**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS *(Universidad del Perú, Decana de América)***

**FACULTAD INGENIERIA INDUSTRIAL**



**PROYECTO**

**Prototipo de un modelo de machine learning para la predicción de LA CALIDAD de agua en el perú**

**PRESENTADO POR:**

* ***Oliva Moya, Ronnie Osmar 15170068***
* ***Huasacca León, Lyshet Estefani 15170024***
* ***Armas Benavide, Raul Marcelo 15170111***
* ***Egües Mendoza, Italo 07170168***

**LIMA – PERÚ**

**2020**

**Índice**

[1](#_Toc50459103)

[1 Introducción 1](#_Toc50459104)

[2 Planteamiento del problema 1](#_Toc50459105)

[1.1 Descripción de la realidad problemática 1](#_Toc50459106)

[2.1 Antecedentes 3](#_Toc50459107)

[2.1.1 Antecedentes Nacionales 3](#_Toc50459108)

[2.1.2 Antecedentes Internacionales 4](#_Toc50459109)

[Los desafíos del agua 4](#_Toc50459110)

[El derecho al agua 5](#_Toc50459111)

[El agua y los Objetivos de Desarrollo Sostenible 5](#_Toc50459112)

[Agua, saneamiento e higiene 6](#_Toc50459113)

[Saneamientos mejorados y beneficios económicos 6](#_Toc50459114)

[2.2 Justificación e importancia de la investigación 6](#_Toc50459115)

[2.3 Formulación del Problema 7](#_Toc50459116)

[2.4 Objetivo general 7](#_Toc50459117)

[2.5 Objetivos específicos 7](#_Toc50459118)

[2.6 Limitaciones de la investigación 7](#_Toc50459119)

[2.7 Hipótesis 8](#_Toc50459120)

[3 Marco Teórico 8](#_Toc50459121)

[3.1 Marco Legal 8](#_Toc50459122)

[3.2 Los problemas de la gestión del agua y sus objetivos. 10](#_Toc50459123)

[1.1.1 Sistema de alcantarillado 19](#_Toc50459124)

[3.3 Aplicación de la inteligencia artificial y el agua 21](#_Toc50459125)

[3.3.1 Machine learning: Definición 21](#_Toc50459126)

[3.3.1.1 Tipos de Machine Learning 24](#_Toc50459127)

[3.4 Deep learning 28](#_Toc50459128)

[4 Metodología de la investigación 31](#_Toc50459129)

[4.1 Tipo de la investigación 31](#_Toc50459130)

[4.2 Diseño de la investigación 31](#_Toc50459131)

[4.2.1 Instrumentos 32](#_Toc50459132)

[4.2.1.1 Azure Machine Learning 32](#_Toc50459133)

[4.2.1.2 Excel 32](#_Toc50459134)

[4.3 Técnicas instrumentos de recolección de datos 32](#_Toc50459135)

[4.3.1 Procedimientos 32](#_Toc50459136)

[4.3.2 Análisis de datos 32](#_Toc50459137)

[4.3.2.1 Procedimiento de recolección de datos 32](#_Toc50459138)

[4.4 Presupuesto de la investigación 32](#_Toc50459139)

[4.4.1 Fuentes de financiamiento 32](#_Toc50459140)

[4.5 Población y muestra 33](#_Toc50459141)

[4.6 Operacionalización de variables 33](#_Toc50459142)

[4.6.1 Calificación del impacto ambiental 33](#_Toc50459143)

[4.7 Recolección de datos 33](#_Toc50459144)

[4.8 Normalización 39](#_Toc50459145)

[4.9 Análisis de Correlación 40](#_Toc50459146)

[4.10 Fase de entrenamiento 42](#_Toc50459147)

[4.10.1 Algoritmos de regresión 42](#_Toc50459148)

[4.11 Fase de pruebas 43](#_Toc50459149)

[4.12 Resultados de algoritmos de regresión 43](#_Toc50459150)

[5 Conclusiones 44](#_Toc50459151)

[6 Referencias bibliográficas 45](#_Toc50459152)

[7 Anexos 46](#_Toc50459153)

[7.1 Matriz de consistencia 46](#_Toc50459154)

# Introducción

Este trabajo de investigación trata sobre la problemática que se está dando a nivel nacional sobre los recursos hídricos y cómo esto está afectando a la población peruana, ver la relación que hay con la gestión ambiental que se lleva a cabo en el Perú y establecer una relación con la calidad del agua en nuestro territorio nacional.

Nuestro trabajo consta de una primera parte introductoria a lo que es la problemática que vamos a abarcar, que son principalmente los problemas que existen en la actualidad sobre 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.

En una segunda instancia pasaremos al marco teórico en el cuál veremos a ciencia cierta, revisando otros autores que ya han investigado, de forma específica, las causas por las cuales se da esta contaminación a nuestros recursos hídricos. Esto nos permitirá tener una idea clara de cuál es el mejor camino que tomaremos para dar una solución a esta problemática.

Luego realizaremos un modelo de inteligencia artificial para analizar los componentes para determinar el grado de calidad de 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 6 de Agua limpia y saneamiento.

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

# Planteamiento del problema

## Descripción de la realidad problemática

El Banco Mundial (BM), ubica al Perú en el puesto 14 a nivel de Sudamérica en acceso al agua. Según datos del Ministerio del Ambiente el Perú cuenta con 106 cuencas hidrográficas; sin embargo, por acción de la naturaleza, la distribución de los recursos hídricos es muy desigual.

Según el Ministerio de Vivienda, construcción y saneamiento el 93.4% de sus residentes tiene acceso al agua potable en el ámbito urbano y solo el 63.2% en el rural. Uno de los primeros eventos para discutir la situación sobre el agua se realizó en noviembre del año 2000, se denominó “Perú Agua para el siglo XXI de la visión a la acción”. Esto se dio luego del II Foro Mundial del Agua para definir los objetivos del siglo XXI en materia de manejo del agua. Paradójicamente, aunque el Perú tenga una oferta hídrica abundante, el país se considera un país carente de agua. En las ciudades en la costa, como Lima, la contaminación del agua se produce en forma directa, a través de acequias, ductos y desagües. Esto ha sido una consecuencia de una poca, casi nula intervención política en materia de cuidado del agua a lo largo de los últimos siglos, tal como lo señala el escritor Jorge Lossio en su libro **Acequias y Gallinazos**. Por ejemplo, el río Rímac, sirve casi exclusivamente para atender a la población limeña. En 1996, la extracción bordeaba alrededor de 400 millones de m3/año. El río Rímac ha ido descendiendo poco a poco. En la sierra, la calidad de agua se ha visto afectada por actividades extractivas y actividades productivas como la agricultura (por el uso de abonos sintéticos, plaguicidas y fertilizantes).

En la sierra, el agua está muy contaminado, esto se puede apreciar en las subcuencas de los ríos San José- Anticona, San Juan, Yauli, Huayhuay, laguna de Huascacocha, lago de Junín, de los Reyes o Chinchaycocha; cuenca del río Mantaro, zona agrícola del Valle del Mantaro, etc., especialmente por los relaves, botaderos de desmontes, sedimentos, rebose de las canchas de relaves, agua ácida de las minas, aguas servidas, dispersión de los contaminantes secos por el aire y otros tóxicos, que eliminan sin ningún tratamiento a los recursos hidrográficos nombrados, existen 17 compañías mineras en actividad, la refinería de la Oroya y las 67 mineras abandonadas.

Esta situación también se repite en la selva donde los ríos del departamento de Madre de Dios, en la selva sur del país, son los más contaminados del país por la presencia de metales pesados, entre ellos el mercurio, utilizados por la minería ilegal, advirtió hoy la Autoridad Nacional del Agua (ANA).

Betty Chung, titular de la Dirección de Gestión de Calidad de los Recursos Hídricos de la ANA, dijo que un monitoreo arrojó que los ríos Huepetuhe, Caichihue, Dos de Mayo y la quebrada Jayave, que forman parte de la cuenca del Inambari, evidencian gran contaminación.

La evaluación concluyó en que dichos ríos contienen diversos metales pesados como arsénico, mercurio, plomo, cobre, níquel, entre otros.

## 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.

#### Los desafíos del agua

* 2,1 billones de personas carecen de acceso a servicios de agua potable gestionados de manera segura (OMS/UNICEF 2017).
* 4,5 billones de personas carecen de servicios de saneamiento gestionados de forma segura (OMS/UNICEF 2017).
* 340 000 niños menores de cinco años mueren cada año por enfermedades diarreicas (OMS/UNICEF 2017).
* La escasez de agua ya afecta a cuatro de cada 10 personas (OMS).
* El 90% de los desastres naturales están relacionados con el agua (UNISDR).
* El 80% de las aguas residuales retornan al ecosistema sin ser tratadas o reutilizadas (UNESCO, 2017).
* Alrededor de dos tercios de los ríos transfronterizos del mundo no tienen un marco de gestión cooperativa (SIWI).
* La agricultura representa el 70% de la extracción mundial de agua (FAO).
* Aproximadamente el 75% de todas las extracciones de agua industrial se utilizan para la producción de energía (UNESCO, 2014).

#### El derecho al agua

Uno de los hitos recientes más importantes ha sido el reconocimiento por parte de la Asamblea General de las Naciones Unidas del [derecho humano al agua y al saneamiento](http://www.un.org/spanish/waterforlifedecade/human_right_to_water.shtml) en julio de 2010. La Asamblea reconoció el derecho de todos los seres humanos a tener acceso a una cantidad de agua suficiente para el uso doméstico y personal (entre 50 y 100 litros de agua por persona y día) y que sea segura, aceptable y asequible (el coste del agua no debería superar el 3% de los ingresos del hogar), y accesible físicamente (la fuente debe estar a menos de 1.000 metros del hogar y su recogida no debería superar los 30 minutos).

#### El agua y los Objetivos de Desarrollo Sostenible

[El Objetivo de Desarrollo Sostenible (ODS) 6](http://www.un.org/sustainabledevelopment/es/water-and-sanitation/) es «Garantizar la disponibilidad de agua y su gestión sostenible y el saneamiento para todos». Las metas de este objetivo cubren tanto los aspectos del ciclo del agua como los sistemas de saneamiento, y la consecución de este objetivo se ha planteado para que contribuya en el progreso de otros ODS, principalmente en salud, educación, crecimiento económico y medio ambiente.

Las Naciones Unidas llevan mucho tiempo abordando la crisis mundial derivada de un abastecimiento de agua insuficiente y la creciente demanda de agua para satisfacer las necesidades humanas, comerciales y agrícolas.

