

# POSGRADOS

MAESTRÍA EN

---

## ELECTRÓNICA Y AUTOMATIZACIÓN

RPC-SO-19-No.277-2018

Opción de Titulación:

Artículos profesionales de alto nivel

Tema:

Prototipo móvil IoT para la predicción de la calidad del aire a través de Machine Learning

Autor(es)

Andreina Nathaly Santana Andrade

Director:

Paulina Adriana Morillo Alcívar

QUITO – Ecuador  
2022

Autor(es):



Andreina Nathaly Santana Andrade  
Ingeniera Electrónica, Mención Sistemas Industriales  
Candidata a Magíster en Electrónica y Automatización Mención  
Informática Industrial por la Universidad Politécnica Salesiana –  
Sede Quito.  
naty199316@yahoo.es

**Dirigido por:**



Paulina Adriana Morillo Alcívar  
Ingeniera en Electrónica y Telecomunicaciones  
Politécnica Nacional (2012)  
Magíster en Gestión de la Información  
pmorillo@ups.edu.ec

Todos los derechos reservados.

Queda prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra para fines comerciales, sin contar con autorización de los titulares de propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual. Se permite la libre difusión de este texto con fines académicos investigativos por cualquier medio, con la debida notificación a los autores.

DERECHOS RESERVADOS

2022 © Universidad Politécnica Salesiana.

QUITO– ECUADOR – SUDAMÉRICA

**Andreina Nathaly Santana Andrade**

***Prototipo móvil IoT para la predicción de la calidad del aire a través de Machine Learning***

# Prototipo móvil IoT para la predicción de la calidad del aire a través de Machine Learning

Andreina Nathaly Santana Andrade and Paulina Adriana Morillo Alcívar

Universidad Politécnica Salesiana - Ecuador

---

## Resumen

Existe gran relevancia y preocupación sobre la condición del entorno para la sociedad, ya que los contaminantes producidos por las actividades cotidianas de los hombres no solo causan daños al medio ambiente, sino también riesgos a la salud humana. Por otro lado, el control de parte de las agencias gubernamentales, ha permitido establecer niveles de contaminación aprobados y alertas de perjuicio para la salud. Sin embargo, los equipos convencionales presentan altos costos de implementación y mantenimiento, lo que dificulta que todas las ciudades tengan acceso a estos equipo y realicen un control adecuado. Por consiguiente, este proyecto propone construir un dispositivo IoT de bajo costo que facilite el monitoreo de contaminantes como CO y a través de otras variables como humedad y temperatura pueda predecir el índice de calidad del aire.

La implementación se realizó con el hardware NodeMCU IoT que permitió la recolección y visualización de las variables. A su vez, se construyó un *dataset* que fue usado para establecer tres modelos de series temporales con los que se entrenó tres modelos de redes neuronales recurrentes encargados de la predicción del IAQ. Conforme a los resultados, el modelo con menor error corresponde al modelo entrenado con series temporales multivariable con emdeddings, en lo que respecta al promedio del error cuadrático *MSE<sub>test</sub>* para este modelo es menor al (7%). En cuanto al modelo ST1V univariable presenta un error del (12,93%) y el modelo STMV con multivariables obtiene un error del (15,71%). Por lo que se puede manifestar que el modelo que utiliza embeddings ofrece un modelo óptimo en razón que el *dataset* de la serie temporal utiliza al dato "fecha" como dato categórico.

**Palabras clave**— Calidad del Aire, monóxido de carbono, regresión, series temporales

---

### Abstract

There is great relevance and concern about the condition of the environment for society, since pollutants produced by the daily activities of mankind not only cause damage to the environment, but also risks to human health. On the other hand, control by governmental agencies has made it possible to establish permitted levels of contamination and warnings of harm to health. However, conventional equipment presents high implementation and maintenance costs, which makes it difficult for all cities to have access to this equipment and to carry out adequate monitoring. Therefore, this project proposes to build a low-cost IoT device that facilitates the monitoring of pollutants such as CO pollutants such as CO and through other variables such as humidity and temperature can predict the air quality index.

The implementation was done with the NodeMCU IoT hardware that allowed the collection and visualization of the variables. In turn, a *dataset* was built which was used to establish three time series models with which to train three recurrent neural network models in charge of IAQ prediction. According to the results, the model with the lowest error corresponds to the model trained with multivariate time series with emdeddings, with respect to the *MSE\_test* square error for this model is less than (7 %). As for the univariate ST1V model it presents an error of (12, 93 %) and the STMV model with multivariables obtains an error of (15,71%). Therefore, it can be The model using embeddings offers an optimal model because the time series *dataset* uses the data "date" as categorical data.

---

**Palabras clave**— Air Quality Index, carbon monoxide, regresion, temporal series

# 1 Introducción

En muchos países, tiene índole crítica la condición del aire debido a la prevalencia de contaminación permanente en sus ciudades. Donde intervienen las labores diarias, el sector industrial, la liberación de gases tóxicos, entre otros. Este aumento en los niveles de contaminación ha proliferado padecimientos en el tracto respiratorio mayoritariamente. (Romero Placeres et al., 2006), por esta razón, la Agencia de Protección Ambiental de Estados Unidos (EPA, por sus siglas en inglés) ha previsto para el Índice de Calidad del Aire (IAQ, por sus siglas en inglés) valores aceptables, a fin de salvaguardar la salud de los residentes.

El índice de la condición del aire es un valor adimensional, con el que se pretende indicar la categoría del aire, este número describe la cifra de contaminación ambiental. De esta forma, si se supera un valor de 100 cifras únicas de concentración, la calidad del aire se considera insalubre (Seguel et al., 2017). Por lo tanto, es importante establecer un monitoreo eficaz para determinar índices de calidad del aire adecuados (Aguilar Ferraro, 2021).

En diversos componentes medibles en el medio ambiente el cálculo de la calidad del aire está basado como: dióxidos y partículas, etc (Trelles Motte, 2018). El modelamiento de la condición del aire incluye información de las variadas fuentes de contaminación en el aire presentes con el fin de obtener información amplia de los contaminantes y ligar las afecciones permanentes en el aire con mayor trascendencia. De modo, que se puedan tomar decisiones para disminuir la contaminación.

En el caso del Ecuador, pocas ciudades disponen de sistemas de monitoreo encargadas de medir la solidificación de contaminación. Por ejemplo, en la ciudad de Quito se tiene la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ) que esta conformada por nueve terminales de recepción distribuidas alrededor de la ciudad en sectores de mayor afectación atmosférica que adquieren información instantánea del grado de los contaminantes y material particulado, son 66 instrumentos de la familia Thermo Environmental Instruments (TEI) y TELEDYNE API (ebay, 2021), cada una, con un costo aproximado de 4850,00 USD, lo que representa una inversión inicial alta que se suma al costo de mantenimiento de las estaciones (de Ambiente del MDMQ, 2016).

De igual manera, en Cuenca se exhibe la red dirigida por la Empresa Pública de Movilidad, Tránsito y Transporte (EMOV-EP) instituida con 20 servicios de control distribuidos en diferentes espacios y con tres estaciones automáticas que miden los niveles de contaminación atmosférica, bajo los índices permitidos por la normativa nacional e internacional. Los sensores de calidad del aire implementados en los distintos puntos también son de la marca TELEDYNE modelos M300E, 100E, 200E, 400E (EMOV, 2018). La EMOV-EP registra cada segundo, los valores de varios elementos contaminantes atmosféricos de elementos particulados (Walden, 2017).

