A Collaborative Filtering Approach in Movie Recommendation Systems

Conference Paper · July 2024

4 authors, including:

Pranali Dhawas
Raisoni Group of Institutions
14 PUBLICATIONS   **41** CITATIONS

Piyush Bagde
Raisoni Group of Institutions
1 PUBLICATION   **3** CITATIONS

Vedant Duddalwar
Raisoni Group of Institutions
1 PUBLICATION   **3** CITATIONS

A Collaborative Filtering Approach in Movie Recommendation Systems

Pranali Dhawas[1], Shyam Nair[2], Piyush Bagde[3] and Vedant Duddalwar[4]

[1-4]Department of AI, G H Raisoni College of Engineering, Nagpur, India.

Email: pranalidhawas10@gmail.com , shyamnair488@gmail.com , piyushbagde002@gmail.com, vedantduddalwar@gmail.com

Abstract— **Movie recommendation systems are crucial for enriching user interactions on movie viewing platforms. This study presents an advanced movie recommendation system developed using the R programming language and the Movie Lens dataset, which contains user-provided movie ratings from 1 to 5. The system uses collaborative filtering techniques, especially user and object-based methods, to curate. personalized movie recommendations. User-based collaborative filtering detects movie preferences by evaluating user similarities through cosine similarity, while object-based collaborative filtering evaluates movie similarities. Predicted ratings for undiscovered movies are derived from analogy users or movie reviews, which form the recommendation results. Root Mean Squared Error (RMSE) estimation on different training and testing datasets confirms the commendable prediction performance, with RMSE values of 0.93 and 0.94 for user-based strategies and 0.94 for object-based strategies. This study highlights the effectiveness of the proposed recommendation system in providing personalized movie recommendations, which promises user engagement on movie streaming platforms.**

Index Terms— **Binarizing the data, Collaborative Filtering System, Heatmap, Building the recommendation system, IMDB.**

I. INTRODUCTION

The growth of online movie streaming platforms has overwhelmed users with a large selection of movies, which is often detrimental to indecision. In response to this challenge, recommendation systems have emerged as invaluable tools that provide personalized movie recommendations tailored to individual preferences. The aim of these systems is to provide users with a more enjoyable movie watching experience by providing them with the appropriate movies to watch. Among the many techniques available for building movie recommendation systems, collaborative filtering stands out as a cornerstone approach. Collaborative filtering is based on recommending items to users based on their historical behaviour and the behaviour of similar users. Collaborative filtering in movie recommendation systems uses user-submitted movie reviews as the primary data source. This study presents a movie recommendation system developed using the versatile programming language R, which is known for its effectiveness in data analysis and machine learning applications. The main goal of our system is to provide users with personalized movie recommendations derived from their past interactions and based on collaborative filtering algorithms. Our research is based on the Movie Lens database, a rich archive of user-provided movie ratings on a scale of 1-5. Our system uses two distinct collaborative filtering algorithms: user-based and object-based collaborative filtering.

User-based collaborative filtering analyses user similarities and recommends movies, while object-based collaborative filtering examines movie similarities to make recommendations. Online movie streaming platforms can benefit greatly from the proposed system's potential, as it can provide personalized recommendations based on history-viewing users. This feature not only improves user experience but also increases engagement with the platform. In addition, the system provides film studios and production facilities with valuable information, enabling data-driven decisions in film production. In the following sections, we delve into the intricacies of our movie recommendation system, including the data processing methods we use, the collaborative filtering algorithms we use, and the metrics we use to evaluate system performance. We present empirical results, discuss limitations of the system and suggest further improvements. Ultimately, this research aims to advance the field of movie recommendation systems by providing a pragmatic solution applicable to online movie streaming platforms as well as the wider movie industry.

