MOVIE RECOMMENDATION SYSTEM

ABSTRACT:

A personalized recommendation system analysis user preferred movie name to provide personalized suggestions. This helps users discover new movies that align with their interests, increasing their overall satisfaction with the movie-watching experience. Recommendation systems can introduce users to movies they might not have otherwise discovered. A recommendation system simplifies the process by presenting users with a curated list of relevant movies, saving them time and effort. Movie recommendation systems also bring advantages to businesses. By providing personalized recommendations, platforms can enhance customer loyalty, attract new users. We are going to use Naive Bayes algorithm to implement this recommendation system. Naive Bayes is a simple and probabilistic machine learning algorithm based on Bayes theorem, which calculates the probability of an event based on prior knowledge. The "naive" assumption in Naive Bayes refers to the assumption of feature independence, meaning that each feature is considered to be independent of the others, given the class label. Naive Bayes is a machine learning algorithm that can be used in movie recommendation systems due to its simplicity, efficiency, and effectiveness in certain scenarios. In movie recommendation systems, where there can be a vast number of movies and users, Naive Bayes can process the data quickly and make predictions in real-time. In movie recommendation systems, this can help explain why certain movies are recommended to a user based on the probabilities associated with specific features or attributes. The movie recommendation system consists of three main components: data collection and pre-processing, recommendation algorithm, and user interface. The data collection and pre-processing phase involves gathering movie-related data such as genre, cast, director, user ratings, and reviews. This data is then pre-processed to extract relevant features and create a comprehensive movie database. Evaluation of the movie recommendation system is crucial to assess its performance. Common evaluation metrics include precision, recall, accuracy, and mean average precision. These metrics measure how well the system predicts movies that the user is likely to enjoy, as well as its ability to retrieve relevant movies from a large database.

KEYWORDS: Naïve Bayes, cosine similarity, movie recommender system, user preference, movie genre prediction, data mining.

INTRODUCTION:

In the recent days of covid, people are restricted to home and prohibited in going for theatres. As OTT that provides television and film content over the internet at the request and to suit the requirements of the individual consumer, recommendation system is much needed in the OTT platforms. It can be difficult to choose what to watch with so many films accessible today. Movie recommendation systems examine user preferences, viewing history, ratings, and other pertinent information. These algorithms help people save time and effort by suggesting films that are specific according to their tastes. So, movie recommendation system plays a crucial role in the OTT (over-the-top). Users are introduced to movies by movie recommendation systems that they might not have otherwise seen. This broadens user perspectives for movie-watching by introducing them to different media and genres that they might like. Movie recommendation systems can be advantageous for streaming services, movie theatres, and other companies in the film industry. These technologies aid in keeping customers, increase customer happiness, and potentially increase income by making individualised recommendations. Additionally, information gathered from user interactions with the recommendation system can assist in informing commercial choices and marketing plans. In order to achieve this recommendative system model we are going to use naïve-bayes supervised algorithm. A technique based on supervised machine learning called Naive Bayes is frequently used for classification problems. It is based on the Bayes theorem, which determines the likelihood that an event will occur given the likelihood that related events will also occur. The features are thought to be independent of one another according to naive Bayes. Naive Bayes is a lightweight algorithm that can handle large datasets and high-dimensional feature spaces efficiently. By leveraging the Naive Bayes algorithm, the recommendation system can classify movie reviews into positive or negative sentiments, enabling personalized recommendations based on user preferences and sentiments. Naïve Bayes algorithm in a movie recommendation system can help infer the user's choice for movies based on their characteristics and ratings, enabling the system to propose specific movies that match their interests and preferences. Naive Bayes can be useful in addressing the "cold start" problem in movie recommendation systems. The cold start problem occurs when there is limited or no data available for a new user or a new movie. Naive Bayes is a simple yet powerful machine learning algorithm based on Bayes' theorem. It is primarily used for classification tasks, including text classification, spam filtering, sentiment analysis, and recommendation systems. Cosine similarity can be used in a movie recommendation system to gauge how similar two films are based on their feature vectors or content qualities. A movie recommendation system can find movies with identical features or characteristics to the movies a user has expressed interest in by using cosine similarity. Instead of depending exclusively on collaborative filtering or user ratings, it assists in providing personalised suggestions based on content similarities. Cosine similarity is a flexible metric that can be used in conjunction with Naive Bayes method to improve the precision and applicability of movie recommendations for viewers. Convolutional Neural Networks are generally made for computer vision and object recognition jobs that need image processing. Movie posters are an important visual element that shapes viewers' first impressions and can provide important details about the genre, style, or theme of a film. By extracting visual features and learning representations that capture their visual content, CNNs can be used to analyse movie posters. In order to provide suggestions for films based on their content, these representations can then be utilised to assess how similar or pertinent movie posters are to one another.