Por el contrario, otras ciudades principales del país, como Guayaquil, no cuentan actualmente con una medición periódica de la contaminación del aire. Aunque, a finales de 2021 se propuso la contratación de un servicio para monitorear esta variables en 23 puntos de la ciudad, el costo estimado del servicio sería de aproximadamente 160.000 USD (Ramos, 2021). Como evidencian los hechos, a pesar de que las autoridades de la ciudades hacen esfuerzos por controlar los niveles de contaminación ambiental, establecer redes de monitoreo utilizando equipos especializados o sus equivalentes, tiene costos iniciales muy elevados y requiere de una cantidad significativa de recursos adicionales para su mantenimiento y calibración. Por lo tanto, persiste el menester de ligar tecnologías de operación y monitoreo asequibles para intensificar la medición del índice de calidad en otras ciudades (Cremades et al., 2013).

Incluso en las áreas industriales presentes en el Ecuador carecen de un control de la condición ambiental, se presentan aproximaciones por modelos de estudio generados por el Banco Mundial, organismo internacional, por parte del departamento manufacturero (Ecuador, 2008). Los encargados de gestionar estas instalaciones enfrentan varios desafíos, especialmente el control de los gases peligrosos en los espacios de trabajo (Valle, 2017) y la limitada disponibilidad de recursos para equiparar gastos de instrumentos de medición, su instalación y operatividad. Ante esta perspectiva, los sensores de bajo costo aparecen como una alternativa atractiva para resolver los problemas de contaminación. Sin embargo, los sistemas comerciales equipados con estos sensores siguen siendo manufacturados en el exterior y por su costo (ej. Aeroqual de 500 a 2500 euros) aún distan mucho de ser accesibles en el contexto económico local (Cremades et al., 2013).

Este trabajo presenta el desarrollo de un dispositivo móvil IoT (Internet de las cosas) para la predicción de la calidad del aire en espacios. El dispositivo desarrollado está basado en sensores electrónicos de bajo costo presentes en el mercado local. Además, dispone de una interfaz de usuario para visualizar la condición del aire en tiempo real. Después de la construcción del dispositivo, se realiza la medición de diferentes variables con el fin de entrenar tres modelos de *machine learning* que permitan

predecir el IAQ. Finalmente, se validan los modelos y se compararan los resultados.

## 2 Materiales y Métodos

El proyecto esta constituido en tres fases (ver Figura 1). La primera fase consiste en el diseño y construcción del prototipo IoT, la segunda fase reside en la adquisición y medición de variables como temperatura, humedad, concentración de CO necesarias para calcular el IAQ. En la última fase, se procede con la construcción de los modelos de *Machine Learning* para la predicción del IAQ.

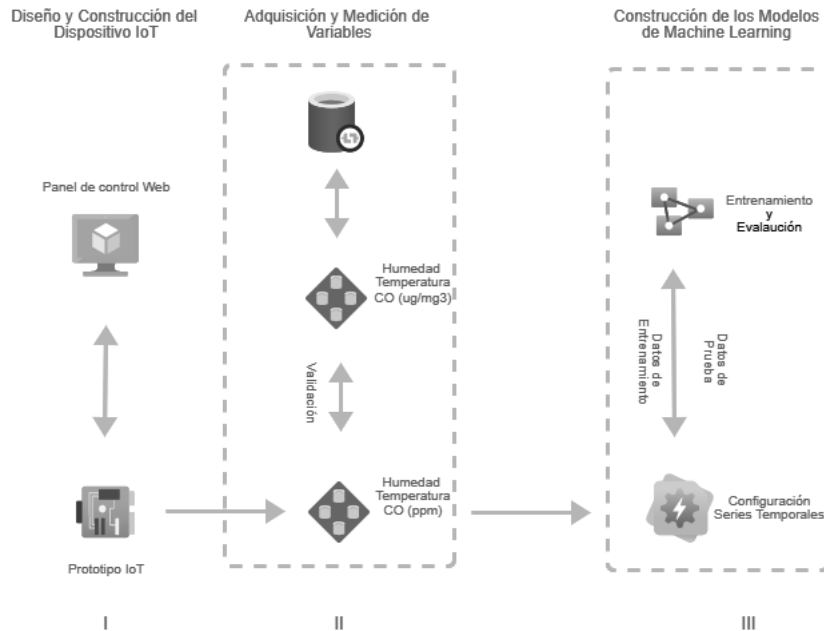


Figura 1: Metodología de construcción de un dispositivo y un modelo de *Machine Learning* para la medición y predicción del IAQ, respectivamente.

### 2.1 Diseño y construcción del prototipo IoT

La fase I del proyecto consiste en la construcción del prototipo, donde se incorporan elementos electrónicos de bajo costo y herramientas de software de código abierto, que se describen en la Tabla 1. Destaca el sensor MQ-135, que es usado en equipos de control de calidad aéreo, en edificaciones industriales.

En la Figura 2 se definen las magnitudes y variables que serán considerados en el cálculo de la condición aérea, estas son: Monóxido de carbono (CO), Temperatura (T) y Humedad (H). Tanto la temperatura como la humedad son variables que afectan directamente a la interpretación, por ejemplo, influye en el período de exposición en distintas épocas del año ya que podría tener una elevación el cual afecta a la troposfera. Por lo tanto, los elementos climáticos de Temperatura (T) y Humedad (H) son factores característicos, los períodos de estabilidad, y la dilución de gases en el aire. (Romero et al., 2004)

El prototipo se construyó sobre la placa electrónica NodeMCU plataforma IoT, esta puede conectarse a distintos elementos y permite la transmisión de datos, su principal función en el prototipo es capturar las lecturas de los sensores MQ135-sensor de calidad y DHT11-sensor de temperatura y humedad. Los datos se envían a través de internet a un servidor, con el fin de visualizarlos mediante la plataforma IoT de Ubidots.

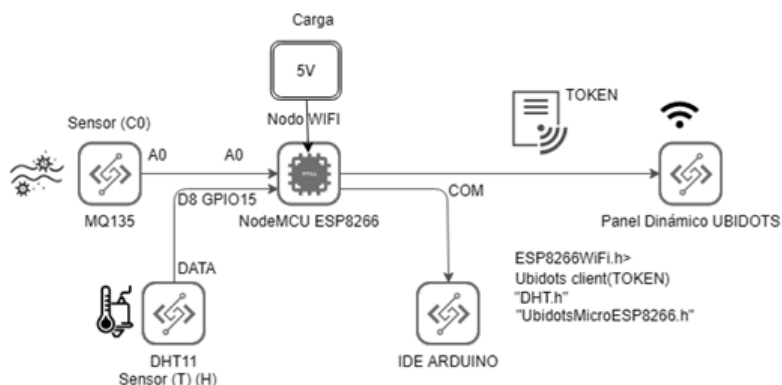


Figura 2: Esquema diseño prototipo IoT.

