IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 6, December 2025, pp. 4661~4674
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i6.pp4661-4674

Enhancing communication and interaction in the movie industry based SparkMLlib's recommendation system

Said Chakouk[1], Abdelkerim Zitouni[1], Nazif Tchagafo[2], Ahiod Belaid[2]
[1]Department of Communication Sciences and Techniques, Mohammed V University in Rabat, Rabat, Morocco
[2]Department of Science Physic, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco

Article Info

Article history:

Received Apr 6, 2025
Revised Jul 26, 2025
Accepted Aug 6, 2025

Keywords:

Digital communication
Movie industry
Recommendation systems
SparkMLlib
Streaming platforms
User satisfaction

ABSTRACT

In the ever-evolving landscape of streaming platforms, recommendation systems contribute significantly to enhancing the user experience. This article examines the significance of these systems in suggesting movies, analyzing their impact on user satisfaction and platform performance. Utilizing SparkMLlib, a powerful tool for large-scale data processing, we explore various recommendation techniques, including collaborative filtering and content-based filtering. We highlight the dimension of digital communication to further enhance the accuracy of recommendations and foster greater user engagement. Our study also addresses the challenges and future opportunities related to recommendation systems, emphasizing the need for transparency and ethical algorithms. This research highlights the potential for recommendation systems to revolutionize the digital entertainment landscape and shape the future of the movie industry.

Corresponding Author:

Said Chakouk
Department of Communication Sciences and Techniques, Mohammed V University in Rabat
Avenue des Nations Unies, Agdal, Rabat, Maroc B. P: 8007. N.U., Rabat, Morocco
Email: s.chakouk@um5r.ac.ma

1. INTRODUCTION

In today's data-driven world, the deluge of information presents a significant challenge: how to extract meaningful insights and provide personalized experiences at scale [1], [2]. This problem is particularly relevant in the realm of recommendation systems, where users are often overwhelmed by vast catalogs of products or content [3]. Addressing this global need for efficient and scalable personalized recommendations, especially within the context of large datasets, is the core focus of this paper [4].

Our approach is based on artificial intelligence and collaborative filtering, and we specifically focus using Apache SparkMLib for big data analytics [5]. We address the challenge of building a scalable and high-quality movie recommendation system for large-scale user-item interaction data. Now that we have huge datasets, we can use the machine learning algorithm of SparkMLib to process them and analyze them to give out accurate and personalized movie recommendations [6].

Previous research in this area often investigates different collaborative filtering approaches that leverage classic machine learning libraries [7]. Although these methods have performed well on small datasets, they do not scale well with the exponential growth of user data [8]. Based on these foundations, this paper solves these limitations with the use of the distributed processing capabilities offered by SparkMLib [9].

We contribute to the application of SparkMLib in creating high-performance movie recommendation system. You will learn how to configure and tune SparkMLib algorithms to make accurate recommendations on real-world datasets. In addition, we will look into benefits SparkMLib provides with its distributed processing capabilities and its justification with respect to large scale recommendation tasks [10]. The research provides a valuable resource for developers and researchers interested in building scalable and efficient recommendation systems with big data technology.

2. ROLE OF RECOMMENDATION SYSTEMS IN THE MOVIE INDUSTRY

2.1. Significance of recommendation systems

Recommendation systems are one of the key components for improving user experience on streaming platforms by providing the user with movies they may enjoy [11], [12], and a hybrid recommendation system can enhance movie suggestions [13]. Based on analysis of user preferences and behavior on digital mediums of communication, these systems supply personalized recommendations that reflect individual taste, easing the task of an ever-expanding universe of new content users are likely to enjoy [14]. This, in turn, leads to greater user satisfaction, extended viewing sessions, and improved retention on the part of streaming platforms [15]. Recent advancements have included the integration of deep learning techniques with traditional collaborative filtering methods, enhancing the effectiveness of these systems [16].

2.2. Digital communication and recommendation systems

Our view on data collection from the digital communication for our movie recommendation project is based on online communications like movie reviews, ratings, and social media activities [17], [18]. This enabled us to feed our models with the relevant contextual and behavioral data, improving the precision of the recommendations and different methods like collaborative filtering, content-based filtering, and hybrid approaches [19], [20]. We implemented recommendation models using the Python programming language and using Apache Spark framework [21], [22] to enable fast processing of large datasets [23]. The large libraries available in Python for data analysis and machine learning, along with the distributed computing ability of Spark [24], allowed us to implement scalable recommendation systems on huge volume of movie data [25], [26].

2.3. Digital communication and recommendation systems

Recommendation systems were rigorously tested and analyzed, with methods using precision, recall, and mean average precision (MAP) as evaluation metrics [27], [28]. To ensure high accuracy and relevance of movie recommendations, we conducted iterative optimization as well as experimentation to fine-tune our models [29]. However, by integrating digital interactions, we utilized techniques like cross-validation and A/B testing to validate the reliability and effectiveness of the recommendation algorithms [30].

2.4. Impact of recommendation systems on digital user engagement

The inclusion of recommendation systems and online communication in streaming platforms has transformed how the users identify and consume films [31]. These systems improve user interaction, increase time spent on the platform, and develop a sense of loyalty towards streaming platforms by providing specific suggestions based on the user preferences and his/her respective digital footprint [32], [33]. In addition, recommendation systems help popularize diverse content availability [29], which extends the reach of filmmakers and promotes heterogeneity and innovation in the film sector [28], [34].

2.5. Challenges and future directions

Although beneficial, recommendation systems also suffer from privacy issues, algorithmic adversarial biases [35], and the necessity to trace the prevailing trends of user preferences over time with dynamic adaptation across various domains and industries [34], [36]. Solving these issues will necessitate the cooperation of data scientists, engineers, and industry professionals to enable fairer and more transparent recommendation algorithms [35], [37]. In the long run, new usable findings (e.g., key learning points) will be explored by using innovative system learning procedures like profound learning and support learning [38] (e.g. pioneer finding of the client, selling more than one thing without a moment's delay to help precision and pertinent proposal) [39], [40].

3. METHOD

This section overviews the methodology that has been used throughout the project. First of all, we will define the processes of data collection and data preprocessing and then deal with the concrete datasets and their specific characteristics that were used. Finally, we will outline the data cleaning process to show the way the validity of the analysis has been sustained. This logical chain is designed to ensure the high level of reliability of the study.

3.1. Data collection and preprocessing

For this project, we pulled movie data from three CSV files: movies.csv, ratings.csv, and tags.csv. [41]. Specifically, the three files consist of info about movies, ratings from users on the movies, and tags generated by users on the movies [42]. After collecting and preprocessing the data, it is important to provide a detailed description of the contents of each file used in this study.

