UNIVERSIDADE D
COIMBRA

Nuno Pires

Intelligent system for localising and monitoring forest fires

Dissertation in the context of the Master in Informatics Engineering, specialization in Information Systems, advised by Professor Alberto Cardoso and Professor Jacinto Estima and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

January 2024

DEPARTAMENTO DE
ENGENHARIA INFORMÁTICA
FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Nuno Pires

INTELLIGENT SYSTEM FOR LOCALISING AND MONITORING FOREST FIRES

Dissertation in the context of the Master in Informatics Engineering, specialization in Information Systems, advised by Professor Alberto Cardoso and Professor Jacinto Estima and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

January 2024

DEPARTAMENTO DE
ENGENHARIA INFORMÁTICA
FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Nuno Pires

SISTEMA INTELIGENTE PARA LOCALIZAÇÃO E MONITORIZAÇÃO DE INCÊNDIOS FLORESTAIS

Dissertação no âmbito do Mestrado em Engenharia Informática, especialização em Sistemas de Informação, orientada pelo Professor Alberto Cardoso e Professor Jacinto Estima e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Janeiro 2024

Abstract

Fire can have disastrous consequences. Decision-support systems play a central role in dealing with forest fires. Its early warning capacity and real-world impact help to protect forests, species, and communities from wildfire.

The presented work proposes a system for forecasting and monitoring forest fires using multiple data sources. Data fusion, aggregation, and enhancement techniques are also mentioned.

The main purpose of the system is to provide important information for emergency decision-making, such as the geolocation, severity, and temporal evolution of a wildfire. It will employ statistical and machine learning methodologies to predict and determine fire occurrence, susceptibility, and risk.

Finally, the system, with the help of data visualisation tools, will show findings and insights.

The document also presents current approaches and obstacles to forest fire prediction, as well as the suggested methodology and analysis of risk.

Keywords

Decision support system, Fire management, Fire forecasting, Machine learning, Spatial and temporal prediction

Resumo

Os incêndios podem ter consequências desastrosas. Os sistemas de apoio à decisão desempenham um papel central na luta contra os incêndios florestais. As suas capacidades de alerta e o seu impacto no mundo real ajudam a proteger as florestas, as espécies e as comunidades.

O trabalho apresentado propõe um sistema de previsão e monitorização de incêndios florestais que utiliza fontes diversas de dados. Onde são utilizadas técnicas de fusão, agregação e melhoramento de dados.

O principal objetivo do sistema é fornecer informações importantes para a tomada de decisões de emergência, tais como a geolocalização, a gravidade e a evolução temporal de um incêndio florestal. O sistema empregará metodologias estatísticas e de aprendizagem automática para prever e determinar a ocorrência, a suscetibilidade e o risco de incêndio.

Finalmente, com a ajuda de ferramentas de visualização de dados, o sistema será capaz de apresentar informações e resultados.

No documento também são analisadas as abordagens actuais e os obstáculos à previsão de incêndios florestais, bem como a metodologia sugerida e a análise de risco.

Palavras-Chave

Sistema de apoio à decisão, Gestão de incêndios, Previsão de incêndios, Aprendizagem automática, Previsão espacial e temporal

Contents

List of Figures

List of Tables

Chapter 1

To-be named Chapter

1.1 Study Area

The study area encompasses the geographic mainland region of Portugal, extending 89,000 km^2 located between latitude 42.3°N to 36.7°N and longitude 9.8°W to 6.0°W. Portugal experiences a Mediterranean climate influenced by the Atlantic Ocean and characterized by a mostly wet and cool season followed by a dry summer (Living and Overseas, 2023; Marques et al., 2011; Mora and Vieira, 2020).

Altitudes range from sea level to 2000 metres, and higher elevations are more prevelent central and northern regions (Marques et al., 2011).

The northern region is characterised by lower temperatures and higher values of precipitation than the southern region (Marques et al., 2011). The average annual temperature ranges from 7 to 18°C, while the annual rainfall ranges from 400 to 2,800 mm.

Portugal's vegetation is a blend of Atlantic, European, Mediterranean, and African species (Encyclopædia Britannica, Inc., 2024), and four tree species account for 80% of all the forest area: *Pinus pinaster*, *Eucalyptus globulus*, *Quercus suber*, and *Quercus rotundifolia* (Marques et al., 2011).

