

Personalized Recommender System for Arabic News on Twitter

Bashaier Almotairi, Mayada Alrige, Salha Abdullah

Department of Information Systems
Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah, Saudi Arabia

Abstract—Reading online news is the most popular way to read articles from news sources worldwide. Nowadays, we have observed a mass increase of information that is shared through social media and specially news. Many researchers have proposed different techniques that focus on providing recommendations to news articles, but most of these researches focused on presenting solution for English text. This research aimed to develop a personalized news recommender system that can be used by Arabic newsreaders; to display news articles based on readers' interests instead of presenting them only in order of their occurrence. To develop the system we have created an Arabic dataset of tweets and a set of Arabic news articles to serve as the source of recommendations. Then we have used CAMEL tools for Arabic natural language processing to preprocess the collected data. After that, we have built a hybrid recommender system through combining two filtering approaches: First, using a content-based filtering approach to consider the user's profile to recommend news articles to the user. Second, using collaborative filtering approach to consider the article's popularity with the support of Twitter. The system's performance was evaluated using two evaluation metrics. We have conducted a user experimental study of 25 respondents to perform an assessment to get the users' feedbacks. Also, we have used Mean Absolute Error (MAE) metrics as another way to evaluate the system accuracy. Based on evaluation results we found that hybrid recommender systems would recommend more relevant articles to users compared to the other two types of recommender system.

Keywords—Hybrid recommender system; online social network; Arabic news recommendation

I. INTRODUCTION

The magnitude of data generated on the Internet by different business communities, administrations, industrial sectors, scientific research and general data has increased immeasurably [1]. The research of [2] has reported that the world produces about 2.5 quintillion bytes of internet data daily, and almost 90% of this data is unstructured. Similarly, on social media sites, there is a massive bulk of un-related data and other non-authentic news which cause the users several issues related to privacy, psychological problems and many more. Organizations and some users need different new methods to process the extensive, massive data, on social media, into meaningful comprehensions. A report by [3] stated that more than half of the world's population is presently using the Internet and more than 250 million population use Arabic.

Understanding Arabic words and their vowels and consonants is still lacking in machine learning and getting the proper theme of the words written in Arabic. This situation is due to the significant number of Arabic languages subcategories with different accents and meanings, which cause the listener to misconception the idea [4]. The text categorization to assign a category to Arabic language content on the Internet and in social media lags behind the actual outcomes, since there are no firm rules for Arabic language understanding by machine learning systems [5]. Part of these massive Arabic data on social media and other platforms is news. The relation between journalists and audience is shifting towards the use of new technologies and algorithms in digital news. Several platforms display your content in this digital world as YouTube, Instagram, Facebook and different news sites and journals. Different other platforms are being used in news field, but not all of these platforms are authentic and reliable for the news and data collected.

This research's motivation is the need for a valuable tool or system to help readers in such information overload situations. Thus, the news at any site can be displayed in an order based on the reader's interests, instead of using the same order of presentation for all readers based on a publisher's opinion. The motivation for the solution comes from [1], where the researcher has developed a news recommender system with the help of micro-blogging services. News articles are sorted based on the popularity of the article, which is identified using tweets from the public timeline of Twitter. The researcher has also built the user profile based on the user's interests, so the news articles are sorted by matching the user profile's characteristics. The research of [6] has developed a personalized recommender system for calculating the closeness between Twitter users in a social circle. The system suggests topics or interests that users may have to analyze social information. This solution and other similar propositions have already been presented English medium, but not in the Arabic language.

The research aims at developing an Arabic recommender system to display news or articles to each reader based on their interests instead of presenting them only in the order of their occurrence using Twitter.

The research is organized as follows. Section 2 gives a comprehensive overview of recommender systems and discusses other researchers' work in this field. Section 3 discusses the process of data gathering and preprocessing to move to Section 4, where it describes the design and implementation of our Arabic news recommender system.

Section 5 discusses the evaluation of the proposed system's result using one of recommendation systems evaluation metrics. Finally, Section 6 summarizes the research and suggests future works.

II. LITERATURE REVIEW

A. Recommendation Systems

The rapid increase in the amount of accessible digital data and the vast numbers of Internet users has generated a possible knowledge overload problem that sometimes hinders timely accessibility to online information or needed items. Data gathering systems, including Google, Altavista and DevilFinder, have slightly fixed the issue, but the personalization and prioritization of the information have been missing. It has generated a much higher demand for recommendation systems compared to older times. These systems are considered as knowledge retrieval systems used to tackle the issue of overloading information [7]. This is done via filtering essential fragments of information from a substantial amount of dynamically produced information as per user interest, actions or preferences noticed regarding specific items [8]. Using information from user profiles, recommendation systems possess the potential to predict if a user will choose something or not [9].

