

CALIDAD DEL AGUA SUPERFICIAL

Elmer Ali Payró Costilla A01014943
María Fernanda Prieto Vargas A01794329



OBJETIVO DE INVESTIGACIÓN

Analizar el estado del agua de superficie de México y sus estados, con el propósito de evidenciar su calidad, tomando como insumos la base de datos donde se plasman toda las variables claves para dar respuesta a este objetivo



An aerial photograph of Mexico City, showing a dense urban landscape with numerous buildings, streets, and green spaces. The city extends to the horizon under a clear sky. A semi-transparent blue rectangular box is overlaid on the lower-left portion of the image, containing white text.

ANÁLISIS 1

Limpieza, Análisis y Kmeans

LIMPIEZA Y ANÁLISIS

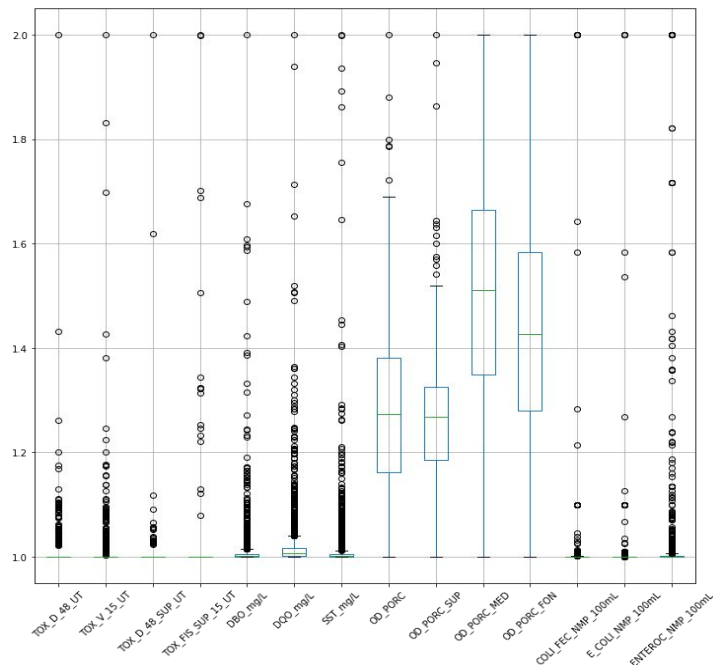


La base de datos cuenta con 3494 registros y 54 Variables

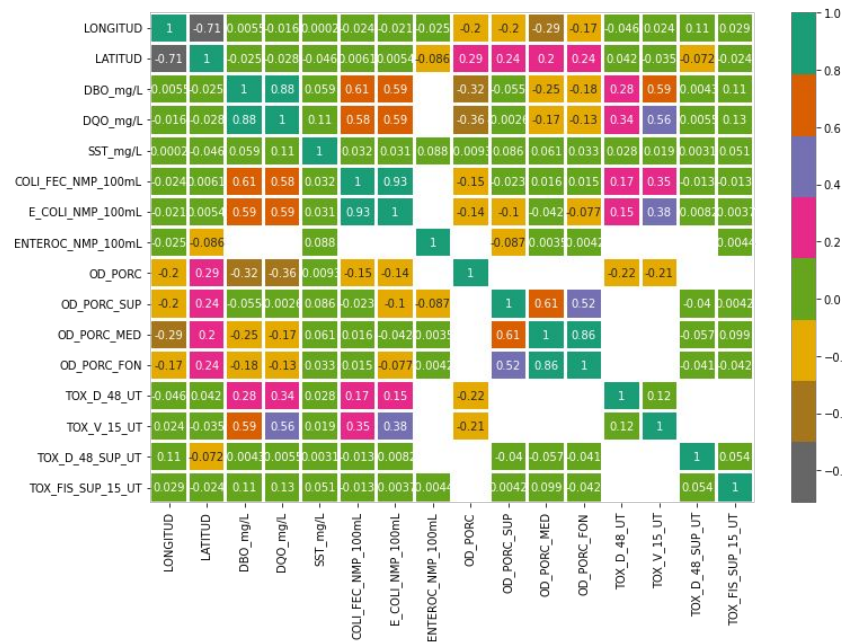
Para la limpieza de los datos se realizó:

- Se **eliminaron solo el 9%** de las variables ya que no contaban con ninguna información por lo tanto no proporcionaban información para el análisis del modelo.
- Un **reemplazo de notación matemática** y transformación las variables a tipo double, lo cual permitió contar con el 26 % de las variables tipo numéricas y el 76% tipo Categorías.
- Se **reemplazaron algunos valores categóricos** que estaban dentro de una variables numérica, como por ejemplo valores descritos tales como '<2', '<10'. Esto fue para el 18% de las variables.
- Se aplicó el **método 'bfill'** el cual reemplaza los valores NULL con los valores de la siguiente fila esto se hizo para el 5% de los datos
- Se aplicó un **Tresh de 20**, ya que fue el valor óptimo para que las variables de longitud, latitud y semáforo en este punto no tenían valores faltantes.

MEDIDAS Y CUARTILES

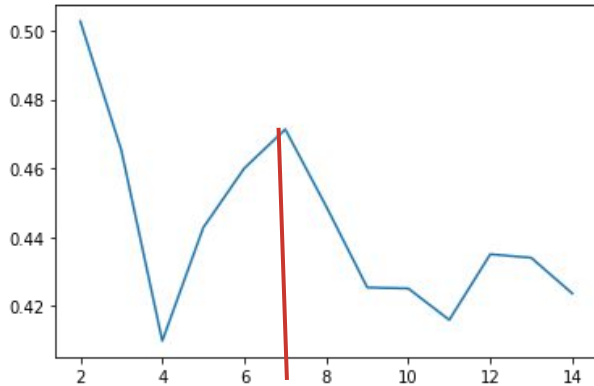
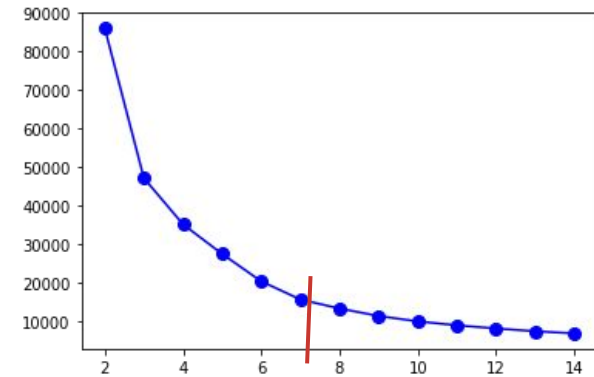


En este gráfico de Bigotes, permite identificar valores atípicos y comparar las distribuciones entre las variables numéricas del modelo.

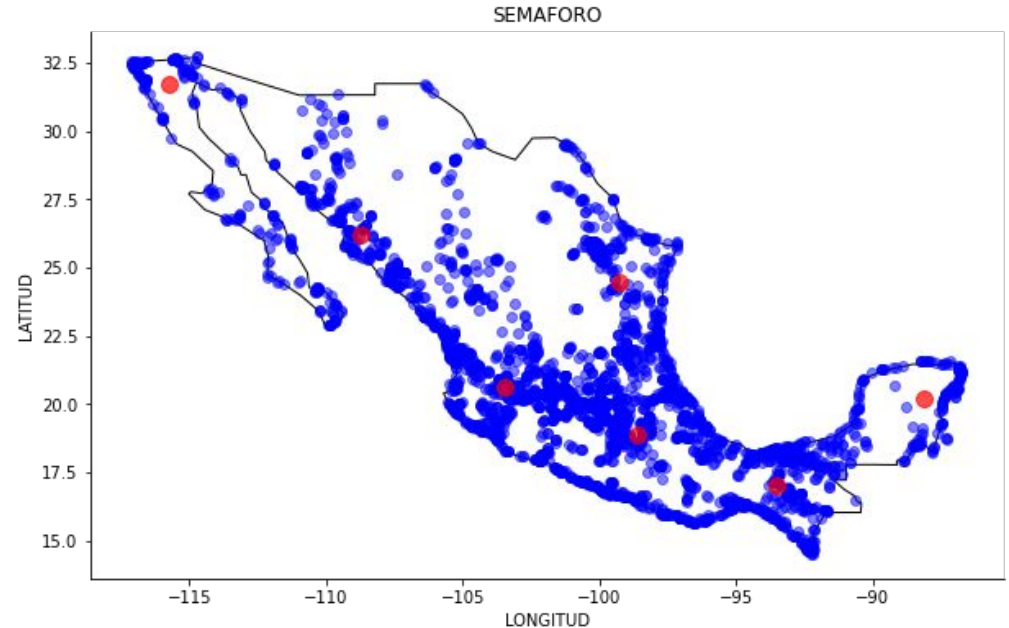


En este gráfico de correlación, permite evidenciar que no hay una correlación muy marcada entre las variables, pero se evidenciaba una correlación entre la latitud y resultados numéricos de medición del agua

