Finding Reasons of Forest Fire in Algeria: Utilising of Random Forest

Volkan Mazlum
Çankaya University
Department of Computer Engineering

***Abstract* —One of the major problems is the forest fires in our world. This is one of the most serious issues for Algeria that is desert country. Solving forest fires is one of the main objectives of scientists. Our goal in this study is to find out what wildfires are connected to and to find solutions from the data available. Basically, datasets contain data such as rain, humidity, FFMC, and DC. These final data are based on the Random Forest (tree) method, which is one of the decision tree methods. With this method, it is intended to be able to find out exactly what forest fires are associated with. The accuracy for the region of Sidi-Bellabas is 97.54%, and the accuracy for the region of Béjaïa is 93.44%. The possibility of fire depends on the temperature, heat, FFMC, DC, and DMC for the Béjaïa region. It has been seen to depend on RH values, and for the Sidi-BelAbbes region, it has been discovered that only loyalty to FFMC value has been made. This is thought to be since the Sidi-BelAbbes region is warmer and drier than the Béjaïa region. In the data set where both sets are located, the possibility of fire is clearly observed, regardless of region-based, depending on FFMC, FWI, and temperature. At the end of the study, there is also a comparison with the accuracy rates of different methods.**

Key words—Forest Fire, Random Forest, Fire Prediction

1. Introduction

Forest fires are one of the biggest threats to every country in the world. Every year, millions of hectares of land are burned for different reasons. Algeria is one of the countries in the world, most affected by the forest fires. According to [1], more than 320409 hectares of forest in Algeria were destroyed by fires between 2008 and 2017. Also, according to this article [2], Algeria is the fourth most affected country by forest fires among the countries mentioned in the article. By looking at these data, it can be seen how devastating forest fires are.

Considering the increasing forest fires in recent years, it is desired by everyone to reveal the causes of forest fires and to prevent forest fires before they occur. In addition, new methods and methods that will solve them are gaining importance. Camera surveillance systems, helicopters, and similar methods are used by most countries. Research continues for more efficient ones.

The data from two Algerian regions with slightly different climates were examined in this article. I tried to find out what caused the forest fires in both regions. The Random Forest algorithm was used for these operations. To explain the random forest algorithm, this algorithm is used in both classification and regression problems like the decision tree. The working logic is to create more than one decision tree and take the average value of these decision trees when they will produce a result.

The following parameters are available from the dataset: day, month, year, temperature, RH, Ws, rain, FFMC, DMC, DC, ISI, BUI, FWI, Classes. Meteorologically, there are three basic features that affect fires: rain, humidity, and wind speed. The Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), Buildup Index (BUI), and FWI are the main three features used. There are three codes produced: Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), and Drought Code (DC). Figure-1 shows which parameters these data consist of.

These are the 6 key elements that make up the Canadian Forest Fire Weather Index (FWI) System. The Fine Fuel Moisture Code (FFMC) shows the dryness of the smallest forest fuels (surface litter, leaves, needles, twigs, etc.). DMC represents the dryness of medium-sized surface fuels and upland duff layers (approximately 2 to 10 cm). DC shows the dryness (approximately 10+ cm) of the largest surface fuels and deep duff layers.

An ISI is a score that associated with the spread of fire velocity, while BUI is a relative measure of the amount of fuel available for combustion. It is derived from DC and DMC. The FWI index is an indicator of fire density and combines the two previous components. Higher values indicate more severe combustion conditions, although different scales are used for each of the FWI elements. More detailed data are shown in Figure-2.

Hazard Rating	**FFMC** Fine Fuel Moisture Code	**DMC** Duff Moisture Code	**DC** Drought Code	**ISI** Initial Spread Index	**BUI** Build Up Index	**FWI** Fire Weather Index	**HFI** Head Fire Intensity
Low	0-76	0-21	0-79	<1.5	0-24	0-4	1-2
Moderate	77-84	22-27	80-189	1.5-4.0	25-40	5-10	3
High	85-88	28-40	190-299	4.1-8.0	41-60	11-18	4
Very High	89-91	41-60	300-424	8.1-15.0	61-89	19-29	5
Extreme	92+	61+	425+	>15.0	90+	30+	6

Fig.1 Canadian Forest Fire Weather Index (FWI) System

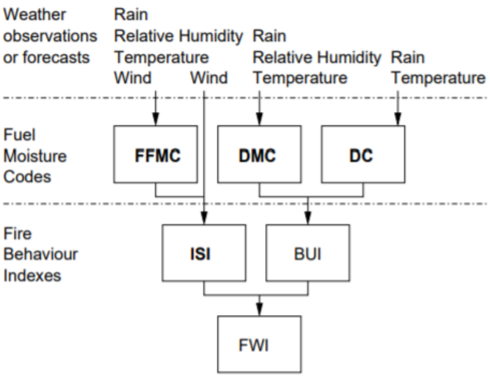

Fig.2 Different scales for FWI [4]

