



ORIGINAL

Implementation and Evaluation of a Hybrid Recommendation System for the Real Estate Market

Implementación y evaluación de un sistema híbrido de recomendaciones para el mercado inmobiliario

Carlos Henriquez Miranda¹  , German Sanchez-Torres¹  

¹Universidad del Magdalena, Grupo de Investigación y Desarrollo en Sistemas y Computación. Santa Marta, Colombia.

Cite as: Henríquez Miranda C, Sánchez-Torres G. Implementation and Evaluation of a Hybrid Recommendation System for the Real Estate Market. Data and Metadata. 2024; 3:.426. <https://doi.org/10.56294/dm2024.426>

Submitted: 07-02-2024

Revised: 24-05-2024

Accepted: 15-09-2024

Published: 16-09-2024

Editor: Adrián Alejandro Vitón-Castillo 

Corresponding author: Carlos Henriquez Miranda 

ABSTRACT

Introduction: the real estate market has been transformed by digital technologies, especially Industry 4.0, which has made property searching and evaluation more efficient, improving its accuracy with the use of advanced algorithms. Traditional methods have been replaced by online platforms using modern machine learning (ML) algorithms, leading to the need for personalized recommendation systems to improve user experiences.

Method: this study designed and implemented a hybrid recommendation system that combines collaborative and content-based filtering techniques. The development process involved four phases: literature review, technology selection, prototype implementation, and system deployment.

Results: the proposed hybrid model effectively addressed challenges such as data sparsity and the cold start problem, improving recommendation accuracy. In the evaluation, users indicated high satisfaction with the system's ability to offer personalized property recommendations.

Conclusion: thus, hybrid recommendation systems can significantly improve the property search experience by offering personalized recommendations. However, further research into the applicability of the system in different domains remains a need for further exploration.

Keywords: Hybrid Recommendation System; Collaborative Filtering; Content-Based Filtering; Real Estate Market; Machine Learning; User Satisfaction.

RESUMEN

Introducción: el mercado inmobiliario se ha transformado gracias a las tecnologías digitales, especialmente la Industria 4.0, que ha hecho que la búsqueda y evaluación de propiedades sea más eficiente, mejorando su precisión con el uso de algoritmos avanzados. Los métodos tradicionales han sido reemplazados por plataformas en línea que utilizan algoritmos modernos de aprendizaje automático (ML), lo que ha llevado a la necesidad de sistemas de recomendación personalizados para mejorar las experiencias de los usuarios.

Método: este estudio diseñó e implementó un sistema de recomendación híbrido que combina técnicas de filtrado colaborativo y basadas en contenido. El proceso de desarrollo implicó cuatro fases: revisión de la literatura, selección de tecnología, implementación del prototipo y despliegue del sistema.

Resultados: el modelo híbrido propuesto abordó eficazmente desafíos como la escasez de datos y el problema del arranque en frío, mejorando la precisión de las recomendaciones. En la evaluación, los usuarios indicaron una alta satisfacción con la capacidad del sistema para ofrecer recomendaciones de propiedades personalizadas. Así, los sistemas de recomendación híbridos pueden mejorar significativamente la experiencia de búsqueda de propiedades al ofrecer recomendaciones personalizadas.

Conclusiones: sin embargo, queda abierta la necesidad de una mayor investigación sobre la aplicabilidad del sistema en diferentes dominios como áreas para una futura exploración.

Palabras clave: Sistema de Recomendación Híbrido; Filtrado Colaborativo; Filtrado Basado en Contenido; Mercado Inmobiliario; Aprendizaje Automático; Satisfacción del Usuario.

INTRODUCTION

The real estate market has been experiencing a major transformation, mainly due to a large number of changes that have been experienced by digital technologies related to Industry 4.0. This evolution has significantly impacted the design of property search and evaluation processes, making them more accessible and intuitive.

In the past, the real estate industry relied largely on solutions built with traditional methods, such as print ads or direct interaction with customers. Currently, the volume of information available on the Internet has steadily increased, driving technological advancements and paving the way for sophisticated tools like Big Data and Machine Learning.⁽¹⁾ This field has become increasingly significant, developing models that uncover hidden patterns within complex systems, thereby enhancing organizational decision-making. Analyzing and processing the vast amounts of data generated by various devices is essential. For Big Data to be effective, the data must be abundant, diverse, and rapidly processed to deliver meaningful results.^(2,3) In this context, recommendation systems have emerged, leveraging algorithms and artificial intelligence techniques to offer personalized suggestions for products, services, or content based on data analysis, machine learning, and information theory.⁽⁴⁾ These systems have revolutionized various sectors—from e-commerce and entertainment to healthcare and education—by improving user experiences and optimizing decision-making processes.⁽⁵⁾

Online platforms with modern Machine Learning (ML) algorithm technologies have significantly transformed the real estate sector, giving users new facilities to access large property catalogs. In context, in India alone, the real estate sector has more than 3 million active users and 1,1 million listed properties, demonstrating the large scale of the market and the challenges associated with user engagement and satisfaction.⁽⁶⁾ In several regions, including the United States, the digital transformation is evident. Platforms such as Zillow and Redfin have become central to the home-buying process, relying on tools to estimate prices, analyze neighborhoods, and offer virtual tours. The increasing participation in real estate investments has generated a growing need for personalized and efficient property recommendations, showing the importance of effective recommendation systems (RS) to improve user experiences and assist in informed decision-making.

When reviewing works in the area, it is possible to notice that different approaches have been investigated to create recommendation systems specifically designed for the real estate industry. In Nguyen et al.⁽⁷⁾ a session-based recommendation system was presented that uses contextual data to address the cold start problem, offering useful recommendations for new users. Similarly, RE-RecSys, which ranks users based on their interaction history and uses a combination of collaborative and content-based filtering techniques to provide personalized property recommendations.⁽⁶⁾ Other approaches, attempting to leverage the strengths of neural networks and graph-based algorithms In Li et al.⁽⁸⁾ the combination of graph neural networks (GNN) and deep mutual learning techniques was proposed as a way to improve multimodal recommendation systems, attempting to diminish the effect of biases arising from feature fusion.^(9,10,11)

This field still needs further studies to address some challenges, one evident weakness is not considering broader applicability and generalizability across various contexts and domains. Additionally, the academic community in the area will always benefit from studies that critically analyze the existing literature by synthesizing the proposed methodologies and identifying general trends and gaps in the field of real estate recommendation systems.⁽¹²⁾ Specifically, one glaring weakness in this field is the little existing literature related to combining collaborative and content-based filtering techniques into a unified hybrid model that can effectively address the specific challenges of the real estate market.

