

ITEC 2600 Group Project: Algerian Forest Fires Dataset

Raj Patel - 2181078217
Abdul Raqeeb Khatri - 217979824
Shivraj Singh Chordia -219037084
Detharrsan Sutharsan-218131615

Data set

Algerian Forest Fires Dataset Data Set

<https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++>

Introduction

- Our team chose the Algerian forest fires data file. The data file consists of 244 instances divided between two regions in Algeria namely, Bejaia region and Sidi Bel-abbes region (with 122 instances each). The data file consists of 11 attributes and 1 output attribute. The 11 attributes are : Date, temp, relative humidity, wind speed, rain, fine fuel moisture code index, duff moisture code index, drought code index, initial spread index, buildup index and fire weather index. The output attribute is whether there was a fire or no fire. To begin with we must split the dataset by 2, in order to do a statistical analysis for each data set. However in the second data set(Sidi Bel-abbes region) we had to omit a whole row(171 in excel and 168 in matlab) as the data set had a missing value in the classes column and in the fire weather index column there is an error as the row is not numeric. Therefore the whole column would give us Nan after doing a statistical analysis for the fire weather index.

Data Analysis

Comparison of Bejaia Region and Sidi Bel-abbes

The mean temperature is slightly more in the Sidi Bel-abbes region(33.24°C) in comparison to the Bejaia region where the mean temperature is 31.18°C hence between June to September 2012 the Sidi Bel-abbes region was slightly warmer. Despite the Sidi Bel-abbes region being warmer the average relative humidity in the Bejaia region is significantly more. The average relative humidity for the Bejaia region is at 67.97%, however the average humidity in the Sidi Bel-abbes region is at 55.83%. On average the Bejaia region is 12.14% more humid everyday. Apart from the Bejaia region being more humid, the average rainfall was also more in 2012. Approximately the Bejaia region had 0.8426mm of rain per day whereas the Sidi Bel-abbes region had only 0.6826mm of rain per day. On top of the Bejaia region having a higher average rainfall and humidity they also have a higher wind speed at 16 km/h in comparison to 15.083 km/h in the Sidi Bel-abbes region. However the dataset does have a few outliers, the maximum wind speed in the dataset occurs in the Sidi Bel-abbes region at 29 km/h. Despite the Bejaia region average being higher the maximum windspeed

occurred in the Sidi Bel-abbes region. Additionally in the dataset the maximum humidity occurs in the Sidi Bel-abbes region at 90%, despite the Bejaia region being 12.14% more humid the maximum humidity in a day occurred in the Sidi Bel-abbes region.

Comparison for both the regions Month wise

During the month of June 2012 the mean temperature for both the regions was relatively less than the sample mean, for the Bejaia region it was 29.8°C and for the Sidi Bel-abbes region 31.8°C. Despite the mean temperature in June 2012 being lower than the average temperature for all four months, both the regions had a higher relative humidity %. In the Bejaia region in June 2012 the average relative humidity was 70.4%, which is 2.42% more humid than the general average. In addition the Sidi Bel-abbes region had a relative humidity of 60.1667% in June 2012, that is 4.332% more humid than the average of the 4 months in 2012. The maximum humidity in June 2012 occurred in the Sidi Bel-abbes region at 90%, this shows that there is another outlier in the dataset. As the average humidity per day is 10.233% more in the Bejaia region. Apart from the temperature in June 2012 the average wind speed was also less than the mean wind speed throughout the 4 months. During June in the Bejaia region the average wind speed was 15.76 km/h and 14.9 km/h in the Sidi Bel-abbes region. In spite of the mean rainfall being greater in the Bejaia region, during June it significantly rained more in Sidi Bel-abbes region at an average of 1.133mm per day in comparison to just 0.8167mm in the Bejaia region. To conclude during June 2012 the temperature and wind speed in both the Algerian regions was relatively lower than the average of the whole sample. In the Bejaia region the average rainfall is slightly less than the total mean, in contrast the average rainfall in the Sidi Bel-abbes region during June 2012 is 0.4507mm higher than its average rainfall. As well as throughout the four months it has rained more on average in the Bejaia region however during June 2012 the average rainfall in the Sidi Bel-abbes region is significantly more(0.3167mm). Also the humidity in June 2012 in both the regions is more than the average relative humidity.

In the course of July 2012 the weather in the Algerian regions got relatively warmer in comparison to the average of the dataset. In the Sidi Bel-abbes region the average temperature rose to 34.16°C(0.927°C more than the average) and in the Bejaia region the average temperature increased to 31.83°C, which is 0.6584°C more than the average of the dataset. In the Sidi Bel-abbes region the relative humidity fell to 55.6%, which is similar to the average humidity (55.83%) however there is a significant difference in comparison to the previous month. The average humidity fell by 4.56% in July 2012. However in the Bejaia region there was a slight drop from June in the relative humidity to 69.58%, despite that the humidity is still 1.60% more than the average in the Bejaia region. The maximum humidity took place in the Sidi Bel-abbes region at 87% during the 27th July, as well as the minimum humidity (29%) during the 21st of July. This perhaps indicates that the Sidi Bel-abbes region has more extreme forms of data, as at an average the Bejaia region was 14% more humid but the maximum humidity in the month occurred in the Sidi Bel-abbes region. Apart from a rise in the temperature, there has also been a rise in the wind speed in both the regions. In the Bejaia region the wind speed increased to 16.35 km/h, which is 0.58 km/h more than the

