

Predicción de incendios forestales mediante modelos de Machine Learning

Ana Martínez Saucedo¹ y Pablo Ezequiel Inchausti¹

Universidad Argentina de la Empresa (UADE). Instituto de Tecnología (INTEC).
Buenos Aires, Argentina
{anmartinez,pinchausti}@uade.edu.ar

Resumen La severidad de los incendios forestales ha llegado a niveles preocupantes tanto a nivel internacional como nacional. No obstante, gracias al avance de la tecnología es posible predecir su ocurrencia y magnitud a través de modelos de Machine Learning especialmente desarrollados para tal fin. En línea con diversas investigaciones realizadas en materia de predicción espaciotemporal de incendios forestales, en el presente trabajo el objetivo fue desarrollar un modelo de Machine Learning que contribuya a la prevención de incendios forestales en el Partido de Pinamar. Para ello se entrenaron diversos modelos utilizando registros de incendios históricos de la zona, alcanzando una sensibilidad del 88.4 % para predecir la ocurrencia de incendios forestales a través de un árbol de decisión. Gracias al desarrollo de un pipeline de datos y el entrenamiento automatizado de modelos se sentaron las bases necesarias para posibilitar la predicción de incendios forestales en localidades vecinas.

Palabras claves: machine learning, aprendizaje supervisado, incendios forestales, medioambiente.

1. Introducción

Las consecuencias ambientales, económicas y sociales que los incendios forestales provocan en el mundo llevaron a gobiernos e investigadores a estudiar diversas maneras de prevenirlos, sobre todo en los últimos tiempos donde cada vez se torna más difícil controlarlos. Tan solo en el 2020 en Argentina se quemaron más de 1,1 millones de hectáreas a causa de incendios forestales. Además, se registraron en ese año más de 74.113 focos activos, una cifra récord que representa un incremento del 251,9 % con respecto al año anterior [1].

Esta tendencia nacional también se vio reflejada en distintos puntos del país. La ciudad balnearia de Pinamar, conocida por sus bosques de pino y dunas de arena, ha perdido según datos de incendios provistos por los bomberos locales más de 3,5 kilómetros cuadrados de bosques en los últimos seis años a causa de incendios forestales. Esta cifra representa aproximadamente el 10 % de la superficie total cubierta por vegetación del partido, confirmando así la tendencia creciente de incendios forestales en la zona detectada por bomberos y autoridades locales. Si bien en Argentina el 95 % de los incendios forestales son originados

por el hombre [2], su magnitud y desarrollo dependen en gran medida de las condiciones climáticas y ambientales del momento y lugar donde se desarrollan.

En la literatura se han desarrollado modelos de Machine Learning (ML) que utilizan todas o combinaciones de algunas de las variables que conforman el denominado Triángulo de Comportamiento del Fuego (TCF) para predecir incendios forestales, siendo estas la topografía, el combustible y la meteorología. En efecto, las mismas definen bajo qué condiciones es más propenso que se produzca un incendio forestal y cómo se desarrollará el mismo. Además, el Índice Meteorológico de Peligro de Incendio (FWI) [3] permite estimar la probabilidad de que se produzca un incendio a partir de variables meteorológicas y de combustible [4,5]. Asimismo, el Índice de Vegetación de Diferencia Normalizada (NDVI) ha sido utilizado en varios estudios ya que es un indicador de la salud de la vegetación [6,7]. Los investigadores han recurrido a fuentes como satélites, sensores remotos, mapas y estaciones meteorológicas para obtener estos datos y entrenar modelos predictivos. Entre los algoritmos aplicados para modelar la ocurrencia de incendios forestales se encuentran árboles de decisión, máquinas de vectores de soporte, redes neuronales artificiales y redes bayesianas [8].

No obstante, la mayor dificultad a la hora de estudiar la predicción de ocurrencia y magnitud de incendios forestales radica en la obtención de datos históricos. En Argentina son pocos los cuarteles de bomberos que cuentan con un registro informatizado de incendios, y las investigaciones realizadas en el país han recurrido al uso de datasets de otros países, como el del parque Montesinho de Portugal que abarca los incendios ocurridos en el mismo entre los años 2000 y 2003 [9]. Por ello, y en línea con diversos trabajos realizados con el fin de determinar la relación que existe entre la superficie final quemada a causa de incendios forestales y las condiciones ambientales circundantes, el presente trabajo tuvo como objetivo desarrollar y evaluar modelos de ML para predecir la ocurrencia y magnitud de incendios forestales en la ciudad de Pinamar, provincia de Buenos Aires, utilizando datos de incendios locales recopilados por bomberos de la zona y no sólo variables meteorológicas, sino también topográficas y de combustible.

