



UNIVERSIDAD SANTO TOMÁS  
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

T U N J A



# ingenio Magno

# 8

*No. 1*



OPEN ACCESS  
descarga gratuita

<http://revistas.ustatunja.edu.co/index.php/ingeniomagno>

ISSN (versión impresa)

2145-9282

ISSN (en línea)

2422-2399

Enero-junio de 2017, vol. 8, no. 1

Universidad Santo Tomás

Tunja, Boyacá

Publicación semestral

Hecho el depósito que establece la ley

© Derechos reservados

Universidad Santo Tomás

Suscripción y canje

Unidad de Investigación

Cll. 19 No. 11-64

Universidad Santo Tomás, Tunja-Colombia

PBX: 744 0494

Línea gratuita: 018000 932340

desde cualquier lugar del país

<http://revistas.ustatunja.edu.co/index.php/ingeniomagno>

Los conceptos expresados en los artículos son de exclusiva responsabilidad de sus autores y no comprometen a la Institución o a la publicación.

---

División de Arquitectura e Ingenierías

---

INGENIO MAGNO	Tunja Colombia	Vol. 8 No. 1	pp. 1-172	Enero- junio	2017
---------------	-------------------	-----------------	-----------	-----------------	------



UNIVERSIDAD SANTO TOMÁS  
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA  
T U N J A



## Directivos

P. Jorge Ferdinando RODRÍGUEZ RUIZ, O.P.  
Rector

P. José Antonio BALAGUERA CEPEDA, O.P.  
Vicerrector Administrativo-Financiero

P. Javier Antonio CASTELLANOS, O.P.  
Vicerrector Académico

P. Samuel Elías FORERO BUITRAGO, O.P.  
Decano de División de Ingeniería y Arquitectura

Jimena BOHÓRQUEZ HERRERA, Ph.D.  
Directora Unidad de Investigaciones

Édgar Andrés GUTIÉRREZ CÁCERES, Esp.  
Director Centro de Investigaciones en Ingeniería  
San Alberto Magno - (CIAM)

## Editor

Fredy Andrés Aponte Novoa, M.Sc.  
Centro de Investigación en Ingeniería Alberto Magno  
(CIAM)  
[ingeniomagno@ustatunja.edu.co](mailto:ingeniomagno@ustatunja.edu.co)

## Equipo Editorial

Edwin Blasnilo Rúa Ramírez, M.Sc.  
Coeditor Universidad Santo Tomás, Tunja, Colombia  
edwin.rua@usantoto.edu.co

José Ricardo Casallas Gutiérrez, M.Sc.  
Coeditor Universidad Santo Tomás, Tunja, Colombia  
jose.casallas@usantoto.edu.co

John Fredy Guzmán Vargas  
Profesional en Filosofía y Letras  
Corrector de estilo, redactor  
fredyguzmanvargas@gmail.com

Albany Milena Lozano Násner, M.Sc.  
Traducción español-portugués  
mlnasner@gmail.com

Departamento de Idiomas  
Traducción español-inglés  
Universidad Santo Tomás, Tunja, Colombia  
coordinacion.idiomas@ustatunja.edu.co

David Enrique González Camargo  
Administrador OJS  
Universidad Santo Tomás, Tunja, Colombia  
ing.investigacion@ustatunja.edu.co

## Comité Científico

Antonio Moreira Teixeira, Ph.D.  
Universidad de Aberta (Lisboa, Portugal)  
antonio.teixeira@uab.pt

Electo Eduardo Silva Lora, Ph.D.  
Univerdidad Federal de Itajubá (Itajubá M.G., Brasil)  
electo@unifei.edu.br

María Julia Mazzarino, Ph.D.  
Universidad de Buenos Aires (Buenos Aires, Argentina)  
mmazzari@crub.uncoma.edu.ar

Antonio Rico Sulayes, Ph.D.  
Universidad de las Américas Puebla (Puebla, México)  
antonio.rico@udlap.mx

Carlos Enrique Montenegro Marín, Ph.D.  
Universidad Distrital Francisco José de Caldas (Bogotá D.C., Colombia)  
cemontenegrom@udistrital.edu.co

César Darío Guerrero Santander, Ph.D.  
Universidad Autónoma de Bucaramanga (Bucaramanga - Santander, Colombia)  
cguerrer@unab.edu.co

## Comité Editorial de la Revista

Antonio José Bula Silvera, Ph.D.  
Universidad Católica del Norte (Barranquilla - Atlántico, Colombia)  
abula@uinorte.edu.co

Wilson Javier Pérez Holguín, Ph.D.  
Universidad Pedagógica y Tecnológica de Colombia (Tunja - Boyacá, Colombia)  
wilson.perez@uptc.edu.co

José Carlos Escobar Palacios, Ph.D.  
Investigador Universidad Federal de Itajubá (Itajubá M.G., Brasil)  
jocesobar@unifei.edu.br

Camilo Andrés Lesmes Fabian, Ph.D.  
Universidad Santo Tomás (Tunja - Boyacá, Colombia)  
camilo.lesmes@usantoto.edu.co

## Impresión

Editorial Jotamar Ltda.  
Calle 57 No. 3-39  
Tel.: (8) 745 7120  
editorialjotamar@yahoo.com  
Tunja - Boyacá - Colombia

# Índice de calidad del agua recolectada en el río Bogotá: un análisis mediante la computación cognitiva Watson

## Index of the quality of water collected from Bogotá river: An analysis using Watson cognitive computation

## Índice de qualidade da água do Rio Bogotá: uma análise usando a computação cognitiva Watson

Para citar este artículo / To reference this article / Para citar este artigo: Barrera Williams, C., Montenegro Marín, C. E. y Gaona García, P. A. (2017). Índice de calidad del agua recolectada en el río Bogotá: un análisis mediante la computación cognitiva Watson. *Ingenio Magno*, 8(1), 104-117.

