36105 iLab 2 - Spring 2023

Enhancing Wildfire Management
in Greece
By
Team 3 (Pyro Vision)

Final Project Report and Individual contributions

Ahmed Khursheed

Ashutosh Patil

Nutan Thapa

Shivani Nandkishor Nipane

Table of Contents

1. Executive Summary

In this executive summary, we offer a succinct yet comprehensive glimpse into our endeavor to combat the Wildfire crisis using data, systems, and tools. Our project encapsulates a multidisciplinary approach to understanding, predicting, responding to, and recovering from wildfires, thereby offering invaluable support to the client organization.

2. Introduction

2.1 Background on the Wildfire Crisis and framework choosen

Since World War II, emergency management has primarily focused on preparedness, often geared toward enemy attack scenarios. Community preparedness for all types of disasters involves identifying resources and expertise in advance and planning for their utilization during a disaster. Preparedness is one of four key phases in modern emergency management.

The four phases of emergency management are:
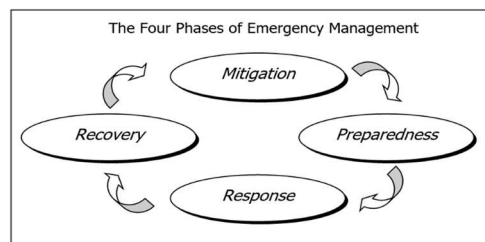
Given this framework, we have chosen to address the wildfire crisis problem and apply data science tools to develop a comprehensive plan covering all four phases of emergency management.

Why choose Greece ?

The phenomenon of wildfires is not new to our planet. However, recent years have witnessed an alarming increase in the frequency and intensity of these fires across the globe. From the vast rainforests of the Amazon to the bushlands of Australia, no region has remained untouched. The consequences are multifaceted, affecting the environment, economy, and human settlements.

Greece: A Nation Under Siege

- ➢ Greece, with its Mediterranean climate and extensive forested areas, has always been susceptible to wildfires.
- ➢ However, the past decade (2013-2023) has seen an unprecedented surge in these destructive events.
- ➢ Each year, news headlines relay harrowing tales of villages evacuated, ancient forests reduced to ash, and sadly, lives lost.
- ➢ Notably, in August 2023, Greece faced the largest wildfire in the European Union.

This recent development in August 2023 is a key reason why our group, Pyro Vision, has chosen Greece as the focal point for our wildfire research efforts. The choice of Greece is ideal not only because of its vulnerability to wildfires but also because of the urgent need for effective solutions, given the recent escalation in wildfire incidents.

2.2 Purpose and Scope of the Report

The purpose of this report is to outline the comprehensive approach and multifaceted strategies we have employed to address the Wildfire crisis, particularly in the context of Greece. Our mission is to harness the power of data, systems, and advanced tools to enhance preparedness, response, and recovery efforts in the face of this escalating challenge. The report serves several key purposes:

> ➢ **Documenting Our Approach**: This report documents the methodologies, techniques, and tools utilized by our team, known as Pyro Vision, to address the Wildfire crisis. It provides an in-depth account of our problem-solving strategies and the rationale behind their selection.

> ➢ **Communicating Key Findings**: We aim to communicate our key findings and recommendations clearly and concisely. These findings are the result of extensive research, analysis, and application of data science and technology to mitigate the impacts of wildfires.

> ➢ **Providing Insights to the Client Organization:** Our primary audience is the client organization that has engaged our services. We intend to offer valuable insights that can inform their decision-making processes and enhance their capacity to respond to wildfire crises.

> ➢ **Sharing Lessons and Best Practices**: We also seek to contribute to the broader field of emergency management by sharing our experiences, lessons learned, and best practices. This report can serve as a resource for other organizations and individuals working to address similar challenges.

> ➢ **Guiding Future Initiatives**: The scope of this report extends to offering recommendations for the client organization and, by extension, other stakeholders involved in wildfire crisis management. These recommendations are designed to guide future initiatives and improvements in the field.

3. Wildfire Prediction Model

The primary objective of this model is to ascertain the likelihood of wildfire incidents transpiring across Greece on a monthly basis, spanning from January 2022 to August 2023. The geographical area of Greece is divided into grid segments characterized by one-degree precision along both latitude and longitude coordinates. The aim of these predictions is to facilitate more focused and rigorous monitoring efforts within potential fire-prone regions and to enable rapid response mechanisms, ensuring early intervention in the event of fire initiation.

3.1 Data Collection

To facilitate wildfire prediction, we engaged in comprehensive data collection, targeting both historical wildfire data and external factors that significantly contribute to wildfire occurrence. An essential component of this data collection process involved trend analysis, a valuable tool for wildfire prediction, as it sheds light on the evolving patterns in these incidents. Of particular importance is the unpredictable nature of temperature fluctuations in recent times, making it a paramount external factor influencing wildfires.

While numerous external factors, such as human negligence, wind speed, and flora moisture, play pivotal roles in wildfire dynamics, they are often challenging to obtain in their entirety. In light of this, our model primarily focused on temperature, a variable that possesses high impact and is particularly challenging to predict accurately. Temperature fluctuations are a critical aspect of wildfire occurrence, and by incorporating this factor into our model, we aimed to enhance the precision and effectiveness of our wildfire predictions.

(i). Wildfire data collection

For the purpose of wildfire data collection, we employed a valuable resource known as NASA Firms, specializing in the distribution of Near Real-Time (NRT) active fire data. This dataset is made available within a remarkably short span of three hours following satellite observations conducted by the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard the Aqua and Terra Satellites.

Our data collection endeavor spanned from January 2013 to August 2023. It's noteworthy that, due to the limitations of the platform, only one year's worth of data could be requested at a time. Consequently, we made ten separate requests, each focusing on a different yearly timeframe.

It's important to mention that NASA Firms provides access to data originating from various satellite observations, including the Visible Infrared Imaging Radiometer Suite (VIIRS) aboard S-NPP and NOAA 20. In the process of data acquisition, we meticulously evaluated the completeness and reliability of data from these sources. After thorough comparison and assessment, we arrived at the decision to predominantly rely on data obtained from MODIS satellite observations for our wildfire prediction model.

Dataset received:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	latitude	longitude	brightness	scan	track	acq_date	acq_time	satellite	instrument	confidence	version	bright_t31	frp	daynight	type
2	41.092	22.2347	303.4	1	1	03/01/2013	1133	Aqua	MODIS	56	6.03	282.8	7.4	D	C

The datasets were provided in the form of CSV files, and after successfully receiving data for each of the ten years from the MODIS source, a consolidation effort was undertaken. These ten individual CSV files were merged into a singular, comprehensive Wildfire dataset. This consolidated dataset was then poised for the subsequent stages of data preparation and exploration, which are essential steps in our wildfire prediction model development.

(ii). Temperature data collection

Acquiring temperature data in a format that could be effectively integrated with wildfire data, considering specific coordinates, months, and years, proved to be a complex and time-consuming endeavor. To accomplish this, we sourced temperature datasets from various stations and coordinates across multiple sites. Our data collection process involved the following platforms:

A. Berkeley Earth (**https://berkeleyearth.org/data/**)

Initially, we attempted to procure temperature data from Berkeley Earth. However, the temperature data for Greece obtained from this source did not meet our requirements in terms of accuracy and comprehensiveness. It is possible that their data may be more suitable for other countries, but for our specific needs in Greece, it fell short.

B. European Climate Assessment and Dataset (**https://www.ecad.eu/**)

➢ Our alternative data source was the European Climate Assessment and Dataset. Navigating to the data section on the website, we accessed valuable temperature datasets. This resource provided the necessary information to meet our data requirements, ultimately allowing us to retrieve comprehensive temperature data for integration into our wildfire prediction model.
➢ The dataset obtained from the European Climate Assessment and Dataset was quite extensive; however, it posed a challenge due to the presence of numerous invalid rows containing temperature values that could not be relied upon.

C. Weather and Climate (https://weatherandclimate.com/greece#citytemper)

➢ As part of our rigorous data collection process for the Greece case, we also turned to the website Weather and Climate (https://weatherandclimate.com/greece#citytemper). This particular source proved to be highly valuable as it provided temperature data for Greece on a monthly basis, down to the state and city levels.
➢ What makes this source particularly advantageous is its granularity, offering detailed temperature information not only at the national level but also for various cities within Greece. This enriched dataset facilitated our ability to collect accurate and relevant temperature data, which is indispensable for the success of our wildfire prediction model.

While the Weather and Climate website provided valuable temperature data for Central Greece, it's important to note that not all cities were covered, and critical information such as latitudes and longitudes was missing for the cities listed. To address this limitation, our team proactively identified and supplemented the missing geographical coordinates of these cities. This enabled us to replace invalid data with accurate temperature information.

Despite our efforts to enhance the dataset, some locations remained unrepresented. In response to this challenge, we embarked on further investigation, eventually discovering an additional data source that held the potential to extract the missing data needed to bolster our temperature dataset for Greece.

D. Weather and Climate Central Greece(https://weatherandclimate.com/greece/central-greece)

➢ The ultimate solution to our data collection endeavor came in the form of this website . This resource proved to be a game-changer as it offered historical temperature data with detailed information regarding years and months. Moreover, it notably provided latitudes and longitudes for specific locations within Greece.

