



Kharazmi University



Prediction of Users' Movie Genre Preferences Based on Their Personality and Demographic Features

Mahsa Rezaei¹ | Marjan Kaedi^{2✉}

1. Master's Student, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran. **Email:** mah.rezaei@eng.ui.ac.ir
2. Corresponding author, Associate Professor, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran. **Email:** kaedi@eng.ui.ac.ir

Article Info	ABSTRACT
Article type: Research Article	Purpose: Online movie streaming platforms provide access to a wide range of content and have changed the way audiences select and watch movies. Predicting users' preferences for different movie genres can improve personalization and the performance of movie recommender systems. Given the significant relationship between users' personality traits and their preferences for movie genres, this study aims to investigate the effectiveness of using personality and demographic characteristics to predict users' genre preferences.
Article history: Received 24 September 2025 Received in revised form 31 January 2026 Accepted 12 february 2026 Published online 26 Mardh 2026	Method: In this study, movies were characterized according to 18 main genres, while users were represented by personality traits based on the Five-Factor Model, including conscientiousness, extraversion, agreeableness, neuroticism, and openness to experience, along with demographic characteristics including age, gender, and location. Various machine learning models, including Support Vector Regression and Random Forest Regression, were trained separately for each genre to learn the relationship between user characteristics and their genre preferences. The models were evaluated on a version of the MovieLens dataset using Mean Absolute Error (MAE) as the primary evaluation metric.
Keywords: Genre preferences, Machine learning, Personality traits, Big-five personality traits, Modeling	Findings: The results showed that the models for predicting preferences for drama, action, and crime achieved the best performance, with MAE values of 0.18, 0.21, and 0.22, respectively. In contrast, the models for musical and fantasy genres obtained the highest error values, with MAEs of 0.34 and 0.33, respectively.
	Conclusion: The findings indicate that combining personality and demographic characteristics can effectively predict users' genre preferences and can provide useful information for improving personalization and the performance of movie recommender systems.

Cite this article: Rezaei, M., Kaedi, M. (2026). Prediction of Users' Movie Genre Preferences Based on Their Personality and Demographic Features. *Human-Information Interaction*, 13(1), 76-97. <https://doi.org/10.22034/hii.2026.115620.0>



© The Author(s). Publisher: University of Kharazmi.



Kharazmi University

Journal of Human-Information Interaction

Online ISSN: 2423-7418

<https://hi.khu.ac.ir/>



Introduction

The rapid growth of online movie streaming platforms has provided users with access to a large and diverse range of content. Although recommender systems facilitate content selection, the large number of available movies and diversity of user preferences make accurate personalization challenging. Predicting users' preferences for different movie genres can provide valuable information for improving the personalization and effectiveness of recommender systems. Previous studies have reported meaningful relationships between personality traits and media or movie preferences. However, most existing studies have focused on identifying these relationships rather than developing predictive models that directly estimate genre preferences from personality characteristics. Demographic characteristics can also provide complementary information about individual differences in movie preferences. Therefore, this study investigates whether personality and demographic characteristics can be used to predict users' preferences for different movie genres. Specifically, separate prediction models are developed for 18 movie genres to identify the most informative characteristics for each genre. The findings can provide an additional source of user information for recommender systems, particularly when users have limited behavioral histories.

Methods

The study uses the MovieLens 100K dataset provided by GroupLens Research, containing 100,000 ratings from 1,000 users for approximately 1,700 movies. Personality information was previously collected using a questionnaire based on the Big Five personality model, including openness, conscientiousness, extraversion, agreeableness, and neuroticism. Demographic characteristics, including age, gender, and geographical location, are also incorporated. Movie-related information is supplemented using IMDb. The proposed methodology consists of three main stages, as illustrated in Figure 1. First, users' genre preferences are estimated from their movie ratings. Eighteen genres are considered: action, adventure, animation, children, comedy, crime, documentary, drama, fantasy, film-noir, horror, musical, mystery, romance, sci-fi, thriller, war, and western. For each user and genre, the average rating assigned to movies belonging to that genre is calculated as the preference score, ranging from 1 to 5. When a user has not rated a movie belonging to a particular genre, the corresponding value is treated as inactive rather than as a preference score. Because the original movie information does not directly provide a single dominant genre for every movie, semantic similarity is used to determine the main genre. Movie overviews are transformed into 384-dimensional embeddings using the SentenceTransformer model, all-MiniLM-L6-v2. Descriptions of the 18 genres are embedded using the same model. Cosine similarity is calculated between each movie overview and the genre descriptions, and the genre with the highest similarity is assigned as the dominant genre. This procedure is a semantic genre-assignment process rather than supervised classification.

In the second stage, relevant personality and demographic features are selected separately for each genre. Since the target variable is a numerical preference score, feature selection is formulated as a regression problem. Four methods are employed: Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Sequential Floating Forward Selection (SFFS), and Recursive Feature Elimination (RFE). Feature selection is performed only on the training data to prevent information leakage. Five-fold cross-validation is conducted within the training set, and the feature subset producing the lowest Mean Absolute Error (MAE) is selected for each genre. In simple terms, this stage identifies which personality and demographic characteristics provide more useful information for estimating users' interest in each genre. In the third stage, separate prediction models are developed for the 18 genres using Support Vector Regression (SVR), Random Forest Regression (RF), Gradient Boosting (XGB), K-Nearest Neighbors (KNN), and Feedforward Neural Networks (FNN). For each genre, 80% of the data are used for training and 20% for independent testing.



Kharazmi University

Hyperparameters are optimized using grid search. Model performance is evaluated using MAE, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 , with MAE used as the primary evaluation criterion. In simple terms, the models use the selected personal characteristics to estimate how much a user is likely to prefer each genre, and their predictive performance is subsequently evaluated.

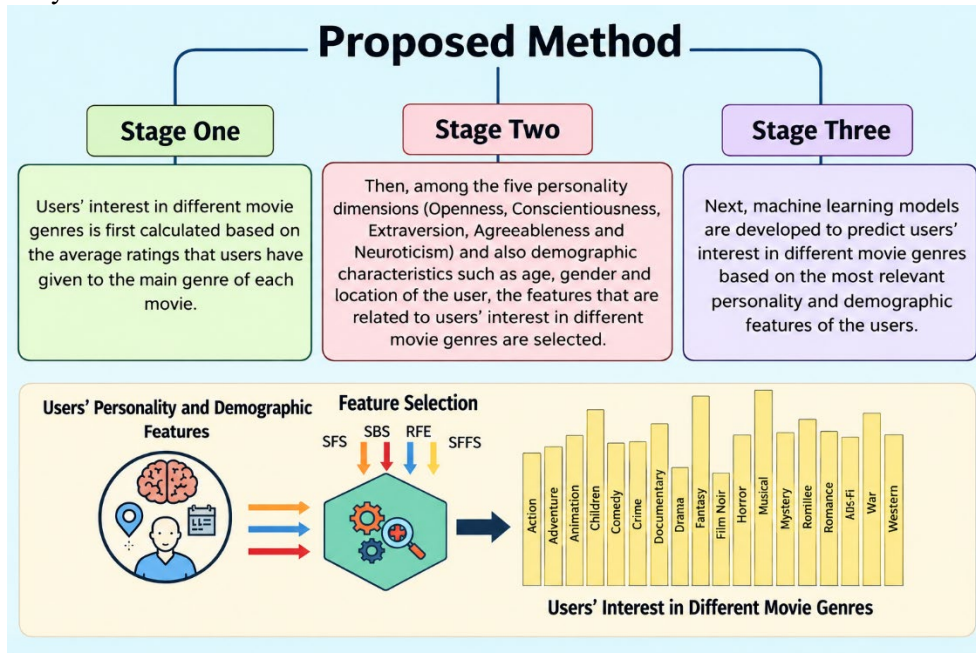


Figure 1: Proposed framework for predicting users' movie genre preferences based on personality and demographic features.

Resultss and Discussion

The feature-selection results demonstrate that the most informative combination of personality and demographic characteristics differs across movie genres. This finding indicates that the characteristics associated with users' preferences are not identical for all types of movie content. For example, the prediction of action preference involves a broader combination of personality and demographic characteristics, whereas comedy can be predicted using a smaller subset of personality traits. Drama also involves a combination of personality and demographic characteristics. These differences indicate that individual characteristics contribute differently to the prediction of users' preferences across genres. The prediction results further demonstrate that the performance of the models varies across genres. No single machine-learning algorithm consistently performs best for all genres. Instead, different algorithms provide better predictions for different genres after feature selection and hyperparameter optimization. This finding supports the use of genre-specific prediction models rather than applying a single model to all movie genres. In other words, the predictive relationship between users' individual characteristics and their movie preferences depends on the genre being considered.

As illustrated in Figure 1, the proposed process first uses users' personality and demographic characteristics as input information, then identifies the most relevant features, and finally uses these features to estimate users' interest in different movie genres. This step-by-step structure makes the data-mining process more understandable by showing how individual characteristics are transformed into genre-preference predictions. The findings indicate that personality and demographic information can provide meaningful predictive signals, although their usefulness varies depending on the genre. For additional evaluation, the proposed approach was compared with the method of Gupta et al. (2025), which also uses the MovieLens dataset and the same set of movie genres.



Kharazmi University

Journal of Human-Information Interaction

Online ISSN: 2423-7418

<https://hi.khu.ac.ir/>



However, the two studies differ in their modeling objectives and input features. Gupta et al. focus on descriptive movie features and neural-network-based prediction of movie ratings, whereas the present study predicts users' genre preferences from personality and demographic characteristics. Therefore, this comparison is intended to provide a reference under a common dataset and genre setting rather than to claim that the two approaches perform the same prediction task.

Conclusions

The findings demonstrate that personality and demographic characteristics can provide useful information for predicting users' preferences across different movie genres. This result is consistent with previous research reporting associations between personality traits and media or movie preferences, while extending this line of research by using individual characteristics as inputs to predictive models. The differences observed in the selected features and prediction performance across genres further indicate that the relationship between individual characteristics and movie preferences is genre-dependent. Therefore, genre-specific modeling can better reflect differences in users' interests than a single general prediction model. From an application perspective, predicted genre-preference scores can be incorporated into recommender systems as an additional layer of user modeling. These predictions can be combined with users' behavioral histories and movie-content information to provide more personalized recommendations. Such information may be particularly useful for users with limited behavioral histories, where conventional behavior-based recommendation methods face the cold-start problem. In shared-account environments, estimating the potential genre preferences of individual users based on their personal characteristics may also provide complementary information for distinguishing different interests and generating more personalized recommendations. The study has several limitations. First, personality information may not normally be available on real-world streaming platforms. Its practical use may therefore require users' voluntary participation or reliable indirect estimation from available behavioral information. Second, the MovieLens dataset is limited in terms of size and cultural and geographical diversity, which may restrict the generalizability of the findings. In addition, assigning a dominant genre based on semantic similarity may not completely represent the complexity of movies that belong to multiple genres.

From a practical perspective, recommender systems can use predicted genre preferences as supplementary user-profile information rather than replacing behavioral data. Personality characteristics could also be estimated indirectly from users' interactions, where technically and ethically appropriate, reducing dependence on direct questionnaires. Future research should validate the proposed framework using larger and more culturally and geographically diverse datasets. Integrating personality information with richer behavioral data, such as viewing time, search history, and social interactions, may further improve user modeling. Deep-learning and attention-based approaches can also be investigated to improve prediction accuracy and generalizability. Overall, the findings suggest that combining personality and demographic characteristics with genre-specific prediction is a promising approach for modeling users' movie preferences. The proposed framework can provide complementary information for recommender systems and contribute to the development of more personalized movie recommendation services.

Ethical Considerations

Compliance with Research Ethics

The study was conducted in accordance with all ethical principles and standards governing scientific research.



Kharazmi University

Journal of Human-Information Interaction

Online ISSN: 2423-7418

<https://hii.khu.ac.ir/>



Authors' Contributions

First Author: Conceptualization, data preparation and processing, data collection, implementation, experiments, statistical analysis, interpretation of results, and manuscript drafting and revision. **Second Author:** Conceptualization and research design, methodology development, supervision of the research process, validation and interpretation of results, scientific guidance, and critical review and revision of the manuscript.