While several strategies have been established in the past, research is still ongoing as it is mainly used in many apps that configure suggestions and deal with overloading information [10]. These systems benefit both service providers and users [11], they substantially reduce transaction costs for detecting and choosing products in an online retail environment [12]. These systems have been shown to enhance the reliability of the decision-making process. When setting up e-commerce, the recommended systems boost sales since they successfully sell more products [11]. Throughout libraries science, these systems assist consumers by encouraging them to step past index searches. Recommending programs assist consumers of research libraries by encouraging them to pass beyond catalogue searches. Consequently, the need to utilize precise and effective recommendations strategies inside a program that can offer reliable and relevant advice to consumers could not be overemphasized [9].

B. Recommendation Systems' Approaches

Using systems of recommendation that are both reliable and efficient is necessary for a program that aims to offer excellent and practical guidance to its user, demonstrating the value of knowing the characteristics and potential of different domain approaches. Classifying recommendation system is usually based on the rating estimation [10]. Generally, there are three approaches to recommendation systems:

- Collaborative filtering approach.
- Content-Based filtering approach.
- Hybrid Filtering approach.

Collaborative filtering is a domain-independent predictive strategy where the information cannot be conveniently and accurately classified via metadata, e.g., music and movies. This approach operates by constructing a database with user

preferences for specific items. Afterwards, it groups the users with related preferences and interests to make recommendations by calculating how similar their profiles are [13]. Content-based filtering technique uses an algorithm which is dependent on domain and relies primarily on evaluating the properties of objects for producing predictions. It is most suitable where documents such as publications, web pages and news, have to be recommended. Throughout this method, recommendations are provided using user identities formed by using features that have been derived from the contents of those items that previously have been reviewed by the particular user, so the system will recommend the user of items that are so related to the items he/she liked before [14,15].

To understand the difference, Content-Based filtering requires information about items' features, instead of using user's interactions and feedback. Good examples are movie attributes such as actor genre, year, director or textual articles content. On the other side, Collaborative filtering doesn't require anything else except the user's historical preference on a set of items to recommend from. So it assumes that user who has agreed in the past will also tend to agree in the future [14,15].

Traditional filtering techniques have distinct weaknesses and strengths. For instance, Collaborative filtering suffers from cold start and sparsity issues, whereas content-based suffers from need and narrowness explanation. Nevertheless, hybrid solution that uses one method to render recommendations where the other fails lead to a more reliable recommendation framework [16,17]. Hybrid filtering combines various recommendation techniques to optimize the system's optimization to avoid specific difficulties and challenges with the systems of recommendation [18]. By using more than one filtering technique, the limitations can be minimized and the recommendations will be more accurate in a hybrid model [19].

C. Current Researches in Recommendation Systems

Some researchers have proposed the hybrid recommendation system technique as an excellent solution to solve or enhance their research issue. The research [20] has discussed the impact of visual information, i.e., customers' photos and put on some blogs, to predict favourite restaurants for any given user. By considering the visual information as an intermediate, the researcher suggested integrating two common recommender system approaches, collaborative filtering and content-based filtering, to show the proposed hybrid system's effectiveness with considering visual information. Another research of [21] has described a recommendation system built on a probabilistic programming language and discussed the benefits and challenges of explaining the generated recommendations to users. Using an online user survey, the research evaluated the explanations for hybrid algorithms in a set of text, visual and graph formats, which are either new designs or derived from existing hybrid recommendation systems. Moreover, the research showed that hybrid systems demonstrate better accuracy than recommendation strategies that use a single source.

A study of [22] had focused on creating an engine for a product recommendation, which can be accessed by third-party software to obtain certain product recommendations using some input data. The study's overall aim was to make novel contributions on how the user events can be gathered and processed and how they can be used by the methods of data mining to make product recommendations. The study has solved product recommendations appropriate for big data by investigating methods to process user event data.

D. News Recommendation Systems

The relation between journalists and audiences is shifting towards new technologies and algorithms in digital news [23]. Several platforms display the digital world's content as YouTube, Instagram, Facebook and different news sites and journals. Different approaches are being used in the news field, but not all platforms are authentic and reliable for the news and data collected. These news fields serve the expanded eyes of the viewers or listeners. The study of [24] showed how the news topics could be used to recommend English news articles. The researcher used supervised learning methods such as Naive Bayes, linear regression and logistic regression. All of these machine learning models have a different nature to determine the user ratings for an article. The research of [25] has presented a robust system for providing English real-time news recommendations to the user based on the user's history of the last visits to the website, popularity of stories and current user's context.