### **Carolina Barrera-Williams**

Universidad Distrital Francisco José de Caldas,  
Ingeniería de Sistemas, Bogotá, Colombia  
[cbarreraw@correo.udistrital.edu.co](mailto:cbarreraw@correo.udistrital.edu.co)

### **Carlos Enrique Montenegro-Marín**

Universidad Distrital Francisco José de Caldas,  
Ingeniería de Sistemas, Bogotá, Colombia  
[cemontenegrom@udistrital.edu.co](mailto:cemontenegrom@udistrital.edu.co)

### **Paulo Alonso Gaona-García**

Universidad Distrital Francisco José de Caldas,  
Ingeniería de Sistemas, Bogotá, Colombia  
[pagaonag@udistrital.edu.co](mailto:pagaonag@udistrital.edu.co)



## Resumen

En este artículo se analizan los datos sobre la calidad del agua del río Bogotá en el periodo 2008-2015 proporcionados por la Corporación Autónoma Regional de Cundinamarca (CAR); para ello se aplica la computación cognitiva, con el fin de establecer los cambios más significativos que se han presentado. Además, se hace un análisis espacio-tiempo para las variables relacionadas en el recurso hídrico en la cuenca alta, para así contribuir con un mecanismo que facilite la comprensión del comportamiento, a través del tiempo, del estado de la calidad del agua. Esto hará posible que las entidades territoriales tengan nuevos criterios y visiones a la hora de formular o reformular nuevos planes. Como herramienta de análisis se utiliza IBM Watson.

**Palabras clave:** análisis de datos, computación cognitiva, índice de calidad del agua (ICA), río Bogotá, variables de impacto, Watson.

## Abstract

This article analyzes the data on the water quality of the Bogotá river during the period 2008-2015, provided by the Autonomous Regional Corporation of Cundinamarca (CAR); For this purpose cognitive computation is applied, in order to establish the most significant changes that have been presented. In addition, a space-time analysis is performed of the related variables in the water resource in the upper basin, in order to contribute to a mechanism that facilitates the understanding of the behavior, over time, of the state of water quality. This will make it possible for territorial entities to have new criteria and visions when formulating or reformulating new plans. IBM Watson is used as an analysis tool.

**Keywords:** data analysis, cognitive computation, water quality index (ICA), Bogotá river, impact variables, Watson.

## Resumo

Este artigo analisa os dados da qualidade da água do rio Bogotá no período 2008-2015, fornecidos pela Corporação Regional Autônoma de Cundinamarca (CAR), na Colômbia; Para este fim, a computação cognitiva é aplicada, com o objetivo de estabelecer as mudanças mais significativas que têm se apresentado. Além disso, uma análise espaço-temporal foi realizada para as variáveis relacionadas no recurso hídrico na bacia superior, a fim de contribuir com um mecanismo que facilite a compreensão do comportamento do estado da qualidade da água, ao longo do tempo. Isso permitirá que as entidades territoriais tenham novos critérios e visões para formular ou reformular novos planos. Como uma ferramenta do análise, o IBM Watson foi usado.

**Palavras chave:** análise de dados, computação cognitiva, índice de qualidade da água (ICA), rio de Bogotá, variáveis de impacto, Watson.

## I. Introducción

Señala la Alcaldía Mayor de Bogotá (2007) sobre el río Bogotá:

El río Bogotá nace a unos 3400 m.s.n.m., en el Alto de la Calavera, municipio de Villapinzón, desemboca en el río Magdalena a una altura de 280 m.s.n.m., en el municipio de Girardot. En este tramo, drena una superficie de 599.561 hectáreas, siendo sus principales contribuyentes los ríos San Francisco, Sisga, Siecha, Tibitó, Teusacá, Chicú, Juan Amarillo, Fucha, Tunjuelito, Balsillas, Soacha y Muña en la cuenca alta, Calandaima y Apulo en su cuenca baja.

Uno de los principales elementos del sistema hídrico de la capital del país es el río Bogotá. Este río se encuentra en un estado muy avanzado de contaminación, debido a la incesante recepción de aguas residuales, industriales y domésticas de sus respectivas cuencas de drenaje. Las muestras recolectadas presentan valores muy altos de carga orgánica y carencia de oxígeno disuelto en sus aguas, en las desembocaduras (Alcaldía Mayor de Bogotá, 1997).

Por todo lo anterior, la Corporación Autónoma Regional de Cundinamarca (CAR), entidad encargada de velar por el adecuado uso los recursos naturales del departamento, realiza el monitoreo de la calidad de agua a lo largo de toda la cuenca, en cada una de las 81 estaciones de monitoreo. Este monitoreo es realizado dos veces al año con el fin de contribuir con información, la cual ha sido utilizada como base para la formulación de planes o proyectos en la cuenca alta, en cada uno de sus tramos. De acuerdo con lo anterior, esta investigación pretende analizar el estado de la calidad de agua de los últimos 7 años en la cuenca alta del río Bogotá, comprendida entre

Villapinzón y el Puente de la Virgen, en Cota, para así determinar el estado de la calidad de agua de los puntos de monitoreo por año. Para ello, se aplicaran técnicas de la computación cognitiva para su análisis y se realizará una comparación entre los resultados de calidad de agua versus lo recomendado por la normatividad ambiental vigente en relación con este tema.

Este estudio se realizará por medio de la computación cognitiva, la cual permite que un sistema de tecnología pueda entender las cosas de una manera muy similar a como lo hacen los humanos: aprendizaje y experiencia. Estos sistemas permanecen en un aprendizaje continuo a través de cada interacción, por lo que ganan valor y conocimiento a través del tiempo. Como herramienta de la computación cognitiva se usará Watson, un “sistema informático de inteligencia artificial que es capaz de responder a preguntas formuladas en lenguaje natural desarrollado por IBM” (The DeepQA Research Team - IBM, 2013).

“IBM Watson es la primera plataforma de tecnología de computación cognitiva abierta. Forma parte del proyecto de investigación DeepQA, liderado por David Ferrucci. Lleva su nombre en honor del fundador y primer presidente de IBM, Thomas J. Watson” (IBM, s. f.). Watson estructura una base de datos almacenada localmente, mediante el análisis de datos estructurados y no estructurados, hasta el punto de preguntas complejas y presentar respuestas simples y soluciones predictivas.