➢ This extensive dataset proved instrumental in plugging the gaps within our temperature data. This meticulous data acquisition process ensured that our records were not only comprehensive but also highly accurate. The incorporation of latitudes and longitudes further elevated the richness of our dataset, which serves as a cornerstone for the success of our wildfire prediction model. As a result of these efforts, our final dataset was meticulously crafted:
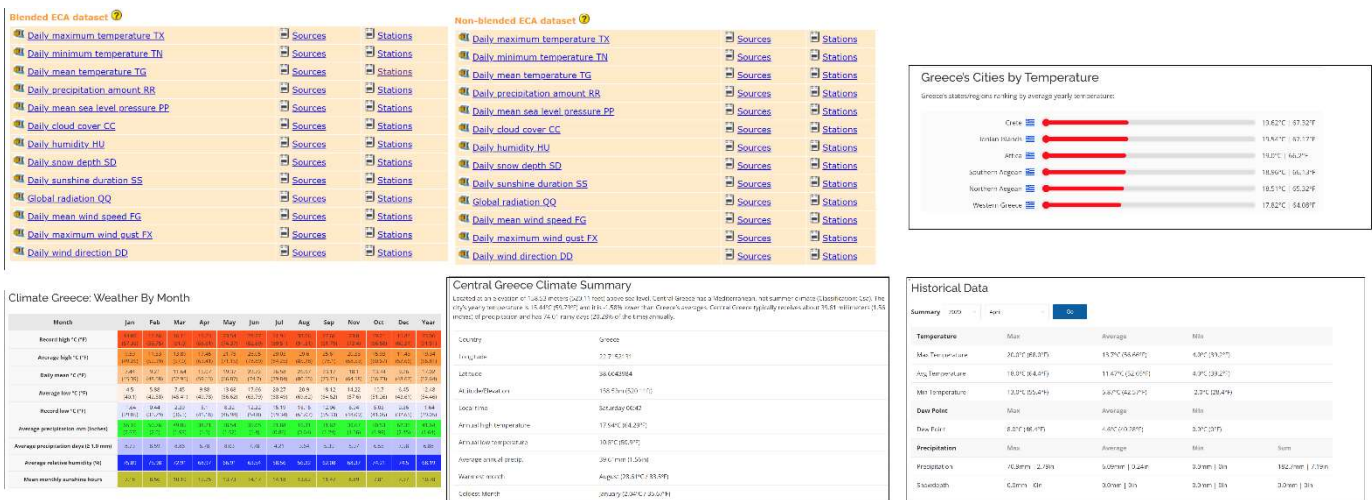
3.2 Data Understanding

(i). Wildfire data

The initial wildfire data provide to use consist of 15 columns: latitude ,longitude, brightness, scan, track, acq_date, acq_time, satellite, instrument, confidence, version, bright_t31, frp, daynight and type.

But all these columns were not required. Relevant columns were selected by filtering the wildfire data to include records within specified latitude (34 to 42) and longitude (19 to 29) ranges. Subsequently, the data was organized into groups based on latitude, longitude, acquisition date, satellite, and instrument. Within each group, the highest confidence value was retained while records with low confidence (confidence < 50) were excluded. The resulting dataset of daily fire records was then collected in the 'chunks' list for subsequent analysis and use.

	A	B	C	D	E	F
1	latitude	longitude	acq_date	satellite	instrument	confidence
2	34.81	24.12	08/05/2018	Terra	MODIS	62
3	34.94	26.13	26/08/2023	Aqua	MODIS	60

(ii). Temperature data

The final temperature dataset after changing coordinates and cleaning data looks like:

	A	B	C	D	E	F	G
1	latitude	longitude	year	month	temperature_min	temperature_avg	temperature_max
2	35	25.7	2018	3	7	12	16.5
3	37.9	23.8	2016	10	15.5	20	24
4	38.9	22.4	2016	4	8.1	13.3	18.3

There were seven columns present in this dataset now: latitude, longitude, year, month, temperature minimum, temperature maximum and temperature average.

3.3 Wildfire Data processing

The wilfire dataset merged file was processed in following order:

- ➢ The merge-wilfire data was uploaded to a code file.
- ➢ Then the data was organised into groups based on latitude, longitude, acquisition date, satellite, and instrument. Within each group, the maximum confidence value was retained. Records with low confidence (confidence < 50) were excluded from the dataset.
- ➢ The resulting dataset consisted of aggregated daily fire records.
- ➢ The final dataset was saved as an csv with name wildfire processed data.

	A	B	C	D	E	F
1	latitude	longitude	acq_date	satellite	instrument	confidence
2	34.81	24.12	08/05/2018	Terra	MODIS	62
3	34.94	26.13	26/08/2023	Aqua	MODIS	60

3.4 Temperature Data Processing

In the context of processing temperature data, the dataset created after scraping and requesting information from various sources was found to be in the appropriate format for our work. Upon uploading this dataset, a critical step involved ensuring that the latitude and longitude values fell within the predefined ranges established for the wildfire dataset.

```
print(f"Latitude Range: {min_latitude} to {max_latitude}")
print(f"Longitude Range: {min_longitude} to {max_longitude}")

Latitude Range: 35.0 to 40.8
Longitude Range: 19.9 to 28.1
```

The examination yielded the following results:
Latitude Range: 35.0 to 40.8 and Longitude Range: 19.9 to 28.1

These findings confirmed that the temperature data was well-aligned with the specified latitude and longitude ranges, ensuring its relevance to the wildfire dataset.

Subsequently, a meticulous check for null values was conducted to verify the absence of unavailable data. This precaution was taken to avert potential issues in our model's performance and analysis in the future.

3.5 Data Creation

Before the merger of temperature and wildfire datasets, certain essential columns were required within the wildfire dataset.

- ➢ The process commenced with the loading of wildfire data and the conversion of the 'acq_date' column into a datetime format.
- ➢ The data was meticulously refined to ensure its quality and relevance. High-confidence wildfire records were retained based on a defined confidence threshold (CONFIDENCE_THRESHOLD), and the dataset was focused further by narrowing it down to records within predefined geographic boundaries (LAT_RANGE and LON_RANGE). This stringent selection process guaranteed that the dataset encompassed pertinent and dependable wildfire incidents within the specified geographical area.
- ➢ Temporal attributes were enhanced by creating new 'year' and 'month' columns from the 'acq_date.' Simultaneously, spatial attributes were fine-tuned by rounding latitude and longitude coordinates to a specified precision level (PRECISION). These preparations equipped the dataset for comprehensive analysis and modeling, aligning both temporal and spatial attributes to optimize wildfire predictions and management strategies.
- ➢ Following these preparatory steps, the dataset was further structured. It was grouped based on 'latitude,' 'longitude,' 'year,' and 'month,' summarizing wildfire occurrences within these specific temporal and spatial dimensions. The resulting dataset included columns denoting 'latitude,' 'longitude,' 'year,' 'month,' and **'fire_cnt' (indicating wildfire incident counts)**, enhancing its suitability for advanced analysis and modeling.

This **'fire_cnt'** column, denoting the count of wildfire incidents, was systematically generated as part of the data preparation process. Initially, the dataset was organized by grouping it based on specific attributes, namely 'latitude,' 'longitude,' 'year,' and 'month.' These groupings created distinct subsets of data, each representing a unique combination of geographical location and temporal intervals. Within each of these groupings, an aggregation operation was applied, counting the number of records associated with each specific combination. This count essentially represented the frequency of wildfire incidents occurring at a particular 'latitude' and 'longitude' during a given 'year' and 'month.' Subsequently, the computed counts were consolidated into a new column termed 'fire_cnt.' This column served as a quantitative indicator of the prevalence of wildfire incidents, offering a valuable metric for in-depth analysis and modeling. By generating the 'fire_cnt' column, the dataset became better equipped for detailed examination and prediction of wildfire patterns and trends, enhancing its utility for the project's objectives.

➢ After this, a comprehensive grid of coordinates was systematically generated to facilitate spatiotemporal analysis. This involved the creation of latitude and longitude coordinates within the specified geographic range defined by **'LAT_RANGE'** and **'LON_RANGE'**. The precision of these coordinates was determined by the project's 'PRECISION' parameter. Additionally, unique **'year'** and **'month'** values were extracted from the wildfire dataset, forming the temporal dimensions of the analysis.

➢ Subsequently, two separate sets of unique combinations were created using these coordinates. The first set combined latitude and longitude coordinates, producing a 'coords' DataFrame with columns for 'latitude' and 'longitude.' The second set comprised combinations of 'year' and 'month,' forming a 'times' DataFrame with columns 'year' and 'month.' Both DataFrames were enhanced with a constant value, 'one,' to facilitate their integration into a unified framework.

➢ The 'base' DataFrame was then created by merging the 'coords' and 'times' DataFrames. This operation was executed as an outer join on the common 'one' column, ensuring that all possible spatial and temporal combinations were represented. Subsequently, the 'history' DataFrame was formed by merging the 'base' DataFrame with the wildfire data. The merger was based on shared attributes, including 'latitude,' 'longitude,' 'year,' and 'month.' This structured dataset, 'history,' laid the groundwork for comprehensive spatiotemporal wildfire analysis, enabling the project to progress with advanced modeling and predictions and to make it proper the null values was replace with 0.