The main objective of this study was *to design and implement a hybrid recommendation system that combines collaborative filtering and content-based techniques and to evaluate its effectiveness in enhancing the real estate search experience*. To achieve this, the main research question guiding this study was *How can a hybrid recommendation system be designed, implemented, and evaluated to improve the search experience in the real estate market by integrating collaborative and content-based filtering techniques, and what is its impact on user satisfaction?* For us, an important topic regarding improving processes is to increase user satisfaction. Therefore, a methodological approach was adopted that involves an exploratory review of existing recommendation systems, their methodologies, and tools. The focus was on hybrid models that integrate collaborative filtering and content-based techniques.

In the following section, we describe in detail the methodology used to develop and evaluate this system.

METHOD

The development of the project was divided into four main phases (see figure1):

- Phase 1, consisted of a review of the existing literature in order to identify the techniques and tools utilized in recommendation systems within the real estate industry.
- Phase 2, the focus was on selecting technologies and designing the system architecture. Our primary goal was to determine the appropriate technologies and develop the system architecture. During this phase, we selected key technologies for data extraction, analysis of the extracted data, and the recommendation engine. These technologies include Web Scraping for data extraction, Natural Language Processing (NLP) for data analysis, and machine learning techniques.
- Phase 3, we focused on the Prototype Implementation, where our main objective was to develop a fully functional prototype of the recommendation system. The system’s essential modules have been developed, which include the Web Scraper, NLP engine, recommendation engine, and updater. In this phase, the database was constructed and the frontend was developed using contemporary web technologies.
- Phase 4, the system was deployed and evaluated in an actual environment. The goal was to implement the system and assess its effectiveness. The prototype was deployed on a web server and acceptance tests were conducted with actual users. Furthermore, collaborative filtering techniques and hybrid models were employed to refine recommendations by taking into account recorded interactions and preferences.

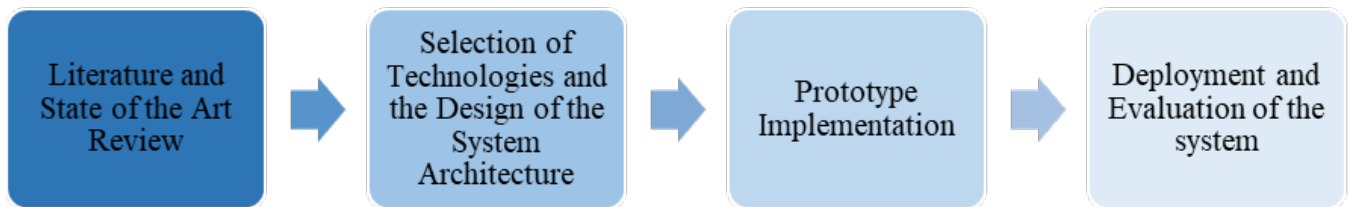


Figure 1. Overview of the general methodology employed

RESULTS**Literature and State of the Art Review***Exploratory review design*

Study Design: The methodology follows an exploratory review design, which aims to synthesize existing literature on hybrid recommendation systems specifically tailored to the real estate market. This design is aligned with Grant and Booth⁽¹³⁾, who define exploratory reviews as those that allow for an initial mapping of the literature, identifying key themes, gaps, and potential areas for further in-depth review. It permits the identification of best practices and gaps in the current literature, contributing to the development of a hybrid recommendation system.

Search Strategy: Searching process was conducted across two primary databases: Scopus and ScienceDirect, selected for their extensive coverage of peer-reviewed articles in the domains of computer science and real estate. These databases were chosen due to their robust indexing of high-quality research outputs, ensuring to capture relevant literature. The temporal range for the literature search was defined from January 2019 to October 2024, ensuring the inclusion of the most recent advancements and trends in hybrid recommendation systems, thus enhancing the relevance of the findings to current market dynamics.

Search Equation Development: The search equation includes relevant keywords and phrases that encapsulate the core themes of the review. The final search equation is as follows: “Title ((‘recommender systems’ OR ‘hybrid recommendation systems’ OR ‘collaborative filtering’ OR ‘content-based filtering’)) AND Title ((‘real estate’ OR ‘property market’)) AND Title ((‘system implementation’ OR ‘system architecture’ OR ‘personalized recommendation’))”. Key terms were selected based on their relevance to the objective, with a particular focus on system implementation and architecture, which are critical for understanding the practical implications of the research.

Selection Criteria: Inclusion criteria for this review encompassed: relevance to the review’s objective, peer-reviewed status, and publication in journals or conference proceedings. Conversely, exclusion criteria included non-relevance to the review’s objective, non-English publications, conference abstracts, books, articles aimed at non-academic audiences, studies focusing on unrelated fields, and purely literature reviews without new contributions. These aspects ensure that the findings synthesized are both credible and applicable to the real estate market.

Screening and Selection Process: The screening process was conducted into two stages, beginning with an initial review of titles and abstracts making a preliminary relevance analysis to identify relevant articles. Followed by a comprehensive full-text assessment of these articles.

Data Extraction: Data extraction was aimed to capture key variables relevant to the review's objectives. These variables included study design, sample size, methodology employed, key findings, and implications for practice.

Data Synthesis and Analysis: We used a narrative synthesis. The narrative synthesis explicitly links the findings to practical implications, emphasizing how the results can inform the development and implementation of hybrid recommendation systems in the real estate market. At this stage a taxonomy of relevant categories was identified.