LITERATURE SURVEY:

Harsh Khatter et.al [1]**-** Providing precise and personalized recommendations by analysing the thoughts of people using Cosine similarity and API Calls.

P. Li et.al [2] – To quickly produce high quality recommendations, even for very large-scale information resources used decision tree classification.

G. Adomavicius et.al [3] – In this paper a wide range of work is reviewed in the field of a recommender system for movies where dataset source, methods used and accuracy are compared to deduce best one
and future scope for improvement in this area is analysed by means of KNN, Content-based Filtering, Collaborative Filtering, Matrix Factorization, Hybrid Clustering.

R. Lavanya et.al [4] – Organizing all the significant flaws found in the most widely used commercial movie recommendation systems using Hybrid system of Collaborative filtering and Content based approaches.

Bodgen Walek et.al [5] – Based Recommender System for Online Stores Using Expert System using Content-based
Filtering.

Minjae Kim et.al [6] – Categorizing consumers with similar movie tastes, testing user similarity based on movie rating data, and learning the consumption patterns of each similar user group to predict and recommend by Recurrent neural network and collaborative filtering algorithm.

Ashrf Althbiti et.al [7] – Addressing Data Sparsity in Collaborative Filtering Based Recommender Systems Using Clustering and
Artificial Neural Network.

C. Christakou et.al [8] – Constructing a movie recommendation system merging content and collaborative information by Semi-supervised Clustering.

S.Sathiya Devi et.al [9] – Constructed a new hybrid method using Naïve Bayesian Classifier along with Gaussian correction and feature engineering.

Debashish Roy et.al [10] – Integrating trailer data as movie features using sentimental scores derived from trailer comments as rating matrix and integrate with rating data treating as features by Matrix Factorization and Deep
Neural Network Models.

Mohammed Nazim Uddin et.al [11] – The weight of the movies has a significant impact on the films. Weighted movie ratings are obtained by strongly correlating with user information or preferences from their activities by a formula. Recommending movies with the help of this weighted value by the use of Kmeans algorithm. And then using those weighted values to create movie clusters. The customer is then advised to watch the movie with the highest mean movie rating.

Sajal Halder et.al [12] – Finding the group of a user depending on the movie genre by movie swarm mining, frequent item mining.

Bagher Rahimpour Cami et.al [13] – Performing modelling and construction phases of proposed content-based movie recommender system by Naïve Bayesian framework

Muppana Mahesh Reddy et.al [14] - Analysis of Movie Recommendation Systems; with and without considering the low rated by Pearson correlation coefficient, Logistic Regression.

Pooja Jain et.al [15] – Computing similarity between the different movies in the given dataset efficiently and in least time and to reduce computation time of the movie recommender engine using cosine similarity measure by Content-Based Filtering and Collaborative Filtering.

