## **MOVIES RECOMMENDER SYSTEM IN R**

The system recommends the 5 movies to users based on user's watched movies genres. Cosine Similarity function is used to determine the similarity between the movies.

## **R** Code

```
movies <- read.csv('C:/Users/Taranpreet/Desktop/movies.csv')
ratings <- read.csv('C:/Users/Taranpreet/Desktop/ratings.csv')</pre>
  ratings_user_570 <- ratings[ratings$userId == 570,]
movie_ids <- movies[movies$movieId %in% c(18,231,592,914,1225,2000,3210,3801,54997,88812) ,]
#movie_ids <- movies[movies$movieId %in% c(5612, 943, 7358, 25, 2108, 8371, 2699, 8923, 2487, 102800) ,]</pre>
  x <- c("Action", "Adventure", "Animation", "Children", "Comedy", "Crime", "Documentary", "Drama", "Fantasy", "Film-Noir", "Horror", "IMAX", "Musical", "Mystery", "Romance", "Sci-Fi", "Thriller", "War", "Western", "(no
  genres listed)")
  colnames(df_user_prof) <- x
  colnames(df_movie_prof) <- x
x <- c("MovieId", "MovieName", "Similarity")
  x <- c("MovieId", "MovieName
colnames(df_similarity) <- x
  library(dplyr)
  movies_user_570 <- movies %>%
  filter(movieId %in% ratings_user_570$movieId)

df_movies_gen <- data.frame(do.call("rbind",strsplit(movies_user_570$genres, "|",fixed = TRUE)))

df_movies_570_gen <- data.frame(do.call("rbind",strsplit(movie_ids$genres, "|",fixed = TRUE)))
  x <- c(1:ncol(df_movies_gen))</pre>
  y <- c(1:nrow(ratings_user_570))
  vec <- numeric()
for(i in y){
   for(j in x){
      df_user_prof[i,df_movies_gen[i,j]] = 1
      df_movie_prof[2,df_movies_570_gen[2,j]] = 1
      df_movie_prof[2,df_movies_570_gen[2,j]] = 1
      df_movie_prof[3,df_movies_570_gen[3,j]] = 1
      df_movie_prof[4,df_movies_570_gen[4,j]] = 1
      df_movie_prof[5,df_movies_570_gen[5,j]] = 1
      df_movie_prof[6,df_movies_570_gen[6,j]] = 1
      df_movie_prof[7,df_movies_570_gen[8,j]] = 1
      df_movie_prof[8,df_movies_570_gen[8,j]] = 1
      df_movie_prof[9,df_movies_570_gen[0,j]] = 1
      df_movie_prof[10,df_movies_570_gen[10,j]] = 1
}
  df_movie_prof[is.na(df_movie_prof)] <- 0</pre>
  colsum_user <- colsums(df_user_prof, na.rm = T)</pre>
  vec <- append(vec,colsum_user/nrow(ratings_user1))</pre>
\label{eq:sum} \begin{array}{ll} \text{--my.cosine} <- \text{ function}(x,y) \{ \\ \text{sum}(x^*y)/(\text{norm}(x, \text{ type="2"}) \ * \text{ norm}(y,\text{type = "2"})) \end{array}
       df_similarity[i,] <- c(movie_ids[i,1],movie_ids[i,2], format(round(my.cosine(vec,df_movie_prof[i,]),4), nsmall = 4))</pre>
 . }
  cat("User ID 570 chose the following 10 movies: 18,231,592,914,1225,2000,3210,3801,54997,88812\n") cat("Of these, the following 5 movies are recommended:\n") print(head(df_similarity[with(df_similarity, order(-xtfrm(Similarity))), ],n=5))
```

## **Output**

	Movield <chr></chr>	MovieName <chr></chr>	Similarity <chr></chr>
3	592	Batman (1989)	0.7344
6	2000	Lethal Weapon (1987)	0.6545
9	54997	3:10 to Yuma (2007)	0.5805
10	88812	30 Minutes or Less (2011)	0.5593
2	231	Dumb & Dumber (Dumb and Dumber) (1994)	0.4026