iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.54258

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Moviebox: A Movie Recommendation System

Manan Mahajan[1], Abhedya Mishra[2], Praveen Kumar[3], Ms. Sapna Gupta[4]
[1, 2, 3]*Student, ECE Department,* [4]*Asst. Professor, Maharaja Agrasen Institute of Technology, New Delhi, India*

Abstract: ***In the vast world of cinema, numerous critically acclaimed movies often go unnoticed, despite their significant artistic and cultural value. Recognizing this disparity, our research endeavors to bridge the gap by developing a comprehensive movie recommendation system that highlights these "Out of the box" films. In the initial phase, we meticulously collected a bespoke dataset by scraping data from IMDb, encompassing a wide range of movies from various genres and regions. In the subsequent phase, we constructed an advanced algorithm utilizing content-based filtering techniques. This algorithm analyzes both user behavior and movie features to provide personalized recommendations that align with users' unique preferences. By embracing this approach, our research aims to enhance the discoverability and appreciation of lesser-known yet meaningful movies, empowering users to explore a diverse array of cinematic experiences.***
Keywords: Movie recommendation system, research, IMDb

I. INTRODUCTION

The movie industry has grown tremendously in recent years with the introduction of various genres and film styles and viewership options. With such a vast amount of content, it has become increasingly challenging for viewers to find relevant and high-quality movies to watch. To this day, the caste and crew of OTT releases and critically acclaimed movies remain obscure to many. In response, movie recommendation systems have been developed to assist viewers in discovering movies that match their interests.

This research paper focuses on the design and implementation of a movie recommendation system. The model utilizes content-based filtering techniques to provide personalized movie recommendations to users. The algorithm recommends movies based on the similarity of their features, such as genre, director, and actors using cosine similarity.

The system has been developed using Python programming language and its various libraries such as pandas, NumPy, and scikit-learn. The dataset used for the system has been collated from IMDb's website using Web Scraping. It contains more than 500 movies from all across the globe and 15 relevant parameters to facilitate the content-based filtering approach.

The system was scrutinized on the basis of prevalent accuracy parameters for such models, to check its effectiveness and veracity all of which demonstrate that the proposed system provides highly accurate and relevant recommendations for a wide range of users.

Overall, the research highlights the potential of movie recommendation systems to assist viewers in finding movies that suit their interests, while also contributing to the development of the field of recommendation systems.

II. LITERATURE SURVEY

The research landscape surrounding movie recommendation systems has witnessed significant advancements, driven by the goal of promoting the discovery of meaningful and lesser-known films. This section provides a concise overview of the existing literature, highlighting key studies and methodologies employed in the field.

Researchers have acknowledged the prevalence of critically acclaimed movies that often remain unnoticed by audiences. Addressing this issue, our research aims to develop a comprehensive movie recommendation system that brings these "Out of the box" films to the forefront. To achieve this objective, our study draws upon content-based filtering techniques, which have proven effective in personalized recommendation systems. By analyzing user behavior and movie features, the algorithm generates tailored recommendations aligned with individual preferences. This approach allows users to explore a diverse range of cinematic experiences and discover movies that resonate with their unique tastes.

Prior literature has emphasized the importance of data collection and dataset construction. In our research, a bespoke dataset was curated by scraping data from IMDb, encompassing a wide range of movies across different genres and regions. This dataset forms the foundation for training and testing our advanced recommendation algorithm.

Overall, our research builds upon existing literature in the field of movie recommendation systems, aiming to address the underrepresentation of meaningful movies and provide users with an enhanced viewing experience. By leveraging content-based filtering techniques and a meticulously curated dataset, our study contributes to the broader research discourse and provides valuable insights for future advancements in the domain.

In conclusion, research on movie recommendation systems has explored techniques such as collaborative filtering, content-based filtering, matrix factorization, factorization machines, and deep learning. Ongoing research aims to refine and combine these approaches to overcome challenges like cold-start problems and data sparsity. Future advancements will enhance the precision and personalization of movie recommendations, improving the overall movie-watching experience.