➢ Following the establishment of this framework, the project embarked on the creation of additional columns to enhance the dataset's analytical capabilities. One pivotal addition was the **'fire'** column, which served as a binary indicator. This column was determined based on the comparison of **'fire_cn**t' (fire incident count) against a predefined minimum threshold, denoted as **'MIN_FIRE_RECORDS**.' The 'fire' column helped discern and highlight significant fire incidents within the dataset, adding a valuable dimension to the analysis.

➢ Subsequently, the project sought to gain deeper insights into the dataset's temporal and spatial attributes. To achieve this, two new DataFrames, 'yearly' and 'monthly,' were thoughtfully formulated. The 'yearly' DataFrame was crafted by grouping data based on 'latitude,' 'longitude,' and 'year.' Within these groups, the total counts of 'fire_cnt' and 'fire' were meticulously calculated, presenting an annual perspective of fire incidents within each spatial unit. In a similar vein, the 'monthly' DataFrame was curated through groupings extending to include 'month' as an additional attribute, offering a monthly perspective of fire incidents across the spatial grid.

➢ An additional DataFrame, 'last_year,' was derived from the 'yearly' DataFrame, representing data from the preceding year. This was accomplished by duplicating the 'yearly' data and incrementing the 'year' attribute by one. The columns in 'last_year' were thoughtfully adapted to include **'fire_cnt_last_year'** and **'fire_last_year,'** signifying the counts and occurrences of fire incidents from the previous year.

➢ In parallel, a historical perspective was introduced through the creation of the 'past' DataFrame. This DataFrame originated from the 'yearly' data and included an additional 'one' column. The process involved a series of strategic steps that incorporated only relevant data from the 'history' dataset and facilitated the computation of fire counts and occurrences in previous years. Consequently, the 'past' DataFrame featured columns representing **'latitude,' 'longitude,' 'year,' 'fire_cnt_before,'** and **'fire_before,'** offering a historical context for each spatial unit within the dataset.

- ➢ These thoughtful data transformations, encompassing the creation of multiple new columns and derived DataFrames, enriched the dataset's temporal and spatial attributes significantly. This enhancement served as a solid foundation for more comprehensive spatiotemporal wildfire analysis and predictive modeling, amplifying the project's capabilities for informed decision-making and management.
- ➢ After these meticulous data transformations, including the creation of multiple new columns and derived DataFrames, the final wildfire dataset was enriched with enhanced temporal and spatial attributes, setting the stage for comprehensive spatiotemporal wildfire analysis and predictive modeling. The subsequent step involved integrating these enriched attributes into the main dataset 'X' by merging **'history,' 'past,' 'last_year,' and 'last_year_month' data**. Redundant columns were thoughtfully removed, streamlining the 'X' dataset and making it ready for advanced analytical and modeling tasks.

The final wildfire data was made and printed:

latitude	longitude	year	month	fire_cnt	fire	fire_cnt_before	fire_before	fire_cnt_last_year	fire_last_year	fire_cnt_last_year_same_month	fire_last_year_same_month
34.0	19.0	2013	7	0.0	0	NaN	NaN	NaN	NaN	NaN	NaN
34.0	19.0	2013	5	0.0	0	NaN	NaN	NaN	NaN	NaN	NaN

The Nan values here are because fire is 0 and in case of fire equals to 1 which impiles fire occurs these columns contains some valuebased on that.

Merging Final Wildfire and Temperature data.

- ➢ Following the integration of historical and spatial attributes into the dataset 'X,' the next step involved merging this dataset with the **temperature dataset, 'temp_df.'** This merge operation was executed using common attributes such as **'month,' 'year,' 'latitude**,' and **'longitude,'** performed as an inner join to retain only matching rows between the datasets. The resulting DataFrame, 'X,' displayed the combined dataset with the merged temperature information, providing a comprehensive foundation for advanced spatiotemporal wildfire analysis and modeling.
- ➢ The Nan values after merging dataset was replaced with 0 and this dataset was saved as a csv file **named final_data**.
- ➢ So we have a final data file which is shown below:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	latitude	longitude	year	month	fire_cnt	fire	fire_cnt_before	fire_before	fire_cnt_last_year	fire_last_year	fire_cnt_last_year_same_month	fire_last_year_same_month	temperature_min	temperature_avg	temperature_max
2	35	25.7	2013	7	0	0	0	0	0	0	0	0	21.6	26.4	31.6

The final dataset has new columns and some old. The pipeline that will be use in model for predicton has columns given below along with their decsription.

Columns	Description
Latitude	Angular distance north or south of earth equator.
Longitude	Angular distance of place east or west of Greenwich meridian.
Month	Month of wildfire predict period.
Fire Report count Before	Number of fire report in past for specific latitude – longitude pair.
Fire count (Before)	The number of fires with probability with probabiltiy of occuring in the past for a specific latitude – longitude pair.
Fire Report (last year)	Number of fire report in last year for specific latitude – longitude pair.
Fire Count (last year)	The number of fires with probability of occuring in the last year for a specific latitude – longitude pair.
Fire Report (Same month last year)	Number of fire report in same month of the last year for specific latitude – longitude pair.
Fire Count (Same month last year)	The number of fires with probability of occuring in same month of last year for a specific latitude – longitude pair.
Min. Temp.	Lowest Temperature during month
Max. Temp.	Highest Temperature during month
Avg. Temp.	Average Temperature during month.

> ➤ The final stage of data creation involved partitioning this final dataset into three distinct sets: a training set, a validation set, and a test set. This partitioning is essential for evaluating the predictive performance of the model. The years were chosen as temporal boundaries to define these sets.

> ➤ The training set, covering the years from 2013 to 2018, serves as the foundation for training the predictive model. During this period, the model learns from historical data to recognize patterns and relationships between wildfire incidents and various factors.

> ➤ The validation set spans the years from 2019 to 2021 and acts as an intermediate evaluation stage. Models trained on the training data are assessed on the validation set to fine-tune their parameters and ensure their generalization to new data.

> ➤ The test set covers the years 2022 and 2023 and is reserved for the final evaluation of the model's performance. It represents unseen data, allowing us to assess how well the model can make predictions on new wildfire incidents.

> ➤ This division into distinct sets, aligned with specific timeframes, is a crucial step in ensuring the model's ability to make reliable predictions and evaluate its effectiveness. It helps prevent overfitting to historical data and provides a robust foundation for further analysis and prediction of wildfire incidents. And they were saved as a csv file at last.

3.6 Data Analysis

The wildfire and temperature datasets underwent separate analyses, each involving an array of graphical representations to gain insights into the data's characteristics and distributions. These preliminary analyses were crucial in providing a clear understanding of the datasets before embarking on the development of predictive models. A variety of graphs, including histograms, scatter plots, time series plots, and heatmaps, were crafted to explore the data's structure, patterns, and anomalies. These visualizations not only revealed the temporal and spatial variations within each dataset but also helped identify potential correlations, outliers, and any data-related nuances. This comprehensive exploration of the data served as a foundational step, informing subsequent model development and ensuring that any underlying patterns and trends were fully considered in the predictive processes.
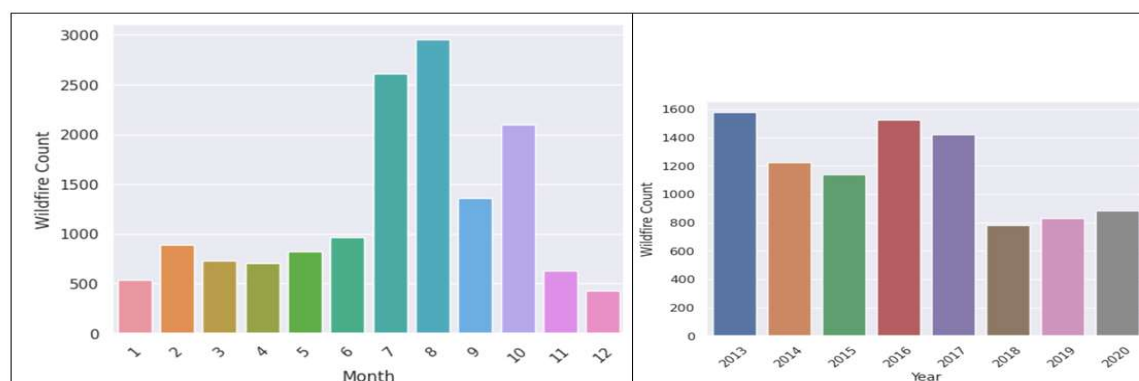
Some of the visuals are given below
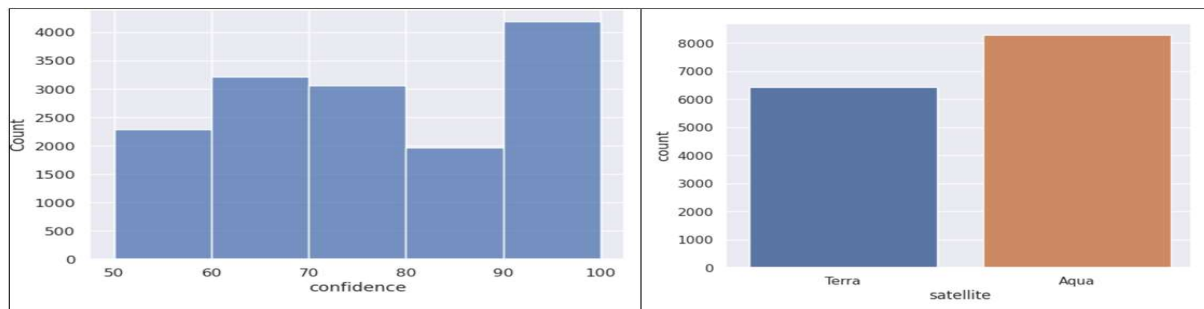
A. Wildfire-process data