Publication	Author	Technique	Task	Dataset	Advantages
IEEE 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)	Harsh Khatter et al [1] 2021	Cosine similarity and API Calls.	Providing precise and personalized recommendations by analyzing the thoughts of people.	Movie dataset.	Delivers the user with precise, effective, and personalized recommendations.
IEEE – Conference on Cybernetics and Intelligent Systems, 2004.	P. Li et al [2] 2004	Decision tree classification	To quickly produce high quality recommendations, even for very large-scale information resources.	Movie datasets, review texts.	Inspects the effectiveness of technology that produce high quality recommendations.
IEEE - Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)	G. Adomavicius et al [3] 2019	KNN, Content-based Filtering, Collaborative Filtering, Matrix Factorization, Clustering, Hybrid	In this paper a wide range of work is reviewed in the field of a recommender system for movies where dataset source, methods used and accuracy are compared to deduce best one and future scope for improvement in this area is analyzed.	Movie Lens dataset	Enhances the overall user experience on movie platforms, leading to increased user engagement and potentially higher revenue.

IEEE survey on International Conference on Artificial Intelligence and Smart Systems (ICAIS) 2021	R. Lavanya et al [4] 2021	Hybrid system of Collaborative filtering (CF) and Content-based approaches	Organizing all the significant flaws found in the most widely used commercial movie recommendation systems.	Movie datasets	Solves problems by analyzing scalability, the cold start problem, data sparsity, realistic usage feedback and accuracy and temporal complexity.
IEEE - 2018 IEEE First International Conference on Artificial Intelligence And Knowledge Engineering (AIKE)	Bodgen Walek et al [5] 2018	Content-based filtering	Based Recommender System for Online Stores Using Expert System.	User profile data, item features	Users receive personalized product recommendations that align with their interests. Increases the sales production of the less popular products.
2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)	Minjae Kim et al [6] 2019	Recurrent neural network and collaborative filtering algorithm	Categorizing consumers with similar movie tastes, testing user similarity based on movie rating data, and learning the consumption patterns of each similar user group to predict and recommend.	Movie rating data	Compares accuracies of predictions of RNN and collaborative filtering Algorithm.

IEEE - 2021, 11th Annual Computing and Communication Workshop and Conference (CCWC)	Ashrf Althbiti et al [7] 2021	Clustering and Artificial Neural Network Based Collaborative Filtering	Addressing Data Sparsity in Collaborative Filtering Based Recommender Systems Using Clustering and Artificial Neural Network	Four different datasets from four popular domains (Books, music, jokes, and movies)	CANNBCF effectively solves the sparsity issue, improves the quality of recommendations, and demonstrates promising prediction accuracy
IEEE survey on International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce	C. Christakou et al [8] 2005	Semi-supervised Clustering	Constructing a movie recommendation system merging content and collaborative information.	Movie lens dataset	Yielding recommendations of high accuracy

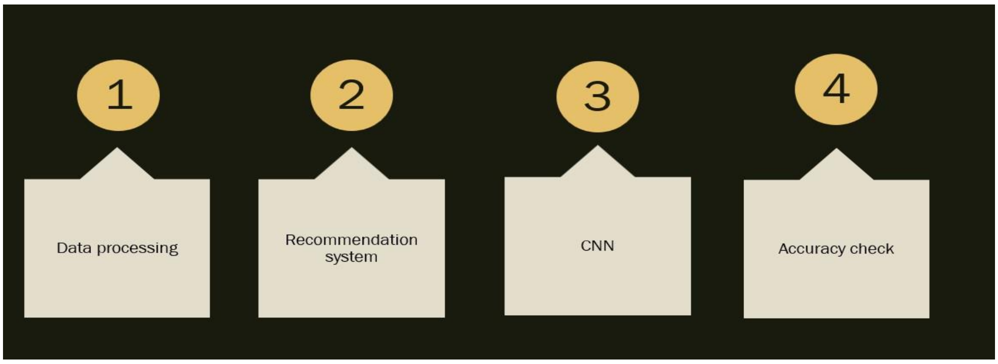
IEEE 2018 Second International Conference on Inventive Communication And Computational Technologies (ICICCT)	S.Sathiya Devi et al [9] 2018	Hybrid approach system using Feature Engineering	Constructed a new hybrid method using Naïve Bayesian Classifier along with Gaussian correction and feature engineering	Movie lens 100K data set	Improved prediction performance and better results in comparison with others.
2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)	Debashish Roy et al [10] 2020	Matrix Factorization (MF) and Deep Neural Network (DNN) Models	Integrating trailer data as movie features using sentimental scores derived from trailer comments as rating matrix and integrate with rating data treating as features	YouTube Movie Trailer Data as the Side Information	Accurate result improved as trailer feedback data is integrated as movie features.
2017 International Conference on Information and Communication Technology Convergence (ICTC)	Mohammed Nazim Uddin et al [11] 2017	Weight based movie recommendation system using K-means algorithm	The weight of the movies has a significant impact on the films. Weighted movie ratings are obtained by strongly correlating with user information or preferences	Users' data from the user's activity	Efficient way to suggest