1.2 Dataset Sources

1.2.1 Wildfire Occurrences

Historical records of fire occurrences were taken from (Central de Dados Github Repository, 2017; centraldedados, 2017) and (ICNF, 2024). The records featured in (Central de Dados Github Repository, 2017; centraldedados, 2017) span across 35 years, from 1980 to 2015, and hold 791453 instances of wildfires. Its most relevant features are encapsulated in table 1.1, and table 1.2 holds the fire records for each year.

The records from (ICNF, 2024) span across 10 years, from 2013 to 2023, and were retrieved with an API endpoint. It features 140514 entries, and its most important fields are described in table 1.2, while table 1.4 holds the number of occurrences individually by year.

Table 1.1: Field description of Historical fires from 1980 to 2015 (Central de Dados Github Repository, 2017; centraldedados, 2017)

Variable	Description
ano	Year of fire occurrence
codigo_sgif	Unique identifier for the fire occurrence
tipo	Kind of wildfire. The available options are: forest, slash-and-burn, false alarm, and agricultural.
distrito	District of fire occurrence.
concelho	Municipality of fire occurrence.
freguesia	Parish of fire occurrence.
local	Location of fire occurrence.
ine	Not described or explained anywhere.
x | y	Wildfire location.
data_alerta	Wildfire warning date.
hora_alerta	Wildfire warning hour.
data_extincao	Wildfire complete extinguishment date.
hora_extincao	Hour of wildfire complete extinguishment.
data_primeira_intervencao	Date of first fire intervention
hora_primeira_intervencao	Hour of first fire intervention
fonte_alerta	Authority or group of people who reported the fire first.
nut	A Unique identifier for a given nomenclature of territorial units for statistics.
area_povoamento	Burnt settlement area.
area_mato	Burnt bush area.
area_agricola	Burnt agricultural area.
area_pov_mato	Sum of burned area from the burnt settlement area and burnt agricultural area.
area_total	Total burnt area.
reacendimento	Describes if a given fire is a re-ignition from a previous wildfire.
queimada	Identifies if a fire is a slash-and-burn.
falso_alarme	Identifies if it is a false alarm.
fogacho	Identifies if it is a specific type of fire named a blaze.
incendio	Identifies if it is a fire.
causa	Numerical identifier for the fire cause.
tipo_causa	Description of fire cause. The available options are unknown, deliberate, natural, negligent fire, and undefined.

Table 1.2: Number of Fire occurrences between 1980 and 2015

Year	Number of Occurrences
1980	2346
1981	6727
1982	3625
1983	4536
1984	7355
1985	8439
1986	5036
1987	7703
1988	6130
1989	21895
1990	10743
1991	14327
1992	14951
1993	14799
1994	19983
1995	31116
1996	28626
1997	23494
1998	34675
1999	25473
2000	34107
2001	31582
2002	33697
2003	30345
2004	34722
2005	50364
2006	31445
2007	31122
2008	23139
2009	34979
2010	32357
2011	35941
2012	30740
2013	27372
2014	11387
2015	23175
Total	791453

Table 1.3: Field description of Historical fires from 2013 to 2023 (ICNF, 2024)

Variable	**Description**
CODIGO id	Unique identifier for the fire occurrence
DISTRITO	District of fire occurrence.
TIPO	Kind of wildfire. The available options are: forest and agricultural fire.
ANO	Year of fire occurrence
AREAPOV	Burnt settlement area.
AREAMATO	Burnt bush area.
AREAAGRIC	Burnt agricultural area.
AREATOTAL	Total burnt area.
REACENDIMENTOS	Boolean value for reignition.
FOGACHO	Boolean value for small fire.
Incendio	Boolean value for fire.
Agricola	Boolean value for agricultural fire.
DATAALERTA	Wildfire warning date.
HORAALERTA	Wildfire warning hour.
LOCAL	Location of fire occurrence
CONCELHO	Municipality of fire occurrence.
FREGUESIA	Parish of fire occurrence
FONTEALERTA	Authority or group of people who reported the fire first.
DIA	Day of the fire occurrence.
MES	Month of fire occurrence.
HORA	Fire hour of occurrence.
CAUSA	Numerical identifier for the fire cause.
TIPOCAUSA	Description of fire cause. The available options differ from year to year.
DHINICIO	Firefighting starting date and time
DHFIM	Ending of firefighting efforts
DURACAO	Number of minutes it took to extinguish the fire
HAHORA	Not described
DATAEXTINCAO	Fire extinguish date
HORAEXTINCAO	Fire extinguish hour
QUEIMA	Boolean value for intentional small fire
LAT	Latitude coordinate of fire occurrence
LON	Longitude coordinate of fire occurrence
TEMPERATURA	Value for temperature at the time of fire occurrence
HUMIDADERELATIVA	Relative Humidity value
VENTOINTENSIDADE	Wind velocity
VENTODIRECAO_VETOR	Direction of wind
PRECIPITACAO	Value for rainfall
FFMC	Fine fuel moisture code
DMC	Duff moisture code
DC	Drought code
ISI	Initial Spread index
BUI	Build Up Index
FWI	Fire Weather Index
DSR	Daily Severity Rating
ALTITUDEMEDIA	Mean altitude
DECLIVEMEDIO	Mean Slope

