Prototipo de un modelo de Machine Learning para la predicción de la calidad de agua en el Perú

Raúl Armas Benavides[[1]](#footnote-1)

Lyshet Huasacca León2

Ronnie Oliva Moya3

Italo Egues Mendoza4

**RESUMEN**

En el presente artículo de investigación se abarca la problemática existente con respecto a la contaminación del agua en nuestro país, para ello se aplica un modelo de inteligencia artificial con el cual analizaremos componentes que determinen el grado de la calidad del agua mediante la elaboración de un prototipo en Machine Learning. Se empleará el uso de redes neuronales en el tratamiento de datos obtenidos a través de los sistemas de monitoreo de agua, para con ello, lograr la optimización en la gestión de los recursos hídricos en la aplicación de las diferentes industrias de manera rápida. Este prototipo servirá como guía para profesionales que deseen tratar problemas con la calidad del agua.

**Palabras clave:** Machine Learning, inteligencia artificial, redes neuronales.

**ABSTRACT**

This research article covers the existing problem with water pollution in our country, for this topic, an artificial intelligence model is applied to analyze components that determine the degree of water quality through the development of a prototype in Machine Learning. The use of neural networks will be really important in the treatment of data obtained through water monitoring systems, in order to achieve optimization in the management of water resources in the application of different industries quickly. This prototype will serve as a guide for professionals who wish to deal with water quality problems.

**Key words:** Machine Learning, artificial intelligence, neural networks.

# INTRODUCCIÓN

Este trabajo de investigación trata sobre cómo podemos brindar una solución, mediante herramientas tecnológicas innovadoras, a la problemática que se está dando a nivel nacional sobre los recursos hídricos y cómo esto está afectando a la población peruana, veremos la relación que existe entre la gestión ambiental que se lleva a cabo en el Perú y la calidad del agua en nuestro territorio nacional.

Nuestro trabajo consta de una primera parte avocada a la problemática que vamos a tocar, que son principalmente la contaminación de nuestros recursos hídricos, esto visto desde la perspectiva de la gestión ambiental, para así lograr una mejor gestión del agua en el Perú, que sea sostenible, viable y que pueda abarcarse en los próximos años mediante la herramienta de redes neuronales del Machine Learning.

En una segunda instancia pasaremos a profundizar en el análisis de los datos que nos permitirán tener una idea clara de cuál es el mejor camino que tomaremos para dar una solución a esta problemática; para ello realizaremos un modelo de inteligencia artificial con la cual determinaremos el grado de calidad de agua mediante el tratamiento de datos obtenidos de ciertas estaciones de monitoreo del agua.

Las oportunidades que la inteligencia artificial presenta para el sector agua en el Perú son amplísimas, desde la protección de fuentes de agua hasta la optimización de sistemas avanzados de tratamiento de aguas residuales. Si bien esta tecnología no resolverá los problemas del agua, nos puede ayudar a avanzar rápidamente a soluciones innovadoras, contribuyendo así al logro del Objetivo de Desarrollo Sostenible de Agua limpia y saneamiento.

En última instancia veremos los resultados del prototipo de Inteligencia Artificial que nuestro equipo de trabajo ha planteado para solucionar los problemas que están actualmente amenazando nuestros recursos hídricos.

## Antecedentes

### Antecedentes Nacionales

La existencia y gravedad de los problemas ambientales son reconocidas en diversas escalas, de allí el interés de muchos países en institucionalizar instrumentos que incorporen la variable ambiental; sin embargo, según la UNESCO, la gravedad y complejidad de esta problemática ha continuado, lo que motivó la Segunda Conferencia Mundial sobre Medio Ambiente y Desarrollo en 1992.

Al problema de contaminación de las aguas costeras, se le ha dedicado atención mundial primordialmente atención en los países desarrollados de Europa, en los Estados Unidos de América, etc. Ello reviste mayor gravedad por el gran auge de las industrias y la elevada densidad de la población.

En el Perú, entre otros trabajos relacionados a la contaminación marina destacan quienes dan a conocer que las fuentes más importantes de contaminación, la constituyen las descargas industriales y domésticas, las cuales ocasionan un fuerte impacto en el ambiente receptor (el mar). El impacto de la contaminación sobre la pesquería se refleja en la pérdida del mercado de recursos pesqueros, por las regulaciones sobre los niveles aceptables de metales pesados y otros contaminantes. Asimismo, los diversos trabajos de investigación coinciden que las aguas costeras y playas del litoral peruano se encuentran en estado de contaminación.

Existen altas concentraciones de metales pesados, con contenidos de 27.0 a 7.7 ppb de Fierro, 27.0 a 2.3 ppb de cobre soluble y 0.00 ppb a 68.0 ppb de plomo soluble total en las aguas de mar del área del Callao. El conocimiento del estado del medio marino costero, como resultado de las actividades antropogénicas en el litoral peruano fue analizado en el marco del Plan de Acción para la Protección del Medio Marino y Áreas Costeras del Pacífico Sudeste, por Carmen Conopuma, en 1991, quien manifiesta que el crecimiento de la población y los centros industriales vienen afectando seriamente importantes zonas del litoral peruano, particularmente las zonas cerradas.

En el Perú, las descargas domésticas y mineras son las que ocasionan mayor contaminación en las aguas receptoras, la primera por el volumen de sus vertimientos, y la minería por el tipo de vertimiento de sustancias tóxicas. Los principales procesos de modificación de la línea de costa son la erosión y la acreción, que generan importantes procesos morfológicos, activados por la socavación de los ríos, transporte de sedimentos que se incrementaron en el año 1983 en casi veinte veces el promedio movilizado en años anteriores, llegando a más de treinta millones de metros cúbicos. Es probable que estas acciones ocurran también en el río Chancay y en otros ríos costeros, alterando la calidad de las aguas costeras del litoral.

### Antecedentes Internacionales

El agua está en el epicentro del desarrollo sostenible y es fundamental para el desarrollo socioeconómico, la energía y la producción de alimentos, los ecosistemas saludables y para la supervivencia misma de los seres humanos. El agua también forma parte crucial de la adaptación al cambio climático, y es el vínculo crucial entre la sociedad y el medioambiente.

El agua es, además, una cuestión de derechos. A medida que crece la población mundial, se genera una necesidad creciente de conciliar la competencia entre las demandas comerciales de los recursos hídricos para que las comunidades tengan lo suficiente para satisfacer sus necesidades. En concreto, las mujeres y las niñas deben tener acceso a instalaciones de saneamiento limpias y que respeten la privacidad para que puedan manejar la menstruación y la maternidad con dignidad y seguridad.

Para el desarrollo del ser humano, el agua y los sistemas de saneamiento no pueden estar separados. Ambos son vitales para reducir la carga mundial de enfermedades y para mejorar la salud, la educación y la productividad económica de las poblaciones.

## Formulación del Problema

¿Cómo un modelo de machine learning ML puede utilizarse para predecir la calidad del agua?

## Objetivo General

Desarrollar un prototipo basado en algoritmos de Machine Learning que permita predecir la calidad del agua en las estaciones de monitoreo a menor costo.

