Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 117−124
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5761

Optimizing News Recommendations: Utilizing POS-Tagging and Content-Based Methods to Enhance Personalization in News Recommendations

Arga Kusuma Wiratama, Z K A Baizal*

Faculty of Informatics, Departement of Informatics, Telkom University, Bandung
Jl. Telekomunikasi. 1, Terusan Buahbatu - Bojongsoang, Telkom University, Sukapura, Kec. Dayeuhkolot, Kabupaten Bandung, Jawa Barat, Indonesia
Email: [1]argakw@student.telkomuniversity.ac.id, [2,*]baizal@telkomuniversity.ac.id
Correspondence Author Email: baizal@telkomuniveristy.ac.id
Submitted: **09/08/2024;** Accepted: **05/09/2024;** Published: **14/10/2024**

Abstract−Access to information continues to experience significant developments. With the rapid advancement of the internet, the amount of news content available on digital platforms is also increasing rapidly. Internet users can quickly and easily access news and information from various sources. However, this also brings new challenges for internet users, especially digital news readers. With the vast amount of available news, readers often receive news recommendations that are irrelevant to their interests. This is due to the different preferences of each user. Additionally, each user may have more than one preference, leading to the appearance of random and unwanted news recommendations. Therefore, this research aims to enhance the personalization of news recommendations by utilizing POS-Tagger technology to analyze news content. Additionally, the content-based filtering method is used to match news with user preferences based on previously consumed content. The news matching is done after calculating vectors using TF-IDF, followed by matching using cosine similarity calculation. The recommender system demonstrates a good ability to provide recommendations that are relevant to user preferences. The performance evaluation showed satisfactory results. F1-score showed an average result of 90% from the three users, and high cosine similarity value with an average from the three users of 8% of the overall recommendation results indicating a high relevance between the recommendations and the news that users have read.

Keywords: Content-Based Filtering; Cosine Similarity; News Recommender System; Pos-Tagger; TF-IDF

1. INTRODUCTION

In the rapidly evolving digital age, access to information has undergone significant development, Internet has helped people to access news and information from various sources quickly and easily. However, continuously increasing amounts of news content available on digital platforms has also brought new challenges. Users often find it challenging to sift through and find news that matches their interests and preferences amidst the abundance of information. To provide more personalized and relevant experience for users, news recommender systems emerged as a promising solution[1]. In this system, we propose analytical technologies such as POS-Taggers, to analyze user's news preferences. POS-Tagger is a natural language processing tool capable of identifying and labeling key parts of sentences, such as nouns, verbs, and adjectives. Taggers are commonly used in applications or blogs to allocate data[2]. By analyzing the sentence structure of news content, POS-Taggers can be utilized to better understand the content and context of articles. In addition, content-based filtering approach is also one of the methods of news recommender systems. The focus of content-based filtering in recommender systems is on the properties of the item itself[3]. Content-based filtering considers the characteristics of the news content itself including keywords, topics, and language used. In content-based filtering, users or items are considered as atomic units to create more personalized recommendations[3]. By analyzing the news content before it is provided to users, the systems can match the latest news with the user's preferences based on the content they have read before[4]. Combination of POS-Tagger and content-based filtering method in a news recommender system has the potential to help users find news that matches their interests, reduce workload of sorting news, and introduce diverse news viewpoints. Therefore, the system not only improves the quality of the user's experience, but also ensures that they stay informed according to their personal preferences. Thus, the purpose of this research is to evaluate the relevance of news recommendations to users. By utilizing POS-Tagger to understand the sentence structure and news content and applying content-based filtering to consider the characteristics of the content itself, can provide news recommendations that match the interests and preferences of users.

News recommendations systems are created to make it easier for users to choose news that matches the user's preferences. A recommender system can be distinguished from an information retrieval system based on its user interaction semantics[5]. The results of a recommender system are understood as recommendations, options worth considering; the results of an information retrieval system are interpreted as matches to the user's query. Numerous studies on news recommender systems have emerged using various approaches, systems have emerged using various approaches, such as content-based filtering[2], collaborative filtering[6], and knowledge graph[7] with the aim of building a news recommender systems that can provide news according to user's preferences. (Darvishy et al., 2020), proposing news recommendation framework, HYPNER, using a combination of collaborative filtering and content-based filtering. This framework can solve the problem of news clustering, diverse users modeling, news ranking, and diverse news selection. A personalized news recommendation

