



INGENIERIA DE SISTEMAS SISTEMAS INTELIGENTES II PRIMER PROYECTO

SEGMENTACIÓN DE PELÍCULAS BASADAS EN SUS CARACTERÍSTICAS

El objetivo principal de este proyecto es descubrir patrones ocultos y segmentar películas utilizando datos históricos que abarcan información detallada de más de 5000 películas. Se deben aplicar técnicas de Análisis Exploratorio de Datos (EDA), Análisis de Componentes Principales (PCA) y algoritmos de clustering para descubrir las características que definen distintos tipos de películas en la industria cinematográfica, identificando grupos o segmentos de películas basados en atributos clave como presupuesto, ingresos, popularidad, duración, puntuación de los usuarios y géneros. Al segmentar las películas de esta manera, se busca entender mejor las tendencias actuales del cine, identificar qué características están asociadas con el éxito comercial o crítico, y ofrecer recomendaciones valiosas para productores y directores.

Desarrollo del Proyecto:

Para el desarrollo del proyecto se requiere una **exploración exhaustiva del dataset**, familiarizarse con las diversas variables disponibles, tales como el presupuesto de producción, los ingresos generados, la duración de las películas, la popularidad medida a través de diferentes métricas, y las puntuaciones otorgadas por los usuarios. Este paso inicial es crucial para identificar posibles inconsistencias, valores faltantes o atípicos que podrían influir en el análisis posterior.

Una vez que los datos han sido limpiados y preprocesados, se deberá realizar un **Análisis Exploratorio de Datos (EDA)**. Mediante visualizaciones como histogramas, diagramas de dispersión y matrices de correlación, en los que se podrán identificar relaciones interesantes entre variables, como por ejemplo, la correlación entre presupuesto e ingresos, o cómo la duración de una película puede influir en su popularidad.

Se deberá aplicar un **Análisis de Componentes Principales (PCA)** para reducir la dimensionalidad del conjunto de datos, para facilitar la identificación de patrones subyacentes sin la complejidad añadida de múltiples variables.

Finalmente, utilizando los datos transformados por PCA, se deben emplear el **algoritmo de clustering** que justificadamente considere para segmentar las películas en grupos distintivos. A través de este proceso, descubrirán categorías como películas de bajo presupuesto con alta popularidad, producciones de alto presupuesto con ingresos



INGENIERIA DE SISTEMAS SISTEMAS INTELIGENTES II PRIMER PROYECTO

moderados, o películas de géneros específicos que tienden a recibir puntuaciones elevadas de los usuarios.

Finalmente, se deben presentar los hallazgos en un informe detallado y una presentación visual. Estos resultados proporcionarán insights valiosos sobre qué factores contribuyen al éxito de una película, ofreciendo recomendaciones prácticas para la industria del cine.

Consideraciones:

- La implementación deberá hacerse en un notebook de Python documentando cada franja de código (no cada línea de código). Así, cada franja tendrá definido el propósito y una descripción explicativa del código.
- Las visualizaciones deberán ser descriptivas per se.
- Para los clusters, deberá contarse con visualización bi y tridimensional. La visualización tridimensional debe ser interactiva.
- Pueden conformar grupos de 3 estudiantes, máximo.
- Fecha de sustentación: 5 de noviembre de 2024.
- El informe debe ser detallado, como se indica en el enunciado, esto corresponde a un documento bien estructurado y descriptivo que también contenga un análisis de resultados y conclusiones derivadas de esta experimentación.