## II. El problema de la calidad del agua en el río Bogotá

La falta de planeación urbana alrededor de las cuencas altas, media y baja del río Bogotá ha llevado a que el recurso hídrico se vea alterado de manera negativa. Aunque si bien existen

planes y estrategias para mejorar la calidad del agua para las cuencas media y alta, la cuenca baja tiene características ambientales y sociales diferentes que hace que la población cercana al cauce principal del río se vea afectada de forma diferente a la población en la parte superior de la cuenca.

El deterioro en la calidad del agua en el río Bogotá ha originado problemas ambientales evidentes que es preciso atender de manera apropiada y oportuna; sin embargo, es necesario tener en cuenta que las necesidades del recurso hídrico para cada uno de los tramos de la cuenca son diferentes; por tanto, el manejo y las estrategias que se puedan formular para cada tramo del río deben ser únicas y particulares, atendiendo así las necesidades que se presenten.

Por esto, realizar un análisis espacio-tiempo permitirá establecer con mayor certeza las estrategias adecuadas para una administración adecuada del recurso natural, de manera tal que las características ambientales y naturales

del río permanezcan en el tiempo, como también que se logre un balance entre la oferta y la demanda del recurso en beneficio de cada uno de los actores involucrados.

El Decreto 1575 del 2007 del Ministerio de la Protección Social, “por el cual se establece el Sistema para la Protección y Control de la Calidad del Agua para Consumo Humano”, estipula los parámetros del recurso hídrico para garantizar la salud de los consumidores. Realizar el pertinente análisis en la calidad del agua, de acuerdo con estos parámetros establecidos, permitirá consolidar una buena evidencia para el diseño de estrategias oportunas en aras de promover un mejor uso del recurso (Ministerio de la Protección Social, 2007).

La CAR, organismo encargado de la gestión integral de este recurso hídrico, ha establecido una serie de puntos de monitoreo de calidad de agua para realizar un debido seguimiento y control; para efectos de esta investigación, los 32 puntos por estudiar corresponden a la cuenca alta del río y son descritos en la tabla 1.

**Tabla 1.** Puntos de monitoreo en la cuenca alta río Bogotá

No. Punto	Nombre	Norte	Este	Abscisado	Cota
1	Aguas Arriba Villapinzón	1055836	1070203	K0+000	2802
2	Puente Villapinzón	1053523	1068988	K4+146	2771
3	Aguas arriba Q. Quincha	1053282	1068706	K4+506	2769
4	Q. Quincha	1053598	1068453	K4+886	2723
5	Estación LM Chingacio	1051139	1066295	K7+842	2787
6	Agregados Chocontá	1045598	1062363	K15+460	2648
7	Río Tejar	1044769	1061491	K16+810	2627
8	Puente Vía Telecom	1044134	1061555	K17+340	2625
9	Descarga Municipio Chocontá	1042676	1060785	K19+100	2626
10	Aguas abajo Municipio Chocontá	1042532	1060812	K19+260	2643

11	Estación LG Saucio	1041097	1056901	K24+840	2634
12	Descarga Embalse Sisga	1038962	1057069	K27+130	2630
13	Estación LM Santa Rosita	1036255	1056612	K30+270	2622
14	Puente Santander	1031591	1055100	K35+515	2566
15	Descarga Municipio Suesca	1031102	1055197	K36+015	2568
16	Aguas abajo Municipio Suesca	1031155	1054269	K37+065	2561
17	Descarga Embalse Tominé	1031692	1051077	K39+755	2565
18	Aguas arriba descarga Papeles y Molinos	1026455	1049171	K48+155	2556
19	Estación LG – Puente Florencia	1024543	1048112	K51+515	2554
20	Descarga Municipio de Gachancipá	1021962	1043839	K58+565	2553
21	Aguas abajo Municipio Gachancipá	1021344	1043552	K59+495	2559
22	Estación LM – Tocancipá	1017904	1041499	K64+365	2572
23	Descarga Municipio Tocancipá	1017744	1041206	K64+755	2553
24	Aguas arriba Terkoziza	1015132	1042109	K69+415	2570
25	Descarga Terkoziza	1014658	1042304	K70+075	2570
26	Hacienda El Triunfo	1012820	1041746	K72+615	2553
27	Río Neusa	1012270	1044152	K75+230	2548
28	Estación LG – El Espino	1011628	1044139	K75+750	2552
29	Río Negro	1010323	1044001	K77+250	2559
30	Aguas Abajo Río Negro	1010165	1043622	K77+550	2560
31	Río Teusacá	1008465	1038475	K83+850	2547
32	Estación LG – Puente Vargas	1007836	1035833	K87+320	2545

Fuente: CAR (2013).

### III. Estado del arte

En el mundo se han desarrollado diversos métodos de análisis de datos que permiten obtener información de las diversas variables presentes en los ríos, ya que el estudio de estas es de valiosa importancia en la conservación de los recursos naturales (Hammond *et al.*, 1995). La correcta identificación de las condiciones de calidad del agua en un sistema fluvial, basado en observaciones, es una labor fundamental para alcanzar los objetivos de la gestión ambiental (Chang, Chen y Ning, 2001).

La CAR, como autoridad ambiental competente en el área de estudio, ha sido el ente encargado del monitoreo tanto de la calidad hídrica del río Bogotá como de sus afluentes. Con el pasar de los años se ha venido ejecutando un seguimiento a las variables físico-químicas del río, las cuales se han organizado en los *Boletines de Calidad Hídrica*, que son publicaciones anuales en las que se muestran las acciones realizadas en dirección, evaluación y seguimiento ambiental.



El muestreo ambiental es muy complejo y costoso, y requiere invertir una muy buena cantidad de tiempo. También es poco probable que se obtengan datos continuos a largo plazo, con propiedades completas, en todos los lugares de muestreo en un sistema fluvial (Chang *et al.*, 2015). En consecuencia, es necesario utilizar programas o algoritmos de predicción que pueden ser muy útiles para realizar la extracción de conocimiento en muchos ámbitos gracias a su función, ya que ello permitirá extraer información importante y relevante a partir de una base de datos (Roland, Uhrmacher y Saha, 2009), utilizando métodos para encontrar patrones y diseñar modelos con dicha información y, simultáneamente, aumentar la fiabilidad de los modelos (Rosado y Verjel, 2015).