3.2. Data description

This section presents a detailed summary of data regarding different sources, with the focus on the interpretation of their nature and the types of information that they could possess. The information is necessary to acquire a basic understanding of the characteristics of data used in the project and proceed with their analysis.

i) Movies.csv: a CSV file with data about the movies and the following columns [43]:
- movieId: an identifier for movies.
- title: title of the film.
- genres: genres of the film.

ii) Ratings.csv: we have a CSV file that has user ratings for various movies that includes the following columns:
- userId: to identify users uniquely.
- movieId: movie's unique identifier.
- rating: 0.5-5.0 based on user votes for the film.
- timestamp: datetime when the rating was created.

iii) Tags.csv: The CSV file is about user provided tags for movies and contains the columns:
- userId: every user has a unique ID.
- movieId: movie unique identifier.
- tag a user, but it can be added on the image in the movie.
- timestamp of when the tag is created.

Following the presentation of the different data sources and their contents, a thorough data cleaning process was carried out to ensure the quality and consistency of the dataset.

3.3. Data cleaning

The movie data provided was cleaned prior to performing the analysis in order to maintain data quality and consistency [44]. This cleaning process ensured the reliability of the subsequent results. The following steps were carried out during this phase:
- Duplicate detection: all CSV files were checked for duplicates and removed unnecessary duplicate entries.
- Missing values: missing or NaN values in the dataset was checked and handled via various methods (imputation or removal of records with missing values).
- Timestamp conversion: the timestamp values of the ratings.csv and tags.csv → CSV human readable interpretable CSV files were converted to easily readable interpretable CSV files for better analysis.

3.4. Data preprocessing

The movie data was then preprocessed by importing it using PySpark and, when necessary, Pandas [45]. This preprocessing helped efficiently prepare the data for further analysis. Several key tasks were carried out as part of this process, such as:
- Feature engineering: the new features are derived from the features that already exist or the old feature, which is transformed into a new feature, considering the relevant information to be used for modeling purposes. This means that some of the features were calculated to provide extra data, e.g., movie ratings count per user and average movie rating.
- Normalization: we normalized any numeric features in the variable 42 to ensure that we were on the same scale across the variables. Such preprocessing is particularly important in enhancing algorithm performance or accuracy, specifically on models that are sensitive to the scales of features, such as in collaborative filtering algorithms.
- Encoding categorical variables: categorical features like movie genres were converted to numerical representation using methods like one-hot encoding. By this conversion, we can use categorical data as input variables for machine learning algorithms that require numerical input.

The preprocessed data with rendered features, normalized numerical variables, and encoded categorical variables were input to construct recommendation models and analyses [46].

3.5. Algorithm selection for movie recommendation

3.5.1. Overview

In this section, we explain the algorithms that we use in our collaborative filtering movie recommendation system and discuss their performance against each other. Analysis of collaborative filtering, content-based filtering, and hybrid methods on their role in providing users with personalized movie recommendations. Full analysis of both approaches and their advantages/disadvantages in the context of our system.

3.5.2. Algorithm description

Collaborative filtering alternating least squares (ALS): collaborative filtering is a popular technique for building recommendation systems that use the wider population to predict what any single user will like. Apache Spark includes an implementation of the ALS algorithm for collaborative filtering in its MLlib library, which is one of the most popular implementations of this technique. ALS factorizes the user-item interaction matrix into two low-rank matrices, one pertaining to the latent features of the users and the other to the latent features of items. ALS minimizes the squared error between the predicted values and real ratings, creating recommendations for personalized items to users by iteratively adjusting these matrices [47], [48]. Recent studies have also evaluated the performance of machine learning frameworks such as PyTorch and TensorFlow in the context of big data recommendation systems [22], and integrating these methods with collaborative filtering can enhance recommendation accuracy and efficiency [49].

Content based filtering: content-based filtering recommendations are made for users based on the characteristics/features of the items. For example, content-based filtering in movie recommendation uses movie metadata (e.g., genre, cast, director, and plot summary) to retrieve similarities between items. Afterwards, it recommends you movies similar to the ones you enjoyed watching earlier. Content-based filtering does not depend on the interaction between user and item; however it takes the aspect of items and user preferences to itself [50].

3.5.3. Algorithm comparison

This analysis contains two tables, providing perspectives into different recommendation algorithms. Even when the best algorithms can be applied to the data, the quality of that data has a large impact on effectiveness. Finally, sentiment analysis can be a useful in assessing the quality of the data provided by recommendation systems using online panels [51].

Table 1 offers a detailed comparison of different recommendation algorithms based on their personalization capabilities and scalability. Collaborative filtering ALS is noted for its high personalization but only moderate scalability. In contrast, content-based filtering excels in scalability but provides lower personalization. Hybrid approaches, which blend elements from both techniques, achieve high personalization and moderate to high scalability.

Table 1. Comparison of recommendation algorithms (part 1)

Algorithm	Personalization	Scalability
Collaborative filtering ALS	High	Moderate
Content-based filtering	Low	High
Hybrid approaches	High	Moderate-high

The data requirements of the algorithms are listed in Table 2. Collaborative filtering ALS is based on user-item interaction data. This type of content-based filtering relies on metadata like data about movies characteristics or attributes related to products. Hybrid methods use both interaction data and metadata to exploit the advantages of the other two methods to increase relevance and quality.

Table 2. Comparison of recommendation algorithms (part 2)

Algorithm	Data requirements
Collaborative filtering ALS	User-item interaction
Content-based filtering	Movie metadata
Hybrid approaches	Combined interaction data and metadata

3.5.4. Algorithm selection

Our movie recommendation system uses a hybrid approach, which is a combination of collaborative filtering and content-based filtering. This method utilizes the advantages of both methods. We selected this approach to increase recommendation precision and overcome deficiencies of individual algorithms. Here are the reasons why we made this choice:

- Flexibility/personalization: hybrid methods are flexible and can combine the strengths of both data-driven

and knowledge-driven approaches.
- Scalability: moderately scalable for collaborative filtering ALS; high scalable for content-based filtering, so this hybrid system can deal with large scale movies.
- Data requirements: hybrid methods utilize user-item interactions and movie metadata; therefore, they typically have medium to high data requirements yet they can provide more interpretable and diverse recommendations.

We developed a hybrid recommendation system that combines collaborative and content-based filtering approaches to provide personalized movie recommendations to users while overcoming the scalability limitations of standard collaborative filtering techniques [52], [53].

