



UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA MECÁNICA ELÉCTRICA,
ELECTRÓNICA Y SISTEMAS
ESCUELA PROFESIONAL DE INGENIERÍA ELECTRÓNICA



REGRESIÓN MULTIPARAMÉTRICA PARA UN SISTEMA DE
MONITORIZACIÓN DE RADIACIÓN UV MEDIANTE IOT EN LA
CIUDAD DE PUNO

TESIS

PRESENTADA POR:

Bach. YAMIR GONZALO COTRINA QUISPE

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ELECTRÓNICO

PUNO – PERÚ

2022



DEDICATORIA

Dedicado en especial a mi familia, a mis padres Milka y Rafael, mis hermanos Álvaro y Shande, quienes han creído en mí dándome ejemplos de superación, humildad y sacrificio. Por la confianza y todo lo que me han dado en mi vida, especialmente por sus sabios consejos y a mi esposa Alejandra por estar a mi lado como guía y apoyo en todo este proceso y a las personas cercanas que me brindaron el apoyo incondicional.

Yamir Cotrina



AGRADECIMIENTOS

A Dios por cuidarme en todo este tiempo y haberme permitido llegar a este momento de mi vida. A la Universidad Nacional del Altiplano por haberme brindado la oportunidad de formarme como futuro profesional dentro de las aulas. Así también a los docentes de la Escuela Profesional de Ingeniería Electrónica por todos los conocimientos, experiencias y enseñanzas que contribuyen a mi formación profesional. A mis amigos y compañeros por su constante apoyo.

Yamir Cotrina



ÍNDICE GENERAL

DEDICATORIA

AGRADECIMIENTOS

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

ÍNDICE DE ACRÓNIMOS

RESUMEN 11

ABSTRACT..... 12

CAPÍTULO I

INTRODUCCIÓN

1.1 OBJETIVO GENERAL 14

1.2 OBJETIVOS ESPECÍFICOS 14

CAPÍTULO II

REVISIÓN DE LA LITERATURA

2.1 ANTECEDENTES DEL PROYECTO 15

2.1.1 A Nivel Local 15

2.1.2 A Nivel Nacional 16

2.1.3 A Nivel Internacional 18

2.2 MARCO TEÓRICO 19

2.2.1 Internet de las cosas (IOT)..... 19



2.2.1.1 Tendencias tecnológicas y de mercado que están impulsando IoT	21
2.2.1.2 Características principales de internet de las cosas.....	22
2.2.1.3 Arquitectura global para internet de las cosas	23
2.2.2 Introducción a la regresión multivariable.....	40
2.2.2.1 Análisis de regresión.....	41

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 UBICACIÓN GEOGRÁFICA DEL ESTUDIO.....	54
3.2 PERIODO DE DURACIÓN DEL ESTUDIO	54
3.3 PROCEDENCIA DEL MATERIAL UTILIZADO.....	55
3.4 POBLACIÓN Y MUESTRA DEL ESTUDIO	55
3.5 DISEÑO ESTADÍSTICO	56
3.5.1 Diseño de la investigación.....	56
3.5.2 Nivel de la investigación	57
3.6 PROCEDIMIENTO.....	57
3.6.1 Descripción del hardware y software del prototipo.....	57

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1 RESULTADOS DEL DISEÑO E IMPLEMENTACIÓN DEL PROTOTIPO DE RECOLECCIÓN DE DATOS MEDIANTE UN SISTEMA IOT DE LA RADIACIÓN UV EN LA CIUDAD DE PUNO	65
---	-----------



4.2	RESULTADOS PARA EL MODELO DE REGRESIÓN MULTÍPARAMÉTRICA PARA LOS DATOS RECOLECTADOS MEDIANTE EL SISTEMA IOT EN LA CIUDAD DE PUNO	67
V.	CONCLUSIONES.....	75
VI.	RECOMENDACIONES	76
VII.	REFERENCIAS BIBLIOGRÁFICAS.....	77
	ANEXOS.....	84

Área: Telecomunicaciones

Tema: Comunicación IoT

Fecha de sustentación: 26 de enero del 2022



ÍNDICE DE FIGURAS

Figura N° 1: Representación de internet de IoT	20
Figura N° 2: Arquitectura global de IoT	23
Figura N° 3: Adquisición de datos	24
Figura N° 4: Características de termocuplas.....	25
Figura N° 5: Detector infrarrojo	26
Figura N° 6: Actuadores hidráulicos.....	26
Figura N° 7: Actuadores neumáticos	27
Figura N° 8: Actuadores eléctricos	28
Figura N° 9: Nivel de Comunicaciones	30
Figura N° 10: Relación cobertura ancho de banda para Low power short Range	31
Figura N° 11: Nivel de Procesamiento.	33
Figura N° 12: Nivel de aplicaciones.	39
Figura N° 13: Modelo de regresión lineal.....	43
Figura N° 14: Ciudad de Puno.	54
Figura N° 15: Módulo utilizado.....	56
Figura N° 16: Sensor GYML8511 y su diagrama interno.	58
Figura N° 17: Arduino UNO y Ethernet Shield W5100.....	58
Figura N° 18: Interfaz de programación IDE.	60
Figura N° 19: Tecnología LTE utilizada.	61
Figura N° 20: Gestión de base de datos.	62



Figura N° 21: Valores obtenidos de medición de la radiación UV.....	63
Figura N° 22: Intensidad UV promedio.....	64
Figura N° 23: Interfaz gráfica de usuario.	64
Figura N° 24: Monitoreo de radiación UV.	65
Figura N° 25: Precisión de los modelos realizados.	68
Figura N° 26: Error absoluto medio de los modelos realizados.	69
Figura N° 27: Error cuadrático medio de los modelos realizados.	70
Figura N° 28: Coeficiente de determinación de los modelos realizados.	71
Figura N° 29: Coeficiente de determinación ajustado de los modelos realizados.	72
Figura N° 30: Tiempo de entrenamiento de los modelos realizados.	73
Figura N° 31: Tiempo de test o prueba de los modelos realizados.....	74



ÍNDICE DE TABLAS

Tabla N° 1: Duración del estudio.....	54
Tabla N° 2: Lista de materiales.....	55
Tabla N° 3: Materiales utilizados para el módulo de adquisición de datos.	55
Tabla N° 4: Data obtenida mediante el monitoreo de radiación UV.	66
Tabla N° 5: Parámetros obtenidos de los tres modelos de regresión multiparamétrica.	67



ÍNDICE DE ACRÓNIMOS

ANSI	(American National Standards Institute)
BW	(Bandwidth)
GPL	(General Public License)
I ² C	(Inter Integrated Circuit)
IDE	(Integrated Development Environment)
IoT	(Internet of Things)
IP	(Internet Protocol)
ISM	(Instrumentation Scientific and Medical)
LoRa	(Long Range)
LoRaWan	(Low-power wide-area network)
LPWAN	(Low Power Wide Area Network)
LTE	(Long Term Evolution)
SSM	(Spread Spectrum)
SENAMHI	(Servicio Nacional de Meteorología e Hidrología del Perú)
SPI	(Serial Peripheral Interface)
UV	(Ultraviolet Radiation)
WSN	(Wireless Sensor Network)
WAN	(Wide Area Network)
OLS	(Ordinary Least Squares)



RESUMEN

El presente trabajo de tesis se desarrolló desde junio a noviembre del 2021 en la ciudad de Puno, tiene como objetivo realizar un modelo de regresión multiparamétrica para un sistema de monitorización de radiación UV mediante IoT empleando tres modelos de regresión teniendo como problemática que a nivel mundial el cáncer de piel es el más común entre todos los tipos de cáncer siendo el melanoma el 1% de los casos de cáncer de piel, pero causante de la mayoría de las muertes por este tipo de cáncer. Una manera de evitar esta enfermedad, es que las personas eviten la exposición de radiación UV, entonces la realización de un sistema de monitorización ayudará a evitar la exposición excesiva. Para realizar este trabajo se utilizaron métodos experimentales y para ello se empleó tres modelos de regresión, tomando tres puntos de medición como lugar de control y obtención de datos en un periodo de tres meses desde agosto a octubre. Primero se realizó un sistema de recolección de datos basado en IoT, para que con los mismos, realizar tres modelos de regresión multiparamétrico, para determinar con mayor exactitud la variable a determinar partiendo de las otras variables que se encuentran en el sistema de pronóstico que son: OLS; Ridge y Bayesian Ridge. Habiéndose realizado los tres modelos de regresión multiparamétrica OLS, Ridge y Bayesian Ridge para la radiación UV en la ciudad de Puno podemos mencionar que teniendo como resultados más resaltantes un error cuadrático medio para OLS de 0.02288, para Ridge de 0.0230 y para Bayesian Ridge 0.02288. El coeficiente de determinación ajustado para OLS de 0.9997, para Ridge de 0.9997 y para Bayesian Ridge de 0.997. Siendo como el mejor modelo de regresión el Bayesian Ridge.

Palabras Clave: radiación UV, sistema de monitoreo IOT, regresión multiparamétrica, OLS, RIDGE, BAYESIAN RIDGE.



ABSTRACT

The present thesis work was developed from June to November 2021 in Puno city. Its objective is to carry out a multiparametric regression model for a UV radiation monitoring system using IoT with three regression models having as a problem that skin cancer is the most common worldwide problem of all types of cancer. Melanoma being only 1% of skin cancer cases, but it causes the vast majority of deaths from this type of cancer. One way to avoid this disease is for people not to expose themselves, as much as possible, to UV radiation exposure. So carrying out a monitoring system will help avoid excessive exposure. To carry out this work, experimental methods were used and for this, three regression models were used, taking three measurement points as a place of control and obtaining data in a period of three months from August to October. First, a data collection system based on IoT was carried out, so that with them, performing three multiparametric regression models, to determine with greater accuracy the variable to be determined starting from the other variables that are in the forecasting system that are : OLS; Ridge and Bayesian Ridge. Having carried out the three multiparametric regression models OLS, Ridge and Bayesian Ridge for UV radiation in the city of Puno, we can mention that resulting in more outstanding results a mean square error for OLS of 0.02288, for Ridge of 0.0230 and for Bayesian Ridge 0.02288 . The coefficient of determination adjusted for OLS of 0.9997, for Ridge of 0.9997 and for Bayesian Ridge of 0.997. The Bayesian Ridge being the best regression model.

Keywords: UV radiation, IOT monitoring system, multiparametric regression, OLS, RIDGE, BAYESIAN RIDGE.



CAPÍTULO I

INTRODUCCIÓN

A nivel mundial se tiene el problema de que el cáncer de piel es por mucho, el más común entre todos los tipos de cáncer. El melanoma conforma solo el 1% de los casos de cáncer de piel, pero causa la gran mayoría de las muertes por este tipo de cáncer. Los melanomas se desarrollan sobre todo en personas de piel y ojos claros con dificultad para broncearse y que han sufrido quemaduras solares, fundamentalmente en la infancia o adolescencia. La exposición puntual, excesiva e intermitente al sol se relaciona con su aparición, lo que explica que los melanomas se localicen preferentemente en zonas no expuestas al sol de forma habitual, como es el caso de la espalda y las piernas. Además de la acción de las radiaciones solares, el riesgo que una persona tiene de padecer un melanoma aumenta con: la existencia de antecedentes familiares o personales de este tumor, la presencia en la piel de gran número de lunares (más de 50 o 60), el padecimiento de quemaduras solares con ampollas durante la infancia y adolescencia y la edad avanzada. Por estas razones se utilizaron tres modelos de regresión multiparamétrico para un sistema de monitorización de radiación UV mediante IoT en la ciudad de Puno. Para ello se realizó un sistema de recolección de datos basado en IoT, para que con los mismos, realizar una comparación de estos modelos y determinar con mayor exactitud el pronóstico de esta variable de radiación UV, a partir de las otras variables que se encuentran en el sistema. El estudio se realizó en la ciudad de Puno, ciudad que tiene características especiales respecto a los niveles de radiación: altitud, distancia del Ecuador, entre otros.



Es por ello que con el presente proyecto de tesis se pretende contribuir al cuidado de la salud de los pobladores de la ciudad de Puno y también ser una antecedente teórico para futuras investigaciones de estudiantes de la escuela profesional de ingeniería electrónica y quieran abordar temas y ampliar sus conocimientos en sistemas de monitorización empleando IoT. Este sistema de pronóstico será de utilidad para pronosticar la radiación UV y así evitar la exposición innecesaria a este tipo de radiación, pues la incidencia de cáncer de piel debido a la radiación UV está presente en todo el Perú. El desarrollo de esta investigación está estructurado de la siguiente forma:

En el capítulo I se desarrollará la introducción, el objetivo general y los objetivos específicos, en el capítulo II se incluye la revisión de literatura de la investigación, en el capítulo III se establece la metodología, diseño de la investigación, materiales e instrumentos utilizados para el presente proyecto, mientras que en el capítulo IV se muestran los resultados obtenidos por las mediciones y que servirán para el análisis de los datos; para luego finalizar con las conclusiones y recomendaciones del proyecto de investigación.

1.1 OBJETIVO GENERAL

- Realizar un modelo de regresión multiparamétrica para un sistema de monitorización de radiación UV mediante IoT en la ciudad de Puno.

1.2 OBJETIVOS ESPECÍFICOS

- Recolectar los datos mediante un sistema IoT de la radiación UV en la ciudad de Puno.
- Realizar un modelo de regresión multiparamétrica para los datos recolectados mediante el sistema IoT en la ciudad de Puno.



CAPÍTULO II

REVISIÓN DE LA LITERATURA

2.1 ANTECEDENTES DEL PROYECTO

2.1.1 A Nivel Local

En la localidad de UPINA, centro poblado perteneciente a la provincia de Carabaya del departamento de Puno se realizaron mediciones monitoreadas por el SENAMHI, para luego caracterizar la radiación ultravioleta y construir un modelo matemático de cálculo del índice ultravioleta en dicha localidad. El análisis de la radiación ultravioleta tipo A (UVA), dio por resultado que el promedio acumulado de intensidad mensual es 22.04 W/m² registrada mensualmente presentando un nivel de dispersión 1.7 indicando que el promedio de radiación ultravioleta no varía con gran intensidad durante todo el año, pero en comparación de los valores máximos registrados cada mes se encontró un incremento del 64% en cuanto al primera valor máximo registrado el mes de Mayo, y al analizar la radiación ultravioleta tipo B, se observa un comportamiento similar a la radiación tipo A en cuanto al promedio acumulado de radiación , encontrándose un promedio de intensidad 2.15 W/m² con un nivel de dispersión de 0.12 y pero en cuanto al análisis de valores máximos es muy diferente donde se observa un incremento de del 35.7 % respecto al valor máximo registrado en Mayo . Para la formulación del modelo matemático se partió de simples ecuaciones, evaluando los datos calculados por el modelo y los datos obtenidos por la estación meteorológica se obtuvo una correlación promedio de 0.96 considerando aceptable el modelo (Huaraya, 2018).



Así también la investigación propone el diseño de una red FOG basado en internet de las cosas IoT, para monitorear la contaminación en la bahía del Lago Titicaca, monitoreo de las variables ambientales: temperatura, acidez y algunos parámetros de contaminación del Lago Titicaca, desde un abordaje enfocado al concepto de la “Internet de las Cosas”, Computación en la Nube, y Redes de Sensores Inalámbricos (WSN). La red consiste en unir de manera remota sensores en la bahía del lago mediante radio enlaces todas fijadas en el Lago con capacidad de registrar eventos de contaminación en las aguas del Lago Titicaca, haciendo el correcto uso de las Tecnologías de la información y la comunicación (TIC), la red permitirá reducir tiempo de análisis, optimizar costos, mejorar la calidad de los servicios de obtención de datos en tiempo real , disminuir riesgos y ampliar la cobertura de monitoreo del Lago Titicaca para mitigar efectos de contaminación (Ttacca, 2017).

2.1.2 A Nivel Nacional

Actualmente los requerimientos para el despliegue del internet de las cosas sobrepasan los modelos de redes tradicionales tales como Wi-Fi, Bluetooth, etc, por ello están surgiendo nuevas tecnologías que son más acore a las necesidades que se presenta. En la irrigación Majes – Sigvas I, el sistema de riego que se tiene es en su mayoría es de forma manual, quiere decir que los agricultores y/o microempresarios de la zona recorren el campo abriendo y cerrando válvulas para cumplir con la tarea, sin tener valores de humedad y temperatura de sus cultivos. En el sistema que se desarrolló se toma como referencia el 60% de humedad, quiere decir que, si el cliente recopila un valor menor a 60% de humedad, el actuador (electro válvula) tiende a encenderse y si en caso opuesto la humedad sea mayor del 60% el actuador tiende a apagarse. En el proyecto no se estudian las propiedades de los cultivos ya que no es nuestra prioridad. Los sensores instalados en los módulos cliente realizan la recopilación de datos de humedad y



temperatura en diferentes horarios y diferentes puntos del campo experimental para su posterior envío a través de LORA por el modulo cliente hacia el modulo servidor, una vez que los datos se encuentren en el servidor éste lo procesa y toma la decisión de abrir o cerrar una electroválvula para el respectivo riego del cultivo dependiendo de la humedad (Robles & Camayo, 2019).

En la ciudad de Arequipa se realizó un estudio comparativo y modelación de la radiación solar global y ultravioleta, trabajando con datos de radiación solar global y ultravioleta UV datos que fueron proporcionados el Servicio Nacional de Meteorología e Hidrología (SENAMHI) registrados en la estación “la Pampilla”. Se estudió y analizó el comportamiento de la radiación UV, y radiación solar global para diferentes días del año 2013 donde se encontró que existe un comportamiento de curvas similar y de proporción directa entre estas variables, así se obtuvo que entre las 12 y 13 horas se obtienen valores de 9 a 13 índice de radiación ultravioleta y 700 a 1000 W/m² de radiación solar global, estos resultados muestran valores de UV que superan altamente los mostrados por la OMS. Se estableció la correlación entre radiación ultravioleta y radiación solar global con datos registrados por la estación de SENAMHI, donde se obtuvo la ecuación lineal empírica para el año 2012 de $y = 0,00117 + 0,00021 x$, con un coeficiente de correlación lineal de Pearson $R = 0.95$, y para el año 2013 de $y = 0,00057 + 0,00023 x$, con un valor de $R = 0.96$, verificándose en ambos casos una muy alta correlación lineal. Se encontró también para datos del periodo 2012-2013 en conjunto la ecuación empírica $y = 0,00096 + 0,00021 x$, con coeficiente de determinación $R = 0.95$, lo que indica correlación positiva muy alta, que demuestra que este es un método puede ser aplicado para este tipo de estudios. Teniendo datos de radiación solar global, datos proporcionados por SENAMHI, se logró aplicar el modelo matemático empírico de regresión lineal para el cálculo de la variable UV, graficando la radiación UV para 1 y 2 de enero del 2014; con valores



medidos y calculados; se observa un comportamiento similar entre ambas curvas, validando de ésta manera la ecuación empírica (Mayhua, 2019).

