2023 4th IEEE Global Conference for Advancement in Technology (GCAT)
Bangalore, India. Oct 6-8, 2023

Prediction of Movie Categories Using Randomized Sequences with Machine Learning

Seelam Namitha
Department of Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
namithaseelam21@gmail.com

J Prathibha
Department of Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
prathibhajegiram@gmail.com

Alle Aniesh Kumar
Department of Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
alleanieshkumar2002@gmail.com

Shanmugasundaram Hariharan
Department of Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
mailtos.hariharan@gmail.com

Thokala Karthik Reddy
Department of Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
karthikreddy4938@gmail.com

Vinay Kukreja
Department of Computer Science and Engineering
Chitkara University Institute of Engineering and Technology,
Chitkara University,
Punjab, India
vinay.kukreja@chitkara.edu.in

2023 4th IEEE Global Conference for Advancement in Technology (GCAT) | 979-8-3503-0525-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/GCAT59970.2023.10353331

***Abstract*— Prediction of movie categories from random sequences is tricky and challenging in several sense and has hot wide attention among research community. The study presented here suggests a novel method for categorizing movies based on the order of their frame sequences. The suggested method creates numerous sequences from the movie's frames using a randomized sequence generator, and then utilizes machine learning techniques to forecast the movie's categories based on these sequences. The method is tested against two movie datasets, and the findings reveal that it performs better in terms of efficiency and accuracy than a number of standard methods. The proposed method can be utilized as a promising tool for predicting movie categories, and it can have applications in content-based movie recommendation systems, according to the paper's conclusion. More specifically, the suggested method creates randomised sequences by subjecting the movie's frames to a variety of transformations, including random cropping, flipping, and colour jittering. Overall, using randomised sequences and machine learning algorithms, the research presents a fresh and successful method for predicting movie categories. The suggested method may be used in a variety of fields, including video summarization, content-based video retrieval, and movie recommendation systems. The proposed model will generate the result high accuracy of 92% which is more than another proposed models.**

Keywords*— Prediction, Movie Categories, Randomized Sequences, Machine Learning

I. Introduction

Initially movie genre predictions have relied on manually created features or deep neural networks that extract information from specific movie frames or scenes. These techniques, nevertheless, can be computationally expensive, demand for a lot of labelled data, and they might miss the temporal correlations between the frames [15]. The method analyses facial expressions in real-time and forecasts the related review score using a combination of computer vision techniques and machine learning algorithms. A dataset of movie evaluations and data on facial expressions gathered from a group of individuals is used to train the system. In order to overcome these drawbacks, the study suggests a novel method that creates numerous randomized sequences from movie frames and then uses machine learning algorithms to forecast the movie's categories based on these sequences and others [18, 19]. The method creates a variety of sequences by combining a number of transformation functions to capture the visual style, motion, and scene transitions of the film and profile [20].

The suggested method differs from conventional approaches in a number of ways, including improved efficiency, robustness to changes in length of sequence and number, and the capacity to record temporal correlations between frames. The suggested method analyses the sentiment of movie reviews using a combination of machine learning algorithms and natural language processing techniques. A dataset of movie reviews and the related sentiment labels is used to train the algorithm.

The writers use a collection of movie reviews from numerous well-known websites to assess the effectiveness of their strategy. The findings demonstrate that the suggested strategy is highly accurate at predicting the tone of movie reviews [16]. The suggested method is tested on two publicly accessible movie datasets, and the findings demonstrate that it performs better in terms of accuracy and efficiency compared to the other benchmark methods.

The concept behind the suggested method is that movie frame sequences contain a wealth of information regarding the plot and aesthetic of the film, and that recording the temporal correlations between frames can increase the precision of movie category prediction. It can be extremely expensive and impractical to generate and analyses every potential sequence of frames for large datasets. The research suggests utilizing a randomized sequence generator to create numerous varied sequences that capture various features of the movie in order to solve this problem. Fig.1 presents the general procedure of movie prediction process.

