**Movie Recommendation System with Machine Learning**

**Student name**

**Reg. Number**

# Abstract

Movies are a form of entertainment to enjoy and refresh one’s mind away from the busy schedule. This need has led to development of the industry from actors to video production professionals. Movie streaming platforms have also been on the rise making billions of dollars by providing entertainment conveniently. Competition is on the rise with the streaming platforms trying to stay on top of their game to retain viewers who directly correlate with profits. These is a huge market and getting to know the preferences of users is a huge step in the right direction to help movie streaming giants to retain customers. With the large chunks of user data available analysis can be done to make the appropriate decisions that maximize on profit and keep the users entertained. The main insights are the time spent by different users, the genre and the ratings of the movies only to name a few. These insights can bring about impactful change and drive great profits as each user is provided with what he/she consumes rather than a catalogue of movies which may be exhaustive to search through and still not end up to watch a movie. These research focuses on building a machine and deep learning movie recommendation system to provide the streaming giants with a powerful system for user retention. The system will make movie recommendations based on the user history or previous watch-list and employ machine learning models to predict which movies the user may be interested in. The research will also factor in some inputs from the user such as length of a movie, genre, ratings and favorite actors. These questions can be asked prior to try and make our model more effective in predicting the movie which the user is most likely to watch. The system will also include an element of recommending based on friend’s preferences. These can come in handy especially if the taste in movies is the same with the friends. The report outlines the steps taken and the machine learning models that have been used to make the movie recommendation system a success.

**Keywords**

Machine learning, content-based filtering, Regression, Support vector machine, recommendation system, Algorithms.

# Table of Contents

[Abstract 2](#_Toc149892058)

[Table of Contents 4](#_Toc149892059)

[Chapter 1 5](#_Toc149892060)

[Introduction 5](#_Toc149892061)

[Research Background 6](#_Toc149892062)

[Research aim 6](#_Toc149892063)

[Research objectives 6](#_Toc149892064)

[Chapter 2: 8](#_Toc149892065)

[Literature review 8](#_Toc149892066)

[References 15](#_Toc149892067)

# Chapter 1

# Introduction

The age of new technology has brought about change in form of entertainment and consumption of media. The rapid development has made it possible for more creatives to make content and provide it for consumption to the users. Movies are a form of rich media with lots of them being produced every day. This has led to the development of movie streaming platforms such as Netflix and Hulu which provide platforms to watch these movies and be entertained by them. However, the nature of the world is that everyone is diverse and don’t take interest in the same form of media. Provision of a catalogue of movies which one can choose from without recommendation can stagger growth. This is due to the fact that some users may get bored not because it is a bad movie or media but due to the fact that it doesn’t pick their interest. Recommendation systems come in handy to solve this problem and provide users a way to explore the catalogue by choosing what is interesting to them. It leads to user retention which directly correlates to revenue as the more users stream the more the company gets revenue based on their revenue model. A recommendation system is a complex program that uses data and machine learning algorithms to predict an outcome based on previous learned behavior that are driven by data (Kang et al., 2019). Recommendation systems come in handy and can be tweaked based on users and regions to maximize on their efficiency. They become effective with more usage and data for learning. They can make sense of the large datasets or catalogue which to the user may seem daunting to sift through and provide what the user needs.

Recommendation systems complement search functionality in an app and can provide the highest rated movies which is geared towards improving and ensuring customer satisfaction. The results are retention of customers and seamless user experience of viewers with long hours of viewing which ensures profits are made (Furtado & Singh, 2020). There is a huge difference in sales when there is use of a recommendation system and where one is not used. Therefore, this approach is futuristic and is geared towards ensuring a win situation. We are using the TBDM movie database data as our movie dataset. It consists of a myriad of movies which we can play around with and try to predict which movie is best based on provided metrics of a user.

# Research Background

Traditional movie recommendation systems make use of ratings only to suggest movies. These are only based on user feedback and can’t narrow down to what a user may need. A movie can be highly rated but not interesting to a specific user. These recommendation systems only factor in what’s popular and think it’s a one size fits all situation. Other metrics can be added as a way to provide a more accurate prediction. These metrics include length, genre, ratings and starring actors all which affect the choice of a movie. These metrics are given weights so as to influence the final suggestions based on our machine learning algorithm.

