



Universidad del Azuay

Facultad de Ciencias de la Administración

Escuela de Ingeniería de Sistemas y Telemática

**APLICACIÓN DE LA CIENCIA DE DATOS A
JUEGOS SERIOS ORIENTADOS A LA
ATENCIÓN Y MEMORIA.**

Trabajo de titulación previo a la obtención del
grado de Ingeniero de Sistemas y Telemática.

Autor:

Reinerio Alexandro Rodríguez García

Director:

Marcos Orellana Cordero

Cuenca – Ecuador

2022

DEDICATORIA

El presente trabajo investigativo lo dedico principalmente a Dios, por darme fuerza, paciencia, sabiduría y constancia para permitirme culminar este proceso de obtener uno de los anhelos más deseados en mi vida.

A mis padres Jaime Rodriguez y Leonor García, por su confianza, trabajo y sacrificio en todos estos años, gracias a ustedes he logrado llegar hasta aquí y poder culminar este proceso. A mis hermanos y hermana por estar siempre presentes apoyándome en esta etapa de mi vida.

A todas las personas que me han apoyado y han hecho que el trabajo se realice con éxito en especial a aquellos que me abrieron las puertas y compartieron sus conocimientos.

AGRADECIMIENTO

Primeramente, quiero agradecer a la Universidad del Azuay, directivos y maestros y en especial a mis tutores de este trabajo de titulación, ingeniero Marcos Orellana e ingeniera María Inés Acosta, quienes con sus conocimientos y apoyo me guiaron a través de cada una de las etapas de este proyecto para alcanzar los resultados que buscaba.

Por último, quiero agradecer a todos mis compañeros y a mi familia, por apoyarme aun cuando mis ánimos decaían. En especial, quiero hacer mención de mis padres, que siempre estuvieron ahí para darme palabras de apoyo y un abrazo reconfortante para renovar energías.

Muchas gracias a todos.

ÍNDICE

| | |
|---|-----|
| DEDICATORIA | I |
| AGRADECIMIENTO | II |
| ÍNDICE..... | III |
| RESUMEN | VI |
| ABSTRACT | VII |
| CAPÍTULO 1 | 1 |
| 1. INTRODUCCIÓN | 1 |
| 1.1 Motivación | 1 |
| 1.2 Problemática..... | 1 |
| 1.3 Objetivos generales | 2 |
| 1.4 Objetivos y específicos..... | 2 |
| 1.5 Preguntas de investigación | 2 |
| CAPÍTULO 2 | 4 |
| 2. TRABAJOS RELACIONADOS | 4 |
| CAPÍTULO 3 | 6 |
| 3. MARCO TEÓRICO | 6 |
| 3.1 Datos ex situ y caja negra..... | 6 |
| 3.2 Datos in situ y caja blanca..... | 6 |
| 3.3 Minería de datos | 7 |
| 3.4 Consideraciones de investigación del comportamiento | 7 |
| 3.5 Ciencia de los datos..... | 7 |
| 3.6 Clustering con K-means | 8 |
| 3.7 Métricas para validación de Clustering..... | 9 |
| 3.8 Normalización de los datos | 11 |
| CAPÍTULO 4 | 12 |
| 4 REVISIÓN SISTEMÁTICA DE LA LITERATURA (SLR) | 12 |
| 4.1 Método de investigación | 12 |
| 4.2 Etapa de Planificación | 12 |
| 4.3 Etapa de ejecución de la revisión | 17 |
| 4.4 Etapa de reporte..... | 27 |
| CAPÍTULO 5 | 28 |
| 5. APLICACIÓN DE LA CIENCIA DE LOS DATOS..... | 28 |
| 5.1 Metodología de desarrollo..... | 28 |
| CAPÍTULO 6 | 46 |
| 6. ANÁLISIS DE RESULTADOS..... | 46 |
| CAPÍTULO 7 | 47 |
| 7. ESTADO FINAL..... | 47 |
| BIBLIOGRAFÍA | 49 |
| ANEXOS | 52 |

Índice de tablas

| | |
|--|----|
| Tabla 1 Bibliotecas e indexadores seleccionados..... | 13 |
| Tabla 2 Cadena de búsqueda automática..... | 14 |
| Tabla 3 Evaluación de la calidad | 15 |
| Tabla 4 Criterios de extracción de los datos..... | 16 |
| Tabla 5 Valoración del Índice Kappa | 18 |

| | |
|---|----|
| Tabla 6 Matriz de resultados | 19 |
| Tabla 7 Tabla de atributos y descripción de datos recopilados del juego de pares. | 30 |
| Tabla 8 Clasificación de datos mediante la edad..... | 31 |
| Tabla 9 Estadística de puntajes de participantes | 33 |
| Tabla 10 Distribución de los datos de participación en las diferentes categorías. | 33 |
| Tabla 11 Resultado de diagrama de caja y bigotes referentes a edades categorizadas y puntos obtenidos en el juego | 34 |
| Tabla 12 Número de intentos y cantidad de personas de la categoría Infancia..... | 35 |
| Tabla 13 Número de intentos y cantidad de personas de la categoría Adolescentes..... | 36 |
| Tabla 14 Número de intentos y cantidad de personas de la categoría Adultos. | 37 |
| Tabla 15 Número de intentos y cantidad de personas de la categoría Adultos mayores | 37 |
| Tabla 16 Cantidad de datos clasificados por edad y Puntajes | 38 |
| Tabla 17 Cantidad de datos clasificados por edad y tiempo (Primera fase)..... | 39 |
| Tabla 18 Cantidad de datos clasificados por edad y tiempo (segunda fase) | 39 |
| Tabla 19 Cantidad de datos clasificados por edad y tiempo (tercera fase)..... | 40 |
| Tabla 20 Cantidad: Edad y tiempo (cuarta fase) | 41 |
| Tabla 21 Centroides de datos normalizados y clasificados. | 42 |
| Tabla 22 Abreviaturas utilizadas y sus significados..... | 42 |
| Tabla 23 Centroides de datos normalizados y clasificados del primer intento del juego. | 43 |
| Tabla 24 Resultados de datos por categorías de la Figura 25..... | 43 |
| Tabla 25 Centroides de datos normalizados y clasificados del intento dos del juego.... | 44 |
| Tabla 26 Resultados de datos por categorías del intento dos (Figura 26) | 44 |
| Tabla 27 Centroides de datos normalizados y clasificados del intento tres del juego.... | 44 |
| Tabla 28 Resultados de datos por categorías del intento tres (Figura 27)..... | 45 |

Índice de figuras

| | |
|---|----|
| Figura 1 Método de codo..... | 11 |
| Figura 2 Proceso de selección de estudios primarios | 19 |
| Figura 3 Relevancia de los artículos por pregunta de investigación (RQ) | 20 |
| Figura 4 Áreas de aplicación usando preprocesamiento de datos | 22 |
| Figura 5 Tipos de usuarios en los que se ha usado preprocesamiento de datos | 22 |
| Figura 6 Áreas de aplicación en los que se ha usado alguna clasificación de los datos. 24 | |
| Figura 7 Tipos de usuarios en los que se ha usado algún método/técnica de la ciencia de los datos | 25 |
| Figura 8 Áreas de aplicación en donde se han aplicado algún algoritmo de agrupamiento. | 26 |
| Figura 9 Tipos de usuarios en los que se ha aplicado algún algoritmo de agrupamiento. | 26 |
| Figura 10 Clasificación de los jugadores en donde se han aplicado algún algoritmo de agrupamiento. | 27 |
| Figura 11 Metodología de desarrollo según el modelo de transferencia tecnológica ... | 29 |
| Figura 12 Metodología para encontrar patrones de conocimiento en datos provenientes de juegos serios..... | 30 |
| Figura 13 Frecuencia de puntajes obtenidos por los participantes | 33 |
| Figura 14 Diagrama de caja y bigotes de categoría y puntajes de los participantes..... | 34 |
| Figura 15 Intentos realizados de los participantes de la categoría Infancia. | 35 |

| | |
|--|----|
| Figura 16 Intentos realizados de los participantes de la categoría adolescentes | 36 |
| Figura 17 Intentos realizados de los participantes de la categoría Adultos..... | 37 |
| Figura 18 Intentos realizados de los participantes de la categoría Adultos Mayores..... | 37 |
| Figura 19 Clúster de edad y puntajes obtenidos por los participantes..... | 38 |
| Figura 20 Clasificación de datos por edad y tiempo de la primera fase..... | 39 |
| Figura 21 Clasificación de datos por edad y tiempo de la segunda fase..... | 39 |
| Figura 22 Clasificación de datos por edad y tiempo de la tercera fase..... | 40 |
| Figura 23 Clasificación de datos por edad y tiempo de la cuarta fase..... | 41 |
| Figura 24 Gráfico radial de la aplicación de clustering con datos de centroides de las variables demográficas..... | 42 |
| Figura 25 Gráfico de centroides agrupados del intento 1 del juego..... | 43 |
| Figura 26 Gráfico de centroides agrupados del intento 2 del juego..... | 44 |
| Figura 27 Gráfico de centroides agrupados del intento 3 del juego..... | 44 |

Índice de anexos

| | |
|--|----|
| Anexo 1 Lista de artículos primarios seleccionados..... | 52 |
| Anexo 2 Resultados de la evaluación a cada artículo con cada pregunta (RQ1) y sus criterios de extracción..... | 55 |
| Anexo 3 Resultados de la evaluación a cada artículo con cada pregunta (RQ2 y RQ3) y sus criterios de extracción..... | 57 |

RESUMEN:

Los juegos serios se caracterizan por estar diseñados especialmente con fines formativos (educativos, entrenamiento e información), permitiendo al usuario aprender de sus errores a través de la retroalimentación inmediata y adquirir experiencia. En este contexto, la Universidad de Cuenca ha desarrollado una serie de juegos en el ámbito de la atención y memoria cognitiva. Estos juegos serios fueron probados por personas de distintas edades y sexo. Dichas aplicaciones, al momento de su ejecución recopilan los datos sociodemográficos de cada jugador. Los datos no han sido tratados para la realización de búsqueda de algún tipo de comportamiento. Frente a estas circunstancias, se realiza una revisión sistemática de la literatura (SLR) y la elaboración y validación de una metodología para la aplicación de técnicas/métodos de la ciencia de los datos con el objetivo de encontrar patrones de comportamiento orientados a la atención y memoria cognitiva. Los resultados obtenidos del análisis de datos servirán para promover el desarrollo y bienestar cognitivo de las personas.

Palabras clave: Revisión sistemática de la literatura (SLR), juegos serios, ciencia de los datos, análisis, minería de datos.

ABSTRACT:

Serious games are characterized by being specially designed for educational purposes (educational, training and information). In this context, the University of Cuenca has developed a series of serious games in the field of attention and cognitive memory. These serious games were tested by people of any age and gender. These applications, at the time of their execution, collect the sociodemographic data of each player. The data has not been processed to carry out a search for any type of behavior. Faced with these circumstances, in the present study a systematic review of the literature (SLR) and the elaboration of a methodology for the application of techniques / methods of data science with the aim of finding patterns of oriented behavior is carried out. to attention and cognitive memory. The results obtained from the data analysis will serve to promote the development and cognitive well-being of people.

Keywords: Systematic literature review (SLR), serious games, data science, data analysis, data mining.



Translated by

Reinerio Rodríguez García



CAPÍTULO 1

1. INTRODUCCIÓN

1.1 Motivación

El término de “juegos serios” fue acuñado en 1970 por el autor del libro *Serious Games* (Nguyen, 2016) en donde se explora las diferentes formas en las que los juegos pueden incluir en el proceso de enseñanza-aprendizaje sin eliminar el entretenimiento. Los juegos serios permiten aprender de los errores a través de retroalimentación inmediata y adquirir experiencia sin poner en riesgo la seguridad e integridad física (Sandí Delgado & Bazán, 2019)

A lo largo de nuestra vida, las personas necesitan estar en un continuo aprendizaje; sin embargo, se tiene etapas en las cuales dicho aprendizaje debe ser adaptado a las circunstancias y a la edad. Actualmente, existen juegos serios que han adquirido gran popularidad en los últimos años y han sido adoptados por instituciones educativas y empresas, especialmente para programas de formación (Nguyen, 2016).

Para evidenciar una correlación entre los datos de los juegos serios, existe una serie de técnicas y métodos de la ciencia de los datos (Hernández et al., 2004) para descubrir patrones de comportamiento basados en un conjunto de datos y en consecuencia la generación de conocimiento. Es posible construir modelos de predicción más precisos con la finalidad de plantear hipótesis o proporcionar conocimiento relacionado a juegos serios orientados a la atención y memoria. Esta información puede ser valiosa para promover el desarrollo y bienestar de la población (Sandí Delgado & Bazán, 2019).

1.2 Problemática

Los datos generados provenientes de juegos serios no tienen mayor significancia si no son analizados, es por eso, que existe la necesidad de que los datos sean examinados con el fin de medir, evaluar y mejorar el rendimiento (Loh et al., 2015). En este entorno, los juegos serios rastrean las acciones y comportamiento de los jugadores.

En este contexto, la Universidad de Cuenca ha desarrollado una serie de juegos para el entrenamiento cognitivo en el ámbito de la atención y memoria (Prado et al., 2020). Dichas aplicaciones, al momento de su ejecución recopilan los datos del test (evaluación), los datos del usuario (información personal) e información relacionada al tiempo de

juego, como, por ejemplo: la precisión de aciertos y la cantidad de veces que se ha ejecutado la aplicación; información que es almacenada en una base de datos. Con el fin de analizar y extraer información de este conjunto de datos, existe la necesidad de relacionarlos con la ciencia de datos para hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados (Hernández et al., 2004).

El presente estudio, busca dar apoyo al tratamiento de información orientada a la salud integral de la población, a través del análisis de datos provenientes del uso de juegos serios orientados a la atención y memoria cognitiva, con el objetivo de buscar patrones recurrentes con las variables de los jugadores y los resultados obtenidos en cada juego; información que puede utilizarse para promover el desarrollo y bienestar cognitivo de la población.

1.3 Objetivos generales

Elaborar y validar una metodología en el ámbito de la ciencia de los datos para extraer patrones de comportamiento en base a datos provenientes de juegos serios orientados a la atención y memoria.

1.4 Objetivos y específicos

1. Realizar una revisión sistemática (SLR) para estructurar el estado del arte sobre estudios de algoritmos, métodos y técnicas de la ciencia de los datos aplicado al dominio de juegos serios.
2. Diseñar una metodología para identificar patrones en datos de juegos serios orientados a la atención y memoria cognitiva.
3. Desarrollar un modelo computacional para identificar patrones de comportamiento en datos de juegos serios orientados a la atención y memoria cognitiva.
4. Evaluar la propuesta de solución.
5. Concentrar y escribir los resultados obtenidos de la técnica utilizada.

1.5 Preguntas de investigación

- ✓ ¿Cómo se abordan los juegos serios en la ciencia de los datos?

- ✓ ¿En qué dominios se ha implementado la ciencia de datos orientada a juegos serios?
- ✓ ¿Qué técnicas/métodos se han implementado y cuáles son las más efectivas para el tratamiento de datos provenientes de juegos serios orientados a la atención y memoria cognitiva?

CAPÍTULO 2

2. TRABAJOS RELACIONADOS

Estudios previos han evaluado el proceso de aprendizaje de los estudiantes en juegos educativos o simulaciones (Barab et al., 2009), existe una escasa investigación empírica sobre análisis en profundidad de patrones de comportamiento.

Según la investigación (Benmakrelouf et al., 2015), se centra en la identificación de datos relevantes y métodos apropiados de minería de datos, a través de regresión lineal múltiple y técnica K-means; la cual como resultado explora las relaciones entre las características de los jugadores y su desempeño.

Existe un estudio que usa datos de juegos para examinar patrones de comportamiento de aprendizaje (Kang et al., 2017), cuyo objetivo es mejorar el rendimiento del aprendizaje mediante la resolución de problemas. El mismo, pone en práctica la técnica de minería de patrones secuenciales, análisis estadístico y un enfoque analítico. Este estudio se realizó a estudiantes escolares, lo que hace que esté enfocado a un público cerrado.

Según el estudio realizado por el autor Slimani et al. (2018) se pone en práctica un juego serio, el cual, es utilizado en el área médica con la finalidad de enseñar a los estudiantes de biología la técnica inmunológica para la determinación de anticuerpos del virus VIH. El estudio tiene como objetivo describir, analizar y aumentar el interés del aprendizaje y evolución del desempeño de los usuarios a través de juegos serios, el mismo que pone a consideración dos métodos y enfoques de minería de datos educativos aplicando el algoritmo de maximización de expectativas y K-Means.

Se han utilizado diferentes métodos de minería de datos aplicando analítica de juegos serios. Sin embargo, estos métodos y técnicas aplicados tienen limitaciones en el contexto de la atención y memoria, a la vez, están orientados a un público específico. Dados estos desafíos, se ve la necesidad de un procesamiento y búsqueda de conocimiento más exhaustivo mediante la aplicación de la ciencia de los datos para encontrar patrones de comportamiento en nuevos datos provenientes en nuevas intervenciones con pacientes que realizan entrenamiento cognitivo, y así dar la posibilidad de búsqueda de soluciones y nuevas aplicaciones que sirvan de soporte al profesional, solución que puede

extrapolarse a otros dominios y juegos creados de ayuda a personas que requieran este tipo de entrenamiento.

Los juegos serios también ofrecen aprendizaje, adquisición de conocimiento, cambios efectivos de comportamiento, así como la adquisición de habilidades y destrezas cognitivas (Boyle et al., 2016)

CAPÍTULO 3

3. MARCO TEÓRICO

3.1 Datos ex situ y caja negra

Los datos ex situ se recopilan "fuera del sistema", por ejemplo, los datos que se obtiene de las encuestas de usuarios (datos demográficos, comentarios) pertenecen a esta categoría porque normalmente se recopilan en el mundo real y no dentro de los entornos del juego. Por lo general, los datos de investigación se recopilan ex situ por conveniencia o debido a limitaciones.

