



**Universidad Abierta  
Interamericana**

Plataforma Virtual de Educación STEAM

Evaluación Estadística y Minería de Rendimiento

Tutoría técnica: Dra. Daniela López De Luise

Profesora de Trabajo Final: Dra. Marcela Samela

Alumno: Sergio Miguel Leiva

Trabajo Final de Carrera presentado para obtener el título de

Lic. en Gestión de Tecnología Informática

Diciembre, 2022

### Resumen

El Museo Virtual STEAM (acrónimo del inglés Science, Technology, Engineering, Art y Mathematics) es una plataforma tecnológica creada para facilitar el acceso a las distintas actividades formativas promovidas por este modelo educativo bajo la modalidad STEAM. A través de la misma se logró recabar, integrar y analizar los datos de la propia plataforma, de la interacción del usuario con el sistema y los surgidos de la implementación de las actividades.

Los datos de cada cápsula se encontraban distribuidos en distintas locaciones, a cargo de diferentes actores y sin ser explotados en beneficio de la educación fomentada por la plataforma. En consecuencia, dichos datos fueron seleccionados, almacenados, clasificados, integrados y relacionados de manera tal que permitan una explotación más profunda de la información. A partir de allí, se realizó un análisis estadístico y se aplicó minería de datos a dicha información, utilizando los softwares Infostat y Weka, respectivamente. Dichos procesos permitieron la evaluación y explotación de datos de forma más exhaustiva.

Esta investigación posibilitó la obtención sistemática y el estudio por minería de datos de conocimientos que mejoran el entendimiento de las cápsulas y su interrelación. Como consecuencia, se comprende mejor al proceso educativo a través de la plataforma, brindando posibilidades de mejora para la misma, para sus cápsulas, actividades, contenido y navegabilidad. Así, se colabora con su evolución y mejora en pos del beneficio de cada participante.

**Palabras clave:** base de datos, educación STEAM, estadística, minería de datos, museos virtuales

### **Abstract**

The Virtual STEAM Museum (an acronym for Science, Technology, Engineering, Art and Mathematics) is a technological platform created to facilitate access to the different training activities promoted by this educational model under the STEAM mode. Through it, it was possible to collect, integrate and analyse the data from the platform itself, the interaction of the user with the system and the data arising from the implementation of the activities.

The data of each capsule was distributed in different locations, in charge of other actors and without being exploited for the benefit of education fostered by the platform. As a consequence, such data was selected, stored, classified, integrated and related in such a way as to allow deeper exploitation of information. As a result, a statistical analysis was carried out and data mining was applied to such information, using the Infostat and Weka software, respectively. Those processes allowed the evaluation and exploitation of data in a more exhaustive way.

This investigation allowed us to systematically obtain and study through data mining the knowledge that improves the understanding of the capsules and their interrelationship. As a consequence, the educational process is better understood through the platform, offering possibilities of improvement for it, for its capsules, activities, content and navigability. In that way, collaborating with its evolution and improvement in pursuit of the benefit of each participant.

**Keywords:** data mining, databases, statistics, STEAM education, virtual museums

## **Dedicatoria**

A Sofi.

## **Reconocimientos**

En esta obra se resumen todos los momentos vividos en mi trayecto académico. Cada clase, trabajo, examen, victoria, derrota, alegría, frustración, viaje caluroso, noche fría, reunión familiar o de amistad ausente. Cada “no puedo”, cada disculpa, cada mensaje, cada espera y cada festejo.

Sin dudas, todos ellos precisaron principalmente de mucha dedicación. Pero no existe esfuerzo alguno suficiente, nada hubiera alcanzado para llegar a la meta si no hubiera contado con el apoyo de muchas personas que hicieron esto posible.

En primer lugar, quiero agradecerle a mi familia por su apoyo incondicional. Siempre atentos y dispuestos a ayudarme en cada detalle que estuviera a su alcance para hacer mis días más fáciles y felices.

A Laura, por su comprensión, amor, tiempo y atención. Cada paso era compartido y tomaba sentido al ser acompañado por ella. Logrando que en ningún momento me sienta solo en esta experiencia.

A mis compañeros por toda ayuda, apoyo y trabajos que resolvimos juntos. Especialmente a mis compañeras del proyecto de investigación de tesis, Sofía R. y Cecilia S. Aprendimos y recorrimos el camino juntos, y seguramente continúe siendo así.

A Marcela, mi profesora de Trabajo Final, por sus enseñanzas y recomendaciones para que cada línea de esta obra sea mejor. Gracias por tus infaltables palabras de apoyo y por buscar y encontrar siempre lo mejor de mí. Gracias también a toda la comunidad que forma parte de la universidad, por darme el espacio para desarrollarme y crecer.

Por último, un agradecimiento especial a mi tutora, la Dra. Daniela López De Luise. Gracias por tu guía en este camino, por tantas enseñanzas, reuniones, correos y recomendaciones. Tu paciencia y responsabilidad para con mi trabajo fue incondicional. Siempre con la dedicación y vocación por enseñar que te caracteriza.

A todos ellos, y a todos los que formaron parte y olvido mencionar les digo: Este trabajo lleva mi nombre, pero sin duda alguna, es un poco de todos. Muchas gracias.

## Índice general

RESUMEN .....	1
ABSTRACT.....	2
DEDICATORIA .....	3
RECONOCIMIENTOS .....	3
ÍNDICE DE FIGURAS.....	6
ÍNDICE DE TABLAS .....	6
<b>1. INTRODUCCIÓN .....</b>	<b>8</b>
1.1    PLANTEAMIENTO DEL PROBLEMA .....	9
1.2    PREGUNTAS DE LA INVESTIGACIÓN.....	9
1.3    HIPÓTESIS .....	10
1.4    JUSTIFICACIÓN DE LA PROPUESTA.....	10
1.5    OBJETIVOS .....	11
1.5.1    Objetivos Generales .....	11
1.5.2    Objetivos Específicos.....	11
1.6    CONTRIBUCIONES PRINCIPALES .....	12
1.7    ESTRUCTURA GENERAL DEL TRABAJO DE INVESTIGACIÓN.....	12
1.7.1    Capítulo 1 - Introducción .....	12
1.7.2    Capítulo 2 – Marco Teórico.....	13
1.7.3    Capítulo 3 - Desarrollo Técnico.....	13
1.7.4    Capítulo 4 - Prototipo.....	14
1.7.5    Capítulo 5 – Informe Final.....	14
1.7.6    Capítulo 6 - Conclusiones .....	15
1.7.7    Capítulo 7 – Líneas Futuras De Investigación.....	15
1.7.8    Capítulo 8 – Anexos.....	15
1.7.9    Capítulo 9 – Acrónimos .....	16
1.7.10    Capítulo 10 – Referencias .....	16
<b>2. MARCO TEÓRICO.....</b>	<b>16</b>
2.1    MODELO EDUCATIVO STEAM .....	16
2.2    MUSEOS FÍSICOS Y VIRTUALES .....	23
2.3    LA TECNOLOGÍA COMO APOYO PARA EL ANÁLISIS DE DATOS.....	27

3.	DESARROLLO TÉCNICO.....	37
3.1	INTRODUCCIÓN .....	37
3.2	METODOLOGÍA DE INVESTIGACIÓN.....	38
3.2.1	Técnicas de Recolección de Datos.....	39
3.2.2	Datos Analizados .....	40
3.2.3	Materiales Utilizados .....	43
3.2.4	Métodos Aplicados .....	45
3.3	DESARROLLO DE LA INVESTIGACIÓN .....	46
3.3.1	Etapa 1: Pre-procesamiento .....	47
3.3.2	Etapa 2: Procesamiento De Datos.....	49
4.	PROTOTIPO.....	68
4.1	INTRODUCCIÓN .....	69
4.1.1	Apariencia De La Plataforma.....	69
4.1.2	Base De Datos.....	73
4.2	TESTING.....	74
5.	INFORME FINAL.....	74
6.	CONCLUSIONES .....	76
7.	LÍNEAS FUTURAS DE INVESTIGACIÓN.....	77
8.	ANEXOS .....	79
8.1	ANEXO I – DIAGRAMAS ENTIDAD RELACIÓN.....	79
8.2	ANEXO 2 – ESTRUCTURA DE CAMPOS DE CADA DATASET .....	81
9.	ACRÓNIMOS.....	82
10.	REFERENCIAS.....	84

### Índice de Figuras

Figura 1 Buffer Weka de la ejecución 1 .....	52
Figura 2 Buffer Weka de la ejecución 2 .....	53
Figura 3 Buffer Weka de la ejecución 3 .....	53
Figura 4 Buffer Weka de la ejecución 4 .....	55
Figura 5 Buffer Weka de la ejecución 5 .....	56
Figura 6 Buffer Weka de la ejecución 6 .....	56
Figura 7 Buffer Weka de la ejecución 7 .....	57
Figura 8 Buffer Weka de la ejecución 8 .....	57
Figura 9 Buffer Weka de la ejecución 9 .....	58
Figura 10 Buffer tracking de usuarios .....	59
Figura 11 Buffer tracking de usuarios reducido .....	60
Figura 12 Resultado Infostat: Tablas de frecuencia con el dataset completo .....	62
Figura 13 Resultado Infostat: Gráfico de dispersión con el dataset completo.....	63
Figura 14 Resultado Infostat: Gráfico de dispersión por edades recomendadas .....	63
Figura 15 Resultado Infostat: Histograma con el dataset por tema de cápsula .....	64
Figura 16 Tabla de frecuencias tracking de usuarios.....	65
Figura 17 Frecuencias relativas por región de acceso .....	65
Figura 18 Accesos por elemento.....	66
Figura 19 Gráfico accesos cápsula-administrativo .....	66
Figura 20 Prototipo - Pantalla de Inicio del Museo Virtual STEAM .....	69
Figura 21 Prototipo - Pantalla Nosotros.....	70
Figura 22 Prototipo - Pantalla Participe.....	71
Figura 23 Prototipo - Pantalla Blog .....	72
Figura 24 DER inicial de la plataforma virtual.....	79
Figura 25 DER de entidades incorporadas.....	80

### Índice de Tablas

Tabla 1 Datos principales de cápsulas y actividades .....	50
Tabla 2 Identificadores del dataset cápsulas y actividades completo .....	51
Tabla 3 Matriz de confusión con dataset edades recomendadas.....	54
Tabla 4 Resultado Infostat: Medidas resumen con el dataset completo .....	61
Tabla 5 Datos de formularios post actividad .....	67
Tabla 6 Estructura de campos: Dataset cápsulas y actividades .....	81

Tabla 7 Estructura de campos: Dataset Post Implementación .....81

Tabla 8 Estructura de campos: Dataset Tracking de Usuarios .....82

## 1. Introducción

La educación STEAM es un modelo educativo que pretende acercar al estudiante a las ciencias, desarrollar su vocación y lograr, por ejemplo, disminuir la deserción académica del mismo (Moreira-Mora, 2007).

Algunos de los establecimientos encargados históricamente de brindar educación son los museos. Los mismos, se reconocen como lugares dedicados al estudio científico, al desarrollo de la ciencia, a la promoción y divulgación del conocimiento. Estas características los enlazan directamente con la enseñanza y aprendizaje de las personas, principalmente cuando transcurren por medio de este tipo de instituciones.

Acercándonos a un contexto actual, a partir de la pandemia COVID 19 en el año 2020, las instituciones educativas no estuvieron exentas de la necesidad de enfocar sus actividades hacia una modalidad mucho más virtual que la practicada hasta el momento. El carácter de presencialidad que requerían las actividades habituales ya no era viable, se necesitaba ampliar el rumbo de la enseñanza.

Teniendo en cuenta los cambios y restricciones de circulación en la sociedad a partir de la pandemia y el incesante avance de la tecnología, se potenciaron los entornos virtuales para todo tipo de actividad, incluyendo la educativa.

Todas estas modalidades de enseñanza generan datos que, de ser analizados y trabajados adecuadamente, producen un mejor conocimiento de la educación y sus procesos (Chen y otros, 2021; Tabuenca y otros, 2021).

Para la difusión a distancia de la educación denominada STEAM, se creó su Museo Virtual. A través de esta plataforma se presentan las distintas actividades y los recursos necesarios para poder implementarlas. Todos los cuales, se enmarcan bajo el concepto de cápsula: material teórico, videos, imágenes, información de recursos humanos participantes, elementos necesarios de aplicación, etc.

El presente trabajo propone la integración de los datos provenientes de la plataforma, de la interacción de los usuarios con la misma, y de los resultados de las actividades. Todo ello en un modelo único y reproducible que permita la evaluación del proceso de enseñanza. De esta forma, se logra obtener la retroalimentación necesaria para la mejora y crecimiento de todo el modelo en su conjunto. Dicha información permite una actualización constante de cada etapa involucrada.

Con ese propósito, la presente investigación aborda inicialmente un marco conceptual que describe los principales conocimientos actuales en la materia, desde el punto de vista de diferentes autores. Se puntualiza sobre algunos trabajos destacados sobre educación, tecnología y herramientas para la evaluación de los procesos involucrados. A partir de ese contexto, se expone el desarrollo técnico de la investigación, sus correspondientes resultados, conclusiones y sugerencias sobre posibles trabajos futuros.

### **1.1 Planteamiento del Problema**

Para una aplicación virtual del proceso educativo presentado se desarrolla el Museo Virtual STEAM, primer museo destinado a esta modelo educativo y caracterizado principalmente por su condición virtual. Esta plataforma permitirá el examen de los todos los datos involucrados referentes a alumnos, profesores, colaboradores, navegabilidad y actividades realizadas.

Todos los procesos y personas implicadas en el circuito de la plataforma generan un gran volumen de tráfico de datos en cada etapa. Esos datos se encuentran dispersos en distintas locaciones, almacenados en distintos tipos de fuentes y formatos, y a cargo de diferentes referentes. Asimismo, actualmente no se cuenta con un soporte que integre todos estos datos ni procedimientos de análisis de los mismos. Esto último, permitiría extraer, mediante la inteligencia computacional, información valiosa para el control, evaluación y actualización del modelo educativo propuesto.

En el contexto mencionado, se presenta la necesidad e importancia de poder integrar y analizar dicha información. De esta forma, se podrá lograr encontrar estándares y/o tendencias que posibiliten retroalimentar y mejorar el rendimiento de cada actividad, del proceso y, consecuentemente, del alumno. Toda la información disponible debe poder ordenarse, analizarse, estudiarse y explotarse para asistir y mejorar al proceso mismo, a la vez que se evalúan los procedimientos aplicados y los rendimientos de las actividades para favorecer su desarrollo.

### **1.2 Preguntas de la Investigación**

El análisis necesario mencionado, al ser enfocado en la fuente de información de la plataforma da lugar a las siguientes preguntas:

¿Es información suficiente la suministrada y almacenada en la base de datos de la plataforma como soporte de análisis de resultados?

¿Qué conocimiento aporta el análisis estadístico y la minería de datos de la información obtenida?

¿Qué proporciona al proceso de mejora de las actividades STEAM la definición de métricas?

¿Es posible la evaluación tecnológica de actividades STEAM desarrolladas en condiciones de carencias tecnológicas?

¿Se obtienen resultados favorables y suficientes al aplicar este modelo de evaluación aun cuando los contextos sociales, económicos y tecnológicos de las actividades realizadas son muy distintos?

### **1.3 Hipótesis**

A partir de lo hasta aquí expuesto, y de los interrogantes surgidos en consecuencia, se plantean las siguientes aseveraciones:

- La obtención de estadísticas y la aplicación de minería de datos a la información resultante de la Plataforma Virtual STEAM favorece la evaluación y evolución del proceso educativo.
- La aplicación de procesos tecnológicos de evaluación STEAM con resultados favorables es posible aún en estudiantes sumidos en un contexto de carencias tecnológicas.
- La definición de métricas respecto al uso de la plataforma y el resultado de sus actividades permite mejorar el contenido y rendimiento de la Plataforma Virtual STEAM en forma continua.

### **1.4 Justificación de la Propuesta**

Las actividades STEAM se realizan desde su concepción de forma presencial entre los distintos actores. A medida que los procesos educativos y las tecnologías de la información avanzan y se masifican surgen nuevas alternativas de aplicación, como las virtuales.

Con el apoyo conjunto del Museo Histórico Sarmiento, la Sociedad Científica Argentina, la escuela Manuel Dorrego de Buenos Aires y CI2S Labs<sup>1</sup> (del inglés Computational

---

<sup>1</sup> Equipo de investigadores formados en 2013, abocados a la investigación y desarrollo referida a la inteligencia computacional.

Intelligence & Information Systems Labs) se propone esta investigación. En la misma, a través del proyecto denominado Lincievís y en mi rol como integrante colaborador del CAETI, se plantea un modelo de análisis y evaluación que abarque de forma íntegra a toda actividad referida al Museo Virtual STEAM y sus extensiones presenciales.

Dicha plataforma web, pretende dar una nueva perspectiva al trasladarle a esa virtualidad un contexto que tome las características y virtudes de los museos físicos, pero traspolándolos al mundo digital.

En este escenario, la principal fuente de información necesaria para registrar y analizar los datos obtenidos se encontrará en la base de datos que da soporte tecnológico a dicha plataforma. La misma, no solo almacena los datos propios de la plataforma y todo el material educativo que exhibe, sino que es plausible de adaptaciones. Estas actualizaciones permiten la incorporación de nuevos datos. En el caso de la competencia del presente trabajo, eso significa la posibilidad de incorporar, por ejemplo, los datos de navegación de los usuarios en la web del museo virtual, y los resultados de las implementaciones de las actividades.

De esta forma, se podrá obtener el conocimiento integral y necesario que permita el perfeccionamiento de cada cápsula y cada actividad. A su vez, permitirá modelar el progreso de los alumnos y alcanzar la versatilidad requerida en cada etapa del proceso para su evolución. Este conocimiento puede ser extraído a través de la obtención de estadísticas y la aplicación de minería de datos sobre toda la información disponible.

## **1.5 Objetivos**

### ***1.5.1 Objetivos Generales***

El presente trabajo posee como principal objetivo evaluar el rendimiento y performance del modelo educativo propuesto por el Museo Virtual STEAM.

Asimismo, a través de dichos procesos, se pretende asistir a la mejora y perfeccionamiento del proceso educativo mencionado. Esto podrá ser realizado a partir de la obtención del conocimiento extraído de los datos analizados.

### ***1.5.2 Objetivos Específicos***

En cuanto a los objetivos específicos perseguidos, se puede mencionar:

Crear un modelo de datos que brinde un soporte tecnológico íntegro al almacenamiento de los datos del Museo Virtual STEAM y a la implementación de sus actividades.

Con dicho modelo creado y adaptado a las necesidades antes mencionadas, se realizará un análisis estadístico y emplearán procesos de minería de datos para la evaluación de su rendimiento y performance. Dicho análisis abarca cápsulas, navegación web y actividades educativas STEAM.

A partir de allí, se analizarán los resultados obtenidos con el propósito de identificar las oportunidades de mejora en cada etapa del proceso.

## **1.6 Contribuciones Principales**

Este trabajo aporta un análisis sistemático de toda la información recabada a partir de la plataforma Museo Virtual STEAM y sus actividades.

En el mismo se captura y analiza la información resultante de dichos procesos colaborando con la mejora y constante evolución del modelo educativo.

Los procesos descritos son de carácter iterativo e incremental. Esto alude a que luego del análisis de los datos mediante herramientas de inteligencia computacional, se procede a la obtención de resultados que permitan una retroalimentación del propio sistema. De esta forma, luego de cada iteración y de los procedimientos de mejora consecuentes, el modelo propuesto por el Museo Virtual STEAM es superior a su propia versión anterior.

Este modelo original de análisis y evaluación propuesto significa una mejora en el aprendizaje de ciencias a través del ámbito virtual de la plataforma, incorporando conocimientos a los procesos existentes a partir de los procesos y herramientas expuestas.

Asimismo, y desde otra perspectiva, todas las mejoras llevadas a cabo en la plataforma web, sus procesos y actividades, devengarán en una mejora de la enseñanza STEAM. Con lo cual, los estudiantes serán los principales beneficiados al formar parte de un modelo enriquecido y perfeccionado de forma continua.

## **1.7 Estructura General del Trabajo de Investigación**

A continuación, se precisa un breve resumen del contenido de cada capítulo.

### **1.7.1 Capítulo 1 - Introducción**

El capítulo I comienza con la presentación y caracterización de la plataforma del Museo Virtual STEAM. A partir de allí se hace un recorrido exponiendo todos los datos que son generados por y como efecto de la misma. Teniendo en cuenta esta gran cantidad de datos,

surge la necesidad de su explotación, en pos de obtener una retroalimentación que mejore al propio modelo en sus distintas etapas.

Como consecuencia de este planteamiento, se manifiestan las preguntas que guiarán la investigación y serán contrastadas a lo largo de la obra por las hipótesis también explicitadas. Luego, se expone como resultante, la justificación de la investigación.

Por último, son detallados tanto los objetivos generales como específicos del trabajo, finalizando con las contribuciones principales que el mismo aporta.

### ***1.7.2 Capítulo 2 – Marco Teórico***

El capítulo II se inicia abordando el marco teórico del trabajo. En el mismo, se brindan las nociones y beneficios más importantes por conocer del modelo educativo denominado STEAM.

A partir de allí, se continúa profundizando en los trabajos relacionados e investigaciones recientes. En esta sección se exponen trabajos de diferentes autores referidos a la educación, la tecnología, herramientas de análisis de datos, etc. De esta forma, se realiza una introducción al conocimiento y tecnología actual utilizada para ahondar en la información de algunas disciplinas como la educación.

Luego, se exponen conceptos relativos a los museos y cómo estos fueron transformándose para adaptarse al mundo virtual que la pandemia COVID 19 promovió.

Se continúa con la explicación de algunas tecnologías que serán de mucha valía para la explotación de los datos concernientes a la plataforma propuesta. De esta forma, se versan temas como la inteligencia artificial, minería de datos, aprendizaje automático y estadística, entre otros.

Con los conceptos hasta aquí expuestos, se puede abordar el siguiente capítulo, referido específicamente a la arista técnica de la presente obra.

### ***1.7.3 Capítulo 3 - Desarrollo Técnico***

El capítulo III comienza con la exposición de la metodología de investigación que enmarca esta obra. Para la misma, se puntualiza, en primer lugar, sobre las técnicas de recolección de datos utilizadas.

Luego, como las técnicas mencionadas deben aplicarse sobre un conjunto de datos en particular, se explicita cuáles son las cápsulas escogidas que brindarán su información para el análisis ulterior.