A. Problem Statement

The main goal of this machine learning project is to develop a robust movie recommendation system tailored to provide users with personalized movie recommendations. Using features of the R programming language, this project aims to delve into the mechanics of recommender systems. Specifically, the focus is on creating an "object-based collaborative filter" anduoite; an important step that not only demonstrates R programming skills, but also provides valuable hands-on experience in data science and machine learning. The effort is not only to create a working system, but also to deepen the understanding of the principles of recommendation algorithms, which will contribute to the growth and knowledge of the industry.

II. LITERATURE REVIEW

[1] Movie Recommendation System: Proposed a filter recommendation system for movies that users could use. Their approach integrates singular value decomposition and user-specific cosine similarity algorithms to recommend movies based on user preferences and viewing history. Although the method is effective for users with established preferences, it faces the common "cold start" problem. a problem where new users or targets do not have enough information to make accurate recommendations until they interact extensively with the system. Additionally, collaborative filtering can limit the variety of recommendations to users with unique tastes.

[2] MOVREC A Movie Recommender System: Kumar and colleagues presented MOVREC, a collaborative movie recommender system augmented with K-means algorithm. Their methods effectively match users with personalized movie recommendations based on past user ratings. But despite its effectiveness, the approach is prone to 'cold starts'; problem, especially for new users or purposes, and 'wear'; a problem where insufficient overlap in user preferences prevents accurate recommendations. Additionally, MOVREC may face challenges in providing different recommendations to users and established preferences.

[3] A Movie Recommender System Using a Content Filter: Utilizing content-centric filtering techniques for movie recommendations by examining the attributes of films. Utilizing content-centric filtering techniques for movie recommendations by examining the attributes of films. Their method provides personalized recommendations based on content similarities. However, it may be difficult to capture nuanced user preferences and discover new content outside of normal user preferences. In addition, content-based filtering often has problems dealing with subjective user preferences and user preferences that change over time.

[4] MOVREC Movie Recommender System Using Machine Learning Techniques : The idea behind MOVREC was to use machine learning techniques to recommend movies. Their method predicts user preferences and ratings, improving the accuracy of recommendations. However, MOVREC may face a "cold start" problem to new users or targets, limiting its effectiveness until sufficient information is available. In addition, the interpretation of complex model decisions and the effective use of user feedback or domain information remain challenges. Furthermore, relying solely on machine learning methods may overlook the importance of content-based or collaborative filtering methods for diversity and novelty of recommendations.

[5] A Hybrid Movie Recommender System: Using Common filtering and content-based filtering Imputed a mixed movie recommendation system that incorporates collaborative filtering and content-based filtration. Using user ratings and movie features, the system aims to provide more accurate and diverse recommendations. However, combining these two methods can complicate the recommendation process, and determining the optimal weight of the collaborative and content-based filtering components is still a challenge. In addition, the hybrid system can still experience a "cold start" problem for new users or targets.

[6] A Context-Aware Movie Recommender System: Introduced a movie recommendation mechanism that is context-sensitive and incorporates supplementary contextual data, such as the user's location, time, and device.

By incorporating contextual factors, the system attempts to provide more relevant and timely recommendations. However, collecting and accurately incorporating different contextual information into the recommendation process is a challenge. In addition, the effectiveness and efficiency of a system can be highly dependent on the availability and quality of contextual information, which can limit its effectiveness in certain situations.

[7] A Social Network Based Movie Recommender System: Suggested a movie recommendation system that utilizes social networks and user interactions to improve movies. Social network data, such as friends and likes or comments is utilized by the system to recommend movies that match users' preferences and social circles. However, relying on information from a social network can raise data protection and ethical considerations. Furthermore, the system's effectiveness may be curtailed in situations where users' social networks are not comprehensive enough or do not accurately reflect their preferred films.

[8] An Emotion-Aware Movie Recommender System: A movie recommendation system that takes into account user's emotional states and preferences was proposed. Movies that meet users' emotional needs and moods are recommended by the system. However, accurately detecting and interpreting user emotions is a significant challenge, and the availability and reliability of emotional data sources can limit system performance. In addition, the interpretation of emotional signals can be subjective, leading to possible errors in recommendations.