El enfoque ex situ significa que los investigadores tratarán los juegos serios como una caja negra y recopilan datos del usuario antes y después de que los usuarios interactúen con los juegos serios. Si bien aún se puede detectar un cambio en el desempeño a través de medios estadísticos, los hallazgos de los datos ex situ no pueden explicar de manera efectiva cómo los contenidos del juego afectan los cambios en el desempeño en los aprendices del juego (Loh, 2012).

3.2 Datos in situ y caja blanca

Al contrario de ex situ, el término in situ significa "en el lugar" en latín. Por lo tanto, los datos in situ se recopilan "dentro del hábitat o sistema natural" en el ámbito en el que se está investigando un evento. Los investigadores de juegos expertos en programación ven los juegos serios como una aplicación de software más. Por lo tanto, se acercan a los juegos serios como una caja blanca, abierta a la manipulación de contenido y la recopilación de datos generados por el usuario in situ. Un programador, puede incluso crear agentes de software y usarlos para automatizar el proceso de recopilación de datos in situ. Obviamente, los métodos de recopilación de datos in situ son preferibles a los ex situ porque eliminan la gran cantidad de "datos subjetivos" obtenidos de encuestas, entrevistas y autoinformes, que simplemente no se pueden utilizar en investigaciones de alto nivel (Quellmalz et al., 2009). Una vez que las acciones de los aprendices del juego han sido capturadas desde dentro de los juegos serios, los investigadores pueden rastrear qué acciones realizaron los jugadores dentro del juego, visualizar sus rutas de navegación y dar sentido a patrones interesantes en ellos (see Loh, 2012b; Scarlatos & Scarlatos, 2010; Thawonmas & Iizuka, 2008).

3.3 Minería de datos

La minería de datos es el proceso de analizar datos desde diferentes perspectivas y resúmenes, marcando los resultados como información útil. Se ha definido como el proceso de identificar válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles patrones en los datos. El proceso de minería de datos educativos implica varios métodos para el análisis del aprendizaje, tales como predicción, descubrimiento de estructuras, minería de relaciones y descubrimiento con modelos (Fayyad et al., 1996).

Existen diferentes tecnologías que pueden usar datos, extraer información que no se puede detectar a simple vista. Una de estas tecnologías es la denominada minería de datos, que combina tecnología semiautomática como inteligencia artificial, análisis estadístico, visualización de bases de datos y gráficos para obtener información que no esté representada o tomada en cuenta en los datos. Mediante la aplicación de minería de datos se encuentran comportamientos atípicos, relaciones, tendencias y patrones, así apoyando el proceso de toma de decisiones con un soporte de conocimiento.

3.4 Consideraciones de investigación del comportamiento

Debido a que el propósito es investigar el nivel de atención y memoria de los jugadores de juegos serios, no se toma en cuenta autoinformes y las evaluaciones de tipo encuesta porque se sabe que los participantes informan lo que creen que el investigador quiere escuchar en lugar de sus propias creencias (también conocido como sesgo de deseabilidad social según el autor Paulhus (1991)). Además, se ha descubierto que las acciones de las personas difieren de lo que dicen / afirman que harían, ya sea por reclamar en exceso (see Roese & Jamieson, 1993) u otras razones (ver Fan et al., 2006; Hoskin, 2012). Un enfoque más objetivo rastrea lo que los jugadores realmente hacen directamente dentro del entorno del juego (es decir, medición in situ) como evidencia empírica.

3.5 Ciencia de los datos

En la revisión sistemática (Fernández et al., 2019), menciona que existe algoritmos y técnicas de ciencia de datos utilizados en los estudios revisados con respecto a juegos serios y se pueden agrupar en tres categorías principales:

- ✓ **Algoritmos supervisados:** regresión lineal y logística, árboles de regresión y decisión, máquinas de vectores de soporte, redes bayesianas, redes neuronales, Bayes ingenuo y rastreo del conocimiento Bayesiano.
- ✓ **Algoritmos no supervisados:** correlación, agrupamiento, análisis factorial.
- ✓ **Técnicas de visualización:** visualización de rutas de juego, métricas de rendimiento, curvas de aprendizaje, mapas de calor de interacciones, uso de herramientas del juego (frecuencia o duración).

3.6 Clustering con K-means

El análisis de conglomerados (Tan, Steinbach & Kumar, 2006) es un método de análisis de datos exploratorio para agrupar objetos similares en categorías. Cada objeto se puede ver en un espacio n -dimensional de modo que se puedan calcular las distancias entre los objetos, donde n es el número de características (variables) en las que se basa la agrupación. Por ejemplo, las ciudades se pueden agrupar según en la información del censo, como la densidad de población, los ingresos, la edad, entre otros; para identificar ciudades con datos demográficos similares características. Al mismo tiempo, las variables del censo se pueden agrupar para encontrar variables con distribuciones similares en las ciudades. Allí hay diferentes tipos de medidas de distancia (por ejemplo, distancia euclidiana, distancia euclidiana al cuadrado y distancia de Manhattan) para calcular las distancias entre objetos. También hay diferentes reglas de vinculación (p. Ej., Vinculación única, vinculación completa, y promedio de grupos de pares no ponderados) que, junto con las medidas de distancia, se utilizan para calcular las distancias entre racimos. Se pueden distinguir diferentes tipos de análisis de conglomerados según sus propósitos y algoritmos (p. Ej., árbol de agrupación, unión bidireccional, agrupación de k -medias y agrupación de expectativas y maximización). Sin embargo, todo tipo de análisis de conglomerados tiene un objetivo común, el cual es minimizar la relación entre las distancias dentro del conglomerado y las distancias entre los conglomerados.

Kerr, Chung e Iseli (2011) aplicaron el análisis de conglomerados para identificar las estrategias de los estudiantes al resolver problemas de suma de fracciones en *Save Patch*, un juego serio diseñado para enseñar la suma de fracciones. En *Save Patch*, el personaje del juego, llamado *Patch*, necesita rebotar obstáculos para llegar a su casa; Se requiere que los estudiantes apliquen conceptos subyacentes a la suma racional para completar tareas.

Dada la amplia cobertura y la clara interpretabilidad de los patrones de acción identificados por los análisis de conglomerados en su estudio, Kerr et al(2011) concluyó que el análisis de conglomerados puede ser una herramienta válida para analizar los datos de procesos generados a partir de la resolución de problemas complejos en juegos serios o simulaciones, para detectar estrategias de aprendizaje y resolución de problemas significativas. A su vez, los resultados se pueden utilizar para diagnosticar los errores de los estudiantes y proporcionar una rápida corrección e instrucción personalizada a través de juegos o simulaciones.

K-means es un algoritmo de clasificación (agrupamiento) no supervisado, que divide los objetos en k grupos de acuerdo con sus características. La agrupación de los datos se realiza mediante la minimización de la suma de las distancias entre cada objeto y el centroide de su grupo. A menudo se utiliza la distancia euclidiana.

El algoritmo consiste en seguir estos tres pasos:

Inicialización: Una vez seleccionado el número de grupos k , se establecen los k centroides en el espacio de datos, por ejemplo, se seleccionan aleatoriamente.

Asignar objetos a centroides: cada objeto de los datos se asigna a su centroide más cercano.

Actualización del centroide: actualice la posición del centroide de cada grupo y use la posición promedio de los objetos que pertenecen al grupo como el nuevo centroide. Repita la asignación de objetos a los centroides y actualizar el centroide hasta que el centro de masa no se mueva o se mueva por debajo de la distancia de umbral en cada paso.

3.7 Métricas para validación de Clustering

Existen dos categorías de validaciones de clustering más relevantes: externa y la validación interna. La primordial diferencia es si se utiliza o no información externa para la validación, o sea, información que no es producto de la técnica de agrupación usada.

La validación interna mide sólo el clustering fundamentadas en la información de los datos. Evaluando que tan buena es la composición del clustering sin necesidad de información ajena al propio algoritmo y su resultado.

La validación externa mide la calidad del agrupamiento conociendo información externa, es utilizada primordialmente para elegir un algoritmo de clustering óptimo sobre un conjunto de datos en concreto. Las métricas de validación interna tienen la posibilidad de usarse para elegir el mejor algoritmo de clustering, así como el número de clúster óptimo sin ningún tipo de información adicional.

Método de codo para el valor óptimo de k en K-Means

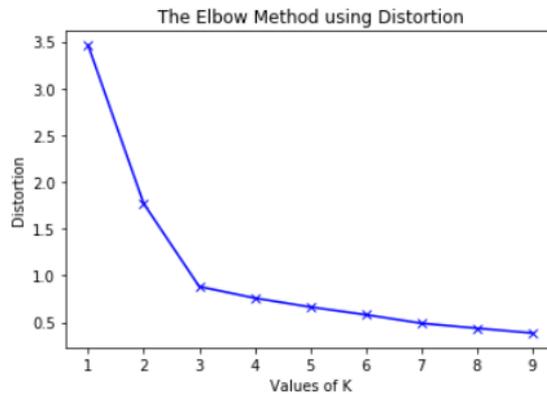
Los algoritmos no supervisados deben necesariamente establecer un número óptimo de agrupaciones en las que tienen la posibilidad de agrupar los datos. El procedimiento del codo pertenece a los procedimientos más populares para decidir este costo óptimo de k . Si el gráfico de líneas se asemeja a un brazo, entonces el "codo" (el punto de inflexión en la curva) es una buena indicación de que el modelo subyacente se acomoda mejor en aquel punto.

Pasos para crear el método de codo

- ✓ **Distorsión:** Este valor se obtiene mediante el cálculo del promedio de las distancias al cuadrado a partir de los centros de los respectivos grupos de datos. Comúnmente, se usa la métrica de distancia euclidiana.
- ✓ **Inercia:** Se obtiene mediante la suma de las distancias al cuadrado de las muestras a su centro de conglomerado más cercano.

Para decidir el número óptimo de grupos de datos, tenemos que elegir el costo de k en el "codo", o sea, el punto a partir del cual la distorsión / inercia empieza a reducir o incrementar de manera lineal. Por consiguiente, el número óptimo de conglomerados para el costo de k de la figura 1 es el costo de tres.

Figura 1
Método de codo



Fuente: <https://aprendeia.com>

3.8 Normalización de los datos

La normalización de los datos es el proceso de estandarización de los datos, permitiendo que los datos sean comparables, debido a que originalmente tienen la posibilidad de tener diferentes dimensiones y escalas. Existiendo la posibilidad de convertirlos o transformarlos para poder generar agrupaciones de buena calidad y mejorar la precisión de los algoritmos de agrupación.

3.8.1 Métodos

Aquí se proporcionan los métodos más utilizados para normalizar los datos.

Transformación Z: Esta normalización realiza la resta de la media de los datos de todos los valores y después realiza una división por la desviación estándar. A continuación, la distribución de los datos tiene una media de cero y una varianza de uno. Esta técnica conserva la distribución original de los datos y está menos influenciada por los valores atípicos.

Transformación de rango: la transformación de rango normaliza todos los valores de atributo a un rango de valores especificado. Cuando se escoge este método, se puede establecer dos parámetros (min, max). Entonces, el valor más grande se establece en “máx” y el valor más pequeño se establece en “min”. Todos los demás valores se escalan, por lo que se ajustan al rango dado. Este método puede ser influenciado por valores atípicos, porque los límites se mueven hacia ellos. Por otro lado, este método mantiene la distribución original de los puntos de datos, por lo que también se puede usar para el anonimato de datos.

CAPÍTULO 4

4 REVISIÓN SISTEMÁTICA DE LA LITERATURA (SLR)

4.1 Método de investigación

Las etapas que tendrá el mapeo sistemático son las siguientes

- Planificación
- Realización
- Informes

4.2 Etapa de Planificación

La etapa de planificación consta de los siguientes pasos:

- Establecer las preguntas y las sub-preguntas de investigación
- Definir la estrategia de búsqueda
- Seleccionar los estudios primarios
- Realizar una evaluación de la calidad
- Definir una estrategia de extracción de datos
- Seleccionar los métodos de resumen

4.2.1 Pregunta de investigación

Las preguntas de investigación son el soporte que servirá a lo largo de este proceso de estudio que permitirán encontrar datos relevantes y dotar de contribución a la investigación. Con este fin, se involucraron tres variables: los juegos serios, la ciencia de los datos y la atención y memoria de las personas.

- ✓ ¿Cómo se está llevando la investigación de soluciones de ciencia de los datos aplicadas a los juegos serios orientados a la atención y memoria?
- ✓ ¿Qué tipo de información es requerido para analizar juegos serios?
- ✓ ¿Cómo se abordan los juegos serios en la ciencia de datos?
- ✓ ¿En qué área de desarrollo (industria, empresa, academia) se aplican juegos serios?

4.2.2 Estrategia de búsqueda

Las bibliotecas digitales a usar para la búsqueda de información se presentan en la Tabla 1, estas bibliotecas fueron elegidas puesto que cubren gran cantidad de artículos relacionados a los juegos serios orientados a la atención y memoria.

Tabla 1
Bibliotecas e indexadores seleccionados

| Biblioteca digital | URL |
|------------------------|---|
| ACM – Digital Library. | https://dl.acm.org/ |
| IEEE Xplore | https://ieeexplore.ieee.org/Xplore/home.jsp |
| Taylor & Francis | https://www.tandfonline.com/ |
| Scopus | https://www.scopus.com/search/form.uri?display=basic#basic |
| Hinari (OARE) | https://www.who.int/hinari/es/ |
| We of science | https://www.webofknowledge.com |
| Springer Link | https://link.springer.com/ |
| Science direct | https://www.sciencedirect.com/ |
| EBSCO | https://www.ebsco.com/es |

Cadena de búsqueda

Las cadenas de búsqueda (Tabla 2) se ajustaron a la sintaxis de cada una de las bibliotecas digitales antes citadas, y se aplicó la cadena de búsqueda a los mismos metadatos, es decir, título, resumen y palabras clave de cada artículo para todos los orígenes (la sintaxis de la cadena de búsqueda se adaptará para que se aplique en cada biblioteca digital). Estos términos de búsqueda también se tuvieron en cuenta en las otras fuentes que se inspeccionarán manualmente para realizar una búsqueda coherente.

Para constituir la cadena de búsqueda, se utilizaron los términos más relevantes, mediante la utilización de conectores “AND” y “OR”, obteniendo como consecuencia la siguiente cadena, la misma que se adecuará dependiendo de cada biblioteca:

Abstract: Serious games AND (data mining OR text mining OR data science OR learning analytics)

Tabla 2*Cadena de búsqueda automática*

| Término relevante | Conectores |
|--------------------------|-------------------|
| Serious games | AND |
| Data mining | OR |
| text mining | OR |
| Data science | OR |
| learning analytics | |

4.2.3 Criterios de exclusión e inclusión

Para la selección de los artículos más importantes de los resultados obtenidos, se requiere establecer criterios de inclusión y exclusión, esto permite disminuir la cantidad de estudios primarios seleccionados. Se consideraron los siguientes criterios de inclusión para decidir si un artículo es viable:

- Artículos que describen la aplicación de la ciencia de los datos orientado a juegos serios.
- Artículos que describen los juegos serios orientados a la atención y memoria.
- Artículos que describen las técnicas de minería de datos a juegos serios.
- Artículos que describan las la metodología para la ciencia de datos orientados a los juegos serios.
- Artículos que tengan relación con el tema y que hayan sido publicados desde el año 2002. Aquel periodo se conoció como la ola de los juegos serios, debido al lanzamiento y descarga masiva de dos juegos serios: el primero fue *Improving public policy through game-based learning and simulation report* (Sawyer & Rejeski, 2002), el segundo, un juego de guerra encargado por el ejército de los Estados Unidos para mostrar la vida militar como una experiencia interesante, informativa y entretenida (Loh et al., n.d.) , (Chalodgeridis & Tsiatsos, 2020).

Los criterios de exclusión establecidos que se utilizarán para descartar a un artículo de la lista son:

- Artículos duplicados del mismo estudio en diferentes fuentes.
- Documentos introductorios para temas especiales, libros y talleres.

- Artículos que no estén escritos en el idioma inglés.
- Artículos que solo están disponibles como presentaciones, resúmenes o estén incompletas.
- Artículos sin diseño de investigación como workshops, encuestas u otros sin preguntas de investigación bien definidas.
- Publicaciones que no se han sometido a un proceso de revisión formal, o informes técnicos.
- Publicaciones duplicadas referidos a un mismo estudio en otras bibliotecas digitales.
- Publicaciones cortas de menos de cinco páginas.
- Literatura que no disponen de un DOI.

4.2.4 Evaluación de la calidad

Para la evaluación de la calidad de los artículos encontrados, se realizó una serie de preguntas sobre los artículos que permita realizar una clasificación de su calidad por la asignación de puntos dependiendo de la respuesta a cada pregunta.

Las preguntas a realizarse, junto con sus respectivas respuestas, está representada en la siguiente tabla.

Tabla 3

Evaluación de la calidad

| No. | Preguntas | Respuestas y puntajes |
|-----|---|--|
| QA1 | El estudio presenta temas sobre minería de datos (ciencia de datos) en juegos serios. | De acuerdo (+1) Parcialmente (0) En desacuerdo (-1) |
| QA2 | El estudio se ha publicado en una revista o conferencia relevante. | Muy relevante (+1) Relevante (0) No relevante (-1) |
| QA3 | El estudio ha sido citado por otros autores. | Sí, más de 5 (+1) Parcialmente, de 1 a 5 (0) No, no tiene citas (-1) |

4.2.5 Estrategia de extracción de datos

Se aplicarán los criterios de extracción de la Tabla 4 a los artículos, con el fin de facilitar la clasificación y la sintetización de los resultados obtenidos.

