1. REFLEXIONES GENERALES

La sequía es una de las consecuencias más directas del calentamiento global que existe hoy en día, el cual supone un aumento de las temperaturas que hacen que los entornos más secos aumenten el riesgo de incendios forestales.

Esta situación está haciendo que cada vez más empresas se interesen en trabajar para crear recursos tecnológicos que ayuden a la prevención y detección de incendios forestales.

Algunos ejemplos de estas empresas son:

* Telefónica, que en colaboración con otros organismos (UC3M, Divisek y Dronitec) ya ha presentado un piloto en el que combina drones con IoT para avanzar en la localización del fuego.
* IBM, que está trabajando con la empresa portuguesa Compta en el desarrollo de la solución Bee2FireDetection. Esta herramienta combina la tecnología de inteligencia artificial de IBM Watson con otras técnicas y herramientas para detectar incendios forestales en un radio de 15 kilómetros.

Si bien estas herramientas se basan en la detección del incendio una vez éste se ha producido, también es de interés anticiparse a la causa.

Una herramienta para este fin es el análisis del histórico de incendios producidos en la región que se quiere estudiar, en este caso Galicia, para el análisis de las causas de incendios a través de modelos predictivos. Para ello, como se ha podido ver a lo largo de este TFM se han llevado a cabo el desarrollo de una serie de pasos como han sido el preprocesado, EDA y división del dataset combinando dos BBDD diferentes (con el fin de complementar al máximo la información disponible) para terminar el proceso llevando a cabo la aplicación de diferentes modelos con el fin de obtener resultados viables.

En la visualización se ha hecho uso de programas diferentes a los empleados en el máster para aportar más valor a lo aprendido en este curso y optimizar los conocimientos de los componentes del grupo.

El preprocesado se ha realizado en R, ya que es un programa de análisis científico con mucha potencia para estos fines. En base a los módulos cursados con la utilización de R Studio, se ha considerado el mejor candidato para la limpieza del dataset, y el tratamiento de valores extremos y valores perdidos.

Se ha aplicado en medida de lo posible el aprendizaje adquirido en el máster para la justificación de cada una de las decisiones tomadas. Aun así, algunas de las transformaciones aplicadas a las variables pueden sostenerse en un criterio más subjetivo, primando en este caso la mejora de la interpretabilidad.

También ha habido algunas cuestiones críticas. La mayoría de observaciones presentaba un porcentaje elevado de valores extremos y casos concretos, dada la naturaleza de la variable, se han considerado como "puntos de interés" y se han conservado.

Respecto a las observaciones perdidas, el tratamiento no ha sido tan exhaustivo dado que esta problemática se había atajado previamente por medio de la discretización de variables y en algunos casos la creación de categorías específicas para designar las observaciones con falta de información.

En materia de la modelización predictiva, se han explorado bastantes posibilidades en cuanto a la predicción de la variable de las causas del incendio.

El comienzo de esta sección establece el reducido número de observaciones de las que se dispone respecto a la dimensión del dataset original, pero aun así es suficiente. Mediante la codificación one - hot en algunas de las variables se elimina el establecimiento de un orden numérico poco informativo.

Los modelos generados a través de machine learning con selección de variables basada en el estadístico de F para hallar las variables más significativas del propio modelo concluyen que es posible llegar a predecir hasta cerca de un 90% de las causas de los incendios con muy poca varianza entre los distintos resultados obtenidos a través de distintas muestras , sin embargo , el principal problema que alberga esta base de datos son las diferencias de los pesos de las distintas categorías de nuestra variable objetivo, es decir, aquellos incendios provocados por un rayo, por ejemplo , sería muy difícil de predecir ya que son muy pocas las situaciones en las que se ha producido el fuego por dicha causa. En nuestro estudio hemos intentado paliar ese problema mediante técnicas de balanceo que mejoran la predicción de las causas con menor peso.

1. POSIBLES MEJORAS

Al trabajar con datos siempre se ha tenido presente que las posibilidades y la creatividad son infinitas y el proyecto se ha adaptado a las restricciones temporal que conllevaban las fechas de entrega.

* Disponer de otras BBDD para complementar el estudio ( ver anexo)
* En estudios posteriores , podríamos añadir métodos más complejos de balanceo para seguir mejorando nuestra predicción, o transformar nuestra variable en dicotómica uniendo las categorías poco representadas, creando con ello un modelo para predecir si un incendio ha sido intencionado o no.
* Disponer de algoritmos o técnicas más fiables para las imputaciones de valores extremos u observaciones perdidas con un margen menor de error

Mirar la importancia de haber incluido esas variables de temperatura, focos, viento

**ANEXOS**

Este documento reúne la información adicional, de interés, que no ha podido ser incluida en el informe inicial dadas las restricciones de espacio.