III. METHODOLOGY

The initial phase of our research involved careful selection and identification of the parameters to be incorporated into our movie recommendation algorithm. This critical decision was reached through extensive discussions and thorough examination of the available parameters provided by the IMDb website. The chosen parameters were deemed essential in capturing relevant aspects of movies that would contribute to the accuracy and effectiveness of our recommendation system. This meticulous process ensured that our research was built upon a solid foundation and enabled us to create a robust algorithm capable of delivering valuable and tailored movie recommendations to users.

Next, we conducted web scraping to collect comprehensive data on approximately 600-700 films from the IMDb website. This meticulously curated dataset served as the foundation for training and testing our machine learning algorithm. By utilizing this rich and diverse collection of movie information, we ensured the effectiveness and reliability of our algorithm in generating accurate recommendations. The dataset acquisition process involved adhering to ethical guidelines and terms of service provided by the IMDb website, ensuring the integrity and legality of our research. The robustness and breadth of our dataset played a crucial role in enabling our algorithm to learn patterns and correlations, ultimately enhancing the quality and precision of our movie recommendations.

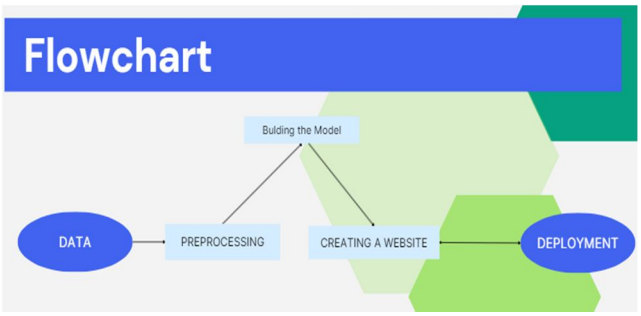


Fig 1. WorkFlow

After Scrapping we decided how we will process the data

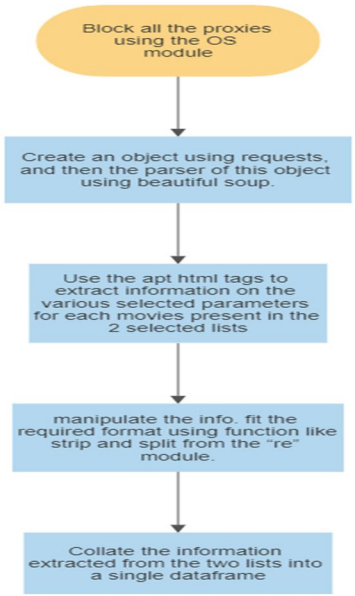


Fig2. Scrapping of Data

1) Conducted an in-depth study of the IMDb website to identify the attributes required for the data frame. Analyze the HTML code structure using Google Developer Tools to understand how these attributes are represented on the website.
2) Implement a proactive security measure by utilizing the OS module to block all proxies. This prevents unauthorized access and strengthens network security by restricting the use of proxy servers.
3) Utilize the requests library to create an object for accessing web content. Use Beautiful Soup, a popular parsing library, to parse and extract relevant information from the created object. This enables efficient data extraction and manipulation for further analysis and processing.
4) Use the apt html tags to extract information on the various selected parameters for each movie present in the 2 selected lists and manipulate it fit the required format using function like strip and split from the “re” module.
5) Collate the information extracted from the two lists into a single dataframe.

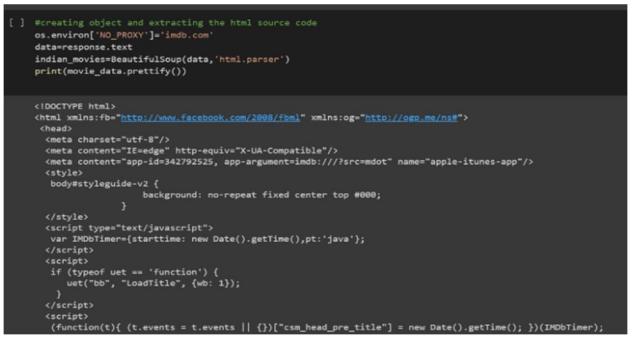


fig. 3. Web Scrapping

6) The ultimate objective is to develop a robust machine learning model using the collected data. This model will undergo training using a carefully selected subset of the data to effectively predict a specific outcome or make meaningful inferences. By leveraging advanced algorithms and techniques, the model will learn from patterns and relationships within the data, enabling it to make accurate predictions or draw insightful conclusions. This process involves optimizing the model's parameters and fine-tuning its performance to ensure reliable and high-quality results. The trained machine learning model will serve as a powerful tool to uncover valuable insights, drive informed decision-making, and unlock new opportunities in the given domain.