Esta lista de estrategias de extracción de datos se muestra en la Tabla 4, la cual se desarrolló basado en la información que debería incluirse para responder las preguntas de investigación con el objeto de brindar un control de calidad por cada estudio. Además, cada pregunta fue respondida dividiéndola en sub-preguntas, cada una de las sub-

preguntas fue reconocida con el prefijo EC ##. Los estudios que hayan aprobado los criterios de exclusión e inclusión se establecen por medio de una codificación determinada en el siguiente formato: [A + número secuencial Creador Título].

Tabla 4

Criterios de extracción de los datos

| RQ1: ¿Qué tipo de información es requerido para analizar juegos serios? | | | |
|--|------------------------------------|--|---|
| EC1 | Lugar de despliegue | [Consola, app, web, otras] | (Loh et al., 2015) (Chalodgeridis & Tsiatsos, 2020) |
| EC2 | Plataformas de despliegue | [Computadora, teléfono, Tablet] | (Zhang & Lu, 2014) (Chalodgeridis & Tsiatsos, 2020) |
| EC3 | Área de aplicación | [Salud, educación, comercio, otros] | (Bente & Breuer, 2010) (Chalodgeridis & Tsiatsos, 2020) |
| EC4 | Rango de edades | [5-17 años, 18-64 años, más de 65 años] | (Bente & Breuer, 2010) |
| EC5 | Tipos de herramienta de desarrollo | [Metodología, programa, framework, otros] | (Slimani, Elouaai, Elaachak, Yedri, & Bouhorma, 2018) |
| EC6 | Modos de interacción | [Multijugador, un jugador, co-tutoría, otros] | (Bente & Breuer, 2010) |
| EC7 | puntajes | [Si, No] | (Loh et al., n.d.) (Bente & Breuer, 2010) |
| EC8 | Clasificación de jugadores | [Puntaje, edades, otros] | (Hou, 2015), (Loh et al., 2015) |
| RQ2: ¿Cómo se abordan los juegos serios en la ciencia de datos? | | | |
| EC9 | Preprocesamiento de datos | [Medida de distancia, muestreo, reducción de dimensionalidad] | (Loh et al., 2015) |
| EC10 | Clasificación | [KNN (máquinas de vector soporte), árboles de decisión, redes bayesianas, SVN (vecinos más cercanos), ANN (redes neuronales artificiales)] | (Duarte, Berns, Yañez, & Dodero, 2015) (Hernandez, Duarte, & Dodero, 2017) |
| EC11 | Clustering | [k-means, basado en densidad, paso de mensajes Jerárquico] | (Zhang & Lu, 2014) |
| EC12 | Cómo se almacenan los datos | [Json, SQL, no SQL] | (Loh et al., n.d.) |
| RQ3: ¿En qué área de desarrollo se aplican juegos serios? | | | |
| EC13 | Áreas de despliegue | [Industria, empresa, academia, medicina] | (Duarte, Berns, Yañez, & Dodero, 2015) (Chalodgeridis & Tsiatsos, 2020) |
| RQ4: ¿Cómo está la investigación? | | | |
| EC14 | Tipo de evaluación del estudio | [Análisis, diseño, implementación, pruebas] | (Duarte, Berns, Yañez, & Dodero, 2015) (Chalodgeridis & Tsiatsos, 2020) |
| EC15 | Tipo de validación | [Experimento controlado, caso de estudio, prueba de conceptos] | (Duarte, Berns, Yañez, & Dodero, 2015) |
| EC16 | Ámbito de desarrollo | [Industria, academia] | (Duarte, Berns, Yañez, & Dodero, 2015) |
| EC17 | Continuidad del estudio | [Nueva, extensión] | |

4.2.6 Métodos de síntesis

Se aplica métodos de síntesis cuantitativos y métodos de síntesis cualitativos.

La síntesis cuantitativa se basa en el conteo del número de estudios primarios que se pueden clasificar en cada respuesta a las sub-preguntas de investigación. Se definen

gráficos de burbuja que indican la frecuencia entre resultados de las sub-preguntas de investigación y contar el número de artículos encontrados en cada fuente bibliográfica por año. Los diagramas de burbujas representan un procedimiento de síntesis eficaz para dar un mapa y ofrecer una perspectiva instantánea de un campo de indagación (Petersen, Feldt, Mujtaba & Mattson, 2008)

La síntesis cualitativa se basa en la inclusión de varios estudios representativos para cada sub-pregunta de investigación, considerando los resultados de la evaluación de calidad.

4.3 Etapa de ejecución de la revisión

Con la aplicación del protocolo de revisión, se generan resultados primarios, los cuales se seleccionan mediante los criterios de inclusión definidos anteriormente. Con dichos resultados, se genera una tabla que permita describir de manera sintetizada los artículos encontrados y sus criterios de extracción. Cuando se encuentren artículos duplicados, se seleccionará siempre el artículo que tenga mayor nivel de completitud en caso de que existan diferencias entre las distintas publicaciones del mismo artículo.

Esta sección tiene el objetivo de explicar las actividades que se llevando a cabo relacionadas con la ejecución del protocolo de la revisión. Para lo cual se establecen las siguientes etapas: i) selección de los estudios primarios; y, ii) extracción de datos.

4.3.1.1 Selección de estudios primarios

En este periodo, se usó los criterios de inclusión, exclusión y calidad definidos en el plan de revisión para determinar, seleccionar y evaluar la investigación principal. El programa se divide en las siguientes actividades:

- a) **Búsqueda sistemática:** Para realizar esta actividad, primero se adaptó la cadena de búsqueda de acuerdo con la sintaxis del motor de cada una de las bibliotecas digitales seleccionadas. Posteriormente, se ejecutaron las búsquedas, obteniendo como resultado un total de 595 estudios primarios. Adicionalmente, se eliminaron 57 estudios duplicados, quedando finalmente, para su posterior análisis 540 estudios primarios.
- b) **Primera selección:** Ya que en algunas bibliotecas e indexadores tiene un alto número de resultados, se realizó una lectura del título, resúmenes y palabras

claves de los estudios primarios recopilados en el paso anterior con respecto a los criterios de inclusión y exclusión definidos en el protocolo, con el fin de descartar trabajos que no tengan una relación evidente al tema y alcance de esta investigación. Como resultado se obtuvo un repositorio con 60 estudios primarios.

- c) **Evaluación de la calidad:** Como actividad final, se ordenaron los estudios primarios seleccionados de acuerdo a su nivel de relevancia científica, mediante la aplicación de los criterios de calidad incluidos en el protocolo. En esta etapa se aplicó la técnica de Kappa de Fleiss (llamado así por Joseph L. Fleiss), la cual es una medida estadística para evaluar la confiabilidad del acuerdo entre un número fijo de evaluadores al asignar validez o descarte de cada uno de los artículos encontrados, garantizando la veracidad de los datos.

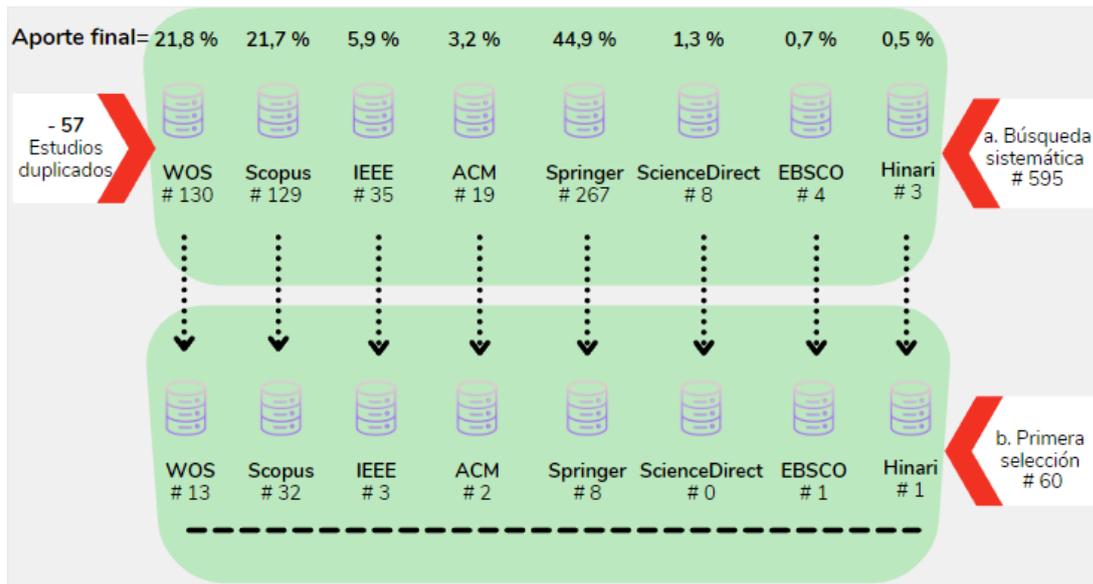
La evaluación de la confiabilidad y calidad del estudio se consiguió mediante la lectura de los artículos realizadas por el grupo de investigación, la aprobación de cada artículo se da cuando el índice kappa supere el 0,61 hasta el valor de uno que corresponde a una concordancia buena y muy buena como se ve reflejada la información en la Tabla 5. El resumen de este proceso de selección y aprobación de los artículos se muestra en la Figura 2

Tabla 5
Valoración del Índice Kappa

| Valor de k | Fuerza de la concordancia |
|--------------|---------------------------|
| < 0.20 | Pobre |
| 0.21 – 0.40 | Débil |
| 0.41 – 0.60 | Moderada |
| 0.61 – 0.80 | Buena |
| 0.81 – 1.00 | Muy buena |

Fuente: (Landis y Koch, 1977)

Figura 2
Proceso de selección de estudios primarios



4.3.1.2 extracción de datos.

En este proceso se extrajeron los datos y posteriormente se sintetizaron los estudios primarios seleccionados según los criterios de extracción previamente identificados y definidos.

En la Tabla 6 se refleja la matriz de resultados del proceso de extracción con cada una de las opciones, número de estudios y porcentaje de estudios encontrados por cada opción.

Tabla 6
Matriz de resultados

| Pregunta | Criterios de extracción | Opciones | Nro. de estudios | % |
|---|--|---|------------------|------|
| RQ1: ¿Qué tipo de información es requerido para analizar juegos serios? | EC1 Lugar de despliegue | [Consola, App, Web] | 52 | 28.9 |
| | EC2 Plataformas de despliegue | [Computadora, Teléfono, Tablet] | 56 | 31.1 |
| | EC3 Área de aplicación | [Salud, Educación, Comercio, otros] | 57 | 23.8 |
| | EC4 Rango de edades | [5-17 años, 18-64 años, más de 65 años] | 41 | 22.8 |
| | EC5 Tipos de herramienta de desarrollo | [Metodología, Programa, Framework] | 56 | 31.1 |
| | EC6 Modos de interacción | [Multijugador, Un jugador, Co-tutoría, otros] | 38 | 15.9 |
| | EC7 puntajes | [Si, No] | 47 | 39.2 |
| | EC8 Clasificación de jugadores | [Puntaje, Edades, Otros] | 25 | 13.9 |
| RQ2: ¿Cómo se abordan los juegos serios en la ciencia de datos? | EC9 Preprocesamiento de datos | [Medida de distancia, muestreo, reducción de dimensionalidad] | 33 | 18.3 |
| | EC10 Clasificación | [KNN (máquinas de vector soporte), árboles de decisión, | 23 | 7.7 |

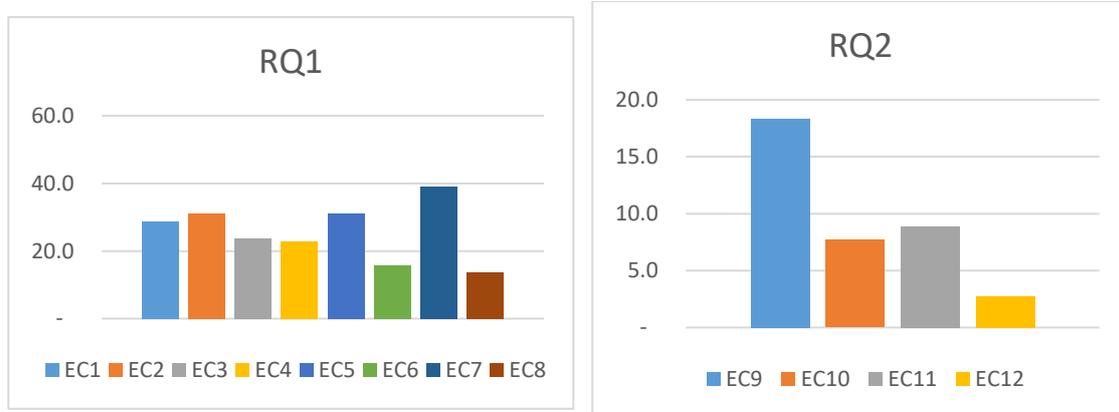
| | | | | |
|---|-------------------------------------|--|----|------|
| | | redes bayesianas, SVN (vecinos más cercanos), ANN (redes neuronales artificiales)] | | |
| | EC11 Clustering | [k-means, Basado en densidad, Paso de mensajes Jerárquico] | 16 | 8.9 |
| | EC12 Como se almacenan los datos | [Json, SQL, No SQL] | 5 | 2.8 |
| RQ3: ¿En qué área de desarrollo se aplican juegos serios? | EC13 Áreas de despliegue | [Industria, Empresa, Academia, Medicina] | 57 | 23.8 |
| RQ4: ¿Cómo está la investigación? | EC14 Tipo de evaluación del estudio | [Análisis, Diseño, Implementación, Pruebas] | 64 | 26.7 |
| | EC15 Tipo de validación | [Experimento controlado, caso de estudio, Prueba de conceptos] | 53 | 29.5 |

4.4 Etapa de reporte de resultados

Una vez evaluado la calidad de los estudios aplicando los criterios de extracción presentados en la Tabla 3, los resultados obtenidos se encuentran en la Figura 3, en la cual consta los porcentajes de estudios de cada pregunta de investigación. Siendo de relevancia conocer el área de aplicación, tipos de usuarios, algoritmo de agrupamiento, clasificación y preprocesamiento de datos que han utilizado en los estudios realizados.

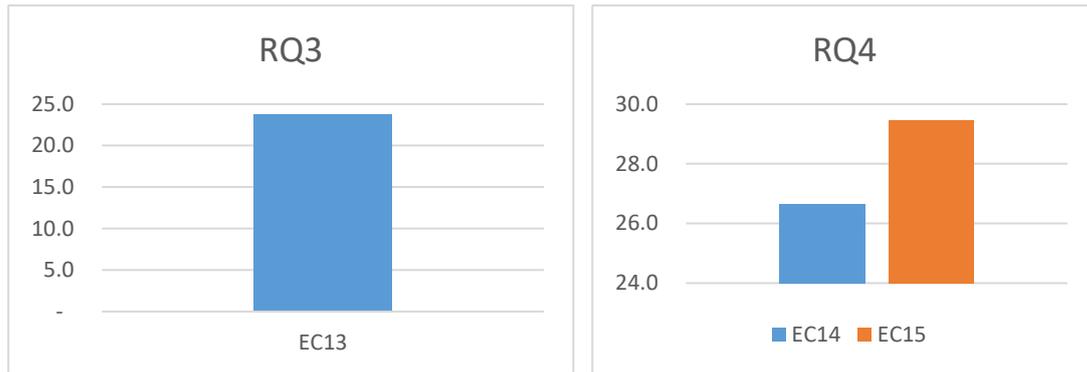
Figura 3

Relevancia de los artículos por pregunta de investigación (RQ)



RQ1: ¿Qué tipo de información es requerido para analizar juegos serios?

RQ2: ¿Cómo se abordan los juegos serios en la ciencia de datos?



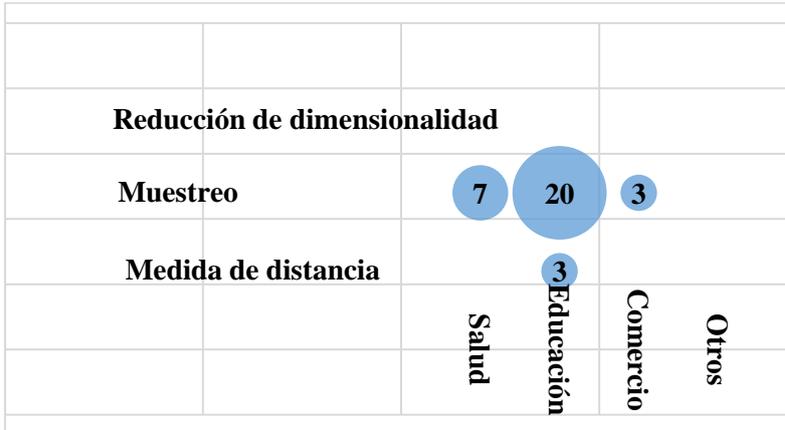
RQ3: ¿En qué área de desarrollo se aplican juegos serios? RQ4: ¿Cómo está la investigación?

4.4.1 Variables de áreas de aplicación usando preprocesamiento de datos

En la Figura 4 se evidencia que en distintas áreas se ha aplicado preprocesamiento de datos. Los resultados son los siguientes: siete estudios han utilizado un muestreo de datos aplicado en la salud y tres estudios se han elaborado en el área del comercio; En el área de la educación tres estudios han aplicado medidas de distancia y 20 empleando un muestreo de datos. Los estudios que se centran en el área de la educación son los siguientes: [A001, A003, A004, A005, A006, A008, A009, A010, A014, A015, A021, A024, A026, A032, A034, A035, A036, A051, A056, A060] (Anexo 1). Además, se puede evidenciar que no existen estudios aplicando el método de reducción de dimensionalidad para el preprocesamiento de datos en ninguna área de aplicación que se analizan en este trabajo.

En el estudio A003, el autor Wang y Huang (2021) realiza un preprocesamiento en datos provenientes de estudiantes con el objetivo de probar su efectividad del aprendizaje en los alumnos.