# Research aim

* To create a movie recommendation system using machine learning and deep learning algorithms
* Research on common pitfalls and traditional recommendation systems to make an improvement on them

# Research objectives

1. To build a movie recommendation system using machine learning algorithms
2. To develop a machine learning model that performs analysis and recommends a movie based on watch history and other metrics
3. To research on available movie recommendation systems and ways to improve them to be more effective and implement in my model

# Research deliverables

A machine learning model which predicts movies based on watch history and other metrics. The model will be written in python and can be integrated with any backend system of a movie streaming platform. A technical paper outlining the approach from problem statement to design and implementation of the system.

# Chapter 2:

# Literature review

Wu, Garg, & Bhandary (2018) proved on their research that recommendation systems drive user and business decisions and have proven to be an integral part of movie streaming business. They provide information to the user on the best movies based on their taste by gauging the similarity. The similarity could be based on the cast available, genre, release date and even up votes count. These systems have been used to personalize the movie experience of the users increasing overall satisfaction and user retention within the systems.

Research done by Lee et al. (2022) indicated that a content-based approach favored least popular movies than the in-demand ones and a more user centric or preference approach was to be followed. Traditional and most common approach used was to recommend the popular ones. This approach seemed futile as even though the movies may be popular, they are not indicative of personal preference. Interacting with relevant content based on one’s taste ensures that they are engaged and stay within our platform. This strategy has proven to be the leading source of revenue for the tech giants with 30 to 35% of their revenue coming from recommendation systems. Content based movie recommendation systems are based on what the user is watching. They take note of the users watch history. The recommendation system will then recommend similar movies to the ones that have been previously watched. The recommendation system recommends movies based on similarity score may it be by actors or the genre of the movie which the person was watching. Collaborative movie recommendation systems are based on the user’s past data. They take a cluster of users who have shown similarity and recommend movies which are similar but haven’t been watched by the current user. This method is effective in grouping users and giving recommendations based on what others are watching. Data is stored in a collaborative matrix which is indicative of the movies a user has watched and similar movies are suggested as long as they haven’t been watched by the current user. This approach clusters users together and certain patterns are used to determine which movie is going to be recommended based on the user groups.

The paper “Movie recommendation system with sentimental analysis using cosine similarity technique” by Javed et al. (2022) focuses on the use of cosine similarity after factoring in both user ratings of the movie and emotions to provide a tailored recommendation. This recommendation is based on user ratings and also experience with the movie which is a more effective approach. Cosine similarity is a concept that relates two variables and is bound within a range of 0 to 1. When the value is close to zero it implies orthogonality and hence less similarity between the entities. If the values are close to a range of 1 then it implies that they are similar to each other and their angle in vector space is acute. This concept is applied in movies dataset within the matrix and after cosine similarity based on a particular threshold provided. With the above techniques explained we provide room for accurate predictions of the movie that a user or user groups would be interested in. The clustering of users into accurate user groups also ensures lesser computing operations as there is similarity among users as movies are based on genres. Genres are the basis of movies and therefore used as a metric in grouping and provide a great basis of clustering users. This is based on if two users have similarity and one has watched this movie and the other hasn’t then we can recommend the movie to the one who hasn’t watched. We conclude that they may be interested in the movie. Traditional recommendation systems only consider the user ratings of the movies as a way of providing recommendation. This is a technique based on popular rating and doesn’t consider the experience and the opinion of the individual user about a movie. A weighted approach that gives weight to all the factors mentioned is a great approach and a step towards making movie recommendation systems even more effective (Kumar, De & Roy, 2020). Taking users opinions and also other metrics as how many users quit the movie midway can be a great indicator of the quality of a movie and weighted also as a way to improve the overall recommendation. These emotions or metrics can be obtained in machine learning by performing sentiment analysis. We can then use the technique of cosine similarity to determine the relationship between the movies and determine which movie to recommend.

