Building SMART Recommendation Systems for Medical Facilities in Uganda using Content-Based and Case-Based Approaches

Rashid Kisejjere, Arthur Kakande
rashid[at]pollicy.org, arthur[at]pollicy.org
Research and Data Products
Pollicy Data Institute

October 10, 2024

Abstract

Uganda faces significant health challenges, including high rates of infectious diseases and maternal mortality, prompting the government to implement digital tools under its Health Information and Digital Health Strategic Plan 2020/21-2024/25. While digital applications like MatHelp and Ask RHU have improved access to reproductive health and HIV/AIDS services, there remains a critical need for a customised medical recommendation system to help users locate general healthcare services. In this study, we utilise content-based filtering and knowledge-based approaches to building recommendation systems tailored to Uganda's healthcare system. The content-based filtering approach uses data preparation techniques like vectorization along with cosine similarity to match users with healthcare facilities best suited to their needs. The knowledge based approach uses case-based reasoning implemented with a novel open-source framework "intellikit" alongside structured case representations to identify the best facility based on user preferences specified in a query. This paper also introduces a novel dataset for medical facilities compiled from various online sources and the Google Maps Application Programming Interface (API) curated with multiple features such as location, and opening hours, suitable for further studies on medical facilities in Uganda. Altogether, this study demonstrates the potential of intelligent algorithms in enhancing healthcare access by matching user requirements with appropriate facilities. It also represents an innovative step towards improving healthcare accessibility and quality in Uganda through advanced digital solutions.

Keywords: Recommendation Systems, AI, Case-Based Reasoning

1 Introduction

Uganda has a high prevalence of infectious diseases such as malaria, HIV/AIDS, and tuberculosis, as well as other health issues like neonatal and maternal-related deaths, mental health, road injuries, and respiratory infections which puts a strain on the healthcare system. Additionally, the healthcare system is still struggling to provide meaningful access to health services, especially in rural areas and congested areas in the country(Turyamureeba, Yawe and Bosco, 2024). Uganda, just like many other African countries, access to quality healthcare is not proportional to the structure of healthcare facilities in the country (Mwesigwa, Wahid and Sohheng, 2021)., With over 6,940 health facilities including government-owned, private for-profit, and private non-profit establishments (Turyamureeba et al., 2024). Navigating this complex ecosystem to find suitable healthcare providers tailored to individual needs can be challenging. Uganda's healthcare system faces significant challenges including inadequate health financing which has a profound impact on the quality of medical personnel, shortage of equipment and medicines as well as the overall healthcare services provided(Mugenyi, Oduoye and Akilimali, 2024). This, as a result, leads to potential diagnostic flaws and inappropriate treatment by medical professionals.

Against this backdrop, more and more people rely on the internet to find health information. Search engines like Google receive over a billion health-related queries daily, accounting for 7% of all queries (Drees, 2019). According to a Pew Research survey, eight out of ten internet users have looked for health-related information online (such as food, exercise, medications, health insurance, therapies, physicians, and hospitals)(PEW Research Center, 2005). But typically, this online data is not customized to meet the unique requirements of every patient (Carter, Nunlee-Bland and Callender, 2011). Additionally, users' levels of health literacy differ, as some require proficiency to comprehend medical jargon, assess the true significance of the data that has been retrieved, or verify the reliability of the information sources(Hardey, 1999). Recommendation systems that use computer-based intelligent mechanisms can help reduce any kind of information overload by tailoring results to what is most relevant to the user. Significant work has been done in the past on healthcare recommendation systems such as the hospital recommendation system using machine learning(WARSE The World Academy of Research in Science and Engineering, 2020), the med-recommender system for predictive analysis of hospitals and doctors(Swarnalatha, Kesavarthini, Poornima and Sripriya, 2019), and the health recommendation system using deep learning-based collaborative filtering(Chinnasamy, Wong, Raja, Khalaf, Kiran and Babu, 2023). In this project, we contributed to research by developing a medical recommendation system tailored to Uganda's health care system with a major focus on health facilities in Kampala city and the neighboring Wakiso district.

User ID	Hospital ID	Ratings
123	323	4
343	233	3
545	342	4.5
400	231	3.1

Table 1: Table showing sample data of collaborative filtering

2 Approaches For Recommendation Systems

Recommendation systems utilize a diverse number of techniques with the most prominent ones being content-based filtering, collaborative filtering, case-based filtering, and hybrid approaches. Depending on the application, each approach varies in its efficacy and accuracy. Therefore, it's critical to identify the optimal technique to incorporate into our system based on each method's unique characteristics. For our health facility recommendation system to work, we identified and tested out multiple of these approaches. The choice among these approaches depended on the nature of the available data. Below, we delved into each approach to determine the most suitable one for the dataset at hand.