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 117−124
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5761

framework, HYPNER, has been proposed and the results showed that HYPNER achieved 81.56% improvement in terms of F1-score and 5.33% in terms of diversity as compared to an existing recommender system called SCENE[8]. (Raza et el., 2020), emphasized that a news recommender system is not only measured by accuracy in providing recommendations for users, but also paying attention to the diversity of recommendations. The system builds using a combined method of latent factor models with Generalized Linear Model (GLM) regression with the aim of making an accurate and diverse news recommender system[9]. Then the research conducted by (Amara et al., 2020), the use of POS-Tagger is utilized in the system built. POS-Tagger is implemented in the dataset for data categorization. A user profile model is also created based on tags that have been created from the dataset to provide better recommendations. The model shows improved precision when compared to personalize model and content-based[2]. In addition to building a recommender system to match user preferences, a user model needs to be created to store data on articles that have been read. The data has an influence because it relates to the user's own preferences. (Zhu et al., 2017), proposed a new approach to personalized news recommendation by introducing a user profile model that considers user preferences from multiple perspectives. They used the BAP (Behavior and Popularity) method, which assigns weights to historical news based on user behavior and news popularity. This method improves the accuracy of user profiling, especially in short-term profiling, where a time function is used to adjust preferences for all historical news, instead of focusing only on recent interactions[10]. From these studies, the development of news recommender system is a tool that can help users in choosing news according to their respective preferences. With so much news available, this system will help news readers not to bother looking for or choosing news to read while maintaining the diversity of the news provided.

To help solve the problem we are facing, we propose the method using POS-Tagger and content-based filtering. Part-Of-Speech Tagger, or commonly abbreviated as POS-Tagger is a part of Natural Language Processing (NLP) that aims to assign a label to each word in sentence or text, POS-Tagger is an important tool in many natural language processing applications[11]. POS-Tagger is widely used on websites to perform data categorization and search function[2]. In the recommender system, POS-Tagger is used to analyze and extract text, then give a label for each word or token[2], [4]. The labeling process will help in identifying necessary features such as content or title[4]. The use of POS-Tagger here is relevant to the topic of news recommender system, because the data obtained will be in the form of textual content. Other than that, a deep understanding of news articles through NLP techniques is essential for news recommendation. Both effective text representation methods and pre-trained language models can contribute to the improvement of news recommendation performance[12]. In the research that has done, the use of POS-Tagger makes the process more efficient in labeling words or tokens[2], [4], [13]. In a recommender system, content-based filtering provides recommendations by comparing the features of items[3]. The systems analyze documents or preferences given by a particular user and attempt to build a model around this data[14]. A common approach for this content-based is to represent users and items in the same feature, then the similarity score between users and items can be calculated, recommendations will be made based on the similarity score between users and items[6]. Users can be represented by a user profile. User profile creation is an important process in a personalized news recommendation system, where the system tries to understand the preferences of individuals based on their past activities[15]. One way to represent items in a feature is by vectorization. The vectorization calculation itself is done using TF-IDF (Term Frequency-Inverse Document Frequency)[16]. The TF-IDF algorithm is used to evaluate the importance of words in a text corpus, where the importance of a word will be directly proportional to the number of occurrences of the word in the document and inversely proportional to the frequency of occurrence of the word in the corpus[17]. In the case of a news recommender system, the system here will calculate the similarity between the newly published news and the content-based profile of the user and then give a similarity score[8]. To calculate the similarity between the two items, one way is the cosine similarity method. Cosine similarity between two objects measures the cosine angle between the two objects. It compares two documents on a normalized scale, which can be done by finding the dot product between two identity vectors[16]. The use of content-based methods has several advantages such as, users can build their own preference profiles as well as being able to recommend items that have never been placed by any user, this is an advantage for new users who have not built their preference profiles[18].

2. RESEARCH METHODOLOGY

2.1 Research Stages

We developed research flow to help the process of conducting this research. This flow describes the workflow of the research.

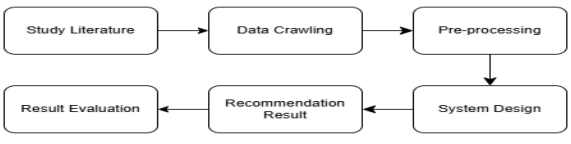

Figure 1. Research flow

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 117−124
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5761

From figure 1 above, in the study literature stage, we collected research journals, books, or articles written by other researchers related to the problem and methods in this research. This way, we obtained relevant theories as references to help address the problems in this research. For the news recommender system we developed, we used a dataset obtained through data crawling. We conducted data crawling on the online news website Detik.com, which is one of the commonly accessed news websites by readers in Indonesia. Afterward, we applied a pre-processing stage to the obtained dataset, including lowercasing, tokenization, and normalization. We used python programming language for the system design stage. This system design consists of three processes. First is POS-Tagging for each word in the news content. Then, calculate vector values for each tagged word using TF-IDF vectorization. The final process is the content-based filtering method, where we use cosine similarity to compare the vector values of similar words in the news read by user with the entire news data. The resulting news recommendations are based on the news articles in the user's data, representing the news read by users. Then we evaluate the recommendation results using accuracy, precision, recall, and F1-score metrics.

2.2 System Design

The following is the system flow on the recommender system that we built. The figure below shows how the process in building the system.

