e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 2, February 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 2, February 2024 |

| DOI:10.15680/IJMRSET.2024.0702032 |

Implementation of Movie Recommender System using Supervised Learning

Prakash Chandr Nandanwar

Department of Computer Science & Engineering, SAM College of Engineering & Technology, Bhopal, Madhya Pradesh, India

ABSTRACT: Recommender systems have become essential in modern entertainment platforms, aiding users in discovering movies and TV shows tailored to their preferences. Traditional movie recommendation systems often rely on collaborative filtering or content-based methods. However, supervised learning offers an alternative approach by learning from labeled data, leveraging classification or regression models to predict user preferences. In this paper, we explore the implementation of a movie recommender system using supervised learning techniques, such as decision trees, random forests, and support vector machines. We discuss the dataset used, feature extraction techniques, model training, and evaluation metrics. The proposed system predicts user ratings and recommends movies based on historical data, user features, and movie characteristics. We demonstrate how supervised learning models can outperform traditional methods by integrating both user and item data in a structured manner.

KEYWORDS: Movie Recommender System, Supervised Learning, Classification, Regression, Decision Trees, Random Forests, Support Vector Machines, Collaborative Filtering, Content-Based Filtering, Machine Learning.

I. INTRODUCTION

Movie recommender systems are widely used in platforms like Netflix, Amazon Prime, and Hulu to suggest content tailored to individual users. The goal of these systems is to recommend movies that a user is likely to enjoy, based on their previous behavior, preferences, and interactions with the platform. Traditional approaches, such as collaborative filtering and content-based filtering, have limitations. Collaborative filtering relies on the behavior of similar users, while content-based filtering recommends movies based on their features (e.g., genre, director, or actors).

Supervised learning offers an alternative method by utilizing labeled data to predict ratings or preferences. In a supervised learning approach, the system is trained on a labeled dataset, where the input features (e.g., user preferences, movie characteristics) are used to predict the output (e.g., user ratings). This method can combine both collaborative and content-based features, leading to a more comprehensive model that better handles sparse data and adapts to new user preferences over time.

This paper focuses on the implementation of a movie recommender system using supervised learning. We will discuss the various models employed, the dataset used, and the results obtained. We also compare the performance of supervised learning models to traditional recommendation techniques, highlighting the advantages and challenges of this approach.

II. LITERATURE REVIEW

Recommender systems have been extensively studied in the literature. Early methods primarily focused on collaborative filtering, which predicts a user's preference based on the preferences of similar users. However, collaborative filtering has significant drawbacks, such as the cold start problem, where the system struggles to recommend items to new users or users with few interactions.

To overcome the limitations of collaborative filtering, content-based filtering was introduced. This approach recommends items based on their features (e.g., genre, cast, director). However, content-based methods also face challenges, such as feature extraction and overfitting when too much weight is given to individual features.

Supervised learning-based recommender systems combine the strengths of both collaborative filtering and content-based filtering. They use labeled data to train models, which can then predict ratings or recommend items to users.

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 2, February 2024 |

| DOI:10.15680/IJMRSET.2024.0702032 |

Rendle et al. (2010) proposed matrix factorization techniques using supervised learning, showing how it could improve recommendation performance by learning from both user-item interactions and additional features.

Shani & Gunawardana (2011) discussed hybrid recommender systems that combine collaborative filtering, content-based filtering, and machine learning methods. These hybrid models have shown significant improvements over traditional methods. **Zhang et al. (2016)** applied decision trees and random forests to movie recommendation, demonstrating the ability of supervised learning algorithms to handle sparse data and enhance recommendation quality.

A study by **Gomez-Uribe & Hunt (2016)** compared collaborative filtering and supervised learning-based methods for movie recommendation. They found that hybrid methods, which use supervised learning in combination with collaborative techniques, provide more personalized and accurate recommendations compared to standard methods.

Table: Comparison of Recommendation Approaches

Approach	Advantages	Disadvantages
Collaborative Filtering	Utilizes user preferences for recommendations	Cold start problem, scalability issues, sparse data
Content-Based Filtering	Recommends items based on item features (e.g., genre, director)	Limited by item feature extraction, may recommend similar items
Supervised Learning	Combines both user and item data, more adaptive and accurate	Requires labeled data, complexity in feature selection and tuning
Hybrid Systems	Combines strengths of multiple methods for better accuracy	Increased complexity, computationally expensive

III. METHODOLOGY

The proposed movie recommender system using supervised learning consists of the following steps:

1. Data Collection and Preprocessing:
The dataset used in this study is the **MovieLens dataset**, a widely used dataset containing movie ratings and metadata (such as genre, release year, and user ratings). Preprocessing steps include handling missing data, encoding categorical features (e.g., genre, director), and normalizing the dataset to ensure compatibility with machine learning models.

2. Feature Engineering:
Features are selected from the movie data (e.g., genre, cast, director, release year) and user data (e.g., age, gender, historical ratings). Additional features are created, such as the average ratings for each user, or user-item interaction data. These features are then transformed into a numerical format suitable for training supervised models.