2.1.3 A Nivel Internacional

Se presenta también un proyecto que se dá por la necesidad de monitorear el índice de radiación ultravioleta (IUV) en la ciudad de Popayán, derivada por el continuo aumento de los casos de cáncer y otras afecciones en la piel y afecciones oculares. El objetivo primordial de este proyecto fue de generar alarmas y sugerencias cuando el nivel de IUV fue dañino para los habitantes de la ciudad durante los meses octubre, noviembre y enero. La visualización fue diseñada por intermedio de una aplicación móvil, que recoge información enviada por un sensor UV (UVM-30A) a una base de datos a través de un módulo ESP32. Este prototipo, por su configuración sencilla y de fácil acceso podría expandirse en varios puntos de la región y del país para ampliar la información, convirtiéndose en un gran apoyo para la prevención de enfermedades cutáneas, oculares y de salud pública (Herrera & Ojeda, 2019).

En la actualidad la tecnología se vea inmersa a base de programaciones y control, por lo que en esta propuesta tecnológica se pretende implementar un sistema que sea capaz de monitorear la radiación ultravioleta, en la Universidad Técnica de Cotopaxi, exponiendo datos en tiempo real reflejados en una página web, una aplicación móvil y una pantalla numérica para referenciar los niveles de radiación solar, con las recomendaciones de acuerdo a lo establecido en la Organización Mundial de la Salud, además dispondrá de una base de datos que se podrá hacer uso mediante una página web o directamente de la memoria extraíble, la misma que registrara periódicamente las mediciones realizadas cada media hora, el sistema contara con dos paneles solares y su banco de dos baterías conectadas en serie para la respectiva alimentación de los



elementos. Este dispositivo será ubicado en un lugar estratégico, donde la afluencia de personas sea la mayor, con el propósito de que todo aquel individuo que concurra por dicho espacio pueda conocer cuál es el nivel de radiación al que está expuesto y de igual manera relacionar con las medidas de protección que debe tomar según sea el caso, de esta manera concientizar sobre el cuidado de la salud y aportar con un registro de datos históricos del comportamiento de la radiación UV en el laboratorio de la carrera de ingeniería eléctrica y de esta manera puedan realizar prácticas de evaluación de la potencia estimada en el diseño de sistemas fotovoltaicos u otra utilización que deseen dar a los datos reales extraídos (Cañizares, 2019).

2.2 MARCO TEÓRICO

2.1.1 Internet de las cosas (IOT)

Alineándonos con el primer objetivo específico, que es de recolectar datos mediante un sistema IoT, desarrollamos los conceptos y el modelo para la implementación de esta tecnología para la recolección de datos de radiación UV en la Ciudad de Puno.

Existe muchas definiciones de Internet de las cosas (IoT), pero de manera simple podemos mencionar que este concepto nace cuando: “sencillamente el punto en el tiempo en el que se conectaron a Internet más “cosas u objetos” que personas” (Evans, 2011).

Se puede mencionar que IoT está compuesta por una colección dispersa de redes diferentes y con distintos fines, como por ejemplo: cada automóvil que consta de diversas redes para controlar sus sistemas de funcionamiento: motor, transmisión, seguridad. Así también los edificios comerciales y residenciales para el control de su funcionamiento y seguridad. A medida que IoT evoluciona, estas redes y muchas otras estarán conectadas

con la incorporación de capacidades de seguridad, análisis y administración descrita en la Figura N° 1, (Evans, 2011).

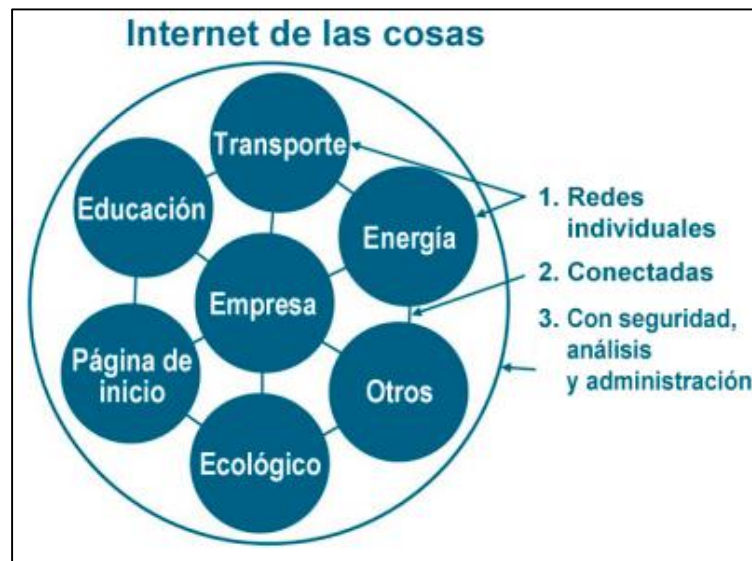


Figura N° 1: Representación de internet de IoT.

Fuente: <https://www.goconqr.com/es/p/6252265>.

En resumen podemos mencionar que IoT se refiere a escenarios en los que la conectividad de red y la capacidad de cómputo se extienden a objetos, sensores y artículos de uso diario que habitualmente no se consideran computadoras, permitiendo que estos dispositivos generen, intercambien y consuman datos con una mínima intervención humana (Rose, Eldridge & Chapin, 2015).

Este concepto no está lejos del proporcionado por la Unión Internacional de Telecomunicación – ITU (International Telecommunication Union, 2012), el Instituto de Ingeniería Eléctrica y Electrónica – IEEE (Minerva, Biru, & Rotondi, 2015) y el Grupo de Trabajo de Ingeniería de Internet – IETF (Internet Engineering Task Force, 2019), que indica: “Red global que conecta objetos identificables con capacidades de detección, actuación, procesamiento y comunicación con otros dispositivos, centro de datos o nubes



computacionales empleando tecnologías de comunicación e información, con el propósito de obtener aplicaciones avanzadas e información productiva”.

2.2.1.1 Tendencias tecnológicas y de mercado que están impulsando IoT

Entre los muchos factores existentes podemos destacar los siguientes:

- **Conectividad ubicua.**- el bajo costo y las redes de banda ancha que proporcionan alta velocidad junto con la tecnología inalámbrica facilitan la conexión de los dispositivos (Rose, Eldridge & Chapin, 2015).
- **Adopción generalizada de redes basadas en el protocolo Ipv6.**- al contar este protocolo con numerosas direcciones disponibles, al mismo tiempo ordenadas, permiten fácilmente la identificación debido a la creación de redes que ofrecen plataformas definidas que pueden interconectar fácilmente cualquier número de dispositivos IoT (Rose, Eldridge & Chapin, 2015).
- **Economías en la capacidad de cómputo.**- Actualmente la industria en general: comunicaciones, producción, comercio impulsan las inversiones tanto en investigación como en producción de elementos destinados para IoT (Rose, Eldridge & Chapin, 2015).
- **Miniaturización.**- Los avances tecnológicos en la tecnología microelectrónica, permiten actualmente contar con elementos que se acomodan a los requerimientos actuales de IoT: tamaño reducido, bajo costo, alta potencia computacional. Lo que permite el desarrollo de sensores y actuadores con las capacidades que exige IoT (Rose, Eldridge & Chapin, 2015).
- **Avances en el análisis de datos.**- los servicios virtualizados en la nube junto con el desarrollo de software para el análisis de datos, permiten trabajar con grandes



cantidades de datos con la capacidad de extraer información y conocimiento (Rose, Eldridge & Chapin, 2015).

- Surgimiento de computación en la nube.- junto con la computación en la niebla, permiten aprovechar los recursos informáticos remotos, lo que proporciona poder de almacenaje, procesamiento y gestión de los datos obtenidos, así como la distribución de los datos dando seguridad a los mismos (arquitectura distribuida) aumentando el poder de procesamiento y disminuyendo el costo de producción (Rose, Eldridge & Chapin, 2015).

2.2.1.2 Características principales de internet de las cosas

- a) **Detección y actuación:** Formado por sensores y actuadores que poseen cierta “inteligencia” para recolectar valores de las variables a medir, así como actuar sobre ellas (Rose, Eldridge & Chapin, 2015).
- b) **Comunicación:** Dichos sensores y actuadores tiene la capacidad de comunicarse o vincularse a través de protocolos y estándares definidos con otros objetos o aplicaciones para estar a disposición de los usuarios (Rose, Eldridge & Chapin, 2015).
- c) **Procesamiento:** Los datos obtenidos son trasladados hacia bases de datos presentes en la nube de internet o en la niebla, para su posterior procesamiento proporcionando información o conocimiento a los usuarios finales (Rose, Eldridge & Chapin, 2015).
- d) **Aplicación:** finalmente la información es presentada a los usuarios a través de interfaces adecuadas sean en forma de ases de datos tradicionales o mediante interfaces gráficas que pueden ser visualizadas en sistemas computacionales fijos o en sistemas móviles (Rose, Eldridge & Chapin, 2015).

Por lo que de la definición, tendencias, características principales y donde se aplica se logró una comprensión detallada de esta tecnología para poder entender los conceptos básicos y posteriormente aplicarlos en la recolección de datos.

2.2.1.3 Arquitectura global para internet de las cosas

En esta sección detallaremos la arquitectura referencial de IoT que se utilizó para el cumplimiento del primer objetivo específico. Como con la definición, existen muchos modelos de arquitectura para Internet de las cosas, presentamos el que mejor se acomoda al presente proyecto:

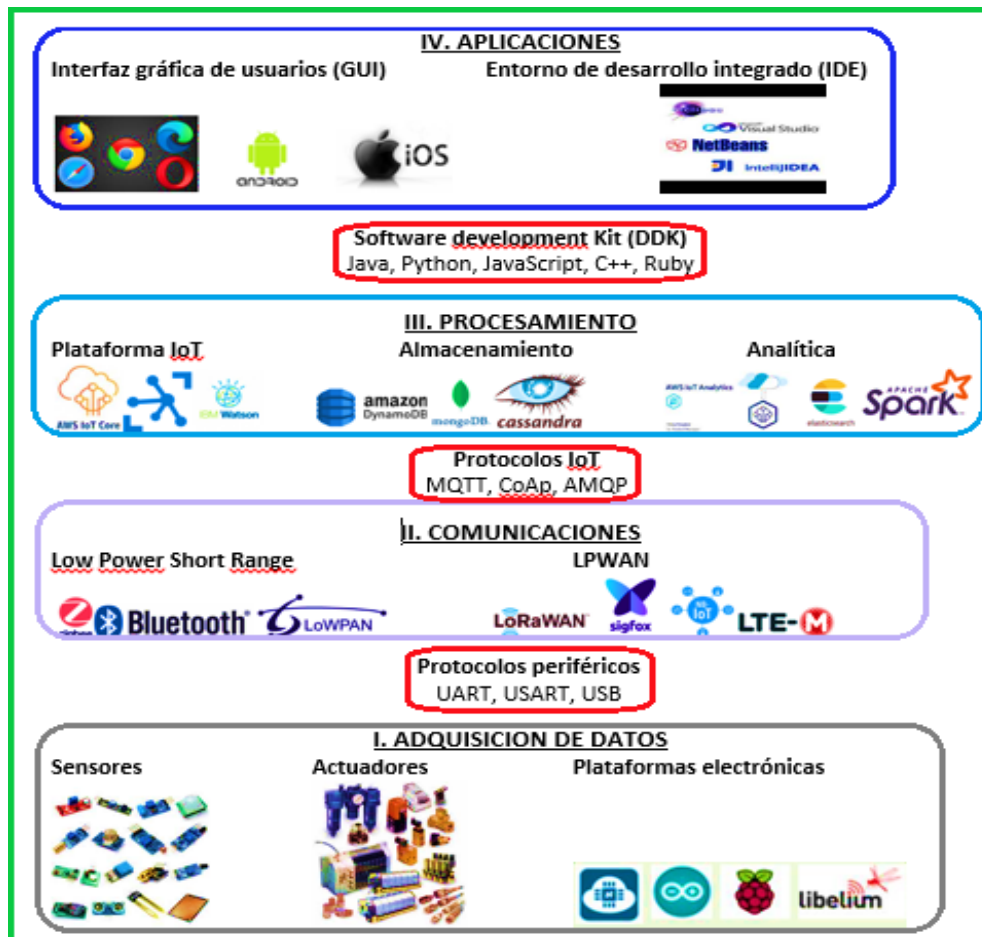


Figura N° 2: Arquitectura global de IoT.

Elaboración propia.

Nivel I: Adquisición de datos

En el primer nivel de la arquitectura IoT, se encuentran los sensores, actuadores y plataformas electrónicas encargadas de la adquisición de datos de las variables a ser monitoreadas. La Figura N° 3 representa el nivel de adquisición de datos de la arquitectura IoT.



Figura N° 3: Adquisición de datos.

Elaboración propia.

Este nivel básicamente está formado por sensores y actuadores y las plataformas electrónicas. Los sensores, objetos físicos que permiten la obtención de datos, y los actuadores que inciden sobre el fenómeno a controlar o medir. Se presentan algunos ejemplos de sensores como: termocuplas que son dispositivos empleados para medir temperatura que producen pequeñas señales, generalmente en el orden de los micro voltios.

Constan de dos materiales que al medir una misma temperatura producen un diferencial de voltaje independiente uno del otro. La diferencia entre los voltajes tiene una relación no lineal con la temperatura. La magnitud del voltaje depende del metal elegido. Son adecuados para la medición de amplios rangos de temperatura (Lea, 2018). En la Figura N° 4, se muestran las caracterizaciones para diferentes tipos de termocuplas.

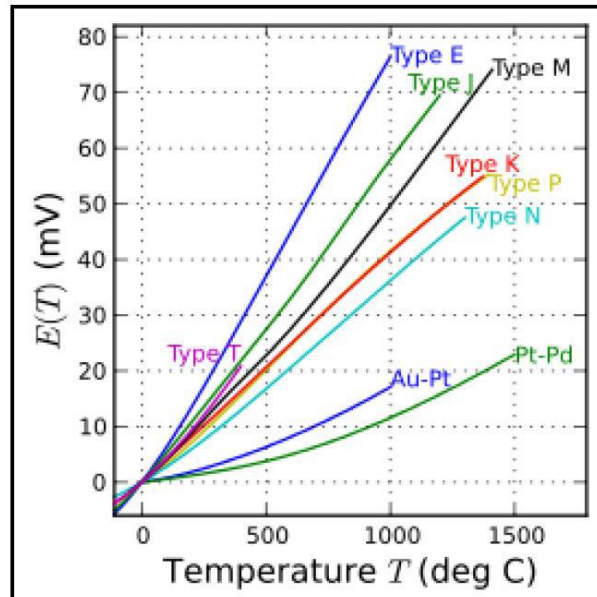


Figura N° 4: Características de Termocuplas.

Fuente: Lea, 2018.

Así también se tiene los detectores de infrarrojo no disperso (NDIR) que son empleados para la medición de concentraciones de óxidos de carbono (CO y CO₂). Su operación se fundamenta en la medida del total de luz infrarroja que absorbe una muestra de gas a una determinada frecuencia.

De acuerdo a la cantidad de luz infrarroja absorbida, el detector determina la concentración volumétrica del gas en la muestra. Se utiliza una rueda calibrada frente al detector para corregir el desplazamiento en la medición y la ganancia del analizador (Cambustion Ltd, 2019). En la Figura N° 5, se muestran las partes de un detector de infrarrojo no disperso.

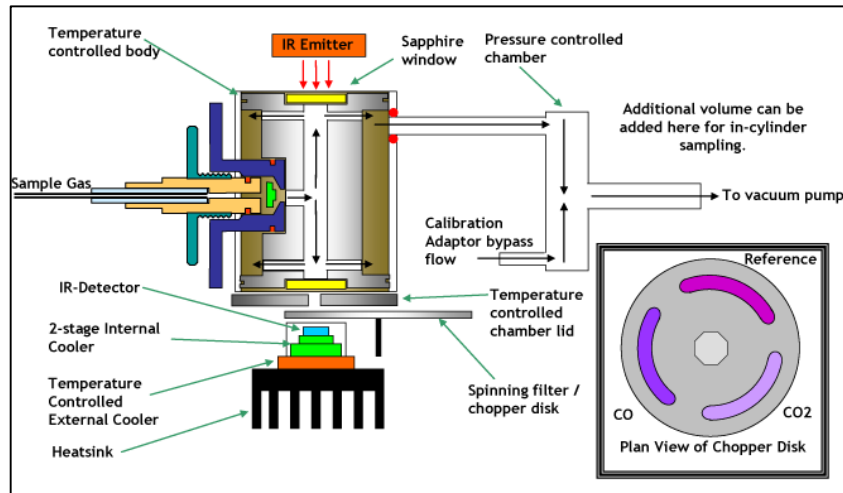


Figura N° 5: Detector infrarrojo.

Fuente: Cambustion Ltd, 2019.

En lo referente a actuadores podemos clasificarlos en tres grandes grupos: Actuadores hidráulicos, que son los de mayor antigüedad, y que pueden ser clasificados de acuerdo con la forma de operación, funcionan en base a fluidos a presión. Existen tres grandes grupos: cilindro hidráulico, motor hidráulico y motor hidráulico de oscilación (Núñez & Páez, 2016). En la Figura N° 6 se presentan algunos modelos.



Figura N° 6: Actuadores hidráulicos.

Fuente: <http://actuadoreshidraulicos.blogspot.com/2013/09/actuadores-hidraulicos.html>.

La siguiente clase de actuadores son los neumáticos, que convierten la energía del aire comprimido en trabajo mecánico. Aunque en esencia son idénticos a los actuadores hidráulicos, el rango de compresión es mayor en este caso, además de que hay una pequeña diferencia en cuanto al uso y en lo que se refiere a la estructura, debido a que estos tienen poca viscosidad (Núñez & Páez, 2016)



Figura N° 7: Actuadores neumáticos.

Fuente: <https://www.litem.info/actuadores-neumaticos/>

Por último se encuentran los actuadores eléctricos, cuya estructura es simple, ya que sólo requieren de energía eléctrica como fuente de poder. Es altamente versátil y prácticamente no hay restricciones respecto a la distancia entre la fuente de poder y el actuador. Existe una gran cantidad de modelos y es fácil utilizarlos con motores eléctricos estandarizados según la aplicación (Núñez & Páez, 2016).