Climate data for Béjaïa [hide]													
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Record high °C (°F)	27.7 (81.9)	32.0 (89.6)	37.0 (98.6)	33.0 (91.4)	37.3 (99.1)	42.8 (109.0)	44.7 (112.5)	47.6 (117.7)	42.5 (108.5)	40.0 (104.0)	37.4 (99.3)	33.0 (91.4)	47.6 (117.7)
Average high °C (°F)	16.4 (61.5)	16.8 (62.2)	17.7 (63.9)	19.3 (66.7)	22.0 (71.6)	25.3 (77.5)	28.7 (83.7)	29.3 (84.7)	27.8 (82.0)	24.3 (75.7)	20.3 (68.5)	16.9 (62.4)	22.1 (71.7)
Daily mean °C (°F)	12.1 (53.8)	12.3 (54.1)	13.1 (55.6)	14.7 (58.5)	17.6 (63.7)	21.0 (69.8)	24.0 (75.2)	24.8 (76.6)	23.2 (73.8)	19.7 (67.5)	15.8 (60.4)	12.7 (54.9)	17.6 (63.7)
Average low °C (°F)	7.7 (45.9)	7.6 (45.7)	8.5 (47.3)	10.1 (50.2)	13.1 (55.6)	16.6 (61.9)	19.3 (66.7)	20.2 (68.4)	18.5 (65.3)	15.0 (59.0)	11.2 (52.2)	8.4 (47.1)	13.0 (55.4)
Record low °C (°F)	−1.0 (30.2)	−4.0 (24.8)	−0.1 (31.8)	2.0 (35.6)	5.8 (42.4)	7.8 (46.0)	13.0 (55.4)	11.0 (51.8)	11.0 (51.8)	8.0 (46.4)	1.6 (34.9)	−2.4 (27.7)	−4.0 (24.8)
Average precipitation mm (inches)	99.7 (3.93)	85.9 (3.38)	100.4 (3.95)	70.7 (2.78)	41.2 (1.62)	16.2 (0.64)	5.8 (0.23)	13.0 (0.51)	40.4 (1.59)	89.5 (3.52)	99.7 (3.93)	135.0 (5.31)	797.5 (31.39)
Average relative humidity (%)	78.5	77.6	77.9	77.9	79.9	76.9	75.0	74.6	76.4	76.3	75.3	76.0	76.9

Fig.3 Climate Data for Béjaïa [5]

Climate data for Sidi Bel Abbès [hide]													
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Average high °C (°F)	14 (57)	15 (59)	18 (64)	20 (68)	24 (75)	29 (84)	34 (93)	35 (95)	30 (86)	24 (75)	18 (64)	14 (57)	22 (71)
Daily mean °C (°F)	8 (46)	9 (48)	11 (51)	13 (55)	16 (60)	21 (69)	25 (77)	25 (77)	21 (69)	17 (62)	12 (53)	8 (46)	15 (59)
Average low °C (°F)	1 (33)	2 (35)	4 (39)	6 (42)	8 (46)	12 (53)	15 (59)	15 (59)	13 (55)	9 (48)	6 (42)	2 (35)	7 (44)
Average precipitation mm (inches)	61 (2.4)	48 (1.9)	46 (1.8)	41 (1.6)	38 (1.5)	10 (0.4)	2.5 (0.1)	5.1 (0.2)	15 (0.6)	38 (1.5)	43 (1.7)	64 (2.5)	410 (16.1)

Fig.4 Climate Data for Sidi Bel Abbes [6]

With these data, we can reveal whether there will be a forest fire using the random forest algorithm. This algorithm was applied after preprocessing for each data set. As a result, 93.44% accuracy occurred for the Béjaïa region. Thanks to the resulting decision tree, it has been seen that forest fire is dependent on temperature, ISI, FFMC, DC, DMC, and RH properties. Considering that the Béjaïa region has a Mediterranean climate, we can see that this result is consistent. An accuracy rate of 97.51% was obtained for the Sidi-BelAbbes region. In contrast to the Béjaïa region, forest fires in the Sidi-BelAbbes region were found to depend only on the FFMC value. The reason for this can be shown as the fact that the Sidi-BelAbbes region has a drier climate than the Béjaïa region. The temperature tables of the regions are in Figure-3 and Figure-4. As a result of this examination, it is revealed that surface litter which is the basis of forest fires in both regions and that these threats should be eliminated. The study was carried out on the abovementioned data. This article is organized as follows: There will be a brief introduction to the purpose of the study in Chapter 3. Comparisons with different studies are presented in Chapter 6. In Chapter 7, the methods and algorithms for carrying out the study are introduced. The results are in Chapter 8.

2. Relations between Nominal Data and Other Data

We'll look at the link between numerical and nominal data in this part, as the title indicates. We'll go through each piece of information one by one before attempting to anticipate the outcome of the random forest technique we'll employ in this part. There is no objective to achieve a certain outcome. This component is just provided to serve as a foundation for future parts. Fire is represented by red, whereas no fire is represented by blue.