Conflict of Interest

According to the authors, this article has no conflict of interest.

Funding

Not applicable.

Acknowledgments

This article is derived from a master's thesis at University of Isfahan. The authors would like to express their sincere gratitude to the anonymous reviewers for their valuable comments and constructive suggestions, which significantly improved the quality of this manuscript.

تخمین علاقه کاربران به ژانرهای فیلم بر اساس شخصیت و ویژگی‌های جمعیت‌شناختی آن‌ها

مهسا رضائی^۱، مرجان کائدی^۲

۱. دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران.
رایانامه: mah.rezaei@eng.ui.ac.ir
۲. نویسنده مسئول، دانشیار گروه مهندسی فناوری اطلاعات، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران.
رایانامه: kaedi@eng.ui.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	هدف: پلتفرم‌های پخش آنلاین فیلم با فراهم کردن دسترسی به طیف گسترده‌ای از محتوا، نقش مهمی در نحوه انتخاب و تماشای فیلم ایفا می‌کنند. در این میان، پیش‌بینی ترجیحات کاربران نسبت به ژانرهای مختلف می‌تواند به بهبود شخصی‌سازی و عملکرد سیستم‌های توصیه‌گر کمک کند. با توجه به وجود رابطه معنادار میان ویژگی‌های شخصیتی کاربران و علاقه‌مندی آن‌ها به ژانرهای فیلم، هدف این پژوهش، بررسی قابلیت استفاده همزمان از ویژگی‌های شخصیتی و جمعیت‌شناختی برای پیش‌بینی ترجیحات ژانری کاربران است.
تاریخ دریافت: ۱۴۰۴/۰۹/۰۲	روش: در این پژوهش، فیلم‌ها بر اساس ۱۸ ژانر اصلی و کاربران بر اساس ویژگی‌های شخصیتی مدل پنج‌عاملی شامل وظیفه‌شناسی، برون‌گرایی، توافق‌پذیری، روان‌رنجوری و گشودگی به تجربه، همراه با ویژگی‌های جمعیت‌شناختی شامل سن، جنسیت و موقعیت مکانی، توصیف شدند. سپس، برای هر ژانر، مدل‌های مختلف یادگیری ماشین از جمله رگرسیون بردار پشتیبان و جنگل تصادفی برای یادگیری رابطه میان ویژگی‌های کاربران و میزان علاقه آن‌ها به ژانرهای مختلف آموزش داده شدند. مدل‌ها بر روی نسخه‌ای از مجموعه‌داده MovieLens ارزیابی شدند و معیار میانگین خطای مطلق برای مقایسه عملکرد آن‌ها به کار رفت.
تاریخ بازنگری: ۱۴۰۴/۱۱/۱۱	یافته‌ها: نتایج نشان داد مدل‌های پیش‌بینی علاقه به ژانرهای درام، اکشن و جنایی، به‌ترتیب با میانگین خطای مطلق ۰/۱۸، ۰/۲۱ و ۰/۲۲، بهترین عملکرد را داشتند. در مقابل، ژانرهای موزیکال و فانتزی با مقادیر ۰/۳۴ و ۰/۳۳، ضعیف‌ترین نتایج را به خود اختصاص دادند.
تاریخ پذیرش: ۱۴۰۴/۱۱/۲۳	نتیجه‌گیری: نتایج نشان می‌دهد که ترکیب ویژگی‌های شخصیتی و جمعیت‌شناختی می‌تواند رویکردی مؤثر برای پیش‌بینی ترجیحات ژانری کاربران باشد و از این اطلاعات می‌توان در راستای بهبود شخصی‌سازی و عملکرد سیستم‌های توصیه‌گر فیلم استفاده کرد.
تاریخ انتشار: ۱۴۰۵/۰۱/۰۵	
کلیدواژه‌ها: علاقه‌مندی‌های ژانری، یادگیری ماشین، ویژگی‌های شخصیتی، مدل پنج عاملی شخصیت، مدل‌سازی	

استناد: رضائی، مهسا؛ کائدی، مرجان (۱۴۰۵). تخمین علاقه کاربران به ژانرهای فیلم بر اساس شخصیت و ویژگی‌های جمعیت‌شناختی آن‌ها. *تعامل انسان و اطلاعات*، ۱۳(۱)، ۷۶-۹۷. <https://doi.org/10.22034/hii.2026.115620.0>



مقدمه

با توجه به تنوع گسترده فیلم‌ها در پلتفرم‌های پخش فیلم نظیر نتفلیکس و آمازون، انتخاب فیلم مناسب برای تماشا فرآیندی پیچیده است. این پلتفرم‌ها با بهره‌گیری از الگوریتم‌های پیشرفته، تجربه کاربری را به‌طور قابل‌توجهی بهبود بخشیده‌اند. سیستم‌های توصیه‌گر فیلم، به‌عنوان یکی از اجزای کلیدی این پلتفرم‌ها، با هدف ارائه محتوای متناسب با سلیقه و ترجیحات کاربران طراحی شده‌اند (ریچی و همکاران^۱، ۲۰۲۱، شفیلو و همکاران^۲، ۲۰۲۴). با این حال، با وجود ارائه محتوای متنوع و الگوریتم‌های پیشرفته، انتخاب فیلم مناسب همچنان به صورت فرآیندی چالش‌برانگیز باقی مانده است. عوامل متعددی از جمله حجم بالای محتوا، سلیقه‌های متفاوت کاربران و ویژگی‌های فردی آن‌ها، تصمیم‌گیری را دشوار می‌کند. بنابراین، توسعه روش‌ها و ابزارهایی که بتوانند تجربه کاربری را بهبود بخشیده و انتخاب محتوای مناسب را تسهیل کنند، اهمیت فراوانی یافته است (وانگ و همکاران^۳، ۲۰۱۴). پیش‌بینی علایق کاربران نسبت به ژانرهای مختلف فیلم، به‌عنوان یکی از مهم‌ترین مؤلفه‌های عملکرد سیستم‌های توصیه‌گر فیلم، می‌تواند دقت توصیه‌ها را به شکل چشمگیری افزایش دهد و تجربه کاربری را بهبود بخشد (گائو و همکاران^۴، ۲۰۲۵).

مطالعات پیشین نشان داده‌اند که انتخاب‌ها و ترجیحات کاربران تا حد قابل‌توجهی تحت تأثیر ویژگی‌های روان‌شناختی، شخصیتی و جمعیت‌شناختی آن‌ها قرار دارد (صدیق‌زاده و کائدی^۵، ۲۰۲۲، نگارش و همکاران^۶، ۲۰۲۳). شخصیت به عنوان مجموعه‌ای از صفات روان‌شناختی پایدار تعریف می‌شود که الگوهای پایدار اندیشه، هیجان و رفتار فرد را شکل می‌دهند. این صفات تأثیری مستقیم بر علایق و ترجیحات شخصی دارند و تعیین می‌کنند که افراد چگونه در محیط‌های دیجیتال با محتوا، محصولات و دیگران تعامل می‌کنند (بوسین و همکاران^۷، ۲۰۲۵، صالحی و همکاران^۸، ۲۰۲۵). نشان داده شده است که ارتباط معناداری بین ویژگی‌های شخصیتی کاربران و علاقه آن‌ها به ژانرهای مختلف فیلم، از جمله درام، کمدی، فانتزی و عاشقانه، وجود دارد (هوانگ و همکاران^۹، ۲۰۱۶). با این حال این مطالعات پیشین، صرفاً وجود چنین ارتباطاتی را تایید کرده‌اند. ولی از این ارتباطات برای پیش‌بینی علاقه کاربران به ژانرهای مختلف فیلم استفاده نکرده‌اند. بر این اساس، در پژوهش پیش‌رو، مدل‌هایی ارائه می‌گردد که با استفاده از ویژگی‌های شخصیتی و جمعیت‌شناختی کاربران به‌عنوان منبع اطلاعات ارزشمند برای پیش‌بینی علاقه‌مندی‌های ژانری آن‌ها استفاده می‌کند. به این منظور، در این پژوهش ۱۸ ژانر مختلف فیلم (مانند کمدی، درام، اکشن، و غیره) در نظر گرفته شده است و ویژگی‌های شخصیتی کاربران نیز بر اساس مدل پنج عاملی شخصیت شامل وظیفه‌شناسی، برون‌گرایی، توافق‌پذیری، روان‌رنجوری و گشودگی به تجربیات جدید در نظر گرفته می‌شوند. علاوه بر این، اطلاعات جمعیت‌شناختی کاربران شامل سن، جنسیت و موقعیت مکانی (کشور) به‌عنوان ویژگی‌های تکمیلی در تحلیل‌ها مورد استفاده قرار می‌گیرد. سپس ۱۸ مدل یادگیری ماشین مجزا توسعه می‌یابد که هر یک به پیش‌بینی ترجیحات کاربران نسبت به یکی از ژانرهای فیلم اختصاص دارد. ورودی هر مدل، مجموعه ویژگی‌های شخصیتی و جمعیت‌شناختی کاربران است و خروجی آن، میزان علاقه‌مندی پیش‌بینی‌شده کاربر به ژانر مربوطه است. در این پژوهش از مجموعه‌داده MovieLens استفاده شده است. مدل‌های آموزش‌دیده قادر خواهند بود به‌عنوان زیرساختی برای ارتقای سیستم‌های توصیه‌گر به کار روند.

پیشینه

پیشینه این پژوهش شامل دو حوزه اصلی است: ابتدا سیستم‌های توصیه‌گر فیلم که در آنها، پیش‌بینی علاقه‌مندی‌های ژانری کاربران در بهبود دقت توصیه‌ها و ارتقای تجربه کاربری نقش دارد، بررسی می‌شوند؛ سپس مطالعات روان‌شناختی که به بررسی رابطه بین ویژگی‌های شخصیتی کاربران و علاقه‌مندی آن‌ها به فیلم پرداخته‌اند، مرور می‌شوند.

¹ Ricci et al.² Wang et al.³ Shafiloo et al.⁴ Gao et al.⁵ SadighZadeh & Kaedi⁶ Negaresh et al.⁷ Buseyne et al.⁸ Salehi et al.⁹ Hwang et al.

• اهمیت پیش‌بینی علاقه‌مندی‌های ژانری کاربران در ارتقای سیستم‌های توصیه‌گر فیلم

در این بخش پژوهش‌هایی مرور می‌شوند که ژانر مورد علاقه کاربران را برای توصیه کردن فیلم به آنها استفاده می‌کنند. با این حال، همانطور که در ادامه مرور می‌شود، در هیچ یک از این پژوهش‌ها، ویژگی‌های شخصیتی کاربران برای پیش‌بینی علایق ژانری مورد استفاده قرار نگرفته است. مرور این پژوهش‌ها نشان می‌دهد که پیش‌بینی علایق کاربران به ژانرها، اقدامی ارزشمند و موثر برای بهبود توصیه‌گرهای فیلم است.