Table 1.4: Number of Fire occurrences between 2013 and 2023

Year	Number of Occurrences
2013	24479
2014	10286
2015	19669
2016	16131
2017	21074
2018	12296
2019	10909
2020	9713
2021	6996
2022	1335
2023	7626
Total	140514

1.2.2 Weather Variables

Historical weather variables were obtained via the *Open-Meteo API* (Zippenfenig, 2023). Which includes daily and hourly weather data extraction from all across the world. The API compares the provided location to several reanalysis datasets and returns the most optimal result based on location. It includes a full database of historical hourly weather conditions dating back to 1940 from multiple sources (table 1.5).

The API incorporates observations from weather stations, aeroplanes, buoys, radar, and satellites and it fills gaps in data using mathematical models to estimate the values of different weather variables (Zippenfenig, 2023).

The variables extracted from *Open-Meteo* are on table 1.6 as well as the unit for each weather variable.

Table 1.5: API Weather Sources (Hersbach et al., 2023; Muñoz Sabater, 2019; Schimanke et al., 2021; Zippenfenig, 2023)

Dataset	Region	Resolution	Availability
ECMFWF IFS	Global	9km, Hourly	2017 to present
ERA5	Global	25km, Hourly	1940 to present
ERA5-Land	Global	11km, Hourly	1950 to present
CERRA	Europe	5km, Hourly	1985 to June 2021

Table 1.6: Hourly weather variables from *Open-meteo*

Variable	Unit	Description
Temperature	°C	Air temperature 2 metres above ground.
Relative Humidity	%	Relative humidity 2 metres above ground.
Dew	°C	Dew point 2 metres above ground.
Apparent Temperature	°C	Apparent temperature is the result of a wind chill factor, relative humidity, and solar radiation.
Pressure	hPa	Atmospheric air pressure reduced to mean sea level.
Surface Pressure	hPa	Surface pressure reduced to mean sea level.
Precipitation	mm	Sum of preceding hour precipitation including rain, showers, and snow.
Rain	mm	Preceding hour of liquid precipitation.
Snowfall	cm	Preceding hour of snowfall amount.
Cloud cover low	%	Fog and low level clouds up to an altitude of 2 kilometres.
Cloud cover mid	%	Clouds floating at a medium level with altitudes ranging from 2 kilometres to six kilometres.
Cloud cover high	%	Clouds floating at an altitude of 6 kilometres.
Shortwave radiation	W/m^2	Shortwave solar radiation.
Direct radiation	W/m^2	Direct solar radiation.
Direct normal irradiance	W/m^2	Direct solar irradiance.
Diffuse radiation	W/m^2	Diffuse solar radiation.
Global tilted irradiance	W/m^2	Total radiation received on a tilted pane.
Sunshine duration	Seconds	Duration of sunshine in seconds.
Wind speed at 10m	km/h	Speed of the wind, 10 metres above ground.
Wind speed at 100m	km/h	Speed of the wind, 100 metres above ground.
Wind direction at 10m	°	Wind direction at 10 metres above ground.
Wind direction at 100m	°	Wind direction at 100 metres above ground.
Wind gusts	km/h	Wind gusts at 10 metres above ground.
Evapotranspiration	mm	Evapotranspiration value for the required irrigation for plants calculated from temperature, wind speed, humidity, and solar radiation.
Weather code	WMO code	Numeric codes for weather conditions.
Snow depth	meters	The depth of snow on the ground.
Vapour pressure deficit	kPa	Vapour pressure deficit in kilopascal.
Soil temperature	°C	Average soil temperature ranging from 0 to 7cm, 7 to 28cm, 28 to 100cm, and 100 to 255cm below ground.
Soil moisture	m^3/m^3	Average soil moisture ranging from 0 to 7cm, 7 to 28cm, 28 to 100cm, and 100 to 255cm depths.