Está estructurado de la misma manera que lo está el índice, incluyendo la información correspondiente a cada apartado y las explicaciones para su entendimiento.

No se ha generado en R Markdown para evitar la entrega de más archivos con código ya que desvía la importancia del código principal.

1. **ARCHIVOS DE LA INTRODUCCIÓN**

En lo referente al apartado de la introducción, hay un conjunto de tablas que aportan información relativa a la configuración de las variables originales que debían figurar en el contenido del TFM.

Esta tabla refleja la relación entre la codificación establecida entre el total de provincia del territorio español y el número que se les ha asignado. Esta conexión ha sido “ clave” en cuanto a la configuración de la base de datos concatenada en la que se ha basado el trabajo , dado que el criterio de anexión por provincia ha sido uno de los dos relevantes ( siendo el otro la coincidencia temporal reflejado en las fechas).

* *Comunidad* y *Provincia:*

|  |  |  |  |
| --- | --- | --- | --- |
| **CODAUTO** | **Comunidad Autónoma** | **CPRO** | **Provincia** |
| 1 | Andalucía | 4 | Almería |
| 1 | Andalucía | 11 | Cádiz |
| 1 | Andalucía | 14 | Córdoba |
| 1 | Andalucía | 18 | Granada |
| 1 | Andalucía | 21 | Huelva |
| 1 | Andalucía | 23 | Jaén |
| 1 | Andalucía | 29 | Málaga |
| 1 | Andalucía | 41 | Sevilla |
| 2 | Aragón | 22 | Huesca |
| 2 | Aragón | 44 | Teruel |
| 2 | Aragón | 50 | Zaragoza |
| 3 | Asturias, Principado de | 33 | Asturias |
| 4 | Balears, Illes | 7 | Balears, Illes |
| 5 | Canarias | 35 | Palmas, Las |
| 5 | Canarias | 38 | Santa Cruz de Tenerife |
| 6 | Cantabria | 39 | Cantabria |
| 7 | Castilla y León | 5 | Ávila |
| 7 | Castilla y León | 9 | Burgos |
| 7 | Castilla y León | 24 | León |
| 7 | Castilla y León | 34 | Palencia |
| 7 | Castilla y León | 37 | Salamanca |
| 7 | Castilla y León | 40 | Segovia |
| 7 | Castilla y León | 42 | Soria |
| 7 | Castilla y León | 47 | Valladolid |
| 7 | Castilla y León | 49 | Zamora |
| 8 | Castilla-La Mancha | 2 | Albacete |
| 8 | Castilla-La Mancha | 13 | Ciudad Real |
| 8 | Castilla-La Mancha | 16 | Cuenca |
| 8 | Castilla-La Mancha | 19 | Guadalajara |
| 8 | Castilla-La Mancha | 45 | Toledo |
| 9 | Cataluña | 8 | Barcelona |
| 9 | Cataluña | 17 | Girona |
| 9 | Cataluña | 25 | Lleida |
| 9 | Cataluña | 43 | Tarragona |
| 10 | Comunitat Valenciana | 3 | Alicante/Alacant |
| 10 | Comunitat Valenciana | 12 | Castellón/Castelló |
| 10 | Comunitat Valenciana | 46 | Valencia/València |
| 11 | Extremadura | 6 | Badajoz |
| 11 | Extremadura | 10 | Cáceres |
| 12 | Galicia | 15 | Coruña, A |
| 12 | Galicia | 27 | Lugo |
| 12 | Galicia | 32 | Ourense |
| 12 | Galicia | 36 | Pontevedra |
| 13 | Madrid, Comunidad de | 28 | Madrid |
| 14 | Murcia, Región de | 30 | Murcia |
| 15 | Navarra, Comunidad Foral de | 31 | Navarra |
| 16 | País Vasco | 1 | Araba/Álava |
| 16 | País Vasco | 48 | Bizkaia |
| 16 | País Vasco | 20 | Gipuzkoa |
| 17 | Rioja, La | 26 | Rioja, La |
| 18 | Ceuta | 51 | Ceuta |
| 19 | Melilla | 52 | Melilla |

Esta segunda tabla contiene la información relacionada con el contenido de la variable de causa, que va a ser el núcleo del trabajo. En el preprocesado, las causas 2 y 3 se sintetizan en una misma, ya que por su contenido , es el tratamiento lógico que se le debería de aplicar.