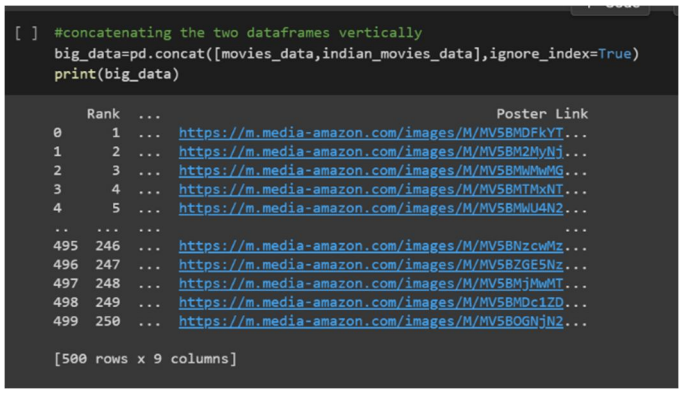


Fig 4. Concat two dataframes

7) Cosine similarity is utilized in our recommendation system to compute the similarity between two movies. It measures the cosine of the angle between vectors representing the movies in a multi-dimensional space. The resulting score falls between 0 and 1, where higher values indicate a stronger similarity.
8) Cosine similarity considers both direction and magnitude, making it suitable for comparing movies with varying attributes such as genres, budgets, or cast sizes. By leveraging cosine similarity, our recommendation system provides personalized recommendations based on users' preferences and interests. This approach enhances the overall viewing experience by introducing users to a diverse range of movies that share similar characteristics.

9) Once a machine learning model is constructed, trained, and evaluated, it empowers the generation of predictions on new data, such as the anticipation of movie success before its official release. By incorporating pertinent data points such as genre, cast, director, budget, marketing strategies, and historical performance, the model produces valuable predictions. This valuable predictive capability enables stakeholders to make well-informed decisions, harnessing the potential of the model's predictions to optimize outcomes in the dynamic landscape of the film industry.

10) To enhance the accessibility and reach of the machine learning model, we have developed a dedicated website. This website serves as a platform to showcase the capabilities of the model and empowers users to leverage its predictive abilities. Through the website, users can conveniently input their own data, such as movie attributes and metrics, to receive personalized predictions and insights. This user-friendly interface expands the model's usage beyond industry experts, allowing filmmakers, enthusiasts, and industry professionals to make data-driven decisions and gain valuable insights about the potential success of their movies.

11) Once the website is built, our next step is to deploy it on a server, ensuring its public accessibility. To achieve this, we will utilize the services of a cloud service provider. Cloud platforms offer a reliable and scalable infrastructure that allows us to host the website securely and efficiently. By leveraging the resources provided by the cloud service provider, we can ensure high availability and optimal performance for users accessing the website. This deployment strategy enables seamless access to the predictive capabilities of the machine learning model from anywhere in the world, fostering widespread adoption and facilitating data-driven decision-making in the film industry.