			from their activities by a formula. Recommending movies with the help of this weighted value by the use of k-means algorithm. And then using those weighted values to create movie clusters. The customer is then advised to watch the movie with the highest mean movie rating.		
2012 Second International Conference on Cloud and Green Computing	Sajal Halder et al [12] 2017	Movie swarm mining, frequent item mining	Finding the group of a user depending on the movie genre.	Movie dataset	Delivers and find the popular genres of movies at timestamps t.
2017 3rd Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)	Bagher Rahimpour Cami et al [13] 2017	Naïve Bayesian framework	Performing modeling and construction phases of proposed content-based movie recommender system	Movie lens dataset	improved the accuracy of movie recommendation.

2020 International Conference on Emerging Trends in Information Technology and Engineering (ie-ETITE)	Muppana Mahesh Reddy et al [14] 2020	Pearson correlation coefficient, Logistic Regression	Analysis of Movie Recommendation Systems; with and without considering the low rated	Movie-Lens-100K.	User can get to know what are the low rated movies liked by .and comparison of high rated and low rated movies.
International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017	Pooja Jain et al [15] (2017)	Content-Based Filtering and Collaborative Filtering	Computing similarity between the different movies in the given dataset efficiently and in least time and to reduce computation time of the movie recommender engine using cosine similarity measure.	1.Movie Lens 1M dataset 2. Movie Lens Latest Small dataset 3. Movie Lens 10M dataset	improving the scalability and quality of the movie recommendation system

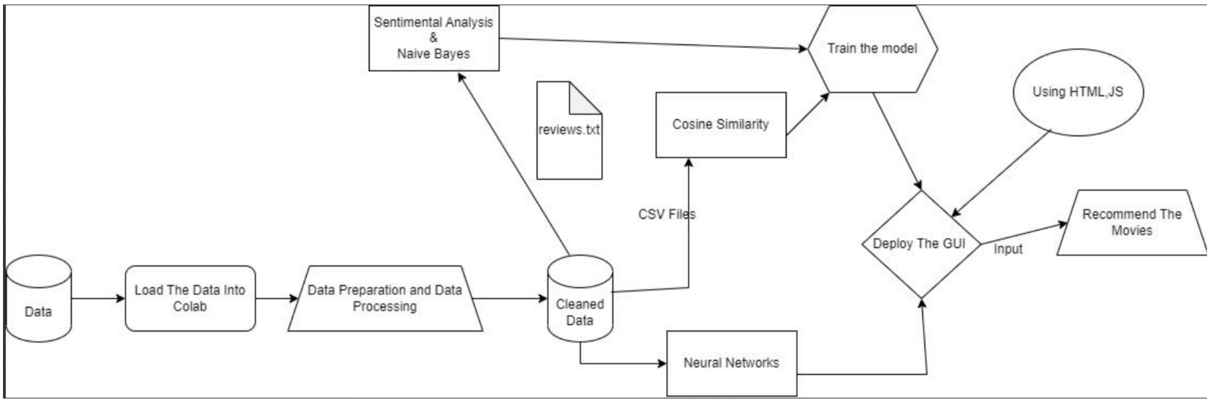
MATERIALS AND METHODS:

With the abundance of movies available today, users often face difficulty in finding movies that align with their preferences. The sheer volume of options can lead to decision paralysis and a frustrating user experience. Movie recommendation systems address this problem by filtering and suggesting movies based on user preferences, saving time and effort in searching for relevant content.