El presente artículo se organiza de la siguiente manera: en la sección 2 se delimita el área de estudio del trabajo y las variables explicativas utilizadas. En la siguiente sección 3 se describen los distintos algoritmos de ML utilizados y las métricas empleadas para su evaluación. Luego, en la sección 4 se explica la metodología adoptada en el trabajo. A continuación, en la sección 5 se expone y analiza el rendimiento de los modelos entrenados. Por último, en la sección 6 se resumen los resultados obtenidos y se presentan futuras líneas de investigación.

2. Área de estudio y datos utilizados

El partido de Pinamar está ubicado en la zona sudeste de la provincia de Buenos Aires y se caracteriza por una gran predominancia de coníferas implantadas a lo largo de 40 kilómetros cuadrados, representando el 63 % de la superficie total del partido. Según los datos provistos los bomberos locales, los incendios forestales de mayor magnitud (>1 hectárea) ocurren en primavera y verano durante la franja horaria de 8 de la mañana a 3 de la tarde. Coincidentemente, las

zonas donde mayor cantidad de incendios forestales han ocurrido corresponden a aquellas donde la densidad poblacional o concentración de actividades turísticas en temporada alta y fines de semana largos es mayor.

Esta información permitió determinar qué variables pueden explicar la ocurrencia de incendios forestales en Pinamar. Para construir el dataset se desarrolló un pipeline de datos automatizado que extrae, transforma y preprocesa las variables meteorológicas, topográficas y de combustible como el FWI, el código de sequía (DC), el código de humedad del mantillo (DMC), la velocidad de propagación del incendio (ISI) y el combustible disponible (BUI) (Tabla 1) para un rango de fechas y par de coordenadas establecidas. En el presente estudio el dataset creado corresponde a los incendios forestales ocurridos en el área descrita por las coordenadas 56°57' O, 37°12' S, 56°48' O, 37°3' S y desde el 01/01/2015 hasta el 01/01/2020.

Tabla 1. Descripción de variables utilizadas en el dataset creado.

Descripción	Variables	Origen	Resolución espacial	Resolución temporal	Cobertura geográfica
Incendios forestales	Día	Asociación	N/A	Horaria	Partido de Pinamar
	Mes	Bomberos			
	Día no laboral	Voluntarios de Pinamar			
	Hora				
	X, Y				
Meteorología	Temperatura	Servicio Meteorológico Nacional	N/A	Horaria	Partidos de Pinamar y Villa Gesell
	Humedad				
	Viento				
	Precipitaciones				
Temperatura de superficie de suelo (LST)	LST	NASA / MYD11C1 v006 [10]	0.05° × 0.05°	Diaria	Global
Índice de Vegetación de Diferencia Normalizada	NDVI	NASA / MYD13Q1 v006 [11]	250 metros × 250 metros	Quincenal	Partido de Pinamar
Índice meteorológico de peligro de incendio e índices derivados	DC	NASA / GFWED GEOS-5 - GPM Late v5 [12]	0.1° × 0.1°	Diaria	60° S - 60° N
	DMC				
	FFMC				
	ISI				
	BUI				
	FWI				

3. Métodos

Tal como se describió anteriormente, en la literatura se ha abordado la predicción de ocurrencia y magnitud de incendios forestales utilizando diversos algoritmos. En particular, en este estudio se han implementado algunos de ellos para la predicción de tanto la ocurrencia de incendios forestales (clasificación),

como de la superficie final quemada (regresión). A su vez se han variado los hiperparámetros utilizados en cada modelo hasta encontrar una configuración que provea los mejores resultados. A continuación se describe brevemente cada uno.

3.1. Regresión logística o *Logistic Regression* (LR)

Los modelos de LR explican la relación que existe entre una variable dependiente y varias independientes o explicativas. Este es uno de los algoritmos de clasificación supervisados más sencillos y utilizados en ML. Los modelos de LR son lineales, por lo que se obtienen buenos resultados cuando los datos son linealmente separables a través de una frontera de decisión. No obstante, para modelos más complejos en donde hay múltiples fronteras de decisión los modelos de LR no pueden capturar la complejidad de las relaciones en los datos [13].

