

Aprendizaje Automático

TP4: Métodos de Aprendizaje no supervisado

1. Considere el conjunto de datos movie data.csv que contiene las siguientes variables:
 - budget: Presupuesto de la película en dólares.
 - genres: Género de la película.
 - imdb id: ID en IMDB.
 - original title: Título de la película.
 - overview: Descripción de la película.
 - popularity: Número arbitrario que indica la popularidad, más alto, más popularidad.
 - production companies: Cantidad de compañías que produjeron la película.
 - production countries: Cantidad de países que produjeron la película.
 - release date: Fecha de lanzamiento.
 - revenue: Ganancia de la película en dólares.
 - runtime: Duración de la película en minutos.
 - spoken languages: Cantidad de idiomas que se hablan en la película.
 - vote average: Votación promedio, número entre 0 y 10.
 - vote count: Cantidad de votos que tiene la película.

Realizar:

- a) Un análisis univariado de las variables.
- b) Aplicar los métodos de aprendizaje no supervisado de k-medias, Agrupación jerárquica y Red de Kohonen (mostrar las variables por separado y la matriz U) para agrupar películas e interpretar los resultados.
- c) Considerar el subconjunto de películas cuyo género es "Action", "Comedy" y "Drama" y utilizar los tres métodos para tratar de predecir el género.