1.2.3 Fuel Variables

Fuel variables encompass two sources: Copernicus Climate Change Service (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019) and Forestry Inventory 2015 (GBIF.Org User, 2024; Uva et al., 2021)

Copernicus Climate Change Service (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019)

The Copernicus Climate Change Service contains historical reconstructions of fire danger indices, i.e., variables that emphasize conditions suitable for the origin, spread, and sustainability of naturally occurring fires. The fire danger indices are obtained from historical simulations and weather forecasts, and the available data starts in January 1940 and extends all the way through 2023. Variables contained in the dataset are expressed in the table 1.7. From this source, all variables that belong to the fire weather index system were extracted except the fire daily severity index, which is not necessary for the calculation of fwi.

Table 1.7: Fire danger indices from historical data (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019)

Variable	Unit	Description
Build-up index	Dimensionless	Weighted combination of the Duff moisture code and Drought code.
Drought code	Dimensionless	Component representing fuel availability, and the influence of recent temperatures and rainfall events on fuel availability.
Duff moisture code	Dimensionless	Moisture content in loosely-compacted organic layers of moderate depth. Duff moisture code fuels are affected by rain, temperature and relative humidity.
Fine fuel moisture code	Dimensionless	Moisture content in litter. Representative of the top litter layer less than 1-2 cm deep.
Fire daily severity index	Dimensionless	A numerical assessment of the difficulty of controlling flames.
Fire weather index	Dimensionless	Combination of Initial spread index and Build-up index. Numerical rating of the potential fire intensity.
Initial spread index	Dimensionless	Combines fine fuel moisture code with weed speed to measure the expected rate of fire spread.

Forestry Inventory 2015 (GBIF.Org User, 2024; Uva et al., 2021)

The Forestry Inventory 2015 contains 579422 occurrences of forest tree species on mainland Portugal. The data was gathered using aerial images and ground surveys. The most important features of the dataset are described in Table 1.8. From this source, the only variable taken was the name of the tree species closest to the wildfire area.

Table 1.8: Forest Inventory 2015

Variable	Description
gbifID | datasetKey | ocurrenceID	Identifiers for the occurrence of trees and the dataset.
kingdom	Kingdom classification of a given Tree.
phylum	Phylum classification of a given Tree.
class	Taxonomic class.
order	Taxonomic Order of a Tree.
genus	Tree genus.
species	Data containing the species of a given tree.
taxonRank	Data containing the highest taxonomic rank available for a given tree group.
scientificName | verbatimScientificName	Scientific name for the available taxonomic classification.
verbatimScientificNameAuthorship	Scientific name authorship for the available taxonomic classification.
countryCode	Contry code of Portugal.
locality	Name of a locality containing a given tree.
stateProvince	Name of a district containing a given tree.
occurrenceStatus	Describes if a tree is still present.
decimalLatitude	Latitude for the tree occurrence.
decimalLongitude	Longitude for the tree occurrence
coordinateUncertaintyInMeters	Uncertainty for a given tree location in metres.
eventDate | year	Year of event record.
taxonKey	Taxonomic key for the highest available classification for a tree
speciesKey	Individual key for a given tree species if available.
speciesKey	Individual key for a given tree species if available.
institutionCode	Unique identifier for ICNF.
collectionCode	Unique collection identifier for the institutionCode.

1.2.4 Topography Variables

Topography includes data from three distinct sources. The mean elevation was taken from GMTED2010 (Danielson and Gesch, 2011). The land class type was extracted from Corine Land Cover (Agency, 2020), and the slope, aspect, and roughness were extracted from GLO-30 (European Space Agency and Sinergise, 2021).