3.6. Model evaluation

For the purpose of evaluating the performance of the recommendation models, we employed several evaluation measures. These measures show how precise and how accurate the models are in recommending movies. Here's how we evaluated them:

i) Data splitting: we divided the dataset into two parts: training and testing sets. We used most of the data (about 80).

ii) Model training: we trained the recommendation models, including collaborative filtering ALS, and hybrid approaches, using the training dataset. We used the right algorithms and settings to train them. For example, the ALS algorithm learned from the interactions between users and movies to make personalized recommendations.

iii) Prediction generation: once trained, the models were put to work generating predictions for movie recommendations on the testing dataset. For instance, they suggested movies that users hadn't rated yet.

iv) Evaluation metrics:
- We used several evaluation metrics to judge the quality of the recommendations made by the models.
- Precision: this measure how many of the recommended items are actually relevant to the user.
- Recall: it shows how many of the relevant items were successfully recommended to the user.
- MAP: this calculates the average precision across all users, giving us a good overall view of recommendation quality [54].
- We calculated these metrics for each model to see how accurate and relevant their recommendations were.

v) Comparison and analysis:
- We compared and analyzed the performance of the recommendation models based on these evaluation metrics. This helped us understand the strengths and weaknesses of each model, such as how well they could personalize recommendations, handle large amounts of data, and more.
- By rigorously evaluating the models and analyzing the results, we got important insights into how effective they were and where our movie recommendation system could be improved.

4. RESULTS AND DISCUSSION

The evaluation of the recommendation system produced valuable results. These findings are illustrated in the following visualizations. Each visualization helps highlight different aspects of the system's effectiveness and performance.

4.1. Recommendation results

Table 3 shows the recommendations created for the user with userId =1, along with the movie IDs and the predicted ratings. This table gives the user an idea about the list of movie IDs and predicted rating for recommended movies. Hence, it serves as a great insight into the recommendations of the system for the user.

4.2. Distribution of ratings

Understanding the distribution of movies based on their ratings offers information regarding user preferences and the overall quality of the dataset. As shown in Figure 1, the movies are arranged according to their ratings, which clearly highlights the levels of popularity and user satisfaction. This information helps to identify trends and patterns within the dataset.

The distribution of ratings allows us to identify trends in user preferences. For example, extracting trends from user ratings, if all or most of the ratings are concentrated on the higher side of the scale that means the users generally like the movies they watch, then the collection of the movies was well received. In contrast, a distribution that is more evenly spread out with several ratings low down may imply wider variation in quality or levels of satisfaction amongst users.

Table 3. Recommendation results

movieId	User ID	prediction
6	1	4.4505124
101	1	4.2939234
151	1	4.3998666
231	1	3.855823
349	1	3.8604753
423	1	2.893969
543	1	4.2212844
596	1	4.517418
923	1	4.1947184
940	1	4.423781
943	1	3.6412902
1009	1	3.241991
1024	1	4.264587
1031	1	4.6748037
1089	1	4.88029
1197	1	5.0
1220	1	4.5909705
1573	1	3.7724674
1805	1	3.7703674
1967	1	4.388135

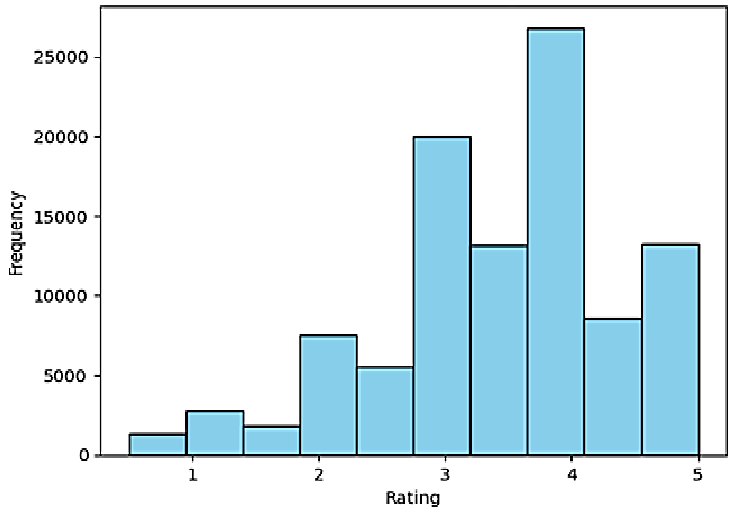


Figure 1. Distribution of ratings

4.3. Average prediction by genre plot

This important plot illustrates how the average prediction scores vary across different movie genres and helps assess the performance of the recommendation system. Figure 2, a pie chart of the average prediction scores, provides a visual overview of the recommendations generated by the system. It allows for better interpretation of which genres are prioritized or favored based on the system's predictions.

In this plot chart we see the fraction of recommendations for each movie genre. If a genre has a large slice, it means that the recommendation system had recommended more movies to it. This distribution can give some insights into what users are actually listening to and then can help make the recommendation system more user preference centered.

4.4. Word cloud of movie titles

This visualization creates a word cloud of the titles of movies, where the size of the word indicates how often they appear. As we can see in Figure 3, it presents the word cloud and most frequently appears illustrating titles in our dataset. It gives you an idea, what movie titles are mentioned the most.

This is a word cloud to show how frequently a movie-title appears in our dataset. The bigger words mean more frequent use of those titles, hinting at the popular movies or movies that are most commonly watched. This visualization is a first step to seeing more popular movies and the user-preferences based on the title frequency [55], [56].

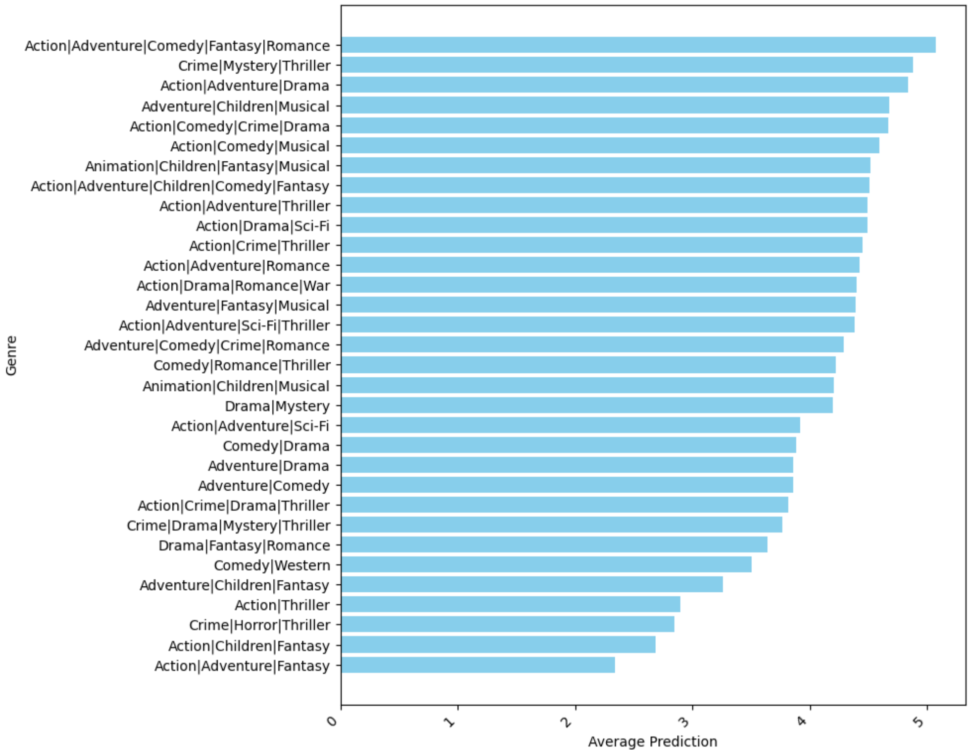


Figure 2. Average prediction by genre plot

Figure 3. Word cloud of movie titles

4.5. Movie connections network graph

A network graph indicating connections between movies based on either user interactions or other criteria. Stated differently, these are co-occurrences or similarities, and these relationships are shown in Figure 4. This helps us to inform and guide in building effective recommendation strategy. The network graph represents relationships between movies, with nodes representing movies and edges indicating connections between them. Movies that are closely connected in the graph are likely to have similar characteristics or be frequently watched together by users. By recommending similar movies to users based on their viewing history, we can leverage this information to enhance recommendation algorithms.