❖ **Total Wildfire counts for latitude and longitude intervals**
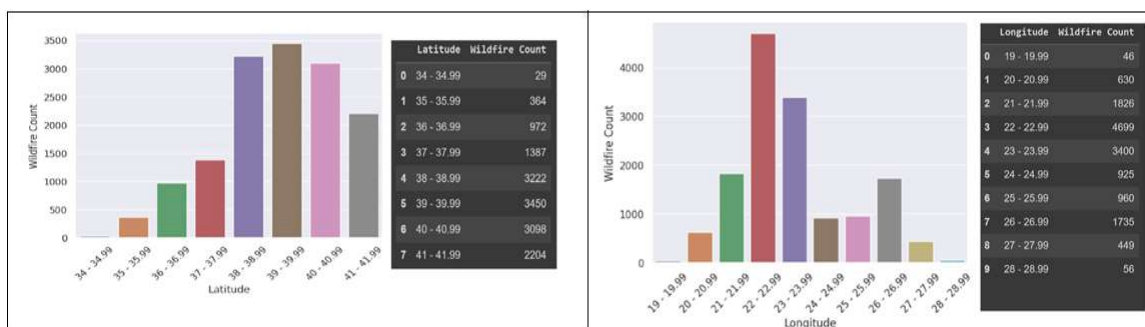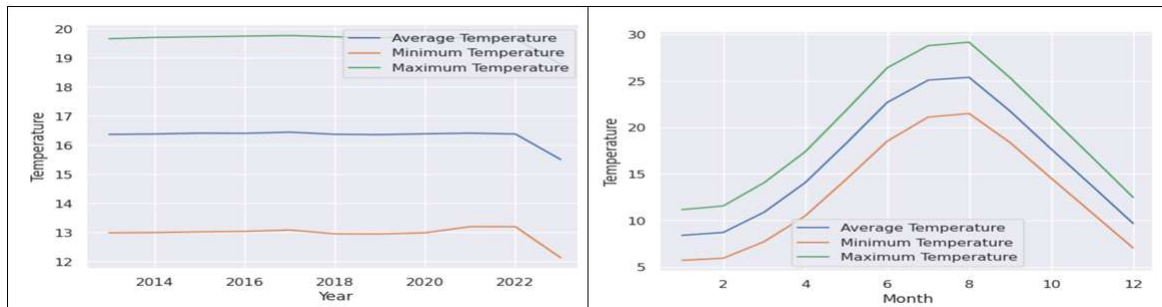

❖ **Total wildfire count by month and year**

10

❖ **Distribution of detection confidence and Sample count of satellite types**

B. Temperature data

❖ **Yearly temperature means for Greece**
❖ **Minimum, maximum and average temperatures for each month**

3.7 Model Tuning

In the model tuning phase, we focused on optimizing the hyperparameters of our predictive model to enhance its performance. We employed Optuna, a hyperparameter optimization framework, to systematically search for the best set of hyperparameters. Our primary objective was to maximize the ROC-AUC score, a crucial metric for assessing the classifier's performance. The following steps were taken:

A. Data Loading

We started by reading the training, validation, and test datasets to prepare our data for the model optimization process.

B. Feature Selection

We defined a set of relevant feature columns to be used in the model optimization. These features included geographical and temporal attributes, past fire incident counts, historical data, and temperature-related information.

C. Objective Function

We defined the objective function for optimization. Our goal was to maximize the ROC-AUC score. The objective function was designed to take different hyperparameters as input, and it utilized LightGBM, a gradient boosting framework, for classification tasks.

D. Hyperparameter Optimization

Using Optuna, we systematically searched for the best hyperparameters for our model. The parameters considered included regularization terms (lambda_l1 and lambda_l2), feature subsampling (colsample_bytree and subsample), learning rate, maximum depth of trees (max_depth), number of leaves (num_leaves), feature fraction, bagging settings, and other parameters. The optimization process was conducted over multiple trials (n_trials=100) in parallel for efficiency (n_jobs=-1).

3.8 Model Training and Evaluation

Data Preparation:

- ❖ We began by reading the training, validation, and test datasets, which contained the necessary data for our predictive model.
- ❖ A set of essential feature columns was defined, encompassing geographical coordinates, temporal attributes, past fire incident counts, historical data, and temperature-related information.

LightGBM Model Configuration

- ❖ Dataset objects were created for the LightGBM model (an advance decision tree alogrithm), using the defined features and target variable.
- ❖ Optimized hyperparameters for the LightGBM model were set. These hyperparameters were obtained through the optimization process described earlier and included parameters for binary classification, regularization terms (reg_alpha and reg_lambda), and the number of leaves in decision trees (num_leaves).
- ❖ Early stopping was enabled during training to prevent overfitting, with a maximum of 20 rounds without improvement.

Model Training

- ❖ The model was trained using the training dataset with the specified hyperparameters and early-stopping criteria.
- ❖ A fixed number of training rounds (500) was conducted to ensure the model's convergence.

Model Evaluation:

- ❖ Predictions were generated on the test dataset using the trained model. To obtain a binary classification, predictions with a probability threshold of 0.1 were considered as "fire" incidents.
- ❖ A confusion matrix was plotted to visualize the model's performance in terms of true positives, true negatives, false positives, and false negatives.
- ❖ A classification report was presented, providing detailed metrics such as precision, recall, and F1-score for both "fire" and "not fire" classes.
- ❖ The ROC-AUC score was calculated to assess the model's ability to distinguish between positive and negative classes.
- ❖ An ROC/AUC curve was plotted to visualize the trade-off between true positive rate and false positive rate.

Decision Tree Visualization

A decision tree from the trained LightGBM model was visualized, providing insights into how the model makes predictions.
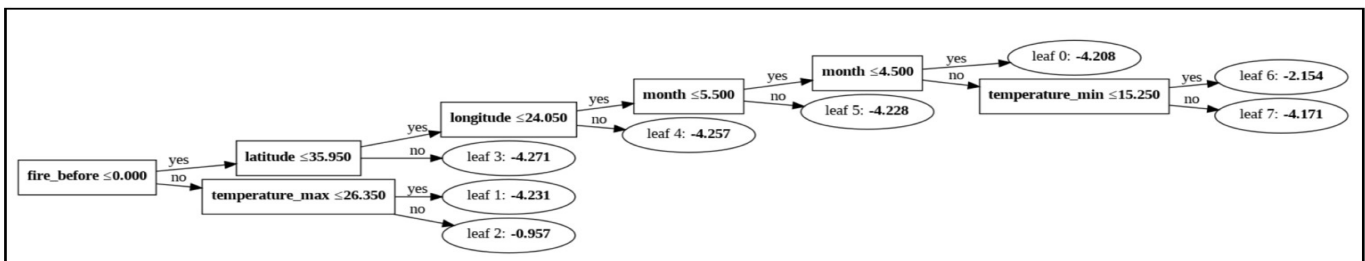
3.9 Key Finding

- ❖ **Geographical Hotspots:** The highest wildfire counts were observed in a specific geographical region, primarily in the latitude range of 39 – 39.99 and the longitude range of 22 – 22.99. This area demands special consideration and focused wildfire management efforts to mitigate the recurring incidents.
- ❖ **Temperature Fluctuations**: While some fluctuations in temperature were noted, it's challenging to attribute these fluctuations as the sole major cause of wildfires. It's plausible that other contributing

factors may have a more substantial impact on the occurrence of wildfires. Further investigation and data analysis are needed to uncover these potential influencers.

