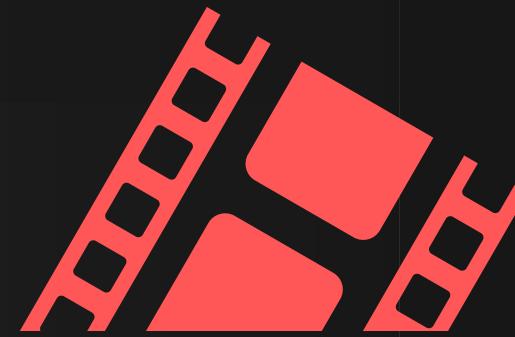


Movie Apps.



MOVIE GENRE

PREDICTION

David Hernando Avila Jimenez , Erica Marcela Martínez Silva, Lina Marcela Rivasa,
Raúl Eduardo Vásquez Duarte, Sebastián Amaya Porrasa



Contexto

- Transformación del consumo de películas
- Auge en disponibilidad de plataformas de Streaming
- Modelo de negocio basado en la personalización
- Segmentación de películas por género
- Inclusión de algoritmos de predicción
- Transformación del marketing cinematográfico





OBJETIVOS

Negocio

- Ofrecer a las plataformas de streaming de películas, la oportunidad de fortalecer sus mecanismos de categorización de contenido, a través de una herramienta que precise correctamente las clasificaciones y mejore así la experiencia al cliente.
- Brindar una herramienta que sea un puente entre el marketing y publicidad digital cinematográfico y las plataformas de streaming de películas, al permitir la categorización de estas a partir de una pieza digital.

Criterio de éxito

- Realizar la categorización correcta de al menos el **70%** de un catálogo que contiene 3.383 películas .



Movie Genre Prediction



OBJETIVOS

Minería de datos

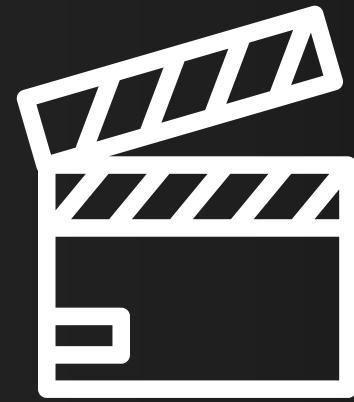
Desarrollar un modelo basado en el procesamiento de lenguaje natural, que prediga el género de las películas a partir del contenido que hay en sus posters.

Criterio de éxito

Se espera obtener un AUC mayor o igual a **0.89**, a través del uso de un modelo que utilice técnicas de vectorización y representaciones distribuidas (embeddings), para la predicción del género de las películas.



TRATAMIENTO DE LOS DATOS



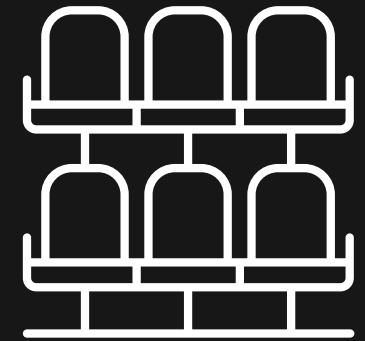
ENTENDIMIENTO DE LOS DATOS

SET ENTRENAMIENTO

- Formato **.csv**
- **7,895** películas registradas (ID, año de estreno, título, resumen, géneros y rating)
- 24 Categorías de género
- Películas lanzadas entre los años **1894 y 2015**

SET VALIDACIÓN

- Formato **.csv**
- **3,383** películas registradas
- Películas lanzadas entre los años **1983 y 2015**