E. Arabic Language and Recommendation Systems

Arabic language is one of the most widespread languages globally [26], spoken by more than 422 million people. Its speakers are distributed in the region known as the Arab world and many other neighboring regions such as Turkey, Chad, Mali, Senegal and Eritrea. Arabic language considered as one of the Semitic languages that are rich in syntax (words arrangement to make phrases and sentences) and its morphology (the structure of internal word) [27]. There are three forms of Arabic language: classical Arabic [28], which is the language of Islam's Holy Book (Qur'an) that is not used in daily life, modern standard Arabic (MSA) as well as dialect Arabic. MSA is the standard form that is used in official news, education and media.

Unfortunately, the Arabic language's Recommendation Systems did not acquire enough attention [29]. It happened due to the limited number of tools and other challenges related to the language's nature. According to [30], the Arabic language became a challenge for researchers and machine learning developers due to language richness, language complexity, ambiguous structure of Arabic, and available Arabic types dialects.

F. The Microblogging Service Twitter

Microblogging is a general term of any web service that allows the users to broadcast short messages to other users of the service quickly from mobile devices. One of the first microblogging services is Twitter.com [31]. Twitter is one of the digital platforms where users can read news. It is a microblogging or a social networking service on which users can post and interact with messages known as tweets [32].

Registered users can like, post and retweet any tweets, while unregistered users can read them only. Any user can access Twitter through the website interface, through short message service SMS or its mobile application.

Now-a-days, Twitter has become one of the most popular services with students, academics, politicians, policymakers and the general public. In the past, many users find some difficult in understanding what Twitter is and how they can use it. However, now it has becomes the social media platform of choice for many.

In collaboration with research company DB5, the American Press Institute with Twitter has created a new study that explores and examines the relationship between news use and the Twitter environment [33]. The study involved an online survey of more than 4,700 active social media users. It found that Twitter users tend to be heavier newsreaders than other social media users. The study also found that reading news is one of the main activities they engage in on the network.

While these users on the service in general, or sometimes do so just as a way of passing the time, they behave differently when following breaking news. They participate, comment, post and share at moments more when events are moving fastest. These signals can lead news publishers to make more effective use of social media in general and Twitter in more particular.

Due to the limitation discussed above, we propose a hybrid personalized Arabic news recommender system that recommends interesting articles or news to users using Twitter service. The proposed tool sorts the news in two ways: Firstly, using a content-based filtering algorithm to consider the user's profile to recommend news articles to the user. Secondly, using collaborative filtering algorithms to consider the article's popularity with tweets from Twitter's timeline. We propose this approach to help users find interesting news articles to read by combining the above two filtering techniques of sorting articles [34].

III. DATA GATHERING AND PREPROCESSING

A. Data Collection

1) *Obtaining news articles:* In order to collect Arabic News articles, we accessed RSS (Really Simple Syndication) feeds of online Arabic news sites such as Alarabiya (<https://www.alarabiya.net/>) and CNNArabic (<https://arabic.cnn.com/>). RSS is an XML file format created by news website editors for distributing and sharing web content, such as news headlines and summaries. Using an RSS reader, users can view data feeds from many news sources [35]. Alarabiya and CNNArabic organize their news articles by categories, e.g., Politics, Economics, Sport, etc. Therefore we know the associated topic for every article in the collection. The article collection includes 150 different Arabic news articles collected from six different main categories (Sports, Economics, Technology, Entertainment, Health and Politics).

2) *Obtaining tweets dataset:* In order to obtain a dataset of tweets, a streaming API is set up to store incoming tweets as

soon as they are posted in the Twitter public timeline using Tweepy Library. The data returned from Twitter streaming API are formatted using JSON (JavaScript Object Notation). In this research we only need the text attribute that contains the actual tweet text, the other attributes in the tweet object were ignored. The dataset for this research is a collection of 100,000 of last published tweets which are collected along with the news articles dataset on the same day.

B. Data Preprocessing

To create all data preprocessing steps for Arabic text, we decide to use CAMeL tools [36]. It is an open-source toolkit for Arabic natural language processing in Python. CAMeL tools offer utilities for preprocessing, dialect identification, sentiment analysis, named entity recognition and morphological modelling. What makes CAMeL the most suitable tool for our proposed system is that it is implemented in Python which is our research programming language, whereas other available tools are only supporting Java. The tools also provide APIs and command line interfaces in order to cover CAMeL's utilities, where others provide only command-line tools. This can lead to overhead writing glue code that includes interfaces to different packages. Through using CAMeL tools, we have followed the steps below in order to preprocess the obtained data.

1) *News articles cleaning*: The RSS news articles need to be preprocessed before making them ready for implementation. We needed to remove unnecessary content such as numbers, special characters and HTML tags.