کیم و همکاران^۱ (۲۰۲۱) سیستم‌های توصیه‌گر فیلم ارائه کرده‌اند که در آن، روشی مبتنی بر خوشه‌بندی کاربران با الگوهای مشابه استفاده شده و به جای توصیه مستقیم فیلم، الگوریتمی برای پیش‌بینی ژانر فیلم بعدی مورد علاقه کاربر را به کار می‌برد. ورودی این مدل شامل داده‌های رفتاری کاربران، ترتیب مشاهده فیلم‌ها و ژانرهای فیلم‌های قبلی است و الگوریتم با در نظر گرفتن علاقه‌مندی‌های ژانری کاربران و حذف ژانرهای غیرضروری در برخی خوشه‌ها، دقت توصیه‌ها را افزایش می‌دهد. در پژوهشی دیگری، وانگ (۲۰۱۸) سیستم توصیه‌گری ارائه کرده که با هدف پیش‌بینی علاقه‌مندی‌های ژانری کاربران، از داده‌های جمعیت‌شناختی، رفتاری و ترجیحات کاربران استفاده می‌کند. در این روش، الگوریتم‌هایی مانند ماشین بردار پشتیبان و مدل‌های تجمیعی نظیر الگوریتم تقویت تطبیقی و همچنین انتخاب ویژگی‌های مؤثر به کار گرفته شده‌اند تا صحت پیش‌بینی ژانرهای مورد علاقه کاربران افزایش یابد. در پژوهش سو و همکاران^۲ (۲۰۲۱) نیز استفاده از ترجیحات ژانری کاربران برای بهبود سیستم‌های توصیه‌گر مورد توجه قرار گرفته است؛ به این صورت که یک سیستم توصیه‌گر گروهی مبتنی بر ژانر ارائه شده است که در آن، بردار علایق ژانری کاربران به عنوان ورودی مدل برای خوشه‌بندی گروه‌ها و وزن‌دهی به ترجیحات آیت‌ها مورد استفاده قرار گرفته است. این رویکرد موجب کاهش چشمگیر زمان خوشه‌بندی شده و دقت توصیه‌ها را نیز حفظ کرده است. ویدیانیگ‌تاس و همکاران^۳ (۲۰۲۱) الگوریتم شباهت مبتنی بر همبستگی پروفایل کاربر را برای سیستم‌های پالایش مشارکتی معرفی کردند که علاوه بر داده‌های ژانر، اطلاعات پروفایل کاربر شامل سن، جنسیت، شغل و موقعیت مکانی را برای تعیین وزن شباهت بین امتیازدهی و رفتار کاربران در نظر می‌گیرد و باعث کاهش معیار خطای میانگین مطلق به میزان ۱/۶۴ درصد می‌شود. پال و همکاران^۴ (۲۰۲۰) سیستمی را معرفی کرده‌اند که از ژانرهای فیلم‌های مختلف و امتیازات فیلم‌ها به عنوان ورودی برای آموزش مدل‌های یادگیری ماشین استفاده می‌کند. مدل‌ها شامل الگوریتم‌های تجمیعی و شبکه‌های عصبی هستند که در این میان، شبکه عصبی عمیق بهترین عملکرد را در معیار ریشه میانگین مربعات خطا با مقدار ۰/۰۶ ارائه داده است به گونه‌ای که این سیستم امکان ارائه پیشنهاد‌های دقیق‌تر و هدفمند را فراهم می‌کند. در ادامه، موندال و همکاران^۵ (۲۰۲۳) در پژوهش خود، یک سیستم توصیه‌گر مبتنی بر سازوکار توجه^۶ را برای مجموعه داده فیلم‌های هندی چندرسانه‌ای ارائه کرده‌اند که بازخورد کاربران را در سه کلاس (نپسندیدن، پسندیدن، خنثی/تماشا نشده) پیش‌بینی می‌کند. ورودی مدل شامل ویژگی‌های صوتی و تصویری استخراج شده از تریلر فیلم‌ها بود و عملکرد آن با استفاده از امتیاز تطابق ژانر ارزیابی شد که بیانگر میزان هم‌راستایی ژانر فیلم با علایق ژانری کاربران است. نتایج نشان داد که هم‌راستایی ژانر فیلم با ژانرهای مورد علاقه کاربر باعث می‌شود مدل بهتر بتواند پسندیدن/نپسندیدن را پیش‌بینی کند و ترکیب اطلاعات صوتی و تصویری نیز موجب بهبود عملکرد سیستم شده است. ملیکا و کارویی^۷ (۲۰۲۰) هم یک سیستم توصیه‌گر مبتنی بر زنجیره‌های مارکوف و خوشه‌بندی ژانرها ارائه کرده‌اند که تغییرات کوتاه‌مدت ترجیحات کاربران را در نظر می‌گیرد. ورودی مدل شامل تعاملات اخیر کاربران و ژانرهای فیلم‌های مشاهده شده است. این سیستم با استفاده از زنجیره‌های مارکوف، اقدامات بعدی کاربر را بر اساس تعاملات اخیر او پیش‌بینی می‌کند و با الگوریتم خوشه‌بندی مبتنی بر چگالی، علاقه‌مندی‌های هر کاربر را به طور دقیق شناسایی می‌کند. ارزیابی این روش بر روی مجموعه داده فیلم‌های پلتفرم محبوب نتفلیکس نشان داد که این روش قادر است پیشنهاد‌هایی هدفمندتر و

^۱ Kim et al.

^۲ Seo et al.

^۳ Widiyaningtyas et al.

^۴ Pal et al.

^۵ Mondal et al.

^۶ Cross-attention-based recommender

^۷ Mlika & Karoui

مرتبط با نیازهای واقعی کاربران ارائه دهد. راج^۱ و همکاران (۲۰۲۴) یک رویکرد نوین برای شناسایی فیلم‌های موردعلاقه کاربران، با استفاده از مدل‌های زبان بزرگ ارائه کردند که شامل پیش‌بینی ژانر فیلم و ارائه پیشنهادها برای شخصی‌سازی شده است. آن‌ها زیرنویس صوتی تریلرهای فیلم را به‌عنوان ورودی برای مدل‌های زبانی و تولید خلاصه فیلم‌ها استفاده کرده‌اند و این اطلاعات را برای پیش‌بینی ژانر و ارائه پیشنهاد به کاربران به کار گرفتند. ارزیابی روی مجموعه داده MovieLens-100K نشان داد که مدل‌های زبان بزرگ قادر به درک دقیق ترجیحات کاربران و ویژگی‌های فیلم‌ها هستند و پیشنهادهای دقیق‌تری نسبت به مدل‌های پایه ارائه می‌دهند.

همان‌طور که مشاهده شد، پیش‌بینی ژانر مورد علاقه کاربران در پژوهش‌های مختلف مورد توجه قرار گرفته است. با این حال، در هیچ یک از این پژوهش‌ها ویژگی‌های شخصیتی کاربران برای پیش‌بینی علایق ژانری مورد استفاده قرار نگرفته است. این موضوع نوآوری پژوهش حاضر محسوب می‌شود.

• ارتباط میان ویژگی‌های شخصیتی کاربران و علاقه‌مندی‌های آن‌ها در پلتفرم‌های پخش فیلم

در این بخش، پژوهش‌هایی مرور می‌گردند که نشان داده‌اند ارتباطی بین ویژگی‌های شخصیتی کاربران و علاقه‌مندی‌های آن‌ها به ژانرهای فیلم وجود دارد. با این حال، در هیچ یک از این پژوهش‌ها، از وجود این ارتباط برای پیش‌بینی علایق کاربران به ژانرها استفاده نشده است. بلکه این پژوهش‌ها، صرفاً تاییدی بر وجود چنین ارتباطی هستند.

گلبک و نوریس^۲ (۲۰۱۳) نشان داده‌اند که ویژگی‌های شخصیتی، به ویژه وظیفه‌شناسی و برون‌گرایی، نقش مهمی در نحوه تعامل کاربران با پلتفرم‌های ارائه‌دهنده فیلم دارند؛ افرادی که این ویژگی‌ها در آن‌ها برجسته‌تر است، معمولاً امتیازهای بیشتری برای فیلم‌ها ثبت می‌کنند. علاوه بر این، شخصیت کاربران می‌تواند بر فرآیند تصمیم‌گیری آن‌ها در انتخاب فیلم تأثیرگذار باشد؛ به طور مثال، ویژگی توافق‌پذیری بر نحوه امتیازدهی به فیلم‌ها و میزان مصرف محتوای کاربران اثرگذار است (کارومور^۳ و همکاران، ۲۰۱۶). مطالعات دیگری نیز نشان داده‌اند که نوع فیلم‌های مورد علاقه افراد می‌تواند بازتاب‌دهنده ویژگی‌های شخصیتی آن‌ها باشد (نال‌مپانتیس و تیورتجیس^۴، ۲۰۱۷، وو و چن^۵، ۲۰۱۵)؛ به عنوان نمونه، افراد برون‌گرا گرایش بیشتری به فیلم‌های اکشن و کمدی دارند، در حالی که کاربران با سطح توافق‌پذیری بالاتر، به فیلم‌های درام و عاشقانه تمایل نشان می‌دهند. همچنین کسانی که تجربه‌پذیری بالایی دارند، به فیلم‌های موزیکال و انیمیشن علاقه بیشتری دارند. هال^۶ (۲۰۰۵) نشان داده است که برون‌گرایی، روان‌رنجوری و روان‌پریشی بر ترجیحات رسانه‌ای افراد تأثیر قابل توجه دارد؛ برای مثال، افراد برون‌گرا به تماشای فیلم‌های اکشن و موسیقی کلاسیک علاقه دارند، در حالی که افراد روان‌رنجور بیشتر به موسیقی پاپ-راک گرایش دارند. مطابق یافته‌های چن^۷ و همکاران (۲۰۱۶)، ویژگی‌های شخصیتی کاربران بر تنوع علاقه‌مندی‌های آن‌ها نیز اثرگذار است؛ افراد با وظیفه‌شناسی پایین‌تر، به ژانرهای متنوع‌تری علاقه نشان می‌دهند و افراد تجربه‌پذیر، به بازیگران مختلف توجه دارند. خان^۸ و همکاران (۲۰۲۰) و نیو^۹ و همکاران (۲۰۲۰)، با تحلیل داده‌های رسانه‌های اجتماعی و امتیازات ثبت‌شده برای فیلم‌ها، نشان داده‌اند که ارتباط قوی بین ویژگی‌های شخصیتی و انتخاب ژانر وجود دارد؛ افراد با گشودگی بالا به تجربیات جدید، به فیلم‌های کمدی علاقه دارند، افراد روان‌رنجور به درام تمایل دارند، علاقه‌مندان به ژانر اکشن معمولاً برون‌گرا هستند و طرفداران فانتزی و علمی‌تخیلی به تجربه‌های نوآورانه گرایش دارند. با وجود اینکه مطالعات روان‌شناختی وجود رابطه میان ویژگی‌های شخصیتی و ترجیحات و رفتارهای کاربران در انتخاب و مصرف فیلم را تأیید کرده‌اند، تاکنون مدل‌های یادگیری ماشین برای پیش‌بینی علایق ژانری کاربران توسعه نیافته‌اند. پژوهش حاضر با هدف پر کردن این شکاف انجام شده است.

¹ Raj

² Golbeck & Norris

³ Karumur

⁴ Nalmpantis & Tjortjijis

⁵ Wu & Chen

⁶ Hall

⁷ Chen

⁸ Khan

⁹ Nave

روش پژوهش

مطالعات پیشین، تایید می‌کنند که ارتباط معناداری بین ویژگی‌های شخصیتی کاربران و علاقه‌مندی آن‌ها به ژانرهای مختلف فیلم وجود دارد. این پژوهش از این ارتباط به‌عنوان مبنایی برای پیش‌بینی میزان علاقه کاربران به ژانرهای مختلف فیلم بهره می‌برد. به‌منظور افزایش دقت مدل‌ها، ویژگی‌های جمعیت‌شناختی کاربران (نظیر سن، جنسیت و مکان جغرافیایی) نیز در کنار ویژگی‌های شخصیتی مورد استفاده قرار می‌گیرند تا با ترکیب این دو دسته ویژگی، دقت برآورد علایق ژانری کاربران بهبود یابد. روش پیشنهادی شامل سه گام اصلی است که در شکل ۱ مشاهده می‌شود. در ادامه، ابتدا داده‌های مورد استفاده معرفی خواهد شد و سپس به معرفی هر یک از گام‌ها پرداخته خواهد شد.

