

# Índice general

<b>1. Introducción</b>	<b>3</b>
1.1. Objetivos . . . . .	7
<b>Bibliografía</b>	<b>9</b>



# Capítulo 1

## Introducción

### El fuego y los incendios

El fuego es un factor natural clave en los ecosistemas terrestres. Tan solo necesita de tres elementos básico que se encuentran en abundancia en la superficie de la Tierra: oxígeno, combustible y calor, por lo que no sorprende que su presencia en este planeta se remonte muy atrás en el tiempo. Hay pruebas de la existencia de fuego hace 400 millones de años y desde hace 350 millones de años se producen incendios en la Tierra de forma frecuente (Scott y Glasspool [12]).

El fuego ha condicionado la evolución y la dispersión de plantas, el desarrollo de los biomas, la formación de suelos y los ciclos hidrológicos y erosivos. Se trata, por tanto, de uno de los procesos platenarios clave. La presencia del fuego en los ecosistemas terrestres ha dado lugar a numerosas adaptaciones en los seres vivos (Mataix Solera y Cerdà [9]). En el caso de los ecosistemas mediterráneos, donde los incendios son frecuentes, la vegetación ha desarrollado adaptaciones que le permiten adaptarse al fuego. Un buen ejemplo de esta adaptación al fuego lo encontramos en el pino carrasco (*Pinus halepensis*), cuyas piñas que cuelgan de las ramas de la copa solo se abren al calor de las llamas.

El control del fuego fue clave para el desarrollo de la humanidad. Cocinar alimentos, abrir zonas de cultivos, facilitar el traslado de la población, quemar restos de cosechas o eliminar plagas son solo algunos de los usos que se le ha dado al fuego desde que los primeros homínidos lo descubrieron hace 1,5 millones de años (James et al. [6]). Sin embargo, la relación del hombre con el fuego parece haber cambiado radicalmente en los últimos años.

La industrialización de las sociedades modernas, la sustitución del uso de biomasa por combustibles fósiles y el éxodo rural junto con el abandono de la agricultura han provocado el cese de la explotación de los montes, generando grandes acumulaciones de biomasa que actúan como combustible para los incendios forestales (Mataix Solera y Cerdà [9]). De esta forma, el fuego ha pasado de ser una herramienta para el hombre, como ha sido a lo largo de la historia, a convertirse en un gran problema ambiental.

A todo esto hay que sumar los efectos del cambio climático, que a través del aumento de las temperaturas y la disminución de las precipitaciones, ha provocado que se alargue la estación de incendios y ha aumentado las situaciones de riesgo alto [10].

---

## Una visión desde Andalucía

En España, los incendios forestales figuran en el Plan Estratégico Estatal del Patrimonio Natural y de la Biodiversidad a 2030 como una de las principales presiones y amenazas para el patrimonio natural y la biodiversidad en España, considerándose “el principal elemento de degradación de los ecosistemas forestales, con importantes repercusiones sobre bienes e incluso vidas humanas”. En 2022, España fue el país más afectado por los incendios forestales en Europa (excluyendo a Ucrania), con un total de 315.705 ha quemadas (SANMIGUEL-AYANZ et al. [11]).

En Andalucía este problema toma una dimensión especial al tratarse de la segunda comunidad de España con más terreno forestal (cuenta con 4.325.378 ha de suelo forestal que suponen el 49.37 % de su superficie) y contar además con el *hotspot* de biodiversidad de Sierra Nevada, uno de los enclaves con mayor diversidad del continente.

En 2022, 15.786,64 ha fueron afectadas por el fuego en Andalucía, prácticamente el doble de la media anual de los 10 años anteriores, 8873,68 ha. Ese mismo año, el 91.91 % de las actuaciones forestales<sup>1</sup> con causa conocida fueron de origen antrópico. De estos, el 48.20 % fueron debidos a negligencias, el 38.40 % fueron intencionados y el 13.40 % fueron accidentales. Tan solo el 5.25 % de las actuaciones forestales que ocurrieron en Andalucía en 2022 fueron debida a causas naturales [1]. Estas cifras ponen de manifiesto la importancia de considerar el factor humano en el estudio de los incendios forestales.