IV. PROJECT OUTCOMES

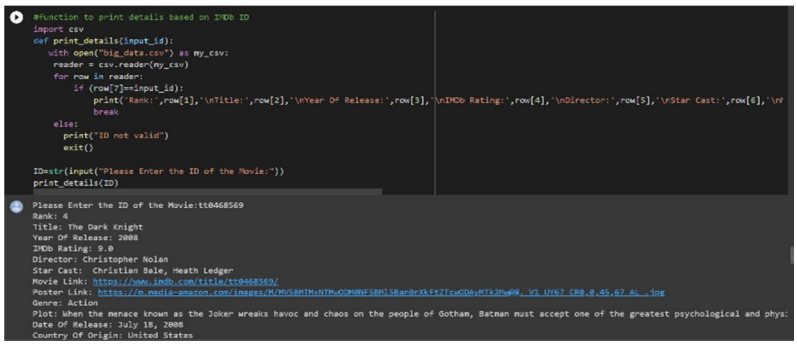


fig. 6. Output screen for Web Scrapping

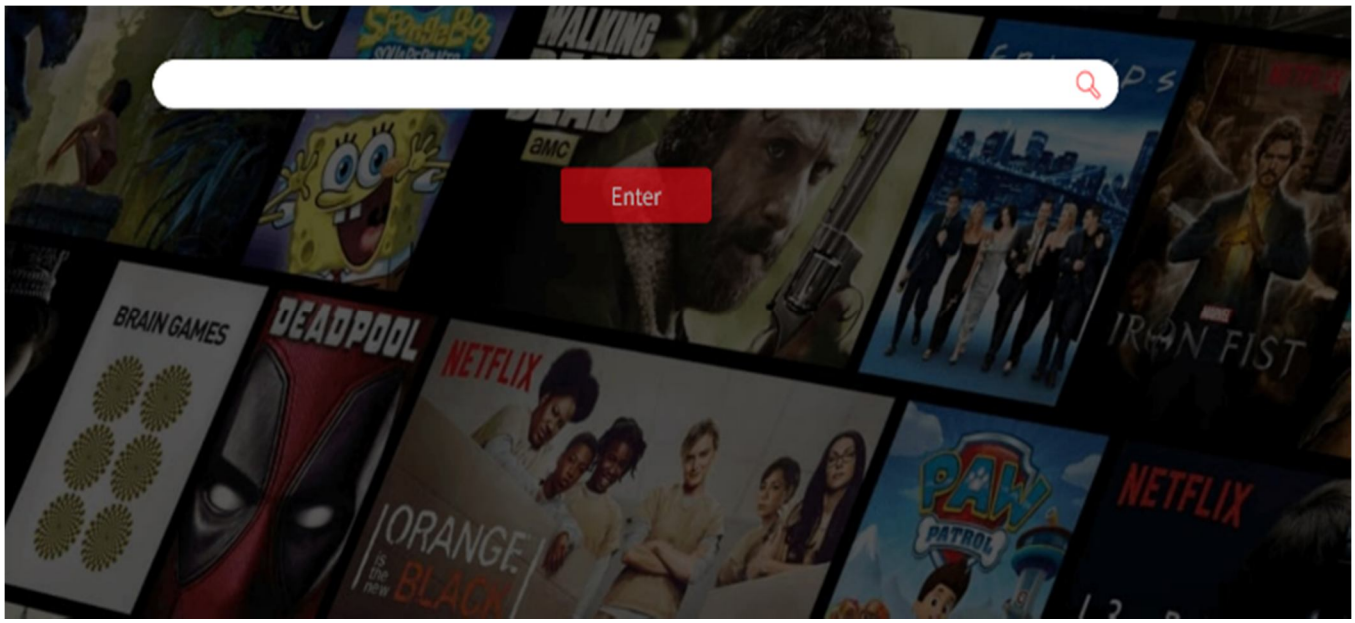


fig. 7. Output screen for User

V. CONCLUSION

The predictive power of the model can help filmmakers, production studios, and distributors make informed decisions and formulate effective strategies. For instance, it can assist in determining the optimal release date, identifying the target audience, tailoring marketing campaigns, and even influencing production choices during the filmmaking process. This enables stakeholders to allocate resources wisely and mitigate risks in a highly competitive and financially demanding industry. By utilizing the recommendation system, filmmakers can increase the visibility of their movies and expand their reach beyond the traditional multi-starrer box office productions.This is particularly significant for the countless artists and individuals working behind the scenes who often remain unnoticed and their careers go unnoticed. The recommendation system provides an opportunity to level the playing field and reduce the opportunity cost associated with movies that may not have received widespread attention. It opens doors for the appreciation of diverse forms of art and gives a chance for lesser-known movies to gain recognition and viewership.