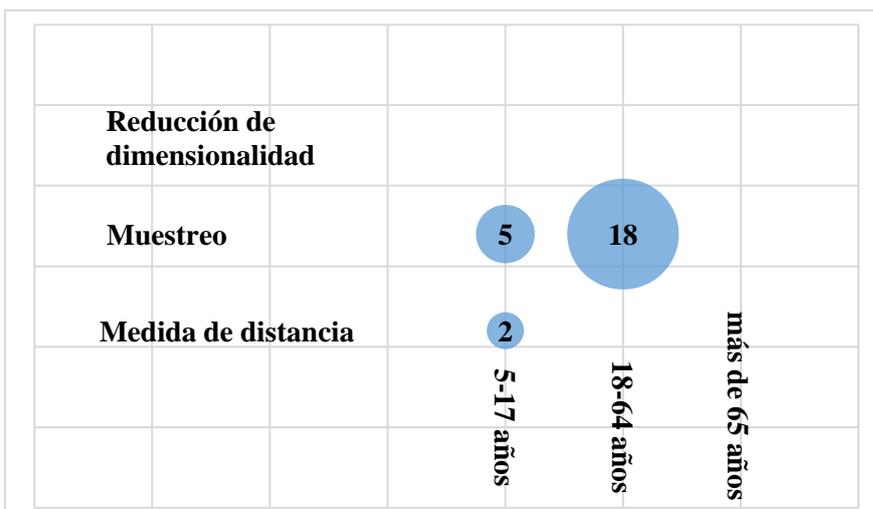
Figura 4
Áreas de aplicación usando preprocesamiento de datos



4.4.2 Tipos de usuarios en los que se ha usado preprocesamiento de datos

En la Figura 5 se muestra los tipos de usuarios identificados y las distintas técnicas/métodos de preprocesamiento de datos. Se puede evidenciar que existe mayor número de estudios que utilizaron un muestreo de datos en la categoría de usuarios de edad entre 18 y 64 años [A001, A003, A006, A008, A014, A021, A026, A027, A029, A033, A034, A035, A036, A050, A051, A053, A054, A056] (Anexo 1). Además, evidenciando que no existe estudios realizados con datos de personas mayores a los 65 años. Asimismo, no existe estudios realizados mediante la técnica/método de preprocesamiento denominada reducción de dimensionalidad.

Figura 5
Tipos de usuarios en los que se ha usado preprocesamiento de datos



4.4.3 Áreas de aplicación en los que se ha usado clasificación de los datos

En la Figura 6 se muestra la información de los algoritmos de clasificación de los datos usados en las distintas áreas de aplicación. La información relevante se encuentra en el área de la educación, en donde se aplica clasificación mediante los árboles de decisión y redes bayesianas, contando con cinco estudios de aplicación cada uno, también se ha aplicado cuatro estudios con redes neuronales artificiales y uno con máquinas de vectores de soporte. También se puede evidenciar que existe déficit de estudios en el área del comercio y escasos estudios que han utilizado una clasificación de los datos en el área de la salud.

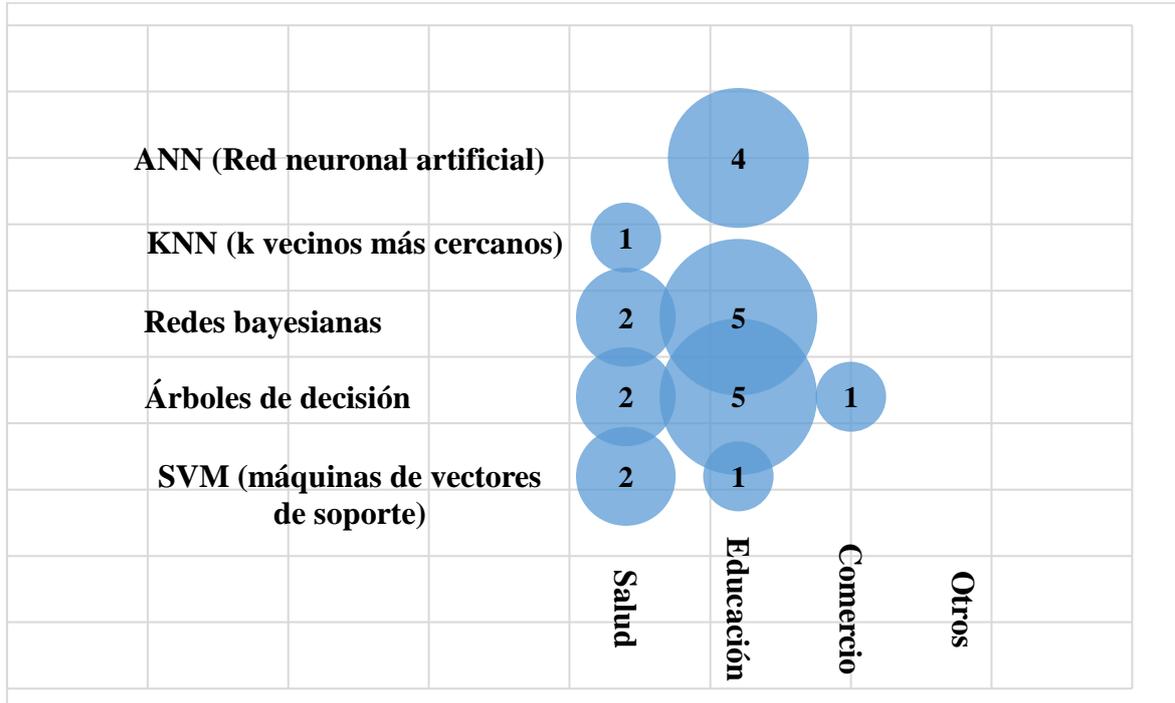
En el área de la educación, el objetivo de estos estudios es detectar la conducta de los estudiantes y descubrir aquellos alumnos que tienen algún tipo de problema o comportamiento inusual como: acciones erróneas, baja motivación, mal uso, trampas, abandono, fallas a nivel académico, etc. El autor del artículo Suhirman y col. (2014) aplica varias técnicas de minería de datos principalmente clasificación y la agrupación ha utilizado para revelar este tipo de estudiantes para brindarles la ayuda adecuada.

Además de estas técnicas de clasificación, los autores utilizan técnicas estadísticas tradicionales como el análisis factorial y el análisis de ajuste del modelo para examinar los datos y la estructura del modelo. Finalmente, los autores desarrollan un modelo computacional del proceso cognitivo utilizando una red neuronal artificial (ANN) que permite examinar los mecanismos subyacentes de la cognición.

Los estudios en el ámbito de la educación que han aplicado el algoritmo de árboles de decisión son: [A006, A018, A023, A037, A056] (Anexo I). Y Los estudios que han aplicado el algoritmo de redes bayesianas son: [A004, A028, A037, A049, A060] (Anexo I).

Figura 6

Áreas de aplicación en los que se ha usado clasificación de los datos



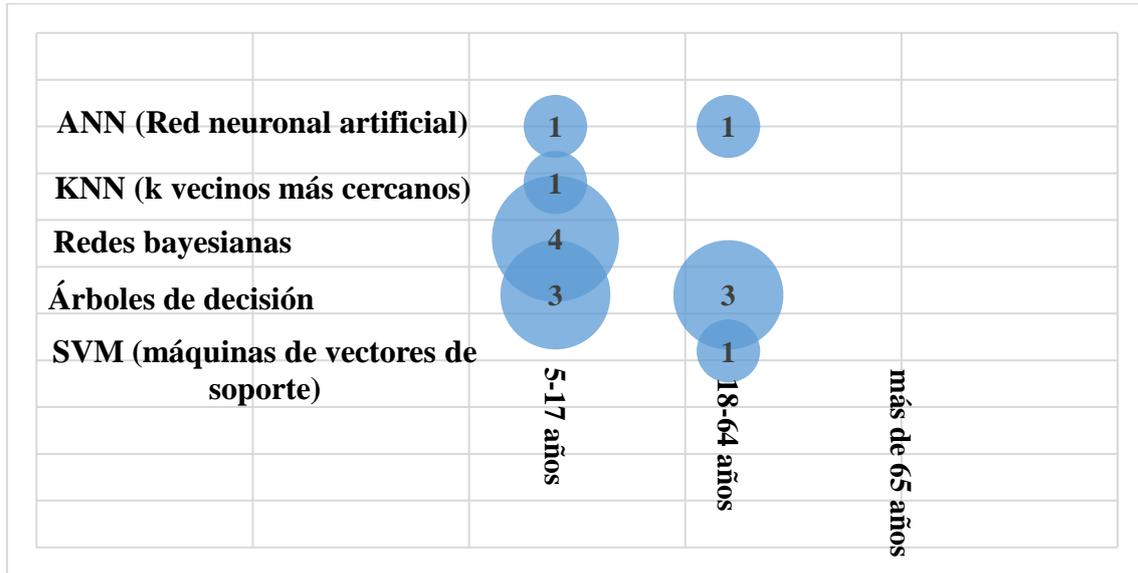
4.4.4 Tipos de usuarios en los que se ha usado algún método/técnica de la ciencia de los datos

Los estudios más relevantes se reflejan en la Figura 7. En el cual que han aplicado redes bayesianas conjuntamente con la clasificación de edades entre cinco y 17 años son: [A028, A037, A041, A060] (Anexo 1). También, se puede evidenciar que no existen estudios que hagan uso de la ciencia de los datos en la categoría de adultos mayores.

En el estudio realizado por Yuhana y col. (2017), se analiza el comportamiento de los participantes en un juego matemático denominado *Math*, un juego serio de evaluación que se desarrolla para evaluar la habilidad matemática de niños con necesidades especiales. Este juego adopta un plan de estudios de matemáticas de Indonesia para estudiantes de escuela primaria. El juego se presenta en forma de cuestionario en múltiples opciones con cuatro opciones a niños, y realiza una clasificación y la predicción de datos se pueden utilizar para determinar el contenido del aprendizaje, preguntas de evaluación, así como una alerta temprana.

Figura 7

Tipos de usuarios en los que se ha usado algún método/técnica de la ciencia de los datos



4.4.5 Áreas de aplicación en donde se ha aplicado algún algoritmo de agrupamiento.

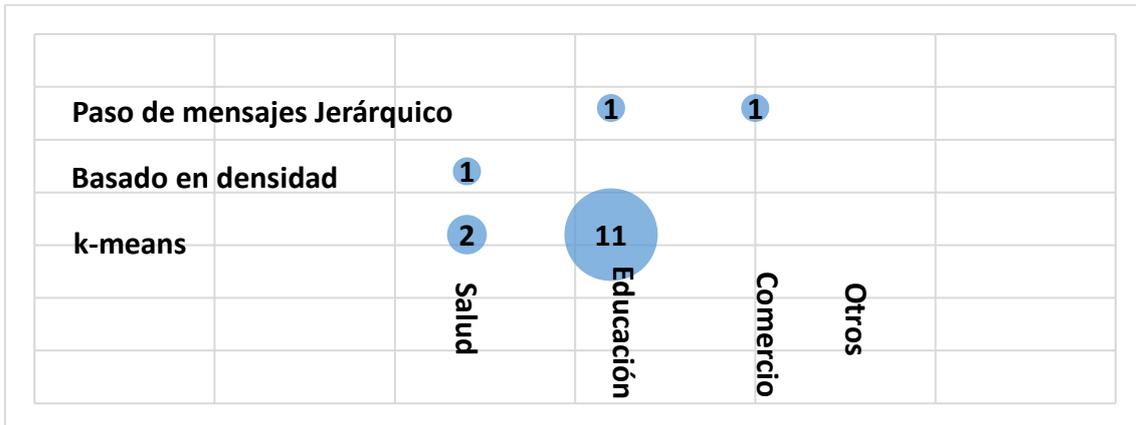
En la Figura 8 se evidencia los algoritmos de agrupamiento de datos conjuntamente sus áreas de aplicación, donde existe mayor número de estudios en el ámbito educativo aplicando el algoritmo de agrupamiento k-means. Existiendo de igual manera, escasez de estudios en el área del comercio y salud.

Este artículo del autor Benmakrelouf y col. (2015) aplica métodos y enfoques de minería de datos k-Means para discutir el análisis de aprendizaje a través de juegos serios, y luego se proporciona un análisis de los datos de experiencia del jugador recopilados del juego educativo. El estudio reveló que existen tres formas de participación de jugadores: principiante, intermedio y avanzado, que interactúan con el juego de acuerdo con su experiencia.

Los estudios que involucran la educación con los algoritmos de agrupamiento son: [A006, A008, A012, A016, A018, A024, A026, A028, A035, A036, A037]

Figura 8

Áreas de aplicación en donde se han aplicado algún algoritmo de agrupamiento.

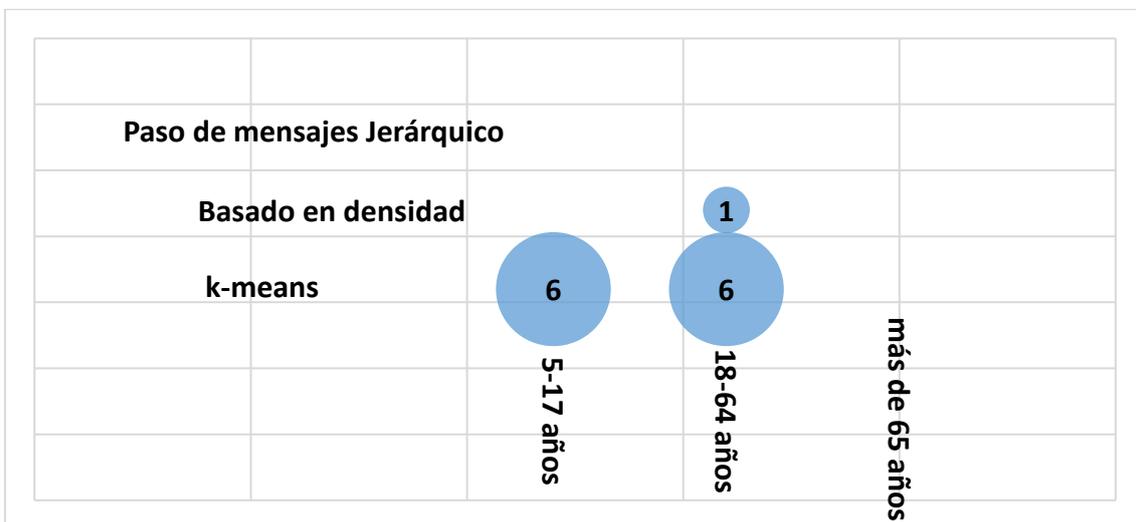


4.4.6 Tipos de usuarios en los que se ha aplicado algún algoritmo de agrupamiento.

En la Figura 9 se evidencia que el algoritmo de agrupamiento más utilizado es el K-mean [A012, A016, A018, A024, A028, A037]. No existen estudios que apliquen el algoritmo de agrupación denominado paso de mensajes jerárquicos. De igual manera, se puede evidenciar que no existe estudios que aplique algoritmos de agrupación en la categoría de adultos mayores (más de 65 años).

Figura 9

Tipos de usuarios en los que se ha aplicado algún algoritmo de agrupamiento.



4.4.7 Clasificación de los jugadores en donde se han aplicado algún algoritmo de agrupamiento.

En la Figura 10 se evidencian los diferentes algoritmos de clasificación de los datos y las distintas formas de agrupamiento. En el estudio realizado por el autor Benmakrelouf et al. (2015), realiza una clasificación de los datos con el objetivo de identificar los

CAPÍTULO 5

5. APLICACIÓN DE LA CIENCIA DE LOS DATOS.

5.1 Metodología de desarrollo

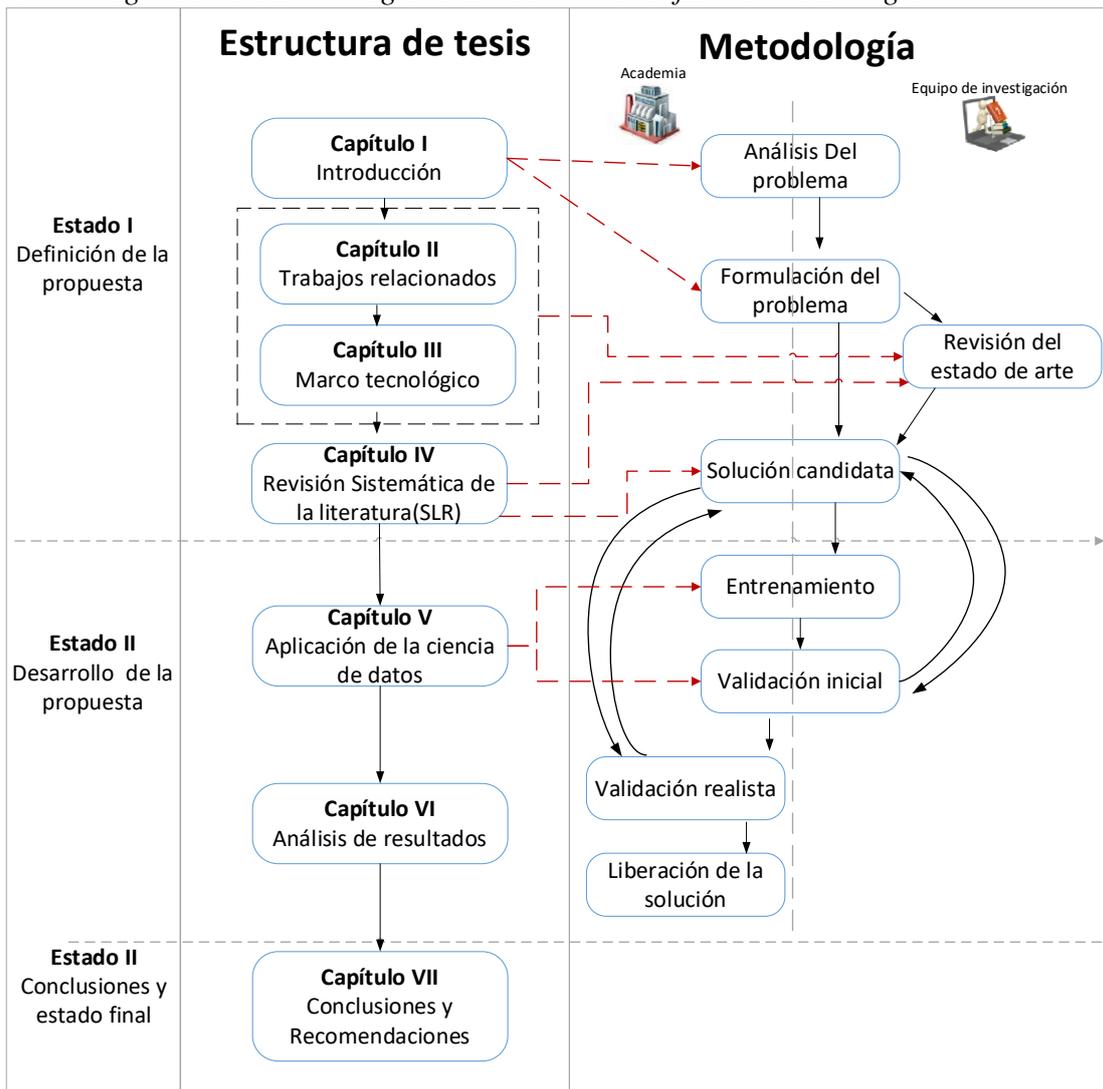
En este apartado es necesario seguir una serie de pasos que permitan gestionar de mejor manera el desarrollo y aplicación de la ciencia de datos; la metodología de este proyecto está basada en el modelo de Gorscheck (2006), la cual consta de las siguientes etapas las mismas que se van desarrollando en los distintos capítulos (Figura 11).

