Design Approaches and Tools for the Implementation of a Medicinal Cocktails Recommendation System

Sviatoslav Tyskyi [a]*, Solomiia Liaskovska[b], Andy T. Augousti[c]

[a] *Department of Artificial Intelligence, Lviv Polytechnic National University, Kniazia Romana Street, 5, Lviv, 79905, Ukraine*

[b] *Department of Artificial Intelligence, Lviv Polytechnic National University, Kniazia Romana Street, 5, Lviv, 79905, Ukraine*

[c] *Faculty of Engineering, Computing and the Environment, Kingston University, Kingston, London, Room RV MB 215, Main Building (RV), Roehampton Vale, United Kingdom*

Abstract

The development of a recommendation system for medical cocktails is becoming increasingly relevant in today's world, where health and well-being are becoming a higher priority. Such a system allows patients to receive personalized recommendations for the use of therapeutic beverages, contributing to the improvement of their health and quality of life. The aim of the article is to create a Recommendation System that provides functionality for convenient search of Modern Medicinal Cocktails recipes, based on the user's available ingredients and his tastes. The object of the research is a system of recommendations for a small sample of data and the application of this system. As a result of the research, a Recommendation System was created that allows users to quickly and efficiently find a recipe of Medicinal Cocktails with personalized health recommendations involves a deep understanding of health and medical information, and it's important to ensure the advice given is accurate and safe for users. The Recommendation System also provides personalized recommendations based on the user's preferences. The obtained results confirm the effectiveness of the proposed approach and allow us to recommend the use of the developed Telegram bot for quick and convenient search of cocktail recipes with personalized recommendations.

Keywords 1

Recommendation System, Medicinal Cocktails Recipes, personalized recommendations, Python, data processing, data analysis, artificial intelligence.

1. Introduction

Many people use a Recommendation System to read the actual information about healthcare, life being, entertainment and other useful services. At the same time all information about healthcare and healthy lifestyle have become increasingly popular, especially among people with high activity life. So a Recommendation System for Medicinal Cocktails will became popular among dietitians, nutritionists and people how spent healthy life.

Developing a Recommendation System for Medicinal Cocktails recipes with intelligent recommendations can have a significant impact on the health and life balance. In many countries, the culture of preparing drinks for health that involves in person self-feeling, energy is very important, and learning recipes and how to make cocktails is part of that culture.

Thanks to the Recommendation System, people should know more about Modern Medicinal Cocktails culture and the right way to consume it. Recommendations and advice on the selection of ingredients and preparation methods can help users understand how to properly enjoy cocktails without harming their health and combine the ingredients correctly[1-3].

IDDM'2023: 6th International Conference on Informatics & Data-Driven Medicine, November 17 - 19, 2023, Bratislava, Slovakia

EMAIL: sviatoslav.tyskyi.mknssh.2023@lpnu.ua (A. 1); solomiya.y.lyaskovska@lpnu.ua (A. 2); augousti@kingston.ac.uk (A. 3)
ORCID: 0009-0006-7773-6560 (A. 1); 0000-0002-0822-0951 (A. 2); 0000-0003-3000-9332 (A. 3)

In addition, the Recommendation System can become a means of popularizing non-alcoholic cocktails, which can be included in the culture of alcohol consumption and change people's attitude towards alcoholic beverages[]. Increasing user awareness of soft drinks can help reduce alcohol consumption and the production of alcoholic cocktails, thereby reducing the health effects of alcohol and improving people's quality of life.

Therefore, it can be said that the development of such system has the potential to influence the culture of drinking no alcoholic beverages, make this process more known and conscious, and provide a healthier and more appropriate lifestyle.

The main purpose of this paper is to create a Telegram bot for cocktail recipes with intelligent recommendations based on machine learning [4-6]. The implementation of such a product allows users to quickly and efficiently find a cocktail recipe by name, ingredients or type of drink, as well as receive personalized recommendations based on their search history and preferences. To achieve the goal, the following tasks must be solved:

1. Collection and processing of cocktail recipe data, including name, ingredients, proportions, description, and images [7,8].
2. Development of an algorithm for searching and filtering recipes by name, ingredients or type of drink.
3. Implementation of a machine learning model to analyze the user's search history and preferences and provide personalized recommendations [9-11].
4. Integration with Telegram API to provide interaction with users and give them access to work functions.
5. Development of a user interface that ensures the convenience of interaction with the bot and allows you to find a cocktail recipe, get recommendations, and save your favorite recipes.
6. Checking and adjusting the operation to ensure its stable and trouble-free operation.
7. Testing and improvement of work based on feedback received from users.