El calentador fijado en la estructura del MQ135 proporciona las condiciones para activar el sensor, este recepta directamente las partes por millón del gas CO mediante el pin analógico AO del NodeMCU desde el pin de salida que posee el dispositivo, los dos pines restantes son para obtener la energía necesaria para calentar el microtubo. Al alimentar el sensor con 5v existe un cambio en el valor de las resistencias del sensor,  $R_s$  de 30 a 200  $k\Omega$  y la resistencia de carga  $R_L$  de 10 a 47  $k\Omega$ .

Para este caso se conecta el pin partida del sensor a un pin analógico del NodeMCU, para acceder y escalar en varios rangos la condición del gas respecto al requerimiento.

El sensor de temperatura trabaja en conjunto con el sensor de MQ135 enviando las lecturas de humedad y temperatura hacia el NodeMCU, este sensor utiliza su pin digital para enviar la información.

Tabla 1: Elementos usados en el Prototipo (IAQ)

Elemento	Costo	Parámetros	Características	Lectura
<b>Hardware</b>				
Sensor MQ-135	P.U 7 USD	Monóxido de carbono (CO)	Sensibilidad ajustable para detectar el gas (CO), humo y otros	Salida Dual Lectura Análogica
Sensor DHT11	P.U 12 USD	Humedad(H) Temperatura (T)	Requiere su propio protocolo. Rango de operación entre 0°C hasta 50°C para la (T), y para la (H) entre el 20 % a 90 % de humedad relativa	Lectura Digital
NodeMCU	P.U 35 USD	Módulo WiFi ESP8266	Tarjeta libre Software y de Hardware	Integra GPIO, I2C, PWM, 1-Wire ADC
<b>Software</b>				
Arduino IDE		Código abierto	Adquisición de datos uso librerías (include "ESP8266WiFi.h") (include "DHT.h")	Lectura de variables

### 2.1.1 Caracterización sensor MQ135

La concentración del gas CO, medido por el sensor en partes por millón (*ppm*) requiere ser escalado, ya que los valores obtenidos frente a los datos reales no presentan características directas. Por lo tanto, se

estima la curva a partir de las características de sensibilidad que se encuentra en el *datasheet* del sensor, la Figura 3 muestra para el sensor los rangos exponenciales comunes en varios gases.

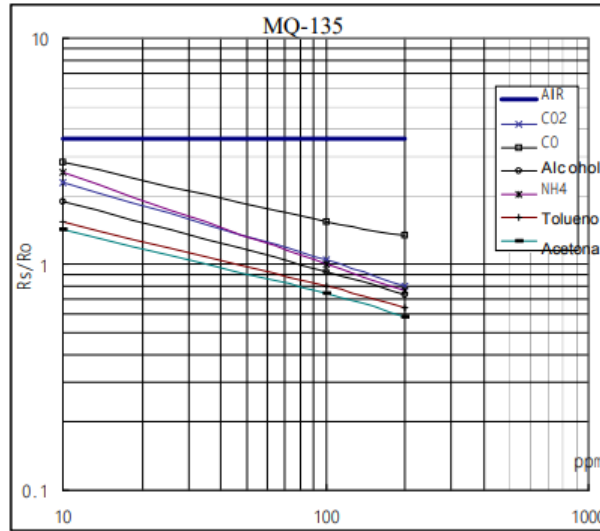


Figura 3: Tendencia exponencial MQ135.

Para caracterizar las mediciones del gas CO, se equipara las lecturas del prototipo con las de la Red Metropolitana de Monitoreo Atmosférico de Quito (REMMAQ). La REMMAQ funciona las 24 horas del día, el período anual, propicia cada diez minutos promedios de la condición climática, esta información se encuentra con apertura en <http://www.quitoambiente.gob.ec/ambiente/>. Desde junio del 2003 los datos son presentados en formatos como archivos separados por comas, con el fin de facilitar su revisión y manipulación (CORPAIRE, 2008).

Mediante software se realiza la calibración del sensor MQ135, tomando en cuenta los puntos correspondientes al gas Monóxido de Carbono, su comportamiento se transforma en un esquema exponencial, en la que la polución del gas en función de  $R_s/R_o$  se encuentra en *ppm* como se observa en (1). A continuación, se calcula el coeficiente de determinación ( $R^2$ ) igual a 0.987, lo que indica que el ajuste es adecuado.

$$X = \frac{R_s}{R_o} \quad (1)$$

Se obtiene la curva de ajuste como se observa en (2) y al aplicarla se obtiene los valores aproximados de concentración del gas CO en *ppm*.

$$y = 933,56X^{-4,355} \quad (2)$$

En efecto se escala el dato analógico de entre 0 y 1023 bits, siendo 1.2V (*VRL*) el voltaje límite en la salida del sensor a considerar, con un voltaje de entrada de 5V (*Vc*). Se calcula  $R_s$  en base a los fundamentos de construcción del divisor del sensor con  $R_L$  de un  $1k\Omega$ .

Para la adquisición de los datos emitidos por los sensores se utiliza el panorama de elaboración integrado (IDE) de Arduino aplicando el plugin ESP2866 que permite programar el NodeMCU desde el IDE de Arduino.

El módulo NodeMCU se puede configurar con varios entornos de desarrollo, incluido Arduino IDE, pero para realizar este proceso es necesario añadir al entorno un archivo con la configuración básica del módulo. Esta configuración está relacionada con la comunicación entre el entorno y NodeMCU, estableciendo el puerto de comunicación COM5 de 115200 bits por segundo (*baudios*) a la agilidad de remisión de datos.

Posteriormente, se incorporan las librerías ESP8266WiFi.h y DHT.h, para la programación del NodeMCU y la vinculación de los sensores (monóxido de carbono, temperatura y humedad).



Para transmitir datos se utiliza la librería DHT.h, por medio de esta se conecta el sensor DHT11 (dht) con pocas pérdidas. El sensor DHT-11 mide la temperatura y la humedad, a través de los métodos `dht.readTemperature()` y `dht.readHumidity()`, respectivamente.

Por otra parte, para conectar a la red WIFI con la unidad ESP8266 mediante la `ESP8266WIFI.h` permite la transmisión de los datos de varios captadores. Al trabajar bajo un cifrado WEB y WPA2 con una operación cliente servidor.

De acuerdo al incremento o decremento de la concentración de humo/gas varía el voltaje de salida del captador directamente proporcional. Sin embargo, se debe considerar que las distintas unidades de la serie MQ presentan sensibilidad a múltiples gases en diferentes proporciones. Por lo tanto, para obtener los valores de la concentración del gas CO se plantea diferentes rangos de escalabilidad que se obtienen de la ficha técnica del sensor.

