

TRABAJO FINAL INTEGRADOR
ESPECIALIZACIÓN EN INGENIERÍA
DE SISTEMAS DE INFORMACIÓN

Título:
“Aplicación de la ciencia de datos en el
modelamiento hidrológico”

Autor: María Claudia Jurado
Tutora: Paola Britos

Buenos Aires – 02/2022

RESUMEN

El agua es uno de los recursos naturales esenciales para la vida del ser humano y de todos los seres vivos que habitan nuestro planeta, su cuidado y buen uso es un tema que siempre preocupó ya que a medida que la población mundial crece, se incrementa también la demanda de agua. En este sentido, la Hidrología cumple un papel muy importante a la hora de la planificación de los recursos hídricos y apoya a otras áreas como ser la ingeniería, meteorología, agricultura, etc. Actualmente el comportamiento de las cuencas hidráulicas es descrito generalmente por modelos hidrológicos que son diseñados en forma manual por los ingenieros, es muy poca la utilización de herramientas informáticas como apoyo a este trabajo.

Por este motivo, en el presente trabajo se pretende establecer el estado de situación actual del modelamiento hidrológico en cuencas semi-urbanas específicamente de la provincia de Buenos Aires y plantear como un aporte de investigación de la utilización de nuevas tecnologías en las ciencias hidrológicas un modelo de procesos de explotación de información desarrollado con tecnologías de sistemas inteligentes.

PALABRAS CLAVE

Explotación de la información. Minería de datos. Patrones. Modelo hidrológico.

Cuenca.

Índice de Contenidos

Índice de Contenidos	2
Índice de Tablas.....	4
1. INTRODUCCIÓN	6
2. NATURALEZA DE LA INVESTIGACIÓN	9
2.1. DEFINICIÓN DEL PROBLEMA	9
2.2. OBJETIVOS DEL TRABAJO DE ESPECIALIDAD	12
2.2.1. OBJETIVO GENERAL	12
2.3. OBJETIVOS ESPECÍFICOS.....	12
2.3.1. ALCANCE	13
2.4. METODOLOGÍA	13
3. DESARROLLO	14
3.1. ÁMBITO DE LA HIDROLOGÍA	14
3.1.1. CONCEPTO DE HIDROLOGÍA	14
3.1.2. CICLO HIDROLÓGICO.....	16
3.1.3. COMPONENTES DEL CICLO HIDROLÓGICO	19
3.1.3.1. PRECIPITACIÓN.....	19
3.1.3.2. EVAPORACIÓN Y EVAPOTRANSPIRACIÓN.....	20
3.1.3.3. INFILTRACIÓN, PERCOLACIÓN Y AGUAS SUBTERRÁNEAS	21
3.1.3.4. ESCURRIMIENTO SUPERFICIAL, CUENCA HIDROGRÁFICA Y AGUAS SUPERFICIALES	22
3.1.4. ANÁLISIS HIDROLÓGICO	23
3.1.4.1. RECOLECCIÓN DE DATOS	23
3.1.4.2. ORDENAMIENTO DE DATOS.....	23
3.1.4.3. EXTENSIÓN DEL REGISTRO.....	24
3.1.4.4. ANÁLISIS DE LOS DATOS	24
3.1.4.5. VERIFICACIÓN DE LA CALIDAD DE LOS DATOS.....	25
3.1.5. SISTEMAS HIDROLÓGICOS	25
3.1.6. MODELAMIENTO HIDROLÓGICO.....	27
3.1.7. INFORMACIÓN HIDROLÓGICA EN ARGENTINA	27
3.2. ÁMBITO DE LAS CIENCIAS DE DATOS	29
3.2.1. DEFINICIÓN DE CIENCIA DE DATOS.....	29
3.2.2. INTELIGENCIA DE NEGOCIOS	30
3.2.3. MINERÍA DE DATOS	31
3.2.3.1. EL PROCESO DE DESCUBRIMIENTO DEL CONOCIMIENTO.....	33
3.2.3.2. TÉCNICAS DE MINERÍA DE DATOS	35
3.3.1. EXPLOTACIÓN DE LA INFORMACIÓN	37
3.3.2. PROCESOS DE EXPLOTACIÓN DE LA INFORMACIÓN	38
3.3.2.1. DESCUBRIMIENTO DE REGLAS DE OMPORTAMIENTO	39
3.3.2.2. DESCUBRIMIENTO DE GRUPOS.....	39
3.3.2.3. PONDERACIÓN DE INTERDEPENDENCIA DE ATRIBUTOS	39
3.3.2.4. DESCUBRIMIENTO DE REGLAS DE PERTENENCIA A GRUPOS	40

3.3.2.5.	PONDERACIÓN DE REGLAS DE COMPORTAMIENTO O DE LA PERTENENCIA A GRUPOS	40
3.3.3.	METODOLOGÍAS DE EXPLOTACIÓN DE INFORMACIÓN	40
3.3.3.1.	CRISP-DM	41
3.3.3.2.	SEMMA	42
3.3.3.3.	KDD	43
3.3.3.4.	TDSP.....	44
3.3.3.5.	ASUM.....	46
4.	ESTADO DEL ARTE.....	47
4.3.	APLICACIÓN DE LA INFORMATICA EN LA HIDROLOGÍA	47
4.5.	APLICACIÓN DE LAS CIENCIA DE DATOS EN LA HIDROLOGÍA	52
5.	CONCLUSIONES	54
6.	REFERENCIAS	56

Índice de Tablas

Tabla 1. Cantidades Estimadas de Agua en la Tierra	15
Tabla 2. Etapas del proceso de descubrimiento del conocimiento KDD	35
Tabla 3. Clasificación de las técnicas de data mining	36

Índice de Figuras

Figura 1. Temas relacionados con la hidrología	16
Figura 2. Ciclo hidrológico en la naturaleza	17
Figura 3. Pluviómetro	19
Figura 4. Pluviógrafo	20
Figura 5. Medidores de nivel de agua subterránea	22
Figura 6. Representación de un Sistema Hidrológico	26
Figura 7. Estaciones de Red Hidrológica Nacional	28
Figura 8. Disciplinas que componen la ciencia de datos	30
Figura 9. Información utilizada para herramientas de minería de datos	33
Figura 10. Etapas de la metodología CRISP-DM	41
Figura 11. Ciclo de etapas de la metodología SEMMA	42
Figura 12. Etapas de la metodología KDD	44
Figura 13. Ciclo de vida de la Metodología TDSP	46
Figura 14. Fases de la tecnología ASUM	47
Figura 15. Enfoque de la informática en ciclo hidrológico	49

1. INTRODUCCIÓN

La hidrología en su concepto más simple es la ciencia que estudia la distribución, cuantificación y utilización de los recursos hídricos que están disponibles en el globo terrestre. Estos recursos se distribuyen en la atmósfera, la superficie terrestre y las capas de suelo (Silva Medina, 2003). El ciclo del agua implica un cambio continuo de grandes masas de agua de un estado físico a otro y su transporte de un lugar a otro. Al volumen de agua que se transporta de un depósito a otro a lo largo de un año se denomina Balance Hídrico Global. La ecuación de continuidad, o de balance hidrológico, es la ley más importante en Hidrología, y aunque su expresión es muy simple, la cuantificación de sus términos es normalmente complicada, principalmente por la falta de mediciones directas en campo y por la variación espacial de la evapotranspiración, de las pérdidas profundas (a acuíferos) y de las variaciones del agua almacenada en una cuenca (Ordoñez Gálvez, 2011). La predicción de flujos de agua en una cuenca requiere el uso de modelos, los cuales deben ser evaluados para diferentes condiciones ambientales... Los modelos hidrológicos buscan capturar los mecanismos físicos de generación de escorrentía superficial y la recarga a los acuíferos subterráneos a partir de la evaporación y la evapotranspiración (Ocampo & Vélez, 2013).

Se entiende como ciencia de datos al proceso de descubrir patrones interesantes y conocimiento de grandes cantidades de datos. Las fuentes de datos pueden incluir bases de datos, la web, otros repositorios de información o datos que se transmiten a un sistema de forma dinámica (Han, Kamber, & Pei, 2012). Durante ese proceso, se aplican técnicas y herramientas para extraer y presentar el conocimiento implícito, previamente

desconocido, potencialmente útil y comprensible. Uno de los objetivos de la minería de datos es que mediante el uso de sus herramientas se pueden predecir futuras tendencias y comportamientos, permitiendo tomar decisiones proactivas y conductivas por un conocimiento a partir de la información (Herrera Prado, Uribe Agundis, & Ruiz Luna, 2015). Los Sistemas Inteligentes constituyen el campo de la Informática en el que se estudian y desarrollan algoritmos que implementan algún comportamiento inteligente y su aplicación a la resolución de problemas prácticos... Esto resulta una alternativa de solución a problemas que no pueden ser resueltos mediante algoritmos tradicionales, entre los cuales podemos mencionar especificación de condiciones asociadas a diagnósticos técnicos o clínicos, identificación de características que permitan reconocimiento visual de objetos, descubrimiento de patrones o regularidades en estructuras de información (en particular en bases de datos de gran tamaño), entre otros (Britos, Procesos de Explotación de Información Basados en Sistemas Inteligentes, 2008).

Con el desarrollo de la tecnología de base de datos, se están utilizando diversas técnicas de análisis de datos y herramientas de extracción de conocimientos para recopilar datos de series de tiempo en organizaciones científicas y comerciales... La naturaleza y la calidad de los datos recopilados en hidrología son extremadamente importantes y todas las características de dichos datos deberían ser las mejores analizadas. La minería de datos en hidrología depende de los datos hidrometeorológicos, que generalmente toman la forma de series temporales. Las series de tiempo hidrológico son conjuntos de varios valores de registros de datos hidrológicos que varían con el tiempo (Mishra, Dwiveki, Saravanan, & Pathak, 2013).

En el presente trabajo, como un aporte de investigación se pretende realizar un estudio sobre la realidad actual del proceso de modelamiento hidrológico, y la utilización de herramientas tecnológicas como apoyo al mismo, y más específicamente herramientas y modelos de procesos de la ciencia de datos. En base a esta situación, se enmarcan los objetivos generales y específicos.

Posteriormente, se pretende desarrollar un estado del arte para poder determinar si existe un área de vacancia de la aplicación de herramientas, técnicas y procesos de ciencias de datos en el modelamiento hidrológico.

2. NATURALEZA DE LA INVESTIGACIÓN

En esta sección del trabajo, primeramente, se describe el problema a ser resuelto (Sección 2.1). A continuación, se define el objetivo principal (sección 2.2) y los objetivos específicos (sección 2.2) a alcanzar para resolver el problema descrito. Luego se realiza una descripción detallada de la metodología (sección 2.4) a utilizar en el marco de la presente investigación. Por último, se especifica el alcance del presente trabajo de fin de especialidad (sección 2.4).

2.1. DEFINICIÓN DEL PROBLEMA

La hidrología en su concepto más simple es la ciencia que estudia la distribución, cuantificación y utilización de los recursos hídricos que están disponibles en el globo terrestre. Los proyectos de estudios hidrológicos recolectan y procesan información histórica obteniendo resultados que producen información sobre distintos aspectos como ser características climatológicas, capacidad de las fuentes que suministran caudales, eventos extremos como sequías o inundaciones. Es por este motivo que la hidrología se considera una ciencia de uso intensivo de datos. Los métodos de modelado dirigidos a mejorar la comprensión o la capacidad predictiva pueden requerir grandes cantidades de datos de observación en el modelo (Spate, Croke, & A.J., Data Mining in Hydrology, 2016). Cabe destacar que esta información se registra constantemente en repositorios oficiales de datos, mismos que pueden ser utilizados por la Minería de Datos. La Minería de Datos, es una de las técnicas más utilizadas actualmente para analizar y extraer la información útil de grandes bases de datos. Se fundamenta en varias disciplinas, como la

estadística, las técnicas de visualización de datos, los sistemas para tomas de decisión, el aprendizaje automático o la computación paralela y distribuida, con la finalidad de extraer patrones, describir tendencias, predecir comportamientos y sobre todo, producir beneficios a aquellas entidades que posean amplias bases de datos de, aparentemente, escasa utilidad (Herrera Prado, Uribe Agundis, & Ruiz Luna, 2015)

Uno de los problemas inminentes entre las comunidades científicas y técnicas es la dificultad en el acceso a la información para poder realizar diferentes estudios, el posterior procesamiento y la mantención de los datos tanto para la vigilancia como para el pronóstico del tiempo, la vigilancia hidrológica, la predicción agrometeorológica o para reducir el riesgo climático. Por este motivo (Aguirre-Munizaga, Gomez, Aviles, Vasquez, & Recalde-Coronel, 2016) propusieron un modelo para la gestión de la información meteorológica a través de la computación en la nube utilizando la tecnología GPRS (General Packet Radio Service) transferencia de datos entre estaciones meteorológicas automáticas para mejorar el monitoreo y predicción de la atmósfera y el comportamiento de las aguas continentales.

En los últimos años surgieron estudios de aplicación de la tecnología en la ciencia hidrológica como apoyo a los científicos de esta área al momento de realizar modelamiento hidrológico y predicciones. (Spate, Croke, & A.J., Data Mining in Hydrology, 2016) realizaron un estudio comparativo de técnicas de minería de datos y su aplicación en hidrología con el objetivo de proporcionar una visión general rápida de algunas técnicas de extracción de datos que pueden ser útiles para los investigadores en la comunidad de modelos. Técnicas como Series de Tiempo fueron aplicadas con el

objetivo de desarrollar una aplicación de minería de datos utilizando tecnología de información moderna para descubrir la información o los patrones ocultos detrás de los datos hidrológicos ... y también buscar una nueva forma de cumplir con los requisitos hidrológicos (Satanand, C., V. K., & K. K., 2013). Un nuevo planteamiento se desarrolló en (Herrera Prado, Uribe Agundis, & Ruiz Luna, 2015) donde se utilizaron y se probaron algoritmos de predicción, basados en los 3 métodos de aprendizaje supervisado del tipo de: a) Máquina de Soporte Vectorial, b) Regresión Lineal y c) Perceptrón Multicapas utilizando datos hidrológicos de una cuenca mexicana para demostrar su hipótesis.

También se realizaron estudios más específicos sobre la optimización en el aprovechamiento del agua subterránea como un recurso esencial en zonas donde existe escasez hídrica. Para esto, se identificaron e investigaron nuevos enfoques utilizando métodos estadísticos y geoestadísticos para la optimización de la red de monitoreo de aguas subterráneas con el fin de mejorar las estrategias de monitoreo. Se analizaron los factores que influyen en los métodos de optimización de monitoreo de aguas subterráneas y se integraron métodos nuevos y mejorados mediante la prueba de aquellos con la optimización de la red de monitoreo de aguas subterráneas en sitios contaminados (Krishna Thakur, 2015). Posteriormente, en (Nourani, Davanlou Tajbakhsh, & Molajou, 2018) se realizó un estudio, como una estrategia novedosa, se introdujo un nuevo modelo de Wavelet-M5 y se aplicó a los datos hidrológicos de una cuenca iraní y se verificó que la precisión del modelo propuesto se mejoró debido al beneficio de las series de tiempo descompuestas en lugar del único árbol modelo M5.

2.2. OBJETIVOS DEL TRABAJO DE ESPECIALIDAD

Los objetivos del presente trabajo de especialidad se dividen en un objetivo general a alcanzar (sección 2.2.1) y los objetivos específicos (sección 2.2.2) que delimitan los pasos a seguir para alcanzar el objetivo general. Por último, el alcance de dichos objetivos específicos (sección 2.2.3).

2.2.1. OBJETIVO GENERAL

En el presente trabajo se pretende realizar un relevamiento del modelado del ciclo hidrológico y la aplicación actual de herramientas tecnológicas, y más específicamente de la ciencia de datos como apoyo para su desarrollo.

2.3. OBJETIVOS ESPECÍFICOS

Los objetivos específicos que se pretende alcanzar son:

- Identificar y obtener conocimientos acerca del proceso del modelamiento del ciclo hidrológico y el definir las técnicas convencionales para su desarrollo, a través de revisiones sistemáticas de la literatura.
- Describir las herramientas de la ciencia de datos y los procesos de explotación de información. Obtener una definición detallada acerca de las principales herramientas que se están utilizando actualmente, a través de revisiones sistemáticas de la literatura.

- Identificar la utilización de herramientas informáticas y más específicamente referidas a la ciencia de datos en el modelamiento de ciclo hidrológico, a través de revisiones sistemáticas de la literatura.

2.3.1. ALCANCE

El presente trabajo tiene como alcance obtener conocimiento acerca del proceso del modelamiento del ciclo hidrológico, para luego determinar cómo se utilizan herramientas informáticas como apoyo en su desarrollo. Por último, profundizar sobre la aplicación de la ciencia de datos en el proceso de modelado del ciclo hidrológico, determinar si actualmente se están utilizando y en qué medida.

2.4. METODOLOGÍA

Este trabajo se guiará por la metodología post-positivista ya que se planteará inicialmente una hipótesis y a medida que se desarrolle el trabajo, se verificará la misma con los resultados obtenidos. El post-positivismo es un paradigma de investigación científica cuantitativo que tiene como objeto explicar el fenómeno estudiado para luego predecirlo y controlarlo. Al utilizar este paradigma, el investigador puede formar parte del fenómeno de interés, el objeto de estudio influencia al investigador y viceversa, la teoría e hipótesis que sustenta investigación influye en el desarrollo de la misma (Ramos, 2015). El propósito del paradigma post-positivista es obtener medidas cuantitativas que permitan entender y predecir la respuesta social ante el problema (Catalán Vázquez & Jarillo Soto, 2010).

3. DESARROLLO

En este capítulo se desarrollan las definiciones principales que nos permitirán establecer posteriormente un estado del arte de la utilización de técnicas de minería de datos aplicados al modelamiento hidrológico. En el apartado 3.1 se describen los principales conceptos hidrológicos. En el apartado 3.2 se define el ámbito de la ciencia de datos. Posteriormente, en el capítulo 4, desarrollaremos un análisis de la aplicación de herramientas informáticas en el modelamiento del ciclo hidrológico, precisamente el área de estudio de la hidroinformática. Y por último, realizaremos un estado del arte de la aplicación actual de la ciencia de datos en el área hidrológica para determinar si existe un área de vacancia en este campo.

3.1. ÁMBITO DE LA HIDROLOGÍA

3.1.1. CONCEPTO DE HIDROLOGÍA

La hidrología es la ciencia destinada al estudio del agua sobre la tierra, su ocurrencia, distribución y circulación, sus propiedades químicas y físicas y su reacción con el medio ambiente, incluyendo su relación con los seres vivos; en la *Tabla 1* se encuentran detalladas las cantidades estimadas de agua sobre la tierra. La hidrología es una ciencia interdisciplinaria, ya que se vincula con otras ramas de la ciencia como ser la física, química, geología, mecánica de fluidos, matemáticas, estadística (Gallardo Terrones, 1994).

Agua	Área (10 ⁶ km ²)	Volumen (km ³)	% Agua Total	% Agua Dulce
Océanos	361,3	1.338.000.000,00	96,5	
Agua Subterránea Dulce	134,8	10.530.000	0,76	30,1
Agua Subterránea Salada	134,8	12.870.000	0,93	
Humedad de Suelo	82	16.500	0,0012	0,05
Hielo Polar	16	24.023.500	1,7	68,6
Hielo no polar y nieve	0,3	340.600	0,025	1
Lagos Dulces	1,2	91.000	0,007	0,26
Lagos Salinos	0,8	85.400	0,006	
Pantanos	2,7	11.470	0,0008	0,03
Ríos	148,8	2.120	0,0002	0,006
Agua Biológica	510	1.120	0,0001	0,003
Agua Atmosférica	510	12.900	0,001	0,04
Agua Total	510	1.385.984.610	100	
Agua Dulce	148,8	35.029.210	2,5	100

Tabla 1. Cantidades Estimadas de Agua en la Tierra (Gallardo Terrones, 1994)

Según (Şen, 2015) la hidrología es la ciencia que estudia la ocurrencia del agua, su movimiento y transporte a través de la naturaleza. Sus aplicaciones proporcionan leyes básicas, ecuaciones y algoritmos, procedimientos y modelos para usos prácticos de la actividad humana. También cumple las aplicaciones prácticas y de campo utilizando cálculos simples y racionales que conducen a una gestión adecuada. En la *Figura 1* se muestran las aplicaciones de la hidrología desde el punto de vista de la ingeniería y la ciencia de la tierra.

Destacando su importancia con respecto al aprovechamiento de los recursos hidráulicos, y su aprovechamiento, podemos citar la definición que presenta la Organización Meteorológica mundial: “Hidrología es la ciencia que trata de los procesos

que rigen el agotamiento y recuperación de los recursos de agua en las áreas continentales de la tierra y en las diversas fases del ciclo hidrológico” (Campos Aranda, 1998)

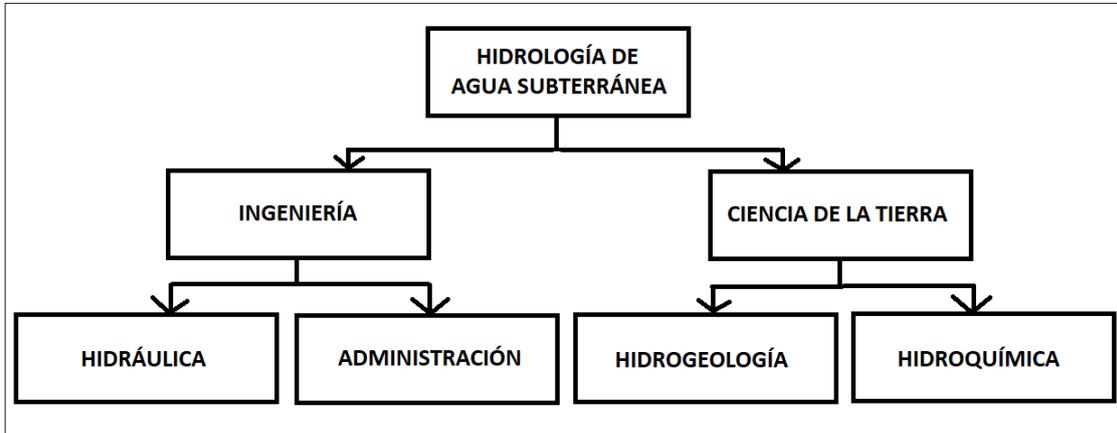


Figura 1. Temas relacionados con la hidrología (Şen, 2015)

La hidrología puede ser aplicada a campos relacionados con el desarrollo y gestión de recursos hidráulicos, de ahí viene su importancia, puesto que es un instrumento imprescindible para la planificación, diseño y explotación de proyectos hidráulicos en diversas áreas como ser ingeniería civil, ingeniería agronómica, ingeniería forestal, geología, climatología, etc. (Jurado Coronel, 2004)

3.1.2. CICLO HIDROLÓGICO

El agua circula en la hidrósfera a través de un laberinto de caminos que componen el Ciclo Hidrológico, el cual se constituye en el foco central de la hidrología. El ciclo hidrológico no tiene principio ni fin y sus procesos ocurren en forma continua (Jurado Coronel, 2004).

Se llama ciclo hidrológico al fenómeno que explica la circulación global de agua en nuestro planeta que se produce principalmente por influencia del sol, la fuerza de gravedad terrestre y la rotación de la Tierra. Se comienza describiendo a partir de la recepción de energía solar con la generación de vapor de agua hacia la atmósfera producto de la evaporación del agua desde lagos, ríos, océanos y mares y por medio de la evapotranspiración desde los suelos y vegetación. Este vapor de agua se condensa cambiando nuevamente de estado, lo que produce la formación de nubes. Por último, a través de la dinámica de las masas de aire se concreta la principal transferencia de agua denominada precipitación (Gallardo Terrones, 1994).

En la *Figura 2* se ilustra en forma esquemática el ciclo hidrológico, representando de esta manera que no tiene principio ni fin, sus diversos fenómenos ocurren de manera continua.

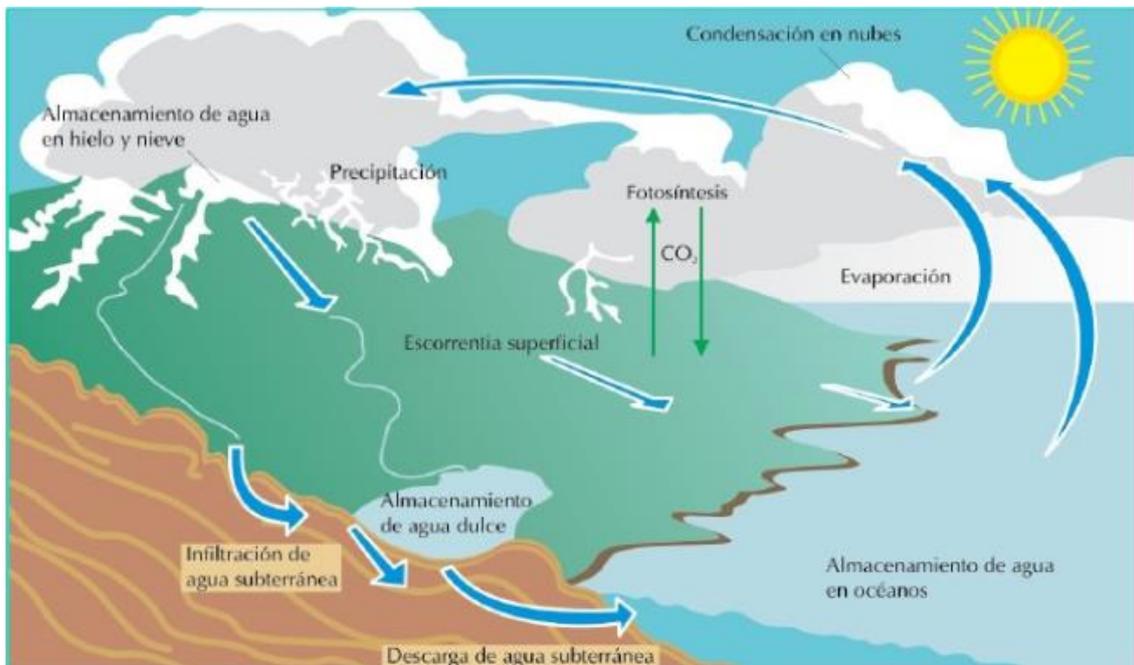


Figura 2. Ciclo hidrológico en la naturaleza (Gallardo Terrones, 1994)

Los principales reservorios dentro del ciclo hidrológico incluyen los océanos, la atmósfera, criósfera y el agua subterránea. Existen modelos matemáticos para describir el ciclo hidrológico como ser la “ecuación de continuidad hidrológica”. Para describir y predecir las variaciones dentro del ciclo hidrológico se ha invertido un gran esfuerzo en el desarrollo de modelos numéricos basados en computadoras. Cada componente del sistema climático tiene sus propios modelos, y dentro de las distintas disciplinas hay diferentes modelos utilizados. (Pagano & Sorooshian, 2002)

En resumen, podemos definir el ciclo hidrológico como la circulación general del agua en la superficie terrestre, que se define como: “Una sucesión de etapas que atraviesa el agua al pasar de la atmósfera a la tierra y volver a la atmósfera: evaporación desde el suelo, mar o aguas continentales, condensación de nubes, precipitación, acumulación en el suelo o masas de agua y reevaporación”. (Campos Aranda, 1998)

Podemos suponer que el ciclo hidrológico se inicia cuando el agua de los océanos se evapora y este vapor de agua se traslada a la atmósfera debido a las masas de aire en movimiento, luego, se forman las nubes las cuales dan origen a las precipitaciones. El agua que llega al suelo, una parte se evapora, otra parte es retenida por la vegetación, pero existe también una parte que circula por la superficie terrestre y se concentra en arroyos o pequeños surcos de agua que desembocan en ríos los cuales conducen a embalses, lagos y océanos. Por último, hay una tercera parte que se infiltra en el terreno, de la cual, una parte desciende para desembocar en agua subterránea y otra, que por medio de la saturación, es evaporada y vuelve a la atmósfera.

3.1.3. COMPONENTES DEL CICLO HIDROLÓGICO

3.1.3.1. PRECIPITACIÓN

La precipitación puede darse en forma de lluvia, nieve y otros procesos mediante los cuales el agua cae a la superficie terrestre. Se forma cuando se produce la elevación de una masa de agua en la atmósfera de tal manera que se enfríe y parte de su humedad se condense. En ciencias como hidrología y meteorología existe la necesidad de realizar mediciones de la cantidad de precipitación que se produce en una zona dada, que se realiza mediante un instrumento llamado *pluviómetro* (ver *Figura 3*) y se registran estas cantidades en un instrumento llamado *pluviógrafo* (ver *Figura 4*), las cantidades son expresadas en láminas de agua, generalmente en milímetros (mm) (Gallardo Terrones, 1994).



Figura 3. Pluviómetro. (Pepa, 2013)



Figura 4. Pluviógrafo (Baylina, 2003)

3.1.3.2. EVAPORACIÓN Y EVAPOTRANSPIRACIÓN

La evaporación es un proceso físico por el cual el agua pasa de estado líquido a gaseoso permitiendo que se transporte hacia la atmósfera. La evaporación se produce debido a la radiación solar que brinda la energía necesaria para que las moléculas de agua cambien de estado. Existen distintas maneras de evaporación como ser evaporación de superficie de agua libre como ser lagos, tanques, ríos, océanos, etc.; evaporación del agua del suelo y transpiración de plantas, y vienen a conformar lo que se denomina evapotranspiración. Entonces, la evaporación se constituye por la evaporación de superficie de agua libre y por la evapotranspiración (Gallardo Terrones, 1994).

3.1.3.3. INFILTRACIÓN, PERCOLACIÓN Y AGUAS SUBTERRÁNEAS

Se denomina infiltración al proceso de entrada de agua a través del suelo proveniente de precipitaciones, de irrigación o acumulación de agua en depresiones. Para que exista infiltración se debe presentar un abastecimiento de agua y también un receptor de esta infiltración, es decir el suelo a través de su capacidad de almacenamiento. El movimiento del agua una vez que ha ingresado a la superficie del suelo continúa, este proceso se denomina percolación y constituye la distribución lateral del agua en el suelo. Puede producirse una saturación de agua en el suelo una vez que se llenan completamente los espacios (poros) existentes en los materiales del suelo, generándose lo que se denominan aguas subterráneas. En la superficie del suelo se generará un proceso de evapotranspiración por efecto de los rayos solares. Existe un proceso por el cual se mide la profundidad del nivel de agua por efecto de la infiltración realizada con una sonda (*Ver Figura 5*) que permite tener un registro sistemático de las profundidades que aporta información importante sobre la disponibilidad de agua en los reservorios subterráneos y las variaciones en el movimiento de agua de los acuíferos (Gallardo Terrones, 1994).



Figura 5. Medidores de nivel de agua subterránea (Montoya, 2019)

3.1.3.4. ESCURRIMIENTO SUPERFICIAL, CUENCA HIDROGRÁFICA Y AGUAS SUPERFICIALES

Una cuenca o área de captación es el resultado del drenaje de un río. Los límites de esta área están determinados por zonas más elevadas del terreno que se constituyen como divisorias del agua. El conjunto de todos los cursos de agua como ser ríos, afluentes, arroyos que confluyen en un curso de agua principal se denominan red de drenaje superficial.

El caudal se mide en la cantidad de agua expresada en volumen que circula por unidad de tiempo. Existen varios instrumentos que se utilizan para estimar el caudal que está circulando por un tramo específico de un río (Gallardo Terrones, 1994).

3.1.4. ANÁLISIS HIDROLÓGICO

El objetivo del análisis hidrológico es obtener una caracterización de la cuenca y del río, determinar el régimen hídrico (estimar caudales mínimos y máximos en el sitio de un proyecto hidráulico) y definir el caudal de diseño para el dimensionamiento de dicho proyecto (Jurado Coronel, 2004).

El proceso de análisis hidrológico puede resumirse en las siguientes etapas: 1) Recolección de datos. 2) Ordenamiento de los datos. 3) Verificación de la calidad de los datos. 4) Extensión del registro. 5) Análisis de los resultados.

3.1.4.1. RECOLECCIÓN DE DATOS

Para comenzar con la ejecución de un análisis hidrológico, se deben tomar los datos de caudales, precipitaciones, temperatura, evaporación, etc de cierta cuenca, que fueron medidos con los instrumentos descritos anteriormente.

3.1.4.2. ORDENAMIENTO DE DATOS

Una vez obtenidos los datos, el primer paso que debe realizar el hidrólogo es ordenar los datos según los resultados que pretende alcanzar, esto es ya sea para obtener valores extremos como caudales máximos o mínimos o precipitaciones máximas o mínimas, o valores medios de caudales medios o mínimos. El resultado es un vector de

datos, los cuales corresponden al valor máximo, medio o mínimo de los datos hidrológicos para un periodo de tiempo.

3.1.4.3. EXTENSIÓN DEL REGISTRO

Todo proceso hidrológico depende de la cantidad y calidad de los datos hidrológicos, los cuales deben de ser confiables, consistentes y de extensión conveniente. Tener y trabajar en hidrología con datos de mala calidad, es lo mismo que no tenerlos. Por este motivo, el hidrólogo debe verificar si los datos son consistentes y están completos. Por lo tanto, si existen datos faltantes en una estación, se debe completarlos utilizando algún método como ser el de regresión lineal, tomando como base una estación índice. Luego debe aplicarse un método como ser el de la curva de doble masa para determinar si los datos son consistentes.

3.1.4.4. ANÁLISIS DE LOS DATOS

Cuando se trabaja, en general los datos de las estaciones no son suficientes para realizar un análisis completo, es decir, se genera la necesidad de obtener más datos a partir de otros. Existen varios métodos para realizar esta operación, aunque el más utilizado es el de la recta de regresión. Se debe determinar el coeficiente de correlación entre estas dos variables, para poder determinar la magnitud de la relación entre ambas y decidir si realizará una correcta estimación de los datos generados.

La extensión del registro debe realizarse para atrás, es decir, deben adicionarse datos anteriores al primer registrado. Se debe tener cuidado de no utilizar la extensión de

datos cuando se tiene una muestra demasiado pequeña, porque se generaría un margen de error muy grande.

3.1.4.5. VERIFICACIÓN DE LA CALIDAD DE LOS DATOS

Existen dos maneras de afrontar la realización de un análisis hidrológico. La primera, en la cual, teniendo caudales o precipitaciones, se realiza un tratamiento estadístico al conjunto de datos y se determinan probabilidades de ocurrencia de determinados valores de caudales o precipitaciones, según sea la información introducida. La segunda, se utiliza en el caso de tener solamente datos de precipitaciones, si se desea obtener valores de caudales, se debe aplicar la fórmula racional después de realizar el proceso estadístico.

3.1.5. SISTEMAS HIDROLÓGICOS

Actualmente, los hidrólogos consideran el ciclo hidrológico como un gran sistema, de esta manera tener una visión más allá de la cualitativa del mismo. Tener una visión cuantitativa del ciclo hidrológico permite conocer la cantidad de agua involucradas en cada una de sus fases y de esta manera, poder predecir la cantidad de recursos hidráulicos disponibles.

Se puede definir un sistema hidrológico como: “Un conjunto de elementos o procesos físicos unidos a través de una forma de interdependencia, que actúa sobre un

grupo de variables de entrada para convertirlas en las de salida. En estos sistemas, cada uno de los elementos o procesos integrantes es el resultado de complicadas interrelaciones de muchos factores de gran variabilidad espacial y temporal, cuyas características físicas prácticamente no son medibles y por ello no son calculables” (Campos Aranda, 1998).

En la *Figura 6*. Se muestra una representación del sistema hidrológico, que representa como un conjunto de partes diferenciadas que interactúan como un todo. Se puede considerar que los principales componentes de un sistema hidrológico son: precipitación, evaporación, escorrentía, y las otras fases del flujo (Ordoñez Galvez, 2011)

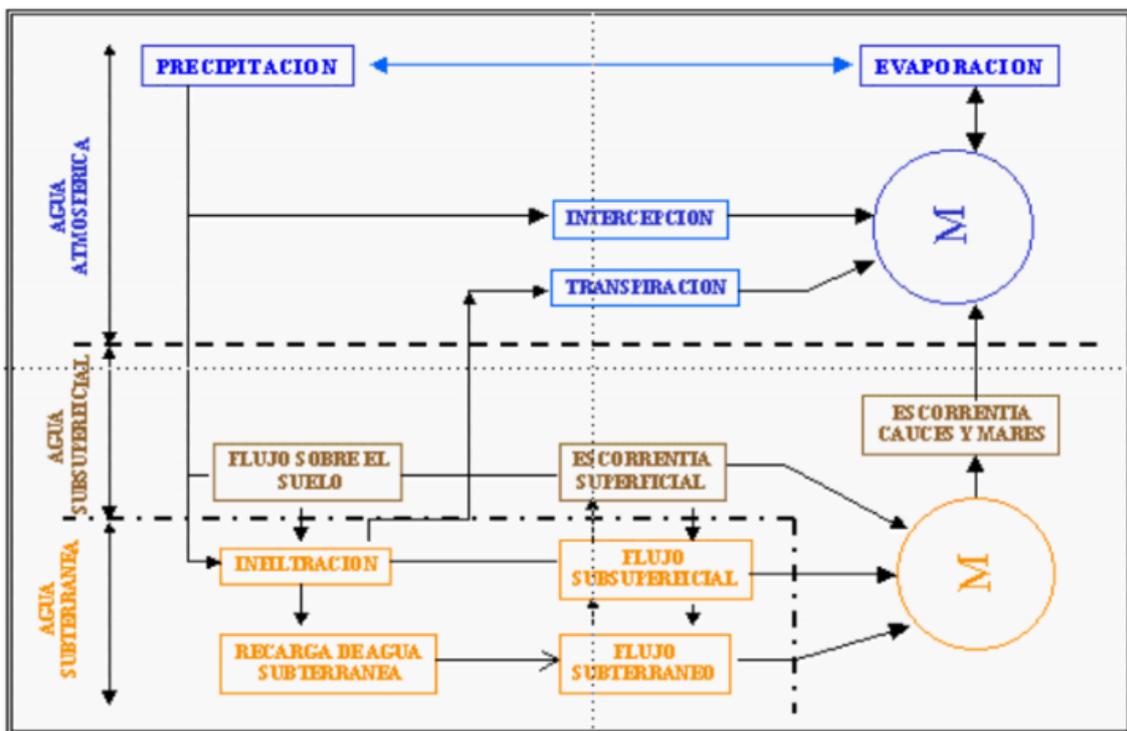


Figura 6. Representación de un Sistema Hidrológico. Fuente (Ordoñez Galvez, 2011)

3.1.6. MODELAMIENTO HIDROLÓGICO

La modelación hidrológica es una herramienta de gran importancia para el estudio de avenidas que se ha extendido por todo el mundo, fundamentalmente en países desarrollados. En la actualidad, con el empleo de estos modelos, se realiza el análisis y la prevención de las inundaciones; además, es posible manejar hipótesis suficientemente realistas o previsibles que ofrezcan un cierto grado de confianza para la toma de decisiones, ya sea en la ordenación del territorio en torno a los ríos o para exigir criterios de diseño de obras e infraestructuras capaces de soportar y funcionar adecuadamente en situaciones de emergencia (Estrada Sifontes & Pacheco Moya, 2012)

3.1.7. INFORMACIÓN HIDROLÓGICA EN ARGENTINA

El Ministerio del Interior, Obras Públicas y Vivienda a través de la Secretaría de Infraestructura y Política Hídrica intervienen en la ejecución de la política hídrica nacional y de la política relativa a los servicios públicos de abastecimiento de agua potable y saneamiento. A través de su página web (Red Hidrológica Nacional, 2019) proporcionan la mayor fuente de información hidrológica del país, su función es recabar los datos básicos necesarios para la formulación de proyectos y la administración regional de recursos hídricos. Cuentan con una base de datos actualizada de parámetros hidrológicos clasificados por cuencas y por fecha de aforo, los cuales están disponibles para cualquier persona u organismo que los necesite.

La Red Hidrológica Nacional (RHN) cuenta con 369 puntos de medición en todo el país, en muchos de los cuales se presentan tanto parámetros hidrológicos como meteorológicos. En la *Figura 7* se muestra un mapa de Argentina con el detalle de las estaciones de la RHN.

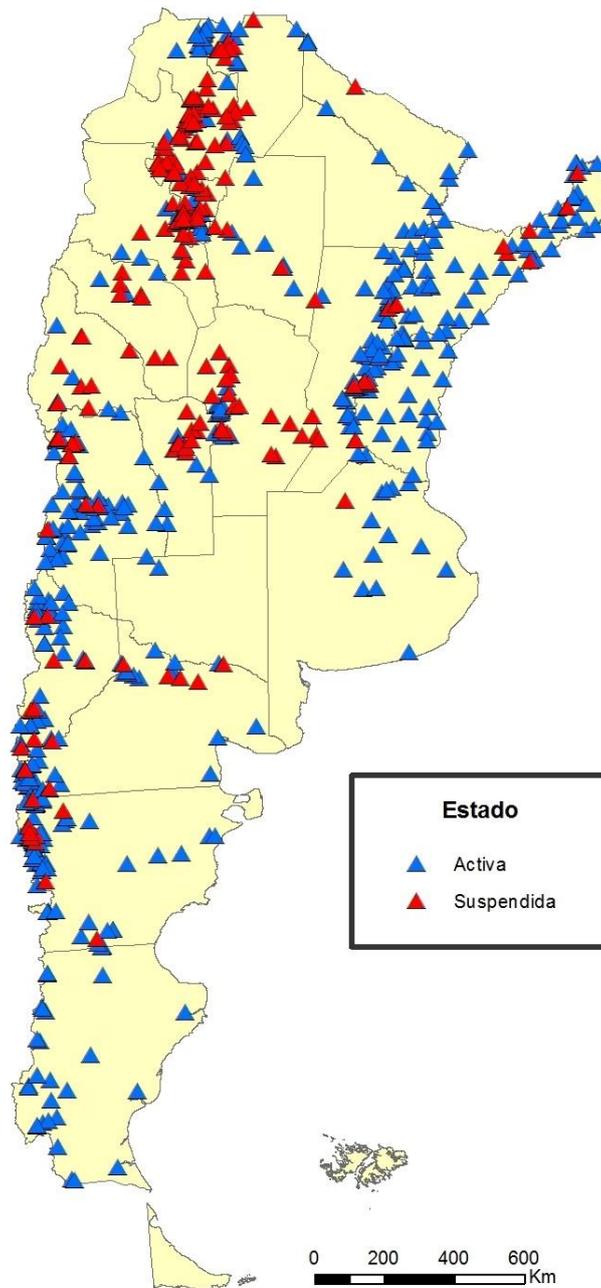


Figura 7. Estaciones de Red Hidrológica Nacional. (Red Hidrológica Nacional, 2019)

3.2. ÁMBITO DE LAS CIENCIAS DE DATOS

3.2.1. DEFINICIÓN DE CIENCIA DE DATOS

Podemos definir la Ciencia de Datos, en un nivel general, como un conjunto de principios fundamentales que apoyan y guían la extracción de información y conocimiento a partir de datos. Es probable que el concepto más relacionado con la ciencia de datos es la *minería de datos*, que puede definirse con la extracción real de conocimiento de los datos a través de tecnologías que incorporan estos principios. La minería de datos cuenta con una gran cantidad de técnicas y algoritmos. Pero la ciencia de datos involucra mucho más que algoritmos de ciencias de datos, ya que implica visualizar el problema del área de aplicación que se está estudiando desde una perspectiva de los datos. Una perspectiva de la ciencia de datos proporciona a los profesionales una estructura y principios, que le dan al científico de datos un marco para tratar sistemáticamente los problemas de extraer conocimiento útil a partir de los datos (Provost & Fawcett, 2013).

Para (Pujol Menendez & Porven Rubier, 2018), la ciencia de datos fue desarrollándose a partir de la evolución de ciencias como la estadística destinadas a analizar datos hacia una nueva disciplina combinadas con la matemática y la informática y en conjunto con la ciencia del comportamiento, posteriormente se agrega el aporte de la minería de datos, la minería de procesos y el aprendizaje automático. Un elemento más que se incorpora es el dominio de la aplicación específica, que puede llegar a tener conocimientos muy variados de acuerdo con el conocimiento que se quiera llegar a obtener. Esta relación entre disciplinas se muestra en la *Figura 8* que representa el

diagrama de Venn, el cual nos proporciona una visión de cómo se relacionan estas disciplinas entre sí. Cada una de las disciplinas que aportan a la ciencia de datos proporcionan teorías y técnicas para poder construir soluciones.

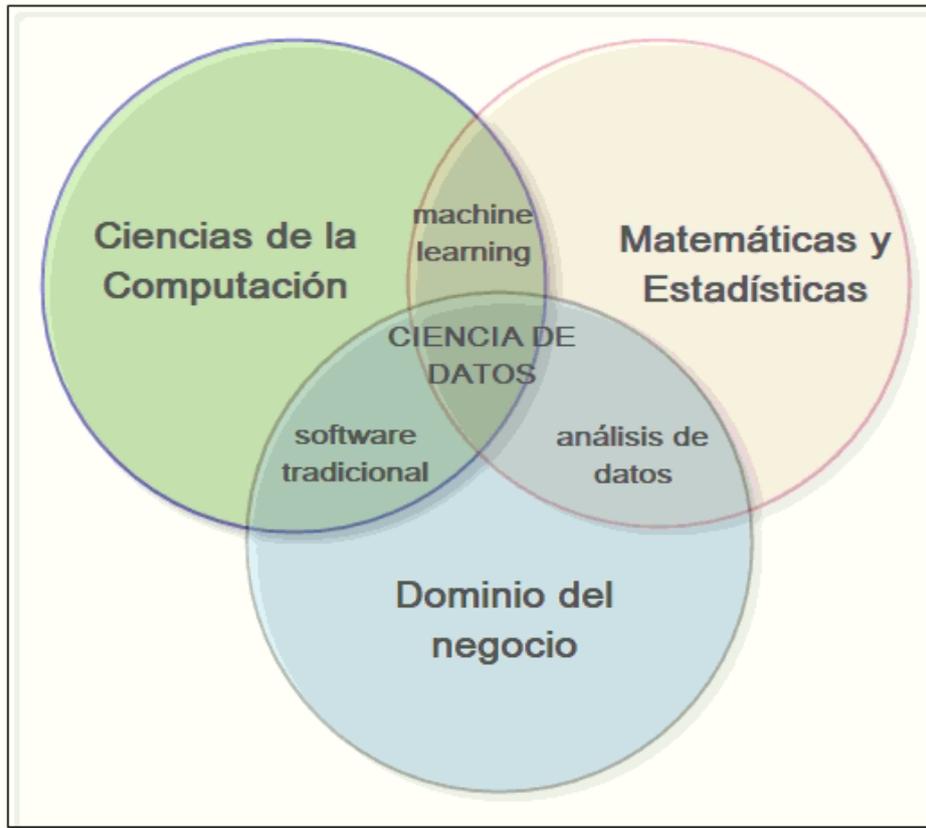


Figura 8. Disciplinas que componen la ciencia de datos. (Pujol Menendez & Porven Rubier, 2018)

3.2.2. INTELIGENCIA DE NEGOCIOS

La inteligencia de negocios es una herramienta que permite la toma de decisiones en base a información precisa y oportuna y a su vez, garantiza la generación del conocimiento necesario que permita seleccionar las alternativas más convenientes en pos del éxito del proyecto. Mediante el uso de metodologías, aplicaciones y tecnologías, se puede reunir, depurar y transformar datos y aplicar en ellos técnicas de extracción de

conocimiento de manera de obtener oportunidades de solución a problemas y ser aprovechados oportunamente. (Rosado Gomez & Rico Bautista, 2010)

La inteligencia de negocios, según (Rouhani, Asgari, & Mirhosseini, 2012) se define como un conjunto de habilidades, herramientas, técnicas y soluciones que proporcionan un apoyo para la resolución y entendimiento en situaciones de negocio. La implementación de inteligencia de negocios está basada en objetivos específicos al problema a resolver realizando una identificación de la información y agrupando conceptos ocultos y relevantes con el objetivo de mejorar el conocimiento y de esta forma dar un soporte a la toma de decisiones.

3.2.3. MINERÍA DE DATOS

La disponibilidad de grandes volúmenes de datos y la utilización de herramientas informáticas transformó el análisis de datos orientándolo al desarrollo de técnicas especializadas que conforman la minería de datos. Estas técnicas persiguen el descubrimiento automático del conocimiento que está contenido en la información de un área específica que está almacenada de forma ordenada en bases de datos. Estas técnicas buscan descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas de análisis de datos (Han, Kamber, & Pei, 2012).

Se define minería de datos como el proceso de seleccionar, explorar, modificar, modelizar y valorar grandes cantidades de datos con el objetivo de descubrir conocimiento. El proceso debe ser automático o semi-automático, los modelos encontrados deben de ser significativos, demostrando cierto patrón o regla de comportamiento. Se utiliza la explotación de información sobre todo cuando es necesario algún tipo de predicción (Rosado Gomez & Rico Bautista, 2010).

En la *Figura 9* se explica gráficamente cómo son utilizados los distintos repositorios de datos en la aplicación de las herramientas de minería de datos. Podemos ver que es necesaria la integración de todo tipo de información almacenada en distintos tipos de repositorios y en varios formatos como ser base de datos relacionales y archivos planos, ya sean de origen externo o interno, en un almacén de datos o DataWarehouse; una vez que la información es centralizada, se utilizan las herramientas de minería de datos, y a través de una interfaz de usuario, permitir que se represente de forma clara y resumida a los usuarios.

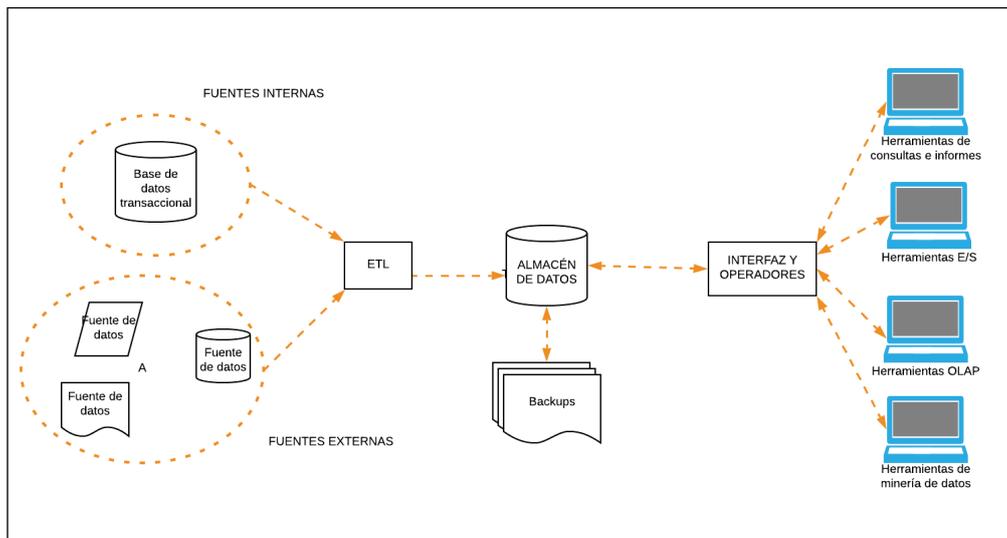


Figura 9. Información utilizada para herramientas de minería de datos (Pérez López & González, 2008)

3.2.3.1. EL PROCESO DE DESCUBRIMIENTO DEL CONOCIMIENTO

La minería de datos viene a constituir solamente una parte del proceso de extracción de conocimiento (KDD por sus siglas en inglés *Knowledge Data Discovery*) que se define como la extracción no trivial de información potencialmente útil a partir de un gran volumen de datos y el descubrimiento de patrones o relaciones en base a técnicas de aprendizaje, estadística y base de datos. Entre las tareas más comunes de datamining se encuentran los problemas de clasificación y clustering, el reconocimiento de patrones, el modelado descriptivo y la detección de dependencias (Valcarcel Ascencios, 2004).

Este proceso contiene varias etapas como ser la preparación de los datos, minería de datos propiamente dicha, evaluación e interpretación de los patrones, y por último,

difusión del nuevo conocimiento. Cuando se comienza con la preparación de los datos se realiza una preparación, limpieza y transformación previa de los datos a utilizar, y determina que las fases sucesivas del proceso sean capaces de extraer conocimiento válido y útil a partir de la información original. En la etapa de minería de datos, se desarrollan los modelos y se realiza un análisis de los datos, además se decide qué tarea se realizará, clasificar, agrupar, etc y se elige la técnica descriptiva o predictiva que se va a realizar. En la fase de evaluación e interpretación de los patrones, los expertos los analizan y de ser necesario, se vuelven a las fases anteriores para una nueva interpretación.

Por último, en la fase de difusión del conocimiento generado se expone el resultado a todos los usuarios interesados (Pérez López & González, 2008). En la *Tabla 2* se muestra en forma resumida estas etapas anteriormente explicadas.

KDD	Selección	Recopilar e integrar las fuentes de datos existentes		
		Identificar y seleccionar las variables relevantes en los datos		
		Aplicar las técnicas de muestreo adecuadas		
	Exploración	Utilizar las técnicas de análisis exploratorio de datos		
		Deducir la distribución de los datos, simetría y normalidad		
		Analizar las correlaciones existentes en la información		
	Limpieza	Detectar y tratar la presencia de valores atípicos (outliers)		
		Imputar la información faltante o valores perdidos (datos missing)		
		Eliminar datos erróneos e irrelevantes		
	Transformación	Utilizar técnicas de reducción y aumento de la dimensión		
		Aplicar técnicas de discretización y numerización		
		Realizar escalado simple y multidimensional		
	Minería de datos	Utilizar técnicas predictivas	Regresión y series temporales	
			Análisis discriminante	
			Métodos bayesianos	
			Algoritmos genéticos	
			Árboles de decisión	
			Redes neuronales	
		Utilizar técnicas descriptivas	Clustering y segmentación	
			Escalamiento	
			Reglas de asociación y dependencia	
			Análisis exploratorio	
			Reducción de la dimensión	
Evaluación e interpretación de los resultados				
Evaluación e interpretación de los resultados	Intervalos de confianza			
	Bootstrap			
	Análisis ROC			
	Evaluación del modelo			
Difusión y uso de modelos	Visualización			
	Simulación			

Tabla 2. Etapas del proceso de descubrimiento del conocimiento KDD (Pérez López & González, 2008) .

3.2.3.2. TÉCNICAS DE MINERÍA DE DATOS

Pueden distinguirse las técnicas de minería de datos entre técnicas descriptivas y técnicas predictivas las cuales están enfocadas en el descubrimiento del conocimiento embebido en los datos. Las *técnicas predictivas* desarrollan un modelo a partir de un conocimiento teórico previo y contiene variables de entrada que pueden ser dependientes

o independientes. El modelo obtenido debe ser contrastado con las premisas previamente asumidas, y una vez contrastado se acepta como válido. En las *técnicas descriptivas* las variables iniciales no tienen ninguna asignación previa ni tampoco se asume la existencia de variables dependientes o independientes. Se crean modelos a partir de reconocimiento de patrones. Existen también las llamadas *técnicas auxiliares* que son herramientas de apoyo más superficiales y limitadas, las cuales son basadas en técnicas estadísticas y descriptivas, consultas e informes, enfocados a la verificación de los resultados. En la *Tabla 3* se detallan estas técnicas especificando su clasificación (Pérez López & González, 2008).

Técnicas	Predictivas	Regresión		Descubrimiento	
		Análisis de Varianza y covarianza			
		Series temporales			
		Métodos bayesianos			
		Algoritmos genéticos			
		Clasificación Ad Hoc	Discriminantes		
	Árboles de decisión				
	Redes neuronales				
	Descriptivas	Clasificación post hoc	Clustering		Verificación
			Segmentación		
		Asociación			
		Dependencia			
		Reducción de la dimensión			
		Análisis exploratorio			
	Escalamiento multidimensional				
	Técnicas auxiliares	Proceso analítico de transacciones (OLAP)			
		SQL y herramientas de consulta			
Reporting					

Tabla 3. Clasificación de las técnicas de data mining (Pérez López & González, 2008)

3.3.1. EXPLOTACIÓN DE LA INFORMACIÓN

La explotación de la información es una subdisciplina de la informática que aporta a la inteligencia de negocios las herramientas y técnicas para la transformación de la información en conocimiento. En general, se define como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información. Esta disciplina engloba un conjunto de técnicas cuyo objetivo es la extracción de conocimiento procesable implícito en los repositorios de datos (Britos, Procesos de Explotación de Información Basados en Sistemas Inteligentes, 2008).

El conocimiento en un área específica es desconocido previamente, el cual puede ser útil para la toma de decisiones para un experto o para un responsable de sistemas de información. Los datos en sí no son los más relevante, sino el conocimiento que se encierra en sus relaciones, fluctuaciones y dependencias. Si el conjunto de relaciones refleja la realidad, aportan algo novedoso y útil para la toma de decisiones (Pollo-Cataneo, 2017).

Según (Basso, 2014) se define como explotación de información es una subdisciplina de los sistemas de información relacionada a la inteligencia de negocios que aporta las herramientas de análisis y síntesis para extraer conocimiento a partir de un conjunto de información referente a un área específica, este conocimiento se encuentra implícito en los datos disponibles en diferentes fuentes. Por lo tanto, la explotación de la información es el conjunto de herramientas que permiten transformar la información en conocimiento permitiendo el descubrimiento de correlaciones, patrones y tendencias

significativas dentro de grandes volúmenes de información provenientes de distintos orígenes y formatos utilizando tecnologías de reconocimiento de patrones, técnicas matemáticas y estadísticas.

3.3.2. PROCESOS DE EXPLOTACIÓN DE LA INFORMACIÓN

Un proceso de explotación de información puede definirse como un conjunto de tareas que están relacionadas lógicamente y que previamente tienen un valor para un problema o cuestión y que, luego de ser ejecutadas, se obtiene un segundo conjunto de información con mayor grado de relevancia y valor. Cada proceso de explotación de información está definido por un conjunto de información de entrada, un conjunto de transformaciones y un conjunto de información de salida. Un proceso de explotación de información puede estar conformado por un conjunto de procesos y a su vez, ser parte de un proceso de explotación de información mayor (Britos, Procesos de Explotación de Información Basados en Sistemas Inteligentes, 2008).

Un proceso de explotación de información se puede definir como un conjunto de tareas relacionadas lógicamente y que engloba un conjunto de técnicas de minería de datos y con su aplicación poder extraer conocimiento procesable que está implícito en el almacén de datos del área de estudio. La base de estas técnicas está en ciencias como la estadística y sistemas inteligentes y se abordan problemas de predicción, clasificación y segmentación (Basso, 2014).

En (Britos, Procesos de Explotación de Información Basados en Sistemas Inteligentes, 2008) se detalla una propuesta de procesos de explotación de información que se focalizan en el descubrimiento de reglas de comportamiento, descubrimiento de grupos, descubrimiento de atributos significativos por ponderación de reglas de comportamiento o de pertenencia a grupos y descubrimiento de reglas de pertenencia a grupos. A continuación, se describe cada uno de ellos.

3.3.2.1.DESCUBRIMIENTO DE REGLAS DE COMPORTAMIENTO

Este proceso se aplica cuando el objetivo es identificar las condiciones para obtener determinados resultados en el dominio del problema. Es decir, que el proceso de descubrimiento de reglas de comportamiento se utilizará cuando se requiere encontrar algunos factores o características por los cuales se presenta el problema identificado.

3.3.2.2.DESCUBRIMIENTO DE GRUPOS

Este proceso se utilizará cuando se necesita identificar algunos agrupamientos en la masa de información previamente obtenida, de acuerdo con el dominio del problema. El proceso de descubrimiento de grupos se utiliza cuando se necesita clasificar la información en determinados grupos respecto a características específicas, comunes a cada uno de ellos.

3.3.2.3.PONDERACIÓN DE INTERDEPENDENCIA DE ATRIBUTOS

Este proceso se aplica cuando se requiere identificar los factores de mayor incidencia sobre el resultado de un problema. El proceso de ponderación de

interdependencia de atributos necesita determinar la frecuencia de ocurrencia de algunos factores que determinan la ocurrencia del problema a analizar.

3.3.2.4.DISCUBRIMIENTO DE REGLAS DE PERTENENCIA A GRUPOS

Este proceso se utiliza para identificar condiciones por las cuales existe una pertenencia de acuerdo con cada clase de una partición que es desconocida previamente, pero se encuentra disponible dentro de la información del dominio del problema. El proceso de descubrimiento de reglas de pertenencia a grupos determina una clasificación de tipologías, clases o segmentos de acuerdo con ciertas características de la información.

3.3.2.5.PONDERACIÓN DE REGLAS DE COMPORTAMIENTO O DE LA PERTENENCIA A GRUPOS

Este proceso se utiliza cuando se necesita identificar las condiciones sobre la obtención de un resultado en el dominio del problema, estas condiciones serán las que definan la incidencia sobre el comportamiento o la que mejor defina la pertenencia a un grupo. El proceso de ponderación de reglas de comportamiento va a permitir determinar cuál es el factor o rasgo preponderante en un grupo.

3.3.3. METODOLOGÍAS DE EXPLOTACIÓN DE INFORMACIÓN

Los procesos de explotación de información anteriormente explicados deben apoyarse en metodologías para organizar y planificar las actividades. Algunas de las metodologías más utilizadas son CRISP-DM, SEMMA y KDD, y posteriormente se desarrollaron las tecnologías TDSP y ASUM. A continuación, se detalla cada una de ellas.

3.3.3.1. CRISP-DM

La metodología CRISP-DM (Proceso Standard Cruzado para Minería de datos, por sus siglas en inglés) proporciona un marco uniforme y directrices para realizar el proceso de minería de datos. Consta de 6 fases bien estructuradas y definidas, las cuales son: 1) Comprensión del negocio. 2) Comprensión de los datos. 3) Preparación de los datos. 4) Aplicación de técnicas de modelado. 5) Evaluación e interpretación del modelo generado. 6) Despliegue y presentación de los conocimientos adquiridos (Shafique & Haseeb, 2014). En la *Figura 10* se muestra la interacción de cada una de estas etapas y cómo se presenta como un proceso iterativo.

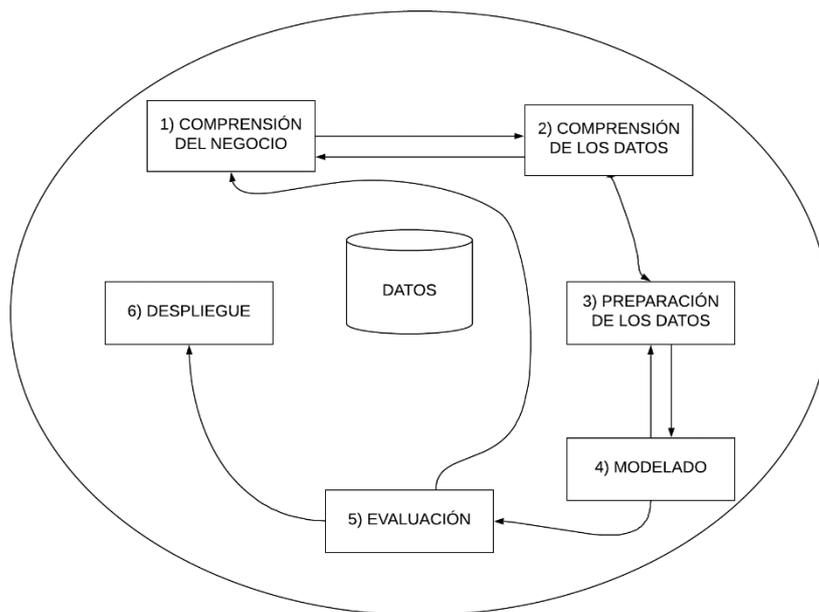


Figura 10. Etapas de la metodología CRISP-DM (Shafique & Haseeb, 2014)

Esta secuencia de seis etapas no es rígida, como se muestra en la *Figura 10* están debidamente organizadas, estructuradas y definidas, permitiendo que un proyecto pueda ser fácilmente comprensible, completo y documentado (Azevedo & Santos, 2008).

3.3.3.2.SEMMA

La metodología SEMMA (Muestra, explora, modifica, modela y accede por sus siglas en inglés) permite la comprensión, organización, desarrollo y mantenimiento de proyectos de minería de datos. Tiene un ciclo de cinco etapas o pasos: 1) Muestreo de los datos en la cual se toma un conjunto significativo de datos para poder comenzar con el proceso. 2) Exploración de datos que ayuda a comprenderlos y perfeccionar el proceso de descubrimiento mediante la búsqueda de tendencias o anomalías. 3) Modificación de los datos mediante la creación, selección y transformación de variables para enfocarse en el proceso de selección del proceso. 4) Modelado de datos utilizando diferentes técnicas y desarrollando un modelo indicado para cada tipo de situación. 5) Evaluación donde se evalúa la confiabilidad de los hallazgos y estimación de su rendimiento (Shafique & Haseeb, 2014). En la *Figura 11* se detalla el ciclo de etapas o pasos y su interacción entre ellos.