Figura N° 8: Actuadores eléctricos.

Fuente: <https://www.interempresas.net/Robotica/Articulos/245870-EleCylinder-los-actuadores-electricos-middle-end.html>.

Por último existen diferentes plataformas electrónicas en el ecosistema tecnológico para el despliegue de soluciones fundamentadas en el Internet de las Cosas:

- a) **Arduino:** es una plataforma de creación de electrónica de código abierto, la cual está basada en hardware y software libre, flexible y fácil de utilizar para los creadores y desarrolladores. Esta plataforma permite crear diferentes tipos de microordenadores de una sola placa. Sus tarjetas electrónicas utilizan los microcontroladores Atmel AVR y son fabricadas por Arduino.cc y Arduino.org. Para extender las entradas, salidas y funcionalidades se utilizan placas adicionales denominadas “shields” (Kurniawan, 2016).
- b) **Raspberry Pi:** es un ordenador de bajo coste y tamaño reducido, tanto es así que cabe en la palma de la mano, pero puedes conectarle un televisor y un teclado para interactuar con ella exactamente igual que cualquier otra computadora. Creado por Eben Upton para fines educativos. Sus principales tarjetas electrónicas son los modelos Pi 3 B y Pi Zero. El Pi 3 B pertenece a la 3° genealogía de la familia Raspberry, el cual contiene un microprocesador Quad-Core de 64 bits, un módulo de comunicaciones Wi-Fi y Bluetooth Low Energy (BLE) (Kurniawan, 2016).



- c) **Microcontrolador ESP8266:** es un chip de bajo costo Wi-Fi con un stack TCP/IP completo y un microcontrolador, fabricado por Espressif, una empresa afincada en Shanghái, China. Existe una variedad de placas basadas en este microcontrolador. Una de ellas es NodeMCU, la cual fabrica kits de desarrollo con firmware de código abierto capaz de ser configurados mediante el lenguaje de programación Lua. Es un chip diseñado para el desarrollo del IoT, IIoT y la Industria 4.0, gracias a su bajo coste, su bajo consumo y, cada vez más, su facilidad de uso (Kurniawan, 2016).
- d) **Waspnote:** es una plataforma modular open source para construir redes de sensores inalámbricas de muy bajo consumo con código abierto para el manejo de una gran variedad de placas de sensores para medir eventos físicos y parámetros necesarios en el smart metering. Por lo que es adecuado para aplicaciones de Internet de la Cosas. Contiene numerosas placas de comunicaciones tales como ZigBee, Wi-Fi, Bluetooth, LoraWAN, etc. Asimismo, tiene la posibilidad de añadir sensores variados para diferentes aplicaciones tales como ciudades inteligentes, medioambiente inteligente, seguridad, control industrial, agricultura, (Kurniawan, 2016).

Nivel II: Comunicaciones

Uno de los principales habilitadores de un proyecto de Internet de las Cosas son las redes de comunicaciones que permiten conectar dispositivos, máquinas, sensores o “cosas” que generan datos desde cualquier punto geográfico del planeta. Es decir, cuando hablamos de un proyecto de IoT no tenemos por qué pensar en miles de sensores conectados a la vez. Con dos “generadores de datos”, que no se encuentren próximos, conectados de alguna forma a Internet, y que dispongan de una red de comunicaciones

que les permita mandar datos a Internet para su posterior tratamiento, ya tendríamos la base para un proyecto de Internet de las Cosas. Las redes de comunicaciones han ido evolucionando hacia el sector del IoT que, aunque actualmente no compite con el sector de la telefonía móvil a nivel comercial, sí que ha despertado el interés y la inversión de numerosas empresas en este sector (efor, 2017).



Figura N° 9: Nivel de Comunicaciones.

Elaboración propia.

Todas estas tecnologías se pueden clasificar en dos categorías principales:

- a) **Low Power Wide Area Network (LPWAN):** caracterizadas por proveer una gran cobertura sobre anchos de banda reducidos utilizando poca energía. Los nodos LPWAN logran un alcance en el orden de los kilómetros, dependiendo del área urbana o rural donde se despliegan. Los 2 principales componentes que hacen posible que las LPWAN logren una comunicación de largo alcance entre los nodos transmisores y los nodos receptores, son el uso de una topología de red del tipo estrella y técnicas de modulación en el enlace inalámbrico tales como espectro ensanchado pulsado (CSS) y banda ultra estrecha (UNB). Pueden desplegarse sobre bandas de frecuencias licenciadas (bandas de telefonía móvil) y no licenciadas (rango de frecuencias para actividad industrial, científica y médica - ISM) (Ismail, Abdul Latiff, Rokhani, & Abdul Aziz, 2019).
- b) **Low Power Short Range:** caracterizadas por otorgar una distancia de cobertura de hasta los 100 metros y consumir muy poca energía en su operación. Las

topologías de red que pueden adoptar son variadas, siendo los tipos estrella, malla y P2P las principales. Se despliegan sobre bandas de frecuencia no licenciadas (2400, 868 y 915 MHz). Las principales técnicas de modulación que emplean son espectro ensanchado por salto de frecuencia (FHSS) y espectro ensanchado de secuencia directa (DSSS). Las principales tecnologías correspondientes a esta categoría son Zigbee, Bluetooth Low Energy (BLE) y LoWPAN (Ghamari et al., 2016).

La Figura N° 10 revela un esquema comparativo de los alcances y velocidad de datos de las LPWAN, low power short range y redes celulares.

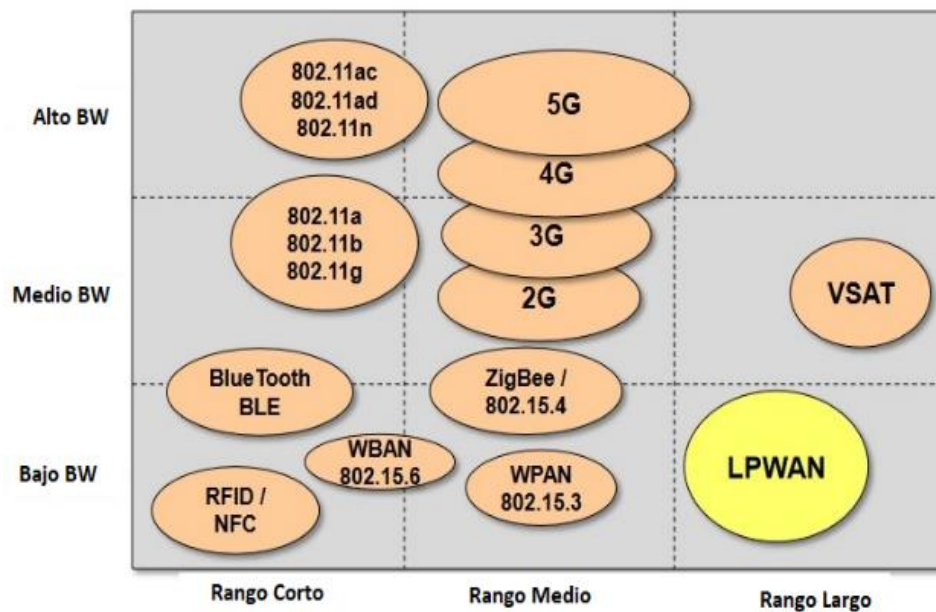


Figura N° 10: Relación cobertura ancho de banda para Low power short Range.

Fuente: <https://telos.fundaciontelefonica.com/la-cofa/las-tecnologias-lpwan-un-internet-de-las-cosas-low-cost/>.

- c) **Tecnologías LPWAN:** existen diversas tecnologías LPWAN para aplicaciones del Internet de las Cosas. Las principales son las siguientes (Mekki, Bajic, Chaxel, & Meyer, 2019):



- **LoRaWAN:** es una LPWAN que puede lograr un alcance en zonas urbanas de hasta 5 kilómetros y en zonas suburbanas hasta los 20 kilómetros de distancia. Puede lograr una velocidad de datos en bits hasta de 50 kbps. No tiene limitaciones en la cantidad de mensajes a enviar por día tanto en el uplink como en el downlink. El servicio de la red puede ser realizado mediante un despliegue propio de infraestructura o contratado a una empresa particular (Mekki, Bajic, Chaxel, & Meyer, 2019).
- **Sigfox** es una LPWAN creada por una compañía de Francia que opera en rangos de frecuencia que no requieren licencia y tiene un alcance en zonas urbanas de hasta 10 kilómetros. Utiliza modulación BPSK en anchos de banda muy estrechos de cien hercios. Alcanza velocidades de datos de 100 bps y la carga útil en el uplink es de 12 bytes y en el downlink es de 8 bytes. La cantidad máxima de mensajes que se puede enviar por día en el uplink es de 140 y en el downlink es de 4. Para el empleo de la red es necesaria la existencia de un operador en el ámbito geográfico de interés y la contratación de los respectivos servicios (Mekki, Bajic, Chaxel, & Meyer, 2019).
- **Narrow-Band IoT (NB-IoT):** es normada por el Proyecto de Asociación de 3^o Generación y opera en bandas de frecuencia licenciadas asignadas a la telefonía móvil. El alcance que logra es de hasta 1 kilómetro en zonas urbanas y utiliza la modulación QPSK. Puede desplegarse en anchos de banda de 200 kHz de forma autónoma, en bandas de guarda o dentro de una banda de una portadora LTE. Puede lograr velocidades de hasta 200 kbps, una carga útil de 1600 bytes y para su utilización se requiere la presencia de una empresa de comunicaciones móviles que brinde el servicio en la zona geográfica de interés (Mekki, Bajic, Chaxel, & Meyer, 2019).

- **LTE-M:** es otra LPWAN normada por la 3GPP. Logra un alcance superior a NB-IoT, llegando hasta los 5 kilómetros en zonas urbanas. Utiliza las modulaciones QPSK y 16QAM sobre un espectro de 1.4 MHz. La tasa de transmisión que logra es de 1 Mbps y al igual que NB-IoT, el servicio es brindado por un operador de telefonía móvil (Mekki, Bajic, Chaxel, & Meyer, 2019).

Nivel III: Procesamiento

En este nivel se encuentran las plataformas computacionales IoT, las bases de datos para el registro histórico y las herramientas de analítica para adquirir conocimiento valioso de la data obtenida. La Figura N° 11 muestra el nivel del procesamiento de la arquitectura detallada.



Figura N° 11: Nivel de Procesamiento.

Elaboración propia.

- a) **Computación en la nube:** es una tecnología que permite acceder remotamente, de cualquier lugar del mundo y en cualquier momento, a softwares, almacenamiento de archivos y procesamiento de datos a través de Internet, sin la necesidad de conectarse a un ordenador personal o servidor local. Utiliza la conectividad y gran escala de Internet para hospedar los más variados recursos, programas e información, y permite que el usuario ingrese a ellos través de cualquier ordenador, tablet o celular. Los principales atributos de la computación basada en la nube son (Mell & Grance, 2011):



- **Servicio de acuerdo a demanda:** El cliente es capaz de proveerse de funcionalidades computacionales automáticamente sin intervención humana del proveedor de servicio.
 - **Amplio acceso a red:** Se dispone de funcionalidades en la red y son accesibles mediante procedimientos que habilitan el empleo de plataformas clientes diversos.
 - **Medios compartidos en común:** Los entornos de cómputo de la compañía proveedora están concentrados para asistir a muchos usuarios mediante un marco multiusuario, con distintas entidades físicas y virtuales asignadas aleatoriamente y reasignadas en base al requerimiento del cliente.
 - **Elasticidad acelerada:** Las funcionalidades son capaces de ser provistas y liberadas para aumentar ágilmente en correspondencia con la demanda.
 - **Servicio medido:** Los entornos de la nube rigen y perfeccionan el empleo de recursos automáticamente aprovechando un entorno de control apropiado para distintas categorías de servicio.
- b) **Plataformas IoT:** son la base para que dispositivos estén interconectados y se genere un ecosistema propio. Dicho de otra forma una plataforma web integrada al Internet of Things (IoT) es el software que conecta hardware, puntos de acceso y redes de datos a lo que generalmente suele ser la aplicación de la que disfruta el usuario. Por lo que se puede indicar que una plataforma IoT es un conjunto de servicios computacionales que administra la operación de los dispositivos remotos, gestiona las comunicaciones y mecanismos de seguridad, procesa los datos obtenidos y remite la información a múltiples aplicaciones o servicios computacionales. Sus bloques principales son (Mell & Grance, 2011):



- Conectividad y normalización: con diferentes protocolos y diferentes formatos de datos en una interfaz de “software” garantiza la precisa transmisión de datos y la interacción con todos los dispositivos.
- La gestión de dispositivos: asegura que todas las “cosas” conectadas están funcionando correctamente.
- Base de datos: almacenamiento escalable de datos del dispositivo basados en la nube a un nuevo nivel en términos de volumen de datos, variedad, velocidad y veracidad.
- Procesamiento y gestión de la acción: aporta datos basados en reglas de acción de evento-disparadores que permitan la ejecución de las acciones “inteligentes” basados en datos específicos del sensor.
- Analítica: lleva a cabo una serie de análisis complejo de la agrupación de datos básicos y de aprendizaje automático.
- Visualización: permite a los seres humanos observar las tendencias de cuadros de mando de visualización de datos, donde se retrata vívidamente a través de gráficos.
- Herramientas adicionales: la IoT permiten a los desarrolladores de prototipos, probar y comercializar para visualizar, gestionar y controlar los dispositivos conectados.
- Interfaces externas: se integran con los sistemas de 3^a parte y el resto del ancho de TI en los ecosistemas a través de una función de interfaces de programación de aplicaciones (API), kits de desarrollo de software (SDK), y puertas de enlace (Mell & Grance, 2011).



- c) **Bases de datos para IoT:** la proliferación de los datos del IoT ha creado desafíos únicos para los sistemas de gestión de bases de datos, lo que requiere una capacidad de procesamiento e integración sólida, un rendimiento de clase empresarial y un bajo costo total de propiedad. Los medios de almacenamiento para aplicaciones IoT son del tipo NoSQL, las cuales procesan con rapidez grandes volúmenes de datos de diferentes tipos distribuidos geográficamente en comparación con bases de datos tradicionales. Presentamos algunas de ellas:
- **DynamoDB:** es una base de datos NoSQL ofrecida como servicio perteneciente a Amazon Web Services, la cual provee soporte para estructura de datos del tipo valor-clave y documentales. Permite crear tablas para almacenar cualquier tipo de datos las cuales pueden expandirse de forma rápida y ser modificadas sin afectar el rendimiento (Mell & Grance, 2011).
 - **MongoDB:** es también una base de datos NoSQL de fuente abierta que utiliza esquemas basados en documentos similares al lenguaje JSON y que puede ser instalado en distintas plataformas computacionales. Su instalación y configuración es compleja, lo cual requiere de usuarios experimentados (Mell & Grance, 2011).
 - **Cassandra:** base de datos NoSQL de fuente abierta basada en familia de columnas desarrollado por la Fundación de Software Apache. Puede ser implementado en diversas nubes computacionales y otorga redundancia completa (Mell & Grance, 2011).
- d) **Plataformas de analítica** para aplicaciones de IoT son utilizadas para obtener información valiosa del análisis de grandes volúmenes de datos generados por los dispositivos conectados. La información obtenida permite que los usuarios



puedan ejecutar mejores disposiciones con mayor precisión sobre los sistemas que están siendo supervisados. Algunos ejemplos de plataformas:

- **AWS IoT Analytics:** es un servicio de analítica para aplicaciones de IoT desarrollado por Amazon Web Services. Realiza la examinación de la data extraída por los elementos de adquisición, almacenándolos en esquemas de datos para su análisis a través de consultas SQL, análisis de series temporales o mediante la aplicación de aprendizaje de máquina para efectuar predicciones precisas. Soporta una amplia variedad de lenguajes para el desarrollo de aplicaciones de clientes (Amazon Web Services, 2019).
 - **ElasticSearch:** es un motor de búsqueda y analítica de código abierto desarrollado por la empresa Elasticsearch B.V. desde el año 2010. Ofrece una interfaz web para la búsqueda exhaustiva de textos empleando documentos JSON libre de esquemas. Puede utilizarse para un gran número de aplicaciones incluyendo inteligencia de negocios, análisis de registros, IoT, entre otros (Elasticsearch B.V., 2019).
 - **Apache Spark:** es una herramienta de analítica para la examinación de grandes volúmenes de datos auspiciado por la Fundación de Software Apache. Emplea una búsqueda de datos relacional, puede utilizarse sobre otros sistemas tales como Hadoop, Apache Mesos, Kubernetes, de forma independiente o en una nube computacional y soporta el desarrollo de aplicaciones en lenguaje Java, Scala, Python, R (Apache Software Foundation, 2019).
- e) **Protocolos de comunicación para IoT:** existen distintos protocolos de mensajería para aplicaciones de IoT, los cuales ofrecen características eficientes,



seguras y escalables en el manejo de datos. Se presenta a continuación, los principales protocolos de mensajería IoT:

- **Transporte de telemetría de cola de mensajes (MQTT):** protocolo para circulación de mensajes ligero muy sencillo fundamentado en suscribir y publicar, creado para terminales restringidos y conexiones no confiables, con alta latencia y bajo ancho de banda. Busca reducir el uso de los recursos de red y los requerimientos de servicios del terminal mientras asegura confiabilidad y algún grado de certeza en la entrega. Fue inventado en 1999 por Andy Stanford Clark y Arlen Nipper. Desde el año 2014 se encuentra estandarizado por la Organización para el Avance de las Normas de Información Estructurada (OASIS, por sus siglas en inglés). Los puertos TCP/IP 1883 se encuentran reservados para el protocolo MQTT por la Autoridad de Asignación de Números de Internet (IANA, por sus siglas en inglés) (mqtt.org, 2019).
- **Protocolo de aplicación restringida (CoAP):** es un protocolo de intercambio empleado en dispositivos y redes limitadas pertenecientes al ecosistema IoT. Está basado en el RFC 7252 y utiliza el modelo de Transferencia de Estado Representacional (REST). Dispone de una cabecera de 4 bytes y los formatos de datos soportados son XML, JSON, CBOR, entre otros. Hace uso del protocolo UDP sobre IP y utiliza el protocolo Seguro en el Nivel de Transporte de Datagrama (DTLS) como mecanismo de seguridad (Bormann, 2019).
- **Protocolo de colas de mensaje avanzado (AMQP):** es un esquema estandarizado por la Organización Internacional de Estandarización (ISO) con la numeración ISO/IEC 19464. Es un protocolo de comunicación orientado a mensajes con control de flujo. El protocolo envía los datos en forma de tramas de

manera serial a través de canales virtuales asociados a los clientes. Cada mensaje remitido consta de un identificador único (Lea, 2018).