Bias Consideration: We identified potential sources of bias in the review process, including source publication bias, language bias, and selection bias.

Exploratory review result

The analysis of the literature allowed us to establish a taxonomy that groups the studies with a more specific focus, allowing us to extract recommendations and conclusions in related subfields. The resulting taxonomy is shown graphically in Fig. 2. We considered that literature review works are important to the field, however, we grouped these works into a specific subcategory due to not a practical or methodological contribution is given.

Real Estate Recommendation Systems: Real Estate Recommendation Systems (RERS) are specialized computational frameworks designed to assist users in identifying and selecting properties that align with their preferences and needs, leveraging the facilities of algorithms and data analytics to offer personalized property suggestions, enhancing the decision-making process of investments and purchases. These systems can handle vast amounts of heterogeneous data, including user preferences, property attributes, and market trends, to offer tailored recommendations. Face up to the cold-start problems, where new users or properties lack sufficient interaction history, and the integration of multi-modal data sources to improve recommendation accuracy are relevant topics for these systems.⁽¹⁴⁾ More frequent methodologies encompass content-based filtering, which recommends properties based on the similarity of property features to user preferences; collaborative and hybrid approaches, which combine user-item interaction data to generate recommendations; and multi-modal recommendation systems, which utilize various data modalities such as text, images, and graphs to enhance recommendation quality.^(6,7,8,15,16)

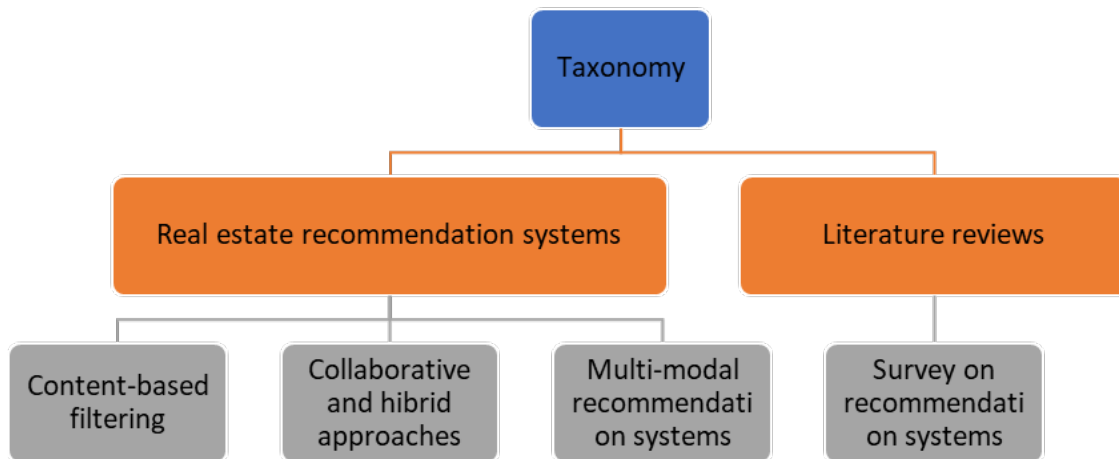


Figure 2. Taxonomy of the exploratory review

- **Content-Based Filtering:** This is a recommendation system methodology that leverages the attributes of items and user preferences to suggest relevant content. Works in this field cover the analysis of user behavior, the integration of contextual data, and the development of algorithms that improve the accuracy of recommendations.

In Nguyen et al.⁽⁷⁾, the authors propose a recommendation system tailored for real estate that addresses the cold-start problem using a session-based approach. The system integrates contextual data and employs a weighted cosine similarity method to identify suitable candidates. The results recommend that incorporating contextual information significantly improves recommendation accuracy. The conceptual contribution was demonstrating the efficacy of session-based approaches. However, it acknowledges limitations related to potential biases in the dataset and the generalizability of findings across different real estate markets.

Similarly, Badriyah et al.⁽¹⁵⁾ develops a web-based recommendation system for property searches using CBF, that utilizes the CBF method and the Apriori algorithm for association rule mining. This approach allows for a systematic analysis of user behavior and property characteristics. Practical implications include improved user satisfaction and increased efficiency in property searches. Some limitations, such as potential biases in user behavior data and the cold-start problem for new users, are recognized by author.

Both works recognize the significance of personalized recommendations in enhancing user satisfaction and engagement. Future research, includes the integration of hybrid systems and the incorporation of user feedback.

- **Collaborative and Hybrid Approaches:** This group consists of multiple methodologies with the same goal of improving recommendation accuracy. These approaches address the limitations of systems that rely on a single method, such as data sparsity and scalability issues. Central topics in this area include the combination of collaborative filtering, content-based filtering, and rule-based systems to tailor recommendations based on user behavior, preferences, and historical data.

In Venkatesh et al.⁽⁶⁾, a comprehensive real estate recommendation system is presented that classifies users into four distinct groups: cold-start users, short-term users, long-term users, and short-long-term users. This categorization allows the recommendation approach to be customized based on the user's interaction history. That is, cold-start users rely on a rule-based engine, while short-term users benefit from content filtering models. For long-term and short-long-term users, recommendations are generated using a hybrid approach that combines collaborative and content filtering. Using these multiple layers addresses the problem of sparse interactions between user and properties and ensures low latency in real-world applications.

Similarly, in Nallamala et al.⁽¹⁶⁾, a comprehensive review of various recommendation techniques, including collaborative filtering, content-based filtering, and hybrid methods, is provided. The study emphasizes that hybrid approaches can overcome the limitations of individual methods, specifically by managing scalability and adaptation to diverse user needs. The comparative analysis of existing algorithms in this paper shows the strengths and weaknesses of each approach, and serves as a valuable resource for designing effective recommendation systems in different industries. The study concludes that hybrid systems present promising solutions to the challenges of knowledge overload.

This sub-area highlights the need to integrate multiple recommendation techniques to meet diverse user needs and the importance of personalized recommendations to improve user engagement. Both papers suggest the practical benefits of hybrid approaches, significantly improving performance.

