

# Predicción de la clasificación de una película chilena en IMDB

Juan Carlos Barría Rival

## ARTICLE INFO

### Keywords:

Películas  
Predicción  
IMDb  
CineChile  
Chile  
Películas Chilenas  
Aprendizaje automático  
Random Forest

## ABSTRACT

La industria del cine no es sólo una industria o un centro de entretenimiento, sino que ahora es un centro de negocios global. Todo el mundo está entusiasmado con el éxito de una película en la taquilla, su popularidad, etc. En este caso de estudio se utilizará como referencias las películas y documentales chilenos. Para este proyecto se ha creado un *dataset* con la lista de las películas y/o documentales mejores de todos los tiempos, según CineChile, junto con su calificación en IMDb de cada una, en esta se pueden ver las valoraciones de estas.

## 1. Introducción

En la industria del cine cada país crea sus propias producciones, ya sea para enseñarlas al público nacional o para mostrarlas a un público internacional y lograr un gran salto y ser nominados, por ejemplo, para ganar un premio Óscar. Conocidos son los casos de "Historia de un Oso" de 2014 que logró ganar este premio a Mejor Cortometraje Animado el año 2016 y actualmente la nominación de "El Agente Topo" largometraje estrenado en 2020, nominado a Mejor Largometraje Documental para los Oscars 2021.

Según datos del Ministerio de Cultura del Gobierno de Chile, en cuanto al tipo de películas exhibidas en espacios de cine durante el 2019, se exhibieron un total de 1.568 filmes, entre largometrajes y cortometrajes, de las cuales un 15,37% corresponden a cine chileno. Sin embargo, del total de espectadores registrados el 2019, 359.627 fue a ver películas de cine chileno, lo cual equivale a un 1,2% de las audiencias. Finalmente, a nivel del comportamiento por regiones, es interesante observar que la mayor cantidad de espectadores se concentran en las regiones Metropolitana, de Valparaíso y del Biobío, pero que, sin embargo, al analizar estos datos en términos per cápita, las regiones de Arica y Parinacota y de Antofagasta, pasan a cobrar relevancia dentro de las que más consumen cine en Chile. [1]

IMDb es una base de datos en línea que almacena información relacionada con películas, con sus características, además de la valoración por parte del público con nota de 1 a 10.

## 2. Desarrollo

El cine es la actividad cultural más popular en Chile, en la medida que cuenta con los índices de participación más altos por parte de la población, en relación a otras actividades de este tipo, según la última Encuesta Nacional de Participación Cultural (2017), la cual concluye que un 43% de las personas ha asistido al cine durante los últimos 12 meses. De este modo, el cine no solo es relevante en tanto creación artística y cultural que pone a disposición una serie de contenidos que fluctúan entre lo educativo y lo entretenido, lo emotivo y lo informativo, entre otras posibilidades dentro

de su amplia oferta; sino que además representa una manifestación cultural de masas, que llega a amplias audiencias y atrae a un público diverso que comprende a personas de todas las edades, diferentes geografías y sectores sociales. También se puede tomar en cuenta que en este último año las películas chilenas ya acumulan cerca de medio millón de espectadores, cifra inesperada para una temporada difícil. Destacan "Pacto de fuga", estrenada entre estallido y pandemia, y "Tengo miedo torero", que acosada por el cierre de salas se lanzó a las plataformas y vendió 50 mil entradas en dos días. [2]

Por lo que en este proyecto se busca predecir las clasificaciones de las producciones chilenas, con el objetivo de conocer las características que hacen que estas sean populares en nuestro país con la idea de exportarlas a otros mercados y sean vistas en otros lugares del mundo, ya sea a través del cine o de plataformas de streaming.

### 2.1. Conjunto de datos creado

Al no haber encontrado en sí un *dataset* con películas chilenas se creó uno desde cero. En 2016, el sitio web especializado **CineChile.cl** hizo un top 50 con las mejores películas chilenas de todos los tiempos, para conformar esta lista, enviaron un formulario a un universo amplio de especialistas: cineastas, actores, investigadores, académicos, periodistas y críticos de cine. Ellos podían elegir cualquier película, ya sea largometraje de ficción, documental, cortometraje o animación, y se debían elegir 10 títulos, por orden de preferencia. También, se les dio la opción de si querían que sus preferencias fueran publicadas. Finalmente recibieron 77 respuestas, de ellas, 44 están con la autorización de ser publicadas.