A continuación, se detallan los materiales o herramientas tecnológicas principales utilizadas para llevar a cabo los procedimientos estadísticos y de minería de datos. En este punto, son declaradas las aplicaciones y las características más importantes que se tuvieron en cuenta y justifican su elección.

Posteriormente, son descriptos los métodos aplicados en cada etapa del proceso. Dichos procedimientos, de naturaleza estadística y de minería de datos, son llevados a cabo y descriptos hacia el final del capítulo.

Una vez finalizados los procesos detallados anteriormente, ya se cuenta con los resultados disponibles para su análisis y posterior extracción de conclusiones.

#### ***1.7.4 Capítulo 4 - Prototipo***

En el capítulo IV se presenta el prototipo que sirve de base para el desarrollo del trabajo. Del mismo, se presentan las capas más importantes a considerar. En este caso, presentación al usuario y base de datos.

En cuanto a la presentación al usuario, se describen las 4 pantallas más importantes para la captura de datos de navegación del usuario. De cada una, se ilustra su diseño y se especifica la importancia de su incorporación.

En referencia a la base de datos, se detallan tanto la estructura de tablas original de la plataforma como las nueva creada para esta investigación. A partir de la conjunción de estas estructuras, el sistema se encuentra preparado para recibir la información que nutrirá los procedimientos estadísticos y de data mining.

Por último, se detallan las modificaciones sufridas por la estructura de datos finalmente utilizada. Estas pruebas permitieron la puesta a punto de la nueva distribución de componentes.

#### ***1.7.5 Capítulo 5 – Informe Final***

En el capítulo V, se presenta el informe final con los resultados y recomendaciones destacadas a partir de los resultados obtenidos en cada proceso realizado.

En el mismo, se tratan temas como la información de las cápsulas, el nivel de dificultad de las actividades realizadas y los inconvenientes que se presentan a la hora de ser implementadas. Por otro lado, se analizan los resultados a partir de la perspectiva de las entidades participantes y de la navegación de los usuarios en la plataforma.

### ***1.7.6 Capítulo 6 - Conclusiones***

En el capítulo VI, se realizan las conclusiones de toda la investigación realizada.

En el mismo, se comienza exponiendo los beneficios encontrados en la implementación de la propuesta y la integración de la información en una misma base de datos. Luego, se mencionan las conclusiones que se extraen de los datos y métricas utilizadas. Esto da lugar a las menciones sobre las cápsulas de forma general y luego, la puntualización en las de tipo extremo.

De esta forma, se versa de forma integral sobre las preguntas de la investigación, las hipótesis propuestas, objetivos y beneficios del trabajo.

Finalmente, se mencionan las limitaciones identificadas durante el desarrollo de la investigación.

### ***1.7.7 Capítulo 7 – Líneas Futuras De Investigación***

En el capítulo VII, se exponen las líneas futuras de investigación identificadas.

Las mismas, responden principalmente a las limitaciones encontradas en la obra. De esta forma, se exponen posibles futuras investigaciones referidas a la educación del tipo STEAM extremo, la automatización de tareas y la minería textual.

### ***1.7.8 Capítulo 8 – Anexos***

En el capítulo VIII, se presentan los anexos que acompañan a la investigación.

En los mismos, en primer lugar, se pueden visualizar los diagramas entidad relación que dan soporte tecnológico para el almacenamiento de toda la información tratada. Luego, se exhiben las estructuras de datos que corresponden a cada dataset utilizado en las ejecuciones con los softwares Weka e Infostat.

### **1.7.9 Capítulo 9 – Acrónimos**

En el capítulo IX, se describen los distintos acrónimos utilizados durante la investigación. Para cada uno, se expone su significado y una breve explicación para su entendimiento.

### **1.7.10 Capítulo 10 – Referencias**

En el capítulo X, se detalla la bibliografía correspondiente a cada una de las citas utilizadas en la presente obra.

## **2. Marco Teórico**

En este capítulo se desarrolla el concepto STEAM, su historia y evolución. Luego, se realiza una revisión de cómo algunas entidades vinculadas al conocimiento evolucionaron hacia el mundo virtual. Finalmente, se estudian las herramientas tecnológicas que darán soporte al análisis de datos presentado en esta investigación.

### **2.1 Modelo Educativo STEAM**

#### ***Historia y Concepto de la Educación STEAM***

El modelo educativo STEM se construye con el objetivo de brindar a estudiantes la posibilidad de conocer y entender las ciencias. El mismo nace en 2001 a manos de Ramaley, con el objeto de integrar distintos conceptos educativos en la enseñanza de ciencias (López De Luise & Ruiz Tabarez, 2020b). Los autores mencionados comentan que los resultados de esta integración no solo se observan en modo teórico o con ensayos en laboratorio, sino que influyen directamente en el accionar de las personas. Esto se debe a que el modelo STEAM toma principal importancia en su formación educativa, ya que para un entendimiento básico del mundo resultan muy importantes las matemáticas y las ciencias. Mientras que, desde otro punto de vista, para poder interactuar con el mismo son esenciales la tecnología y la ingeniería (Christenson, 2011).

Inicialmente, la educación STEM (acrónimo del inglés Science, Technology, Engineering and Mathematics) en su concepción no incluía a las artes, pero hacia el año 2014 se las incorpora y pasan a formar parte de STEAM. Dicho acontecimiento, colabora incentivando aún más a los niños a pensar como científicos, y se sustenta en que las relaciones

y uniones entre el arte y la tecnología cobran cada vez más importancia en la sociedad y en su desarrollo (Chung, 2014).

A fines de la década del 2000, como una nueva evolución del modelo, comienza a desarrollarse el concepto de STEAM+H. Dicho acrónimo resulta de añadir a la sigla ya conocida, la H de Humanidades.

Este enfoque dota a la educación STEAM de un nuevo punto de vista, el cual intenta formar personas que orienten sus conocimientos de ciencias y objetivos en pos de la resolución de “retos globales, como el cambio climático y la pobreza, los alumnos tendrían que contar con las capacidades (conocimientos, competencias, valores y actitudes) para construir un mundo más justo, pacífico, tolerante, inclusivo, sostenible y seguro” (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2021).

Con esta nueva perspectiva, ya se desarrollan a nivel internacional distintas conferencias que promueven los aspectos mencionados. Por ejemplo, una que se destaca por su poder de convocatoria e influencia es la realizada en México 2021, la cual fue promovida principalmente con la visión de por qué es “necesaria la educación STEAM+H para cultivar un pensamiento y habilidades transformadoras, innovadoras y creativas para avanzar hacia un desarrollo sostenible” (UNESCO, 2021).

De esta manera, y avanzando a cada momento, se conforma el modelo completo como se conoce hoy en día, el cual continúa evolucionando y generando beneficios educativos a los alumnos ya sea con actividades educativas del tipo curricular o extracurricular.

### *Beneficios del Modelo Educativo*

Las actividades STEAM no solo colaboran con la formación educativa del estudiante sino con el desarrollo del carácter del mismo. Paralelamente a la incorporación de conocimientos, el estudiante también desarrolla su responsabilidad, su capacidad para enfrentar y superar obstáculos que dificulten su tarea, ser más positivos en su vida diaria, compartir y formar parte de un equipo (Flowers, 2019).

Además, según Xueyun & Zihui (2010), los estudiantes pueden obtener mejores resultados en las autoevaluaciones a través de las actividades educativas STEAM. Esto también se debe a que las mismas promueven la responsabilidad y el trabajo autónomo de los mismos (Xueyun y Zihui, 2010; Zawieska y Duffy, 2015).

Por otro lado, la evaluación exhaustiva de su desarrollo puede permitir una detección temprana de un posible abandono académico, brindando la información necesaria para actuar a tiempo y evitar la deserción. (López De Luise & Ruiz, 2018).

Teniendo en cuenta la concepción e irrupción del modelo STEAM en la educación de ciencias y los efectos en el aprendizaje de los estudiantes, se han realizado diferentes trabajos relacionados. Al respecto, se detallan a continuación algunos que se destacan en función de la presente obra.

### *Trabajos Relacionados - Investigaciones Recientes*

Desde su aparición, han sido desarrollados diferentes estudios sobre el modelo educativo STEAM. Muchos de los cuales, fueron o son guiados por entidades reconocidas en la comunidad científica de nuestro país.

En tal sentido, resulta interesante para este trabajo evaluar en detalle algunos casos particulares enfocados en el aprendizaje, actividades STEAM y su relación con herramientas tecnológicas de análisis.

Se analiza en primer lugar, un trabajo llevado a cabo con el apoyo de entidades como la SCA (acrónimo de Sociedad Científica Argentina), IEEE (acrónimo del inglés Institute of Electrical and Electronics Engineers), Games Technical Committee de Argentina y Lincievís Lab. Dichas instituciones cumplen un rol fundamental en el desarrollo del presente proyecto debido a su experiencia y conocimiento en el ámbito educativo y tecnológico abarcado en el mismo.

En ese sentido, la SCA se ocupa de diversas actividades relacionadas a la ciencia. Desde su fundación, en el año 1872 se responsabiliza de la ininterrumpida comunicación y promoción de ciencia e industrias relacionadas. Eso se lleva a cabo con distintos tipos de experiencias como eventos, proyectos de investigación, publicaciones, congresos, etc. (López De Luise & Ruiz Tabarez, 2020b).

Actualmente, con sus 150 años, “es una de las entidades locales más tradicionales, permanentes y altamente representativas del interés por la ciencia vernácula e internacional” (Sociedad Científica Argentina [SCA], 2021).

La SCA brinda un aporte sustancial a la presente obra basado en su experiencia relacionada al desarrollo del ámbito científico tecnológico. Desde sus albores, promueve una

política que vincule estos ámbitos con un desarrollo social de forma prioritaria. Asimismo, aseveran que una relación normada y clara con el sistema productivo favorece al desarrollo propio del país (SCA, 2021).

Por su parte, la IEEE “es la asociación profesional más grande del mundo dedicada al avance de la innovación tecnológica y excelencia en beneficio de la humanidad” (Institute of Electrical and Electronics Engineers [IEEE], s.f.).

A través de su sección IEEE Games Technical Committe de Argentina, fomenta la “creatividad, el desarrollo y la integración de diferentes ciencias, a compartir y aplicar los avances en las tecnologías de la información, de consumo, electrónica y ciencias en general para beneficio de la humanidad y de los mismos profesionales” (IEEE, 2022).

Como su principal objetivo lo dicta, “el IEEE brinda especial atención a los estudiantes y estimula sus actividades con el propósito de perfeccionar su formación profesional” (IEEE, s.f.).

Con respecto al programa Lincievis Lab (Laboratorio de Investigaciones Científicas en Video Juegos y STEAM), el mismo se constituye como un laboratorio centrado en tópicos referidos a las ciencias y los videojuegos (Virvou y otros, 2022).

El primer trabajo que se considera importante analizar es el de (López De Luise & Ruiz Tabarez, Evaluación STEAM a partir de métricas específicas, 2020a). En el mismo, se definieron tres formularios para protocolizar las actividades a realizar: El primero con información de las entidades colaboradoras, el segundo para recolectar la información de cada actividad luego de ser implementada, y un último formulario con información referida al alumno para su perfilamiento. En este último caso, algunas variables tenidas en cuenta son la edad, grado educativo, familia, comportamiento, resultados de actividades, etc. Esto brinda datos importantes principalmente sobre todo el contexto del proceso.

En ese aspecto, es importante destacar la heterogeneidad de niveles educativos, zonas geográficas, culturas y economías para cada actividad realizada. Lo que permite una evaluación más integral del modelo abarcando distintas realidades de los estudiantes. Esta información también permite adaptar las actividades, por ejemplo, dependiendo del nivel educativo de los alumnos.

Se trabajaron con ejercicios sobre temas como células, herramientas de computadora, diseño de logos, robótica, narración, etc. Los mismos tuvieron una duración de entre 2 y 4 hs.

y no solo incluían ejercicios con computadoras, sino también, actividades manuales o artísticas. La dinámica de cada tarea fue adaptada de acuerdo a las competencias afectadas y al contexto geográfico de cada grupo de alumnos.

Una vez finalizadas las actividades, los autores capturan la información resultante de las mismas y la incorporan a una base de datos. Adicionalmente, con las variables incluidas en los formularios se formularon métricas iniciales para su posterior evaluación en términos de eficacia y eficiencia del proceso, para lograr generar un modelo de carácter sistemático para dicho análisis.

A partir de los datos recabados, el próximo paso fue aplicar Machine Learning (Aprendizaje Automático) para obtener un modelo basado en variables que permiten medir a partir de los datos mencionados la influencia y resultado de las actividades STEAM en los alumnos, evaluar el rendimiento del alumno e identificar en algunos casos posibles deserciones.

Como resultado, se evidenció que no siempre se cuenta con la información necesaria para las variables claves, por lo que se generó un nuevo modelo que permite inferirlas a partir de las existentes, aunque su calidad de predicción disminuye en cuanto a las originales. Estos ajustes necesarios son parte de las evaluaciones y resultados parciales que se obtienen en el proceso educativo, por lo que todas forman parte de una evolución lógica del modelo (López De Luise & Ruiz Tabarez, 2020a).

En este primer trabajo destacado se puede observar una clara tendencia al análisis de los resultados de las actividades STEAM a través de distintas herramientas tecnológicas vigentes, principalmente la denominada Machine Learning.

Como siguiente instancia, se analiza el trabajo realizado por (Dima y otros, 2012).

Esta investigación, denominada “Búsqueda de la calidad del aprendizaje mediante herramientas de inteligencia artificial” utiliza como herramientas la inteligencia artificial y la minería de datos. Lo que pretende detectar son posibles aspectos que influyan en el aprendizaje matemático de los estudiantes, y que no estén directamente relacionados con lo académico.

En el trabajo, los investigadores explican que la inteligencia artificial es seleccionada como tecnología ya que resulta un camino interesante de apoyo para complementar las diferentes formas aprendizajes con el uso de computadoras (Dima y otros, 2012).

Su objetivo principal no es solo la detección y diagnóstico de diferentes elementos que dificulten el aprendizaje, sino también, generar una retroalimentación que favorezca su desarrollo y mejora. Para el proceso se utilizó un sistema desarrollado en el lenguaje de programación JAVA que toma como entrada la información de los estudiantes, sus datos socioeconómicos, sus respuestas, y la definición con las respuestas correctas para cada ejercicio (Dima y otros, 2012).

Toda esta información es procesada por el módulo de evaluación de datos del sistema. Para cada pregunta se puede considerar una respuesta correcta, varias respuestas correctas de forma en que no importa el orden, o varias respuestas correctas donde sí importa el orden de selección. Los autores comentan que el resultado de esta evaluación es procesado por el módulo de minería de datos o módulo inteligente (construido a partir de librerías del software Weka). El mismo clasifica los datos provenientes del módulo de evaluación de acuerdo a características similares de los alumnos y a su vez los distingue de otros grupos heterogéneos. Por ejemplo, teniendo en cuenta edades, sexo, tipo de problemas a resolver, etc. De aquí se obtienen las variables más influyentes en el resultado final.

Por último, los datos son tratados por el sistema experto y se genera un reporte, tanto individual como grupal, donde se exponen las recomendaciones en cada caso. En este paso se evalúan los datos, se generan reglas y se devuelve un informe con las recomendaciones de mejora para los estudiantes (Dima y otros, 2012).

Los autores destacan como primeros resultados el bajo rendimiento generalizado de los estudiantes, más específicamente en conceptos de modelizado.

De esta forma, la investigación muestra su aporte en cuanto a la evaluación de rendimiento académico de los estudiantes, e incluye resultados específicos sobre cuáles son las áreas que presentaron mayores dificultades.

A continuación, resulta pertinente analizar el trabajo de Chango Sailema (2022).

El mismo refiere a la predicción del rendimiento académico de estudiantes utilizando datos provenientes de diferentes fuentes. Para lo cual, se emplean técnicas de minería de datos y de fusión de datos. La fusión de datos se entiende como la combinación de datos de diferentes fuentes de forma que se pueda obtener un

conocimiento mayor que la suma de conocimientos que aportan individualmente.  
(Chango Sailema, 2022, págs. 21-22)

El estudio aborda distintas modalidades de enseñanza, como son la presencial, virtual y mixta. También hay que destacar, que en estos casos se hizo foco mayormente en el alumno, dejando para trabajos futuros un abordaje más integral con la inclusión del docente. En el estudio de Chango, se observa que, en cuanto al tipo de datos trabajado, los mismos varían. Asimismo, se destacan principalmente los cuantitativos de rendimiento académico y grabaciones realizadas a los alumnos. No siendo parte del caso, por ejemplo, los datos sociales de los participantes. Éste último punto resulta importante para el presente trabajo, ya que el mismo abarca también el análisis de casos donde las condiciones sociales, económicas y/o tecnológicas no son las favorables para el desarrollo de proyectos científico tecnológicos.

Como destaca el autor, se evidencia que en investigaciones de esa índole, la técnica de fusión de datos más empleada es la denominada *early fusion* o fusión temprana. La misma consiste en la integración de datos de diferentes fuentes, trabajándolos de forma independiente al resto. Los mismos se filtran y se obtienen características propias de cada uno. Esto resulta útil “ya que a este nivel los datos presentan la mayor riqueza en cuanto información estadística sobre el evento a detectar” (Soriano Tolosa, 2013, pág. 123).

A pesar de la mayor utilización de la técnica de fusión temprana, el trabajo arroja resultados más precisos cuando los datos son tratados con la técnica *late fusión* o fusión tardía. La misma, consiste en realizar una primera clasificación de los datos, para luego fusionar la predicción de cada grupo (Chango y otros, 2022).

Con respecto a los algoritmos de clasificación de minería de datos, el autor asevera que se no logra distinguir, en el contexto de su investigación, cuál se destaca por ofrecer mejores resultados. Por tal motivo, no se puede aseverar cuál será más conveniente utilizar en la evaluación del rendimiento académico.

En cuanto a los diferentes modelos de predicción utilizados en su investigación, el autor refiere que algunas de las fuentes más importantes de información fueron: La interacción de los alumnos en la plataforma web educativa utilizada (Moodle) y el seguimiento ocular a través de un reconocimiento facial automático. El hallazgo de la importancia de la interacción de los alumnos con la web resulta especialmente notable ya que el presente trabajo también comprende una evaluación de datos provenientes de la plataforma virtual propuesta.

Para finalizar, y en cuánto a las modalidades de enseñanza futuras, el autor concluye que urgen “entornos de aprendizaje híbridos que puedan responder a circunstancias muy cambiantes” (Chango Sailema, 2022, pág. 13). Este tipo de modalidad de enseñanza mixta, es precisamente la que se desprende de la implementación de actividades que propone el Museo Virtual STEAM.

Con estos trabajos destacados, se pueden representar muchos otros que desde la concepción misma del modelo STEAM se realizan integrando la tecnología como principal medio de apoyo, particularmente enfocadas a las actividades propuestas y al aprendizaje obtenido.

Hasta aquí se observa cómo nace y va creciendo la educación STEAM, sus beneficios como modelo educativo, y algunos trabajos relacionados que brindan su aporte a partir de algunas herramientas tecnológicas. Pero bien, aquí se debe tener en cuenta algunos aspectos tangenciales a este tópico y preguntarnos, ¿qué ocurre con los edificios o instituciones que históricamente se dedicaban a brindar conocimiento? ¿qué rol cumplen a partir del desarrollo de este nuevo modelo? ¿se pudieron adaptar o desaparecieron por falta de flexibilidad?

Para responder estas preguntas, se verá cómo fue la evolución de estos organismos y su relación inevitable con la educación STEAM.

## **2.2 Museos Físicos y Virtuales**

### *Los Museos y su Transformación*

Desde su origen, los museos se presentan como portadores del conocimiento y de la responsabilidad de su transmisión.

Así como como sucedió oportunamente con las bibliotecas, las personas se han acercado a lo largo de la historia a estos lugares en busca de respuestas, de saber y de inspiración. Por eso, los museos siempre formaron parte del vasto mundo de la ciencia, desde su carácter prácticamente privativo para unos pocos elegidos, hasta su apertura al público en general que se encuentre interesado.

En consecuencia, se puede afirmar que un nuevo tipo de ciencia nació a partir de los museos, cerca del año 1800 (Winsor y otros, 2008), y de allí su importancia hasta la actualidad.

Si se analiza al museo como un sistema compuesto de partes que interactúan entre sí y con su exterior, podemos destacar una característica que resulta interesante para el presente trabajo, la física.

La parte física de un edificio representa el vocabulario que se ve. Por eso se debe brindar especial atención a que el mismo pueda mostrar las colecciones basadas en una estética que acompañe al usuario en ese lugar (Forgan, 2005). Toda esa estética es pensada en detalle y seleccionada por razones específicas, ya que conformaba su interacción principal con las personas.

Esta concepción va cambiando con el devenir de nuevas tecnologías, como internet. Desde la red tecnológica más grande del mundo, el visitante, devenido en usuario, ahora puede acceder a mucha de la información requerida sin moverse ni un solo paso desde prácticamente cualquier lugar del mundo.

Estos cambios en los paradigmas de las fuentes de información, su acceso y su transmisión llevaron a los museos a mutar, pero no solo para adaptarse, sino para seguir cumpliendo su rol primordial, hacer y transmitir la ciencia. En este punto es donde muchos educadores se preguntan cómo fomentar la educación STEAM, incluso muchos ven difícil el entender por dónde comenzar el proceso (Kemble, 2018).

De esta forma, en medio de cambios e interrogantes, surgen los museos virtuales. Los mismos se erigen con los mismos principios y objetivos que sus pares físicos, pero con elementos y formas de interacción diferentes para con su principal interlocutor, ni más ni menos que el usuario. En un museo virtual, la arquitectura y su expresión abandonan la colección, de esta manera es el diseñador del sitio web el encargado de elegir qué será replicado al mundo virtual y cómo. Como menciona López De Luise (2020), de esa manera toman gran importancia la navegación del sitio, los tamaños de elementos, formas, colores, etc., ya que todo esto influye en la experiencia del cibernauta o usuario de la web. Todo el diseño de estos elementos se vuelve relevante, así como la propia interacción entre ellos.

Si nos enfocamos en dicha interacción, la autora refiere a que en un museo virtual la tecnología capta la reacción de los visitantes a través de diferentes elementos o eventos. Los mismos pueden ser: información de formularios, interacciones con las actividades presentadas virtualmente, solicitudes a la entidad organizadora y, por ejemplo, distintas actividades generadas en redes sociales. Toda esta información del movimiento del usuario en el sistema

resulta muy útil al momento de ser analizada con estadísticas y métricas con el afán de mejorar su propio funcionamiento.