En este trabajo se aplicó la regularización L2 para evitar que el modelo de LR se sobreajuste a los datos de entrenamiento y se varió el hiperparámetro C en 100 valores espaciados en el intervalo comprendido entre 0.00002 y 1.

3.2. Máquinas de Vectores Soporte o *Support Vector Machines* (SVM)

SVM busca determinar los puntos (o vectores de soporte) que separan al máximo dos clases, aunque se puede generalizar a múltiples clases. Para ello cada ejemplo se representa como un vector en un espacio n -dimensional, donde n es la cantidad de atributos o características. Los vectores de soporte definen de esta forma un hiperplano que separa los datos linealmente [14].

Entre los hiperparámetros que se calibraron para tanto el clasificador como el regresor se encuentran la función kernel (lineal, polinómico y *Radial Basis Function* o RBF), el valor gamma del kernel (4 valores espaciados entre 0.0001 y 1) y C (10 valores espaciados entre 0.1 y 10000).

3.3. Árboles de decisión o *Decision Tree* (DT)

Un DT se representa mediante un árbol binario donde cada nodo representa un atributo de entrada. En caso de que la variable de entrada sea numérica, el nodo también es un punto de división sobre esa variable. Por otro lado, las hojas del árbol representan las salidas, que pueden ser tanto etiquetas para problemas de clasificación como un valor continuo para problemas de regresión. El objetivo del mismo es capturar las relaciones entre las variables de entrada y salida mediante el árbol más pequeño posible, de forma tal de evitar caer en un estado de sobreajuste de los datos de entrada [15].

En el presente estudio se variaron hiperparámetros como la profundidad máxima del árbol (10, 6, 3, sin límite), la cantidad máxima de características a considerar (33, 26, 20, 10 o 5), el número mínimo de ejemplos en el cual un nodo se considera hoja (1, 3, 5, 9 o 10), y el criterio de ganancia de información (Gini y Entropía en el caso de los clasificadores, y MAE y Friedman para los regresores).

3.4. Redes Neuronales Artificiales o *Artificial Neural Networks* (ANN)

Una ANN es un modelo que simula el comportamiento de una red neuronal biológica. La misma está compuesta por capas de nodos, particularmente una capa de entrada, una o más capas ocultas y una capa de salida. Cada nodo está conectado con otro, y posee un peso y umbral, especificando si el mismo está activado y envía datos a la siguiente capa de la red. Por el contrario, si está desactivado no se envían datos a la siguiente capa [16].

La arquitectura de red adoptada en los experimentos llevados a cabo es de dos capas ocultas, variando la cantidad de nodos de cada una entre 4 y 25 según la experiencia manifestada en la literatura [6,17]. Otros hiperparámetros ajustados fueron la tasa de aprendizaje α (5 valores espaciados entre 0.00001 y 0.001), la dilución (6 valores espaciados entre 0.15 y 0.5) y la función de activación (ReLU, Tangente Hiperbólica y Sigmoide).

3.5. Métodos de ensamble

Los métodos de ensamble de algoritmos de ML combinan las predicciones de múltiples modelos con el fin de obtener predicciones más precisas. Entre las técnicas de ensamble más comunes se encuentran:

Bagging: se construyen independientemente varios estimadores (ya sean clasificadores o regresores) utilizando el mismo algoritmo y distintos subconjuntos del conjunto de datos de entrenamiento. Una vez entrenados, las predicciones de cada modelo se promedian (en problemas de regresión) o votan (en problemas de clasificación) [18]. En este trabajo se utilizó el modelo de bosques aleatorios o *Random Forests* (RF) que conforman un ensamble de distintos árboles de decisión. Para ello se ajustaron hiperparámetros como la profundidad máxima de cada estimador (80, 90, 100, 110, 120 o 150), la cantidad máxima de características a considerar (4, 6, 12, 24 o 33), el número mínimo de ejemplos para dividir un nodo interno (8, 10 o 12), el número mínimo de ejemplos para determinar si un nodo es hoja (3, 4 o 5), y la cantidad de árboles en el bosque (100, 200, 500 o 1000).