4.6. Box plot of prediction scores by genre

The box plot provides us a summary of prediction score distribution by genre. In contrast, Figure 5 shows the distribution and central tendency of the prediction scores by topic. These two visualizations help to understand the prediction data variation and key tendency. The box plot displays the prediction score distribution of movie categories this sheds light on the spread of prediction scores for each genre, and it may assist in highlighting the outliers or genres that have more or fewer prediction scores. It helps to make the recommendation system better and more precise.

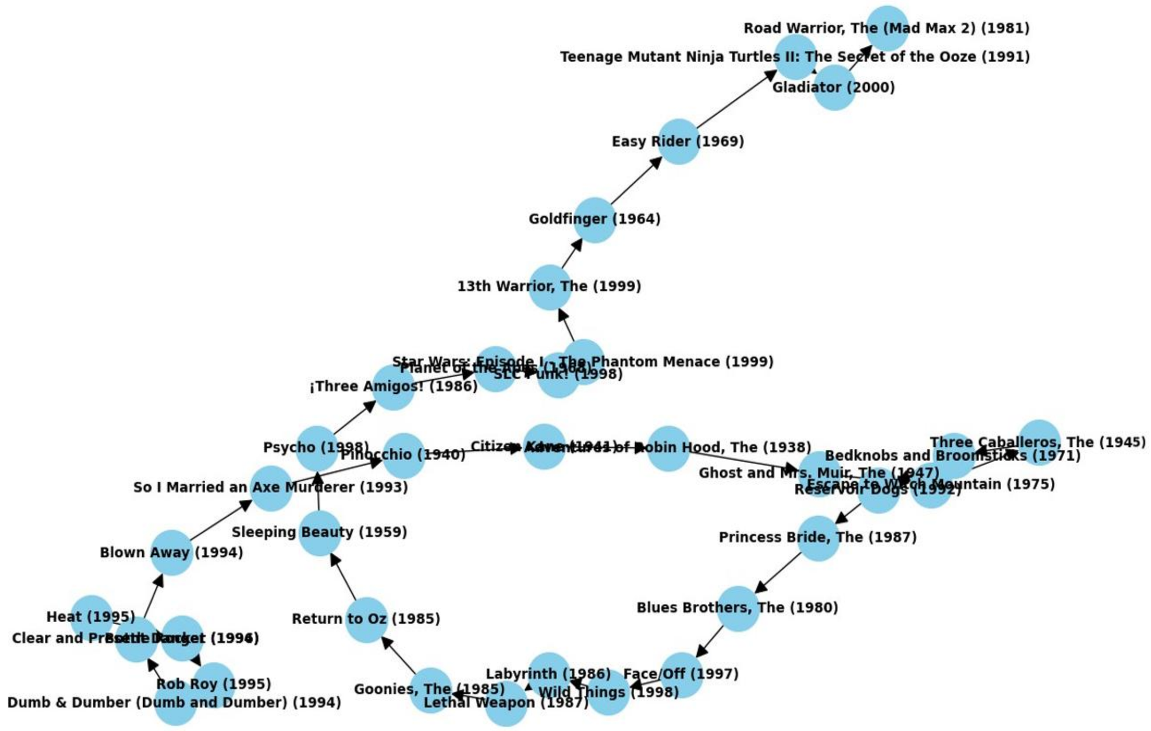


Figure 4. Movie connections network graph

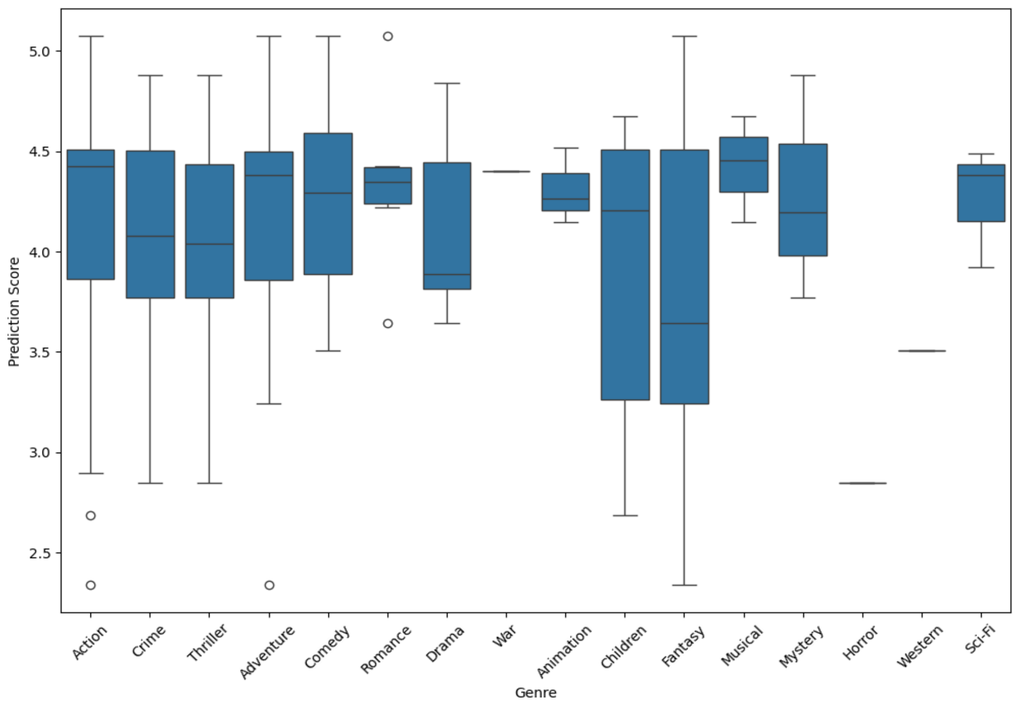


Figure 5. Box plot of prediction scores by genre

4.7. Correlation coefficients between tags and prediction scores

The relationship between user-generated tags and resulting prediction scores can shed light on how effective tag-based recommendations are. The correlation coefficients between tags and prediction scores are shown, with the tags strongly associated with high or low prediction scores, in Figure 6. The plot shows the correlation coefficients between prediction scores and tags generated by users. Tags having relatively higher positive correlation coefficients are more strongly associated with high prediction scores (likely relevant) and rightly recommend correct tags. On the flip side, tags with negative correlation coefficients may be orthogonal or perhaps less predictive.