Global Multi-Resolution Terrain Elevation Data 2010 (Danielson and Gesch, 2011)

GMTED2010 is a source that provides raster elevation data. For every wildfire location, the mean elevation was extracted with a resolution of 7.5 arcseconds.

Corine Land Cover (Agency, 2020)

Corine Land Cover is a dataset that provides European land classification from 44 thematic classes. From this source, the land class corresponding to the location of a wildfire event was extracted.

Copernicus GLO-30 Digital Elevation Model (European Space Agency and Sinergise, 2021)

The GLO-30 Digital Elevation Model is a digital surface model. It encompasses the earth's surface with its buildings, infrastructure, and vegetation.

From this source, slope, aspect, and roughness were extracted with a resolution of 30 meters for every wildfire occurrence.

1.2.5 Dataset variables

All the variables used in the dataset are listed in Table 1.9.

Table 1.9: Topography, Weather and Fuel variables

Variable Type	Variable Name	Resolution
Topography	Mean Elevation (m)	_,7.5 arc-seconds, _
	Slope (0-90°)	30m, _, _
	Roughness	30m, _, _
	Aspect (0-360°)	30m, _, _
Weather	Temperature (°C) Relative Humidity (%) Dew (°C) Apparent Temperature (°C) Pressure (hPa) Surface Pressure (hPa) Precipitation (mm) Rain (mm) Snowfall (cm) Cloud Cover Low/Mid/High (%) Shortwave Radiation (W/m^2) Direct Radiation (W/m^2) Direct Normal Irradiance (W/m^2) Diffuse Radiation (W/m^2) Global Tilted Irradiance (W/m^2) Sunshine Duration (Seconds) Wind Speed 10m/100m (km/h) Wind Direction 10m/100m (°) Wind Gusts 10m (km/h) Et0 Evapotranspiration (mm) Weather Code (WMO code) Snow Depth (metres) Vapour Pressure Deficit (kPa) Soil Temperature 0 to 255cm (°C) Soil Moisture 0 to 255cm (m^3/m^3)	0.25/0.1/9km/5km, _ , Hourly
Fuel	Fine fuel moisture code Duff moisture code Drought code Initial Spread index Build Up Index Fire weather index	0.25 , _ , Daily
	Tree Species	-

1.2.6 Additional sources of Data

The python library geopy (Community, 2023) was used to geolocate multiple locations, resolving district, parish, municipalities, and localities to sets of co-ordinates. Geopy utilises multiple geocoding web services like OpenStreetMap Nominatim and Google Geocoding API to resolve locations.

The Google Maps service (Google, 2024) was used to manually check if samples extracted from Open-Meteo corresponded to the intended location. It was also used to analyse some errors that were found in the location of some entries.

1.3 Data Preparation of historical wildfire sites

This section covers the time frame selected for study and the selection of wildfire ignition causes.

1.3.1 Selecting the time frame of wildfire occurrences

The dataset described in 1.1 is composed of multiple files describing historical occurrences since 1980 until 2015. Prior to 2001, the fields from each file became unstandardized, and there's no explicit parameter mentioning a natural wildfire cause. Therefore, the time frame considered was from 2001 to 2012. The latter years were rejected due to the fact that entries from 1.1 do not contain any explicit latitude and longitude. They rely on territorial entities such as districts, municipalities, parishes, and NUTS to describe locations.

The second historical wildfire dataset 1.3 is also composed of multiple files. Its time frame is from 2013 until 2023. Unlike dataset 1.1, entries do contain an explicit latitude and longitude values. It also features descriptive territorial entities.

1.3.2 Ignition cause selection

Historical wildfire sites whose cause was set as deliberate or negligent fire were excluded. The remaining causes left in the dataset, in order of importance, are: natural, reignition, unknown, and undefined. For the creation of the first dataset, all these causes were encompassed. The undefined cause differed from the unknown cause due to the fact that their cause field was left blank, and entries that had unknown causes were explicitly described as unknown. Since the remaining causes aren't an object of study and their importance relies solely on inventory, their analysis won't be assessed in this thesis. Table 1.10 contrasts reignition, unknown, and undefined causes with only natural causes. It is possible to observe that fires whose cause is set as natural have a tiny proportion compared to the outside scope of unknown, undefined, and reigniting causes of ignition. The undefined causes were labeled as *NC* (as for non-characterised) in the dataset.