Boosting: se desarrollan secuencialmente varios *weak learners* o modelos sencillos, buscando corregir en cada paso los errores del modelo predecesor. Un ejemplo de esta técnica es el algoritmo *Gradient Boosted Trees* (GB), en el que varios árboles de decisión individuales simples y con pocas ramificaciones se ajustan secuencialmente [19]. Además de los hiperparámetros ajustados en el modelo de DT, en GB se han ajustado la tasa de aprendizaje (0.01, 0.025, 0.05, 0.075, 0.1, 0.15 o 0.2), la función de calidad de división (Friedman, RMSE), y la cantidad de etapas de boosting (10, 50, 100, 200, 500 o 1000).

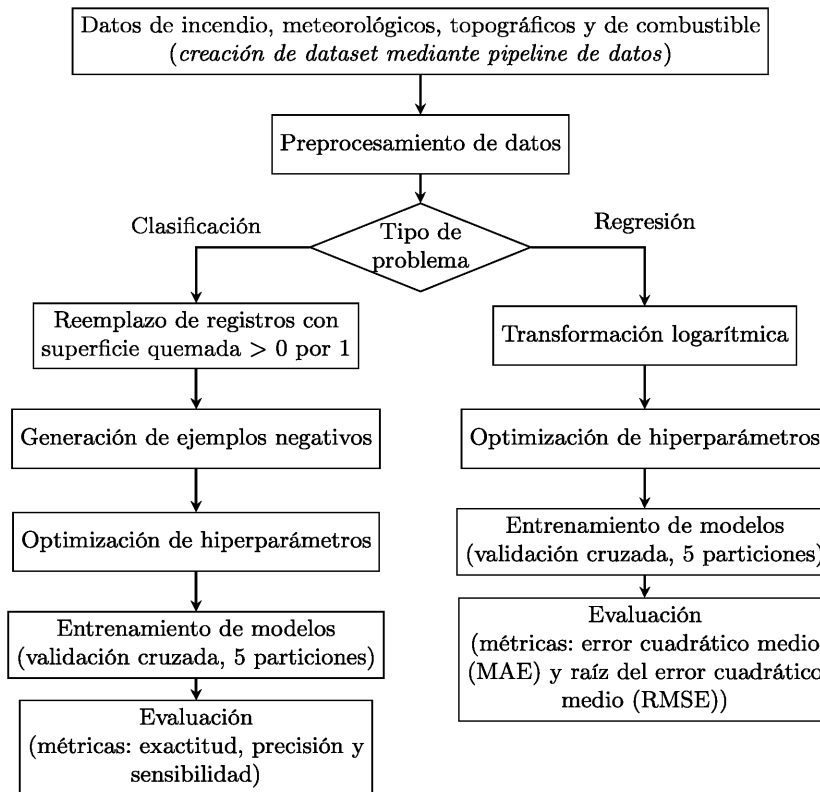
4. Metodología de modelado

La metodología adoptada en el presente trabajo se detalla en la Figura 1. En primer lugar se generó el dataset de incendios forestales de los años 2015 a 2020 para el área de interés. El mismo posee un total de 750 registros, aunque luego de la etapa de preprocesamiento sólo 597 se utilizaron para entrenar modelos

mientras que los restantes 163 registros se descartaron por no contar con alguno de los atributos de la Tabla 1 requeridos. Entre las tareas de preprocesamiento que se llevaron a cabo se destacan:

- La conversión de coordenadas geográficas (latitud y longitud) a coordenadas (X, Y) de 250 metros x 250 metros.
- La aplicación de técnicas de escalado de datos como *MinMax*.
- La limpieza de registros con datos meteorológicos, topográficos o de combustible faltantes.
- La conversión de variables categóricas a numéricas.

Figura 1. Diagrama de flujo de la metodología propuesta.



La experiencia manifestada en la literatura y en el presente trabajo demuestran que la distribución de la magnitud de incendios forestales es altamente despareja: la mayoría de los incendios forestales ocurridos en Pinamar han sido pequeños, aunque representan el 37 % de la superficie quemada históricamente. Por el contrario, el 61 % de la superficie quemada fue producida por el 6 % del total de incendios forestales. Esta disparidad afecta directamente el rendimiento de modelos que, en este tipo de distribuciones, se verán influenciados por lo que se

refleja en la mayoría de los casos, ya sea la cantidad de incendios o el promedio de superficie quemada.