2.1 Collaborative Filtering.

This approach operates on a user-centric approach, where recommendations are tailored to individuals based on the preferences of similar users. By identifying groups of individuals with similar preferences within their user base, the system offers personalized recommendations. Figure 1 illustrates sample data suitable for a collaborative recommendation system.

The system evaluates the similarity between the target user and others in the dataset first and once the similarities are established, the system recommends the hospital most frequently utilized by the most similar user. This approach has the following limitations;

User-Centricity. This method heavily relies on user-centered data, which introduces difficulties in data collection due to associated costs.

Cold-start problem Introducing a new user to the system presents challenges in effectively clustering them to identify potential similarities with existing users. The initial lack of user data complicates the recommendation process.

Data Sparsity. Collaborative filtering relies on large datasets, making it suitable for platforms like Netflix recommendation systems(Sütçü, Kaya and Erdem, 2021), which benefit from extensive databases. However, for hospital recommendation systems in Uganda, gathering such voluminous data is arduous and costly.

These factors collectively render this approach less feasible for implementation in this particular project.

2.2 Content-Based Filtering (CBF)

Unlike collaborative filtering, content-based filtering recommends an item by primarily analyzing the item's intrinsic characteristics (hence the name content-based). It identifies other items that are similar to a specific item based on their attributes and their close alignment with user preferences or needs(Son and Kim, 2017). Content-based filtering recommendation carries out keyword extraction first, which involves identifying details about the specific item under consideration. In the context of recommending hospitals, this involves extracting crucial information such as the hospital's name, location, payment methods, and the services provided. The extracted features are then converted into vector representations. This can be through vector extraction techniques like Term Frequency Inverse Document Frequency(TF-IDF)(Roelleke and Wang, 2008) afterward representing the hospital as vector distribution which is used to calculate the similarity. Content-based filtering is a good approach for this dataset as it tracks items and makes recommendations based on their similarities.

2.3 Knowledge-based recommenders and Case-Based Reasoning (CBR)

The content-based approach shares great similarities with case-based reasoning, an AI approach that solves new problems by utilizing solutions from cases in a case base. The core concept in CBR is that similar problems have similar solutions and once a similar problem is identified, the solution from that problem can be adapted to the new problem(Bergmann, Althoff, Minor, Reichle and Bach, 2009). Within the case-based reasoning approach, items or products are represented in a structured manner, such as an attribute-value case representation. The preferences of a user are then utilized to identify a suitable recommendation. Case-based reasoning is a prevalent technique in knowledge-based recommenders, which make recommendations based on the user's explicit qualitative knowledge. This approach generates recommendations by analyzing user choices and matching them with similar past cases. In this project, each hospital is represented as a unique case, and the user's selected preferences serve as the query. We then apply CBR to identify the most similar case that matches the user's preferences. The case-based reasoning cycle involves four main stages:

Retrieve: During this stage, the system searches the case base to find cases that are similar to the user's query. This involves comparing the user's preferences with the attributes of existing cases to identify the most relevant matches.

Reuse: In this phase, the system adapts the retrieved cases to fit the new problem context. This may involve modifying certain aspects of the retrieved cases to better align with the user's specific needs or preferences.

Revise: After reusing the case, the system tests the proposed solution in the real-world context. If the solution is not satisfactory, the system adjusts and refines it based on feedback and further analysis.

Retain: Finally, the system updates its case base by adding the new solution and the associated problem. This helps improve the system's performance over time by learning from new experiences and enriching the case base with diverse examples.

By following the CBR cycle, the knowledge-based recommendation system continually enhances its ability to provide accurate and personalized recommendations, leveraging past experiences to inform future decisions. The CBR cycle is il-

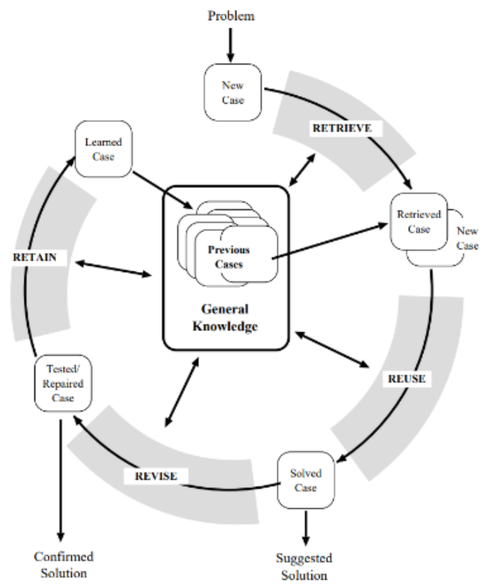


Figure 1: Image illustrating the CBR cycle

lustrated in Figure 2.