- 1. Análisis del problema:** Se enfoca en entender el problema de los juegos serios orientados a la atención y memoria.
- 2. Formulación del problema:** Una vez que el problema está identificado, éste es formulado de manera más precisa.
- 3. Revisión del estado del arte:** Se realiza una revisión sistemática de la literatura, así como de las soluciones de la ciencia de los datos disponibles para identificar hasta qué punto las soluciones han sido abordadas y cuáles son los problemas abiertos a resolver con la investigación a desarrollar.
- 4. Solución candidata:** Se idean una o más soluciones potenciales, las cuales posteriormente serán depuradas mediante distintas actividades de refinamiento.
- 5. Entrenamiento:** Fase en donde se aplican técnicas y métodos de la ciencia de los datos para el tratamiento, preprocesamiento y experimentación con los datos.
- 6. Validación inicial:** Se realiza una evaluación preliminar de las soluciones llevando a cabo estudios controlados conjuntamente con la colaboración de expertos.
- 7. Validación realista y validación de la solución:** Estas etapas no se llevarán a cabo ya que quedará pendiente para estudios futuros.

Cabe indicar que estos procesos no se llevan a cabo de manera secuencial, sino recursivo y acorde a los resultados que se vayan obteniendo.

Figura 11

Metodología de desarrollo según el modelo de transferencia tecnológica



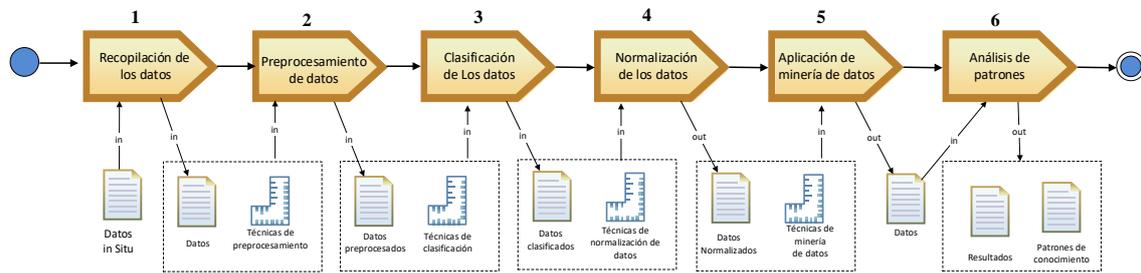
Fuente: Gorscheck (Gorscheck, T., Wohlin, C., Garre, P & Larsson, 2006)

5.2 Metodología para encontrarla solución

La metodología propuesta se muestra en la Figura 12, a través de un diagrama trazado con el proceso de software y sistemas metamodelo de ingeniería 2.0 (SPeM 2.0) según la guía propuesta por los autores Ruiz y Verdugo (2008). Las tareas involucradas en la metodología describen la entrada y salida de artefactos en cada etapa de desarrollo del modelo.

Figura 12

Metodología para encontrar patrones de conocimiento en datos provenientes de juegos serios



La metodología propuesta se dividió en las siguientes actividades: 1) Recopilación de los datos, 2) Preprocesamiento de datos, 3) Clasificación de los datos, 4) Normalización de los datos, 5) Aplicación de técnicas de minería de datos y 6) Análisis de datos.

1) Recopilación de los datos: aquí implica la extracción de datos de los juegos serios de los distintos repositorios mediante el criterio in Situ. En donde se ingresará datos provenientes del juego serios:

El juego de pares consiste en encontrar el par de cada ficha, para esto las fichas se muestran un determinado tiempo y posición, posteriormente se ocultan, con el fin de que la persona recuerde la posición inicial (Pacheco et al. 2021).

Tabla 7

Tabla de atributos y descripción de datos recopilados del juego de pares.

| Nombre del atributo | Descripción |
|---------------------|---|
| Cedula | Identificación del participante. |
| Nombre | Nombre del participante. |
| Ciudad | Ciudad de origen del participante. |
| Tiene discapacidad | True en caso de tener discapacidad y false cuando no hay discapacidad |
| Discapacidad | Se debe especificar la discapacidad que tiene la persona |
| Edad | Edad del participante |
| Teléfono | Teléfono del participante |
| Fecha de nacimiento | Fecha de nacimiento del participante |
| Tiempo Final | Tiempo en el que se tardó el jugador en resolver todo el juego |
| Tiempo nivel 1 | Tiempo en el que tardó en culminar el nivel 1 |
| Tiempo nivel 2 | Tiempo en el que tardó en culminar el nivel 2 |
| Tiempo nivel 3 | Tiempo en el que tardó en culminar el nivel 3 |
| Tiempo nivel 4 | Tiempo en el que tardó en culminar el nivel 4 |
| Puntaje obtenido | Considerando que el puntaje perfecto es 2000 y se reduce según el número de equivocaciones y el tiempo. |

Los datos fueron obtenidos de una sola tabla de una base de datos en línea, donde consta el registro de los datos de todos los jugadores. El juego de pares está disponible en la siguiente dirección: <https://proyectoia2020.web.app/>

2) Preprocesamiento de los datos: describe el conjunto de datos sin procesar y su preparación previa. Aplica técnicas de preprocesamiento como la eliminación de datos faltantes, inconsistencias, valores extremos, etc. Con el objetivo de obtener un conjunto de datos limpio.

En los datos se encontraron valores inconsistentes como personas con nombres inexistentes, valores faltantes como la edad o la fecha de nacimiento, tiempos de solo un nivel del juego. Por último, los valores extremos fueron el tiempo registrando sobrepasando la media hora para terminar los distintos niveles, como consecuencia afectaba directamente al puntaje del participante. Estos registros fueron eliminados con el objetivo de tener datos limpios y obtener resultados eficaces.

3) Clasificación de los datos

Según el estudio realizado por Hou, H. T. (2015) existe una clasificación de los datos, con el objetivo de categorizar los resultados, ya que se prevé que la atención y memoria de las personas van cambiando según la edad. Así como también el autor (Slimani et al., 2018) aplica una clasificación de valores según el puntaje obtenido de los participantes, con ello, se logra una agrupación mediante el grado de desempeño.

Tabla 8
Clasificación de datos mediante la edad

| Clasificación | Edad |
|-----------------|-------------------|
| Niños | hasta los 11 años |
| Adolescentes | 12 – 20 años |
| Adultos | 21 – 59 |
| Adultos mayores | 60 en adelante |

Fuente: Hou, H. T. (2015)

4) Normalización de los datos

Para este proceso se aplicó mediante el programa RapidMiner (RapidMiner Inc., 2020), con el objetivo de transformar los valores de puntajes y edades en una escala común para estos dos atributos. Aquí se utilizó un mínimo de cero y un máximo de uno. Los valores resultantes tomarán este rango de valores.

5) Aplicación de técnicas de minería de datos

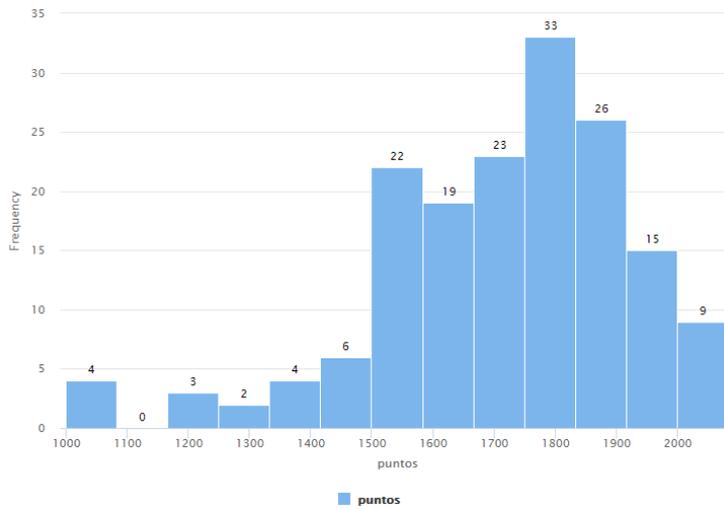
Aquí se aplicarán las técnicas de minería de datos de clustering con K-means y redes neuronales. Con el objetivo de analizar y verificar cuál de estas técnicas se puede encontrar patrones de comportamientos más eficaces.

En esta sección se utiliza el algoritmo de clustering K-means que ofrece RapidMiner (RapidMiner Inc., 2020), el cual encuentra el k número de clústers para satisfacer un criterio determinado (Li & Wu, 2012). El algoritmo k-means es el método de agrupación que permite trabajar con una gran cantidad de datos numéricos de alta dimensionalidad, capaz de proporcionar una clasificación eficaz de datos similares (Hou et al., 2015).

5) Análisis de datos

En esta etapa se analizan los datos y los resultados obtenidos luego de aplicar las distintas técnicas de minería de datos propuestas.

Los datos se analizan estadísticamente para entender de mejor manera la distribución. En la Figura 13 se puede visualizar la cantidad de valores de puntos obtenidos por los participantes, donde cuyo valor más frecuente es alrededor 1,800 puntos, el mínimo puntaje obtenido es de 1,000, el valor máximo con 2,000 puntos. La media de todos los participantes es de 1,705. Un valor considerablemente aceptable teniendo en cuenta que la puntuación perfecta del juego serio es 2,000. La desviación estándar con un valor de 208 valor de relevancia para analizar y cuantificar la dispersión de los datos.

Figura 13*Frecuencia de puntajes obtenidos por los participantes***Tabla 9***Estadística de puntajes de participantes*

| Nombre | Valor |
|---------------------|----------|
| Mínimo | 1,000 |
| Máximo | 2,000 |
| Media | 1,705.57 |
| Desviación estándar | 208.62 |

A continuación, se presenta en la Tabla 10 la distribución de los datos en las diferentes categorías ya identificadas anteriormente. Donde se puede denotar que la mayoría de datos son provenientes de las personas de la categoría adultas (21-59 años) con un total de 112 participaciones equivalentes al 67%. Posteriormente se tiene los valores de adolescentes (12-20 años) con un total de 32 participaciones equivalentes a 19%. En la categoría infancia (menores a 11 años) con 18 participaciones equivalentes a 11%. Por último, con 4 participaciones los adultos mayores (mayor a 65 años) equivalentes al 2%

Tabla 10*Distribución de los datos de participación en las diferentes categorías.*

| Categoría | Cantidad | Porcentaje |
|--------------------------------|----------|------------|
| Infancia (menor a 11 años) | 18 | 10.78 % |
| Adolescencia (12-20 años) | 32 | 19.16 % |
| Adulto (21-59 años) | 112 | 67.66 % |
| Adulto Mayor (mayor a 60 años) | 4 | 2.40% |

En la Figura 14 se presenta el diagrama de caja y bigotes de las edades categorizadas y los puntos obtenidos de los participantes. Esta representación gráfica simboliza los datos a través de los cuartiles, de igual manera los valores atípicos encontrados.

Figura 14

Diagrama de caja y bigotes de categoría y puntajes de los participantes

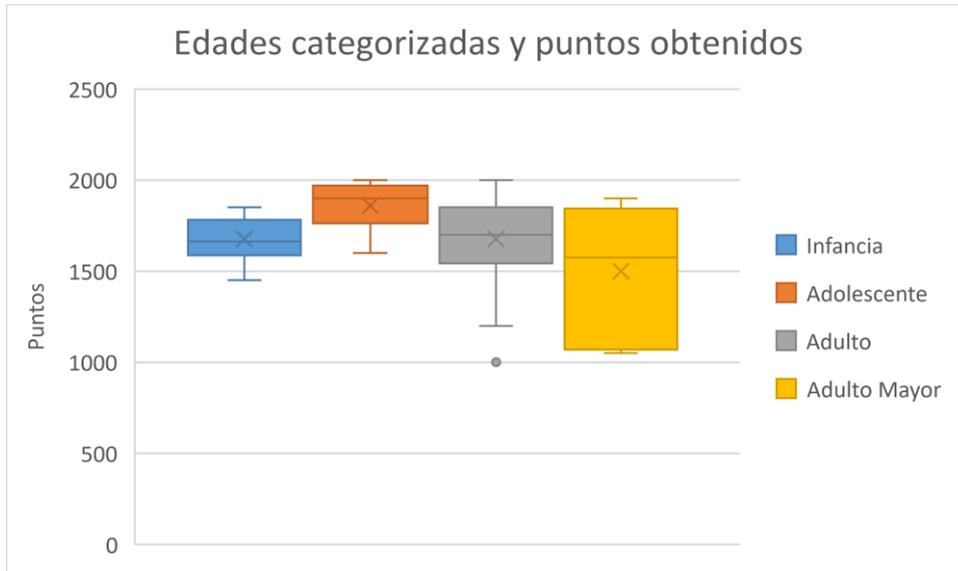


Tabla 11

Resultado de diagrama de caja y bigotes referentes a edades categorizadas y puntajes obtenidos en el juego

| Infancia | | Adolescencia | |
|------------------|----------|------------------|----------|
| Nombre | Valor | Nombre | Valor |
| Máximo | 1,850 | Máximo | 2,000 |
| Cuartil Superior | 1,781.25 | Cuartil Superior | 1,968.75 |
| Media | 1,662.5 | Media | 1,900 |
| Cuartil inferior | 1,587.5 | Cuartil inferior | 1,762.5 |
| Mínimo | 1,450 | Mínimo | 1,600 |

| Adulto | | Adulto Mayor | |
|------------------|----------|------------------|----------|
| Nombre | Valor | Nombre | Valor |
| Máximo | 2,000 | Máximo | 1,900 |
| Cuartil Superior | 1,850 | Cuartil Superior | 1,843.75 |
| Media | 1,700 | Media | 1,575 |
| Cuartil inferior | 1,543.75 | Cuartil inferior | 1,068.75 |
| Mínimo | 1,200 | Mínimo | 1,050 |
| Valor atípico | 1,000 | | |

En la Tabla 11 se presentan los resultados de cada categoría de edades y su diferentes cuartiles superiores, inferiores, máximo, media y mínimo. Logrando interpretar que categoría de edades puede sacar mayor puntaje en el juego serio.

Se puede apreciar que en el diagrama de caja de la categoría infancia, cuya representación es compacta y no tan dispersa, de igual manera para la clasificación de adolescentes y adultos. En cambio, en la categoría de adultos mayores existe mayor diversidad de puntajes, teniendo como resultado el cuartil inferior y el valor mínimo menor con respecto a las demás categorías. Este resultado se puede dar ya que existen pocas muestras en esta categoría, si un puntaje es bajo, afectaría directamente a los resultados. Por otro lado, la categoría de adolescentes tiene mayor puntaje y menor dispersión de los datos, Según estos resultados se puede evaluar que los adolescentes tienen mayor atención y memoria, logrando un desempeño mejor con respecto a las demás categorías.

En la Figura 15 se reflejan los intentos realizados por los participantes de la categoría infancia (0-11 años). El cual cinco personas realizaron tres intentos, dos personas cuatro intentos y una persona hizo cinco intentos de cada una de las cuatro fases del juego serio. Se puede apreciar que en cada fase y según vaya intentando más veces, las personas van mejorando el tiempo en terminar cada fase del juego. Por ende, poniendo mayor atención y memoria. En la primera fase del juego se puede visualizar que existe un leve crecimiento en el tiempo, sin embargo, este crecimiento no es mayormente relevante.

Figura 15

Intentos realizados de los participantes de la categoría Infancia.