[La Conferencia de las Naciones Unidas sobre el Agua (1977)](http://www.who.int/water_sanitation_health/unconfwater.pdf), [el Decenio Internacional del Agua Potable y del Saneamiento Ambiental (1981-1990)](https://undocs.org/es/A/RES/35/18), [la Conferencia internacional sobre el agua y el medio ambiente (1992)](http://www.wmo.int/pages/prog/hwrp/documents/english/icwedece.html) y la [Cumbre para la Tierra (1992)](https://sustainabledevelopment.un.org/milestones/unced) — se centraron en este recurso vital. En concreto, el Decenio ayudó a unos 1.300 millones de personas de países en desarrollo a conseguir acceso a agua potable.

El [Decenio Internacional de Acción "Agua para la Vida" 2005-2015](http://www.un.org/spanish/waterforlifedecade/) contribuyó a que alrededor de 1,3 billones de personas en los países en desarrollo obtuvieran acceso al agua potable e impulsó el progreso en materia de saneamiento como parte del esfuerzo por alcanzar los Objetivos de Desarrollo del Milenio.

Los últimos acuerdos clave incluyen la [Agenda 2030 para el Desarrollo Sostenible](http://www.un.org/sustainabledevelopment/es/la-agenda-de-desarrollo-sostenible/), el [Marco de Sendái para la Reducción del Riesgo de Desastres 2015-2030](http://www.preventionweb.net/files/resolutions/N1516720.pdf), la [Agenda de Acción de Addis Abeba 2015 sobre la Financiación para el Desarrollo](https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=2051&menu=35), y el [Acuerdo de París 2015 dentro del Marco de la Convención de las Naciones Unidas sobre el Cambio Climático](http://unfccc.int/files/essential_background/convention/application/pdf/spanish_paris_agreement.pdf).

#### Agua, saneamiento e higiene

Las aguas contaminadas y la falta de saneamiento básico obstaculizan la erradicación de la pobreza extrema y las enfermedades en los países más pobres del mundo.

En la actualidad, 2,3 billones de personas no disponen de instalaciones básicas de saneamiento, como baños o letrinas. Según el [Programa Conjunto OMS/UNICEF de Monitoreo del Abastecimiento de Agua y del Saneamiento](https://washdata.org/), al menos 1800 millones de personas en todo el mundo beben agua que no está protegida contra la contaminación de las heces. Un número aún mayor bebe agua que se distribuye a través de sistemas vulnerables a la contaminación.

El agua no potable y el saneamiento deficiente son las causas principales de la mortalidad infantil. La diarrea infantil -asociada a la escasez de agua, saneamientos inadecuados, aguas contaminadas con agentes patógenos de enfermedades infecciosas y falta de higiene- causa la muerte a 1,5 millones de niños al año, la mayoría de ellos menores de cinco años en países en desarrollo.

#### Saneamientos mejorados y beneficios económicos

La relación entre la falta de agua y saneamiento y los objetivos de desarrollo es obvia y solucionar el problema es rentable. Los estudios muestran que cada dólar invertido se traduce en un beneficio de 9. Ese retorno de la inversión beneficia específicamente a niños pobres y comunidades desfavorecidas que más lo necesitan.

## 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.

## 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.
* 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

## 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 quizo delimitar la investigación.

## Hipótesis

Hipótesis General:

Desarrollar el prototipo de un modelo de machine learning predictivo, alimentada con los datos de un sensor ambiental, que permita predecir la calidad de agua todo esto enfocado en la localidad de Lima.

Hipótesis específicas:

* Se normalizan los datos recopilados para alimentar las redes neuronales de las estaciones de monitoreo del agua en Lima.
* Encontrar una arquitectura de diseño que nos brinde la cardinalidad de capas ocultas del modelo de red neuronal.
* Con un algoritmo de entrenamiento podremos calcular los pesos de la red del modelo de red neuronal.
* Se corrobora la efectividad del modelo mediante los resultados obtenidos en el prototipo implementado.
* Se logra reducir el costo del sistema de monitoreo y su complejidad.

# Marco Teórico

En esta sección se elabora una revisión bibliográfica de los conceptos generales a partir de los cuales se sustenta el análisis de las redes neuronales artificiales y el agua.

## Marco Legal

Con el cuidado del agua, el Perú ha tenido avances en su marco legal.

* Creación de la Autoridad Nacional del Agua (2008).
* Ley de Recursos hídricos, Ley N° 29338 y su reglamento.
* Gestión Integral de Recursos Hídricos (GIRH).
* Política de Estados sobre los Recursos Hídricos – Política 33 del Acuerdo Nacional (2012).
* Creación del Consejo Directivo del ANA (2012).
* Política y Estrategia Nacional de Recursos Hídricos (5 políticas y 16 estrategias; en proceso de aprobación).
* Política Nacional del Ambiente, con 4 ejes de política (2009).
* Plan Nacional de Gestión de Recursos Hídricos (11 estrategias, 30 programadas; en proceso de aprobación).
* Planes de Gestión de Recursos Hídricos de cuenca (6 cuencas piloto en 2013; Programa de Modernización de gestión de recursos hídricos en cuencas).

En el Perú ya se ha establecido políticas de los recursos básicos que se encuentran establecidos en el Plan Nacional de Recursos Hídricos (D. S. N° 013-2015-MINAGRI), el cual se enmarca en cinco Políticas, que son: gestión de la cantidad, gestión de la calidad, gestión de la oportunidad, gestión de la cultura del agua, y adaptación al cambio climático y eventos extremos.

El Plan Nacional de Recursos Hídricos ha utilizado los instrumentos de la planificación que marca la Ley de Recursos Hídricos, donde se establecen las políticas y estrategias, que se orientan a conseguir los objetivos que se definieron en la política y estrategia nacional de recursos hídricos (D. S. N° 006-2015-MINAGRI). Dichas políticas y estrategias son las siguientes:

· Lograr la conservación de los ecosistemas y los procesos hidrológicos, así como la determinación y planificación de la oferta y disponibilidad hídrica para optimizar la atención de la demanda.

· Recuperar y proteger la calidad de los recursos hídricos y fiscalizar los agentes contaminantes.

· Atender de manera oportuna la demanda de recursos hídricos para garantizar el acceso al agua como derecho humano.

· Promover una cultura del agua por la paz para lograr la gestión integrada de los recursos hídricos con un enfoque solidario.

· Identificar la variedad climática y sus impactos sobre los recursos hídricos para promover la adaptación al cambio climático y disminuir la vulnerabilidad frente a los eventos extremos.

Con este marco conceptual se han identificado 30 programas de medidas, que se han estructurado a partir de los instrumentos de planificación indicados; es decir, se han organizado por estrategias y políticas, de manera que las soluciones estén alineadas con la legislación de aguas vigente.

## Los problemas de la gestión del agua y sus objetivos.

|  |  |  |
| --- | --- | --- |
| **Dimensión** | **Problemas** | **Propuestas** |
| Equilibro ecosistémico | Contaminación por parte de actividades económicas informales. | * Mayor control por parte del Estado estableciendo mayor control militar en zonas críticas donde se realizan las principales actividades económicas informales. |
| Eficiencia económica | Gestión ineficaz de la demanda. | * Establecer un planeamiento estratégico para que se pueda dar una reorganización y pueda llevarse a cabo una distribución y evitar las pérdidas económicas que se dan por esta actual organización. * Priorizar el uso del agua en aquellas actividades que produzcan mayores beneficios al país. * La desalación del agua de mar ofrece una alternativa de ejecución relativamente rápida. |
| Equidad social | Existen muchos peruanos sin acceso al agua potable.  Escasa cultura del agua | * Llevar a cabo programas sociales por parte del Estado para hacer llegar a todos los peruanos el conocimiento de la importancia de lo que es el cuidado del agua, estos deben ser supervisados por las municipalidades, para que de este modo tenga un mayor abarque entre toda la población. |



Reducir la huella hídrica

La huella hídrica es un [indicador](https://es.wikipedia.org/wiki/Indicador_ambiental) del uso de agua dulce que hace referencia tanto al uso directo del agua de un consumidor o productor, como a su uso indirecto.  La huella hídrica puede considerarse como un indicador integral de la apropiación de los recursos de agua dulce**,** ya que se utiliza para medir el volumen total de agua dulce usado para producir los bienes y servicios producidos por una empresa, o consumidos por un individuo o comunidad.

En el país, recientemente ha habido esfuerzos por medir la huella hídrica a nivel nacional. Hay algunos estudios donde colaboraron empresas grandes. Se tuvo como resultado datos de relevancia como:

En promedio, la huella hídrica total del arroz en el Perú es 6496,04 m3/ton.



**Figura 4.1. La huella hídrica**

Optimizar el uso y re-uso del agua Compartir los beneficios económicos, sociales y ambientales de los ríos, lagos y acuíferos transfronterizo.

**ESTRATEGIAS**

Para cumplir los objetivos de esta política, se determinaron las dos estrategias siguientes, que se justifican de la siguiente manera:

· **Estrategia para la mejora de la calidad de las aguas.**

El deterioro de la calidad del agua no solo afecta los ríos para diferentes usos, sino que también produce daños ecológicos y disminuye el valor del agua como bien económico. La mejora de la calidad de las aguas es, por tanto, una tarea ineludible e inaplazable, para lo que, como primera actuación, es necesario conocer su calidad actual. Los programas que se han de llevar a cabo en el marco de esta estrategia son los siguientes:

– Programa de mejora del conocimiento de la calidad de las aguas superficiales.

– Programa de mejora del conocimiento de la calidad de las aguas subterráneas.

– Programa de supervisión y fiscalización de vertimientos de aguas residuales.

– Programa de regulación normativa de la calidad de las aguas y buenas prácticas en el uso del agua.

· **Estrategia para la mejora y ampliación de la cobertura de los servicios de saneamiento.**

Estos servicios incluyen el suministro de agua potable y el alcantarillado, así como el tratamiento de las aguas residuales. El Estado debe garantizar el acceso de toda la población a estos servicios, a través del reconocimiento de la importancia que tienen para el cuidado de la salud pública, la superación de la pobreza, la dignidad humana, el desarrollo económico y la protección del ambiente, tanto en las poblaciones urbanas como en las rurales. Los avances logrados en el territorio peruano sobre la prestación de los servicios de saneamiento son todavía lentos, debido a la organización y estructura de la industria, la disponibilidad de los recursos, las dificultades en la aplicación del sistema tarifario, las políticas no consolidadas y tampoco asumidas por todos, y la ausencia de una visión a largo plazo, entre otras causas.

Los programas por llevar a cabo en el marco de esta estrategia son los siguientes:

– Programa de aumento de la cobertura de agua potable.

– Programa de aumento de la cobertura de alcantarillado.

– Programa de aumento de la cobertura de tratamiento de aguas residuales.

