

CIENCIA DE DATOS COMO HERRAMIENTA DE SOPORTE EN LA GESTIÓN PÚBLICA DE CALIDAD DEL AGUA

J. J. Bolano, M. G. Rey, U. Ramirez, J. G. A. Pautsch, E. Zamudio, H. D. Kuna

Instituto de Investigación Desarrollo e Innovación en Informática (IIDII)
Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones

hdkuna@gmail.com

Resumen

El agua es un recurso vital y como tal requiere una adecuada gestión de su calidad para la subsistencia de la población en todas sus dimensiones. En la **provincia de Misiones**, existen varias organizaciones que intervienen de alguna manera en la gestión de la calidad del agua, aunque se encuentran limitados respecto de la generación de información adecuada para la gestión, pese a esfuerzos individuales de algunas de estas organizaciones. La **gestión de calidad del agua** es un proceso que requiere datos. Estos datos son generados a un **alto costo**, ya que cada estudio implica un traslado al lugar donde se realizan las muestras, gastos en recursos humanos, y gastos en el análisis de las propiedades del agua para determinar su calidad, entre otros. En esta línea de investigación se aborda la aplicación de procesos de **ciencia de datos** que podrían contribuir en la generación de información dirigida a la gestión de calidad del agua. A través de un análisis **descriptivo** se determinaron parámetros asociados a la calidad del agua, destacándose la temperatura, turbiedad, conductividad, pH, P-ortoFosfatos, sólidos disueltos totales, oxígeno disuelto, sólidos suspendidos totales y coliformes fecales. Asimismo, se evidencia la necesidad de ampliar el **desarrollo de conjuntos de datos** que integren variables relevantes para la gestión de la calidad del agua.

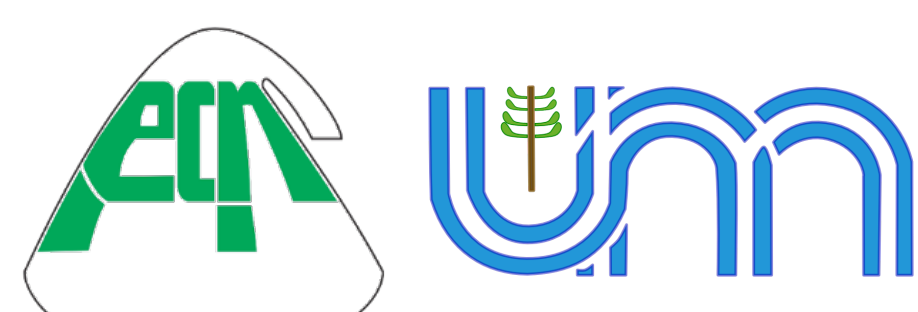
Líneas de I+D

Esta línea de trabajo aborda el desarrollo de procesos de ciencia de datos que permitan la identificación de patrones y generación de información dirigida a la toma de decisiones en relación a la gestión de la calidad del agua, incluyendo los siguientes objetivos.

- Identificar las **necesidades** específicas en relación a la gestión pública de la calidad del agua en el ámbito de la provincia de Misiones.
- Determinar las **fuentes y las características** de los datos asociados a la gestión pública de la calidad del agua.
- Desarrollar **mecanismos** de procesamiento de datos disponibles que permitan su uso en modelos de ciencia de datos dirigidos a la gestión de la calidad del agua.
- Desarrollar **estrategias** de ciencia de datos, incluyendo el entrenamiento de modelos de aprendizaje automático, para dar soporte a los requerimientos de información de la gestión de la calidad del agua.
- Evaluar el **impacto** de la implementación de los procesos desarrollados de ciencia de datos en organizaciones intervinientes en la gestión de calidad del agua.

Formación de Recursos Humanos

Las líneas de investigación presentadas cuentan con doce integrantes relacionados con las carreras de Ciencias de la Computación de la UNaM. El grupo de investigación desarrolla dos tesis de grado articulando sus trabajos con una beca de Estímulo a las Vocaciones Científicas del Consejo InterUniversitario Nacional (CIN) y una beca UNaM; dos tesis de maestría en curso y una finalizada. Asimismo, las líneas de investigación y sus integrantes se vinculan con grupos de la Universidad de Castilla-La Mancha, España y la Universidad de Sonora, México.



Actividades y Resultados

Se trabajó con el procesamiento de un conjunto de datos de calidad del agua del Río Paraná censados en el intervalo de tiempo para los años 1980 a 1990.

La ausencia de un marco de referencia para poder analizar el comportamiento de los datos incentivó a tomar la iniciativa en la **reconstrucción de los datos históricos** del Río Paraná a través de la articulación de los datos disponibles que disponen las organizaciones que se encargan de medir la calidad del agua del río Paraná.

Otro desafío que enfrenta este proyecto es la propia **naturaleza del Río Paraná**, que posee un gran nivel de auto-depuración debido a su tamaño y fluidez.

A partir de una revisión de la literatura relacionada con la predicción de la calidad del agua haciendo uso de técnicas de inteligencia computacional (IC), se detecta que la mayor proporción de trabajos estudian **métodos híbridos** que involucran más de una tecnología de IC para la construcción de un modelo predictivo de calidad de agua.

Contexto

Esta línea de investigación se desarrolla en el ámbito Programa de Investigación en Computación (PICom), perteneciente al **Instituto de Investigación Desarrollo e Innovación en Informática** de la Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones (**IIDII, FCEQyN, UNaM**). El PICom desarrolla líneas de investigación relacionadas con la explotación de información y la robótica.

El grupo de trabajo aborda principalmente las áreas relacionadas con el tratamiento de datos y el descubrimiento de información a partir de éstos mediante técnicas de ciencia de datos, en distintos campos de aplicación. Este contexto ha permitido contribuir con estrategias para la detección de datos anómalos, expansión de consultas, algoritmos de ranking, generación automática de perfiles y desambiguación de entidades, recomendación y selección de grupos de expertos.