The object of study:

The object of the study is a Recommendation System for Medicinal Cocktails [10] with smart recommendations.

The subject of study:

The subject of research is the development and implementation of machine learning algorithms for the selection of personalized recommendations, as well as interaction with API and process processing and input data analysis [11-14] to create an effective and convenient Telegram bot for Medicinal Cocktails.

2. Content-Based Filtering

Content-Based Filtering (CBF) is one of the approaches to solving the problem of recommendations [1]. This approach was designed to provide personalized recommendations based on information about items that the user has liked in the past.

2.1. The mechanism of content-oriented filtering

The CBF method basically uses an item's description (or its "content") to recommend similar items. So, for example, if a user liked certain movies with a specific actor in the past, the CBF system will recommend other movies with that actor to the user.

Content-based filtering makes recommendations using keywords and attributes assigned to objects in the database (such as products in an online marketplace) and matches them to a user's profile. A user profile is created based on data obtained from user actions such as purchases, ratings (likes and dislikes), downloads, searching for products on the website or adding them to the cart, and clicking on links and products.

The basic concept behind CBF is that a user will like items that are similar to those he has already rated highly in the past. Therefore, the main task of CBF is to identify items that are similar to those that the user liked.

Similarity can be determined using similarity metrics [2], which are mathematical measures used to determine how similar vectors are to each other. Different similarity metrics are used in different contexts. Here are some of the main ones:

Euclidean distance: This metric defines the distance between two points in space. The smaller the Euclidean distance, the more similar the objects.

$$||p-q|| = \sqrt{\sum_{i=1}^{n}(p_i-q_i)^2} \tag{2.1}$$

where $p = (p_1, p_2 \dots p_n)$, $q = (q_1, q_2 \dots q_n)$ - vectors.

Cosine similarity: These metric measures the angle between two vectors and is used to determine similarity in a high-dimensional space. The closer the angle is to 0, the greater the similarity.

$$cos\,\theta = \frac{AB}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{2.2}$$

where A, B – vectors.

Manhattan distance: This metric, also known as the L1-norm, measures the sum of the absolute differences between the coordinates of two points.

$$d_1(p,q) = ||p-q||_1 = \sum_{i=1}^{n}|p_i-q_i| \tag{2.3}$$

where $p = (p_1, p_2 \dots p_n)$, $q = (q_1, q_2 \dots q_n)$ - vectors.

Jacquard measure: This metric is used to determine the similarity between two sets and measures the size of the intersection of the sets divided by the size of the union of the sets.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.4}$$

where A, B – sets.

Pearson's metric: This metric is used to measure the statistical correlation between two vectors. It takes values from -1 to 1, where 1 means full positive correlation, -1 means full negative correlation, and 0 means no correlation.

$$r_{xy} = \frac{\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i-\bar{x})^2\sum_{i=1}^{m}(y_i-\bar{y})^2}} \tag{2.5}$$

where $\bar{x}$ та $\bar{y}$ – sample averages.

$$r_{xy} \in [-1; 1]$$

2.1.1. Usage of content-oriented filtering

CBF is often used in recommender systems, such as movie, music, book, news, search engine recommendations, etc. This approach can be very effective when detailed item descriptions are available.

Once a user has performed a few searches, browsed a few items, or made a few purchases, the content-based filtering system can start making relevant recommendations. This makes it ideal for businesses that don't have a large user base to collect data from. It also works well for sellers who have many users but few user interactions in specific categories or niches.

2.1.2. Advantages and disadvantages of content-oriented filtering

The recommendations correspond to the interests of the user. Recommender systems based on content can be adapted to the interests of the user, including recommendations for niche products, since the method is based on comparing the characteristics or attributes of the database object with the user's profile. No input from other users is required to start issuing recommendations. Unlike collaborative filtering, content-based filtering does not require input from other users to generate recommendations.

For example, content-based filtering recognizes a user's specific preferences and tastes, such as hot sauces made in Texas with organic Scotch Bonnet peppers, and recommends products with the same attributes. Content-based filtering is also valuable for businesses with a large number of products of the same type, such as smartphones, where recommendations must be based on many different characteristics.

