



Tipo de artículo: Artículos originales
Temática: Inteligencia Artificial
Recibido: 18/12/2022 | Aceptado: 03/02/2023 | Publicado: 30/03/2023

Identificadores persistentes:
ARK: ark:/42411/s11/a83
PURL: 42411/s11/a83

Predicción de la clasificación ESRB para videojuegos según su contenido usando árboles de decisión

Predicting ESRB ratings for video games by content using decision trees

Rodrigo Sebastián Huamán Maque¹, Graciela Condori Anahua², Fátima Gigi Rojas Carhuas³ Rodolfo Robert Quispe Huacho⁴

¹ Universidad Nacional de San Agustín. Arequipa, Perú. rhuamanma@unsa.edu.pe

² Universidad Nacional de San Agustín. Arequipa, Perú. gcondoria@unsa.edu.pe

³ Universidad Nacional de San Agustín. Arequipa, Perú. frojasca@unsa.edu.pe

⁴ Universidad Nacional de San Agustín. Arequipa, Perú. rquispehuach@unsa.edu.pe

Resumen

Diversos estudios han comprobado que los niveles de violencia en los videojuegos pueden influir negativamente en el desarrollo de los niños, especialmente en la adolescencia y es por ello que se debe tener cuidado en que la clasificación sea la adecuada según el contenido presente. Para el análisis se utilizó la clasificación ESRB, que contiene 7 diferentes categorías, junto con la implementación de un modelo de árboles de decisión, que es una técnica de minería de datos capaz de representar gráficamente la relación entre las variables. Los resultados arrojaron que el nivel de precisión para un árbol de nivel 6 no supera el mínimo requerido.

Palabras clave: Árbol de decisión, ESRB, Minería de datos.

Abstract

Various studies have proven that the levels of violence in video games can negatively influence the development of children, especially in adolescence and that is why care must be taken that the classification is appropriate according to the content present. For the analysis, the ESRB classification was used, which contains 7 different categories, together with the implementation of a decision tree model, which is a data mining technique capable of graphically representing the relationship between the variables. The results showed that the precision level for a level 6 tree does not exceed the minimum required.

Keywords: Decision tree, Data mining, ESRB.

Introducción

Los videojuegos se han convertido en un mercado crecimiento año a año, llegando a facturar más que la industria cinematográfica y de música a nivel global; desde su creación en 1962 por Steve Rusell, en el Massachusetts Institute of Technology inventando el Spacewar [1], ha ido mejorando en gráficos y mecánicas nuevas en busca de nuevos jugadores, pero también a todo su largo recorrido fue acompañado de polémicas por el tipo de contenido sin control que ofrecían.

Los videojuegos se definen [2] como programas informáticos que mantienen a los usuarios interactuando a través de imágenes que se muestran en dispositivos con pantallas de varios tamaños, a través de un incentivo implícito para ganar. Actualmente se están desarrollando videojuegos que pueden ser controlados por voz o movimiento, siendo un avance a como antes solo se podía controlar con los dedos u otros instrumentos adicionales (como guitarras, rifles y pistolas).

Los sistemas de clasificación previo a 1994 no existían, sin embargo después del lanzamiento de algunos videojuegos en los años 90 como Funk, Buchman y principalmente por la polémica de Mortal kombat, causada por los elementos violentos que siempre tuvo la saga, se crea en Estados Unidos la Entertainment Software Rating Board (ESRB) con el fin de garantizar contenido adecuado a partir de diferentes parámetros [3].

Categorías de clasificación y su sugerencia de nivel de edad [4]:

- EC Early Childhood: Contenido apto para niños de tres a diez años, no contiene material inapropiado.
- E Everyone: Contenido apto para personas de seis años en adelante, contiene una mínima violencia, travesuras y lenguaje obsceno.
- E10+ Everyone 10+: Contenido apto para mayores de diez años. Puede contener mayor cantidad de violencia de caricatura o temas sugestivos.
- T Teen: Contenido apto para personas de trece años en adelante. Puede contener violencia, humor grosero y temas sugestivos con un mínimo de sangre.

- M Mature: Contenido apto para personas de diecisiete años en adelante. Puede contener temas sexuales y lenguaje o violencia más intensa con derramamiento de sangre.
- AO Adults Only: Contenido apto sólo para adultos de 18 años a más. Puede incluir contenido sexual gráfico o escenas prolongadas de violencia.
- RP: La clasificación se encuentra en espera.