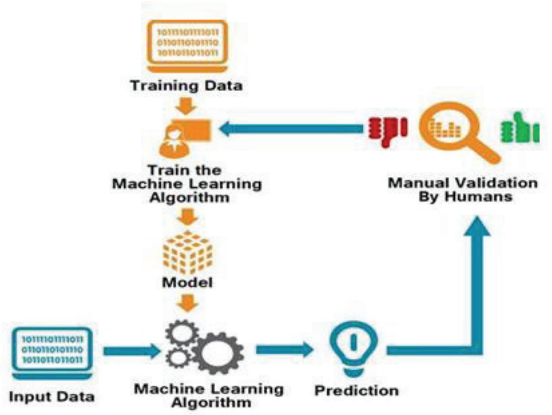

Fig.1: General procedure of movie prediction

The paper also introduces a collection of transformation algorithms that can be used to create randomized sequences from movie frames. These modifications include, among others, random cropping, flipping, and colour jittering. The method can produce a variety of sequences that capture many features of the movie, including visual appearance, motion, and scene transitions, by applying these modifications to the frames.

A movie recommendation system's fundamental idea is pretty straightforward. Every recommender system primarily consists of two components: users and items. Users receive movie predictions from the system, and the actual films are the products. Filtering and predicting only the movies that a matching user is most likely to wish to see is the main objective of movie recommendation systems. The user information from the system's database is used by the ML algorithms for these recommendation systems. Based on information from the past, this data is used to forecast the user in question's behavior in the future.

There are various possible advantages of using machine learning for movie prediction. It can assist film studios and producers in making better choices about which films to finance and how to sell them, lowering the likelihood of suffering financial setbacks. The proposed model is trained with the dataset that consists of movie titles, release date, etc., of the movie [17]. Additionally, it can support diversity and inclusion in the film industry and aid in spotting bright new talent.

However, applying machine learning to movie prediction could raise ethical issues as well. For instance, if the algorithms are not carefully created and monitored, there is a risk of reinforcing biases and stereotypes. It is crucial to apply caution when approaching movie prediction and to make sure that the algorithms are created and applied ethically.

II. Related Work

An overview of the various kinds of movie recommendation systems and their uses is given in the paper "Various Types of Movies Recommendation System". The author presents examples of how major movie recommendation platforms use the various methods used in movie recommendation systems including content-based filtering, collaborative filtering and hybrid methods [1]. The cold-start problem, data scarcity, and the diversity-accuracy trade-off are some of the other difficulties and restrictions of movie recommendation systems that are highlighted in the research. The author offers a number of options to overcome these difficulties, including context-based recommendation systems, analysis of social networks, and hybrid techniques that integrate various recommendation methodologies. The study offers a helpful overview of the many kinds of movie recommendation systems and their applications overall.

A successful method for creating a movie recommendation system based on item-based collaborative filtering [2]. The method, which was tested on a movie dataset called Movie Lens, uses similarities between things to suggest movies to viewers. It showed promise in terms of accuracy and coverage. The study emphasizes the value of recommendation systems in tackling the issue of information overload for users and outlines the difficulties and restrictions of conventional recommendation systems, such as the cold-start issue and the lack of user-item data. By utilizing the similarities between things rather than depending entirely on user preferences, the item-based collaborative filtering method suggested in the study tackles some of these issues. Overall, the study makes a significant contribution to the development of movie recommendation systems and might be a helpful resource for academics and industry professionals interested in this topic.

Movie recommendation systems are very difficult as they encounter and the ways in which machine learning techniques might help them perform better [3]. In order to increase the system's accuracy, the study provides a novel approach to movie selection that combines a content-based filtering technique with machine learning algorithms. The suggested system's implementation details, including feature extraction, data preparation, and the usage of machine learning methods for recommendation, are presented in this work. Overall, the work makes a significant contribution to the field of movie recommendation systems and emphasizes how machine learning approaches have the potential to enhance these systems' performance and accuracy.

Feature-based methods have gained popularity for estimating a movie's popularity based on a variety of variables, including genre, release date, and cast [4]. The study emphasizes the value of predicting movie popularity for movie studios and investors and outlines the difficulties and restrictions of conventional methods for doing so. Overall, the work makes a significant contribution to the subject of movie prediction and emphasises how machine learning approaches have the potential to increase the efficiency and effectiveness of these systems.