Nivel IV: Aplicaciones

En este nivel se encuentran las aplicaciones para los usuarios que actúan como interfaces gráficas de usuario (GUI) desarrolladas en entornos de desarrollo integrado (IDE) mediante el uso de herramientas de elaboración de software (SDK). La Figura N° 12 esquema el nivel de aplicaciones de la arquitectura propuesta.



Figura N° 12: Nivel de aplicaciones.

Elaboración propia.

- a) **Interfaz gráfica de usuario (GUI):** sería el equivalente a la capa de Aplicación del modelo OSI, es decir, una forma visual de interactuar con un terminal computacional utilizando ventanas, gráficos, animaciones, entre otros. Esto faculta a los usuarios a tener acceso a los datos suministrados por el sistema de manera amigable y sencilla, facilitando el acceso y la comprensión del contenido de la información. En tal sentido, las interfaces gráficas de usuario para IoT pueden ser las aplicaciones web para navegadores de internet tales como Chrome, Microsoft Edge, Firefox, entre otros. Asimismo, las aplicaciones desarrolladas para sistemas operativos de terminales móviles tales como Android o iOS, son otro ejemplo de GUI (Aguilar, 2020).



b) **Entorno de desarrollo integrado (IDE):** permiten la creación de aplicativos de software en base a un lenguaje de programación haciendo uso de un editor de código, librerías, instrumentos de compilación, prueba y depuración de código. Facilitan la creación de aplicaciones IoT mediante la elaboración de páginas web o aplicativos celulares por medio del empleo de kits de desarrollo de software (SDK) para un lenguaje de programación o sistema operativo en particular (Aguilar, 2020). En esta sección se describió en detalle la arquitectura referencial de 6 capas del modelo IoT, el cual sirvió de base para el cumplimiento del primer objetivo específico, cabe mencionar que este tipo de modelo nos indica una “referencia”, no una camisa de fuerza o un estándar (Aguilar, 2020).

2.2.2 Introducción a la regresión multivariable

Para el cumplimiento del segundo objetivo específico, que es el de realizar el modelo de regresión multiparamétrica, daremos los conceptos necesarios, ventajas, tipos de regresión a ser utilizados: OLS, Ridge y Bayesian-Ridge, indicando sus principales características.

Por lo que se indica que en el mundo actual, los datos son generados a un ritmo sorprendente y de diversas fuentes, estos datos que pueden ser estructurados y no estructurados, deben ser manipulados, almacenados y tratados para obtener información que sea de utilidad a los usuarios y a partir de ellos obtener el conocimiento para su beneficio. Para su tratamiento se utilizarán técnicas estadísticas que se incluyen en el campo del Machine Learning que permitirán describir y visualizar, reducir, revisar, resumir y evaluar datos en información útil que proporciona un mejor contexto para los datos. Este análisis de datos juega actualmente un papel fundamental en la toma de decisiones de la industria, comercio del aparato privado y estatal de los países. Este tratamiento de datos en este momento se le denomina ciencia de los datos, que no es sino



combinar muchos métodos, metodologías científicas, procesos, algoritmos y herramientas para extraer de grandes cantidades de datos (no necesariamente) información necesaria de los datos estructurados y no estructurados (Aguilar, 2020).

2.2.2.1 Análisis de regresión

Se define el análisis de regresión como una herramienta de frecuente uso en estadística, la cual permite investigar las relaciones entre diferentes variables cuantitativas, mediante la formulación de ecuaciones matemáticas. Este método es uno de los métodos más buscados utilizados en el análisis de datos. Sigue un algoritmo de aprendizaje automático supervisado. El análisis de regresión es un método estadístico importante que nos permite examinar la relación entre dos o más variables en el conjunto de datos. Si nos guiamos por el método científico responde a las preguntas: ¿las variables importantes? ¿Cuáles se pueden ignorar? ¿Cómo interactúan entre sí? Y lo más importante es ¿qué tan seguros estamos de estas variables? Generalmente se tiene una variable dependiente, el factor principal que estamos tratando de comprender o predecir, junto con una o varias variables independientes: los factores que creemos tienen un impacto en la variable dependiente. Por lo que de manera resumida mencionamos que la regresión lineal simple es un modelo de regresión que estima la relación entre una variable dependiente y una variable independiente usando una línea recta. De otro lado la regresión lineal múltiple estima la relación entre dos o más variables independientes y una variable dependiente. La diferencia entre estos dos modelos es el número de variables independientes. Al ser un modelo matemático los modelos de regresión mencionados anteriormente no funcionarán. Esto debido a que en el mundo real, hay una gran cantidad de situaciones en las que muchas variables independientes se ven influenciadas por otras variables, por lo que tenemos que buscar otras opciones en lugar de un modelo de regresión único que solo puede funcionar con una variable independiente (NCSS, 2021).



a) ¿Qué es la regresión multivariable?

La regresión multivariable es un algoritmo de aprendizaje automático supervisado que involucra múltiples variables de datos para su análisis. Una regresión multivariable es una extensión de la regresión múltiple con una variable dependiente y múltiples variables independientes. Con base en el número de variables independientes, intentamos predecir la salida. La regresión multivariable intenta encontrar una fórmula que pueda explicar cómo los factores en las variables responden simultáneamente a los cambios en otros. Existen numerosas áreas en las que se puede utilizar la regresión multivariable. Veamos algunos ejemplos para comprender mejor la regresión multivariable: por ejemplo si se quiere estimar el precio de una casa, se recopilará detalles como la ubicación de la casa, el número de dormitorios, el tamaño en pies cuadrados, las comodidades disponibles o no; en base a estos detalles, se puede predecir el precio de la casa y cómo se interrelacionan cada una de las variables. Por otro lado un científico agrícola quiere predecir el rendimiento total de los cultivos que se espera para el verano. Recopiló detalles de la cantidad esperada de lluvia, los fertilizantes que se utilizarán y las condiciones del suelo. Al construir un modelo de regresión multivariable, los científicos pueden predecir el rendimiento de su cultivo. Con el rendimiento del cultivo, el científico también intenta comprender la relación entre las variables. Otro ejemplo es si una organización quiere saber cuánto tiene que pagar a un nuevo empleado, tendrá en cuenta muchos detalles, como el nivel de educación, el número de experiencia, la ubicación del trabajo, tiene una habilidad específica o no. En base a esta información se puede predecir el salario de un empleado, cómo estas variables ayudan a estimar el salario. Si un economista pueden usar la regresión multivariable para predecir el crecimiento del PIB de un estado o un país en función de parámetros como el monto total gastado por los consumidores, el gasto de importación, las ganancias totales de las exportaciones, el ahorro total, (Forero, 2020).

b) Ecuaciones matemáticas

El modelo lineal de regresión simple representa una línea recta, lo que significa que y es una función de x .

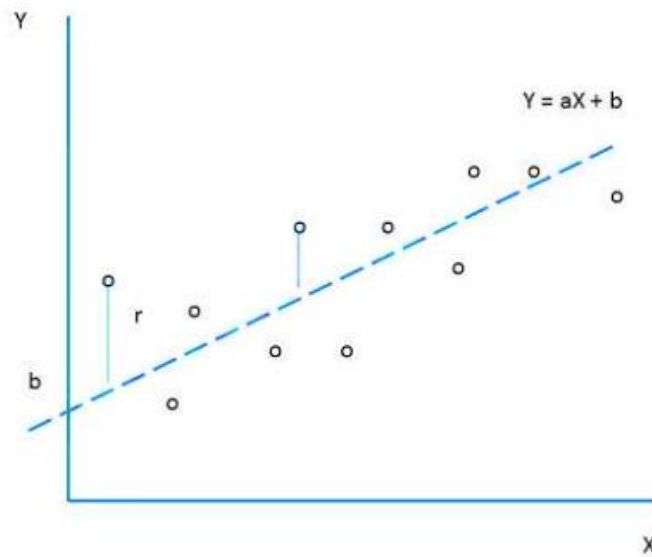


Figura N° 13: Modelo de regresión lineal.

Fuente: Nicosia, 2020.

La ecuación de regresión lineal ahora se puede expresar como:

$$Y = a \cdot X + b \quad (1)$$

Donde:

Y es la variable dependiente, es decir, la variable que debe predecirse.

X es la variable independiente. Es la primera entrada.

a es la pendiente de X . Nos permite conocer el ángulo de la recta (x).

Generalizando y tomando una notación estándar, la ecuación para un modelo con dos variables de entrada se puede escribir como:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 \quad (2)$$

De manera análoga si tenemos más de tres variables, no se puede visualizar pero si se puede generar el modelo matemático y la ecuación del modelo con tres variables de entrada se puede escribir como:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 \quad (3)$$

Si queremos generalizar entonces el modelo de regresión multivariable quedará como:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots \beta_n \cdot x_n \quad (4)$$

Donde n representa el número de variables independientes, $\beta_0 \sim \beta_n$ representa los coeficientes y $x_1 \sim x_n$, es la variable independiente.

El modelo multivariado nos ayuda a comprender y comparar los coeficientes de la salida. Aquí, la función de costo pequeño hace que la regresión lineal multivariable sea un mejor modelo (Nicosia, 2020).

c) Función de coste

La función de costo (MSE) es una función que permite asignar un costo a las muestras cuando el modelo difiere de los datos observados. Esta ecuación es la suma del cuadrado de la diferencia entre el valor predicho y el valor real dividido por el doble de la longitud del conjunto de datos. Un error cuadrático medio más pequeño implica un mejor desempeño. Aquí, el costo es la suma de los errores al cuadrado (Nicosia, 2020).

Por lo que el costo de la regresión múltiple estará representado por:

$$MSE = \frac{1}{2m} \sum (h_{\theta}(x^i) - y^i)^2 \quad (5)$$

d) Pasos del análisis de regresión multivariable

No necesariamente se tomarán los siguientes pasos de manera obligatoria, pero son una buena guía para el desarrollo del modelo a implementar.



- **Selección de características:** es un paso importante en la regresión multivariable. La selección de funciones también se conoce como selección de variables. Es importante para nosotros elegir variables significativas para una mejor construcción del modelo, se realizará en función del conocimiento del experto o especialista en el área.
- **Normalización de características:** necesitamos escalar las características, ya que mantiene la distribución general y las proporciones en los datos. Esto conducirá a un análisis eficiente. El valor de cada característica también se puede cambiar.
- **Selección Función de pérdida e Hipótesis:** la función de pérdida predice siempre que haya un error. Es decir, cuando la predicción de la hipótesis se desvía de los valores reales. Aquí, la hipótesis es el valor predicho de la característica o variable.
- **Establecer parámetros de hipótesis:** el parámetro de hipótesis debe establecerse de tal manera que reduzca la función de pérdida y prediga bien.
- **Minimizar la función de pérdida:** la función de pérdida debe minimizarse mediante el uso de un algoritmo de minimización de pérdidas en el conjunto de datos, que ayudará a ajustar los parámetros de hipótesis. Una vez que se minimiza la pérdida, se puede utilizar para otras acciones. El descenso de gradientes es uno de los algoritmos comúnmente utilizados para minimizar las pérdidas.
- **Prueba la función de hipótesis:** la función de hipótesis también debe verificarse, ya que predice valores. Una vez hecho esto, debe probarse con datos de prueba (NCSS, 2021).



e) Ventajas y desventajas de la regresión multivariable

La ventaja más importante de la regresión multivariable es que nos ayuda a comprender las relaciones entre las variables presentes en el conjunto de datos. Esto ayudará aún más a comprender la correlación entre variables dependientes e independientes. La regresión lineal multivariable es un algoritmo de aprendizaje automático ampliamente utilizado. Como desventajas podemos mencionar: las técnicas multivariadas son un poco complejas y requieren un alto nivel de cálculo matemático, el resultado del modelo de regresión multivariable no es fácil de interpretar a veces, porque tiene algunos resultados de pérdida y error que no son idénticos; por último este modelo no tiene mucho margen para conjuntos de datos más pequeños. Por tanto, no se les puede aplicar lo mismo. Los resultados son mejores para conjuntos de datos más grandes (Forero, 2020).

f) Regresión OLS

La regresión de mínimos cuadrados ordinarios (OLS) es un método de análisis estadístico que estima la relación entre una o más variables independientes y una variable dependiente; el método estima la relación minimizando la suma de los cuadrados en la diferencia entre los valores observados y predichos de la variable dependiente configurada como una línea recta. Se puede realizar la pregunta: “¿Cuál es la relación entre la variable X y la variable Y o conjunto de variable Y_i ?”. Si la relación se supone que es lineal, regresión multivariable puede usarse para abordar esta cuestión mediante el ajuste de una línea recta a un diagrama de dispersión de las observaciones de la variable X y las variables Y_i . El enunciado más simple de tal relación entre una variable independiente, etiquetada como X , y una variable dependiente o conjunto de variables independientes etiquetadas como Y_i . Por lo que OLS es el método de estimación más común para modelos lineales, y eso es cierto por una buena razón. Siempre que su modelo

satisfaga los supuestos de OLS para la regresión lineal, es muy probable que obtenga las mejores estimaciones posibles. La regresión es un análisis poderoso que puede analizar múltiples variables simultáneamente para responder preguntas de investigación complejas. Sin embargo, si no cumple con los supuestos de OLS, es posible que no pueda confiar en los resultados (Kapoor, 2020).

El análisis de regresión es como otras metodologías inferenciales. Su objetivo es extraer una muestra aleatoria de una población y usarla para estimar las propiedades de esa población. En el análisis de regresión, los coeficientes de la ecuación de regresión son estimaciones de los parámetros reales de la población. Se necesita que estas estimaciones de coeficientes sean las mejores estimaciones posibles. Las estimaciones deberían tender a acertar en el objetivo. No deben ser sistemáticamente demasiado altas o demasiado bajas. En otras palabras, deben ser imparciales o correctos en promedio. Al reconocer que las estimaciones casi nunca son exactamente correctas, desea minimizar la discrepancia entre el valor estimado y el valor real. Evitando que estas diferencias tengan un valor muy alto (Towards Data Science, 2019).

Cuando su modelo de regresión lineal satisface los supuestos de OLS, el procedimiento genera estimaciones de coeficientes insesgadas que tienden a ser relativamente cercanas a los valores reales de la población (varianza mínima). De hecho, el teorema de Gauss-Markov establece que OLS produce estimaciones que son mejores que las estimaciones de todos los demás métodos de estimación de modelos lineales cuando los supuestos son ciertos.

Este modelo de regresión puede ser representado por la siguiente ecuación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (6)$$

Donde ε es el error aleatorio.



Mencionaremos que tanto las variables independientes como el término de error se asumen lineales (Bruin, 2006).

g) **Regresión Ridge**

Ridge Regression es una técnica para analizar datos de regresión múltiple que sufren de multicolinealidad. Cuando ocurre la multicolinealidad, las estimaciones de mínimos cuadrados son insesgadas, pero sus varianzas son grandes, por lo que pueden estar lejos del valor real. Al agregar un grado de sesgo a las estimaciones de regresión, la regresión Ridge reduce los errores estándar.

- **Multicolinealidad**

La multicolinealidad, o colinealidad, es la existencia de relaciones casi lineales entre las variables independientes. Durante los cálculos de regresión, esta relación provoca una división por cero que a su vez provoca que los cálculos se cancelen. Cuando la relación no es exacta, la división por cero no ocurre y los cálculos no se cancelan. Sin embargo, la división por una cantidad muy pequeña todavía distorsiona los resultados. Por lo tanto, uno de los primeros pasos en un análisis de regresión es determinar si la multicolinealidad es un problema (Salmerón, 2017).

- **Efectos de la multicolinealidad**

La multicolinealidad puede crear estimaciones inexactas de los coeficientes de regresión, inflar los errores estándar de los coeficientes de regresión, desinflar las pruebas t parciales para los coeficientes de regresión, dar valores p falsos, no significativos (Salmerón, 2017).

- **Fuentes de multicolinealidad**

Para hacer frente a la multicolinealidad, debe poder identificar su origen. La fuente de la multicolinealidad afecta el análisis, las correcciones y la interpretación del



modelo lineal. Son cinco fuentes (Salmerón, 2017):

Recolección de datos. En este caso, los datos se han recopilado de un subespacio estrecho de las variables independientes. La multicolinealidad ha sido creada por la metodología de muestreo; no existe en la población. La obtención de más datos sobre un rango ampliado solucionaría este problema de multicolinealidad.

Restricciones físicas del modelo lineal o población. Esta fuente de multicolinealidad existirá independientemente de la técnica de muestreo que se utilice. Muchos procesos de fabricación o servicios tienen restricciones sobre variables independientes (en cuanto a su rango), ya sea física, política o legalmente, lo que creará multicolinealidad.

Modelo sobre definido. Aquí hay más variables que observaciones. Esta situación debe evitarse.

Elección o especificación del modelo. Esta fuente de multicolinealidad proviene del uso de variables independientes que son potencias o interacciones de un conjunto original de variables. Cabe señalar que si el subespacio muestral de las variables independientes es estrecho, cualquier combinación de esas variables aumentará aún más el problema de multicolinealidad.

Valores atípicos. Los valores extremos o valores atípicos en el espacio X pueden causar multicolinealidad y ocultarla. A esto lo llamamos multicolinealidad inducida por valores atípicos. Esto debe corregirse eliminando los valores atípicos antes de aplicar la regresión ridge.

- **Detección de multicolinealidad**

Existen varios métodos para detectar la multicolinealidad. Mencionamos algunos. Empiece por estudiar diagramas de dispersión por pares de pares de variables independientes, buscando relaciones casi perfectas. También eche un vistazo a la



matriz de correlaciones para correlaciones altas. Desafortunadamente, la multicolinealidad no siempre aparece cuando se consideran las variables de dos en dos (Salmerón, 2017).

Considere los factores de inflación de la varianza (VIF). Los VIF superiores a 10 indican variables colineales (Salmerón, 2017).

Los valores propios de la matriz de correlación de las variables independientes cercanas a cero indican multicolinealidad (Salmerón, 2017).

En lugar de mirar el tamaño numérico del valor propio, use el número de condición. Los números de condición grandes indican multicolinealidad (Salmerón, 2017).