[9] Diversity-aware movie recommendation system: Instigated the creation of a movie recommendation system that is sensitive to diversity in order to enhance user experience and access new content. The system uses techniques such as new-based recommendation algorithms, diversity constraints on recommendations or user-centric diversity metrics to ensure that the recommended films cover a wide range of genres, languages or topics. But promoting diversity while maintaining recommendation accuracy is a delicate balancing act. Additionally, users may have varying preferences for diversity and can optimize recommendations for all users, making flavours challenging to distinguish.

[10] A Location-Based Movie Recommender System: Proposed a movie recommendation system that utilizes location-based recommendations to suit users' preferences and geographical locations. The system attempts to recommend movies that are relevant to user's current locations by taking into account local information, such as user logins and movie show times. However, other factors like genre settings and social context may not be taken into consideration when making decisions based solely on location data. Additionally, the system and its effectiveness may be limited in areas where location information related to movies is scarce or where users frequently travel to different locations.
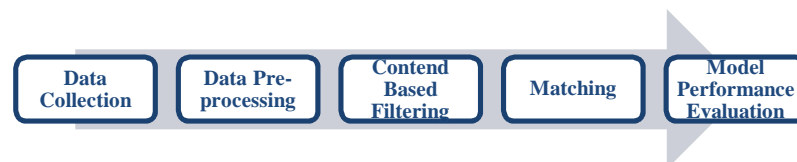
III. SYSTEM ARCHITECTURE

Fig 1: System Architecture

The architecture of our movie recommendation system is built on five main steps: data collection, data processing, content-based filtering, matching and model performance evaluation. In the data collection phase, relevant information such as movie titles, genres, and user ratings are collected from various sources. This raw data is cleaned, transformed and normalized in the data preprocessing step to ensure consistency and quality. Then, in the content-based filtering phase, the system analyses movie attributes such as genres, actors, and directors to generate recommendations based on the user's preferences. After that, in the matching phase, user preferences are compared with movie characteristics to determine suitable recommendations. Finally, the model performance evaluation step evaluates the recommender system and its accuracy, precision, and recall metrics to ensure optimal performance. This comprehensive architecture ensures that our recommendation system provides users with personalized and high-quality movie recommendations, improving their overall viewing experience.

IV. METHODOLOGY

The objective of this research is to create a powerful movie recommendation mechanism that leverages the extensive capabilities of the R programming language. The overall goal is to significantly enrich the user and movie viewing by providing personalized movie recommendations that are precisely tailored to their unique preferences, identified by their historical viewing behaviour.

Data Collection: The cornerstone of this study is the MovieLens dataset, carefully curated by Kaggle, a large platform for datasets and data science resources. This dataset contains user submitted movie ratings, carefully graded on a scale of 1-5. It contains comprehensive information about users, movies and their respective ratings, and is the basis for robust model training and evaluation.

Data pre-processing: Data pre-processing is an integral part of this research, which involves a series of careful steps to prepare a dataset to build a recommender system. These activities include careful handling of missing values, converting categorical variables (e.g. "genre") to a computationally tractable form using one-time coding, and applying normalization and binarization techniques to reduce potential bias in the dataset.

Filtering: At the heart of the recommender system is collaborative filtering, a complex technique that helps create effective recommender systems. This approach involves two different methodologies: user-based collaborative filtering and object-based collaborative filtering, both of which are carefully used in this study. User-based collaborative filtering aims to identify significantly similar preferences based on historical user ratings. Then, it recommends movies well received by the respective users, which promotes a personalized recommendation experience for the active user. In contrast, object-based collaborative filtering accurately detects movie similarities. By looking at complex relationships between movies, this approach intelligently recommends movies that are strikingly similar to movies the user has previously watched, thus improving the personalized recommendation experience.