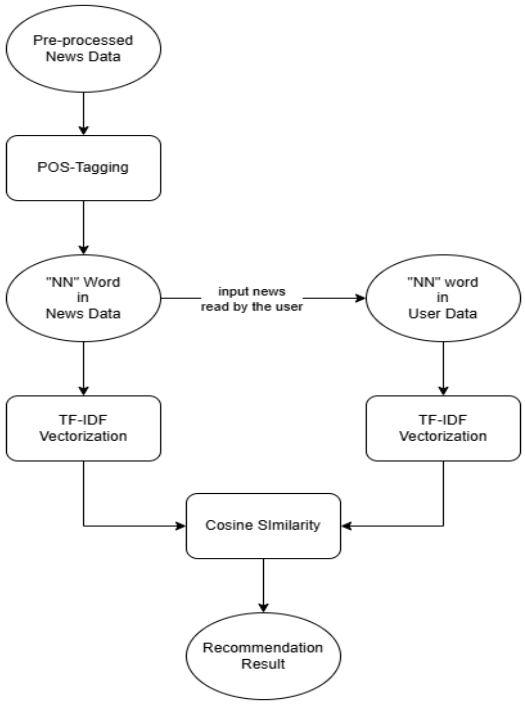

Figure 2. System flow

After going through the pre-processing stage, the contents of the content column in the news data will be tagged using POS-Tagger. We use the CRFTagger library from NLTK and the Indonesian POS-Tagger model that has been previously trained[19]. In this process, each word will have its own tag. Then feature extraction is done so that only words with the tag "NN" (NOUN) or nouns will be used in the next process. The selection of nouns is because they have a more stable meaning when compared to adjectives or verbs. The collection of nouns in the form of a list will be united into one whole text for each news. This process is done to facilitate the vectorization process.

Noun_Tokens	Noun_Text
['ketua', 'partai', 'demokrat', 'yudhoyono', '...	ketua partai demokrat yudhoyono koalisi prabow...
['kota', 'batu', 'pengalaman', 'belanja', 'kul...	kota batu pengalaman belanja kulineran pasar a...
['pelarian', 'engelbertus', 'lowa', 'soda', 'f...	pelarian engelbertus lowa soda frater calon pa...
['teknologi', 'nopember', 'its', 'kampus', 'su...	teknologi nopember its kampus surabaya berikut...
['calon', 'jemaah', 'umrah', 'asal', 'tejoasri...	calon jemaah umrah asal tejoasri laren lamonga...
...	...
['presiden', 'joko', 'widodo', 'ketum', 'nasde...	presiden joko widodo ketum nasdem surya paloh ...
['powerboat', 'danau', 'toba', 'akhir', 'pekan...	powerboat danau toba akhir pekan logistik bala...
['guru', 'garda', 'terdepan', 'edukasi', 'masy...	guru garda terdepan edukasi masyarakat literas...
['kepolisian', 'as', 'bocah', 'tahun', 'wilaya...	kepolisian as bocah tahun wilayah utah penemba...
['industri', 'pemain', 'produk', 'fitur', 'hal...	industri pemain produk fitur hal produsen mobi...

Figure 3. POS-Tagging result

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 117−124
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5761

User data is created after getting the results of the POS-Tagging process. This data represents users who have read news articles. We created three users with each user having read 200 news articles. The news articles are randomly inserted into the user data taken from the news data that has gone through the POS-Tagging process. This randomization is intended to represent users who have more than one preference. In this way, we can assess the recommendation results for users who read more than one type of news preference. After building the user data, we proceed with the vectorization process. We use the TF-IFP library to perform vectorization. In this process, each word will have its vector value calculated based on how often it appears in other news articles. After obtaining vector results for each news article in the news data and user data, the vectors will then be compared using cosine similarity. We set a threshold of 0.6 for the minimum cosine similarity score, so that only recommendation results with scores equal to or above the threshold are displayed. Furthermore, we classify recommendation results that are highly relevant to user preferences with a score of ≥ 0.9.

3. RESULT AND DISCUSSION

3.1 Data Crawling

The first process we did was data collection. Data collection was done by data crawling on the online news website Detik.com[20]. With this method, we obtained 6707 news articles consisting of news headlines, news publication dates, news links, and news content covering themes such as politics, economy, sports, automotive, health, technology, and traveling. From the news data obtained, we only use the news title as identification to ensure there are no duplicate news articles and news content, which we then process in the pre-processing stage. Figure 2 below shows an example of the data we obtained through data crawling.