**a. Política 1: Gestión de la cantidad de agua**

Para cumplir los objetivos de esta política, se han determinado las tres estrategias que se indican a continuación, que se justifican de la siguiente manera:

· **Mejora del conocimiento de los recursos y las demandas.**

Un conocimiento insuficiente de los recursos y las demandas hídricas conlleva graves dificultades para elaborar balances hídricos rigurosos. Asimismo, dificulta el otorgamiento de derechos de uso de agua y el cobro de retribuciones económicas y tarifas.

Por otra parte, crea falsas expectativas sobre disponibilidad de recursos hídricos, las cuales suelen desembocar en conflictos sociales.

Entre otras consecuencias, se da una sobreexplotación de acuíferos y se dificulta la previsión de las infraestructuras necesarias para atender los diferentes usos. Por tanto, los programas por llevar a cabo en el marco de esta estrategia son los siguientes:

– Implantación de una Red hidrometeorológica nacional.

– Aumento del conocimiento de las aguas subterráneas.

– Implantación del Sistema nacional de información de la cantidad.

**· Mejora de la eficiencia del uso del agua y gestión de la demanda.**

A escala nacional, la eficiencia media del riego está en torno al 35 %, mientras que la eficiencia media en el abastecimiento de agua potable es del orden del 45 %. Para evitar que el uso ineficiente conlleve pérdida de recursos hídricos, acceso no equitativo al agua, pérdidas económicas en los sectores productivos, límite de la expansión agrícola y salinización de los suelos por exceso de riego, de manera que genere conflictos, es necesario adoptar medidas de diversa tipología: técnicas, económicas, administrativas y culturales, de las que las tres últimas se desarrollan en otras políticas. Las medidas de tipo técnico o estructural necesarias para la mejora de la eficiencia del uso del agua se distribuyen en los siguientes programas:

– Control y medición de la demanda.

– Mejoramiento de los sistemas de conducción y distribución del agua.

– Tecnificación del riego.

– Ampliación sostenible de la frontera agrícola.

**· Aumento de la disponibilidad del recurso.**

En esta estrategia se han desarrollado los siguientes programas:

– Incremento de la regulación superficial de los recursos hídricos y la transferencia de recursos entre cuencas.

– Reforestación de las cabeceras de cuencas vertientes a embalses.

– Eliminación de la sobreexplotación de acuíferos.

– Reutilización de aguas residuales tratadas y desalinización del agua de mar.

**b. Política 2: Gestión de la calidad del agua**

Aunque el agua superficial disponible en el Perú es abundante, su calidad es crítica en algunas regiones del país. Este deterioro de la calidad del agua es uno de los problemas más graves, ya que es un impedimento para lograr un uso eficiente del recurso, y compromete el abastecimiento —tanto en calidad como en cantidad—, la salud de las personas, del ganado, la producción agrícola y la conservación del medio ambiente. Las fuentes naturales ven afectada su calidad por diferentes fuentes contaminantes, que merman el potencial de su utilización para usos poblacionales y productivos demandados aguas abajo. El crecimiento de las ciudades está originando el aumento de la contaminación del agua en las fuentes naturales causada por el vertimiento de aguas residuales sin tratar.

Por otro lado, se presenta la descarga de agua contaminada procedente de pasivos ambientales históricos, minería informal e ilegal, y otras actividades productivas, cuyo destino final son las fuentes naturales de agua.

Por todo ello, el objetivo general de esta política se ha enfocado en:

– Promover la mejora del conocimiento de la calidad de las aguas, su recuperación y protección.

– Establecer una vigilancia, fiscalización y mitigación de las fuentes contaminantes.

– Mejorar y ampliar la cobertura de los servicios de saneamiento.

**c. Política 3: Gestión de la oportunidad**

Por gestión de la oportunidad se entiende el “Atender de manera oportuna la demanda de los recursos hídricos, mejorando su distribución, temporal y espacial; garantizando el derecho al agua, en el marco de la seguridad hídrica y seguridad alimentaria priorizándose el desarrollo de infraestructura hidráulica para satisfacer la demanda hídrica poblacional y agraria en zonas de mayor vulnerabilidad.

Así, la gestión de la oportunidad es un proceso centrado en la gestión integrada de los recursos hídricos por cuencas hidrográficas —incluyendo las transfronterizas—, el fortalecimiento de aspectos administrativos de la gestión del agua y la promoción de inversiones públicas y privadas para el desarrollo de infraestructura hidráulica, con prioridad en zonas de pobreza.

Para avanzar en una gestión integrada de los recursos hídricos que permita su uso racional para maximizar el bienestar económico y social, sin comprometer su sostenibilidad y la de los ecosistemas asociados, se han definido estrategias y cuatro programas.

**· Estrategia 6 Implementación de la gestión integrada de los recursos hídricos**

– Programa 19: Fortalecimiento institucional de la GIRH.

– Programa 20: Fortalecimiento administrativo y económico de la GIRH.

– Programa 21: Implementación de la GIRH en cuencas transfronterizas.

**· Estrategia 7: Desarrollo de riego y saneamiento con prioridad en zonas de pobreza**

– Programa 22: Desarrollo de riego y saneamiento en zonas de pobreza.

**d. Política 4: Gestión de la cultura del agua**

La política de la gestión de la cultura del agua se entiende como un proceso de concientización de los actores y de la sociedad en general respecto a la importancia de la gestión integrada de los recursos hídricos. Implica dar a conocer la necesidad de conservar los recursos hídricos en cantidad y calidad adecuada, potenciar su utilización eficiente, conocer y pagar los costos que lleva su disponibilidad, y de impulsar el concepto de hidrosolidaridad entre los usuarios, para construir escenarios participativos de desarrollo sostenible.

Para dar cumplida respuesta a la LRH y a la PENRH, así como a los problemas identificados en la fase de diagnóstico, se han desarrollado dos estrategias y dos programas de medidas para cada una de ellas:

**· Estrategia de coordinación institucional y gobernanza hídrica:**

– Programa de coordinación institucional del SNGRH.

– Programa de hidrosolidaridad y gobernanza hídrica.

**· Estrategia de educación ambiental y cultura del agua:**

– Programa de gestión del conocimiento y cultura del agua por la paz.

– Programa de comunicación, difusión y sensibilización de los actores de la GIRH.

**e. Política 5: Adaptación al cambio climático y eventos extremos**

La política de adaptación al cambio climático y eventos extremos se entiende como un proceso centrado en la investigación y elaboración de estudios sobre el clima y sus efectos sobre los recursos hídricos de las cuencas, así como en la programación de una serie de acciones para disminuir los perversos efectos de los fenómenos extremos como las inundaciones y sequías.

Para desarrollar esta política se proponen dos estrategias y dos programas:

**· Estrategia para la adaptación al cambio climático:**

– Programa de mejora del conocimiento de los efectos del cambio climático.

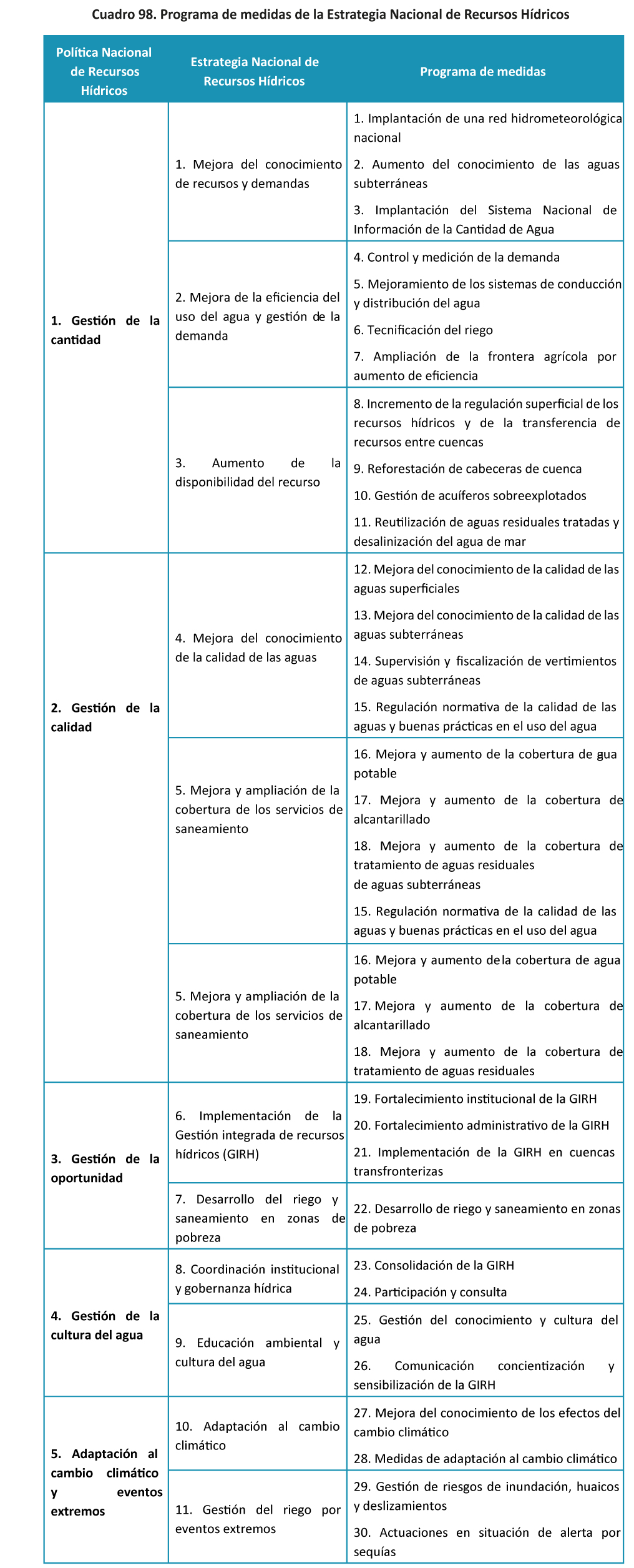
– Programa de medidas de adaptación al cambio climático.