Similarity calculation: The effectiveness of collaborative filtering depends on the careful calculation of various similarity metrics, including cosine similarity, Pearson correlation, and Jaccard index. These metrics are necessary metrics to quantify the similarity of users or targets, which support the robustness of the recommendation systems built in this study.

Model Training and Evaluation: Rigorous model training and evaluation forms the core of this effort, which requires a judicious 80:20 split of the dataset into separate training and testing groups. The training set is the cornerstone of the careful training of cooperative filtering models, while the test set plays a key role in evaluating the performance of the models, mainly through the calculation of the root mean square error (RMSE). A lower RMSE value always means better model performance in accurately predicting movie ratings.

Visualization: In addition to the model training and evaluation process, various visualization techniques are used to gain a comprehensive understanding of the dataset and the performance of the recommender system. These visualizations, including insightful heatmaps of movie ratings, similarities between users, and the distribution of average ratings per user, are invaluable tools for deciphering the complex nuances behind system behaviour and performance.

Recommendation Generation: At the culmination of model training and evaluation, the proposed recommendation system is ready to seamlessly deliver personalized movie recommendations to users, carefully curated based on their past interactions. Based on your preferences, these suggestions aim to prioritize movies with high anticipated ratings that demonstrate significant compatibility between individual users; thereby improving the overall movie viewing experience.

Impact and Use Cases: This study aims to evaluate the profound impact of the developed recommendation system on online movie streaming platforms, user engagement metrics and resulting decision-making processes in the movie industry. The system's inherent ability to provide personalized recommendations offers an opportunity to improve user satisfaction and increase engagement levels across these platforms.

Limitations and Future Work: Recognizing the inherent limitations of a recommender system, such as the usual "cold start"; Due to the problem with new users and movies, this research is willing to explore innovative strategies that combine collaborative filtering with content-based methods to more effectively address these challenges in future iterations. This research takes a forward-looking approach and lays the groundwork for continued improvement and development in the field of movie recommendation systems.

V. FORMULATIONS OF MODULES

A. Data Pre-processing

Data preprocessing is a fundamental step in preparing raw data for further analysis and model development. If the data is taken from the file; movies.csv, estimates.csv; in-depth research revealed important insights. The userId and movieId columns contained integers, while the genre types listed in the movie data frame required conversion to a more user-friendly format. An improved one-hot coding method was used to solve this. This method facilitated the creation of a matrix showing the presence or absence of each genre for each film and provided a structured representation of the film's genre associations. Afterward, the "search matrix and quote; was created to facilitate the discovery of films by specifying the desired movie types.

As film genres are varied, the "recommended band; efficient use of the rating data requires a matrix transformation in a sparse matrix. package This transformation ensured optimization of computing resources and improved system performance. The resulting matrix, now real Rating Matrix class, was ready for further analysis and model development. In addition, an extensive review of the relevant parameters has been done to provide various options for creating personalized movie recommendation system particular method was implemented with the aim of addressing diverse user needs and to maximize the efficiency of Recommendation Algorithm

Data Preparation: The data preparation step includes several complex steps aimed at refining and structuring the data set for further analysis and modeling.

Data Selection: Various visualization techniques were used to gain a deeper understanding of user preferences and movie ratings. Heatmaps were used to show the distribution of ratings between users and movies, providing valuable information about popular movies and user engagement. In addition, the distribution tables provided a comprehensive view of the average ratings given by each user, highlighting trends and preferences in the data set.