## Objetivos Específicos

* Normalizar los datos recopilados de las estaciones de monitoreo del agua en Lima, para alimentar las redes neuronales de nuestro prototipo.
* Seleccionar una arquitectura de diseño para escoger la cardinalidad de las capas ocultas del modelo de red neuronal artificial.
* Seleccionar un algoritmo de entrenamiento para calcular los pesos de la red del modelo de red neuronal.
* Validar el prototipo implementado mediante los resultados obtenidos para corroborar la efectividad del modelo.
* Reducir la complejidad y el costo del sistema de monitoreo

## Justificación e importancia de la investigación

Este trabajo de investigación es importante para informar y concientizar a la sociedad civil, al empresariado y a la comunidad científica de la importancia de la calidad de agua en el país. Este trabajo aporta una detallada información de la actual situación problemática de nuestro país y presenta propuestas de estrategias para optimizar la gestión de los recursos hídricos en las aplicaciones en diferentes industrias. La preocupación por el agua surge a partir de la ambivalente situación de que somos un país con escases de agua potable, pero a la vez un país con abundante recurso hídrico.

## Limitaciones de la investigación

El curso de la presente investigación se desarrolló en el semestre 2020-1 en la facultad de ingeniería industrial de la Universidad Nacional Mayor de San Marcos durante la pandemia por el Covid 19. Esto limitó al trabajo de campo y la investigación in situ en las centrales de monitoreo del río Chillón, donde se quiso delimitar la investigación.

# Metodología de la investigación

En un principio los modelos de redes neuronales requieren una serie de datos como fuente entrada, agrupados en un dataset, para poder realizar el análisis de los mismo y poder ejecutar un determinado cálculo que permita realizar una predicción. Para la obtención de los datos de entrada se utilizarán los informes realizados por la ANA y sus reportes mensuales sobre la calidad del agua en la ciudad, primero se seleccionarán los datos comprendidos dentro de un periodo de tiempo, luego se escoge el modelo de red neuronal artificial que se va a utilizar para realizar la predicción. A este modelo, se le realiza todo el proceso previsto de enseñanza de una red neuronal con sus respectivas etapas de adecuación de los datos de entrada, selección de capas, aprendizaje y verificación de salida y ajuste de pesos. Finalmente se realizará el análisis respectivo sobre los resultados obtenidos y performance del modelo utilizado.

## Tipo de la investigación

Este proyecto busca aportar un modelo viable para la predicción de la contaminación del agua en la ciudad de Lima, esto mediante la utilización de una red neuronal artificial. Con esto se busca tener un estudio como base que muestre los beneficios que tiene tener un modelo computacional que realice una predicción aproximada sobre una variable de entorno, que en esta ocasión es la de un contaminante del agua.

## Diseño de la investigación

Los datos utilizados para esta investigación se obtuvieron de Internet y se limpiarán realizando un análisis de diagrama de caja, discutido en esta sección. Después de limpiar los datos, se normalizarán usando normalización del valor *q* para convertirlos al rango de 0–100 para calcular el WQI usando seis disponibles parámetros Una vez que se calculó el WQI, todos los valores originales se normalizaron utilizando la puntuación *z*, por lo que estaban en la misma escala. El procedimiento completo se detalla a continuación:

### Instrumentos

#### Azure Machine Learning

Azure Machine Learning es un servicio de análisis predictivo en la nube que permite crear e implementar rápidamente tanto modelos predictivos como soluciones de análisis.

#### Excel

Excel es un programa informático desarrollado y distribuido por Microsoft Corp. Se trata de un software que permite realizar tareas contables y financieras gracias a sus funciones, desarrolladas específicamente para ayudar a crear y trabajar con hojas de cálculo.

## Aplicación de la inteligencia artificial y el agua

### MACHINE LEARNING o APRENDIZAJE AUTOMÁTICO

Machine Learning Machine Learning es una forma de IA que permite a un sistema aprender de los datos en lugar de a través de la programación explicita. Sin embargo, el aprendizaje automático no es un proceso simple. El aprendizaje automático utiliza una variedad de algoritmos que aprenden iterativamente de datos para mejorar, describirlos mejor y predecir resultados. A medida que los algoritmos ingieren datos de entrenamiento, es posible producir modelos más precisos basados en esos datos. Un modelo de aprendizaje automático es la salida generada cuando entrena su algoritmo con datos. Después del entrenamiento, cuando proporciona un modelo con una entrada, se le dará una salida. Por ejemplo, un algoritmo predictivo creara un modelo predictivo. Luego, cuando proporciona datos del modelo predictivo, recibirá una predicción basada en los datos que entrenaron al modelo. Machine Learning se ha convertido en uno de los temas más importantes dentro de las organizaciones de desarrollo que buscan formas innovadoras de aprovechar los activos de datos para ayudar a la empresa a obtener un nuevo nivel de comprensión. Con los modelos de aprendizaje automático adecuados, las organizaciones tienen la capacidad de predecir continuamente los cambios en el negocio, buscando predecir lo que sigue. A medida que se agregan datos constantemente, los modelos de aprendizaje automático aseguran que la solución se actualice constantemente. El valor es sencillo: si utiliza las fuentes de datos más adecuadas y en constante cambio en el contexto del aprendizaje automático, tiene la oportunidad de predecir el futuro. Los algoritmos de Machine Learning se dividen en tres categorías, siendo las dos primeras las más comunes:

* Aprendizaje supervisado: estos algoritmos cuentan con un aprendizaje previo basado en un sistema de etiquetas asociadas a unos datos que les permiten tomar decisiones o hacer predicciones. Un ejemplo es un detector de spam que etiqueta un e-mail como spam o no dependiendo de los patrones que ha aprendido del histórico de correos (remitente, relación texto/imágenes, palabras clave en el asunto, etc.).
* Aprendizaje no supervisado: estos algoritmos no cuentan con un conocimiento previo. Se enfrentan al caos de datos con el objetivo de encontrar patrones que permitan organizarlos de alguna manera. Por ejemplo, en el campo del marketing se utilizan para extraer patrones de datos masivos provenientes de las redes sociales y crear campañas de publicidad altamente segmentadas.
* Aprendizaje por refuerzo: su objetivo es que un algoritmo aprenda a partir de la propia experiencia. Esto es, que sea capaz de tomar la mejor decisión ante diferentes situaciones de acuerdo a un proceso de prueba y error en el que se recompensan las decisiones correctas. En la actualidad se está utilizando para posibilitar el reconocimiento facial, hacer diagnósticos médicos o clasificar secuencias de ADN.

**Redes neuronales artificiales:**

Las redes neuronales artificiales son aproximadores no lineales a la forma en que funciona el cerebro; por lo tanto, no deben compararse directamente con el cerebro ni confundir los principios que fundamentan el funcionamiento de las redes neuronales artificiales y el cerebro, ni pensar que las redes neuronales se basan únicamente en las redes biológicas ya que solo emulan en una parte muy simple el funcionamiento del cerebro humano. Además, se debe considerar que las redes biológicas son generadoras de procesos neurobiológicos en que se establecen relaciones de complejidad muy alta, las cuales no se puede lograr con redes mono-capas ni con redes multicapas. Las RNA pueden estudiarse como aproximadores universales desde el punto de vista matemático.