Los métodos tradicionales para realizar la medición de la calidad del agua no son fáciles de usar; lo que se necesita son técnicas de aprendizaje inteligente. En estudios anteriores se han demostrado diferentes categorías de técnicas de análisis como redes neuronales artificiales (RNA), sistemas difusos y diferentes algoritmos basados en ellos, orientando su aplicación hacia la gestión de la calidad del agua (Chau, 2006). Las perspectivas tradicionales habitualmente utilizan conceptos de conjunto clásicos, pero, en la vida real, en lugar de tener valores fijos, los parámetros tienen rango de valores y conceptos establecidos, de manera que muchas veces conducen a resultados erróneos. Para superar este problema, Prato (2005) ha demostrado la aplicabilidad de proposiciones difusas para los atributos.

En Lobbrecht, Dibike y Solomatine (2002) se han estudiado las aplicaciones de redes neuronales, lógica difusa y los enfoques híbridos hacia la gestión de la calidad del agua. Los

autores han reconocido la aplicabilidad de los diferentes modelos en fuentes hídricas. Por su parte, la Corporación Colombia Digital (2015) ha indicado:

Los sistemas de computación cognitiva integran ciencias como nanotecnología, neurociencias y supercomputación para hacer realidad el nivel de análisis esperado, facilitando la toma de decisiones a partir de patrones y relaciones directas. Así, un proyecto se puede orientar al futuro para ofrecer una solución real para diversos proyectos ambientales, empresariales, entre otros.

Como se evidencia en la investigación realizada, no existe un análisis mediante computación cognitiva que emplee Watson a fuentes hídricas y estudie calidad del agua en los puntos de monitoreo a través del espacio-tiempo.

#### IV. Desarrollo

Watson es un sistema desarrollado por IBM para la búsqueda de respuestas. La Corporación lo describe de este modo:

Una aplicación de tecnologías avanzadas diseñadas para el procesamiento de lenguajes naturales, la recuperación de información, la representación del conocimiento, el razonamiento automático, y el aprendizaje automático al campo abierto de búsquedas de respuestas, [...] [que es] construido empleando la tecnología DeepQA de IBM para la generación de hipótesis, la recopilación de pruebas masivas, el análisis y la calificación (The DeepQA Research Team - IBM, 2013).

Se decide utilizar Watson como un sistema avanzado que es capaz de procesar cantidades masivas de datos estructurados y no estructurados, y producir conocimientos adicio-

nales de forma rápida y precisa. “Watson está optimizado para realizar un análisis complejo, este se hace posible por la integración de procesadores masivamente paralelos de POWER7 y el software DeepQA” (IBM Watson, 2011).

IBM Watson, creado sobre Linux on Power, aprovecha el profundo análisis de contenido y las deducciones lógicas basadas en pruebas

para acelerar y mejorar las decisiones, reducir los costes y optimizar los resultados. Watson utiliza el lenguaje natural para la generación de hipótesis y el aprendizaje basado en pruebas. Adicionalmente, Watson, combinado con las técnicas de procesamiento paralelo masivo, genera una nueva forma para encontrar soluciones rápidas a los problemas (IBM, 2013).

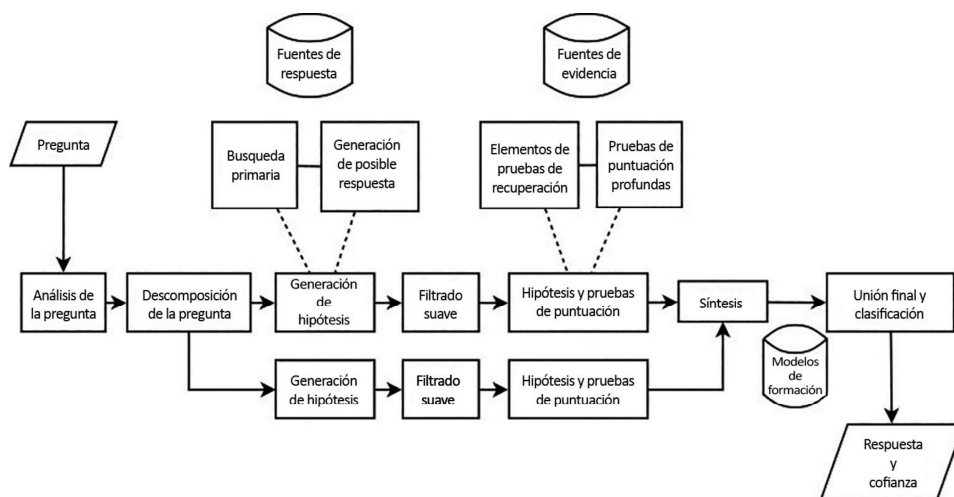


Figura 1. La arquitectura de alto nivel empleada por la tecnología DeepQA, utilizada específicamente por OBM para Watson

Fuente: Ferrucci *et al.* (2010).

La arquitectura utilizada por DeepQA puede observarse en la figura 1; esta es la que se emplea para Watson. El primer paso en cualquier aplicación de DeepQA para resolver un problema es la identificación y recopilación de contenido que se usará para las fuentes de respuesta. Esta adquisición de contenidos es una combinación de pasos manuales y de pasos automáticos. En el proceso de adquisición de contenidos es importante identificar y recopilar estos recursos, que incluyen bases de datos, taxonomías, ontologías, etc. El segundo paso es analizar las preguntas, el espacio del problema, la descripción de los tipos de consultas que deben ser contestadas y una caracterización del dominio de aplicación. El análisis del

sistema intenta entender lo que pide la pregunta inicial y realizar los estudios que determinan cómo será procesada por el resto del sistema (McCord, 1990).