Se ve, a partir de lo expuesto, como los museos comienzan su mutación, de museos físicos a virtuales, de privados a públicos, de un modelo educativo tradicional a nuevas líneas de educación acorde a los tiempos actuales.

A partir de aquí, se fusionan los conceptos del museo como el portador y transmisor de conocimientos, y la educación STEAM.

### *Museo Virtual STEAM*

Antes de la pandemia, el MHS realizaba sus actividades de forma presencial, pero una vez producido el aislamiento preventivo para sortear el impacto del COVID 19, la entidad se vio obligada a ampliar su horizonte hacia el mundo virtual. Esta transformación abarcó a toda área y entidad promotora de conocimientos, incluso fomentada por la UNESCO, quien recomendó la utilización del aprendizaje no presencial (UNESCO, 2020).

Los cambios tecnológicos y políticos sumados a la variación cultural ocasionada por la pandemia generan cambios en las formas en que los museos interactúan con sus comunidades.

Para adaptarse de forma más flexible a esta nueva forma de comunicar, y teniendo en cuenta la tendencia virtualizadora en la transmisión del conocimiento en la sociedad, la SCA y el MHS impulsan el desarrollo del Museo Virtual STEAM. El mismo tiene como misión el desarrollo y difusión de la educación STEAM, ya sea con o sin soporte tecnológico:

Esta suerte de museo virtual tiene como objetivo brindar a los estudiantes la posibilidad de conocer y entender las ciencias. El análisis de la información emergente del mismo ante cada interacción permite la creación de un modelo conceptual para conocer y entender a los estudiantes, logrando una simbiosis y retroalimentación constante y evolutiva. (Leiva y otros, 2022a, pág. 5)

Esta transformación resultó muy positiva ya que posibilitó difundir la comunicación a más comunidades, ampliar y modernizar el contenido transmitido.

Para comenzar, se definió la estética que reflejará la identidad del mismo y el ciclo de vida de cada módulo: definición de responsables, facilitadores, recursos necesarios para su

implementación, target objetivo, lugar y momento de realización, entre otros (López De Luise, 2020).

El Museo Virtual STEAM fue desarrollado bajo una plataforma Wordpress. La misma permite la creación, edición y publicación de sitios web de forma fácil y rápida (Wordpress, 2022).

El sitio del museo posee información relacionada a sus objetivos, las distintas experiencias STEAM desarrolladas (imágenes, descripciones, videos, etc.), calendario de próximas actividades (presenciales o virtuales), blog, medios de contacto, etc.

Adicionalmente, permite buscar actividades, proponerlas o suscribirse para ser contactado por el área correspondiente para poder llevarlas a cabo. Las mismas, luego sirven a la plataforma de sus resultados, obtenidos por medios físicos o virtuales.

En cuanto al museo virtual, resulta importante destacar el concepto de cápsula. El mismo hace referencia al conjunto de elementos intelectuales y técnicos necesarios para llevar adelante determinada actividad. Esto incluye, por ejemplo: materiales a utilizar, información necesaria para su aplicación, fotos, videos, capacitación, paso a paso de actividades, consultorías, objetivos, formularios, eventos, entidades participantes, etc. (López De Luise, 2020). Es decir, todos los elementos necesarios para poder alcanzar los objetivos específicos de cada módulo.

Una cápsula transcurre el ciclo de vida de explotación, en el que se llevan a cabo principalmente la definición de los elementos mencionados, determinación de entidades participantes, adaptación al módulo correspondiente e implementación, entre otras. La plataforma está compuesta por un conjunto de cápsulas que pueden adaptarse a distintas comunidades educativas, cada una con sus objetivos específicos.

Cada una de ellas educa y transmite procedimientos pedagógicos del modelo STEAM, y por lo tanto “pueden representar actividades en el museo, en una institución educativa, o en el mundo virtual” (Leiva y otros, 2022a, pág. 6).

Estas cápsulas transcurrirán un proceso de análisis de datos a través del análisis estadístico y de minería de datos propuesto en este trabajo. A continuación, a partir de la información resultante, se derivarán propuestas de mejora a los procesos involucrados. Estas actualizaciones permitirán una retroalimentación al sistema de cápsulas, lo que

generará una evolución en el modelo educativo en cada iteración realizada. (Leiva y otros, 2022b)

### **2.3 La Tecnología Como Apoyo Para el Análisis de Datos**

#### *El Análisis Tecnológico Como Base Para la Evolución*

A partir de los cambios en las vías de enseñanza de las ciencias mencionados, la transformación del aspecto físico al virtual de los museos y el devenir tecnológico generalizado en la sociedad surgió el Museo Virtual STEAM. Un espacio virtual que funciona como disparador de muchas otras actividades que acercan al público en general a las ciencias.

Todo este espacio genera un gran volumen de información, la cual puede ser procesada para obtener valiosas conclusiones que sirvan para la mejora del modelo. Aquí es donde surge el proyecto de análisis y evaluación mediante herramientas tecnológicas a toda información resultante de los mecanismos mencionados.

En la actualidad, esta información proviene tanto de fuentes o actividades presenciales como virtuales. Todas ellas brindan la posibilidad de análisis, y esa explotación combinada desde distintos orígenes se enmarca en lo que se conoce como Multimodal Learning Analytics (MMLA). Como su nombre lo indica, esta orientación tecnológica tiene sus bases en la recopilación y combinación desde datos de diferentes fuentes educativas, para lograr un entendimiento integral del aprendizaje (Sharma & Giannakos, 2020).

Este trabajo contempla herramientas basadas en el análisis estadístico y la aplicación de minería de datos a toda la información concerniente a la interacción del usuario con el Museo Virtual STEAM y la surgida de la implementación de sus cápsulas. (Leiva y otros, 2022a, pág. 6)

Esto permite la corrección, mejora y evolución del proceso, del accionar docente, del aprendizaje de los estudiantes y de la interacción entre todas estas variables. La utilización de estas herramientas del orden de la inteligencia computacional, aseguran además su objetividad y permiten alcanzar un estándar de procesos STEAM que faciliten su control a través de métricas y procedimientos establecidos (López De Luise & Ruiz Tabarez, 2020a).

En cuanto a su lado metodológico, Leiva y otros (2022) distinguen que, a través de la recopilación y análisis de datos referidos a la interacción de los usuarios con el sistema y los resultados de sus actividades, se pueden entender aún más las necesidades de los estudiantes, de los docentes y de las distintas cápsulas, mejorando su interacción integral y la comunicación de conocimientos. Incluso, se destaca que “al ser de carácter sistemático y reproducible, brinda una base sólida para la reproducción en otras instituciones y la evaluación de los procesos involucrados” (pág. 6).

Así es como las actividades STEAM, su promulgación y articulación a través de plataformas virtuales, surgen como una herramienta importante y novedosa que se suma a las ya existentes para la continua mejora y desarrollo de cada etapa de la educación basada en STEAM.

En definitiva, como menciona Ortega y Gasset en “Meditación de la técnica y otros ensayos sobre ciencia y filosofía”, la técnica hace más humano al medio para hacerlo más habitable para las personas. Por tal motivo, una mejora en la técnica repercute en una mejora de la humanidad (2000).

Esto, además, se complementa perfectamente con el concepto de que la sociedad puede mejorar la cultura aplicando la ciencia y, en consecuencia, también la economía (Neressini & Pellegrini, 2008). Por tal motivo, la intervención científico tecnológica a la información resultante de todas estas actividades, procesos y sus medios de aplicación resultan tomar tanta relevancia.

A continuación, se describe en primer lugar a qué se refiere cuando se habla de información. Y luego, qué herramientas tecnológicas de las mencionadas se aplicarán directa o indirectamente en el presente trabajo para tratarla.

### *La Información*

La información puede significar alguna intervención útil y significativa en cualquier proceso. Por ello, “los datos son un recurso valioso. Puede conducir a nuevos conocimientos y, en entornos comerciales, a las ventajas competitivas” (Witten & Frank, 2005).

Este concepto de información sencillo, también se puede analizar un poco más en detalle si se estudia la conceptualización propuesta por (Peirce, 1931-1935), donde refiere a la información como la integración de 3 partes en igual consideración: un mensaje, un sistema de referencia y un receptor.

Desde ese punto de vista se puede determinar que:

La información, si de veras es tal, va más allá del mensaje, a través de un receptor, hasta un sistema de referencia. Dicho de otro modo, la información no está ubicada en el mensaje, ni en receptor, ni en el sistema de referencia, sino en la relación triádica ente ellos. (Alfredo, 2020)

Entonces, se ve como el dato no se vale por sí mismo, y la información adquiere valor recién cuando se genera un efecto en el receptor.

A raíz de lo expuesto, surge la necesidad del análisis de los datos con los que cuenta la plataforma STEAM y sus actividades, para lograr obtener información que permita ser contextualizada y genere una evolución en el sistema de educación mencionado, en la enseñanza del mismo y, por supuesto, en el aprendizaje de cada alumno o docente receptor.

Llegado este punto, se debe considerar con qué herramientas se puede realizar un análisis exhaustivo de los datos, qué tecnologías existen actualmente para estudiarlos y cómo se articularán las mismas para obtener el mayor beneficio en sus resultados.

A continuación, se conceptualizan los elementos más importantes a tener en cuenta para la comprensión de los distintos procesos involucrados que forman parte de la presente obra.

### *Métricas de Software*

Como se ha expuesto, la información es un capital muy valioso en toda organización. Se puede entender cómo logra ser la fuente de conocimientos que puede marcar la diferencia en la toma de decisiones, marcar rumbos de acción correctos o equivocados, o servir de apoyo en la evaluación de procesos.

De la misma forma, en la disciplina referida como ingeniería de software, los análisis de esos datos resultan de gran importancia para la evaluación de cualquier sistema. Entonces, la primera pregunta que se debe responder es ¿Cómo medir el rendimiento de un software? Para ello se utilizan, por ejemplo, las métricas.

Cada una de ellas, independientemente del objeto de estudio, se puede definir como “un indicador cuantitativo o punto de referencia de un atributo, que es una característica o propiedad de una entidad” (Imoize y otros, 2019, pág. 57).

Para el caso que compete a esta obra, se enfocará la definición a lo que alude específicamente al desarrollo de software. En ese sentido, los autores explican que:

Las métricas de software se refieren a la medida o grado en que un componente, proceso, módulo o el sistema posee un atributo dado. Proporcionan medios para la medición cuantitativa y validación de los diferentes modelos utilizados en el ciclo de vida del desarrollo de software. (pág. 57)

Una vez definidas las métricas, toma importancia su utilización en cualquier equipo dedicado al desarrollo, ingeniería o tratamiento de software. En ese sentido, los redactores afirman:

La medición adecuada de las métricas del software es esencial para desarrollar modelos para el proceso de desarrollo de software. Permite establecer expectativas para los stakeholders, rastrear y medir el éxito/fracaso del proyecto. Las métricas y modelos desarrollados también son herramientas invaluableles utilizadas para estimar costos, medir la productividad y la calidad del producto. (pág. 57)

Según lo hasta aquí enunciado, los datos que forman parte de los distintos procesos juegan un rol fundamental en cualquier entidad. De igual forma, y en referencia al ámbito del software y las estadísticas, la posibilidad de medir y obtener información de esos datos también se vuelve indispensable para la evolución de cualquier sistema.

A continuación, se exponen las diferentes herramientas computacionales que darán soporte al análisis, medición y evaluación de la Plataforma Virtual STEAM.

### *Inteligencia Artificial (IA)*

El concepto de inteligencia artificial tiene su origen en el año 1956. El mismo fue propuesto por John McCarthy durante la conferencia ofrecida en Dartmouth, Estados Unidos.

Como menciona Saint-Pierre (2018) en el artículo web publicado por Fundación Chile, la inteligencia artificial se logra programando computadoras para que simulen los procesos cognoscitivos de las personas. Entre los mismos se destacan la capacidad de aprendizaje o resolución de conflictos. También alude que no es un desarrollo que recién comienza a

explorarse en términos de tiempo, ya que desde la década de 1970 que se comenzaron a construir los primeros sistemas expertos con base en la IA.

Desde sus comienzos, la IA ya deslumbraba por su reconocido potencial. Tal es así que una alusión conocida desde aquellos años es la de (Simon):

La manera más simple de decirlo es que en el mundo hay máquinas que piensan, crean y aprenden. Incluso, esas capacidades crecerán tan velozmente hasta un futuro que se puede predecir. En el mismo, las máquinas tendrán la capacidad de solucionar problemas a la par que la capacidad de una mente humana. (1957)

Actualmente, este tipo de programación se da paso entre las actividades cotidianas en gran medida, y muchas veces sin darnos cuenta o saber que estamos tratando con ella.

Algunos ejemplos que denotan dicha influencia según Saint-Pierre, pueden ser: El reconocimiento facial que utilizan algunos softwares de computadora o celular, el reconocimiento de voz que nos permite solicitar la ejecución de una acción en algún dispositivo o, en una escala mayor, la conducción de autos en forma autónoma (2018).

La inteligencia artificial también es desarrollada para desempeñarse en otros ámbitos no académicos o científicos. Por ejemplo, los juegos. En ese sentido, se puede destacar que el caso Deep Blue de IBM. Este fue el primer sistema de ajedrez en vencer a un campeón mundial, Garry Kasparov. El resultado fue 3.5 a 2.5 en un juego de exhibición (Goodman & Keene, 1997).

Por otro lado, desde una mirada más relacionada al ámbito del presente documento, (Russel & Norvig, 2004) aseveran que “la IA ya forma parte del ámbito de los métodos científicos” (pág. 58). Dicha afirmación también es apoyada por los autores (Cohen & Page). Ellos complementan que para que las hipótesis sean aceptadas, deben ser expuestas a exhaustivos estudios empíricos. A partir de allí, sus resultados deben ser analizados estadísticamente para determinar su importancia (1995).

Por lo visto hasta aquí, se observa como las grandes compañías reconocen el enorme potencial de la IA para la gestión de información que mejore e innove sus procesos y resultados.

A partir de la implementación de la inteligencia artificial, surge la importancia de lograr que “la máquina pueda ir aprendiendo a través de la experiencia, a esto se le llama aprendizaje

automático que puede ser supervisado (algoritmos de clasificación) o no supervisado (algoritmos de clusterización)” (Saint-Pierre, 2018). Estos conceptos se analizan más en detalle en una sección siguiente.

Ahora que se ha conceptualizado la IA, se verá como una de sus implementaciones puede favorecer al estudio de la información de todo el proceso del modelo educativo tratado en la presente obra. En este caso, se habla de la minería de datos.

### *Minería de Datos (Data Mining)*

Para situar en contexto a la minería de datos, resulta oportuna la de definición de (González) en cuanto a su relación con el denominado proceso KDD (acrónimo del inglés Knowledge Discovery in Databases):

El proceso KDD, es nuestra propuesta para la preparación de información antes de llegar a una técnica de minería que se ajuste a nuestros requisitos. Con este proceso, el analista de datos podrá generar información limpia y centrarse en las variables necesarias para aplicar la técnica de minería más ajustada a la situación, para llevar esos resultados a un entorno real de producción, además de todo esto nos ayuda a estandarizar el trabajo, con pasos ya definidos. (2021)

En este marco, para lograr una sencilla pero efectiva noción de lo que es el data mining, se procede a comentar una analogía frecuente, en esta oportunidad expuesta por Redacción España:

Se sabe que los mineros, en la vida real, se sumergen en las minas y trabajan hasta poder encontrar los minerales buscados. Paralelizando, se puede entender que en el caso del data mining, las minas (como fuente que otorga una materia prima) son las bases de datos de un sistema. En otras palabras, toda aquella herramienta que almacene datos sobre un proceso u operación. (2020)

El equipo de redacción complementa que, en lugar de los mineros, como encargados de buscar el o los elementos deseados, se encuentran los algoritmos de software basados en IA. Los mismos, serán ejecutados en busca de información relevante en esa mina de datos con la

que cuenta el sistema. Y por último, en cuanto a los minerales, entendidos como el objetivo o fruto de esa búsqueda, se encuentran los patrones hallados y sus relaciones.

Por otro lado, es importante tener en cuenta que la sociedad actual atraviesa la denominada era de la información o era digital. La misma “hace referencia a un período de tiempo en el que el movimiento de la información se volvió más veloz que el movimiento físico, debido al desarrollo de las nuevas tecnologías digitales y la comunicación, las TICs” (Epitech, 2022).

Todo este movimiento genera información que puede resultar muy valiosa para quien sepa gestionarla adecuadamente. La misma puede, por ejemplo, colaborar en la reducción de costos operativos, minimizar la posibilidad de compras innecesarias, etc. (González, 2021). A partir de allí, se logra entender por qué cada vez son más los canales por los que las empresas buscan hacerse de datos relativos a sus procesos. Toda esta información es almacenada y representa una riqueza proveniente de cada etapa en la que se vieron involucrados sus procedimientos. Desde la adquisición de elementos de trabajo, trato con proveedores, relación con clientes, hasta la venta de productos o servicios generan información valiosa para cualquier organización.

Los algoritmos de software mencionados buscan clasificar y segmentar estos datos, y a través de los patrones “se pueden elaborar hipótesis basadas en modelos probabilísticos que sirven para la toma de decisiones en diferentes campos, desde el empresarial hasta el de la investigación o la salud, por ejemplo” (Redacción España, 2020). Concluyendo, el data mining se encarga del análisis y explotación de datos para transformarlos en información útil que nutra y acompañe la toma de decisiones de cualquier organismo o entidad.

Luego de los conceptos esbozados, para ejemplificar un posible uso del data mining, (Witten & Frank, 2005) proponen:

Patrones del comportamiento de antiguos clientes pueden analizarse para identificar características distintivas de aquellos que probablemente cambien de producto y aquellos que probablemente permanezcan leales. Una vez se encuentran tales características, se pueden poner a trabajar para identificar a los clientes actuales que es probable que abandonen el barco. Este grupo puede ser objeto de un tratamiento especial, tratamiento demasiado costoso para aplicar a la base de clientes en su conjunto. (pág. 5)

En el caso del presente trabajo, esta información referirá a la forma de mejorar el modelo educativo en cuestión, a través de su implementación por medio de la Plataforma Virtual STEAM.

### *Aprendizaje Automático (Machine Learning)*

El aprendizaje automático es una de las aristas más importantes en los procesos mencionados. Dicho concepto es la raíz del potencial que alcanza la minería de datos.

Cada algoritmo se interna en un mar de datos y obtiene resultados. Luego, dichos algoritmos se modifican a sí mismos e incorporan lo obtenido, logrando perfeccionarse antes de realizar una nueva iteración en busca de resultados (Redacción España, 2020).

Por ello, el equipo de Redacción España también afirma que:

De esta forma, la minería de datos es cada vez más sofisticada. Las decisiones extraídas del procesamiento de la información a través de la búsqueda de patrones incrementan su nivel de precisión a medida que avanza el tiempo y el número de datos procesados. Así, la minería de datos no solo sirve para tomar mejores decisiones, sino que también hace mejor a la propia minería. (2020)

Como se comentó en una sección anterior, según sus características, el machine learning puede clasificarse como supervisado o no supervisado (Saint-Pierre, 2018).

El aprendizaje supervisado consiste en que el programador enseñe al algoritmo los datos de entrada ya etiquetados y los resultados esperados. Dicho de otra forma, la salida del proceso ya se conoce (TIBCO, 2022).

Por ejemplo, se puede brindar al algoritmo un conjunto de imágenes etiquetadas como plantas acompañándolas de algunas características propias de ellas. Luego, enseñarle que todas ellas representan plantas. De esta forma, una vez aprendida la lección, el algoritmo ya será capaz de reconocer por sí mismo si en una imagen aparece o no una planta.

En cuanto al aprendizaje no supervisado, en el mismo no se conocen los resultados ni existen datos de entrenamiento para tomar como referencia. En este caso, el algoritmo solo dispone del conjunto de datos de entrada y una guía de operaciones lógicas para operar sobre él (EDS Robotics, 2021).

Continuando con el ejemplo anterior, el algoritmo deberá ser capaz de reconocer formas similares entre los elementos de las imágenes y agrupar, por ejemplo, a las plantas. Esto lo realizará sin saber ciertamente que son plantas, pero sí detectando similitudes en sus características. Cada uno de esos grupos son etiquetados y tomados luego como referencia para futuros exámenes.

En este punto, según (Saint-Pierre) surge la necesidad de poder clasificar los datos teniendo en cuenta algunos patrones de referencia. A las observaciones y su clasificación se las denomina conjunto de aprendizaje (2018). Complementando, el autor explica que se debe tener en cuenta que estos métodos de análisis no siempre cuentan con información completa o certera, para ello también se han creado funciones basadas en probabilidades. Las mismas disminuyen la posibilidad de realizar un análisis incorrecto.

A partir de lo expuesto, se observan claramente los beneficios de la minería de datos y su aprendizaje automático al tratar con datos históricos de cualquier índole para obtener información antes oculta. Entonces, la pregunta que puede surgir es: ¿la minería de datos puede ser aplicada directamente al comportamiento de un usuario en un sistema? ¿puede el machine learning darnos más conocimientos sobre los internautas de una plataforma web?

A continuación, se describe una aplicación particular de la minería de datos. La misma es orientada exclusivamente al conocimiento del usuario y su desempeño en la utilización de un software web.

### *Perfiles de Usuario (User Profiling)*

En primer lugar, corresponde hacer referencia al concepto de web mining o minería web. El mismo puede entenderse como el proceso que permite “procesar grandes cantidades de información de sitios web” (Medrano & Barriento, 2021). Esto permite encontrar patrones potencialmente útiles y el examen de información oculta en los elementos o actividades de la web (Heider & Sánchez, 2008).

Dicho de otra forma, como menciona Cuevas Palma, es la minería de datos aplicada a todo componente o suceso ocurrido en una web, con el fin principal de encontrar patrones en su organización, contenido y utilización (2010).

Formando parte del web mining se encuentra el user profiling, también denominado en algunas oportunidades como user mining o minería de usuarios. Con respecto a sus objetivos específicos se pueden mencionar la “búsqueda de información relevante o relacionada,

creación de nueva información a partir de la existente, personalización de la información y generar conocimientos a partir de los comportamientos de los usuarios web” (Cuevas Palma, 2010, pág. 26).