Las RNA se basan en una estructura de neuronas unidas por enlaces que transmiten información a otras neuronas, las cuales entregan un resultado mediante funciones matemáticas. Las RNA aprenden de la información histórica a través de un entrenamiento, proceso mediante el cual se ajustan los parámetros de la red, a fin de entregar la respuesta deseada, adquiriendo entonces la capacidad de predecir respuestas del mismo fenómeno. El comportamiento de las redes depende entonces de los pesos para los enlaces, de las funciones de activación que se especifican para las neuronas, las que pueden ser de tres categorías: lineal, de umbral (o escalón) y sigmoideal, y de la forma en que propagan el error.

Las redes neuronales artificiales (RNA) son sistemas de aprendizaje inspirados en el funcionamiento del cerebro humano. De esta forma simulan e imitan sistemas permitiendo establecer relaciones no lineales entre las variables de entrada y salida. Su principal ventaja que consiste en procesar información en paralelo en tiempo real ha permitido su aplicación en la clasificación y reconocimiento de patrones en sistemas complejos.

El principal potencial de las RNA es detectar no-linealidades en series temporales por lo que han sido de gran utilidad en la predicción de datos económicos y financieros. En un sistema con RNA los nodos se conectan por medio de sinapsis, esta estructura de conexión determina el comportamiento de la red. La estructura más utilizada es el Perceptrón Multicapa.

El concepto *machine learning* (ML) hace referencia al aprendizaje automático o aprendizaje de máquinas y, como tal, se engloba dentro del campo de la inteligencia artificial. La teoría del funcionamiento del ML es simple, las máquinas toman datos y aprenden por ellos mismos.

Los sistemas de aprendizaje automático permiten que un sistema aprenda a reconocer patrones por sí mismos y a hacer predicciones, a diferencia de otras técnicas más antiguas donde se requiere la codificación manual de un programa de software con instrucciones específicas para completar una tarea.

Mediante esta técnica, un sistema tiene la capacidad de aplicar rápidamente el conocimiento obtenido de grandes conjuntos de datos para conseguir distintos objetivos en poco tiempo como pueden ser, por ejemplo, un reconocimiento facial o de voz.

## Recolección de datos

Lo primero que debe hacerse es recolectar la data de muestras de cantidad de agua.

Parámetros:

**Tabla 2**

*Datos de muestras de agua*

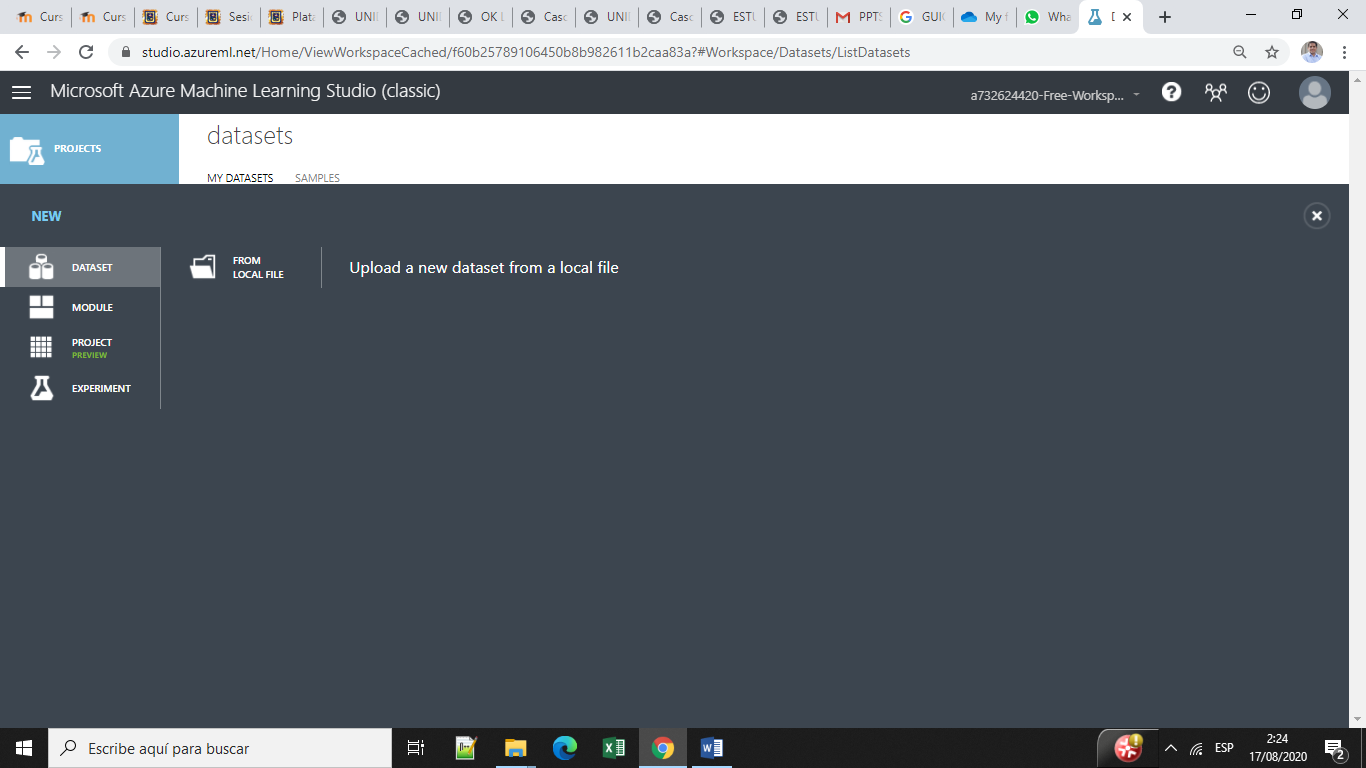
|  |  |  |  |
| --- | --- | --- | --- |
| **Símbología** | **Parámetro** | **Unidades** | **Factor de Ponderación** |
| **OD** | Oxígeno Disuelto | % Sat | 0.17 |
| **C.term** | Coliformes termotolerantes | NMP/100mL | 0.16 |
| **Ph** | Indice pH | unidad | 0.11 |
| **DBO** | Demanda Bioquímica de Oxígeno | mg/L | 0.11 |
| **T°** | Cambio de Temperatura | °C | 0.10 |
| **PO4** | Fosfatos Totales | mgPO4/L | 0.10 |
| **NO3** | Nitratos | mgNO3/L | 0.10 |
| **NTU** | Turbidez | NTU | 0.08 |
| **STD** | Sólidos Totales Disueltos | mg/L | 0.07 |
|  |  |  | 1.00 |

La calidad del agua se puede también determinar por un número de análisis cuantitativos en el laboratorio, tales como pH, sólidos totales (TS), la conductividad y la contaminación microbiana.  
El pH es el valor que determina si una sustancia es ácida, neutra o básica, calculado el número de iones de [hidrógeno](https://www.lenntech.es/periodica/elementos/h.htm) presentes. Se mide en una escala a partir de 0 a 14, en la cual, en el medio, es decir 7 la sustancia es neutra. Los valores de pH por debajo de 7 indican que una sustancia es ácida y los valores de pH por encima de 7 indica que es básica. Cuando una sustancia es neutra el número de los átomos de hidrógeno y de oxhidrilos es igual. Cuando el número de átomos de hidrógeno (H+) excede el número de átomos del oxhidrilo (OH-), la sustancia es ácida.