2) *Tweets cleaning*: On the other side the tweets dataset need to be cleaned as well to eliminate unwanted noise and preserve textual content only, this can be done by addressing the issues of character replication, abbreviations, hashtag, URLs, usernames, whitespace and removing repeated tweets.

3) *Tokenization*: The first step after data cleaning is tokenization; this step consists of splitting the text into tokens or words separated by punctuation characters or whitespaces. The result of this phase is a set of words.

4) *Normalization*: It is the process to change all the forms of a word into a common form. To get texts common forms, the added normalizer removes the ‘tatweel’ character ‘_’ such as changing the word “واسع” which means “wide” to look like “واسع”, and tashkeel such as “هذة” which means “truce” to make it as “هذه”. There are also other steps include removing symbols and special characters such as ! ? () , numbers such as ٦ which is 6 in English , non-Arabic words such as طبعاً “NO”, and Al “التعريف” which is used as the definite article ‘the’ in English. We will also replace the letters “إ” and “ئ” with the letter “إ” and the letter “ئ” with the letter “ئ”.

5) *Stop words removal*: Stop words such as prepositions and pronouns are used frequently in tweets and news articles dataset. We decided to remove them since they are not significant in the research. For instance, these words have been removed: “من” of, “على” on, “انت” you and so on.

Fig. 1 shows the general dataset cleaning and preprocessing phases.

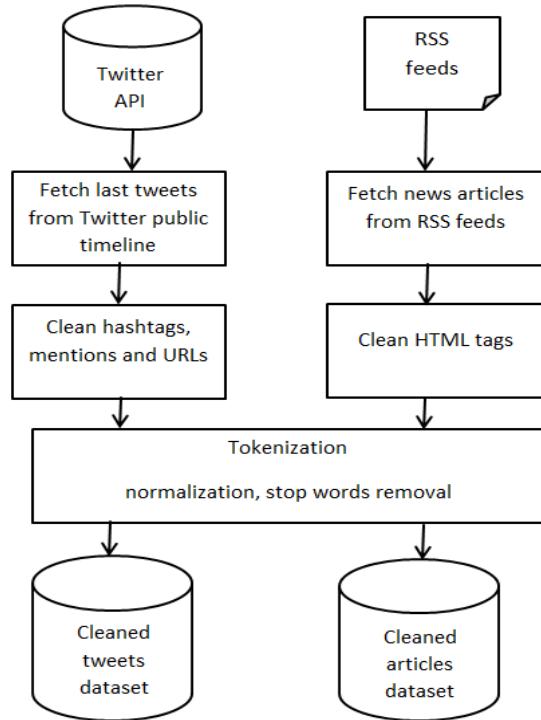


Fig. 1. Dataset Cleaning and Preprocessing.

IV. DESIGN AND IMPLEMENTATION

A. High-Level Design

Our proposed system consists of three parts: collaborative recommender system which is popularity-based, content-based recommender system which is profile-based and hybrid recommender system which is the combination of the two previous systems.

The first recommender selects news articles based on the popularity of the article. In this research, the article popularity is identified with the help of the Twitter service. So tweets are gathered from Twitter public timeline and preprocessed to identify the articles that users around the world are tweeting. The tweets are then compared to the news articles based on the co-occurring terms in the articles and the tweets. Articles that mentioned frequently in the tweets will be considered popular or hot.

The second recommender sorts the news articles based on their similarity to the user's profile. In this research, the users build their profiles by providing input about their level of interest in each of six news categories. The incoming news articles are classified into same set of categories using K Nearest Neighbors text classifier. This classifier works by finding the K nearest matches in the training data and then using the label of closest matches for predicting [37]. The articles are then sorted based on the similarity between the categories the user showed interest in his/her profile and the categories which the article belongs to.

Finally the hybrid recommender combines the results from the two recommenders to recommend news articles to the user. It sorts the new articles based on the combination of their popularity ranking and the similarity to the user's profile.