Según describe el autor, el proceso de aplicación de web mining se puede dividir en la serie de 4 etapas consecutivas que se detallan a continuación.

En primer lugar, la selección de datos: Para comenzar se define qué datos serán útiles para nuestro objeto de estudio y se lleva a cabo la recopilación de los mismos.

En una segunda etapa, se realiza el tratamiento previo de datos: En este momento se limpian, ordenan y organizan los datos obtenidos. También se eliminan los datos que no estén completos o que contengan errores hasta lograr su homogeneización.

En tercer lugar, en su etapa más significativa de desarrollo, se ejecutan los algoritmos en busca de patrones de comportamiento y correlaciones entre ellos. Al finalizar esta etapa, los datos se transforman para lograr obtener información de ellos.

En la cuarta y última etapa, comenta el autor, toma protagonismo el humano, ya que será el responsable, mediante el análisis apoyado en distintas herramientas tecnológicas, de poder entender e interpretar los patrones que brindarán la información deseada.

Hasta aquí se ha contextualizado y explicado los principales procesos que conforman la minería de datos. En ese sentido, restaría mencionar una última cara que la conforma y concierne puntualmente a este proyecto. Se trata a continuación la relación del data mining con la estadística.

### *Estadística y Minería de Datos*

Antes de hacer referencia a la estrecha relación entre la estadística y la minería de datos, se definirán algunos conceptos para su completo entendimiento.

Ya en una de sus concepciones más antiguas, Sir R. Fisher describía a la estadística como la integración de tres conceptos: El estudio de poblaciones, el estudio de la variabilidad de los fenómenos, y el estudio de los métodos de reducción de la información comprendida en los datos (1925).

Su clasificación puede distinguirse en estadística descriptiva o inferencial. En relación a la estadística descriptiva, la misma “muestra las características que acompañan a un objeto o acontecimiento en estudio” (Hurtado & Gómez, 2022, pág. 5). Trata específicamente de la

recopilación y descripción numérica de los datos. Los mismos son resumidos para facilitar su interpretación (Vargas Sabadías, 1995).

Acerca de la estadística inferencial, (Hogg y otros) aseveran que la misma busca establecer inferencias en base al análisis de una muestra, realizando interpretaciones y pronosis (2021).

Se puede afirmar que la primera de ellas se ocupa de “ofrecer soluciones algorítmicas con un coste computacional aceptable, mientras que la estadística se ha preocupado más del poder de generalización de los resultados obtenidos, esto es, poder inferir los resultados a situaciones más generales que la estudiada” (Aluja, 2001, pág. 481).

En cuanto a su relación con la minería de datos, según (Witten & Frank), en realidad no se debe buscar marcar una frontera clara entre la estadística y el aprendizaje automático, porque entre ambas representan una continuidad de técnicas de análisis de datos (2005). Los autores también refieren que, por ejemplo, la minería de datos utiliza una buena cantidad de métodos y pruebas estadísticas. En este último caso, para evaluar y validar modelos de machine learning.

En cuanto a establecer una diferencia importante entre ambos conceptos, aunque algo simplista, ellos aseveran que las estadísticas pueden verse más enfocadas en la prueba de una hipótesis, mientras que el aprendizaje automático busca tratar al proceso de generalización como una búsqueda por medio de hipótesis tentativas.

De esta forma, se plantean las principales herramientas tecnológicas y estadísticas que forman parte de esta investigación. Cada una de ellas, destinada a dar soporte y respuestas a los problemas, cuestionamientos y conjeturas que se detallan en las siguientes secciones.

A continuación, se plantean los procedimientos llevados a cabo con la información y conceptos expuestos.

### **3. Desarrollo Técnico**

#### **3.1 Introducción**

En este capítulo se describen de forma pormenorizada todas las variables y procesos que forman parte de esta investigación. Para comenzar, se aborda todo elemento referido a la

metodología de investigación utilizada. Ello incluye a los datos, sus técnicas de recolección, los materiales utilizados y los métodos que se aplicaron en cada etapa.

Por último, se explicita el procesamiento de información realizado que da lugar a los resultados y su posterior análisis para la obtención de conclusiones.

### **3.2 Metodología de Investigación**

En primer lugar, cabe señalar la apreciación de (Hernández Sampieri y otros) en 2010. Ellos mencionan que una investigación es una serie de procedimientos sistemáticos, cruciales y experimentales que sufre un fenómeno al ser estudiado (pág. 4). Tal es el caso del presente documento, el cual puede catalogarse de las formas expuestas a continuación:

Se pueden realizar diferentes clasificaciones según los criterios que se deseen evaluar. El primero a destacar es el carácter cuantitativo de esta investigación. Como los autores indican, dicha clasificación se debe a estar formada por una serie de pasos sucesorios y que se pueden probar. Este orden es riguroso, por lo que no puede ser alterado.

Asimismo, los autores también argumentan que el enfoque cuantitativo se basa en una hipótesis e intenta probarla a través de la cosecha de datos, sustentándose en comprobaciones numéricas y el análisis estadístico. De esta forma obtiene patrones de comportamiento y logra probar sus teorías (pág. 4). Justamente estos procedimientos de recolección de datos y posteriores análisis a partir de procesos estadísticos son los utilizados en este trabajo.

Desde otro punto de vista, también se puede clasificar esta obra como una investigación de transferencia. Según (Secundo y otros, 2017) y (Calafell y otros, 2014), la investigación académica fomenta nuevos conocimientos y descubrimientos. Estos sirven como impulsores para el desarrollo de la sociedad, su economía y cultura, transfiriendo sus avances a las empresas. Estos avances transferidos a la sociedad también son destacados por (Cardozo-Acosta y otros, 2015) y (Compagnucci & Spigarelli, 2020).

Por otro lado, también corresponde añadir, a partir de los distintos tipos de investigación existentes, la denominación de aplicada. Esto se debe a su intención de resolver los problemas (Hernández Sampieri y otros, 2010) planteados en secciones anteriores. Para ello, se basa en la alguna teoría conocida del tema abordado, se elaboran hipótesis, establecen métricas y seleccionan ecuaciones a implementar. De estos procesos se obtienen resultados, los cuales se analizan de forma cualitativa en las secciones finales.

### 3.2.1 *Técnicas de Recolección de Datos*

En esta apartado, se detallan las principales fuentes de datos utilizadas. Entre ellas, se destacan la plataforma web junto a sus correspondientes logs de interacción, y la base de datos del sistema. De ambos componentes se describen sus principales características.

#### *La plataforma web y sus logs de interacción*

La primera fuente de información para este trabajo es la Plataforma Virtual STEAM. Como se mencionó con anterioridad, existe un prototipo funcional para la misma. Dicho prototipo (estudiado en profundidad en el apartado homónimo) fue actualizado de manera tal que permita el flujo de datos necesario para nutrir a esta investigación.

Una de las modificaciones más importantes realizadas fue la incorporación de un logueo de interacción de los usuarios con la plataforma. Para el caso, se utilizó un logueo de información con cada clic del usuario en pantalla. Al producirse dicho evento, se captura principalmente información de la fecha y hora del suceso, dirección IP<sup>2</sup> de la cual se accedió, ubicación geográfica del usuario y elemento en el que está realizando el clic.

Con respecto al componente seleccionado, se pueden identificar grupos como: links, imágenes, formularios, botones, íconos de acceso a redes sociales, entre otros.

De esta forma, todo el recorrido del internauta es registrado o trackeado por el sistema. Sus preferencias de navegación son almacenadas para nutrir los análisis de minería de datos que retroalimentarán al propio modelo.

#### *La base de datos*

Todas las implementaciones de actividades realizadas de forma presencial cuentan con un recabado de información post acontecimiento. Dicho procedimiento se realiza a través de encuestas utilizando tanto formularios físicos como digitales. Los mismos, previamente sufren un proceso de formalización y protocolización que les permite contar con la información completa y adecuada que funcione como datos entrada para las siguientes etapas.

Una vez obtenidos todos los datos necesarios, son integrados en una planilla de cálculo (también previamente normalizada) junto a los datos propios que el sitio web muestra al

---

<sup>2</sup> Las direcciones IP son utilizadas para reconocer a cualquier dispositivo que se encuentra conectado a una red IP (Cisco, 2021).

usuario. Como último paso, toda la información es cargada a la base de datos del sistema, dejándola preparada para su explotación.

### ***3.2.2 Datos Analizados***

En esta sección, se detallan los datos de entrada más importantes que serán utilizados para el análisis estadístico y de minería de datos. Los mismos pueden ser agrupados en tres cápsulas, pormenorizadas a continuación.

#### ***Cápsula 1: Programación y Robótica***

Esta cápsula fue implementada los días 10 y 11 de Diciembre de 2019 en el Colegio del Salvador (Jujuy). La misma estuvo a cargo de la UCSE DASS (acrónimo de Universidad Católica Santiago del Estero - Departamento Académico San Salvador).

Es estas ocasiones, la duración del evento fue de entre 90 y 105 minutos. Tiempo en el cual se desarrollaron 2 actividades por día. Las mismas se describen principalmente por las siguientes características.

La primera actividad, denominada “Introducción a la Programación”, promueve la generación de habilidades mentales al aprender programación sin poseer conocimientos previos. Para la misma, se utiliza Scratch<sup>3</sup>. Esta tecnología facilita el aprendizaje de los estudiantes ya que utiliza una programación guiada por eventos, agrupándolos en distintos segmentos.

En cuanto a las edades recomendadas para su aplicación, los organizadores definieron que la misma se desarrolle con alumnos a partir de los 10 años, ya que posee un nivel de dificultad intermedio.

En cuanto a la segunda actividad, llamada “Robótica y electrónica extracurricular” fue realizada en el mismo establecimiento y con el mismo grupo de alumnos que la primera, ya que cuenta con gran similitud en dificultad y en su edad recomendada de aplicación.

---

<sup>3</sup> Scratch es la comunidad de codificación para niños más grande del mundo. Asimismo, es un lenguaje de codificación que posee una interfaz visual simple, permitiendo a los más jóvenes la creación de historias, juegos y animaciones digitales (Scratch, 2019).

Dicha actividad, denominada “Robótica y Electrónica extracurricular”, tiene como principal objetivo introducir al adolescente a la robótica y electrónica. Para ello, se utiliza una plataforma CAD1 llamada Tinkercad®<sup>4</sup>.

### *Cápsula 2: Modelización de células*

Esta cápsula, denominada “Modelización de células”, fue coordinada por la UADER FCyT - Concepción del Uruguay. La misma, fue desarrollada el 15 de Mayo de 2019 y también contó con la realización de dos actividades.

La primera ellas, referida como “Modelización de la membrana plasmática”, es una propuesta pedagógica-didáctica que tiene por objetivo consolidar la enseñanza y el aprendizaje en distintos niveles de educación. Basándose en sus conocimientos previos y en lo transmitido durante la jornada, los alumnos construyeron distintos modelos didácticos. Los mismos simulan la estructura de la membrana plasmática de una célula. Esta actividad, fue recomendada para estudiantes a partir de los 15 años ya que cuenta con un nivel de dificultad medio-alto.

Por otro lado, la segunda actividad, denominada “Modelización de una célula”, tuvo lugar en el mismo establecimiento y se desarrolló con el mismo grupo de alumnos que la primera actividad. En este caso, los objetivos son muy similares, solo que el enfoque no está dado en la estructura de la membrana plasmática, sino en la célula en su totalidad.

### *Cápsula 3: Sarmiento Gamificado*

La siguiente cápsula contó con el desarrollo de una sola actividad. La misma, denominada “Juegos de Mesa” fue realizada el 04 de Septiembre de 2022 en el Museo Histórico Sarmiento (Ciudad Autónoma de Buenos Aires). Dicho evento estuvo organizado por la institución homónima y contó con una duración de 150 minutos.

En esta ocasión, su objetivo fue la difusión de conocimientos referidos a los sistemas inteligentes. Dicha difusión, se llevó a cabo a través de una propuesta que incluyó juegos de mesa, de rol y videojuegos, todos enmarcados en un proceso de gamificación. Este último concepto es importante ya que, según (Perotta y otros), la gamificación puede ser una

---

<sup>4</sup> Tinkercad es una aplicación del tipo web utilizada para el diseño 3d, codificación y electrónica. La misma es gratuita y utilizada por millones de personas alrededor del mundo (Autodesk Inc, 2022).

herramienta que permita el aprendizaje a través del disfrute, y esto llevar a la persona a una sensación de inmersión total (2013).

#### *Cápsula 4: Trivias de intercambio Chaco – Jujuy*

La última cápsula que se analiza comenzó a desarrollarse en Septiembre de 2021, y tuvo lugar durante una serie de diez encuentros semanales. En esta ocasión, participaron 31 alumnos del Colegio Secundario de Arte N°30 (Casira, Pcia. de Jujuy) y 30 alumnos de la Escuela de Educación Secundaria 129 "Roque Romero" del Paraje de Pampa Cejas, Pcia. de Chaco. En ambos casos, los estudiantes formaban parte del ciclo secundario de su colegio.

La experiencia consiste en armar una historia con caminos alternativos, al estilo elige tu propia aventura, utilizando Scratch®. Para la misma serán tenidos en cuenta algunos componentes como personajes, escenarios, historias y música, entre otros. Las actividades propuestas permiten la enseñanza de programación utilizando bifurcaciones que construyan las habilidades básicas necesarias para su resolución. Se emplean conceptos como representación, abstracción, descomposición y modularización. Como objetivo conjunto se encuentra la realización de un intercambio entre escuelas mediante el aprendizaje de habilidades de programación.

La importancia de analizar este proyecto radica en las condiciones sociales, económicas y ambientales en la que se encontraban inmersos los participantes. Ambos establecimientos corresponden a escuelas de las denominadas rurales. El grupo, en general, no contaba con todas las posibilidades tecnológicas necesarias para este tipo de actividades. Por ejemplo, muchos alumnos no poseían computadora o notebook y en muchos casos tampoco dispositivos celulares. Otra gran dificultad es la falta o débil conexión a internet. Esto les impedía conectarse a los encuentros sincrónicos pautados a realizarse a través de la plataforma de videoconferencias Zoom. Como consecuencia, muchas veces tampoco contaban con la posibilidad de realizar las actividades propuestas. Sumado a ello, se pueden destacar los problemas eléctricos para la utilización de dispositivos electrónicos, y los ambientales. Éstos últimos, impedían el traslado de alumnos y capacitadores a las instituciones correspondientes, por encontrarse, en ocasiones, la zona anegada.

Teniendo en cuenta los casos hasta aquí mencionados de desarrollo de actividades STEAM, este resulta el más claro en cuanto a grupos de estudiantes que se encuentran en condiciones de carencias tecnológicas, edilicias, económicas y/o ambientales para realizar las

actividades planteadas. Para los fines de esta investigación, esta cápsula es catalogada como STEAM extremo.

Como resultado del trabajo se obtuvo que muchos de los encuentros no pudieron llevarse a cabo por los inconvenientes antes descritos, no se contaba con todos los participantes disponibles, o se presentaban fallas de conexión a internet en forma habitual. A pesar de ello, tanto estudiantes como capacitadores destacan como positiva la interacción con otros grupos en las mismas condiciones, demostrando mucho interés y logrando avanzar principalmente en las tareas de índole artística. Estos aprendizajes llevaron a los organizadores a generar alternativas de actividades, de medios de comunicación e incluso de objetivos para poder alcanzar las metas de enseñanza básicas propuestas.

Todos los datos relacionados al trabajo realizado en ambas escuelas son tomados en cuenta e incorporados a la base de datos de la plataforma, bajo el concepto de cápsula. Se incluye información sobre sus posibilidades de acceso a la tecnología, el contexto sociocultural, los conocimientos previos, rango etario, elementos utilizados, inconvenientes detectados, etc. De esta forma, la estadística y minería de datos aplicada podrá generar resultados que incluyan a grupos de trabajo STEAM en las condiciones previamente detalladas.

Hasta aquí se han expuesto los principales datos referidos a las cápsulas que han sido seleccionadas para formar parte del presente trabajo. En las mismas encontramos temas y disciplinas diversas, como son la programación y robótica, la biología y los juegos de rol.

Asimismo, de cada cápsula se han capturado resultados referidos a sus implementaciones. Los mismos, fueron recabados a partir de formularios tanto físicos como digitales, y gracias a la colaboración de los capacitadores y organizadores. Dichos resultados fueron cargados a la base de datos de la plataforma, para permitir posteriormente su análisis a través de la estadística y la minería de datos.

A continuación, se presentan los materiales o herramientas utilizadas para el desarrollo de la presente obra.

### ***3.2.3 Materiales Utilizados***

En esta sección se detallan las principales herramientas tecnológicas utilizadas en las distintas etapas de la investigación.

#### ***Prototipo***

Como se mencionó anteriormente, este trabajo se basa en un prototipo web ya desarrollado y denominado Museo Virtual STEAM. En el mismo recayeron los ajustes necesarios para permitir el almacenamiento integral de todos los datos que servirán como fuente de entrada para el desarrollo de la presente obra. Esto incluye modificaciones en su diseño y almacenamiento.

La utilización de dicho prototipo se puede destacar desde dos perspectivas. Por un lado, la interfaz gráfica que permite la interacción del usuario. Gracias a ella, se pueden registrar los movimientos de los usuarios en la plataforma. Por otro lado, su base de datos. A partir de la cual, se obtienen los datos de forma unificada tanto de lo ocurrido en el sitio web como los resultados cargados de las actividades realizadas.

Todo análisis posterior está basado en lo que el prototipo pueda ofrecer luego de las actualizaciones mencionadas. En ese punto radica la vital importancia de dicho artefacto de software.

#### *Infostat (versión 2020)*

Para la realización de los análisis estadísticos se utilizó el software denominado Infostat. El mismo, según (Di Rienzo y otros), posee la ventaja de contar con una sencilla interfaz que permite operaciones profesionales en el manejo de datos. Adicionalmente, cuenta con módulos preparados para la enseñanza de la estadística y una versión en español para el usuario habla hispana (2020). Además, comentan que la misma puede ser ejecutada en cualquier computadora con sistema operativo Windows y conectarse con R<sup>5</sup> (software muy reconocido para el desarrollo de algoritmos estadísticos)

La elección de dicho software reside en la utilización simple que ofrece, su versión en idioma español, el soporte para el sistema operativo Windows y la robustez de sus funcionalidades. Adicionalmente, la utilización de este software fomenta y engrandece la tecnología nacional, ya que fue creado por investigadores y docentes de la Universidad Nacional de Córdoba.

#### *Weka (versión 3.9.6)*

---

<sup>5</sup> R es un software con distribución libre que ofrece un entorno para el desarrollo de la computación estadística y generación de gráficos. El mismo puede ser ejecutado, por ejemplo, en plataformas UNIX, Windows o MacOS (R Project, 2022).

Weka es un software utilizado para la aplicación de minería de datos y aprendizaje automático. Dicha aplicación reviste de carácter libre y abierto, distribuido bajo la licencia GNU-GPL<sup>6</sup>. Esto permite su acceso gratuito por cualquier entidad. Adicionalmente, se pueden extender sus funcionalidades sin mayor complejidad (López De Luise, 2020).

La elección de Weka radica principalmente en su gran aceptación en la comunidad científica tanto para estudios profesionales como académicos, además de la posibilidad de utilización de forma libre y gratuita.

Una vez enunciados los materiales utilizados, se procede a exponer los métodos empleados en la investigación.

### ***3.2.4 Métodos Aplicados***

A continuación, se describen las principales características de los métodos aplicados a los distintos dataset utilizados. Los mismos, fueron extraídos de la base de datos de la plataforma.

#### ***Clustering***

El clustering, también conocido como análisis por agrupamiento, utiliza algoritmos que logren una clasificación del dataset en grupos menores (clusters). Lo realiza a partir del hallazgo de características que resulten similares en cada sub grupo. Teniendo en cuenta determinados criterios, los grupos distintos no compartirán esas mismas propiedades o características (Heynz Roberth & Ticona Gonzáles, 2020). En cuanto a la clasificación del tipo de estudio, según los autores, pertenece al grupo de aprendizaje conocido como no supervisado.

Algunos de los algoritmos principales para realizar clustering son los denominados K-Means y Expectation Maximization.

#### ***K-Means***

Como se mencionó, K-Means es uno de los algoritmos utilizados al realizar clustering. En el mismo, se debe indicar antes de la ejecución la cantidad (k) de clusters que se desean generar (Choez Franco, 2022). Cada subgrupo tendrá un valor medio conocido como centroide, y se busca determinar la distancia de cada objeto hacia ese valor medio ya definido (Vintimilla

---

<sup>6</sup> GNU-GPL: Un software se considera libre si fue publicado bajo una licencia de software libre. La mayoría de las veces se utiliza la Licencia Pública General de GNU (GPL de GNU) (Free Software Foundation, Inc., 2022).

y otros, 2017). De esta forma, se va completando cada subgrupo según la cantidad solicitada por el usuario.

### *Expectation Maximization (EM)*

Otro de los algoritmos utilizados al realizar clustering es el denominado Expectation Maximization. El mismo, define una distribución de probabilidad de cada elemento. Dicha distribución, informa la probabilidad de que el elemento pertenezca a cada uno de los subgrupos o clústeres. Al realizar este estudio, el usuario puede especificar cuántos clústeres desea generar o dejar que lo resuelva el propio algoritmo.

### *Classify*

Los métodos de clasificación pertenecen al machine learning y son utilizados para clasificar elementos de acuerdo a su salida o resultado. Dicha salida será discreta. La misma, cuenta con un número finito de posibilidades, generalmente dos. Por ejemplo, podrá responder a la pregunta si una cápsula será elegida para ser implementada o no. Para resultados que sean de carácter continuo, podrán utilizarse los algoritmos denominados de regresión. Estos últimos no forman parte de la presente investigación.

Dentro de los algoritmos de clasificación se encuentra el denominado J48. El mismo se describe a continuación.

### *J48*

Este heurístico tiene la habilidad de generar un modelo de cómo se clasifican las instancias basándose en el concepto de Entropía, término acuñado originalmente por (Shannon & Weaver, 1949). Según los autores, el mismo refiere a que los símbolos utilizados, por ejemplo en un texto, brindan una cierta información promedio. Para ejemplificar, pueden entenderse como más valiosos los sustantivos que los artículos que integran un texto, ya que brindarán mayor y mejor información del sentido del mismo.

## **3.3 Desarrollo de la Investigación**

A continuación, se describen los procesos llevados a cabo con los datos y métodos vistos en las secciones anteriores. Los mismos se dividen en dos etapas: Pre-procesamiento y Procesamiento de Datos. Este último, cuenta con un análisis cualitativo de los resultados obtenidos en cada ejecución.