Recommendations are transparent to the user. Highly relevant recommendations create a sense of openness for the user, strengthening their level of trust in the recommendations offered. For example, in collaborative filtering, there are more cases where users do not understand why they see certain recommendations. For example, let's say a user group that bought an umbrella also buys down coats.

Collaborative filtering creates a potential "cold start" situation where a new website or community has few new users and lacks connections between users. So it is avoiding the "cold start" problem. Although content-aware filtering requires some initial input from users to start making recommendations, the quality of early recommendations is usually better than a collaborative system that needs to add and correlate millions of data points before becoming optimized.

Content-based filtering systems are usually easier to create. The technical implementation of creating a content-based filtering system is relatively simple compared to collaborative filtering systems designed to simulate user-to-user recommendations.

On the other hand, CBF has several disadvantages. First, it is generally considered less accurate than collaborative filtering methods. Second, it may miss some relevant items if they are not similar enough to items already rated by the user. Third, it can lead to "personal tumor", when the system recommends only very similar items, and the user does not get a new experience.

Given this, CBF may be a better choice in some scenarios, especially when detailed item descriptions are available or when user interaction data is lacking. However, to maximize the accuracy and variety of recommendations, CBF is often combined with other approaches, such as collaborative filtering, in hybrid recommender systems.

2.2. Using Collaborative Filtering for creating a Recommendation system

Collaborative filtering is one of the widely used methods of recommender systems. This technology is designed to solve the problems of information personalization in conditions of a large amount of data, in particular, on the Internet.

2.2.1. The working mechanism of collaborative filtering

Collaborative filtering looks exclusively at historical interactions between customers and the products they've used to recommend new products. The details of the element itself are not of great importance, because all information about how the user interacts with this element is stored in a special repository - the matrix of the user's interaction with the element.

Collaborative filtering is based on the idea that if two people have agreed on some issues in the past, they are more likely to have the same opinion in the future. Technology can be divided into two main types: user-based and item-based. User-based collaborative filtering looks for users who have similar ratings to the current user and uses their ratings for recommendation. Item-based collaborative filtering, on the other hand, analyzes the items that have been rated by the current user and looks for items that are similar to those that the user has rated highly.

2.2.2. Usage of collaborative filtering

Collaborative filtering is widely used in various fields, including e-commerce, online streaming services, social networks, and many others. For example, Amazon uses item-based collaborative filtering to recommend products, and Netflix uses it to recommend movies and series.

Collaborative filtering works best when there is enough data about user behavior. It is also better suited to situations where user interests change over time because it can adapt to those changes. If it is important to take into account the interaction between users, collaborative filtering with its algorithms can be a very useful tool.

2.2.3. Matrix factorization

Matrix factorization [3] is a popular approach in the field of Recommendation system, which is commonly used in collaborative filtering. The main idea of this method is to decompose a large matrix into two smaller matrices, which, when multiplied, will be as close as possible to the original matrix.

For example, in the case of a recommender system, a large user-product matrix might be created, where the rows correspond to users, the columns correspond to products, and the values in the matrix represent the ratings that users have given to products. Since many users rated only a small fraction of the products, this matrix is usually very sparse.

Matrix factorization decomposes this matrix into two smaller matrices: the user-factor matrix and the product-factor matrix. Each row of the user-factor matrix represents a "factor profile" of the user, and each row of the product-factor matrix represents a "factor profile" of the product. "Factors" here are latent (or hidden) properties that are determined during the factorization process, and may represent abstract concepts that reflect product properties and user interests.

The rating that the user gives to the product is modeled as a scalar product of the corresponding factor profiles. This model is able to fill in the gaps in the original matrix by predicting scores for user-product pairs that were not previously scored.

2.2.4. Advantages and disadvantages of collaborative filtering

There are some advantages and disadvantages of using collaborative filtering. Let’s analyze the advantages of collaborative filtering. Firstly it can help users discover new interests by recommending new items similar to their interests. Also, it does not require detailed characteristics and contextual data of products or items. All that is required is the user-item interaction matrix to train the matrix factorization model.

There are some disadvantages: data sparsity can make it difficult to recommend new products or users because recommendations are based on historical data and interactions. As the user base grows, the algorithms face a performance problem due to the large amount of data and lack of scalability.
The second disadvantage is “Lack of diversity in the long term”. Because the algorithms work based on historical ratings, they will not recommend items with little or limited data. Popular products will become even more popular in the long run and there will be a lack of new options. Suffers from the 'cold start' problem is the third disadvantage that we analyzed.

