

Índice general

1. Aplicación de los modelos	3
1.1. Visión general del desempeño de los modelos	3
1.2. Caso estudio	5

Capítulo 1

Aplicación de los modelos

En esta sección se pretende ilustrar el funcionamiento de los modelos construidos en la sección anterior, valorar su desempeño al ser aplicados en la realidad y conocer sus limitaciones.

1.1. Visión general del desempeño de los modelos

Dado que los modelos se han construido y evaluado sobre un conjunto de datos que es inevitablemente limitado y que se ve fuertemente influenciado por decisiones metodológicas, surge la incógnita de si las métricas de rendimiento obtenidas están realmente reflejando la capacidad de generalización de los modelos al ser aplicados en la realidad o de si, en cambio, tan solo están reflejando sesgos introducidos en el conjunto de datos a través del proceso de selección. Para entender mejor el funcionamiento de los modelos y conocer si realmente esto sirven para predecir incendios forestales en la vida real, se ha decidido adoptar el siguiente enfoque.

En primer lugar, se ha construido una malla de puntos con una resolución de 10km por 10km cubriendo toda la extensión de Andalucía (en este caso, se entiende como resolución la distancia entre los puntos en la dirección Este-Oeste y Norte-Sur) (Figura 1.1).

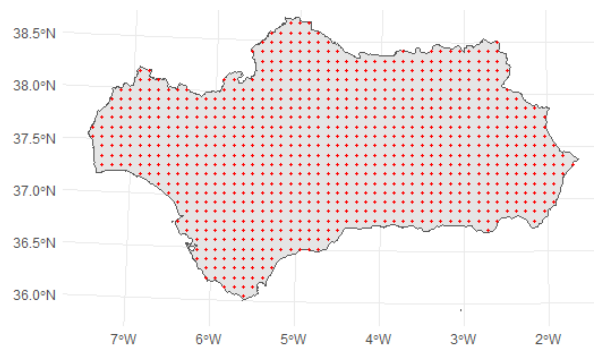


Figura 1.1: Malla de puntos con una resolución de 10km por 10km.

A continuación, se ha asociado a cada uno de los puntos de la malla el valor de todas las variables predictoras el día 15 de cada mes del año 2022 en esa localización, usando los métodos de preprocesamiento y depuración ya descritos. Estos datos se han utilizado

para predecir el riesgo de incendio forestal en cada uno de los puntos de la malla el día 15 de cada mes, utilizando para ello los modelos finales de la sección anterior que mejor rendimiento mostraron sobre los datos test. Los resultados se muestran en las Figuras 1.2 (Regresión logística con penalización) y 1.3 (SVM lineal) y 1.4 (Random Forest).