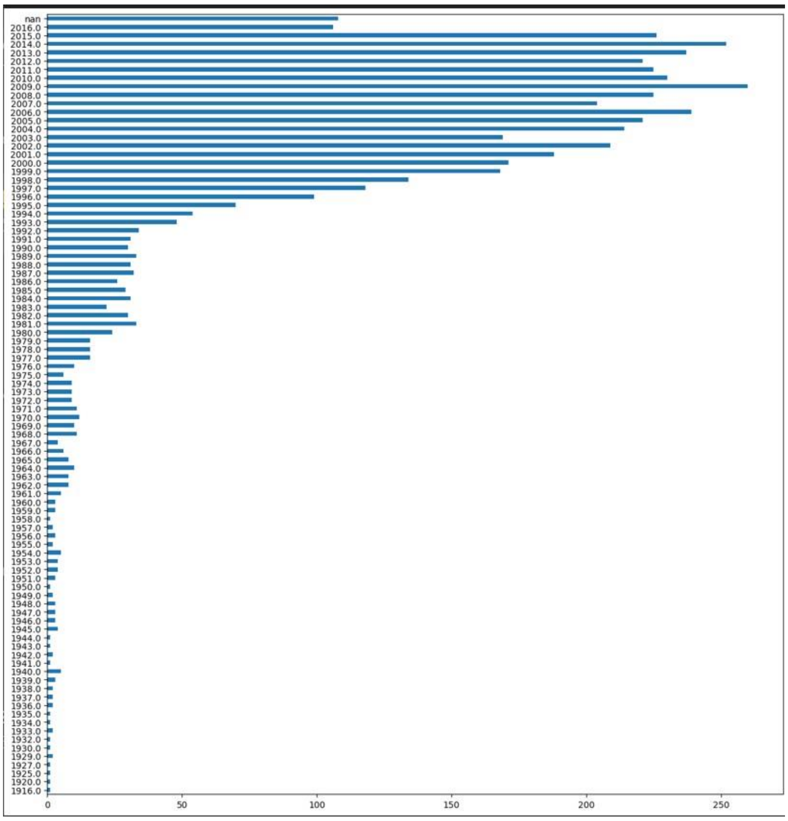
The data processing module of a movie recommendation system is responsible for collecting, organizing, and pre-processing movie-related data to facilitate accurate and efficient movie recommendations. This module plays a crucial role in extracting relevant features from the data and creating a comprehensive movie database. The system gathers movie-related data from various sources such as streaming platforms, movie databases, user ratings, reviews, and external APIs. The data may include movie titles, genres, cast and crew information, release dates, user ratings, reviews, and other relevant attributes. Data Cleaning the collected data is often noisy and inconsistent. Data cleaning involves removing duplicate entries, handling missing values, correcting errors, and standardizing the data format. This ensures the quality and integrity of the dataset. The movie recommendation system needs to handle regular updates to keep the database current and reflect the evolving movie landscape. This involves implementing mechanisms to update the database with new movie releases, user ratings, and reviews. The extracted features are transformed into a suitable representation format for the recommendation algorithms. This could involve converting categorical data into numerical representations, creating vectors to represent movie attributes, or using matrix representations for collaborative filtering algorithms. The recommendation system plays a central role in a movie recommendation system by generating personalized and relevant movie suggestions, filtering options based on user preferences, promoting movie discovery, encouraging exploration, and continuously learning from user interactions to enhance the overall user experience. It utilizes advanced algorithms and techniques to analyse user preferences, movie attributes, and historical data to generate personalized movie recommendations. The recommendation system aims to provide personalized movie suggestions to individual users based on their unique preferences, viewing history, and behaviour. With a vast number of movies available, users may feel overwhelmed by the sheer volume of options. The recommendation system filters the movie database based on user preferences, eliminating irrelevant or less likely appealing movies. Convolutional Neural Networks (CNNs) are primarily used in image recognition and computer vision tasks. However, in the context of a movie recommendation system, CNNs can be employed for certain aspects of the recommendation process. Movie posters contain valuable visual information that can influence a user's decision to watch a movie. CNNs can be used to analyse and extract features from movie posters. By training a CNN on a large dataset of movie posters, it can learn to recognize visual patterns and attributes such as colours, composition, genre-specific elements, and actor/actress appearances. These extracted features can then be used to recommend movies with similar visual characteristics to users who have shown interest in specific movie posters. CNNs can aid in content-based filtering by analysing visual attributes of movies. By training a CNN on frames or clips from movies, it can learn to extract visual features that are indicative of genres, settings, or visual styles. Accuracy check in a movie recommendation system is essential to evaluate the performance and effectiveness of the recommendation algorithms. It helps measure how well the system predicts movies that users are likely to enjoy and assesses its ability to retrieve relevant movies from a large database. These metrics can be computed using test datasets, where the true user preferences are known. The evaluation process involves comparing the recommended movies with the actual preferences of the users to assess the accuracy of the system. It is important to consider a diverse set of users and their preferences to ensure the accuracy evaluation is representative of the system's performance across different user profiles.