- ❖ **August Wildfire Occurrence**: The data highlighted a distinct pattern with the majority of wildfires occurring during the month of August. This observation aligns with background information provided, as August 2023 witnessed one of the most significant wildfires in Greece, marking it as one of the largest wildfire events in Europe. This emphasizes the temporal significance of August and underscores the need for heightened preparedness during this month.
- ❖ **High Precision for Fire Predictions** The model exhibits remarkable precision for identifying fire incidents, with a precision score of 1.00 for "not fire" class and 0.25 for the "fire" class. This indicates that when the model predicts a fire incident, it is correct in 25% of the cases, and when it predicts "not fire," it is correct in 100% of the cases.
- ❖ **Recall Highlights for Fire** The model demonstrates excellent recall for the "fire" class, achieving a score of 1.00. This implies that it effectively captures all instances of actual fire incidents. However, it's important to note that the "not fire" class also has a recall of 0.99, meaning that it correctly identifies most non-fire instances.
- ❖ **F1-Score Balance** The F1-score, which balances precision and recall, stands at 0.41 for the "fire" class. While this indicates relatively lower performance in correctly identifying fire incidents, the balance between precision and recall is generally robust.
- ❖ **Outstanding Overall Accuracy** The model attains an impressive overall accuracy of 99%, signifying its proficiency in classifying both fire and non-fire instances.
- ❖ **Roc Auc Score** The Receiver Operating Characteristic Area Under the Curve (ROC AUC) score is an outstanding 0.997, which showcases the model's exceptional ability to distinguish between fire and non-fire instances.

```
              precision    recall  f1-score   support

    not fire       1.00      0.99      1.00     13240
        fire       0.25      1.00      0.41        31

    accuracy                           0.99     13271
   macro avg       0.63      1.00      0.70     13271
weighted avg       1.00      0.99      1.00     13271
```

```
[ ] test_auc = roc_auc_score(test.fire, test_predictions)
    test_auc

    0.9965634441087613
```

Not Fire	Implication	Fire	Implication
Precision (1.00)	Model predicts an instance as "not fire," it is correct 100% of the time	Precision (0.25)	Model predicts an instance as "fire," it is correct 25% of the time
Recall (0.99)	Model effectively captures 99% of the actual "not fire" instances. It misses 1% of the "not fire" cases (false negatives).	Recall (1.00)	Model effectively captures all (100%) of the actual "fire" instances.
F1-Score (1.00)	Indicating a balance between precision and recall for the "not fire" class	F1-Score (0.41)	The F1-score for the "fire" class is 0.41, reflecting a balance between precision and recall.
Support (13240)	The support represents the number of instances in the "not fire" class in the test dataset, which is 13,240.	Support (31)	The support represents the number of instances in the "fire" class in the test dataset, which is 31.

4. SMOKE AND FIRE DETECTION

Object detection is a crucial computer vision task that involves identifying and localizing objects within images or video frames. In this report, we will explore the essential aspects of data understanding and data creation for our object detection tasks. We will discuss the importance of datasets, their types, and the steps involved in creating, preparing and using the data for our goal of smoke and fire detection.

4.1 Types of Datasets

Datasets play a pivotal role in training and evaluating object detection models. Ultralytics offers a diverse range of datasets tailored to various computer vision tasks, including:

- ➢ **Detection Datasets**: These datasets are designed for bounding box object detection, where the goal is to detect and localize objects by drawing bounding boxes around them. Examples include COCO, Objects365, and VOC.
- ➢ **Instance Segmentation Datasets**: Instance segmentation goes a step further by not only detecting objects but also segmenting each object at the pixel level. COCO is a prominent dataset for instance segmentation.

- ➤ **Pose Estimation Datasets**: Pose estimation involves determining the pose of objects relative to the camera or the world coordinate system. COCO and Tiger-pose are examples of datasets for this task.
- ➤ **Classification Datasets**: Image classification tasks involve categorizing images into predefined classes. Datasets like CIFAR-10, ImageNet, and MNIST are widely used for image classification.
- ➤ **Oriented Bounding Boxes (OBB) Datasets**: OBB is a technique for detecting angled objects in images using rotated bounding boxes. DOTAv2 is a popular dataset for this purpose.
- ➤ **Multi-Object Tracking Datasets**: Multi-object tracking involves detecting and tracking multiple objects over time in video sequences. Datasets like Argoverse and VisDrone support multi-object tracking tasks.

4.2 Data Understanding for Effective Object Detection

The success of any object detection model is inherently tied to the quality and comprehensiveness of its dataset. A thorough understanding of this dataset is paramount, and here's a step-by-step elucidation:

- ➤ **Data Collection**: This is the cornerstone of our dataset preparation. The goal was to source images that capture the breadth and variability of real-world scenarios. While repositories like COCO, ImageNet, and Pascal VOC offer a vast array of labeled datasets, niche requirements might demand bespoke collections. When embarking on custom data collection, it's essential to ensure diversity in terms of object orientations, lighting conditions, object scales, and varied backgrounds to foster model generalization.
- ➤ **Annotation**: Once our data pool is established, the critical task of annotation commences. This entails demarcating objects within each image and categorizing them. Annotations serve as the 'ground truth', stipulating not just the presence of an object, but also its precise location, often represented as bounding boxes, and its class or category.
- ➤ **Data Organization**: To streamline the model training process, the dataset needs a structured organization. A common practice is to bifurcate data into distinct directories for training and validation. Within these primary divisions, there should be dedicated subdirectories for images and their corresponding annotations. This hierarchical arrangement aids in seamless data access during model training and validation.
- ➤ **Data Format**: Different object detection frameworks have varied annotation format preferences. It's crucial to transform our annotations to align with the chosen framework. A popular choice among many is the YOLO format, which encapsulates details like class IDs and object coordinates and dimensions.
- ➤ **Data Split**: The final preparatory step is segmenting the dataset. A typical split includes training, validation, and testing sets. Ensuring this division is representative and balanced is crucial. A good split facilitates effective model training, iterative validation during the training process, and an unbiased evaluation of the model's performance post-training.

In sum, a robust and comprehensive dataset, meticulously prepared, lays the foundation for the successful training of object detection models.

4.3 Data Creation

The prowess of an object detection model can often be traced back to the quality of its dataset. The genesis of this lies in dataset creation, with a pivotal component being the annotation of the data. In this segment, we will illuminate the diverse annotation techniques at our disposal and elucidate our rationale for selecting Roboflow as our preferred annotation tool.

Annotation Methods:

- ➤ **Manual Annotation**: This is a hands-on approach where human annotators painstakingly outline objects in images using bounding boxes, subsequently labeling them. Its hallmark is accuracy, but this method's labor-intensive nature can render it suboptimal for expansive datasets, not to mention the susceptibility to human errors.

- ➢ **Semi-Automatic Annotation**: This method amalgamates human intuition with technological assistance. An annotator might initiate a bounding box, only for the tool to suggest potential refinements or label recommendations. This hybrid method expedites the annotation process without entirely sidelining human judgment.
- ➢ **Fully Automatic Annotation**: Here, sophisticated algorithms take the helm. They autonomously discern and annotate objects in images. Their rapidity is commendable, but their precision can fluctuate based on task intricacies and the caliber of the algorithms deployed.

Our Choice - Roboflow:

Our decision to harness Roboflow stemmed from its myriad of compelling features:

- ➢ **User-Centric Design**: Roboflow's interface is emblematic of user-centered design. Its intuitive layout demystifies the annotation process, mitigating the need for extensive training.
- ➢ **Format Versatility**: With support for a plethora of annotation formats like Pascal VOC, COCO, and YOLO, Roboflow ensures compatibility with predominant object detection architectures. This versatility promises a seamless transition from annotation to model training.
- ➢ **Holistic Data Management:** Beyond mere annotation, Roboflow avails tools for holistic data governance, be it organization, versioning, or augmentation. This integrated approach simplifies the continuum from data preparation to model deployment.
- ➢ **Collaborative Ecosystem & Quality Assurance**: The platform is a boon for teamwork. Multiple annotators can annotate concurrently, and the platform facilitates meticulous review mechanisms to guarantee the fidelity of annotations.
- ➢ **Preprocessing & Augmentation Capabilities:** Roboflow is adept at dataset enhancement. Its repertoire includes preprocessing and augmentation techniques, vital for cultivating a diverse and resilient dataset.
- ➢ **Adaptable Export Capabilities**: With export configurations tailored for esteemed deep learning frameworks like YOLO, TensorFlow, and PyTorch, Roboflow ensures your data is primed for your preferred ML ecosystem.
- ➢ **Optimized Workflow**: Efficiency is intrinsic to Roboflow. Tasks such as data uploads, versioning, and exporting are streamlined, culminating in significant time and resource conservation.

4.4 Model selection

Once the dataset was ready we moved ahead with our task and we chose Ultralytics's YOLO model for our task.Let's delve into the reasons for selecting YOLO as the preferred model for the task of smoke and fire detection. For the specialized task of smoke and fire detection in imagery, the choice of model is pivotal. Among the various contenders, YOLO stands out for a multitude of reasons:

- ➢ **Real-time Detection**: YOLO is renowned for its real-time object detection capabilities. In emergency situations, where detecting smoke and fire promptly can be life-saving, the speed of YOLO is a significant advantage.
- ➢ **Unified Architecture**: Unlike some other detectors which apply classifiers to an array of region proposals, YOLO detects and classifies objects in a single pass, making it inherently faster and more streamlined.
- ➢ Accuracy with Overlapping Objects: Fires and smoke plumes can often overlap or be closely intertwined in imagery. YOLO's design reduces false negatives in such scenarios, as it considers the entire image context during detection.
- ➢ **Adaptability**: YOLO can be fine-tuned for specific tasks. With a dataset curated for smoke and fire, YOLO can be adapted to recognize the subtle nuances and variations of smoke and fire patterns, enhancing its detection precision for this task.

- ➢ **Scalability**: The YOLO architecture is scalable. Depending on the deployment scenario (e.g., real-time surveillance cameras vs. post-event analysis), one can choose between different versions of YOLO (like YOLOv3 or YOLOv4) to strike the right balance between speed and accuracy.
- ➢ **Rich Feature Extraction**: YOLO's deep architecture enables it to extract intricate features, which is crucial when differentiating between smoke, fire, and other similar-looking phenomena in varied lighting conditions and backgrounds.