### 3.3.1 Etapa 1: Pre-procesamiento

En primer lugar, se procede a generar consultas SQL<sup>7</sup> que extraigan de la base de datos todos los datos que se desean analizar. En este caso, se utilizarán tres consultas específicas:

1. Consulta SQL referida a los datos de las cápsulas seleccionadas y detalladas en la sección anterior.
2. Consulta SQL para los resultados de las implementaciones de las actividades.
3. Consulta SQL que contenga los datos de navegación de los usuarios en la plataforma.

Luego de la obtención de esos resultados, se procede a la exportación de los mismos. La misma debe ser en un formato .csv separado por “;” (punto y coma) que permita su apertura en una aplicación del tipo planilla de cálculo. Posteriormente, se verifica su correcta exportación para poder avanzar con la etapa de pre-procesamiento.

En la etapa de pre-procesamiento, se realizan todos los procedimientos previos a la minería a de datos. El objetivo del pre-procesamiento, pormenorizado en las líneas siguientes, es la obtención de un conjunto de datos, o dataset, más limpio o pulido. Esta característica refiere a eliminar todos los datos que puedan generar errores o simplemente no generen un valor agregado al análisis.

Para dicha limpieza de datos, normalmente se utilizaría un software especializado, por ejemplo, el ya mencionado R. Con el mismo, pueden ser ejecutadas distintas librerías de software que recorren el dataset completo y realizan su limpieza. Sin embargo, en este caso, al tratarse de una cantidad de datos manipulable, se procede de forma manual, lo que asegura un pulido más racionalizado.

A continuación, se realizan las verificaciones columna por columna, realizando las correcciones necesarias en cada caso. Para comenzar, se verifica que todos los registros de una misma columna sean del mismo tipo. Por ej., se chequea que no permanezcan en una misma columna, algunos datos almacenados como números y otros del tipo texto.

---

<sup>7</sup> SQL (lenguaje estructurado de consulta) es un lenguaje de computación que permite la manipulación de conjuntos de datos y sus relaciones. SQL es un estándar de carácter internacional, reconocido por ISO y ANSI (Microsoft, 2022).

Luego, se comprueba que todos los registros cuenten con el número total de columnas. Es decir, que a ninguno le falten datos. Asimismo, se eliminan todos los valores “null”, ya que al ser procesados por algunos softwares, pueden ser interpretados erróneamente como valores concretos en lugar de la falta de valor que se espera.

Otro chequeo que se realiza normalmente, es referido a la aparición de caracteres especiales (símbolos, signos, etc.). Los mismos deberían ser eliminados o reemplazados para obtener un texto más limpio, pero, en este caso, no se quitarán. Esto se debe a que no generan inconvenientes en la aplicación que será utilizada para su análisis.

En cuanto a los valores contenidos, y contextualizándolos en su utilización en algoritmos de minería de datos, se pueden distinguir, a modo general, en datos métricos y nominales. Los datos métricos o numéricos deben ser controlados para evitar ocurrencias de valores que se distancien de manera considerada de la normal de datos. Por ejemplo, si la mayoría de los datos rondan el valor 5, no se considerará algún registro que figure como 100000. Dicha consideración se debe tomar en cuenta para las columnas que representen un valor en sí misma y en comparación con otros semejantes. Por ejemplo, si la numeración de calles de una dirección difiere considerablemente, no significará ningún inconveniente, ya que los números de calle no forman parte del análisis propuesto.

En cuanto a los valores nominales o cadenas de texto, también denominadas en informática Strings, se verificará la posibilidad de ser reemplazadas por algún carácter más corto. Por ejemplo, si al referirse a la edad de los alumnos aparecen de forma repetitiva textos como “Desde los 10 años”, entonces, el mismo será reemplazado, por ejemplo, por la letra “A”. Esta modificación permite que los datos puedan ser procesados de manera más eficiente por el software en cuestión. De esta forma, se libera a la memoria del almacenamiento y procesamiento de cadenas de textos más largas.

Otra manipulación que se le realiza a los datos para favorecer su sistematización, es la de acortar los valores a su mínima expresión, sin perder su significado. Por ejemplo, en el caso de una columna que refiere a la duración de una actividad y existen valores del tipo “Duración 6 horas”. Dicho valor será reemplazado por “360”, entendiéndolo como minutos.

Un último examen que se realiza es el de completar los datos faltantes y eliminar los que no generen valor agregado. Este último caso puede ocurrir por falta de valor o por contener todos los registros exactamente el mismo dato.

A modo de revisión general, se evalúa la consistencia de datos. Esto refiere al control de que no figuren datos fuera de contexto, que hayan sido relacionados de forma incorrecta, o que, por conocimiento del tema tratado, se consideren erróneos.

Hasta aquí se expusieron los diferentes tratamientos que sufrieron los datos al ser preparados para el procesamiento con el software seleccionado. Dicho proceso se explica a continuación, y fue realizado para cada una de las tres consultas SQL mencionadas al comienzo de esta sección.

### ***3.3.2 Etapa 2: Procesamiento De Datos***

#### ***Weka***

Para comenzar a trabajar con este software, en primer lugar, se abre el archivo con el conjunto de datos y se lo formatea nuevamente para que sea tomado exitosamente por Weka. Esto incluye ajustar el separador correcto (reemplazando “;” por “.” y “;” por “,”), arrobas (reemplazados por “.at.”), eliminar saltos de líneas incorrectos, modificar caracteres especiales como comillas, etc.

Con las modificaciones realizadas se carga el archivo en Weka.

A partir de este momento, se puede observar como análisis preliminar el porcentaje de registros de cada columna que no poseen datos, cuántos datos únicos posee el dataset, cuántos valores distintos, tipos de datos existentes, etc.

Para un análisis más exhaustivo, se comienza por definir cuál o cuáles de los atributos pertenecientes al dataset nos aporta información valiosa, ya sea por sí mismos o por una precalificación en clases. En ese sentido, primero se deben especificar el Método de Búsqueda y el Atributo Evaluador que se utilizarán. El primero de ellos, es un mecanismo para buscar los atributos más relevantes y cuáles se relacionan con ellos. El atributo Evaluador, es un procedimiento que nos permitirá definir el grado de eficiencia de esa agrupación. A través del mismo podemos observar si en verdad presentan un comportamiento distintivo que puede aportar información valiosa, o no.

Otro elemento a definir es el modo de selección de atributos. El mismo define el mecanismo mediante el cual serán procesados los datos. Podrá ser “Use full training set” o “Cross validation”. El primero, recoge todos los datos cargados y los valida entre sí, de acuerdo al criterio de selección elegido. Este método es conocido como “optimista” y suele ser utilizado

cuando la carga de datos no es tan pesada. En cuanto a “Cross validation”, conocido como “pesimista”, divide el dataset en 10 (por ejemplo) partes o folds utilizando números al azar y una “semilla”. Luego, utiliza 9 de ellos para la construcción del modelo y el restante (semilla) lo utiliza para validar. Posteriormente repite el procedimiento hasta cubrir la totalidad de 10 partes. Al final de las 10 validaciones se promedia el resultado. En este caso, el modo de selección de atributos elegido será “Use full training set”.

El siguiente elemento a definir es si utilizaremos clases. Las mismas pueden agrupar la información de acuerdo a algún criterio predefinido. Por ejemplo, puedo tener un conjunto de datos de inversores, y cada registro seteado con una “B” para “buen inversor” y “M” para “mal inversor”. Entonces, Weka nos permite realizar el análisis teniendo en cuenta la/s clase/s que podamos haber definido para una columna o, simplemente, no utilizar clases.

Antes de poder realizar el análisis con Weka, el dataset de entrada debe ser compatible con el método de búsqueda elegido, el atributo evaluador seleccionado y la clase definida (si hubiera). De lo contrario, no se podrá realizar dicho análisis.

### Dataset: Cápsulas y actividades

En la Tabla 1, se observan los datos más relevantes que presenta el dataset que refiere, principalmente, a los datos de las cápsulas y actividades que se exponen en la plataforma.

**Tabla 1**

*Datos principales de cápsulas y actividades*

Datos de cápsulas y actividades	
Nombre de cápsula	Datos del patrocinador
Elementos disponibles	Kit de elementos utilizados
Voluntarios participantes	Patrocinadores
Datos geográficos de los participantes	Nombre/tipo/descripción de las actividades propuestas
Tiempos de desarrollo	Palabras claves
Edades recomendadas	Nivel de dificultad
Tema estudiado	Ubicación geográfica de la implementación

Asimismo, al evaluar específicamente los identificadores que presenta el dataset, los cuales serán tratados en diversas ejecuciones, se puede observar en la Tabla 2:

**Tabla 2***Identificadores del dataset cápsulas y actividades completo*

ID	Atributo
1	Id_capsula
3	Id_capsula_patrocinador
4	Id_patrocinador
12	Id_kit_elemento_kit
14	Id_elemento_kit
16	Id_capsula_voluntario
17	Id_voluntarioCapacitador
18	Id_categoria_voluntarioCapacitador
26	Id_tipo_voluntarioCapacitador
28	Id_actividad_capsula
29	Id_actividad
35	id_actividad_keywords
35	Id_keyword
38	id_actividad_entidad_organizadora
39	Id_entidad_organizadora

Para el conjunto de datos completo, se proponen una serie de combinaciones de método de búsqueda, atributo evaluador y clase elegida. Cada una de estas combinaciones, representa una perspectiva diferente de análisis, por lo que la evaluación de cada resultado proveerá de valor a la investigación. A continuación, se detallan cada una de las 18 ejecuciones realizadas, sus resultados y análisis correspondientes.

### **Ejecución 1**

Se procede con la carga del dataset completo y la siguiente combinación de opciones.

Atributo Evaluador: ClassifierSubsetEval: Utiliza un clasificador o algoritmo de aprendizaje para evaluar los diferentes conjuntos que se obtienen con el método de búsqueda. Este método no realiza validación cruzada con otros datos externos (Troncoso Lora, 2017).

Método de búsqueda = GreedyStepwise: Tiene por objetivo encontrar un grupo de atributos que en conjunto den un resultado útil para el análisis (Rodríguez Manzanares, 2018).

Class: Sin clase. Al no utilizar clases, implica que buscará las características generales de las cápsulas clasificándolas de acuerdo a su propia naturaleza, y no en base a una tipificación

realizada previamente por una persona. De esta forma, generará subconjuntos de comportamientos y a partir de ahí recomendará los atributos del dataset que permitirán trabajarlos más exhaustivamente.

A continuación, en la Figura 1, se puede observar el buffer o resultado de la ejecución.

### Figura 1

#### *Buffer Weka de la ejecución 1*

```
Evaluation mode:    evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Greedy Stepwise (forwards).
  Start set: no attributes
  Merit of best subset found:    0.283

Attribute Subset Evaluator (supervised, Class (nominal): 40 Descripcion):
  Classifier Subset Evaluator
  Learning scheme: weka.classifiers.rules.ZeroR
  Scheme options:
  Hold out/test set: Training data
  Subset evaluation: classification error

Selected attributes:
```

Como se puede observar, el grupo “Selected attributes” permanece vacío. Esto implica que el dataset otorgado, teniendo en cuenta los datos y la cantidad de registros existentes en el mismo, no se puede estudiar de forma genérica (sin clase seleccionada). Dicho de otra forma, no encuentra comportamientos significativos para clasificar las cápsulas de forma general. Las cápsulas contienen características distribuidas de forma tan preponderantemente homogénea que no permitió encontrar atributos seleccionables.

### Ejecución 2

Atributo Evaluador: ClassifierSubsetEval

Método de búsqueda = GreedyStepwise

Class: id\_keyword. En este caso, el atributo es del tipo numérico. Este campo refiere al identificador de cada palabra clave de una cápsula. Por lo tanto, en este caso se trabajará para averiguar cuáles son las variables que se deberán tener en cuenta para poder identificar o mejorar las cápsulas de acuerdo al tema o tipo de contenido de cada una.

A continuación, en la Figura 2, se puede observar el buffer o resultado de la ejecución.

### Figura 2

#### *Buffer Weka de la ejecución 2*

```
Evaluation mode:    evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Greedy Stepwise (forwards).
  Start set: no attributes
  Merit of best subset found:    2.961

Attribute Subset Evaluator (supervised, Class (numeric): 36 id_keyword):
  Classifier Subset Evaluator
  Learning scheme: weka.classifiers.rules.ZeroR
  Scheme options:
  Hold out/test set: Training data
  Subset evaluation: RMSE

Selected attributes:
```

En este caso, también se puede observar que los “Selected attributes” permanecen vacíos. Por lo tanto, tampoco se cuenta con suficiente información para trabajar las cápsulas por contenido temático.

### Ejecución 3

Atributo Evaluador: CfsSubseteval: Este evaluador se basa en las correlaciones existentes en un subconjunto de atributos. Se destacan las correlaciones entre clases y se le resta preponderancia a la intercorrelación del subconjunto con sus pares (Troncoso Lora, 2017).

Método de búsqueda BestFirst: Este método realiza una búsqueda solo en el espacio que pertenece a los subconjuntos de atributos (Martin Ramos y otros, 2012).

Class: edades\_recomendadas. El atributo edades\_recomendadas es del tipo nominal ya que presentaba datos del tipo “A partir de los 10 años”. Entonces, en el pre-procesamiento antes descripto, esas frases fueron reemplazadas por caracteres únicos. Por ejemplo, “A”. Esta variable se selecciona para poder estudiar las cápsulas, conocerlas o inferir su comportamiento en función de las edades objetivo.

A continuación, en la Figura 3, se puede observar una previsualización del buffer o resultado de la ejecución.

### Figura 3

#### *Buffer Weka de la ejecución 3*

```
Selected attributes: 1,2,3,10,13,14,17,20,23,24,25,29,33,35,36,40 : 16
Instance_number
id_capsula
desc_capsula
localidad_patrocinador
id_kit_elemento_kit
cod_kit
id_capsula_voluntario
tipo
provincia_voluntarioCapacitador
localidad_voluntarioCapacitador
mail_voluntarioCapacitador
id_actividad_capsula
nivel_dificultad
tiempo_desarrollo
id_actividad_keywords
id_entidad_organizadora
```

Al estudiar el conjunto de datos de esta forma, se obtienen 16 atributos que deberán ser seleccionados para un análisis más profundo de las cápsulas. Son los campos que, estudiados de la forma realizada, contienen mayor preponderancia con respecto al resto para la caracterización de cápsulas.

A partir del resultado obtenido se procede a reducir el dataset de entrada. Para el caso, se eliminan todos los atributos que no hayan sido resultado en la sección “Selected attributes” de la Figura 3. Posteriormente se guarda el nuevo conjunto de datos, el cual será tratado luego con Clasificación J48 e Infostat.

**Ejecución 4**

Tomando el buffer o resultado de la ejecución anterior (edades recomendadas), se procede a realizar Clasificación utilizando J48.

El conjunto de datos sufrió la nominalización de los atributos ID 3, 4, 12, 16, 17, 18, 26, 28, 29, 36, 38, 39. La clasificación se realiza en base al campo id\_entidad\_organizadora, que es la entidad que organiza la cápsula.

La matriz de confusión (evalúa los datos reales vs los datos predichos) obtenida es:

**Tabla 3**

*Matriz de confusión con dataset edades recomendadas*

Matriz de confusión						
A	B	C	D	E	F	← classified as
10	0	0	0	0	0	A = 2
0	12	0	0	0	0	B = 3
0	0	2	0	0	0	C = 4

0	0	0	45	0	0	D = 5
0	0	0	45	0	0	E = 6
0	0	0	45	0	0	F = 7

Se aprecia que el modelo puede distinguir perfectamente las cápsulas del tipo A (IEEE JUJUY UCSE DASS - Programación y Robótica), B (Trivias de intercambio Chaco-Jujuy), C (Introducción a la Programación) y D (Juegos de Mesa), pero no las E (Modelización de la membrana plasmática) y F (Modelización de una célula). Estas últimas evidencian tener información muy similar entre sí y con las del tipo D.

El árbol podado se puede observar a continuación en la Figura 4. Ahí se aprecia el efecto de solapamiento de las categorías e y f. Debido a esta falta de discernimiento entre subcomportamientos es que el Kappa statistics (medida de confiabilidad del modelo), ha quedado con un valor bajo (0.25). A fin de mejorar estos resultados, se separan del dataset los casos de cápsulas e y f.

#### Figura 4

*Buffer Weka de la ejecución 4*

```

J48 pruned tree
-----
desc_capsula = A: 2 (10.0)
desc_capsula = B: 3 (12.0)
desc_capsula = W: 4 (2.0)
desc_capsula = C: 5 (135.0/90.0)

Number of Leaves :    4
Size of the tree :    5
Time taken to build model: 0 seconds

=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds

=== Summary ===
Correctly Classified Instances      69           43.3962 %
Incorrectly Classified Instances    90           56.6038 %
Kappa statistic                     0.2452
Mean absolute error                 0.1887
Root mean squared error            0.3071
Relative absolute error             75.1785 %
Root relative squared error        86.8744 %
Total Number of Instances         159

```

De lo anterior se puede concluir que las entidades son fácilmente definibles a partir de las características de sus cápsulas (válido para las del tipo a, b, c y d). En el caso de las otras dos cápsulas (e y f) esto no es tan concluyente. Probablemente, esto se debe a que las entidades generadoras tienen mayor capacidad para realizar eventos de diversa índole.

#### Ejecución 5

Atributo Evaluador: CfsSubseteval.

Método de búsqueda BestFirst.

Class: id\_actividad. El mismo, representa al identificador de cada actividad perteneciente a una cápsula en la plataforma.

A continuación, en la Figura 5, se puede observar una previsualización del buffer o resultado de la ejecución.

### Figura 5

*Buffer Weka de la ejecución 5*

```
Selected attributes: 1,28,30,31,35,38 : 6
    id_capsula
    id_actividad_capsula
    titulo
    Resumen
    id_actividad_keywords
    id_actividad_entidad_organizadora
```

A partir del resultado obtenido se procede a reducir el dataset de entrada. Para el caso, se eliminan todos los atributos que no hayan sido resultado en la sección “Selected attributes” de la Figura 5. Posteriormente, se guardan el nuevo conjunto de datos, el cual será tratado con clustering y más adelante con Infostat.

### Ejecución 6

Tomando el buffer o resultado de la ejecución anterior, se procede a realizar clustering con EM. Los datos en este análisis corresponden a la nominalización de los atributos 1, 3, 4, 12, 14, 16, 17, 18, 26, 28, 29, 35, 36, 38, 39. Los mismos fueron seleccionados en base a su capacidad para discriminar por id\_actividad. El resultado se puede observar a continuación en la Figura 6.

### Figura 6

*Buffer Weka de la ejecución 6*

```
Clustered Instances
0      135 ( 85%)
2       12 (  8%)
4       12 (  8%)

Log likelihood: -3.55028
```

Los 3 clusters hallados indican que, desde la perspectiva de datos, las cápsulas pueden agruparse por 3 tipos distintos. El más popular (85%) corresponde al cluster con ID 0, típicamente con actividad ID 13 (Aprender a programar con Scratch). Mientras que los clusters 2 y 4 tienen dos o tres tipos de actividades. El cluster 0 tiene como actividades típicamente los números 12 y 13 (Juegos de Mesa). En el cluster 2 predominan las actividades 10 (Modelización de la membrana plasmática) y 11 (Modelización de una célula), con algunas del tipo 9

(Robótica y Electrónica extracurricular). El cluster 4 es prácticamente actividad ID 7 (Introducción a la Programación).

### **Ejecución 7**

Para este caso, en primer lugar, se nominalizan todos los atributos que sean identificadores de registros en la tabla de base de datos. Por ejemplo, `id_capsula`, `id_actividad`, etc. Luego, se selecciona la siguiente combinación de opciones.

Atributo Evaluador: `CfsSubsetEval`.

Método de búsqueda = `GreedyStepwise`.

Class: Sin clase.

A continuación, en la Figura 7, se puede observar el resultado de la ejecución.

### **Figura 7**

*Buffer Weka de la ejecución 7*

```
Selected attributes: 1,38,39 : 3
                    id_capsula
                    id_actividad_entidad_organizadora
                    id_entidad_organizadora
```

Como se mencionó previamente, los datos no tienen un sesgo de información específica como en las ejecuciones previas. El resultado indica que las cápsulas son suficientemente diversas, por lo que sus principales características diferenciadoras son la entidad que las realiza y los tipos de actividad declarados. Fuera de eso no son relevantes otras características (lugar de residencia, edad, nivel de dificultad, etc.). De igual forma, estos resultados serán utilizados para la próxima ejecución.

### **Ejecución 8**

Con el resultado obtenido de la ejecución 7, se procede a realizar agrupamientos mediante clustering con EM. Dicho proceso nos permite identificar la forma en que se distingue una cápsula de otra, de forma natural de acuerdo a sus características.

En la Figura 8, se pueden observar 6 tipos de cápsulas distintivas en sus propiedades.

### **Figura 8**

*Buffer Weka de la ejecución 8*

```

=== Model and evaluation on training set ===

Clustered Instances

0      45 ( 28%)
1      45 ( 28%)
2      45 ( 28%)
3       8 (  5%)
4       6 (  4%)
5      10 (  6%)

Log likelihood: -2.08288

```

Se pueden observar 3 agrupaciones dominantes con 28%, correspondientes a los clusters 0, 1 y 2. Analizando los eigenvectores (valores no nulos y numéricos) que representan a los mismos, se observa que corresponden a las actividades 9 (Robótica y Electrónica extracurricular), 7 (Introducción a la Programación) y 8 respectivamente, realizadas por las entidades 7 (SCA), 5 (Ministerio de Educación, Cultura, Ciencia y Tecnología de la provincia de Chaco) y 8 (Museo Histórico Sarmiento) en cada caso. Los clusters 3 y 4 corresponden a actividades 4 y 5 (Juegos de mesa) combinadas y todas realizadas por la entidad 3 (UCSE DASS). El cluster id 5 representa a la cápsula id 2, realizada por la entidad id 2 (UADER FCyT), caracterizada por las actividades id 2 (Modelización de una célula) y 3 (Modelización de la membrana plasmática).

### Ejecución 9

En este caso, se parte nuevamente del resultado obtenido en la ejecución 7. Se procede a remover los atributos que no son seleccionados por el algoritmo, y luego, se realiza clustering. En esta oportunidad, se utiliza SimpleKMeans. En la Figura 9 se puede observar una previsualización del resultado de dicha ejecución.