ARCHITECTURE DIAGRAM:

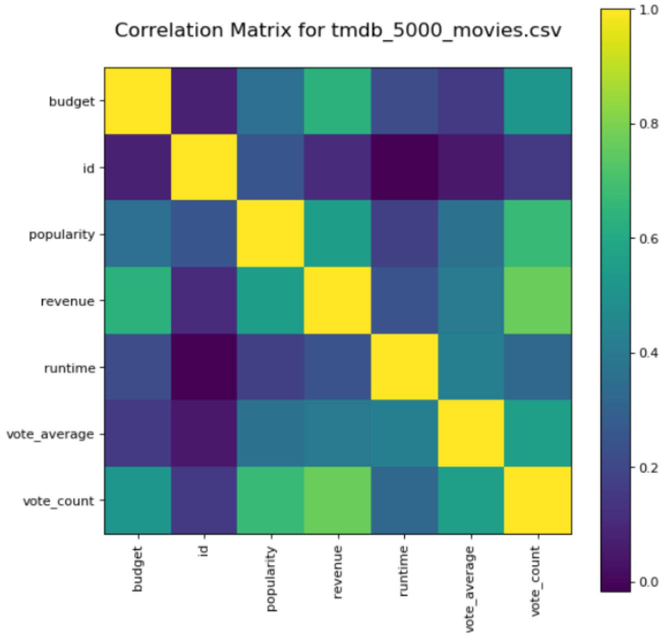


RESULTS AND DISCUSSIONS:

Naive Bayes is a probabilistic algorithm that relies on Bayes' theorem to make predictions. In the context of movie recommendation, Naive Bayes can be used to predict whether a user will like a particular movie based on their previous preferences and the features of the movie.



Whereas Content-based recommendation systems focus on the characteristics and attributes of items themselves, rather than relying on user preferences.



CONCLUSION:

In this prediction we can easily get suggestion of a movie with our interest genres in movie. In conclusion, a movie recommendation system is a powerful tool that addresses the challenges of information overload, personalization, discoverability, user retention, revenue generation, and efficient utilization of user data in the movie-watching landscape. By leveraging data collection, pre-processing, recommendation algorithms, and user interfaces, these systems enhance the user experience by providing personalized and relevant movie suggestions. The recommendation system plays a central role in a movie recommendation system by generating accurate and tailored movie recommendations based on user preferences and movie attributes. It filters movie options, promotes movie discovery, encourages exploration, and continuously learns from user interactions to improve recommendation accuracy over time. These systems enhance user engagement, increase platform retention, and drive revenue by suggesting movies that align with individual user taste.