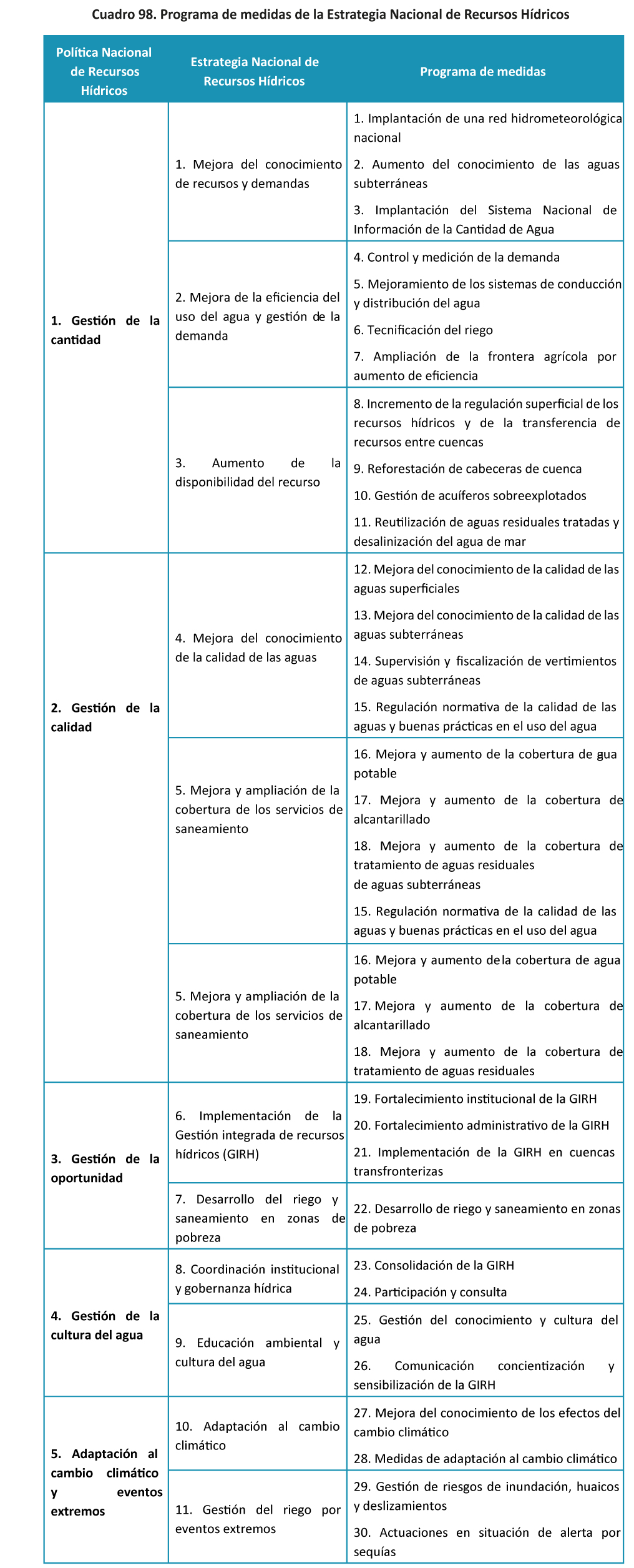
**· Estrategia para la gestión del riesgo por eventos extremos:**

– Programa para la gestión de los riesgos de inundación, huaicos y deslizamientos.

– Programa de actuación en situaciones de alerta por sequía.

A modo de resumen, el siguiente cuadro explica cada uno de los programas incluidos en el PNRH.

****

****

**Fuente: Ministerio de Agricultura y Riego (Minagri). Autoridad Nacional del Agua (ANA).  
Elaboración: Equipo técnico EsDA-MINAM**.

Como se puede visualizar en el cuadro, los objetivos y metas de la gestión de los recursos hídricos, establecidos en el Plan Nacional de Recursos Hídricos, se encuentran estructurados dentro de la Política Nacional de Recursos Hídricos. La presente investigación se centrará en la calidad del agua.

### Sistema de alcantarillado

Saneamiento básico es la tecnología de más bajo costo que permite eliminar higiénicamente las excretas y aguas residuales y tener un medio ambiente limpio y sano tanto en la vivienda como en las proximidades de los usuarios. El acceso al saneamiento básico comprende seguridad y privacidad en el uso de estos servicios. La cobertura se refiere al porcentaje de personas que utilizan mejores servicios de saneamiento, a saber: conexión a alcantarillas públicas; conexión a sistemas sépticos; letrina entre otros. (INEI, 2018)

En el Perú, la provisión de agua y alcantarillado se encuentra a cargo exclusivamente del Estado. En el ámbito urbano, las encargadas son las EPS, mientras que, en el ámbito rural, las responsables son las organizaciones comunales.

Actualmente, existen 50 EPS en el país y muestran resultados operativos y financieros heterogéneos. Al cierre del 2017, más de la mitad de las EPS operaron con pérdidas financieras. De ellas, a la fecha, 13 presentan severas deficiencias financieras y operativas, por lo cual se encuentran bajo el Régimen de Apoyo Transitorio (RAT) a través de la OTASS.

Los diversos problemas que enfrentan actualmente las EPS se encuentran relacionadas con ineficiencias internas y los escasos recursos económicos que manejan. Estos recursos provienen del cobro de la tarifa -establecida por la Superintendencia Nacional de Servicios de Saneamiento (SUNASS)- que pagan los usuarios.

Sin embargo, la capacidad de las EPS de facturar el agua provista es limitada, y el monto de la tarifa entre aquellos que sí pagan es insuficiente para cubrir tanto los costos de operación como de nuevas inversiones destinadas a mejorar el servicio. Ante ello, entre el 2017 y 2018 se transfirió, aproximadamente, S/500 millones a las EPS con el fin de fortalecer sus capacidades de gestión, pero manteniendo sus limitaciones institucionales de origen, posiblemente perpetuando el problema.

Según el último Plan Nacional de Infraestructura de AFIN, la brecha de infraestructura del sector saneamiento para el período 2016-2025 ascendía a US$12 mil millones, de los cuales el 79% corresponden al acceso a alcantarillado.

La priorización de gasto en el sector ha sido clara: mientras que la inversión pública nacional creció en 68% entre 2009 y 2018, la del sector saneamiento aumentó en 84%. Así, de los S/10,244 millones adicionales invertidos en total en dicho período, el 15% corresponde a la función de agua y alcantarillado, solo por detrás de Transportes (35%) y Educación (30%).

La inversión de este sector es responsabilidad principalmente de los gobiernos locales, con participación en las EPS. Sin embargo, durante el período 2015-2018, las municipalidades ejecutaron menos de dos tercios de su presupuesto para inversión en infraestructura de saneamiento.

Si bien es cierto que las vertientes con mayor aumento de uso del agua por todos los tipos de uso son las vertientes del Amazonas y del Titicaca, la situación de vulnerabilidad hídrica de esta última es mucho mayor por la escasez de agua imperante. Mientras que la vertiente del Titicaca ha cuadriplicado su uso agrícola del agua, la vertiente del Pacífico lo ha reducido, pasando de usar para fines agrícolas y pecuarias 14,228.00 hm³ en 1992 a 13,196.17 hm³ en 2016.

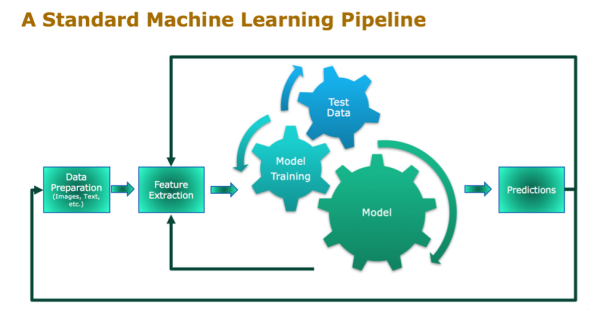
Preocupante es el aumento del uso del agua en la vertiente del Amazonas. El aumento de población ha sido mayor en la vertiente del Pacífico (en cifras absolutas) y aparentemente, por la disponibilidad de agua existente en esta vertiente, no existen todavía esfuerzos tangibles para lograr una eficiencia del manejo del agua.

## Aplicación de la inteligencia artificial y el agua

### Machine learning: Definición

Machine Learning Machine Learning es una forma de IA (Inteligencia Artificial) 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 mas 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 mas 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.

#### Tipos de *Machine Learning*

Los diferentes tipos de aprendizaje automático (*machine* *learning*) se dividen principalmente en tres grandes categorías: aprendizaje supervisado (*supervised* *learning*), aprendizaje no supervisado (*unsupervised* *learning*) y aprendizaje por refuerzo (*reinforcement* *learning*).

1. **Aprendizaje supervisado**

En este tipo de aprendizaje, los algoritmos utilizan datos que han sido previamente etiquetados (*labeled data*), buscando dar con una función que, dependiendo de los datos de entrada (*input data*), les asigne la etiqueta de salida que sea adecuada.

Además, el algoritmo es capaz de aprender de sus propias decisiones anteriores, es decir, utiliza un histórico de datos donde se encuentran todos los utilizados en ocasiones anteriores y de esta forma aprende a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, es capaz de predecir el valor de salida.

Dentro de los principales métodos de aprendizaje supervisado nos encontramos con los siguientes:

* **Árbol de decisión (Decision tree):** consiste en una herramienta que se encarga de realizar predicciones en base a un conjunto de datos, construyendo para ello una serie de diagramas lógicos que sirven para representar las condiciones que se suceden progresivamente, para finalmente dar solución a un determinado problema. Estos árboles de decisión están formados por nodos, vectores de números, flechas y etiquetas.

* **Clasificador bayesiano ingenuo (*Naive* *Bayes* *Classifier*)**: se trata de un clasificador probabilístico basado en el teorema de Bayes, que vincula la probabilidad del evento ‘A’ dado ‘B’ con la probabilidad de ‘B’ dado ‘A’, junto con otros tipos de hipótesis simplificadoras. De forma resumida, se podría decir que este algoritmo asume que la presencia o ausencia de una característica no depende de la presencia o ausencia de cualquier otra característica, por cuanto se trata de una clase variable.

* **Mínimos cuadrados ordinarios (*Ordinary* *Least* *Squares o OLS*)**: también conocido como mínimos cuadrados lineales, consiste en un método para encontrar los parámetros ‘β’ y estadísticas poblacionales en un modelo de regresión lineal, que sirve para aproximar la relación de dependencia entre una variable dependiente ‘Y’, otra independiente ‘X’, y un término aleatorio ‘ε’.

* **Regresión logística (*Logistic* *regression*)**: se trata de un tipo de análisis de regresión empleado para predecir la solución de una variable categórica en base a otras variables independientes o predictorias. Las probabilidades que representan el posible resultado de un ensayo concreto se modelan, como una formula de variables explicativas, utilizando una función logística. Este algoritmo es útil para analizar datos distribuidos binomialmente.

* **Máquinas de soporte vectorial (*Support* *Vector* *Machines* *o* SVM):** se trata de un conjunto de algoritmos relacionados con problemas de clasificación y regresión. Su funcionamiento consiste en el entrenamiento de una SVM a través de etiquetar las clases de una serie de supuestos a modo de ejemplo, consiguiendo así que la máquina, ante un nuevo supuesto, sea capaz de predecir la clase a la que se corresponde. La SVM representa, mediante un hiperplano, los distintos puntos de muestra en el espacio, realizando una amplia separación entre dos espacios, y permitiendo así poder definir la clase de la que se trata en función de la mayor o menor proximidad que tenga con cada una de las clases.