En la siguiente cita, extraída de Moreno et al. [10], los autores enfatizan la necesidad de considerar el factor humano en la predicción de incendios forestales, e indican la existencia de factores que influyen en que unas determinadas zonas tengan un mayor riesgo de verse afectas por incendios forestales que otras, entre los que mencionan el tipo y la configuración de la vegetación:

*A pesar de la importancia de la meteorología en los incendios, la capacidad predictiva de la ocurrencia de incendios [...] en base a variables meteorológicas [...] suele ser baja. [...] Esto es debido a que la mayor parte de los incendios en España es de de origen humano, lo que dificulta su predictibilidad. Así, las igniciones no ocurren al azar, ni en el espacio ni en el tiempo. [...] El territorio no se quema de manera aleatoria, siendo normal que unas zonas ardan más que otras. [...] Unos tipos de vegetación suelen arder más frecuentemente que otros. [...] La probabilidad de que un incendio se propague se ve favorecida por la configuración espacial de las manchas de vegetación que conforman el paisaje.*

## La estadística, el *Machine Learning* y los incendios forestales

Los incendios forestales son un proceso sumamente complejo en el que se interrelacionan una gran cantidad de factores como la fuente de ignición, la composición del combustible, las condiciones meteorológicas o la orografía del terreno, además de la ya mencionada acción humana. Desde el estudio de los procesos de combustión que ocurren a escala molecular al estudio de la propagación de los incendios forestales, el estudio de los mismos puede abordarse desde numerosos puntos de vista y con distintos enfoques que van desde

---

<sup>1</sup>Según se indica en la Estadística General de Incendios Forestales (EGIF), se usa el término actuaciones forestales para englobar a conatos (de extensión inferior a 1 ha) e incendios forestales (de extensión superior a 1 ha)

---

la ecología a la física, pasando por la estadística. Pero cuando se trata de construir los modelos a gran escala necesarios para llevar a cabo la gestión de los incendios forestales, las limitaciones computacionales, la cantidad y calidad de los datos requeridos y la interacción de una enorme cantidad de factores hacen que la modelización físico-matemática no sea, en muchos, casos un enfoque factible. Es por ello, que los modelos empíricos y estadísticos han tomado cada vez más peso en el estudio de los incendios forestales, aunque su utilidad depende en muchos casos de la calidad y cantidad de los datos disponibles, así como de la capacidad los modelos de representar relaciones no lineales presentes en los datos (Jain et al. [5]).

A esto se deben sumar otros factores. En primer lugar, la creciente disponibilidad de datos provenientes de satélites y sensores remotos que permiten el monitoreo de los incendios forestales a partir de la recolección continua de datos geoespaciales y climáticos, además del desarrollo de los Sistemas de Información Geográfica (GIS), que han permitido manipular de forma eficiente estos conjuntos de datos espaciales. El aumento de la capacidad computacional de los equipos que, unido al desarrollo de las tecnologías de la información, motivó el auge por el *Machine Learning*(ML) desde la década de los 90, lo que se manifestó en el desarrollo de nuevos algoritmos como *Support Vector Machine* o *Random Forest*. Y el creciente interés de los gobiernos en la recopilación sistemática y detallada de información relativa a los incendios forestales producidos con el fin de disponer de información relevante para el análisis y la toma de decisiones<sup>2</sup>.

Todo ello ha propiciado la aplicación del ML en el estudio de los incendios forestales desde la década de los 90 en seis dominios claves: caracterización de combustibles, detección y mapeo de incendios; clima y cambio climático; ocurrencia, susceptibilidad y riesgo de incendios; predicción del comportamiento del fuego; efectos del fuego; y gestión de incendios. En Jain et al. [5] se revisan las aplicaciones del ML en la ciencia y gestión de incendios forestales, dentro de los dominios de aplicación mencionados. El estudio identifica 300 publicaciones hasta finales de 2019, mostrando el uso frecuente de algoritmos de ML como *Random Forests*, *MaxEnt*, redes neuronales artificiales, árboles de decisión, SVM y algoritmos genéticos. La revisión enfatiza las ventajas y limitaciones de estos métodos y subraya la necesidad de combinar la experiencia en la ciencia del fuego con técnicas avanzadas de ML para poder construir modelos realistas y útiles.

