

Regresión y *machine learning* para predicción de áreas de incendio forestal

Sebastián André Lizarraga Calderón
Computer Science
UTEC
Lima, Perú
sebastian.lizarraga@utec.edu.pe

Carlos Esteban Guerrero Robles
Computer Science
UTEC
Lima, Perú
carlos.guerrero@utec.edu.pe

I. INTRODUCCIÓN

Según un informe de la UNEP (*United Nations Environment Programme*, Programa de las Naciones Unidas para el Medio Ambiente) el cambio climático generado por la actividad humana ocasionará que los incendios forestales aumente un 30% para hasta el 2050 [1]. Además se sugiere que para mitigar el daño de estos incendios es necesario un cambio en las estrategias actuales basándolas en prevención y preparación acompañada de presupuestos previos, pues los costes de reconstrucción suele estar fuera del alcance de países con rentas bajas. [1].

Siendo Perú uno de estos países, es necesario tomar acciones al respecto. En especial con el registro de un incidente de este tipo hace menos de un mes y que según el informe del INDECI (Instituto Nacional de Defensa Civil) tuvo dos personas damnificadas y una vivienda destruida [2].

En este sentido, la predicción del área y posible daño que puede ocasionar un incendio forestal se vuelve de vital importancia para prevenir este tipo de desastres. Dicha predicción podría ser realizada haciendo uso de *machine learning* basada en regresiones, pues existen magnitudes relacionadas a los lugares del suceso (i.e. humedad relativa, temperatura, etc.) que presentan correlación con el área afectada en dichos siniestros.

II. MARCO TEÓRICO

II-A. Regresión lineal

La regresión lineal plantea la hipótesis de existencia de una función lineal que aproxima con el menor error posible todos los puntos conocidos de una función real e incluso predice la localización de un punto futuro basándose en sus variables independientes.

Por ejemplo, el caso más sencillo es se presenta cuando las variables independientes tienen una sola dimensión. En este caso la regresión plantea una hipótesis sobre la existencia de una recta que aproxima la función real de la siguiente manera:

$$h(x_i) = x_i * w + b \quad (1)$$

Donde:

- w es el peso de dicha variable independiente
- b es el *bias* de la recta

II-A1. Regresión lineal multivariable: Sin embargo, muchas veces una variable dependiente es producto de varias variables independientes. En ese sentido debemos extender el concepto planteado por la Ecuación 1 para que sea capaz de soportar la existencia de múltiples variables.

Así mismo, para facilitar la representación añadiremos una variable independiente más a cada entrada y cuyo valor será 1 para todos los casos, con la finalidad de abstraer el *bias*. Así la ecuación quedaría de la siguiente forma:

$$h(x_i) = \sum_{j=0}^k x_i^j w_j \quad (2)$$

Donde:

- k es la cantidad total de variables independientes de nuestro sistema

III. IMPLEMENTACIÓN

Este proyecto fue implementado haciendo uso de un Jupyter Notebook en Datalore. La versión más reciente de ese Notebook puede ser accedido en el siguiente link.

Adicionalmente, se cuenta con un <https://github.com/201710424-utec/ia-CS2601-proy1> donde se almacena una copia del Notebook trabajado y una versión exportada en PDF. Además de incluir una carpeta con el dataset usado para el proyecto.

REFERENCIAS

- [1] UNEP, *Spreading like Wildfire – The Rising Threat of Extraordinary Landscape Fires*. United Nations Environment Programme, 2022. dirección: https://wedocs.unep.org/bitstream/handle/20.500.11822/38372/wildfire_RRA.pdf.

- [2] INDECI, *INFORME DE EMERGENCIA N° 1338 - 30/8/2022 / COEN - INDECI / 16:15 HORAS (Informe N° 1)*. Instituto Nacional de Defensa Civil, 2022, pág. 2. dirección: <https://portal.indeci.gob.pe/wp-content/uploads/2022/08/INFORME-DE-EMERGENCIA-N%C2%BA-1338-30AGO2022-INCENDIO-FORESTAL-EN-EL-DISTRITO-DE-CURAHUASI-APUR%C3%8DMAC-1.pdf>.