The used user categorization and rule-based engine for cold-start users is an important contribution to the area. It is directly related to the generation of low latency, making it attractive for real-world application. Furthermore, the comprehensive analysis of existing algorithms offers a valuable reference for future research and development in this field.⁽¹⁶⁾ However, potential biases in user interaction data and challenges in generalizing findings across different contexts are acknowledged. Future research should investigate the integration of advanced machine learning techniques, such as deep learning, to further improve recommendation accuracy and adapt to evolving user behaviors over time.

- **Multi-Modal Recommendation Systems:** Multimodal recommender systems are advanced frameworks designed to improve the accuracy and relevance of recommendations by integrating multiple types of data modalities such as visual, textual, and acoustic features. They are important in areas such as financial decision making and multimedia content services, where user preferences are influenced by a variety of data sources of different nature. Graph neural networks (GNNs) and deep mutual learning techniques, which are effective in capturing and utilizing the rich semantic information embedded in different modalities, are used.

In Li et al.⁽⁸⁾, GNNMR is presented, a multi-modal recommendation framework that integrates GNNs, that face up to the inherent bias among different modalities during feature fusion, consequently leading to sub-optimal embeddings for items with multi-modal features. The main innovation of GNNMR is its integration of GNNs with deep mutual learning, which employs mutual knowledge distillation to collaboratively train multiple unimodal bipartite user-item graphs. Each GNN is trained on a specific unimodal user-item bipartite graph, separated from the original multimodal user-item bipartite graph, to generate unimodal embeddings. These embeddings then serve as mutual supervision signals, allowing the model to discover and synchronize latent semantic relationships between different modalities.

Experimental results show that GNNMR outperforms other multimodal recommendation methods in the Top-K recommendation task on two real-world datasets.⁽¹⁷⁾ This demonstrates the benefits of multimodal learning in recommendation systems, emphasizing the importance of effectively capturing semantic relationships between different modalities to improve recommendation performance. Therefore, it is suggested that improved recommendation systems can lead to better investment decisions and higher

user engagement on multimedia platforms.

However, biases in the datasets and questions about the generalizability of the findings across different domains are some of the most outstanding limitations. This indicates an area for future research, which could explore the integration of GNNMR with various late fusion methods and test its performance in different application domains to further validate its effectiveness.

Literature Reviews: Literature reviews are indispensable in academic research. They establish the foundation for future research by identifying trends, gaps, and methodological approaches. Systematic reviews, meta-analyses, and narrative syntheses are among the standard methodologies that guarantee a thorough and impartial analysis of the literature.

- *Surveys on Recommendation Systems:* In Gharahighehi et al.⁽¹²⁾, an extensive analysis of how recommendation systems are used in this context is offered, identifying distinctive challenges such as the cold start problem and domain-specific limitations that differentiate real estate recommendation systems from those in other fields. 26 key research papers were evaluated, organizing them based on their methodological approaches and the specific challenges they address. The prevalence of collaborative filtering as a preferred method in real estate recommendation systems was identified. However, the findings show the need for hybrid models that can integrate geographic and temporal factors to improve recommendation accuracy, reflecting broader trends and recognizing the limitations of relying solely on single-method approaches.

Furthermore, the importance of addressing domain-specific topics is highlighted, in particular the cold start problem, which constitutes a major problem in the real estate market due to the high variability and infrequency of real estate transactions. The development of advanced feature extraction techniques from unstructured data and the creation of multi-stakeholder recommender systems that take into account the diverse preferences of users, such as buyers, sellers, and agents, are suggested. As limitations, the lack of publicly available data sets for benchmarking hinders the development and evaluation of new recommender system methodologies. Thus, better and more standardized benchmarking practices and new data sources are desirable in this field.

Technology Selection and System Architecture Design

Our main objective was to design the architecture of the hybrid recommendation system for the real estate market and select the most suitable technologies. This includes identifying the technological tools that we consider critical for the extraction, analysis, and development of the recommendation engine.

Technology Selection

Based on the research question, the user satisfaction was an important topic. As well as mitigating some of the limitations found in the exploratory literature review. Managing substantial amounts of real estate data and offering personalized recommendations to consumers were aspects that were considered in this stage. Therefore, three main technologies were selected:

- *Web Scraping:* We implemented web scraping-based procedures to extract real estate data from a variety of online sources, as input for training, testing, and validation. Technologies such as BeautifulSoup, Scrapy, and Selenium were evaluated for their ability to efficiently collect data, including property details, prices, and locations, and their ability to handle dynamic web pages.⁽¹⁸⁾
- *Natural Language Processing (NLP):* NLP techniques were employed to analyze textual data extracted from property descriptions and user reviews. To enhance the system's understanding of user preferences and property characteristics, libraries such as NLTK, SpaCy, and Scikit-Learn were employed to process and extract meaningful insights from unstructured text data.
- *Machine Learning:* To analyze user interactions and anticipate future preferences, the recommendation engine employed machine learning algorithms. The system's accuracy and ability to deliver personalized recommendations were improved by integrating techniques such as content-based filtering and collaborative filtering. Models were developed and trained using frameworks such as Scikit-Learn and TensorFlow.

System Architecture Design

The system architecture was designed to be modular and scalable, allowing each component to operate independently while ensuring seamless integration. The architecture consisted of four main components (see figure 3):

- a. *Data Collection Web Scraper:* The first module of the system is responsible for the automated extraction of data from various online sources. This component is crucial for ensuring that the system relies on updated and relevant information. The scraping process involves not only data collection but

also content classification and filtering to eliminate noise and ensure data quality. This initial phase is critical, as the accuracy and relevance of the recommendations directly depend on the quality of the collected data.

b. The content analysis module focuses on interpreting and understanding the collected text. This component goes beyond simple keyword extraction to delve into the semantics of the language, capturing deeper nuances and contexts. Sentiment analysis, entity detection, and other advanced techniques enable the system to identify not only explicit preferences but also more subtle inferences that may not be immediately apparent. This level of textual understanding is essential for building a more comprehensive and accurate user profile.

c. The recommendation engine constitutes the core of the system, where multiple methodological approaches are integrated to maximize the accuracy and relevance of suggestions. The combination of content-based methods, collaborative filtering, and knowledge-based techniques allows for capturing current user preferences as well as predicting future interests and needs. This hybrid approach addresses the need for personalization in complex systems and ensures that recommendations are both accurate and adaptive as user behavior evolves.

d. Updater: The update module acts as a dynamic mechanism that adjusts recommendations in real-time based on the user's most recent interactions. This continuous feedback process is essential for maintaining the relevance and personalization of the system. The ability to adapt suggestions to changes in user behavior not only enhances the user experience but also reinforces the system's efficiency by preventing the obsolescence of recommendations. This ongoing update approach ensures that the system remains aligned with the current expectations and needs of users.