Table 1.10: Proportion of natural fires compared to other causes from 2001 to 2023

Year	Reignition ⁞ Unknown ⁞ Undefined	Natural Fires
2001	25938	44
2002	25637	13
2003	25042	96
2004	21173	16
2005	34575	3
2006	19108	67
2007	15565	50
2008	9877	28
2009	17293	106
2010	14293	138
2011	14811	102
2012	11785	56
2013	11822	77
2014	3795	38
2015	8293	138
2016	7715	67
2017	10308	104
2018	5445	114
2019	3912	128
2020	3858	95
2021	2507	103
2022	3925	115
2023	2427	72
TOTAL	299104	1770

1.4 Historical Meteorological Data Extraction

1.4.1 Retrieving historical meteorological data

Extracting historical meteorological data for the dataset described in 1.3 was done with ease with a Python script. It went through each historical fire location and downloaded hourly weather data about the entire day regarding the wildfire occurrence. For each wildfire site, a file was created with the date of the occurrence and fire coordinates. All files followed this convention. A latter approach dismissed this method of individually creating files for each site as it proved to be less efficient than having multiple sites in one batch of occurrences (further explanation on this topic will be featured in section 1.6, which covers the creation of the natural fires dataset).

As for the dataset in 1.1, it was necessary to geocode the position of each occurrence individually. For both datasets, all the variables described in table 1.6 were extracted.

1.4.2 Geocoding places from 2001 to 2012 historical wildfire locations

The dataset entries featured in 1.1 contain no direct field leading up to the real site location coordinates. To tackle this issue, an algorithm with the help of the geopy library (Community, 2023) was made to resolve the names of historial wildfire places to a set of coordinates. It was necessary to try multiple combinations to obtain the best results (see table 1.11 for the list of combinations), since the services provided by the geopy library aren't as robust as *Google Maps*. The district, municipality, parish, and local (if available) of each entry were utilized for the purpose of geocoding. Sometimes, the name of the exact wildfire locality was enclosed in brackets, requiring processing using strings to extract it.

Table 1.11: Combinations for local geocoding

Combination
Local, District
Local, Parish, District
Local, Parish, Municipality
Local, Parish, Municipality, District
Parish, Municipality, District
Local, Parish, District

These combinations caused errors in the location of some entries because the geocoders returned coordinates in other countries, such as Spain and Brazil, due to similar names in some locations. The entries that produced errors underwent recalculation, with the addition of "Portugal" at the end. An example of this usage is *Parish, District, Portugal*.

After each entry was resolved, their latitude and longitude were added as values in the columns *LAT* and *LON* of each corresponding wildfire site in the file containing the historical wildfire data.

A very minor sample of entries couldn't be geocoded using this method. Therefore they were manually geocoded from the *Google Maps* service.

1.5 Crafting the First dataset

At this point, wildfire sites from the source 1.1 were geocoded, and the weather data corresponding to the day of occurrence for each wildfire was extracted. It was now necessary to go through each downloaded file and extract the most relevant attributes. Each file contained attributes that were not necessary to explain the problem of fires.

For example, the variables had a separate column with their respective units. Columns that featured units were dropped. The rest of the columns that were not considered were: *latitude and longitude* (these coordinates refer to the weather

station location and not to the wildfire site), *generationtime_ms*, *utc_offset_seconds*, *timezone*, and *timezone_abbreviation*.

These methods were conducted repeatedly each year. At this stage, each year had a separate file with its historical wildfire sites, and the dataset had the following variables: *year, date, district, municipality, parish, local, latitude, longitude, cause, elevation* (this variable comes with the weather data, but since it is not optionally chosen, it hasn't been mentioned before) plus all the weather variables from *Open-Meteo* (table 1.6).

After extracting the weather data for the day of each wildfire occurrence, it was time to match each site with the surrounding tree species.

This dataset does not contain any topographical data and fuel data, except for the surrouding tree species, corresponding to the location of the fires. It only features weather variables at the hour the fire ignited. Although the object of study is fires of natural origin, this dataset was created in order to have a historical reference for fires that occurred between 2001 and 2023 and whose cause could not be ascertained. So that in the future it would be possible to determine the origin of fires of uncertain origin and characterise them. From this dataset, all wildfires whose origin was only natural were extracted.