Investigue los signos de los coeficientes de regresión. Las variables cuyos coeficientes de regresión son de signo opuesto al que cabría esperar pueden indicar multicolinealidad (Salmerón, 2017).

- **Corrección por multicolinealidad**

Dependiendo de cuál sea la fuente de multicolinealidad, las soluciones variarán. Si la multicolinealidad ha sido creada por la recopilación de datos, recopile datos adicionales en un subespacio X más amplio. Si la elección del modelo lineal ha aumentado la multicolinealidad, simplifique el modelo utilizando técnicas de selección de variables. Si una observación o dos ha inducido la multicolinealidad, elimine esas observaciones. Sobre todo, tenga cuidado al seleccionar las variables desde el principio. Cuando estos pasos no sean posibles, puede intentar la regresión Ridge (Cruz, 2021).

h) Regresión Bayesian Ridge

La regresión es una tarea de aprendizaje automático para predecir valores continuos (números reales), en comparación con la clasificación, que se utiliza

para predecir valores categóricos (discretos). Cuando escuche la palabra 'Bayesiano', podría pensar en Naive Bayes. Sin embargo, los principios bayesianos también se pueden utilizar para realizar la regresión.

En primer lugar, debe comprender que Bayesiano es solo un enfoque para definir y estimar modelos estadísticos. La regresión bayesiana puede ser muy útil cuando no tenemos datos suficientes en el conjunto de datos o los datos están mal distribuidos. El resultado de un modelo de regresión bayesiana se obtiene a partir de una distribución de probabilidad, en comparación con las técnicas de regresión regulares donde el resultado se obtiene simplemente a partir de un valor único de cada atributo. La salida, 'y' se genera a partir de una distribución normal (donde la media y la varianza están normalizadas). El objetivo de la regresión lineal bayesiana no es encontrar los parámetros del modelo, sino más bien encontrar la distribución 'posterior' de los parámetros del modelo. No solo la salida y, sino que también se supone que los parámetros del modelo provienen de una distribución (Cruz, 2021). La expresión para Posterior es:

$$\mathbf{Posterior} = \frac{\mathbf{Likelihood*Prior}}{\mathbf{Normalizacion}} \quad (7)$$

Donde:

Posterior: Es la probabilidad de que ocurra un evento; digamos, H, dado que otro evento; digamos, E ya ha ocurrido, es decir, P (H | E).

Prior: Es la probabilidad de que un evento H haya ocurrido antes de otro evento, es decir, P (H)

Likelihood: Es una función de verosimilitud en la que se margina alguna variable de parámetro.

En realidad, esto es equivalente al teorema de Bayes que dice:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (8)$$

Donde:

A y B son eventos, P (A) es la probabilidad de ocurrencia de A

P (A | B) es la probabilidad de que ocurra A dado que el evento B ya ocurrió.

P (B), la probabilidad de que ocurra el evento B no puede ser 0 porque ya ocurrió.

Al observar la fórmula anterior, podemos ver que, en contraste con LSO, tenemos una distribución posterior para los parámetros del modelo que es proporcional a la probabilidad de los datos multiplicada por la probabilidad previa de los parámetros. A medida que aumenta el número de puntos de datos, el valor de probabilidad aumentará y será mucho mayor que el valor anterior. En el caso de un número infinito de puntos de datos, los valores de los parámetros convergen a los valores obtenidos de LSO. Entonces, comenzamos nuestro proceso de regresión con una estimación inicial (el valor anterior). A medida que comenzamos a cubrir más puntos de datos, nuestro modelo se vuelve menos erróneo. Por lo tanto, para la regresión ridge bayesiana, se necesita una gran cantidad de datos de entrenamiento para que el modelo sea preciso (Cruz, 2021).

- **Ventajas de la regresión bayesiana:**

Muy eficaz cuando el tamaño del conjunto de datos es pequeño.

Particularmente adecuado para el aprendizaje en línea (los datos se reciben en tiempo real), en comparación con el aprendizaje por lotes, donde tenemos todo el conjunto de datos en nuestras manos antes de comenzar a entrenar el modelo. Esto se debe a que la regresión bayesiana no necesita almacenar datos.

El enfoque bayesiano es un enfoque probado y comprobado y es muy sólido, matemáticamente. Entonces, uno puede usar esto sin tener ningún conocimiento



previo adicional sobre el conjunto de datos.

- **Desventajas de la regresión bayesiana:**

La inferencia del modelo puede llevar mucho tiempo. Si hay una gran cantidad de datos disponibles para nuestro conjunto de datos, el enfoque bayesiano no vale la pena y el enfoque frecuentista regular hace un trabajo más eficiente (Nicosia, 2020).

En esta sección se explicó de manera resumida los conceptos básicos de los tres modelos realizados: OLS. Ridge y Bayesian-Ridge para realizar el modelo multiparámetro de los datos obtenidos de radiación UV en la ciudad de Puno.

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 UBICACIÓN GEOGRÁFICA DEL ESTUDIO

La investigación se desarrolló en el distrito de Puno, Provincia de Puno, departamento de Puno, que tiene las siguientes coordenadas geográficas: 15°50'36"S 70°01'25"O a 3818msnm. Los experimentos se desarrollaron en tres ubicaciones de la ciudad de Puno.



Figura N° 14: Ciudad de Puno.

Elaboración propia.

3.2 PERIODO DE DURACIÓN DEL ESTUDIO

Tabla N° 1: Duración del estudio.

Actividad	Mes						
	J	J	A	S	O	N	D
Análisis y búsqueda de información	X	X					
Diseño de la solución		X	X	X			
Implementación del sistema			X	X	X		
Pruebas del diseño				X	X		
Obtención del resultado					X	X	
Redacción del trabajo final						X	X

Elaboración propia.

3.3 PROCEDENCIA DEL MATERIAL UTILIZADO

Tabla N° 2: Lista de materiales.

Descripción	Unidad de medida	Costo unitario (S/.)	Cantidad	Costo total (S/.)
Ordenador	Unidad	3 000.00	1	3 000.00
Sistema de monitorización	Sistema	2 000.00	3	6 000.00
Herramientas de programación	Herramientas	100.00	3	300.00
Útiles de escritorio		500.00	1	500.00
Otros		1 000.00	1	1 000.00
Total				10 800.00

Elaboración propia.

Tabla N° 3: Materiales utilizados para el módulo de adquisición de datos.

N°	Cantidad	Descripción
1	3	Arduino UNO
2	3	Shield W5100
3	3	The ultraviolet (UV) sensor GYML8511
4	3	Toma de alimentación regulada
5	3	Acceso a Internet
6		Cables y conectores diversos

Elaboración propia.

3.4 POBLACIÓN Y MUESTRA DEL ESTUDIO

La población para el trabajo de investigación está constituida por las muestras representativas de las tres ubicaciones en estudio, donde podemos mencionar en resumen la siguiente data obtenida de datos analizados:

- 01/09/2021 - 31/09/2021
- Datos recolectados: $14880 \times 3 = 44640$

- Datos procesados (se eliminaron filas): $13346 \times 3 = 40038$

En tal resumen se realizó una etapa de pre procesamiento en el cual están los prototipos utilizados, que se muestran en la Figura N° 15.

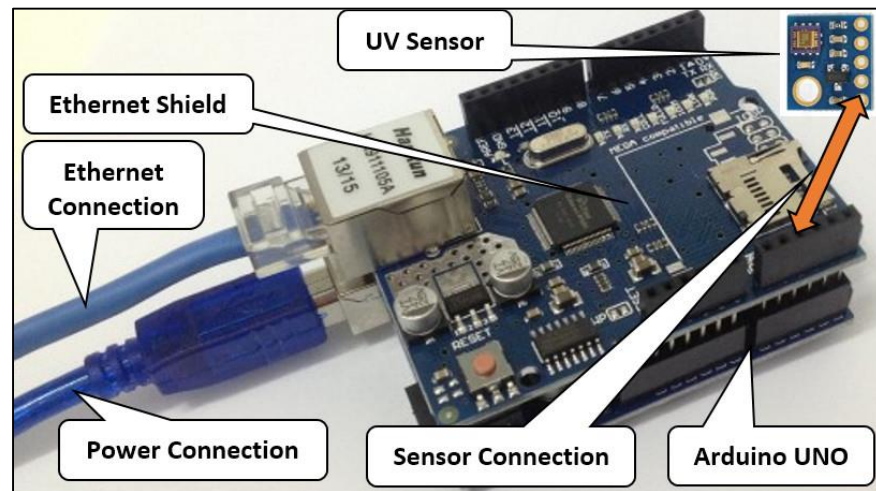


Figura N° 15: Módulo utilizado.

Elaboración propia.

Se tomará toda la población como muestra debido a que se necesitan muchos valores para que la regresión sea fiable y así obtener los resultados deseados.

3.5 DISEÑO ESTADÍSTICO

3.5.1 Diseño de la investigación

Esta investigación es experimental, porque es un tipo de investigación que utiliza la lógica y los principios encontrados en las ciencias y que serán implementados en un prototipo. La esencia de esta concepción de experimento es que requiere la manipulación intencional de una acción para analizar sus posibles resultados. Los experimentos pueden ser llevados a cabo en el laboratorio o en la vida real. Aquí se involucran un número relativamente pequeño de personas o equipos y abordan una pregunta bastante enfocada.



Los experimentos son más efectivos para la investigación explicativa y frecuentemente están limitados a temas en los cuales se puede manipular la situación en la cual las personas se hallan. Esta investigación es experimental, se centra en la medición de la radiación UV y su regresión multiparamétrica (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010).

3.5.2 Nivel de la investigación

El nivel de investigación se refiere a la profundidad del conocimiento que se busca lograr con la investigación, por tanto, el nivel de la presente investigación es exploratorio, señalando que las investigaciones exploratorias buscan abrir nuevos caminos en el desarrollo del conocimiento humano.

3.6 PROCEDIMIENTO

3.6.1 Descripción del hardware y software del prototipo

En la presente tesis se utilizó el siguiente hardware y software:

a) Nivel de Adquisición de datos:

Se utilizó el siguiente sensor para la medición de radiación UV: GYML8511 que es un sensor UV adecuado para adquirir intensidad UV en interiores o exteriores. El sensor utiliza la tecnología SOI-CMOS por lo que se utilizó para la presente tesis, Funciona a partir del principio de que la fotocorriente se puede convertir en voltaje, el sensor puede detectar la intensidad de los rayos UV, por lo que se escogió por su buen funcionamiento en el exterior. Finalmente indicaremos que se genera un voltaje analógico proporcional a la intensidad de los rayos ultravioleta. En la Figura N° 16 se muestra el sensor y su diagrama interno

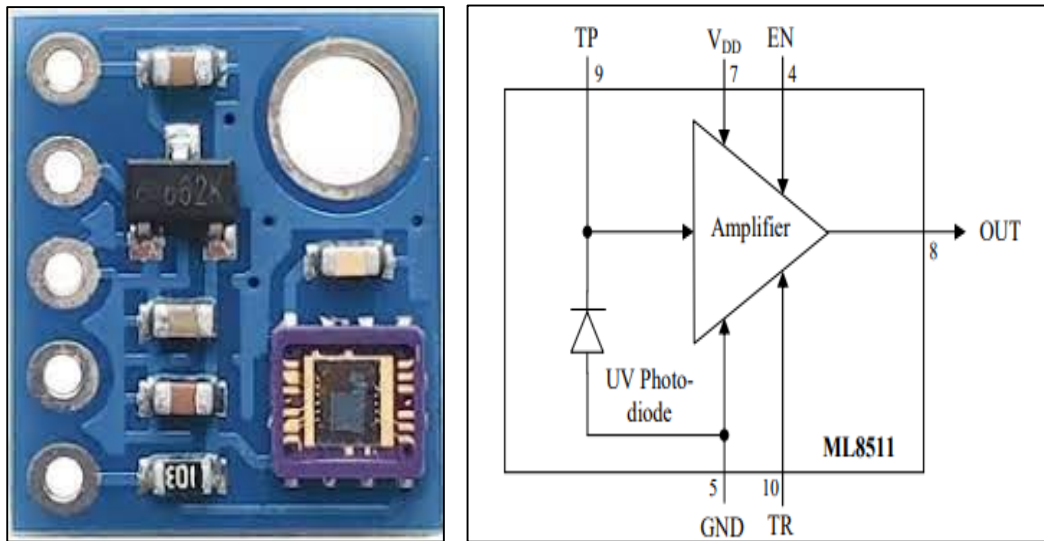


Figura N° 16: Sensor GYML8511 y su diagrama interno.

Fuente: http://wiki.sunfounder.cc/index.php?title=GYML8511_UV_Sensor.

Así mismo en este nivel se utilizó en lo que respecta a la plataforma electrónica los siguientes componentes: Arduino Uno con una placa Ethernet Shield W5100, que se muestran en la figura N° 17.

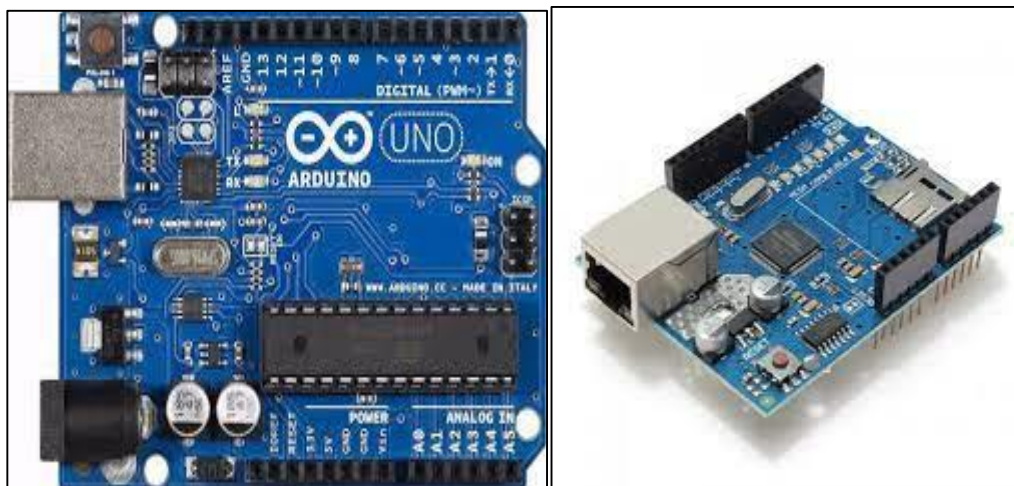


Figura N° 17: Arduino UNO y Ethernet Shield W5100.


Fuente: http://wiki.sunfounder.cc/index.php?title=GYML8511_UV_Sensor.



El Shield Ethernet W5100 es un complemento que permite controlar el Arduino a través de Internet o de la LAN domiciliaria. Algunos de los campos donde se aplica este shield serán: domótica, automatización, Internet de las cosas (IoT), control y monitoreo remoto, etc. Es compatible con Arduino Uno, Mega y Leonardo. Además las librerías Ethernet y SD vienen incluidas por defecto en el IDE de Arduino, por lo que no hay necesidad de descargar librerías adicionales. Así mismo incluye un slot para tarjetas micro-SD, el cual puede ser empleado para almacenar archivos que podrás poner disponibles a través de la red. Otra característica es que este modelo contiene un controlador de reset, esto es para asegurarse que el módulo W5100 Ethernet inicie correctamente al ser energizado. El Shield Ethernet tiene una interfaz digital SPI, por lo que debemos revisar cuales son los pines SPI en el modelo de Arduino a utilizar, En Arduino Uno son los pines digitales 11, 12 y 13, y el pin 10 es empleado para seleccionar el W5100 y el pin 4 para la tarjeta SD. Debemos tomar en cuenta que esos pines solo se usaran para SPI y ya no estarán disponibles para otras tareas.

El shield posee un conector RJ45 estándar para Ethernet. El botón de reset inicializa tanto el shield como el Arduino. Una gran ventaja de este shield es que es apilable por lo que podrás disponer de todos sus pines en otros shields.

Para lograr el funcionamiento es necesario realizar la programación de los dispositivos en el entorno de desarrollo integrado IDE Arduino pues el hardware Arduino Uno y el Shield W5100, utilizan este lenguaje, la programación se trata de un conjunto de instrucciones del lenguaje C, estas instrucciones son planteadas para facilitar el funcionamiento del desarrollo del programa y el uso de la plataforma en conjunto hardware y software, como se muestra en la Figura N° 18.



```
COM5
Initialize System
W5100 initialized
IP Address: 192.168.1.2
GET / HTTP/1.1
GET /favicon.ico HTTP/1.1
GET / HTTP/1.1
GET /favicon.ico HTTP/1.1
GET / HTTP/1.1
GET /favicon.ico HTTP/1.1
GET / HTTP/1.1
GET /favicon.ico HTTP/1.1
GET / HTTP/1.1
GET /favicon.ico HTTP/1.1
GET / HTTP/1.1
```

Figura N° 18: Interfaz de programación IDE.

Elaboración Propia.

b) Nivel de comunicaciones

En este nivel utilizamos la tecnología LTE (acrónimo de Long Term Evolution) para la comunicación, pues se utilizará un servidor en Internet para el almacenamiento y tratamiento de los datos. LTE es un estándar para comunicaciones inalámbricas de transmisión de datos de alta velocidad para teléfonos móviles y terminales de datos. El 3GPP está definida por unos como una evolución de la norma 3GPP UMTS (3G) y por otros como un nuevo concepto de arquitectura evolutiva (4G). LTE se destaca por su interfaz radioeléctrica basada en OFDMA, para el enlace descendente (DL) y SC-FDMA para el enlace ascendente (UL). La modulación elegida por el estándar 3GPP hace que las diferentes tecnologías de antenas (MIMO) tengan una mayor facilidad de implementación. La operadora Movistar (Telefónica del Perú) proporcionó el servicio a través del sistema AWS (Clase 4) con la frecuencia 1700/2100 MHz (bloque A) con 20MHz de ancho de banda.



Figura N° 19: Tecnología LTE utilizada.

Fuente: <https://www.con-cafe.com/2013/10/14/telefonica-del-per-despliega-lte-4g/>.

c) Nivel de procesamiento

Se utilizó MariaDB que es un sistema de gestión de bases de datos derivado de MySQL con licencia GPL (General Public License). Es desarrollado por Michael Widenius —fundador de MySQL—, la fundación MariaDB y la comunidad de desarrolladores de software libre. Introduce dos motores de almacenamiento nuevos, uno llamado Aria que reemplaza a MyISAM y otro llamado XtraDB en sustitución de InnoDB—. Tiene una alta compatibilidad con MySQL ya que posee las mismas órdenes, interfaces, API y bibliotecas, siendo su objetivo poder cambiar un servidor por otro directamente. Este SGBD surge a raíz de la compra de Sun Microsystems compañía que había comprado previamente MySQL por parte de Oracle.