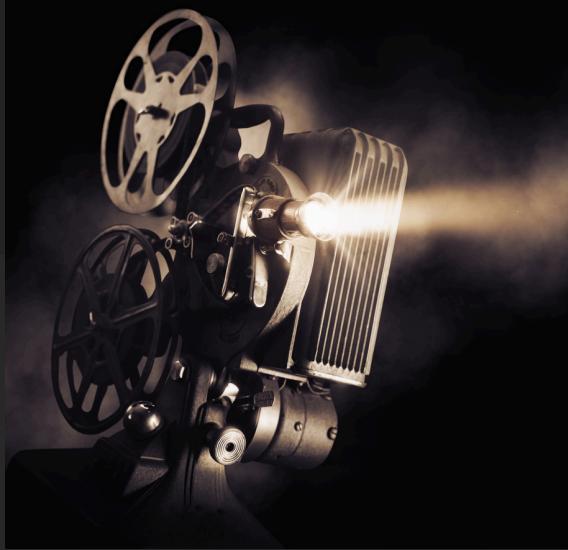


PREPARACIÓN & LIMPIEZA

- Limpieza de símbolos
- Conversión a minúsculas
- Tokenización sencilla (espacios)
- Eliminación de Stopwords
- Lemmatización y Stemming
(algunos casos)



MODELAMIENTO



BOW & RN Sencilla



**N-Grams ajustes &
Random Forest**

Preprocesamiento: Lematización

Vectorización: Bag of Words

Red neuronal:

- 1 capa oculta (428 neuronas)
- Func. activación: RELU
- Capa de salida: #etiquetas (24) & activación 'sigmoid'.

AUC: 0.872

Vectorización: Ngrams (1-3),
max features (900K)

Random Forest:

- N. árboles: 200
- Min ramas: 10
- Min hojas: 2
- Profundida max: None.

AUC: 0.877



**Regresión Logística /
Random Forest**

Preprocesamiento: Lematización

**Unión 'plot' + 'Title' como data de
entrenamiento**

Vectorización: Ngrams (1-3)

Reg. Logística: C: 1.0, Solver: 'lbfgs',
max_iter: 2000

Random Forest: árboles: 500, min
ramas: 2, min hojas: 4, profundidad
max: 30, 'bootstrap': False

AUC: 0.879



**BOW + GloVe & Red
Neuronal Sencilla**



**Unión 'plot' + 'Title' como
data de entrenamiento**

Vectorización: BOW

Embedding: GloVe

Red neuronal: misma que
modelo 1, con 400 neuronas.