معرفی داده‌ها: در این پژوهش فرض بر این است که ویژگی‌های شخصیتی کاربران در پروفایل آن‌ها ثبت شده و در دسترس است. در این راستا، از مجموعه‌داده MovieLens100K (گروپلنز ریسرچ^۱، ۲۰۱۷، هارپر و کانستان^۲، ۲۰۱۷) استفاده شده است. این مجموعه‌داده شامل ۱۰۰۰۰۰ امتیازدهی ثبت‌شده توسط ۱۰۰۰ کاربر به ۱۷۰۰ فیلم است. به‌منظور جمع‌آوری اطلاعات شخصیتی این کاربران نیز، نگوین^۳ و همکاران (۲۰۱۸) از گروه پژوهشی گروپلنز^۴ پرسشنامه‌ای مبتنی بر پژوهش گسلینگ^۵ و همکاران (۲۰۰۳) را در اختیار کاربران این مجموعه داده‌ی MovieLens قرار داد. در این پرسشنامه، کاربران به مجموعه‌ای از سؤالات ارزیابی شخصیت بر اساس مقیاس لیکرت^۶ پاسخ دادند که دامنه پاسخ‌ها از ۱ (کاملاً مخالف) تا ۷ (کاملاً موافق) متغیر بود. بر اساس این پاسخ‌ها، برای هر کاربر، امتیازی در بازه ۱ تا ۷ برای هر یک از پنج بُعد اصلی شخصیت محاسبه شد. افزون بر این، در این نسخه از داده‌ها، اطلاعات جمعیت‌شناختی کاربران نیز در دسترس است. همچنین، به‌منظور استخراج اطلاعات تکمیلی مربوط به فیلم‌ها، از مجموعه‌داده IMDb (سینگ، ۲۰۲۱) استفاده شده است. این مجموعه‌داده شامل اطلاعات جامعی از فیلم‌ها از جمله نام، تاریخ انتشار، امتیاز کاربران، ژانرها، شرح کوتاه داستان فیلم، بازیگران و عوامل تولید، عناوین اصلی، وضعیت تولید، زبان‌های اصلی، بودجه، میزان درآمد و کشورهای مبدأ است. در پژوهش ما از این اطلاعات، به‌ویژه خلاصه‌های داستان، برای تشخیص ژانر اصلی فیلم‌ها استفاده می‌شود.

حريم خصوصی و محرمانگی اطلاعات: در این پژوهش از داده‌های عمومی مجموعه‌داده MovieLens، شامل اطلاعات شخصیتی و جمعیت‌شناختی کاربران، استفاده شده است. این اطلاعات پیش‌تر توسط پژوهشگران جمع‌آوری و منتشر شده و در این پژوهش هیچ داده شخصی جدیدی از کاربران دریافت نشده است. همچنین، تحلیل‌های انجام‌شده در این پژوهش بر داده‌های موجود و شناسه‌های پژوهشی انجام شده و از انتشار یا گزارش اطلاعاتی که بتواند کاربران را به‌صورت مستقیم شناسایی کند، خودداری شده است. داده‌های مورد استفاده صرفاً برای اهداف پژوهشی و در چارچوب شرایط استفاده از مجموعه‌داده‌های مذکور به کار گرفته شده‌اند.

گام اول: محاسبه میزان علاقه کاربران به ژانرها بر پایه امتیازدهی آن‌ها به فیلم‌ها

مجموعه‌داده MovieLens حاوی امتیازات کاربران به فیلم‌های مختلف است که میزان علاقه کاربران به فیلم‌ها را در قالب اعدادی در بازه ۱ تا ۵ نشان می‌دهد. با این حال، در این پژوهش، به جای علاقه کاربران به فیلم‌ها، میزان علاقه‌ی کاربران به ژانرهای سینمایی مورد نیاز است. بدین منظور، ابتدا ۱۸ ژانر اصلی مختلف فیلم در نظر گرفته می‌شود که عبارتند از ژانرهای اکشن، ماجراجویی، انیمیشن، کودکانه، کمدی، جنایی، مستند، درام، فانتزی، سیاه، ترسناک، موزیکال، معمایی، عاشقانه، علمی-تخیلی، هیجان‌انگیز، جنگی و وسترن. ژانر اصلی هر یک از فیلم‌های موجود در مجموعه‌داده، به روشی که در ادامه معرفی می‌شود، از بین این ۱۸ ژانر انتخاب می‌گردد. سپس برای هر کاربر و هر ژانر، میانگین امتیازات کاربر به فیلم‌هایی با آن ژانر اصلی، محاسبه می‌شود و به عنوان میزان علاقه آن کاربر به آن ژانر در نظر گرفته می‌شود. اگر کاربر هیچ فیلمی از یک ژانر خاص را امتیازدهی نکرده باشد، مقدار علاقه آن کاربر به آن ژانر، یک مقدار بی‌اثر^۷ در نظر گرفته می‌شود؛ به این

^۱ GroupLens Research

^۲ Harper & Konstan

^۳ Nguyen

^۴ GroupLens research

^۵ Gosling

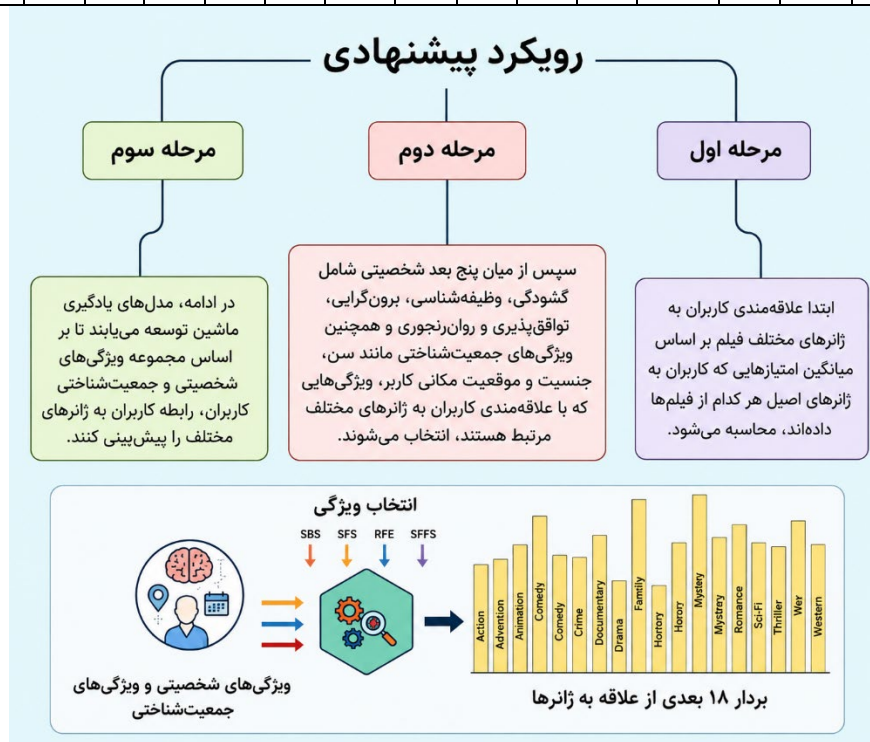
^۶ Likert scale

^۷ Null

معنا که برای آن ژانر اطلاعات کافی برای تعیین میزان علاقه کاربر وجود ندارد. بنابراین، مقادیر معتبر علاقه ژانری در بازه ۱ تا ۵ قرار دارند و مقدار بی‌اثر نشان‌دهنده فقدان سابقه امتیازدهی به آن ژانر است و بیانگر میزان علاقه کاربر نیست. در نتیجه، مجموعه داده‌ای شامل امتیازهای کاربران به هر ژانر ایجاد می‌شود که در آن، برای هر کاربر یک بردار ۱۸ مؤلفه‌ای از علاقه او به ژانرها ایجاد شده است و هر مؤلفه بردار، نشان‌دهنده میزان علاقه‌ی کاربر به یکی از ژانرها است و در بازه ۱ تا ۵ قرار دارد. عناوین ۱۸ ژانر در نظر گرفته شده در این پژوهش به همراه نمونه‌ای از امتیازات کاربران به این ژانرها در جدول ۱ مشاهده می‌شود.

جدول ۱: مثالی از امتیازات سه نمونه کاربر به ژانرهای مختلف فیلم در بازه ۱ تا ۵

Western	War	Thriller	Sci-fi	Romance	Mystery	Musical	Horror	Film-Noir	Fantasy	Drama	Documentary	Crime	Comedy	Children	Animation	Adventure	Action	User_id
۳/۰	۴/۰	۱/۰	۲/۲	۴/۲	۳/۱	۲/۴	۴/۸	۱/۶	۲/۷	۳/۳	۳/۷	۱/۴	۴/۴	۲/۹	۳/۵	۳/۰	۲/۹	۱۴۷
۳/۰	۳/۵	۲/۲	۱/۷	۱/۵	۲/۰	۱/۵	۴/۶	۳/۴	۲/۶	۰	۴/۵	۰	۲/۰	۲/۹	۴/۰	۳/۵	۳/۲	۲۲۶
۲/۵	۵/۰	۲/۰	۱/۴	۲/۳	۴/۰	۰	۱/۵	۵/۰	۳/۵	۲/۶	۲/۸	۰	۳/۴	۴/۳	۳/۹	۴/۰	۴/۱	۷۸۳



شکل ۱: شمای کلی روش پیشنهادی برای پیش‌بینی علاقه‌مندی‌های ژانری کاربران

شناسایی ژانر اصلی فیلم‌ها: در این مرحله، ابتدا اطلاعات مربوط به هر فیلم از مجموعه داده IMDB استخراج می‌شود. سپس، برای تعیین ژانر اصلی هر فیلم، از روش تطبیق معنایی بهره گرفته می‌شود. برای این منظور، خلاصه داستان یا شرح کلی فیلم‌ها^۱ بهره‌گیری از مدل قدرتمند (all-MiniLM-L6-v2) SentenceTransformer، به بردارهای عددی با طول ۳۸۴ تبدیل می‌گردند تا امکان مقایسه عمیق معنایی متن‌ها فراهم آید (رایمرز و گورویچ، ۲۰۱۹). برای هر ژانر نیز توضیحی

¹ Overview

² Reimers & Gurevych

مختصر از وبسایت ویکی‌پدیا^۱ جمع‌آوری و به بردار عددی با طول ۳۸۴ تبدیل می‌گردد. در نهایت، با محاسبه شباهت کسینوسی بین بردار فیلم و بردارهای ژانرها، ژانری که بیشترین شباهت معنایی را داراست، به عنوان ژانر غالب یا اصلی انتخاب می‌گردد. این فرایند یک روش انتساب ژانر مبتنی بر شباهت معنایی است و به دلیل نبود داده‌های آموزشی دارای برچسب مرجع، یک مسئله طبقه‌بندی نظارت‌شده محسوب نمی‌شود؛ بنابراین، ارزیابی این مرحله بر مبنای معیارهای متداول طبقه‌بندی انجام نمی‌شود.

گام دوم: انتخاب ویژگی‌های مرتبط با علایق ژانری

در این گام، ویژگی‌های شخصیتی و جمعیت‌شناختی مرتبط با پیش‌بینی علاقه کاربران به هر یک از ژانرهای فیلم شناسایی می‌شوند. هدف از انتخاب ویژگی، کاهش ابعاد داده‌های ورودی، حذف ویژگی‌های کم‌ارتباط و فراهم کردن مجموعه ویژگی مناسب برای آموزش مدل‌های پیش‌بینی است. این فرایند می‌تواند از پیچیدگی مدل‌ها و احتمال بیش‌برازش نیز بکاهد. از آنجا که متغیر هدف در این پژوهش میزان علاقه کاربر به هر ژانر است که به صورت یک مقدار عددی در بازه ۱ تا ۵ تعریف می‌شود، مسئله انتخاب ویژگی در چارچوب یک مسئله رگرسیون انجام می‌شود. ویژگی‌های مورد بررسی شامل پنج ویژگی شخصیتی گشودگی به تجربه، وظیفه‌شناسی، برون‌گرایی، توافق‌پذیری و روان‌رنجوری، به همراه ویژگی‌های جمعیت‌شناختی شامل سن، جنسیت و موقعیت مکانی هستند. برای جلوگیری از ورود اطلاعات مجموعه آزمون به فرایند انتخاب ویژگی، داده‌های هر ژانر ابتدا به دو بخش آموزش و آزمون تقسیم شدند. انتخاب ویژگی صرفاً بر روی مجموعه آموزش انجام شد و مجموعه آزمون تا مرحله ارزیابی نهایی مدل مورد استفاده قرار نگرفت. در این پژوهش، چهار روش انتخاب ویژگی شامل جستجوی پیش‌رونده^۲ یا SFS، جستجوی پس‌رونده^۳ یا SBS، جستجوی شناور^۴ یا SFFS و حذف بازگشتی ویژگی‌ها^۵ یا RFE به کار گرفته شدند. این روش‌ها با هدف یافتن ترکیبی از ویژگی‌ها که بتواند عملکرد مدل رگرسیون را بهبود دهد، به صورت جداگانه برای هر ژانر اجرا شدند.