La concentración de CO en *ppm* se estima a través de la Ecuación 2, que está programada en el terminal IDE Arduino, y para poder comparar este valor con el CO estimado por la REMMAQ se debe hacer una conversión de unidades de *mg/m<sup>3</sup>* a *ppm*, usando la Ecuación 3.

$$y = \frac{p * m * x}{R * 1 * (273 + t)} = \frac{760 * 28,01 * x}{62,36 * 1 * (273 + 25)} \quad (3)$$

Donde:

- y*: Cifra de CO en *mg/m<sup>3</sup>*
- p*: Presión de la atmosfera en *mmHg* (760)
- m*: Peso en mol del CO en *g/mol* (28,01)
- x*: Cifra de CO en *ppm*
- t*: Temperatura en *°C* (25)
- R*: Constante de los gases ideales *62.36 mmHgL/mol°K*

### 2.1.2 Configuración del Tablero Dinámico

El prototipo se conecta a Ubidots, una plataforma de IoT que permite la toma de datos de los sensores y los transmite a la nube de Ubidots, a través del protocolo de conexión HTTP. Así, el usuario cuenta con un cuadro de mando web (*dashboard*) para visualizar la información de los sensores, con una características relevante en tiempo real, central y remota. Además, por medio de la aplicación de Ubidots Explorer se pueden acceder y visualizar los datos, desde cualquier dispositivo móvil Android.

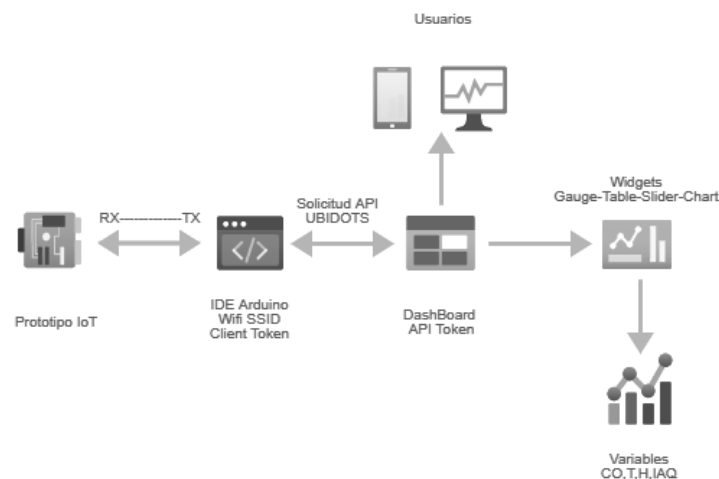


Figura 4: Esquema de operación del dispositivo con la plataforma IoT de Ubidots.

Ubidots también permite la exportación de reportes de las variables, por medio de un evento creado en Add Reports, el cual envía automáticamente un informe hacia un correo electrónico.

En la Figura 4 se muestra la operación del dispositivo con la nube de Ubidots. De esta forma, para que el NodeMCU se conecte con el *dashboard* de la aplicación, es necesario, por un lado, establecer la configuración de la red WIFI, y por otro lado, conectarse a la API de Ubidots a través de un token único, generado por la plataforma. Una vez conectado el dispositivo se pueden acceder a las mediciones de las diferentes variables a través de cuadros, tablas, indicadores, mapas, métricas y widgets del panel de control.

En este caso, el IAQ y el Monóxido de Carbono (*ppm*) se muestran sobre un indicador de tipo gauge, mientras que los valores distintivos de temperatura y humedad se reflejan en widgets de control. Las tablas dinámicas se usan para obtener reportes más detallados, que pueden incluir la fecha y hora de cada medición y los gráficos de líneas y los widgets de doble eje se usan combinados para visualizar los datos de series de tiempo de una o más variables a la vez, con un máximo de 30 variables.

Después del registro de las variables y la configuración de los parámetros en la página de Ubidots se personaliza el tablero de la interfaz gráfica (ver Figura 5).

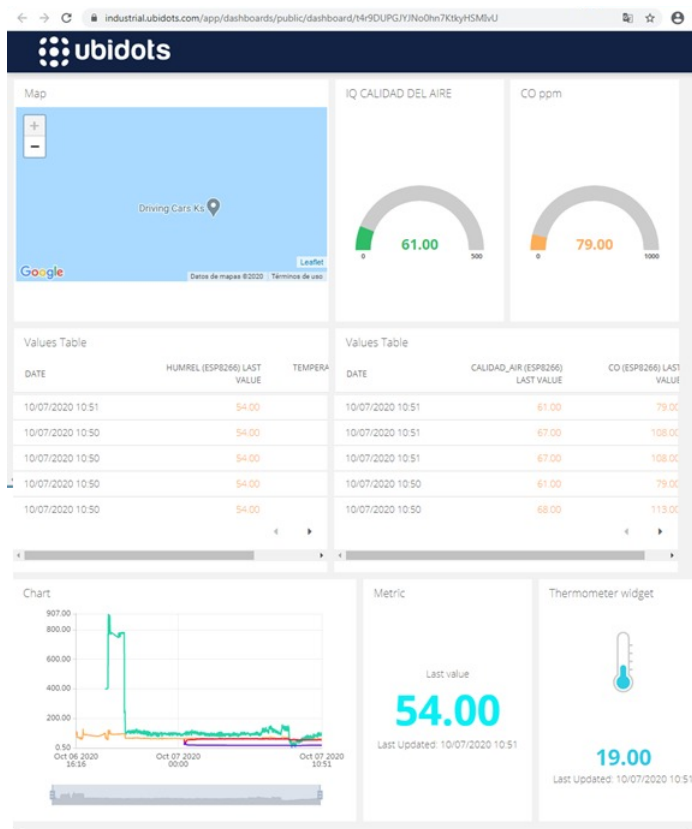


Figura 5: *Dashboard* de la aplicación de medición de IAQ

## 2.2 Adquisición y medición de variables

La fase II del proyecto corresponde a la adquisición de datos (ver Figura 2). En este caso, se construyen dos conjuntos de datos, el primero se extrae de los reportes mensuales de CO de la REMMAQ y el segundo se construye directamente de los datos adquiridos por el dispositivo IoT. Los datos de la REMMAQ se usaron para validar las mediciones del dispositivo IoT, mientras que los datos del dispositivo se usan para el entrenamiento y la evaluación de los modelos de *machine learning*.

El *dataset* que se extrae de la REMMAQ expresa los valores de concentración del CO en  $ug/m^3$ , para ser comparables con los valores de CO del prototipo, tienen que convertirse en primera instancia a unidades de  $mg/m^3$ . Por el contrario, los valores de CO censados por el dispositivo IoT se calculan directamente en *ppm* y se guardan en un fichero en formato .csv.

Los datos recolectados en ambos *datasets* se miden en un período de tiempo de  $n$  días y para cada día se realizan las captaciones de las variables, en ventanas de tiempo de una hora, desde las 00h00

hasta las 23h00. Después de la adquisición de los datos se realiza una depuración para eliminar datos faltantes e inconsistentes. Luego, se comparan los valores de las variables de ambos conjuntos de datos para determinar el correcto funcionamiento del dispositivo.

Posteriormente, los datos censados por el prototipo se dividen en tres franjas horarias por día de 00h00 a 08h00, de 08h00 a 16h00 y de 16h00 a 24h00, respectivamente. De estas franjas horarias se escoge el valor máximo de concentración de CO y el valor promedio de las variables temperatura y humedad, es decir, por cada día se obtiene tres instancias con tres atributos cada una. Para el caso del monóxido de carbono se hace una conversión adicional a *ppm*, para estimar el valor del IAQ a través de la Ecuación 4.

$$IAQ = \frac{(Ih_i - Il_o)}{(BPH_i - BP_l o)}(C_p - BP_l o) \quad (4)$$

Donde:

$IAQ$  = Identificativo Índice de calidad del aire.

