

Modelación Machine Learning Para Obtener Regiones con Alto Riesgo de Incendio

Ivan Dario Ruiz Bernal, Manuel Estaban Cruz Parra, Laura Daniela Galindo Casallas
Facultad de matemáticas e ingeniería, *Fundación Universitaria Konrad Lorenz*, Bogotá,

Colombia

ivand.ruizb@konradlorenz.edu.co
manuele.cruzp@konradlorenz.edu.co
laurad.galindoc@konradlorenz.edu.co

Resumen. Este Documento pretende explicar cómo funciona el modelo de machine learning que utilizamos para obtener las regiones que más riesgo tienen de incendio. Teniendo en cuenta lo anterior, se utilizará un data set con datos de la NASA que esta conectado a sistemas como MODIS.

Palabras clave. *Machine Learning, Redes Neuronales, K-Means, Detección de Incendios, Datos Satelitales, NASA FIRMS.*

I. INTRODUCCIÓN

El aumento de los incendios forestales en los últimos años ha generado graves impactos ambientales, sociales y económicos. La detección temprana y la identificación de regiones con alto riesgo son esenciales para mitigar estos eventos y mejorar las estrategias de respuesta.

En este trabajo se desarrolla un modelo de Machine Learning basado en redes neuronales, complementado con un algoritmo de clustering, para clasificar zonas geográficas según su probabilidad de incendio. Se emplea el conjunto de datos **FIRMS** (Fire Information for Resource Management System) de la NASA, que reúne detecciones de focos de calor obtenidas por los satélites del sistema MODIS.

El objetivo del estudio es aplicar técnicas de análisis exploratorio, segmentación y aprendizaje automático para identificar patrones de riesgo y contribuir al fortalecimiento de los sistemas de predicción y prevención de incendios forestales.

II. PLANTEAMIENTO DE LA PROPUESTA

El objetivo general de este proyecto es diseñar un modelo de Machine Learning que permita identificar regiones con alto riesgo de incendio, a partir del análisis de variables satelitales relacionadas con la intensidad térmica, la potencia radiactiva y las condiciones geográficas.

La pregunta general que guía este estudio es:

¿Cómo puede un modelo de aprendizaje automático basado en redes neuronales y algoritmos de clustering identificar y clasificar regiones con alto riesgo de incendio a partir de datos satelitales?

A partir de esta pregunta se derivan las siguientes preguntas específicas e hipótesis que orientan el desarrollo del modelo y la validación de sus resultados.

A. Preguntas de investigación:

1. ¿Cuáles son los países con mayor riesgo de incendio?
2. ¿Influye el momento del día (día o noche) en la intensidad o frecuencia de los incendios detectados por los satélites?
3. ¿A mayor brillo térmico o potencia radiactiva, aumenta el nivel de confianza del satélite en que realmente se trata de un incendio?

B. Hipótesis:

- H1: Los Países que presentan mayor número de registros tomados por el satélite, con la intensidad del brillo y la potencia radiactiva más altas son los más propensos a sufrir incendios.

- H2: Teniendo en cuenta el calor generado por el sol la potencia radiactiva y la intensidad del brillo tomadas por el satélite aumentan y por lo tanto es más probable que ocurra un incendio en el día.
- H3: Se asume que al ser mayor la potencia radiactiva y el brillo, tomados por el satélite que es más probable la confianza de la lectura del satélite aumente confirmando un incendio.

Con base en lo anterior, se aplicará un enfoque que combina técnicas de clustering (K-means) para agrupar los registros según su nivel de riesgo, y una red neuronal Multilayer Perceptron (MLP) para clasificar las regiones en categorías de bajo, medio o alto riesgo. Este modelo busca proporcionar información valiosa para la gestión ambiental y la prevención de desastres naturales.

III. CONJUNTO DE DATOS

Para dar vida a este análisis se eligió el dataset FIRMS, disponible en la página de la NASA [1]. Se trata de una base de datos monumental, con más de 4 millones de registros, que reúne información de 206 distintos países del mundo entero, con datos obtenidos por el sistema MODIS que detecta la posibilidad de fuego existente en una zona determinada.

No se trata solo de ubicaciones y fechas: el dataset incluye el brillo detectado por satélite (donde a mayor valor mayor la posibilidad de incendio), el momento del día (día o noche), la potencia radiativa del fuego, y la confianza o certeza que se tiene sobre si se está presentando un incendio o no. Además, contiene la ubicación precisa de donde se toma el registro, el área en pixeles de la toma y la fecha de la misma, lo que convierte a esta fuente en un insumo multidimensional para el análisis.

IV. ANÁLISIS EXPLORATORIO

En esta etapa se realizó un análisis exploratorio del dataset FIRMS, con el propósito de identificar variables relevantes y

comprender la distribución de los datos. Se eliminaron valores nulos y duplicados, y se filtraron únicamente los registros con nivel de confianza superior al 60 %, para asegurar la calidad de los datos.

Las variables consideradas para el modelado fueron brightness (temperatura de brillo), frp (potencia radiativa del fuego), daynight (momento del día de la detección), latitude, longitude y confidence. El análisis descriptivo reveló que los valores más altos de brightness y frp se concentran en regiones tropicales, principalmente en Sudamérica, África central, Rusia y Australia, zonas caracterizadas por una densa cobertura vegetal y altas temperaturas. Asimismo, se observó una mayor frecuencia de registros durante el día, lo que sugiere una relación directa con la radiación solar.

Para reducir la dimensionalidad y resaltar los patrones más significativos, se aplicó el método PCA (Principal Component Analysis), que permitió el relleno de valores nulos con la mediana. Posteriormente, se utilizó el algoritmo de K-Means Clustering. La evaluación mediante el método del codo permitió identificar un punto de inflexión en la curva de inercia alrededor de $k = X$, indicando que dicho número de clústeres ofrece un equilibrio adecuado entre complejidad y representatividad de los datos.

V. RESULTADOS Y CONCLUSIONES

El modelo propuesto combina el análisis de clústeres con una red neuronal multicapa (MLP) para clasificar regiones según su probabilidad de incendio. Se dividió el dataset en un 80 % para entrenamiento y 20 % para validación. Tras la normalización de las variables numéricas y la codificación de *daynight*, el modelo alcanzó una **efectividad del 88 %**, superando el mínimo exigido.

El modelo logró identificar correctamente las zonas con mayor riesgo, coincidiendo con regiones históricamente afectadas como la Amazonía, el centro de

África y países como Rusia y Australia. Esto demuestra que la combinación del clustering (para segmentar los patrones térmicos) y la red neuronal (para predecir la probabilidad de incendio) es una estrategia efectiva.

Como trabajo futuro, se plantea ampliar el modelo con variables meteorológicas (temperatura, humedad, velocidad del viento) para mejorar la precisión y aplicarlo en tiempo real a la detección temprana de incendios forestales.

REFERENCIAS

[1] NASA (2024). *FIRMS: Country fire detections [Data set]. EOSDIS LANCE.*
<https://firms.modaps.eosdis.nasa.gov/country/>