El análisis que realiza Watson en las bases de datos existentes depende de la capacidad de estudiar y detectar sus relaciones. Otro paso es la generación de hipótesis, que toma los resultados del análisis de la pregunta y produce posibles respuestas con cierto grado de confianza. El proceso que se realiza en este trabajo utiliza la técnica de modelado originado por los datos, pues inicialmente se tienen aproximadamente 15.000 registros de la cuenta alta en documentos tipo PDF, con diversas variables contenidas en el agua en diferentes

cantidades y distribuidas en diversas zonas por las que pasa la fuente hídrica. En este sentido, Watson Analytics es una herramienta de análisis de datos ejecutada sobre el sistema computacional Watson, basados en el proyecto DeepQA de IBM.

### A. Obtención de los datos

Los datos utilizados en este estudio fueron extraídos a partir de los resultados de las mediciones obtenidas del río Bogotá y sus afluentes durante el periodo 2008-2015, los cuales son organizados en registros semestrales que presentan los niveles encontrados de cada una de las variables medidas, en cada uno de los puntos de medición correspondientes, en los cuales existen los sensores de información remota o por formas de medición in situ para aquellas variables que requieren de análisis en laboratorio.

A partir de la recopilación documental de los *Boletines de Calidad Hídrica* del río Bogotá, se procedió a establecer la conversión del formato de documentos PDF a un formato de .xlsx o .xls, desde el cual se realiza el análisis de los datos.

### B. Selección de los datos

Se cuenta con una base de datos de 51 variables, que son parámetros físico-químicos contenidos en el agua del río Bogotá. Todas estas variables tienen un porcentaje de afectación sobre la calidad del agua; es decir, la existencia en mayor o menor medida de algunas de estas variables define si el agua en ese punto de monitoreo es buena o mala para los diferentes usos, como el consumo doméstico, agricultura, ganadería, entre otros. Ahora bien, el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM, 2011) indica:

Para el cálculo del Índice de calidad del agua ICA en Colombia se utiliza una expresión numérica agregada y simplificada surgida de la sumatoria aritmética equiponderada de los valores que se obtienen al medir la concentración de cinco o seis variables físico-químicas básicas en las estaciones de monitoreo y que evalúan la calidad del agua en las corrientes superficiales.

La fórmula del cálculo del índice de la calidad de agua (ICA) se muestra en la ecuación [1]:

$$ICA_{njt} = \left( \sum_{i=1}^n W_i * I_{ikjt} \right) \quad [1]$$

Donde  $ICA_{njt}$  es el índice de calidad del agua de una determinada corriente superficial en la estación de monitoreo  $j$ , en el tiempo  $t$ , evaluado con base en  $n$  variables.  $W_i$  es el ponderador o peso relativo asignado a la variable de calidad  $i$ .  $I_{ikjt}$  es el valor calculado de la variable  $i$  (obtenido de aplicar la curva funcional o ecuación correspondiente), en la estación de monitoreo  $j$ , registrado durante la medición realizada en el trimestre  $k$ , del periodo de tiempo  $t$ . Por último,  $n$  es el número de variables de calidad en el cálculo del indicador;  $n$  es igual a 5 o 6, según la medición del ICA seleccionada (para este caso,  $n$  es igual a 5).

De acuerdo con el IDEAM (2011), las cinco variables seleccionadas para el cálculo del ICA son:

1. *Oxígeno disuelto (OD)*. Define la presencia o ausencia potencial de especies acuáticas.
2. *Sólidos suspendidos totales (SST)*. La presencia de sólidos en suspensión en los cuerpos de agua indica cambio en el estado de las condiciones hidrológicas de la corriente. Dicha presencia puede estar relacionada con procesos erosivos, vertimientos industriales, extracción de materiales y disposición de escombros. Tiene una relación directa con la turbiedad.

3. *Demanda química de oxígeno (DQO)*. Refleja la presencia de sustancias químicas susceptibles de ser oxidadas a condiciones fuertemente ácidas y alta temperatura, como la materia orgánica, ya sea biodegradable o no, y la materia inorgánica.
4. *Conductividad eléctrica (CE)*. Está íntimamente relacionada con la suma de cationes y aniones determinada en forma química; refleja la mineralización.
5. *pH*. Mide la acidez; valores extremos pue-

den afectar la flora y fauna acuáticas.

“Los valores calculados se comparan con los que se establecen en las tablas de interpretación y de esta forma se permite clasificar la calidad del agua de forma descriptiva en una de cinco categorías (buena, aceptable, regular, mala o muy mala), que a su vez se asocian a un color (azul, verde, amarillo, naranja y rojo, respectivamente)” (IDEAM, 2011). En la tabla 2 se registra la relación entre valores y calificación.

**Tabla 2.** Calificación de la calidad del agua según los valores que tome el ICA

Categorías de valores que pueden tomar el indicador	Calificación de la calidad del agua	Señal de alerta
0,00 – 0,25	Muy mala	Rojo
0,26 – 0,50	Mala	Naranja
0,51 – 0,70	Regular	Amarillo
0,71 – 0,90	Aceptable	Verde
0,91 – 1,00	Buena	Azul

Fuente: IDEAM (2011).

Dada la importancia que tiene la calidad del agua, que el índice de calidad es afectado por otras variables y que con base en este índice se determina el uso del agua, se seleccionará como objetivo predecir únicamente la calidad de agua en diferentes puntos de monitoreo.

### C. Depuración de los datos

Una vez se han calculado los ICA de toda la red de monitoreo del río Bogotá, desde el primer semestre de 2008 hasta el primer semestre de 2015, se observa que los datos se encuentran ordenados alfabéticamente y no en el orden geográfico en el que deberían estar; por ello, se opta por realizar una clasificación de los 81 puntos de monitoreo en su respectiva cuenca alta, media o baja, y se procede a agregar el orden geográfico de todos los puntos de monitoreo. Por último, se procede a filtrar los

valores que pertenecen a la cuenca alta del río para su respectivo análisis.

A partir de esta nueva organización se puede inferir que la mayor cantidad de puntos de monitoreo se encuentran en la cuenca media, con 34 puntos, seguido de la cuenca alta, con 32 puntos, y, por último, la cuenca baja, con 15 puntos. Asimismo, se observa que en la cuenca alta hay una mayor calidad del agua, ya que es donde nace la fuente hídrica.