در روش SFS، فرایند با یک مجموعه خالی آغاز می‌شود. در هر مرحله، هر یک از ویژگی‌های موجود که هنوز انتخاب نشده‌اند، به صورت موقت به مجموعه فعلی اضافه شده و عملکرد مدل ارزیابی می‌شود. ویژگی‌ای که بیشترین بهبود را در عملکرد ایجاد کند، به مجموعه اضافه می‌شود و این فرایند تا دستیابی به بهترین ترکیب ادامه می‌یابد. در مقابل، در روش SBS، فرایند با مجموعه‌ای شامل تمام ویژگی‌ها آغاز می‌شود. در هر مرحله، حذف هر یک از ویژگی‌های موجود به صورت جداگانه بررسی شده و ویژگی‌ای که حذف آن کمترین افت یا بیشترین بهبود را در عملکرد مدل ایجاد کند، حذف می‌شود. این فرایند تا دستیابی به مجموعه مناسب ادامه پیدا می‌کند.

روش SFFS برای رفع محدودیت جستجوی صرفاً پیش‌رونده SFS طراحی شده است. این روش ابتدا مانند SFS ویژگی‌ها را به مجموعه اضافه می‌کند، اما پس از هر بار اضافه شدن یک ویژگی، بررسی می‌کند که آیا حذف یکی از ویژگی‌های انتخاب‌شده می‌تواند عملکرد مدل را بهبود دهد یا خیر. در صورت وجود چنین حالتی، ویژگی نامناسب حذف می‌شود و جستجو ادامه پیدا می‌کند. در نتیجه، این روش امکان بازنگری در انتخاب‌های قبلی را فراهم می‌کند. در روش RFE، ابتدا مدل با مجموعه ویژگی‌های موجود آموزش داده می‌شود و اهمیت نسبی ویژگی‌ها تعیین می‌گردد. سپس کم‌اهمیت‌ترین ویژگی حذف شده و مدل مجدداً با ویژگی‌های باقی‌مانده ارزیابی می‌شود. این فرایند به صورت تکراری ادامه می‌یابد تا مجموعه‌ای با تعداد ویژگی مطلوب حاصل شود.

در این روش‌ها برای مقایسه مجموعه‌های مختلف ویژگی از اعتبارسنجی متقابل پنج‌بخشی^۶ در مجموعه آموزش استفاده شد؛ به این صورت که داده‌های آموزش به پنج بخش تقسیم شده و در هر مرحله، مدل با چهار بخش آموزش و با بخش باقی‌مانده ارزیابی می‌شد و این فرایند برای هر پنج بخش تکرار می‌گردید. در نهایت، عملکرد حاصل از این ارزیابی‌ها برای

^۱ https://en.wikipedia.org/wiki/Film_genre

^۲ Sequential Forward Selection

^۳ Sequential Backward Selection

^۴ Sequential Floating Forward Selection

^۵ Recursive Feature Elimination

^۶ 5-fold cross-validation

مقایسه مجموعه‌های ویژگی مورد استفاده قرار گرفت. هر یک از این روش‌ها برای هر ژانر به صورت جداگانه اجرا می‌شوند. عملکرد حاصل از مجموعه‌های ویژگی مختلف بر اساس معیار میانگین خطای مطلق (MAE) ارزیابی شده و مجموعه ویژگی دارای کمترین خطای پیش‌بینی به عنوان مجموعه ویژگی منتخب برای ژانر مربوطه در نظر گرفته می‌شود. نتایج این فرآیند در بخش یافته‌ها ارائه شده است.

گام سوم: مدل‌سازی پیش‌بینی علاقه‌مندی‌های ژانری

پیش‌بینی علاقه کاربران به ژانرهای گوناگون فیلم، به عنوان یک مسئله رگرسیون در یادگیری ماشین در نظر گرفته شده است. هدف این مسئله، برآورد یک مقدار کمی (میزان علاقه‌مندی به هر ژانر) در بازه ۱ تا ۵، بر اساس ویژگی‌های شخصی و جمعیت‌شناختی هر کاربر است. در این رویکرد، برای هر یک از ۱۸ ژانر فیلم، یک مدل یادگیری ماشین جداگانه آموزش می‌دهیم. به این ترتیب ۱۸ مدل مستقل از یکدیگر ایجاد می‌شوند. این مدل‌های یادگیری ماشین شامل الگوریتم‌هایی مانند رگرسیون بردار پشتیبان^۱ (SVR)، رگرسیون جنگل تصادفی (RF)، گرادیان بوستینگ (XGB)، رگرسیون k نزدیک‌ترین همسایه^۲ (KNN) و شبکه عصبی پیش‌خور^۳ (FNN) هستند. پس از تعیین ویژگی‌های منتخب در مرحله قبل، داده‌های مربوط به هر ژانر با نسبت ۸۰ درصد برای آموزش و ۲۰ درصد برای آزمون تقسیم شدند. مدل‌ها با استفاده از مجموعه آموزش یاد گرفتند رابطه میان ویژگی‌های ورودی و میزان علاقه کاربران به ژانر مورد نظر را برآورد کنند و سپس عملکرد آن‌ها با استفاده از مجموعه آزمون مستقل ارزیابی شد. هر مدل با تنظیم پارامترهای کنترلی مناسب، بهینه‌سازی شده است. این ۱۸ مدل، بر اساس منتخبی از ویژگی‌های شخصی و جمعیت‌شناختی (که در مرحله انتخاب ویژگی انتخاب شدند)، علاقه‌مندی‌های ژانری کاربران را پیش‌بینی می‌کنند. به عنوان مثال، همانطور که در ردیف اول جدول ۲ مشاهده می‌شود، مدل مربوط به ژانر اکشن با دریافت ویژگی‌های شخصی (گشودگی، توافق‌پذیری، روان‌رنجوری، وظیفه‌شناسی) و دموگرافیک (سن، جنسیت، موقعیت مکانی) یک کاربر، میزان علاقه‌مندی آن کاربر به ژانر اکشن را در قالب یک مقدار عددی پیوسته در بازه ۱ تا ۵ پیش‌بینی می‌کند.

معیارهای ارزیابی و تنظیم مدل‌ها: برای ارزیابی عملکرد مدل‌های پیش‌بینی، از چهار معیار متداول در مسائل رگرسیون شامل میانگین خطای مطلق (MAE)، میانگین مربعات خطا (MSE)، ریشه میانگین مربعات خطا (RMSE) و ضریب تعیین^۴ (R^2) استفاده می‌شود. معیار MAE میانگین قدرمطلق اختلاف میان مقادیر واقعی و مقادیر پیش‌بینی شده را اندازه‌گیری می‌کند. MSE میانگین مربع خطاهای پیش‌بینی را محاسبه کرده و RMSE نیز ریشه دوم MSE است. در مقابل، R^2 میزان توانایی مدل در توضیح تغییرات متغیر هدف را نشان می‌دهد. در معیارهای MAE، MSE و RMSE، مقدار کمتر نشان‌دهنده عملکرد بهتر مدل است، در حالی که مقدار بیشتر R^2 بیانگر عملکرد بهتر است. برای هر ژانر، عملکرد مدل‌های مختلف با استفاده از این معیارها مقایسه شده و مدلی که عملکرد بهتری داشته باشد، به عنوان مدل منتخب آن ژانر تعیین می‌شود.

یافته‌های پژوهش

در این بخش، نتایج حاصل از اجرای روش پیشنهادی ارائه می‌شود. ابتدا نتایج فرآیند انتخاب ویژگی برای هر یک از ۱۸ ژانر بررسی شده و سپس عملکرد مدل‌های مختلف یادگیری ماشین در پیش‌بینی علاقه‌مندی‌های ژانری مقایسه می‌شود. در ادامه نیز عملکرد روش پیشنهادی با روش پیشین مورد مقایسه قرار می‌گیرد.

^۱ Support Vector Regression

^۲ K-Nearest Neighbor regression

^۳ Feedforward Neural Network

^۴ Coefficient of determination

نتایج فرآیند انتخاب ویژگی

پس از اجرای چهار روش SBS، SFS، SFFS و RFE برای هر یک از ۱۸ ژانر، مناسب‌ترین مجموعه ویژگی برای هر ژانر بر اساس مقدار MAE تعیین شد. به بیان دیگر، در این مرحله مشخص شد کدام ویژگی‌های شخصیتی و جمعیت‌شناختی، اطلاعات بیشتری برای پیش‌بینی علاقه به هر ژانر فراهم می‌کنند. نتایج نشان می‌دهد که مجموعه ویژگی بهینه در ژانرهای مختلف یکسان نیست و ترکیب متفاوتی از ویژگی‌های شخصیتی و جمعیت‌شناختی برای پیش‌بینی هر ژانر به دست آمده است. این موضوع نشان می‌دهد که نقش ویژگی‌های فردی کاربران در پیش‌بینی علاقه، به نوع ژانر وابسته است. همان‌طور که در جدول ۲ مشاهده می‌شود، در برخی ژانرها ترکیبی از ویژگی‌های شخصیتی و جمعیت‌شناختی انتخاب شده است، در حالی که برای برخی ژانرها تنها تعدادی از ویژگی‌های شخصیتی یا ترکیبی محدود از ویژگی‌های شخصیتی و جمعیت‌شناختی بیشترین توان پیش‌بینی را داشته‌اند. این موضوع نشان می‌دهد که نقش ویژگی‌های کاربران در پیش‌بینی ترجیحات ژانری، وابسته به ژانر هدف است. برای مثال، در ژانر اکشن، ویژگی‌های گشودگی، توافق‌پذیری، روان‌رنجوری و وظیفه‌شناسی به همراه سن، جنسیت و کشور انتخاب شده‌اند. در مقابل، برای ژانر کمدی، تنها ویژگی‌های روان‌رنجوری، وظیفه‌شناسی و برون‌گرایی به‌عنوان ویژگی‌های منتخب شناسایی شده‌اند. همچنین، برای ژانر درام، ویژگی‌های گشودگی، روان‌رنجوری، سن، کشور و جنسیت انتخاب شده‌اند.

