

Índice general

1. Conclusiones, aportaciones y trabajo futuro	3
1.1. Conclusiones	3
1.2. Aportaciones	4
1.3. Trabajo futuro	5

Capítulo 1

Conclusiones, aportaciones y trabajo futuro

En este último capítulo se va a realizar una recapitulación de las conclusiones extraídas en cada una de las secciones del presente trabajo de fin de estudios. En primer lugar, se presentarán de forma resumida las conclusiones obtenidas a lo largo del trabajo. A continuación, se detallarán las aportaciones realizadas en el campo de la predicción de incendios forestales. Por último, se indicarán algunas líneas de investigación, dentro del campo de la predicción de incendios forestales y la inteligencia artificial, que permitirían profundizar en el desarrollo de la metodología presentada de cara a obtener mejores modelos con utilidad práctica.

1.1. Conclusiones

El control y extinción de los incendios forestales requiere del desplazamiento de un gran número de efectivos, y el éxito en la operación depende en muchos casos de la velocidad en la respuesta. En el presente trabajo se ha desarrollado una metodología completa para dotar de herramientas que permitan predecir eficazmente las zonas en riesgo de verse afectadas por un incendio forestal en la Comunidad Autónoma de Andalucía, facilitando así la toma de decisiones y permitiendo una asignación de los recursos más eficiente.

A lo largo de esta memoria se ha desarrollado una metodología completa para la construcción de modelos de predicción de incendios forestales en la Comunidad Autónoma de Andalucía mediante el uso de técnicas de *Machine Learning* y procesamiento de datos geoespaciales, adoptando un enfoque dinámico y global. Dinámico, pues se predice el riesgo de incendio forestal para una localización específica en un día concreto. Global, pues se han considerado 27 variables que abarcan las 5 dimensiones principales : antropológica, demográfica, meteorológica, topográfica y de vegetación.

Se ha comenzado introduciendo el problema, justificando su relevancia y estableciendo 3 tareas claras para alcanzar el objetivo del trabajo. La primera de las tareas implicaba la construcción de un conjunto de datos adecuado para el análisis estadístico y la construcción de modelos de ML. Esta tarea se abordó en el capítulo 2, donde no solo se ha generado un conjunto de 20.000 muestras sobre el que se ha desarrollado el trabajo, si no que se ha implementado un algoritmo para tomar muestras aleatorias de casos positivos y negativos

dentro del marco del estudio y asociar a cada observación los valores correspondientes de todas las variables consideradas. Se ha explorado el uso de estratificación por mes en la selección de la muestra, con el objetivo de entrenar modelos que sean capaces de detectar relaciones más profundas en los datos relacionadas con la aparición de los incendios forestales. Se ha visto que, por ejemplo, en el caso del *Random Forest* esta estratificación ha conducido a que la variable *uso_suelo* haya tomado demasiada importancia en la clasificación, conduciendo a modelos que no parecen comportarse adecuadamente. ||| Queda, por tanto, abierta la pregunta de cuál sería la mejor manera de tomar la muestra de casos negativos de forma que se encuentre un equilibrio entre la sensibilidad del modelo a las variables meteorológicas sin perder la influencia de las demás variables. |||

A continuación, se ha analizado en profundidad el conjunto de datos generado, recurriendo principalmente a técnicas gráficas, aunque también se han usado algunos métodos numéricos. La complejidad de esta fase radica en que para llegar a obtener información relevante es necesario tener en cuenta las dimensiones espacial y temporal de los datos. Esta fase ha permitido llegar a un mayor conocimiento acerca del conjunto de datos, caracterizado por una gran presencia de valores *outlier*, por distribuciones asimétricas hacia la derecha en casi todas las variables numéricas y correlaciones generalmente pequeñas en valor absoluto. A nivel gráfico, se ha observado que las variables que muestran mayores diferencias entre ambas clases son *PRECTOTCORR* (el total diario de precipitaciones), *WS10M* (la velocidad del viento a 10 m) y *uso_suelo* (la clasificación de uso de suelo). Un buen reflejo de la complejidad del conjunto de datos es que para explicar el 80 de la varianza de las 18 variables numéricas en la muestra se necesitan 11 componentes principales.

La siguiente tarea planteada se ha desarrollado en el capítulo 5, donde se han construido distintos modelos de ML de clasificación binaria. Se ha usado el flujo de trabajo propuesto por el paquete de R *tidymodels* para preprocesar los datos, entrenar los modelos y ajustar los valores de los hiperparámetros. Los modelos considerados han sido: regresión logística con penalización, regresión logística con penalización usando PCA, *k-Nearest Neighbours*, SVM lineal, SVM radial, árbol de decisión y *Random Forest*. Se ha usado una partición temporal en entrenamiento-validación-test para entrenar los modelos, ajustar los parámetros y evaluar su capacidad de generalización sobre nuevos datos. Al entrenar los modelos sobre el conjunto de datos generado en las etapas anteriores con la estratificación por mes,

Tanto en el conjunto de validación como sobre el conjunto test, los modelos de regresión logística con penalización, SVM lineal y SVM radial mostraron resultados muy similares

A continuación se detallarán los aspectos más relevantes de cada sección del trabajo, pero es relevante mencionar que probablemente las conclusiones más importantes no sean las que queden en recogidas en estas páginas, si no a las que se podría llegar a partir de las herramientas desarrolladas, extendiéndolas y profundizando en ellas, con mayor recursos.

1.2. Aportaciones

Enfoque dinámico y global METODOLOGÍA COMPLETAMENTE REPRODUCIBLE Y ADAPTABLE (a otras regiones, más variables,...)!

- recopilación de toda la información

- Metodología para construir conjuntos de datos útiles para la predicción de incendios forestales
- Amplio estudio de las variables predictivas consideradas en la Comunidad de Andalucía.
- Modelos “dinámicos” de predicción de riesgo de incendios
- Incluir el factor humano, todas las dimensiones
- Estratificación por mes en el muestreo y uso de una partición temporal. Evitar estimaciones de la capacidad de generalización de los modelos excesivamente optimistas, como ocurre en la mayoría de estudios considerados.

1.3. Trabajo futuro

- Base de datos exhaustiva, que incluya todos los incendios, no solo los mayores de 100ha. Explorar el uso de la EGIF
- Datos de estaciones meteorológicas en lugar de datos satelitares: Mayor detalle, calidad y sensibilidad
- Explorar el uso de modelos de Deep Learning como Redes Convolucionales (CNN), que permitirían considerar la estructura espacial de los datos.
- Estudiar el método de muestreo que mejor pueda adecuarse al problema que se pretende abordar
- Área quemada