2.3. Hybrid filtering and it’s using for the Recommendation system

Hybrid filtering was invented to overcome the limitations of individual recommender system approaches, such as content-based and collaborative filtering. It combines the characteristics of both approaches, trying to use their advantages and avoid their disadvantages.

2.3.1. Types of hybrid filtration

There are several types of hybrid [4] systems, depending on how they combine content-oriented and collaborative approaches. For example, they may include weighted, overlapped, switched, mixed or hybrid models. Let’s analyze each of type:

Combination method: In this method, content-oriented and collaborative predictions are generated separately and then combined. This can be important if you have a large amount of data that can be processed in parallel.

Fusion: In this method, the attributes obtained from each method are combined and used to support recommendations. This can be useful if you want to combine the benefits of both methods.

Cascade method: One method is used to generate a ranked list of recommendations, and then a second method is used to refine that list. This can be useful if the first method is able to generate a rough list quickly, while the second method can take longer but provide more accurate recommendations.

Feature method: One of the representations (content-oriented or collaborative) is used as input data for the other. For example, a content-oriented system can be used to create a set of characteristics for each user, which are then used in a collaborative manner.

Blended recommendations: Collaborative and content-aware filtering recommendations are simply mixed together. This can be useful if you want to present a wide range of recommendations to users.

Switching: The system switches between collaborative filtering and content-oriented filtering depending on the situation. For example, if the system has enough information about the user for collaborative filtering, it can use this method. Otherwise, it can use content-based filtering.

Ensemble: In this method, content-based and collaborative predictions are combined in the model training phase, usually using machine learning techniques such as decision trees, neural networks, or regression.

2.3.2. Advantages and disadvantages of hybrid filtering

Let's consider the advantages of hybrid filtering:

- Overcoming the "cold start" problem that occurs in the absence of user interactions.
- Ability to recommend more diverse items.
- Greater accuracy of recommendations compared to individual methods.
- Ability to work with different types of data.

The main disadvantages of hybrid filtering: complexity of implementation, as this approach requires knowledge of various filtering methods; the possibility of overloading the system, since a significant amount of data needs to be processed. The last disadvantage of hybrid filtering is loss of transparency in the decision-making process, as different methods may have conflicting recommendations.

3. The effective methods of implementing the system of recommendations for Modern Medicinal Cocktails

The experimental part, in which filtering methods are applied to real data and the results obtained are analyzed. The most effective methods of implementing the system of recommendations for cocktails were determined, and their implementation was evaluated in the context of the specified metrics.

3.1. Content-Based Filtering

Using Python and the Pandas library for data processing, a content-oriented recommender system model was created. In this model, each object is represented by a vector characterizing its attributes. The similarity metric is used to compare vectors of different objects. Two main similarity metrics were implemented to compare their performance: Euclidean distance and sinusoidal similarity.

Using these metrics to determine the similarity between objects, recommendation lists were obtained for each user and the recommendations were evaluated using the accuracy metrics RMSE, MAE, MSE, MAPE, MRE (Table 1).

Table 1
Comparison of similarity metrics

	Sinusoidal similarity	Euclidean distance
MAE	0.7749	0.7725
MSE	0.9468	0.9471
RMSE	0.9730	0.9731
MAPE	0.2568	0.2567
MRE	0.002568	0.002567

Based on the accuracy metrics, sinusoidal similarity (2.2) was determined as the best. It will be used in the future.

3.1.1. Collaborative filtering

After implementation and analysis of content-oriented filtering, experiments with collaborative filtering were conducted. In this part of the study, matrix factorization was used to implement collaborative filtering.

It was used to compare different matrix factorization algorithms, such as SVD, SVD++, NMF, KNN Basic, KNN Means, KNN ZScore, CoClustering, to implement collaborative filtering and compare their effectiveness. Each algorithm has its own characteristics and may be better suited for certain types of data.

Using the RMSE and MAE evaluation metrics, the quality of the recommendations generated by each algorithm was evaluated and the one that best fit the model was selected. After selecting the best matrix factorization algorithm, it will be integrated with the content-oriented model to create a hybrid recommender system.

In order to determine the best algorithm, it is first necessary to choose the hyperparameters so that the algorithms give the best result.