#### Figura 9

*Buffer Weka de la ejecución 9*

```

=== Model and evaluation on training set ===

Clustered Instances

0      114 ( 72%)
1       45 ( 28%)

```

Este resultado confirma la tendencia a dos tipos de cápsulas: una mayoritaria (las que en la ejecución 8 correspondían a los clusters 0, 1 y 2) y otras minoritarias (el resto). Se confirma con el análisis de los eigenvectores, ya que en cluster id 0 (mayoritario) domina la entidad

5 y la actividad 7, como en la ejecución 8. La diferencia en los porcentajes se justifica porque en los criterios usados por EM existe la hipótesis de distribución normal sin solapamientos, en cambio en k-means estas hipótesis no existen. La cantidad de iteraciones es baja (2) por lo que el algoritmo ha convergido por estabilidad del error, confirmando también que las asignaciones son por características definidas y no hay oscilaciones en las mismas.

### **Dataset: Tracking de usuarios**

Una primera versión del prototipo de la plataforma fue publicada y distribuida en la comunidad académica y científica. La misma fue tomada con gran aceptación y, al cabo de siete días, se lograron obtener 946 registros de usuarios. Estos incluían accesos de distintas ciudades de Argentina, por ejemplo: Buenos Aires, Villa La Angostura, La Falda, Santa Fe, Concepción del Uruguay, San Luis y Resistencia, entre otras. Asimismo, también se registraron accesos de distintas ciudades del mundo, por ejemplo: South Houston, Bogotá, Santiago de Chile, París, Atlanta, Dearing, Washington y Berlín, entre otras.

#### **Ejecución 1**

Atributo Evaluador: CfsSubseteval.

Método de búsqueda BestFirst.

Class: elemento\_principal. El atributo elemento\_principal representa al tipo de elemento accedido por un usuario. Por ejemplo: link, imagen, botón, menú, formulario, etc.

En la Figura 10, se puede observar una previsualización del buffer obtenido con esta ejecución.

#### **Figura 10**

*Buffer tracking de usuarios*

```
Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 27
  Merit of best subset found: 0.722

Attribute Subset Evaluator (supervised, Class (nominal): 8 elemento_principal):
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 3,4 : 2
                    fecha_hora
                    url
```

Como se observa en la figura, la selección del heurístico indica que la determinación del elemento principal depende de, esencialmente, dos variables: fecha\_hora y URL. Ello evidencia que los accesos no dependen del origen geográfico (país, región, ciudad), sino que el

criterio de navegación es universal (transversal al origen del individuo). Como los datos en estudio corresponden al tipo de contenido accedido, indica que ciertos contenidos tienen relevancias similares para todos los usuarios.

## Ejecución 2

Partiendo del resultado anterior, se procede a generar un nuevo dataset. En el mismo, se eliminan todos los atributos que no hayan sido seleccionados por el software y tampoco sea el atributo elegido como clase. Por lo tanto, se trabaja con los atributos `fecha_hora`, `url` y `elemento_principal`.

Con este nuevo grupo de datos, se realiza clustering con EM. Una previsualización del resultado se puede observar a continuación en la Figura 11.

### Figura 11

*Buffer tracking de usuarios reducido*

Clustered Instances	
0	343 ( 36%)
1	427 ( 45%)
2	111 ( 12%)
3	5 ( 1%)
4	59 ( 6%)

Log likelihood: -9.29652

Se observa la distinción en 5 clusters, el análisis de los mismos es el siguiente:

Cluster 1: Se caracteriza por el acceso (en orden descendente de prioridad) a los ítems: Imagen, Avanzar/Retroceder carrusel, Lupa, Link. Describiendo el comportamiento de quienes desean conocer los contenidos de las cápsulas. En total 427, representando el 45%.

Cluster 0: Se caracteriza por el acceso (en orden descendente de prioridad) a los ítems: Participe, Blog. Describiendo el comportamiento de quienes se interesan por la parte estratégica y administrativa. En total 343, representando el 36%.

Cluster 2: Se caracteriza por el acceso (en orden descendente de prioridad) a los ítems: Patrocinadores, Kits, Capacitadores y voluntarios. Describiendo el comportamiento del tipo de filiaciones y entidades que participan. En total 111, representando el 12%.

Cluster 4: Se caracteriza por el acceso (en orden descendente de prioridad) a los ítems: Cápsula - Modelización de una célula, Entidades, Claves, Actividad. Describiendo el comportamiento de quienes desean conocer tipo de cápsulas ofrecidas. En total 59, representando el 6%.

Cluster 3: Se caracteriza por el acceso (en orden descendente de prioridad) a los ítems: Kits, Capacitadores y voluntarios. Describiendo el comportamiento de quienes se interesan por el perfil del grupo humano en cada cápsula. En total 5, representando el 1%.

A partir de aquí, se continúa con el estudio de los distintos datasets con el software Infostat.

### *Infostat*

#### **Dataset: Cápsulas y actividades**

Para continuar, se procesará con el software Infostat algunos de los archivos anteriormente tratados con Weka.

#### **Ejecución 1**

En este primer caso, se trabaja con estadísticas descriptivas para permitir su ordenamiento, caracterización y presentación. Específicamente, se trata el dataset completo y se evalúan las siguientes medidas resumen: media, desvío estándar (evalúa la dispersión de valores respecto al promedio), mínimo y máximo. En la Tabla 4, se observa el resultado de la ejecución.

**Tabla 4**

*Resultado Infostat: Medidas resumen con el dataset completo*

<b>Medidas resumen</b>					
<b>Variable</b>	<b>N</b>	<b>Media</b>	<b>D.E.</b>	<b>Mín</b>	<b>Máy</b>
id_capsula	159	4,65	0,87	2	5
id_capsula_patrocinador	159	6,50	1,44	3	8
id_patrocinador	159	6,50	1,44	3	8
id_kit_elemento_kit	159	14,97	2,89	5	18
cod_kit	159	5,65	0,87	3	6
id_elemento_kit	159	18,43	4,76	1	22
id_categoria_voluntarioCapacitador	159	1,15	0,36	1	2

Teniendo en cuenta el dataset seleccionado y los resultados obtenidos, se puede confirmar que el id de patrocinador dominante es el 7 (Ministerio de Educación, Ciencia y Tecnología de la provincia de Chaco), aunque existe también, en menor medida, intervención de otros. La cápsula dominante es la id 5 (Trivias de intercambio Chaco-Jujuy) con un id de elemento dominante entre el 14 y el 15, presentando una pronunciada kurtosis hacia el valor máximo (señal

de una tendencia hacia los id más elevados). La distribución de id categoría voluntario no muestra id especiales.

### Ejecución 2

Para este caso, contando con el dataset completo, se procede a generar una tabla de frecuencias. La misma, permitirá obtener de forma ordenada el número de repeticiones de cada valor. En la Figura 12, se puede observar su resultado.

**Figura 12**

*Resultado Infostat: Tablas de frecuencia con el dataset completo*

Variable	Clase	LI	LS	MC	FA	FR
id_kit_elemento_kit	1	[ 5,00	6,86 ]	5,93	4	0,03
id_kit_elemento_kit	2	( 6,86	8,71 ]	7,79	4	0,03
id_kit_elemento_kit	3	( 8,71	10,57 ]	9,64	8	0,05
id_kit_elemento_kit	4	( 10,57	12,43 ]	11,50	7	0,04
id_kit_elemento_kit	5	( 12,43	14,29 ]	13,36	28	0,18
id_kit_elemento_kit	6	( 14,29	16,14 ]	15,21	54	0,34
id_kit_elemento_kit	7	( 16,14	18,00 ]	17,07	54	0,34

Variable	Clase	LI	LS	MC	FA	FR
id_elemento_kit	1	[ 1,00	4,00 ]	2,50	9	0,06
id_elemento_kit	2	( 4,00	7,00 ]	5,50	0	0,00
id_elemento_kit	3	( 7,00	10,00 ]	8,50	0	0,00
id_elemento_kit	4	( 10,00	13,00 ]	11,50	4	0,03
id_elemento_kit	5	( 13,00	16,00 ]	14,50	10	0,06
id_elemento_kit	6	( 16,00	19,00 ]	17,50	55	0,35
id_elemento_kit	7	( 19,00	22,00 ]	20,50	81	0,51

Variable	Clase	LI	LS	MC	FA	FR
id_categoria_voluntarioCap..	1	[ 1,00	2,00 ]	1,50	126	0,79
id_categoria_voluntarioCap..	2	( 2,00	3,00 ]	2,50	0	0,00
id_categoria_voluntarioCap..	3	( 3,00	4,00 ]	3,50	2	0,01
id_categoria_voluntarioCap..	4	( 4,00	5,00 ]	4,50	2	0,01
id_categoria_voluntarioCap..	5	( 5,00	6,00 ]	5,50	2	0,01
id_categoria_voluntarioCap..	6	( 6,00	7,00 ]	6,50	0	0,00
id_categoria_voluntarioCap..	7	( 7,00	8,00 ]	7,50	27	0,17

Variable	Clase	LI	LS	MC	FA	FR
cod_kit	1	[ 3,00	3,43 ]	3,21	10	0,06
cod_kit	2	( 3,43	3,86 ]	3,64	0	0,00
cod_kit	3	( 3,86	4,29 ]	4,07	12	0,08
cod_kit	4	( 4,29	4,71 ]	4,50	0	0,00
cod_kit	5	( 4,71	5,14 ]	4,93	2	0,01
cod_kit	6	( 5,14	5,57 ]	5,36	0	0,00
cod_kit	7	( 5,57	6,00 ]	5,79	135	0,85

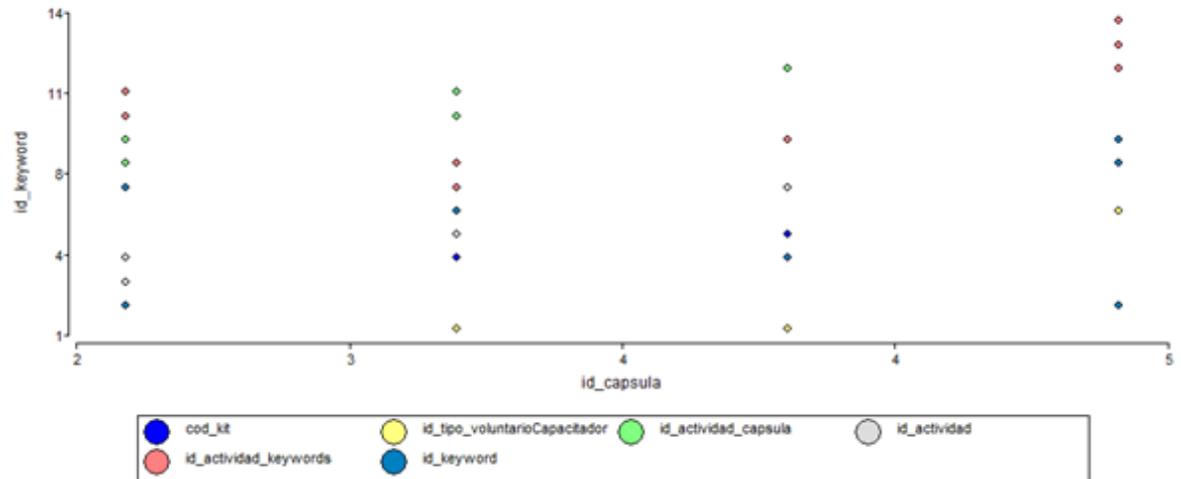
De las tablas se aprecia que las clases se diferencian principalmente por los elementos del kit, ya que los LI (límites inferiores) y LS (límites superiores) no se solapan, aunque hay algunos kits de mayor frecuencia, como son el caso del cod\_kit 7 (software Scratch, software Zoom, conexión a internet), y los id\_elemento\_kit 5 a 7 (proyector, pc, pizarrón).

### Ejecución 3

A continuación, se prosigue generando un gráfico de dispersión, nuevamente con el conjunto de datos en su totalidad. En la Figura 13, se puede observar el resultado obtenido.

**Figura 13**

*Resultado Infostat: Gráfico de dispersión con el dataset completo*



La gráfica confirma lo observado en la Ejecución 3. Hay otros datos que resultan dominantes, pero en menor grado. Ahora también se incluye la asociación id\_actividad y sus correspondientes id\_keyword.

**Ejecución 4**

En la Ejecución 3 del tratamiento de archivos con Weka, se estudió un subconjunto del dataset. El mismo respondía a los atributos sobresalientes al analizar los datos en función de las edades recomendadas.

Para esta ejecución, se toma ese nuevo conjunto de datos, se lo procesa con Infostat, y se obtiene una nueva tabla de frecuencias. En la Figura 14, se puede observar su resultado.

**Figura 14**

*Resultado Infostat: Gráfico de dispersión por edades recomendadas*

Variable	Clase	LI	LS	MC	FA	FR
id_capsula	1	[ 2,00	2,43 ]	2,21	10	0,06
id_capsula	2	( 2,43	2,86 ]	2,64	0	0,00
id_capsula	3	( 2,86	3,29 ]	3,07	12	0,08
id_capsula	4	( 3,29	3,71 ]	3,50	0	0,00
id_capsula	5	( 3,71	4,14 ]	3,93	2	0,01
id_capsula	6	( 4,14	4,57 ]	4,36	0	0,00
id_capsula	7	( 4,57	5,00 ]	4,79	135	0,85

Variable	Clase	LI	LS	MC	FA	FR
id_actividad_capsula	1	[ 8,00	8,71 ]	8,36	5	0,03
id_actividad_capsula	2	( 8,71	9,43 ]	9,07	5	0,03
id_actividad_capsula	3	( 9,43	10,14 ]	9,79	6	0,04
id_actividad_capsula	4	( 10,14	10,86 ]	10,50	0	0,00
id_actividad_capsula	5	( 10,86	11,57 ]	11,21	6	0,04
id_actividad_capsula	6	( 11,57	12,29 ]	11,93	2	0,01
id_actividad_capsula	7	( 12,29	13,00 ]	12,64	135	0,85

Variable	Clase	LI	LS	MC	FA	FR
nivel_dificultad	1	[ 1,00	1,43 ]	1,21	135	0,85
nivel_dificultad	2	( 1,43	1,86 ]	1,64	0	0,00
nivel_dificultad	3	( 1,86	2,29 ]	2,07	2	0,01
nivel_dificultad	4	( 2,29	2,71 ]	2,50	0	0,00
nivel_dificultad	5	( 2,71	3,14 ]	2,93	10	0,06
nivel_dificultad	6	( 3,14	3,57 ]	3,36	0	0,00
nivel_dificultad	7	( 3,57	4,00 ]	3,79	12	0,08

Variable	Clase	LI	LS	MC	FA	FR
id_actividad_keywords	1	[ 7,00	8,00 ]	7,50	12	0,08
id_actividad_keywords	2	( 8,00	9,00 ]	8,50	2	0,01
id_actividad_keywords	3	( 9,00	10,00 ]	9,50	5	0,03
id_actividad_keywords	4	( 10,00	11,00 ]	10,50	5	0,03
id_actividad_keywords	5	( 11,00	12,00 ]	11,50	45	0,28
id_actividad_keywords	6	( 12,00	13,00 ]	12,50	45	0,28
id_actividad_keywords	7	( 13,00	14,00 ]	13,50	45	0,28

En este caso se observa que el nivel de dificultad predomina (ver distribución FA y FR en nivel\_dificultad). Hay mayores eventos en nivel básico de dificultad, lo que es lógico ya que muchas de las actividades se realizaron en primarios. Le siguen los niveles 5 y 7 (típicos para adultos que han avanzado por algunos cursos previos). Como es de esperar, las keywords también tienen una distribución específica a cada caso (ver distribución id\_actividad\_keywords), con marcada incidencia de los ID 5 a 7.

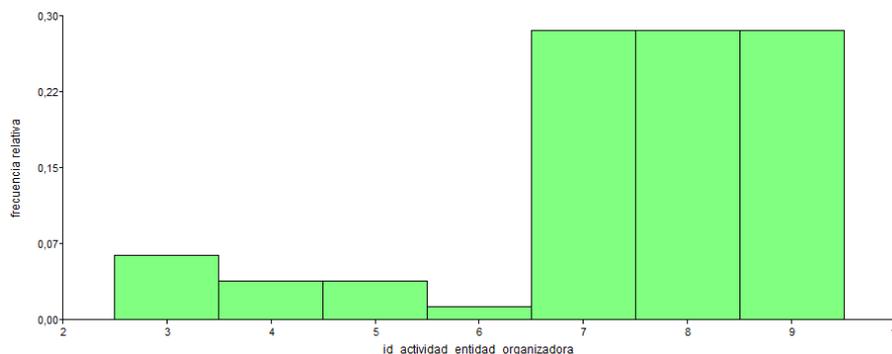
### Ejecución 5

En la ejecución 3 del tratamiento con Weka, se estudió el dataset en función del tema de cada cápsula. En esta ocasión se partirá desde el resultado de dicho estudio, para generar el histograma que se observa a continuación.

En la Figura 15, se visualiza su resultado.

### Figura 15

*Resultado Infostat: Histograma con el dataset por tema de cápsula*



En este gráfico es sencillo observar la diferencia entre las entidades con capacidad organizadora sistemática, ya que se observan con mayores frecuencias que el resto. Se lo puede apreciar en los índices relativos a los ID 7, 8 y 9 (SCA, IEEE GTC y MHS respectivamente). Se puede concluir que existen entidades que realizaron actividades STEAM como evento de excepción, mientras que otras las llevan a cabo de forma sistemática y asumida.

### Dataset: Tracking de usuarios

Para el dataset donde se almacenaron los registros de solicitud de contenido de los usuarios en la plataforma, se llevaron a cabo los procesos que se detallan a continuación.

### Ejecución 1

La primera ejecución se realiza con el fin de identificar los principales sitios geográficos de acceso a la plataforma. En la Figura 16, se observa el resultado al analizar la tabla de frecuencias absolutas obtenida:

**Figura 16**

*Tabla de frecuencias tracking de usuarios*

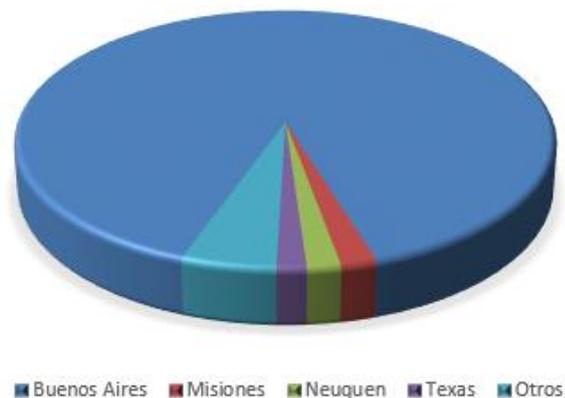
Variable Clase	Categorías	FA	Variable Clase	Categorías	FA
region	1 Alabama	1	region	13 Neuquen	17
region	2 Berlin	1	region	14 San Luis	1
region	3 Buenos Aires	16	region	15 Santa Fe	5
region	4 Buenos Aires F.D.	831	region	16 Santiago Metropolitan	3
region	5 Chaco	3	region	17 Saxony	1
region	6 Cordoba	5	region	18 Texas	15
region	7 Entre Rios	1	region	19 Thuringia	1
region	8 Georgia	2	region	20 Uusimaa	2
region	9 Illinois	1	region	21 Virginia	5
region	10 Kansas	2	region	22 Washington. D.C.	2
region	11 Mendoza	9	region	23 Åžle-de-France	3
region	12 Misiones	18			

En la figura se observa como principales locaciones de acceso a Buenos Aires, Misiones, Neuquén y Texas.

Si tenemos en cuenta las frecuencias relativas solo de los sitios con mayor acceso, podemos visualizar el resultado graficado a continuación en la Figura 17.

**Figura 17**

*Frecuencias relativas por región de acceso*



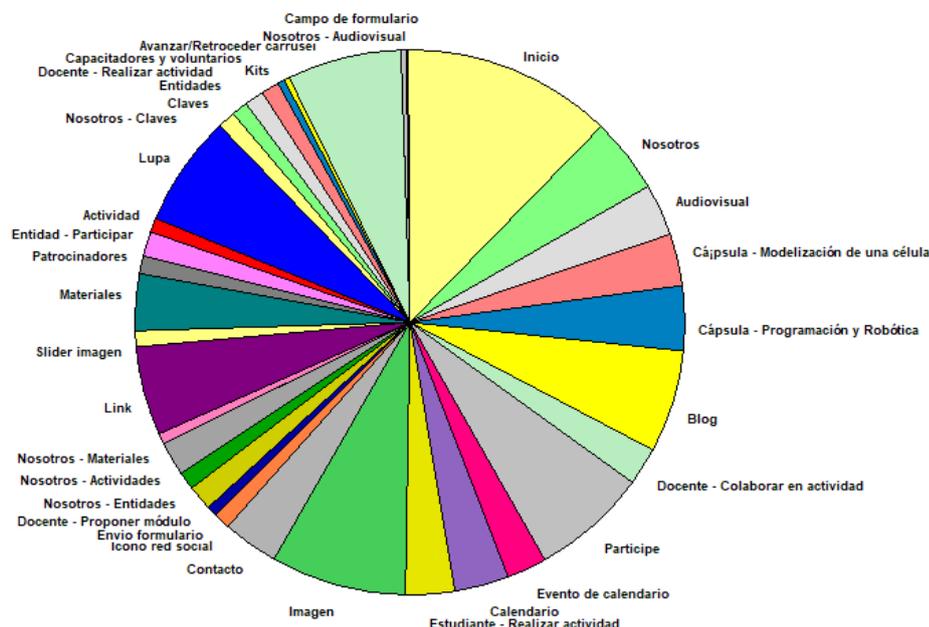
Nuevamente, la región de Buenos Aires fue la más activa, seguida por el conjunto de ciudades alrededor del mundo con una participación menor a Misiones, Neuquén y Texas.

## Ejecución 2

En esta ejecución, se realiza el mismo estudio del caso previo, pero teniendo en cuenta los elementos seleccionados por cada usuario en la plataforma. Estos incluyen links, imágenes, menús, formularios, etc. De esta forma, también se contribuye a la obtención de información a través del user profiling explicado en el capítulo 2.

A partir de las frecuencias relativas obtenidas y sus respectivas representaciones en porcentajes, se obtiene el resultado que se observa en la Figura 18.