2.4 Hybrid

Combining the capabilities of collaborative filtering and content-based filtering, the hybrid approach provides more realistic and reliable recommendations ensuring high-accuracy recommendations. It's used by major companies like Netflix(Aslanian, Radmanesh and Jalili, 2016) to recommend movies to watch based on the user ratings (collaborative filtering) and also the movie's features like genre and release date (content-based approach). Although this method is effective, it still doesn't apply to the current task as the dataset doesn't contain information regarding the user's tastes and preferences.

In summary, collaborative filtering, CBF, CBR, and hybrid are widely used in building recommendation systems. In this paper, we experiment with CBR and CBF on our compiled dataset of Uganda healthcare facilities.

3 Data Collection

The initial web search carried out revealed no public repositories containing the required data about the healthcare facilities in Uganda. So we curated a new dataset with a major portion of the data from the Ministry of Health's 2018 list of over 6,000 healthcare facilities in Uganda(Ministry of Health, 2020). We then narrowed this data to only Kampala and Wakiso Regions resulting in 1,662 healthcare facilities. Additional data was obtained from the health facility websites and the Google Map's API(Google, 2024) which provided more information about the health facilities like their services, ratings, and concise locations. The collected data was publicly available through web scraping and included no personally identifiable information (PII). The collected features included the healthcare facility's name, rating, services offered, location, coordinates, time of operations, the form of payment (no payment, cash, or insurance), hospital type, hospital contact phone number, and the website URL.

4 Implementation

This section explains in more detail the different approaches that were used when implementing the recommendation system.

4.1 Content-Based Filtering Approach

To implement the content-based filtering approach we used cosine similarity, which is a measure between two non-zero vectors(Steck, Ekanadham and Kallus, 2024). It represents the cosine of the dot product of two vectors divided by the product of their lengths, as illustrated in the equation below. This measure is ideal for high-dimensional sparse data because the magnitude of the vectors does not influence the similarity measure as similarity is based solely on the directional alignment.

$$\text{Cosine Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

The recommendation process we adopted consisted of two stages; Filtering based on services first and then filtering based on the remaining features to arrive at the recommendation. This guaranteed that the suggestions were relevant to the services the user needed. In step one, the services were vectorized using the OHE (One Hot Encoding) vectorization that returns a 1 where the health facility has a given service and a 0 otherwise as illustrated below. The cosine similarity of a given encoded user service was then calculated with all the vectorized data to narrow the data to only the top 4 health facilities with the highest similarity to the user's needed services and thereafter proceed to step 2

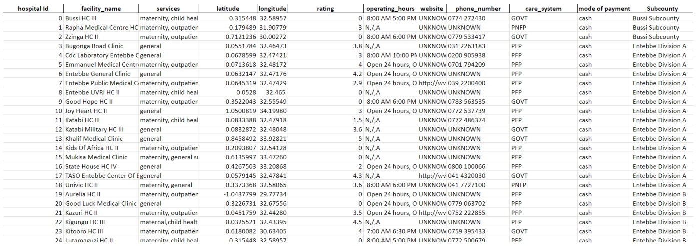
	Service A	Service B	Service C	Service D
Hospital A	1	0	0	1
Hospital B	0	0	1	1
Hospital C	0	1	0	1

Table 2: Table showing OHE Vectorization

In step two, similarity calculation was based on the other features operating time, location, rating, care system, and payment type. In the vectorization process, The OHE vectorization was carried out on the operating hours and for the care system, the data was numerically encoded since it only has two values of Public and Private. This numerical encoding was applied to the mode of payment feature as well. The numerical features "rating", "latitude" and "longitude" didn't require any preprocessing and were left as is. The encoded data was then used to calculate the cosine similarity of the filtered health facility and the top 3 most similar health facilities returned.