Figura N° 20: Gestión de base de datos.

Fuente: <https://universo-digital.net/aumentar-tamano-redo-log-mysql-mariadb/>.

d) Nivel de aplicación

Para el entorno de desarrollo integrado se utilizó el IDE (sigla en inglés de integrated development environment), compuesto por un conjunto de herramientas de programación. Puede dedicarse en exclusiva a un solo lenguaje de programación o bien puede utilizarse para varios. Este IDE es un entorno de programación que ha sido empaquetado como un programa de aplicación, es decir, que consiste en un editor de código, un compilador, un depurador y un constructor de interfaz gráfica (GUI). Además en el caso de Arduino incorpora las herramientas para cargar el programa ya compilado en la memoria flash del hardware a través del puerto serie. Junto con el manejador de base de datos MariaDB proporcionará resultados que se muestran en la Figura N° 21.

UV Level	Output Voltage	UV Intensity	Max UV Intensity	Sensor ID	Client IP
275.00	1.34	2.72	6.19	S1AEP421PPP	179.7.226.16
293.00	1.57	4.57	9.39	S1CJCC31PPP	179.7.224.34
258.00	1.42	3.35	7.47	S1JJG167PPP	179.7.225.198
278.00	1.34	2.74	6.23	S1AEP421PPP	179.7.226.16
291.00	1.54	4.31	9.03	S1CJCC31PPP	179.7.224.34
257.00	1.41	3.31	7.4	S1JJG167PPP	179.7.225.198
275.00	1.34	2.75	6.25	S1AEP421PPP	179.7.226.16
290.00	1.53	4.27	8.97	S1CJCC31PPP	179.7.224.34
257.00	1.41	3.31	7.4	S1JJG167PPP	179.7.225.198
273.00	1.33	2.65	6.02	S1AEP421PPP	179.7.226.16
290.00	1.53	4.21	8.89	S1CJCC31PPP	179.7.224.34
256.00	1.41	3.26	7.31	S1JJG167PPP	179.7.225.198
275.00	1.32	2.61	5.93	S1AEP421PPP	179.7.226.16
290.00	1.55	4.41	9.17	S1CJCC31PPP	179.7.224.34
255.00	1.40	3.22	7.23	S1JJG167PPP	179.7.225.198
275.00	1.33	2.66	6.05	S1AEP421PPP	179.7.226.16
289.00	1.54	4.29	9	S1CJCC31PPP	179.7.224.34
255.00	1.40	3.22	7.23	S1JJG167PPP	179.7.225.198
271.00	1.32	2.58	5.86	S1AEP421PPP	179.7.226.16

Figura N° 21: Valores obtenidos de medición de la radiación UV.

Elaboración propia.

Acerca de la interfaz gráfica de usuario, también conocida como GUI (Graphical User Interface), es un programa que hace las veces de intermediario entre usuario y máquina, es decir es el software que muestra de forma visual todas las acciones posibles en una plataforma, así como la información disponible, para que los usuarios puedan interactuar con mayor facilidad y sin necesidad de disponer de profundos conocimientos de informática para nuestro caso la información recolectadas será visualizada en cualquier navegador, facilitando así la portabilidad del programa. Esto se muestra en la Figura N° 22.



Figura N° 22: Intensidad UV promedio.

Elaboración Propia.

El promedio es hallado en base a los resultados más recientes mostrados a continuación Figura N° 23.

Max UV Intensity
8.97
7.4
6.02
8.89
7.31
5.93
9.17

Figura N° 23: Interfaz gráfica de usuario.

Elaboración Propia.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1 RESULTADOS DEL DISEÑO E IMPLEMENTACIÓN DEL PROTOTIPO DE RECOLECCIÓN DE DATOS MEDIANTE UN SISTEMA IOT DE LA RADIACIÓN UV EN LA CIUDAD DE PUNO

La figura N° 24 muestra el módulo de adquisición de medidas de radiación UV utilizado, tanto en hardware como en el software utilizado (programación de la adquisición, transmisión de los datos).

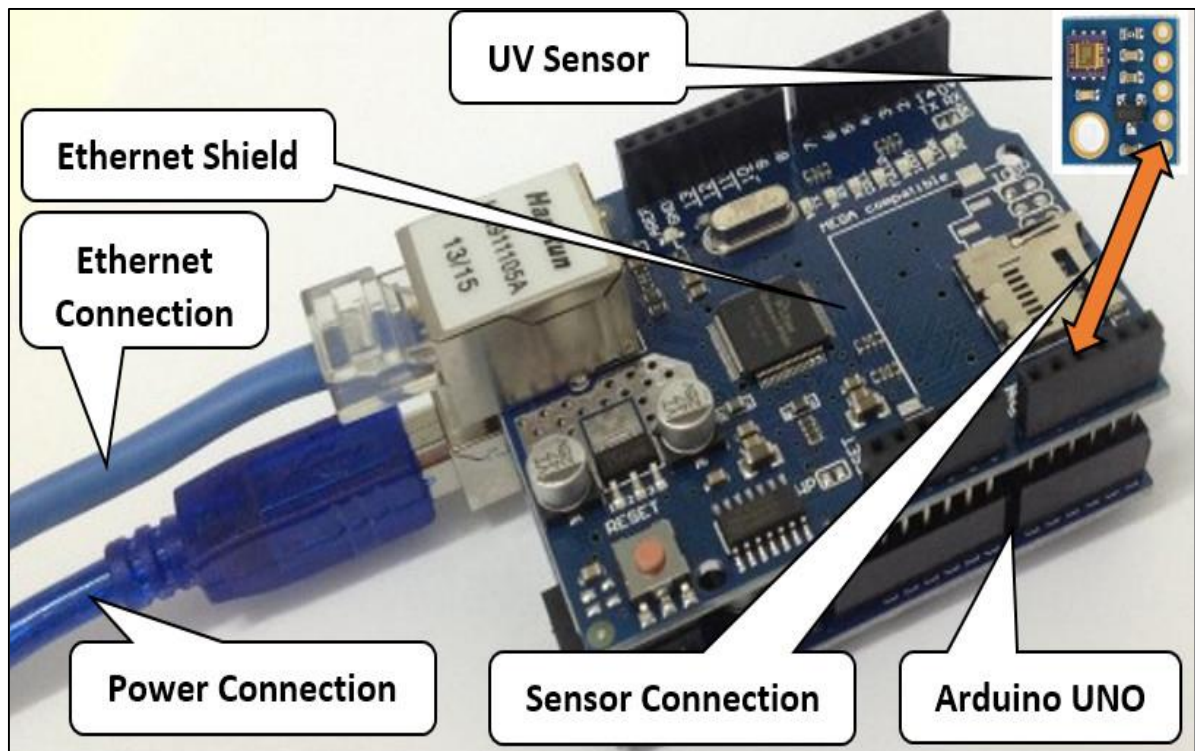


Figura N° 24: Monitoreo de radiación UV.

Elaboración propia.



A continuación se muestra parte de la data obtenida para realizar el modelo de regresión.

Tabla N° 4: Data obtenida mediante el monitoreo de radiación UV.

UV Level	Output Voltage	UV Intensity	Max UV Intensity	Sensor ID	Client IP	Date Time
275.00	1.34	2.72	6.19	S1AEP421PPP	179.7.226.16	2021-09-13. 10:36:01
293.00	1.57	4.57	9.39	S1CJCC31PPP	179.7.224.34	2021-09-13. 10:35:19
258.00	1.42	3.35	7.47	S1JJG167PPP	179.7.225.198	2021-09-13. 10:35:17
278.00	1.34	2.74	6.23	S1AEP421PPP	179.7.226.16	2021-09-13. 10:35:01
291.00	1.54	4.31	9.03	S1CJCC31PPP	179.7.224.34	2021-09-13. 10:34:18
257.00	1.41	3.31	7.4	S1JJG167PPP	179.7.225.198	2021-09-13. 10:34:16
275.00	1.34	2.5	6.25	S1AEP421PPP	179.7.226.16	2021-09-13. 10:34:00
290.00	1.53	4.27	8.97	S1CJCC31PPP	179.7.224.34	2021-09-13. 10:33:18
257.00	1.41	3.31	7.4	S1JJG167PPP	179.7.225.198	2021-09-13. 10:33:15
273.00	1.33	2.65	6.02	S1AEP421PPP	179.7.226.16	2021-09-13. 10:32:59
290.00	1.53	4.21	8.89	S1CJCC31PPP	179.7.224.34	2021-09-13. 10:32:17
256.00	1.41	3.26	7.31	S1JJG167PPP	179.7.225.198	2021-09-13. 10:32:15
275.00	1.32	2.61	5.93	S1AEP421PPP	179.7.226.16	2021-09-13. 10:31:59
290.00	1.55	4.41	9.17	S1CJCC31PPP	179.7.224.34	2021-09-13. 10:31:16
255.00	1.4	3.22	7.23	S1JJG167PPP	179.7.225.198	2021-09-13. 10:31:14
275.00	1.33	2.66	6.05	S1AEP421PPP	179.7.226.16	2021-09-13. 10:30:58
289.00	1.54	4.29	9	S1CJCC31PPP	179.7.224.34	2021-09-13. 10:30:16
255.00	1.4	3.22	7.23	S1JJG167PPP	179.7.225.198	2021-09-13. 10:30:13
271.00	1.32	2.58	5.86	S1AEP421PPP	179.7.226.16	2021-09-13. 10:29:57

Elaboración propia.

En resumen podemos mencionar acerca de la data obtenida, se realizó una etapa de pre procesamiento que se puede resumir en los siguientes tres puntos:

- Datos analizados: 01/09/2021 - 31/09/2021
- Datos recolectados: $14880 \times 3 = 44640$
- Datos procesados (se eliminaron filas): $13346 \times 3 = 40038$

4.2 RESULTADOS PARA EL MODELO DE REGRESIÓN MULTÍPARAMÉTRICA PARA LOS DATOS RECOLECTADOS MEDIANTE EL SISTEMA IOT EN LA CIUDAD DE PUNO

Con la finalidad de cumplir el objetivo de modelar mediante una regresión multiparamétrica, se realizaron tres modelos distintos:

- Ordinary Least Squares (OLS)
- La regresión Ridge o regularización L2
- Por último se presenta un modelo denominado Bayesian Ridge.

Los valores obtenidos mediante la regresión para los tres modelos utilizando Python como lenguaje de programación, se muestran en la Tabla N° 5:

Tabla N° 5: Parámetros obtenidos de los tres modelos de regresión multiparamétrica.

Parámetro	Modelo		
	OLS	Ridge	Bayesian Ridge
Error absoluto medio	0.01975674	0.01967830	0.019756736
Error cuadrático medio	0.02288033	0.02306340	0.022880322
Coefficiente de determinación	0.99978077	0.99977724	0.999780766
Coefficiente de determinación ajustado	0.99978072	0.99977719	0.999780717
Tiempo de entrenamiento	0.03226900	0.00930700	0.009166
Tiempo de test	0.00129900	0.00126900	0.001965

Elaboración propia.

A continuación se muestran gráficos comparativos de los modelos propuestos así como la discusión de resultados para cada uno de los siguientes parámetros: precisión, error absoluto medio, error cuadrático medio, coeficiente de determinación, coeficiente de determinación ajustado, tiempo de entrenamiento y tiempo de test o prueba:

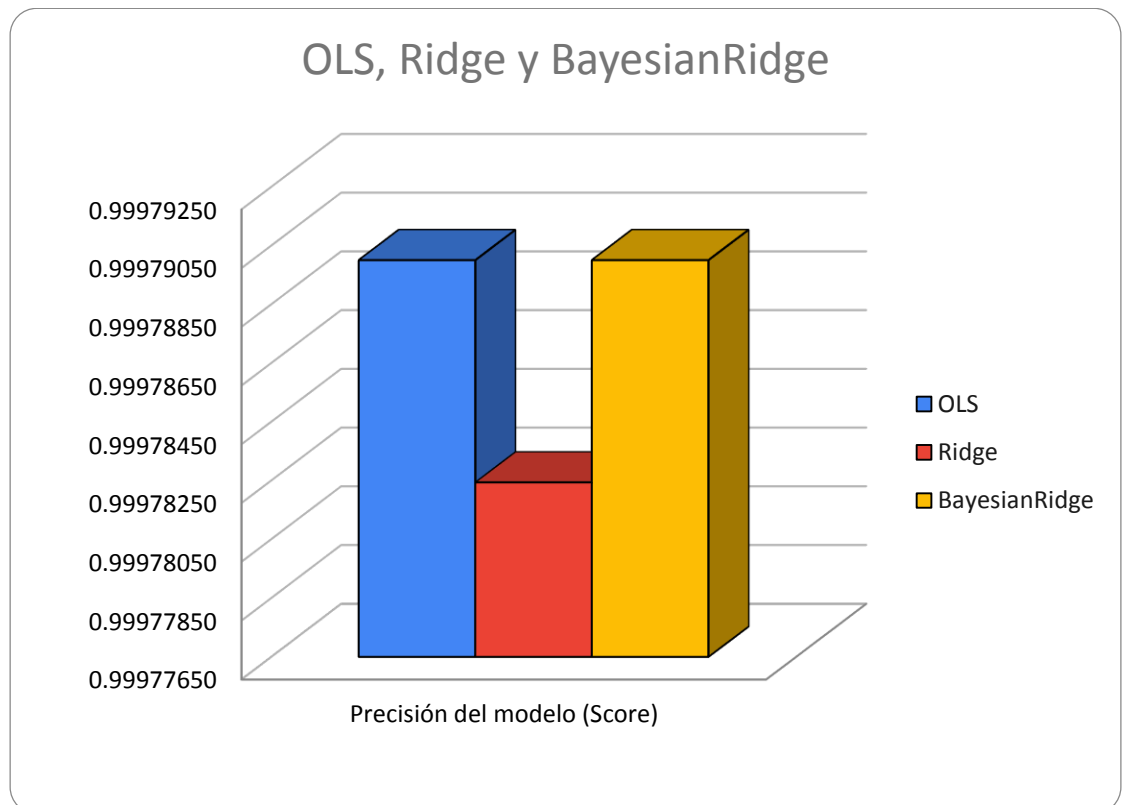


Figura N° 25: Precisión de los modelos realizados.

Elaboración propia.

Con la métrica de precisión podemos medir la calidad del modelo de Machine Learning, llamada también valor predictivo positivo, es la fracción de instancias relevantes entre las instancias recuperadas. De la Figura N° 25 notamos que OLS y Bayesian Ridge tienen una mayor precisión: 0.99979250 mientras que Ridge tiene una precisión de 0.99978250.

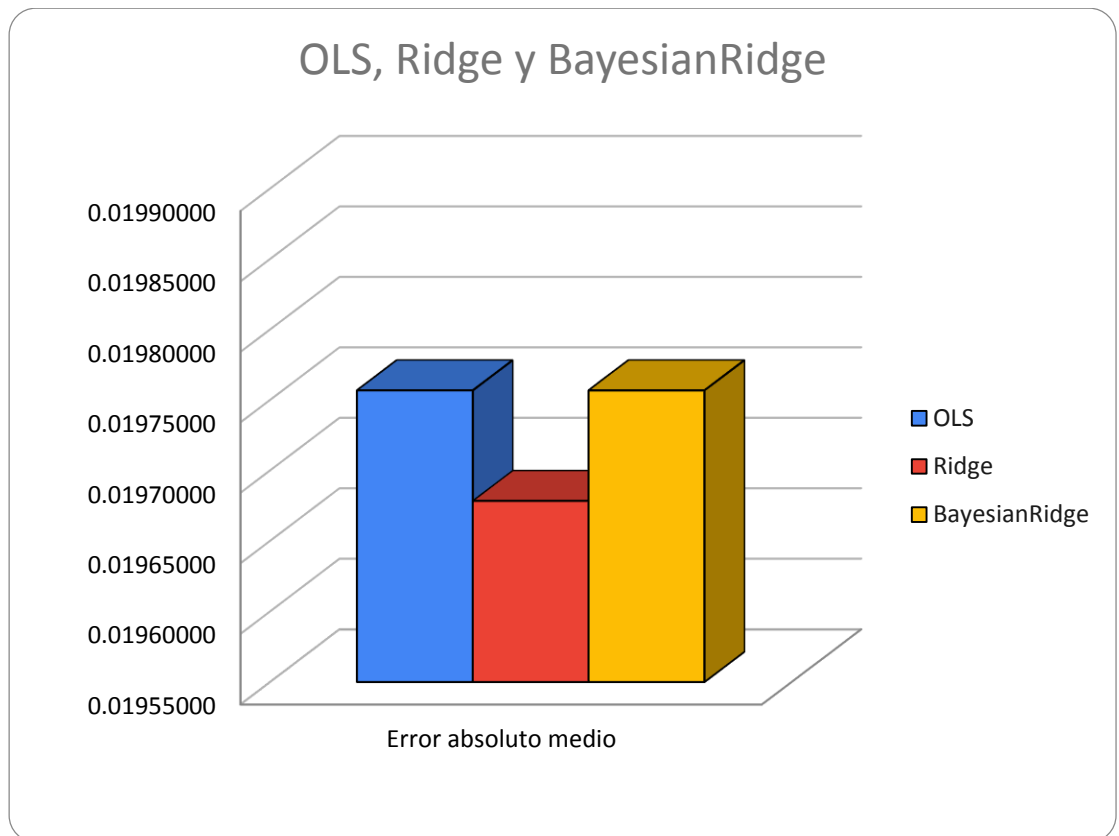


Figura N° 26: Error absoluto medio de los modelos realizados.

Elaboración propia.

El error absoluto medio o MAE se calcula como un promedio de diferencias absolutas entre los valores objetivo y las predicciones. El MAE es una puntuación lineal, lo que significa que todas las diferencias individuales se ponderan por igual en el promedio. Lo importante de esta métrica es que penaliza errores enormes que no tan mal como lo hace el error cuadrático medio. Por lo tanto, no es tan sensible a los valores atípicos como el error cuadrático medio. De la Figura N° 26 notamos que Ridge es el que tiene menor MAE 0.01965, mientras que OLS y Bayesian Ridge tienen un MAE de 0.0198.