First, I'd mention to go over the months 6, 7, 8, and 9 in the data collection. As shown in the table, there is insufficient information to conclude there will be a definite fire when this month arrives. Only in the eighth month did more fires break out than in the preceding months.

When we examine the temperature data, we can observe that beyond 30 degrees, a fire happens as a percentage. Even someone who has never read scientific research may predict this. The relative humidity level is indicated by the RH value. It is common knowledge that when humidity is low, the risk of fire is increased. This may be seen in the table below. The risk of fire is significantly decreased with humidity levels exceeding 45 percent. The wind speed is represented by the WS value. These data do not appear to have a clear relationship.

A rain value of less than 1.68 appears to enhance the likelihood of a fire. We might claim that this is one of the reasons why there are so many fires throughout the summer months. A FFMC value of more than 80 indicates that a fire is imminent. The FFMC value, on the other hand, is made up of rain, humidity, and ws value. It's not a standalone metric. The fact that the FWI value is lower than 3.11 seems to prevent the fire to a large extent. When it is higher than 3.11, it is clearly seen that there is mostly fire. However, this value is not dependent on only 3 values like the FFMC value. It depends on all the others except the values such as day, month, year seen in the table. It does not represent anything by itself, it can be called a generalization of other parameters. Therefore, the presence of FWI indicates the presence of other parameters. They are impossible to ignore. The remaining parameters have a very close connection with the nominal data. The likelihood of fire grows after a particular pace. Indeed, our investigation reveals that the FFMC value is critical, and that humidity, rain, and ws values have an indirect impact on it. In addition, there is a heat map figure-6, which shows the relationship of each parameter with each other. Generally, the correlation of the data in Figure-2 with each other was higher than the others. These tables were made with the discretization method, which is one of the preprocessing methods.

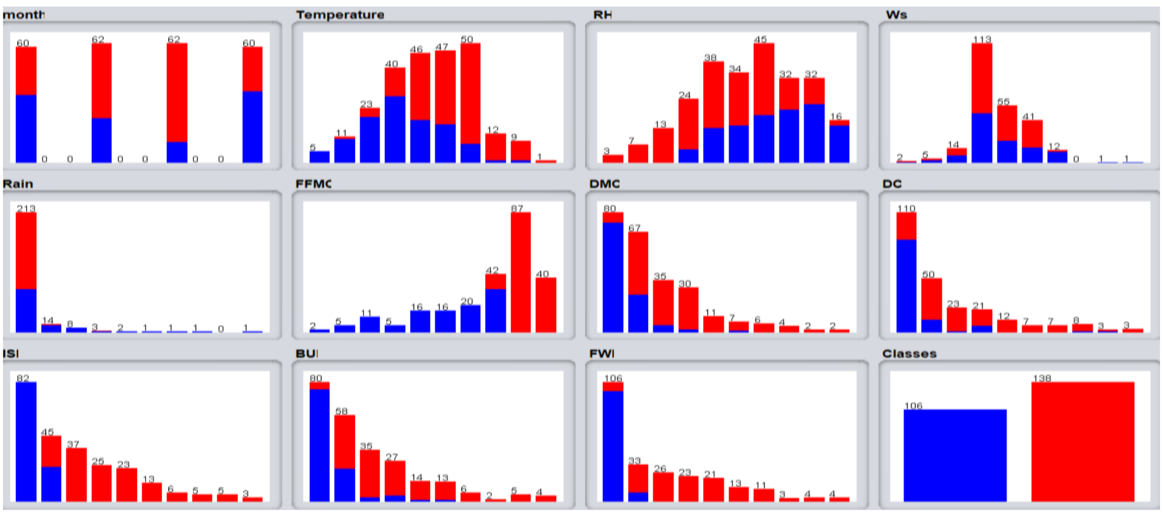

Fig.5 Relations between Nominal Data and Other Data

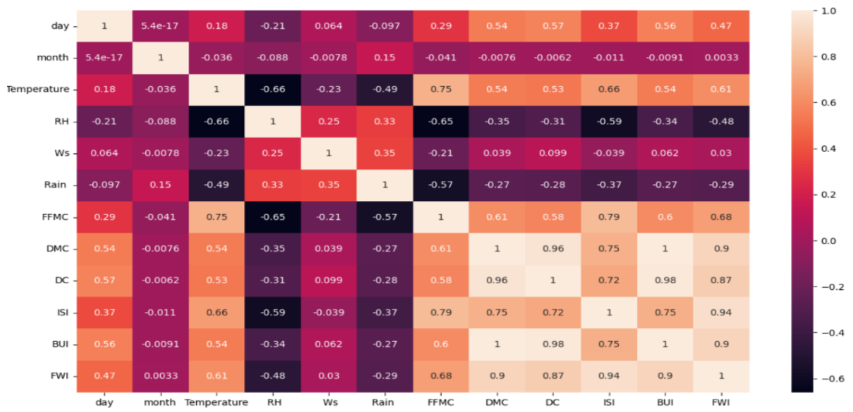

Fig.6 Heatmap with Correlation

3. Purpose of Paper

The results of this study, which seeks a solution to one of the world's biggest problems, are very valuable. Each relationship found between the data reveal a new area that needs to be examined. In this study, FFMC and FWI values stand out. If the FFMC value is higher than 80 and the FWI value is more than 6%, it is revealed that it prepares the ground for the fire. Of course, it is not possible to generalize according to every region at the moment, but this conclusion can be reached in future studies by using the methods described in this article. Relationships can be revealed with even more attributes. For this reason, I believe that our work will have a great role in preventing forest fires.