4.2 Case-Based Reasoning Implementation

When implementing this approach, the system's goal was to find a health facility that best matches a user's specific requirements provided in a query. We stored health facility



hospital Id	facility_name	services	latitude	longitude	rating	operating_hours	website	phone_number	care_system	mode of payment	Subcounty
0	Bussi HC III	maternity, child heal	0.315448	32.58957	0	8:00 AM 5:00 PM,	UNKNOW	0774 272430	GOVT	cash	Bussi Subcounty
1	Rapha Medical Centre HC	maternity, outpatier	0.179489	31.90779	3	N,/,A	UNKNOW	UNKNOWN	PNFP	cash	Bussi Subcounty
2	Zzinga HC II	maternity, outpatier	0.7121236	30.00272	0	8:00 AM 6:00 PM,	UNKNOW	0779 533417	GOVT	cash	Bussi Subcounty
3	Bugonga Road Clinic	general	0.0551784	32.46473	3.8	N,/,A	UNKNOW	031 2263183	PFP	cash	Entebbe Division A
4	Cdc Laboratory Entebbe C	general	0.0678599	32.47421	3	8:00 AM 10:00 PM	UNKNOW	0200 905938	PFP	cash	Entebbe Division A
5	Emmanuel Medical Centr	maternity, outpatier	0.0713618	32.48172	4	Open 24 hours, O	UNKNOW	0701 794209	PFP	cash	Entebbe Division A
6	Entebbe General Clinic	general	0.0632147	32.47176	4.2	Open 24 hours, O	UNKNOW	UNKNOWN	PFP	cash	Entebbe Division A
7	Entebbe Public Medical C	maternity, outpatier	0.0645319	32.47429	2.9	Open 24 hours, O	http://ww	039 2200400	PFP	cash	Entebbe Division A
8	Entebbe UVRI HC II	maternity, child heal	0.0528	32.465	0	N,/,A	UNKNOW	UNKNOWN	PFP	cash	Entebbe Division A
9	Good Hope HC II	maternity, outpatier	0.3522043	32.55549	0	8:00 AM 6:00 PM,	UNKNOW	0783 563535	GOVT	cash	Entebbe Division A
10	Joy Heart HC II	general	1.0500819	34.19980	3	Open 24 hours, O	UNKNOW	0772 537739	PFP	cash	Entebbe Division A
11	Katabi HC III	maternity, child heal	0.0833388	32.47918	1.5	N,/,A	UNKNOW	0772 486374	PFP	cash	Entebbe Division A
12	Katabi Military HC III	general	0.0832872	32.48048	3.6	N,/,A	UNKNOW	UNKNOWN	GOVT	cash	Entebbe Division A
13	Khalif Medical Clinic	general	0.8458492	33.92821	5	N,/,A	UNKNOW	UNKNOWN	GOVT	cash	Entebbe Division A
14	Kids Of Africa HC II	maternity, outpatier	0.2093807	32.54128	0	N,/,A	UNKNOW	UNKNOWN	PFP	cash	Entebbe Division A
15	Mukisa Medical Clinic	maternity, general su	0.6135997	33.47260	0	N,/,A	UNKNOW	UNKNOWN	PFP	cash	Entebbe Division A
16	State House HC IV	general	0.4267503	33.20868	2	Open 24 hours, O	UNKNOW	0800 100066	PFP	cash	Entebbe Division A
17	TASO Entebbe Center Of E	general	0.0579145	32.47841	4.3	N,/,A	http://ww	041 4320030	GOVT	cash	Entebbe Division A
18	Univic HC II	maternity, general	0.3373368	32.58065	3.6	8:00 AM 6:00 PM,	UNKNOW	041 7727100	PNFP	cash	Entebbe Division A
19	Aurelia HC II	maternity, outpatier	-1.0437799	29.77734	0	Open 24 hours, O	UNKNOW	UNKNOWN	PFP	cash	Entebbe Division B
20	Good Luck Medical Clinic	general	0.3226731	32.67556	0	Open 24 hours, O	UNKNOW	0779 063702	PFP	cash	Entebbe Division B
21	Kazuri HC II	maternity, outpatier	0.0451759	32.44280	3.5	Open 24 hours, O	http://ww	0752 222855	PFP	cash	Entebbe Division B
22	Kigungu HC III	maternal,child healt	0.0325521	32.43395	4.5	N,/,A	UNKNOW	UNKNOWN	PFP	cash	Entebbe Division B
23	Kitooro HC III	maternity, outpatier	0.6180082	30.63405	4	7:00 AM 6:30 PM,	UNKNOW	0759 395433	GOVT	cash	Entebbe Division B
24	Lutamaguzi HC II	maternity, child heal	0.315448	32.58957	0	8:00 AM 5:00 PM	UNKNOW	0772 500679	PFP	cash	Entebbe Division B

Figure 2: Figure shows some of the healthcare facilities in the dataset

features as the case characterization, while elements describing the health facility were captured as the solution descriptions. However, there was no clear-cut distinction between the problem and solution descriptions, as seen in other approaches. For instance, features such as location could be both specified by the user in the query and required to be shown in the solution when the recommended health facility is presented.

