

UNIVERSIDAD PERUANA UNIÓN
FACULTAD DE INGENIERIA Y ARQUITECTURA
Escuela Profesional de Ingeniería Ambiental



**Modelo predictivo con machine learning para variables
ambientales y PM 2.5 en Huachac Junín - Perú**

Tesis para obtener el Título Profesional de **Ingeniero Ambiental**

Autor:

Emery Elsa Olarte Suarez
Jhonatan Michael Gutierrez Gutierrez
Gwayne Kelly Roque Párraga

Asesor:

Ing. Orlando Alan Poma Porras

Lima, Abril de 2024

DECLARACIÓN JURADA DE ORIGINALIDAD DE TESIS

Yo, Orlando Alan Poma Porras, docente de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería Ambiental, de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **“MODELO PREDICTIVO CON MACHINE LEARNING PARA VARIABLES AMBIENTALES Y PM 2.5 EN HUACHAC JUNIN - PERU”** del (los) autor (autores) Emery Elsa Olarte Suarez, Jhonatan Michael Guterrez Gutierrez y Gwayne Kelly Roque Parraga, tiene un índice de similitud de 16 % verificable en el informe del programa Turnitin, y fue realizada en la Universidad Peruana Unión bajo mi dirección.

En tal sentido asumo la responsabilidad que corresponde ante cualquier falsedad u omisión de los documentos como de la información aportada, firmo la presente declaración en la ciudad de Lima, a los 30 días del mes de abril del año 2024.



Ing. Orlando Alan Poma Porras

ACTA DE SUSTENTACIÓN DE TESIS

En Lima, Ñaña, Villa Unión, a 23 día(s) del mes de abril del año 2024 siendo las 7:32 horas, se reunieron los miembros del jurado en la Universidad Peruana Unión Campus Lima, bajo la dirección del (de la) presidente(s):

Mg. Milda Amparo Cruz Huamanga el (la) secretario(a): Mg. Liliana Del Carmen Gutiérrez Rodríguez y los demás miembros: Mg. Joel Hugo Ferrández Rojas Mg. Jackson Edgardo Pérez Caspio y el (la) asesor(a) Ing. Orlando Alan Porras Porras

con el propósito de administrar el acto académico de sustentación de la tesis titulado: "Modelo predictivo con machine learning para variables ambientales y PM2.5 en Huachac, Junín - Perú"

del(los) bachiller(es): a) Gwayne Kelly Roque Parraga b) Jonathan Michael Gutiérrez Gutiérrez c) Emery Elsa Olarte Suarez

conducente a la obtención del título profesional de: Ingeniero Ambiental (Denominación del Título Profesional)

El Presidente inició el acto académico de sustentación invitando al (a la) / a (los) (las) candidato(a)s hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del jurado a efectuar las preguntas, y aclaraciones pertinentes, las cuales fueron abueeltas por al (a la) / a (los) (las) candidato(a)s. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Bachiller (a): Gwayne Kelly Roque Parraga

| CALIFICACIÓN | ESCALAS | | | Mérito |
|--------------|-----------|---------|-------------|---------------|
| | Vigesimal | Literal | Cualitativa | |
| Aprobado | 18 | A- | Muy bueno | sobresaliente |

Bachiller (b): Jonathan Michael Gutiérrez Gutiérrez

| CALIFICACIÓN | ESCALAS | | | Mérito |
|--------------|-----------|---------|-------------|---------------|
| | Vigesimal | Literal | Cualitativa | |
| Aprobado | 18 | A- | Muy bueno | sobresaliente |

Bachiller (c): Emery Elsa Olarte Suarez

| CALIFICACIÓN | ESCALAS | | | Mérito |
|--------------|-----------|---------|-------------|---------------|
| | Vigesimal | Literal | Cualitativa | |
| Aprobado | 18 | A- | Muy bueno | sobresaliente |

(*) Ver parte posterior

Finalmente, el Presidente del jurado invitó al (a la) / a (los) (las) candidato(a)s a ponerse de pie, para recibir la evaluación final y concluir el acto académico de sustentación procediéndose a registrar las firmas respectivas.

Presidente/a

Asesor/a

Bachiller (a)

Miembro

Bachiller (b)

Miembro

Bachiller (c)

[Firma]
Secretaria

* Esta sustentación fue realizada de manera virtual de forma sincrónica conforme al Reglamento General de Grados y Títulos.