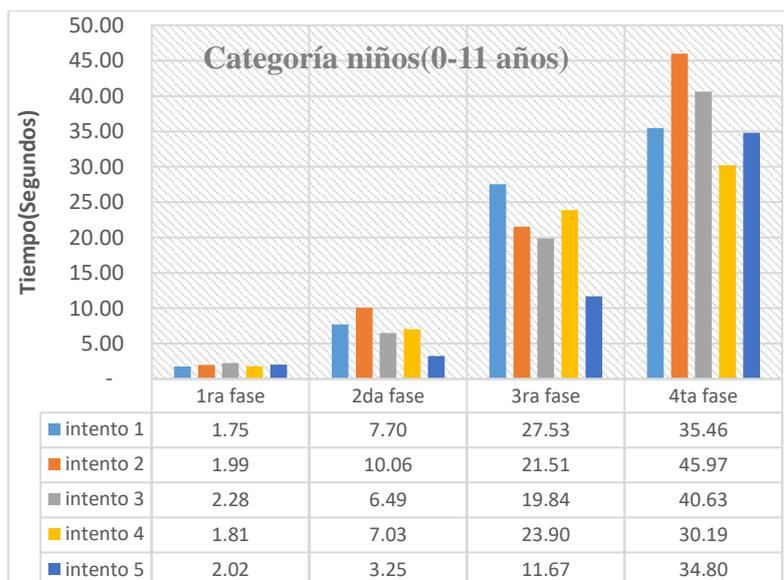


Tabla 12

Número de intentos y cantidad de personas de la categoría Infancia

| Nro. de intentos | Nro. de personas |
|-------------------------|-------------------------|
| 1 | 5 |
| 2 | 5 |
| 3 | 5 |
| 4 | 2 |
| 5 | 1 |

En la categoría Adolescentes (12-20), En la Figura 16 se refleja los intentos realizados por los participantes. El cual 30 personas realizaron un intento, y dos personas realizaron dos intentos de cada una de las cuatro fases del juego serio. Se puede apreciar que en cada fase y según vaya intentando más veces, las personas van mejorando el tiempo en terminar cada fase del juego.

Figura 16

Intentos realizados de los participantes de la categoría adolescentes

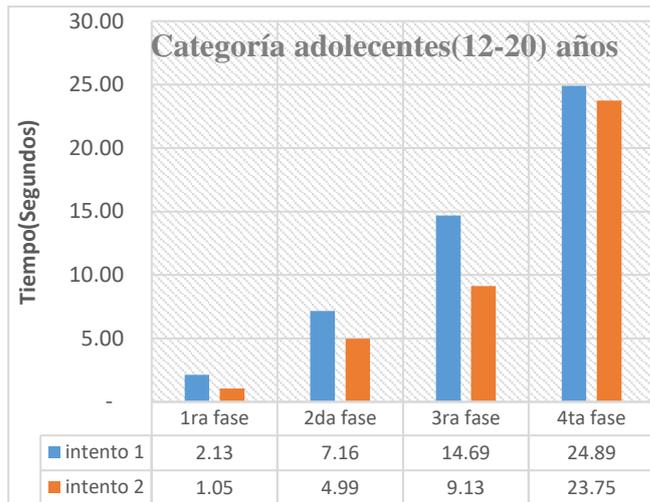


Tabla 13

Número de intentos y cantidad de personas de la categoría Adolescentes

| Nro. de intentos | Nro. de personas |
|-------------------------|-------------------------|
| 1 | 30 |
| 2 | 2 |

En la categoría Adultos (21-59 años), En la Figura 17 se refleja los intentos realizados por los participantes. El cual 73 personas realizaros un intento, 15 personas realizaron dos intentos, siete personas realizaron tres intentos, dos personas realizaron hasta siete intentos y una persona realizo hasta 11 intentos de cada una de las cuatro fases del juego serio. Se puede apreciar que en cada fase de igual manera que en las anteriores, según vaya intentando más veces, las personas van mejorando el tiempo en terminar cada fase del juego.

Figura 17

Intentos realizados de los participantes de la categoría Adultos

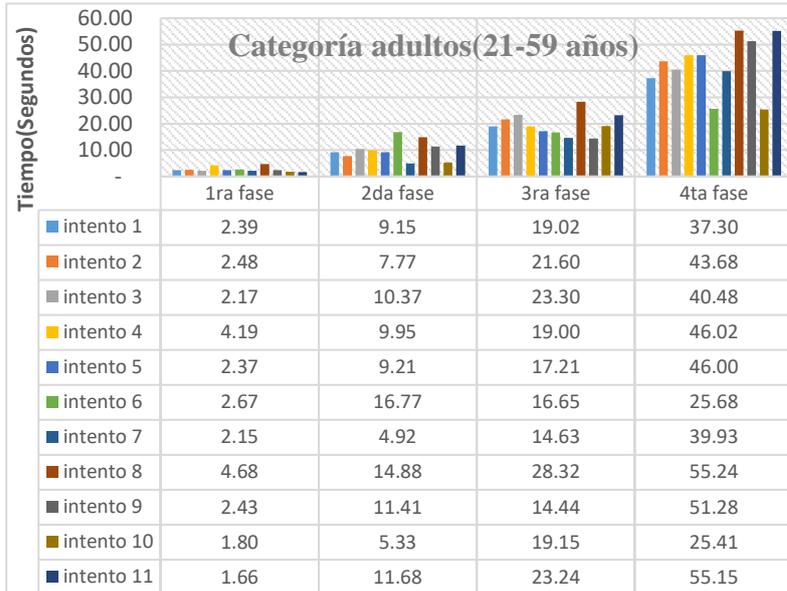


Tabla 14

Número de intentos y cantidad de personas de la categoría Adultos.

| Nro. de intentos | Nro. de personas |
|------------------|------------------|
| 1 | 73 |
| 2 | 15 |
| 3 | 7 |
| 4 | 5 |
| 5 | 2 |
| 6 | 2 |
| 7 | 2 |
| 8 | 1 |
| 9 | 1 |
| 10 | 1 |
| 11 | 1 |

En la categoría adultos mayores (mayor a 59 años), En la Figura 18 se refleja los intentos realizados por los participantes. El cual tres personas realizaron un intento y dos personas realizaron dos intentos de cada una de las cuatro fases del juego serio. Se puede apreciar que en cada fase y según vaya intentando más veces, las personas van mejorando el tiempo en terminar cada fase del juego.

Figura 18

Intentos realizados de los participantes de la categoría Adultos Mayores.

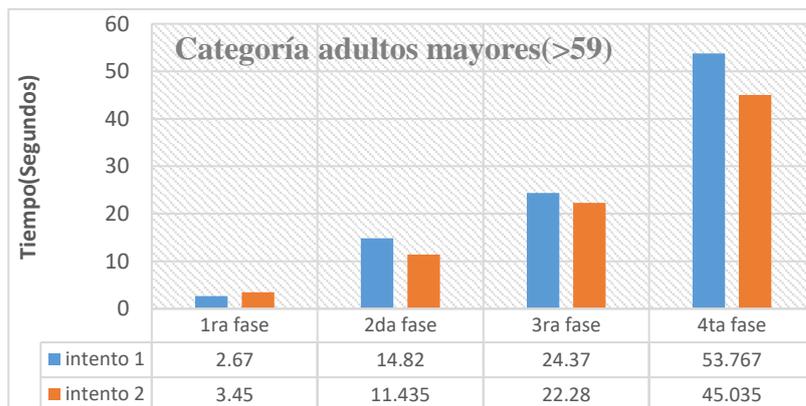


Tabla 15

Número de intentos y cantidad de personas de la categoría Adultos mayores

| Nro. de intentos | Nro. de personas |
|------------------|------------------|
| 1 | 3 |
| 2 | 2 |

Aplicación de clústers

Para la aplicación de clasificación de los datos se ha elegido la distancia euclídea como medida de asociación, los grupos formados contendrán individuos parecidos de

forma que la distancia entre ellos tiene que ser corta. En la Figura 19 se puede apreciar la que al momento de clasificar el conjunto de datos de edades y puntajes de los participantes se dividen en grupos definidos por valores similares, pudiendo determinar según su edad y su puntaje se van definiendo los grupos, además, se ha identificado el clúster cinco como disjunto, es decir, existe dos valores atípicos. Gran parte de los valores se encuentran concentrados en el rango de menores a 41 años y superiores a los 1,350 puntos; pudiendo determinar que mayormente las personas menores a 40 años tienen mayor puntaje, después de esa edad existen otros factores que pueden afectar el desempeño y rendimiento para culminar las etapas del juego.

Figura 19
Clúster de edad y puntajes obtenidos por los participantes

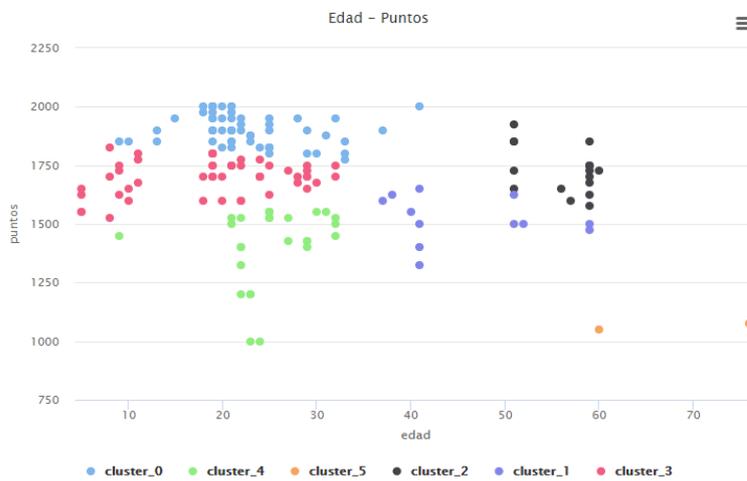


Tabla 16
Cantidad de datos clasificados por edad y Puntajes

| Clúster | Cant. | Porcentaje |
|---------|-------|------------|
| 0 | 57 | 34% |
| 1 | 12 | 7% |
| 2 | 17 | 10% |
| 3 | 51 | 31% |
| 4 | 27 | 16% |
| 5 | 2 | 1% |

En la Figura 20 se puede apreciar en el momento de clasificar el conjunto de datos de edades y tiempo de la primera etapa del juego, los participantes se dividen en grupos definidos por valores similares, pudiéndose definir según su edad y su tiempo de en cuatro grupos, además, se ha identificado en el clúster dos como disjunto, es decir, existe un valor atípico. Gran parte de los valores se encuentran concentrados en el clúster cero con 126 datos que equivalen al 76% de los datos, es decir, las personas que son menores a los 40 años terminan la primera etapa antes de los cuatro segundos.

Figura 20

Clasificación de datos por edad y tiempo de la primera fase

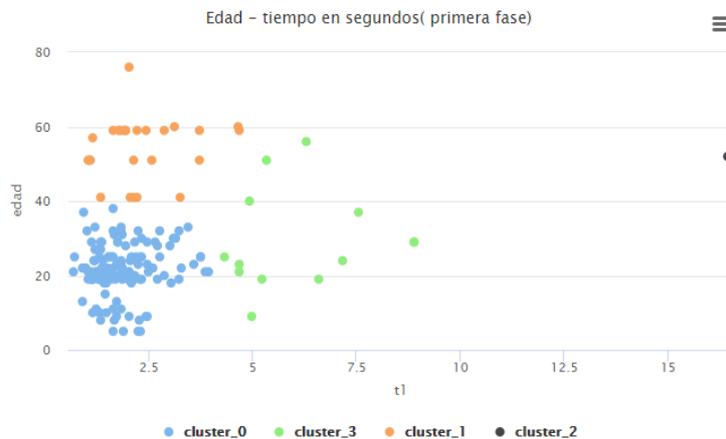


Tabla 17

Cantidad de datos clasificados por edad y tiempo (Primera fase)

| Clúster | Cant. | Porcentaje |
|---------|-------|------------|
| 0 | 126 | 76% |
| 1 | 26 | 16% |
| 2 | 1 | 1% |
| 3 | 13 | 8% |

En la Figura 21 se puede apreciar que, al momento de clasificar el conjunto de datos de edades y tiempo de la segunda etapa del juego, los participantes se dividen en grupos definidos por valores similares, pudiéndose definir según su edad en los cuatro grupos, además, se ha identificado en el clúster dos como disjunto, es decir, existe un valor atípico. Gran parte de los valores se encuentran concentrados en el clúster cero con 120 datos que equivalen al 72%, es decir, las personas que son menores a los 37 años terminan la segunda etapa antes de los 14 segundos. Posteriormente existen valores dispersos en el clúster uno pertenecientes a personas de entre 20 y 60 años que superaron los 14 segundos

Figura 21

Clasificación de datos por edad y tiempo de la segunda fase

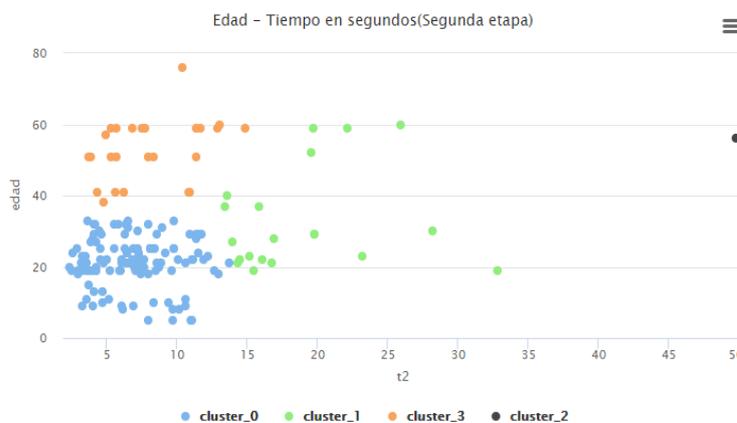


Tabla 18

Cantidad de datos clasificados por edad y tiempo (segunda fase)

| Clúster | Cant. | Porcentaje |
|---------|-------|------------|
| 0 | 120 | 72% |
| 1 | 20 | 12% |
| 2 | 1 | 1% |
| 3 | 25 | 15% |

En la Figura 22 se representa la clasificación de los datos por edades y tiempos de la tercera fase del juego, donde se puede divisar la dispersión de los datos en cuatro grupos

de clúster, con mayor concentración de los datos en el clúster cero que representa el 60%, denotando que las personas hasta los 40 años terminan la tercera fase en un tiempo menor a los 22 segundos. Posteriormente, se encuentran muy seguidos en el clúster uno, que está representado por las personas mayores a los 40 años y que tienen un tiempo estimado entre 12 y 38 segundos. Por otra parte, se encuentra el clúster dos que está involucrando personas con edades entre cinco y 38 años, con un tiempo entre 20 y 35 segundos. Consecutivo a los anteriores está el clúster tres con el 5% que representa a las personas entre 5 y 40 años con un tiempo para culminar el reto entre 35 y 60 segundos.

Figura 22
Clasificación de datos por edad y tiempo de la tercera fase

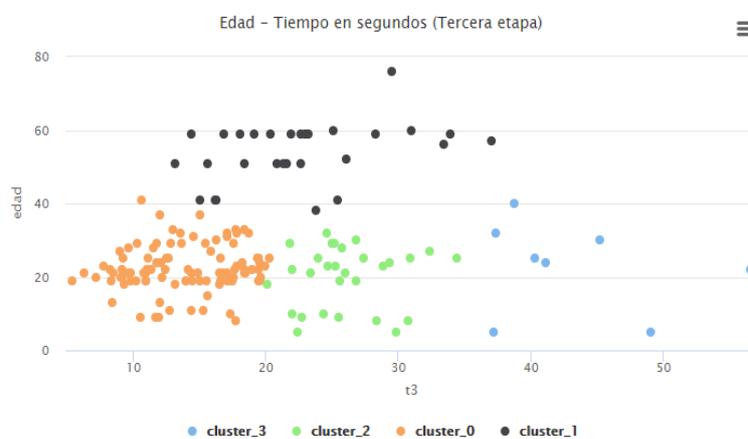


Tabla 19
Cantidad de datos clasificados por edad y tiempo (tercera fase)

| Clúster | Cant. | Porcentaje |
|---------|-------|------------|
| 0 | 99 | 60 % |
| 1 | 29 | 17% |
| 2 | 30 | 18% |
| 3 | 8 | 5% |

En la Figura 23 se representa la clasificación de los datos por edades y tiempos de la cuarta fase del juego, donde se puede divisar la dispersión de los datos en cuatro grupos de clúster, con mayor concentración de los datos en el clúster cero que representa el 63%, expresando que las personas hasta los 41 años terminan la cuarta fase en un tiempo menor a los 45 segundos. Posteriormente, se encuentra el clúster uno que está representado con un 16% por las personas entre 39 y 60 años que tienen un tiempo estimado entre 25 y 45 segundos. Por otra parte, se encuentra el clúster dos con un 2% de los datos que está involucrando a personas con edades entre 55 y 68 años, con un tiempo entre 60 y 85 segundos. Consecutivo a los anteriores está el clúster tres con el 19% que encasilla las personas entre cinco y 39 años con un tiempo para culminar el reto entre 45 y 95 segundos.

Figura 23

Clasificación de datos por edad y tiempo de la cuarta fase

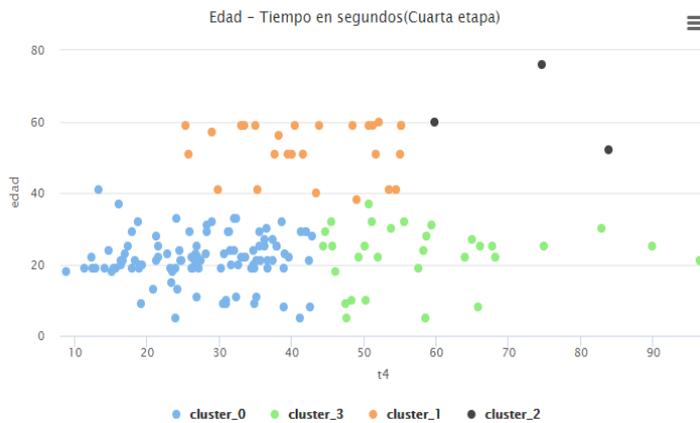


Tabla 20

Cantidad: Edad y tiempo (cuarta fase)

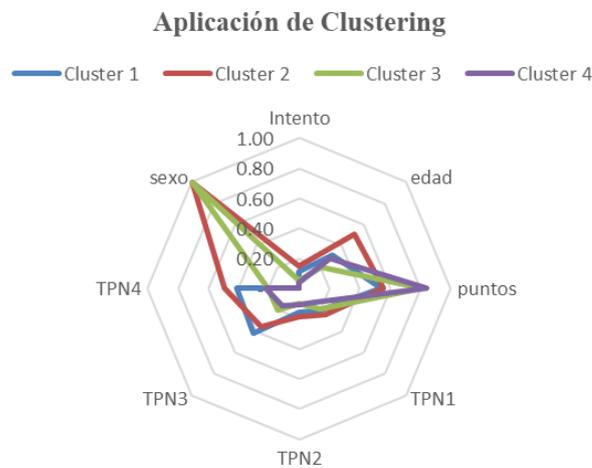
| Clúster | Cant. | Porcentaje |
|---------|-------|------------|
| 0 | 105 | 63% |
| 1 | 27 | 16% |
| 2 | 3 | 2% |
| 3 | 31 | 19% |

Para obtener datos en una misma escala se aplicó la técnica de normalización de los datos con el objetivo de estandarizar los valores y ponerlos en una misma escala las variables de tiempos de cada uno de los niveles de juego (TPN1, TPN2, TPN3, TPN4), edades, puntos y sexo. Con la finalidad de dar un análisis de datos más eficiente.