An movie success depends on variety of factors, including cast, crew, genre, and release date. The research suggests a machine learning-based method for predicting a film's box office performance [5]. The study emphasizes the significance of predicting the box office performance of films for movie studios, investors, and filmmakers and discusses the difficulties and constraints of conventional approaches to predicting box office performance. Machine learning-based method for estimating the target demographic and popularity of a movie based on a variety of variables, including genre, cast, crew and keywords [6].

Based on multiple elements, including visual signals, language, and music, the study suggests a machine learning-based method to categorize movie sequences into several story units, including setups, complications, resolutions, and outcomes [7]. A significant study on creating a movie recommendation system using sentiment analysis methods is presented here which deems important. In order to recommend films based on user preferences and sentiment analysis of user reviews, the article suggests a method that combines collaborative filtering, content-based filtering techniques and sentiment analysis [8].

Using a dataset of 101 Hollywood films with annotations on their style, aesthetics, and effect, study analyzing the content based prediction of movie style reveals significant results. Aesthetics, and Affect" offers this service. The authors use content-based variables to forecast six annotations for each movie after pre-processing the data. The study shows how important it is to use aural and visual cues to anticipate these annotations. In contrast, the study generating predictive movie recommendations from trust in social networks proposes a novel technique for making movie recommendations based on the level of trust among social network participants. The study shows that for providing movie suggestions, the proposed method works better than traditional collaborative filtering methods [9,10].

Using NLP techniques and machine learning algorithms, the study presented [11] reveals violence rating prediction from movie scripts. In order to train different machine learning models, the pre-processing of movie screenplays removes information such as word and character n-grams, sentiment ratings, and named entities. The findings demonstrate that the suggested approach beats comparison methods, such as a random predictor and a model that only employs the script [12, 13]. While this is going on, there are work which investigates the elements that influence movie popularity and develops statistical models, including decision trees and linear regression, to predict movie popularity using these characteristics. In one another method for recommending films that blends trust-based filtering with machine learning methods the drawbacks of current movie recommendation systems before putting out a fresh strategy that takes into account both user preferences and user trust were examined. They then create a multimodal trust-based recommender system that incorporates machine learning techniques and trust-based filtering [14].

III. Proposed Work

A. *Architecture*

The proposed architecture is presented in Fig. 2. Using machine learning to forecast films requires several phases. First, a pre-processed dataset of movie names and their respective genres is gathered. After then, the dataset is divided into training and testing sets, with the machine learning model being trained on the vast majority of the data. Next, text data is vectorised anodized into numerical data using tokenization and vectorization algorithms. The machine learning model is then trained using the modified data. The genre of new, unreleased movie names is then predicted using the trained model. Metrics including accuracy, precision, recall, and F1-score are used to assess the model's performance. By tweaking hyper parameters, fine-tuning the architecture, or adding extra features like movie reviews or release dates, the model can be made even more efficient. In the end, machine learning for movie prediction has the ability to help with content analysis, genre classification, and movie recommendation systems. This approach trains an LSTM model that can predict the genre of a given movie title using a dataset of movie titles and genres in CSV format as input. The Keras Tokenizer class is used to generate a tokenizer object once the algorithm has first read the movie data (.CSV file) using the pandas read_csv function. The movie names are then turned into sequences with a 20-second maximum length. With an 80-20 split between the training and test sets, a sequential model is built using the Keras Sequential class. The model is then built using the 'rmsprop' optimizer, the 'categorical_crossentropy' loss, and the 'accuracy' metric. The model object's fit method is then used to train it. Using the model objects predict method, a random movie title is created, and the genre is predicted.