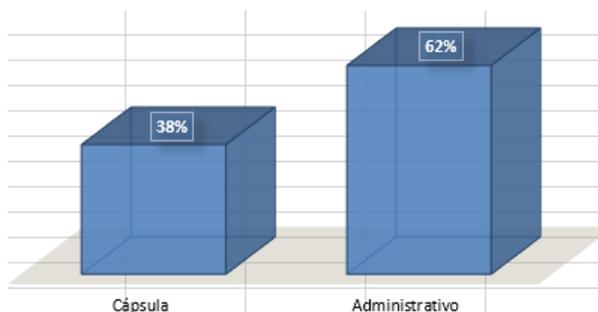
**Figura 18**  
*Accesos por elemento*



En el gráfico, se puede observar rápidamente como los elementos más elegidos fueron: la página principal (acceso al home del sitio) y la visualización de imágenes. Completan los mayores accesos el módulo Participo (incluyen formularios para envío de información) y Blog (fotos, videos e información de la implementación de cápsulas). A modo de resumen, los elementos más atractivos de acuerdo a su contenido para las personas, son los gráficos y los que presentan la posibilidad de interacción administrativa.

Siguiendo esta última interpretación, se agrupó los diferentes elementos en dos categorías: Cápsulas (para todo elemento referido a una cápsula) y Administrativo (para todo elemento referido al sitio, contactos e información extra cápsulas). Al evaluar los datos de esa forma, se obtienen los porcentajes de acceso como se observan a continuación en la Figura 19.

**Figura 19**  
*Gráfico accesos cápsula-administrativo*



Los accesos a controles web administrativos predominan por sobre los referidos a cápsulas ya realizadas. De este análisis, puede inferirse un interés mayor por información y comunicación con los distintos organismos, que por la visualización del material de implementaciones ya realizadas. Este análisis corrobora el resultado obtenido en procesos previos al evaluar el mismo dataset con el software Weka, realizando clustering con expectation maximization.

### *Análisis cualitativo*

#### **Dataset: Formularios post actividad**

Como se comentó anteriormente, luego de la implementación práctica de una cápsula, se realiza una recolección de datos resultantes. Los mismos se obtienen a través de los participantes de la actividad. Ellos pueden ser estudiantes, docentes, coordinadores, capacitadores, etc.

Para el presente trabajo, se cuenta con la información recolectada de cuatro cápsulas distintas. Dicha información, es recopilada a través de formularios web y, luego, exportados a formato .csv para su estudio cualitativo o cuantitativo, según se considere oportuno.

En la Tabla 5, se pueden observar algunos de los datos más relevantes que forman parte del dataset obtenido.

**Tabla 5**

*Datos de formularios post actividad*

Datos de implementación de actividades	
Cápsula	Duración de la actividad
Actividad	Organizador/es
Nivel de dificultad	Coordinador/es
Edades recomendadas	Comentarios
Tiempo de ejecución	Datos del encuestado
Fecha de implementación	¿Qué te pareció la actividad?
Establecimiento anfitrión	Dificultades surgidas

Para esta fase del trabajo, se considera que la cantidad de registros del conjunto de datos es escasa, pero amerita ser estudiada mediante un análisis cualitativo. En ese sentido, se lleva a cabo el examen de información, y se pueden mencionar las siguientes observaciones:

De acuerdo a las respuestas obtenidas en los formularios, se obtiene que, según los propios participantes, la mayoría de las actividades poseen un grado bajo de dificultad. Es decir,

no presentan inconvenientes para ser completadas. Asimismo, de este subgrupo de experiencias, se observa que las mismas son destinadas principalmente a adolescentes, teniendo una valoración preponderantemente “Regular”. Este análisis, en resumen, permite evidenciar que los adolescentes no estiman como “Buenas” o “Muy buenas” actividades las que no exigen un grado mayor de conocimiento o utilización de sus facultades mentales para la resolución.

Según los resultados, también se observa que la gran mayoría de las actividades puede ser resuelta en un tiempo máximo de 90 minutos, y ninguna de estas fue votada como “Mala experiencia” ni se recibieron observaciones al respecto. Por lo que se puede inferir que el tiempo de ejecución es el adecuado.

Por último, y en cuanto a las dificultades surgidas durante las implementaciones, se destacan principalmente las clasificadas como tecnológicas y el estado climático a la hora de efectuar las actividades. De aquí se puede realizar una primera aproximación sobre la complejidad de realizar STEAM en locaciones donde el acceso a la tecnología informática presente complicaciones de forma regular. Asimismo, muchas de estas dificultades también se sumaron a las generadas por las inclemencias climáticas. Las mismas, actuaron como obstáculos para el arribo de los participantes o para la ejecución de las tareas. De aquí se puede evaluar, por ejemplo, la posibilidad de llegar físicamente a más lugares con cada experiencia propuesta. Este tópico, también será tratado más en detalle en el capítulo referido a las conclusiones del trabajo.

#### **4. Prototipo**

En este capítulo se presenta el prototipo que es tomado como base para el desarrollo de la presente investigación. Dicho prototipo, se encuentra implementado sobre una plataforma Wordpress. La misma, permite la creación de páginas web de cualquier tipo de forma rápida y sencilla (Wordpress, 2022).

En primer lugar, se expondrán las pantallas y funcionalidades principales inherentes a este trabajo. Luego, se detallarán los elementos más importantes de la base de datos de la plataforma. Finalmente, se diserta sobre las pruebas realizadas a los efectos de lograr una óptima ejecución de los procesos involucrados.

## 4.1 Introducción

La aplicación de los métodos de análisis y evaluación propuestos para el Museo Virtual STEAM se aplican al actual prototipo vigente. El mismo fue desarrollado en el año 2020 con el objetivo de alentar la difusión de las cápsulas educativas STEAM (López De Luise, 2020).

A continuación, se presentan las pantallas y características principales de dicho prototipo, mencionando los ajustes necesarios realizados en cada etapa. Estas adaptaciones posibilitaron la obtención de todos los datos necesarios para llevar a cabo la investigación.

### 4.1.1 Apariencia De La Plataforma

La plataforma cuenta con una sección superior que resulta transversal a cada una de sus pantallas. En la misma, se puede observar el logo de la página, una barra de menús y los accesos a las distintas redes sociales. Estos componentes revisten de importancia ya que favorecen la dinámica en la interacción entre el usuario y el sistema en la plataforma. Dichos elementos se encontrarán fácilmente disponibles en cualquiera de las pantallas del sistema.

Con respecto a los menús disponibles, los mismos son: Inicio, Nosotros, Calendario, Participe, Blog y Contacto. A continuación, se expondrán los más importantes en cuanto a la competencia de esta investigación.

#### *Pantalla Inicio*

En la Figura 20, se visualiza la pantalla principal de la plataforma virtual.

#### **Figura 20**

*Prototipo - Pantalla de Inicio del Museo Virtual STEAM*



*Nota: Home Page del prototipo de la plataforma virtual (Leiva y otros, 2022b).*

En esta pantalla, se puede observar en primer lugar, un conjunto de imágenes del tipo carrusel. El mismo contiene una imagen representando a cada elemento de STEAM y se visualiza unos segundos para luego dar paso al siguiente elemento. Debajo del mismo, se ubica una leyenda que incentiva al usuario a recorrer la plataforma en busca de los distintos materiales disponibles. Como su nombre lo indica, es la pantalla inicial o principal, el recibimiento y primera presentación hacia el usuario web.

### *Pantalla Nosotros*

A continuación, en la Figura 21, se visualiza la pantalla Nosotros.

### **Figura 21**

#### *Prototipo - Pantalla Nosotros*



*Nota: Pantalla Nosotros del prototipo de la plataforma virtual*  
(Leiva y otros, 2022b)

Una vez que el usuario se encuentra situado aquí, podrá visualizar una breve reseña a modo de presentación de la página, sus objetivos, entidades organizadoras, etc.

Adicionalmente, se presentan los distintos elementos que pueden ser consultados y la forma de acceso a ellos. Entre los cuales, se pueden observar el Material teórico, Audiovisual, Entidades organizadoras, Actividades y Claves<sup>8</sup>.

---

<sup>8</sup> Las palabras clave puede servir para representar un trabajo, documento, o como en este caso, un espacio web (López De Luise, 2020).

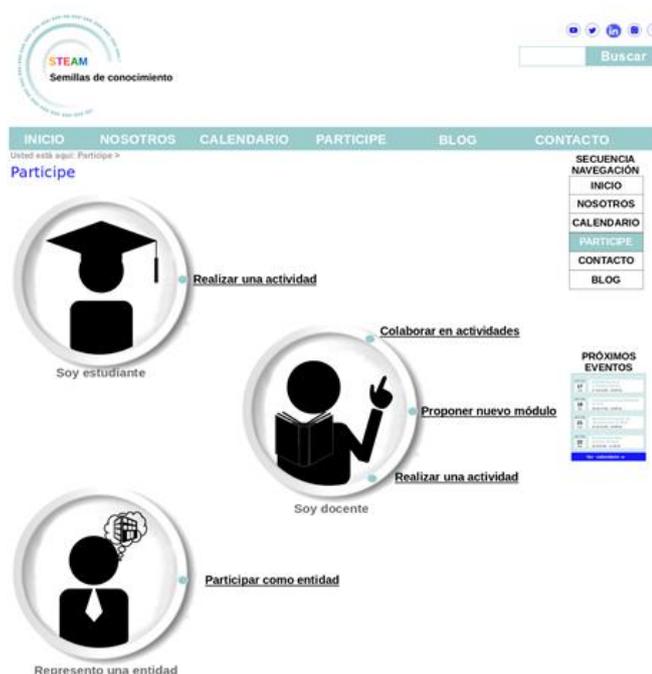
Por otro lado, algunos componentes a tener en cuenta a la hora de destacar la interacción del sistema con el usuario son: el calendario de próximos eventos donde se implementarán las actividades propuestas, y las piezas de navegación visual para que el usuario pueda ubicarse y reconocer el sector del sitio en el que se encuentra.

### *Pantalla Participe*

Como siguiente distinción, en la Figura 22, se observa la pantalla Participe.

#### **Figura 22**

##### *Prototipo - Pantalla Participe*



*Nota: Pantalla Participe del prototipo de la plataforma virtual  
(Leiva y otros, 2022b)*

En esta pantalla, se pueden discriminar 3 grupos de funcionalidades. Las mismas refieren a las necesidades y posibilidades de acción de acuerdo al rol del internauta.

Como estudiante, podrá acceder a la información adecuada para realizar distintas actividades. Como docente, podrá realizar o colaborar en actividades, además de contar con la posibilidad de proponer un nuevo módulo. En cuanto al usuario de tipo entidad, el mismo podrá participar en alguna de las actividades ofrecidas.

Esta pantalla es una de las consideradas relevantes para el presente trabajo. A través de la misma, se podrá identificar el tipo de usuario que está haciendo uso de la plataforma, qué acciones decide efectuar y qué rutas de navegabilidad traza de acuerdo a su propio interés. En

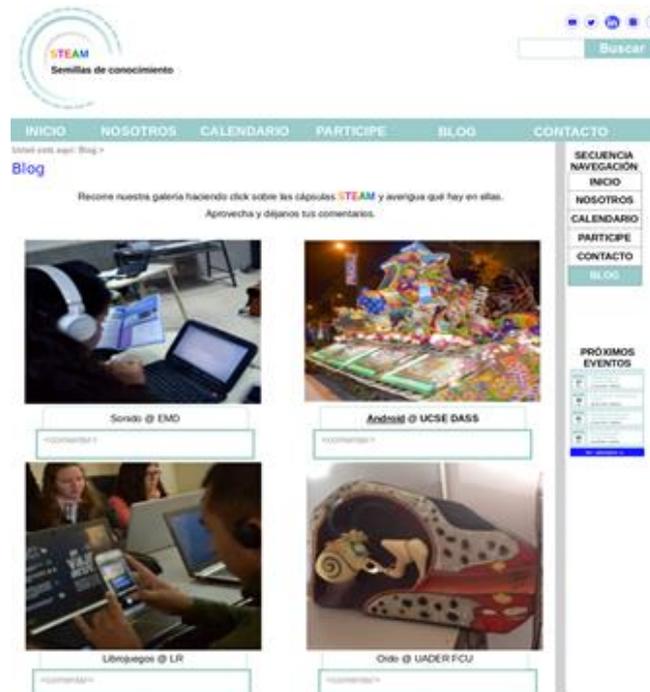
cualquiera de los casos antes mencionados, el cibernauta contará con la oportunidad de enviar consultas o solicitar información a través de formularios web. Los mismos, también forman parte del análisis de datos que se realizó y fue especificado en el capítulo anterior.

### *Pantalla Blog*

A continuación, en la Figura 23, se observa la pantalla Blog.

### **Figura 23**

#### *Prototipo - Pantalla Blog*



*Nota: Pantalla Blog del prototipo de la plataforma virtual (Leiva y otros, 2022b)*

La página Blog es otra de las consideradas importantes para la investigación. Esto se debe a que la misma posee cuantiosa información que el usuario puede consultar o generar. Dichas acciones decantan en un conjunto de datos de navegación del internauta, todos plausibles de análisis y evaluación.

Al ingresar a esta página, el sistema presenta una galería con imágenes de las distintas cápsulas STEAM implementadas. Para cada una de ellas, el usuario cuenta con la posibilidad de agregar un comentario o ingresar para más información. Si decide ingresar, el sistema mostrará nuevas opciones de selección, entre las cuales se encuentran: Audiovisual, Materiales, Capacitadores, Documentos explicativos, Eventos realizados de la cápsula, etc.

De esta forma, el usuario podrá navegar en todas las cápsulas ofrecidas, en cada una de sus actividades, y en cada evento donde hayan sido implementadas. En todo momento contará con la información audiovisual ilustrativa y la funcionalidad de enviar sus propios comentarios.

Hasta aquí se destacaron las pantallas y funcionalidades más importantes de la plataforma, en cuanto a su relevancia para este trabajo.

Seguidamente, se analizará otro aspecto del sistema. El mismo reviste carácter de máxima importancia debido a que es la herramienta tecnológica que oficia como fuente de información principal para el presente trabajo. A continuación, se examina la base de datos del sistema.

#### **4.1.2 Base De Datos**

Como se comentó anteriormente, la plataforma del Museo Virtual STEAM está basada en la estructura de contenidos web ofrecida por el sitio Wordpress. Al instalarse, se pueden observar a nivel base de datos las características que se detallan a continuación.

El proceso de instalación crea 12 tablas, las cuales almacenan todos los datos necesarios para el uso básico a partir de una nueva implantación. Además de los datos propios del sistema, Wordpress tiene la particularidad de que cada una de sus páginas también son almacenadas en las distintas tablas de su base de datos. Las mismas son traducidas y presentadas al usuario según sean solicitadas (Tellado, 2021).

Si el administrador de la plataforma decide la incorporación de algún elemento de software que expanda las funcionalidades básicas (plugin), el mismo instalará también sus propias tablas de gestión. Todas serán ubicadas en la misma base de datos, por lo que la información se concentra en un solo lugar. El diagrama entidad relación (DER) inicial se puede apreciar en la Figura 24 de la sección Anexo 1 Diagramas Entidad Relación.

A partir de allí, se procedió a la creación y articulación de nuevas tablas. Los datos provenientes de la propia plataforma, de la interacción de los usuarios en la misma, y de los resultados de las actividades serán concentrados en estas nuevas tablas. El diagrama entidad relación generado a partir de dicha intervención se puede observar en la Figura 25 de la sección Anexo 1 Diagramas Entidad Relación.

Por otro lado, para finalizar la articulación de las pantallas del sitio y su nueva base de datos, se procedió a configurar la plataforma. En este caso, lo restante fue realizar la

vinculación necesaria entre ambos componentes para que toda interacción del usuario sea persistida en las tablas anteriormente creadas.

A partir de este punto, la base de datos del sistema ya cuenta con toda la información necesaria y la flexibilidad suficiente para servir como fuente integral de datos. Los mismos, serán empleados en el análisis estadístico y de minería de datos propuestos en esta investigación.

## **4.2 Testing**

Todos los tratamientos de datos realizados en la presente obra fueron llevados a cabo con datos reales. Esto incluye los datos de cápsulas que expone la plataforma, los datos capturados luego de cada implementación de actividades, y los referentes a la navegación de los usuarios en el sistema.

La estructura que sufrió diferentes cambios a partir de las pruebas realizadas fue la referida al formato de cada conjunto de datos o dataset utilizado. Se partió de una base de dicha estructura que contemplaba, inicialmente, lo necesario para poder abordar los procesos estadísticos y de minería de datos planteados. A partir de allí, se realizaron diferentes análisis y pruebas cualitativas para, posteriormente, actualizar la estructura de acuerdo a lo observado. Finalmente, se obtuvo una estructura de datos operativa para cada dataset. La misma permitiría contar con todos los datos necesarios y suficientes para realizar los análisis que la investigación requiere.

Para más detalle, en el Anexo 2, se describe la estructura de campos final para cada uno de los tres datasets utilizados.

## **5. Informe Final**

Una vez llevada a cabo la aplicación de los distintos procesos detallados, se desatacan los siguientes resultados y recomendaciones.

Al referirnos a la información de las cápsulas, se encuentra que la diferenciación entre cada una de ellas se da principalmente en dos grupos: tecnología (predominantemente programación) y ciencias naturales. Al tener en cuenta este resultado, se sugiere la incorporación de cápsulas enfocadas principalmente en disciplinas como el arte y las

matemáticas. De esta forma, el conocimiento transmitido por la plataforma será más holístico en cuanto a la educación STEAM en su totalidad.

En referencia a las actividades, se hallan muchas de las mismas con niveles básicos de dificultad. Dichos niveles, resultan beneficiosos para los estudiantes de primaria, pero poco atractivos para alumnos de mayor edad. Por lo tanto, sería provechoso para el objetivo de la plataforma, la incorporación de actividades con mayor nivel de dificultad cuando se destinen a alumnos adolescentes.

Con respecto a las dificultades de implementación de actividades, se destacan la tecnología y el acceso al lugar físico de la actividad. De aquí se desprende la necesidad de incorporar equipamiento tecnológico que se pueda proveer a los estudiantes. Asimismo, será beneficioso para los alumnos la posibilidad de realizar las mismas actividades en distintas locaciones en simultáneo, incluso de forma virtual. Esto brindará una mayor posibilidad de acceso a todos los estudiantes.

En cuanto a las cápsulas, de acuerdo a su tipo, suelen tener una entidad en particular que las desarrolla. Esto se destaca como positivo ya que cada entidad en cuestión posee mayor experiencia en el tema tratado. De igual forma, y en referencia a los 90 minutos promedio de actividad, se encontró que es el tiempo indicado para cada ejercicio.

Al evaluar el trackeo de usuarios realizado en la plataforma virtual, se observa que el contenido preferido de acceso no depende del lugar de procedencia del usuario. Esto lleva a establecer un acceso de contenido de tipo universal, de importancia e interés similar independientemente del lugar de procedencia del cibernauta. Adicionalmente, se detectó de forma mayoritaria, el acceso a contenido informativo o administrativo. El mismo, superó al contenido referido a cápsulas. De aquí, se observa fundamentalmente como mejora, la posibilidad de brindar mayores canales de comunicación de acuerdo a los distintos tipos de usuarios existentes.

Cada uno de estos resultados contribuyó al hallazgo de tendencias o estándares que favorecen la toma de decisiones en pos de promover la educación a través de la plataforma. Cada ocasión de mejora fue mencionada oportunamente. Las más destacadas fueron los tipos de cápsulas a incorporar, diversidad de actividades a proponer, niveles de dificultad preferidos a considerar y el tipo de información que resulta atractiva para los usuarios.

## 6. Conclusiones

A partir de la aplicación de los métodos expuestos en esta investigación, se obtienen procesos de carácter sistemático y reproducible de evaluación de las actividades educativas promovidas desde el Museo Virtual STEAM. De esta forma, se obtiene información que permite la evolución del propio sistema como impulsor de educación.

El modelo de almacenamiento generado integra en una misma base de datos toda la información concerniente a la plataforma web, a la interacción del usuario con la misma, y los resultados de las implementaciones de cada cápsula. Es decir, toda la información necesaria para realizar los procesos antes detallados.

En cuanto a la cantidad de datos utilizados en el trabajo, la misma resultó conveniente para todos los procesos llevados a cabo. No se presentaron inconvenientes bloqueantes referidos a la falta de información. Asimismo, la calidad de los datos fue completamente suficiente para cada análisis realizado. Los mismos representaron a un grupo mayor de datos que no se tuvieron en cuenta por pertenecer a otras perspectivas de análisis no alcanzadas por esta investigación.

A la hora de evaluar los datos antes mencionados, tomó expresa importancia la utilización de métricas. Los softwares y técnicas utilizadas se basan en distintas métricas, como por ejemplo: kappa statistics, log likelihood y matrices de confusión, entre otras. Todas ellas, métricas tradicionales en la evaluación de los procesos de minería de datos. Gracias a las mismas, se obtuvieron los resultados antes mencionados, por lo que representan la base de cada tratamiento y otorgan un aporte fundamental a esta obra.

En cuanto a las cápsulas utilizadas, una mención especial amerita el análisis realizado a la denominada “Trivias de intercambio Chaco – Jujuy”. La misma, fue estudiada como caso único de STEAM extremo. A partir de ella, se llevó a cabo una descripción de las características distintivas encontradas. Las recomendaciones al respecto, incluyen la facilitación de elementos tecnológicos necesarios a los alumnos cuando fuera necesario, y la posibilidad de llegar a mayores locaciones (incluso de forma virtual y simultánea). Esto último, permitirá fomentar el acceso de los estudiantes, ya sea de manera física o digital.

Al finalizar la totalidad de estudios realizados, se observa como el análisis estadístico y la aplicación de minería de datos llevados a cabo en esta investigación, permitieron la evaluación de la totalidad del modelo educativo propuesto en la plataforma web. De esta forma, el sistema se retroalimenta de forma continua a si mismo con información cada vez más

enriquecida. En consecuencia, se favorece la evaluación de su rendimiento, su evolución y se producen mejoras en: la plataforma, su contenido, la navegabilidad del usuario, las actividades propuestas, la enseñanza y aprendizaje, y, en definitiva, la experiencia de cada participante.

Por último, se pueden mencionar algunas limitaciones encontradas en el transcurso de la investigación.

Por un lado, si se analiza una posible clasificación de las cápsulas trabajadas, se evidencia la necesidad de explotar otras cápsulas del tipo STEAM extremo. En este caso, se analizó solo una con estas características. Esto no permitió una evaluación que decante en explicaciones más profundas de los sucesos, ni tampoco la generación de resultados estadísticos.

