

# Examen práctico 1

**Apertura:** martes, 10 de diciembre de 2024, 18:30

Tras los daños sufridos por la DANA en Valencia, urge controlar la potabilidad del agua que llega a los grifos y también encontrar nuevas fuentes de agua potable que puedan suministrar a la población mientras se restablece un suministro seguro y estable. Con este propósito, un grupo de voluntariado de desarrollo de sistemas inteligentes cuenta con nosotros para terminar de diseñar un sistema que permita predecir, en base a cuatro características, si un posible punto de suministro de agua (o el agua que sale por los grifos en los diferentes municipios) es (1) recomendable para el consumo, (2) apto, pero no recomendable, o (3) no apto.



Tras cientos de mediciones y análisis en el laboratorio, han intentado simplificar los parámetros de análisis de forma que el prototipo de sistema inteligente "**OK Water**" pueda hacer una predicción instantánea de la potabilidad del agua en base a 4 características de la muestra: alcalinidad (pH), dureza, turbidez y presencia de metales. Nos han facilitado un *dataset* (adjunto a este enunciado) que presenta algunos problemas, y cuentan con nosotros para solucionarlos. A continuación se describen las tareas que nos han encomendado:

1. Desde el laboratorio de análisis nos han hecho llegar una hoja de resultados en base a los cuatro parámetros pero faltan algunos valores que debemos completar (hay datos que no están recogidos). Elegiremos una estrategia para imputarlos y la llevaremos a cabo sobre nuestro conjunto de datos. Justificaremos porqué hemos elegido una determinada estrategia.
2. Creemos que también hay errores en algunos datos, debidos a fallos en los sensores o en el laboratorio. Revisaremos el conjunto de datos en busca de valores atípicos y adoptaremos alguna solución al respecto que documentaremos adecuadamente, justificando las acciones llevadas a cabo.
3. Para completar el análisis exploratorio de los datos que nos han proporcionado vamos a representarlos:
  - En la primera gráfica enfrentaremos pH, dureza y turbidez.
  - En la segunda gráfica pH, dureza y metales.
  - En la tercera gráfica pH, turbidez y metales.
  - En la cuarta gráfica dureza, turbidez y metales.

- Por último, para observar las posibles correlaciones obtendremos un mapa de calor.
  - Comenta lo observas en el conjunto de visualizaciones.
4. Intenta obtener un clasificador basado en el algoritmo KNN, prueba con diferentes parámetros y compara las matrices de confusión de cada clasificador.
5. Busca ahora obtener información del *dataset* para construir un árbol de decisión, represéntalo gráficamente y muestra también la matriz de confusión resultado.
6. Por último trata de obtener un clasificador basado en una red neuronal, prueba diferentes diseños (número de neuronas y capas) teniendo en cuenta la configuración de los datos. Puede que obtengas mejores resultados si aplicas alguna transformación previa al *dataset*, como escalar cada variable.

Utiliza para el diseño de los clasificadores un *split* 70/30.


El resultado del ejercicio se entregará en un cuaderno de Colab, descargado y subido a esta entrega. El nombre del archivo ha de tener esta forma: **apellido1\_nombre.ipynb**

Los criterios de evaluación se muestran en la rúbrica de esta tarea.

 [dataset\\_ok\\_water.csv](#) +

9 de diciembre de 2024, 22:46

Estado de la entrega


Estado de la entrega	Enviado para calificar
Estado de la calificación	Calificado
Última modificación	martes, 10 de diciembre de 2024, 21:04
Archivos enviados	<div><div> <a href="#">marin_daniel.ipynb</a> +</div><div>10 de diciembre de 2024, 21:04</div></div> <div><a href="#">Exportar al portafolios</a></div>
Comentarios de la entrega	<div>► <a href="#">Comentarios (0)</a></div>

Criterios de calificación

Missing Values	No se tratan. <i>0 puntos</i>	Se prescinde de las filas o se sustituyen por una constante. <i>5 puntos</i>	Se aplica algún método de imputación pero no se explica ni se defiende su idoneidad. <i>10 puntos</i>	Se aplican uno o más métodos de imputación justificando su idoneidad. <i>15 puntos</i>
Outliers	No se detectan. <i>0 puntos</i>	Se detectan pero no se hace nada al respecto. <i>10 puntos</i>	Se detectan y se transforman en base a alguna decisión argumentada. <i>15 puntos</i>	
Visualización	No se representa ninguna de las gráficas solicitadas. <i>0 puntos</i>	Se representan algunas gráficas pero no hay comentarios o los comentarios son muy escuetos. <i>10 puntos</i>	Se representan al menos cuatro gráficas y se comentan. <i>20 puntos</i>	Se representan todas las gráficas solicitadas y se comentan los hallazgos. <i>25 puntos</i>