جدول ۲. ویژگی‌های منتخب برای مدل‌های پیش‌بینی‌کننده میزان علاقه به ژانرها در مقیاس ۱ تا ۵

ژانر هدف	روش	ویژگی‌های منتخب	MAE
اکشن	SBS	گشودگی، توافق‌پذیری، روان‌رنجوری، وظیفه‌شناسی، سن، جنسیت، کشور	۰/۲۱
ماجراجویی	SFS	گشودگی، توافق‌پذیری، روان‌رنجوری، وظیفه‌شناسی، کشور	۰/۲۲
انیمیشن	SBS	توافق‌پذیری، روان‌رنجوری، برون‌گرایی، کشور	۰/۳۵
کودکانه	SBS	گشودگی، توافق‌پذیری، کشور، وظیفه‌شناسی، جنسیت	۰/۳۹
کمدی	RFE	روان‌رنجوری، وظیفه‌شناسی، برون‌گرایی	۰/۲۲
جنایی	SBS	گشودگی، توافق‌پذیری، روان‌رنجوری، برون‌گرایی، سن، کشور، جنسیت	۰/۲۲
مستند	RFE	توافق‌پذیری، وظیفه‌شناسی، برون‌گرایی	۰/۳۸
درام	SFFS	گشودگی، روان‌رنجوری، سن، کشور، جنسیت	۰/۱۸
فانتزی	RFE	روان‌رنجوری، وظیفه‌شناسی، برون‌گرایی	۰/۳۲
سیاه	SFS	گشودگی، وظیفه‌شناسی، برون‌گرایی، سن، کشور	۰/۴۰
ترسناک	SFFS	گشودگی، روان‌رنجوری، برون‌گرایی، وظیفه‌شناسی، سن، جنسیت	۰/۳۴
موزیکال	SBS	توافق‌پذیری، جنسیت، سن، کشور	۰/۴۵
معمایی	SFS	توافق‌پذیری، روان‌رنجوری، برون‌گرایی، کشور	۰/۲۷
عاشقانه	SFFS	توافق‌پذیری، روان‌رنجوری، برون‌گرایی، کشور، سن	۰/۳۰
علمی-تخیلی	RFE	وظیفه‌شناسی، روان‌رنجوری، برون‌گرایی	۰/۲۵
هیجان‌انگیز	RFE	وظیفه‌شناسی، روان‌رنجوری، برون‌گرایی	۰/۲۵
جنگی	SFS	گشودگی، توافق‌پذیری، روان‌رنجوری، برون‌گرایی، جنسیت	۰/۲۹
وسترن	SFS	گشودگی، توافق‌پذیری، روان‌رنجوری، سن، جنسیت، کشور	۰/۳۸

نتایج مدل‌سازی پیش‌بینی علاقه‌مندی‌های ژانری

پس از تعیین ویژگی‌های منتخب برای هر ژانر، مدل‌های RF، SVR، XGB، KNN و FNN با استفاده از ۸۰ درصد داده‌های هر ژانر آموزش داده شدند و ۲۰ درصد باقی‌مانده برای ارزیابی نهایی در نظر گرفته شد. بدین ترتیب، عملکرد هر مدل بر روی داده‌هایی سنجیده شد که در فرایند آموزش آن مدل استفاده نشده بودند. به عبارت دیگر در این مرحله، مدل‌ها با استفاده از ویژگی‌های منتخب تلاش کردند میزان علاقه هر کاربر به هر ژانر را پیش‌بینی کنند و سپس عملکرد آن‌ها با یکدیگر مقایسه شد. عملکرد مدل‌ها با معیارهای MAE، MSE، RMSE و R^2 مورد ارزیابی قرار گرفت و برای هر ژانر، مدلی که بهترین عملکرد را داشت به‌عنوان مدل نهایی انتخاب شد. پارامترهای کنترلی بهینه مدل‌های مختلف با استفاده از

جستجوی شبکه‌ای تعیین شدند. به این معنا که ترکیب‌های مختلف تنظیمات هر مدل بررسی و بهترین ترکیب بر اساس عملکرد آن انتخاب شد. نتایج این فرآیند و بهترین تنظیمات به‌دست‌آمده برای مدل‌های مورد استفاده در جدول ۳ ارائه شده است.

جدول ۳. پارامترهای کنترلی مدل‌های منتخب پیش‌بینی‌کننده علایق ژانری منتخب

ژانر هدف	بهترین مدل	پارامترهای کنترلی
اکشن	رگرسیون جنگل تصادفی	تعداد درخت‌ها = ۲۰۰، حداکثر عمق = نامحدود تعداد ویژگی‌ها در هر تقسیم = ریشه دوم تعداد کل ویژگی‌ها
انیمیشن	رگرسیون جنگل تصادفی	تعداد درخت‌ها = ۲۰۰، حداکثر عمق = ۵۰ تعداد ویژگی‌ها در هر تقسیم = لگاریتم تعداد کل ویژگی‌ها
جنایی	رگرسیون جنگل تصادفی	تعداد درخت‌ها = ۵۰۰، حداکثر عمق = ۲۰ تعداد ویژگی‌ها در هر تقسیم = ریشه دوم تعداد کل ویژگی‌ها
سیاه	رگرسیون جنگل تصادفی	تعداد درخت‌ها = ۱۰۰، حداکثر عمق = ۴۰ تعداد ویژگی‌ها در هر تقسیم = لگاریتم تعداد کل ویژگی‌ها
ترسناک	رگرسیون جنگل تصادفی	تعداد درخت‌ها = ۴۰۰، حداکثر عمق = ۳۰ تعداد ویژگی‌ها در هر تقسیم = لگاریتم تعداد کل ویژگی‌ها
وسترن	رگرسیون جنگل تصادفی	تعداد درخت‌ها = ۵۰۰، حداکثر عمق = نامحدود تعداد ویژگی‌ها در هر تقسیم = لگاریتم تعداد کل ویژگی‌ها
ماجراجویی	شبکه عصبی پیشخور	لایه‌های پنهان = [۶۴ و ۳۲]، نرخ حذف تصادفی = ۰/۲ تعداد دوره‌های آموزش = ۵۰
علمی-تخیلی	شبکه عصبی پیشخور	لایه‌های پنهان = [۴۸ و ۲۴]، نرخ حذف تصادفی = ۰/۱۵ تعداد دوره‌های آموزش = ۷۰
مستند	شبکه عصبی پیشخور	لایه‌های پنهان = [۲۰ و ۴۰]، نرخ حذف تصادفی = ۰/۱ تعداد دوره‌های آموزش = ۸۰
کودکانه	رگرسیون بردار پشتیبان	تابع هسته = شعاعی، $\gamma = ۰/۲$ ، $\epsilon = ۷$ = مقیاس‌بندی‌شده
درام	رگرسیون بردار پشتیبان	تابع هسته = شعاعی، $\gamma = ۰/۱۵$ ، $\epsilon = ۷$ = خودکار
موزیکال	رگرسیون بردار پشتیبان	تابع هسته = شعاعی، $\gamma = ۰/۲۵$ ، $\epsilon = ۷$ = مقیاس‌بندی‌شده
جنگی	رگرسیون بردار پشتیبان	تابع هسته = شعاعی، $\gamma = ۰/۲$ ، $\epsilon = ۷$ = خودکار
کمدی	شبکه عصبی پیشخور	لایه‌های پنهان = [۳۲ و ۶۴ و ۱۲۸]، نرخ حذف تصادفی = [۰/۲ و ۰/۲] تعداد دوره‌های آموزش = ۵۰
فانتزی	شبکه عصبی پیشخور	لایه‌های پنهان = [۳۲ و ۶۴ و ۱۲۸]، نرخ حذف تصادفی = [۰/۲ و ۰/۲] تعداد دوره‌های آموزش = ۵۰
معمایی	شبکه عصبی پیشخور	لایه‌های پنهان = [۳۲ و ۶۴ و ۱۲۸]، نرخ حذف تصادفی = [۰/۲ و ۰/۱] تعداد دوره‌های آموزش = ۶۰
هیجان‌انگیز	شبکه عصبی پیشخور	لایه‌های پنهان = [۳۲ و ۶۴ و ۱۲۸]، نرخ حذف تصادفی = [۰/۱۵ و ۰/۲۵] تعداد دوره‌های آموزش = ۵۰

نتایج کامل عملکرد مدل‌های مختلف برای هر یک از ۱۸ ژانر در جدول ۴ ارائه شده است. در این جدول، بهترین عملکرد هر ژانر مشخص شده و مبنای انتخاب مدل نهایی قرار گرفته است.

مقایسه عملکرد مدل‌های منتخب برای ژانرهای مختلف

پس از مقایسه مدل‌های مختلف، برای هر ژانر بهترین مدل بر اساس معیارهای ارزیابی انتخاب شد. نتایج نشان می‌دهد که میزان خطای پیش‌بینی در ژانرهای مختلف متفاوت است. این تفاوت بیانگر آن است که علاقه به برخی ژانرها از روی ویژگی‌های فردی کاربران آسان‌تر از ژانرهای دیگر قابل پیش‌بینی است. این تفاوت می‌تواند بیانگر آن باشد که رابطه میان

ویژگی‌های شخصیتی و جمعیت‌شناختی کاربران و ترجیحات آن‌ها برای همه ژانرها یکسان نیست. بر اساس معیار MAE، بهترین عملکرد مربوط به پیش‌بینی علاقه کاربران به ژانر درام است که با خطای ۰/۱۸ به دست آمده است. به این معنا که پیش‌بینی علاقه به درام به‌طور متوسط تنها ۰/۱۸ واحد با مقدار واقعی فاصله داشته است. پس از آن، ژانرهای اکشن و جنایی به‌ترتیب با مقادیر MAE برابر با ۰/۲۱ و ۰/۲۲ قرار دارند.

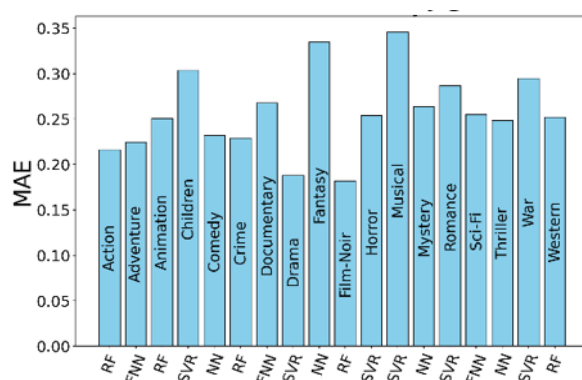
جدول ۴. نتایج ارزیابی عملکرد هر کدام از مدل‌های پیش‌بینی‌کننده میزان علاقه به ژانرها در بازه ۱ تا ۵، بر اساس معیارهای خطا

ژانر	مدل‌های پیش‌بینی‌کننده	معیارهای ارزیابی			
		MSE	MAE	RMSE	R ²
اکشن	رگرسیون بردار پشتیبان	۰/۰۹	۰/۲۳	۰/۳۰	۰/۶۱
	رگرسیون جنگل تصادفی	۰/۰۸	۰/۲۱	۰/۲۹	۰/۶۴
	نزدیک‌ترین همسایه K	۰/۱۱	۰/۲۶	۰/۳۴	۰/۵۲
	گرادیان بوستینگ	۰/۱۰	۰/۲۳	۰/۳۲	۰/۵۶
	شبکه عصبی پیشخور	۰/۱۱	۰/۲۵	۰/۳۴	۰/۵۲
ماجرای جویی	رگرسیون بردار پشتیبان	۰/۰۹	۰/۲۹	۰/۳۱	۰/۵۹
	رگرسیون جنگل تصادفی	۰/۰۹	۰/۲۳	۰/۳۱	۰/۳۱
	نزدیک‌ترین همسایه K	۰/۱۲	۰/۲۶	۰/۳۵	۰/۵۰
	گرادیان بوستینگ	۰/۱۱	۰/۲۴	۰/۳۳	۰/۵۵
	شبکه عصبی پیشخور	۰/۰۸	۰/۲۲	۰/۲۹	۰/۶۴
انیمیشن	رگرسیون بردار پشتیبان	۰/۱۲	۰/۲۶	۰/۳۴	۰/۲۲
	رگرسیون جنگل تصادفی	۰/۱۰	۰/۲۵	۰/۳۲	۰/۳۲
	نزدیک‌ترین همسایه K	۰/۱۳	۰/۲۸	۰/۳۶	۰/۱۵
	گرادیان بوستینگ	۰/۱۱	۰/۲۶	۰/۳۴	۰/۲۳
	شبکه عصبی پیشخور	۰/۱۲	۰/۲۷	۰/۳۴	۰/۲۱
کودکانه	رگرسیون بردار پشتیبان	۰/۱۳	۰/۳۰	۰/۳۷	۰/۲۸
	رگرسیون جنگل تصادفی	۰/۱۴	۰/۳۱	۰/۳۷	۰/۲۶
	نزدیک‌ترین همسایه K	۰/۱۷	۰/۳۴	۰/۴۱	۰/۱۱
	گرادیان بوستینگ	۰/۱۸	۰/۳۴	۰/۴۲	۰/۰۶
	شبکه عصبی پیشخور	۰/۱۴	۰/۳۲	۰/۳۸	۰/۲۴
کمدی	رگرسیون بردار پشتیبان	۰/۱۳	۰/۲۴	۰/۳۶	۰/۶۲
	رگرسیون جنگل تصادفی	۰/۱۴	۰/۲۵	۰/۳۷	۰/۶۰
	نزدیک‌ترین همسایه K	۰/۱۵	۰/۲۶	۰/۳۹	۰/۵۵
	گرادیان بوستینگ	۰/۱۶	۰/۲۶	۰/۴۰	۰/۵۳
	شبکه عصبی پیشخور	۰/۱۲	۰/۲۳	۰/۳۵	۰/۶۴
جنایی	رگرسیون بردار پشتیبان	۰/۰۹	۰/۲۳	۰/۳۰	۰/۶۲
	رگرسیون جنگل تصادفی	۰/۰۹	۰/۲۲	۰/۳۰	۰/۶۱
	نزدیک‌ترین همسایه K	۰/۱۰	۰/۲۵	۰/۳۲	۰/۵۵
	گرادیان بوستینگ	۰/۱۱	۰/۲۴	۰/۳۳	۰/۵۴
	شبکه عصبی پیشخور	۰/۰۹	۰/۲۴	۰/۳۱	۰/۵۹