4. Literature Review

First, I want to start with the methods used in others articles. In the article [7], multiple algorithms such as logistic regression, SVM, KNN, and random forest were applied to forest fire data. As a result, it was decided that the logistic regression method is the most suitable method for the data set containing binary data. A high accuracy rate of 68% has been achieved. The Random Forest algorithm I used showed an average performance. However, before producing this result, Principal Component Analysis (PCA), one of the normalization techniques, was used. If the same algorithms were run on the dataset without using this, gradient boosting would have shown the best result, with 68% accuracy.

In the article [8], a forest fire prediction was made. Weather, location and time, and historical data are partially similar to our data set. A fuzzy inference system is used as a classifier. A 75% accuracy rate has been achieved. A monitoring-based study was conducted using Wirelles sensors.

There are also those who predict forest fires by using satellite-based monitoring instead of using wired sensors. Mallinis and Koutsias [9] aimed to make a map of the burned areas. They used Landsat Thematic Mapper (TM) imagery data. They achieved a very high accuracy of 93% with the SVM algorithm.

Another article using SVM is by Cortez [10]. After collecting meteorological data in northeastern Portugal, a solution based on a support vector machine (SVM) classifier was tried to be produced. The dataset basically includes direct meteorological inputs such as temperature, rain, relative humidity, and wind speed. It has been concluded that the SVM- based solution does a good job of detecting fires that have occurred in small areas but can only make limited predictions for large fires.

Volkan Sevinc [11] studied up to 3231 forest fires near Mugla. A Bayesian Network Model is used to find possible causes. Forest fires in the region were recorded over a 10-year period from 2008 to 2018. Data such as temperature, humidity, wind speed, amount of burned area, distance to agricultural areas, and number of tree species were recorded. The Bayesian Network Model used to predict the possible causes of forest fire said that the biggest causes are hunting, picnics, stubble burning, and shepherd's fire, respectively. Their accuracy rates are as follows: 0.9, 0.89, 0.89, 0.82.

As can be seen, different methods were used in each article, and different results were obtained. Everyone has found the correct result for himself and put it forward. We will perform a study on our own dataset using the Random Forest algorithm. In the Comparison with Existing Works section, we will compare the results with studies that used different algorithms. We will work on an old dataset. We do not have a system that receives data instantly. Our aim is to find the main factor that causes forest fires in the most accurate way.

5. Methodology

In this section, we will examine how wildfire predictions are made using data science techniques and how we arrive at the conclusion. First, we will look at what the decision tree technique is before we move on to the method we use. Later, we will examine the random forest technique. In the Findings section, we will compare some techniques with the Random Forest technique, but these techniques will not be introduced. Only their efficiency and accuracy will be compared.

5.1. Description of Decision Tree

Any business may utilize a decision tree [12] to comprehend objectives and gains for a given purpose. A supervised learning algorithm is the decision tree. The decision tree's learning process may be used to create classification and regression trees. It can also be described as very functioning.

5.2. Description of Random Forest

Random Forest [13] is a classification and regression approach that may be used to decision trees. Depending on the settings, the working logic generates numerous decision trees. By averaging the decision trees produced at each stage of the process, a tree with common parameters is constructed. Figure-7 is a random forest (tree) example for the Béjaïa region to demonstrate how to use the random forest (tree) technique with the data supplied.

In decision trees, gini index values are calculated for each parameter separately. According to their abundance and scarcity, a nominal value is assigned to the desired value as the initial result. Thanks to these, small small decision trees emerge. The general files created by taking the averages of these are called random forest algorithms. In addition, entropy values are also used. You can see the maths behind the decision tree in the figures below.

$$Gini\,index = 1 - \sum_{i=1}^{n} p_i^2 \quad Entropy = -\sum_{i=1}^{n} p_i * \log(p_i)$$