In summary, YOLO's unique combination of speed, accuracy, and adaptability renders it an ideal choice for the time-sensitive and crucial task of smoke and fire detection. Let's see the results we obtained from our model choice.
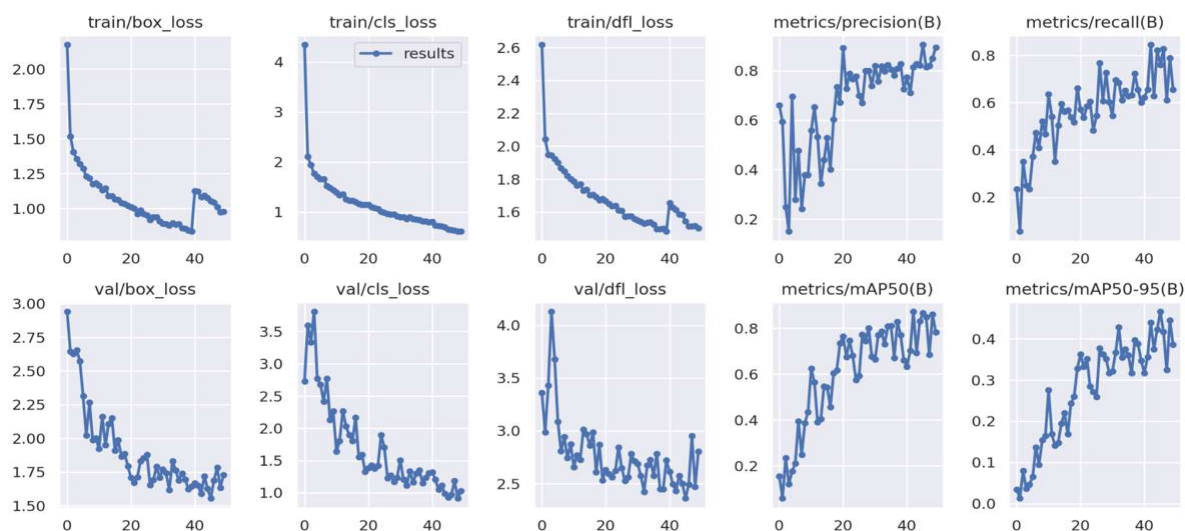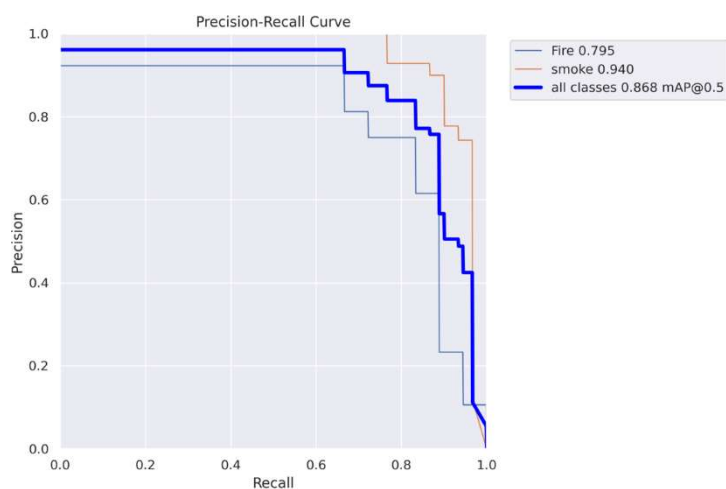

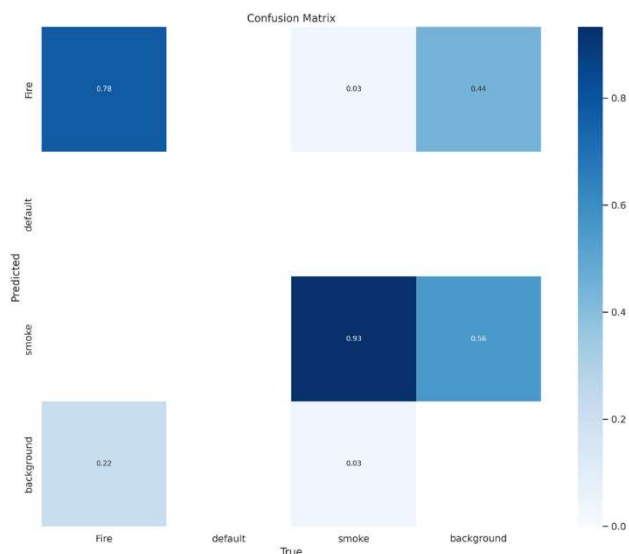Fig.1; Results of model performance

Fig.2: Precision-Recall curve

Fig.3: Confusion Matrix

4.5 Key finding (Model Performance)

Let's analyze the provided model performance metrics in the context of the precision-recall curve for object detection:

- ➢ The Precision-Recall curve visualizes the trade-off between precision (how many detected items are relevant) and recall (how many relevant items are detected).
- ➢ A higher mAP indicates that the area under the Precision-Recall curve is larger, which is desirable. This means the model maintains high precision while achieving high recall, and vice versa.

Performance Analysis Based on mAP@0.5:

Fire (0.795 mAP@0.5):

16

- The model's Mean Average Precision for detecting fire is 0.795 or 79.5%. This suggests that when trying to detect fire instances, the model has an average precision of 79.5% when the predicted bounding boxes have at least a 50% overlap (IoU) with the ground truth.
- While this is a high value, indicating good accuracy, it is slightly lower than the detection rate for smoke, which may imply that there are some challenges or ambiguities in detecting fire compared to smoke in the dataset.

Smoke (0.94 mAP@0.5):

- For smoke detection, the model boasts a 0.94 or 94% mAP. This is notably high, indicating that the model is exceptionally precise in detecting smoke instances, with an IoU threshold of 0.5.
- The higher precision for smoke suggests that the model has been very effective in distinguishing smoke from other objects or phenomena in the images.

All Classes (0.868 mAP@0.5):

This metric aggregates the performance over both the classes (fire and smoke). A score of 0.868 or 86.8% indicates that, on average, the model has an 86.8% precision rate for detecting both fire and smoke when the bounding boxes overlap by at least 50% with the ground truths.

This combined metric provides a holistic view of the model's performance across all categories.

# 5. Wildfire Impact Visualization in Greece

5.1 Data Sources

The data for Area Burned by Wildfires (2013–2023) in Greece and the estimated cost to suppress wildfires per fire in Greece from 2011 to 2018 (in euros) were scraped from Statista. The dataset was stored in Excel, as shown below:

	A	B	C
1	Year	Area Burned by fires in hectares	Estimated Cost to supress wildfire (Euros)
2	2013	19762	1090000
3	2014	14815	835000
4	2015	11613	835000
5	2016	31707	1095000
6	2017	20041	935000
7	2018	12037	1235000
8	2019	10736	
9	2020	14915	
10	2021	130744	
11	2022	22480	
12	2023	52801	

5.2 Data Analysis and Visualization

A. Visualising Area Burned by Wildfires (2013-2023) in Greece
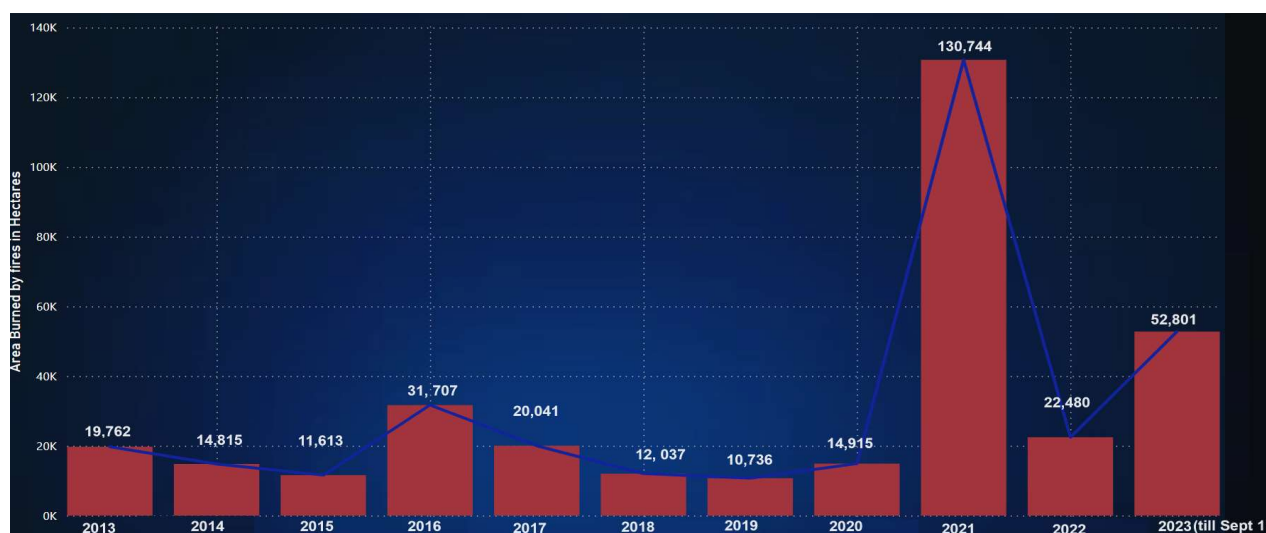
B. Estimated cost to suppress wildfires per fire in Greece from 2011 to 2018 (in euros).
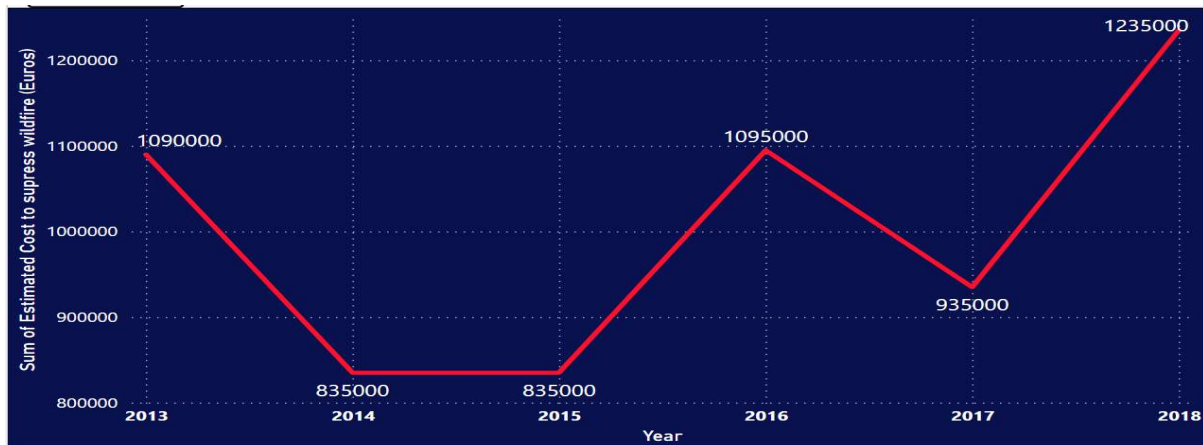
5.3 Key finding