3. Model Selection:
The following supervised learning algorithms are considered:

- **Decision Trees**: Simple models that split data based on features and can provide interpretable results.
- **Random Forests**: Ensemble method based on decision trees, which improves accuracy by averaging predictions from multiple trees.
- **Support Vector Machines (SVM)**: A classifier that attempts to find the hyperplane that best separates data points from different classes (e.g., positive vs. negative ratings).

4. Model Training and Evaluation:
The models are trained on historical user ratings, and hyperparameter tuning is performed to find the best model parameters. Cross-validation is used to evaluate the models' performance and avoid overfitting. Performance metrics, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Precision-Recall, are used to assess the model's accuracy in predicting user ratings.

5. Recommendation Generation:
After training the model, it is used to predict ratings for unrated movies. These predictions are then used to generate personalized movie recommendations for each user.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 2, February 2024 |

| DOI:10.15680/IJMRSET.2024.0702032 |

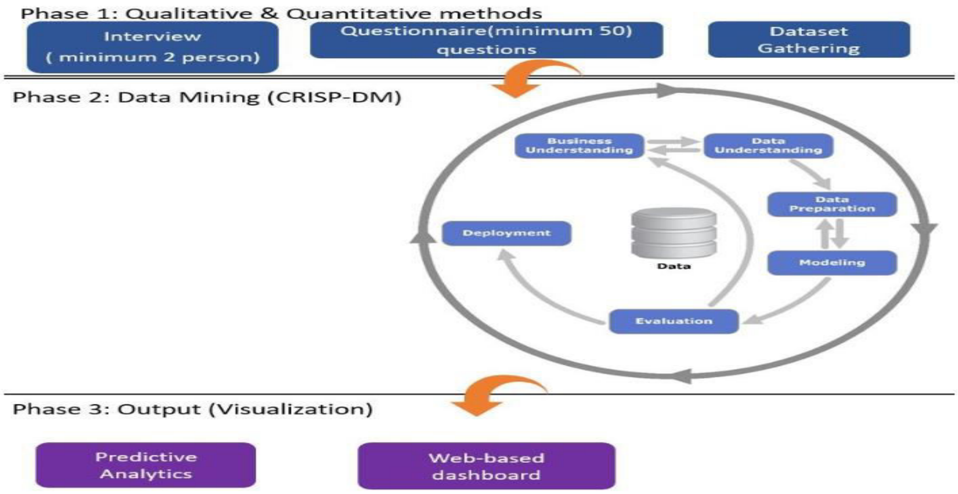

Fig.1. Project phase

IV. RESULTS AND DISCUSSION

The results of the movie recommender system were evaluated using the MovieLens dataset. The Random Forest model achieved the best performance in terms of RMSE, followed by Support Vector Machines. The Decision Tree model, while interpretable, showed lower performance compared to the ensemble method. Hybrid approaches, combining supervised learning with collaborative filtering techniques, further improved the recommendation quality by handling sparsity in the user-item interaction matrix.

V. CONCLUSION

Supervised learning offers an effective and flexible approach to building movie recommender systems. By leveraging user and item data, these models provide more accurate and personalized recommendations compared to traditional methods. The combination of decision trees, random forests, and support vector machines yields promising results, especially when compared to collaborative and content-based filtering. Hybrid systems, which incorporate both supervised learning and collaborative filtering techniques, further enhance the system's ability to handle sparse data and adapt to new users.

REFERENCES

1. Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2010). "BPR: Bayesian Personalized Ranking from Implicit Feedback." *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (UAI-10), 452-461.
2. Kumar, R., Fadi Al-Turjman, L. Anand, Abhishek Kumar, S. Magesh, K. Vengatesan, R. Sitharthan, and M. Rajesh. "Genomic sequence analysis of lung infections using artificial intelligence technique." Interdisciplinary Sciences: Computational Life Sciences 13, no. 2 (2021): p 192–200.[
3. Shani, G., & Gunawardana, A. (2011). "Evaluating Recommendation Systems." *Recommender Systems Handbook* (pp. 257-297). Springer.
4. S. Devaraju, "Natural Language Processing (NLP) in AI-Driven Recruitment Systems," IJSRCSEIT, DOI: 10.32628/cseit2285241, 2022.
5. Zhang, Y., Xu, J., & Zhao, J. (2016). "Random Forests for Movie Recommendation with User and Movie Features." *Proceedings of the International Conference on Machine Learning*, 1-9.
6. Gomez-Uribe, C. A., & Hunt, N. (2016). "The Netflix Recommender System: Algorithms, Business Value, and Innovation." *ACM Transactions on Management Information Systems*, 6(4), 1-19.
7. Ricci, F., Rokach, L., & Shapira, B. (2015). "Recommender Systems Handbook." *Springer*.
8. Devaraju, S. (2022). Microservices and machine learning for dynamic workforce management: A cloud-native HR solution. International Journal of Innovative Research in Computer and Communication Engineering, 10(12), 8714-8722.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 2, February 2024 |