Al realizar el cargue de los datos en Watson, se realiza una depuración más de los datos a partir de un análisis de su calidad en los registros y su rectificación, de acuerdo con los valores límite determinados dentro de los documentos técnicos. El nivel de calidad de los datos mide el grado en que estos son o no adecuados para el análisis predictivo. Los conjuntos de datos

con bajos niveles de calidad simplemente indican que los datos no son adecuados para el análisis, pero pueden ser adecuados para la exploración de datos. Watson Analytics calcula una puntuación de calidad de datos con base en los datos originales, antes de que se produzca la limpieza o transformación de estos. La calidad de datos para cada campo del conjunto de datos se determina por la falta de valores, valores constantes, desequilibrio, valores atípicos, asimetría de una distribución; la asimetría describe cómo los valores se distribuyen a ambos lados del valor central, los valores inconsistentes o nulos, la cantidad de variables y veces que se repiten (Miller, 2016). Para el análisis por realizar se seleccionó manualmente el valor de la cuenca en alta, y para que no hallan incongruencias, se seleccionó del punto 1 al 32, que son los que corresponden a la cuenca alta.

Al depurar manualmente, Watson inmediatamente relaciona la *cuenca* y el *orden río* como datos para tener en cuenta. Gracias al análisis que Watson hace sobre la calidad de los datos y su capacidad de predecir, se pueden ver y cambiar las propiedades de un campo y especificar su papel en una predicción. Además de estos Watson determina que la *campana*, *punto de monitoreo*, *ICA* y *calidad del agua*

son los datos más relevantes y con mayor confiabilidad para realizar el análisis.

## V. Resultados y discusión

Con el apoyo de los datos proporcionados por la CAR y Watson, se desarrollaron diversos métodos para el análisis de datos que permitieron hacer inferencias acerca del comportamiento del ICA y demás variables pertenecientes a la base de datos. Luego de la depuración en la fase 1, se seleccionan los datos que son más relevantes, cuya calidad sea alta e influyan en la calidad del agua del río Bogotá.

Una vez los datos son depurados, se pasa a la etapa 2 o de exploración. Durante esta etapa, la herramienta nos sugiere algunas preguntas a las cuales se puede dar una respuesta según el análisis realizado en la primera etapa; al tiempo, permite realizar preguntas propias, frente a las cuales se sugerirá una respuesta de mayor relevancia.

Una de las primeras cosas que puede observar es el valor promedio del ICA en cada valor descriptivo de la calidad del agua, en cada una de sus cinco categorías (buena, aceptable, regular, mala o muy mala), que a su vez se asocian a un determinado color (azul, verde, amarillo, naranja y rojo, respectivamente) (figura 2).

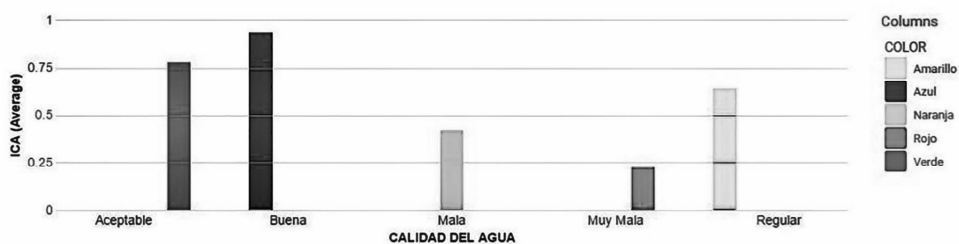


Figura 2. Promedio del ICA en Watson

Puede observarse que la calidad del agua en la mayoría de los puntos, durante los ocho años en la cuenca alta del río Bogotá, ha sido regular, aceptable o buena (figura 2). Esto también

puede inferirse si se saca el promedio general del ICA durante el mismo periodo, lo cual indica que el promedio general durante ocho años es de 0,59, es decir, que su calidad es regular.

Otro punto que permite observar la herramienta es el comportamiento del ICA en cada campaña y en cada uno de los puntos de monitoreo de la cuenca alta del río Bogotá. Uno de los puntos más importantes que muestra Watson es el nivel del ICA durante los ocho

años, en cada uno de los puntos de monitoreo. Este punto permite realizar un análisis detallado del comportamiento del ICA a través de los años, al igual que intuir el comportamiento que tendrá en años futuros (figura 3).

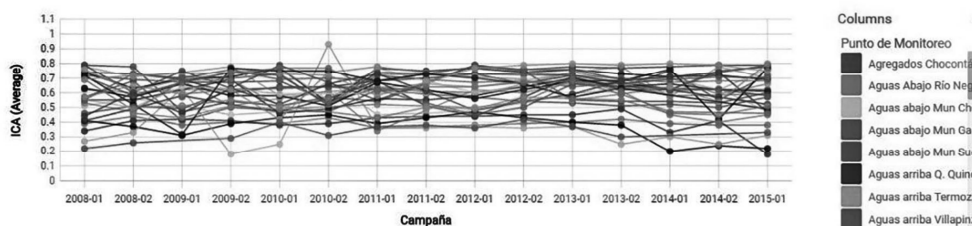


Figura 3. Comportamiento del ICA a través de los ocho años

Del análisis anterior se deriva cuáles son los puntos de monitoreo que han tenido mejor

calidad de agua durante el tiempo de estudio (figura 4).

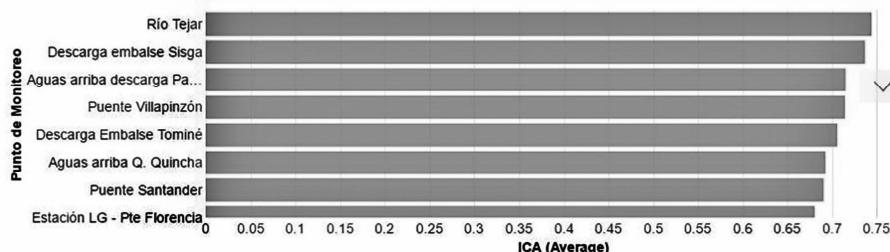


Figura 4. Puntos de Monitoreo con mejor ICA

Los cinco (5) puntos con mejor calidad de agua son:

- Río Tejar
- Descargue Embalse Sisga
- Aguas arriba, descarga Papeles y Molinos
- Puente Villapinzón

- Descarga Embalse Tominé

De igual forma, puede determinarse cuáles son los puntos de monitoreo que han tenido la peor calidad de agua durante el tiempo de estudio (figura 5).