Según se explica en el sitio, El conteo se hizo de la siguiente manera. Se le dio un puntaje a cada posición de la lista, siendo 10 puntos el máximo (para el primero uno de la lista) y 1 punto el mínimo (para el décimo puesto de la lista). Así, se obtuvieron dos cifras para el conteo (cantidad de puntaje y cantidad de menciones), con la cantidad de puntaje como el referente principal para definir el lugar en la lista. Cuando existe igualdad de puntaje, el lugar se define por la cantidad de menciones que la película obtuvo dentro de la encuesta.

El corte de la encuesta se hizo hasta el puesto 50, pero las películas que quedaron fuera de estos puestos igual están mencionadas dentro del top, por lo que el conjunto de datos llega a ser de un total de 60 producciones.

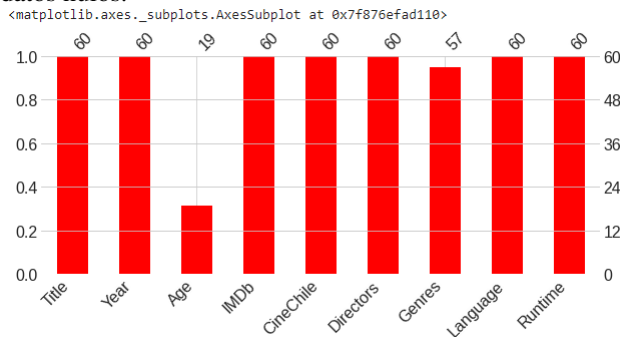
El conjunto de datos creado tiene las siguientes características:

1. ID: Número asignado según su posición en la lista.
2. Title: Título de la película/documental.
3. Year: Año en que se estrenó.
4. Age: Edad a la que esta dirigida.
5. IMDb: Calificación en el sitio IMDb.
6. CineChile: Puntaje asignado por CineChile.
7. Directors: Director/es de la película/documental.
8. Genres: Género/s de la película.
9. Language: Lenguaje/s en el que está disponible.
10. Runtime: Duración en minutos.

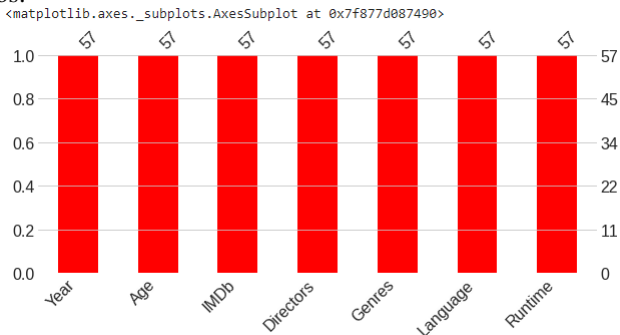
Este conjunto de datos se podría ampliar mucho más añadiendo las películas chilenas populares en las plataformas de streaming debido al impacto que han captado estas en los últimos años.

## 2.2. Implementación, limpieza y preprocesamiento

Se tomó como base el proyecto utilizado anteriormente llamado *IMDb Rating Prediction from a data set of Movies* hecho por el usuario "diptaraj23" subido a la plataforma *Kaggle*. En este se cargó el nuevo conjunto de datos creado de producciones chilenas. Se hizo una búsqueda y limpieza de los datos nulos.



También se eliminaron las columnas que no se toman en cuenta para la predicción, y finalmente quedan ya listo los datos:



Se copian estos datos de entrenamiento y testeo para evitar cambios del dataset original y se aplica *LabelEncoder* para codificar etiquetas de una característica categórica en valores

numéricos entre 0 y el número de clases menos 1. El método *fit* lo entrena y el método *transform* transforma las etiquetas que se incluyan como argumento en los números correspondientes. El método *fit\_transform* realiza ambas acciones simultáneamente. Posteriormente de las columnas que se utilizarán (Year, Runtime, Age, Directors, Genre, Language), se dividirán los datos que se usarán para entrenamiento y prueba.

## References

- [1] Ministerio de las culturas, las artes, y.e.p., 2019. Análisis del cine en Chile y sus audiencias año 2019 .
- [2] González, R., 2020. El cine chileno en el año de la plaga. URL: <https://www.latercera.com/culto/2020/09/20/el-cine-chileno-en-el-ano-de-la-plaga/>.