Probabilidad de incendio estimada el día 15 de cada mes de 2022

Modelo: lr

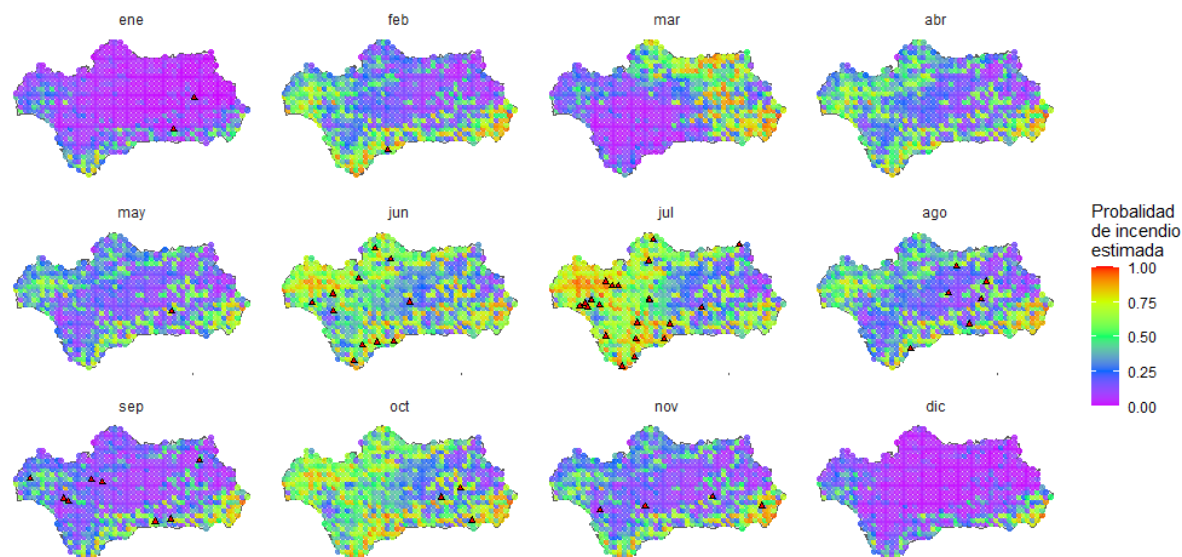


Figura 1.2: Probabilidades de incendios estimadas el día 15 de cada mes de 2022 con el modelo de regresión logística con penalización. Los triángulos indican los incendios de más de 100ha registrados en ese mes.

Se puede observar que las predicciones del modelo de regresión logística con penalización y del SVM lineal son muy parecidos. A diferencia de las predicciones del modelo de Bosque aleatorio, que son bastante similares todos los meses, estos dos modelos muestran bastante variación mensual. Seguramente esta sea la causa de que, si bien el modelo de bosque aleatorio mostraba un buen rendimiento en validación, al evaluar su rendimiento sobre los datos test, este bajó significativamente. Es por ello que, a falta de la opinión de un experto en ecología del fuego, se opta por descartar el modelo de bosque aleatorio ya que no parece reflejar correctamente la variación estacional que se observa en la aparición de incendios forestales.

Se analizan por tanto las predicciones de los otros dos modelos (regresión logística y SVM). Se puede observar una clara componente estacional en las observaciones. En los meses de diciembre y enero se observan los niveles de riesgo más bajos a nivel global, mientras que los niveles de riesgo más elevados se encuentran en los meses de junio y julio, aunque por algún motivo también se observan niveles de riesgo elevado en los meses de marzo y octubre. Se puede observar también como las zonas con una probabilidad alta de incendio forestal varían en función del mes. Es curioso que en marzo ambos modelos den probabilidades de incendio tan elevadas en la zona oriental de la comunidad. Al margen del estudio específico y detallado de los mapas presentados, lo cual correspondería a los expertos en la materia y escapa de los objetivos de este trabajo, se puede observar que prácticamente todos los incendios se producen en zonas con una probabilidad de incendio elevada y que los modelos son capaces de ir más allá del mero estudio de las variables meteorológicas, ya que se observan zonas más o menos aisladas con una mayor