$C_p$  = Valor  $\bar{x}$  de proporción observada y corregida.

$BPH_i$  = Valor de corte  $\geq$  a  $C_p$

$BP_l o$  = Valor de corte  $\leq$  a  $C_p$

$Ih_i$  = Dato del ICA de  $BPH_i$

$Il_o$  = Dato del ICA de  $BP_l o$

Los valores  $BPH_i$ ,  $BP_l o$ , se toman de la tabla de referencia de la norma EPA, para determinar el IAQ a través de la concentración de CO (ver Tabla 2). Sin embargo, la norma EPA permite el calculo de IAQ en función de diferentes contaminantes.

Tabla 2: Intervalos IAQ establecidos por la EPA.

<i>O3(ppb)</i>	<i>PM2,5(g/m3)</i>	<i>PM10(g/m3)</i>	<i>CO(ppm)</i>	<i>SO2(ppb)</i>	<i>NO2(ppb)</i>	<i>IAQ</i>
-	0.0 - 12,0	0 - 54	0.0 - 4.4	0 - 35	0 - 53	0 - 50
-	12,1 - 35,4	55 - 154	4.5 - 9.4	36 - 75	54-100	51 - 100
125 - 164	35,5 - 55,4	155 - 254	9.5 - 12,4	76 - 185	101-360	101 - 150
165 - 204	55,5 - 150,4	255 - 354	12,5 - 15,4	186 - 304	361-649	151 - 200
205 - 404	150,5 - 250,4	355 - 424	15,5 - 30,4	305 - 604	650-1249	201 - 300
405 - 504	250,5 - 350,4	425 - 504	30,5 - 40,4	605 - 804	1250-1649	$\geq$ 300

Finalmente, se califica el IAQ de acuerdo a seis categorías definidas por la norma EPA (ver Tabla 3). Esta tabla y se puede utilizar gracias a que solo se está considerando un solo contaminante (CO), caso contrario sería necesario calcular el índice parcial de cada contaminante y luego el índice global que resulta ser el máximo valor de los índices parciales.

Tabla 3: Cifras de la Calidad de Aire, IAQ.

Rangos	Punto de vista de la Salud	Color de
0-50	Perfecta	Tono Blanco
50-100	Ídonea	Tono Verde
100-200	Afecta enfermos crónicos insalubre.	Tono Gris
200-300	Insalubre (enfermos)	Tono Amarillo
300-400	Insalubre en localidad y dañina.	Tono Naranja
400-500	Grave.	Tono Rojo

Cuando los valores del IAQ están muy por debajo de 100, la calidad del aire, en su conjunto, es adecuada y no existe perjuicio potencial para la salud pública. Al contrario, cuando los valores de IAQ aumentan significativamente por encima de 100, el aire limpio disminuye, convirtiéndose en un riesgo para la salud pública. Mientras que, un valor de IAQ cercano a 100, implica que las concentraciones de los contaminantes se mantienen en el margen de lo permitido.

## 2.3 Construcción de los Modelos de ML

Debido a que la variable de salida (IAQ) que se desea predecir es tipo valor numérica continua, la tarea de aprendizaje automático que se aborda está enmarcada en el paradigma de Aprendizaje Supervisado, específicamente en la tarea de Regresión. Sin embargo, ya que las instancias del conjunto de entrenamiento están estimadas de manera longitudinal en el tiempo es necesario configurar el *dataset* de entrenamiento como series temporales, definiendo la variable a predecir en el tiempo  $t$  en función de variables medidas en instantes previos  $t - 1, t - 2$ , etc.

### 2.3.1 Configuración de Series Temporales

En esta sección se establecen tres modelos de series de tiempo: univariable, multivariable y multivariable con embeddings. Para cada serie se construye un *dataset* mediante el uso de la librería Pandas de Python, ya que esta librería facilita la manipulación de las variables a través del manejo de *dataframes*.

#### Series de tiempo univariable (ST1V)

En este tipo de series, la variable a predecir (IAQ) en el tiempo  $t$  depende exclusivamente de la misma variable en instantes previos  $t - i$  ( $i = 1, 2, \dots, n$ ). En este caso, se escoge una ventana de tiempo  $n = 7$  y se configura el *dataset* para que conste de siete columnas que representan los valores de IAQ en los instantes de  $t - 7$  a  $t - 1$ .

Las mediciones en instantes previos representan los valores máximos de cada franja horaria. Como cada día tiene tres franjas horarias, para predecir el IAQ en  $t$ , se consideran los valores máximos de IAQ de dos días atrás completos y una medición anterior adicional.

#### Series de tiempo multivariable (STMV)

En este tipo de serie de tiempo, se puede usar más de una variable numérica para predecir la variable de salida. En este caso, se usan los atributos de temperatura, humedad e índice de calidad del aire, medidos en tiempos anteriores a  $t$ .

De igual forma, se escoge una ventana temporal de  $n = 7$ , pero esta vez se incluyen dos variables adicionales. Multiplicando por tres los atributos de cada instancia en el conjunto de datos.

#### Series de tiempo con embeddings (STE)

Además de agregar nuevas variables numéricas al conjunto de datos, es posible añadir atributos categóricos por medio del algoritmo *embeddings*. En consecuencia, en *dataset* de la serie univariable se aumentan dos columnas adicionales que corresponden al día y el mes de la medición. La codificación para el día, de lunes a domingo, es de 0 a 6, mientras que para el mes, de enero a diciembre, va de 1 a 12.

### 2.3.2 Entrenamiento y evaluación de los modelos de predicción del IAQ

Luego de configurar los diferentes conjuntos de datos, para los tres tipos de series temporales, se normalizan los valores de los atributos entre -1 y 1, para facilitar el entrenamiento de los modelos. Después, se distribuye el *dataset* en un grupo de datos de entrenamiento (*train set*) y un grupo de datos de prueba (*test set*).

Con el *train set* se entrena una red neuronal recurrente, también llamada RNN por sus siglas en inglés Recurrent Neural Networks. Esta red fue seleccionada a razón debido a que la estructura permite recapitular datos durante rangos o franjas anteriores en el tiempo y facilita la tarea al realizar evaluaciones posteriores utilizando períodos de datos históricos o ventanas de tiempo previos para determinar el siguiente paso. La RNN se configura con la función de activación tangente hiperbólica la cual permite transformar los valores agregados a una escala (-1,1), donde los valores altos tienen de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a -1, la función objetivo *Loss*, el método de optimización Adam mantiene un factor de entrenamiento por parámetro y cada factor de entrenamiento también se ve afectado por la media del momentum del gradiente y dos capas ocultas, cada capa con siete neuronas en razón que al utilizar un gran número de neuronas en una capa oculta no se obtiene una respuesta satisfactoria al problema se aplica dos capas ocultas reduciendo el número de neuronas en cada capa considerando que el entrenamiento es mas lento entre mas capas se aplique. Además, se establecen 200 épocas para el entrenamiento. Por otra parte, una vez que los modelos han sido entrenados se utiliza el *test set* para evaluar su rendimiento a través del cálculo del *Mean Squared Error* (MSE).

### 3 Experimentación y Resultados

En sección se exponen los resultados de los experimentos realizados, con el fin de determinar el correcto funcionamiento del dispositivo IoT y de los modelos de predicción del IAQ.