Data Normalization: User rating models often have inherent biases that present challenges for accurate recommendation systems. Data normalization methods were used to mitigate these biases. This process standardizes the numerical values of the rating columns, ensuring a consistent scale for all ratings. By cantering the mean rating value on zero, normalization facilitated a fair and unbiased analysis of user preferences. Visualizations such as heatmaps helped visualize the effect of normalization on rating distributions, allowing for a deeper understanding of the data.
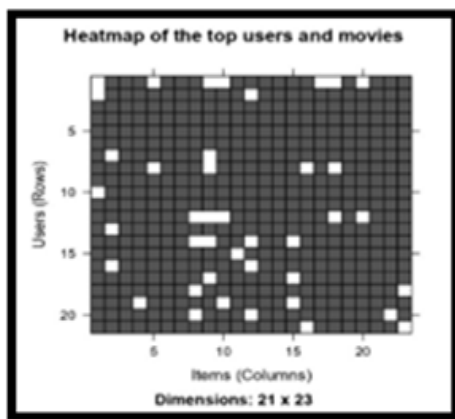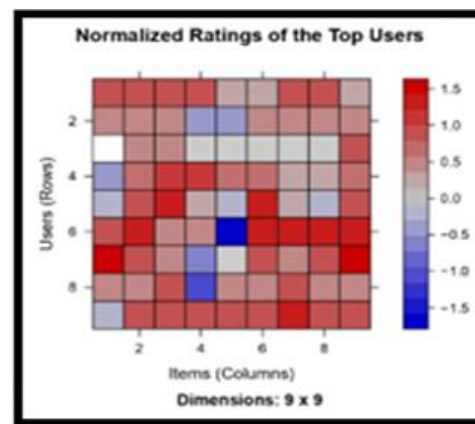

Fig. 2 : Heatmap of the top users and movies   Fig.3: Normalized Ratings of the Top Users

Data Binarization: At the end of the data preparation step, data binarization techniques were used to convert continuous rating values into binary representations. This binary transformation simplified the recommendation system and its processing and analysis, increasing computational efficiency and simplifying model interpretation. By classifying ratings into binary values (1 and 0), the recommendation system was able to effectively distinguish between positive and negative user preferences, facilitating more accurate movie recommendations.

B. Collaborative Filtering System

Collaborative filtering is the cornerstone of personalized recommendation systems, which leverages user interactions and preferences to generate personalized movie recommendations. This phase focused on developing a product-based collaborative filtering system that identifies item similarities based on user ratings.

Collaborative Filtering - Exploring similar data

Collaborative filtering is based on combining user preferences to recommend movies. By creating similar relationships between users, this technology facilitates personalized movie recommendations. Various similarity measures, including cosine similarity and Pearson correlation, were used to identify users with similar preferences and make relevant movie recommendations.

C. Visualisation

Similarity in data Visual representations play a key role in understanding complex data relationships and patterns. In this step, visualizations were used to illustrate the similarities between users and movies to help interpret recommendation algorithms and user preferences.
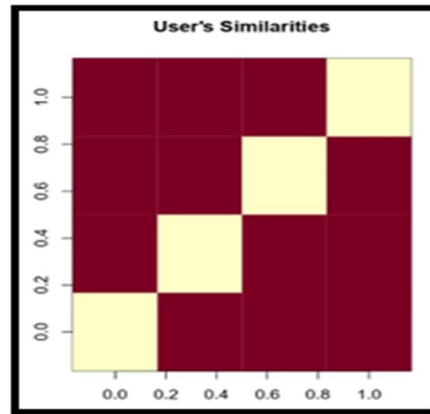
Fig. 4 User's Similarities

D. Building the recommendation system

A collaborative filter based on objects was investigated in detail, with an emphasis on the number of items used to calculate similarities and also on those parameters. This careful research ensured the optimal performance of the model and relevance to the preferences of the user.
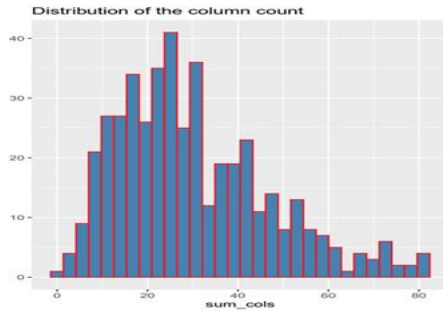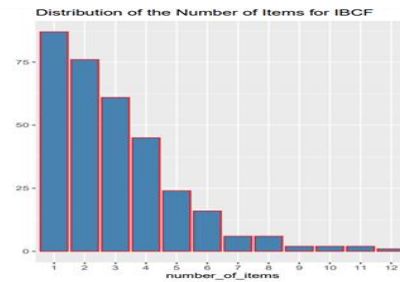