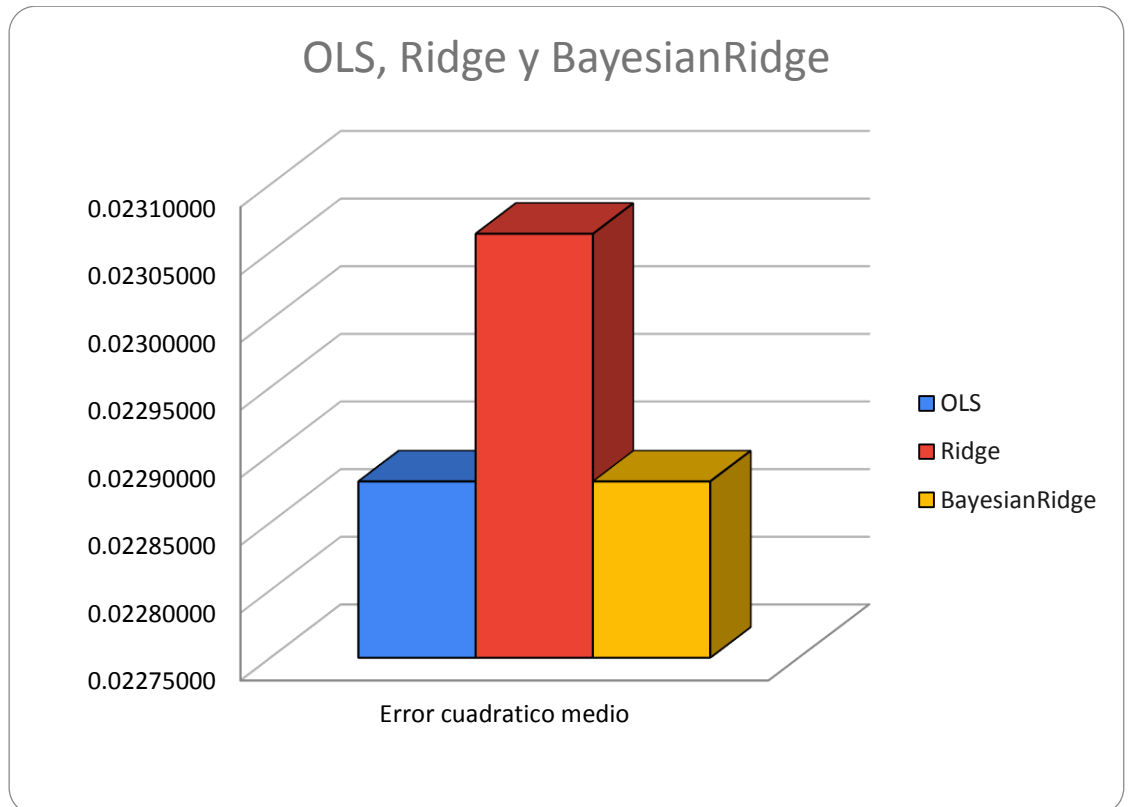


Figura N° 27: Error cuadrático medio de los modelos realizados.

Elaboración propia.

El error cuadrático medio o RMSE es solo la raíz cuadrada de MSE. La raíz cuadrada se introduce para hacer que la escala de los errores sea igual a la escala de los objetivos. Indica el ajuste absoluto del modelo a los datos, cuán cerca están los puntos de datos observados de los valores predichos del modelo. El error cuadrático medio o RMSE es una medida absoluta de ajuste. De la Figura N° 27 notamos que Ridge es la técnica que tiene el mayor valor con 0.02306, mientras que OLS y Bayesian Ridge tienen valores de 0.02288.

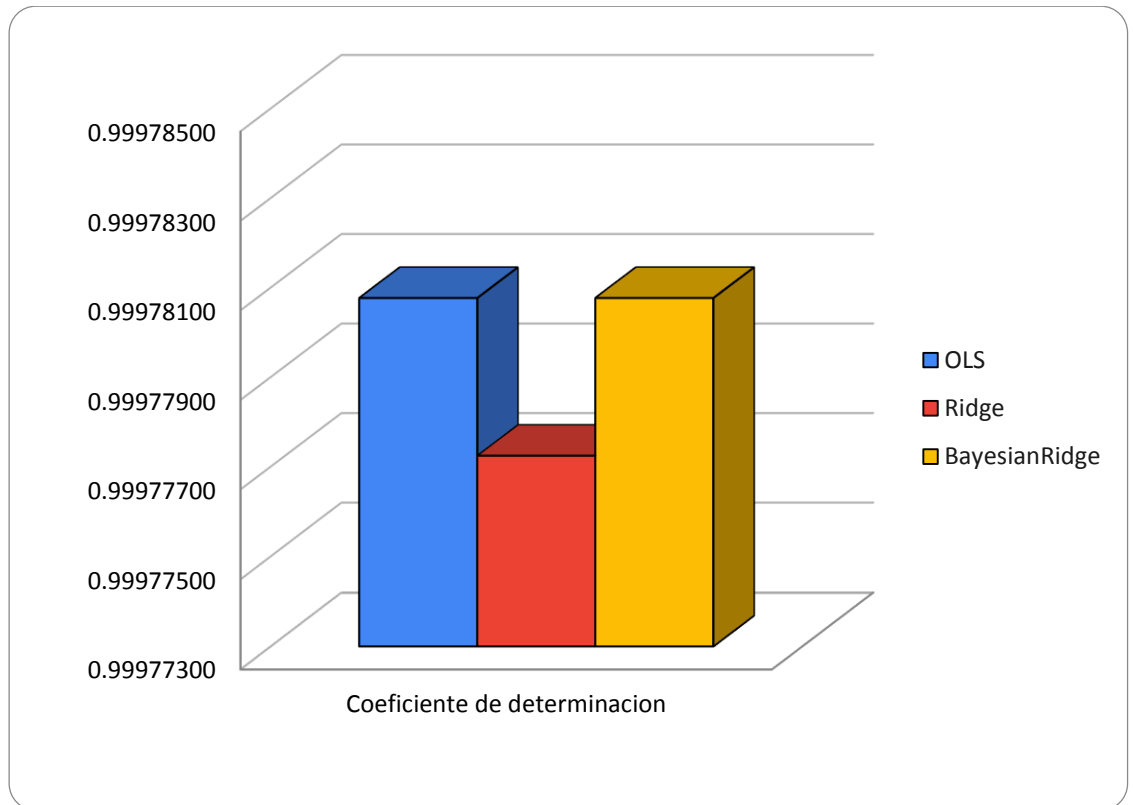


Figura N° 28: Coeficiente de determinación de los modelos realizados.

Elaboración propia.

El coeficiente de determinación se puede entender como una versión estandarizada del MSE, que proporciona una mejor interpretación del rendimiento del modelo. Técnicamente, el R^2 representa la varianza de las respuestas capturadas por el modelo, pero tiene la ventaja de estar libre de escala, no importa si los valores de salida son muy grandes o muy pequeños, el R^2 siempre estará entre $-\infty$ y 1. De la Figura N° 28 notamos que OLS y Bayesian Ridge tienen un mayor valor 0.99978, mientras que Ridge tiene un valor un poco menor 0.999777.

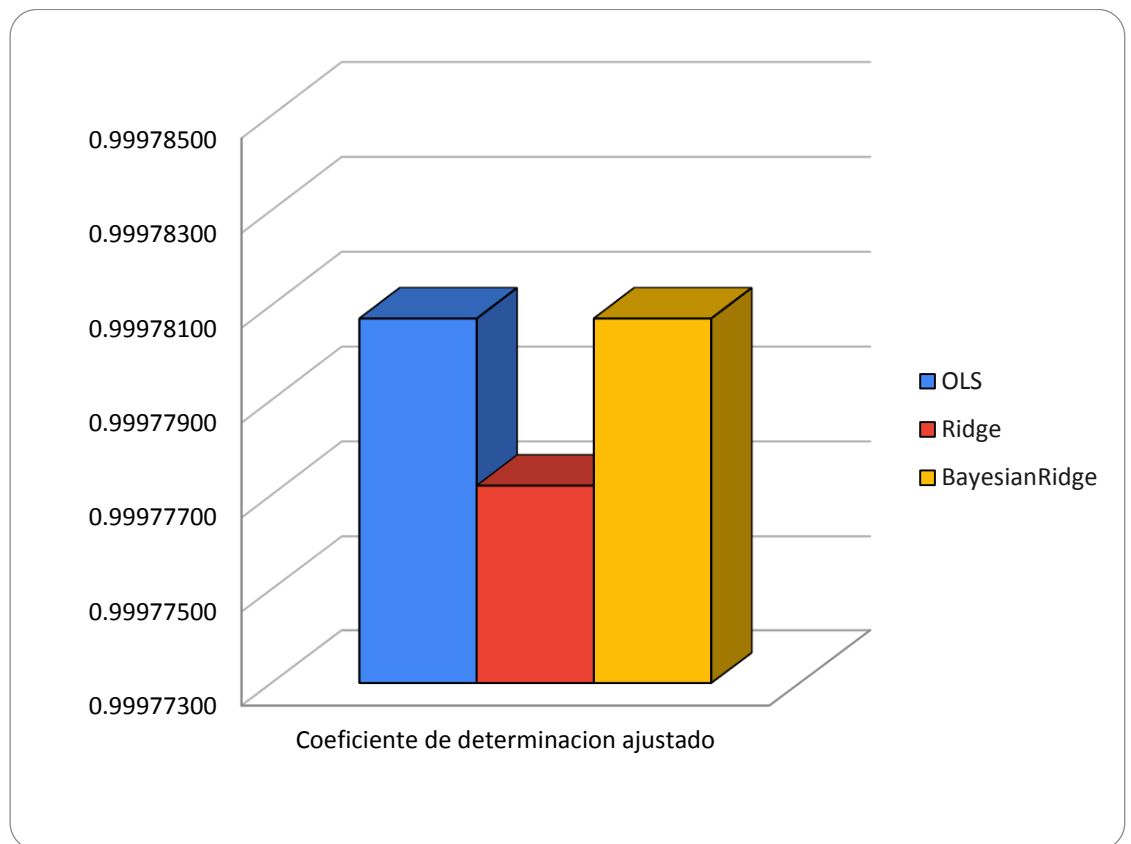


Figura N° 29: Coeficiente de determinación ajustado de los modelos realizados.
Elaboración propia.

El R2 cuadrado ajustado (o coeficiente de determinación ajustado) se utiliza en la regresión múltiple para ver el grado de intensidad o efectividad que tienen las variables independientes en explicar la variable dependiente, es decir, se puede juzgar si los datos en la ecuación de regresión se ajustan bien. A mayor R2 ajustada, mejor es la ecuación de regresión, ya que implica que la variable independiente elegida para determinar la variable dependiente puede explicar la variación en la variable dependiente. De la Figura N° 29 notamos que tanto OLS como Bayesian Ridge tienen un mejor índice de R2 ajustado 0.99978, mientras que Ridge llega a 0.99977.

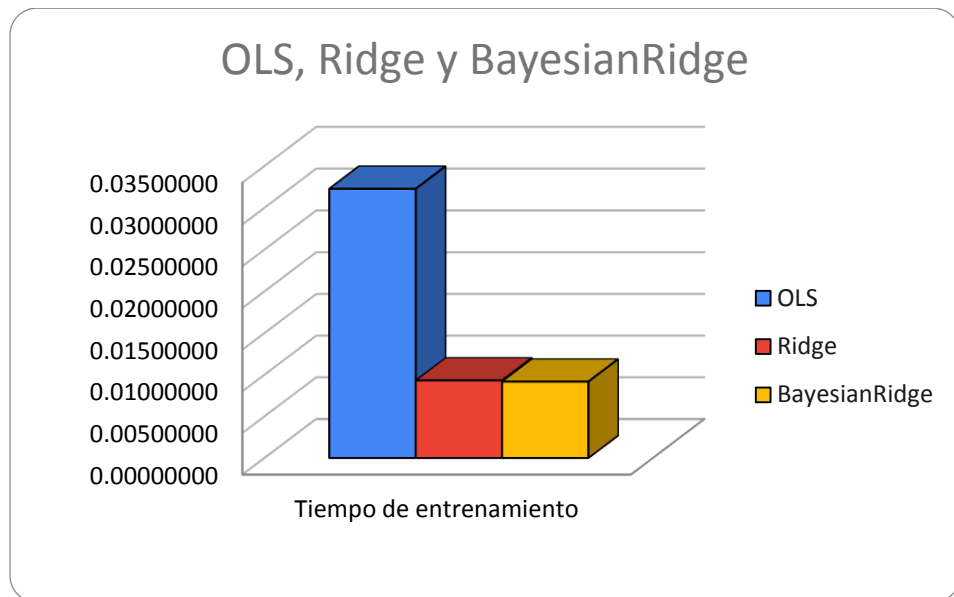


Figura N° 30: Tiempo de entrenamiento de los modelos realizados.

Elaboración propia.

El tiempo de entrenamiento es el tiempo que se da en la fase en el que el sistema trata de aprender tendencias, comportamientos y/o patrones que se ajusten a los datos que forman el conjunto de entrenamiento. Desde un punto de vista matemático, el entrenamiento de un sistema inteligente se corresponde con un proceso de optimización de los parámetros de una función para que su salida sea lo más parecida posible al resultado que queremos obtener. De la Figura N° 30 notamos que el mayor tiempo de entrenamiento lo tiene OLS con 0.032269, mientras que Ridge y Bayesian Ridge tiene un tiempo de entrenamiento de 0.00930 0.00916 respectivamente.

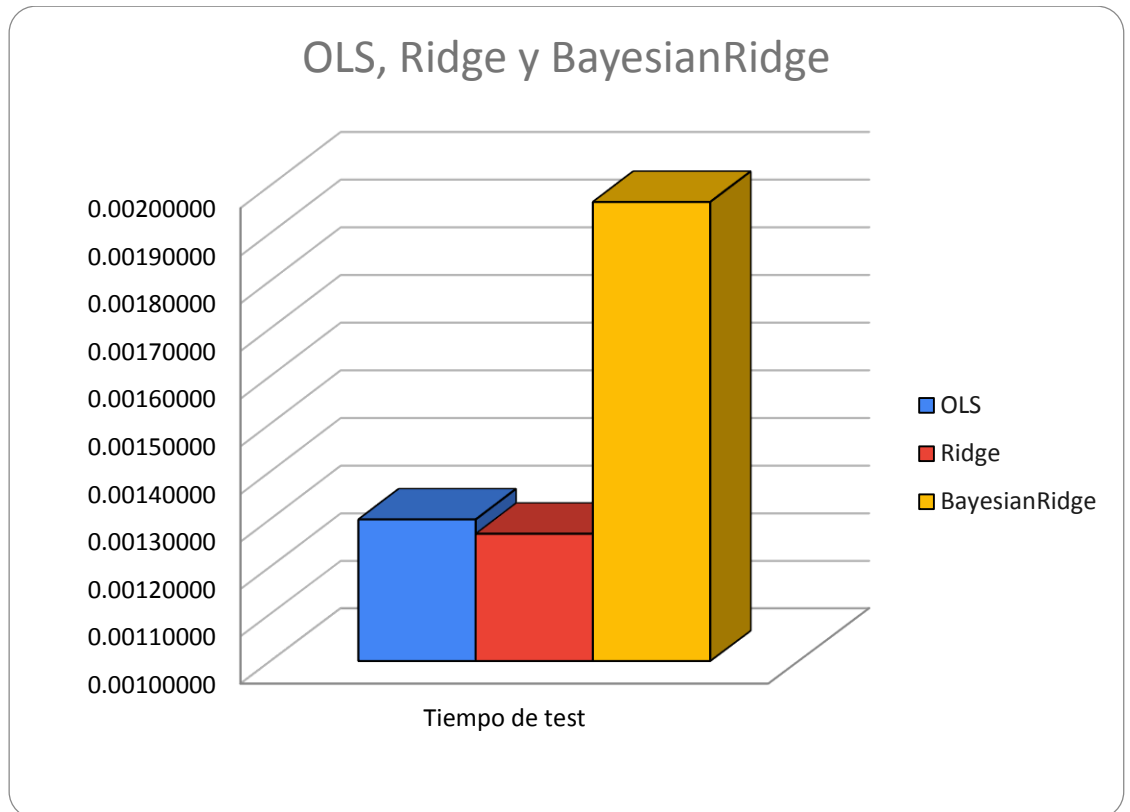


Figura N° 31: Tiempo de test o prueba de los modelos realizados.

Elaboración propia.

Por otro lado el tiempo de test o prueba que es la etapa de predicción que se realiza una vez se ha completado el entrenamiento y consiste en evaluar datos nuevos que no se han tenido en cuenta para el entrenamiento utilizando la función optimizada para obtener un resultado. De la Figura N° 31 el mayor tiempo de test o prueba lo tiene Bayesian Ridge con 0.001965, mientras que OLS y Ridge tienen como tiempos 0.001299 y 0.001269 respectivamente.



V. CONCLUSIONES

PRIMERA: Se logró la recolección de los datos de radiación UV en la ciudad de Puno mediante un sistema IoT cumpliendo con el modelo de cuatro capas del modelo de referencia: adquisición de datos, comunicaciones, procesamiento y aplicación, utilizando en lo posible software no licenciado.

SEGUNDA: Se realizó el análisis de los datos recolectados empleando tres modelos de regresión multiparámetro para la monitorización de la radiación UV en la ciudad de Puno: OLS, Ridge y Bayesian Ridge, teniendo como resultados más resaltantes un error cuadrático medio para OLS de 0.02288, para Ridge de 0.0230 y para Bayesian Ridge 0.02288. El coeficiente de determinación ajustada para OLS de 0.9997, para Ridge de 0.9997 ya para Bayesina Ridge de 0.997. Planteando como el mejor modelo de regresión el Bayesian Ridge.



VI. RECOMENDACIONES

PRIMERA: Se puede mejorar la calidad de los datos obtenidos mediante técnicas de imputación u otros métodos matemáticos, para corregir valores erróneos o valores faltantes, ocasionados por los dispositivos de recolección de datos, dispositivos de transmisión y almacenamiento.

SEFUNDA: Para mejorar la calidad de las predicciones se puede incrementar el número de datos mediante la instalación de nuevos puntos de medida en la ciudad de Puno, así como la instalación de nuevos sensores para la medida de otras variables ambientales tales como temperatura.

TERCERA: Incrementado el presupuesto se pueden adquirir nuevos módulos de sensores para modelar nuevas variables atmosféricas tales como: temperatura, humedad, índices de contaminación, etc.



VII. REFERENCIAS BIBLIOGRÁFICAS

Antón, M. (2007). Modelos Empíricos para la Estimación de la Irradiancia Solar Ultravioleta. [Tesis para optar el grado de Doctor] Universidad de Extremadura.