Una vez ingresado los datos normalizados antes mencionados, se procedió a aplicar la técnica de clustering para clasificar en cuatro grupos. En la Tabla 21 se encuentran los centroides de los cuatro clústers y las distintas variables del juego.

Figura 24

Gráfico radial de la aplicación de clustering con datos de centroides de las variables demográficas.

**Tabla 21**

Centroides de datos normalizados y clasificados.

| | Clúster 1 | Clúster 2 | Clúster 3 | Clúster 4 |
|---------|-----------|-----------|-----------|-----------|
| Intento | 0.11 | 0.15 | 0.05 | 0.03 |
| Edad | 0.31 | 0.51 | 0.20 | 0.28 |
| puntos | 0.53 | 0.55 | 0.80 | 0.83 |
| TPN1 | 0.20 | 0.24 | 0.19 | 0.13 |
| TPN2 | 0.15 | 0.18 | 0.10 | 0.10 |
| TPN3 | 0.4 | 0.35 | 0.19 | 0.16 |
| TPN4 | 0.46 | 0.49 | 0.21 | 0.22 |
| Sexo | 0 | 1 | 1 | 0 |

En la Figura 24 se puede apreciar que las variables de los tiempos de cada nivel del juego, edad y sexo son mayormente relevantes en cada clasificación de clúster. Es por esa razón que se realizará un análisis más detenido de los intentos realizados por los participantes en cada uno de los niveles de juego, con el objetivo de verificar las personas que se encuentran en cada clasificación.

En la Tabla 23, Tabla 25 y Tabla 27 muestran los centroides de los grupos que implican las variables de los tiempos de los diferentes niveles del juego (TPN1, TPN2, TPN3, TPN4) de cada uno de los intentos del juego.

Tabla 22

Abreviaturas utilizadas y sus significados

| Abreviatura | Significado. |
|-------------|--|
| TPN1 | Tiempo promedio para finalizar el nivel uno del juego serio. |
| TPN2 | Tiempo promedio para finalizar el nivel dos del juego serio. |
| TPN3 | Tiempo promedio para finalizar el nivel tres del juego serio. |
| TPN4 | Tiempo promedio para finalizar el nivel cuatro del juego serio. |
| TPH | Tiempo promedio de hombres para finalizar los cuatro niveles del juego serio. |
| TPM | Tiempo promedio de mujeres para finalizar los cuatro niveles del juego serio. |
| PPH | Puntos promedio obtenidos de hombres que finalizan los cuatro niveles del juego serio. |
| PPM | Puntos promedio obtenidos de mujeres que finalizan los cuatro niveles del juego serio. |

En la Figura 25, Figura 26 y Figura 27 se encuentran los gráficos que representa la clasificación de los datos de cada uno de los intentos del juego con cada uno de los niveles de juego. En donde el clúster uno tiene mayor influencia sobre los demás, a continuación,

se analiza de mejor manera los clúster para identificar que categoría de edades de personas están en esta clasificación. La Tabla 24, Tabla 26 y Tabla 28 se reflejan las distintas categorías de edades y la cantidad y porcentaje de hombres y mujeres que se encontraron, el tiempo promedio de hombres (TPH) y tiempo promedio de mujeres (TPM) que transcurre para finalizar los diferentes niveles en los diferentes intentos realizados del juego. Por último, el promedio de puntajes de hombres (PPH) y promedio de puntajes de mujeres (PPM).

Figura 25

Gráfico de centroides agrupados del primer intento del juego.

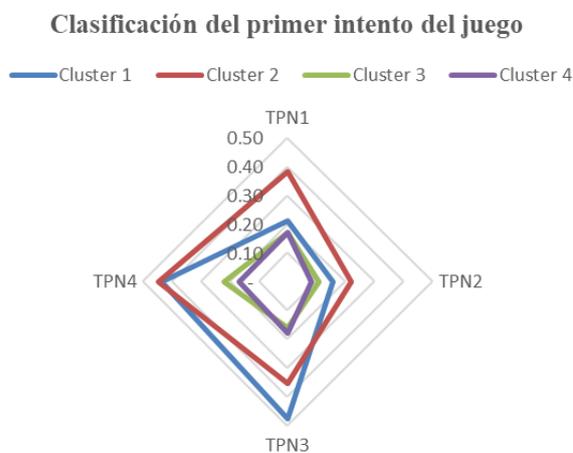


Tabla 23

Centroides de datos normalizados y clasificados del primer intento del juego.

| | Clúster 1 | Clúster 2 | Clúster 3 | Clúster 4 |
|------|-----------|-----------|-----------|-----------|
| TPN1 | 0.21 | 0.38 | 0.16 | 0.17 |
| TPN2 | 0.15 | 0.21 | 0.10 | 0.08 |
| TPN3 | 0.47 | 0.35 | 0.16 | 0.17 |
| TPN4 | 0.43 | 0.44 | 0.21 | 0.16 |

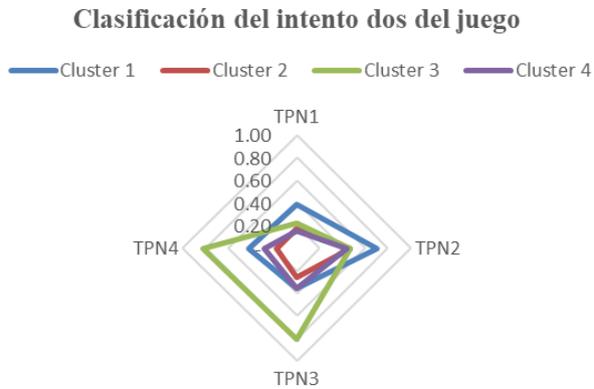
Tabla 24

Resultados de datos por categorías de la Figura 25

| Categoría | Cantidad de datos | Cantidad Hombres | Cantidad Mujeres | TPH (Seg) | TPM (seg.) | PPH | PPM |
|--------------------------------|-------------------|------------------|------------------|-----------|------------|---------|-------|
| Infancia (menor a 11 años) | 5 | 2 | 3 | 16.4 | 19.2 | 1,612.5 | 1,725 |
| Adolescencia (12-20 años) | 30 | 22 | 8 | 12.0 | 12.9 | 1,843.2 | 1,890 |
| Adulto (21-59 años) | 73 | 35 | 38 | 17.8 | 15.6 | 1,649.3 | 1,718 |
| Adulto Mayor (mayor a 60 años) | 2 | 2 | 0 | 29.8 | | 1,062.5 | |
| Total promedio | 110 | 61 | 49 | 18.99 | 15.90 | 1,541.9 | 1,778 |

Figura 26

Gráfico de centroides agrupados del intento dos del juego.

**Tabla 25**

Centroides de datos normalizados y clasificados del intento dos del juego.

| | Clúster 1 | Clúster 2 | Clúster 3 | Clúster 4 |
|------|-----------|-----------|-----------|-----------|
| TPN1 | 0.38 | 0.17 | 0.22 | 0.14 |
| TPN2 | 0.70 | 0.46 | 0.46 | 0.44 |
| TPN3 | 0.35 | 0.25 | 0.80 | 0.35 |
| TPN4 | 0.41 | 0.17 | 0.82 | 0.28 |

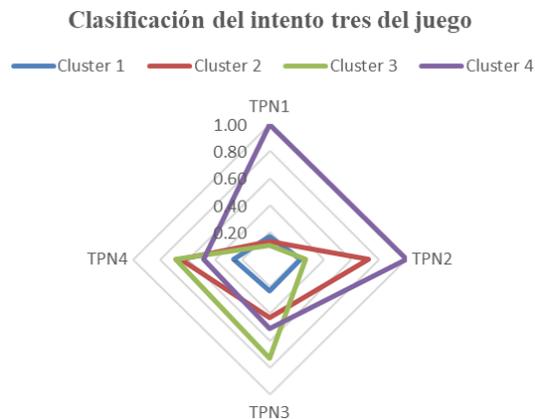
Tabla 26

Resultados de datos por categorías del intento dos (Figura 26)

| Categoría | Cantidad de datos | Cantidad Hombres | Cantidad Mujeres | TPH (Seg) | TPM (seg.) | PPH | PPM |
|--------------------------------|-------------------|------------------|------------------|-----------|------------|----------|----------|
| Infancia (menor a 11 años) | 5 | 3 | 2 | 19.77 | 20.06 | 1,608.30 | 1,650 |
| Adolescencia (12-20 años) | 2 | 1 | 1 | 9.59 | 9.86 | 1,850 | 1,950 |
| Adulto (21-59 años) | 16 | 10 | 6 | 19.94 | 16.92 | 1,725 | 1,579.20 |
| Adulto Mayor (mayor a 60 años) | 1 | 1 | 0 | 23.34 | | 1,725 | |
| Total | 24 | 15 | 9 | | | | |
| Promedio de tiempo(segundos) | | | | 18.2 | 15.6 | 1,727.08 | 1,726.40 |

Figura 27

Gráfico de centroides agrupados del intento tres del juego.

**Tabla 27**

Centroides de datos normalizados y clasificados del intento tres del juego.

| | Clúster 1 | Clúster 2 | Clúster 3 | Clúster 4 |
|------|-----------|-----------|-----------|-----------|
| TPN1 | 0.17 | 0.13 | 0.11 | 1.00 |
| TPN2 | 0.22 | 0.72 | 0.27 | 1.00 |
| TPN3 | 0.23 | 0.43 | 0.73 | 0.51 |
| TPN4 | 0.27 | 0.65 | 0.69 | 0.48 |

Tabla 28*Resultados de datos por categorías del intento tres (Figura 27)*

| Categoría | Cantidad de datos | Cantidad Hombres | Cantidad Mujeres | TPH (Seg) | TPM (seg.) | PPH | PPM |
|-------------------------------|--------------------------|-------------------------|-------------------------|------------------|-------------------|------------|------------|
| Infancia (menor a 11 años) | 5 | 3 | 2 | 13.81 | 22.54 | 1,783.30 | 1,575 |
| Adulto (21-59 años) | 8 | 6 | 2 | 18.67 | 22.74 | 1,716.60 | 1,475 |
| Total | 13 | 9 | 4 | | | | |
| Promedio en tiempo (segundos) | | | | 16.2 | 22.60 | 1,750 | 1,525 |

CAPÍTULO 6

6. ANÁLISIS DE RESULTADOS

Como se puede visualizar en la Tabla 24, Tabla 26 y Tabla 28 contiene los resultados de los datos de los tres intentos realizados por los jugadores en su respectivo orden, donde se puede diferenciar los resultados por categorías o rangos de edades de los participantes, se puede evidenciar en Tabla 24, que representa los resultados del primer intento, en la categoría de infantes, adolescentes y adultos.

El tiempo promedio de hombres (TPH) es menor al tiempo promedio de mujeres (TPM). Sin embargo, los puntos promedio de mujeres (PPM) es mayor a los puntos promedio de los hombres (PPH), es decir, los hombres tardan menor tiempo en terminar el juego serio, pero consiguen menor puntuación; en cambio las mujeres se demoran un mayor tiempo en terminar los diferentes niveles del juego, pero obtienen mayor puntaje, recordando que el puntaje se obtiene mediante el cálculo entre aciertos y errores. Según los resultados obtenidos se puede decir que las mujeres prestan mayor atención y memoria al momento de realizar el juego serio.

La Tabla 26 presenta los datos del segundo intento, el cual permitió analizar que el tiempo promedio de hombres (TPH) es mayor al tiempo promedio de mujeres (TPM). Sin embargo, la diferencia de promedio de los puntos obtenidos en el juego es mínima. Es decir, los hombres tardan mayor tiempo, pero consiguen igual puntuación; en cambio las mujeres se demoran más tiempo en terminar los diferentes niveles del juego, pero obtienen mayor puntaje. Conjuntamente, en la muestra de datos de los hombres adultos indica que mayor es el tiempo que necesita para culminar y obtener igual puntaje en los distintos niveles del juego con respecto a las mujeres.

En la Tabla 28 se logró diferenciar claramente que los hombres de las categorías infancia y adultos, finalizan los niveles en menor tiempo y obtiene mayor puntaje con respecto a las mujeres. Aquí no se encontraron intentos realizados en la categoría adolescentes y adultos mayores.

CAPÍTULO 7

7. ESTADO FINAL

Conclusiones

Este trabajo de titulación aporta con una revisión sistemática de la literatura sobre la aplicación de la ciencia de datos en juegos serios orientados a la atención y memoria, logrando concentrar los estudios y verificar como se abordan los juegos serios en la ciencia de los datos, analizar en que dominios se ha implementado la ciencia de datos y que técnicas/métodos se han implementado y cuáles son las más efectivas para el tratamiento de datos provenientes de juegos serios orientados a la atención y memoria cognitiva. Para lo cual, se han desarrollado actividades dentro de una metodología científica que ha permitido cumplir este objetivo. Obteniendo como resultados información de buena calidad, permitiendo conocer también las brechas de investigación en cada área de aplicación de la ciencia de los datos orientado a los juegos serios.

Principalmente, se ha podido determinar que no existe análisis ni aplicación de la ciencia de los datos orientados a juegos serios dirigidos a adultos mayores. Así como también tiene escasa aplicación de estudios de datos sobre juegos serios orientados al área del comercio y salud.

Posteriormente, se logró analizar los datos mediante la aplicación de la ciencia de los datos, pudiendo descubrir una gran relevancia en las variables de edad, sexo, puntuación y los tiempos de cada uno de los niveles de juego. Es por eso, que se clasificó los datos en categorías por edades, permitiendo analizar el comportamiento de los jugadores en las categorías y contrastar con el sexo, puntajes y tiempos de los jugadores en los tres primeros intentos. Siendo un trabajo riguroso, pero con total satisfacción.

En este sentido, después de la evaluación de los datos estadísticamente se aplicó la ciencia de los datos y se alcanzó a evidenciar que los participantes van adquiriendo experiencia según van realizando más veces el juego. En consecuencia, los tiempos y los puntajes según transcurso de los intentos van mejorando. Además, la categoría de edades influye considerablemente al momento de realizar el juego, no se desenvuelve de la misma manera un adulto mayor que un adolescente, adulto o un niño. Este punto se ha apreciado claramente mediante la normalización y agrupación de los datos.

Recomendaciones

Se recomienda analizar más juegos serios orientados a la atención y memoria utilizando las técnicas de la ciencia de datos empleadas en este estudio para analizar los datos en distintas áreas de aplicación con la finalidad de diversificar el conocimiento y difundir los resultados. Además, utilizando los mismos datos del juego serio expuesto en este artículo, se puede aplicar otras técnicas de la ciencia de datos, con la finalidad de analizar y comparar resultados para evidenciar la información obtenida.