The system's task in health facility recommendation was to find a health facility description that addresses a user's query to the greatest extent possible. Due to the lack of a clear distinction between the problem and solution, only metadata information such as the health facility website and ID were captured as solution descriptions. At the same time, the rest were included in the case characterization.

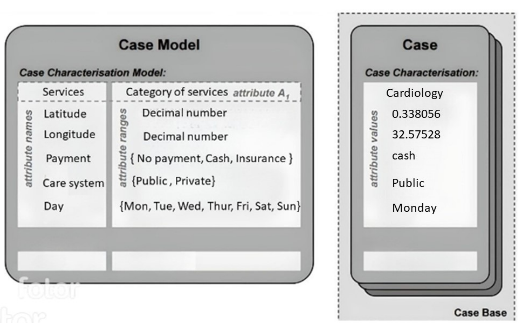


Figure 3: Enter Caption

This data structure was ideal for attribute-value case representation since it did not require relational representations. Key features represented include location, opening hours, closing hours, and services provided.

4.2.1 Similarity Assessment

Case-based recommenders employ structured approaches for similarity evaluation using structured case representations. Consequently, they can handle the retrieval inflexibility (stonewalling) issue that content-based recommenders often encounter. In this implementation, a local similarity measure was provided for each feature based on its data type and domain knowledge regarding its role in selecting a medical facility.

However, since this project was not aimed at comparing which health facility is the best, as is common in other

Features	Similarity Measures	Weights
Location	Euclidean distance	0.3
Services	Levenshtein	0.4
Day	Exact match	0.1
Payment	Exact match	0.1
Type of hospital	Levenshtein	0.1

Table 3: Selected features, similarity computation measures, and weights used

recommendation systems, our primary focus was ensuring that users could access the healthcare services they need as quickly as possible. To achieve this, we introduced a global weighting system. This system assigned higher weights to key features such as location, required services, payment methods, and opening and closing times. This prioritization was designed to enhance access to medical facilities, especially when urgent.

We implemented this approach using "Intellikit," an open-source Python framework for Case-Based Reasoning (CBR) https://arthurkakande.github.io/intellikit. The dataset consisted of 1,662 health facilities, each containing 1,662 columns, including location and other features as listed above in the data collection section. For data processing, the feature "day of the week" was organized into individual cases. Specifically, a health facility that was open 7 days a week was converted into 7 separate cases, each representing a single day. This approach allowed users to access health facilities available on a specific day, accommodating health facilities with irregular opening schedules.

The retrieval employed the Many Are Called Few Are Chosen (MACFAC) retriever, which uses a two-stage retrieval process. During the Many Are Called (MAC) phase (first stage) a simple similarity measure is conducted to reduce the size of the case base; the filtered case base is then subjected to the FAC phase, where the final selection of the suitable solution is conducted. This approach is mostly used for large case bases to improve efficiency. In the MAC phase, the system filters out health facilities based on the service and location, narrowing down the options. Then, in the FAC phase, time is considered on the filtered cases. In both phases, the weighted sum method was used to prioritize the retrieval based on the relevant features. This ensured that the most suitable health facilities were recommended based on the user's specified requirements and the urgency of the needed services.

5 Discussion

The content-based filtering approach handled the structured data available of the health facilities by utilizing one hot encoding (OHE) and cosine similarity which matched the user requirements with the available health facility features. This approach's primary advantage was its ability to recommend health facilities based on the direct comparison of attributes, which ensured that recommendations were relevant to the user's needs. However, its limitation was that it did not consider user preferences beyond the explicitly provided attributes, which could limit the personalization aspect of the recommendations.

The second part of this work involved the implementation of the recommendation system by applying case-based reasoning (CBR) using the Intellikit framework. The system utilized structured case representations and a two-stage retrieval process to identify health facilities that best-matched user preferences expressed in a query. However, the effectiveness of the recommendations could be questioned due to the variability in health facility data and the complexity of user requirements. Nonetheless, this approach serves as an excellent foundation for understanding other features that could enhance the recommendations, such as patient reviews, health facility specializations, and historical patient outcomes. Additionally, incorporating demographic data of users could further refine the recommendations, as it considers personal factors that influence health facility choice.