Jayalakshmi et al. (2022) in their paper “Concepts, challenges and future directions of movie recommendation systems” highlight on the use of appropriate filtering techniques to match movies to a higher degree. They focused on various machine learning algorithms that could be used to design and implement movie recommendation systems. Emphasis is given on metaheuristic algorithms which comprise of algorithms from simple search to complex learning processes. They also highlight on the challenges that have been faced by people making movie recommendation systems and ways to overcome such challenges. They introduce the concept of hybrid filtering which is a combination of content and collaborative filtering. It entails use of both algorithms to better predict the best movie to suggest to a user. Hybrid filtering fills in the gaps in collaborative filtering where the recommendation system lacks information on domain dependencies. It fills in the gaps in content-based filtering where the recommendation system lacks the preferences data of the users to recommend the movie to. The use of both user data and content data in hybrid filtering is a step forward to building effective movie recommendation systems. They focus on high level metaheuristic algorithms from genetic algorithms to PCA which have proven to be high performance algorithms in solving optimization problems with the aim of enhancing similarity. They highlight the challenges faced in movie recommendation systems such as cold start. Cold start is imminent as the algorithms used data from previous user interaction. This is mainly faced in collaborating where the algorithms are based on user data. Absence of user data means no recommendations can take place since collaborative algorithms are based on user data. They provide a solution of using a content-based approach when there is absence of user data and reverting back to collaborative approach after obtaining user data from the interactions with the movie streaming platform. They emphasize on the hybrid approach which uses context filtering when the user is new such as device location and operating system and correlates them with other users so as to provide the first recommendation before the user interacts with the movie streaming platform. Accuracy also is a challenge and is inversely related with the size of the database. However, this is dependent on the type of algorithm used and the search space gets wider with a large dataset of movies which will affect the accuracy. Employment of sophisticated search criteria could improve the accuracy. The paper exhaustively discusses the techniques and challenges with solutions provided as stipulated above.

Rimaz et al. (2019) on their paper “Exploring the power of visual features for recommendation of movies “explore and analyze the power of using visual features to try and predict the appropriate movie to recommend to a user. The features can be explored without user interaction and repetitive analysis of the visual content has proven effective in recommending movies. This method is futuristic and can eliminate the challenges of lack of user interactive data which is associated with collaborative filtering (Rastogi et al., 2022). They point out on the dependency of collaborative and content filtering movie recommendation systems on availability of large datasets and their accuracy also depend on them. They highlight that though semantic analysis on user attributes is key it is not as important as visual cues on implementation of effective recommendation systems. They use exploratory analysis to gather data on movie visual cues which is done by direct analyzing the movies. The backbone of their research is use of visual features to represent movie content as opposed to attributes such as user interaction and genres. The research is based on previous research that there is a close relationship of visual features between movie trailers and the movies and have proven to provide effective recommendation based on visual features. “We have built a pure Content-Based recommender system (CB), which relies solely on semantic item features attributes, i.e genre, tags, or visual features, as well as a similarity metric. The similarity metric is used to measure the similarity among items. Then a model is built, based on the user preferences, exploited to learn the taste of a target user and to recommend to her items that are similar to those that she liked in the past. We have used one of the most common similarity metrics, the Cosine similarity. As baselines, we used the genre and tag attributes.” The models built from each of the baseline attributes uses its own similarity matrix and then uses K-Nearest neighbor to compute the similarity matrix and produces a score which can be used to gauge similarity and provide appropriate recommendation. They conclusively explored and experiment use of visual features which turned out to be effective in the movie recommendation. The visual features provide more accurate results as some movies could be from a genre that is liked by a user but not appropriate for the particular user. Their paper proves futuristic with solving challenges of cold start and scalability as traditional recommendation systems tend to be less accurate and poorly scalable when handling large data sets. Elimination of use of large data sets is a step into the future and picking up on this could prove to be a step forward. The challenge in this approach remains to be development of effective user interface to accommodate such type of development and also the capturing of user expression and emotions to better understand if the movie recommendation system is effective (Anjum, 2022). In general, it is a better and more impactful and cost-efficient approach as it involves minimal user interaction and improves automation hence could be an out of the box solution for new users or even platforms that are starting out and haven’t captured much user data to employ in their algorithms.