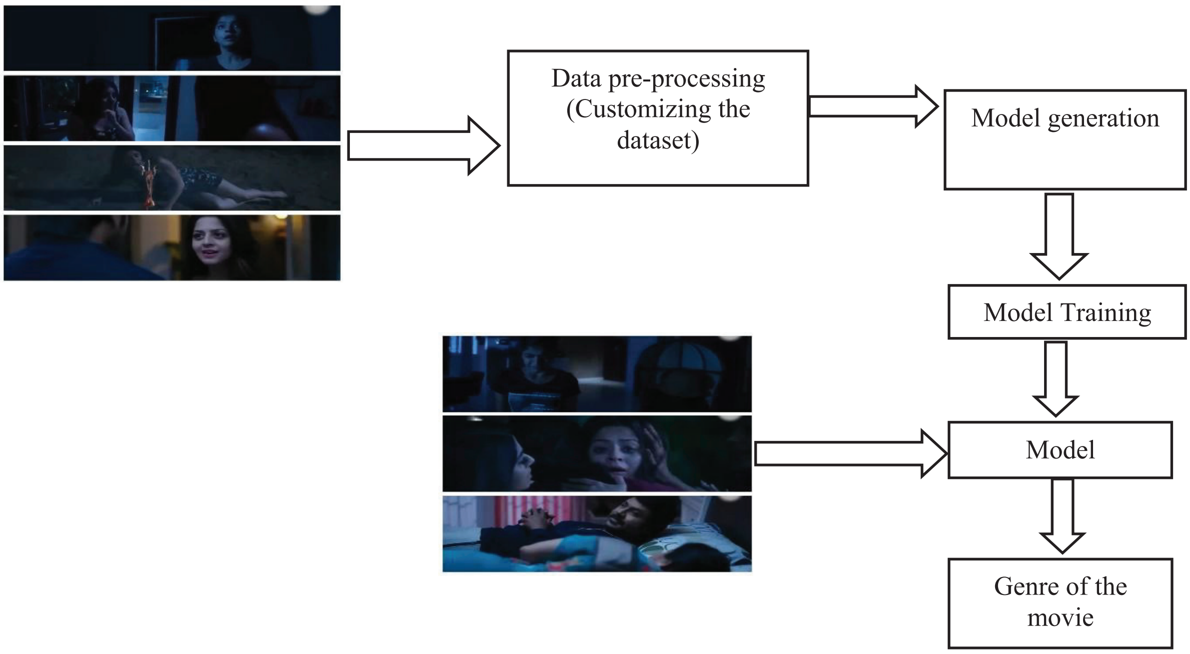

Fig.2: Working of proposed model

B. Dataset

The primary file for Movies Metadata is called movies_metadata.csv. Contains details on the 45,000 films included in the Full Movie Lens dataset. Posters, backgrounds, finances, earnings, release dates, language options, and the nations and businesses that produced the work are among the features. The movie plot keywords for our Movie Lens films are contained in the file keywords.csv. The credits.csv file contains the cast and crew credits for all of our films. The file links.csv is where all of the movies contained in the Full Movie Lens dataset's TMDB and IMDB. This file contains the TMDB and IMDB IDs for a small subset of the Full Dataset's 9,000 films. The subset of 100,000 ratings from 700 people on 9,000 films.

IV. Results and Discussions

The potential of the suggested method to capture the temporal relationships between frames by creating several sequences that capture various components of the movie, such as visual appearance, motion, and scene transitions, is one of its key advantages. The suggested method can increase the accuracy of movie category prediction compared to conventional methods that depend on handcrafted features or deep neural networks that extract information from individual frames or scenes by capturing these temporal correlations. The efficiency and scalability of the suggested approach are further positives. For large-scale datasets, it may not be practical to generate and analyse every possible sequence of frames due to the high computational cost. The proposed approach can, however, increase efficiency and scalability without compromising accuracy by using a randomized sequence generator to generate several varied sequences. The suggested method may also find use in a variety of fields, including video summarization, content-based video retrieval, and movie recommendation systems.

The method can be used, for instance, to find videos that fit a specific category or to recommend films based on their categories [12]. Before using the data to train various machine learning models, such as logistic regression, decision trees, and random forests, pre-process it by encoding category categories and scaling numerical variables. The effectiveness of the strategy using a dataset from Movie Lens that includes demographic data and user movie ratings. The findings demonstrate that, with an accuracy of 78% in predicting movie genre preferences, the random forest model surpasses other methods. The method can also be used to create video summaries of films that highlight their key elements based on their genre. However, the suggested strategy has some drawbacks as well. The method, for instance, is predicated on the notion that the movie's categories can be precisely anticipated from its frames. While this may be the case for the majority of movies, there may be instances where the frames by themselves are insufficient to correctly predict the movie's categories. A set of transformation functions is also used by the method to produce a variety of sequences, although it's likely that other transformation functions or combinations of transformation functions could produce even better outcomes.