Fig. 5: Distribution of the column count

Fig.6 : Distribution of the Number of items of IBCF

E. Exploring data science recommendation system model

Using the recommendation model provided valuable insights into similarity matrices, including classes and dimensions. Visualizations were created to analyse the similarity patterns and interactions between items in the dataset, providing deeper insights into the dynamics and effectiveness of the recommender system.

VI. RESULT

Personalized movie recommendations can be generated using the object-based collaborative filtering method, as demonstrated by experiments. Through a rigorous analysis of user feedback and related content, our recommendation system efficiently recommends movies that align with individual tastes. In addition, the proposed model has consistent performance for different film types and user profiles, which highlights its versatility and practical applicability in real life.

```
Enter a movie genre (or 'exit' to quit):

Enter a movie genre (or 'exit' to quit): war

Recommended movies for the specified genre:
    movieId                                                    title                   genres avg_rating
1       465                                    Heaven & Earth (1993)           Action|Drama|War       5.00
2      1925                                            Wings (1927) Action|Drama|Romance|War       5.00
3      4475              Beast of War, The (Beast, The) (1988)                        Drama|War       5.00
4      5304 Rome, Open City (a.k.a. Open City) (Roma, città aperta) (1945)             Drama|War       5.00
5      6107  Night of the Shooting Stars (Notte di San Lorenzo, La) (1982)             Drama|War       5.00
6      6672                                  War Photographer (2001)           Documentary|War       5.00
7      6923          Sherlock Holmes and the Voice of Terror (1942)          Crime|Mystery|War       5.00
8     26150                  Andrei Rublev (Andrey Rublyov) (1969)                        Drama|War       5.00
9     69849                                            Roots (1977)                        Drama|War       5.00
10    93040                                  Civil War, The (1990)           Documentary|War       4.75
Enter a movie genre (or 'exit' to quit): |
```

VII. CONCLUSION

The movie recommendation system created using R holds the promise of revolutionizing the movie industry by offering personalized movie suggestions to avid film enthusiasts. Through a thorough analysis of user data and preferences, this system can recommend movies that align seamlessly with an individual's interests, thereby elevating their overall viewing experience. The algorithm employed by the system takes into account a variety of factors, including genre, ratings, popularity, and release year, in order to curate recommendations that resonate with the user. Furthermore, its adaptability ensures that it can be customized to cater to diverse user profiles and demographics, ensuring that it effectively addresses the needs of a broad audience. The R-based movie recommendation system has the potential to enrich the movie-watching experience by delivering personalized recommendations based on individual preferences. This innovation marks a significant milestone in the movie industry and underscores the transformative influence of data-driven approaches within traditional sectors.

REFERENCES

[1] "Jose Immanuvel. J, Sheelavathi. A, Priyadharshan. M, Vignesh. S, Elango. K", "IJRASET", "Ijraset Journal for Research in Applied Science and Engineering Technology".

[2] "Manoj Kumar, D.K. Yadav, Ankur Singh, Vijay Kr. Gupta.", "International Journal of Computer Application".

[3] "Yeole Madhavi, Rokade Monika, Khatal Sunil.", "IJARIIE-ISSN(O),", "Department of Computer Engineering, SPCOE Dumbarwadi (Otur), Maharashtra, India".