Con el fin de abordar este problema, y en línea con trabajos como el de [20], el enfoque adoptado es el de dividir en dos pasos la predicción de incendios forestales: en primer lugar, se predice por cada coordenada de la grilla si se producirá o no un incendio forestal (clasificación binaria); y en segundo lugar, para aquellas coordenadas en las que se haya predicho la ocurrencia de un incendio forestal (con una probabilidad mayor a 0.5), se predicen las hectáreas que podrían quemarse (regresión). A partir de esta división, dependiendo del tipo de problema los pasos en la metodología difieren. Por un lado, para el problema de clasificación se reemplazan los registros cuyo valor de superficie quemada es mayor a 0 por 1. Luego, como los datos que se cuentan corresponden enteramente a la clase 1 (Incendio), se genera aleatoriamente una cantidad proporcionada de registros de la clase 0 (No Incendio) (Algoritmo 1) en concordancia con algoritmos propuestos en la literatura [21]. El objetivo de este tipo de algoritmos es que las ubicaciones de los puntos en donde se produjeron incendios y donde no estén espacial y temporalmente relacionados con los atributos. Por otro lado, para abordar el problema de regresión se aplica una transformación logarítmica sobre la variable objetivo (superficie quemada) para reducir la asimetría de la distribución. A continuación, en ambos problemas se optimizan los hiperparámetros de los distintos algoritmos seleccionados según lo descrito en la Sección 3 utilizando la técnica de búsqueda exhaustiva [22], y se entrenan modelos aplicando la técnica de validación cruzada utilizando 5 particiones sobre el conjunto de datos de entrenamiento. Por último, se evalúan los modelos entrenados a través de las métricas correspondientes al tipo de problema.

Algoritmo 1 Generación de ejemplos negativos ("no incendio")

```

1: dias  $\leftarrow$  3
2: kilometros  $\leftarrow$  1,5
3: P  $\leftarrow$   $\emptyset$  ▷ Puntos de "no incendio"
4: for puntoDeIncendio in incendios do
5:   fechaIncendio  $\leftarrow$  incendio.fecha
6:   incendiosRecientes  $\leftarrow$  seleccionarIncendiosEnLapso(dias, fechaIncendio)
7:   R  $\leftarrow$   $\emptyset$  ▷ Región donde se produjeron incendios
8:   for incendioReciente in incendiosRecientes do
9:     coordenadas  $\leftarrow$  incendioReciente.coordenadas
10:    R  $\leftarrow$  R  $\cup$  crearRegionEn(kilometros, coordenadas)
11:   end for
12:   P  $\leftarrow$  P  $\cup$  seleccionarCoordenadasEnArea(R')
13: end for

```

5. Resultados y discusión

Los resultados obtenidos en los modelos de clasificación y regresión se presentan en las Tablas 2 y 3 respectivamente. En lo que respecta a la predicción de incendios forestales (problema de clasificación), el modelo con mayor exactitud es el de RF con un valor del 82.4 %. Coincidentemente este modelo fue también el que obtuvo el valor de precisión más alto (82 %). No obstante, el modelo con

mayor sensibilidad es el de DT con un 88.4 %, convirtiéndolo en aquel que mejor distingue la ocurrencia de incendios. Tanto los modelos correspondientes a LR como SVM han tenido un rendimiento inferior en comparación al resto de los modelos.

La disparidad observada en el rendimiento de los distintos modelos puede tener relación con la complejidad y alta dimensionalidad del dataset, sumado a que la distribución de los datos no es normal. En efecto, DT es un algoritmo no paramétrico que no asume la distribución de los datos. No obstante, los algoritmos no paramétricos requieren datasets con mayor cantidad de registros, por lo que los resultados podrían mejorar en caso de añadir al dataset los incendios forestales registrados en los últimos dos años.

Tabla 2. Resultados de la evaluación de modelos de predicción de incendios forestales.

Modelo	Exactitud	Precisión	Sensibilidad	F1 score
ANN	74.2 %	71.6 %	79.3 %	75.3 %
SVM	61.5 %	62.4 %	56.2 %	59.1 %
GB	77.9 %	77.7 %	77.7 %	77.7 %
LR	62.3 %	63.6 %	56.2 %	59.7 %
DT	80.7 %	76.4 %	88.4 %	82 %
RF	82.4 %	82 %	82.6 %	82.3 %

En lo que respecta a modelos de predicción de superficie quemada, los mejores modelos en términos de valores de MAE y RMSE son ANN. Particularmente, el MAE más bajo corresponde a un valor de 0.255 mientras que el mejor RMSE es de 0.178. No obstante, los hiperparámetros de ambas redes difieren entre sí: mientras la red n°1 tiene 18 nodos en la capa oculta n°1 y 4 en la n°2, la red n°2 posee 9 y 18 respectivamente. Por otro lado, la red n°1 utiliza como función de activación la tangente hiperbólica, mientras que la n°2 utiliza la ReLu. También se puede observar que el RMSE obtenido a partir de los modelos de ANN es considerablemente mejor a los demás modelos.