Probabilidades de incendio estimada el día 15 de cada mes de 2022

Modelo: svm_linear

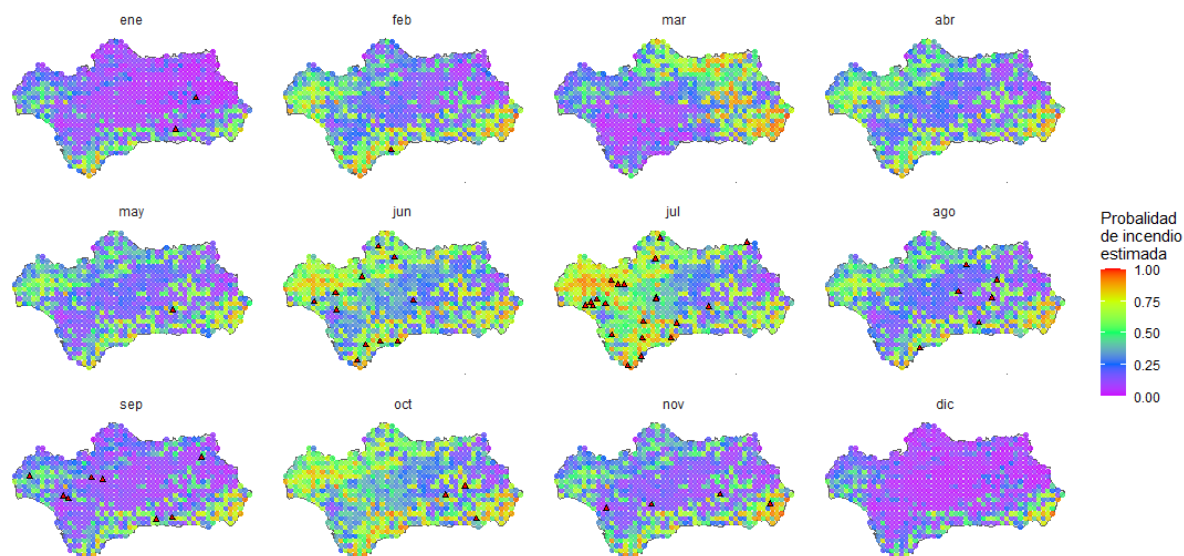


Figura 1.3: Probabilidades de incendios estimadas el día 15 de cada mes de 2022 con el modelo de SVM lineal. Los triángulos indican los incendios de más de 100ha registrados en ese mes.

probabilidad de incendio que se corresponden con las zonas en las que se ha observado un incendio. Esto indica que son otros factores los que el modelo está considerando para indicar riesgo de incendio, ya que la resolución espacial de las variables meteorológicas es bastante baja (50km), por lo que las variaciones a un mayor detalle son debidas a otros factores.

Este es, sin embargo, un enfoque bastante pobre, pues solo se está considerando el día 15 de cada mes, lo que podría llevar a conclusiones erróneas a la hora de evaluar los modelos (debidas, por ejemplo, a valores atípicos en ese día concreto). Esto es debido a las limitaciones computacionales del equipo disponible. Pese a ello, se ha podido ilustrar, aunque sin entrar en detalle, el desempeño de los modelos al aplicarlos para evaluar el riesgo de incendio en la realidad.

1.2. Caso estudio

A continuación, se pondrá a prueba el modelo de regresión logística construido con un caso real, el incendio de Sierra Bermeja, que se originó el 8 de septiembre de 2021 en el municipio de Jubrique en la provincia de Málaga (Figura 1.5). Se ha elegido este incendio por dos motivos. En primer lugar, porque fue el mayor incendio que hubo en España en el año 2021, con una superficie total afectada de 8607ha y una duración de 46 días hasta su extinción. Y en segundo lugar, porque fue un incendio intencionado, por lo que permitirá reflejar el comportamiento del modelo en incendios causados por el hombre.

Para analizar la capacidad de predicción del modelo para este incendio, se ha llevado a cabo el siguiente enfoque. Primero, se ha construido una malla de puntos con una resolución de 1km por 1km, cubriendo todo el *bounding box* de un *buffer* de 10km alrededor