previous month and 0.35 km/h more than the average mean. In the Sidi Bel-abbes region the wind speed increased to 15.76 km/h, that is 0.867 km/h more than the previous month and 0.75 km/h more than the average wind speed. In comparison to the previous month there is a significant decrease in the rain fall in Sidi Bel-abbes region, in July the average rainfall is 0.69mm which is 0.44 less than the previous month and just 0.0107mm more than the mean rainfall of the region. Regarding the Bejaia region there is a sharp decline in rainfall in comparison to the mean of the rainfall in the previous month and the dataset. In July there has only been an average of 0.2258mm of rainfall, this is 0.591mm less than the average of the previous month and 0.617mm less than the average rainfall for all the months. In comparison to the previous month(June), you can see a decline in the humidity and the rainfall in both the regions however there is an increase in temperature and wind speed.

In the month of August, the temperature is on a steady rise with the mean temperature in the Bejaia region escalating to 34.4°C, 3.20°C more than the average temperature in the dataset and 2.50°C more than the average temperature of July. Also the mean temperature in the Sidi Bel-abbes region rose to 36.29°C, 3.05°C more than the average temperature in the dataset and 2.12°C more than the average temperature of July. Apart from an increase in temperature there is also a sharp increase in the rainfall of the Bejaia region, being at 0.629mm. In comparison to the previous month the average rainfall has increased by 0.40mm. However in comparison to the dataset it is still 0.21mm below the average rainfall in the region. In contrast the Sidi Bel-abbes had a relatively dry month; the average rainfall in the region was 0.2677mm, in comparison to the mean of the dataset this was 0.415mm below the average and 0.4256mm less than July's mean. After the first month (June) the rainfall in the Sidi Bel-abbes has been drying up in comparison to the mean . This might indicate the dataset has outliers, since June has been the only month with significant rain. Using Matlab commands, I decided to find the number of outliers present in the dataset. Referring to z1 as the variable, we decided to use $z1 = \text{isoutlier}(\text{rain2})$, this shows a binary array. 1 being outlier and 0 being not an outlier. Furthermore I used the function $\text{nnz}(z1)$ then it showed me the dataset has 54 outliers. The relative humidity has also fallen in the Bejaia region during this month at an average of 62%, 7.58% less than the previous month and 5.97% less than the mean humidity. As well, there has been a decline in the relative humidity in the Sidi Bel-abbes region during this month at an average of 45.96%, which is 9.62% less than the previous month's humidity and 9.86% less than the total average.

Through the month of September it is evident that the weather cools down. In the Bejaia region the average temperature fell to 28.56°C, that is 5.82°C less than the mean temperature of the previous month and 2.63°C less than the average temperature. In addition to a decline in the temperature in the Bejaia region, the Sidi Bel-abbes region temperature fell to 30.16°C. A 6.12°C fall in the daily average temperature in comparison to August and a 3.07°C fall in daily average temperature during the course of the 4 months. Despite a fall in temperature both the regions had an increase in the relative humidity, the Bejaia region humidity increased to 70.067% in comparison to the 62% humidity in August and there has been a 2.09% increase in the humidity in comparison to the average humidity in the dataset. Also the Sidi Bel-abbes region had an increase of relative humidity to 62.83%, 16.86% more humid

than the previous month and nearly 7% more humid than the average humidity in the Sidi Bel-abbes region. After seeing the average temperature fall and the relative humidity rise, I determined the relation between both the variables is inversely proportional. However to make sure I used the matlab function `corr(temp,rh)` for both the data sets. The correlation between the temperature and the relative humidity in the Bejaia region is -0.6769, since it is less than 0 the prediction is true. Both the variables have a negative association with each other. For the Sidi Bel-abbes region the correlation is -0.4264, which yet again indicates a negative association. As the mean temperature in both the regions is falling, the average rainfall has significantly increased in comparison to other months. In the Bejaia region the average rainfall during September is 1.7267mm, this happens to be more than twice the average rainfall during the 4 month period (0.8426mm) and 1.098mm more rain than the daily average in August. For the Sidi Bel-abbes region the average daily rainfall increased to 0.65mm, 0.38mm more rainfall in comparison to August. However it was still 0.0326mm below the average rainfall in Sidi Bel-abbes region. As the temperature is falling, it is evident in both the regions the average rainfall is also. If the correlation is less than 0 this would indicate a negative association. Using the same method as I did earlier I was able to determine that in the Sidi Bel-abbes region there is a positive relation(0.0202) between temperature and rainfall. However in the Bejaia region there is a negative association(-0.7522) that supports my hypothesis in contrast to the relation between rainfall and temperature in the Sidi Bel-abbes region. At last the wind speed in the Bejaia region had a daily average of 15.8 km/h, 0.2 km/h less than the mean wind speed and 0.265 km/h reduction in the wind speed daily average of the previous month. This does not show a major change in the wind speed in comparison to the mean of the previous month and the data set. For the Sidi Bel-abbes region the wind speed had a minor difference from the previous month. In September the average wind speed was 14.7 km/h, 0.1194 km/h more than the daily average wind speed in August and 0.3083 km/h less than the average wind speed in Sidi Bel-abbes region.