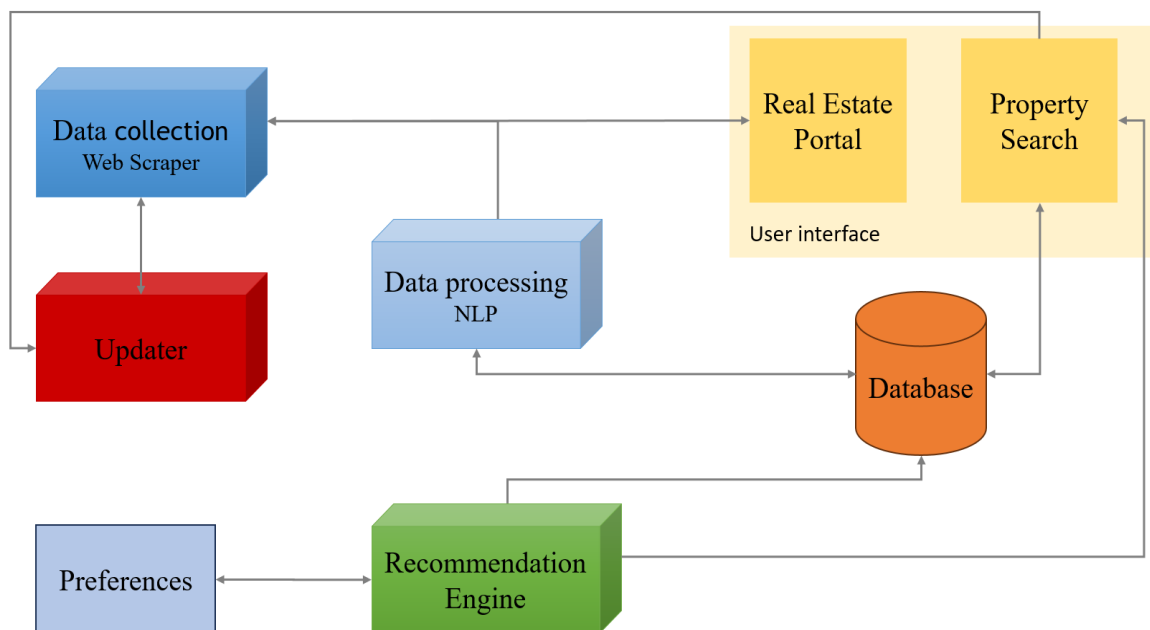


Figure 3. System architecture design

Prototype Implementation

The system's core components were created, which encompassed the Web Scraper, Natural Language Processing (NLP) engine, recommendation engine, and updater. In this stage, the database was constructed (see figure 4) and the frontend was created utilizing modern web technologies (see figure 5). The process of constructing a prototype involved ongoing testing and feedback in order to improve and optimize the performance of the system.

Deployment and Evaluation of the system

The recommendation system module operates through process that begins with the user's initial registration. During registration, the user is prompted to select their preferences for property features. These preferences serve as one of the key interaction points within the application. Additional interactions include marking a property as a favorite, rating a property, tracking the number of clicks on each property, and selecting specific property preferences. Each of these interactions is recorded and used to tailor the recommendations.