REFERENCE:

H. Khatter, N. Goel, N. Gupta and M. Gulati, "Movie Recommendation System using Cosine Similarity with Sentiment Analysis," *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2021, pp. 597-603, doi: 10.1109/ICIRCA51532.2021.9544794.

P. Li and S. Yamada, "A movie recommender system based on inductive learning," *IEEE Conference on Cybernetics and Intelligent Systems, 2004.*, Singapore, 2004, pp. 318-323 vol.1, doi: 10.1109/ICCIS.2004.1460433.

R. Lavanya, U. Singh and V. Tyagi, "A Comprehensive Survey on Movie Recommendation Systems," *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, India, 2021, pp. 532-536, doi: 10.1109/ICAIS50930.2021.9395759.

B. Walek and P. Spackova, "Content-Based Recommender System for Online Stores Using Expert System," *2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, Laguna Hills, CA, USA, 2018, pp. 164-165, doi: 10.1109/AIKE.2018.00036

. M. Kim, S. Jeon, H. Shin, W. Choi, H. Chung and Y. Nah, "Movie Recommendation based on User Similarity of Consumption Pattern Change," *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, Sardinia, Italy, 2019, pp. 317-319, doi: 10.1109/AIKE.2019.00064

Althbiti, R. Alshamrani, T. Alghamdi, S. Lee and X. Ma, "Addressing Data Sparsity in Collaborative Filtering Based Recommender Systems Using Clustering and Artificial Neural Network," *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, NV, USA, 2021, pp. 0218-0227, doi: 10.1109/CCWC51732.2021.9376008.

P. D. Yoo, M. H. Kim and T. Jan, "Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation," *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, Vienna, Austria, 2005, pp. 835-841, doi: 10.1109/CIMCA.2005.1631572.

Shubhada Labde, Vishesh Karan, Shubham Shah, Dhruv Krishnan, "Movie Recommendation System using RNN and Cognitive thinking", *2023 4th International Conference for Emerging Technology (INCET)*, pp.1-7, 2023

O. Chunhapran, C. Phromsuthirak, M. Hama and M. Maliyaem, "Movie Recommendation System Using Director-Based," *2022 26th International Computer Science and Engineering Conference (ICSEC)*, Sakon Nakhon, Thailand, 2022, pp. 151-155, doi: 10.1109/ICSEC56337.2022.10049347.

M. Akter, N. Sultana, M. F. Hossain and S. R. H. Noori, "Factors that has Effect on Early Movie Recommendation: A Systematic Literature Review," *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2021, pp. 1-4, doi:10.1109/ICCCNT51525.2021.9579861.

R. Suguna, P. Sathishkumar and S. Deepa, "User Location and Collaborative based Recommender System using Naive Bayes Classifier and UIR Matrix," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1-5, doi: 10.1109/ICECA49313.2020.9297589.

A. Sinha, P. Mishra, M. Ramish, H. R. Mahmood and K. K. Upadhyay, "Employing Unsupervised Learning Algorithm for Stock Market Analysis and Prediction," 2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT), Meerut, India, 2021, 79, doi:10.1109/ICACFCT53978.2021.9837372

M. E. Karim and S. Ahmed, "A Deep Learning-Based Approach for Stock Price Prediction Using Bidirectional Gated Recurrent Unit and Bidirectional Long Short Term Memory Model," 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2021, pp. 1-8, doi: 10.1109/GCAT52182.2021.9587895.

M. Wu, "Financial Transaction Forecasting using Neural Network and Bayesian Optimization," 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Fuzhou, China, 2021, pp. 193-196, doi: 10.1109/CEI52496.2021.9574535.

NAME: D.RUSHMA

Gmail: rushma.dodda@gmail.com