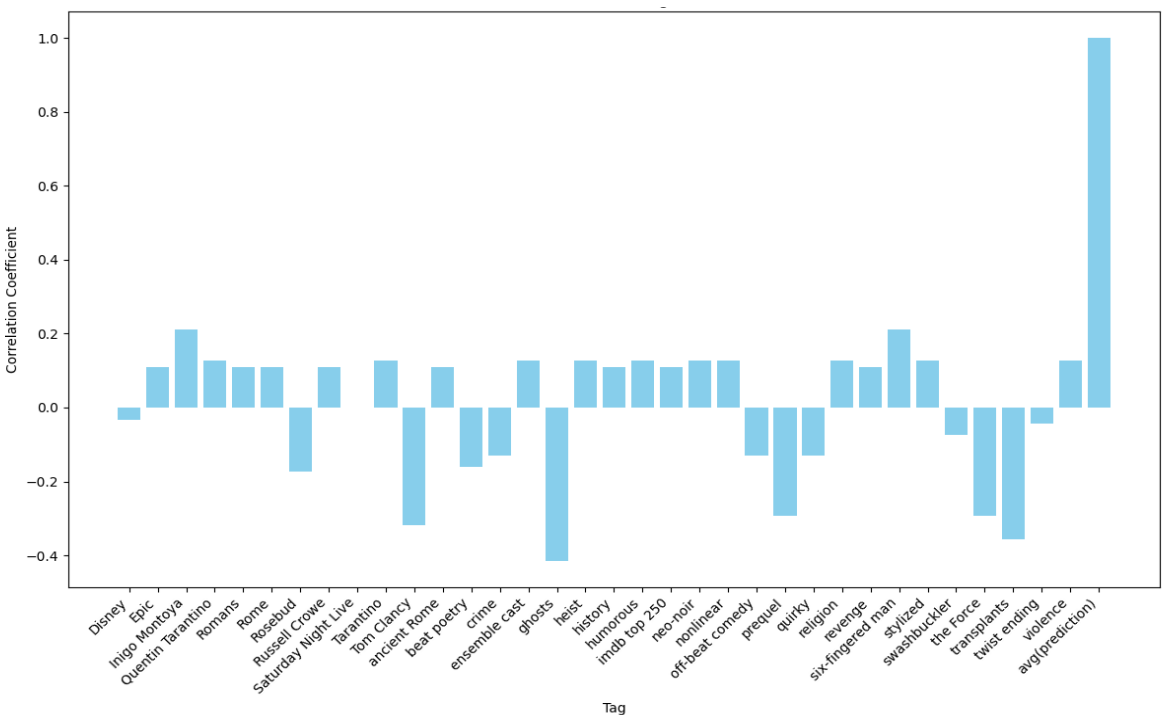


Figure 6. Correlation coefficients between tags and prediction scores

4.8. Distribution of movie ratings by hour of the day

The distribution of movie ratings according to time of day unveils some fascinating information about how users consume their content and their preferences. This distribution shows where views peak and when engagement drops as can be seen in Figure 7. With these insights, content release timing can be maximized, and user experience can be optimized.

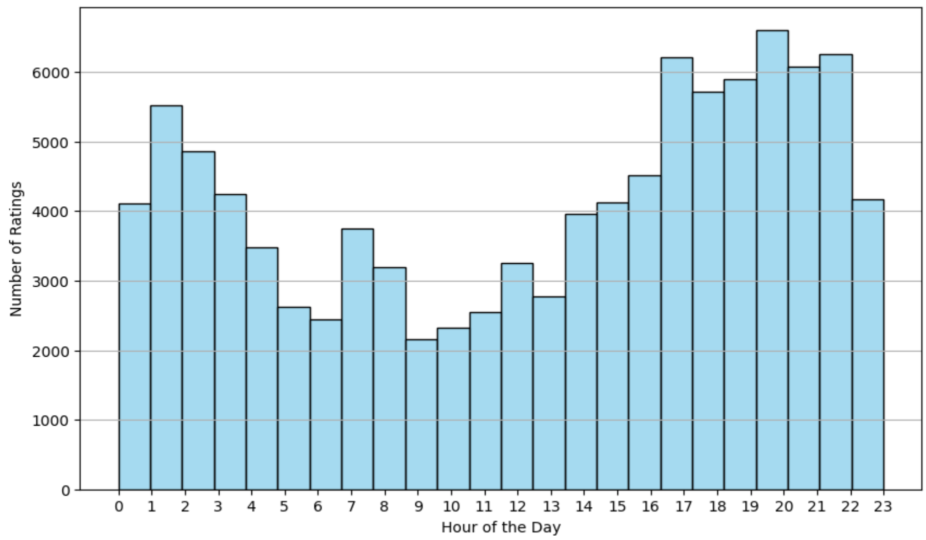


Figure 7. Distribution of movie ratings by hour of the day

This visualization shows how movie ratings are distributed by the time of day in which users watched movies. This distribution allows us to identify what the peak hours are for watching television, as well as how users' viewership behavior is during different parts of the day. This information is extremely valuable when determining when to release content or how to frame marketing efforts so that people are more likely to engage with or be converted by your service. In addition, unlike other models that solely depend on existing user data for generating the model, this recommendation system is not dependent on any of these data. The system can also suggest personalized movies for even nonregistered users if they are not already present in the database, the suggestion can be based on the time of the day, genre preference or the popularity of the movie. Such ability adds to the enjoyment of a movie and may increase user satisfaction and engagement [53]. Recommendation systems are powerful tools for harnessing the vast amounts of data produced in the big data era and creating personalized user experiences. This chapter gives us an idea about how recommendation algorithms function under the hood, an overview of their importance, and a basic understanding of their future within big data.

4.9. Understanding user preferences: the backbone of recommendation systems

At its core, a recommendation system is reading between the lines and making intelligent predictions based on a granular pattern hidden under big oceans of data. These systems identify patterns across massive data sets comprising users' interactions, viewing history, and comments to detect more nuanced patterns of behavior and preference with the aid of big data analytics. Recommendation systems perform a large-scale analysis of user behaviors using the advanced machine learning wizardry of collaborative filtering and deep learning and predict with high accuracy the next piece of content the user is likely to consume.

4.10. Driving business outcomes: the power of personalization

The recommendation systems using big data analytics for Netflix, Disney, and Amazon Prime Video giants are not just a technological marvel but a business grow-or-die strategic asset. With access to mountains of data, created by millions of users around the world, these companies can provide hyper-personalized content recommendations based on personal interests and viewing histories. Such a high degree of personalization increases user activity, retention and, in turn, the bottom line, allowing them to continue asserting market leadership in the highly competitive arena of digital entertainment [51], [52].