Para esto, se tomaron en cuenta las mediciones realizadas por un equipo de la REMMAQ ubicado en el sector del Camal de la ciudad de Quito. De igual forma, el dispositivo IoT se ubicó en la Terraza del Hospital Patronato Municipal San José del Sur, también en el Sector Camal, muy cerca a la estación REMMAQ. Se realizaron mediciones de temperatura, humedad e IAQ durante 153 días consecutivos, del 1 de Agosto al 31 de Diciembre de 2020, considerando 24 mediciones por día, una por cada hora.

El código del desarrollo de este proyecto, así como los grupos de datos utilizados en el entrenamiento y evaluación de los modelos de ML se encuentran disponibles en <https://github.com/AndreinaSantana/Serie-Temporal-IQC.git>.

#### 3.1 Funcionamiento del Prototipo IoT

Después de implementar el dispositivo e instalarlo en la locación de prueba, se encendió el prototipo y se dejó pasar 24 horas antes de la toma de mediciones, ya que de acuerdo a las indicaciones técnicas, este período de tiempo es necesario para estabilizar el sensor MQ135. Una vez que se estabilizó el sensor, se midieron las variables de humedad, temperatura y concentración de CO y estos valores se enviaron a la nube de Ubidots. También se captó de cada uno de los componentes el dispendio de energía para compararlos con las estimaciones de consumo hechas por el fabricante (Tabla 4). Con estos valores se determinó el período de soporte de carga de la pila igual a 3600 mAh. Las características de la batería consideradas en este análisis fueron: Batería de Polímero de litio montada en células con capacidad hasta 22.2 V y 6 celdas, están abastecidas con alta robustez a la descarga, lo que ofrece a permanecer cargas de alta corriente y reducir la resistencia. El peso y tamaño aproximado es de 577g y 137x43x43 mm, respectivamente. De igual forma, el enchufe de carga es el JST-XH y el enchufe de descarga XT-60, soportando 19 horas de carga.

Tabla 4: Consumo de energía de los elementos del prototipo vs. las estimaciones del fabricante

Componente	Prototipo (mA)	Fabricante (mA)
Microcontrolador	45	12
Nodo	14	14
Sensor	125	119

Respecto a la conexión a internet, las características de la red empleada para conectar el dispositivo IoT con la nube de Ubidots fueron: ancho de banda de 35Mbps y tiempo de latencia igual a 96ms. De esta forma, no se registraron pérdidas de datos o fallos en la conexión de la red, los datos se actualizaban en la nube cada dos segundos.

Por otro lado, dado que el fin de este trabajo fue desarrollar un dispositivo móvil para la predicción del IAQ, de bajo coste, de acuerdo a los precios comerciales de los elementos usados en la construcción del dispositivo (Tabla 1) y al costo de desarrollo incluyendo horas de investigación y tiempo de construcción, el precio referencial de este dispositivo estaría alrededor de 250 USD. Se debe considerar que este valor podría variar de acuerdo a las funcionalidades que presente el prototipo.

#### 3.2 Datasets de entrenamiento

Para la validación de las mediciones de temperatura, humedad y concentración de CO censadas por el prototipo, se compararon estas variables con las estimaciones de la REMMAQ en el mismo período de tiempo. Así, los conjuntos de datos estuvieron conformados por 3672 instancias que luego de la depuración, terminaron en 3071 instancias. Con estos valores se calculó el error cuadrático medio (MSE) para cada variable (Tabla 5).

Tabla 5: Comparación de las variables del *dataset* construido con el dispositivo vs. el *dataset* REMMAQ

	Dispositivo			REMMAQ			MSE
	Min	Max	Media	Min	Max	Media	
Temperatura ( $^{\circ}C$ )	9.32	14.89	5.57	9.22	15	5.78	31.0249
Humedad (%)	58.97	82.81	23.84	59	83	24	568.3456
CO (ppm)	0.415501607	0.446926098	0.031424491	0.425501607	0.456926098	0.031424491	0.000987499

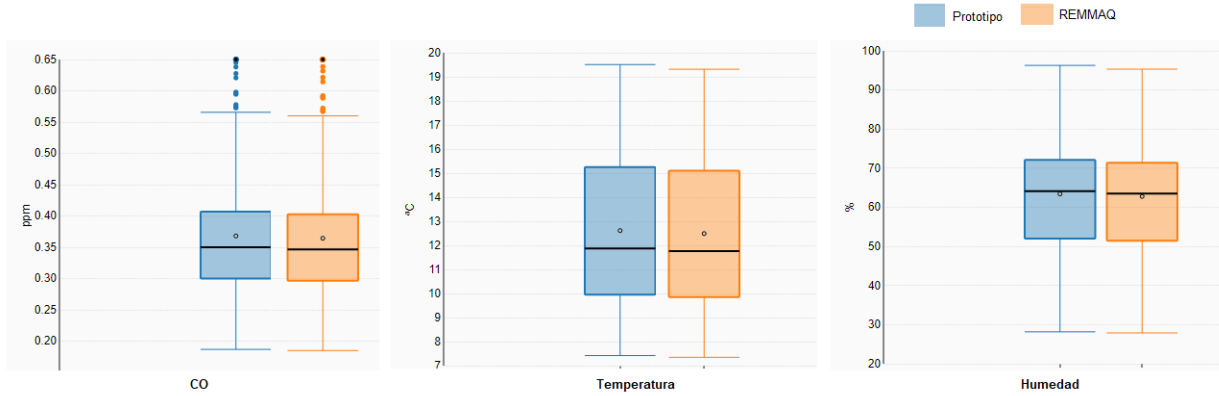


Figura 6: Diagrama de cajas y bigote datos Dispositivo vs REMMAQ.

La Figura 6 evidencia el diagrama de caja y bigote de los datos medidos por el dispositivo y los datos de la REMMAQ, en el cual se observa que para las tres variables no se presentan datos atípicos ya que estos no se muestran después de los valores máximos o mínimos. Los valores (*max/min*) para la temperatura se encuentra en (15,02/9,31) con una desviación estándar mínima del  $\sigma=1.96$  para la variable humedad (82,81/60,85) y la desviación en  $\sigma=7,29$  y para CO se presentan los valores entre (0,40/0,22) con una desviación del  $\sigma=0.051$  presentando una menor dispersión de los datos. El rango intercuartílico o *RIC* se encuentra en 3.05 para el dispositivo frente a 3.4 que corresponde a los datos de la REMMAQ. Además respecto a la temperatura no presenta variabilidad fuera del cuartil superior al contrario de las variables humedad y CO que si presenta una variabilidad.

En los gráficos no se sitúa la mediana en el centro de las cajas por lo que se tiene una distribución asimétrica es decir los datos no se encuentran distribuidos normalmente, y se toma en consideración si un lado de la caja es más largo que otro, indica un rango más amplio, por lo que los datos estarán más dispersos y un rango menos amplio, muestra que los datos están más próximos.

Después de la validación de las mediciones, se segmentaron los datos en tres franjas octohorarias. Esto redujo el tamaño del *dataset* a 146 instancias. De las cuales, las 116 primeras se usaron para el entrenamiento y las 30 restantes en la evaluación. La Tabla 6 presenta una muestra de estos los valores.