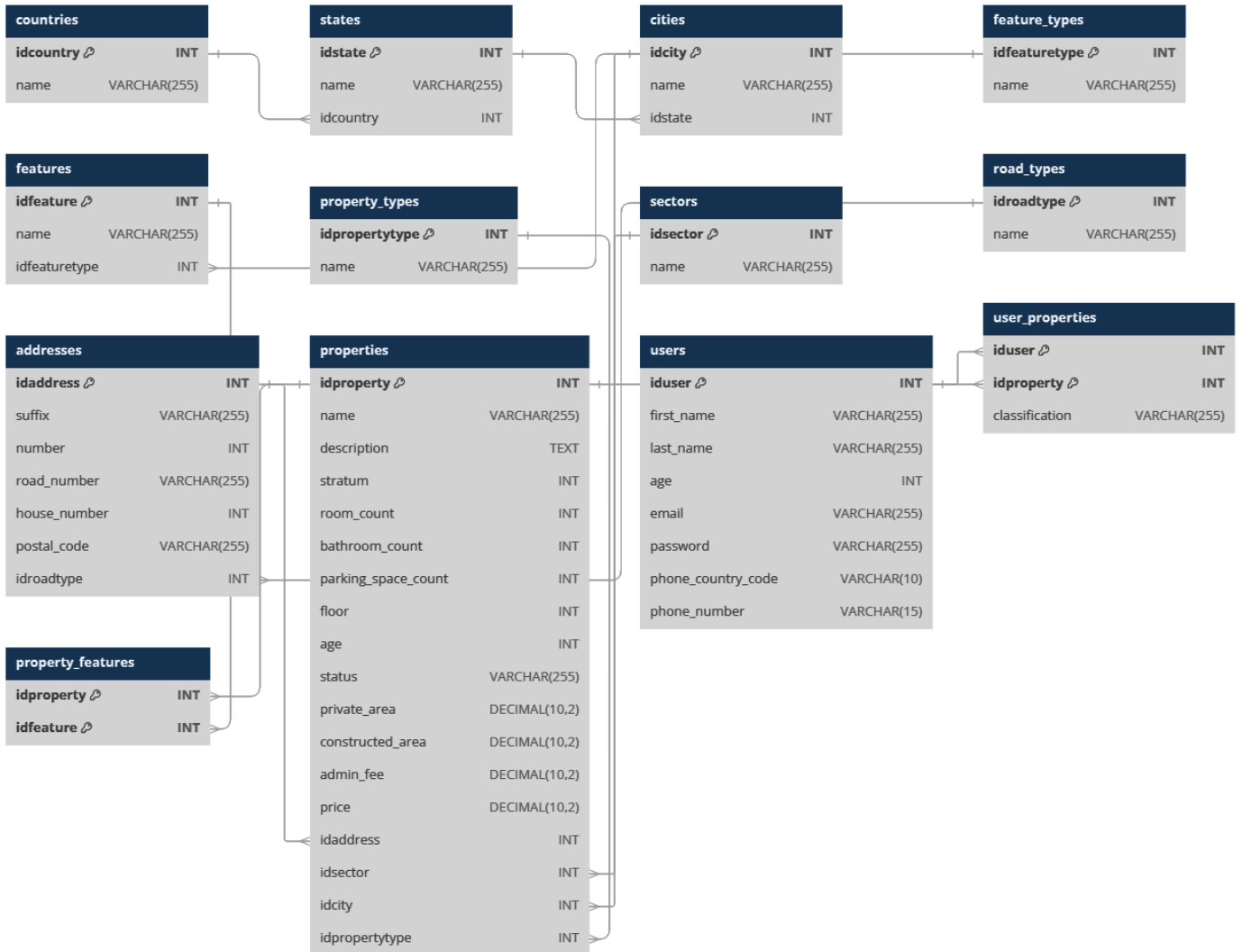


Figure 4. Basic database model of the implemented system

Given the variety of interactions and their differing scales, such as binary values for favorites, a 1-5 scale for ratings, and an open-ended count for clicks, a standardized scoring system is necessary. The score for each interaction is calculated using the following equation:

$$valuePerPreference = maximumScore / numberOfPreferences \quad (1)$$

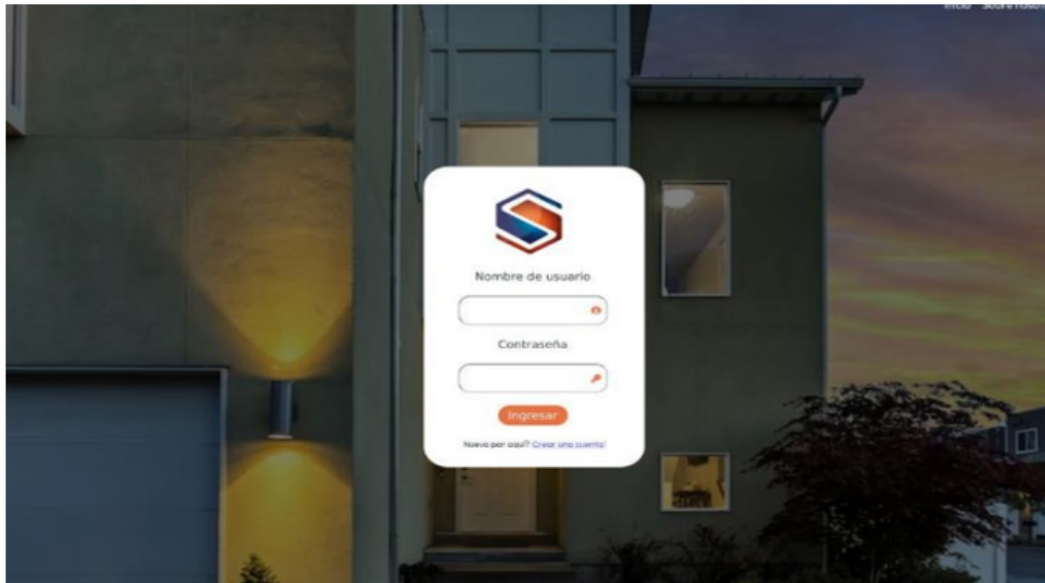
Where the maximum score is set to 1,0 to maintain a scale between 0 and 1. The number of preferences is user-specific, meaning that the value per preference will vary depending on how many preferences a user has selected.

For each property, a score is computed based on how many of the user’s preferences it meets. The score for each property is then calculated as:

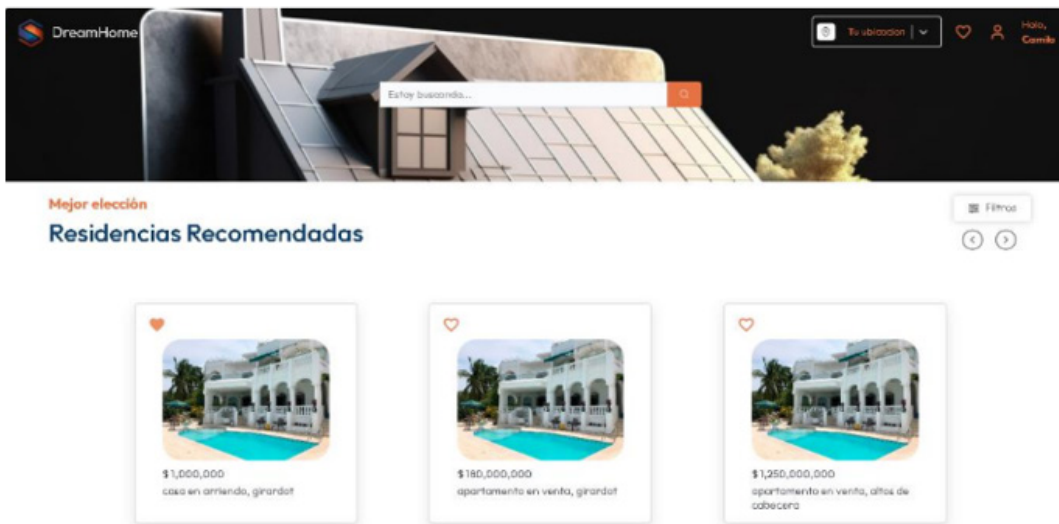
$$preferenceScorePerProperty = preferencesMet \times valuePerPreference \quad (2)$$