4.11. Business impact of recommendation systems

The influence of big data analytics-powered recommendation systems deployed widely by digital entertainment companies on firms in and around the digital entertainment space is very profound. Personalized recommendations go past convenience (Mobile payment APIs lead to greater delight); on a macro level, personalized recommendations have profound power over consumer behavior, driving how digital communication takes place across millions of users. A custom content catalog driven by big data analytics will bring companies closer to consumers, increasing subscription renewals, customer lifetime value, and brand loyalty. Additionally, recommendation systems are powerful content discovery tools that uplift niche titles and promote diverse content consumption across different global audiences [52].

4.12. Charting the path forward: future directions in recommendation systems

As we chart the course forward in the realm of big data- driven recommendation systems, a myriad of possibilities beckons us to explore new frontiers and push the boundaries of innovation. While the current state of recommendation algorithms stands as a testament to their efficacy, there is sample room for further enhancement and evolution. Looking ahead, several key areas warrant attention and investment:

– Enhanced personalization and contextual understanding: continuously refining algorithms to further analyze user-preferences, incorporating contextual information such as time, location, and device type to deliver hyper-personalized recommendations tailored to individual preferences and circumstances.
– Multi-modal recommendations: embracing a multimodal approach that transcends traditional boundaries, combining text, image, audio, and other forms of data to offer richer, more immersive content recommendations that resonate with diverse audiences across global markets.
– Interpretable and transparent models: developing recommendation models that are not only accurate but also interpretable, providing users with insights into the rationale behind each recommendation, thereby fostering trust, transparency, and user engagement at scale.
– Continuous evaluation and feedback loop: instituting robust mechanisms for continuous evaluation of recommendation performance, gathering user feedback, and iteratively refining the system to adapt to evolving user preferences, cultural nuances, and industry trends within the dynamic landscape of big data [52], [53].

By embracing these forward-looking strategies and harnessing the power of big data analytics, recommendation systems will continue to redefine the digital entertainment landscape, enriching user experiences and driving business success on a global scale through enhanced communication, advanced digital approaches, and increased interactivity [53]. In conclusion, the integration of digital communication into recommendation systems has revolutionized the movie-watching experience, setting the stage for future innovations that will continue to enhance digital interaction and engagement. By addressing the prevailing challenges and exploring new avenues in big data analytics, these systems will remain at the forefront of advancing the digital entertainment industry, ultimately bringing about a more personalized and engaging experience for users worldwide.

4. CONCLUSION

We propose the movie recommendation system by applying Apache SparkMLib with proven high performance, which allows the construction of a movie recommendation system of massive datasets. This blog demonstrated how we can use SparkMLib to overcome the limitations of traditional recommendation methods by harnessing the distributed processing capabilities and machine learning algorithms from SparkMLib and provide a scalable and efficient solution to the detrimental issue of information overload. The knowledge acquired through optimizing SparkMLib settings and evaluating its performance provides important directions for developers and researchers, who are important to develop a solid recommendation system. The ability to provide right and personalized recommendations in a big data setting has wide implications not only for improving user experiences in areas such as entertainment and e-commerce but also for any areas where personalized information retrieval is vital. In conclusion, this work highlights supporting future and intelligent/personalized digital experiences solving real-world problems using big data technologies such as SparkMLib.

FUNDING INFORMATION

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Said Chakouk	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓
Abdelkerim Zitouni	✓			✓			✓			✓				✓
Nazif Tchagafo		✓			✓	✓	✓	✓		✓	✓			✓
Ahiod Belaid	✓				✓		✓	✓	✓	✓				✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [SC], upon reasonable request.

REFERENCES

[1] S. Ahmed, T. Sharif, D. H. Ting, and S. J. Sharif, "Crafting emotional engagement and immersive experiences: Comprehensive scale development for and validation of hospitality marketing storytelling involvement," *Psychology and Marketing*, vol. 41, no. 7, pp. 1514–1529, 2024, doi: 10.1002/mar.21994.

[2] Y. Deng, W. Zhou, A. U. Haq, S. Ahmad, and A. Tabassum, "Differentially private recommender framework with dual semi-autoencoder," *Expert Systems with Applications*, vol. 260, Jan. 2025, doi: 10.1016/j.eswa.2024.125447.

[3] I. Hariyale and D. M. Raghuwanshi, "Advanced hybrid recommender systems: Enhancing precision and addressing cold-start challenges in e-commerce and content streaming," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1737–1753, 2024, doi: 10.12785/ijcds/1601129.

[4] E. Yalcin and A. Bilge, "A novel target item-based similarity function in privacy-preserving collaborative filtering," *The Journal of Supercomputing*, vol. 80, no. 13, pp. 19433–19461, Sep. 2024, doi: 10.1007/s11227-024-06221-7.

[5] A. Zouhri, A. Ez-Zahout, S. Chakouk, and M. El Mallahi, "A numerical analysis based internet of things (IoT) and big data analytics to minimize energy consumption in smart buildings," *Journal of Automation, Mobile Robotics and Intelligent Systems*, vol. 18, no. 2, pp. 46–56, Jun. 2024, doi: 10.14313/JAMRIS/2-2024/12.

[6] J. A. Barbero-Aparicio, A. Olivares-Gil, J. J. Rodríguez, C. García-Osorio, and J. F. Díez-Pastor, "Addressing data scarcity in protein fitness landscape analysis: A study on semi-supervised and deep transfer learning techniques," *Information Fusion*, vol. 102, Feb. 2024, doi: 10.1016/j.inffus.2023.102035.

[7] I. Rabiu, N. Salim, A. Da'u, and A. Osman, "Recommender system based on temporal models: a systematic review," *Applied Sciences*, vol. 10, no. 7, 2020, doi: 10.3390/app10072204.

[8] T. Xia *et al.*, "Rethinking data selection at scale: Random selection is almost all you need," *arXiv:2410.09335*, 2024.

[9] J. Yan, "Research on data analysis application based on Spark computing," *European Journal of AI, Computing & Informatics*, vol. 1, no. 2, pp. 23–29, 2025, doi: 10.71222/6n9hhf79.

[10] M. Mezati and I. Aouria, "Machine learning in big data: A performance benchmarking study of Flink-ML and Spark MLlib," *Applied Computer Science*, vol. 21, no. 2, pp. 18–27, 2025, doi: 10.35784/acs_7297.

[11] S. Mukhopadhyay, A. Kumar, D. Parashar, and M. Singh, "Enhanced music recommendation systems: A comparative study of content-based filtering and k-means clustering approaches," *Revue d'Intelligence Artificielle*, vol. 38, no. 1, pp. 365–376, 2024, doi: 10.18280/ria.380138.

[12] R. Raman *et al.*, "Solar energy measurement and monitoring model by using internet of things," *Electric Power Components and Systems*, vol. 52, no. 10, pp. 1796–1807, Jun. 2024, doi: 10.1080/15325008.2023.2246481.

