



Tecnológico de Monterrey

Reto Análisis del contexto y la normatividad

Inteligencia artificial avanzada para la ciencia de datos I

Alumno

Jackeline Conant Rubalcava A01280544

Profesores

Antonio Carlos Bento

Alfredo Esquivel Jaramillo

Mauricio González Soto

Julio Antonio Juárez Jiménez

Frumencio Olivas Álvarez

Jesús Adrián Rodríguez Rocha

Hugo Terashima Marín

10 de septiembre del 2024

Reto Análisis del contexto y la normatividad.

Para el análisis de normatividad, utilizaré el [Portafolio 2 - Análisis y Reporte sobre el desempeño del modelo con ayuda de frameworks](#). Esta base de datos se llama “[Dataset of Songs in Spotify](#)” y fue publicada hace 4 años por el autor Andrii Samoshyn. La base de datos anterior detalla varios aspectos sobre diferentes canciones y el tipo de género que tienen en Spotify. En base a las características con las que contaba la base de datos, decidí crear un modelo de predicción que me ayudara a determinar el tipo de género que se estaba escuchando en cada canción.

La base de datos cuenta con **42305 registros y 22 columnas** originalmente, donde se pueden ver:

Datos	Descripción
danceability	Bailabilidad del audio
energy	Energía del audio
key	Keys musicales
loudness	Intensidad
mode	Mode del audio
speechiness	Palabras del audio
acousticness	Acústicalidad
instrumentalness	Instrumentalidad
liveness	Vitalidad
valence	Valencia
tempo	Tempo
type	Tipo de audio
id	Identificación única del audio
uri	Identificación única del audio de spotify

track_href	Link referencia del audio
analysis_url	Link de análisis
duration_ms	Duración del audio en milisegundos
time_signature	Indicación del compás
genre	Género musical
song_name	Nombre del audio
Unnamed: 0	Variable de llave Unnamed: 0
title	Título del álbum

Durante la elaboración de mi proyecto, tuve que limpiar los datos disponibles, eliminando algunas características que no eran necesarias para el análisis. Posteriormente, entrené dos modelos de predicción: XGBoost, que se utiliza para la clasificación de datos basándose en múltiples árboles de decisión mediante boosting y un modelo de búsqueda con un solo árbol de decisión.

La base de datos que utilicé tiene una usabilidad del 100%, lo que significa que los datos están completos, son reales y creíbles, y tienen una gran compatibilidad para la creación de modelos de predicción. Además, cuenta con una licencia [CC0: Public Domain](#) o “No Copyright”, que otorga acceso público a poder ser editado y muestra que el creador de esta base de datos renuncia a sus derechos de autor.

“La persona que ha asociado una obra con este acuerdo ha dedicado la obra al dominio público, renunciando a todos sus derechos sobre la obra en todo el mundo bajo la ley de derechos de autor; incluidos todos los derechos relacionados y conexos, en la medida permitida por la ley.

Puedes copiar, modificar, distribuir y ejecutar la obra, incluso con fines comerciales, todo sin pedir permiso. Consulta la información adicional a continuación.”(Creative Commons, n.d.)

Lo anterior proviene de la licencia CC0: Public Domain, lo cual demuestra que se me permite copiar la base de datos creada por el autor, modificarla para crear mi modelo de predicción, distribuirla o publicarla dentro de mi portafolio del proyecto, e incluso utilizarla en mi trabajo con casi ninguna restricción y todo sin la necesidad de pedirle permiso al autor.

De todas formas, es importante tener en cuenta que existe la [Ley de Protección de Datos Personales \(LFPDPPP\)](#) en México, la cual protege los datos personales que puedan tener las empresas u organizaciones y garantiza el derecho a la privacidad. En caso de incumplimiento, podrían aplicarse sanciones. Por lo tanto, es necesario hacer un uso ético de los datos para no causar daño a la compañía, en este caso Kaggle (de donde se obtuvo la base de datos), y a Spotify (la empresa de donde proviene la información).

“Al utilizar o citar la obra, no se debe implicar la aprobación del autor o del afirmador.”(Creative Commons, n.d.)

Hay que señalar, como en la cita anterior, que dentro de la CC0: Public Domain hay algunas cosas que no respalda y eso es las cuestiones de datos personales y éticos, donde el autor de esta base de datos no tuvo que ver con este proyecto y él no respalda mi trabajo.

“De ningún modo los derechos de patente o marca registrada de cualquier persona se ven afectados por CC0, ni los derechos que otras personas puedan tener sobre la obra o sobre cómo se utiliza la obra, como los derechos de publicidad o privacidad.”(Creative Commons, n.d.)

Durante la creación de mis modelos predictivos, tomé en cuenta los datos disponibles y **eliminé aquellos que pudieran estar relacionados con datos personales** o enlaces a otras páginas web. Además, durante y después de finalizar mi portafolio, me aseguré de no proporcionar información falsa o dañina en contra de la empresa Spotify y de respetar los datos del autor.

En caso de no seguir el código de ética al utilizar datos muy personales, como nombres de usuarios, identificaciones personales y ubicaciones, y causar daños a la compañía a través de contenido falso y engañoso, estaría en incumplimiento de la Ley de Protección de Datos Personales (LFPDPPP) y podría enfrentar una multa por difamación a la compañía.

En conclusión, mi trabajo SÍ cumple con la normatividad de uso de bases de datos y SI es utilizado de forma ética y legal, y está respaldado por la licencia CC0: Public Domain. Además, no se ha utilizado con ninguna intención maliciosa y no muestra ningún dato personal ni información falsa. El único propósito de este proyecto es predecir el género de una canción en función del tipo de características que posee.

Bibliografía:

- Kaggle. (n.d.). Dataset of songs in Spotify. Retrieved September 10, 2024, from <https://www.kaggle.com/datasets/mrmorj/dataset-of-songs-in-spotify/data>
- Creative Commons. (n.d.). CC0 1.0 Universal (CC0 1.0) public domain dedication. Retrieved September 10, 2024, from <https://creativecommons.org/publicdomain/zero/1.0/#ref-publicity-rights>
-
- Cámara de Diputados del H. Congreso de la Unión. (2010). Ley Federal de Protección de Datos Personales en Posesión de los Particulares. Retrieved September 10, 2024, from <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>