El parámetro WQI se correlaciona con siete parámetros, a saber, temperatura, turbidez, pH, dureza como CaCO3, conductancia, sólidos disueltos totales y recuento de coliformes fecales. Tenemos que elegir un mínimo número de parámetros para predecir el WQI, con el fin de reducir el costo del sistema. Los tres parámetros cuyos sensores están fácilmente disponibles, cuestan lo más bajo y contribuyen claramente al WQI son la temperatura, turbidez y pH, que los considera seleccionados naturalmente. El otro parámetro conveniente es total sólidos disueltos, cuyo sensor también está fácilmente disponible y se correlaciona con la conductancia y el recuento de heces coliformes, lo que significa que seleccionar TDS nos permitiría descartar los otros dos parámetros.

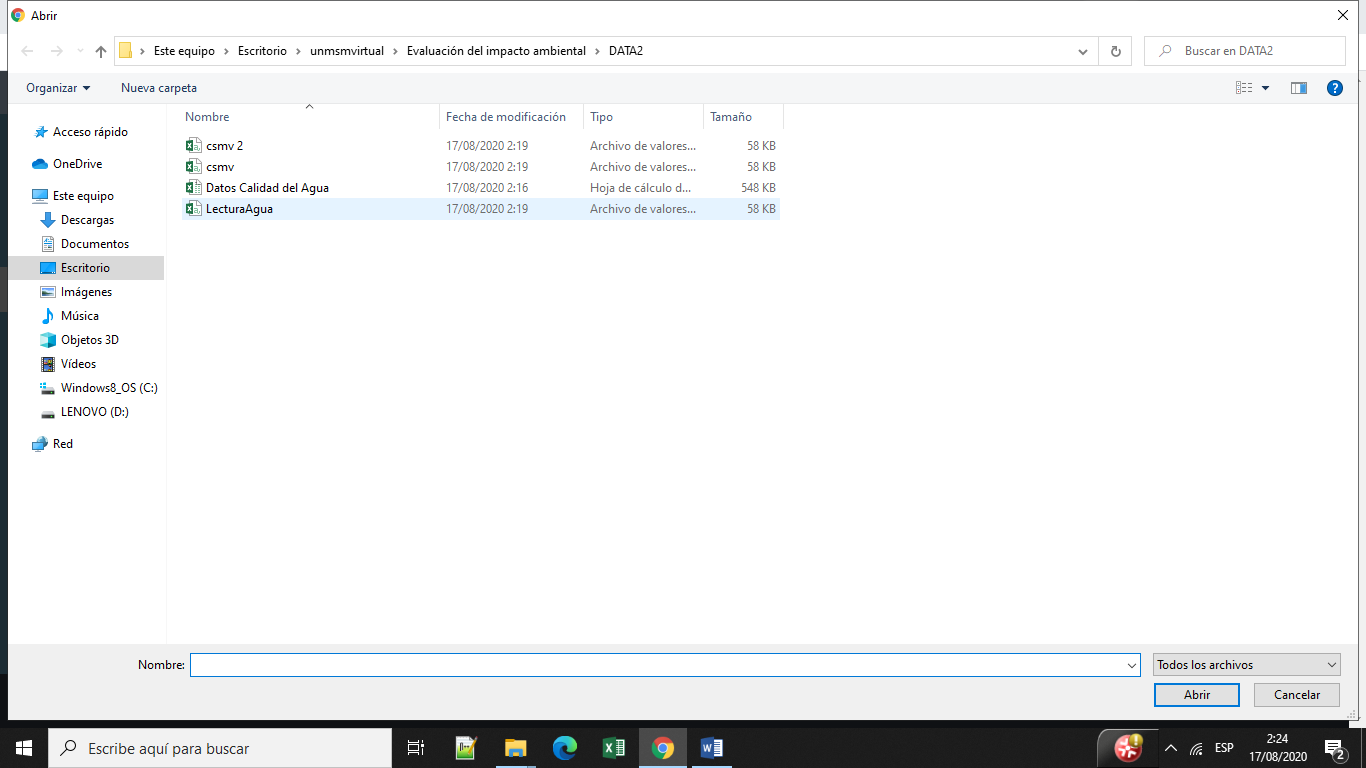
Dejamos el parámetro inconveniente restante, la dureza como CaCO3, porque no es muy correlacionado comparativamente y no es fácil de adquirir.

Ahora que hemos enumerado las observaciones del análisis de correlación, encontramos que nuestra predicción