The final scores for each interaction (favorite, rating, clicks) are normalized to a 0-1 scale using min-max normalization:

$$normalizedValue = value - minValue / (maxValue - minValue) \quad (3)$$



a)



b)

Figure 5. Some views of the implemented system, a) login view, and b) recommended properties view

The final step involves assigning weights to each interaction type to reflect their relative importance. The assigned weights are as follows: 4 for favorites, 2 for ratings, 1 for clicks, and 3 for preferences. The overall score for each property is calculated as:

$$finalScore = (4 \times Favorite) + (2 \times Rating) + (1 \times Clicks) + (3 \times Preferences) \quad (4)$$

This scoring process is repeated for all users and properties, creating a user-property matrix populated with scores. A challenge arises when certain properties have no interactions and therefore a score of zero. This is where collaborative filtering comes into play, utilizing similarities between users to estimate scores for unrated properties.

Collaborative filtering leverages the concept of similarity, which can be measured using methods such as Manhattan distance, Euclidean distance, or cosine similarity.⁽¹⁹⁾ For instance, cosine similarity is used to measure the similarity between two users' interaction vectors. The formula for cosine similarity between two vectors A and B is:⁽²⁰⁾

$$Cosine\ similarity = (A \times B) / (|A| \times |B|) \quad (5)$$

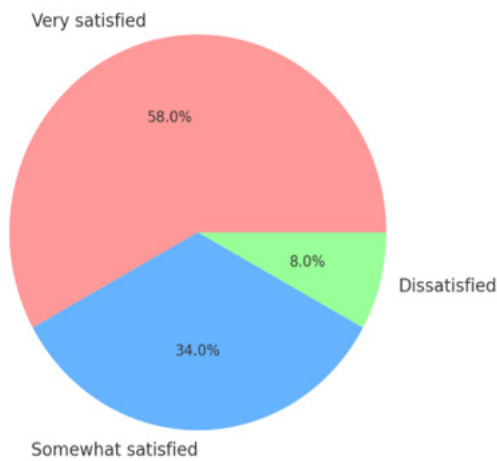
The cosine similarity result ranges from -1 to 1, where values closer to 1 indicate high similarity, and values closer to -1 indicate low similarity. If the result is 0, it means the vectors are orthogonal, indicating no similarity. After computing the similarity scores between the user in question and all other users, the system filters out those with a similarity score greater than 0. The algorithm then returns a list of recommended property IDs based on the highest similarities.

Finally, an endpoint within the system architecture retrieves these recommended property IDs and converts them into JSON objects for the front-end. The front-end then displays the recommended properties first, followed by other properties, enhancing the user’s experience by showing the most relevant options upfront.

The proposed system could potentially serve as a tool to save time and effort, improve the accuracy of suggestions, increase user satisfaction, and optimize sales and profitability for real estate agents and brokers. This study significantly contributes to understanding the evolution of recommendation systems in the real estate sector by presenting a robust and scalable system architecture that could be adapted to different languages, domains, or contexts where similar systems might be needed.

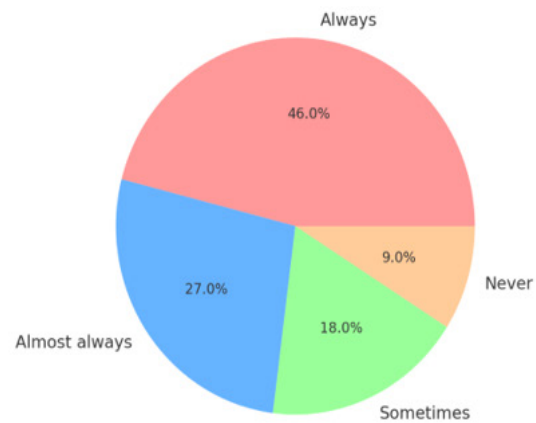
A user experience satisfaction measurement tool was applied to 19 users with the aim of qualitatively assessing the tool. The perception of user satisfaction, the quality of the recommendations, and whether or not they would recommend using the system were the aspects evaluated and are shown in figure 6.

How satisfied are you with the system's usability?



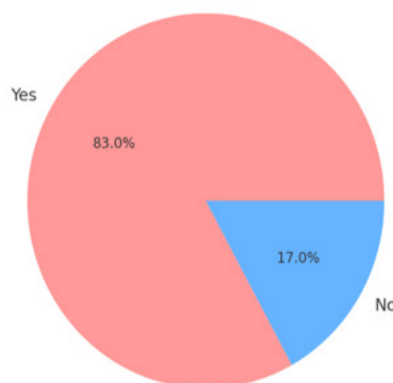
a)

Did the system generate the recommendations you wanted?



b)

Would you recommend the real estate website to others?



c)

Figure 6. Result of some of the perception evaluations carried out by users

CONCLUSIONS

The study’s approach to integrating collaborative filtering with content-based methodologies has shown that hybrid models can meet the unique issues of real estate transactions, notably in terms of user happiness and engagement.

The exploratory examination of the literature showed limitations of typical recommendation systems. It is clear that hybrid systems provide a more robust solution, particularly in contexts with sparse interaction data and the cold-start problem. The proposed model effectively addresses these challenges by combining the characteristics of both filtering techniques.

The deployment and evaluation of the system demonstrated a good response from users, as showed the satisfaction rates and favorable perception assessments as well as the system's effectiveness in providing personalized recommendations. However, one of the most significant constraints identified was the possibility of bias in user interaction data, which could impair the system's generalizability across different markets and user location. Furthermore, while the hybrid model managed successfully in the real state case, its applicability to other sectors needs to be more examined. Future research could look into the integration of advanced machine learning techniques, such as deep learning. Finally, while the system's modular architecture is advantageous for this specific application, it may need to be modified to fit other deployment contexts. The continuous update approach used in the data gathering module, for example, proved helpful in keeping the system relevant to current market trends, but it also presented issues with data consistency and processing overhead.