News Data

	Judul	Tanggal	Link	Konten
0	Hari Terakhir, KPU Ingatkan Peserta Pemilu Sam...	29 Feb 2024 16:46	https://news.detik.com/pemilu/d-7218501/hari-t...	Komisi Pemilihan Umum (KPU) RI mengingatkan pe...
1	Relawan Ganjar Kritik Kursi Komisaris BUMN unt...	29 Feb 2024 16:09	https://news.detik.com/pemilu/d-7218387/relawa...	Ketua Umum Projo Ganjar, Haposan Situmorang, m...
2	Caleg Artis, Modal Sosial, dan Popularitas	29 Feb 2024 15:40	https://news.detik.com/kolom/d-7216558/caleg-a...	Penerapan sistem pemilu proporsional terbuka s...
3	Curahan Hati Atalia Usai Raih Suara Terbanyak ...	29 Feb 2024 15:30	https://www.detik.com/jabar/berita/d-7218070/c...	Atalia Praratya maju sebagai calon legislatif ...
4	Cuma Dapat 1.849 Suara, Anak Amien Rais Teranc...	29 Feb 2024 15:25	https://www.detik.com/jogja/berita/d-7218273/c...	Putra dari tokoh politik Amien Rais, yakni Ahm...
...	...	...	...	...
6702	Eep Saefulloh Launching 'Warga Jaga Suara', Aj...	09 Feb 2024 17:14	https://news.detik.com/pemilu/d-7185096/eep-sa...	CEO PolMark Research Center Eep Saefulloh Fata...
6703	Kemenko Perekonomian Buka Lowongan Kerja buat ...	09 Feb 2024 16:30	https://finance.detik.com/lainnya/d-7184785/ke...	Kementerian Koordinator Bidang Perekonomian (K...
6704	Mengenal Ibnu Sina: Metode Pengobatan dan Kary...	09 Feb 2024 16:00	https://www.detik.com/hikmah/khazanah/d-718371...	Nama Ibnu Sina tidak hanya terkenal di kalanga...
6705	Jurus Bank BUMN Geber Kualitas Pendidikan di RI	09 Feb 2024 13:30	https://finance.detik.com/moneter/d-7184569/ju...	Guru merupakan salah satu garda terdepan untuk...
6706	Relawan Cek Kesehatan Warga, Usul Jadi Bagian ...	09 Feb 2024 12:16	https://news.detik.com/pemilu/d-7184648/relawa...	Relawan pendukung Prabowo Subianto-Gibran Raka...

Figure 4. Data crawling result

3.2 Pre-processing

The pre-processing stage is carried out on the content of the news data that has been obtained. This process consists of several stages such as lowercasing which converts all letters into lowercase letters. Then tokenization to break each text into smaller units. Followed by the normalization process to remove punctuation marks. After that the tokens are put back together into one whole text. This pre-processing is done to facilitate the next tagging process. Below is an example of the results of the pre-processing stage.

Original News Data and Pre-processed News Data

	Original Data	Pre-processed Data
0	Komisi Pemilihan Umum (KPU) RI mengingatkan pe...	komisi pemilihan umum (kpu) ri mengingatkan ...
1	Ketua Umum Projo Ganjar, Haposan Situmorang, m...	ketua umum projo ganjar , haposan situmorang ,...
2	Penerapan sistem pemilu proporsional terbuka s...	penerapan sistem pemilu proporsional terbuka s...
3	Atalia Praratya maju sebagai calon legislatif ...	atalia praratya maju sebagai calon legislatif ...
4	Putra dari tokoh politik Amien Rais, yakni Ahm...	putra dari tokoh politik amien rais , yakni ah...

Figure 5. Pre-processing result

3.3 Recommendation Result

At this stage, recommendations are made for three users based on the results of cosine similarity. Each user has 200 different news articles. From the obtained recommendation, accuracy, precision, recall, and F1-Score will be calculated[21]. Following table 1 is representation of confusion matrix that we used.

Table 1. Recommendation result label

	Read by user	**Not read by user**
Cosine score ≥ 0,9	True Positive (TP)	False Positive (FP)
Cosine score < 0,9	False Negative (FN)	True Negative (TN)

The Following figure 6 and figure 7 are recommendation results and evaluation results for User1 respectively.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 117−124
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5761

```
User1
Recommendation :
Jokowi Rajin Bagi-bagi Bansos, Luhut: Ngapain Sih Ribut? - Score: 0.9894792230150667
Gunung Marapi Kembali Erupsi Pagi Ini, Ketinggian Kolom Abu Tidak Teramati - Score: 0.987297081731191
Pj Bupati Pasuruan Ungkap Duduk Perkara Wajah Gus Irsyad di Cup Kopi Dicoret - Score: 0.9826971549238688
Bela Jokowi soal Rajin Bagi-bagi Bansos, Luhut: Ngapain Sih Ribut? - Score: 0.9801106156429479
Gunung Marapi Kembali Erupsi Selasa Pagi Ini - Score: 0.9798244616405224
Jokowi Rajin Bagi-bagi Bansos, Luhut: Ngapain Sih Ribut? - Score: 0.9776403178183639
Program Makan Siang dan Susu Gratis Dibahas Rapat Kabinet Hari Ini - Score: 0.9757928406585745
Duh! Puluhan Penumpang Ditinggal Whoosh gegara KA Feeder Terlambat - Score: 0.9736183818245367
Bela Jokowi soal Rajin Bagi-bagi Bansos, Luhut: Ngapain Sih Ribut? - Score: 0.971307820041726 2
Bertemu Surya Paloh, Jokowi: Saya Ingin Jadi Jembatan untuk Semua - Score: 0.9712097051627055
```