Además de esta metodología de clasificación de videojuegos ERP existen otras relevantes como son:

- USK es la organización de clasificación de software utilizada en Alemania que viene a ser un tablero de clasificación bastante estricto, en parte debido a la paranoia del país entorno a los videojuegos con violencia, ya que existen distintos videojuegos que han sido prohibidos o han requerido que los desarrolladores editen o quiten ciertas funciones.
- SMECCV es una clasificación mexicana orientada a las especificaciones gráficas para las advertencias, descriptores de contenidos y elementos interactivos que deberán implementar objetos, respecto de los videojuegos distribuidos, comercializados o arrendados en el territorio nacional. Con el objetivo de crear parámetros específicos que apoyen a desarrollar una adecuada categorización tanto el contenido de los mismos, así como de las personas a la que va dirigido.
- CERO Computer Entertainment Rating Organization es una organización que clasifica la edad recomendada de los juegos en Japón, y fue fundada en 2002 como una división de "CESA".
- GSRR es el sistema de clasificación de contenido de videojuegos utilizado en Taiwán, Hong Kong y sureste de Asia. La ley realizada el 6 de julio de 2006 y se enmendó el 29 mayo 2012.

Los padres tienden a preocuparse por el contenido del juego en especial si son violentos, ya que podrían afectar la conducta del niño. Según Russell N. Laczniak, Les Carlson, Doug Walker, y E. Deanne Brocato los ESRB pueden ayudar a que los niños jueguen juegos menos violentos y asimismo reducir la participación de los niños en comportamientos negativos en la escuela [5], [6].

En una investigación sobre consecuencias de los videojuegos en niños y adolescentes en aspectos de la vida social, muestra como posibles efectos psicológicos y fisiológicos pueden sufrir los menores, donde ahonda como el entregar un juego con contenido de acuerdo a su edad le puede evitar comportamiento agresivos, problemas de atención, gasto energético, respuestas hormonales y entre otros el efecto negativo de la exposición prolongada a los videojuegos, tal como la llamada «epilepsia fotosensible» [7], [8].

Los árboles de decisión han sido utilizados en el aprendizaje, sirve para hallar una respuesta o curso de acción, dadas k variables o entradas, donde cada nodo suyo hace una pregunta sobre una variable; también se puede haber más de k preguntas antes de emitir una respuesta, por otro lado, las hojas o nodos finales “dan la respuesta” o decisión a tomar. El método usa una muestra de datos llamada generalmente conjunto de entrenamiento para formar el árbol, por lo cual resulta una buena opción para poder clasificar [9].

Este trabajo tiene como finalidad utilizar los árboles de decisión y la inteligencia que estos proveen para predecir la clasificación de los videojuegos según el contenido que presenten, tales como: sangre, lenguaje explícito, violencia, etc; en las diferentes categorías que tiene la ESRB, con el propósito de brindar ayuda al control de efectos negativos en menores de edad.

Marco Teórico

Modelos de árbol

Son modelos precisos, estables y más sencillos de interpretar principalmente porque construyen unas reglas de decisión, las cuales se pueden representar usando un diagrama similar a un árbol. A diferencia de los modelos lineales conocidos, pueden representar relaciones no lineales para resolver problemas más diversos. En estos modelos, destacan los árboles de decisión y los random forest. Al ser más precisos y elaborados, obviamente ganamos en capacidad predictiva, pero perdemos en rendimiento [10].

Árbol de decisión

Es una forma gráfica y analítica de representar todos los eventos o sucesos que pueden surgir a partir de una decisión asumida en cierto momento. Nos ayudan a tomar la decisión más “acertada”, desde un punto de vista

probabilístico, ante un abanico de posibles decisiones tomando en cuenta cierta precisión. Estos árboles permiten examinar los resultados y determinar visualmente cómo fluye el modelo. Los resultados visuales ayudan a encontrar subgrupos específicos y relaciones que tal vez no se podría hallar con estadísticos más tradicionales [11].