A. Area burned from wildifre

- ➢ **Fluctuating Trends**: The area burned by fires in Greece shows significant year-to-year fluctuations. Some years experience relatively low burn areas, while others see a substantial increase.
- ➢ **Notable Spike in 2021**: The year 2021 stands out with a dramatic increase in the burned area, reaching 130,744 hectares. This suggests a severe wildfire season that year.
- ➢ **Unpredictable Nature**: The data indicates the unpredictable and volatile nature of wildfires, as evidenced by the variation in burned areas from one year to the next.
- ➢ **Recent Increases**: There has been a recent increase in the area burned by wildfires, as evident from the notable figures in 2021, 2022, and 2023.
- ➢ **Mitigation Challenges**: The data highlights the ongoing challenges in wildfire prevention and suppression in Greece, especially given the substantial area burned in recent years.
- ➢ **Need for Preparedness**: These findings underscore the need for improved wildfire preparedness, prevention, and suppression measures in Greece, as well as the importance of investing in strategies to mitigate the impact of wildfires.

B. Estimated cost to supress wildfire.

Increase in Suppression Costs: The data shows a notable increase in the estimated cost to suppress wildfires in Greece over the years. While the costs were relatively lower in 2014 and 2015, there was a substantial increase in subsequent years. This increase in suppression costs suggests that wildfires have become more expensive to manage, potentially due to larger and more severe wildfires, increased resources required for suppression efforts, or changes in firefighting strategies and equipment. It may also indicate the need for additional investments in wildfire prevention and mitigation measures to reduce the economic impact of wildfires.

6 Tools Used

Specific tools that were essential throughout the whole project:

Programming Language	Python
Interactive Development	Jupyter Notebooks & Google Collab
Data Manipulation and Computation	Pandas & Numpy
Data Visualization (Prediction Model)	Matplotlib, Seaborn & Plotly
Machine Learning and Modelling	Scikit-Learn, LightGBM, Optuna and Yolov 8
Data Visualization (Recovery part)	Power BI
Data Collection	Excel

7 Result

Wildfire Preparation

➢ It has been determined that the latitude range of 39 - 39.99 and the longitude range of 22 - 22.99 have the highest wildfire counts, underscoring the need for heightened monitoring and preventative measures in these specific geographic zones.

➢ It is notable that temperature is at its highest in August, coinciding with the highest wildfire counts during this month. This highlights the critical connection between temperature patterns and wildfire risk, emphasizing the need for proactive measures during August to mitigate the potential for wildfire outbreaks.

➢ Recognizing August as a peak wildfire month underscores the importance of heightened preparedness during this period. Authorities can allocate additional resources, manpower, and firefighting equipment in advance.

Response to Wildfires

➢ **High Precision for Fire Detection**: The model's accuracy in identifying fire incidents enables swift response. It can trigger alerts to relevant authorities, leading to the rapid dispatch of response teams to the location of the fire.

➢ **Slightly Lower Fire Detection Precision**: While fire detection precision is slightly lower, it still contributes to early fire identification. The model can assist in distinguishing fire from other objects, facilitating a quicker response.

➢ **High Precision for Smoke Detection**: Exceptional precision in detecting smoke is vital for rapid response. Accurate smoke detection can trigger alerts and help responders assess the situation effectively.

➢ **Balanced Average Precision Across All Classes**: The model's balanced performance across all classes ensures its versatility in handling different object detection tasks during a wildfire response. It can provide real-time decision support to responders, aiding in containment and suppression efforts.

Recovery Strategies from Wildfire

A report on Euronews claims that environmentalists who advocate stronger international action to kerb climate change have accused Greek authorities of spending more funds on extinguishing fires than on prevention.

So, the requirement for recovery is to save money by stopping the wildfire causes, which can result from factors like extreme temperatures or human negligence. By addressing the root causes of wildfires, authorities can save significant amounts of euros, which could be redirected to more effectively manage and mitigate wildfires. These saved funds can be utilized in the following ways to complete the recovery process effectively:

➢ **Invest in Prevention**: Allocate a substantial portion of the saved funds to wildfire prevention measures, including controlled burns, defensible spaces, and public education on fire safety. Prevention is a cost-effective way to reduce the risk of future wildfires.

➢ **Enhance Preparedness**: Strengthen early warning systems, firefighting capabilities, and community preparedness. This ensures that resources are readily available for rapid response and containment when wildfires do occur.

➢ **Ecological Rehabilitation**: Allocate resources to restore and rehabilitate ecosystems that have been affected by wildfires. This supports the recovery of wildlife and helps prevent future catastrophic wildfires.

➢ **Community Support**: Offer financial aid and support to individuals and communities affected by wildfires. This helps residents rebuild their lives and properties and accelerates overall recovery.

- ➢ **Research and Technology**: Invest in research and technology to develop advanced wildfire detection and suppression tools, which can be more cost-effective in the long run.
- ➢ **Education and Awareness**: Promote public awareness and education on responsible land management and fire prevention. This empowers communities to actively contribute to wildfire prevention efforts.
- ➢ **Long-Term Resilience**: Develop long-term resilience plans and policies that address the increasing challenges posed by climate change and the risk of future wildfires.

By taking these steps, authorities can not only recover from the devastating impact of wildfires but also create a more resilient, prepared, and sustainable environment for the future."

8 Challenges

The team encountered numerous obstacles during the completion of this endeavour. Several of the primary obstacles are listed.

Prediction Models

- ➢ **Finding temperature dataset**: Acquiring a comprehensive temperature dataset for Greece was challenging, as it necessitated data with specific attributes like latitude, longitude, date, and temperature, which were not readily available for the entire country, and not provided upon request.
- ➢ **Understanding Wildfire data**: Deciphering the multifaceted wildfire dataset proved demanding, as it encompassed various columns related to a crisis scenario unfamiliar to the team. Substantial research and comprehension were required to make sense of this data.
- ➢ **Data Integration**: Merging the wildfire and temperature datasets presented its own set of challenges. This task required reconciling latitude and longitude information and ensuring the absence of null or invalid values in the merged dataset. To address this, the decision was made to round latitude and longitude to one decimal point, which introduced discrepancies.
- ➢ **Imbalanced Class Distribution**: The final model encountered the issue of an imbalanced distribution of the fire class, which resulted in remarkably high AUC ROC scores.

Smoke and fire detection

- ➢ **Data Annotation**: The meticulous tagging of objects in images or videos for accurate data annotation posed a considerable challenge. This required precise identification and classification of objects within the visual data.
- ➢ **Computation Time**: The algorithms used for smoke and fire detection exhibited lengthy execution times, which added to the complexity of the project.
- ➢ **Software Compatibility**: Conflicts arising from differences in package versions posed compatibility issues and required resolution.

Biggest Challenge – Time Constraints:
The most formidable challenge faced throughout the project was the limitation of time. The crisis solution chosen for the endeavor required a timeline of over six months, while the team had only three months at its disposal. Time management and project completion were critical challenges in this context.

9 Conclusion

This comprehensive report on wildfire prediction, detection, and response in Greece reveals crucial insights, including the identification of geographic hotspots with the highest wildfire counts and a notable emphasis on the significance of the month of August in wildfire occurrence. The YOLO model's exceptional performance in detecting both smoke and fire, along with the rise in wildfire suppression costs, underscores the need for improved prevention strategies and preparedness. The report also emphasizes the importance of recovery strategies, such as fund allocation, research, education, and long-term resilience planning. Despite challenges, the project offers valuable recommendations for enhancing wildfire management, protecting communities, and mitigating the impact of these devastating events.