Figure 6. Top 10 recommendations for User1

```
User1
Cosine Score >= 0.9: 33
Cosine Score < 0.9: 294
Accuracy: 0.9310986964618 25
Precision: 0.8619246861924686
Recall: 0.9809523809523809
F1-Score: 0.9175946547884187
```

Figure 7. Evaluation results for User1

User1 received 327 news recommendations with 33 articles having a score of ≥ 0,9 which means it is highly relevant with User1 preferences. Below are figures 7 showing comparison of cosine similarities score between articles read by User1 and recommendation result for User1.

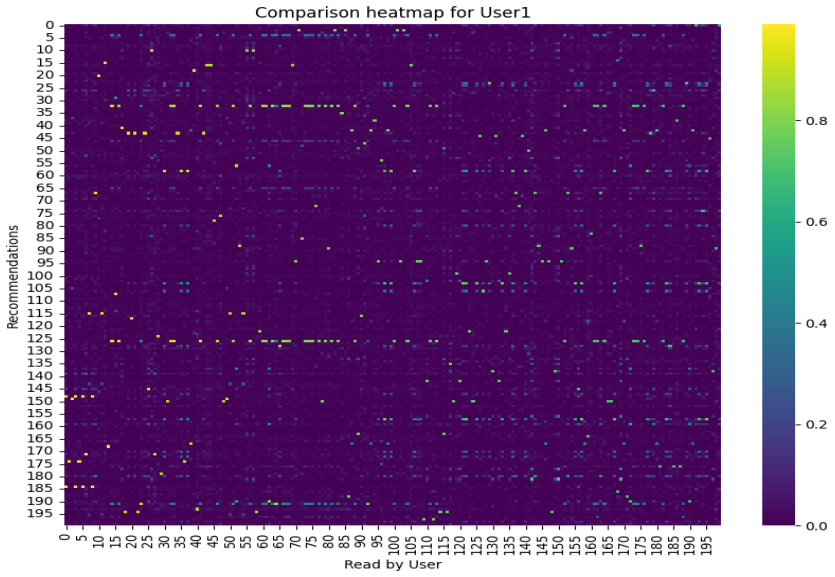

Figure 8. Result comparison for User1

In the figure 8 above, showing the brighter dots are scattered, and some form horizontal patterns that indicating single recommendations are influenced by multiple articles that have been read. The darker dots indicate that the recommendation at those indexes is not influenced by the articles from the corresponding index. Next is the result for User2. The following figures 8 and 9 show the recommendation results and evaluation results respectively.

```
User2
Recommendation :
BPJS Jadi Syarat Urus SKCK, Gimana Jika Belum Daftar atau Statusnya Non Aktif? - Score: 0.9800032170803203
Gus Ipul Sebut Pemakzulan Presiden Bukanlah Tradisi NU - Score: 0.9730892483168724
Kriuk Renyah! Camilan Belalang Goreng dari Ujungjaya Sumedang - Score: 0.9659622538493131
Bulog Buka Lowongan Kerja buat Lulusan SMA hingga S1, Ini Syaratnya - Score: 0.9609925777071291
Alasan Prabowo Tak Pilih Sri Mulyani Jadi Menkeu, Ini Bocorannya - Score: 0.9608069094912315
Pertunangan Nathalie Holscher dengan Ladislao Camara - Score: 0.9591299262627638
Bukan Sri Mulyani, Ini Bocoran Menkeu yang Dibidik Prabowo - Score: 0.9580359273109854
TikToker Ini Keracunan gegara Nekat Cicip Belalang dan Ulat Sagu - Score: 0.9550007231023634
Primadona Baru dari Sumedang: Wisata Air Cigirang - Score: 0.9400665028142285
Marissya Icha Polisikan Oknum Travel karena Jadi Sasaran Dituding Penipu - Score: 0.9343826898764084
```

Figure 9. Top 10 recommendations for User2

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 117−124
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5761

```
User2
Cosine Score >= 0.9: 18
Cosine Score < 0.9: 232
Accuracy: 0.9516483516483516
Precision: 0.9178082191780822
Recall: 0.9804878048780488
F1-Score: 0.9481132075471698
```

Figure 10. Evaluation results for User2

User2 received 250 news recommendations with 18 articles having a score of ≥ 0,9 which means it is highly relevant with User2 preferences. The following figure 10 shows the comparison of cosine similarities between articles read by User2 and recommendation results for User2.