* **Método ensamble (*Ensemble* *Method*):** los métodos de ensamble o métodos de conjunto utilizan algoritmos de aprendizaje múltiples de forma combinada para obtener un mejor rendimiento predictivo que el que se podría obtener de cualquiera de los algoritmos de aprendizaje de forma individual. La evaluación de la predicción de un conjunto de algoritmos generalmente requiere más computación que la evaluación de la predicción por uno solo, por lo que los conjuntos pueden considerarse como una forma de compensar los algoritmos de aprendizaje pobres mediante la realización de una gran cantidad de cálculos adicionales. Los algoritmos rápidos, como los árboles de decisión, se usan comúnmente en los métodos de conjunto, aunque los algoritmos más lentos también pueden beneficiarse de las técnicas de conjunto.

1. **Aprendizaje no supervisado**

La característica principal del método de aprendizaje no supervisado consiste en que el aprendizaje se ajusta y se consigue a través de las meras observaciones, sin necesidad de tener que realizar entrenamientos con datos.

Así, se diferencia del aprendizaje supervisado en que en este caso no existe un conocimiento a priori, sino que se debe aprender directamente de lo que se observa. De esta forma, el aprendizaje no supervisado generalmente trata los datos de objetos de entrada como un conjunto de variables aleatorias, construyendo así un modelo de densidad para el conjunto de datos.

Dentro de los principales métodos de aprendizaje no supervisado nos encontramos con los siguientes:

* Algoritmos Clustering (*Clustering* *Algorithms*): el término clustering hace referencia a la agrupación de un conjunto de elementos de una determinada manera, haciendo que los elementos que se encuentren en un mismo grupo (denominado cluster) sean más similares entre sí (en algún sentido) que aquellos elementos que se encuentran en un grupo distinto. El análisis de los elementos en sí mismo no es un algoritmo específico, sino más bien la tarea general a resolver. Se puede lograr mediante varios algoritmos que difieren significativamente de lo que constituye un cluster.

* Análisis de componentes principales (*Principal* *Component* *Analysis o*PCA): mediante esta técnica se describe un grupo de datos en términos de nuevas variables (componentes) no correlacionadas. Estos componentes se ordenan por la cantidad de varianza original que describen, por lo que la técnica resulta de mucha utilidad para reducir la dimensionalidad de un conjunto de datos. El PCA se emplea sobre todo en análisis explorativo de datos y para construir modelos predictivos. Esta técnica comporta el cálculo de la descomposición en autovalores de la matriz de covarianza, generalmente tras centrar los datos en la media de cada atributo.

* Descomposición en valores singulares (*Singular* *Value* *Decomposition o*SVD): la SVD de una matriz real o compleja es una factorización de la misma con muchas aplicaciones en los campos de la estadística, el procesamiento de señales y otras disciplinas.
* Análisis de componentes independientes (*Independent* *Component* *Analysis o*ICA): consiste en un método computacional que sirve para separar una señal multivariante en sucomponentes aditivos suponiendo que la señal de origen tiene una independencia estadística y es no-Gausiana. Este es un caso especial de «separación ciega de las señales». El ICA es una generalización del análisis de componentes principales (PCA), ya que en ambos casos se practica una transformación lineal de los datos originales, aunque la diferencia básica es que el ICA no requiere que las variables originales tengan una distribución gausiana.

1. **Aprendizaje por refuerzo**

El aprendizaje por refuerzo es un sector que se ocupa del aprendizaje automático orientado hacia la resolución de unos determinados problemas de decisión secuenciales. Generalmente versa sobre problemas relacionados con aplicaciones pertenecientes a campos muy diversos como pueden ser el control automático, la medicina, la investigación operativa o la economía.

Los algoritmos clásicos de este tipo de aprendizaje automático están basados en la teoría matemática de la programación dinámica, donde se asume que el espacio de estados es discreto y que está compuesto de un número manejable de estados.

Sin embargo, en la mayor parte de aplicaciones de interés práctico, el espacio entre estados es continuo, lo que hace que dichos algoritmos clásicos pierdan utilidad. Para poder aplicar esta técnica a espacios continuos se deben dar dos requisitos: generalizar el comportamiento aprendido a partir de un conjunto limitado de experiencias a casos que no hayan sido experimentados anteriormente y representar las políticas de forma compacta.

La combinación de algoritmos de aprendizaje por refuerzo con técnicas de aproximación de funciones es actualmente un área de investigación activa. No obstante, y a pesar de los avances logrados en los últimos años en este campo, todavía existen aspectos que limitan la capacidad del aprendizaje por refuerzo en problemas complejos.

* **Programación dinámica (*Dynamic programming*)**: consiste en una técnica para reducir el tiempo de ejecución de un algoritmo mediante la utilización de subproblemas superpuestos y subestructuras óptimas.
* ***Q-learning***: el objetivo de este método es aprender una política que le indique a un agente qué acción tomar en qué circunstancias. No requiere un modelo del entorno y puede manejar problemas con transiciones estocásticas y recompensas, sin requerir de adaptaciones. Para cualquier proceso de decisión finito de Markov (FMDP), Q-learning finalmente encuentra una política óptima, en el sentido de que el valor esperado de la recompensa total sobre todos los pasos sucesivos, comenzando desde el estado actual, es el máximo alcanzable.

* ***State-Action-Reward-State-Action* (SARSA)**: se trata de un algoritmo empleado para aprender una política de proceso de decisión de Markov. El nombre de SARSA viene dado por cuanto refleja el hecho de que la función principal para actualizar el valor de un elemento (Q) depende del estado actual del agente (S1), la acción que el agente elige (A1), la recompensa que el agente obtiene para elegir dicha acción (R), el estado que ingresa el agente después de tomar esa acción (S2) y, finalmente, la siguiente acción que elige el agente en su nuevo estado (A2).

## Deep learning

El *deep learning* forma parte del *machine learning*, es decir, que el aprendizaje profundo es una rama concreta que se engloba dentro del aprendizaje automático. El aprendizaje profundo se desarrolla a través de modelos informáticos que funcionan de forma similar al cerebro humano, consistente en un sistema de redes artificiales de neuronas que se encargan de analizar los diferentes datos.

En este caso, la máquina o computadora en cuestión evalúa distintos ejemplos e instrucciones para modificar el modelo en caso de que se produzcan errores. De esta forma, el sistema reconoce y aplica patrones que facilitan la solución de problemas de manera muy precisa, es decir, que toma decisiones a partir de los distintos datos que recoge.

Como decimos, el aprendizaje profundo es un subconjunto del aprendizaje automático. Las redes neuronales artificiales constituyen un conjunto de algoritmos que alcanzan nuevos niveles de precisión que antes no existían y que dan solución a una serie de problemas importantes, como el reconocimiento de imágenes, el reconocimiento de sonidos o los sistemas de recomendación.

En otras palabras, este método utiliza algunas técnicas de aprendizaje automático para resolver problemas del mundo real recurriendo a redes neuronales que simulan la toma de decisiones humanas.

El aprendizaje profundo puede ser costoso y requiere de grandes conjuntos de datos para capacitarse. Esto se debe a que hay un gran número de parámetros que deben ser comprendidos por un algoritmo de aprendizaje, que puede producir principalmente muchos falsos positivos.

Por ejemplo, un algoritmo de *deep learning* podría ser entrenado para aprender cómo se ve un perro. Tomaría un enorme conjunto de datos de imágenes para entender los pequeños detalles que distinguen a un perro de un lobo o un zorro, por ejemplo.

Los principales métodos de aprendizaje profundo mediante los que se desarrollan los algoritmos son:

**Redes neuronales profundas (*Deep Neural Networks o*DNN)**: son un tipo de red neuronal artificial (ANN) que consiste en un modelo computacional basado en un gran conjunto de unidades neuronales simples, de forma similar al comportamiento observado en los axones de las neuronas en los cerebros biológicos.

Cada una de las neuronas artificiales está conectada con otras muchas, creando una red neuronal y unos enlaces que pueden incrementar o inhibir el estado de activación de las neuronas adyacentes, ya que cada neurona artificial puede transmitir información a otra.

De esta forma, estos sistemas aprenden y se forman a sí mismos, sin necesidad de ser programados, y sobresalen en áreas donde la detección de soluciones o características es difícil de expresar con la programación convencional.

Una DNN, además, es capaz de encontrar la manipulación matemática correcta para convertir la entrada (input) en la salida (output), ya sea a través de una relación lineal o no.

La red se mueve a través de las capas de entrada y salida calculando la probabilidad de cada salida. Por ejemplo, un DNN que está entrenando para reconocer razas de perros revisará la imagen dada y calculará la probabilidad de que el perro en la imagen sea de cierta raza.

El usuario puede revisar los resultados y seleccionar qué probabilidades debería mostrar la red y devolver la etiqueta propuesta. Cada manipulación matemática como tal se considera una capa, y las redes neuronales profundas tienen muchas capas, de ahí el nombre de redes “profundas”.

* + 1. **Bosque decisión neuronal profunda** **(*Deep Neural Decision Forest*)**: este sistema presenta un enfoque novedoso que combina la técnica del árbol de decisión (Decision tree) con la funcionalidad de aprendizaje de representación conocida a partir de redes neuronales profundas. Para combinar ambos, se crea un modelo de árbol de decisión estocástica y diferenciable que dirige el aprendizaje de la representación generalmente conducido en las capas iniciales de una red neuronal profunda. Este sistema difiere de las redes neuronales profundas convencionales porque proporciona las predicciones finales y asimismo difiere del sistema de los árboles de decisión convencionales porque pretende una optimización conjunta, global y basada en principios de los parámetros de nodo dividido y hoja.

En resumen, el estudio e investigación de la inteligencia artificial, en efecto, se centra en muy diversos campos, entre los que se encuentran el machine learning y el deep learning, siendo además este último una nueva área de investigación del aprendizaje automático.

Por lo tanto, y de forma general, se podría decir que la inteligencia artificial es la tecnología más “básica”, ya que siempre responde igual ante unos mismos parámetros, que ha evolucionado en el machine learning, que es capaz de autoaprender y corregir errores, y en deep learning (la más compleja de las tres) que, además de eso, toma decisiones a partir de los diferentes datos.

Métodos de predicción:

En los métodos de predicción, la predicción es una actividad esencial cuya importancia ha crecido en las últimas décadas, convirtiéndose en imprescindible en la gran mayoría de los procesos de toma de decisiones. La predicción en términos generales es conocer por anticipado algún suceso, basándose en el análisis y consideraciones necesarias dependiendo del campo de aplicación.