In order to figure out if there is a fire or not we have to figure out which attributes necessitate if there would be a fire or not. Through research we found out that the attributes ISI, BUI and FWI dictate if there is a fire or not. In order to figure out the relationship between these attributes we figured out the lowest value for each attribute for every instance there was a fire. We found out that the lowest value for the first dataset ISI to be 2.8, for BUI to be 5.1, and for FWI to be 1.7. For the second data set we found the lowest value that caused the fire is ISI to be 2.7, for Bui to be 6 and FWI to be 2.5. All values must be greater than equal to the values assigned. In order for there to be a fire the instance for these attributes must all be above the lowest value previously mentioned, if two attributes are well over the lowest value but the last attribute does not cross that threshold we found that it does not dictate a fire hence there is a && relationship between these attributes. In order to get the result for both the data sets we created two empty string arrays, one was to simply collect all the values in binary and the other was created to store the values in fire/not fire. After that we converted the classes column to binary, 1 indicating fire and 0 indicating not fire. Furthermore we created another string array displaying the result from the classes column. Then to see whether our results match we used the matlab function `j= strcmp(result1,claresult)` to compare the binary arrays.

For the first dataset we got three inconsistencies, during the 21st, 37th and 43rd row(found this using (find j==0)). Furthermore for the second data set we got 2 inconsistencies using j1 = strcmp (result3, claresult1). The inconsistencies occurred in row 44 and 57.This was done using the same command, (find j1==0).

Conclusion, comments and/or predictions

To conclude with, despite the Sidi Bel-abbes region being warmer the average relative humidity in the Bejaia region is significantly more. The warmest month of the year was August, in the Bejaia region the temperature escalated to 34.4°C and 36.29°C in the Sidi-bel abbes region. The coldest month of the year was a different month for both the region. For the Bejaia region the temperature was lowest in June, the daily average temperature fell to 29.8°C. However for the Sidi-bel abbes region the coldest month occurred during September the average daily temperature was 30.16°C. The maximum wind speed for the Bejaia region occurred in the month of July at an average of 16.35 km/h and the wind speed was lowest in the month of June at an average of 15.76 km/h. For the Sidi-bel abbes region the maximum wind speed occurred in the month of July at an average of 15.76 km/h and the lowest was August 14.58 km/h. As the maximum temperature was during August and the minimum wind speed occurred during that month for the Sidi-bel abbes region. Hence according to my hypothesis both the temperature and wind speed in August for the Sidi-bel abbes region must have a negative relation. The correlation amongst both the relations is -0.4940, which shows a negative association between both the variables. From this I am able to conclude that the average temperature in the Sidi-bel abbes region is relatively higher. However the wind speed is more extreme in the Bejaia region, the month where the wind speed was maximum in the Sidi-bel abbes region it happens to be the minimum wind speed by month in the Bejaia region. In the Sidi-bel abbes region the maximum rainfall occurred during June at an average of 1.133mm daily rainfall, this was the month where the temperature was minimum. Despite the Bejaia region having more average daily rainfall the month where it rained maximum was June in the Sidi-bel abbes region. At an average the relative humidity was 12.14% more humid in the Bejaia region, the maximum humidity occurred in June 2012 for the Bejaia region and the minimum occurred in August. The humidity and temperature are inversely proportional as the minimum humidity occurred when the temperature was maximum in the Bejaia region. The correlation between both the variables is -0.7510, indicating a negative association.

References

<https://www.mathworks.com/help/matlab/ref/pie.html>

<https://www.mathworks.com/matlabcentral/answers/409174-how-to-check-number-of-zeros-in-an-array>

<https://www.mathworks.com/help/matlab/ref/tiledlayout.html>

https://www.mathworks.com/help/matlab/import_export/read-spreadsheet-data-into-table.html

<https://www.mathworks.com/help/matlab/ref/find.html>

<https://www.mathworks.com/help/stateflow/ref/strfind.html>

https://eclass.yorku.ca/pluginfile.php/4245814/mod_resource/content/1/Week4%20ProgrammingUserDefinedFunction.pdf

https://eclass.yorku.ca/pluginfile.php/4337862/mod_resource/content/2/Week9%20DataAnalysis.pdf

https://eclass.yorku.ca/pluginfile.php/4364543/mod_resource/content/1/Week10%20Polyfituncstrings.pdf

Summary of Team collaboration

Task	Finished by
Task 1 : Discussion	All
Task 2 : Introduction	Raj, Detharsan
Task 3 : Data Analysis	Shivraj, Abdul,Raj
Task 4 : Conclusions	Detharsan
Task 5 : Matlab Code	Shivraj, Abdul