BIBLIOGRAFÍA

- A. Rienzo y C. Cubillos, "Playability and player experience in digital games for elderly: A systematic literature review", *Sensors (Switzerland)*, vol. 20, núm. 14, pp. 1–23, 2020, doi: 10.3390/s20143958.
- Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, 141, 103612. <https://doi.org/10.1016/j.compedu.2019.103612>
- Benmakrelouf, S., Mezghani, N., & Kara, N. (2015). Towards the Identification of Players' Profiles Using Game's Data Analysis Based on Regression Model and Clustering. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 1–8. <https://doi.org/10.1145/2808797.2809429>
- Bente, G., & Breuer, J. (2010). Why so serious? On the Relation of Serious Games and Learning. *Eludamos - Journal for Computer Game Culture*, 4(1), 7–24. <http://www.eludamos.org/index.php/eludamos/article/viewArticle/vol4no1-2>
- Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94, 178–192. <https://doi.org/10.1016/j.compedu.2015.11.003>
- Chaldogeridis, A., & Tsiatsos, T. (2020). Implementation and evaluation of a serious game for working memory enhancement. *Applied Sciences (Switzerland)*, 1-16.
- Duarte, M., Berns, A., Yañez, A., & Dodero, J. (2015). Identifying writing profiles in game-based language learning. *ACM International Conference Proceeding Series*, 263-270.
- F. Ruiz and J. Verdugo, "Guía de Uso de SPEM 2 con EPF Composer," *Univ. Castilla-La Mancha Esc. Super. Informática Dep. Tecnol. y Sist. Inf. Grup. Alarcos*, vol. 3, p. 93, 2008.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996, August). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In *KDD (Vol. 96, pp. 82-88)*.
- González Pacheco, A., Pineda Padilla, L., Sánchez Zhunio, C., Acosta Urigén, M., & Orellana, M. (2021). Interfaces Adaptativas en Juegos Serios para el Entrenamiento y Evaluación Cognitiva, 1(1), 1–8.
- Hernandez, J., Duarte, M., & Dodero, J. (2017). An architecture for skill assessment in serious games based on Event Sequence Analysis. *ACM International Conference Proceeding Series*.
- Hou, H. T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for

- science courses: A video-based process exploration. *Computers in Human Behavior*, 48, 424–435. <https://doi.org/10.1016/j.chb.2015.02.010>
- Hernández, O. J., M.José, R. Q., & Ramírez, R. C. F. (2004). Introducción a la Minería de Datos. 3–39. https://www.mendeley.com/catalogue/be5e9ba7-6e5c-33d5-8ac9-53189de6dd62/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_catalog&userDocumentId=%7Be7ecdfe9-89dc-3ad0-b71e-6333dad4bc0%7D
- Hou, H. T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior*, 48, 424–435. <https://doi.org/10.1016/j.chb.2015.02.010>
- Hou, H. T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior*, 48, 424–435. <https://doi.org/10.1016/j.chb.2015.02.010>
- Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, 72, 757–770. <https://doi.org/10.1016/j.chb.2016.09.062>
- Kitchenham, B. A., Mendes, E., & Travassos, G. H. (2007). Cross versus Within-Company Cost Estimation Studies: A Systematic Review. *IEEE Transactions on Software Engineering*, 33(5), 316–329. <https://doi.org/10.1109/tse.2007.1001>
- Kitchenham, B. A. & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering (EBSE 2007-001). Keele University and Durham University Joint Report.
- Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Methodologies for performance measurement, assessment, and improvement. In *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*. <https://doi.org/10.1007/978-3-319-05834-4>
- Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109. <https://doi.org/10.1016/j.phpro.2012.03.206>
- Nguyen, T. (2016). Serious games. In *Significance* (Vol. 13, Issue 6, pp. 14–19). <https://doi.org/10.1111/j.1740-9713.2016.00978.x>
- Prado, D., Cedillo, P., & Cordova, F. (2020). Improving Cognitive Functions in Older People : Stroop Task Solution. *Advances in Intelligent Systems and Computing*.
- P. Fränti and S. Sieranoja, “K-means properties on six clustering benchmark datasets,” *Appl. Intell.*, vol. 48, no. 12, pp. 4743–4759, 2018, doi: 10.1007/s10489-018-1238-7.
- RapidMiner Inc. (2020). RapidMiner _ Best Data Science & Machine Learning Platform. 2020. <https://rapidminer.com/>

- RSJ de Baker y PS Inventado, "Capítulo X: Minería de datos educativos y aprendizaje Analítico."
- Sandí Delgado, J. C., & Bazán, P. (2019). Educational Serious Games as a Service: Challenges and Solutions. *Journal of Computer Science and Technology*, 19(01), e07. <https://doi.org/10.24215/16666038.19.e07>
- Slimani, A., Elouaai, F., Elaachak, L., Yedri, O. B., & Bouhorma, M. (2018). Learning analytics through serious games: Data mining algorithms for performance measurement and improvement purposes. *International Journal of Emerging Technologies in Learning*, 13(1), 46–64. <https://doi.org/10.3991/ijet.v13i01.7518>
- S. S. Yu, S. W. Chu, C. M. Wang, Y. K. Chan, and T. C. Chang, "Two improved k-means algorithms," *Appl. Soft Comput. J.*, vol. 68, pp. 747–755, 2018, doi: 10.1016/j.asoc.2017.08.032.
- Suhrman, S., Herawan, T., Chiroma, H., & Mohamad Zain, J. (2014). Data Mining for Education Decision Support: A Review. *International Journal of Emerging Technologies in Learning (IJET)*, 9(6), 4. <https://doi.org/10.3991/ijet.v9i6.3950>
- T. Gorschek, P. Garre, S. Larsson, and C. Wohlin, "A model for technology transfer in practice," *IEEE Softw.*, 2006.
- Wang, C., & Huang, L. (2021). A Systematic Review of Serious Games for Collaborative Learning: Theoretical Framework, Game Mechanic and Efficiency Assessment. *International Journal of Emerging Technologies in Learning (IJET)*, 16(06), 88. <https://doi.org/10.3991/ijet.v16i06.18495>
- Y. Li and H. Wu, "A Clustering Method Based on K-Means Algorithm," *Phys. Procedia*, vol. 25, pp. 1104–1109, 2012, doi: 10.1016/j.phpro.2012.03.206.
- Zhang, J., & Lu, J. (2014). Using Mobile Serious Games for Learning Programming. ... 2014, *The Fourth International Conference on ...*, c, 24–29. http://www.thinkmind.org/index.php?view=article&articleid=infocomp_2014_1_50_60029

ANEXOS

Anexo 1

Lista de artículos primarios seleccionados.

| Nro. | titulo | Library | DOI |
|------|---|---------|-------------------------------|
| A001 | Improving Learners' Assessment and Evaluation in Crisis Management Serious Games: An Emotion-based Educational Data Mining Approach | WOS | 10.1016/j.entcom.2021.100428 |
| A002 | Knowledge discovery in a serious game for teaching industrial plants: a case study using decision trees | WOS | 10.5335/rbca.v13i1.11378 |
| A003 | A Systematic Review of Serious Games for Collaborative Learning: Theoretical Framework, Game Mechanic and Efficiency Assessment | WOS | 10.3991/ijet.v16i06.18495 |
| A004 | Cognitive Modeling of Learning Using Big Data From a Science-Based Game Development Environment | WOS | 10.4018/IJGBL.2020100102 |
| A005 | Investigating Navigational Behavior Patterns of Students Across At-Risk Categories Within an Open-Ended Serious Game | WOS | 10.1007/s10758-020-09462-6 |
| A006 | On the Use of Soft Computing Methods in Educational Data Mining and Learning Analytics Research: a Review of Years 2010-2018 | WOS | 10.1007/s40593-020-00200-8 |
| A007 | Fueling Prediction of Player Decisions: Foundations of Feature Engineering for Optimized Behavior Modeling in Serious Games | WOS | 10.1007/s10758-018-9393-9 |
| A008 | The impact of learner metacognition and goal orientation on problem-solving in a serious game environment | WOS | 10.1016/j.chb.2019.08.021 |
| A009 | Applications of data science to game learning analytics data: A systematic literature review | WOS | 10.1016/j.compedu.2019.103612 |
| A010 | A framework for applying learning analytics in serious games for people with intellectual disabilities | WOS | 10.1111/bjet.12625 |
| A011 | Learning Analytics Through Serious Games: Data Mining Algorithms for Performance Measurement and Improvement Purposes | WOS | 10.3991/ijet.v13i01.7518 |
| A012 | Is immersion of any value? Whether, and to what extent, game immersion experience during serious gaming affects science learning | WOS | 10.1111/bjet.12386 |
| A013 | Business intelligence serious game participatory development: lessons from ERPsim for big data | WOS | 10.1108/BPMJ-12-2015-0177 |
| A014 | in-Game Raw Data Collection and Visualization in the Context of the "ThimelEdu" Educational Game | SCOPUS | 10.1007/978-3-030-58459-7_30 |
| A015 | Game learning analytics for educators | SCOPUS | 10.1109/EDUCON.2019.8725089 |
| A016 | Improving serious games analyzing learning analytics data: Lessons learned | SCOPUS | 10.1007/978-3-030-11548-7_27 |
| A017 | Cultural heritage, serious games and user personas based on gardner's theory of multiple intelligences: "the stolen painting" game | SCOPUS | 10.1007/978-3-030-34350-7_47 |

| | | | |
|------|---|--------|----------------------------------|
| A018 | Academic results meanings with data mining mediated by the use of serious video game matellogic | SCOPUS | 10.5373/JARDCS/V11SP11/20193061 |
| A019 | Aging facilities prognostic & health management: Data collection, analysis and use | SCOPUS | 10.3303/CET1977155 |
| A020 | An interactive ecosystem based on Borda voting schemes and serious games to support the discovery of aggressiveness and inhibition traits on scholar children | SCOPUS | 10.1109/CONIELECOMP.2018.8327185 |
| A021 | Solving “smart city” transport problems by designing carpooling gamification schemes with multi-agent systems: The case of the so-called “mordor of Warsaw” | SCOPUS | 10.3390/s18010141 |
| A022 | IT-INDUCED COGNITIVE BIASES IN INTELLIGENCE ANALYSIS: BIG DATA ANALYTICS AND SERIOUS GAMES | SCOPUS | 10.2495/SAFE-V8-N3-438-450 |
| A023 | Predicting learning performance in serious games | SCOPUS | 10.1007/978-3-030-02762-9_14 |
| A024 | Storyboard interpretation technology used for value-based STEM education in digital game-based learning contexts | SCOPUS | 10.5220/0006670700780088 |
| A025 | VR games and the dissemination of cultural heritage | SCOPUS | 10.1007/978-3-319-91125-0_35 |
| A026 | An architecture for skill assessment in serious games based on Event Sequence Analysis | SCOPUS | 10.1145/3144826.3145400 |
| A027 | New Approaches for Integration: Integration of Haptic Garments, Big Data Analytics, and Serious Games for Extreme Environments | SCOPUS | 10.1109/MCE.2017.2715519 |
| A028 | Predicting Math performance of children with special needs based on serious game | SCOPUS | 10.1109/SeGAH.2017.7939276 |
| A029 | Integrating Big Data analytics, virtual reality, and ARAIG to support resilience assessment and development in tactical training | SCOPUS | 10.1109/SeGAH.2017.7939256 |
| A030 | Success factors for serious games to enhance learning: a systematic review | SCOPUS | 10.1007/s10055-016-0298-4 |
| A031 | Convolutional neural network for automatic detection of sociomoral reasoning level | SCOPUS | |
| A032 | The antecedents of and associations with elective replay in an educational game: Is replay worth it? | SCOPUS | |
| A033 | Big data analytics for resilience assessment and development in tactical training serious games | SCOPUS | 10.1109/CBMS.2016.64 |
| A034 | Modeling player decisions in a supply chain game | SCOPUS | 10.1109/CIG.2016.7860444 |
| A035 | Learning analytics for a puzzle game to discover the puzzle-solving tactics of players | SCOPUS | 10.1007/978-3-319-45153-4_89 |
| A036 | Identifying writing profiles in game-based language learning using data mining | SCOPUS | 10.1145/2808580.2808620 |
| A037 | Towards the identification of players' profiles using game's data analysis based on regression model and clustering | SCOPUS | 10.1145/2808797.2809429 |
| A038 | Towards a system of guidance, assistance and learning analytics based on multi agent system applied on serious games | SCOPUS | 10.11591/ijece.v5i2.pp344-354 |
| A039 | A method for real-time stimulation and response monitoring using big data and its application to tactical training | SCOPUS | 10.1109/CBMS.2015.44 |

| | | | |
|------|--|----------|-------------------------------|
| A040 | Knowledge discovery, rehabilitation robotics, and serious games: Examining training data | SCOPUS | 10.1109/biorob.2014.6913838 |
| A041 | Jecripe: How a serious game project encouraged studies in different computer science areas | SCOPUS | 10.1109/SeGAH.2014.7067077 |
| A042 | PEGASO: A personalised and motivational ICT system to empower adolescents towards healthy lifestyles | SCOPUS | 10.3233/978-1-61499-474-9-350 |
| A043 | Extracting knowledge in a game-based learning environment from interaction traces | SCOPUS | |
| A044 | Automated detection of mentors and players in an educational game | SCOPUS | |
| A045 | Intelligent feedback polarity and timing selection in the Shufti Intelligent Tutoring System | SCOPUS | |
| A046 | Learning Analytics and Serious Games: Analysis of Interrelation | IEEE | 10.1109/DeSE.2018.00037 |
| A047 | Learning Analytics in Serious Games: a systematic review of literature | IEEE | 10.1109/ARGENCON.2018.8646166 |
| A048 | Multimodal affective state recognition in serious games applications | IEEE | 10.1109/IST.2016.7738265 |
| A049 | Learning Analytics and Serious Games: Trends and Considerations | ACM | 10.1145/2656719.2656729 |
| A050 | Ensemble Classifiers in a Serious Game for Medical Students in Clinical Cases | ACM | 10.1145/3401895.3402068 |
| A051 | A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes | Springer | 10.1007/s11042-017-4467-6 |
| A052 | Hospital-to-Home Transition for Older Patients: Using Serious Games to Improve the Motivation for Rehabilitation – a Qualitative Study | Springer | 10.1007/s12062-020-09274-7 |
| A053 | Using serious games to manage knowledge and competencies: The seven-step development process | Springer | 10.1007/s10796-016-9649-7 |
| A054 | Bigdata Oriented Multimedia Mobile Health Applications | Springer | 10.1007/s10916-016-0475-8 |
| A055 | Comprehensive review and classification of game analytics | Springer | 10.1007/s11761-020-00303-z |
| A056 | Utilizing Game Analytics to Inform and Validate Digital Game-based Assessment with Evidence-centered Game Design: A Case Study | Springer | 10.1007/s40593-020-00202-6 |
| A057 | A Novel Methodology for Extracting and Evaluating Therapeutic Movements in Game-Based Motion Capture Rehabilitation Systems | Springer | 10.1007/s10916-018-1113-4 |
| A058 | Visualizing event sequence game data to understand player’s skill growth through behavior complexity | Springer | 10.1007/s12650-019-00566-5 |
| A059 | Improving evidence-based assessment of players using serious games | EBSCO | 10.1016/j.tele.2021.101583 |
| A060 | Predicting students' knowledge after playing a serious game based on learning analytics data: A case study | Hinari | 10.1111/jcal.12405 |

Anexo 2

Resultados de la evaluación a cada artículo con cada pregunta (RQ1) y sus criterios de extracción.

RQ1 ¿Qué tipo de información se requiere para analizar juegos serios?

| # | Biblioteca | EC01: Lugar de despliegue | | | EC02: Plataformas de despliegue | | | EC03: Área de aplicación | | | | EC04: Tipos de usuarios | | | EC05: Tipos de herramienta de desarrollo | | | EC06: Modos de interacción | | | | EC07: Puntajes | | EC08: Clasificación de jugadores | | |
|------|------------|---------------------------|-----|-----|---------------------------------|----------|--------|--------------------------|-----------|----------|-------|-------------------------|------------|----------------|--|----------|-----------|----------------------------|------------|------------|-------|----------------|----|----------------------------------|--------|-------|
| | | Consola | APP | WEB | Computador | Teléfono | tablet | Salud | Educación | Comercio | Otros | 5-17 años | 18-64 años | más de 65 años | Metodología | Programa | Framework | Multijugador | Un jugador | Co-tutoría | Otros | Si | No | Puntajes | Edades | Otros |
| A001 | WOS | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| A002 | WOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| A003 | WOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| A004 | WOS | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A005 | WOS | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| A006 | WOS | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| A007 | WOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| A008 | WOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| A009 | WOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A010 | WOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A011 | WOS | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A012 | WOS | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| A013 | WOS | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| A014 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| A015 | SCOPUS | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A016 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| A017 | SCOPUS | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | |
| A018 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| A019 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A020 | SCOPUS | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| A021 | SCOPUS | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | | | | | | | |
|------|----------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A022 | SCOPUS | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A023 | SCOPUS | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| A024 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| A025 | SCOPUS | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| A026 | SCOPUS | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| A027 | SCOPUS | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| A028 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| A029 | SCOPUS | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| A030 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A031 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| A032 | SCOPUS | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| A033 | SCOPUS | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| A034 | SCOPUS | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| A035 | SCOPUS | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| A036 | SCOPUS | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| A037 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| A038 | SCOPUS | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A039 | SCOPUS | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| A040 | SCOPUS | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| A041 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| A042 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A043 | SCOPUS | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A044 | SCOPUS | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| A045 | SCOPUS | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| A046 | IEEE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A047 | IEEE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A048 | IEEE | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A049 | ACM | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A050 | ACM | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| A051 | Springer | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| A052 | Springer | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| A053 | Springer | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| A054 | Springer | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|------|----------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A055 | Springer | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| A056 | Springer | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| A057 | Springer | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| A058 | Springer | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| A059 | EBSCO | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| A060 | Hinari | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |

Anexo 3

Resultados de la evaluación a cada artículo con cada pregunta (RQ2 y RQ3) y sus criterios de extracción.

| | | RQ2 ¿Cómo se aborda los juegos serios en la ciencia de los datos? | | | | | | | | | | | RQ3 ¿En qué área de desarrollo se aplican juegos serios? | | | | | | |
|------|------------|---|----------|--------------------------|-------------------------------|---------------------|------------------|---------------------------|---------------------------------|------------------|--------------------|-----------------------------|--|-----|--------|---------------------------|---------|----------|----------|
| # | Biblioteca | EC09: Data Preprocessing | | | EC10: Clasificación | | | | | EC11: Clustering | | | EC12: Como se almacenan los datos | | | EC13: Áreas de despliegue | | | |
| | | Distance measure | Sampling | Dimensionality reduction | SVM (support vector machines) | Arboles de decisión | Redes bayesianas | KNN (k-nearest neighbors) | ANN (Artificial Neural Network) | k-means | Basado en densidad | Paso de mensajes Jerárquico | Json | SQL | No SQL | Industria | Empresa | Academia | Medicina |
| A001 | WOS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A002 | WOS | | | | | | | | | | | | | | | | | | |
| A003 | WOS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A004 | WOS | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A005 | WOS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A006 | WOS | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A007 | WOS | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A008 | WOS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A009 | WOS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

| | | | | | | | | | | | | | | | | | | | |
|------|--------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A010 | WOS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| A011 | WOS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A012 | WOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A013 | WOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| A014 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A015 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A016 | SCOPUS | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A017 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A018 | SCOPUS | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A019 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A020 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A021 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| A022 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| A023 | SCOPUS | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A024 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A025 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A026 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| A027 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| A028 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A029 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A030 | SCOPUS | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A031 | SCOPUS | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A032 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A033 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A034 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A035 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A036 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A037 | SCOPUS | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A038 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A039 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A040 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A041 | SCOPUS | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A042 | SCOPUS | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

| | | | | | | | | | | | | | | | | | |
|------|----------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A043 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A044 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A045 | SCOPUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A046 | IEEE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A047 | IEEE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A048 | IEEE | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A049 | ACM | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A050 | ACM | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A051 | Springer | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A052 | Springer | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A053 | Springer | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| A054 | Springer | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| A055 | Springer | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| A056 | Springer | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A057 | Springer | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| A058 | Springer | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| A059 | EBSCO | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| A060 | Hinari | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |