

# INFORME DIVULGATIVO

## ESTUDIO SOBRE EL PAPEL DE LA DIGITALIZACIÓN EN LA GESTIÓN DE LAS INFRAESTRUCTURAS HÍDRICAS DE LA COMUNITAT VALENCIANA



Càtedra de  
Transformació del  
Model Econòmic  
Economia Circular  
en el Sector de l'Aigua



Xarxa  
Càtedres de  
Transformació  
del Model Econòmic



GENERALITAT  
VALENCIANA  
Conselleria d'Hisenda  
i Model Econòmic



UNIVERSITAT  
DE VALÈNCIA



UNIVERSITAT  
POLITÀCNICA  
DE VALÈNCIA



Universitat d'Alacant  
Universidad de Alicante



UJI UNIVERSITAT  
JAUME I



UNIVERSITAS  
Miguel Hernández

## RESUMEN EJECUTIVO

En la actualidad, las economías se enfrentan a cambios en los patrones climáticos que interrumpen gravemente la vida de las personas y la estructura de la sociedad. El reto del cambio climático y la degradación del medio ambiente son cuestiones donde el potencial de la digitalización puede ayudar a comprender y actuar de manera más eficaz y eficiente. El sector de los recursos naturales siempre ha sido un elemento muy importante en la economía mundial. La gestión de los recursos naturales requiere metodologías innovadoras que permitan monitorizar y entender el ciclo de vida de los productos, bienes y servicios, así como identificar las oportunidades de reconversión de los residuos generados en materias primas para otros procesos industriales. El desarrollo de las tecnologías conduce al avance del uso racional de los recursos naturales y la protección del medio ambiente.

Con tal de avanzar en la sostenibilidad de los recursos se requieren, además de datos, metodologías que permitan estructurarlos y procesarlos con el fin de obtener información de valor. En numerosas ocasiones la cantidad de datos se relaciona con la calidad de información y precisamente, puede ocurrir lo contrario. La información clave puede quedar escondida entre enormes bases de datos de escasa utilidad. Las principales características de los datos de calidad vienen definidas por su exactitud, fiabilidad, relevancia, actualidad, claridad y precisión. Si bien, una vez obtenidos, se requiere además de metodologías capaces de procesarlos, tener muy clara la finalidad o el tipo de decisiones que deseamos tomar. Un objetivo claro, con respecto a la investigación que se desea realizar, permitirá clasificar, ordenar y desechar los datos redundantes con tal obtener finalmente información útil.

La potencialidad de digitalizar y aprender de los datos es muy elevada. Son numerosos los sectores que implementan en sus procesos soluciones tecnológicas, por ejemplo, en el sector agrícola se usan sensores para monitorizar las variables que afectan a la producción, optimizando de este modo la obtención de alimentos, aumentar la precisión de la aplicación de fertilizantes, pesticidas y herbicidas, determinar fechas óptimas de siembra de cultivos, o para ayudar a identificar y eliminar malezas. En el sector pesquero los sistemas satelitales basados en sensores ópticos proporcionan numerosos

parámetros físicos del entorno acuático y localización de embarcaciones pesqueras, los drones submarinos determinan las mejores áreas para la pesca, así como otros dispositivos analizan el volumen y las características biofísicas de los peces capturados generando una elevada riqueza de información en la cadena de suministro. Del mismo modo, en el sector ganadero, los datos recopilados por sensores (como cámaras, micrófonos, acelerómetros, analizadores de gases...) sobre animales o sobre su entorno, junto con técnicas analíticas avanzadas, proporcionan herramientas eficientes para monitorear animales para mejorar su bienestar y optimizar el uso de recursos, como alimentación, agua y tierra. Incluso algunas tareas son realizadas por robots, como por ejemplo el ordeño del ganado o los dispensadores automáticos de alimentación.

La estrategia digital de la UE pretende que esta transformación y adaptación digital funcione tanto para las personas y como para las empresas antes del 2050, contribuyendo, al mismo tiempo, al objetivo de conseguir una Europa climáticamente neutra. Las tecnologías digitales ofrecen toda una serie de beneficios tangibles no solamente económicos, también sociales y ambientales, aumentando aún más las ventajas de su implementación. Esto es debido a que ofrecen un gran abanico de soluciones que a su vez permiten reducir los impactos ambientales generados y reasignar los recursos de una manera más eficiente. Por este motivo, algunos sectores como el de la energía, los residuos o el agua presentan amplias posibilidades en la integración de estas herramientas digitales en sus procesos.

Con respecto al sector del ciclo urbano del agua, el control y la correcta gestión del uso del agua en España es un desafío constante, en el que las distintas administraciones trabajan de manera coordinada para su correcta gestión. El plan de recuperación transformación y resiliencia (PERTE) del gobierno de España impulsa el uso de las nuevas tecnologías de la información en el ciclo integral del agua, lo que permitirá mejorar su gestión, aumentar su eficiencia, reducir las pérdidas en las redes de suministro y avanzar en el cumplimiento de los objetivos ambientales marcados por la planificación hidrológica y las normativas internacionales. Para ello, Financiará programas de ayudas para el impulso a la digitalización de los distintos usuarios del agua. El principal objetivo de estas subvenciones es avanzar en la protección del medio ambiente, la mejora de la gestión de los recursos hídricos y la lucha contra el cambio climático y el reto demográfico.

El ciclo urbano del agua ofrece grandes oportunidades para la implementación de la digitalización a lo largo de todas sus etapas. Un adecuado proceso digital de los flujos de información permite un mayor control de la logística a todos los niveles, desde la predicción de la demanda de agua hasta el control de residuos generados con tal asegurar su reutilización dentro de los canales apropiados. Para garantizar el suministro de agua se requiere de una serie de infraestructuras que permitan almacenar, potabilizar, distribuir y depurar el agua. Al conjunto de procesos que transcurren desde la captación y suministro hasta su depuración y vertido se denomina ciclo urbano del agua.

El proceso de captación y tratamiento de agua permite generar numerosos datos que son captados a través de sensores, tales como caudales, energía, reactivos y mantenimientos requeridos, entre otros. La digitalización permite monitorizar su distribución hasta los hogares, industrias y comercios. Su constante monitorización facilita a los operadores prever los picos de demanda, adecuando de este modo los caudales tratados y maximizando los tiempos de disponibilidad. En este sentido, facilita la detección de posibles fugas en la red, así como posibles desviaciones o anomalías en los consumos de los usuarios, mejorando la eficiencia hídrica del sistema lo que repercute en un menor consumo energético de la red.

Otro ejemplo aplicado en la etapa de potabilización y distribución del agua es el relativo a la instalación de contadores inteligentes y la sectorización de la red de distribución. En materia de tele lectura, los contadores inteligentes permiten a los usuarios consultar el consumo propio de agua en tiempo real a través de las aplicaciones alojadas en la nube. El software ofrece un sistema de alarma por fugas interiores ante un incremento súbito del consumo habitual, detecta posibles ocupaciones en segundas residencias o situaciones de riesgo para personas vulnerables que viven solas, en el caso de detectarse un parón en el consumo. Los contadores permiten una facturación exacta del servicio, eliminando de este modo los consumos estimados por los hogares, industria o comercios, dotando de mayor transparencia y calidad la prestación del servicio.

A su vez, los datos generados están conectados a la plataforma para el abastecimiento (ETAPs), que recopila y analiza datos diarios sobre el relacionados con la cantidad de agua suministrada o los caudales y presiones en la red. Las pérdidas de agua potable en la red de distribución es uno de los mayores problemas para los operadores,

implican elevadas pérdidas económicas además de ser una fuente de ineficiencia. Esto es debido, entre otros aspectos, al elevado consumo energético necesario para su impulsión. La sectorización permite establecer un balance acotado por zonas geográficas, barrios y edificios, identifican y cuantifican las posibles pérdidas de agua potable y, en consecuencia, permite a los operadores actuar en áreas muy delimitadas. En este caso, se combinan herramientas digitales dedicadas a monitorizar los caudales de entrada y salida y la geolocalización de las redes. El posterior análisis de los datos permite establecer alarmas en tiempo real.

En la fase de alcantarillado y depuración de aguas residuales las herramientas digitales ofrecen de igual modo un gran número de posibilidades, por ejemplo, controlar los caudales de aguas residuales vertidos al alcantarillado, ya sea con el fin de detectar grandes fluctuaciones del volumen generado como picos de carga orgánica. Esta información resulta de elevada utilidad para las Estaciones de tratamiento de aguas residuales (EDARs), los datos recogidos permiten planificar las tareas de operación, ajustando los reactivos y la energía necesaria con el fin de garantizar el correcto funcionamiento del proceso.

La monitorización de los tratamientos en las EDARs permite automatizar los procesos con el fin de mejorar la eficiencia en estas infraestructuras. Los beneficios de la digitalización repercuten de manera directa disminuyendo los costes del proceso a la par que maximizan los beneficios obtenidos mediante el control de los residuos con tal de generar recursos para otros usos; fangos para la agricultura, fertilizantes y energía, entre otros.

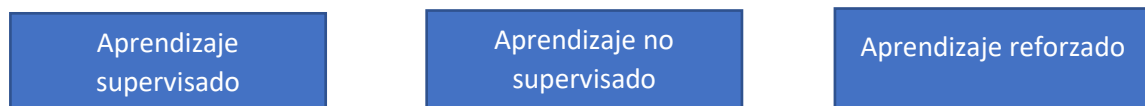
Desde el punto de vista operativo de las instalaciones, la aplicación del sistema de computación en la nube permite integrar volúmenes de datos generados a partir de los sistemas de monitorización y medición para la calidad del agua, el consumo de energía y el uso de reactivos. El sistema de computación en la nube procesa datos y visualiza la operación del proceso de depuración de las aguas residuales, el consumo de energía y el análisis de costes. De este modo, con ayuda de la digitalización, es posible generar un gemelo digital del proceso completo. Los gemelos digitales son generados para simular escenarios tales como; variaciones en los caudales, precipitaciones y cargas orgánicas, entre otros aspectos. Esta simulación permite al operador proyectar distintas alternativas

en el entorno real, favoreciendo de este modo la disminución de los costes y una mayor eficiencia energética a la par que minimiza riesgos de vertido.

Otro aspecto, en el que la digitalización presenta numerosas ventajas es en el de la gestión de los activos físicos de las instalaciones de forma centralizada. Su monitorización constante y la aplicación de algoritmos permite reducir el riesgo de avería y los costes asociados al ciclo de vida a partir de datos históricos, así como una mejor planificación de las tareas de mantenimiento más efectivas y el establecimiento de un plan de renovación e inversión en nuevos equipos e infraestructuras. Una de las metodologías más usadas para procesar y generar información a partir de los datos es el “*Machine learning*”.

El objetivo del aprendizaje automático es proporcionar alguna conclusión o predicción a partir del conjunto de datos disponible utilizando un modelo. Las técnicas de aprendizaje automático (ML) permiten usar estos datos con el fin de ayudar a identificar y clasificar aspectos que pueden resultar de interés. El objetivo es analizar la existencia de posibles patrones que pueden estar relacionados con un objetivo establecido. Principalmente se pueden clasificar en aprendizaje supervisado, aprendizaje no supervisado y aprendizaje reforzado.

*Ilustración 1 Tipologías de aprendizaje automático*



El aprendizaje supervisado tiene un enfoque reactivo debido a que el evento o consecuencia que se desea analizar es conocido de antemano. Por lo tanto, el principal objetivo es construir un modelo de aprendizaje a partir de datos de entrenamiento previamente etiquetados. Las observaciones etiquetadas permiten hacer predicciones futuras junto con el margen de error o tolerancia que el modelo genera. Un mayor número de observaciones permitirá mejorar los resultados del modelo, disminuyendo el margen de error asociado. En este caso, el aprendizaje supervisado puede devolver variables discretas (mediante tarea de clasificación) o valores continuos (regresión).

En el primer de los casos, variables discretas, se usa para predecir las etiquetas de clase categórica de nuevas instancias basadas en observaciones pasadas, por ejemplo, si deseamos conocer si una bomba de impulsión de agua fallará o no, daremos el valor 1 a

Fallo y el valor 0 al no Fallo. Estas etiquetas vienen definidas por una serie de variables explicativas asociadas a las bombas que estamos evaluando, éstas pueden ser las horas de operación, la edad del equipo, el tipo de agua residual, la existencia de cavitaciones etc.... de modo que, a partir de los datos de entrenamiento, el sistema ofrecerá una respuesta de 0 o 1 a partir de la similitud de las nuevas bombas evaluadas con observaciones pasadas. Esta tarea de clasificación permite usar más de etiquetas de clase, por lo que no tiene que ser necesariamente una clasificación binaria. Un ejemplo del uso de más etiquetas podría ser el relacionado con los consumos de agua, pudiendo clasificar el consumo mensual en intervalos que oscilen entre 1 y 10 y que éstos dependan de características relacionadas con las viviendas, superficie, zona residencial, número de baños, número promedio de personas en la vivienda etc...

Por otro lado, el aprendizaje supervisado puede predecir resultados continuos, es lo que conocemos con análisis de regresión. En este caso el objetivo es encontrar una relación entre las variables explicativas que permita predecir un resultado continuo. Para seleccionar la mejor combinación de variables capaces de explicar el modelo se usan técnicas estadísticas.

El aprendizaje no supervisado no cuenta con datos que definan un objetivo concreto, por lo que clasifican principalmente los datos a partir de patrones que permitan separar conjuntos similares. Esta técnica explora la estructura de los datos para extraer información significativa o similitudes entre los datos analizados que nos ayuden a comprender el patrón analizado. Para ello, el análisis recurre a la generación de grupos (clústeres), cada grupo viene definido por un grupo de observaciones que comparten cierto grado de semejanza.

Por último, el aprendizaje reforzado consiste en desarrollar un sistema que mejore su rendimiento, para ello su usa una señal de recompensa. De modo que, a través de la interacción con el entorno, un agente puede utilizar el aprendizaje reforzado para aprender una serie de acciones que maximicen esta recompensa.

Generar un modelo de Machine Learning implica un proceso, en puntos anteriores se menciona la importancia de recolectar y grabar datos para poder ser analizados con

posterioridad. La calidad de los datos resulta de elevada importancia, la siguiente ilustración (1) resume los distintos procesos necesarios.

*Ilustración 1 Fases construcción de un modelo Machine Learning*



1. Importación y muestreo: Los datos se pueden importar de fuentes tal como un sitio web, utilizando una API o una base de datos. Este paso es uno de los más complicados y requiere un tiempo determinado. A continuación, es importante realizar un primer muestreo de la tipología de los valores que incluye, numéricos, ordinales, dicotómicos.

2. Calidad de datos: A partir de un primer muestreo, es necesario asegurar la calidad de los datos para garantizar un rendimiento óptimo del algoritmo. Con este objetivo se lleva a cabo una exploración general y el cálculo de los estadísticos básicos, los nulos y ceros que contiene, posibles valores atípicos y un análisis de coherencia.

3. Transformación de datos: En este punto, se debe depurar valores, discretizar los datos, generar intervalos y seleccionar cuales son los más importantes para el objetivo que hemos diseñado. Seleccionar las variables target y las predictoras del modelo.

4. Modelización: los algoritmos de aprendizaje se alimentan con los datos que se procesaron en las etapas anteriores. La idea es que los algoritmos pueden extraer información útil de los datos iniciales. Paso previo a aplicar un método de clasificación,



es la partición del conjunto de datos en dos conjuntos de datos más pequeños que serán utilizadas con los siguientes fines: entrenamiento y test. El subconjunto de datos de entrenamiento es utilizado para estimar los parámetros del modelo y el subconjunto de datos de test se emplea para comprobar el comportamiento del modelo estimado.

5. Evaluación el algoritmo. Se realizan las pruebas de la información que genera el conocimiento del entrenamiento previo que se obtuvo a través del algoritmo. Se realiza una evaluación sobre la precisión del algoritmo en sus predicciones y, si no está satisfecho con su rendimiento, debe volver a la etapa anterior y continuar entrenando el algoritmo cambiando algunos parámetros hasta que se logre un rendimiento aceptable. Para comparar los distintos modelos generamos una unidad para medir el rendimiento. Esta unidad de medida utiliza la precisión en la clasificación. Un método para evaluar clasificador alternativo a la métrica expuesta es la curva ROC (Receiver Operating Characteristic). La curva ROC es una representación gráfica del rendimiento del clasificador que muestra la distribución de las fracciones de verdaderos positivos y de falsos positivos. La fracción de verdaderos positivos se conoce como sensibilidad, sería la probabilidad de clasificar correctamente a un individuo cuyo estado real sea definido como positivo.

6. Implantación. Finalmente, tras haber seleccionado el modelo óptimo, se adapta el código con tal de asegurar la sincronización con la fuente de datos. Esta etapa garantizará la continua alimentación y aprendizaje del modelo a partir de las nuevas observaciones que se vayan generando.

El documento extenso realiza un ejercicio práctico aplicado a las bombas soplantes, estos equipos son ampliamente utilizados en las plantas de tratamiento de aguas residuales. Su objetivo es proporcionar un flujo de aire continuo al reactor, facilitando así la eliminación de la materia orgánica y los nutrientes que contiene el agua residual. El objetivo es conocer el estado de estos activos para optimizar las distintas tareas de mantenimiento para prolongar la vida útil de los equipos y establecer reemplazos o renovaciones parciales con tal de asegurar un correcto funcionamiento. Si bien, existen numerosos factores que pueden afectar a la condición del activo, muchos de ellos están relacionados con las propias tareas dedicadas a mantener los activos en óptimas condiciones.

El objetivo del estudio es calcular la probabilidad de que una bomba se averíe con la consiguiente parada del equipo, se usa el algoritmo de árboles de decisión. Para entender que causas pueden provocar las averías, se identifican aquellos factores que mayores implicaciones pueden tener el desgaste del equipo y la consecuente rotura. La muestra la componen un total de 400 bombas soplantes a lo largo de los últimos 10 años. El campo objetivo viene definido por una respuesta dicotómica (0) o (1), donde 0 significa que el equipo no ha sufrido avería y 1 que sí ha sufrido alguna avería. Con tal de conseguir una muestra balanceada se revisan los datos, el 43% de los equipos evaluados han sufrido alguna avería.

Los resultados confirman que las bombas que superan el 40% de vida útil según diseño tienen una mayor probabilidad de sufrir una avería (50%). Este resultado viene explicado por el desgaste que provoca el propio funcionamiento de la bomba. Si bien, existen otros factores que pueden acelerar la fatiga y producir una avería. La siguiente variable seleccionada sigue siendo el número de horas acumuladas, en este caso, sitúa la mayor probabilidad de avería para aquellas bombas soplantes que superan el 60% de vida útil. A partir de este punto, el árbol de bifurca en dos nodos, aquellas bombas que se sitúan entre el 40 y 60% de la vida útil puede sufrir una avería con una probabilidad del 27%. En cambio, aquellos equipos que están por encima de esa vida útil tienen una probabilidad de sufrir una avería mucho mayor (70%).

Además, aquellos equipos que tienen una frecuencia de mantenimiento inferior al 20% recomendado por el fabricante tienen una probabilidad del 65% de sufrir una avería. Por otro lado, en los equipos que tienen un mayor mantenimiento la probabilidad de avería disminuye hasta el 51%. El siguiente factor que puede generar mayor probabilidad de fallo es la presencia de arena en las aguas residuales. La arena tiene un efecto abrasivo que aumenta el desgaste de los componentes de la bomba. En este caso, la muestra analizada confirma que una mayor presencia de arena generaría un 52% más de probabilidades de sufrir una avería.

Como se observa en el ejemplo, el uso de Las técnicas de aprendizaje automático (ML) permiten convertir en información una elevada cantidad de datos. El presente documento revisa las potencialidades de la digitalización en el sector de los recursos hídricos, concretamente en el ciclo urbano del agua. Explica brevemente las diferentes

técnicas de aprendizaje e identifica los algoritmos más utilizados. Es importante señalar que estas metodologías son capaces de aprender por iteración, aumentando la calidad en la información que estos modelos son capaces de generar a partir de los datos. El aprendizaje de los modelos no se limita a un aspecto en concreto, la posibilidad de conectar nuevas fuentes de información y generar nuevas variables enriquece los resultados y abre nuevas vías de investigación. En conclusión, La aplicabilidad del “machine learning” en el sector de los recursos hídricos ocupa numerosas posibilidades.