Amazon Web Services. (2019). AWS IoT Analytics Overview. Recuperado el 9 de junio de 2019, de <https://aws.amazon.com/iot-analytics/>

Apache Software Foundation. (2019). Apache Spark™ - Unified Analytics Engine for Big Data. Recuperado el 9 de junio de 2019, de <https://spark.apache.org/>

Bassi, A., Bauer, M., Fiedler, M., & Kramp, T. (2013). Enabling Things to Talk: Designing IoT solutions with the IoT Architectural Reference Model. Guilford: Springer Open.

Blokdyk, G. (2018). *NB-IoT A Complete Guide*. Estados Unidos: 5STARCOOKS.

Bormann, C. (2019). CoAP — Constrained Application Protocol | Overview. Recuperado el 9 de junio de 2019, de <https://coap.technology/>

Cambustion Ltd. (2019). Nondispersive Infrared Detector (NDIR) principle | Cambustion. Recuperado el 3 de junio de 2019, de <https://www.cambustion.com/products/ndir500/operating-principle>

Cañizares, B. (2019). Implementación de un Sistema de Control y Monitoreo para la Radiación Ultravioleta en la Universidad Técnica de Cotopaxi en el Periodo Marzo 2019



– Agosto 2019. [Tesis para optar el título de Ingeniería Eléctrica]. Universidad Técnica de Cotopaxi

<https://www.goconqr.com/es/p/6252265>

Ccora, O. (2015). Radiación ultravioleta B, vulnerabilidad y riesgos en el verano. Lima: SENAMHI.

Chou, T. (2016). Precision: Principles, Practices and Solutions for the Internet of Things. Estados Unidos: CrowdBook.

Cruz, J. et al. (2021). Selection of Characteristics by Hybrid Method: RFE, Ridge, Lasso, and Bayesian for the Power Forecast for a Photovoltaic System. International Conference on Soft Computing and its Engineering Applications icSoftComp 2020: Soft Computing and its Engineering Applications pp 75-87

Elasticsearch B.V. (2019). Open Source Search & Analytics • Elasticsearch | Elastic. Recuperado el 9 de junio de 2019, de <https://www.elastic.co/>

Forero, G. (2020). Modelo de Regresión Lineal Múltiple para el Pronóstico de Ventas de Bolsas Ecológicas para la Empresa Boleco SA, en la Ciudad de Bogotá DC. [Tesis para optar el título de Ingeniero Industrial]. Universidad Cooperativa de Colombia.

Gonzalez L. (2001) Los peligros de la radiación solar. Fotoprotección. Offarm Vol. 20. Núm. 5. Páginas 75-84



Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. (2010). Metodología de la Investigación. México D.F.: McGraw-Hill / INTERAMERICANA EDITORES, S.A. DE C.V.

Herrera, A. & Ojeda, R. (2019). Diseño de un Prototipo para el Monitoreo de Radiación UV Ambiental en el Municipio de Popayán. [Tesis para optar el título de Ingeniero de Sistemas]Fundación Universitaria de Popayan

Herrera, F. Ojeda, R. (2019) Diseño de un Prototipo para el Monitoreo de Radiación UV Ambiental en el Municipio De Popayán. [Tesis para optar el título profesional de Ingeniero de Sistemas] Fundación Universitaria de Popayán

Huaraya, F. (2018). Caracterización y Modelización de la Radiación Ultravioleta Solar Aplicado en Upina, Puno- 2014. [Tesis para optar el título profesional de Licenciado en Ciencias Físico Matemáticas] Universidad Nacional del Altiplano

Huillca, M., & Taipe, C. (2015). Modelo paramétrico de la radiación ultravioleta global para cielos claros en la ciudad de Puno. Puno: Rev. Investigation Altoandina.

International Telecommunication Union. (2012). Overview of the Internet of things. Recuperado de https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.2060-201206-I!!PDF-E&type=items

Internet Engineering Task Force. (2019). The Internet of things. Recuperado el 2 de junio de 2019, de <https://www.ietf.org/topics/iot/>



Ismail, I. S., Abdul Latiff, N. A., Rokhani, F. Z., & Abdul Aziz, S. (2019). A Review on Performances Evaluation of Low Power Wide Area Networks Technology (Vol. 547, pp. 343–349). Springer Singapore.

https://doi.org/10.1007/978-981-13-6447-1_43

json.org. (2019). JSON. Recuperado el 21 de febrero de 2019, de <https://www.json.org/>

Kapoor, D. (2020). Machine learning 101-Linear Regression using the OLS Method. Recuperado el 14 de octubre de 2020, de <https://medium.com/analytics-vidhya/machine-learning-101-linear-regression-using-the-ols-method-299808eab233>

Kurniawan, A. (2016). Making Your IoT Project Smart. En Smart Internet of Things Projects (1a ed., pp. 10–18). Packt Publishing Ltd.

Lea, P. (2018). Internet of Things for Architects. Packt Publishing Ltd. (1a Ed.). Birmingham: Packt Publishing Ltd.

Lucas R, McMichael T, Smith W, Armstrong B. (2006). Solar ultraviolet radiation: global burden of disease from solar ultraviolet radiation. Ginebra: Organización Mundial de la Salud; 2006.

Mayhua, F. (2019). Estudio Comparativo y Modelación de la Radiación Global y Ultravioleta en la Ciudad de Arequipa. [Tesis para optar el título profesional de Licenciado en Física. Universidad Nacional de San Agustín.



Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing (Recommendations of the National Institute of Standards and Technology No. NIST Special Publication 800-145). Special Publication 800-145. Recuperado de <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf>

Minerva, R., Biru, A., & Rotondi, D. (2015). Towards a definition of the Internet of Things (IoT). IEEE Internet of Things, 1, 72–75. Recuperado de https://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_Internet_of_Things_Revision1_27MAY15.pdf

Ministerio de Salud del Perú. (2020). Análisis de la situación del Cáncer en el Perú, 2018. Lima: Ministerio de Salud.

mqtt.org. (2019). FAQ - Frequently Asked Questions | MQTT. Recuperado el 21 de febrero de 2019, de <http://mqtt.org/faq>

NCSS 2021. (2021) Data Analysis & Graphics.

<https://www.ncss.com/software/ncss/ncss-documentation/>

Nicosia G. et al. (2020). Machine Learning, Optimization, and Data Science. 6th

International Conference, LOD 2020, Siena, Italy, July 19–23, 2020 pp 191-202

Perez, M. (2014). Diseño de un radiómetro ultravioleta, para su aplicación en modelos de



radiación uv. México DF, UNAM.

Robles, R. Camayo, C. (2019). Diseño e implementación de un sistema de riego basado en internet de las cosas aplicando tecnologías LPWAN en la irrigación Majes-Siguas I - El Pedregal – Arequipa. [Tesis para optar el título profesional de Ingeniero Electrónico] Universidad Nacional del Altiplano

Serrano, A. (2007). Estudio de Irradiancia de bandas del espectro solar del espectro solar, factores de banda para la caracterización de condiciones atmosféricas en Valencia. Valencia: Tesis doctoral.

Toward data Science. (2019). Understanding the OLS method for Simple Linear Regression. Recuperado el 23 de diciembre de 2019, de <https://towardsdatascience.com/understanding-the-ols-method-for-simple-linear-regression-e0a4e8f692cc>

Ttacca E. (2017). Diseño de una red FOG basado en internet de las cosas para monitorear la contaminación en la bahía del Lago Titicaca. [Tesis para optar el título profesional de Ingeniero Electrónico] Universidad Nacional del Altiplano

Wright, J. (2008). Medición y predicción de la radiación solar global ultravioleta (295 nm - 385 nm). Heredia - Costa Rica: UNICIENCIA.

Evans. (2011). Internet de las cosas “Como la próxima evolución de internet lo cambia todo”



Rose, Eldridge & Chapin. (2015). La internet de las cosas-una breve reseña “Para entender mejor los problemas y desafíos de un mundo más conectado”

Núñez & Páez (2016). La importancia de los actuadores.

IoT & Cloud Computing integration. – ResearchGate.

Aguilar (2020). Tesis: “Diseño de una solución basada en el internet de las cosas (iot) empleando lorawan para el monitoreo de cultivos agrícolas en Perú”

Efor (2017). Tecnologías de comunicación para iot.



ANEXOS

Anexo N° 1: Monitoreo de las variables

COM14
20:40:44.772 -> Rx T = 9.00 H = 79.00
20:40:44.772 -> P.A. = 64814 UV = -1.02
20:40:44.772 -> Vel. V = 0.00
20:40:44.772 -> RSSI -86
20:40:58.028 -> Rx T = 9.00 H = 79.00
20:40:58.028 -> P.A. = 64813 UV = -0.95
20:40:58.028 -> Vel. V = 0.00
20:40:58.028 -> RSSI -86
20:41:11.280 -> Rx T = 9.00 H = 79.00
20:41:11.362 -> P.A. = 64818 UV = -1.03
20:41:11.362 -> Vel. V = 0.00
20:41:11.362 -> RSSI -87
20:41:24.562 -> Rx T = 9.00 H = 79.00
20:41:24.562 -> P.A. = 64812 UV = -1.02
20:41:24.562 -> Vel. V = 0.00
20:41:24.562 -> RSSI -88
20:41:37.844 -> Rx T = 9.00 H = 79.00
20:41:37.844 -> P.A. = 64816 UV = -1.06
20:41:37.844 -> Vel. V = 0.00
20:41:37.844 -> RSSI -86
20:41:51.102 -> Rx T = 9.00 H = 79.00
20:41:51.102 -> P.A. = 64810 UV = -1.05
20:41:51.102 -> Vel. V = 0.00
20:41:51.102 -> RSSI -87
20:42:04.388 -> Rx T = 9.00 H = 79.00
20:42:04.388 -> P.A. = 64819 UV = -1.03
20:42:04.388 -> Vel. V = 0.00
20:42:04.388 -> RSSI -83
20:42:17.630 -> Rx T = 9.00 H = 80.00
20:42:17.665 -> P.A. = 64820 UV = -1.05
20:42:17.665 -> Vel. V = 0.00
20:42:17.665 -> RSSI -82
20:42:30.918 -> Rx T = 9.00 H = 80.00
20:42:30.918 -> P.A. = 64819 UV = -1.02

Elaboración propia.



Anexo N° 2: Muestra de datos obtenidos – Agosto

N	ID	UV- Estático	UV-Móvil	IP	Ubicación	Fecha y Hora
464490	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:00
464491	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:00
464492	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:01
464493	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:01
464494	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:02
464495	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:02
464496	199	1.13	1.07	179.7.224.209	S1JJG167PPP	01/08/2021 00:03
464497	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:03
464498	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:04
464499	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:04
464500	200	1.13	1.1	179.7.224.209	S1JJG167PPP	01/08/2021 00:05
464501	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:05
464502	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:06
464503	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:06
464504	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:07
464505	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:07
464506	199	1.13	1.07	179.7.224.209	S1JJG167PPP	01/08/2021 00:08
464507	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:08
464508	200	1.13	1.12	179.7.224.209	S1JJG167PPP	01/08/2021 00:09
464509	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:09
464510	200	1.13	1.1	179.7.224.209	S1JJG167PPP	01/08/2021 00:10
464511	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:10
464512	200	1.13	1.1	179.7.224.209	S1JJG167PPP	01/08/2021 00:11
464513	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:11
464514	200	1.13	1.1	179.7.224.209	S1JJG167PPP	01/08/2021 00:12
464515	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:12
464516	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:13
464517	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:13
464518	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:14
464519	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:14
464520	199	1.12	1.06	179.7.224.209	S1JJG167PPP	01/08/2021 00:15
464521	207	1.13	1.06	179.7.224.34	S1CJCC31PPP	01/08/2021 00:15

Elaboración propia.



Anexo N° 3: Muestra de datos obtenidos – Setiembre

N	ID	UV- Estático	UV-Móvil	IP	Ubicación	Fecha y Hora
561423	207	1	0.09	179.7.226.16	S1AEP421PPP	01/09/2021 00:01
561424	209	1	0.06	179.7.226.16	S1AEP421PPP	01/09/2021 00:02
561425	208	1	0.09	179.7.226.16	S1AEP421PPP	01/09/2021 00:03
561426	207	1	0.09	179.7.226.16	S1AEP421PPP	01/09/2021 00:04
561427	207	1	0.05	179.7.226.16	S1AEP421PPP	01/09/2021 00:05
561428	207	1	0.08	179.7.226.16	S1AEP421PPP	01/09/2021 00:06
561429	206	0.99	0.02	179.7.226.16	S1AEP421PPP	01/09/2021 00:07
561430	206	1	0.04	179.7.226.16	S1AEP421PPP	01/09/2021 00:08
561431	207	1	0.07	179.7.226.16	S1AEP421PPP	01/09/2021 00:09
561432	209	1.01	0.12	179.7.226.16	S1AEP421PPP	01/09/2021 00:10
561433	206	0.99	0.02	179.7.226.16	S1AEP421PPP	01/09/2021 00:11
561434	207	1	0.07	179.7.226.16	S1AEP421PPP	01/09/2021 00:12
561435	211	1.01	0.17	179.7.226.16	S1AEP421PPP	01/09/2021 00:13
561436	208	1	0.08	179.7.226.16	S1AEP421PPP	01/09/2021 00:14
561437	206	1	0.04	179.7.226.16	S1AEP421PPP	01/09/2021 00:15
561438	206	1	0.06	179.7.226.16	S1AEP421PPP	01/09/2021 00:17
561439	209	1.01	0.13	179.7.226.16	S1AEP421PPP	01/09/2021 00:18
561440	208	1	0.11	179.7.226.16	S1AEP421PPP	01/09/2021 00:19
561441	206	1	0.08	179.7.226.16	S1AEP421PPP	01/09/2021 00:20
561442	207	0.99	0.03	179.7.226.16	S1AEP421PPP	01/09/2021 00:21
561443	209	1	0.09	179.7.226.16	S1AEP421PPP	01/09/2021 00:22
561444	209	1	0.09	179.7.226.16	S1AEP421PPP	01/09/2021 00:23
561445	207	0.99	0.03	179.7.226.16	S1AEP421PPP	01/09/2021 00:24
561446	208	1	0.08	179.7.226.16	S1AEP421PPP	01/09/2021 00:25
561447	208	0.99	0.01	179.7.226.16	S1AEP421PPP	01/09/2021 00:26
561448	209	1	0.1	179.7.226.16	S1AEP421PPP	01/09/2021 00:28
561449	207	0.99	0.02	179.7.226.16	S1AEP421PPP	01/09/2021 00:29
561450	206	1	0.06	179.7.226.16	S1AEP421PPP	01/09/2021 00:30
561451	211	1.01	0.16	179.7.226.16	S1AEP421PPP	01/09/2021 00:31
561452	208	1	0.12	179.7.226.16	S1AEP421PPP	01/09/2021 00:32
561453	209	1	0.11	179.7.226.16	S1AEP421PPP	01/09/2021 00:33
561454	207	0.99	0	179.7.226.16	S1AEP421PPP	01/09/2021 00:34

Elaboración propia.



Anexo N° 4: Muestra de datos obtenidos – Octubre

N	ID	UV- Estático	UV-Móvil	IP	Ubicación	Fecha y Hora
656832	208	1.01	0.13	179.7.225.35	S1AEP421PPP	01/10/2021 00:00
656833	207	1.12	1.03	179.7.224.49	S1CJCC31PPP	01/10/2021 00:00
656834	199	1.12	1.04	179.7.224.209	S1JJG167PPP	01/10/2021 00:00
656835	208	1	0.11	179.7.225.35	S1AEP421PPP	01/10/2021 00:01
656836	207	1.12	1.03	179.7.224.49	S1CJCC31PPP	01/10/2021 00:01
656837	199	1.12	1.03	179.7.224.209	S1JJG167PPP	01/10/2021 00:01
656838	207	1.12	1.05	179.7.224.49	S1CJCC31PPP	01/10/2021 00:02
656839	199	1.12	1.03	179.7.224.209	S1JJG167PPP	01/10/2021 00:02
656840	206	1	0.08	179.7.225.35	S1AEP421PPP	01/10/2021 00:03
656841	207	1.12	1.03	179.7.224.49	S1CJCC31PPP	01/10/2021 00:03
656842	199	1.12	1.04	179.7.224.209	S1JJG167PPP	01/10/2021 00:03
656843	207	1	0.08	179.7.225.35	S1AEP421PPP	01/10/2021 00:04
656844	207	1.12	1.05	179.7.224.49	S1CJCC31PPP	01/10/2021 00:04
656845	199	1.12	1.03	179.7.224.209	S1JJG167PPP	01/10/2021 00:04
656846	206	1	0.04	179.7.225.35	S1AEP421PPP	01/10/2021 00:05
656847	207	1.12	1.05	179.7.224.49	S1CJCC31PPP	01/10/2021 00:05
656848	199	1.12	1.03	179.7.224.209	S1JJG167PPP	01/10/2021 00:05
656849	205	0.99	0.04	179.7.225.35	S1AEP421PPP	01/10/2021 00:06
656850	207	1.12	1.05	179.7.224.49	S1CJCC31PPP	01/10/2021 00:06
656851	200	1.13	1.07	179.7.224.209	S1JJG167PPP	01/10/2021 00:06
656852	208	1	0.06	179.7.225.35	S1AEP421PPP	01/10/2021 00:07
656853	207	1.12	1.03	179.7.224.49	S1CJCC31PPP	01/10/2021 00:07
656854	199	1.12	1.04	179.7.224.209	S1JJG167PPP	01/10/2021 00:07
656855	206	1	0.06	179.7.225.35	S1AEP421PPP	01/10/2021 00:08
656856	207	1.12	1.05	179.7.224.49	S1CJCC31PPP	01/10/2021 00:08
656857	200	1.13	1.07	179.7.224.209	S1JJG167PPP	01/10/2021 00:08
656858	207	0.99	0.02	179.7.225.35	S1AEP421PPP	01/10/2021 00:09
656859	207	1.12	1.05	179.7.224.49	S1CJCC31PPP	01/10/2021 00:09
656860	199	1.12	1.03	179.7.224.209	S1JJG167PPP	01/10/2021 00:09
656861	207	0.99	0.03	179.7.225.35	S1AEP421PPP	01/10/2021 00:10
656862	207	1.12	1.05	179.7.224.49	S1CJCC31PPP	01/10/2021 00:10
656863	200	1.13	1.07	179.7.224.209	S1JJG167PPP	01/10/2021 00:10
656864	207	1	0.06	179.7.225.35	S1AEP421PPP	01/10/2021 00:11

Elaboración propia.