REFERENCES

1. Statista.. Data growth worldwide 2010-2025. <https://www.statista.com/statistics/871513/worldwide-data-created/>
2. Salehi M, Nakhai Kamalabadi I, Ghaznavi Ghouschi MB. Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering. *Education and Information Technologies*. 2014;19(4):713-35. <https://doi.org/10.1007/s10639-012-9245-5>
3. Pence HE. What is Big Data and Why is it Important? *Journal of Educational Technology Systems*. 2014;43(2):159-71. <https://doi.org/10.2190/ET.43.2.d>
4. Lü L, Medo M, Yeung CH, Zhang YC, Zhang ZK, Zhou T. Recommender systems. *Physics Reports*. 2012;519(1):1-49. <https://doi.org/10.1016/j.physrep.2012.02.006>
5. Ledesma F, Malave González BE. Patrones de comunicación científica sobre E-commerce: un estudio bibliométrico en la base de datos Scopus. *Región Científica*. 2022;1(1):202214. <https://doi.org/10.5863/rc202213>
6. Venkatesh C, Oberoi H, Goyal A, Sikka N. RE-RecSys: An End-to-End system for recommending properties in Real-Estate domain. *Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD)*. New York, NY, USA: Association for Computing Machinery; 2024. p. 558-62. (CODS-COMAD '24). <https://doi.org/10.1145/3632410.3632487>
7. Nguyen T, Vu S, Nguyen T, Pham V, Nguyen H. Design a Recommendation System in Real Estate Investment Based on Context Approach. *SCITEPRESS*; 2023. p. 255-63. <http://www.scitepress.org/Papers/2023/122108>
8. Li J, Yang C, Ye G, Nguyen QVH. Graph neural networks with deep mutual learning for designing multi-modal recommendation systems. *Information Sciences*. 2024;654:119815. <https://doi.org/10.1016/j.ins.2023.119815>
9. Corso G, Stark H, Jegelka S, Jaakkola T, Barzilay R. Graph neural networks. *Nat Rev Methods Primers*. 2024;4(1):1-13. <https://doi.org/10.1038/s43586-024-00294-7>
10. Zhang Y, Xiang T, Hospedales TM, Lu H. Deep Mutual Learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. p. 4320-8. https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Deep_Mutual_Learning_CVPR_2018_paper.html
11. Truong QT, Salah A, Lauw H. Multi-Modal Recommender Systems: Hands-On Exploration. *Proceedings of the 15th ACM Conference on Recommender Systems*. New York, NY, USA: Association for Computing Machinery; 2021. p. 834-7. (RecSys '21). <https://doi.org/10.1145/3460231.3473324>
12. Gharahighehi A, Pliakos K, Vens C. Recommender Systems in the Real Estate Market—A Survey. *Applied Sciences*. 2021;11(16):7502. <https://doi.org/10.3390/app11167502>
13. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies.

Health Information & Libraries Journal. 2009;26(2):91-108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>

14. Lika B, Kolomvatsos K, Hadjiefthymiades S. Facing the cold start problem in recommender systems. *Expert Systems with Applications*. 2014;41(4, Part 2):2065-73. <https://doi.org/10.1016/j.eswa.2013.09.005>

15. Badriyah T, Azvy S, Yuwono W, Syarif I. Recommendation system for property search using content based filtering method. 2018 International Conference on Information and Communications Technology (ICOIACT). 2018. p. 25-9. <https://doi.org/10.1109/ICOIACT.2018.8350801>

16. Nallamala SH, Bajjuri UR, Anandarao S, Prasad DDD, Mishra DP. A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems. *IOP Conference Series: Materials Science and Engineering*. 2020;981(2):022008. <https://doi.org/10.1088/1757-899X/981/2/022008>

17. Kweon W, Kang S, Jang S, Yu H. Top-Personalized-K Recommendation. *Proceedings of the ACM Web Conference 2024*. New York, NY, USA: Association for Computing Machinery; 2024. p. 3388-99. (WWW '24). <https://doi.org/10.1145/3589334.3645417>

18. Abodayeh A, Hejazi R, Najjar W, Shihadeh L, Latif R. Web Scraping for Data Analytics: A BeautifulSoup Implementation. 2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU). 2023. p. 65-9. <https://doi.org/10.1109/WiDS-PSU57071.2023.00025>

19. Chiu WY, Yen GG, Juan TK. Minimum Manhattan Distance Approach to Multiple Criteria Decision Making in Multiobjective Optimization Problems. *IEEE Transactions on Evolutionary Computation*. 2016;20(6):972-85. <https://doi.org/10.1109/TEVC.2016.2564158>

20. Xia P, Zhang L, Li F. Learning similarity with cosine similarity ensemble. *Information Sciences*. 2015;307:39-52. <https://doi.org/10.1016/j.ins.2015.02.024>

FINANCING

This project was supported by Universidad del Magdalena.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION:

Conceptualization: Carlos Henriquez Miranda, German Sanchez-Torres.

Data curation: Carlos Henriquez Miranda, German Sanchez-Torres.

Formal analysis: Carlos Henriquez Miranda, German Sanchez-Torres.

Acquisition of funds: Carlos Henriquez Miranda, German Sanchez-Torres.

Research: Carlos Henriquez Miranda, German Sanchez-Torres.

Methodology: Carlos Henriquez Miranda, German Sanchez-Torres.

Project management: Carlos Henriquez Miranda.

Software: Carlos Henriquez Miranda, German Sanchez-Torres.

Supervision: Carlos Henriquez Miranda.

Validation: Carlos Henriquez Miranda, German Sanchez-Torres.

Display: Carlos Henriquez Miranda, German Sanchez-Torres.

Drafting - original draft: Carlos Henriquez Miranda, German Sanchez-Torres.

Writing - proofreading and editing: Carlos Henriquez Miranda, German Sanchez-Torres.