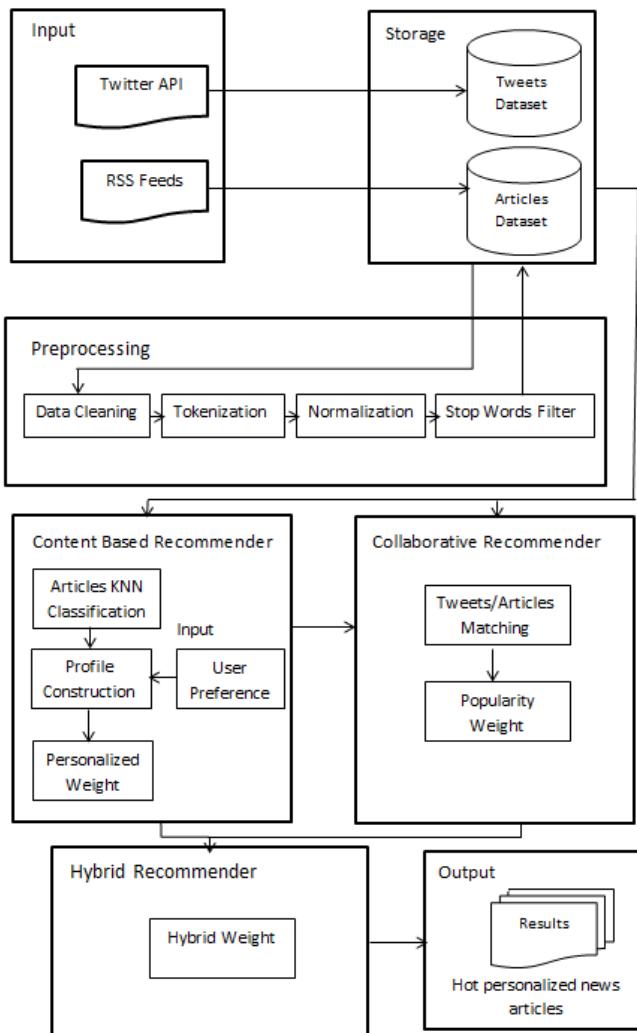


Fig. 2. General Architecture of the System Components.

Fig. 2 shows an architectural diagram of the proposed hybrid recommender system.

B. Collaborative News Recommendation

After the RSS news articles are gathered and preprocessed, the news articles are then indexed using Elasticsearch. Elasticsearch is a free open source search engine built on top of Apache Lucene [38]. It can index many types of content and can be used for several cases such as website search, application search, monitoring application performance, business analytics, etc. It is known for its simple REST APIs which make it flexible to use from any programming language. In addition, it supports 34 languages including Arabic and provides analyzers for every language.

To identify the most popular news articles, the processed tweets are passed as queries to Elasticsearch server. Every tweet is queried against the server to return the articles that

correspond to the tweet content. The server returns the articles associated with the tweet content together with a weight that refers to the similarity between the tweet as query and the article. So the articles are sorted according to their accumulated weights across all tweets. The news articles will be ranked based on their popularity using cosine similarity measures between articles and tweets as shown in (1) [39].

$$\text{Popularity Weight} = \text{CosineSimilarity(Article, Tweet)} \quad (1)$$

C. Content-Based News Recommendation

This recommender uses the same news articles collection as in the collaborative recommender. Although, news articles are placed in just one category by the website editors, they actually might belong to more than one category. To allow for that in the recommender, each news article is classified into seven potential categories using K Nearest Neighbor classifier. With this algorithm we can classify the articles into categories based on the articles similarity to the training documents for each category.

K-Nearest Neighbors classification algorithm [37] identifies K most similar training documents to the article and then uses these similarity scores as a vote for the category that the training documents belong to. Thus the article similarity to each category is the sum of the scores of the article's similarity to the training documents in that category which placed in overall top K most similar documents. So then the categories are sorted through their accumulated score. Then we use Elasticsearch again to create a second index that maps from category IDs to documents IDs and weights. So this index stores news articles based on the category that they belong to instead of the keywords in the article.

Every user creates his profile based on his interest in different categories, every user has an interface to build his profile manually by scoring the categories. Fig. 3 shows the interface that allows the user to create his unique profile. Each user will enter scores in the range of (0-10) so that the total at the end will equal to 10.

نظام توصية نشرات الاخبار العربية

الرجاء اختيار الفئات المفضلة عن طريق ادخال الارقام من ١ إلى ١٠

حيث يكون مجموع الارقام يساوي ١٠ كمثال: (٦ - الرياضة -٣ -الاقتصاد -١ - الصحة)

الفئة	الدرجة
الرياضة	
الاقتصاد	
الصحة	
التقنية	
السياسة	
الترفيه	

إرسال

Fig. 3. User Profile Interface.

After that, this user profile is used to identify the documents which best match his profile. The articles and profiles can be considered as feature vectors where every category is a feature. In (2) we have used cosine similarity measure [39] to get the similarity of each article to the user's profile as shown below.

$$\text{Personalized Weight} = \text{CosineSimilarity(Article, Profile)} \quad (2)$$

D. Hybrid News Recommendation

This system combines the scores result by each of the previous two recommenders to generate a new recommendation that integrates the news articles' match to the user's interest with the news articles popularity everywhere. It calculates the hybrid weight through multiplying popularity weight by personalized weight as shown in (3) [40].

$$H\text{Weight} = \text{Popularity Weight} * \text{Personalized Weight} \quad (3)$$

V. EVALUATION

One of the important phases in building any recommender system is evaluation. There are different metrics to evaluate the performance; the most common are accuracy and experimental studies [41]. Accuracy metrics are divided into decision support and statistical metrics. Decision support metrics show prediction procedure as a binary operation (0 and 1) to distinguish good items from the items that are not good while statistical metrics evaluate the accuracy through comparing the predicted rankings or ratings directly with actual rating.