Como se señala en el artículo mencionado, la predicción de incendios forestales es vital para la planificación, para la preparación del material y del personal, para poder llevar a cabo una gestión eficiente de los recursos, para determinar la distribución de las unidades móviles en el terreno y para asistir la toma de decisiones. Y en este campo, la combinación de los algoritmos de ML y las tecnologías GIS, junto con la gran cantidad de información georreferencia disponible actualmente, ofrece un nuevo abanico de posibilidades. A continuación se mencionan algunos trabajos recientes que ahondan en esta dirección.

En Cortez y Morais [3] los autores evalúan el rendimiento de distintos modelos de ML, como SVM, árboles de decisión, regresión lineal múltiple, Naive Bayes, bosques aleatorios y redes neuronales, para predecir el área quemada por los incendios forestales en el Parque Natural de Montesinho, en el norte de Portugal, a partir de información meteorológica. Los mejores resultados se obtienen para el modelo de SVM considerando 4 variables explicativas: temperatura, humedad relativa, precipitaciones y viento.

---

<sup>2</sup>Una buena muestra de esto es la EGIF, la base de datos nacional de los incendios forestales. Iniciada en 1968, constituye la serie de datos sobre incendios forestales más completa en el ámbito internacional (mit [2]).

---

En la investigación llevada a cabo por Vilar del Hoyo et al. [14] se aplica regresión logística en una cuadrícula de  $1 \times 1$  km en la Comunidad Autónoma de Madrid, utilizando datos socioeconómicos como variables predictoras para representar los factores antropogénicos relacionados con el riesgo de incendio. También se evalúa otro enfoque basado en la predicción de la densidad de puntos de ignición en una cuadrícula de  $10 \times 10$  km, utilizando funciones Kernel. La ocurrencia histórica de incendios de 2000 a 2005 se utiliza como variable de respuesta. El rendimiento de los modelos se evalúa con los incendios ocurridos en 2006 y 2007, obteniendo un AUC de 0.70 y 0.67 para ambos modelos, respectivamente.

En Gutiérrez-Hernández et al. [4] se analiza la superficie afectada por los incendios forestales ocurridos entre 1975 y 2013 en la Comunidad Autónoma de Andalucía en función de 15 variables explicativas: altitud, insolación, pendiente, precipitación invernal, precipitación estival, temperatura estival, velocidad del viento, frecuencia del viento, superficie protegida, superficie de monte público, superficie de usos forestales, distancia a viario, distancia a zonas pobladas, saldo demográfico y saldo ganadero. Se aplica un modelo de regresión lineal múltiple y un Modelo de Regresión Geográficamente Ponderada. Los mejores resultados se obtienen en el segundo modelo, puesto que permite considerar la estructura de correlaciones espaciales presentes en los datos.

En Martínez-Fernández et al. [8] se identifican los factores humanos asociados con un mayor riesgo de incendio forestal en España y se analiza la distribución espacial de la aparición de incendios forestales en el país tomando como unidad de estudio el municipio. Se consideran 108 variables, de las cuales se seleccionan 29 tras un primer estudio exploratorio, las cuales se usan para entrenar un modelo de regresión logística para estimar la probabilidad de una alta o baja ocurrencia de incendios. Finalmente, tan solo resultan significativas 13 de las variables consideradas.

En Liz-López et al. [7] se presenta un nuevo Modelo de Evaluación de Incendios Forestales (WAM, por sus siglas en inglés) que utiliza deep learning para anticipar el impacto de los incendios a partir de información meteorológica satelital y el NDVI en Castilla y León y Andalucía. Las variables respuesta del modelo son el área quemada, el tiempo de control y extinción, y la cantidad de recursos humanos, aéreos y pesados necesarios para la extinción del incendio. Emplea una red convolucional residual que realiza regresiones sobre variables atmosféricas e índice de verdor. El WAM se preentrena con 100,000 ejemplos de datos sin etiquetar y se ajusta con un pequeño conjunto de 445 muestras etiquetadas.

En Stojanova et al. [13] los autores construyen y comparan diversos modelos de clasificación (KNN, naïve Bayes, árboles de decisión, regresión logística, SVM, *Bayesian Networks*, *AdaBoost*, *bagging DT* y *RF*) para estimar el riesgo de incendio en tres regiones de Eslovenia (Kras, la región costera y la Eslovenia continental) a partir de datos georreferenciados, imágenes de teledetección y el modelo de predicción meteorológica ALADIN. Los mejores resultados los obtienen con los modelos *bagging DT* y *Random Forest*.