Herramientas

- *DataSet*: Un conjunto de datos corresponde a los contenidos de una única tabla de base de datos, donde cada columna del dataset representa una variable única, y cada fila está reemplazando a un miembro en concreto del conjunto de datos en la tabla. En un dataset tenemos todos los valores que puede tener cada una de las variables, como por ejemplo la sangre que se pueda visualizar y las agresiones que tengan dentro del juego. Cada uno de estas variables se conoce con el nombre de dato, también puede incluir datos para uno o más miembros en función de su número de filas [14].
- *Google Colab*: Viene a ser una herramienta que nos da la posibilidad de ejecutar y compilar scripts del lenguaje Python a mediante los servidores de Google. Permittiéndonos ejecutar porciones de código similar a un cuaderno de Jupyter Notebook para linux. Orientado para implementar machine learning, ya que al ser una máquina virtual no se limita a los recursos de hardware. Teniendo en cuenta que se puede usar tanto el GPU y la TPU del mismo servidor de Google para poder potenciar el procesamiento del proyecto [15].
- *Python*: Es un lenguaje 100% gratuito. Tratándose de un lenguaje open source siendo disponible para todas las plataformas. Ya que podemos instalarlo y ejecutarlo en diferentes sistemas operativos como Windows, Linux o MacOS [15].

Métodos y Metodología computacional

Descripción de la Aplicación

Se usó para el desarrollo de este clasificador de ESBR usando un árbol de decisión, la cual comprende las siguientes fases:

- La fuente de información fue un Dataset.csv, esto fue sacado de la página Kaggle.com, necesaria para poder construir un árbol de decisión utilizando las librerías sklearn, y se hizo uso de las funciones que nos brinda.

- **Análisis de Datos:** con el dataset que se obtuvo, se vio la cantidad de información que tiene, y las variables. las clasificamos en números para su uso correcto con la librería o lenguaje que se trabajaría.
- **Modelado:** Se escoge una técnica adecuada para poder desarrollar la clasificación.
- **Evaluación:** Tuvimos que corroborar efectivamente que el modelo escogido se ajuste a lo que estamos buscando, en este caso poder clasificar según ESBR y que pueda ser modificado en el futuro.

Trabajos relacionados

A continuación se muestran una serie de artículos que se han analizado para poder desarrollar nuestro trabajo, se procederá a hablar brevemente sobre los puntos que nos sirvieron de apoyo.

Algoritmos de aprendizaje automático para análisis y Predicción de datos

Sandoval [10] expone sobre la técnica de Machine Learning, rama de la inteligencia artificial, como elemento fundamental de la ciencia de datos, así como tipos de aprendizaje, diferentes métodos que son utilizados para la predicción de los mismos y la fase de desarrollo hasta su presentación.

Se obtuvo una fundamentación teórica para el modelado del proyecto, además brinda información sobre cómo cual método de predicción podría ser más factible.

Cómo aplicar árboles de decisión en SPSS

Complemento al trabajo anterior se tiene el realizado por Berlanga, Rubio y Vilà [11] donde explica como funciona un árbol de decisión como parte de la minería de datos, usos del mismo y terminología relacionada; se presenta una información más específica en contraste con el anterior.

Predicción del consumo de cocaína en adolescentes mediante árboles de decisión

Gervilla y Palmer [12] presentan un estudio sobre cómo implementar técnicas de Data Mining, especialmente árboles de decisión, que permitan analizar el valor predictivo de la impulsividad y la búsqueda de nuevas sensaciones sobre consumo de cocaína y/u otras sustancias, principalmente en la etapa de la adolescencia,

partiendo de la suposición que los rasgos anteriormente mostrados evidencian una marcada relación con las conductas de experimentación y la adicción a ciertas sustancias.

Predicción, análisis y pronóstico de covid-19 utilizando un modelo de árbol de decisión

Otro ejemplo desarrollado es la tesis de Castellanos y Haro [13], donde elaboran un modelo de aprendizaje automático basado en IA, utilizando un algoritmo de aprendizaje supervisado, árbol de decisión. Además, la creación y estructuración de una base de datos que les permitió predecir, analizar y pronosticar el grado de afectación en los pacientes contagiados de Covid-19.

Resultados y discusión

Teniendo en cuenta las fases del Machine Learning, se tiene lo siguiente:

Análisis de los datos

El dataset contiene el nombre de 1895 juegos con 34 características de contenido de calificación ESRB con el nombre y la consola como características para cada juego.

Un solo punto de datos se representa como un valor binario 0-1 para la consola y un vector binario para las características del contenido de ESRB.