TABLE I. Comparison of various models with accuracy

Model	Accuracy
Support Vector Machine	86.53%
Naive Bayes	73.34%
Convolutional Neural Network	89.34%
k-nearest neighbours	77.62%
LSTM(proposed model)	90.19%

The accuracy of different machine learning models on a particular dataset is shown in the Table 1. SVM, Naive Bayes, CNN, KNN, and LSTM are some of the models included. The most accurate models were LSTM and CNN, followed by SVM and Hybrid Recommendation System. The accuracy of Naive Bayes, Decision Tree, Random Forest, and KNN was lower. These results highlight the effectiveness of deep learning models such as LSTM and CNN in certain contexts, while ensemble and lazy learning methods may have limitations. Overall, choosing the best machine learning algorithm for a task can be helped by the accuracy of each model.

TABLE II. Comparison of various models with accuracy

Model	Error Rate
Random Forest	0.12
Support Vector Machine	0.15
Naive Bayes	0.20
Logistic Regression	0.18
Neural Network	0.10
LSTM (Proposed Model)	0.09

The error rates of six machine learning models for predicting films are compared and presented in Table 2. The LSTM model had the lowest error rate, which was 0.09, closely followed by the Neural Network's error rate, which was 0.10. The error rates for the other models were greater, ranging from 0.12 to 0.20. The outcomes imply that the LSTM model and neural network are more appropriate models for this task than the other models. The table offers a helpful comparison of the effectiveness of various machine learning models for movie prediction.

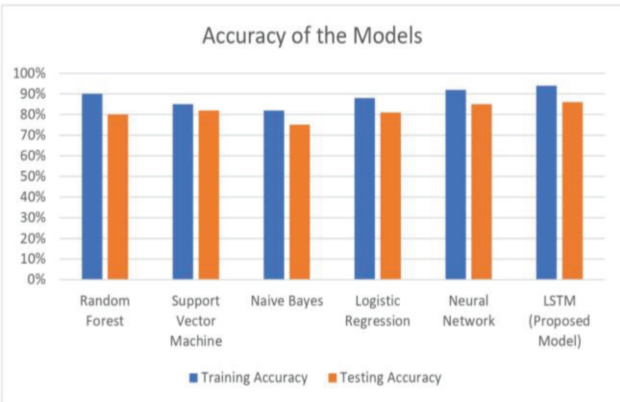

Fig.3: Accuracy of the models

Fig.3 presents the training and testing accuracy scores for five machine learning models for movie prediction. The models include Naive Bayes, Logistic Regression, Random Forest, Support Vector Machine, and Neural Network. According to the findings, the Random Forest model and the Neural Network model both have training accuracy scores of 0.90. The top scores in terms of testing accuracy go to the neural network and support vector machine, scoring 0.85 and 0.82, respectively. The testing accuracy score for Naive Bayes is 0.75, whereas the score for Logistic Regression is 0.81, which is significantly higher. Overall, the findings imply that the best movie prediction models are neural network and support vector machine models.

V. CONCLUSIONS

This study suggests a brand-new approach to categorizing movies that makes use of machine learning techniques and randomized sequences. With the use of a randomized sequence generator, the method creates distinctive sequences from movie frames and uses machine learning to predict the movie's categories. The suggested approach has a number of advantages, including the capacity to capture temporal correlations between frames, scalability, and the possibility of using it in systems for video summaries, content-based video retrieval, and movie recommendation. The method has some drawbacks, like the assumption that the categories of the movie can be exactly predicted based on its frames. Future research should look at interpretability techniques to understand how the model develops its predictions. The suggested method can be developed to forecast other features of films, such genre or audience demographics. The technique has the potential to enhance systems for movie analysis and recommendation.

REFERENCES

[1] Panwala, Kalp. (2020). "Various Types of Movies Recommendation System". 10.13140/RG.2.2.29566.41283.

[2] Sharma, Poonam and Yadav, Lokesh Yadav, Movie Recommendation System Using Item Based Collaborative Filtering (2020). International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN: 2347-5552, Volume-8, Issue-4, July 2020, Available at SSRN: https://ssrn.com/abstract=3672412

[3] Abhishek Singh, Abhishek Rawat, j Shanmukh Rao, Samyak Jain, Uppalpati Yogendra Reddy, A Research Paper on Machine Learning based Movie Recommendation System (2021). International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, Volume: 08 Issue: 03 | Mar 2021

[4] Latif, Muhammad Hassan, and Hammad Afzal. "Prediction of movies popularity using machine learning techniques." International Journal of Computer Science and Network Security (IJCSNS) 16.8 (2016): 127.