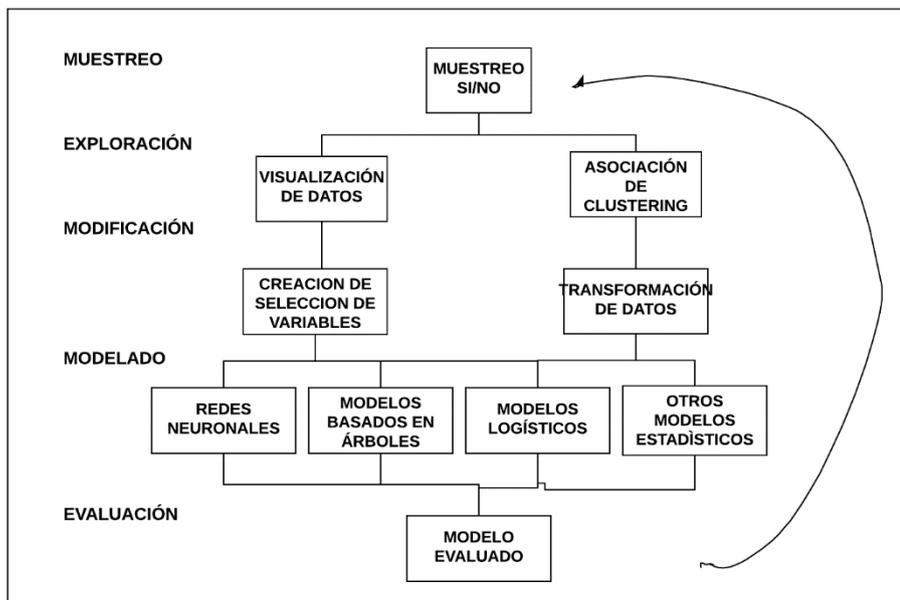


Figura 11. Ciclo de etapas de la metodología SEMMA (Costa & Aparicio, 2020)

SEMMA ofrece un proceso de fácil comprensión que permite un desarrollo organizado y adecuado en el mantenimiento de proyectos de datamining. De esta manera confiere una estructura para su concepción, creación y evolución, contribuyendo a presentar soluciones a problemas empresariales y encontrar los objetivos empresariales de datamining (Azevedo & Santos, 2008).

3.3.3.3.KDD

El modelo KDD (Base de datos de descubrimiento de información por sus siglas en inglés) es el proceso de extraer el conocimiento oculto de grandes repositorios de información. KDD requiere conocimientos previos y comprensión del dominio del cual se está intentando extraer el conocimiento. Esta metodología es interactiva e iterativa y contiene nueve pasos o etapas diferentes: 1) Desarrollo y comprensión del dominio de la aplicación. 2) Creación de un conjunto de datos objetivo. 3) Limpieza y procesamiento de datos. 4) Transformación de datos. 5) Elegir la tarea adecuada de minería de datos. 6) Elegir el algoritmo adecuado de minería de datos. 7) Empleo del algoritmo de minería de datos. 8) Interpretación de los patrones encontrados. 9) Utilización del conocimiento descubierto. En la *Figura 12* se representa cada una de estas fases, su interacción e iteración conjunta (Shafique & Haseeb, 2014).

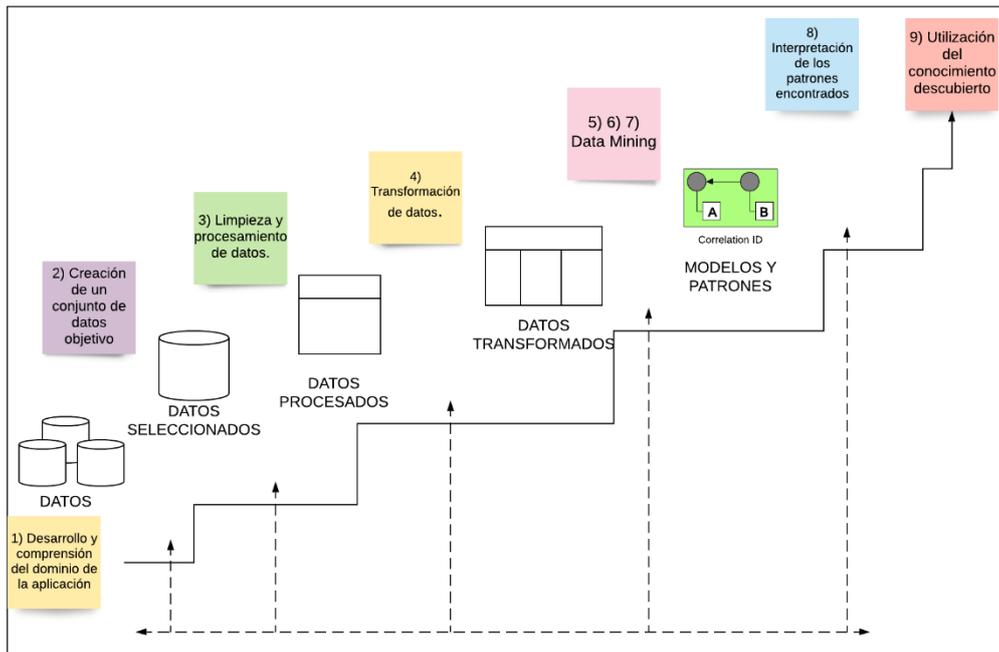


Figura 12. Etapas de la metodología KDD (Shafique & Haseeb, 2014)

La metodología KDD utiliza métodos de datamining para extraer lo que se considera conocimiento de acuerdo con la especificación de medidas y umbrales, utilizando una base de datos con cualquier tipo de preprocesamiento, submuestreo y transformación necesarios. Además, la metodología KDD debe estar precedida por el desarrollo de una comprensión del dominio de la aplicación, el conocimiento previo relevante y los objetivos del usuario final. También debe ser continuado por la consolidación del conocimiento mediante su incorporación al sistema (Azevedo & Santos, 2008).

3.3.3.4. TDSP

TDSP (Proceso de ciencia de datos en equipo, por sus siglas en inglés) es una metodología de ciencia de datos ágil e iterativa que ofrece soluciones de análisis

predictivo y aplicaciones inteligentes de manera eficiente, cuyo objetivo es mejorar el aprendizaje y colaboración de equipos sugiriendo algunos roles que interactúan de manera conjunta. Esta metodología contiene un apartado para desarrollar proyectos de ciencias de datos utilizando metodologías ágiles, dentro del cual se explica cómo planificar un sprint, agregar historias de usuario o bugs y crear tareas para cada ítem (Fois, Agüero Crovella, & Britos, 2020).

TDSP incluye un ciclo de vida sugerido que puede utilizarse para estructurar proyectos de ciencias de datos. Este ciclo de vida describe los pasos que suelen seguir los proyectos cuando se realizan. Como se muestra en la *Figura 10* la metodología TDSP sugiere etapas en paralelo con múltiples posibilidades de intercomunicaciones, las cuales contienen ciertas tareas. A continuación, citaremos cada etapa con sus respectivas tareas son: 1) Entendimiento del negocio que contiene la tarea 1.1) Definición de objetivos. 2) Adquisición de datos y comprensión que contiene las tareas 2.1) Introducción de los datos en el entorno de análisis de destino, 2.2) Exploración de los datos para determinar si su calidad es suficiente para responder a la pregunta y 2.3) Configuración de una canalización de datos para puntuar los datos nuevos o que se canalizan con regularidad 3) Modelado que contiene las siguientes tareas 3.1) Diseño de características, 3.2) Entrenamiento del modelo y 3.3) Determinar si el modelo es adecuado para su uso en producción. 4) Despliegue. 5) Aceptación del usuario que contiene las tareas 5.1) Validación del sistema y 5.2) Entrega del proyecto (Costa & Aparicio, 2020).

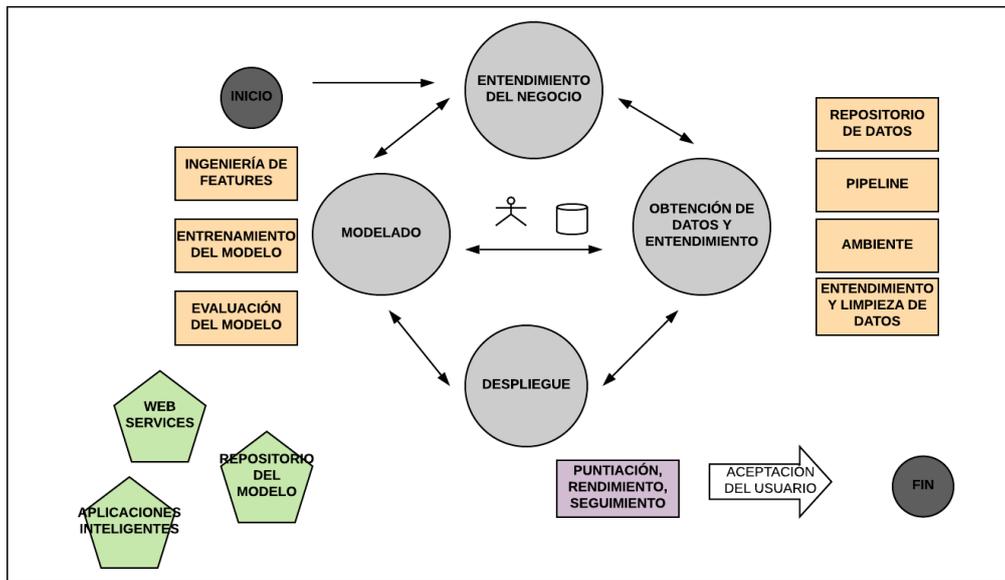


Figura 13. Ciclo de vida de la Metodología TDSP (Costa & Aparicio, 2020)

3.3.3.5.ASUM

ASUM (Método unificado de soluciones analíticas, por sus siglas en inglés) es un proceso interactivo para implementar un proyecto de minería de datos o análisis predictivo que utiliza la implementación de metodologías ágiles y los principios tradicionales de la ingeniería de software para lograr sus objetivos y proporcionar un resultado óptimo (Fois, Agüero Crovella, & Britos, 2020).

ASUM se basa en una metodología CRISP-DM ampliada y refinada, contiene cinco fases: 1) Analizar 2) Diseñar 3) Configurar y construir 4) Implementar 5) Operar y optimizar. Las tres primeras etapas de ASUM se pueden combinar, ya que los proyectos de minería de datos y análisis predictivo son iterativos por naturaleza (Costa & Aparicio, 2020). En la *Figura 14* se explica cada una de las fases de ASUM y su relación entre ellas.

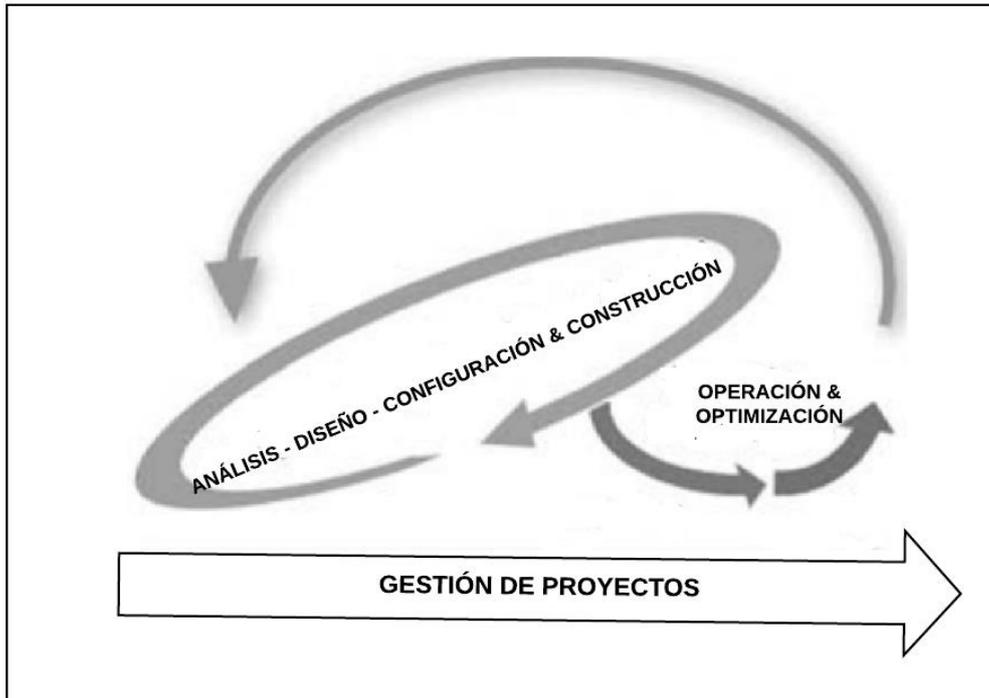


Figura 14. Fases de la tecnología ASUM (Costa & Aparicio, 2020)

4. ESTADO DEL ARTE

4.3. APLICACIÓN DE LA INFORMÁTICA EN LA HIDROLOGÍA

Es un hecho indudable que los sistemas informáticos estén presentes en todos los aspectos de la vida moderna, y más aún, contribuyen sustancialmente en las ramas de la ciencia y tecnología. Un producto de software y la información que maneja constituyen una herramienta imprescindible en la realización de todo proceso ingenieril. Es importante notar que los productos de software son útiles principalmente para obtener de la información almacenada, nueva información derivada que permita tomar decisiones y obrar. En este sentido, es que la informática ha sido aplicada de diversas maneras para facilitar el trabajo de los profesionales del área hidrológica en la realización de un

adecuado análisis y modelamiento hidrológico, proporcionando datos lo más precisos posibles y ahorrándole cálculos tediosos (Jurado Coronel, 2004).