* *Causa:*

|  |  |
| --- | --- |
| **id** | **Significado** |
| 1 | Fuego por rayo |
| 2 | Fuego por accidente o negligencia |
| 3 | Fuego por accidente o negligencia |
| 4 | Fuego intencionado |
| 5 | Fuego por causa desconocida |
| 6 | Incendio reproducido |

Además de la variable de causa estándar, había una variable secundaria que complementaba ampliamente esta información , que más adelante será eliminada. La razón para la eliminación no es otra que la duplicidad de la información que proporcionaba y la invalidez de su uso en futuros modelos. En un primero lugar, se valoró el tratamiento de la síntesis de categorías a un número mucho más reducido, pero esta reducción daba como resultado la creación de una variable muy similar a la de causa, con su contenido correctamente situado en las categorías correspondientes.

* *Causa\_dec:*

|  |  |
| --- | --- |
| **id** | **Significado** |
| 280 | quema de matorral |
| 281 | quema de matorral próximo a edificaciones |
| 282 | quema de matorral para limpieza de caminos o sendas |
| 283 | quema de matorral en focos de animales nocivos |
| 284 | quema de matorral |
| 291 | actividades apícolas |
| 292 | fuegos artificiales |
| 293 | globos |
| 294 | juegos de niños |
| 295 | restos de poda de urbanización |
| 270 | escape de vertedero |
| 260 | quema de basura |
| 250 | fumadores |
| 210 | quema agrícola |
| 211 | quema de rastrojos |
| 212 | quema de restos de poda |
| 213 | quema de lindes y bordes de fincas |
| 214 | quema de bordes de acequias |
| 215 | quema agrícola |
| 220 | quema para regenerar pastos |
| 240 | hogueras |
| 230 | trabajos forestales |
| 223 | quema para regenerar pastos |
| 222 | quema para regenerar pastos |
| 221 | quema para regenerar pastos |
| 340 | maniobras militares |
| 335 | motores y máquinas |
| 334 | motores y máquinas |
| 333 | motores y máquinas |
| 332 | motores y máquinas |
| 331 | motores y máquinas |
| 330 | motores y máquinas |
| 320 | líneas eléctricas |
| 310 | ferrocarril |

1. **ARCHIVOS DE LA VISUALIZACIÓN**
2. **ARCHIVOS DE LA INTRODUCCIÓN**

Respecto al preprocesado realizado en R, no ha sido necesaria la inclusión de archivos adicionales, ya que todo el contenido relevante aparece reflejado en el código. La única cuestión ha sido el hecho de ocultar tanto el código como los resultados arrojados en el informe final, ya que las restricciones de espacio no dan lugar para más. Siguiendo el script se puede comprobar cómo cada uno de los pasos está desarrollado con más detalle del que se visualiza en el informe.

1. **ARCHIVOS DE LA MODELIZACIÓN Y PREDICCIÓN**

La herramienta que se ha utilizado para este propósito ha sido Jupyter Notebook por lo que la información adicional se encuentra en el notebook que contiene la carpeta ZIP entregada, en el que figuran todos los pasos seguidos.

1. **ARCHIVOS DE LA BIBLIOGRAFÍA**

* Bibliografía ( están todas las pag de las 2 bases de datos?)
  + <https://www.mapa.gob.es/es/desarrollo-rural/estadisticas/incendios-decenio-2006-2015_tcm30-511095.pdf>
  + <https://selvicultura.files.wordpress.com/2010/11/incendios-2001-2010.pdf>
  + <https://recyt.fecyt.es/index.php/IA/article/download/2731/2099/0>
  + <https://bigdatamagazine.es/ibm-tira-de-ia-para-hacer-mas-facil-apagar-los-incendios>
  + <https://as.com/meristation/2019/07/19/betech/1563487659_185207.html>
  + <https://www.ciospain.es/industria-y-utilities/la-inteligencia-artificial-entra-en-la-lucha-contra-incendios>
  + <https://es.greenpeace.org/es/trabajamos-en/bosques/incendios-forestales/>
  + <http://www.redruralnacional.es/-/-que-sabemos-de-los-incendios-forestales-en-espana->
  + <https://observatoriosociallacaixa.org/-/incendios-forestales-en-espana-importancia-diagnostico-y-propuestas-para-un-futuro-mas-sostenible>
  + <https://datosclima.es/Aemethistorico/Descargahistorico.html>
  + <https://opendata.aemet.es/centrodedescargas/productosAEMET>

<https://www.ciospain.es/industria-y-utilities/la-inteligencia-artificial-entra-en-la-lucha-contra-incendios>

<https://www.networkworld.es/convergencia/drones-e-iot-para-ayudar-a-detectar-incendios-forestales>