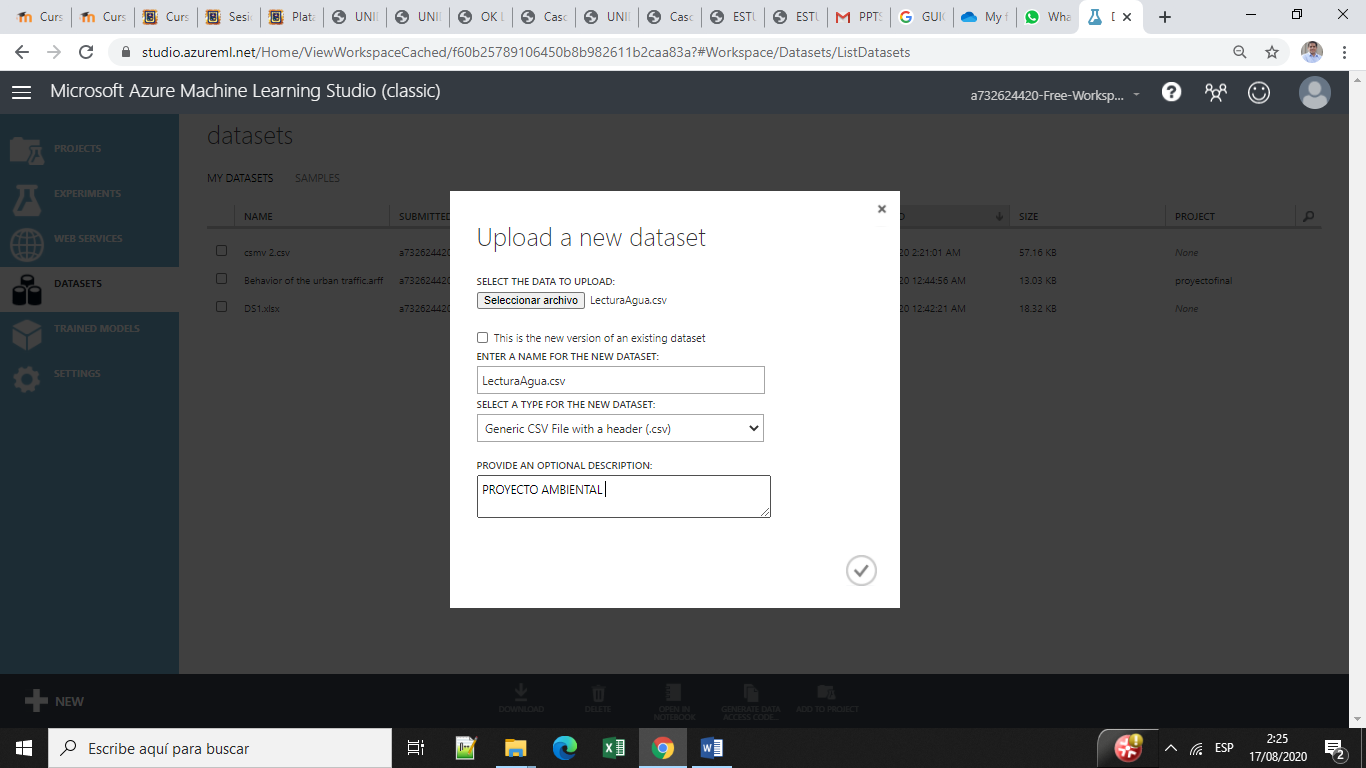


*Figura 1.*

Se seleccionan las lecturas de aguas muestreadas

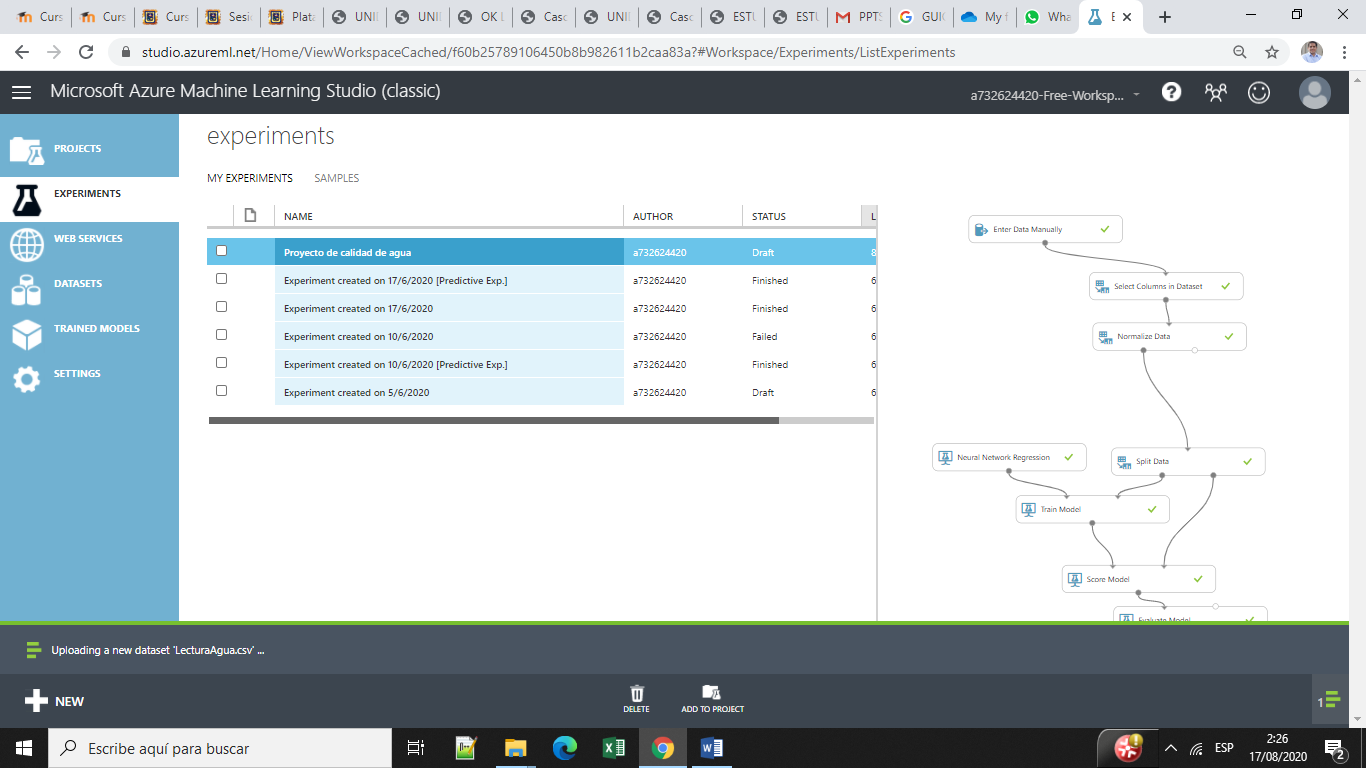


*Figura 2.*

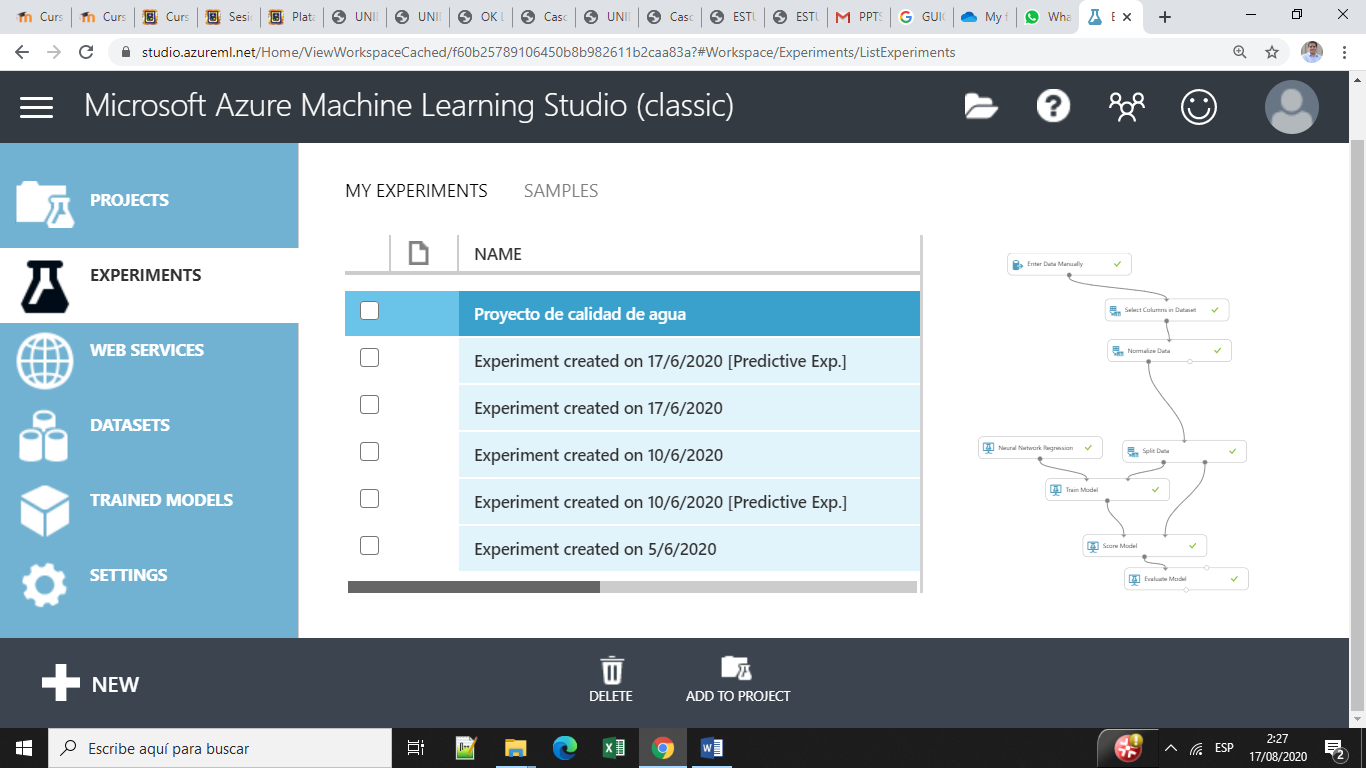


*Figura 3.*

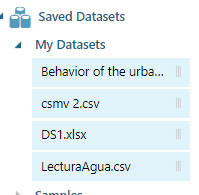
Posteriormente, vamos a la pestaña de experimentos