Tabla 6: IAQ Octohorario Máximo

Fecha	IAQ
2020-08-31	3.840917
2020-09-30	4.472322
2020-10-31	3.634193
2020-11-30	3.190100
2020-12-31	2.863781

De acuerdo a las tres series temporales implementadas univariable, multivariable y mutivariable con *embeddings*, la cantidad de atributos considerado en cada grupos de datos fue 7, 21 y 9, respectivamente.

La Figura 7 muestra la serie temporal de la variable IAQ en el período del 1 de agosto al 31 de diciembre de 2020. Como se puede observar los valores de IAQ son inferiores a 10, esto implica que la codición del aire se encuentra en el rango óptimo, de acuerdo a la norma EPA.

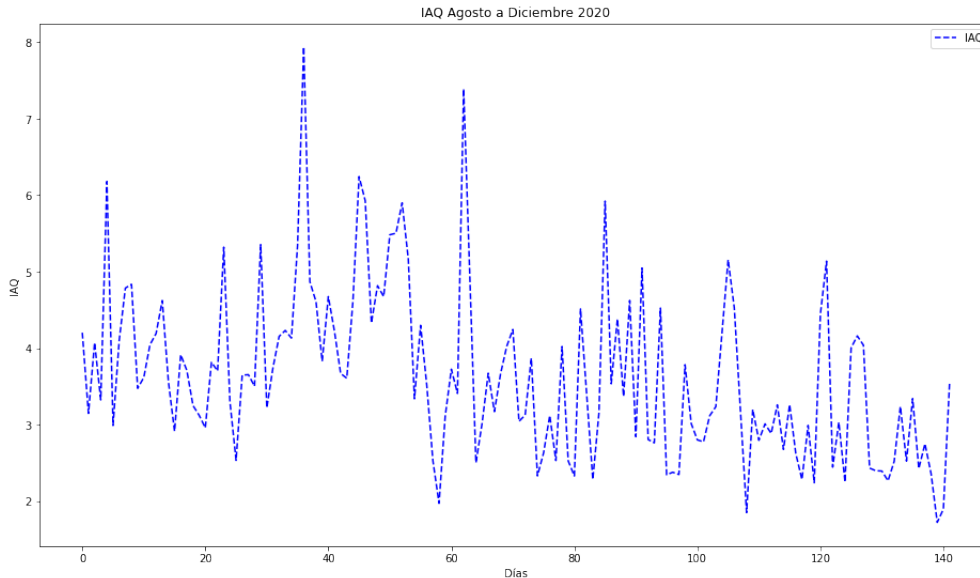


Figura 7: IAQ CO Serie Temporal.

### 3.3 Rendimiento de los modelos de predicción del IAQ

Con los tres conjuntos de datos ST1V, STMV y STE de 116 instancias cada uno, se entrenan tres redes del tipo RNN. Como se visualiza en la Figura 8 se muestran las cifras de la métrica Loss aplicada a los set de entrenamiento y validación. Para el modelo ST1V y STE el valor de *Loss* desciende pero en el modelo que aplica serie temporales con embeddings se mantiene estable el cual se encuentra con un valor de 0.3665. Con respecto al modelo STMV este muestra un valor de *Loss* en 0.3140 pero este no tiene una estabilidad y presenta variaciones a lo largo de la toma de muestras.

Para la evaluación se usan los conjuntos de prueba con 30 instancias cada uno, para cada red se calcula el valor de la tasa de perdidos y el error cuadrático medio. La Tabla 7 muestra los valores de *loss* y MAE tanto en el período de entrenamiento como en el punto de prueba.

Tabla 7: Metricas de las RNN3 Modelos

Modelo	loss_train	loss_test	MSE_train	MSE_test
1)M.ST1V	0.3964	0.3105	0.2452	0.1293
2)M.STMV	0.3140	0.3277	0.2434	0.1571
3)M.STE	0.3665	0.2033	0.2262	0.0716

El modelo ST1V y STMV se muestran similares, es decir que al instaurar las variables de temperatura y humedad no permite una mejora en las métricas. Sin embargo algo más ostensible el modelo con embeddings propicia un mejor trabajo : el *loss\_test* pasa de 0.3277 a 0.2033 y el *MSE\_test* de 0.1571 a 0.0716. En cuanto a la métrica MSE esta ofrece una mejora ya que se presenta una disminución de 0.2452 a 0.2262 y estimula un progreso en el modelo. Las RNN son una clase de redes para analizar datos de series temporales que pueden ser tratadas en la dimensión de “tiempo” por ende al considerar la fecha como dato categórico es el que ofrece un modelo óptimo.