El avance de la tecnología ha permitido recopilar información de diferentes fenómenos a lo largo del tiempo, a esta información se le llama serie de tiempo. Una serie tiempo o cronológica es una secuencia de datos, observaciones o valores medidos en determinados momentos del tiempo, espaciados entre sı de manera uniforme y por tanto ordenados cronológicamente. Uno de los usos más habituales de las series de datos temporales es su análisis para predicción y pronostico. El análisis de series temporales comprende métodos que ayudan a interpretar este tipo de datos, extrayendo información representativa, tanto referente a los orígenes o relaciones subyacentes como a la posibilidad de extrapolar y predecir su comportamiento futuro. Planificadores y decisores disponen de una gran variedad de técnicas para realizar sus predicciones, que van desde las más subjetivas e intuitivas hasta métodos cuantitativos más complejos.

# 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.

## Técnicas instrumentos de recolección de datos

### Procedimientos

Se solicitó información a la consultora RUNA TECNOLOGY.

### Análisis de datos

El análsisis de datos se realizará con métodos de regresión de Machine Learning

#### Procedimiento de recolección de datos

## Presupuesto de la investigación

### Fuentes de financiamiento

El equipo se financió con recursos propios

## Población y muestra

La población tiene como objetivo las aguas del río Chillón, a efectos de complementar el estudio de la consultora Runa Tecnologies.

## Operacionalización de variables

### Calificación del impacto ambiental

Se realizará el cálculo del impacto ambiental a efectos de justificar la investigación.

Contenido de un Estudio Legal de Impacto Ambiental

CAI= Ca x Ro x (I + E + Du + De + Re) /5 )

|  |  |  |
| --- | --- | --- |
| Parámetros | Clasificación | Acción |
| Ca Carácter +1, -1 | 1 | La acción es benéfica |
| Ro Riesgo de Ocurrencia 1…10 | 8 | Muy probable que se dé el impacto |
| I Intensidad 0,1...1 | 1 | muy alta |
| E Extensión 0,1…1 | 1 | Amplia porque abarca todo el Perú |
| Du Duración 0,1…1 | 1 | permanente |
| De Desarrollo 0,1…1 | 0.3 | lento |
| Re Reversibilidad 0,1…1 | 0.7 | parcialmente reversible |

La gestión ambiental de los recursos hídricos genera un impacto ambiental positivo de 6.4

## Recolección de datos

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

Parámetros:

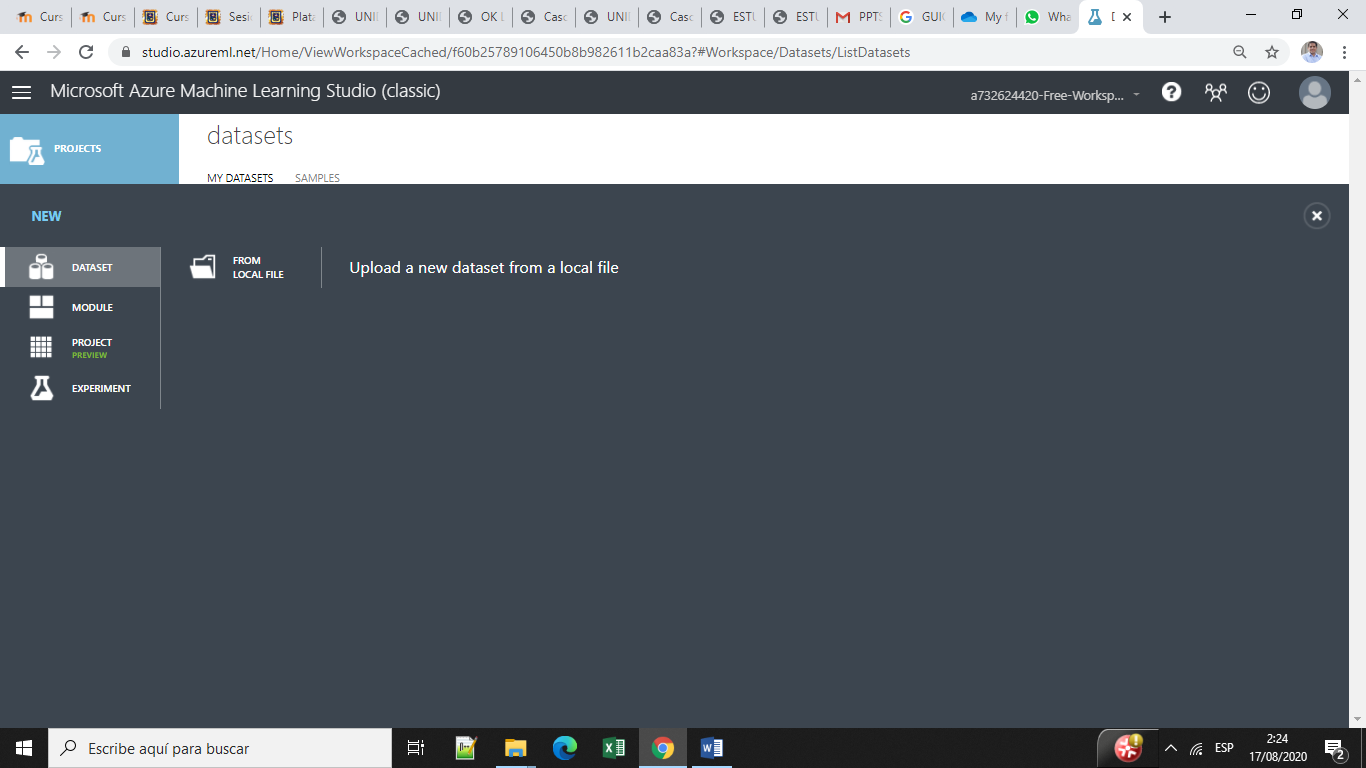
|  |  |  |  |
| --- | --- | --- | --- |
| **Símbología** | **Parámetro** | **Unidades** | **Factor de Ponderación** |
| **OD** | Oxígeno Disuelto | % Sat | 0.17 |
| **C.term** | Coliformes termotolerantes | NMP/100mL | 0.16 |
| **Ph** | Indice pH | unidad | 0.11 |
| **DBO** | Demanda Bioquímica de Oxígeno | 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.

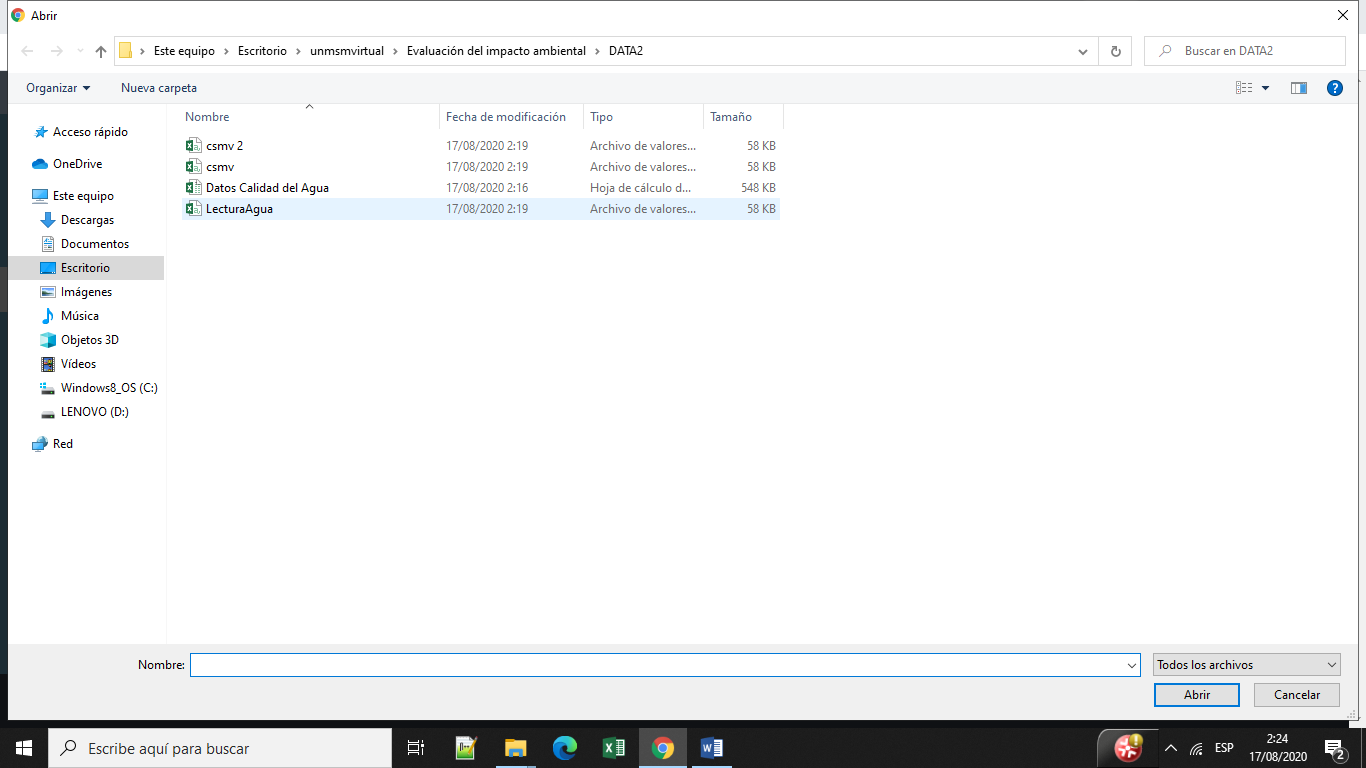
El parámetro WQI se correlaciona con siete parámetros, a saber, temperatura, turbidez, pH, dureza como CaCO3, conductancia, sólidos disueltos totales y recuento de coliformes fecales. Tenemos que elegir lo mínimo número de parámetros para predecir el WQI, con el fin de reducir el costo del sistema. Los 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 las heces recuento de coliformes, lo que significa que seleccionar TDS nos permitiría descartar los otros dos parámetros.