KMEANS



Para la identificación de la cantidad de cluster aplicamos estos dos modelos, pero queremos resaltar que la gráfica del codo es mucho más clara la cantidad de **cluster** sugeridas por este el cual son **7**

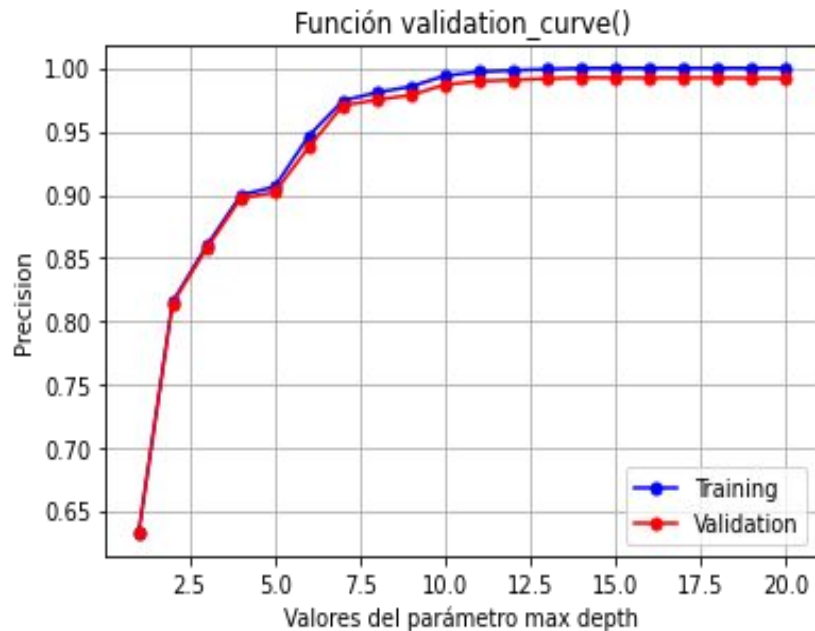


An aerial photograph of Mexico City, showing a dense urban landscape with numerous buildings, streets, and green spaces. The city extends to the horizon under a clear sky. A semi-transparent blue rectangular box is overlaid on the lower-left portion of the image, containing white text.

ANÁLISIS 2

Clasificación y Resultados

DECISION TREE



- (2969, 14) : dimensión de datos de entrada para entrenamiento y validación
- (524, 14) : dimensión de datos de entrada para prueba
- (2969, 1) : dimensión de variable de salida para entrenamiento y validación
- (524, 1) : dimensión de variable de salida para prueba

Como nos muestra la gráfica aplicando la métrica accuracy, vemos que el modelo está sobre-entrenado, pero los resultados de DT muy buenos

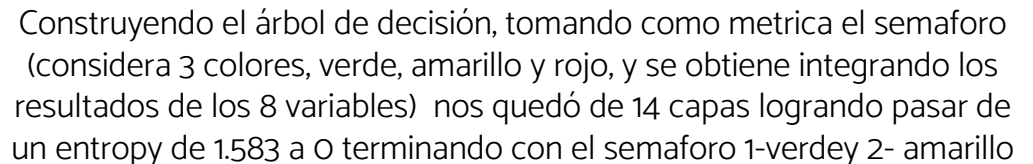
MODELO: Decision Tree

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

1	0.98	0.99	0.99	1066
2	0.99	0.98	0.98	955
3	1.00	0.99	1.00	948

accuracy			0.99	2969
macro avg	0.99	0.99	0.99	2969
weighted avg	0.99	0.99	0.99	2969

Downloaded from <http://ajph.org/> on November 10, 2015



CONCLUSIÓN



Se evidencia que para los estados de Mexico en los que se analizó la calidad del agua, teniendo en cuenta la metrica del semaforo esta en verde el cual representa aproximadamente un 36% de los sitio con este parametro y amarillo un 32%, lo que nos permite concluir que el agua es de buena calidad.

GRACIAS

“El agua es la fuerza motriz de
toda la naturaleza ”
Leonardo Da Vinci