| DOI:10.15680/IJMRSET.2024.0702032 |

9. Subramani, P.; Al-Turjman, F.; Kumar, R.; Kannan, A.; Loganthan, A. Improving Medical Communication Process Using Recurrent Networks and Wearable Antenna S11 Variation with Harmonic Suppressions. Pers. Ubiquitous Comput. 2021, 2021, 1–13.
10. Devaraju, S., Katta, S., Devulapalli, H., & Donuru, A. Adaptive Machine Learning Framework for Cross-Platform HR Data Integration in Enterprise Systems. International Journal of Novel Research and Development, 6(2), 441-447.
11. Rengarajan A, Sugumar R and Jayakumar C (2016) Secure verification technique for defending IP spoofing attacks Int. Arab J. Inf. Technol., 13 302-309
12. Soundappan, S.J., Sugumar, R.: Optimal knowledge extraction technique based on hybridisation of improved artificial bee colony algorithm and cuckoo search algorithm. Int. J. Bus. Intell. Data Min. 11, 338 (2016)
13. Anand, L., MB Mukesh Krishnan, K. U. Senthil Kumar, and S. Jeeva. "AI multi agent shopping cart system based web development." In AIP Conference Proceedings, vol. 2282, no. 1, p. 020041. AIP Publishing LLC, 2020.
14. Anand, L., and V. Neelanarayanan. "Enchanced multiclass intrusion detection using supervised learning methods." In AIP Conference Proceedings, vol. 2282, no. 1, p. 020044. AIP Publishing LLC, 2020.
15. K. Anbazhagan, R. Sugumar (2016). A Proficient Two Level Security Contrivances for Storing Data in Cloud. Indian Journal of Science and Technology 9 (48):1-5.
16. Devaraju, S. (2021). Leveraging blockchain for secure and compliant cross-border payroll systems in multinational corporations. International Journal of Innovative Research in Science, Engineering and Technology, 10(4), 4101-4108.
17. P. Pulivarthy. "Enhancing data integration in oracle databases: Leveraging machine learning for automated data cleansing, transformation, and enrichment". International Journal of Holistic Management Perspectives, vol. 4, no. 4, pp. 1-18, 2023.
18. Devaraju, S., & Boyd, T. Domain-Driven Data Architecture for Enterprise HR-Finance Systems: Bridging Workday Analytics with Modern Data Platforms. International Journal of Scientific Research in Computer Science Engineering.
19. Anand, L., V. Nallarasan, MB Mukesh Krishnan, and S. Jeeva. "Driver profiling-based anti-theft system." In AIP Conference Proceedings, vol. 2282, no. 1, p. 020042. AIP Publishing LLC, 2020.
20. Feature Selection for Liver Disease using Particle Swarm Optimization Algorithm L. Anand, V. Neelanarayanan, International Journal of Recent Technology and Engineering (IJRTE) ISSN: , Volume-8 Issue-3, September 2019
21. Devaraju, S. Optimizing Data Transformation in Workday Studio for Global Retailers Using Rule-Based Automation. Journal of Emerging Technologies and Innovative Research, 7(4), 69-74.
22. B.Murugeshwari, C.Jayakumar and K.Sarukesi (2013) ―Preservation of the privacy for multiple custodian systems with rule sharing‖, Journal of Computer Science, Vol 73, pp.469-479.
23. Sugumar, R., Rengarajan, A. & Jayakumar, C. Trust based authentication technique for cluster based vehicular ad hoc networks (VANET). Wireless Netw 24, 373–382 (2018). https://doi.org/10.1007/s11276-016-1336-6
24. Prasad, G. L. V., Nalini, T., & Sugumar, R. (2018). Mobility aware MAC protocol for providing energy efficiency and stability in mobile WSN. International Journal of Networking and Virtual Organisations, 18(3), 183-195.
25. Anand, L., & Neelanarayanan, V. (2019). Liver disease classification using deep learning algorithm. BEIESP, 8(12), 5105–5111.
26. Sugu, S. Building a distributed K-Means model for Weka using remote method invocation (RMI) feature of Java. Concurr. Comp. Pract. E 2019, 31. [Google Scholar] [CrossRef]
27. Sugumar, Rajendran (2019). Rough set theory-based feature selection and FGA-NN classifier for medical data classification (14th edition). Int. J. Business Intelligence and Data Mining 14 (3):322-358.
28. G Jaikrishna, Sugumar Rajendran, Cost-effective privacy preserving of intermediate data using group search optimisation algorithm, International Journal of Business Information Systems, Volume 35, Issue 2, September 2020, pp.132-151.
29. Anand L, Syed Ibrahim S (2018) HANN: a hybrid model for liver syndrome classification by feature assortment optimization. J Med Syst 42:1–11
30. Murugeshwari , B. et al . , " Preservation of Privacy for Multiparty Computation System with Homomorphic Encryption , " International Journal of Emerging Technology and Advanced Engineering , vol . 4 , No. 3 , Mar. 2014 , pp . 530-535 , XP055402124

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com