Thus, for algorithms of the KNN family, it is necessary to choose the optimal value of the parameter k. For this purpose, several values of the k parameter were selected and evaluated using the RMSE metric and the MAE metric the optimal value of parameter k according to accuracy metrics is 30.

After the selection of hyperparameters, a comparison of matrix factorization algorithms was carried out. The results are recorded in Table 2.

Table 2
Accuracy of matrix factorization algorithms

	The Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
CoClustering	1.211	0.9708
KNN Means	1.1453	0.9226
SVD	0.9953	0.8063
KNN Basic	1.0296	0.8357
SVDpp	1.008	0.8195
NMF	1.257	1.0167
KNN ZScore	1.1591	0.9345
SlopeOne	1.202	0.9572

For clarity, graphs were drawn comparing the accuracy of the algorithms using the RMSE metric (Fig. 1) and the MAE metric (Fig. 2).

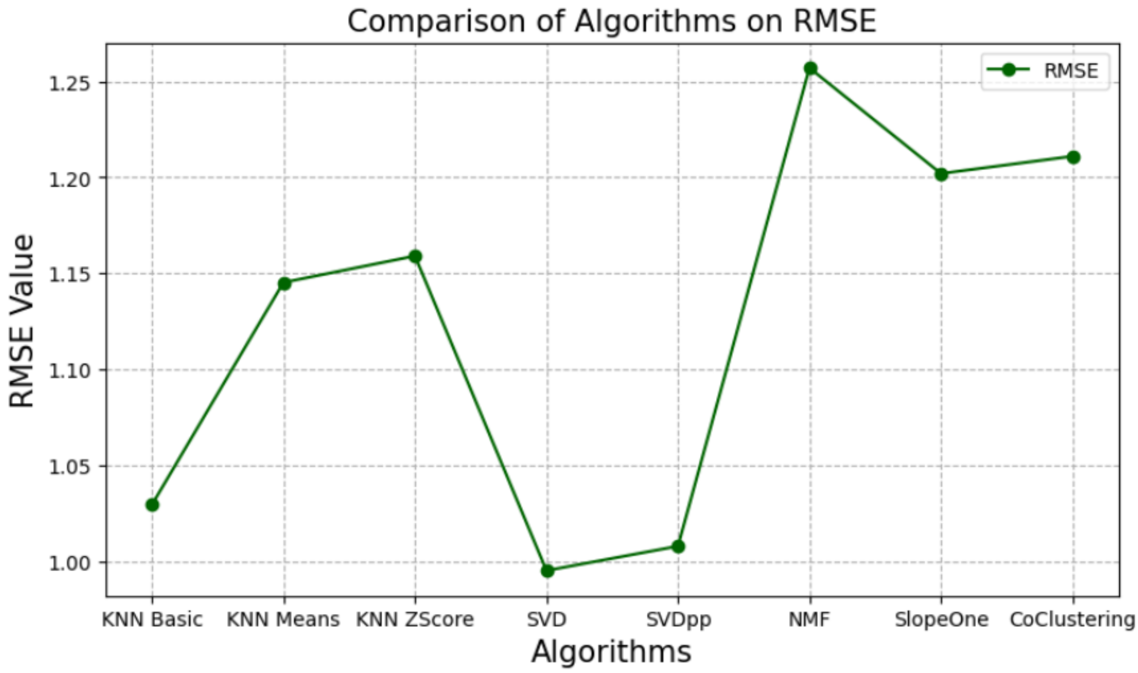

Figure 1: Graph comparing the accuracy of factorization algorithms using the RMSE metric