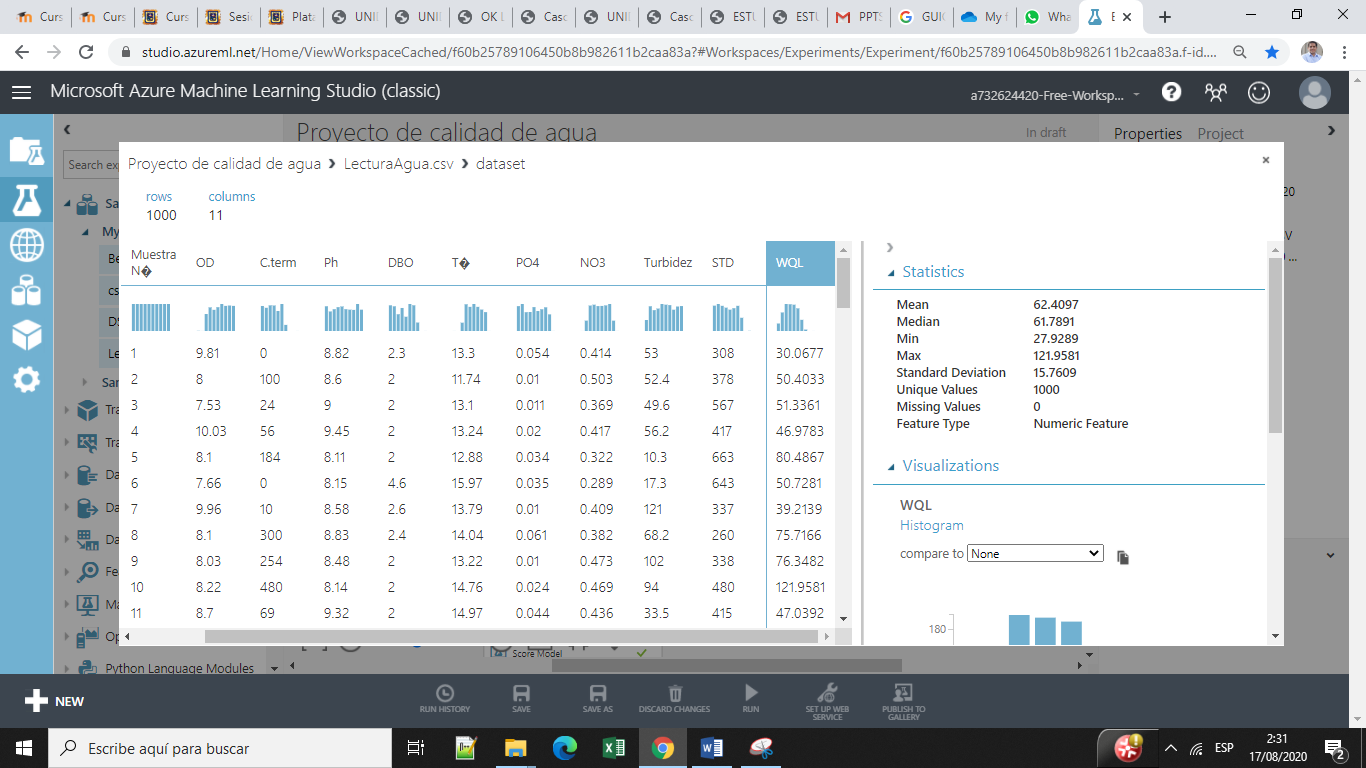
*Figura 4.*



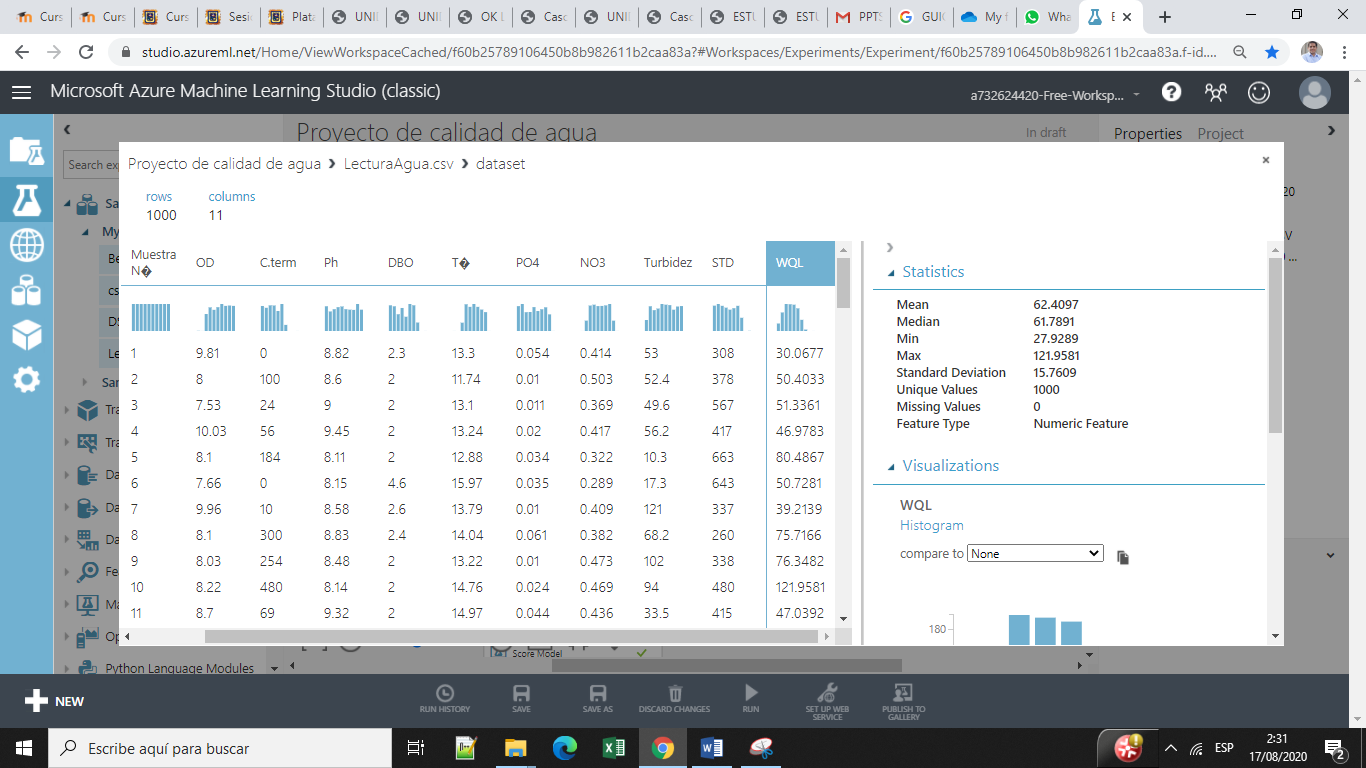
*Figura 5.*



*Figura 6.*



*Figura 7.*



*Figura 8. Boxplot Analysis and Outlier Detection*

## Normalización

1. **Zscore**

* Convierte todos los valores en una puntuación Z.
* Los valores en la columna se transforman utilizando la siguiente fórmula:
* La media y la desviación estándar se calculan para cada columna por separado.
* Se utiliza la desviación estándar de la población.