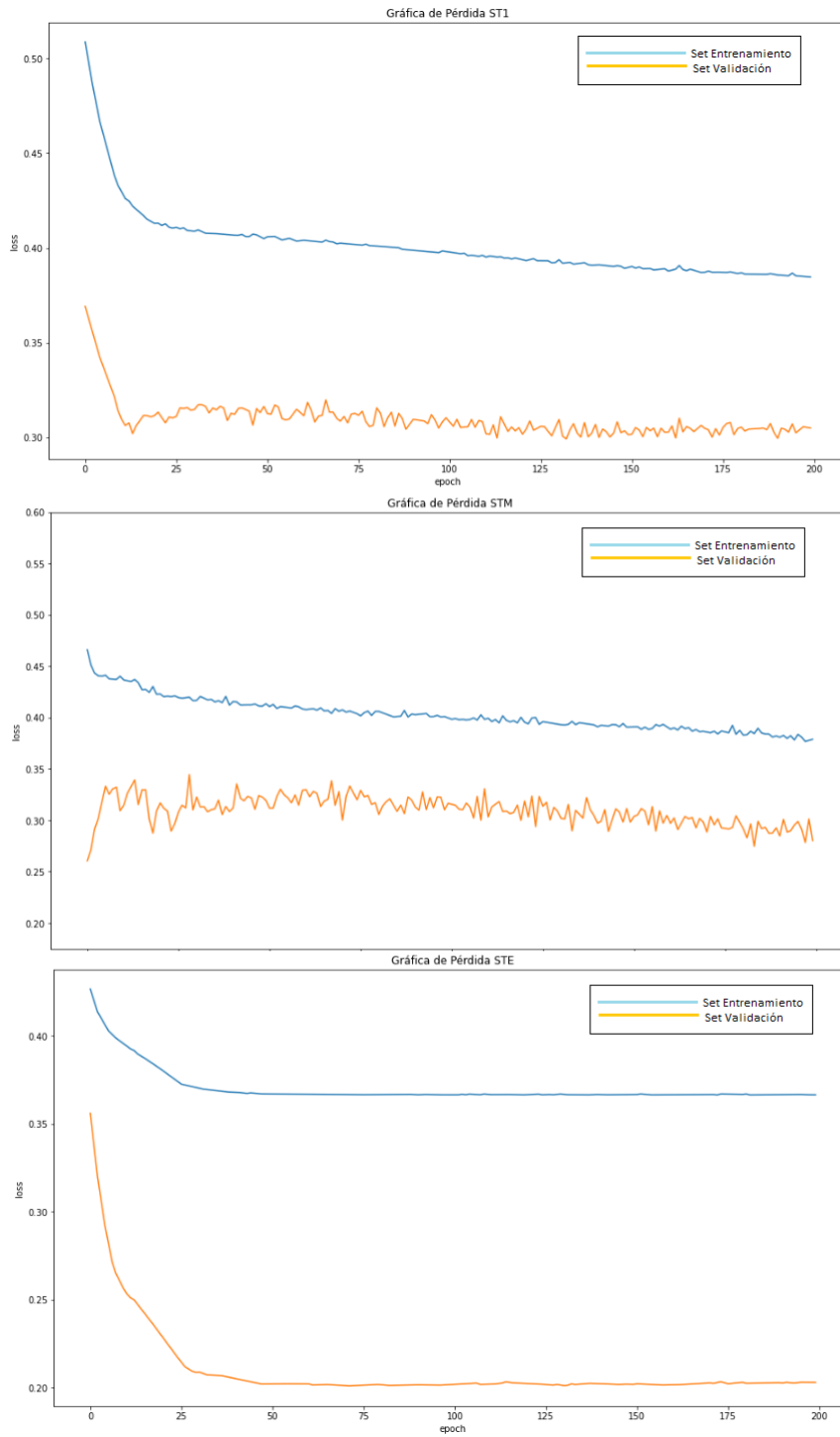


Figura 8: Validación Modelos.



## 4 Conclusiones y Trabajo Futuro

En este trabajo se desarrolló un prototipo móvil IoT para predecir los valores de índice de calidad del aire. El prototipo IoT demostró ser muy eficaz en la medición de variables climáticas como la temperatura y humedad. Así mismo, los elementos electrónicos utilizados en la construcción del dispositivo lograron disminuir los costos, permitiendo que el dispositivo sea asequible.

El uso de series temporales y los algoritmos de *Machine Learning* permiten analizar a profundidad el comportamiento del IAQ en el tiempo. De acuerdo a los resultados de los tres modelos entrenados, el modelo que utilizó el *dataset* construido con la serie multivariable con *Embeddings* obtuvo el menor error. Por el contrario, el modelo entrenado con la serie univariable obtuvo el peor resultado. Por lo tanto, se puede concluir que incorporar variables categóricas de tiempo, pueden ayudar a mejorar el entrenamiento de las redes neuronales recurrentes (RNN).

Como trabajo futuro se recomienda ampliar los experimentos a otras configuraciones de red, otras ventanas de tiempo y el uso de un conjunto de datos de entrenamiento más extenso. Adicionalmente, se podrían incluir variables adicionales de otros contaminantes con el dióxido de carbono, benceno, etc que permitan un monitoreo integral de la calidad del aire.

## 5 Agradecimientos

Al Grupo de Investigación IDEIAGEOCA de la Universidad Politécnica Salesiana en Quito, Ecuador que soporto con información idónea para la ejecución de este trabajo.

## Referencias

S. Aguilar Ferraro. *Estimación del aporte a los niveles de material particulado ambiental, por parte de las centrales térmicas del Gran La Plata, mediante el diseño de una red de monitoreo de calidad de aire*. PhD thesis, Universidad Nacional de La Plata, 2021.

CORPAIRE. Informe anual 2008 la calidad del aire en quito, 2008. URL [http://www.quitoambiente.gob.ec/ambiente/images/Secretaria\\_Ambiente/red\\_monitoreo/informacion/iqca\\_2008.pdf](http://www.quitoambiente.gob.ec/ambiente/images/Secretaria_Ambiente/red_monitoreo/informacion/iqca_2008.pdf). INFORME ANUAL 2008 LA CALIDAD DEL AIRE EN QUITO.

P. Cremades, F. Castro, R. Fernandez, R. Clausen, and S. Puliafito. Desarrollo de un monitor abierto de calidad del aire (maca). 09 2013.

S. de Ambiente del MDMQ. Red de monitoreo atmosférico, 2016. URL <http://www.quitoambiente.gob.ec/ambiente/index.php/71-red-de-monitoreo-atmosferico?start=4>. Antecedentes e institucionalidad.

ebay. Teledyne api t400 uv absorption, 2021. URL <https://www.ebay.com/p/27026729624>. Analizadores de gases o combustión.

G. Ecuador. Estado del aire, 2008. URL <https://biblio.flacsoandes.edu.ec/catalog/resGet.php?resId=16377>. Informe del Estado y Perspectiva del Ambiente.

EMOV. Informe de calidad del aire cuenca 2018, 2018. URL <https://www.emov.gob.ec/sites/default/files/CALIDAD%20DEL%20AIRE%202018.pdf>. EMOV monitorea constantemente la calidad del aire de Cuenca.

X. Ramos. Calidad del aire e impacto en guayaquil, 2021. URL <https://www.eluniverso.com/noticias/informes/conozca-los-23-puntos-de-guayaquil-donde-se-haran-mediciones-periodicas-de-la-c> EL UNIVERSO CALIDAD DEL AIRE.

J. C. H. Romero, D. M. Uribe, and C. M. Méndez. Comportamiento del monóxido de carbono y el clima en la ciudad de toluca, de 1995 a 2001. *CIENCIA ergo-sum, Revista Científica Multidisciplinaria de Prospectiva*, 11(3):263–274, 2004.

- M. Romero Placeres, F. Diego Olite, and M. Álvarez Toste. La contaminación del aire: su repercusión como problema de salud. *Revista Cubana de Higiene y Epidemiología*, 44:0 – 0, 08 2006. ISSN 1561-3003. URL [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1561-30032006000200008&nrm=iso](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1561-30032006000200008&nrm=iso).
- J. M. Seguel, R. Merrill, D. Seguel, and A. C. Campagna. Indoor air quality. *American journal of lifestyle medicine*, 11(4):284–295, 2017.
- R. C. Trelles Motte. Determinación del material particulado (pm10 y pm 2.5), dióxido de azufre (so2), dióxido de nitrógeno (no2) y monóxido de carbono (co) en el distrito de ocoruro-provincia espinar-región cusco. 2018.
- A. Valle. Mejore la calidad del aire en su planta industrial, 2017. URL <https://fierrosindustrial.com/noticias/mejore-la-calidad-del-aire-en-planta-industrial/#comments>. Antecedentes e institucionalidad.
- C. A. S. Walden. Publicación de contaminantes atmosféricos de la estación de monitoreo de la ciudad de cuenca, utilizando servicios estándares ogc. *ACI Avances en Ciencias e Ingenierias*, 9(15), 2017.

## 6 Biografía



**Andreina Nathaly Santana Andrade.** La autora es estudiante del programa de maestría en electrónica y automatización. Ella obtuvo su título de Ingeniera Electrónica, mención Sistemas Industriales en el 2018 por la Universidad Politécnica Salesiana. Actualmente, se desempeña como Ingeniera de Network Operation Center NOC en Telconet Latam.



**Paulina Adriana Morillo Alcívar.** obtuvo su grado de Ingeniería en Electrónica y Telecomunicaciones por la Escuela Politécnica Nacional (2012) y su posgrado en Gestión de la Información por la Universitat Politècnica de València (2016). Actualmente trabaja como profesora e investigadora en la Universidad Politécnica Salesiana (UPS). Posee experiencia en los campos de aprendizaje de máquina, matemática computacional y métodos numéricos.