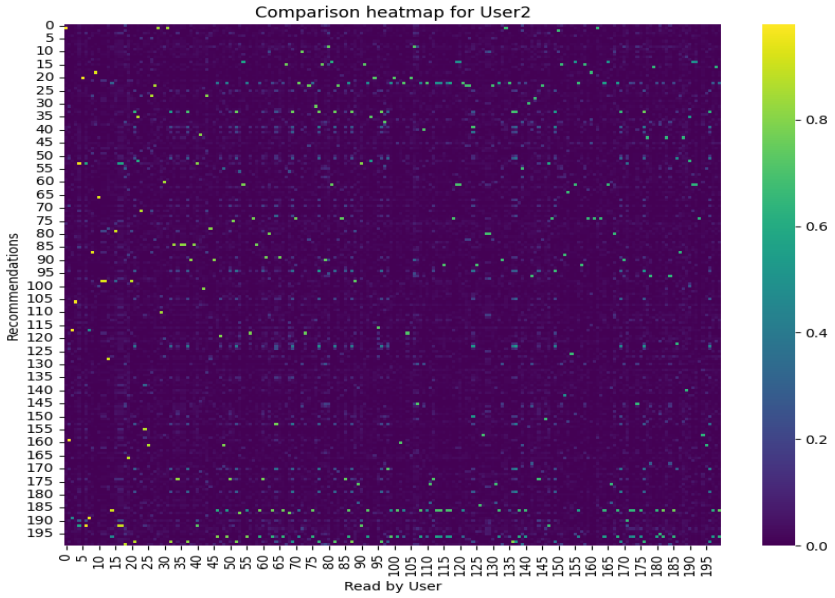

Figure 11. Result comparison for User2

Figure 11 above shows the comparison of cosine similarity scores between the news read by User2 and the results of news recommendations for User2. Compared to the comparison results of User1, the dot pixels with bright colors in User2 are more scattered, indicating that the themes of the news articles read by User2 are more evenly distributed. In addition, there are some horizontal patterns in the recommendations with a low index, which indicates that the recommendations are influenced by several articles that have been read.

Next is the result for User3. The following figures 12 and 13 show the recommendation results and evaluation results respectively.

```
User3
Recommendation :
9 Barang yang Tidak Boleh dan Tak Perlu Dimasukkan ke Kulkas - Score: 0.9994126633033673
Perbedaan Quick Count, Real Count dan Exit Poll dalam Pemilu - Score: 0.9936996665432626
Muhammadiyah Akan Pakai Kalender Hijriah Global Jadi Acuan Lebaran 2025 - Score: 0.9842784249064294
Mengenal Pulau Asu yang Indah di Nias Barat, Pecinta Surfing Wajib Datang! - Score: 0.9825816540085478
Ini Bacaan Doa Malam Nisfu Syaban, Yuk Amalkan! - Score: 0.9802005762942018
4 Bacaan Doa Malam Nisfu Syaban Lengkap Arab, Latin, dan Artinya - Score: 0.9777971138272012
Ngeri! Data Menunjukkan Selfie Lebih Mematikan Dibanding Serangan Hiu - Score: 0.9746105693250455
Jadi Korban AI, Melaney Ricardo Bakal Ambil Langkah Hukum Jika Terus Dirugikan - Score: 0.9712927898859289
Jokowi soal Ketemu Paloh: Baru Awal, Saya Ingin Jadi Jembatan untuk Semua - Score: 0.9710745440386411
Segini Gaji AHY yang Sudah Resmi Jadi Menteri ATR - Score: 0.9687818206809486
```

Figure 12. Top 10 recommendations for User3

```
User3
Cosine Score >= 0.9: 25
Cosine Score < 0.9: 331
Accuracy: 0.915807560137457
Precision: 0.8898678414096917
Recall: 0.8938053097345132
F1-Score: 0.891832229580574
```

Figure 13. Evaluation results for User3

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 117−124
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5761

User3 received 356 news recommendations with 25 articles having a score of ≥ 0,9 which means it is highly relevant with User3 preferences. The following figure 14 shows the comparison of cosine similarities between articles read by User3 and recommendation results for User3.