Por otro lado, una mención con respecto al estudio específico de datos. La obtención, clasificación, formateo, incorporación a la plataforma y posterior tratamiento de los mismos se realizó de forma manual. Esto influía directamente al retrasar los tiempos de realización para culminar cada proceso, así como en un potencial re-tratamiento en caso de modificaciones.

## **7. Líneas Futuras De Investigación**

Como se comentó anteriormente, de esta investigación formó parte un solo caso de los denominados STEAM extremo. En ese sentido, puede extenderse la recopilación de datos referidos a este tipo de cápsulas. Al contar con más actividades similares, la cantidad de datos analizados será mayor, lo que dará paso a otros tipos de análisis, hallazgos y explicaciones de fenómenos de los expresados en esta obra.

Asimismo, todos esos nuevos análisis y resultados mencionados, implicarán la gestión de mayor cantidad de información en cada etapa del proceso. Por lo tanto, también se sugiere la investigación de formas de automatización de cada uno de los procesos involucrados. De esta manera, no solo se permitirá la realización de procedimientos de modo más eficiente, sino que se disminuirá la posibilidad del error humano en el tratamiento de datos.

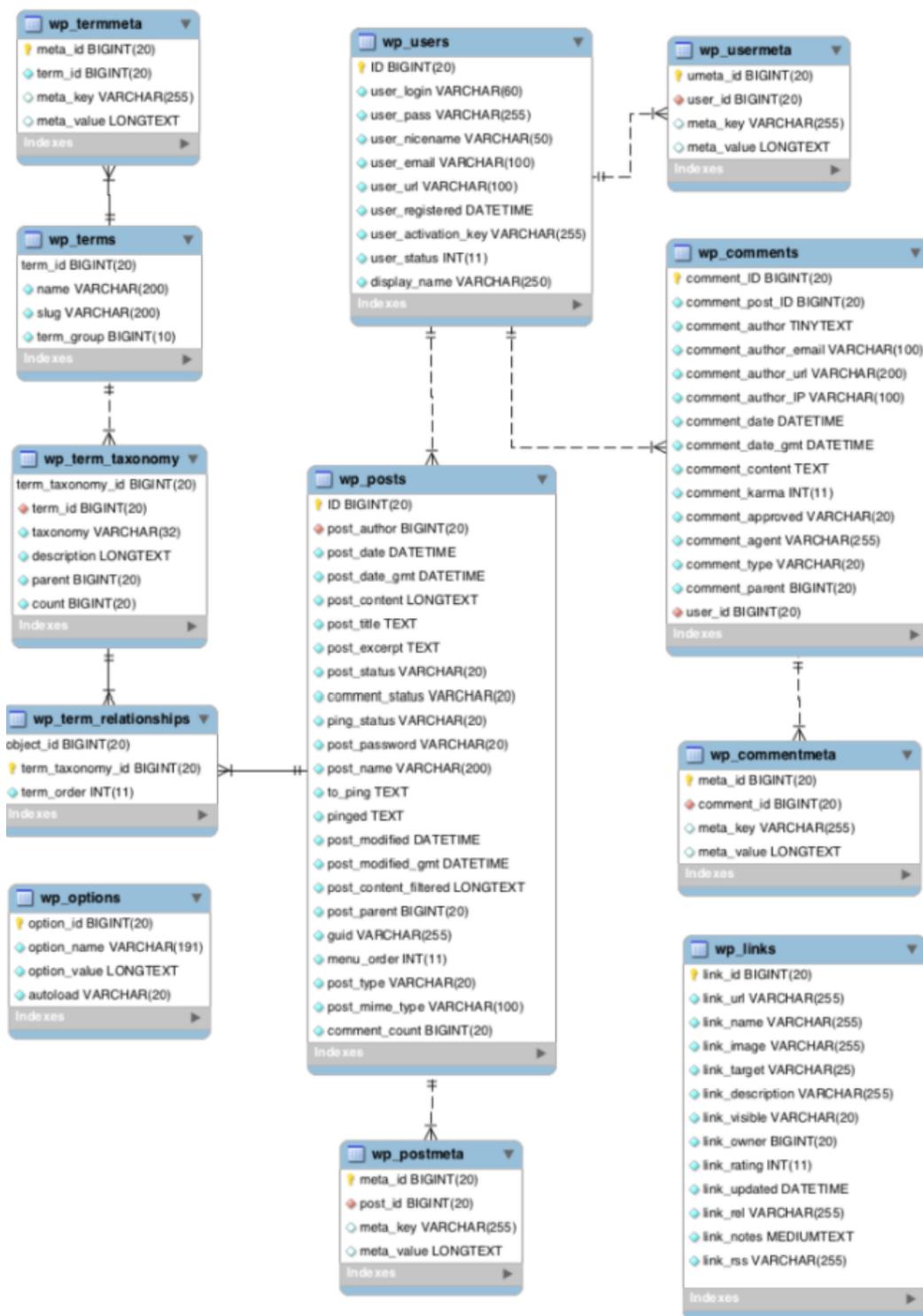
Por último, y haciendo foco en la naturaleza de los datos que serán parte de dicho tratamiento, resultará interesante poder contar con opiniones de los diferentes actores (docentes, alumnos, colaboradores, entidades, etc.) que no sean limitadas por respuestas normalizadas. Esto es, dar la posibilidad de redactar distintas opiniones sobre lo vivido de forma libre. A partir de allí, se podrá realizar minería textual, permitiendo el descubrimiento

de información no explícita en el texto. Esto brindará la posibilidad de incorporar nuevas perspectivas que contribuyan a la continua evolución del modelo educativo STEAM.

## 8. Anexos

### 8.1 Anexo I – Diagramas Entidad Relación

**Figura 24**  
*DER inicial de la plataforma virtual*



*Nota: DER realizado a partir de articulación de tablas inicial de la plataforma*



## 8.2 Anexo 2 – Estructura De Campos De Cada Dataset

### Dataset: Cápsulas y Actividades

En la Tabla 6, se puede observar la estructura de campos final que se utilizó para almacenar los datos referidos a las cápsulas y sus actividades.

**Tabla 6**

*Estructura de campos: Dataset cápsulas y actividades*

Dataset cápsulas y actividades			
id_capsula	desc_capsula	id_capsula_patrocinador	id_patrocinador
nombre_patrocinador	calle	nro	cp
localidad_patrocinador	prov_patrocinador	mail_patrocinador	id_kit_elemento_kit
cod_kit	id_elemento_kit	desc_elemento_kit	id_capsula_voluntario
id_voluntariocapacitador	id_categoria_voluntarioCapacitador	tipo	nombre_voluntarioCapacitador
Apellido	provincia_voluntarioCapacitador	localidad_voluntarioCapacitador	mail_voluntarioCapacitador
desc_categoria_voluntarioCapacitador	id_tipo_voluntarioCapacitador	descripcion	id_actividad_capsula
id_actividad	titulo	Resumen	nivel_dificultad
edades_recomendadas	tiempo_desarrollo	id_actividad_keywords	id_keyword
desc_keyword	id_actividad_entidad_organizadora	id_entidad_organizadora	Descripcion

### Dataset: Post Implementación

En la Tabla 7, se puede observar la estructura de campos final que se utilizó para almacenar los datos referidos a la implementación de cada actividad.

**Tabla 7**

*Estructura de campos: Dataset Post Implementación*

Dataset Post Implementación			
id_evento	id_capsula	desc_capsula	cod_kit

id_actividad	titulo_actividad	resumen	nivel_dificultad
edad_recomendadas_desde	tiempo_desarrollo_HD	id_actividad_capsula	fecha
duracion_minutos	organizador	coordinador	id_comentario
id_persona_comentario	nya	f_nac	id_establecimiento
parecer	agregado	dificultades	experiencia_favorita
desc_establecimiento	id_rol	desc_rol	-

### **Dataset: Tracking de Usuarios**

En la Tabla 8, se puede observar la estructura de campos final que se utilizó para almacenar los datos referidos a la navegación de los usuarios en la plataforma.

**Tabla 8**

*Estructura de campos: Dataset Tracking de Usuarios*

<b>Dataset Tracking de Usuarios</b>	
id_log	ip
fecha_hora	url
elemento_completo	pais
region	ciudad
elemento_principal	-

## **9. Acrónimos**

CI2S LABS                      Computational Intelligence & Information Systems Labs

Equipo de investigadores abocados a la investigación y desarrollo referida a la inteligencia computacional.

DER                              Diagrama Entidad Relación

Modelo gráfico utilizado para representar las entidades en una base de datos

EM                                Expectation Maximization

Algoritmo de tipo agrupamiento utilizado en minería de datos.

IA                                Inteligencia Artificial

Software desarrollado para imitar los procesos cognitivos del ser humano.

IEEE                      Institute of Electrical and Electronics Engineers

Comunidad de orden mundial para el desarrollo de áreas técnicas.

IP                          Internet Protocol (dirección)

Una dirección IP es utilizada para la identificación de un artefacto en una red IP.

MHS                      Museo Histórico Sarmiento

Museo dedicado a la historia argentina.

SCA                      Sociedad Científica Argentina

Comunidad científico tecnológica argentina para la promoción y divulgación de las ciencias.

STEAM                    Ciencia Tecnología Ingeniería Arte Matemáticas

Modelo educativo que acerca al alumno a las ciencias a través de las cinco disciplinas mencionadas

UADER FCyT            Universidad Autónoma de Entre Ríos - Facultad de Ciencia y Tecnología.

Institución educativa de educación superior creada en el año 2000 en la provincia de Entre Ríos

UCSE DASS            Universidad Católica Santiago del Estero - Departamento Académico San Salvador

Institución educativa de educación superior, laica religiosa creada en el año 1960 en la provincia de Santiago del Estero

UNESCO                Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

Organismo que promueve la seguridad y la paz a través de la educación, la ciencia, la cultura y las comunicaciones.

## 10. Referencias

- Alfredo, M. (2020). Información e inteligencia artificial. *Ápeiron. Estudios de Filosofía.*, 12, 73-82.
- Aluja, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial. *QUESTIÓ*, 25(3), 481-482.
- Autodesk, Inc;. (2022). *Autodesk Tinkercad*. <https://www.tinkercad.com/>
- Calafell, G., Bonil, J., & Junyent, M. (2014). Rizoma, an R&D&i proposal in education for Sustainability. *Procedia - Social and Behavioral Sciences*, 551 - 558.
- Cardozo-Acosta, L. A., Álvarez-Londoño, L., Padilla-Muñoz, A., Olis-Barreto, I. M., Sierra-Osopina, O. E., Fajardo-Tiriath, D., . . . Reyes-Bohórquez, P. (2015). Innovación social desde la universidad. *Universidad y sociedad*.
- Chango Sailema, W. G. (2022). *Predicción de rendimiento académico en aprendizaje combinado mediante datos multimodales y técnicas de fusión y minería de datos*. Córdoba: UCOPress.
- Chango, W., Lara, J., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. *WIREs Data Mining Knowl Discovery*.
- Chen, X., Zou, D., Xie, H., & Wang, F. L. (2021). Past, present, and future of smart learning: a topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, 18(1), 1-29.
- Choez Franco, A. (2022). *ANÁLISIS DE LAS CARACTERÍSTICAS DE LOS TIPOS DE ALGORITMOS DE CLUSTERING EN EL APRENDIZAJE NO SUPERVISADO*.
- Christenson, J. (2011). Ramaley coined STEM term now used nationwide. *Winona Daily News*.
- Chung, C. (2014). Integrated STEAM education through global robotics art festival (GRAF). *IEEE Integrated STEM Education Conference*, (págs. 1-6).
- Cisco;. (2021). *Cisco. IP Addressing and Subnetting for New Users*: [https://www.cisco.com/c/es\\_mx/support/docs/ip/routing-information-protocol-rip/13788-3.html](https://www.cisco.com/c/es_mx/support/docs/ip/routing-information-protocol-rip/13788-3.html)

- Cohen, W. W., & Page, C. D. (1995). Learnability in inductive logic programming: Methods and results. *New generation computing*, 13(3-4), 369-409.
- Compagnucci, L., & Spigarelli, F. (2020). The Third Mission of the university: A systematic literature review on potentials and constraints. *Technological Forecasting & Social Change*, 161(120284).
- Cuevas Palma, A. F. (2010). *Web Usage Mining aplicado al estudio del comportamiento*. PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO. FACULTAD DE INGENIERÍA.
- Di Rienzo, J. A., Casanoves, F., Balzarini, M. G., Gonzalez, L., Tablada, M., & Robledo, C. W. (2020). InfoStat versión 2020. Centro de Transferencia InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. <https://www.infostat.com.ar/>
- Dima, L., Urdaneta, R., Sosa, G., Esperón, G., & López De Luise, D. (2012). Búsqueda de la calidad del aprendizaje mediante herramientas de inteligencia artificial. *Congreso TRIC V organizado por IEEE CIS Argentina*.
- EDS Robotics. (2021). *EDS Robotics*. Tipos de aprendizaje en la Inteligencia Artificial: <https://www.edsrobotics.com/blog/tipos-aprendizaje-inteligencia-artificial/>
- Epitech. (2022). *epitech-it*. <https://www.epitech-it.es/caracteristicas-era-informacion/>
- Flowers, J. (2019). Teach Wood Science: As with Project-Based Instruction, Experimentation in the Classroom or Educational Laboratory Can Be Weak If Teachers or Students See the Goal as Completing the Activity Rather Than Focusing on the Underlying Learning. *Technology and Engineering Teacher*, 29-34.
- Forgan, S. (2005). *Museos e Historia de la Ciencia*. Ed. Samuel J. M. M. Alberti.
- Free Software Foundation, Inc. (2022). *El sistema operativo GNU*. <https://www.gnu.org/licenses/licenses.es.html>
- González, R. (2021). *Técnicas de Minería de Datos Aplicadas al Análisis de Rotación de Artículos*.
- Goodman, D., & Keene, R. (1997). Man versus machine: Kasparov versus Deep Blue. *H3 Publications*.
- Heider, H., & Sánchez, E. (2008). *Aplicación en minería de datos*.

- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la investigación* (5 ed.). México: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.
- Heynz Roberth, G., & Ticona Gonzáles, U. A. (2020). Clustering, mediterraneidad y comercio internacional: aplicación empírica de los algoritmos Partitioning Around Medoids y K-means. *Scientific Electronic Library Online*, 95(129).
- Hogg, R., Tanis, E., & Zimmerman, D. (2021). *Probability and Statistical Inference* (Vol. 10th Edition).
- Hurtado, B. H., & Gómez, N. R. (2022). Apuntes estadísticos sobre el procesamiento de datos en un trabajo de investigación cuantitativo de orden organizacional. *Documentos De Trabajo ECACEN*, 23–35.  
<https://hemeroteca.unad.edu.co/index.php/working/article/view/4705/5517>
- IEEE;. (2022). *IEEE Games Technical Committee*. Quiénes somos: <https://r9.ieee.org/ar-cis-gaming/acerca-de-gtc/>
- Imoize, A., Idowu, D., & Bolaji, T. (2019). A Brief Overview of Software Reuse and Metrics in Software Engineering. *World Scientific News. An International Scientific Journal*.
- Institute of Electrical and Electronics Engineers. (s.f.). <http://www.ieee.org.ar/index.asp>
- Kemble, J. (2018). *From Breadcrumbs to Breadboards: How to Spearhead Effective*. Techlearning.: <https://www.techlearning.com/tl-advisor-blog/how-tospearhead->
- Leiva, S. M., López De Luise, D., González, V., Draier, E., Ramírez, S., & Salina, C. (2022a). Comunicación y museos: Proyectos y desafíos en torno a las mediaciones culturales y educativas. *Avatares de la Comunicación y la Cultura. U.B.A.*(23).
- Leiva, S. M., López De Luise, D., Ramírez, S., & Salina, C. (2022b). Sistemas inteligentes en la innovación educativa. Plataforma de cápsulas STEAM. *Jornada sobre innovación educativa*. Buenos Aires.  
<https://www.youtube.com/watch?v=3pDndr5LhJM&t=15102s>
- López De Luise, D. (2020). *Museo Virtual STEAM*. Tesis doctoral FCE UBA.

- López De Luise, D., & Ruiz Tabarez, É. (2020b). Factores relevantes en la educación STEAM: Desarrollo de Métricas y Modelos Automatizados. *IEEE Congreso Bienal de Argentina (ARGENCON)*.
- López De Luise, D., & Ruiz Tabarez, É. A. (2020a). Evaluación STEAM a partir de métricas específicas. *SCA, IEEE Games Technical Committee, UAI – CAETI*.
- López De Luise, D., & Ruiz, É. (2018). Modelo de predicción de deserción de alumnos. *13avo Congreso Colombiano de Computación (13CCC)*.
- Martin Ramos, R., Ramos Palmero, R., Grau Ávalos, R., & García Lorenzo, M. (2012). Aplicación de métodos de selección de atributos para determinar factores relevantes en la evaluación nutricional de los niños. *Gaceta Médica Espirituana*, 9(1).
- Medrano, F., & Barriento, V. (2021). Web Mining y Text Mining: enfoques avanzados para analizar el contenido de grandes cantidades de información. *XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021)*, (págs. 245-249). Chilecito. <http://sedici.unlp.edu.ar/bitstream/handle/10915/120104/Ponencia.pdf-PDFA.pdf?sequence=1&isAllowed=y>
- Microsoft. (2022). *Microsoft Soporte*. Access SQL: conceptos básicos, vocabulario y sintaxis: <https://support.microsoft.com/es-es/office/access-sql-conceptos-b%C3%A1sicos-vocabulario-y-sintaxis-444d0303-cde1-424e-9a74-e8dc3e460671#bm1>
- Moreira-Mora, T. E. (2007). Perfil sociodemográfico y académico de estudiantes en deserción del sistema educativo. *Actualidades en Psicología*, 21(108), 21(108), 145-165.
- Neressini, F., & Pellegrini, G. (2008). Evaluating public communication of science and technology. *Handbook of public communication of science and technology*.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2020). *Education: From disruption to recovery*.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2021). <https://es.unesco.org/news/necesaria-educacion-steamh-cultivar-pensamiento-y-habilidades-transformadoras-innovadoras-y>
- Ortega y Gasset, J. (2000). Meditaciones de la técnica y otros ensayos sobre ciencia y filosofía. *Revista de filosofía*.

Peirce, C. (1931-1935). 5, 484.

Perotta, C., Featherstone, G., Aston, H., & Houghton, E. (2013). Game-based Learning: Latest Evidence and Future Directions. *NFER Research Programme: Innovation in Education, National Foundation for Educational Research*.

R, Project;. (2022). *R Project*. The R Project for Statistical Computing: <https://www.r-project.org/>

Redacción España;. (2020). *B12 Admark*.

Rodríguez Manzanares, C. (2018). Análisis de los Métodos de Predicción Aplicados a los Desvíos en el Sistema Eléctrico Ibérico. Sevilla, España.

Russel, S., & Norvig, P. (2004). *ARTIFICIAL INTELLIGENCE: A MODERN APPROACH*. PEARSON EDUCACIÓN, S.A.

Saint-Pierre, T. d. (2018). *Fundación Chile*. <https://fch.cl/noticianoticia-destacadanoticia-antigua/que-es-la-inteligencia-artificial-y-como-se-puede-aplicar-en-la-mineria/>

Scratch;. (2019). *Scratch*. <https://scratch.mit.edu/about>

Secundo, G., De Beer, C., Schutte, C. S., & Passiante, G. (2017). Mobilising intellectual capital to improve European universities' competitiveness: The technology transfer offices' role. *Journal of Intellectual Capital*, 18(3), 607-624.

Shannon, C., & Weaver, W. (1949). The mathematical theory of communication. *Bell System Technical Journal* , 379-423.

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, 51(5).

Simon, H. A. (1957). *Herbert A. Simon and the concept of rationality: Boundaries and procedures*.

Sociedad Científica Argentina. (2021). <http://cientifica.org.ar/historia/>

Sociedad Científica Argentina. (2021). <http://cientifica.org.ar/institutos/ciencia-para-la-innovacion/>

- Soriano Tolosa, A. (2013). *Fusión de datos estadísticamente dependientes en sistemas de detección*. Valencia.
- Tabuena, B., Serrano-Iglesias, S., Martín, A. C., Villa-Torrano, C., Dimitriadis, Y., Asensio-Pérez, J., & Kloos, C. D. (2021). Affordances and Core Functions of Smart Learning Environments: A Systematic Literature Review. *IEEE Transactions on Learning Technologies*, 14(2), 129-145.
- Tellado, F. (2021). *Ayuda Wordpress*. <https://ayudawp.com/las-tablas-de-la-base-de-datos-de-wordpress-cuales-son-para-que-sirve-cada-una/>
- TIBCO. (2022). *Tibco*. ¿Qué es el aprendizaje supervisado?: <https://www.tibco.com/es/reference-center/what-is-supervised-learning#:~:text=El%20aprendizaje%20supervisado%20es%20una,de%20manera%20expl%C3%ADcita%20d%C3%B3nde%20buscar.>
- Troncoso Lora, A. (2017). Selección de atributos. Series temporales. Sevilla, España.
- Vargas Sabadías, A. (1995). *Estadística descriptiva y diferencial*. Servicios de publicaciones de la Universidad de Castilla - La Mancha.
- Vintimilla, C., Astudillo, F., Severeyn, E., & Encalada, L. (2017). Agrupamiento de K-medias para estimación de insulino-resistencia en adultos mayores de Cuenca. *Maskana*.
- Virvou, M., Tsihrintzis, G., & Jain, L. (2022). *Advances in selected artificial intelligence areas: World outstanding women in artificial intelligence*. Springer Nature.
- Winsor, M. P., Bowler, P. J., & Pickstone, J. V. (2008). Early Modern Science. *The Cambridge History of Science*, 3.
- Witten, I. H., & Frank, E. (2005). *Data Mining. Practical Machine Learning Tools and Techniques*.
- Wordpress. (2022). *Wordpress*. <https://wordpress.com/es/>
- Wordpress;. (s.f.). *Wordpress*. Home Page: <https://wordpress.com/es/>
- Xueyun, J., & Zihui., C. (2010). A new way to software engineering education. *2010 International Conference on Educational and Information Technology*.

Zawieska, K., & Duffy, B. R. (2015). The social construction of creativity in educational robotics. *In Progress in Automation, Robotics and Measuring Techniques* (pp. 329-338). Springer, Cham.