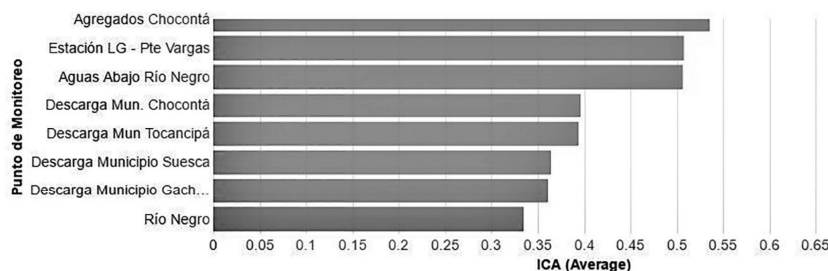


Figura 5. Puntos de Monitoreo con peor ICA

Los cinco (5) puntos con la peor calidad de agua son:

- Río Negro
- Descarga municipio de Gachancipá
- Descarga municipio de Suesca
- Descarga municipio de Tocancipá
- Descarga municipio de Chocontá

De esta forma, se pueden generar múltiples análisis, según la pregunta que se realice o las variables que se definan como importantes en el principio del proceso. Después de estos análisis básicos, se procede a la tercera etapa o predictiva. Esta etapa se basa en los análisis realizados para lograr predecir una situación en el futuro. Cuantos más análisis en la fase 2 se realicen, más aprenderá el sistema y mayor probabilidad de acierto tendrá hacia futuro.

Como para la fase 1, también es necesario establecer que hay una buena calidad en la información, para así poder llevar de forma satisfactoria la parte predictiva. El nivel de la calidad de los datos mide el grado en que estos son adecuados para el análisis predictivo; es un promedio de la puntuación de la calidad de los datos para cada campo en el conjunto de datos, determinado por los valores que faltan, categorías influyentes, valores atípicos, desequilibrio y asimetría. Es necesario tener en cuenta que los conjuntos de datos con bajo nivel de calidad pueden ser adecuados para la exploración de datos, incluso si no son adecuados para el análisis predictivo.

La primera predicción que muestra Watson es que la calidad del agua en la cuenca alta del río Bogotá va a tender a mantenerse en regular, pero se presentará una reducción en el ICA aceptable. A su vez, es posible ver cuál va a ser la calidad del agua por punto de monitoreo. La figura 6 ratifica de forma visual que, en la mayoría de los puntos, la calidad del agua se mantendrá en regular.

Estos valores se hallaron utilizando la técnica estadística del valor de chi-cuadrado. Este es un valor de la distribución teórica. El valor especifica la probabilidad de obtener el valor de chi-cuadrado de la casualidad. Si la probabilidad es menor que el nivel de significación, la prueba se considera que es estadísticamente significativa. El valor de significación, o el valor de  $p$ , es la probabilidad de que un resultado se produzca por casualidad. El valor de significación se compara con un punto de corte predeterminado (el nivel de significación por defecto es 0,05) para determinar si una prueba es estadísticamente significativa.

El valor de significación ayuda a determinar el significado práctico. IBM Watson Analytics utiliza tanto el valor de significación como el tamaño del efecto para determinar si un resultado es lo suficientemente importante como para mostrar. Este número significa que un resultado es estadísticamente significativo con menos de un 5% de probabilidad de ocurrencia del azar.

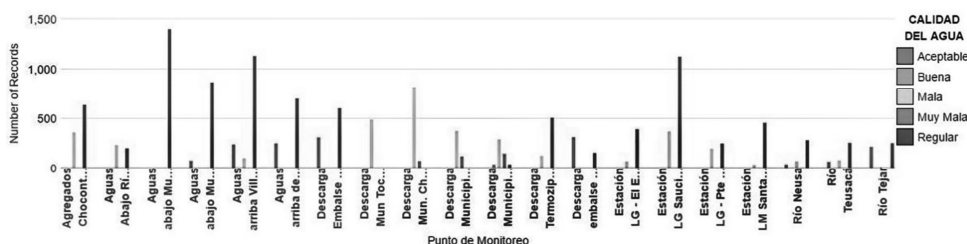


Figura 6. Predicción del ICA en la cuenca alta del río Bogotá

Finalmente, una forma más de ratificar la predicción es por medio de un árbol de decisión estadístico, como se muestra en la figura 7. Se indica que el 63% de los registros totales van a corresponder a un ICA regular, seguido por un 24% de un ICA malo, un 10% de un ICA

aceptable, un 2% de un ICA muy malo y menos del 1% equivaldría a un ICA bueno. En el árbol se indican numéricamente, de mayor a menor, las mejores probabilidades. Se indica así que el camino más probable es el correspondiente al 63%, es decir, al del ICA regular.

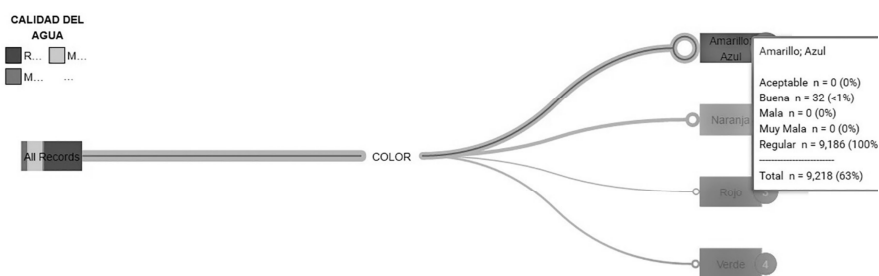


Figura 7. Análisis del árbol de decisión

## VI. Conclusiones

Mediante la computación cognitiva Watson se identifican los patrones de comportamiento en la calidad del agua de la cuenca alta del río Bogotá.