En este sentido, una de las ventajas de las ANN radica en que su rendimiento no se ve alterado por altas correlaciones entre las variables de entrada. Efectivamente, a partir del análisis de datos previo al entrenamiento de modelos se observó una tendencia de que los incendios de mayor magnitud se producen en las primaveras y los veranos (correlación temporal). Asimismo, las localidades donde han ocurrido más incendios forestales corresponden a Ostende y Valeria del Mar, donde la densidad poblacional es mayor (correlación geográfica).

Luego del entrenamiento de estos modelos se pudo analizar cuáles son las variables que demostraron ser importantes a la hora de explicar la ocurrencia de incendios forestales en Pinamar. Correspondientemente a la experiencia manifestada por bomberos locales, la ubicación es un factor importante ya que los incendios son más comunes en zonas de alta densidad poblacional. En efecto, la variable longitud tuvo una importancia del 83 %. En segundo lugar, la variable elevación resultó en una importancia del 8 %. Por último, el NDVI obtuvo una importancia del 6 %. Estos resultados demuestran que el componente humano es el más importante a la hora de predecir incendios en Pinamar.

Tabla 3. Resultados de la evaluación de modelos de predicción de superficie quemada.

Modelo	MAE	RMSE
ANN (n°1)	0.295	0.178
ANN (n°2)	0.255	0.215
SVM	0.274	0.438
DT	0.294	0.46
RF	0.299	0.448

6. Conclusiones

En este estudio se desarrollaron y entrenaron varios modelos de ML para predecir tanto la ocurrencia como magnitud de incendios forestales en Pinamar. Estas predicciones pueden ser de gran importancia para los bomberos, ya que posibilitan estimar tempranamente los recursos humanos y materiales que deben ser empleados para combatir incendios forestales lo más rápido posible y así evitar pérdidas materiales, ecológicas y humanas. Los resultados obtenidos son variados, no obstante los mejores modelos alcanzaron una sensibilidad del 88.4 % (DT, clasificación) y un RMSE del 0.178 (RNA, regresión).

Como futuras líneas de investigación se contempla variar la arquitectura de las RNA para disminuir el RMSE asociado a la magnitud de incendios forestales. Asimismo, incrementar el tamaño del dataset para que incluya una cantidad balanceada de incendios de baja y alta magnitud podría mejorar el rendimiento de los modelos, ya que los datos utilizados en este estudio tienen una distribución despareja. En este sentido, gracias a la automatización de la metodología propuesta realizar estas tareas insumiría poco tiempo, dando lugar a la posibilidad de realizar experimentos más extensos. No obstante, resulta fundamental contar con un sistema informático que permita registrar fácilmente los incendios forestales ocurridos para poder refinar los modelos obtenidos.

Agradecimientos Los autores agradecen a la Universidad Argentina de la Empresa (UADE) y al Instituto de Tecnología (INTEC) por el apoyo brindado en el presente trabajo realizado en el marco del Proyecto Final de Ingeniería en Informática “AQUA: Desarrollo de un modelo de Machine Learning para prevenir incendios forestales en Pinamar”, articulado en el ACyT “Aplicaciones de Machine Learning para mejorar el uso de Recursos Naturales” (A21T03).

Referencias

1. Instituto Nacional de Pesquisas Espaciais: Monitoramento dos Focos Ativos por País - Programa Queimadas - INPE, https://queimadas.dgi.inpe.br/queimadas/portal-static/estatisticas/_países, last accessed 8 May 2021
2. Argentina.gob.ar: Causas de los incendios forestales, <https://www.argentina.gob.ar/sinagir/incendio-forestal/causas>, last accessed 28 Mar 2021
3. Taylor, S.W., Alexander, M.E., Taylor, S.W., Alexander, M.E.: Science, technology, and human factors in fire danger rating: the Canadian experience. *Int. J. Wildland Fire*. 15, 121–135 (2006). <https://doi.org/10.1071/WF05021>
4. Xie, Y., Peng, M.: Forest fire forecasting using ensemble learning approaches. *Neural Computing and Applications*. 31 (2019). <https://doi.org/10.1007/s00521-018-3515-0>
5. Castelli, M., Vanneschi, L., Popovič, A.: Predicting Burned Areas of Forest Fires: an Artificial Intelligence Approach. *fire ecol.* 11, 106–118 (2015). <https://doi.org/10.4996/fireecology.1101106>