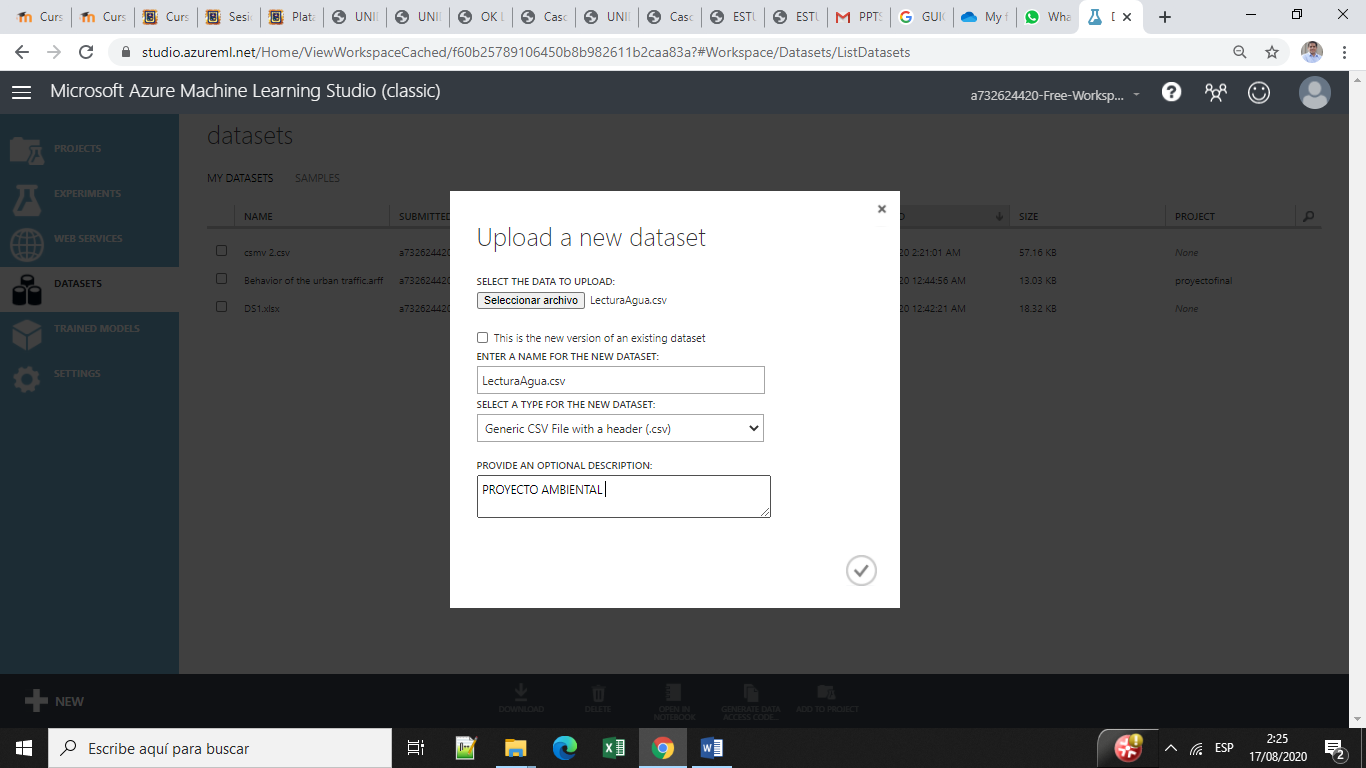
Dejamos el parámetro inconveniente restante, la dureza como CaCO3, porque no es muy correlacionado comparativamente y no es fácil de adquirir.

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

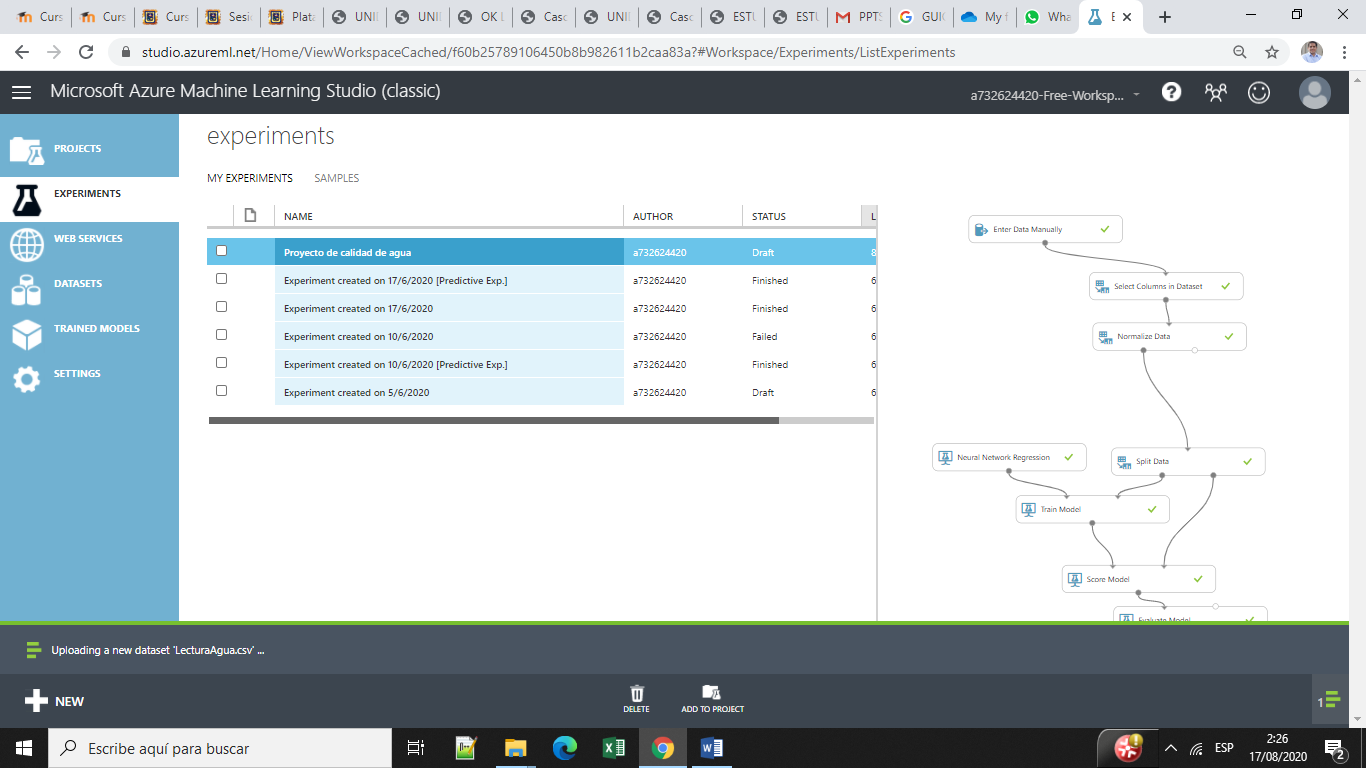


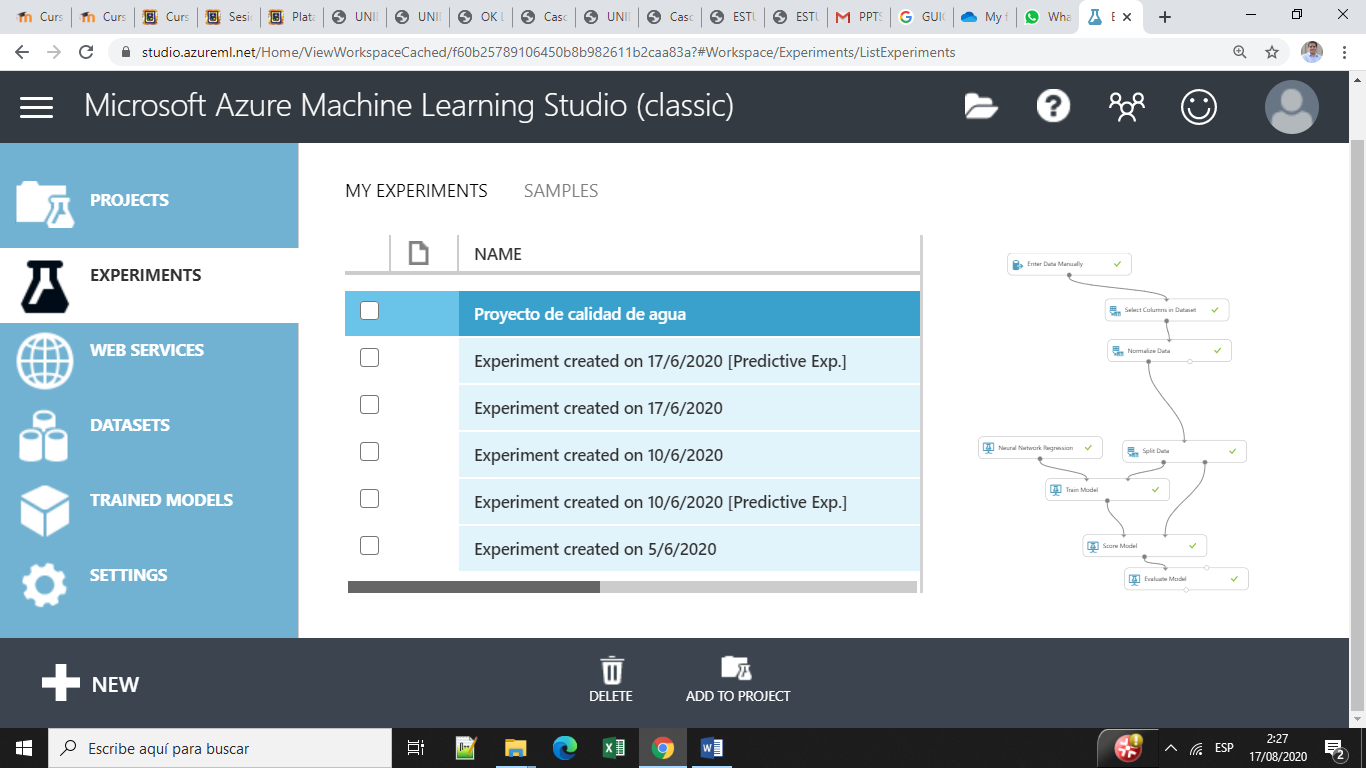
Se seleccionan las lecturas de aguas muestreadas