Kumar et al. (2021) in their paper “A movie recommendation system: MOVREC” focus and dig deep on collaborative filtering approach. It makes use of information provided by users to try and provide accurate recommendations. This can range from preferred genre to age and other user related attributes that are relevant to movie recommendation. The recommendation is based on previous user ratings and is sorted in the list through K-means algorithm. The top user rated movies are the ones most likely to be recommended. Their paper highlights the importance of movie recommendation systems to generate revenue and also provide a more tailored experience. This experience is what keeps users around and therefore has proven to be effective in user retention and finding the right type of movie to watch in a large dataset of movies. They highlight a case study of amazon whose sales increased by up to 35% when they started using a recommender system. This is an upward trajectory which was mainly impacted by tailoring the user experience to suit the needs of each user. The unique approach used in their research is using user feedback as input to try and learn more about the user for future recommendations. This factoring in of user feedback is vital and could solve the issue of cold start quicker with weighted inputs in terms of feedback. In future predictions as per the context of the recommendation system the system will be able to provide a more accurate or closer suggestion which suits the particular user. However, their approach is based on previous user ratings and wouldn’t provide a more current user centric approach. It makes an assumption that if others liked the movie, we are more likely to like the movie too. The K means cluster algorithm ranks the ratings and the movies with the top ratings are suggested first. If movies have the same rating, then the one with more votes is given priority. The advantage of this research is provision of user interface for input which affects weighting of which movies to recommend. Even though ratings are previously obtained from users who’ve watched the movie, the user has the liberty to key in some important metrics which will be applied to the algorithm. These range from length of movie, genre and even actors which affect the movie to be recommended. The user feedback also plays a critical role and impacts future suggestions hence it gives the user control over what the user wants to see but not in an automatic fashion compared to the visual features method (Jozani, Liu & Choo, 2023). This method though traditional is an improvement to the predecessors and opened the way to more complex and complicated algorithms for movie recommendation.

# References

Anjum, I. (2022). Movie Recommendation System Using Collaborative Filtering.

Furtado, F., & Singh, A. (2020). Movie recommendation system using machine learning. *International journal of research in industrial engineering*, *9*(1), 84-98.

Javed, A., Rehman, F., Sarfraz, N., Sharif, H., Khan, R., & Khan, A. M. (2022, December). Movie Recommendation System with Sentimental Analysis Using Cosine Similarity Technique. In *2022 3rd International Conference on Innovations in Computer Science & Software Engineering (ICONICS)* (pp. 1-8). IEEE.

Jayalakshmi, S., Ganesh, N., Čep, R., & Senthil Murugan, J. (2022). Movie recommender systems: Concepts, methods, challenges, and future directions. *Sensors*, *22*(13), 4904.

Jozani, M., Liu, C.Z. and Choo, K.-K.R. (2023) ‘An empirical study of content-based recommendation systems in Mobile App Markets’, Decision Support Systems, 169, p. 113954. doi:10.1016/j.dss.2023.113954.

Kang, K., Park, J., Kim, W., Choe, H., & Choo, J. (2019, November). Recommender system using sequential and global preference via attention mechanism and topic modeling. In *Proceedings of the 28th ACM international conference on information and knowledge management* (pp. 1543-1552).

Kumar, P., Kibriya, S. G., Roy, P. & Ajay, Y. (2021, May). Movie Recommender System Using Machine Learning Algorithms. In *Journal of Physics: Conference Series* (Vol. 1916, No. 1, p. 012052). IOP Publishing.

Kumar, S., De, K., & Roy, P. P. (2020). Movie recommendation system using sentiment analysis from microblogging data. *IEEE Transactions on Computational Social Systems*, *7*(4), 915-923.

Lee, H., Hwang, D., Min, K., & Choo, J. (2022, July). Towards Validating Long-Term User Feedbacks in Interactive Recommendation Systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2607-2611).

Rastogi, S., Agarwal, D., Jain, J., & Arjun, K. P. (2022, January). Demographic Filtering for Movie Recommendation System Using Machine Learning. In *Proceedings of International Conference on Recent Trends in Computing: ICRTC 2021* (pp. 549-557). Singapore: Springer Nature Singapore.

Rimaz, M. H., Elahi, M., Bakhshandegan Moghadam, F., Trattner, C., Hosseini, R., & Tkalčič, M. (2019, June). Exploring the power of visual features for the recommendation of movies. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 303-308).

Wu, C. S. M., Garg, D., & Bhandary, U. (2018, November). Movie recommendation system using collaborative filtering. In *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 11-15). IEEE.