Deciding about the suitable metric depends on the dataset features and the type of tasks that the recommendation system will do. Based on our research and collected dataset we have evaluated our system through the two metrics, calculating the statistical accuracy through using Mean Absolute Error (MAE). The other metrics is done through performing a user experimental study to measure user satisfaction.

A. Mean Absolute Error (MAE)

It is the most commonly used metric in recommender systems [42]. This metric provides measures and illustrates the difference between the actual and estimated prediction over many items and users. It is given in (4):

$$MAE = \frac{\sum_{i=1}^d |a_i - p_i|}{d} \quad (4)$$

Where a is actual observation and p is predicted value divided by d which is the total number of actual ratings or ranks in an item set, the smaller value of MAE the better accuracy we get.

To calculate MAE, we have divided the dataset into training and testing sets. 70% of the data has been used as training set and the rest 30% of data has been used as test set, so variable R in Table I represents the train/test ratio that illustrates the percentage of data used in training set. We have compared MAE results across the three filtering approaches as shown in Table I.

Fig. 4 shows the comparison results of the three filtering approaches, we can observe that the greater value of R is the Lower value of MAE. Besides the most important thing we can

notice is that hybrid filtering has the best MAE results compared to the other two approaches.

B. User Experimental Study

We have created a random sample of 25 respondents (n=5) who are native Arabic Speakers from different educational levels. The respondents have created their user profiles manually by providing their interests in the six news categories to form the user profile. Every respondent has a web page to enter weights for every category and the weights must be sum to ten. We also have made three results sets from news collection for every respondent based on the three filtering approaches. The presentation order of news articles was random so the respondent could not realize which system recommended the articles. The respondent has asked to rate each article in a 3-point likert scale as it is relevant to his interest, somehow relevant or not relevant. When the respondent finishes, all information such as filtering approach, article's rank, article's weight, and respondent's rating are logged into a file. Fig. 5 shows a snapshot example of one of the respondents' response.

TABLE I. MAE COMPARISON RESULTS

Training Ration	Content-Based filtering	Collaborative filtering	Hybrid filtering
R=0.1	0.646	0.618	0.520
R=0.2	0.634	0.597	0.502
R=0.3	0.619	0.569	0.497
R=0.4	0.602	0.547	0.471
R=0.5	0.578	0.524	0.454
R=0.6	0.554	0.502	0.422
R=0.7	0.521	0.478	0.387

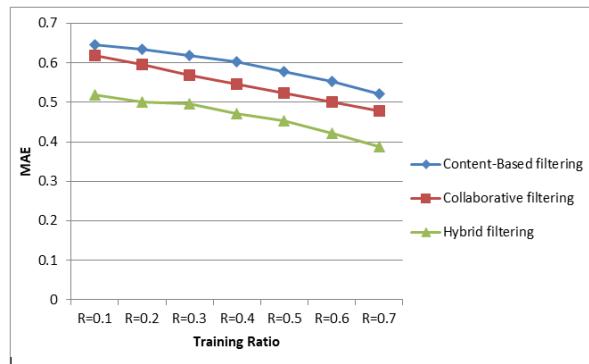


Fig. 4. MAE Comparison Result.

ID	S1 Rank	S1 WT	S2 Rank	S2 WT	S3 Rank	S3 WT	Resp Rating
122	70	0.16634	7	0.91926	20	0.15291	1
621	10	0.47541	138	0.00028	143	0.00013	0
540	25	0.37411	11	0.91634	6	0.34281	2
131	43	0.27787	6	0.92714	13	0.25763	1
618	3	0.68637	0	0	0	0	0
604	6	0.57957	33	0.69677	2	0.40383	2
439	37	0.32415	14	0.91517	10	0.29666	1
531	22	0.41517	5	0.92755	4	0.38509	2
117	8	0.50476	101	0.02954	96	0.01491	0
315	9	0.48671	52	0.68637	8	0.33407	1

Fig. 5. Example of Respondents' Response.

In Fig. 5, S1 Rank refer to where the article was ranked by the collaborative recommender and S1 WT indicate the normalized weight of that article, the same information is available for S2 (the content-based recommender) and S3 (the hybrid recommender). The Resp Rating column shows how the respondent rated each article, where 2 is for relevant, 1 is for somehow relevant and 0 for not relevant. To analyze the result we have calculated the average rank for all respondents over the three filtering approaches as shown in Table II.