Ultimately, this system empowers individuals to make more informed choices, allowing them to explore and appreciate a wide range of movies and art forms. By reducing the reliance on star power and box office success, the system encourages a more inclusive and diverse entertainment industry. It brings attention to the hidden gems and provides a platform for artists and movies that may have otherwise remained inconspicuous. In doing so, it adds value to the livelihoods of millions of people in the entertainment industry by expanding opportunities and fostering a richer cultural landscape globally.

VI. FUTURE SCOPE

The future scope of the movie recommendation system lies in its potential to revolutionize the film industry by providing valuable insights and predictions for stakeholders.

Further advancements in machine learning and data analytics can enhance the accuracy and precision of the recommendation system, enabling more effective decision-making for filmmakers, production studios, and distributors.

Incorporating user feedback and preferences can lead to personalized recommendations, catering to individual tastes and expanding the reach of lesser-known movies. Continuous research and development can address challenges such as data sparsity, cold-start problems, and improving the scalability of the recommendation system.

Collaboration with industry experts and academics can contribute to the refinement and validation of the recommendation system, ensuring its effectiveness and reliability. Exploring partnerships with streaming platforms and movie databases can expand the availability and accessibility of the recommendation system, reaching a wider audience and maximizing its impact.

The future scope also includes exploring the application of the recommendation system in other areas of the entertainment industry, such as TV shows, music to provide comprehensive and personalized content recommendations.

REFERENCES

[1] J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, “A content-based recommendation algorithm for learning resources,” Multimedia Systems, vol. 24, no. 2, pp. 163–173, Mar. 2017, Doi: https://doi.org/10.1007/s00530-017-0539-8.

[2] Kumar, Manoj & Yadav, Dharmendra & Singh, Ankur & Kr, Vijay. (2015). A Movie Recommender System: MOVREC. International Journal of Computer Applications. 124. 7-11. 10.5120/ijca2015904111.

[3] Rakshitha. P, “Recommendation of Movies based on Collaborative Filtering using Apache Spark,” International Journal of Engineering Research and, vol. V9, no. 08, Aug. 2020, doi: https://doi.org/10.17577/ijertv9is080123.

[4] S. Seyednezhad, K. Cozart, J. Bowllan, and A. Smith, “A Review on Recommendation Systems: Context-aware to Social-based,” 2018. Available: https://arxiv.org/pdf/1811.11866.pdf

[5] J. Zhang, Y. Wang, Z. Yuan, and Q. Jin, “Personalized real-time movie recommendation system: Practical prototype and evaluation,” Tsinghua Science and Technology, 2020. https://www.semanticscholar.org/paper/Personalized-real-time-movie-recommendation-system%3A-Zhang-Wang/e5e6abe21daf2385fd09d4c1b55e372a68f1ce19 (accessed Jun. 18, 2023)