1. **MinMax**

* El normalizador min-max re escanea linealmente cada función al intervalo [0,1].
* El cambio de escala al intervalo [0,1] se realiza cambiando los valores de cada característica de modo que el valor mínimo sea 0 y luego dividiéndolo entre el nuevo valor máximo (que es la diferencia entre los valores máximos y mínimos originales).
* Los valores en la columna se transforman utilizando la siguiente fórmula:

1. **Logistic**

* Los valores de la columna se transforman mediante la siguiente fórmula:

1. **LogNormal**

* Esta opción convierte todos los valores a una escala lognormal.
* Los valores en la columna se transforman utilizando la siguiente fórmula:

**)**

* Aquí μ y σ son los parámetros de la distribución, calculados empíricamente a partir de los datos como estimaciones de máxima verosimilitud, para cada columna por separado.

1. **TanH**

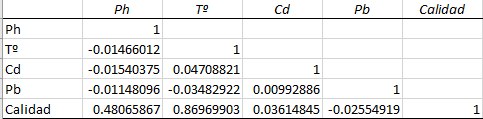
* Todos los valores se convierten en una tangente hiperbólica.
* Los valores en la columna se transforman utilizando la siguiente fórmula:

## Análisis de Correlación

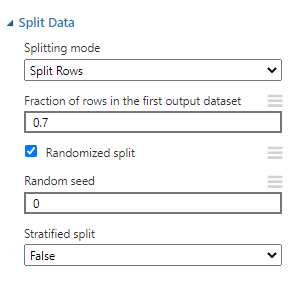
Se realiza el análisis de correlación de las variables de las lecturas

**Tabla 3**

*Resultados de correlación de variables*

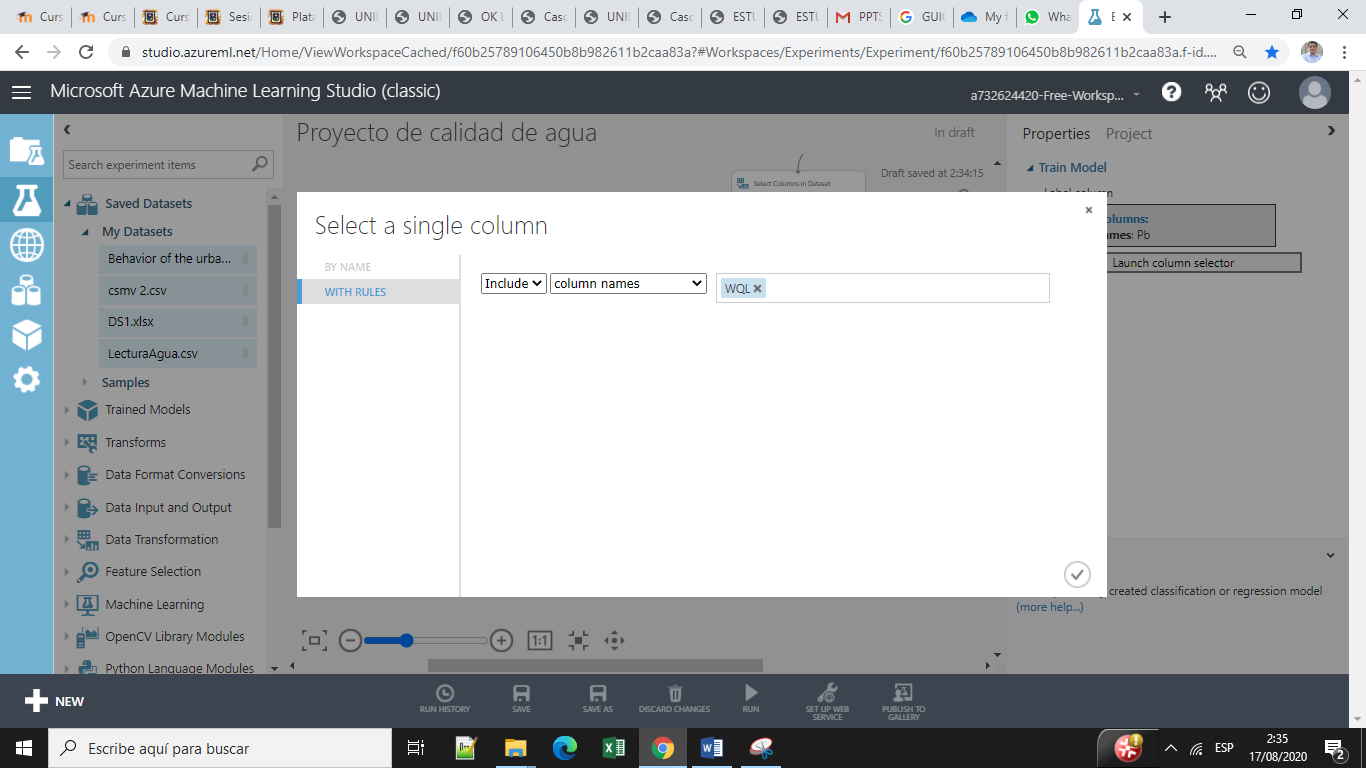


Data Splitting–Cross Validation



*Figura 9.*

Se selecciona al índice WQL como label column del experimento.



*Figura 10.*

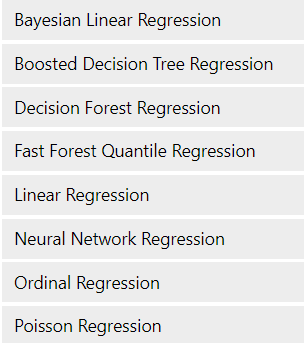
## Fase de entrenamiento

Se realizarán entrenamientos con los algoritmos disponibles en Azure Machine Learning.

### Algoritmos de regresión

Luego pasa por el bloque de tratamiento de datos Select Columns In Data Set para seleccionar las columnas que se desea procesar en la parte inferior se hallan los algoritmos regresión lineal bayesiano, regresión de decisión forestal, regresión lineal, Neural Network Regression, y regresión del árbol de decisión impulsada, en el que se basa la predicción.

Para normalización se usará el método Z score



*Figura 11.*

## Fase de pruebas

## Resultados de algoritmos de regresión

Una vez procesados la data reducida, se tuvieron los siguientes valores.