Variables de entrada

<i>Característica</i>	<i>Tipo</i>	<i>Descripción</i>	<i>Llave</i>
title	string	Nombre del juego.	- - - -
console	int	La consola en la que se comercializo el videojuego.	0 = no 1 = si
Alcohol_Reference	int	Referencia a y/o imágenes de bebidas alcohólicas.	0 = no

			1 = si
Animated_Blood	int	Representaciones de sangre descoloridas y/o poco realistas.	0 = no 1 = si
Blood	int	Representaciones de sangre.	0 = no 1 = si
BloodandGore	int	Representaciones de mutilación o sangre en distintas partes del cuerpo.	0 = no 1 = si
Cartoon_Violence	int	Acciones violentas que involucran situaciones escenográficas y personajes de dibujos animados. Puede incluir violencia en la que un personaje sale ileso después de que se haya infligido la acción.	0 = no 1 = si
Crude_Humor	int	Representaciones o diálogos que involucren travesuras vulgares, incluido el humor de "baño".	0 = no 1 = si
DrugRe_ference	int	Referencia a y/o imágenes de drogas ilegales.	0 = no 1 = si
Fantasy_Violence	int	Acciones violentas de naturaleza fantástica, que involucran personajes humanos o no humanos en situaciones fácilmente distinguibles de la vida real.	0 = no 1 = si
Intense_Violence	int	Representaciones gráficas y realistas de conflictos físicos. Puede incluir sangre, gore, armas y representaciones extremas y/o realistas de heridas y muertes humanas.	0 = no 1 = si
Language	int	Uso moderado de blasfemias.	0 = no

			1 = si
Lyrics	int	Referencias a la blasfemia, la sexualidad, la violencia, el consumo de alcohol o drogas en la música.	0 = no 1 = si
Mature_Humor	int	Representaciones o diálogos que involucren humor "adulto", incluidas las referencias sexuales.	0 = no 1 = si
Mild_Blood	int	Un poco de sangre.	0 = no 1 = si
MildCartoonViolence	int	Algunas acciones violentas que involucran dibujos animados.	0 = no 1 = si
MildFantasyViolence	int	Algunas acciones violentas de naturaleza fantásica.	0 = no 1 = si
Mild_Language	int	Uso leve a moderado de blasfemias.	0 = no 1 = si
Mild_Lyrics	int	Leve referencias a blasfemias, sexualidad, violencia, consumo de alcohol o drogas en la música.	0 = no 1 = si
Mild_Violence	int	Algunas escenas que implican un conflicto agresivo.	0 = no 1 = si
No_Descriptors	int	Sin descriptores de contenido.	0 = no 1 = si
Nudity	int	Representaciones gráficas o prolongadas de desnudez.	0 = no 1 = si

Partial_Nudity	int	Representaciones breves y/o leves de desnudez.	0 = no 1 = si
Sexual_Content	int	Representaciones no explícitas de comportamiento sexual, que posiblemente incluyan desnudez parcial.	0 = no 1 = si
Sexual_Themes	int	Referencias al sexo o la sexualidad.	0 = no 1 = si
Strong_Language	int	Uso explícito y/o frecuente de malas palabras.	0 = no 1 = si
StrongSexualContent	int	Representaciones explícitas y/o frecuentes de comportamiento sexual, posiblemente incluyendo desnudez.	0 = no 1 = si
Suggestive_Themes	int	Referencias o materiales provocativos.	0 = no 1 = si
UseofAlcohol	int	The consumption of alcoholic beverages.	0 = no 1 = si
UseofDrugsandAlcohol	int	El consumo de bebidas alcohólicas y estupefacientes.	0 = no 1 = si
Violence	int	Escenas que implican un conflicto agresivo. Puede contener desmembramiento sin derramamiento de sangre.	0 = no 1 = si

Variable de salida

<i>ESRB_rating</i>	<i>Description</i>
RP	Calificación Pendiente
EC	NIñez temprana
E	Todo público
E 10+	Todo público 10+
T	Adolescente
M	Maduro
A	Adulto