1.5.1 Matching historical wildfire with tree species

In order to match each fire event with the surrounding tree species, it was necessary to separate the events by district. Within each district, the parish or municipality of each occurrence was matched with the parish or municipality of the tree species. For instances without a match within the district, they were matched with a distance function named *Haversine*, which is a method for calculating the distance between two locations on a sphere's surface based on their latitude and longitude (SimonKettle, 2017). This function calculated the distance between the coordinates of the wildfire event and the coordinates of each tree species location.

The distance calculation featured multiple tests. Experiments were done with 120, 500, and 1000 metres. The first distance threshold resulted in a lot of entries without a match. The distance threshold was slightly increased from 500 metres to 1000 metres, due to the fact that the latter distance was able to perform a match with the majority of wildfire events. Species that were in the same parish or municipality as the fire incident were associated without calculating distance. To combat duplicate species, a tree species was only added if it was not already contained in the fire entry.

For the remaining entries that did not match the previous methods, an analysis was made of the species that were closest. In this step, errors were detected, some values tabulated by the ICNF did not correspond to reality, and some values in 1.4.2 were miscalculated.

The forest inventory data source also contained some errors that had to be assessed. For instance, it wrongly stated a locality with a different name from its real-life counterpart. The locality *Ovadas e Panchora* doesn't exist. It was misspelt,

and its name is *Ovadas e Panchorra*.

Other minor adjustments were also made. For example, each district had the word *District* associated with its district name, like in *Bragança District*, and the string *District* was dropped to ensure standardisation between fire incidents and tree species.

Due to the size of the tree dataset, the Python script divided the years into chunks and used multiprocessing to calculate the tree species near the wildfire event.

1.5.2 Locations in the middle of the sea.

Between 2013 and 2023, some of the featured locations provided by the *ICNF* were in the middle of the ocean. Although using their district, municipality, parish, or local field when using services like *Google Maps* yielded in a real-life location. Their coordinates were undeniably wrong, since they were kilometres away from the nearest coast line. For these entries, weather extraction and tree species matching were redone.

These multiple geolocation errors were discovered when trying to pin multiple species of trees (1.5.1) to a single location with the distance function calculator. The algorithm returned values that were outside of the range spectrum previously set at 1500km. Leading up to the manual confirmation of these errors with the help of the *Google Maps service*.

1.5.3 Dataset description

number of entries of each type

1.6 Natural Fires dataset

1.6.1 FWI Calculation

1.7 Variables Explanation

1.7.1 Creating Areas without wildfires

Random Coordinates that were not equal to the ones in Natural Fires. A minimum distance of 25 km for each generated coordinated and the ones in Natural fires. With openstreet map given the range for latitude (36.7, 42.3) and longitude (-9.8, -6) check if the generated coordinate is in Portugal and check if it belongs to a county to ensure it isn't in the middle of the sea.

1.7.2 Creating Areas without wildfires

Random Coordinates that were not equal to the ones in Natural Fires. A minimum distance of 25 km for each generated coordinated and the ones in Natural fires. With openstreet map given the range for latitude (36.7, 42.3) and longitude (-9.8, -6) check if the generated coordinate is in Portugal and check if it belongs to a county to ensure it isn't in the middle of the sea.

1.7.3 Dataset description

1.8 Python libraries used in the conception of the dataset

requests pandas os to check if files already existed. numpy sklearn

1.9 Entry Selection

Specifity how many entries raw files have.

1.10 Models

1.10.1 Model Variables and their importance

porque de escolher soil moisture 7 to 28cm e nao os outros gráfico de correspondencia.

temperature : 8 relative humidity: 8 precipitation': 16 wind speed: 16 soil temperature 28 to 100cm: 8 soil moisture 7 to 28cm: 8 direct normal irradiance: 8

'fwix': 8, 'ffmc': 8, 'dmc': 16, 'dc': 16, 'isi': 8,

roughness', 'aspect', 'slope', 'CLC CODE','scientificNames', 'mean elev'

tree species foram one hot encoded with MultiLabelBinarizer()

variables with multiple days were flattened

Slope has a significant influence on fire behaviour since it speeds its spread (Marques et al., 2011).

