CHAPTER 1
INTRODUCTION

1.1   Background

Movies have become our daily needs in terms of entertainment. It's no secret that several well-known platforms such as Netflix, Youtube, Disney+ Hotstar, and others have more or fewer millions and even tens of millions of films on each of these platforms. Therefore, the recommendation engine is at the heart of various movie provider platforms.

The number of these films is what causes a new problem, namely what film to watch next. Almost all platforms must collect data on movies watched by their users, the problem is if a platform still doesn't have data from that user or a new user. Then what are the movie recommendations that should be presented to the new user if a platform still doesn't have viewing data from that user or still has very little data from that user. Of course, users from each platform want to get movie recommendations that they think are most suitable for them, if not, then it's likely that these users will not use the platform anymore.

However, this problem has been solved with the recommendation engine. There are many algorithms used to create recommendation engines. Each algorithm has its advantages and disadvantages. Some algorithms recommend a film based on data between users, but there is also an algorithm that recommends also based on the relationship from one film to another. Then there is also a combined algorithm of the two, namely user-item, the user-item relation produces more maximum recommendations than relying on only one of the components above.

But of course, there are always shortcomings from the algorithms listed above, such as users who are new and don't have any viewing records. There is also an accuracy that is not optimal due to the low rating of each film. And in this study, the algorithm that I will use is item-based which uses data from ratings per film.

In this study, I tried to make 2 recommendation engines using 2 different algorithms, and then I will measure them concerning MSE and RMSE. And I will try this research with the cosine similarity approach to find the proximity between 1 object to another. From this research, it is expected to see which of the two algorithms has the least error.

1.2   Problem Formulation

1   Which of the Collaborative Filtering and Nave Bayes has the lowest MSE and RMSE values ?

2   What factors can be used as a reference to compare the two algorithms ?

1.3   Scope

The data used is the data that I downloaded from movielends.org. All the data that I use already have all the data records so there are no users who don't have data records at all.

1.4   Objective

Comparing the algorithm between Collaborative Filtering and Naïve Bayes then calculating the RMSE value, and predicting the films that will be recommended to users while maintaining the genre.

Comparing two different algorithms against the film recommendation engine and finding out which of the two algorithms is the most effective, based on the smallest MSE and RMSE values. In collaborative filtering, we use the approach through Cosine Similarity.