## El enfoque de este trabajo

En el siguiente trabajo se abordará el problema de la predicción de incendios forestales en la Comunidad Autónoma de Andalucía desde una perspectiva estadística mediante el uso de (GIS) y modelos de ML. Se adaptará un enfoque dinámico y global, es decir, se buscará predecir el riesgo de que una determinada localización se vea afectada por un incendio forestal en un momento dado a través de 27 covariables que cubren las 5

principales dimensiones desde las que abordar el estudio de los incendios forestales: la antropológica, la demográfica, la meteorológica, la topográfica y la vegetación.

En el capítulo 2 se presentarán los conceptos y herramientas fundamentales que se manejarán en el trabajo. Se hará una introducción a los tipos de datos espaciales, presentando herramientas para manipularlos, y se explicarán los modelos de ML que serán utilizados (Regresión Logística, Árboles de Decisión, KNN, SVM lineal, SVM radial y Bosques Aleatorios).

En el capítulo 3 se construirá la muestra que será usada para entrenar los modelos de clasificación binaria. Se implementará un método para generar muestras aleatorias dentro de los límites del estudio combinando los conjuntos de datos espaciales recopilados, mediante el uso de técnicas específicas de procesamiento de datos espaciales.

En el capítulo 4 se analizará la muestra generada mediante el uso de métodos estadísticos, gráficos y numéricos, considerando las dimensiones espacial y temporal de los datos.

En el capítulo 5 se construirán los modelos mencionados usando una partición temporal de la muestra en entrenamiento-validación-test. Se ajustarán los parámetros de los modelos evaluando el rendimiento sobre el conjunto de validación y se compararán las métricas de rendimiento de los modelos con las configuraciones de parámetros elegidas sobre el conjunto test, reentrenando para ello los modelos sobre el conjunto de entrenamiento y validación.

En el capítulo 6 se pondrán a prueba los mejores modelos en casos prácticos, con el objetivo de evaluar sus desempeños y conocer sus limitaciones.

En el capítulo 7 se repasarán los puntos más relevantes tratados a lo largo del trabajo, se resumirán las contribuciones más importantes del mismo en el contexto de la predicción de incendios forestales y se propondrán líneas de investigación para extender el trabajo y profundizar en ciertos aspectos de interés.

## 1.1. Objetivos

El objetivo de esta investigación es construir modelos de *Machine Learning* que permitan predecir incendios forestales en la Comunidad Autónoma de Andalucía. Para abordar dicho objetivo, se realizarán las siguientes tareas:

1. Construir un conjunto de datos que permita la realización de análisis estadísticos y la posterior construcción de modelos de *Machine Learning* (ML) para la predicción de incendios forestales en Andalucía a partir de un estudio previo del problema.
2. Modelizar el riesgo de incendio forestal usando distintos algoritmos de ML y comparar sus resultados
3. Analizar el desempeño de los modelos en la realidad estudiando potenciales casos de interés.