Probabilidad de incendio estimada el día 15 de cada mes de 2022

Modelo: rf

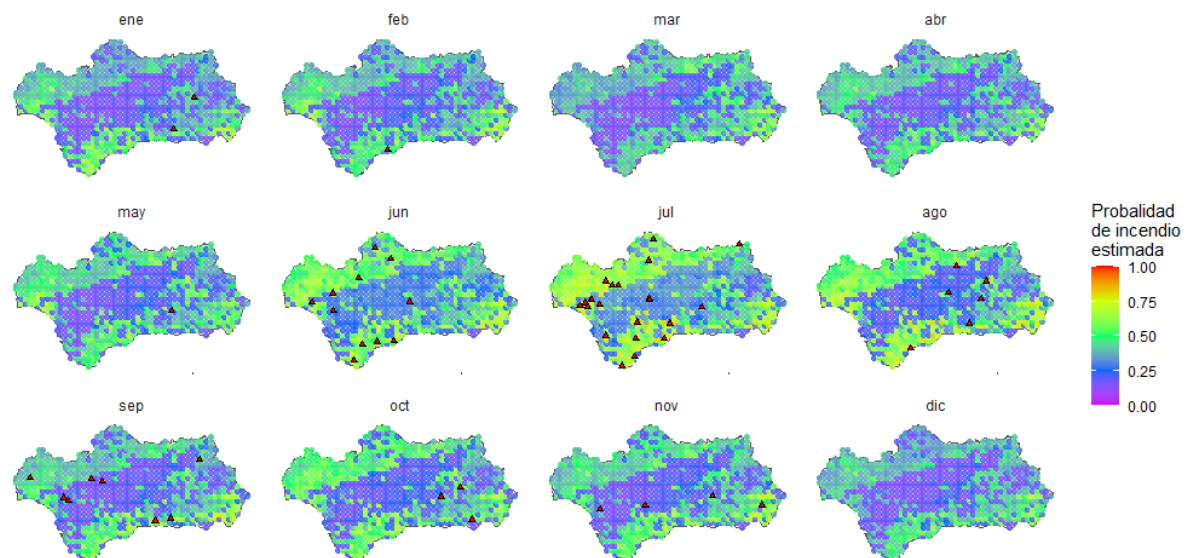


Figura 1.4: Probabilidades de incendios estimadas el día 15 de cada mes de 2022 con el modelo de random forest. Los triángulos indican los incendios de más de 100ha registrados en ese mes.

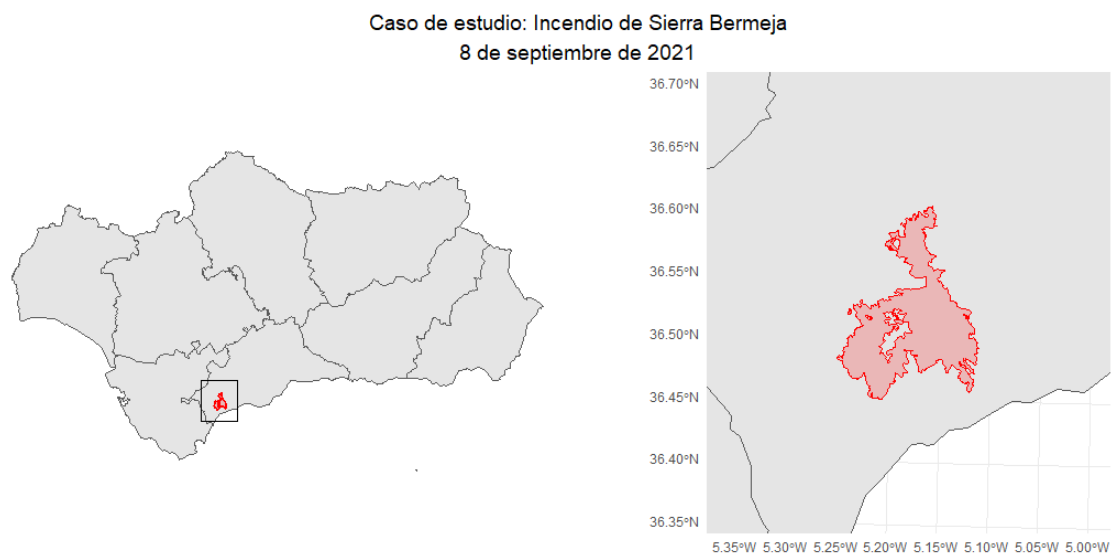


Figura 1.5: Área recorrida por el fuego en el incendio de Sierra Bermeja.

del perímetro del incendio. A continuación, en cada uno de estos puntos se han tomado todas las variables predictoras el día de origen del incendio, 15 y 30 días antes y 15, 30 y 45 días después. Con estos datos se ha utilizado el modelo de regresión logística con penalización para predecir la probabilidad de incendio forestal en cada uno de los días considerados en toda la malla de puntos. Los resultados se muestran en la Figura 1.6.