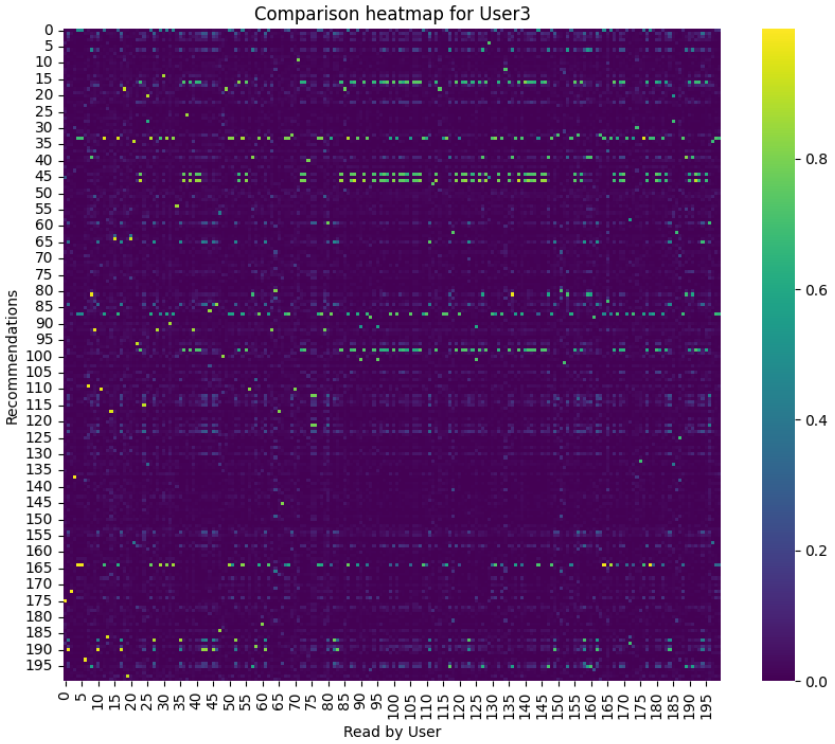

Figure 14. Result comparison for User3

In Figure 14, the comparison results show an increase in the occurrence of horizontal patterns when compared to the results from User1 and User2. This shows that the types of news articles read by User3 are less varied when compared to User1 and User2.

3.4 Discussion

Based on the test results, the system is able to produce a number of recommendations with a high cosine similarity score (≥ 0.9) which indicates that the recommendations are very similar to the news articles that have been read. In addition, the evaluation results from each user also show good results. The recommendations generated are strongly influenced by the articles that have been read by users. The appearance of brightly colored dot pixels that form a horizontal pattern in the heatmap graph indicates if the recommendation has similarities with the news in each user's data represented by its index. The variation in the types of news articles read by users is shown by the spread of dot pixels as depicted in Figure 11, where the more spread out means that the types of news articles read are more evenly distributed. And vice versa, if more dot pixels form a horizontal pattern, it shows that the distribution of news articles read is uneven or in other words there are news article themes that are more numerous than other news article themes as depicted in Figure 14. The distribution of article variations in user preferences also affects the number of recommendations given. User2, who has a more scattered dot pixel result on the heatmap graph has a smaller number of recommendation results compared to User3 whose heatmap graph results show many dot pixels forming a horizontal pattern. Besides that, it also influences the results of the evaluation matrix score, where the more even the distribution of the types of news read by users, the higher the evaluation matrix score. This is related to the number of recommendation results obtained, because the fewer recommendation results obtained, the comparison when calculating the evaluation matrix will be smaller.

4. CONCLUSION

Based on the analysis conducted, it can be concluded that the news recommender system built is quite effective in providing relevant news recommendations to users based on news articles that have been read. The system can generate recommendations with a high cosine similarity score for each user which indicates that the recommendations are very similar to the news articles that have been read. The recommendations generated are

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 117−124
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5761

highly influenced by the articles that have been read by users, as dot pixels that form horizontal patterns appear on the heatmap graph, indicating the similarity between recommendations and some articles that have been read. The distribution of user preferences can also be seen from the distribution of brightly colored dot pixels. Users with more varied news preferences show a more even distribution of dot pixels, while more dot pixels forming horizontal patterns indicate fewer news preferences. In addition, users who read articles with a more even distribution of themes tend to get more diverse recommendations, while users who have a bias towards one preference will get more recommendations. For further development, it is recommended to improve the text processing algorithm with more advanced techniques such as word embeddings, test the system on larger and more diverse datasets, and consider additional data such as user interactions and demographics to provide more personalized recommendations. Evaluation with other metrics such as Jaccard similarity or Dice coefficient can also be considered to see potential improvements in recommendation quality. With these conclusions and suggestions, news recommender systems can be further developed to provide a better and more relevant user experience.