# Bibliografía

- [1] (2023). «Datos Estadísticos Andalucía del 01/01/2022 al 31/12/2022 Plan INFOCA, Centro Operativo Regional». *Informe técnico*, Consejería de Sostenibilidad, Medioambiente y Economía Azul, Junta de Andalucía.  
<https://www.juntadeandalucia.es/medioambiente/portal/documents/20151/55229821/memoria-infoca-Andalucia-2022.pdf/b95604c3-53d7-7564-bbb3-6c8a4f8c332a?t=1713950214344>.
- [2] (2024). «Ministerio para la transición ecológica y el reto demográfico». <https://www.miteco.gob.es/es.html>.
- [3] CORTEZ, PAULO y MORAIS, ANÍBAL DE JESUS RAIMUNDO (2007). «A data mining approach to predict forest fires using meteorological data». <https://hdl.handle.net/1822/8039>.
- [4] GUTIÉRREZ-HERNÁNDEZ, OLIVER; SENCIALES-GONZÁLEZ, J. M. y GARCÍA, LUIS V. (2015). «Los incendios forestales en Andalucía: investigación exploratoria y modelos explicativos».
- [5] JAIN, PIYUSH; COOGAN, SEAN C.P.; SUBRAMANIAN, SRIRAM GANAPATHI; CROWLEY, MARK; TAYLOR, STEVE y FLANNIGAN, MIKE D. (2020). «A review of machine learning applications in wildfire science and management». *Environmental Reviews*, **28**(4), pp. 478–505. doi: 10.1139/er-2020-0019.  
<https://doi.org/10.1139/er-2020-0019>.
- [6] JAMES, STEVEN R.; DENNELL, R. W.; GILBERT, ALLAN S.; LEWIS, HENRY T.; GOWLETT, J. A. J.; LYNCH, THOMAS F.; MCGREW, W. C.; PETERS, CHARLES R.; POPE, GEOFFREY G. y STAHL, ANN B. (1989). «Hominid Use of Fire in the Lower and Middle Pleistocene: A Review of the Evidence [and Comments and Replies]». *Current Anthropology*, **30**(1), pp. 1–26. ISSN 00113204, 15375382.  
<http://www.jstor.org/stable/2743299>.
- [7] LIZ-LÓPEZ, HELENA; HUERTAS-TATO, JAVIER; PÉREZ-ARACIL, JORGE; CASANOVA-MATEO, CARLOS; SANZ-JUSTO, JULIA y CAMACHO, DAVID (2024). «Spain on fire: A novel wildfire risk assessment model based on image satellite processing and atmospheric information». *Knowledge-Based Systems*, **283**, p. 111198. ISSN 0950-7051. doi: 10.1016/j.knosys.2023.111198.  
<https://www.sciencedirect.com/science/article/pii/S0950705123009486>.
- [8] MARTÍNEZ-FERNÁNDEZ, JESÚS; VEGA-GARCÍA, CRISTINA y CHUVIECO, EMILIO (2009). «Human-caused wildfire risk rating for prevention planning in Spain». doi: 10.1016/j.jenvman.2008.07.005.

- [9] MATAIX SOLERA, JORGE y CERDÀ, ARTEMI (2009). «Incendios forestales en España. Ecosistemas terrestres y suelos». En: Càtedra Divulgació de la Ciència Universitat de València (Ed.), *Efectos de los incendios forestales sobre los suelos en España: el estado de la cuestión visto por los científicos españoles*, pp. 25–54. Publicacions de la Universitat de València. ISBN 978-84-370-7653-9.
- [10] MORENO, J. M.; URBIETA, I.R.; BEDIA, J.; GUTIÉRREZ, J.M. y VALLEJO, V.R.. «Los incendios forestales en España ante el cambio climático». [https://www.miteco.gob.es/content/dam/miteco/es/cambio-climatico/temas/impactos-vulnerabilidad-y-adaptacion/cap34-losincendiosforestalesenespanaantealcambioclimatico\\_tcm30-70236.pdf](https://www.miteco.gob.es/content/dam/miteco/es/cambio-climatico/temas/impactos-vulnerabilidad-y-adaptacion/cap34-losincendiosforestalesenespanaantealcambioclimatico_tcm30-70236.pdf).
- [11] SAN-MIGUEL-AYANZ, JESUS; DURRANT, TRACY; BOCA, ROBERTO; MAIANTI, PIERALBERTO; LIBERTA', GIORGIO y OOM, DUARTE; BRANCO ALFREDO; DE RIGO DANIELE; FERRARI DAVIDE; ROGLIA ELENA; SCIONTI NICOLA (2023). «Advance Report on Forest Fires in Europe, Middle East and North Africa 2022». *Informe técnico JRC133215*, Publications Office of the European Union, Luxembourg. doi: 10.2760/091540.
- [12] SCOTT, A.C. y GLASSPOOL, I.J. (2009). «The diversification of Paleozoic fire systems and fluctuations in atmospheric oxygen concentrations», **103**, pp. 10861–10865. doi: <https://doi.org/10.1073/pnas.0604090103>.
- [13] STOJANOVA, DANIELA; KOBLER, ANDREJ; OGRINC, PETER; ŽENKO, BERNARD y DŽEROSKI, SAŠO (2012). «Estimating the risk of fire outbreaks in the natural environment». *Data mining and knowledge discovery*, **24**, pp. 411–442. <https://link.springer.com/article/10.1007/s10618-011-0213-2>.
- [14] VILAR DEL HOYO, L.; ISABEL, MARTÍN; M.P. y MARTÍNEZ VEGA, F.J. (2011). «Logistic regression models for human-caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data». *European Journal of Forest Research*, **130**, pp. 983–996. doi: 10.1007/s10342-011-0488-2.