El uso de las tecnologías de la información y de las comunicaciones en los procesos de gestión han demostrado ser un requisito en la toma de decisiones. En el ámbito de los recursos hídricos, estas tecnologías han proporcionado a los profesionales en el área herramientas de planificación y control y de gestión de recursos de información (Conde, González Vallvé, & Centeno Gutiérrez, 2019).

El enfoque de la ingeniería de sistemas aplicada a la hidrología consiste en identificar los componentes del ciclo hidrológico y el proceso dinámico por el cual ocurre el movimiento, para posteriormente aislarlos, con propósitos de análisis y estudio del ciclo hidrológico, para modelar su simulación de forma posterior. Dentro de este enfoque, se puede determinar dos conceptos: Sistema abierto, el cual opera independientemente de las salidas producidas y Sistema Cerrado, cuando la operación depende de la retroalimentación de todas o parte de sus salidas. (Campos Aranda, 1998). En la *Figura 9* se representa el enfoque de la informática en el ciclo hidrológico.

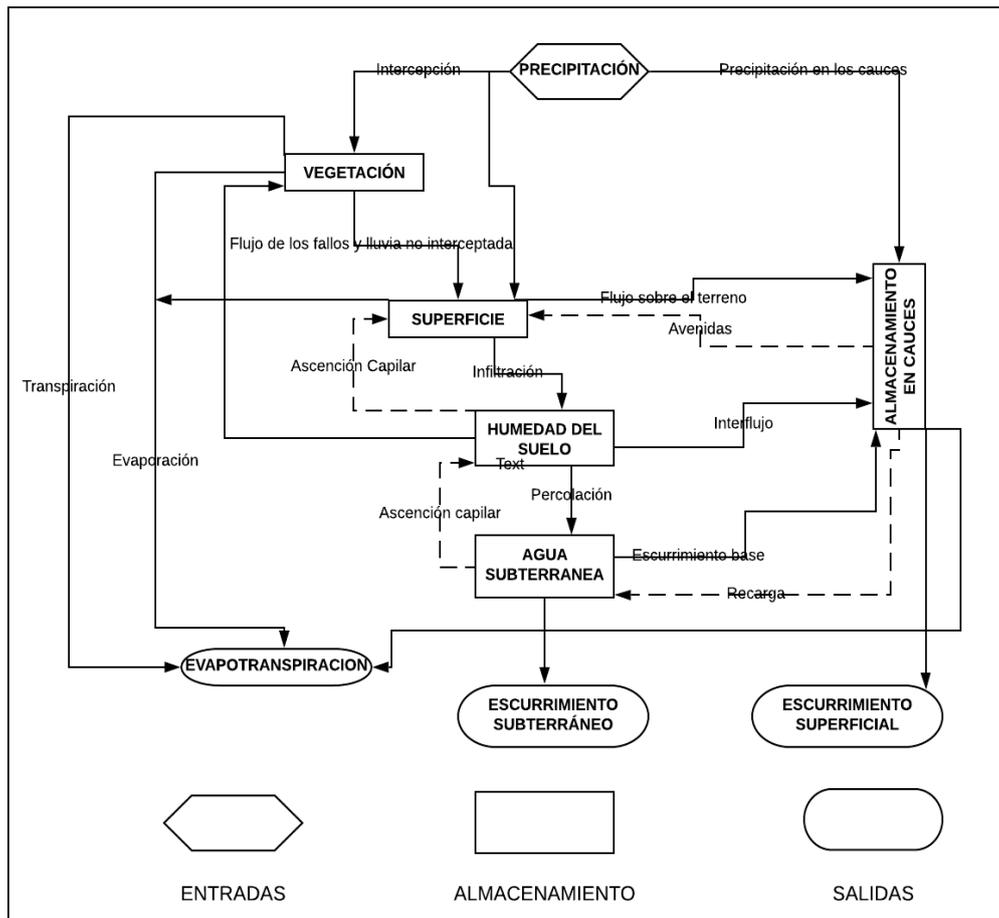


Figura 15. Enfoque de la informática en ciclo hidrológico (Campos Aranda, 1998)

En la actualidad, se está haciendo frecuente el uso de Sistemas de Información Geográfica SIG para el modelamiento hidrológico. Desde un punto de vista global, un SIG puede ser considerado como un conjunto de hardware, software, datos y técnicas eficientemente diseñadas para la captura, almacenamiento, actualización, manipulación, visualización y análisis de información geográficamente referenciada. Los SIG tienen una aceptación generalizada ya que son una herramienta muy productiva y además, tiene un alto grado de aprehensión de los sistemas y modelos hidrológicos, estas dos situaciones legitiman su utilidad su utilización de modelos hidrológicos en los SIG. Por otra parte, la dificultad de integración de series temporales de datos en entornos SIG ha puesto un

obstáculo importante para su difusión entre los investigadores de campo de la hidrología (Morad & Triviño Perez, 2001)

En la actualidad, con el avance de técnicas de inteligencia artificial, se fue incursionando en la aplicación de redes neuronales a series de tiempo hidrológicas. El análisis de series de tiempo hidrológicas a llevado a los investigadores a aprovechar los desarrollos de los últimos 30 o 40 años en estadística matemática, teoría de la probabilidad y teoría de las comunicaciones. El resultado a sido la evolución de un campo de especialización llamado hidrología estocástica (De la Hoz, Calchano, & Sánchez, 2019). Los modelos construidos para diferentes configuraciones de redes neuronales reproducen las principales características estadísticas de una serie temporal histórica con series temporales hidrológicas.

4.4. HIDROINFORMÁTICA

Los sistemas de información ambiental se encargan del manejo de los datos correspondientes a los distintos componentes que interactúan en el ciclo hidrológico, como ser el suelo, el agua, la atmósfera y las especies existentes. La gran demanda de información ambiental y de herramientas apropiadas para su análisis y organización, y la necesidad de que las organizaciones encargadas de este tema cuenten con sistemas de información eficientes, motiva el interés creciente por esta área de investigación. Los principales problemas que se resuelven con los sistemas de información ambiental son modelado de ríos, modelado de cuencas, agua subterránea, transporte de sedimentos, calidad de aguas, predicción de crecidas, riesgos hidrológicos, cambios climáticos,

manejo de suelos y aguas subterráneas y conservación de calidad de aguas (Urciuolo, Iturraspe, Sandoval, & Parson, 2002).

La hidroinformática es un paradigma que surgió como la evolución del desarrollo de una gran cantidad de sistemas de información ambiental que fueron dando origen a un gran número y diversidad de modelos computacionales y que tiene sus bases en la hidráulica computacional. (Urciuolo, Iturraspe, Sandoval, & Parson, 2002) define la hidroinformática como el paradigma que se ocupa del desarrollo y aplicación de modelos matemáticos y tecnología avanzada de computación en la resolución de problemas de hidrología, hidráulica e ingeniería ambiental. Provee sistemas de soporte de decisión basados en computadoras para ser utilizados por ingenieros, autoridades del agua y organismos gubernamentales de la gestión hídrica y ambiental.

La hidroinformática se estableció como campo de estudio durante los últimos años originándose a partir de la evolución de la hidráulica computacional. Originalmente se estableció como el campo de estudio del modelado numérico y flujos de información relacionados con los sistemas acuáticos, su enfoque se ha transformado cada vez más en el estudio de las mejoras de las herramientas de modelado numérico utilizando la velocidad y precisión de las herramientas informáticas, poniendo mucho más esfuerzo en las interfaces de usuario, nuevos tipos de salidas gráficas, controles de parámetros y estructuración de datos. Ha contribuido en gran medida al desarrollo de sistemas de modelado que permitieron aumentar el potencial de especialistas de hidrología, hidráulica y recursos hídricos. Aprovechando los últimos avances tecnológicos en las áreas de instrumentación, transmisión de datos en tiempo real, asimilación de datos e interfaces,

hoy en día se reconoce como un campo que busca desarrollar modelos mejorados, herramientas y servicios. Siguiendo los revolucionarios desarrollos en TIC, se fueron incorporando sistemas de decisiones o evaluación de impacto. En los últimos años la hidroinformática se fue transformando de una ciencia puramente técnica a una disciplina sociotécnica. Tomando en cuenta el impacto de las herramientas y sistemas hidroinformáticos existentes, y en visión de los desafíos de las complejas relaciones actuales entre las sociedades humanas y los recursos hídricos, se está convirtiendo en un campo que asume la responsabilidad de sustentar los nuevos paradigmas emergentes para la gestión del agua. En este sentido es necesaria la comprensión que en la toma de decisiones relacionadas con el agua y los procesos ambientales, lo social y lo técnico están entrelazados hasta tal punto que interactúan simultáneamente.

4.5. APLICACIÓN DE LAS CIENCIA DE DATOS EN LA HIDROLOGÍA

De la lectura bibliográfica realizada, podemos determinar algunas aplicaciones de herramientas de ciencia de datos en hidrología, a continuación, se detallan.

(Belda, Penadés, & García-Haro, 2010) propusieron en su estudio la aplicación de la tecnología OLAP-Mining como un mecanismo de búsqueda de patrones o conocimiento interesante de forma multidimensional y en varios niveles de abstracción, con el objetivo de monitorizar y cuantificar la sequía en diferentes áreas de la península ibérica, utilizando datos climatológicos, de precipitación y evapotranspiración.

(Maier, y otros, 2016) Realizaron un estudio para realizar la modelización hidrológica utilizando el modelado basado en datos dentro de un marco conceptual, como un medio para mejorar el conocimiento requerido para apoyar modelos sociohidrológicos. De esta manera, explorar oportunidades y desafíos para el modelado basado en datos con respecto a aumentar la comprensión del sistema sociohidrológico, mejorar la capacidad para hacer predicciones y proporcionar mejoras en la ciencia de la modelización hidrológica.

En (Minotti & Kandus, 2015) se realizó un estudio de aplicación de modelos predictivos de árboles de decisión de técnicas de data mining para analizar la distribución de humedales en Argentina, con el objetivo de analizar variables relevantes y para evaluar la calidad de los datos de origen de clase objetivo. Se generó una base de datos de las áreas de humedales que ocupan un 22% del territorio argentino y se utilizaron técnicas de clasificación para generar modelos predictivos. La exactitud del modelo se evaluó de manera no espacial mediante curvas ROC las cuales se grafica la tasa de positivos verdaderos en función de la de falsos positivos para un modelo predictivo dado.

(Molero-Castillo & Meda-Campaña , 2011) realizaron un estudio para agrupar zonas o regiones climatológicamente similares en el estado de Michoacán – México basado en técnicas de agrupamiento, basado en técnicas de agrupamiento de minería de datos. Se utilizaron las técnicas Bietápico, K-medias y redes de Kohonen, las cuales fueron aplicadas a una fuente de datos generada a partir de los registros disponibles de las variables de temperatura (observada, máxima y mínima), precipitación y altitud. Este trabajo implicó retos importantes, como el análisis de series temporales, el manejo de

técnicas de agrupamiento para regionalizar zonas homogéneas y el empleo de índices de validación, permitiendo extender la visión de la minería de datos y su aplicación a problemas de diversa índole.

5. CONCLUSIONES

A continuación, se realizará un breve análisis de los aspectos que se tomaron en cuenta para el desarrollo del presente trabajo en base a los objetivos planteados al inicio del mismo.

Primeramente, se gestó un entendimiento de los conceptos generales de la hidrología, el ciclo hidrológico y los elementos que lo componen, esto para adquirir el conocimiento necesario del experto, asimilar el conocimiento necesario del desarrollo del análisis hidrológico y del modelamiento hidrológico, determinar su importancia y analizar su complejidad, para luego determinar cómo se obtienen los datos hidrológicos en Argentina, de esta manera, poder tener una idea clara del tipo de información a procesar y los resultados que es necesario obtener.

A continuación, se verificó la utilización que actualmente se está haciendo de las herramientas informáticas en el modelamiento hidrológico. El concepto de Hidroinformática se presentó al realizar la revisión bibliográfica, y en este sentido, pudimos verificar que existe todo un ámbito de la informática dedicado específicamente a apoyar a los ingenieros hidrólogos, hidráulicos y ambientales a realizar el análisis y modelamiento hidrológico. Son variados los trabajos en los cuales pudimos constatar que

desde hace varios años hasta la actualidad viene aplicándose el uso de las TIC en la hidrología, y su evolución va en conjunto con los avances tecnológicos. Esto nos brinda un panorama optimista, ya que se hace propicia una propuesta de un proyecto de explotación de información como apoyo a los profesionales en hidrología.

