Semester 5
Data Mining Project

Predicting and Preventing Wildfires in Bejaia

Documentation

Team members :
Billal (Student) Chaouche
Amira (Student) Boudaoud
Sarra (Student) Arab
Oumaima (Student) Daif
Hamza (Student) Benzaoui

Introduction:

The Bejaia region in Algeria grapples with a growing threat of wildfires, driven by diverse vegetation and occasional arid conditions. The heightened frequency and intensity of these fires necessitate the development of a predictive model for early detection and intervention. By leveraging historical data and relevant parameters, such a model can provide crucial insights to authorities, enabling proactive measures to mitigate damages and enhance overall preparedness. Given the escalating global risk of wildfires due to climate change, the imperative to develop advanced predictive models in Bejaia aligns with a broader commitment to environmental sustainability and community safety.

- ***First Dataset:***

Initial interpretation and considerations about the columns (features) of the dataset:

Month:

- Considering that the dataset covers months from June to September, this information is crucial. The summer months typically have higher temperatures, lower humidity, and increased fire risk.

Year :

- for the main purpose of the project the year does not seem as a crucial column since the most relevant information is the month, the aim of our model, in more specific way, is predicting fires at any time based on weather and other features regardless of the time the data is collected

Ws (Wind Speed):

- Higher wind speeds can accelerate the spread of fires by carrying embers over longer distances. Wind speed is a critical factor in fire behavior.

Rain:

- Rainfall can have a mitigating effect on fire risk by moistening vegetation. A lack of rain or extended dry periods may increase fire danger.

Fine Fuel Moisture Code (FFMC):

- FFMC is an index from the Fire Weather Index (FWI) system, representing fuel moisture for fine fuels. Low values indicate dry conditions and increased fire risk.

Duff Moisture Code (DMC):

- DMC is another FWI index, measuring moisture for decomposed organic material. Dry conditions in the duff layer can contribute to fire risk.
-

Drought Code (DC):

- : DC is an FWI index indicating long-term drought conditions. High values suggest prolonged dry periods, increasing fire risk.

Initial Spread Index (ISI):

- ISI is an FWI index related to the potential for fire spread. It considers wind, fine fuel moisture, and rate of fire spread.

Buildup Index (BUI):

- BUI is an FWI index representing the total fuel available for combustion. High values indicate an increased potential for intense and difficult-to-control fires

Fire Weather Index (FWI):

- The overall FWI is a composite index based on the previously mentioned FWI components. It provides a comprehensive measure of fire danger.

initial analysis :

Highly Relevant Features:

<u>Month and Temperature: </u>These are likely to be strong predictors, as fire risk often increases in warmer months.

Relative Humidity: Low humidity levels are associated with higher fire risk.
Wind Speed: Higher wind speeds can contribute to the spread of fires.
FFMC, DMC, DC, ISI, BUI, FWI: These FWI indices provide specialized information about fuel moisture, fire spread potential, and overall fire danger.
Potentially Relevant Features:
Rainfall: While generally beneficial for mitigating fire risk, its impact may depend on the region and local fire dynamics.

May Consider Dropping:

Day of the Week: This might not have a significant impact, as patterns could vary widely.
Year: If there is no long-term trend in fire occurrence, the year may not be highly informative.

- These assumptions will be validated by exploring the data, checking correlations, and using feature importance techniques within our machine learning model, with taking in consideration the limited domain knowledge for sure.

About Second Dataset :

Opting for an expanded dataset arose from the identification of potential overfitting issues in the initial dataset. To ensure the generation of authentic data, we maintained our focus on the Bejaia fire prediction task, which involves binary classification. It's worth noting that while the second dataset aligns with the same task, there are slight variations in the columns used. These columns, along with concise explanations for each, are outlined below:

Date: Represents the date in the format yyyy-MM-dd, indicating the day of the observation.
Temperature: Indicates the temperature in Celsius, providing information about the atmospheric heat.
Rain: Represents the amount of rainfall in millimeters, measuring the precipitation during a specific period.
Wd (Wind Direction): Specifies the wind direction in degrees (0 to 360), indicating the compass direction from which the wind is blowing.