De este gráfico pueden extraerse varias conclusiones. Por un lado puede observarse que el día del origen del incendio se produce un aumento drástico de las probabilidades estimadas de incendio, las cuales continúan siendo muy altas 15 días después y, aunque disminuyen de forma general, se mantienen elevadas hasta 45 días después del inicio del fuego. Sin

Incendio de Sierra Bermeja

Model: lr

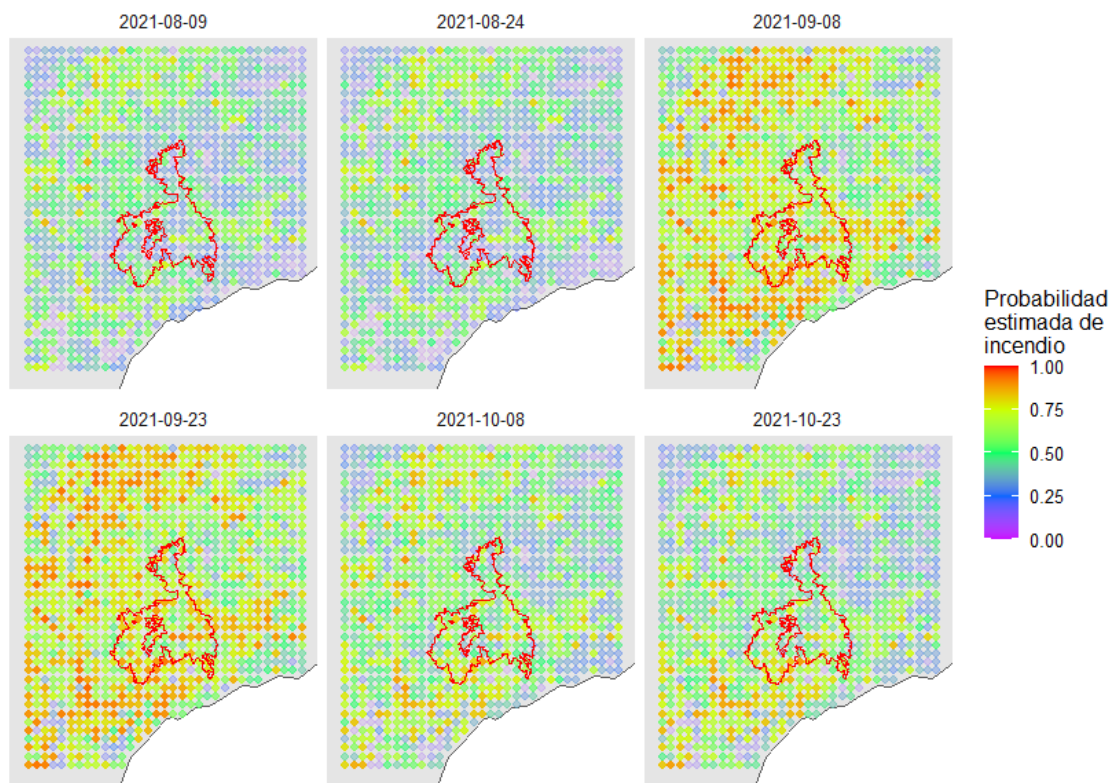


Figura 1.6: Mapa con las probabilidades de incendio estimadas en los días en torno al origen del incendio de Sierra Bermeja. El área total recorrida por el fuego se muestra en rojo.

embargo, si bien es cierto que a nivel global el modelo sí parece aportar información estimando un riesgo muy alto de incendio en la región, al aumentar el nivel de detalle puede observarse que la capacidad discriminadora del modelo disminuye significativamente. Esto es coherente, ya que la resolución de las variables climáticas es de aproximadamente 50km por 50km, por lo que no son adecuadas para trabajar con un nivel de detalle tan reducido.

Cabe mencionar que la variación observada en el riesgo de incendio estimado en las distintas fechas es debida, únicamente, a los cambios en las variables meteorológicas y en el NDVI. Esto es debido a que en el modelo de regresión logística construido no se han considerado las posibles interacciones entre las variables, lo cual podría ser de gran interés dadas las características del problema.