REFERENCES

[1] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi, "Recommender Systems Leveraging Multimedia Content," ACM Comput. Surv., vol. 53, no. 5, 2020, doi: 10.1145/3407190.
[2] S. Amara and R. R. Subramanian, "Collaborating personalized recommender system and content-based recommender system using TextCorpus," 2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020, pp. 105–109, 2020, doi: 10.1109/ICACCS48705.2020.9074360.
[3] C. Feng, M. Khan, A. U. Rahman, and A. Ahmad, "News Recommendation Systems-Accomplishments, Challenges Future Directions," IEEE Access, vol. 8, pp. 16702–16725, 2020, doi: 10.1109/ACCESS.2020.2967792.
[4] Rohit and A. K. Singh, "Accuracy enhancement of collaborative filtering recommender system for blogs using latent semantic indexing," 2017 Conf. Inf. Commun. Technol. CICT 2017, vol. 2018-April, pp. 1–4, 2017, doi: 10.1109/INFOCOMTECH.2017.8340646.
[5] R. Burke, "Hybrid web recommender systems," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 4321 LNCS, pp. 377–408, 2007, doi: 10.1007/978-3-540-72079-9_12.
[6] Z. Lu, Z. Dou, J. Lian, X. Xie, and Q. Yang, "Content-based collaborative filtering for news topic recommendation," Proc. Natl. Conf. Artif. Intell., vol. 1, pp. 217–223, 2015, doi: 10.1609/aaai.v29i1.9183.
[7] D. Lee, B. Oh, S. Seo, and K. H. Lee, "News Recommendation with Topic-Enriched Knowledge Graphs," Int. Conf. Inf. Knowl. Manag. Proc., pp. 695–704, 2020, doi: 10.1145/3340531.3411932.
[8] A. Darvishy, H. Ibrahim, F. Sidi, and A. Mustapha, "HYPNER: A Hybrid Approach for Personalized News Recommendation," IEEE Access, vol. 8, pp. 46877–46894, 2020, doi: 10.1109/ACCESS.2020.2978505.
[9] S. Raza and C. Ding, "A Regularized Model to Trade-off between Accuracy and Diversity in a News Recommender System," Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020, pp. 551–560, 2020, doi: 10.1109/BigData50022.2020.9378340.
[10] Z. Zhu, D. Li, J. Liang, G. Liu, and H. Yu, "A Dynamic Personalized News Recommendation System Based on BAP User Profiling Method," IEEE Access, vol. 6, no. c, pp. 41068–41078, 2018, doi: 10.1109/ACCESS.2018.2858564.
[11] L. P. Manik, A. F. Syafiandini, H. F. Mustika, A. F. Abka, and Y. Rianto, "Evaluating the Morphological and Capitalization Features for Word Embedding-Based POS Tagger in Bahasa Indonesia," 2018 Int. Conf. Comput. Control. Informatics its Appl. Recent Challenges Mach. Learn. Comput. Appl. IC3INA 2018 - Proceeding, pp. 49–53, 2018, doi: 10.1109/IC3INA.2018.8629519.
[12] F. Wu et al., "MIND: A large-scale dataset for news recommendation," Proc. Annu. Meet. Assoc. Comput. Linguist., pp. 3597–3606, 2020, doi: 10.18653/v1/2020.acl-main.331.
[13] M. A. Nugraha, Z. K. A. Baizal, and D. Richasdy, "Chatbot-Based Movie Recommender System Using POS Tagging," Build. Informatics, Technol. Sci., vol. 4, no. 2, pp. 624–630, 2022, doi: 10.47065/bits.v4i2.1908.
[14] S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," Smart Innov. Syst. Technol., vol. 105, pp. 391–397, 2019, doi: 10.1007/978-981-13-1927-3_42.
[15] L. Li, L. Zheng, F. Yang, and T. Li, "Modeling and broadening temporal user interest in personalized news recommendation," Expert Syst. Appl., vol. 41, no. 7, pp. 3168–3177, 2014, doi: 10.1016/j.eswa.2013.11.020.
[16] R. H. Singh, S. Maurya, T. Tripathi, T. Narula, and G. Srivastav, "Movie Recommendation System using Cosine Similarity and KNN," Int. J. Eng. Adv. Technol., vol. 9, no. 5, pp. 556–559, 2020, doi: 10.35940/ijeat.e9666.069520.
[17] M. Chiny, M. Chihab, O. Bencharef, and Y. Chihab, "Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms," no. Bml 2021, pp. 15–20, 2022, doi: 10.5220/0010727500003101.
[18] P. B.Thorat, R. M. Goudar, and S. Barve, "Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System," Int. J. Comput. Appl., vol. 110, no. 4, pp. 31–36, 2015, doi: 10.5120/19308-0760.
[19] W. Y., "POS Tagger Bahasa Indonesia dengan Python," wordpress, 2018. https://yudiwbs.wordpress.com/2018/02/20/pos-tagger-bahasa-indonesia-dengan-pytho (accessed Feb. 04, 2024).
[20] Rina, "Mudah Scraping Berita Online pada Situs detik.com Menggunakan Google Colab," Medium, 2023. https://esairina.medium.com/scraping-berita-online-pada-situs-detik-com-menggunakan-google-colab-3a764981384b (accessed Feb. 06, 2024).
[21] N. K. S., "Confusion Matrix untuk Evaluasi Model pada Supervised Learning," Medium, 2019. https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f (accessed Nov. 30, 2023).