Ws (Wind Speed): Measures the wind speed in kilometers per hour, indicating how fast the wind is moving.
Pres (Air Pressure): Represents the atmospheric pressure in hectopascals (hPa), providing insights into the weight of the air above.
RH (Relative Humidity): Indicates the relative humidity in percentage, representing the amount of moisture in the air relative to the maximum it could hold at the same temperature.

Dew Point Max: Represents the maximum dew point in Celsius, indicating the temperature at which air becomes saturated with moisture.
Dew Point Avg: Represents the average dew point in Celsius, providing a measure of atmospheric moisture.
Dew Point Min: Represents the minimum dew point in Celsius, indicating the lowest temperature at which the air can hold its moisture.
Classes: Binary classification with values 1 and 0, where 1 represents 'fire' and 0 represents 'not fire.' This column is the target variable for classification tasks, indicating the presence or absence of a fire based on the other recorded environmental parameters

Initial interpretation and analysis :

Date:
-Potential Use:Consider temporal patterns for fire occurrences. Seasonal trends or specific days might influence fire likelihood.

Temperature:
- Potential Use: Higher temperatures could contribute to fire risk. Correlation analysis is crucial to assess its impact.

Rain:
- Potential Use:Low rainfall might increase fire risk. Inverse correlation with fire presence should be explored.

WD (Wind Direction):
- Potential Use: Wind direction influences fire spread. Wind from certain directions might escalate or mitigate fire incidents.

Ws (Wind Speed):
- Potential Use:Higher wind speeds can accelerate fire spread. Evaluate its correlation with fire occurrences.

Pres (Air Pressure):
- Potential Use: Air pressure may indirectly impact fire behavior. Explore its relationship with fire incidents.

RH (Relative Humidity):
- *Potential Use:* Low humidity levels may increase fire risk. Inverse correlation with fire presence is expected.

Dew Point Max/Avg/Min:
- Potential Use: Dew point reflects atmospheric moisture. Higher values might indicate less fire risk, but correlation analysis is essential.

Considerations:
- Conduct exploratory data analysis (EDA) to understand distributions, correlations, and potential outliers.
- Feature engineering may involve creating derived features or transforming existing ones to enhance predictive power.
- Machine learning models can then be trained and tuned based on the identified influential features.

- **Remark** : The initial interpretation provides a general overview. Detailed analysis and modeling will refine feature selection and enhance prediction accuracy.

A comparison between the first and second dataset:

The second dataset, despite its larger size, yielded insights through model interpretations and data exploration. The discernment drawn is that the columns utilized in the second dataset lack robust characteristics for effective fire prediction (for example : using Dew Point Max/Avg/Min instead of Fire Weather Indexes (FWI) system). While exhibiting reasonable results, the columns from the first dataset demonstrated superior performance due to the robustness of the criteria considered in the prediction process. This disparity in accuracy underscores the varying influence of each column on the prediction outcome. Future refinement of the model necessitates acquiring additional information and deepening domain knowledge to enhance the features utilized, thereby improving overall accuracy.

Final conclusion and the impact of the model on solving the problematic :

The deployment of predictive models for forecasting wildfires in the Bejaia region of Algeria carries significant real-world implications. The models, particularly Logistic Regression and KNN, showcase the potential to revolutionize fire prevention and response strategies. With high accuracy, precision, and recall, these models offer a reliable tool for early detection, allowing authorities to take swift actions in mitigating fire risks. The positive impact extends to minimizing property damage, safeguarding lives, and optimizing resource allocation. Moreover, the use of such models enables a proactive approach to firefighting efforts, enabling authorities to respond promptly to identified high-risk areas. Beyond immediate applications, the development and enhancement of these models, incorporating advanced techniques like deep learning, hold promise for further refinement. Deep learning can potentially uncover intricate patterns in the data, enhancing the models' predictive capabilities. Continuous refinement and validation against new data can contribute to building a robust and adaptive system, reinforcing its reliability in dynamic real-world scenarios. Overall, the integration of advanced predictive models not only offers a practical solution for wildfire prediction but also signifies a proactive and technologically-driven approach to address environmental challenges, showcasing the positive advantages of ongoing development and enhancement in predictive modeling techniques.