[13] S. Sharma and H. K. Shakya, "Hybrid recommendation system for movies using artificial neural network," *Expert Systems with Applications*, vol. 258, 2024, doi: 10.1016/j.eswa.2024.125194.

[14] S. Milovanović, Z. Bogdanović, A. Labus, D. Barać, and M. Despotović-Zrakić, "An approach to identify user preferences based on social network analysis," *Future Generation Computer Systems*, vol. 93, pp. 121–129, Apr. 2019, doi: 10.1016/j.future.2018.10.028.

[15] M. Ayemowa, R. Ibrahim, and M. M. Khan, "Analysis of recommender system using generative artificial intelligence: a systematic literature review," *SSRN Electronic Journal*, 2024, doi: 10.2139/ssrn.4922584.

[16] N. Li and Y. Xia, "Corrigendum to 'Movie recommendation based on ALS collaborative filtering recommendation algorithm with deep learning model' [Entertain. Comput. 51 (2024) 100715]," *Entertainment Computing*, vol. 52, Jan. 2025, doi: 10.1016/j.entcom.2024.100835.

[17] L. Vaquero *et al.*, "Studying the effects of digital arts-based interventions on teenagers' social media usage, brain connectivity, and mental health: study protocol of the SMART project," *European Journal of Psychotraumatology*, vol. 16, no. 1, 2025, doi: 10.1080/20008066.2025.2550079.

[18] Y. Deldjoo *et al.*, "Movie genome: alleviating new item cold start in movie recommendation," *User Modeling and User-Adapted Interaction*, vol. 29, no. 2, pp. 291–343, Apr. 2019, doi: 10.1007/s11257-019-09221-y.

[19] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *Journal of Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5.

[20] S. Gupta, A. Kumar, and D. Prasad, "Multimodal graph-based recommendation system using hybrid filtering approach," *International Journal of Computing and Digital Systems*, vol. 17, no. 1, pp. 1–15, 2025, doi: 10.12785/ijcds/1571047857.

[21] W. McKinney, *Python for data analysis*, Third. United States: O'Reilly Media, Inc., 2022.

[22] H. K. Omar, M. Frikha, and A. K. Jumaa, "PyTorch and TensorFlow performance evaluation in big data recommendation system," *Ingénierie des systèmes d information*, vol. 29, no. 4, pp. 1357–1364, 2024, doi: 10.18280/isi.290411.

[23] M. A. Fattah, "A model to improve recommendation systems based on Spark platform," *International Journal of Business Information Systems*, vol. 39, no. 1, pp. 328–348, 2022, doi: 10.1504/IJBIS.2022.10046763.

[24] J. Sundaram *et al.*, "An exploration of Python libraries in machine learning models for data science," in *Advanced Interdisciplinary Applications of Machine Learning Python Libraries for Data Science*, IGI Global, 2023, pp. 1–31, doi: 10.4018/978-1-6684-8696-2.ch001.

[25] P. Yu, Z. Xu, J. Wang, and X. Xu, "The application of large language models in recommendation systems," in *International Conference on Artificial Intelligence and Machine Learning Research (CAIMLR 2024)*, 2025, p. 31, doi: 10.1117/12.3062544.

[26] A. M. Fernández, D. Gutiérrez-Avilés, A. Troncoso, and F. Martínez–Álvarez, "Automated deployment of a spark cluster with machine learning algorithm integration," *Big Data Research*, vol. 19–20, Mar. 2020, doi: 10.1016/j.bdr.2020.100135.

[27] B. Amiri, N. Shahverdi, A. Haddadi, and Y. Ghahremani, "Beyond the trends: evolution and future directions in music recommender systems research," *IEEE Access*, vol. 12, pp. 51500–51522, 2024, doi: 10.1109/ACCESS.2024.3386684.

[28] J. A. Konstan and J. Riedl, "Recommender systems: from algorithms to user experience," *User Modeling and User-Adapted Interaction*, vol. 22, no. 1–2, pp. 101–123, 2012, doi: 10.1007/s11257-011-9112-x.

[29] F. Quin, D. Weyns, M. Galster, and C. C. Silva, "A/B testing: A systematic literature review," *Journal of Systems and Software*, vol. 211, May 2024, doi: 10.1016/j.jss.2024.112011.

[30] P. N. Vicente and C. D. Burnay, "Recommender systems and over-the-top services: a systematic review study (2010–2022)," *Journalism and Media*, vol. 5, no. 3, pp. 1259–1278, Sep. 2024, doi: 10.3390/journalmedia5030080.

[31] C. Tryon, *On-demand culture: digital delivery and the future of movies*. Rutgers University Press, 2013.

[32] R. Neyah and M. Vijayakumar, "Advancing personalization and recommendation algorithms in the OTT industry," in *The Rise of Over-the-Top (OTT) Media and Implications for Media Consumption and Production*, IGI Global, 2024, pp. 114–129, doi: 10.4018/979-8-3693-0116-6.ch008.

[33] R. S. Sharma, A. A. Shaikh, and E. Li, "Designing recommendation or suggestion systems: looking to the future," *Electronic Markets*, vol. 31, no. 2, pp. 243–252, Jun. 2021, doi: 10.1007/s12525-021-00478-z.

[34] O. Fouad, R. Fouad, N. Hussen, and I. Abuhadrous, "A comprehensive review of music recommendation systems," *Advanced Sciences and Technology Journal*, vol. 2, no. 1, pp. 1–18, 2025, doi: 10.21608/astj.2025.342474.1017.

[35] S. Wang, X. Zhang, Y. Wang, and F. Ricci, "Trustworthy recommender systems," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 4, pp. 1–20, Aug. 2024, doi: 10.1145/3627826.

[36] J. Hildén, "The public service approach to recommender systems: filtering to cultivate," *Television & New Media*, vol. 23, no. 7, pp. 777–796, Nov. 2022, doi: 10.1177/15274764211020106.

[37] S. Akter, Y. K. Dwivedi, S. Sajib, K. Biswas, R. J. Bandara, and K. Michael, "Algorithmic bias in machine learning-based marketing models," *Journal of Business Research*, vol. 144, pp. 201–216, May 2022, doi: 10.1016/j.jbusres.2022.01.083.

[38] P. S. Varsha, "How can we manage biases in artificial intelligence systems-a systematic literature review," *International Journal of Information Management Data Insights*, vol. 3, no. 1, Apr. 2023, doi: 10.1016/j.jjimei.2023.100165.

[39] A. S. Xiao and Q. Liang, "Spam detection for Youtube video comments using machine learning approaches," *Machine Learning with Applications*, vol. 16, Jun. 2024, doi: 10.1016/j.mlwa.2024.100550.

[40] P. K. Pattanayak, R. M. Tripathy, and S. Padhy, "A semi-supervised approach of short text topic modeling using embedded fuzzy clustering for Twitter hashtag recommendation," *Discover Sustainability*, vol. 5, no. 1, Apr. 2024, doi: 10.1007/s43621-024-00218-1.