**Tabla 4**

*Resultados del procesamiento de datos en el prototipo*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Negative Log Likelihood | Mean Absolute Error | Root Mean Squared Error | Relative Absolute Error | Relative Squared Error | Coefficient of Determination |
| Bayesian Linear Regression | 759.281651 | 0.029701 | 0.035242 | 0.002245 | 0.000005 | 0.999995 |
| Boosted Decision Tree Regresion |  | 1.012617 | 1.302218 | 0.076534 | 0.006834 | 0.993166 |
| Decision Forest Regression | 687.837805 | 1.581507 | 1.997444 | 0.119531 | 0.016079 | 0.983921 |
| Linear Regression |  | 0.000033 | 0.000042 | 0.000003 | 0 | 1 |
| Neural Network Regression |  |  |  |  |  |  |
| Ordinal Regression |  |  |  |  |  |  |
| Poisson Regression |  | 2.370818 | 2.90642 | 0.179187 | 0.034042 | 0.965958 |

*Nota:* Se elige el mejor modelo

# CONCLUSIONES

* La calidad del agua se calcula convencionalmente utilizando parámetros de calidad del agua, que se adquieren a través de análisis de laboratorio que requieren mucho tiempo. Exploramos métodos alternativos de aprendizaje automático para estimarlo y encontramos varios estudios empleándolos. Estos estudios utilizaron más de 10 calidades de agua. Parámetros para predecir el WQI. Ahmad y col. [8] usó 25 parámetros de entrada, Sakizadeh [9] usó 16 parámetros, Gazzaz y col. [4] utilizaron 23 parámetros de entrada en su metodología, y Rankovic et al. [12] utilizó 10 entradas parámetros, que no son adecuados para sistemas económicos en tiempo real. Considerando que, nuestra metodología emplea “solo cuatro parámetros” de calidad del agua para predecir WQI, con un MAE de 1.96, y para predecir la calidad del agua clase con una precisión del 85%. Nuestros resultados hacen una base para una calidad de agua económica en tiempo real sistema de detección, mientras que otros estudios, aunque utilizan aprendizaje automático, utilizan demasiados parámetros para ser incorporado en sistemas de tiempo real.
* Es de vital importancia saber las problemáticas actuales que se están dando actualmente a nivel nacional, ya sea de índole política, social o económica. Esto permite abarcar una solución que pueda estar relacionada con los problemas reales y actuales que se están suscitando actualmente y nos permite obtener una solución acorde a ello, la cual resulta viable y es posible de aplicar al tener en cuenta todos esos aspectos.
* Al hacer uso de solo algunos de los parámetros para el cálculo de la calidad del agua según el WQI, por representar un menor costo en su investigación, estamos dando a relucir que no son necesarios los demás para evidenciar que la calidad del agua examinada no es apta para consumo mas sí para regadíos.
* El análisis manual de la calidad del agua a través de métodos tradicionales es muy engorroso, costoso y consume mucho tiempo cuando el conjunto de datos es demasiado amplio. Por esta razón, este proceso requiere herramientas especializadas que sean apropiadas para un análisis preciso y efectivo de la información
* Forjar la voluntad política para actuar sabiendo que cuando los recursos financieros como naturales son escasos, la atención y el compromiso político es vital para asegurar una buena toma de decisiones y las inversiones necesarias para el desarrollo y el manejo de los recursos de agua. Se tiene que generar concientización a través de programas de educación, investigaciones y proyectos relacionados con la calidad del agua.
* Para mitigar los riesgos en la salud, se debería de incorporar un tratamiento a las aguas residuales de las compañías que hacen uso de ellas para que los vertientes o desembocaduras en los ríos, lagos, canales y demás no se vean tan afectados. Además se puede financiar ideas o iniciativas para proteger los ecosistemas y prevenir la contaminación del agua.
* Un aspecto clave, lo constituye lograr una mayor sensibilización del público sobre la importancia del agua y sobre lo que se debe hacer para conseguir la seguridad hídrica, esto permitirá una participación informada de la sociedad en su conjunto.
* El machine learning está mostrando un gran potencial para la gestión de los recursos hídricos, ya que puede ayudarnos a descubrir relaciones entre variables, permitiendo modelizar sistemas complejos, generando modelos predictivos y apoyando la toma de decisiones. No obstante, no sustituye el criterio del investigador, sino que es capaz de confirmar nuestras intuiciones estudiando todas las variables involucradas y analizando una gran cantidad de datos que no podrían ser analizados en poco tiempo por una sola persona.

# AGRADECIMIENTOS

Tenemos que agradecer al constante apoyo de nuestros docentes en nuestras labores cotidianas de investigación en el desarrollo de los proyectos. Asimismo, debe agradecer a nuestra docente del curso de Evaluación del impacto ambiental, la ingeniera Fiorella Guere Salazar por las consultas sobre la evaluación de la calidad de agua; y a nuestro docente asesor, el doctor Adolfo Acevedo Borrego por evaluarnos con frecuencia y rigurosidad y que, gracias a ello, nos motivó cada día a mejorar en lo académico y en lo profesional.

# REFERENCIAS BIBLIOGRÁFICAS

1. Anónimo. (enero 21, 2019). *¿Agua para todos?: Los desafíos del sistema público de agua y alcantarillado*. Agosto 10, 2019, de El comercio Sitio web: https://elcomercio.pe/economia/peru/agua-desafios-pendientes-sistema-publico-agua-alcantarillado-noticia-599530-noticia/
2. AIDESEP. (junio 13,2014). *El petróleo ha contaminado el Amazonas del Perú durante los últimos 30 años*. Agosto 23, 2019, de Universitat Autònoma de Barcelona Sitio web: https://www.uab.cat/web/sala-de-prensa/detalle-noticia/el-petroleo-ha-contaminado-el-amazonas-del-peru-durante-los-ultimos-30-anos-1345667994339.html?noticiaid=1345672510909
3. INEI. (marzo, 2018). *Peú: Formas de acceso al agua y saneamiento básico*. Setiembre 25, 2019, de INEI Sitio web: https://www. inei.gob.pe/media/MenuRecursivo/boletines/boletín\_agua\_y\_ saneamiento.pdf
4. PLAANA. (2011). *Plan nacional de acción ambiental. Diario Oficial El Peruano, p. 80.*
5. WASH.PE. (2019). *Entre 7 y 8 millones de peruanos no tienen acceso a agua potable*. Agosto 26, 2019, de OXFAM Sitio web: https://peru.oxfam.org/qu%C3%A9-hacemos-ayuda-humanitaria/entre-7-y-8-millones-de-peruanos-no-tienen-acceso-agua-potable

1. Estudiante de ingeniería industrial en la UNMSM y de ingeniería de sistemas de la UIGV. Analista programador .NET en Stefanini IT Solutions: cliente Luz del Sur S.A.A. 2Estudiante de ingeniería industrial, UNMSM. 3Estudiante de ingeniería industrial, UNMSM. Practicante de Infraestructura TI en Scotiabank. 4Estudiante de ingeniería industrial, UNMSM. [↑](#footnote-ref-1)