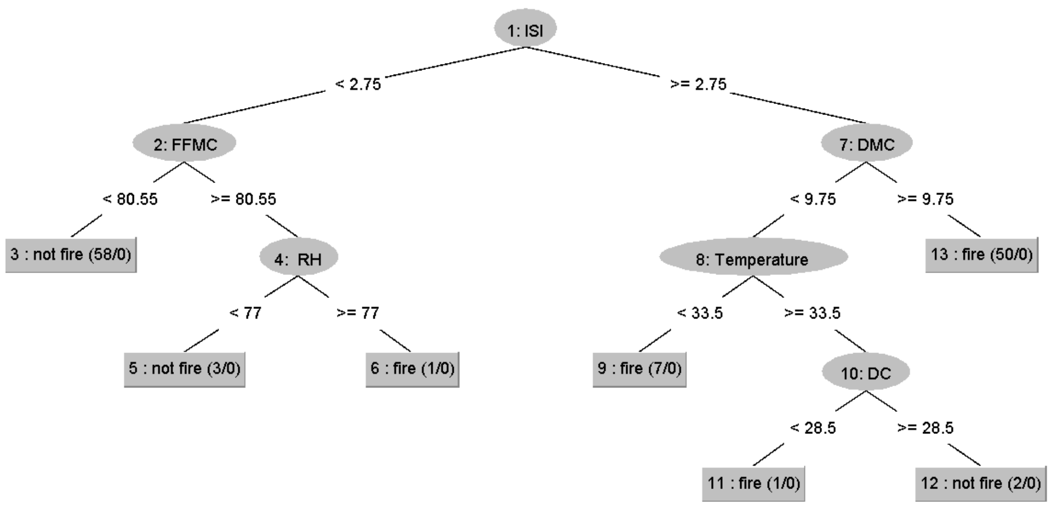

Fig.7 Decision Tree of Béjaïa Region in Algeria

6. Findings

6.1. Description of Dataset

Data sets from Sidi Belabbas in the northwest of Algeria and the Béjaïa area in the northeast of Algeria were utilized in this investigation. Each area has its own set of 122 records. In addition, data for the three months of June to August 2012 were gathered. The data for 2012 was chosen since it was the year in which the most forest fires were documented between 2007 and 2018. In the data set, numerical parameters such as FFMC, rain, DC, and temperature were utilized to construct two predictions: "fire" and "no fire." [14]. The random forest technique we utilized was likewise based on these nominal data. Following figure shows a description of a portion of the dataset.

Attributes name	Description	Values interval Sidi-Bel-Abbes region	Values interval Bejaia region	Values intervals The two regions
Temperature	Temperature in Celsius degrees	24 to 42	22 to 37	22 to 42
Relative humidity	Relative humidity in %	21 to 90	45 to 89	21 to 90
Wind speed	Wind speed in km/h	6 to 29	11 to 26	6 to 29
Rain	Outside rain in mm/m^2	0 to 16.8	0 to 8.7	0 to 16.8

Fig.8 Data Description for two region in Algeria

6.2. Feature Selection in Data Set

The FFMC value for both regions clearly correlates with other features. Table-9 has data for the Béjaïa region. When the tables are inspected one by one, it becomes evident that the system had a fire at some point. In the Sidi Belabbas area, the situation is similar. In truth, the region's lowest limit is a bit lower. This is owing to the fact that the climates in both places are somewhat different. Forest fires have a favorable influence on the regression, and there is a significant connection between them, according to other factors. However, it appears that the most important factor is the above-mentioned dryness of the litter on the forest floor. This is indirectly due to the lack of rain and humidity. In addition, exposure to high temperatures and too much sun in the absence of rain and humidity causes forest litter to catch fire. Actually, as seen here, FFMC is not a problem by itself. Fires occur as a result of a common output. In the introduction part, we talked about what the data consist of. Therefore, what I have said is proof of that. Figure 9 and Figure 10 contain data for these regions

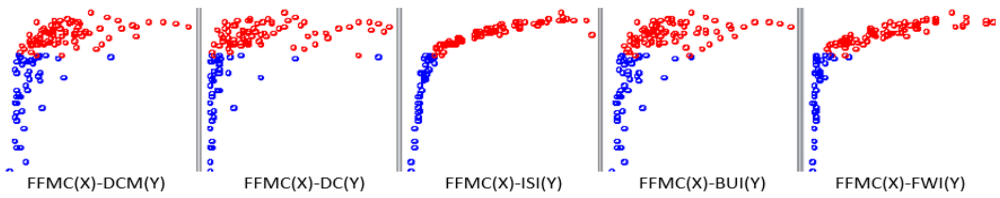

Fig.9 Relations between FFMC and other parametres for Béjaïa Region

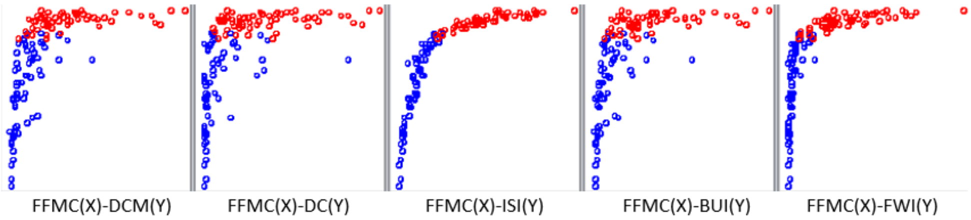

Fig.10 Relations between FFMC and other parametres for Sidi Belabbas

Apart from the actions listed above, we can verify the results by performing feature selection to the dataset including data from the two areas. We utilize it to provide a foundation for our study based on the decision tree.