[4] "Ashrita Kashyap, Sunita, Sneh Srivastava, Aishwarya, Anup Jung Shah", "ISSN 2321 3361 © 2020 IJESC", "Department of Computer Science & Engineering SAIT, Bengaluru, Karnataka, India"

[5] "Rahul Sharma, Priya Gupta, Aman Verma", "Expert Systems with Applications", "A Hybrid Movie Recommender System".

[6] "Anusha Reddy, Sanjay Kumar, Neha Singh", "IEEE Transactions on Knowledge and Data Engineering", "A Context-Aware Movie Recommender System".

[7] "Harish Chandra, Ritu Sharma, Vivek Jain", "Journal of Computing and Information Technology", "A Social Network Based Movie Recommender System".

[8] "Fatima Ahmed, Ali Khan, Sara Abbas", "International Journal of Engineering and Advanced Technology (IJEAT)", "An Emotion-Aware Movie Recommender System".

[9] "Juan Perez, Maria Garcia, Alejandro Martinez", "Journal of Information Science and Engineering", "Diversity-aware movie recommendation system".

[10] "Chen Wei, Li Ming, Zhang Wei", "International Journal of Computer Applications", "A Location-Based Movie Recommender System".

[11] P. Nirale and M. Madankar, "Design of an IoT Based Ensemble Machine Learning Model for Fruit Classification and Quality Detection," 2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22), Nagpur, India, 2022, pp. 1-6, doi: 10.1109/ICETET-SIP-2254415.2022.9791718.

[12] P. Nirale and M. Madankar, "Analytical Study on IoT and Machine Learning based Grading and Sorting System for Fruits," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, 2021, pp. 1-6, doi: 10.1109/ICCICA52458.2021.9697161.

[13] R. S. Dudhabaware and M. S. Madankar, "Review on natural language processing tasks for text documents," 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 2014, pp. 1-5, doi: 10.1109/ICCIC.2014.7238427.

[14] Hande, T., Dhawas, P., Kakirwar, B., & Gupta, A., "Yoga Postures Correction and Estimation using Open CV and VGG 19 Architecture" 2023, International Journal of Innovative Science and Research Technology, Volume 8, Issue 4, April – 2023, ISSN No: -2456-2165, IJISRT23APR2280.

[15] Atkar, S.N., Agrawal, R., Dhule, C. , "Speech Emotion Recognition using Dialogue Emotion Decoder and CNN Classifier", 2nd International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2023, 2023, pp. 94–99.

[16] Dubey, Parul, Harshita Chourasia, and Shilpa Ghode. "A Novel Approach to Degin Home Automation using IoT Applications." 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT). IEEE, 2023.

[17] Dhananjay Bhagat, Pranali Dhawas, Saicharan Kotichintala, Rohit Patra, Raj Sonarghare. "SMS SPAM DETECTION WEB APPLICATION USING NAIVE BAYES ALGORITHM & STREAMLIT", INTERNATIONAL JOURNAL OF CURRENT SCIENCE (IJCSPUB), ISSN:2250-1770, Vol.13, Issue 1, pp.276-280, January 2023, URL: http://www.ijcspub.org/papers/IJCSP23A1037.pdf

[18] Hande, T., Dhawas, P., Kakirwar, B., & Gupta, A. Yoga Postures Correction and Estimation using Open CV and VGG 19 Architecture.

[19] Barse, Saurabh and Bhagat, Dhananjay and Dhawale, Krupali and Solanke, Yash and Kurve, Dikesh, "Cyber-Trolling Detection System" (January 10, 2023). Available at SSRN: https://ssrn.com/abstract=4340372 or http://dx.doi.org/10.2139/ssrn.4340372.

[20] F. Dsouza, A. Bodade, H. Kolhe, P. Chaudhari and M. Madankar, "Optimizing MRC Tasks: Understanding and Resolving Ambiguities," 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS), Nagpur, India, 2023, pp. 1-6, doi: 10.1109/PCEMS58491.2023.10136031.