5.1 Conclusion and Future Work

In this study, we developed an innovative healthcare recommendation system tailored to Uganda's healthcare landscape, specifically for Kampala and Wakiso districts. By utilizing cosine similarity and case-based reasoning (CBR) approaches, our implementations demonstrated how structured data from health facilities in conjunction with intelligent algorithms can match user requirements with appropriate healthcare facilities. For future work, our study could be expanded by implementing a hybrid approach, superseding the current CBF and CBR methods. Recent research has demonstrated that hybrid approaches offer more reliable and accurate recommendations, which would represent a significant enhancement over existing methodologies. Given that Uganda is home to over six thousand healthcare facilities as of 2024, a number that is expected to grow, future research will also necessitate additional data collection of these establishments. This data collection process will be community-based, enabling hospital owners or interested users to submit healthcare facility information such as name, location, services offered, operating time, care system, and payment type. Our team will evaluate the submitted data prior to inclusion in the system. For any healthcare facility to be considered, it must be registered with Uganda's Ministry of Health and compliant. This process will be facilitated through an online platform developed by our team and implemented progressively on a district-by-district basis.

References

Aslanian, E., Radmanesh, M., Jalili, M., 2016. Hybrid recommender systems based on content feature relationship. IEEE Transactions on Industrial Informatics , 1–1.

Bergmann, R., Althoff, K.D., Minor, M., Reichle, M., Bach, K., 2009. Case-based reasoning. KI 23, 5–11.

Carter, E.L., Nunlee-Bland, G., Callender, C., 2011. A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. Perspectives in Health Information Management 8, 1b.

Chinnasamy, P., Wong, W.K., Raja, A.A., Khalaf, O.I., Kiran, A., Babu, J.C., 2023. Health recommendation system using deep learning-based collaborative filtering. Heliyon 9, e22844.

Drees, J., 2019. Google receives more than 1 billion health questions every day. Retrieved June 18, 2024, from https://archive.fo/dJdGu.

Google, 2024. Overview. Google for Developers. Retrieved from https://developers.google.com/maps/documentation/places/web-service/overview.

Hardey, M., 1999. Doctor in the house: The internet as a source of lay health knowledge and the challenge to expertise. Sociology of Health & Illness 21, 820–835.

Ministry of Health, 2020. Complete list of all health facilities in uganda - ministry of health. Retrieved August 6, 2024, from Ministry of Health — Government of Uganda website: https://shorturl.at/ScLRB.

Mugenyi, N.M., Oduoye, M.O., Akilimali, A.M., 2024. Supporting health systems in uganda: A call for action.

Mwesigwa, D., Wahid, K.A., Sohheng, N., 2021. A case study on the quality of healthcare in Uganda: Examining the effectiveness, safety, patient-centred and timeliness of district healthcare facilities .

PEW Research Center, 2005. Health information online. Retrieved June 18, 2024, from https://archive.fo/JK6XS.

Roelleke, T., Wang, J., 2008. Tf-idf uncovered: A study of theories and probabilities, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 435–442.

Son, J., Kim, S.B., 2017. Content-based filtering for recommendation systems using multiattribute networks. Expert Systems with Applications 89, 404–412.

Steck, H., Ekanadham, C., Kallus, N., 2024. Is cosine-similarity of embeddings really about similarity?, in: Companion Proceedings of the ACM on Web Conference 2024. Retrieved from http://dx.doi.org/10.1145/3589335.3651526.

Swarnalatha, S., Kesavarthini, I., Poornima, S., Sripriya, N., 2019. Med-recommender system for predictive analysis of hospitals and doctors, in: 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), IEEE. pp. 1–5.

Sütçü, M., Kaya, E., Erdem, O., 2021. Movie recommendation systems based on collaborative filtering: A case study on netflix. Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi 37, 367–376.

Turyamureeba, M., Yawe, B.L., Bosco, O.J., 2024. Health care delivery system in uganda: A review. Retrieved June 18, 2024, from https://www.ajol.info/index.php/thrb/article/download/231882/23

WARSE The World Academy of Research in Science and Engineering, 2020. Hospital recommendation system using machine learning. *WARSE The World Academy of Research in Science and Engineering - Academia.edu*. Published May 11, 2020.