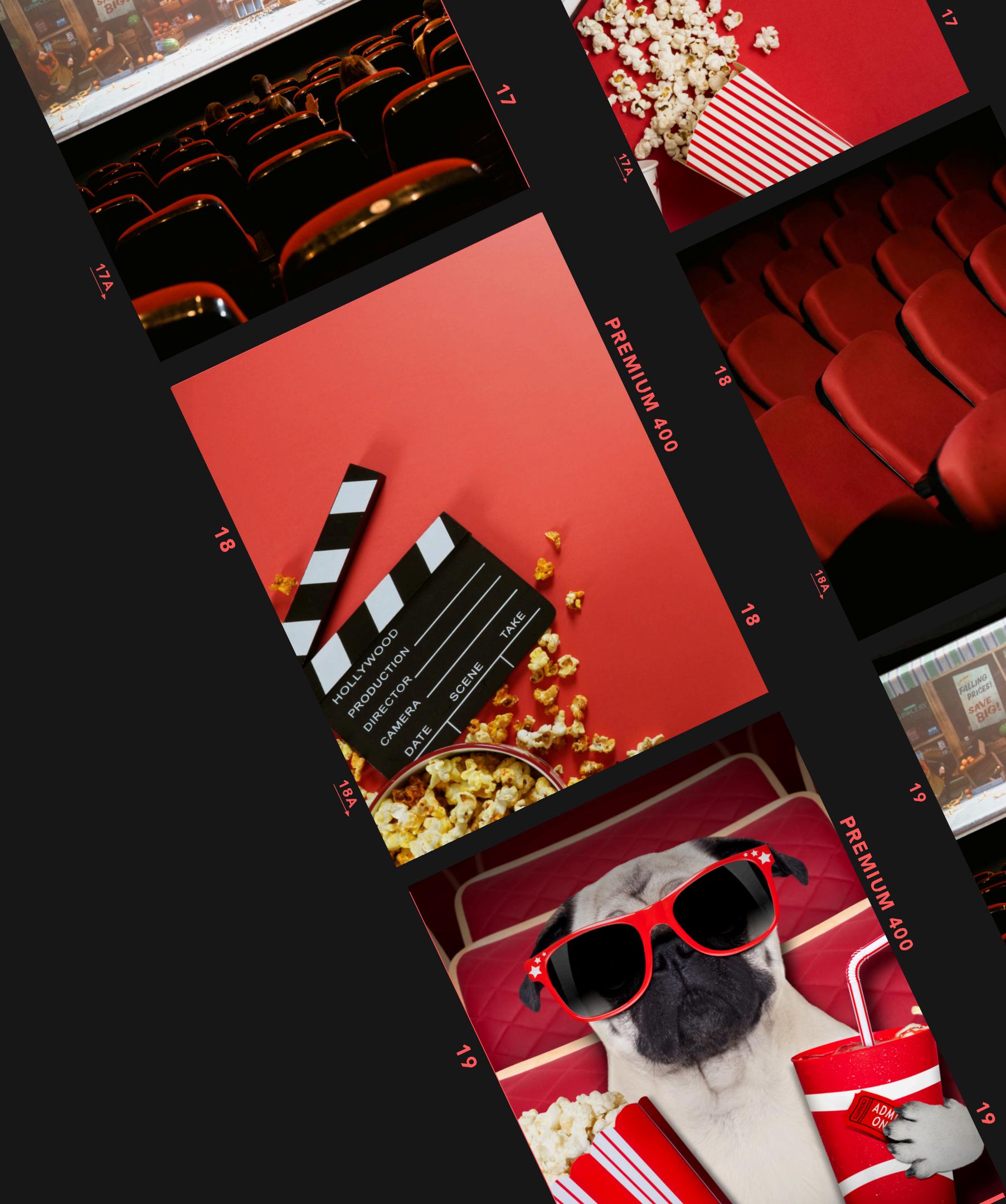
AUC: 0.913



Movie Genre Prediction

DESARROLLO & SIGUIENTES PASOS

-
1. Escenario de prueba.
 2. Siguientes pasos, con el fin de disponer a los medios cinematográficos y de streaming una herramienta sólida que fortalezca el negocio
 3. Estrategias enfocadas en el despliegue.





Movie Genre Prediction

1. Escenarios de prueba

Película				
Género Correcto	Documentary	Adventure, Fantasy, Family, Animation, Musical	Animation, Adventure, Comedy, Family, Fantasy	Action, Adventure
Género Predicción	Documentary	Adventure, Comedy	Adventure, Comedy, Family, Fantasy	Action, Adventure





2. Siguientes pasos

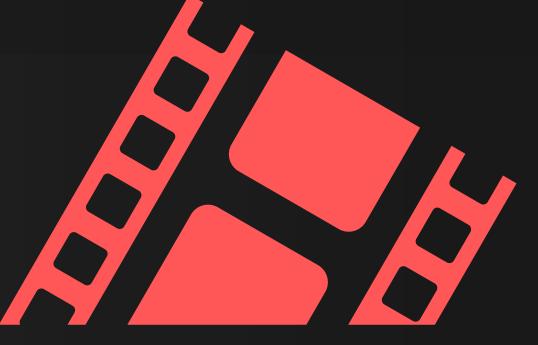
- Optimización de modelos.
- Escalabilidad
- Actualización de catálogo (por ej. datos de IMDB)
- Pruebas piloto.
- Monitoreo y Mantenimiento

3. Estrategias

- API para integrar sistema de clasificación de géneros.
- Conexiones con proveedores de tecnología para mejora en capacidad de procesamiento.
- Expansión del modelo a otros formatos (clasificación series, documentales, entre otros)



Movie Genre Prediction



GRACIAS