TABLE II. AVERAGE RANK

	Not relevant	Somehow relevant	Very relevant
Collaborative	58.3	70.0	59.1
Content-Based	75.6	63.5	57.4
Hybrid	56.1	54.2	40.5

The table displays the average ranking of the top ten articles presented in each approach. For instance, the articles rated by the respondents as very relevant, the collaborative system ranked those articles on average as 59th. The content-based system ranked the very relevant articles 57th while the hybrid ranked them as 40th. So it is shown clearly the hybrid system ranked the articles higher than the other systems. This study supports MAE results where we have seen that hybrid system get less MAE value than other two systems.

VI. DISCUSSION AND CONCLUSION

This research aims to develop an Arabic recommender system to display news or articles to each reader based on their interests instead of presenting them only in the order of their occurrence using Twitter. To achieve that, we started with collecting the news articles and the dataset of tweets from Twitter and preprocessing them to prepare them to be used in building the recommender system. Then we explained the design and implementation of all the components of the proposed recommender system. Lastly, we have evaluated the recommender system through the two metrics, calculating the statistical accuracy through using Mean Absolute Error (MAE). The other metrics is done through performing a user experimental study to measure the user satisfaction.

Based on the result of the evaluation we found that the hybrid system performs better than the collaborative and content-based recommendations. This means that news articles recommended by the hybrid recommender system are more relevant to the user compared to the other two systems.

Although this research has successfully achieved the goal of developing a personalized Arabic news recommender system, it has certain limitations that are inherent in most of the recommender system researches. First, the size of the created dataset was relatively small, which imposes the need to evaluate the robustness of proposed system against large datasets. Second, creating a new application for Twitter developer account was another challenge. Setting up a Twitter API wasn't easy as it used to be, we have tried several times with Twitter team until we got final application approval.

In future, we have many directions to enhance this research. The accuracy of the proposed news recommender system can

be further improved through incorporating other features such as location. In this case the system can recommend news articles based on the user's geographical location. We can increase the efficiency of the system on a wider range of Arabic news articles in order to scale the system and enhance its algorithm so it can apply in different categories of news articles.

REFERENCES

- [1] Y. Wang, D. Yin, L. Jie, P. Wang, M. Yamada, Y. Chang and Q. Mei, "Beyond Ranking: Optimizing Whole-Page Presentation," in Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, 2016.
- [2] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," in Proceedings of the 10th International Conference on World Wide Web, Hong Kong, 2015.
- [3] S. Kemp, "Digital 2020 Global Overview Report," The Next Web, Amsterdam, 2020.
- [4] K. Saranya and S. Sudha, "A Personalized Online News Recommendation System," International Journal of Computer Applications, pp. 6-7, 2012.
- [5] F. Garcin, C. Dimitrakakis and B. Faltings, "Personalized news recommendation with context trees," in Proceedings of the 7th ACM Conference on Recommender Systems, Hong Kong, 2013.
- [6] C. García , V. García-Díaz, D. Meana-Llorián and E. Núñez-Valdez , "Social Recommender System: A Recommender System Tweets for points of interest," in 4th Multidisciplinary International Social Networks Conference, Bangkok, 2017.
- [7] J. Konstan and J. Riedl, "Recommender systems: from algorithms to user experience," User Modeling and User-Adapted Interaction, pp. 101-103, 2012.
- [8] C. Pan and W. Li, "Research paper recommendation with topic analysis," in 2010 International Conference On Computer Design and Applications, Qinhuangdao, 2010.
- [9] F.O. Isinkaye, Y.O. Folajimi and B.A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," Egyptian Informatics Journal, pp. 261-263, 2015.
- [10] B. Bhatt, P. J. Patel and H. Gaudani, "A Review Paper on Machine Learning Based Recommendation System," International Journal of Engineering Development and Research, p. 3955, 2014.
- [11] P. Pu, L. Chen and R. Hu, "A user-centric evaluation framework for recommender systems," in RecSys '11: Proceedings of the fifth ACM conference on Recommender systems, New York, 2011.
- [12] R. Hu and P. Pu, "Acceptance issues of personality-based recommender systems," in RecSys '09: Proceedings of the third ACM conference on Recommender systems, New York, 2009.
- [13] J. Herlocker, J. Konstan, L. Terveen and J. Riedl, "Evaluating collaborative filtering recommender systems," Association for Computing Machinery, pp. 5-7, 2004.
- [14] C. C. Aggarwal, Recommender Systems, Switzerland : Springer, Cham, 2016.
- [15] J. Bobadilla, F. Ortega, A. Hernando and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, pp. 109-111, 2013.
- [16] S. Bostandjiev, J. O'Donovan and T. Höllerer, "TasteWeights: a visual interactive hybrid recommender system," in RecSys '12: Proceedings of the sixth ACM conference on Recommender systems, New York, 2012.
- [17] A. Popescul, L. H. Ungar, D. M. Pennock and S. Lawrence, "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments," arXiv, p. 437, 2013.
- [18] G. Adomavicius and J. Zhang, "Impact of data characteristics on recommender systems performance," ACM Transactions on Management Information Systems, pp. 1-3, 2012.
- [19] D. H. Stern, R. Herbrich and h. Graepel, "Matchbox: large scale online bayesian recommendations," in Proceedings of the 18th international conference on World wide web, New York, 2009.

