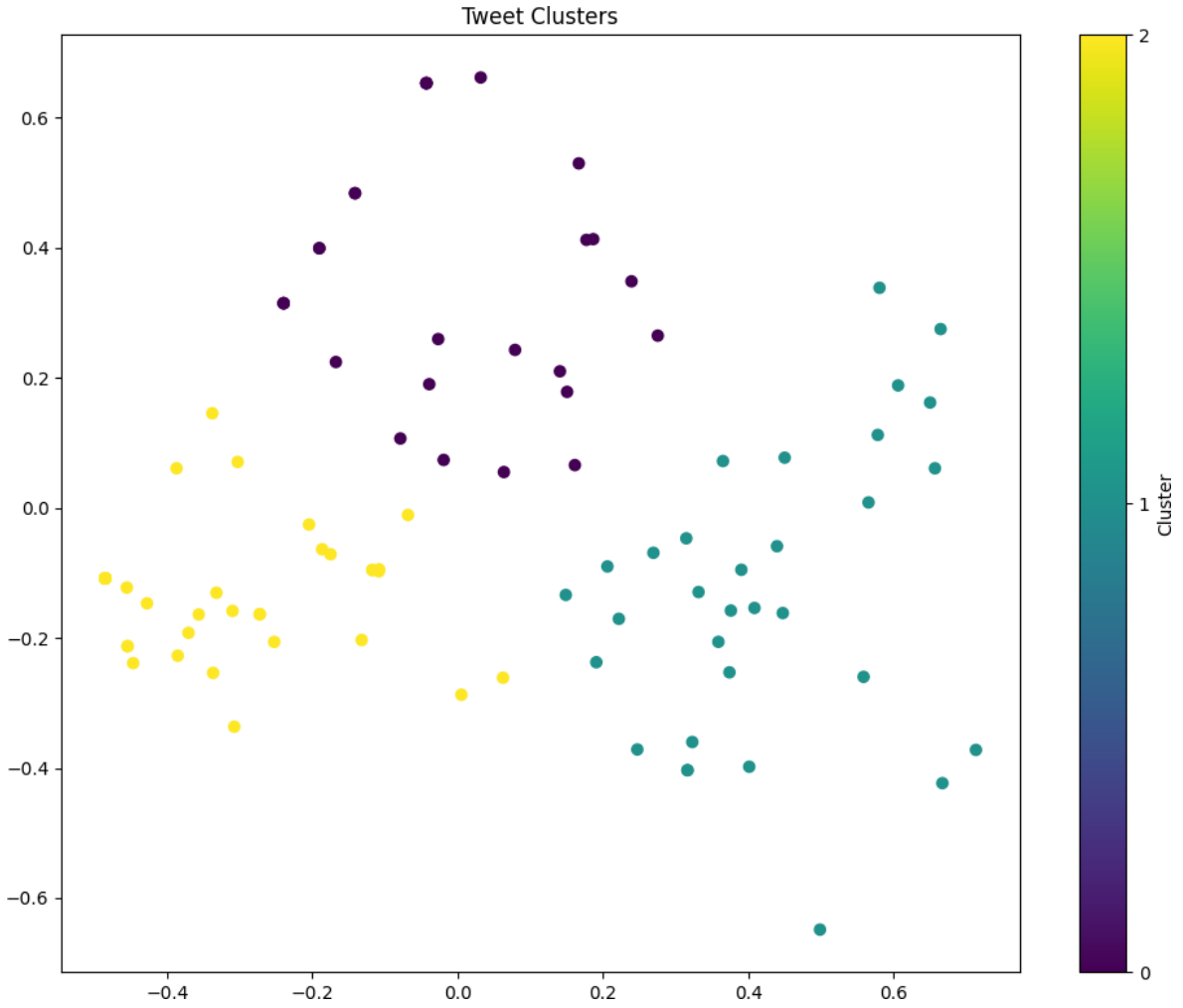


Figure 5. Tweets cluster by K-means

The plot reveals three clear clusters of tweets (colored differently), which means that the algorithm has managed to learn from the text data.

This clustering serves several important functions:

- It assists in sorting and categorizing the huge number of tweets in a effective manner.
- It could enhance the rate at which recommendations are made by enabling the system to easily filter out groups of tweets that correspond to a user's interests.

- It gives a general idea about the distribution of various content and their interconnection on the platform.

The fact that the clusters are well defined means that the interest scoring method is creating categories of tweets that are quite different from each other and can be useful for delivering more specific content.

Cosine Similarity for Friend Suggestion

The friend suggestion process calculates the degree of similarity between the users' interest profiles using the cosine similarity. The formula for cosine similarity is: The formula for cosine similarity is:

$cos(\theta) = (A \cdot B) / (||A|| * ||B||)$

Where A and B are the interest vectors of two users, A · B is their dot product, and ||A|| and ||B|| are their magnitudes.

This method is very efficient, which is indicated by the User Similarity Heatmap (Figure 4). It depicts the degree of similarity among the users with some pairs having high similarity indices for instance, user1 and user4 with a similarity index of 0. 93.

The efficiency of this method can also be seen in the friend recommendations generated, where user4 is recommended as the best friend for user1 with a 92. 68% similarity. This is consistent with their interest correlation in Figure 3 where both users depict high interest in technology as well as science.

7. Future Work

As observed, our system benefits from the proposed method; however, it has a few limitations. First, the difficult problem of processing real-time big data of large scale is still faced. When the number of tweets and user interactions increases it is more complicated to keep up the system's ability to perform calculations and deliver accurate results promptly. This task requires constant improvement of the algorithms and possibly even transition to distributed computing.