Por último, se realizó una revisión bibliográfica a cerca de los conceptos teóricos de la ciencia de datos, más precisamente los procesos de explotación de información y las metodologías utilizadas actualmente, en relación a esto, podemos indicar que existe un proceso de explotación específico para el tipo de modelo a generar y al tipo de problema a afrontar. En este punto nos detenemos más precisamente en el modelo de procesos propuesto por (Britos, Procesos de Explotación de Información Basados en Sistemas Inteligentes, 2008) ya que nos parece adecuados para aplicarlos como propuesta de una alternativa al modelado hidrológico tradicional. Como metodología de explotación de la información se podría utilizar TDSP o ASUM ya que ambas utilizan conceptos ágiles en el desarrollo de proyectos de ciencia de datos.

En conclusión, podemos determinar que es totalmente factible insertar un aporte de la aplicación de un proceso de ciencia de datos como herramienta dentro de la hidroinformática, para dar soporte a ingenieros hidráulicos, hidrólogos y ambientales en el análisis y modelamiento hidrológico. Al revisar las etapas del análisis hidrológico, podemos ver que se asemejan a las fases de varias de las metodologías de explotación de información, en este sentido, podemos predecir que será bienvenido un proyecto de ciencia de datos aplicado a la hidrología.

6. REFERENCIAS

- Aguirre-Munizaga, M., Gomez, R., Aviles, M., Vasquez, M., & Recalde-Coronel, C. (2016). A Cloud Computing Based Framework for Storage and Processing of Meteorological Data. *Conference Paper in Communications in Computer and Information Science*.
- Azevedo, A., & Santos, M. (2008). KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW. *ADIS European Conference on Data Mining 2008*, 24-26.
- Basso, D. M. (2014). Propuesta de Métricas para Proyectos de Explotación de Información.
- Baylina, R. (11 de 11 de 2003). *Rumtor.com*. Obtenido de <https://www.rumtor.com/pluviografo.html>: <https://www.rumtor.com/pluviografo.html>
- Belda, F., Penadés, M., & García-Haro, F. (2010). Técnicas de minería de datos para el análisis de periodos de sequía en España. *Asociación Meteorológica Española*, 20-23.
- Britos, P. (2008). Procesos de Explotación de Información Basados en Sistemas Inteligentes. *Tesis de Doctorado en Ciencias Informáticas. Facultad de Informática Universidad Nacional de La Plata*. La Plata - Argentina.
- Britos, P. (2008). *Procesos de Explotación de Información Basados en Sistemas Inteligentes*. La Plata.
- Campos Aranda, D. (1998). *Procesos del Ciclo Hidrológico*. San Luis Potosí, México: Editorial Universitaria Potosina.
- Catalán Vázquez, M., & Jarillo Soto, E. (2010). Paradigmas de investigación aplicados al estudio de la percepción pública de la contaminación del aire. *Rev. Int. Contam. Ambiental*, 165-178.
- Cogliatti, M., Britos, P., & García-Martínez, R. (2006). Análisis del Viento en el Valle del Río Negro Mediante Mapas Auto Organizados y Algoritmos de Inducción. *Avances en Energías Renovables y Medio Ambiente*. Argentina: ASADES.
- Conde, C., González Vallvé, J., & Centeno Gutiérrez, S. (2019). LOS SISTEMAS AUTOMÁTICOS DE INFORMACIÓN HIDROLÓGICA (SAIH) UNA

INNOVACIÓN QUE SE EXPORTA. *HITOS DE INNOVACIÓN EN INGENIERÍA CIVIL.*

- Costa, C., & Aparicio, J. T. (2020). POST-DS: A Methodology to Boost Data Science. *2020 XV Congreso Ibérico de Sistemas y Tecnologías de la Información (CISTI).*
- De la Hoz, B., Calchano, O., & Sánchez, P. (2019). Redes neuronales para pronóstico de series de tiempo hidrológicas del Caribe colombiano. *Revista I+D en TIC Volumen 10*, 18-31.
- Estrada Sifontes, V., & Pacheco Moya, R. (2012). Modelación hidrológica con HEC-HMS en cuencas montañosas de la región oriental de Cuba. *RIHA vol.33 no.1.*
- Fois, G., Agüero Crovella, G. A., & Britos, P. V. (2020). Evaluación comparativa de las metodologías Team Data Science Process TDSP y Analytics Solutions Unified Method for Data Mining ASUM-DM desde la perspectiva de la ciencia de datos. *Investigación Formativa en Ingeniería*, 264-271.
- Gallardo Terrones, E. (1994). *Conceptos Hidrológicos Básicos.*
- Han, H., Kamber, M., & Pei, J. (2012). En *Data Mining - Concepts and Techniques* (págs. 5, 6, 7). Estados Unidos: Morgan Kaufmann.
- Herrera Prado, A. L., Uribe Agundis, D., & Ruiz Luna, A. (2015). Aplicación de Técnicas de Minería de Datos Para la Generación de Pronósticos de Disponibilidad de Agua en la cuenca Baja del Río San Lorenzo. *XII Encuentro Participación de la Mujer en la Ciencia.* León Guanajuato.
- Herrera Prado, A., Uribe Agundis, D., & Ruíz Luna, A. (2015). Aplicación de Técnicas de Minería de Datos, para la Generación de Pronósticos de Disponibilidad de Agua en la Cuenca Baja el Río San Lorenzo. *XII Encuentro participación de la Mujer en la Ciencia.*
- Jurado Coronel, M. C. (Septiembre de 2004). *Análisis Hidrológico Asistido por Computadora Aplicado al Departamento de Tarija.* Tarija, Bolivia: UAJMS.
- Krishna Thakur, J. (2015). Optimizing Groundwater Monitoring Networks Using Integrated Statistical and Geostatistical Approaches. *Open Access Hydrology.*

- Maier, H., Mount, N., Toth, E., Elshorbagy, A., Solomatine, D., Chang, F.-J., & Abrahart, R. (2016). Data-driven modelling approaches for sociohydrology: Opportunities and challenges within the Panta Rhei Science Plan. *Hydrological Sciences Journal*.
- Materón Muñoz, H., García Vélez, J. L., Arango, D., & Parra, D. F. (2006). Modelación hidrológica e hidráulica aplicada a estudios de inundabilidad en cauces naturales caso de estudio urbanización río Nima. *Redalyc*. Cali - Colombia: Eidenar.
- Minotti, P., & Kandus, P. (2015). APLICACIÓN DE SIG Y DATA MINING PARA ESTABLECER PERFILES ECOLÓGICOS DE LOS HUMEDALES EN ARGENTINA. *Maestría en Minería de Datos y Descubrimiento del Conocimiento*. Universidad de Buenos Aires.
- Mishra, S., Dwiveki, V., Saravanan, C., & Pathak, K. (2013). Pattern Discovery in Hydrological Time Series Data Mining during the Monsoon Period of the High Flood Years in Brahmaputra River Basin. *International Journal of Computer Applications*. Estados Unidos.
- Mita, S., Pal, S. K., & Pabitra, M. (2002). Data Mining in Soft Computing Framework A Survey. *IEEE Computational Intelligence Society*. IEEE Journals.
- Molero-Castillo, G., & Meda-Campana, M. (2011). Climatic Regionalization using Data Mining Techniques: A Case of Study. *2011 VI Congreso Colombiano de Computación (CCC)*.
- Montoya, S. (30 de 10 de 2019). *gidahatari*. Obtenido de <https://gidahatari.com/ih-es/implementacion-sistemas-monitoreo-aguas-subterranas>: <https://gidahatari.com/ih-es/implementacion-sistemas-monitoreo-aguas-subterranas>
- Morad, M., & Triviño Perez, A. (2001). Sistemas de Información Geográfica y Modelizaciones Hidrológicas: Una Aproximación a las Ventajas y Dificultades de su Aplicación. *Boletín de la AGE Nro 31*, 23-46.
- Nourani, V., Davanlou Tajbakhsh, A., & Molajou, A. (2018). Data mining based on wavelet and decision tree for rainfall-runoff simulation. *Hydrology Research*.

- Ocampo, O. L., & Vélez, J. J. (2013). Análisis Comparativo de Modelos Hidrológicos de Simulación Continua en Cuencas de Alta Montaña: Caso del Río Chinchiná. *Revista Ingenierías Universidad de Medellín*. Medellín: Scielo.
- Ordoñez Gálvez, J. J. (2011). Balance Hídrico Superficial. *Global Water Partnership South America*. Lima: Sociedad Geográfica de Lima.
- Ordoñez Galvez, J. J. (2011). *Ciclo Hidrológico*. Lima, Perú: Foro Peruano para el Agua.
- Pagano, T., & Sorooshian, S. (2002). Hydrologic Cycle. En M. MacCracken, & J. Perry, *The Earth system: physical and chemical dimensions of global environmental change* (págs. 450-464).
- Pepa, G. H. (27 de 11 de 2013). *Cordoba Times*. Obtenido de <https://www.cordobatimes.com/el-campo/2013/11/27/sabe-usted-como-se-mide-la-lluvia/>: <https://www.cordobatimes.com/el-campo/2013/11/27/sabe-usted-como-se-mide-la-lluvia/>
- Pérez López, C. S., & González, D. (2008). *Minería de Datos. Técnicas y Herramientas*. Madrid, España: International Thomson Ediciones.
- Pollo-Cataneo, M. (2017). *Modelo de Proceso para Elicitación de Requerimientos en Proyectos de Explotación de Información*. La Plata.
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Mary Ann Liebert Inc*.
- Pujol Menendez, N., & Porven Rubier, J. (2018). Ciencia de datos: Una revisión del estado del arte. *UCE Ciencia: Revista de Postgrado*, Vol 6 (3).
- Ramos, C. A. (2015). Los paradigmas de la investigación científica. *ResearchGate*.
- Red Hidrológica Nacional*. (2019). Obtenido de <https://www.argentina.gob.ar/interior/secretaria-de-infraestructura-y-politica-hidrica/red-hidrologica-nacional>
- Rey-Valencia, D., & Zambrano, J. (2018). Estudio de la respuesta hidrológica en la cuenca urbana de montaña San Luis-Palgrande. *Revista UIS Ingenierías*. Santander: Dialnet.

- Rojas Sigala, D., & González Perez, M. G. (2015). Incertidumbre en los modelos hidrológicos y planeación de obras hidráulicas en México. *Ingeniería Hidráulica y Ambiental*. La Habana: Scielo.
- Rosado Gomez, A. A., & Rico Bautista, D. W. (2010). Inteligencia de Negocios: Estado del Arte. *Scientia et Technica*, 321-326.
- Rouhani, S., Asgari, S., & Mirhosseini, S. V. (2012). Estudio de revisión: Conceptos y enfoques de la inteligencia de negocios. *American Journal of Scientific Research*.
- Satanand, M., C., S., V. K., D., & K. K., P. (2013). Pattern Discovery in Hydrological Time Series Data Mining during the Monsoon Period of the High Flood Years in Brahmaputra River Basin. *International Journal of Computer Applications*.
- Şen, Z. (2015). *Practical and Applied Hydrogeology*.
- Shafique, U., & Haseeb, Q. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 217-222.
- Silva Medina, G. (29 de 7 de 2003). Hidrología Básica. Bogotá: Academia. Obtenido de www.academia.edu.
- Spate, J., Croke, B., & A.J., J. (2016). Data Mining in Hydrology. *The Australian National University*.
- Spate, J., Croke, B., & Jakeman, A. (2016). Data mining in Hydrology. *Department of Mathematics, The Australian National University*. Australia: Canberra Act 0200.
- Stelhi, P. T., Weber, J. F., & Redin Vestena, L. (2016). Aplicación de un modelo hidrológico distribuido en una cuenca urbana. *3er Encuentro de Investigadores en Formación de Recursos Hídricos*. Córdoba - Argentina: IFRH.
- Urciuolo, A., Iturraspe, R., Sandoval, S., & Parson, A. (2002). Estudio de Técnicas apropiadas para modelar aplicaciones de Hidroinformática en el contexto de los Sistemas de Información Ambiental. *Workshop de Investigadores en Ciencias de la Computación*.

Valcarcel Ascencios, V. (2004). Data mining y el descubrimiento del conocimiento. *Revista de la facultad de ingeniería industrial*.

Vázquez Zambrano, R. F. (2010). Modelación hidrológica de una microcuenca Altoandina ubicada en el Austro Ecuatoriano. *Universidad de Cuenca*. Ecuador: Maskana.

Vélez, J. J. (2008). Modelación en hidrología. *III Simposio Nacional. Actualidad y Tendencias de la Ingeniería*. Manizales12: Researchgate.

Weber, J. F., & Jorquera, E. (2011). Estimación de Parámetros y Análisis de Incertidumbre de un Modelo Hidrológico Continuo para dos Cuencas de la Región Serrana de Córdoba. *XXIII Congreso Nacional del Agua*. Córdoba: Conagua.

Zimmerman, E., & Riccardi, G. (2003). Modelo hidrológico superficial y subterráneo desarrollado para la simulacion de sistemas de llanura. *Boletín Geológico y Minero*. Santa Fe: Dialnet.