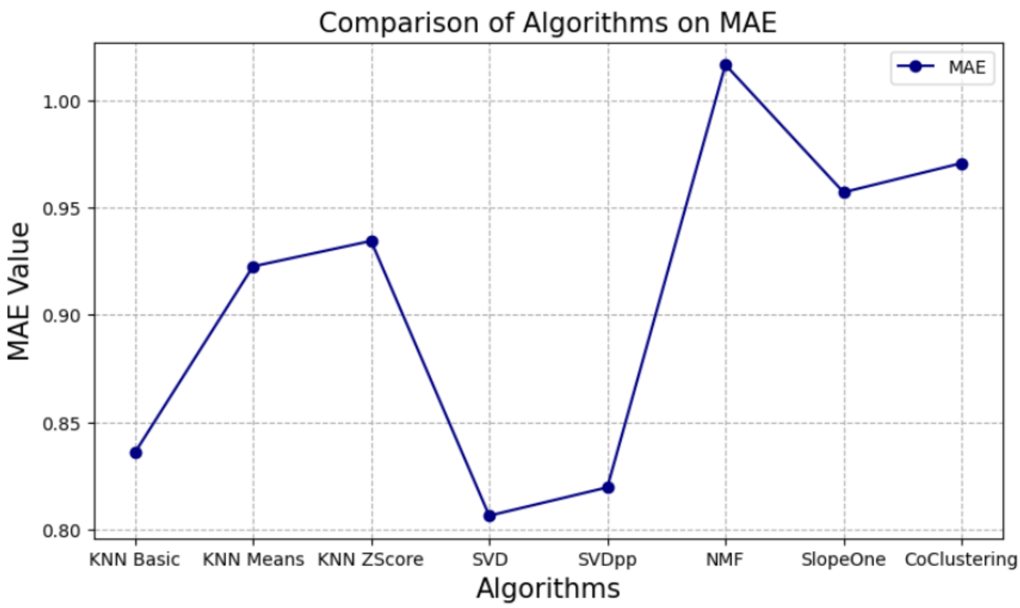

Figure 2: Graph comparing the accuracy of factorization algorithms using the MAE metric

So, from the above graphs, it is possible to determine the SVD algorithm as the most accurate among matrix factorization algorithms. However, even for the SVD algorithm, the error is 0.99, which can be lower due to the combination of collaborative and content-based filters.

3.1.2. Hybrid filtering

After individual implementation of content-oriented and collaborative filtering, both models were combined to create a hybrid recommender system. A hybrid system combines the strengths of both approaches and compensates for their weaknesses

In our case, content-based filtering was used to generate a base list of recommendations based on the characteristics of products that the user had previously evaluated. Thanks to this, the relevance of recommendations to the user's interests is ensured.

However, to ensure diversity of recommendations and take into account the behavior of other users, collaborative filtering is additionally used. To determine the best combination of models, an experiment was conducted in which the accuracy of all models was compared using the RMSE and MAE metrics. The results of the comparison are shown in Table 3.

Table 3
Accuracy of matrix factorization algorithms

	The Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
CoClustering	0.6102	0.3060
KNN Means	0.6207	0.3131
SVD	0.6141	0.3099
KNN Basic	0.6192	0.3122
SVDpp	0.6243	0.3160
NMF	0.6260	0.3135
KNN ZScore	0.6208	0.3129
SlopeOne	0.6244	0.3143

For clarity, the graphs comparing the accuracy of the algorithms using the RMSE metric (Fig. 3) and the MAE metric (Fig. 4) are plotted.