- [20] C. Wei-Ta and T. Ya-Lun , "A hybrid recommendation system considering visual information for predicting favorite restaurants," Springer US, New York, 2017.
- [21] K. Pigi , S. James , P. Jay and O. John , "User Preferences for Hybrid Explanations," in RecSys '17 Proceedings of the Eleventh ACM Conference on Recommender Systems, Como, 2017.
- [22] A. Flodin, "A scalable product recommendation engine suitable for transaction," Mid Sweden University, p. 4, 2018.
- [23] J. Jens, "Interactive television: new genres, new format, new content," in Proceedings of the second Australasian conference on Interactive entertainment, Sydney, 2015.
- [24] C. Akshay, "Recommender System for News Articles Using Supervised Learning," Universitat Pompeu Fabra, Barcelona, 2014.
- [25] B. Fortuna, C. Fortuna and D. Mladenović, "Real-Time News Recommender System," in Proceedings of the 2010 European conference on Machine learning and knowledge discovery in databases, Barcelona, 2010.
- [26] A. Alharsh, "Computer and Learning Arabic language," Mentouri University, pp. 217-218, 2007.
- [27] N. Madi and H. Al-Khalifa, "Error Detection for Arabic Text Using Neural," MDPI, pp. 2-3, 2020.
- [28] A. Wahhab and A. Hassan, "Proposed aspect extraction algorithm for Arabic text reviews," Journal of AL-Qadisiyah for computer science and mathematics, p. 80, 2018.
- [29] B. Hawashin, S. Alzubi and T. Kanan, "An efficient semantic recommender method," Emerald Insight, p. 263, 2019.
- [30] A. Farghaly and K. Shaalan, "Arabic Natural Language Processing: Challenges and Solutions," ACM, pp. 1-2, 2009.
- [31] P. Jackson , "Web 2.0 tools and context," in Web 2.0 Knowledge Technologies and the Enterprise, Manchester, Chandos Publishing, 2010, pp. 11-12.
- [32] J. An, M. Cha, K. Gummadi and J. Crowcroft, "Media Landscape in Twitter: A World of New Conventions and Political Diversity," in Proceedings of the International AAAI Conference on Web and Social Media, Barcelona, 2021.
- [33] T. Rosenstiel, J. Sonderman, K. Loker, M. Ivancin and N. Kjarval, "Twitter and the News: How people use the social network to learn about the world," American Press Institute, U.S, 2015.
- [34] D. Surabhi, "Recommender system in education," in 5th National Conference on E-Learning & E-Learning Technologies, India, 2017.
- [35] S. Agarwal, "Classification of RSS feed news items using ontology," in 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi, 2012.
- [36] O. Ossama, "CAMEL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing," in Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, 2020.
- [37] M. Azam, T. Ahmed, F. Sabah and M. I. Hussain, "Feature Extraction based Text Classification using K-Nearest Neighbor Algorithm," IJCSNS International Journal of Computer Science and Network Security, vol. 18, no. 12, pp. 95-100, 2018.
- [38] C. Gormley and Z. Tong, "CHAPTER 1: You Know,For Search," in Elasticsearch The Definitive Guide: A Distributed Real-Time Search and Analytics Engine, Sebastopol, O'Reilly Media, 2015, pp. 3-4.
- [39] M. B. Magara, S. O. Ojo and T. Zuva, "A comparative analysis of text similarity measures and algorithms in research paper recommender systems," in 2018 Conference on Information Communications Technology and Society (ICTAS) , Durban, 2018.
- [40] Suriati1, M. Dwiaستuti and Tulus, "Weighted Hybrid Technique for Recommender System," in International Conference on Information and Communication Technology (IconICT), 2017.
- [41] F. Isinkaye, B. Ojokoh and Y. Folajimi, "Recommendation systems: Principles, methods and," Egyptian Informatics Journal, p. 270, 2015.
- [42] F. Harrag, A. Al-Salman and A. Alquahtani, "Arabic Opinion Mining Using a Hybrid Recommender System Approach," ArXiv, pp. 17-18, 2020.