1.10.2 Random Forest Classifier

Model with only Weather Variables

Accuracy: 0.8248587570621468 Precision: 0.8532608695652174 Recall: 0.8177083333333334
F1 Score: 0.8351063829787234

1.10.3 Model with weather and topography

Accuracy: 0.9887005649717514 Precision: 0.9973544973544973 Recall: 0.9817708333333334
F1 Score: 0.989501312335958

1.10.4 Model with weather, topography and fire weather

Accuracy: 0.9830508474576272 Precision: 0.9946808510638298 Recall: 0.9739583333333334
F1 Score: 0.9842105263157894

References

Agency, E. E. (2020). Corine land cover 2018 (vector), europe, 6-yearly - version $2020_20u1, may2020$.

Central de Dados Github Repository (2017). Incendios data repository. https://github.com/centraldedados/incendios.

centraldedados (2017). Incêndios. http://centraldedados.pt/incendios/. Accessed: 2024-01-29.

Community, T. G. (2023). Geopy.

Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (2019). Copernicus Climate Change Service, Climate Data Store, (2019): Fire danger indices historical data from the Copernicus Emergency Management Service. https://cds.climate.copernicus.eu/cdsapp#!/dataset/indices-era5-single-levels?tab=overview. Accessed on 12-02-2024.

Danielson, J. and Gesch, D. (2011). Global multi-resolution terrain elevation data 2010 (gmted2010). Open-File Report 2011-1073, U.S. Geological Survey, Reston, VA, USA.

Encyclopædia Britannica, Inc. (2024). Climate of portugal. Encyclopædia Britannica. Retrieved from https://www.britannica.com/place/Portugal/Climate.

European Space Agency and Sinergise (2021). Copernicus global digital elevation model. https://doi.org/10.5069/G9028PQB. Distributed by OpenTopography.

GBIF.Org User (2024). Occurrence download.

Gillies, S. et al. (2013). Rasterio: geospatial raster i/o for Python programmers.

Google (2024). Google maps. https://www.google.com/maps.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N. (2023). Era5 hourly data on single levels from 1940 to present.

ICNF (2024). Iii.12 zonas de risco natural territórios ardidos (Área ardida entre 1975 e 2023). https://geocatalogo.icnf.pt/catalogo_tema5.html. Accessed: 2024-01-30.

Living and Overseas, I. (2023). Portugal weather and climate. https://internationalliving.com/countries/portugal/portugal-weather-and-climate/.

Marques, S., Borges, J. G., Garcia-Gonzalo, J., Moreira, F., Carreiras, J. M. B., Oliveira, M. M., Cantarinha, A., Botequim, B., and Pereira, J. M. C. (2011). Characterization of wildfires in portugal. *European Journal of Forest Research*, 130(5):775–784.

Mora, C. and Vieira, G. (2020). *The Climate of Portugal*, pages 33–46. Springer International Publishing, Cham.

Muñoz Sabater, J. (2019). Era5-land hourly data from 2001 to present.

Schimanke, S., Ridal, M., Le Moigne, P., Berggren, L., Undén, P., Randriamampianina, R., Andrea, U., Bazile, E., Bertelsen, A., Brousseau, P., Dahlgren, P., Edvinsson, L., El Said, A., Glinton, M., Hopsch, S., Isaksson, L., Mladek, R., Olsson, E., Verrelle, A., and Wang, Z. (2021). Cerra sub-daily regional reanalysis data for europe on single levels from 1984 to present.

SimonKettle (2017). Distance on a sphere: The haversine formula. https://community.esri.com/t5/coordinate-reference-systems-blog/distance-on-a-sphere-the-haversine-formula/ba-p/902128. Accessed: 2024-03-05.

Uva, J. S., Onofre, R., Moreira, J., Faias, S. P., Barreiro, S., Santos, E., Capelo, J., Corte-Real, L., Martins, J., Ribeiro, J. R., Cancela, J., Rainha, M., Amaral, N., Santos, C., Perpétua, J., Pinho, J., Araújo, J. M., Reis, L., Canaveira, P., Paulino, J., Pina, A., Binev, Y., and Coelho, P. (2021). Forestry inventory 2015. https://doi.org/10.15468/33hvm4. Accessed via GBIF.org on 2024-04-04.

Zippenfenig, P. (2023). Open-meteo.com weather api.