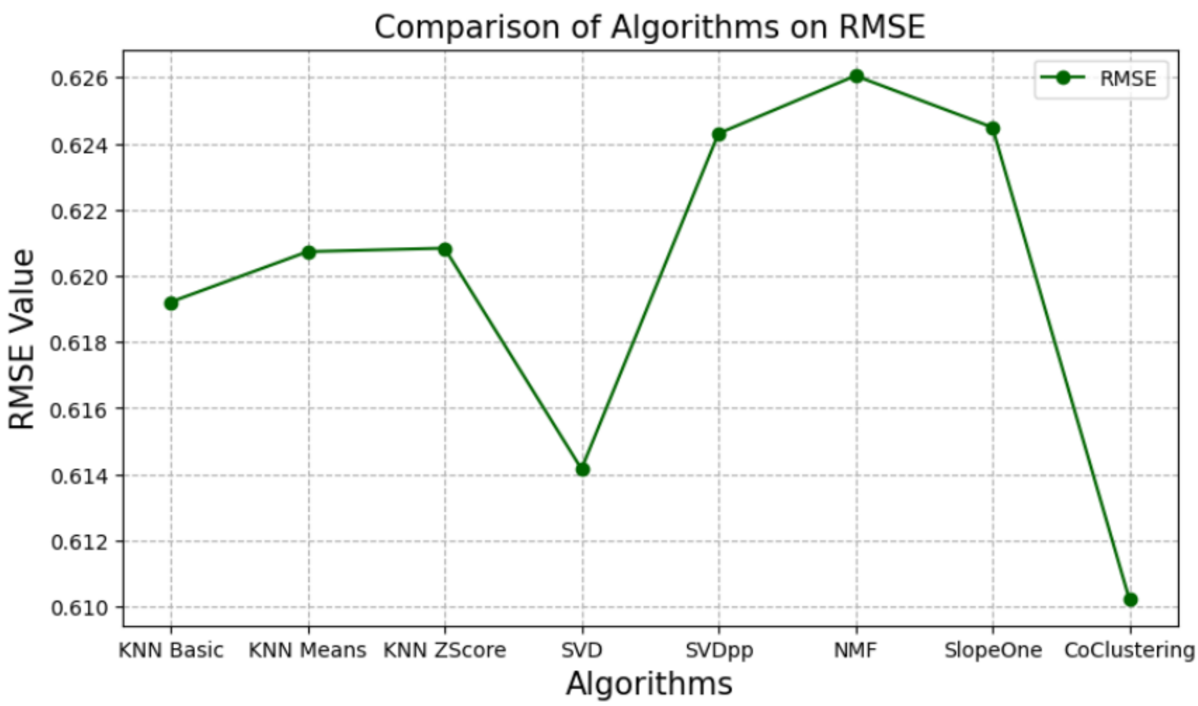

Figure 3: Graph comparing the accuracy of hybrid algorithms using metrics RMSE

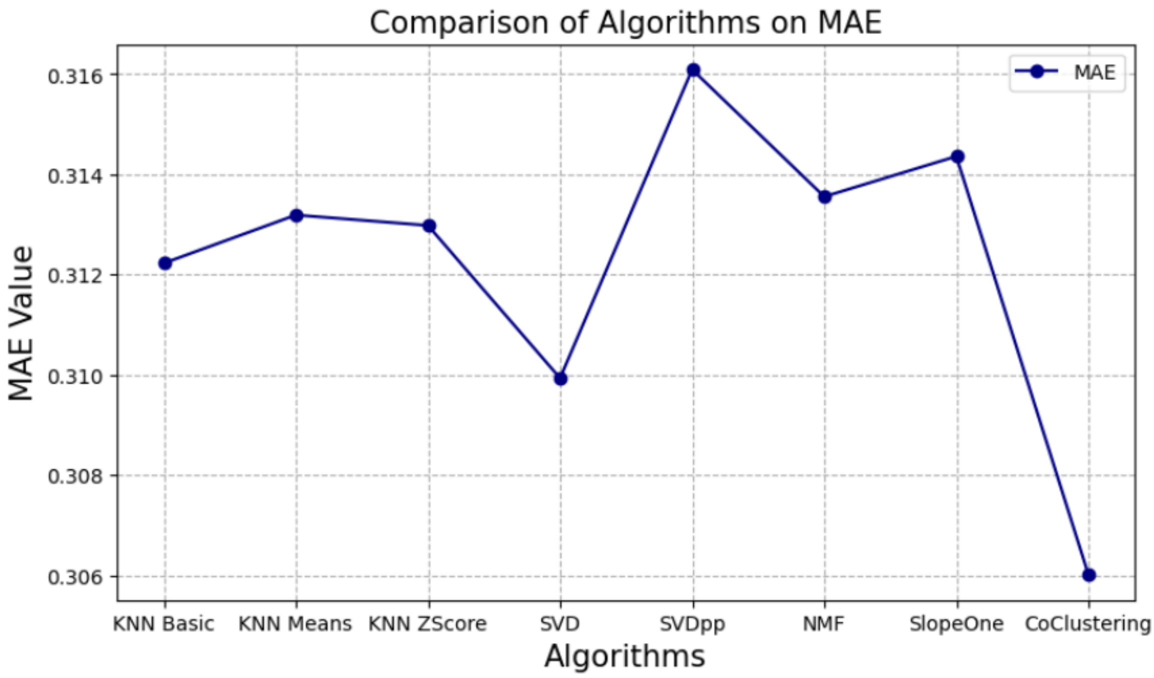

Figure 4: Graph comparing the accuracy of hybrid algorithms using metrics MAE

As a result, recommendations are obtained that more accurately correspond to the user's interests, and at the same time take into account the behavior of other users. The CoClustering algorithm with a content-based filter turned out to be the most effective.