We'll start with one of the feature selection techniques, the correlation attribute. It is based on the concept of examining the differences in correlation between several metrics. The data are divided into three categories: high, medium, and low (between-1,1). As a consequence, the parameters are presented in order of highest to lowest correlation. The table below shows the outcomes of this algorithm. FFMC has the highest correlation, as seen in this table. For forest fires, this characteristic is the most problematic.

Ranked Attributes

0.7701	**FFMC**
0.7361	ISI
0.7194	FWI
0.5845	BUI
0.5842	DMC

We observe that the FWI parameter has the highest value when we do the identical procedures on the Relief F Attribute algorithm. The Relief F algorithm is a method for calculating outcomes based on the Manhattan distance and miss and hit numbers. There are also weights utilized. The results are listed in a table below.

Ranked Attributes

0.16654	**FWI**
0.1632	ISI
0.14944	FFMC
0.06746	BUI
0.06346	DC

The FFMC parameter has the highest value when we do the same procedures on the Gain Ratio Attribute algorithm. The Gain Ratio attribute method is a feature selection technique that favors multivariate values and gain rates when selecting characteristics. The results are listed in a table below.

Ranked Attributes

0.7814	**FFMC**
0.7686	ISI
0.4662	FWI
0.4541	DC
0.3227	DMC

As can be seen, FFMC has the majority of the highest values, while ISI and FWI are not far behind. The values are constantly in close proximity. These findings demonstrate that ISI and FWI have a significant impact on forest fires.

The decision tree that resulted from using the Random Tree (forest) method also via Relief F selection to the data set containing both sets is shown in the Figure-11 below. When the chance of fire is analyzed over all regions, it is evident that it is influenced by the FFMC, ISI, FWI, and temperature.

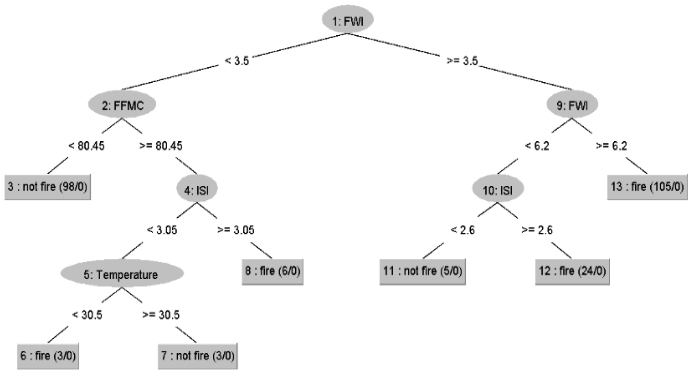

Fig.11 Decision Tree for All Region

7. Performances Evaluation

We have used the Weka tool for the performances evaluation and the Random forest classifier [17]. Figure-12 and Figure-13 were created for the Béjaïa region. Figure-14 and Figure-15 contain the data of the Sidi Belabbas region.

The following values were derived independently for each area. The confusion matrix you've seen is a matrix that shows how well a sequence performs when its true values are known. True positive, negative, and false positive and negative values are included in the data set. Different values can be found in the other table. The ratio of true positives to values is represented by the TP value. It demonstrates how sensitive the categorization algorithm is. Even if the real value is 0, the FP value represents the pace at which it returns 1. It's also known as a by-product. Precision can also be referred to as precision. It demonstrates how well all courses are anticipated. The ROC Area is a scale that depicts the classification algorithm's performance across all classes. ROC Area can be found by "Sensitivity / Specificity". Finally, the ratio of true positive values is represented by the recall value.

To examine the data, the predictions for the Sidi-Bel region, as shown in the tables, came closer to the actual outcome. It's also possible to conclude that the random forest algorithm produces better outcomes in this area. I am confident that the same result will be achieved as a result of the comparisons made in the following chapter.

CLASSES	TP RATE	FP RATE	PRECISION	ROC AREA	RECALL	PRC AREA
NOT FIRE	0.937	0.068	0.937	0.934	0.937	0.910
FIRE	0.932	0.063	0.932	0.934	0.932	0.902
AVERAGE	0.934	0.066	0.934	0.934	0.934	0.906

Fig.12 Performances with Classes for Béjaïa region

A	B	CLASSIFIED MATRIX
59	4	NOT FIRE
4	55	FIRE

Fig.13 Confusion Matrix

CLASSES	TP RATE	FP RATE	PRECISION	ROC AREA	RECALL	PRC AREA
NOT FIRE	0.953	0.013	0.976	0.970	0.953	0.947
FIRE	0.987	0.047	0.975	0.970	0.987	0.971
AVERAGE	0.975	0.035	0.975	0.970	0.975	0.963

Fig.14 Performances with Classes for Sidi Belabbas region

A	B	CLASSIFIED MATRIX
41	2	NOT FIRE
1	78	FIRE

Fig.15 Confusion Matrix

7.1. Performances Comparison between Random Forest, Classification via Regression and Naive Bayes

We will apply my findings to several approaches in this location to evaluate if the method we have picked is appropriate. The Random forest algorithm we utilize may be as good as or better than other methods in terms of efficiency and accuracy. However, all methods generate outcomes that are essentially the same. The outcomes may also differ by geography. Figure-16 shows the results for the Béjaïa region, and Figure-17 shows the results for the Sidi Belabbas region.

According to statistics from the Béjaïa region, each approach has the same accuracy rate. The same cannot be said of the Sidi Belabbas region. In comparison to other methods, the Naive Bayes method is the poorest. The same can be said regarding recall and precision. Instead of using the Random Forest approach, we should use Classification by Regression. With 98.3607 percent accuracy and 0.984 precision, it outperformed Random forest.

As a result, for the data in this region, Classification by Regression should be favored over Random Forest. Finally, it would be preferable to run the data through these algorithms in bulk and then generalize the findings. Problems are plainly visible, and the fact that it is both a general-encompassing and a region-based study will provide subsequent researchers an advantage. In different studies, although data were collected for different regions, which we examined in the Literature view section, data that were combined into a common data warehouse were used for the study. A generalization has been made over the results obtained here. In this study, I did a separate analysis for each region and also tried to reach the result by creating a common data set for generalization without favoring the two regions. Thus, I think I have achieved a more accurate result. The calculation values for the combination of the two data sets are given in Figure-18.

The random forest method provided high accuracy as a result of analyzing both individual data sets and both data sets together. The accuracy of one of the other approaches was higher, whereas the accuracy of the other method was lower. Aside from the fact that the variances between them are minor, the numbers for all three algorithms are extremely high. This demonstrates that every one of them is a good fit for these datasets. If desired, all of them can be utilized.

ALGORITHMS	ACCURACY (%)	PRECISION	RECALL	F-MEASURE
Random Forest	97.541	0.975	0.953	0.975
Classification via Regression	**98.3607**	**0.984**	**0.984**	**0.984**
Naive Bayes	94.2623	0.945	0.943	0.943

Fig.16 Comparison for the Sidi Belabbas region

ALGORITHMS	ACCURACY (%)	PRECISION	RECALL	F-MEASURE
Random Forest (Tree)	**93.44**	0.934	**0.934**	**0.934**
Classification via Regression	93.44	**0.935**	0.934	0.934
Naive Bayes	93.44	0.935	0.934	0.934

Fig.17 Comparison for the Béjaïa region

ALGORITHMS	ACCURACY (%)	PRECISION	RECALL	F-MEASURE
Random Forest (Tree)	**96.72**	0.967	**0.967**	**0.967**
Classification via Regression	96.72	**0.969**	0.967	0.967
Naive Bayes	93.85	0.938	0.939	0.938

Fig.18 Comparison for All Region

7.2. Comparison with Other Works

In this section, we will compare our results with other studies. Random forest algorithm was used in this project. Decision tree (J48) was used for the articles we will compare. The J48 algorithm is a very basic decision tree algorithm found in Weka. By using Shannon's Information Theory, it tries to improve the decision tree and make it fit for the model. To do this, the entropy value I mentioned in the Methodology section is used. The entropy value is a measure of the uncertainty of a variable. The reason for using entropy is precise because of its definition. This method attempts to construct a decision tree by measuring uncertainty changes.

We can understand which method gives better results from the data in Figure-17. The primary goal of these researches is to predict forest fires before they occur. It's also about determining what causes forest fires. The forest fire is likely to have been triggered by a combination of factors. The most crucial parts, however, are FFMC and FWI. Comparing our own research to other studies, in this circumstance, demonstrates the research's correctness. It should be noted, however, that these values are generated to use more than one parameter. Forest fires are not caused by a single factor.

Each study's accuracy and recall values were compared in Figure-19. Only values where a forest fire occurred are taken into account in this table. The values of our own research have the most worth in both cases. Finally, it's worth mentioning that. Variables may lead to differences in data sets, which should be considered.

8. Conclusions and Recommendations

In this study, I used the Random Forest (tree) method to examine data from two Algerian locations. We can see from the data that the main causes of forest fires are FFMC and FWI, and that rain, humidity, and temperature indirectly raise FFMC and FWI values. The parameters, which vary depending on the environment of the region, produced fairly consistent findings when compared to one another, and other algorithms corroborated the same conclusions.

The study's accuracy has been proved by comparisons with other studies, and the results have been concluded that they can be used in the development of new strategies to avoid forest fires.