ژانر	مدل های پیش بینی کننده	معیارهای ارزیابی			
		MSE	MAE	RMSE	R ²
مستند	رگرسیون بردار پشتیبان	۰/۱۱	۰/۲۷	۰/۳۳	۰/۲۲
	رگرسیون جنگل تصادفی	۰/۱۱	۰/۲۷	۰/۳۴	۰/۲۰
	نزدیک ترین همسایه K	۰/۱۱	۰/۲۷	۰/۳۴	۰/۱۹
	گرادیان بوستینگ	۰/۱۴	۰/۲۸	۰/۳۷	۰/۰۵
	شبکه عصبی پیشخور	۰/۱۰	۰/۲۶	۰/۳۲	۰/۲۷
درام	رگرسیون بردار پشتیبان	۰/۰۶	۰/۱۸	۰/۲۵	۰/۶۱
	رگرسیون جنگل تصادفی	۰/۰۶	۰/۱۹	۰/۲۶	۰/۶۰
	نزدیک ترین همسایه K	۰/۰۸	۰/۲۲	۰/۲۹	۰/۵۱
	گرادیان بوستینگ	۰/۰۷	۰/۲۰	۰/۲۷	۰/۵۶
	شبکه عصبی پیشخور	۰/۰۶	۰/۱۹	۰/۲۶	۰/۶۰
فانتزی	رگرسیون بردار پشتیبان	۰/۲۴	۰/۳۴	۰/۴۹	۰/۴۵
	رگرسیون جنگل تصادفی	۰/۲۵	۰/۳۵	۰/۵۰	۰/۴۲
	نزدیک ترین همسایه K	۰/۲۶	۰/۳۶	۰/۵۱	۰/۳۸
	گرادیان بوستینگ	۰/۲۸	۰/۳۷	۰/۵۳	۰/۳۴
	شبکه عصبی پیشخور	۰/۲۲	۰/۳۳	۰/۴۷	۰/۴۹
سیاه	رگرسیون بردار پشتیبان	۰/۰۶	۰/۱۹	۰/۲۴	-۰/۱۵
	رگرسیون جنگل تصادفی	۰/۰۵	۰/۱۸	۰/۲۳	-۰/۰۱
	نزدیک ترین همسایه K	۰/۰۶	۰/۱۹	۰/۲۴	-۰/۱۴
	گرادیان بوستینگ	۰/۰۵	۰/۱۹	۰/۲۴	-۰/۱۱
	شبکه عصبی پیشخور	۰/۰۶	۰/۱۹	۰/۲۴	-۰/۱۷
ترسناک	رگرسیون بردار پشتیبان	۰/۱۱	۰/۲۷	۰/۳۳	۰/۱۸
	رگرسیون جنگل تصادفی	۰/۱۰	۰/۲۵	۰/۳۱	۰/۲۷
	نزدیک ترین همسایه K	۰/۱۳	۰/۲۸	۰/۳۶	۰/۰۷
	گرادیان بوستینگ	۰/۱۲	۰/۲۸	۰/۳۵	۰/۰۹
	شبکه عصبی پیشخور	۰/۱۲	۰/۲۸	۰/۳۳	۰/۱۳
موزیکال	رگرسیون بردار پشتیبان	۰/۱۹	۰/۳۴	۰/۴۴	۰/۳۰
	رگرسیون جنگل تصادفی	۰/۱۹	۰/۳۵	۰/۴۳	۰/۳۲
	نزدیک ترین همسایه K	۰/۲۲	۰/۳۷	۰/۴۷	۰/۲۱
	گرادیان بوستینگ	۰/۲۲	۰/۳۷	۰/۴۷	۰/۲۲
	شبکه عصبی پیشخور	۰/۱۹	۰/۳۵	۰/۴۴	۰/۳۲
معمایی	رگرسیون بردار پشتیبان	۰/۱۳	۰/۲۷	۰/۳۶	۰/۵۲
	رگرسیون جنگل تصادفی	۰/۱۵	۰/۲۷	۰/۳۹	۰/۴۶
	نزدیک ترین همسایه K	۰/۱۶	۰/۳۰	۰/۴۱	۰/۴۱
	گرادیان بوستینگ	۰/۱۷	۰/۲۹	۰/۴۱	۰/۴۰
	شبکه عصبی پیشخور	۰/۱۳	۰/۲۶	۰/۳۶	۰/۵۴

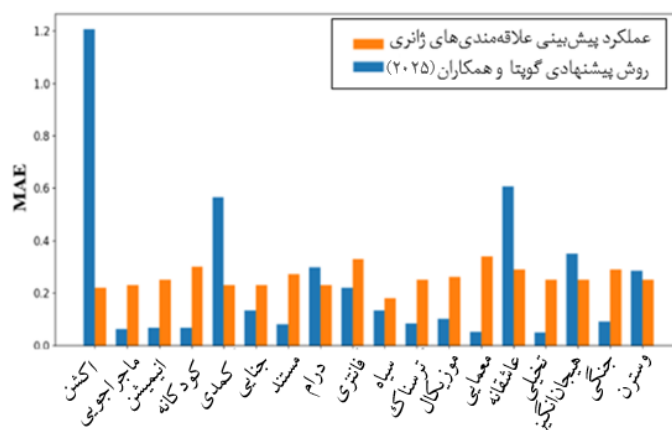
معیارهای ارزیابی				مدل‌های پیش‌بینی‌کننده	ژانر
MSE	MAE	RMSE	R ²		
۰/۱۳	۰/۲۸	۰/۳۷	۰/۳۱	رگرسیون بردار پشتیبان	عاشقانه
۰/۱۴	۰/۲۹	۰/۳۸	۰/۲۸	رگرسیون جنگل تصادفی	
۰/۱۵	۰/۳۰	۰/۳۹	۰/۲۴	نزدیک‌ترین همسایه K	
۰/۱۶	۰/۳۱	۰/۴۰	۰/۱۸	گرادیان بوستینگ	
۰/۱۴	۰/۲۹	۰/۳۷	۰/۲۹	شبکه عصبی پیشخور	
۰/۱۲	۰/۲۶	۰/۳۵	۰/۵۴	رگرسیون بردار پشتیبان	علمی-تخیلی
۰/۱۲	۰/۲۶	۰/۳۵	۰/۵۲	رگرسیون جنگل تصادفی	
۰/۱۳	۰/۲۷	۰/۳۷	۰/۴۹	نزدیک‌ترین همسایه K	
۰/۱۵	۰/۲۹	۰/۳۹	۰/۴۱	گرادیان بوستینگ	
۰/۱۱	۰/۲۵	۰/۳۳	۰/۵۷	شبکه عصبی پیشخور	
۰/۱۲	۰/۲۵	۰/۳۵	۰/۵۵	رگرسیون بردار پشتیبان	هیجان‌انگیز
۰/۱۲	۰/۲۵	۰/۳۵	۰/۵۶	رگرسیون جنگل تصادفی	
۰/۱۴	۰/۲۷	۰/۳۷	۰/۵۱	نزدیک‌ترین همسایه K	
۰/۱۳	۰/۲۷	۰/۳۷	۰/۵۲	گرادیان بوستینگ	
۰/۱۲	۰/۲۴	۰/۳۴	۰/۵۸	شبکه عصبی پیشخور	
۰/۱۶	۰/۲۸	۰/۴۰	۰/۳۸	رگرسیون بردار پشتیبان	جنگی
۰/۱۶	۰/۲۹	۰/۴۰	۰/۳۸	رگرسیون جنگل تصادفی	
۰/۱۷	۰/۳۲	۰/۴۱	۰/۳۳	نزدیک‌ترین همسایه K	
۰/۲۰	۰/۳۲	۰/۴۴	۰/۲۳	گرادیان بوستینگ	
۰/۱۵	۰/۲۹	۰/۳۹	۰/۳۸	شبکه عصبی پیشخور	
۰/۱۱	۰/۲۶	۰/۳۳	۰/۱۵	رگرسیون بردار پشتیبان	وسترن
۰/۱۰	۰/۲۵	۰/۳۱	۰/۲۴	رگرسیون جنگل تصادفی	
۰/۱۲	۰/۲۸	۰/۳۵	۰/۰۹	نزدیک‌ترین همسایه K	
۰/۱۱	۰/۲۶	۰/۳۳	۰/۱۵	گرادیان بوستینگ	
۰/۱۱	۰/۲۷	۰/۳۴	۰/۱۳	شبکه عصبی پیشخور	

این نتایج نشان می‌دهند که علاقه کاربران به این ژانرها با دقت بیشتری بر اساس ویژگی‌های مورد استفاده قابل پیش‌بینی بوده است. در مقابل، بیشترین میزان خطا مربوط به ژانرهای موزیکال، سیاه و کودکانه است. در میان این ژانرها، ژانر موزیکال با MAE برابر با ۰/۴۵ بیشترین خطای پیش‌بینی را دارد. ژانرهای سیاه و کودکانه نیز به ترتیب دارای MAE برابر با ۰/۴۰ و ۰/۳۹ هستند. شکل ۲ مقادیر MAE مدل‌های منتخب برای ۱۸ ژانر را نمایش می‌دهد.



شکل ۲. مقایسه عملکرد بهترین مدل‌های پیش‌بینی‌کننده میزان علاقه‌مندی کاربران به ژانرهای مختلف در بازه ۱ تا ۵

همچنین برای بررسی جایگاه روش پیشنهادی، عملکرد آن با روش ارائه‌شده توسط گوپتا^۱ و همکاران (۲۰۲۵) مقایسه شده است. با وجود تفاوت در رویکردهای مورد استفاده، این مقایسه به دلیل استفاده هر دو پژوهش از مجموعه‌داده MovieLens و ژانرهای یکسان انجام شد و امکان مقایسه نتایج پیش‌بینی را فراهم می‌کند. پژوهش گوپتا و همکاران با استفاده از ویژگی‌های توصیفی فیلم‌ها و مدل‌های مبتنی بر شبکه‌های عصبی، برای ژانرهای مختلف به مدل‌سازی و پیش‌بینی امتیاز فیلم‌ها می‌پردازد، در حالی که پژوهش حاضر از ویژگی‌های شخصیتی و جمعیت‌شناختی کاربران برای پیش‌بینی علاقه‌مندی ژانری استفاده می‌کند. بنابراین، هدف از این مقایسه، مقایسه مستقیم دو روش از نظر رویکرد مدل‌سازی نیست، بلکه بررسی عملکرد پیش‌بینی آن‌ها در شرایطی با مجموعه‌داده و ژانرهای مشابه است. نتایج مقایسه دو روش در شکل ۳ ارائه شده است.