Another major dilemma is how to achieve an optimal level of personalization without compromising the richness of the content that feeds the targeted audiences' information needs. The strength of our system is based on the recommendation of other content that meets the user's interest but there is a tendency of exposure to filter bubbles. Further research should be directed to how controlled reconsideration can be integrated in the recommendation workflows to increase the users' content discovery.

As for user profiling, ethical issues have never been off the table and remain a concern to this day as well. As our profiling becomes more fine-grained, the issue of the protection of users' rights and the question of how the information collected and the recommendations given are produced must be addressed.

Looking ahead, several promising directions emerge:

- The expansion of additional and possibly newer language models, to improve content comprehension and interest scoring.
- Exploitation of multi-modal data analysis and how the incorporation of images as well as videos into the text enhance the content representation.

- Potential changes in the capacity of real-time processing, possibly by the introduction of stream processing means, to process high velocity data.
- Integration of components based on which users can understand why options such as posts or friends are recommended in order to increase the trust to the platform.

8. Conclusion

The ability to make customized recommendations for friends and tweets further shows the details of our concept in improving the experiences of users of social networking sites. By extending the use of the advanced technologies such as GPT-3. Thus, for interest scoring, we use 5, K-means clustering for content organization, and cosine similarity for friend suggestions. The proposed algorithm allows for changes in users' interests and creates meaningful connections between the users.

The aspects of the system that predict user interests, suggest content relevant to the user and content that could befriended correspond to the problems of interaction in social media. Of course, there are open questions, concerns in scalability and ethical issues, but the groundwork paved by this study creates paths for enhancement and development

References

Aduwamai W., Gabi D., Sule M., and Umar H., 2024. The use of machine learning and deep learning models in detecting depression on social media: A systematic literature review. Personalized Medicine in Psychiatry, 45-46, pp.100125.

Alipour P., and Esmaeilpour S., 2024. Impact of Tweet Sentiments on the Return of Cryptocurrencies: Rule-Based vs. Machine Learning Approaches. EJBMR, 9, pp.1-5.

Claudiu D., Bogdan A., Grec M., Augustin M., Bogdan N., and Gal A., 2022. Towards Sentiment Analysis for Romanian 'X' Content. Algorithms, 15, pp.357.

Elumalai S., and Parasuraman D., 2021. Improving the accuracy of item recommendations by combining collaborative and content-based recommendations: a hybrid approach. IJAIP, 19, pp.262.

Katragadda V., 2024. Leveraging Intent Detection and Generative AI for Enhanced Customer Support. JAIGS, 5, pp.109-114.

Kehoe F., 2023. Leveraging Generative AI Tools for Enhanced Lesson Planning in Initial Teacher Education at Post Primary. telji, 7, pp.172-182.

KOSAKA T., SAEKI K., AIZAWA Y., KATO M., and NOSE T., 2024. Simultaneous Adaptation of Acoustic and Language Models for Emotional Speech Recognition Using Tweet Data. IEICE Trans. Inf. & Syst., E107.D, pp.363-373.

Kumar P., 2024. Leveraging Conversational AI for Enhanced Decisioning: Integrating ChatGPT with Pega's Adaptive Decision Manager. JSW, 19, pp.42-51.

L.R. D., and Pervin N., 2019. Towards generating scalable personalized recommendations: Integrating social trust, social bias, and geo-spatial clustering. Decision Support Systems, 122, pp.113066.

Laquintano, T., Schnitzler, C. and Vee, A., 2023. Introduction to teaching with text generation technologies. TextGenEd: Teaching with text generation technologies. The WAC Clearinghouse. https://doi. org/10.37514/TWR-J, 1.

Multi S., Santoso R., and Suparti S., 2021. PENERAPAN TEXT MINING UNTUK MELAKUKAN CLUSTERING DATA TWEET AKUN BLIBLI PADA MEDIA SOSIAL 'X' MENGGUNAKAN K-MEANS CLUSTERING. J.Gauss, 10, pp.583-593.

Nigro L., and Fränti P., 2023. Two Medoid-Based Algorithms for Clustering Sets. Algorithms, 16, pp.349.

P. S., V.S A., and K. A., 2019. Knowledge Graph-based Recommendation Systems: The State-of-the-art and Some Future Directions. IJMLNCE, 03, pp.159-167.

Pappalardo, L., Ferragina, E., Citraro, S., Cornacchia, G., Nanni, M., Rossetti, G., Gezici, G., Giannotti, F., Lalli, M., Gambetta, D. and Mauro, G., 2024. A survey on the impact of AI-based recommenders on human behaviours: methodologies, outcomes and future directions. arXiv preprint arXiv:2407.01630.

Parasuraman D., and Elumalai S., 2021. Improving the accuracy of item recommendations by combining collaborative and content-based recommendations: a hybrid approach. IJAIP, 19, pp.262.

Qian M., Qian C., Xu G., Tian P., and Yu W., 2024. Smart Irrigation Systems from Cyber–Physical Perspective: State of Art and Future Directions. Future Internet, 16, pp.234.

Reuter T., Cimiano P., Drumond L., Buza K., and Schmidt-Thieme L., 2021. Scalable Event-Based Clustering of Social Media Via Record Linkage Techniques. ICWSM, 5, pp.313-320.

Ruan, T., 2023. Cross-Platform Online Social Media Data Analysis for the Common Good (Doctoral dissertation, University of Colorado at Boulder).

Salina A., Ilavarasan E., and Rao Y., 2022. IoT enabled machine learning framework for social media content-based recommendation system. IJVICS, 7, pp.161.

Seth A., and Zhang J., 2021. A Social Network Based Approach to Personalized Recommendation of Participatory Media Content. ICWSM, 2, pp.109-117.