Random forest was used to solve one of the world's major problems, forest fires, with an accuracy rate of 93.44 percent for the Béjaïa region and 97.51 percent for the Sidi-BelAbbes region. The FFMC value of greater than 80 was determined to be the direct cause of the fire in both zones. The Random forest scheme demonstrated that the fires in the Béjaïa region occurred for multiple reasons, partly due to the fact that they have distinct climates. The main difficulty was created by FWI, and a value of more than 6.2 certainly caused forest fire, according to the examination in which both sets were prevalent. When it's less than 6.2, though, it's clear that it's influenced by FFMC, ISI, and temperature. In comparison to FFMC, the number of forest fires triggered by ISI and temperature is quite low.

In the first phase, forest fires are dependent on FWI because the FWI number already reflects fire density.When the presence of fire is combined with other factors, it grows and causes an effective fire. This is basically what is meant here. The results reveal that the Random Forest method delivered results that were in line with the study's objectives, and that the conclusion was attained with adequate precision. I believe that by using various data upgrades or relevant data sets, the accuracy rate can be improved.

In a nutshell, the findings revealed that on days when there is neither rain or humidity, forest debris is the leading cause of forest fires. The cleansing of these waste should be prioritized by social welfare organizations or governments, and it should not be disregarded. I anticipate a significant reduction in the number of forest fires if this issue is resolved.

REFERENCES	ACCURACY (%)	RECALL	F-MEASURE
Our Work	**96.72**	**0.967**	**0.967**
Stojanova [15]	81.2	0.81	0.81
Faroudja [16]	82.89	0.92	0.85

Fig.19 Comparison with Existing Work

References

[1] Benkheira, A.: Ministère de l'agriculture et de développement rurale et de la pèche, direction

des forets Les feux de forêts en Algérie analyse et perspectives. Technical report (2018)

[2] San-Miguel-Ayanz, J., Durrant, T., Boca, R., Libertà, G., Branco, A., de Rigo, D., Ferrari, D., Maianti, P., Vivancos, T.A., Costa, H., Lana, F.: Advance EFFIS report on Forest Fires in Europe, Middle East and North Africa (2017)

[3]Fire Weather Description: https://wildfire.alberta.ca/wildfire-status/fire-weather/understanding-fire-weather.aspx

[4] Summary of FWI: https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi

[5] "Béjaïa, Algeria". Climatebase.ru. Retrieved 11 February 2013.

[6] "Weatherbase: Historical Weather for Sidi-bel-abbes, Algeria". *Weatherbase. 2011.* Retrieved on November 24, 2011.

[7] Cortez, Paulo, and Aníbal de Jesus Raimundo Morais. "A data mining approach to predict forest fires using meteorological data." (2007).

[8] Lin H, Liu X, Wang X, et al. (2018) A fuzzy inference and big data analysis algorithm for the prediction of forest fire based on rechargeable wireless sensor networks. Sustainable Computing: Informatics and Systems 18: 101-111.

[9] Mallinis G, Koutsias N (2012) Comparing ten classification methods for burned area mapping in a Mediterranean environment using Landsat TM satellite data. International Journal of Remote Sensing 33: 4408-4433.

[10] Cortez, P., Morais, A.: Data mining approach to predict forest fires using meteorological data. In: Neves, J., Santos, M.F., Machado, J. (eds.) New Trends in Artificial Intelligence, 13th Portuguese Conference on Artificial Intelligence (EPIA), Guimaraes, Portugal, pp. 512– 523 (2007)

[11] Sevinc V, Kucuk O, Goltas M (2020) A Bayesian network model for prediction and analysis of possible forest fire causes. Forest Ecology and Management 457: 117723.

[12] Decision tree: https://tr.wikipedia.org/wiki/Karar_a%C4%9Fac%C4%B1

[13] Random forest tree: https://medium.com/@ekrem.hatipoglu/machine-learning-classification-decision-tree-random-forest-part-12-8c9515d811b9#:~:text=Random%20Forest%20algoritmas%C4%B1%20decision%20tree,de%C4%9Fer%20al%C4%B1n%C4%B1r%20ve%20sonu%C3%A7%20%C3%BCretilir.

[14] Abid, F., & Izeboudjen, N. (2019, July). Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm. In International Conference on Advanced Intelligent Systems for Sustainable Development (pp. 363-370). Springer, Cham.

[15] Stojanova, D., Panov, P., Kobler, A., Džeroski, S., Taškova, K.: Learning to predict forest fires with different data mining techniques. In: The 9th International Multi Conference Information Society, Jubljana, Slovinia (2006)

[16] Faroudja, A. B. I. D., & Izeboudjen, N. (2020, September). Decision Tree based System on Chip for Forest Fires Prediction. In 2020 International Conference on Electrical Engineering (ICEE) (pp. 1-4). IEEE.

[17] Weka tools: https://www.cs.waikato.ac.nz/ml/weka/

Volkan Mazlum is a student at Çankaya University in the department of Computer Engineer. He is also studying a double major in Mathematics. He worked for a while at STM as an intern. He is currently working as an undergraduate researcher with a professor. He has research interests in bioinformatics, neuroscience, and algorithms.