3.1.3. Conclusion

In the context of collaborative filtering, eight methods of matrix factorization were considered, each of which presents its own characteristics and advantages. Regardless of the choice of a specific method, collaborative filtering demonstrates high efficiency in the presence of a sufficient number of user interactions.

A hybrid system that uses both methods has shown high efficiency, allowing to use the advantages of both approaches and avoid their disadvantages. The hybrid model allows you to take into account the interests of the user, the behavior of other users, and also provide various recommendations.

So, each of the considered approaches has its own strengths and weaknesses, but hybrid filters performed best.

4. References

[1] Roy, A. Introduction To Recommender Systems- 1: Content-Based Filtering And Collaborative Filtering Available online: https://towardsdatascience.com/introduction-to-recommender-systems-1-971bd274f421 (accessed on 21 May 2023).

[2] Chen, D. Recommendation System — Matrix Factorization Available online: https://towardsdatascience.com/recommendation-system-matrix-factorization-d61978660b4b (accessed on 21 May 2023).

[3] Christopher Tierney, Erica Brod Apotheke: Modern Medicinal Cocktails Hardcover, November 24, 2020, 240 p.

[4] Lesia Mochurad, Andrii Dereviannyi, Uliana Antoniv. Classification of X-Ray Images of the Chest Using Convolutional Neural Networks. IDDM 2021 Informatics & Data-Driven Medicine. Proceedings of the 4th International Conference on Informatics & Data-Driven Medicine. Valencia, Spain, November 19 - 21, 2021. 269-282.

[5] Yaroslav Tolstyak, Myroslav Havryliuk "An Assessment of the Transplant's Survival Level for Recipients after Kidney Transplantations using Cox Proportional-Hazards Model", Proceedings of the 5th International Conference on Informatics & Data-Driven Medicine, Lyon, France, November 18 - 20, CEUR-WS.org, 2022. pp. 260-265.

[6] Paolo Giudici, Silvia Figini. Applied Data Mining for Business and Industry. - Wiley, 2009. - 260 p.

[7] Havryliuk, M., Dumyn, I., Vovk, O. Extraction of Structural Elements of the Text Using Pragmatic Features for the Nomenclature of Cases Verification. In: Hu, Z., Wang, Y., He, M. (eds) Advances in Intelligent Systems, Computer Science and Digital Economics. Lecture Notes on Data Engineering and Communications Technologies, vol 158. pp 703–711, 2023, Springer, Cham. https://doi.org/10.1007/978-3-031-24475-9_57

[8] Lesia Mochurad. "Optimization of Regression Analysis by Conducting Parallel Calculations". COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine, 982-996 p.

[9] Liaskovska, S., Martyn, Y. Investigation of Anomalous Situations in the Machine-Building Industry Using Phase Trajectories Method Lecture Notes in Networks and Systems this link is disabled, 463 LNNS, pp. 49–59, 2022.

[10] Yaroslav Tolstyak, Valentyna Chopyak, Myroslav Havryliuk, An investigation of the primary immunosuppressive therapy's influence on kidney transplant survival at one month after transplantation, Transplant Immunology, 2023, https://doi.org/10.1016/j.trim.2023.101832.

[11] Zakir Hossain, A. K. M., Hassim, N. B., Alsayaydeh, J. A. J., Hasan, M. K., & Islam, M. R. A tree-profile shape ultra wide band antenna for chipless RFID tags. International Journal of Advanced Computer Science and Applications, 12(4), 546-550, 2021. doi:10.14569/IJACSA.2021.0120469.

[12] Lesia I. Mochurad " Canny Edge Detection Analysis Based on Parallel Algorithm, Constructed Complexity Scale and CUDA". Vol. 41 No. 4: Computing and Informatics, 957–980, 2022 https://doi.org/10.31577/cai_2022_4_957

[13] Indra, W.A., Zamzam, N.S., Saptari, A., Alsayaydeh, J.A.J, Hassim, N.B., 2020." Development of Security System Using Motion Sensor Powered by RF Energy Harvesting", 2020 IEEE Student Conference on Research and Development, SCOReD 2020 9250984, pp. 254-258.

[14] Zakir Hossain, A. K. M., Hassim, N. B., Alsayaydeh, J. A. J., Hasan, M. K., & Islam, M. R. A tree-profile shape ultra wide band antenna for chipless RFID tags. International Journal of Advanced Computer Science and Applications, 12(4), 546-550, 2021. doi:10.14569/IJACSA.2021.0120469.