شکل ۳. مقایسه عملکرد روش پژوهش پیش‌رو با عملکرد روش گوپتا و همکاران (۲۰۲۵)

بحث و نتیجه‌گیری

در این پژوهش، نقش ویژگی‌های شخصیتی مبتنی بر مدل پنج‌عاملی و اطلاعات جمعیت‌شناختی در پیش‌بینی ترجیحات کاربران نسبت به ۱۸ ژانر فیلم بررسی شد. نتایج نشان داد ترکیب این ویژگی‌ها می‌تواند اطلاعات ارزشمندی برای پیش‌بینی علایق ژانری کاربران فراهم کند. این یافته با مطالعات پیشین درباره ارتباط ویژگی‌های شخصیتی با ترجیحات رسانه‌ای، از جمله کاظمی‌نیا و همکاران (۲۰۱۹) و گلکار و کاندی (۲۰۱۵)، هم‌راستا است و نشان می‌دهد که ویژگی‌های فردی می‌توانند در کنار رفتارهای گذشته، منبع مکملی برای شناخت ترجیحات کاربران باشند. تفاوت عملکرد مدل‌ها در ژانرهای مختلف نیز نشان می‌دهد که رابطه میان ویژگی‌های فردی و علاقه به فیلم برای همه ژانرها یکسان نیست و استفاده از مدل‌های اختصاصی می‌تواند این تفاوت‌ها را بهتر منعکس کند. از نظر کاربردی، خروجی این مدل‌ها می‌تواند به‌عنوان یک لایه پیش‌بینی ترجیحات

^۱ Gupta

ژانری در سیستم‌های توصیه‌گر مورد استفاده قرار گیرد و در کنار سابقه تماشا و ویژگی‌های محتوایی فیلم‌ها، به شخصی‌سازی توصیه‌ها کمک کند. این رویکرد به‌ویژه برای کاربران جدید یا دارای سابقه رفتاری محدود می‌تواند در کاهش مسئله شروع سرد مؤثر باشد. همچنین، در حساب‌های مشترک، شناخت ترجیحات احتمالی افراد بر اساس ویژگی‌های فردی می‌تواند به تفکیک بهتر علایق کاربران و ارائه توصیه‌های شخصی‌تر کمک کند. با وجود این، وابستگی روش به در دسترس بودن اطلاعات شخصیتی و محدودیت اندازه و تنوع داده‌ها، از مهم‌ترین محدودیت‌های پژوهش حاضر است. برای کاربرد عملی، می‌توان ویژگی‌های شخصیتی را به‌صورت ضمنی و بدون دریافت مستقیم پرسشنامه، همانند رویکردهای مطرح‌شده توسط کاظمی‌نیا و همکاران (۲۰۱۹) و گلکار و کائدی (۲۰۱۵)، تخمین زد. با توجه به محدودیت حجم و تنوع داده‌های مورد استفاده در این پژوهش، یکی از مهم‌ترین مسیرهای پژوهش آینده، افزایش حجم و گسترش تنوع مجموعه داده‌ها از نظر فرهنگی و جغرافیایی و اعتبارسنجی نتایج بر روی داده‌های بزرگ‌تر است. همچنین، استفاده از مجموعه داده‌های بزرگ‌تر و متنوع‌تر از نظر فرهنگی و جغرافیایی، همراه با ادغام ویژگی‌های شخصی و داده‌های رفتاری مانند زمان تماشا، سابقه جست‌وجو و تعاملات اجتماعی و بهره‌گیری از مدل‌های عمیق و روش‌های مبتنی بر توجه، می‌تواند دقت و تعمیم‌پذیری نتایج را افزایش دهد. در مجموع، یافته‌ها نشان می‌دهند که ترکیب ویژگی‌های شخصیتی، جمعیت‌شناختی و رفتاری می‌تواند گامی مؤثر در جهت توسعه سیستم‌های توصیه‌گر با قابلیت شخصی‌سازی بیشتر و دقت بالاتر باشد.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

در انجام این پژوهش، تمامی اصول و موازین اخلاقی حاکم بر پژوهش‌های علمی رعایت شده است.

تعارض منافع

نویسندگان اعلام می‌دارند که این مقاله تعارض منافع ندارد.

حامی مالی

این پژوهش به‌صورت مستقل انجام شده و برای انجام آن هیچ‌گونه حمایت مالی دریافت نشده است.

مشارکت نویسندگان

نویسنده اول: مفهوم‌سازی، آماده‌سازی و پردازش داده‌ها، جمع‌آوری داده‌ها، اجرا، آزمایش‌ها، تحلیل آماری، تفسیر نتایج، و نگارش و اصلاح مقاله. نویسنده دوم: مفهوم‌سازی و طراحی تحقیق، تدوین روش‌شناسی، نظارت بر فرآیند تحقیق، اعتبارسنجی و تفسیر نتایج، راهنمایی علمی، و بررسی انتقادی و اصلاح مقاله.

سپاسگزاری

این مقاله مستخرج از یک پایان‌نامه کارشناسی‌ارشد در دانشگاه اصفهان است. نویسندگان از داوران ناشناس بابت نظرات ارزشمند و پیشنهادات سازنده‌شان سپاسگزاری می‌کنند.

منابع

References

- Buseyne, S., Said-Metwaly, S., Van den Noortgate, W., Depaepe, F., & Raes, A. (2026). The relationship between personality and flow: A meta-analysis. *Journal of Personality*, 94(2), 333–350. <https://doi.org/10.1111/jopy.70004>
- Chen, L., Wu, W., & He, L. (2016). Personality and recommendation diversity. In M. Tkalčič, B. De Carolis, M. de Gemmis, A. Odić, & A. Košir (Eds.), *Emotions and Personality in Personalized Services: Models, Evaluation and Applications* (pp. 201–225). Springer, Cham. https://doi.org/10.1007/978-3-319-31413-6_11

- Gao, Y., Zheng, H., & Cui, H. (2025). User preference modeling for movie recommendations based on deep learning. *Scientific Reports*, 15, Article 16592. <https://doi.org/10.1038/s41598-025-00030-5>
- Golbeck, J., & Norris, E. (2013). Personality, movie preferences, and recommendations. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1414–1415). Niagara Falls, ON, Canada: ACM. <https://doi.org/10.1145/2492517.2492572>
- Golkar, I., & Kaedi, M. (2015). Developing a model for estimating the extraversion degree of social network members using the information extracted from the graph structure. *Journal of Modeling in Engineering*, 13(43), 91–106.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528.
- GroupLens Research. (2017). *MovieLens: Non-commercial, personalized movie recommendations*. Retrieved April 20, 2017, from <https://movielens.org/>
- Gupta, S., Adhikari, U., Varshney, S., & Choudhury, T. (2025). Computational analysis of genre effects on movie ratings using MLP algorithms. *Journal of Computer Science*, 21(4), 905–917. <https://doi.org/10.3844/jcssp.2025.905.917>
- Hall, A. (2005). Audience personality and the selection of media and media genres. *Media Psychology*, 7(4), 377–398.
- Harper, F. M., & Konstan, J. A. (2017). *MovieLens 100K Dataset* [Data set]. GroupLens Research, University of Minnesota. Retrieved April 20, 2017, from <https://grouplens.org/datasets/movielens/100k/>
- Hwang, T.-G., Park, C.-S., Hong, J.-H., & Kim, S. K. (2016). An algorithm for movie classification and recommendation using genre correlation. *Multimedia Tools and Applications*, 75, 12843–12858.
- Karumur, R. P., Nguyen, T. T., & Konstan, J. A. (2016). Exploring the value of personality in predicting rating behaviors: A study of category preferences on MovieLens. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys)* (pp. 139–142). Boston, MA, USA: ACM. <https://doi.org/10.1145/2959100.2959140>
- Kazemini, M., Kaedi, M., & Ganji, B. (2019). Personality-based personalization of online store features using genetic programming: Analysis and experiment. *Journal of Theoretical and Applied Electronic Commerce Research*, 14(1), 16–29.
- Khan, E. M., Mukta, M. S. H., Ali, M. E., & Mahmud, J. (2020). Predicting users' movie preference and rating behavior from personality and values. *ACM Transactions on Interactive Intelligent Systems*, 10(3), Article 18.
- Kim, J., Kim, J., & Choi, J. (2021). Sequential movie genre prediction using average transition probability with clustering. *Applied Sciences*, 11(24), Article 11841.
- Mlika, F., & Karoui, W. (2020). Proposed model to intelligent recommendation system based on Markov chains and grouping of genres. *Procedia Computer Science*, 176, 868–877.
- Mondal, P., Kapoor, P., Singh, S., Saha, S., Singh, J. P., & Singh, A. K. (2023). Genre effect toward developing a multi-modal movie recommendation system in Indian setting. *IEEE Transactions on Consumer Electronics*, 70(1), 2517–2526.
- Nalmpantis, O., & Tjortjis, C. (2017). The 50/50 recommender: A method incorporating personality into movie recommender systems. In *Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings* (Communications in Computer and Information Science, Vol. 744, pp. 498–507). Springer, Cham. https://doi.org/10.1007/978-3-319-65172-9_42
- Nave, G., Rentfrow, J., & Bhatia, S. (2020). *We are what we watch: Movie plots predict the personalities of their fans*. PsyArXiv. <https://doi.org/10.31234/osf.io/wsdu8>
- Negaraesh, F., Kaedi, M., & Zojaji, Z. (2023). Gender identification of mobile phone users based on internet usage pattern. *International Journal of Engineering*, 36(2), 335–347. <https://doi.org/10.5829/ije.2023.36.02b.14>

- Nguyen, T. T., Maxwell Harper, F., Terveen, L., & Konstan, J. A. (2018). User personality and user satisfaction with recommender systems. *Information Systems Frontiers*, 20(6), 1173–1189.
- Pal, A., Barigidad, A., & Mustafi, A. (2020). Identifying movie genre compositions using neural networks and introducing GenRec—a recommender system based on audience genre perception. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)* (pp. 1–7). Patna, India: IEEE. <https://doi.org/10.1109/ICCCS49678.2020.9276893>
- Raj, S., Sharma, A., Saha, S., Singh, B., & Pedanekar, N. (2024). Transformative movie discovery: Large language models for recommendation and genre prediction. *IEEE Access*, 12, 186626–186638.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982–3992). Hong Kong, China: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1410>
- Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender systems: Techniques, applications, and challenges. In *Recommender Systems Handbook* (2nd ed., pp. 1–35). Springer, New York, NY.
- SadighZadeh, S., & Kaedi, M. (2022). Modeling user preferences in online stores based on user mouse behavior on page elements. *Journal of Systems and Information Technology*, 24(2), 112–130.
- Salehi, H., Kaedi, M., & Sattari, M. (2025). Estimating the self-esteem of social network users based on their account information. *Soft Computing Journal*, 14(1), 122–133. <https://doi.org/10.22052/scj.2024.254082.1210>
- Seo, Y. D., Kim, Y. G., Lee, E., & Kim, H. (2021). Group recommender system based on genre preference focusing on reducing the clustering cost. *Expert Systems with Applications*, 183, Article 115396.
- Shafiloo, R., Kaedi, M., & Pourmiri, A. (2024). Considering user dynamic preferences for mitigating negative effects of long-tail in recommender systems. *Information Sciences*, 669, Article 120558. <https://doi.org/10.1016/j.ins.2024.120558>
- Singh, A. (2021). *IMDb Movies Dataset* [Data set]. Kaggle. Retrieved April 20, 2024, from <https://www.kaggle.com/datasets/ashpalsingh1525/imdb-movies-dataset>
- Wang, H. (2018). Utilizing imbalanced data and classification cost matrix to predict movie preferences. *International Journal of Artificial Intelligence and Applications*, 9(6), 1–12. <https://doi.org/10.5121/ijai.2018.9601>
- Wang, Z., Yang, Y., He, L., & Gu, J. (2014). User identification within a shared account: Improving IP-TV recommender performance. In *Advances in Databases and Information Systems: 18th East European Conference* (Lecture Notes in Computer Science, Vol. 8716, pp. 219–233). Ohrid, Macedonia: Springer.
- Widiyaningtyas, T., Hidayah, I., & Adji, T. B. (2021). User profile correlation-based similarity (UPCSim) algorithm in movie recommendation system. *Journal of Big Data*, 8(1), Article 52.
- Wu, W., & Chen, L. (2015). Implicit acquisition of user personality for augmenting movie recommendations. In *Proceedings of the 23rd International Conference on User Modeling, Adaptation and Personalization (UMAP)* (Vol. 9146, pp. 302–314). Dublin, Ireland: Springer.