K-NN	No se obtiene el clasificador. <i>0 puntos</i>	Se obtiene un clasificador pero no sus resultados o no se comentan. <i>5 puntos</i>	Se obtiene el clasificador, la matriz de confusión y se comentan los resultados. <i>10 puntos</i>	Se obtienen varios clasificadores, sus matrices de confusión y se comentan y comparan los resultados. <i>15 puntos</i>
Árbol de decisión	No se obtiene el clasificador. <i>0 puntos</i>	Se obtiene un clasificador pero no sus resultados o no se comentan. <i>5 puntos</i>	Se obtiene el clasificador, la matriz de confusión y se comentan los resultados. <i>10 puntos</i>	Se obtienen varios clasificadores, sus matrices de confusión y se comentan y comparan los resultados. <i>15 puntos</i>
RN (MLP)	No se obtiene el clasificador. <i>0 puntos</i>	Se obtiene un clasificador pero no sus resultados o no se comentan. <i>5 puntos</i>	Se obtiene el clasificador, la matriz de confusión y se comentan los resultados. <i>10 puntos</i>	Se obtienen varios clasificadores, sus matrices de confusión y se comentan y comparan los resultados. <i>15 puntos</i>

Comentario

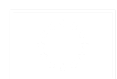
Calificación	85,00 / 100,00
Calificado sobre	jueves, 12 de diciembre de 2024, 12:53
Calificado por	 Jiménez Gómez, Manuel Ángel

Desglose de la calificación

Missing Values	No se tratan. <i>0 puntos</i>	Se prescinde de las filas o se sustituyen por una constante. <i>5 puntos</i>	Se aplica algún método de imputación pero no se explica ni se defiende su idoneidad. <i>10 puntos</i>	Se aplican uno o más métodos de imputación justificando su idoneidad. <i>15 puntos</i>	Sabiendo que había outliers, ¿no es mejor la mediana que la media? Con KNN has logrado una estrategia excelente.
----------------	----------------------------------	---	--	---	--

Outliers	No se detectan. <i>0 puntos</i>	Se detectan pero no se hace nada al respecto. <i>10 puntos</i>	Se detectan y se transforman en base a alguna decisión argumentada. <i>15 puntos</i>	
Visualización	No se representa ninguna de las gráficas solicitadas. <i>0 puntos</i>	Se representan algunas gráficas pero no hay comentarios o los comentarios son muy escuetos. <i>10 puntos</i>	Se representan al menos cuatro gráficas y se comentan. <i>20 puntos</i>	Se representan todas las gráficas solicitadas y se comentan los hallazgos. <i>25 puntos</i>
K-NN	No se obtiene el clasificador. <i>0 puntos</i>	Se obtiene un clasificador pero no sus resultados o no se comentan. <i>5 puntos</i>	Se obtiene el clasificador, la matriz de confusión y se comentan los resultados. <i>10 puntos</i>	Se obtienen varios clasificadores, sus matrices de confusión y se comentan y comparan los resultados. <i>15 puntos</i>

Árbol de decisión	No se obtiene el clasificador. <i>0 puntos</i>	Se obtiene un clasificador pero no sus resultados o no se comentan. <i>5 puntos</i>	Se obtiene el clasificador, la matriz de confusión y se comentan los resultados. <i>10 puntos</i>	Se obtienen varios clasificadores, sus matrices de confusión y se comentan y comparan los resultados. <i>15 puntos</i>	!!!!
RN (MLP)	No se obtiene el clasificador. <i>0 puntos</i>	Se obtiene un clasificador pero no sus resultados o no se comentan. <i>5 puntos</i>	Se obtiene el clasificador, la matriz de confusión y se comentan los resultados. <i>10 puntos</i>	Se obtienen varios clasificadores, sus matrices de confusión y se comentan y comparan los resultados. <i>15 puntos</i>	¡Demasiadas neuronas!



GOBIERNO DE  
JUNTA DE ANDALUCÍA



MINISTERIO  
DE HACIENDA



PARLAMENTO DE ANDALUCÍA



Junta de Andalucía  
Consejería de Desarrollo Educativo  
y Formación Profesional

Reiniciar tour para usuario en esta página