Mediante los estudios y los análisis realizados, puede notarse que el nivel del índice de calidad del agua (ICA) se mantendrá estable en la mayoría de sus puntos, desde que no haya alguna afectación externa grave que pueda complicar la calidad del agua. Aunque en muchos otros puntos de monitoreo tenderá a bajar el ICA, sería una buena opción tomar medidas para mantener y elevar este índice.

Mediante el análisis se logró determinar los puntos de monitoreo con mayor y menor calidad del agua.

Gracias a Watson se logra reducir drásticamente los datos de estudio, primando los de mayor calidad y así facilitando el análisis de estos.

Para poder hallar las variables de mayor impacto en la calidad del agua de la cuenca alta del río Bogotá, sería recomendable continuar con esta

investigación, para su respectivo análisis con Watson.

Para el establecimiento de árboles de decisión y su análisis es más eficiente un conjunto reducido de variables altamente significativas y que ofrezcan un porcentaje alto de acierto, reduciendo así el ruido en el modelo. De esta forma, será posible realizar técnicas de administración de una manera mucho más efectiva, para la toma de decisiones y planteamiento de políticas que se consideren adecuadas para mitigar la producción de contaminantes según el impacto en el nivel de contaminación del río.

## Referencias

- Alcaldía Mayor de Bogotá (1997). Río Bogotá. Recuperado de <http://www.banrepcultural.org/node/19162>
- Corporación Autónoma Regional de Cundinamarca (CAR) (2013). Objetivos de calidad. Recuperado de <https://www.car.gov.co/index.php?idcategoria=64243&download=Y>
- Chang, F. J., Tsai, Y. H., Chen, P. A., Coynel, A. y Vachaud, G. (2015). Modeling water quality in an urban river using hydrological factors –



- Data driven approaches. *Journal of Environmental Management*, 151, 87-96.
- Chang, N. B., Chen, H. W. y Ning, S. K. (2001). Identification of river water quality using the fuzzy synthetic evaluation approach. *Journal of Environmental Management*, 63(3), 293-305.
- Chau, K. (2006). A review on integration of artificial intelligence into water quality modelling. *Marine Pollution Bulletin*, 52(7), 726-733.
- Corporacion Colombia Digital (2015). Cinco usos de la computación cognitiva. Recuperado de <https://www.colombiadigital.net/actualidad/articulos-informativos/item/8602-cinco-usos-de-la-computacion-cognitiva.html>
- Ferrucci, D. et al. (2010). Building Watson: An overview of the DeepQA Project. Recuperado de <http://web.archive.org/web/20120120185642/http://www.stanford.edu/class/cs124/AIMagazine-DeepQA.pdf>
- Hammond, A., Adriaanse, A., Rodenburg, E., Bryant, D. y Woodward, R. (1995). *Environmental indicators: A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development*. Recuperado de [http://pdf.wri.org/environmentalindicators\\_bw.pdf](http://pdf.wri.org/environmentalindicators_bw.pdf)
- IBM (s. f.). ¿Qué es IBM Watson? Recuperado de <http://cognitiva.la/que-es-ibm-watson/>
- IBM (2013). IBM Power Systems: informática cognitiva con IBM Watson compilado sobre Power - Colombia. IBM Corporation. Recuperado de <http://www-03.ibm.com/systems/co/power/solutions/cognitive-computing.html>
- IBM Watson (2011). Un desafío para la historia - Colombia. Recuperado de <http://www-03.ibm.com/systems/co/power/advantages/watson/index.html>
- Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) (2011). Hoja metodológica del indicador Índice de calidad del agua (Versión 1,00). Bogotá: Autor.
- Lobbrecht, A. H., Dibike, Y. B. y Solomatine, D. P. (2002). *Hydroinformatics applications of neural networks and fuzzy logic to integrated water management project report*. Recuperado de [https://www.researchgate.net/publication/268203104\\_Hydroinformatics\\_Applications\\_of\\_Neural\\_Networks\\_and\\_Fuzzy\\_Logic\\_to\\_Integrated\\_Water\\_Management\\_Project\\_Report](https://www.researchgate.net/publication/268203104_Hydroinformatics_Applications_of_Neural_Networks_and_Fuzzy_Logic_to_Integrated_Water_Management_Project_Report)
- McCord, M. C. (1990). Slot grammar: A system for simpler construction of practical natural language grammars. En *Natural Language and Logic: International Scientific Symposium*. Berlín: Srpinger-Verlag.
- Miller, J. D. (2016). *Learning IBM Watson Analytics: make the most advanced predictive analytical processes easy using Watson Analytics with this easy-to-follow practical guide*. Birmingham: Packt Publ. Recuperado de <http://cds.cern.ch/record/2152952>
- Ministerio de la Protección Social (2007). Decreto 1575 de 2007, por el cual se establece el Sistema para la Protección y Control de la Calidad del Agua para Consumo Humano. Recuperado de [http://www.minambiente.gov.co/images/normativa/decretos/2007/dec\\_1775\\_2007.pdf](http://www.minambiente.gov.co/images/normativa/decretos/2007/dec_1775_2007.pdf)
- Prato, T. (2005). A fuzzy logic approach for evaluating ecosystem sustainability. *Ecological Modelling*, 187(2-3), 361-368.
- Roland, E., Uhrmacher, A. y Saha, K. (2009). Data mining for simulation algorithm selection. Recuperado de <http://dl.acm.org/citation.cfm?id=1537633>
- Rosado, A. y Verjel, A. (2015). *Minería de datos aplicada a la demanda del transporte aéreo en Ocaña, Norte de Santander*. *Tecnura*, 19(45), 101-113.
- The DeepQA Research Team - IBM (20132). IBM Corporation. Recuperado de [http://researcher.watson.ibm.com/researcher/view\\_group.php?id=2099](http://researcher.watson.ibm.com/researcher/view_group.php?id=2099)