Modelo predictivo con Machine Learning para variables ambientales y PM2.5 en Huachac Junín – Perú

Emery Olarte¹, Jhonatan Gutierrez², Gwayne Roque³, Orlando Poma Porras^{4*}

¹ Escuela profesional de Ingeniería Ambiental, Universidad Peruana Unión, Lima Perú; elsaolarte@upeu.edu.pe

² Escuela profesional de Ingeniería Ambiental, Universidad Peruana Unión, Lima Perú; jhonatangg@upeu.edu.pe

³ Escuela profesional de Ingeniería Ambiental, Universidad Peruana Unión, Lima Perú; gwayneroque@upeu.edu.pe

⁴ Escuela profesional de Ingeniería Ambiental, Universidad Peruana Unión, Lima Perú; opoma@upeu.edu.pe

* Correspondence: opoma@upeu.edu.pe; Tel.: +51969332640

Abstract: (1) **Antecedentes:** La contaminación por PM2.5 está en aumento provocando problemas de salud, con el objetivo de modelar el comportamiento del PM2.5AQI (índice de calidad de aire) mediante machine learning se aplicó modelos predictivos de regresión Lineal, Lasso, Ridge y Elastic Net considerando 16, 543 registros de la zona de Huachac-Junin en Perú con regresores de humedad en g/m³ y la temperatura en °C. El punto focal del estudio son el material particulado y variables ambientales, donde se identificó un nuevo modelo de Machine Learning ML capaz de predecir el PM2.5 de la zona de Huachac con considerable precisión, basado en variables regresoras estrechamente relacionadas con el material particulado. (2) **Métodos:** Se aplicó el Análisis Exploratorio de Datos (EDA), el cual permitió una visión clara de los datos, realizando un análisis descriptivo, un ajuste de los tipos de variables, la detección y tratamiento de datos faltantes, la identificación de datos atípicos y la correlación de variables, permitiendo tener una data sólida en la predicción de los modelos predictivos de machine Learning aplicados; (3) **Resultados:** El índice de la calidad del aire tiene tendencia a ser más alta en invierno y primavera con medias de 52.6, 36.9 PM2.5AQI respectivamente y con valores más bajos en verano además presenta valor máximo en septiembre y presenta valores mínimos en febrero. El uso de los modelos de regresión arrojó métricas precisas para elegir el mejor modelo para la predicción del PM2.5AQI. La comparación con otras investigaciones resalta la solidez de los modelos de ML elegidos, subrayando el potencial del ML en la ingeniería ambiental. (4) **Conclusiones:** El modelo predictivo encontrado fue $\alpha=0.1111111$ y un valor de Lambda $\lambda=0.150025$ representado por $PM2.5AQI = 83.0846522 - 10.3022000(\text{Humedad}) - 0.1268124(\text{Temperatura})$ con un R² ajustado de 0.1483206 y un RMSE de 25.36203.

Keywords: Lasso; Ridge; Elastic Net; PM2.5AQI; Temperatura, Humedad, Aprendizaje automático

1. Introducción

Uno de los problemas de carácter global es la contaminación atmosférica, siendo uno de los principales riesgos para la salud pública, situándose como la causa más importante de decesos a nivel mundial [1]. Este es un problema global que afecta tanto a los países desarrollados, especialmente en sus grandes ciudades, como a los países en vías de desarrollo, así la contaminación atmosférica puede tener diversos efectos adversos en el entorno natural, destacándose en primer lugar, la lluvia acida, smog, pérdida de biodiversidad, cambio climático y agotamiento de la capa de ozono [2].

La contaminación atmosférica es un problema de índole mundial, donde los principales contaminantes que afectan la atmosfera están los gases (CO, NO₂, SO₂, Pb.) y el material particulado PM₁₀ y PM_{2.5} [3]. Por lo general estos son liberados por fuentes antrópicas y naturales, en donde los estudios de morbilidad respiratorias están implicadas para ciertos tipos de contaminantes del aire conforme a la Organización Mundial de la Salud(OMS), siendo el principal la materia particulada, considerandose el PM_{2.5} [4]. Las partículas PM₁₀ y PM_{2.5} son