Implementación

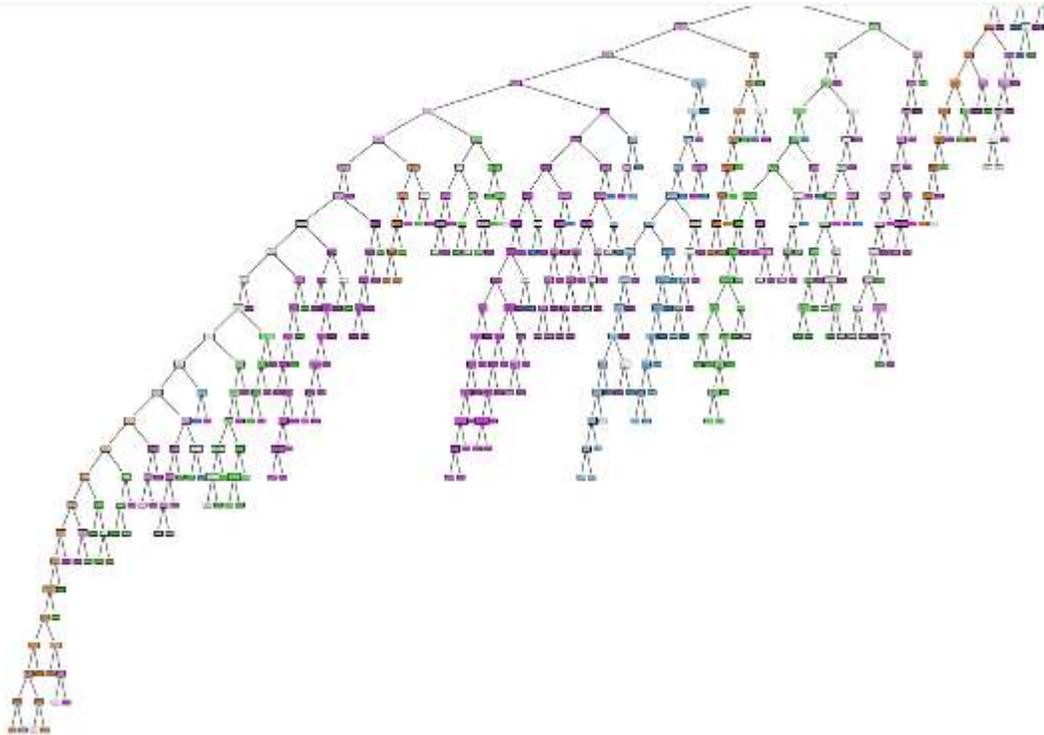
Fase de entrenamiento

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, train_size=0.75, random_state=0)
X_train.info()

[5] arbol = DecisionTreeClassifier();
arbol_cat = arbol.fit(X_train, Y_train)

from matplotlib import pyplot as plt
from sklearn import tree
fig = plt.figure(figsize=(25,20))
tree.plot_tree(arbol_cat, feature_names=list(X.columns.values), class_names=list(Y.values), filled=True)
plt.show()


```



En esta fase se tiene una cantidad enorme de datos, de la cual separamos una parte para entrenar al algoritmo y darle toda esta información para que encuentre los patrones necesarios y después pueda hacer predicciones.

Fase de prueba

El resto de los datos que quedan, se van a utilizar para hacer las pruebas. Así podemos desarrollar algunas preguntas al algoritmo y evaluar si las respuestas están bien o mal, y saber si está aprendiendo o no. Si vemos que no llegan a coincidir los datos, se tiene que añadir más datos o cambiar el método que se está utilizando. Pero si se observa que hay entre un 70% a 90% de respuestas correctas, podemos decir que hay un buen grado de aprendizaje y poder utilizar este algoritmo.

Análisis de resultados

Como se observa en la Figura la precisión después de la fase de entrenamiento es de 0.71 en 6 niveles y completo es de 0.86. A continuación, se muestra una comparación entre los árboles arrojados.

```
[17] from sklearn.metrics import confusion_matrix
      mdc = confusion_matrix(Y_test, Y_pred)
      mdc

array([[ 87,   0,   0,  18],
       [  9,  50,   0,  37],
       [  1,   4,  53,  42],
       [  6,  16,   2, 149]])
```

```
▶ import numpy as np
  pg = np.sum(mdc.diagonal())/np.sum(mdc)
  pg
```

0.7151898734177216

```
[ ] from sklearn.metrics import confusion_matrix
     mdc = confusion_matrix(Y_test, Y_pred)
     mdc
```

```
array([[102,  2,  0,  1],
       [ 4,  78,  0, 14],
       [ 0,  0,  87, 13],
       [ 2,  24, 15, 132]])
```