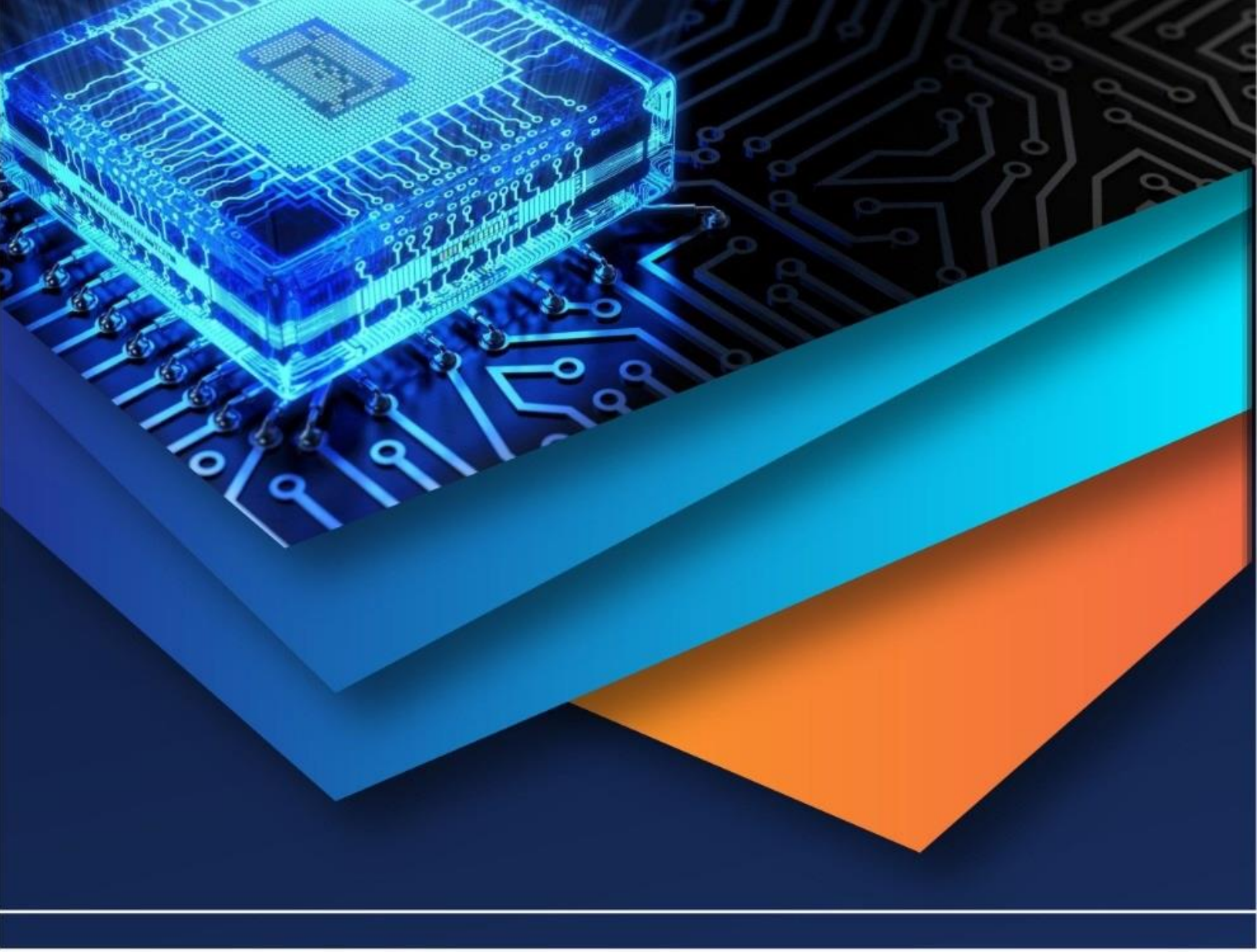
[6] V. Singrodia, A. Mitra, and S. Paul, “A Review on Web Scrapping and its Applications,” IEEE Xplore, Jan. 01, 2019. https://ieeexplore.ieee.org/abstract/document/8821809

[7] “IMPLEMENTATION OF WEB SCRAPING FOR E-COMMERCE WEBSITE,” www.jetir.org. https://www.jetir.org/view?paper=JETIR2106682 (accessed Jun. 18, 2023).

[8] S. Parvattikar and Dr. D. Parasar, “Recommendation System Using Machine Learning,” SSRN Electronic Journal, 2020, doi: https://doi.org/10.2139/ssrn.3702439

[9] A. Fanca, A. Puscasiu, D.-I. Gota, and H. Valean, “Recommendation Systems with Machine Learning,” IEEE Xplore, Oct. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9257290?casa_token=n_cMQ_ihTPAAAAAA:s4jAOFhOeirqFe5ST_UT3bCUHN5rXEIl9Bkiq0-EmgA91J37lhshzE0vLP42pFoD9jrZD_H-zU-rXA (accessed Oct. 28, 2022).

[10] F. Furtado and A. Singh, “Movie recommendation system using machine learning,” International Journal of Research in Industrial Engineering, vol. 9, no. 1, pp. 84–98, Mar. 2020, doi: https://doi.org/10.22105/riej.2020.226178.1128.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY