# Introducción

Se cuenta con información de alrededor de 1.000.000 de votaciones anónimas de aproximadamente 4.000 películas, realizadas por unos 6.000 usuarios que a partir del año 2.000. Se sabe con certeza algunos datos sobre los usuarios, con que puntaje de 1 a 5 calificaron cada pregunta y una etiqueta de texto que algunos le han puesto a cada película como descripción. De las películas solo se dispone del nombre y el o los géneros a los que pertenecen.

El trabajo debe ser realizado en grupos de 2 personas. La fecha de entrega es el 27/06/2014 y la de reentrega el 25/07/2014.

# Datos

Todas las calificaciones se encuentran en el archivo "ratings.txt" y están en el siguiente formato:

Usuario :: MovieID :: Categoría :: Timestamp

* Los identificadores de usuario oscilan entre 1 y 6040
* Los MovieIDs oscilan entre 1 y 3952
* Las puntuaciones se realizan sobre una escala de 5 estrellas
* El Timestamp se representa en segundos desde el primero de enero de 1970 (epoch devuelto por time(2) en UNIX)
* Cada usuario tiene al menos 20 puntuaciones

La información de los usuarios se encuentra en el archivo " users.txt " en el siguiente formato:

Usuario :: Género :: Edad :: Ocupación :: Código Postal

Toda la información demográfica es proporcionada voluntariamente por los usuarios y no se verificó la exactitud. Sólo los usuarios que han proporcionado algunos datos sociodemográficos están incluidos en este conjunto de datos.

El género se denota por una "M" para masculino y "F" para femenino

La edad se elige entre los siguientes rangos:

1. Menor de 18
2. 18-24
3. 25-34
4. 35-44
5. 45-49
6. 50-55
7. 56 +

- La ocupación se elige entre las siguientes opciones:

1. Otra o no especificada
2. Académica / docente
3. Artista
4. Administrativo
5. Universitario / estudiante de posgrado
6. Servicio al cliente
7. Médico / cuidado de la salud
8. Ejecutivo / gerencial
9. Agricultor
10. Ama de casa
11. Estudiantes primarios
12. Abogado
13. Programador
14. Jubilado
15. Ventas / marketing
16. Científico
17. Por cuenta propia
18. Técnico / ingeniero
19. Comerciante / artesano
20. Desempleado
21. Escritor

La información de la película está en el archivo "movies.txt" en el siguiente formato:

MovieID :: Título :: Géneros

* Los títulos son idénticos a los títulos proporcionados por el IMDB (incluyendo año de lanzamiento)
* Los géneros están separados por “|” y se seleccionan de los siguientes:
  1. Action
  2. Adventure
  3. Animation
  4. Children's
  5. Comedy
  6. Crime
  7. Documentary
  8. Drama
  9. Fantasy
  10. Film-Noir
  11. Horror
  12. Musical
  13. Mystery
  14. Romance
  15. Sci-Fi
  16. Thriller
  17. War
  18. Western
* Algunos IDs no corresponden a una película por duplicado accidental de carga y/o cargas de prueba
* Las películas se introdujeron sobre todo a mano, por lo que pueden existir errores e inconsistencias

Todas las etiquetas se encuentran en el archivo "tags.txt" y están en el siguiente formato:

Usuario [Tab] MovieID [Tab] Tag [Tab] Timestamp

* Los identificadores de usuario oscilan entre 1 y 6040
* Los MovieIDs oscilan entre 1 y 3952
* Los Tags suelen ser una o dos palabras que describen la película
* El Timestamp se representa en segundos desde el primero de enero de 1970 (epoch devuelto por time(2) en UNIX)

# Resultados Técnicos Esperados

1. Detalle del preprocesamiento de los datos.
2. Software y algoritmo utilizado
3. Justificación de la elección de los parámetros del algoritmo ( la misma puede ser empírica)
4. Explicación del criterio utilizado para seleccionar los resultados “no técnicos”

# Resultados “No técnicos” Esperados

1. Seleccionar 10 reglas que representen las características más habituales de los gustos sobre películas.
2. Estudiar si existe alguna relación entre las variables demográficas y las películas.
3. Estudiar si existe alguna relación entre las variables demográficas y los géneros.
4. Seleccionar 10 reglas que representen las características más habituales de los gustos sobre películas de las personas que calificaron películas en 2000 y 2001 en dos conjuntos por separado. Verificar el comportamiento de estos dos conjunto con el que resulta de las personas que calificaron películas a partir de 2001.

# Bonus

Se desea implementar un sistema de recomendación de películas basado en el conocimiento que se puede obtener de este set de datos. Seleccionar 10 usuarios que representen diferentes gustos según el conocimiento que ustedes adquirieron anteriormente. A cada uno de estos usuarios se le deben recomendar 2 películas justificando el criterio de recomendación.