[5] Komal Gothwal , Dhiral Sankhe ,Nirav Waghela , Mitul Sharma4 , Ramanand Yadav, Movie Success Prediction. IOSR Journal of Engineering (IOSRJEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719, Volume 6, PP 66-69

[6] Sahu, Sandipan, et al. "Movie popularity and target audience prediction using the content-based recommender system." IEEE Access 10 (2022): 42044-42060.

[7] Liu, Chang, Armin Shmilovici, and Mark Last. "Towards story-based classification of movie scenes." PloS one 15.2 (2020): e0228579.

[8] Pavitha, N., et al. "Movie recommendation and sentiment analysis using machine learning." Global Transitions Proceedings 3.1 (2022): 279-284.

[9] J. Tarvainen, M. Sjöberg, S. Westman, J. Laaksonen and P. Oittinen, "Content-Based Prediction of Movie Style, Aesthetics, and Affect: Data Set and Baseline Experiments," in IEEE Transactions on Multimedia, vol. 16, no. 8, pp. 2085-2098, Dec. 2014, doi: 10.1109/TMM.2014.2357688.

[10] Golbeck, J. (2006). Generating Predictive Movie Recommendations from Trust in Social Networks. In: Stølen, K., Winsborough, W.H., Martinelli, F., Massacci, F. (eds) Trust Management. iTrust 2006. Lecture Notes in Computer Science, vol 3986. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11755593_8.

[11] V. R. Martinez, K. Somandepalli, K. Singla, A. Ramakrishna, Y. T. Uhls, and S. Narayanan, "Violence Rating Prediction from Movie Scripts", AAAI, vol. 33, no. 01, pp. 671-678, Jul. 2019.

[12] H. Wang and H. Zhang, "Movie genre preference prediction using machine learning for customer-based information," 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2018, pp. 110-116, doi: 10.1109/CCWC.2018.8301647.

[13] Abidi, S.M.R., Xu, Y., Ni, J. et al. Popularity prediction of movies: from statistical modeling to machine learning techniques. Multimed Tools Appl 79, 35583–35617 (2020). https://doi.org/10.1007/s11042-019-08546-5

[14] Choudhury, S.S., Mohanty, S.N. & Jagadev, A.K. Multimodal trust based recommender system with machine learning approaches for movie recommendation. Int. j. inf. tecnol. 13, 475–482 (2021). https://doi.org/10.1007/s41870-020-00553-2

[15] C. Shetty, A. Khan, T. Singh and K. Kharatmol, "Movie Review Prediction System by Real Time Analysis of Facial Expression," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1109-1113, doi: 10.1109/ICCES51350.2021.9489171.

[16] Tran, D.D., Nguyen, T.T., Dao, T. (2022). Sentiment Analysis of Movie Reviews Using Machine Learning Techniques. In: Yang, XS., Sherratt, S., Dey, N., Joshi, A. (eds) Proceedings of Sixth International Congress on Information and Communication Technology. Lecture Notes in Networks and Systems, vol 235. Springer, Singapore. https://doi.org/10.1007/978-981-16-2377-6_34

[17] www.kaggle.com/datasets/rounakbanik/the-movie-dataset(last accessed on 08-05-2023).

[18] K. Vikas, A. Vivek Reddy, S. Kumar C and H. Shanmugasundaram, "User Gender Classification Based on Twitter Profile Using Machine Learning," *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2022, pp. 603-608, doi: 10.1109/ICAAIC53929.2022.9793185.

[19] Ravindhar, N. V., Anand, H. S., & Ragavendran, G. W. (2019). Intelligent diagnosis of cardiac disease prediction using machine learning. *International Journal of Innovative Technology and Exploring Engineering*, *8*(11), 1417-1421.

[20] Soumya, T. R., Manohar, S. S., Ganapathy, N. B. S., **Nelson, L.,** Mohan, A., & Pandian, M. T. (2022, September). Profile Similarity Recognition in Online Social Network using Machine Learning Approach. In *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 805-809). IEEE.