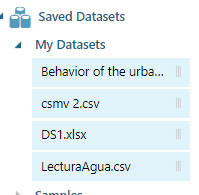


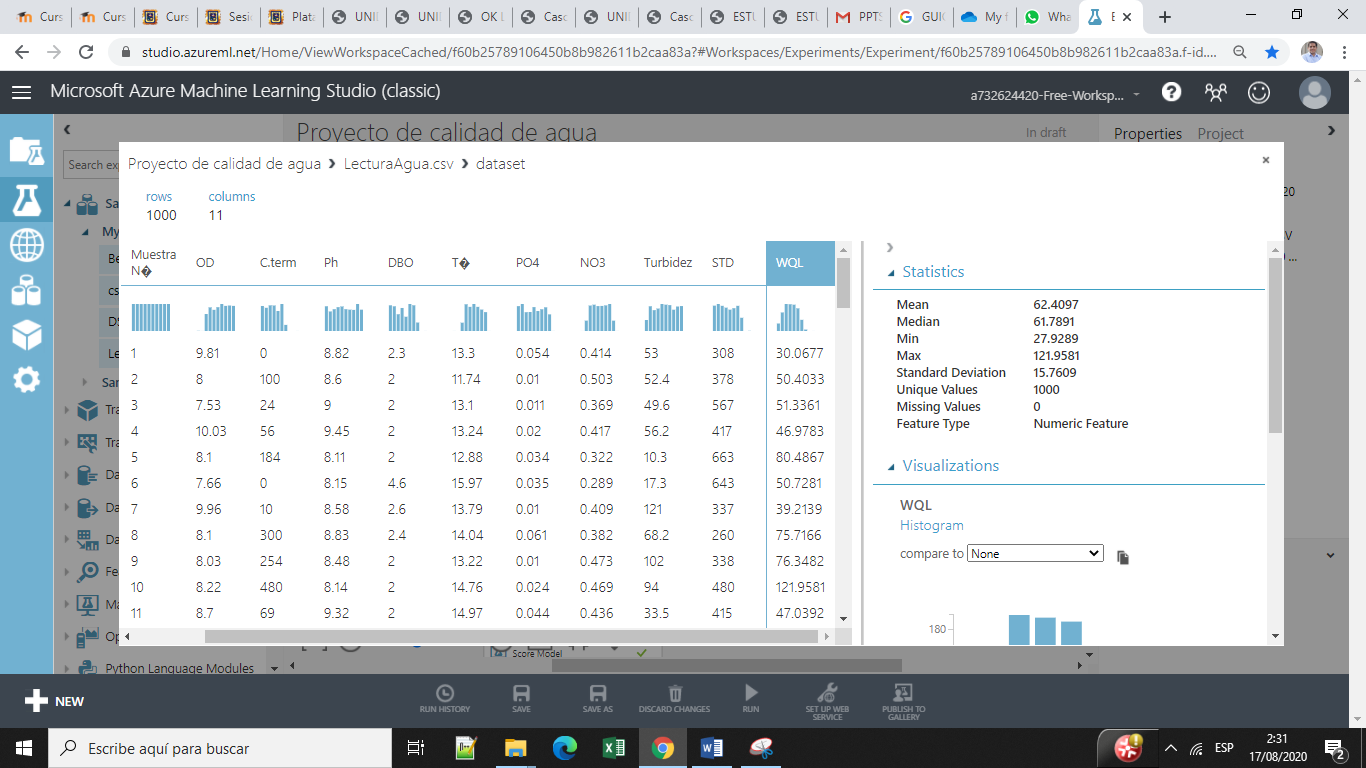


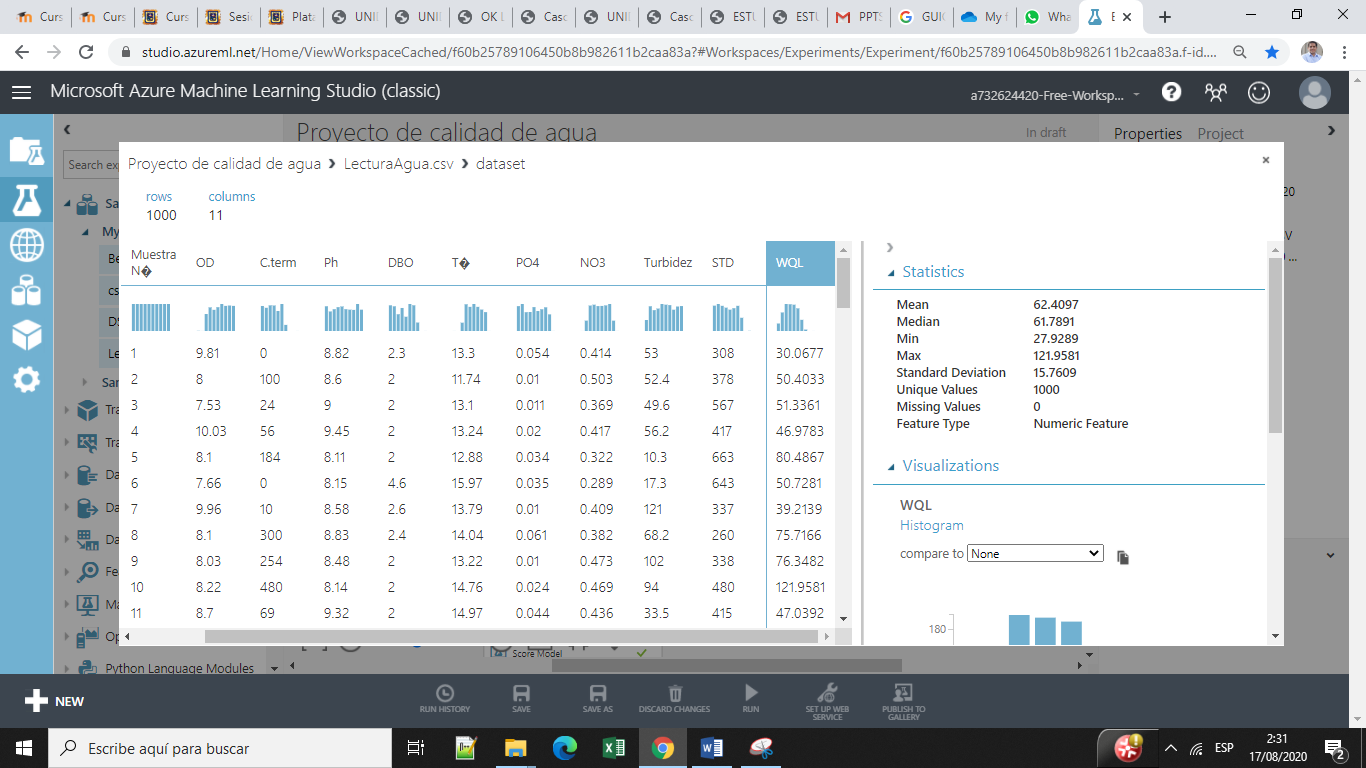
Posteriormente, vamos a la pestaña de experimentos











3.2. 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.

**2) 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:

**3) Logistic**

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

**4) LogNormal**

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

**)**

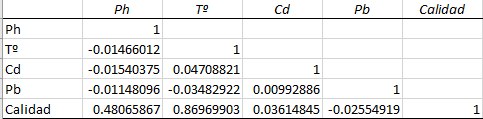
* Aquí μ y σ son los parámetros de la distribución, calculados empíricamente a partir de los datos como estimaciones de máxima verosimilitud, para cada columna por separado.

**5) TanH**

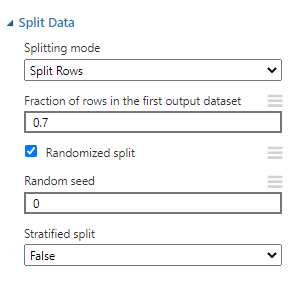
* Todos los valores se convierten en una tangente hiperbólica.
* Los valores en la columna se transforman utilizando la siguiente fórmula:

## Análisis de Correlación

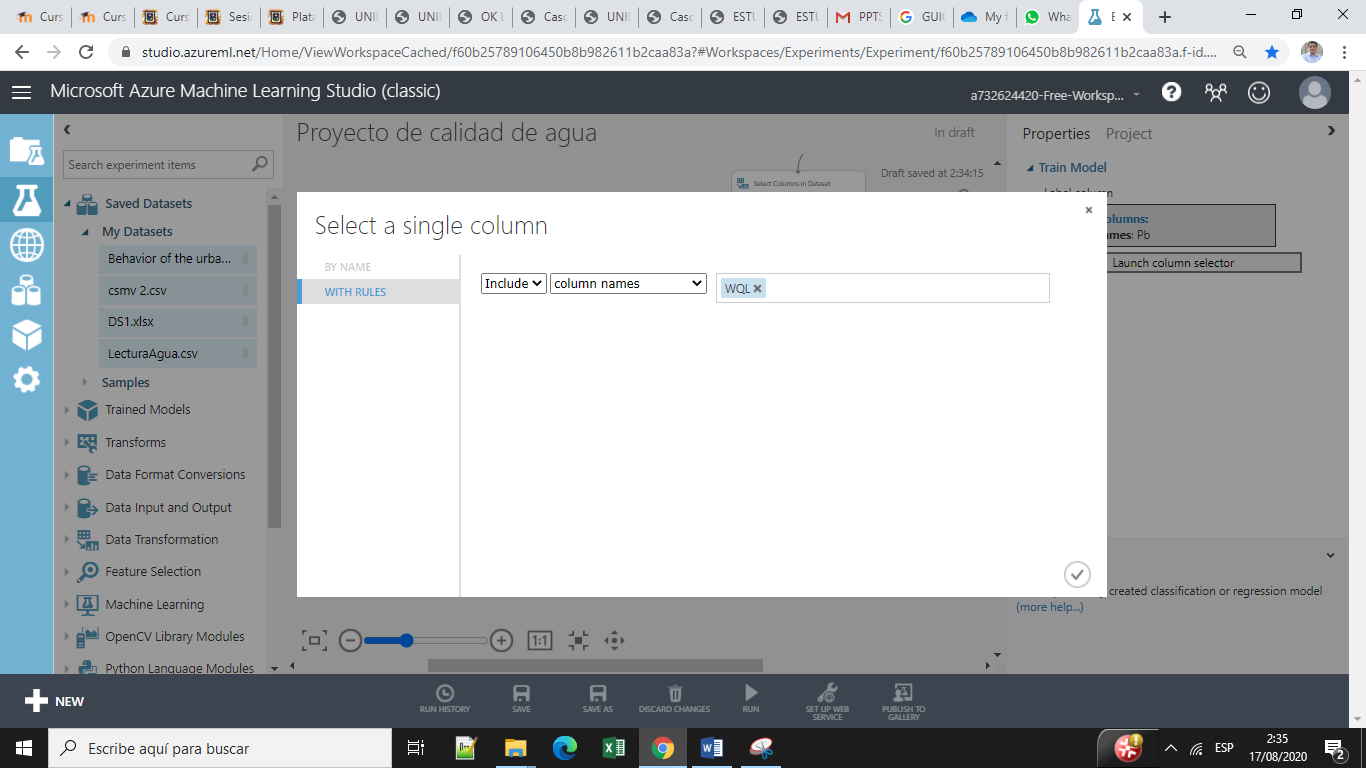
Se realiza el análisis de correlación de las variables de las lecturas



Data Splitting–Cross Validation



Se selecciona al índice WQL como label column del experimento.



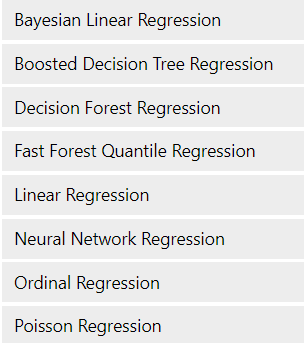
## Fase de entrenamiento

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



## Fase de pruebas

## Resultados de algoritmos de regresión

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 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 |

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.
* 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.
* 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.

# Referencias bibliográficas

* 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/
* 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
* 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/boletin\_agua\_y\_saneamiento.pdf
* *PLAANA. (2011). Plan nacional de acción ambiental. Diario Oficial El Peruano, p. 80.*
* 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

# Anexos

## Matriz de consistencia

