

TALLER: DATASET CALIDAD DEL AGUA



ESTUDIANTE

**JUANA VALENTINA MENDOZA SANTAMARÍA
MIGUEL ÁNGEL AVILA TORRES**

PRESENTADO A:

MARTHA SUSANA CONTRERAS

**MACHINE LEARNING
UNIVERSIDAD SANTO TOMÁS
SECCIONAL TUNJA
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA DE SISTEMAS**

TUNJA, OCTUBRE 04 DE 2021

Contexto

El acceso al agua potable es esencial para la salud, un derecho y un componente de una política eficaz de protección de la salud. Esto es importante como cuestión de salud y desarrollo a nivel nacional, regional y local. En algunas regiones, se ha demostrado que las inversiones en suministro de agua y saneamiento pueden generar un beneficio económico neto, ya que las reducciones de los efectos adversos para la salud y los costos de la atención médica superan los costos de llevar a cabo las intervenciones.

Contenido

El archivo `water_potability.csv` contiene métricas de calidad del agua para 3276 cuerpos de agua diferentes.

1. Valor del PH:

Columna: `ph`

El PH es un parámetro importante en la evaluación del equilibrio ácido-base del agua. También es el indicador de la condición ácida o alcalina del estado del agua. La OMS ha recomendado **el límite máximo permisible de pH de 6,5 a 8,5**. Los rangos de **investigación actuales fueron de 6,52 a 6,83**, que se encuentran en el rango de los estándares de la OMS.

2. Dureza:

Columna: `Hardness`

La dureza es causada principalmente por sales de calcio y magnesio. Estas sales se disuelven a partir de depósitos geológicos a través de los cuales viaja el agua. **El tiempo que el agua está en contacto con el material que produce dureza ayuda a determinar cuánta dureza hay en el agua cruda.** La dureza se definió originalmente como **la capacidad del agua para precipitar el jabón causado por el calcio y el magnesio.**

3. Sólidos (sólidos disueltos totales – TDS) :

Columna: Solids

El agua tiene la capacidad de disolver una amplia gama de minerales o sales inorgánicos y algunos orgánicos, como potasio, calcio, sodio, bicarbonatos, cloruros, magnesio, sulfatos, etc. Estos minerales producen un sabor no deseado y un color diluido en apariencia de agua. Este es el parámetro importante para el uso del agua. **El agua con alto valor de TDS indica que el agua está altamente mineralizada. El límite deseable de TDS es de 500 mg / l y el límite máximo es de 1000 mg / l que se prescribe para beber.**

4. Cloraminas:

Columna: Chloramines

El cloro y la cloramina son los principales desinfectantes que se utilizan en los sistemas públicos de agua. Las cloraminas se forman con mayor frecuencia cuando se agrega amoníaco al cloro para tratar el agua potable. **Los niveles de cloro de hasta 4 miligramos por litro (mg / L o 4 partes por millón (ppm)) se consideran seguros en el agua potable.**

5. Sulfato:

Columna: Sulfate

Los sulfatos son sustancias naturales que se encuentran en minerales, suelo y rocas. Están presentes en el aire ambiente, las aguas subterráneas, las plantas y los alimentos. El uso comercial principal del sulfato es la industria química. **La concentración de sulfato en el agua de mar es de aproximadamente 2700 miligramos por litro (mg / L). Varía de 3 a 30 mg / L en la mayoría de los suministros de agua dulce, aunque se encuentran concentraciones mucho más altas (1000 mg / L) en algunas ubicaciones geográficas.**

6. Conductividad:

Columna: Conductivity

El agua pura no es un buen conductor de corriente eléctrica, más bien es un buen aislante. El aumento de la concentración de sales mejora la conductividad eléctrica del agua. Generalmente, la cantidad de sólidos disueltos en el agua determina la conductividad eléctrica. La conductividad eléctrica (EC) en realidad mide el proceso iónico de una solución que le permite transmitir corriente. **De acuerdo con las normas de la OMS, el valor de CE no debe exceder los 400 $\mu\text{S} / \text{cm}$.**

7. Carbón orgánico:

Columna: Organic_carbon

El carbono orgánico total (TOC) en las fuentes de agua proviene de la materia orgánica natural en descomposición (NOM), así como de fuentes sintéticas. TOC es una medida de la cantidad total de carbono en compuestos orgánicos en agua pura. Según la EPA de EE. UU. **<2 mg / L como TOC en agua tratada / potable y <4 mg / Lit en el agua de origen que se utiliza para el tratamiento.**

8. Trihalometanos:

Columna: Trihalomethanes

Los THM son sustancias químicas que se pueden encontrar en el agua tratada con cloro. La concentración de THM en el agua potable varía según el nivel de material orgánico en el agua, la cantidad de cloro necesaria para tratar el agua y la temperatura del agua que se está tratando. **Los niveles de THM de hasta 80 ppm se consideran seguros en el agua potable.**

9. Turbiedad:

Columna: Turbidity

La turbidez del agua depende de la cantidad de materia sólida presente en estado suspendido. Es una medida de las propiedades emisoras de luz del agua y la prueba se utiliza para indicar la calidad de la descarga de desechos con respecto a la materia coloidal. **El valor medio de turbidez obtenido para Wondo Genet Campus (0,98 NTU) es inferior al valor recomendado por la OMS de 5,00 NTU.**

10. Potabilidad:

Columna: Potability

Indica si el agua es segura para el consumo humano, donde 1 significa potable y 0 significa no potable.

Dataset

- Eliminando los datos NaN

Model	Score	Accuracy	Precision score
Linear	-1.3471437122491992		0.487292
Logistic	63.27543424317618	63.27543424317618	81.5
Decision Tree Classifier	62.779156327543426	62.779156327543426	60.13100552721089
SVC	62.531017369727046	62.531017369727046	31.265508684863523
Neighbors Classifier	61.04218362282878	61.04218362282878	54.98927498927498

- Convertir los datos NaN a la media (ph, Sulfate, Trihalomethanes)

Model	Score	Accuracy	Precision score
Linear	-0.1830838116009259		0.483768
Logistic	62.80487804878049	62.80487804878049	31.402439024390244
Decision Tree Classifier	58.536585365853654	58.536585365853654	55.950562149880525
SVC (Support Vector Classifier)	62.65243902439024	62.65243902439024	31.37404580152672
Neighbors Classifier	56.859756097560975	56.859756097560975	46.950841118360366