6. Jafari Goldarag, Y., Mohammadzadeh, A., Ardakani, A.S.: Fire Risk Assessment Using Neural Network and Logistic Regression. *J Indian Soc Remote Sens.* 44, 885–894 (2016). <https://doi.org/10.1007/s12524-016-0557-6>
7. Sayad, Y.O., Mousannif, H., Al Moatassime, H.: Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal.* 104, 130–146 (2019). <https://doi.org/10.1016/j.firesaf.2019.01.006>.
8. Jain, P., Coogan, S.C.P., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D.: A review of machine learning applications in wildfire science and management. *Environ. Rev.* 28, 478–505 (2020). <https://doi.org/10.1139/er-2020-0019>
9. Cardenas, M., Castillo, J., Medel, R., Casco, O., Navarro, M., Gutierrez, S., Curti, A.: Sistema de predicción de incendios forestales para la provincia de Córdoba. Presented at the Congreso Nacional de Ingeniería en Informática / Sistemas de información , Salta (2016).
10. Wan, Zhengming, Hook, Simon, Hulley, Glynn: MYD11C1 MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 0.05Deg CMG V006, <https://lpdaac.usgs.gov/products/myd11c1v006/>, (2015). <https://doi.org/10.5067/MODIS/MYD11C1.006>
11. Didan, Kamel: MYD13Q1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 250m SIN Grid V006, <https://lpdaac.usgs.gov/products/myd13q1v006/>, (2015). <https://doi.org/10.5067/MODIS/MYD13Q1.006>
12. Field, R.D., Spessa, A.C., Aziz, N.A., Camia, A., Cantin, A., Carr, R., de Groot, W.J., Dowdy, A.J., Flannigan, M.D., Manomaiphiboon, K., Pappenberger, F., Tanpipat, V., Wang, X.: Development of a Global Fire Weather Database. *Natural Hazards and Earth System Sciences.* 15, 1407–1423 (2015). <https://doi.org/10.5194/nhess-15-1407-2015>.
13. Hosmer, D.W., Lemeshow, S.: Introduction to the Logistic Regression Model. In: *Applied Logistic Regression.* p. 1. John Wiley & Sons (2004).
14. Luger, G.: Machine Learning: Connectionist. In: *Artificial Intelligence: Structures and Strategies for Complex Problem Solving.* pp. 482–484. , Boston (2008).
15. Brownlee, J.: Classification and Regression Trees. In: *Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch.* pp. 72–74. Machine Learning Mastery (2016).
16. Haykin, S.: Introduction. In: *Neural Networks and Learning Machines.* pp. 21–24. New York (2008).
17. Vega-Garcia, C.: Applying neural network technology to human-caused wildfire occurrence prediction. *AI Applications.* 10, 9–18 (1996).
18. Casal, R.F., Bouzas, J.C., Fuente, M.O. de la: 3.1 Bagging. *Aprendizaje Estadístico.* (2021)
19. Casal, R.F., Bouzas, J.C., Fuente, M.O. de la: 3.4 Boosting. *Aprendizaje Estadístico.* (2021)
20. Wang, Q., Zhang, J., Guo, B., Hao, Z., Zhou, Y., Sun, J., Yu, Z., Zheng, Y.: CityGuard: Citywide Fire Risk Forecasting Using A Machine Learning Approach. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.* 3, 1–21 (2019). <https://doi.org/10.1145/3369814>.
21. Stojanova, D., Kobler, A., Ogrinc, P., Ženko, B., Džeroski, S.: Estimating the risk of fire outbreaks in the natural environment. *Data Min Knowl Disc.* 24, 411–442 (2012). <https://doi.org/10.1007/s10618-011-0213-2>.
22. Krauß, J.: Fundamentals in the Selection of Hyperparameter Optimization Techniques. In: *Optimizing Hyperparameters for Machine Learning Algorithms in Production.* pp. 41–42. Apprimus Wissenschaftsverlag (2022).