```
[ ] import numpy as np
     pg = np.sum(mdc.diagonal())/np.sum(mdc)
     pg
```

0.8417721518987342

Conclusiones

Los resultados obtenidos con el modelo de clasificación por árboles de decisión, indican que este es capaz de generar modelos consistentes con la realidad observada y el respaldo teórico, basándose únicamente en los datos obtenidos por el DataSet, con una precisión del 0,71 que depende de la cantidad de datos de entrenamiento que se le proporciona.

Como trabajos futuros se plantea utilizar otros clasificadores para comparar con estos resultados, además de ver que tan efectivo sería nuestro modelo, y para aumentar el porcentaje de evaluación se podría utilizar tanto un dataset que contenga más pruebas como otras técnicas de machine learning o aplicar tareas descriptivas de minería de datos como asociación y agrupación, con el fin de encontrar relaciones y similitudes.

Referencias

- [1] S. Belli and C. López, “A brief history of videogames,” *Athenea Digit. Rev. pensam. investig. soc.*, no. 14, p. 159, 2008.
- [2] H. Y. S. P. C. H. P. L. Videojuegos: Conceptos, “Historia y su potencial como herramientas para la educación.”
- [3] A. B. Fernández Revelles, “Sistemas de calificación de video juegos, revisión narrativa,” *ESPHA* 2018, p. 49839, 2018.
- [4] “Guía de clasificaciones,” *ESRB Ratings*. [Online]. Available: <https://www.esrb.org/ratings-guide/es/>. [Accessed: 23-Jun-2022].
- [5] Laczniak, R. N., Carlson, L., Walker, D., & Brocato, E. D. “Parental restrictive mediation and children's violent video game play: the effectiveness of the Entertainment Software Rating Board (ESRB) rating system,” *Journal of Public Policy & Marketing*, vol. 36, no 1, p. 70-78, 2017. [Abstract]. Available: *SageJournals*, <https://journals.sagepub.com/>. [Accessed June 22, 2022].
- [6] Stroud, N. J., & Chernin, A. “Video games and the ESRB: An evaluation of parental beliefs about the rating system,” *Journal of Children and Media*, vol. 2, no 1, p. 1-18, 2008. [Abstract]. Available: *TaylorFrancisOnline*, <https://www.tandfonline.com/>. [Accessed June 22, 2022].
- [7] H. Y. S. P. C. H. P. L. “Educación, VIDEOJUEGOS: CONCEPTO El efecto de los videojuegos en variables sociales, psicológicas y fisiológicas en niños y adolescentes”.

[8] González, M. T., Espada, J. P., & Tejeiro, R. (2016). El uso problemático de videojuegos está relacionado con problemas emocionales en adolescentes. *Adicciones*, 29(3), 180. <https://doi.org/10.20882/adicciones.745>

[9] García, A., & Martínez, G. L. (s/f). CLASIFICACIÓN SUPERVISADA INDUCCIÓN DE ARBOLES DE DECISIÓN, ALGORITMO k-d. Ipn.mx. Recuperado el 29 de junio de 2022, de <https://www.cic.ipn.mx/aguzman/papers/109%20Algoritmo%20KD.%20Clasificacion%20supervisada.pdf>.

[10] L. J. Sandoval Serrano, “Algoritmos de aprendizaje automático para análisis y predicción de datos,” *Revista Tecnológica*, no. 11, 2018.

[11] R. V. Baños and M. Torrado, “CÓMO APLICAR ÁRBOLES DE DECISIÓN EN SPSS. Cómo aplicar los árboles de decisión en SPSS. Applying SPSS decision trees.”.

[12] E. G. García and A. L. P. Pol, “Predicción del consumo de cocaína en adolescentes mediante árboles de decisión,” *Revista de investigación en educación*, vol. 6, no. 1, pp. 7–13, 2009.

[13] K. S. Daza Rosado and M. A. García Reyes, “Predicción, análisis y pronóstico de COVID-19 utilizando un modelo de Machine Learning basado en el análisis forecasting sobre series temporales,” Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería en Sistemas Computacionales., 2021.

[15] D. Blanco Álvarez, *Aprendizaje y evaluación de un dataset para predecir áreas de interés en una secuencia de vídeo*. 2021.

[14] Q. Wu, “geemap: A Python package for interactive mapping with Google Earth Engine,” *Journal of Open Source Software*, vol. 5, no. 51, 2020.

Anexos

<https://colab.research.google.com/drive/1E1K12bAVVTXY77fkcidt6Rb8Gc49N6c5?usp=sharing#scrollTo=pNKE5qVR7sGs>

<https://www.kaggle.com/datasets/imohtn/video-games-rating-by-esrb>

Roles de Autoría

Rodrigo Sebastián Huamán Maqqe: Conceptualización, Curación de datos, Investigación, Metodología, Software, Validación, Redacción - borrador original. **Graciela Condori Anahua:** Conceptualización, Curación de datos, Investigación, Metodología, Software, Validación, Redacción - borrador original. **Fátima Gigi Rojas Carhuas:** Conceptualización, Curación de datos, Investigación, Metodología, Software, Validación, Redacción - borrador original. **Rodolfo Robert Quispe Huacho:** Conceptualización, Curación de datos, Investigación, Metodología, Software, Validación, Redacción - borrador original.