[41] A. Siddique, M. K. Abid, M. Fuzail, and N. Aslam, "Movies rating prediction using supervised machine learning techniques," *International Journal of Information Systems and Computer Technologies*, vol. 3, no. 1, pp. 40–56, 2024, doi: 10.58325/ijisct.003.01.0062.

[42] M. Valera and R. Mehta, "Optimized movie recommendation via sentiment analysis & hyperparameter tuning," *Jurnal Kejuruteraan*, vol. 37, no. 3, pp. 1593–1605, 2025, doi: 10.17576/jkukm-2025-37(3)-40.

[43] J. Zhu, Z. Ge, Z. Song, and F. Gao, "Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data," *Annual Reviews in Control*, vol. 46, pp. 107–133, 2018, doi: 10.1016/j.arcontrol.2018.09.003.

[44] A. Sherif and A. Ravindra, *Apache spark deep learning cookbook: Over 80 recipes that streamline deep learning in a distributed environment with apache spark*, 1st editio. Packt Publishing Ltd, 2018.

[45] A. Ez-Zahout, S. Chakouk, S. Mitouilli, and M. A. El Bouni, "A numerical real time web tracking and scrapping strategy applied to analysing COVID-19 datasets," in *2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC)*, Jul. 2021, pp. 536–542, doi: 10.1109/ICNISC54316.2021.00102.

[46] Y. Rafique, J. Wu, A. W. Muzaffar, and B. Rafique, "An enhanced integrated fuzzy logic-based deep learning techniques (EIFL-DL) for the recommendation system on industrial applications," *PeerJ Computer Science*, vol. 10, 2024, doi: 10.7717/peerj-cs.2529.

[47] M. R. Irfansyah, D. Nurjanah, and H. Nurrahmi, "Music recommendation system using alternating least squares method," *Indonesian Journal on Computing (Indo-JC)*, vol. 9, no. 1, pp. 19–29, 2024, doi: 10.34818/INDOJC.2024.9.1.908.

[48] H.-J. Cha, H.-K. Yang, and Y. Song, "The detection data processing mechanism for vehicular cyber physical system in IoT environment," *Tehnicki vjesnik - Technical Gazette*, vol. 28, no. 3, Jun. 2021, doi: 10.17559/TV-20201209230220.

[49] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: a survey and new perspectives," *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–38, 2020, doi: 10.1145/3285029.

[50] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in *Proceedings of the 24th international conference on Machine learning*, Jun. 2007, pp. 791–798, doi: 10.1145/1273496.1273596.

[51] M. S. Islam *et al.*, "Challenges and future in deep learning for sentiment analysis: A comprehensive review and a proposed hybrid approach," *Artificial Intelligence Review*, vol. 57, no. 3, p. 62, 2024, doi: 10.1007/s10462-023-10651-9.

[52] A. Fischer and C. Igel, "An introduction to restricted Boltzmann machines," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (CIARP 2012)*, Berlin, Heidelberg: Springer Publishing, 2012, pp. 14–36, doi: 10.1007/978-3-642-33275-3_2.

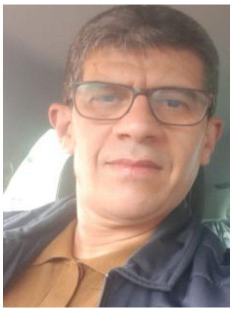
[53] C. C. Aggarwal, *Recommender systems*. Cham: Springer International Publishing, 2016, doi: 10.1007/978-3-319-29659-3.

[54] P. L. Martínez, R. Dintén, J. M. Drake, and M. Zorrilla, "A big data-centric architecture metamodel for Industry 4.0," *Future Generation Computer Systems*, vol. 125, pp. 263–284, Dec. 2021, doi: 10.1016/j.future.2021.06.020.

[55] A. V. Shinde, D. D. Patil, and K. K. Tripathi, "A comprehensive survey on recommender systems techniques and challenges in big data analytics with IOT applications," *Revista de Gestão Social e Ambiental*, vol. 18, no. 2, Mar. 2024, doi: 10.24857/rgsa.v18n2-097.

[56] J. B. N. Penka, S. Mahmoudi, and O. Debauche, "A new Kappa architecture for IoT data management in smart farming," *Procedia Computer Science*, vol. 191, pp. 17–24, 2021, doi: 10.1016/j.procs.2021.07.006.

BIOGRAPHIES OF AUTHORS



Said Chakouk is an Associate Professor in the Department of Information and Communication Sciences at Mohammed V University in Rabat. Ph.D. in Information and Communication Sciences obtained in 2016 from Mohammed V University in Rabat, and agricultural engineer since 1999. He is the author of numerous articles and research works in the digital field, and holder of several international projects. He is a member of the teaching staff for the master's program in Software Development and Decision Engineering, Department of Computer Sciences, Faculty of Sciences of Rabat, and the professional bachelor's degree in networks and telecommunications. He expert for the Agency for Francophone University Cooperation (AUF) in digital pedagogical innovation, an expert for the United Nations online volunteers for the contribution to sustainable development goals program, and the community manager of the Ifker Association for Environmental Education and Sustainable Development. He can be contacted at email: s.chakouk@um5r.ac.ma.

Abdelkerim Zitouni is an Associate Professor of Management and Educational Communication within the Faculty of Education Sciences at Mohammed V University in Rabat (Laboratory of Signals, Distributed Systems, and Artificial Intelligence, ENSET/Mohammedia (SSDIA). He obtained a Ph.D. in management and responsibility from Mohammed V University in Rabat (2017) and a master's degree in communication and training in organizations (2010). He has several publications in the field of organizational communication. He can be contacted at email: zitounikarim2022@gmail.com.

Nazif Tchagafo is currently pursuing a Ph.D. in Computer Science at the Faculty of Sciences, Mohammed V University in Rabat, Morocco. His research focuses on big data analytics for optimization of recommendation systems. He earned his Master of Data Engineering and Software Development, also from the Faculty of Sciences, Mohammed V University, and bachelor in Mathematics from Faculty of Sciences, Chouaib Doukkali University of EL Jadida, Morocco. His background in both mathematics, artificial intelligence, and big data analytics provides him with a unique interdisciplinary perspective. He can be contacted at email: ntchagafo@gmail.com.

Prof. Dr. Ahiod Belaid received his B.S. degree in Applied Mathematics, M.S., and Ph.D. degree in Computer Science from the Faculty of Science of the Mohammed V University in Rabat, Morocco. He is with the Department of Computer Science as a Professor at the Faculty of Science in Rabat. His research activities are carried out, in their majority, in the Laboratory for Computer Science and Telecommunications LRIT (associated CNRST unit (URAC 29)). His research interests include NP-hard combinatorial optimization problems, multi-objective optimization, meta-heuristics, and nature-inspired algorithms. He can be contacted at email: b.ahiod@um5r.ac.ma.