10 Recommendation and Future Scope

Prediction model

- ➤ While the project prioritized the application part due to time constraints, the focus is now shifting towards further enhancements, including the utilization of additional datasets, specifically for features like relative humidity, wind speed, and flora. To strengthen the predictive capabilities of the model, the next steps involve conducting in-depth statistical analysis and exploratory data analysis on these datasets, allowing for more comprehensive and accurate wildfire prediction.
- ➤ Extracting features suitable for time-series analysis, including metrics such as autocorrelation and seasonality, to gain deeper insights into wildfire occurrences.
- ➤ Shifting from location-based detections to generating features that capture areal averages, potentially providing a broader perspective for analysis.
- ➤ Implementing advanced hyper-parameter optimization techniques, such as Optuna or Hyperopt, to fine-tune the model's performance and enhance predictive accuracy.
- ➤ Exploring the potential of different machine learning models beyond LightGBM, including CatBoost, XGBoost, neural networks, and ensemble methods, to determine which one offers the best performance for wildfire prediction and detection.

Smoke Detection and Fire detection

(i). Real-time Deployment on Surveillance Cameras:
- ➤ **Integration with Emergency Systems**: Real-time cameras equipped with the YOLO model can be seamlessly integrated with emergency response systems. When a fire or smoke is detected, the system can immediately alert local fire departments, reducing response times significantly.
- ➤ **Low Latency Processing**: For real-time applications, it's imperative that the model processes video frames swiftly. Future endeavors should focus on optimizing the model for even lower latency, ensuring immediate detection and action.
- ➤ **Multi-camera Coordination**: In large areas like forests multiple cameras can work in coordination. If one camera detects smoke or fire, it can communicate with nearby cameras to triangulate the source and determine the extent of the fire.

(ii). Enhanced Data Collection:
- ➤ **Varied Conditions**: For more robust detection, the model should be trained with data from various lighting conditions, seasons, and environments. For instance, a forest fire's appearance can differ drastically from a house fire.
- ➤ **Synthetic Data Generation**: With the advent of Generative Adversarial Networks (GANs), synthetic data can be created to supplement real data, providing varied scenarios that might be hard to capture otherwise.

(iii). Model Improvements:
- ➤ **YOLO Evolution**: With the continued development of YOLO versions, like the emergence of YOLOv8 or further future iterations, it's prudent to keep the model updated for enhanced accuracy and speed.
- ➤ **Transfer Learning**: Implement transfer learning from related tasks, like general object detection, to improve the accuracy and speed of smoke and fire detection.

(iv). Additional Feature Integration:
- ➤ **Temperature Sensing**: Integrate thermal cameras to detect abnormal heat sources even before visible smoke or flames appear.
- ➤ **Sound Detection**: In many instances, the crackling sound of a fire can be an early indicator. Integrating sound sensors can provide an additional layer of verification.

(v). Continuous Feedback Loop:
- ➤ **Adaptive Learning**: The system can be designed to learn continuously from false positives and false negatives. When an anomaly is flagged by human operators, this data can be fed back into the system, refining its detection capabilities.

(vi). Public Awareness and Integration:
- • **Mobile App Integration**: Provide real-time updates to residents in vulnerable areas through mobile apps, ensuring they are promptly alerted to nearby fire or smoke threats.
- • **Community-based Reporting:** Enable a feature for the community to report fire or smoke, which can be cross-verified with the model's detections. This creates a two-way validation system.

In conclusion, while the current model demonstrates promising results, the above recommendations can further elevate its applicability and efficiency. Implementing these suggestions will not only enhance the detection capabilities but also ensure the safety and well-being of communities and the environment.

11 References

- ➤ Alves, B. (2023) Greece: Wildfire suppression costs, Statista. Available at: https://www.statista.com/statistics/1296865/expenditure-to-suppress-wildfires-in-greece/
- ➤ Salas, E.B. (2023) Greece: Wildfire area burned 2023, Statista. Available at: https://www.statista.com/statistics/1264709/area-burned-by-wildfire-in-greece/
- ➤ Firms (no date) NASA. Available at: https://firms.modaps.eosdis.nasa.gov/download/create.php
- ➤ Data Overview (2023) Berkeley Earth. Available at: https://berkeleyearth.org/data/
- ➤ Home european climate assessment & dataset (no date) Home European Climate Assessment & Dataset. Available at: https://www.ecad.eu/
- ➤ *Greece climate Greece Climate Zone, Weather By Month and Historical Data*. Available at: https://weatherandclimate.com/greece
- ➤ *Central Greece, Greece climate Central Greece, GR Climate Zone, Monthly Weather Averages and Historical Data*. Available at: https://weatherandclimate.com/greece/central-greece
- ➤ *Greece is launching drones to tackle future wildfires euronews*. Available at: https://www.euronews.com/green/2023/09/01/greece-turns-to-tech-to-tackle-wildfires-as-largest-ever-blaze-in-eu-continues-to-burn
- ➤ *The four phases of emergency management*. Available at: https://training.fema.gov/emiweb/downloads/is10_unit3.doc
- ➤ *Fire object detection dataset by Custom Roboflow*. Available at: https://universe.roboflow.com/custom-thxhn/fire-wrpgm

12. Individual Exploration

In this report, I will provide an overview of my individual contributions in overall project. The purpose of this report is to assess my performance throughout the project, and to identify lessons learned for future endeavors.

A. Happening
We decided to have at least one meeting each week either in person or online and chose WhatsApp for the regular discussion of the actions and happenings
Also, for sharing the work files on Microsoft Teams and online metting on Zoom.
Although, all team members contributed equally to the project, some members worked more on specific topics to fulfill their learnings goals for this subject.
Following is the topic each team member wanted to work more individually
- ➤ **Ahmed-** Prediction model (temperature part), Recovery part, Notion Pages update and Report writing.
- ➤ **Shivani -** Prediction model (wildfire part), Recovery part, Notion Pages update and Presentation slides.
- ➤ **Nutan -** Smoke and fire detection.
- ➤ **Ashutosh–** Smoke and fire detection.

Since me and Shivani were working on same parts, so we decided to meet in UTS library one day every weekend because we live near campus.

B. My Contribution to the project

Exploration 1: Prediction Model

➤ **Temperature Processing:**
In this phase of the project, I focused on temperature processing. This included acquiring temperature data relevant to Greece. The challenge here was to find a comprehensive dataset that included crucial attributes like latitude, longitude, date, and temperature.

➤ **Temperature Scraping:**
I delved into web scraping techniques to extract temperature data from various sources. This process involved navigating website restrictions and understanding the specific text data format that would be compatible with our analysis tools.

➤ **Data Creation:**
Once the temperature data was successfully collected, the next step was data creation. I organized, cleaned, and formatted the data to prepare it for analysis. This involved addressing null or invalid values and ensuring that the dataset was ready for further exploration.

➤ **Data Analysis of Temperature**
I conducted a thorough data analysis to understand temperature trends in Greece. This included visualizations and statistical assessments to uncover insights related to temperature fluctuations over time. These insights were crucial for understanding the environmental factors contributing to wildfire occurrences.

➤ **Model Tuning (Me and Shivani)**
For the predictive model, I employed the Optuna hyperparameter optimization framework. My primary objective was to enhance the performance of the model, specifically aiming to maximize the ROC-AUC score.

➤ **Model Training (Me and Shivani)**
Using the tuned hyperparameters, I trained the predictive model with the temperature and wildfire datasets. I set a fixed number of training rounds to ensure model convergence.

Exploration 2: Recovery Visualization

➤ **Visualizing Area Burned in Greece:**
In the recovery phase, I focused on visualizing the area burned by wildfires in Greece. I obtained the dataset from Statista, which included information from 2013 to 2023. This data enabled me to understand the historical trends in wildfire-affected areas.

➤ **Assessing Greece's Actions:**
I examined the actions taken by Greek authorities in response to wildfires, particularly the allocation of resources for suppression. By evaluating the estimated cost to suppress wildfires from 2011 to 2018, I gained insights into how Greece has addressed the economic impact of wildfires.

These explorations provided critical insights into both the predictive aspect of the project, with a focus on temperature and wildfire prediction, and the recovery phase, where understanding the historical impact of wildfires and response strategies in Greece played a vital role.

Notion Page

In our initial meetings, Shivani and I took on the responsibility of managing the Notion page, recognizing the significance of effective note-taking and staying updated on project developments. Throughout the project's lifecycle, I diligently maintained the Notion page, collaborating with team members to gather and input their valuable insights and updates. Shivani also contributed to this essential task, ensuring that our project records were organized and up-to-date.

Report writing

I took the lead in designing and structuring the complete report. When we encountered the challenge of meeting the page limit, I proactively sought solutions by reaching out to our course coordinator and our project partners. Ensuring that our report adhered to the required structure and constraints was a critical part of my role, as I coordinated efforts to present our project in a clear and concise manner.

Interim Presentation

I made a whole PowerPoint presentation for the mid-interim presentation. I was guided and helped by group member Shivani on some steps.

13. Appendx

All codes, notebook , images, videos are available on the github repo

Link: https://github.com/ahmedkhursheed/PyroVision_Greece_Wildfire