Lab Assignment 5 – Decision Tree classifier

Problem Statement:

Forest fire is a disaster that causes economic and ecological damage and human life threat. Thus predicting such critical environmental issue is essential to mitigate this threat. The *Algerian Forest Fire* dataset includes 244 instances that regroup a data of two regions of Algeria, namely the Bejaia region located in the northeast of Algeria and the Sidi Bel-abbes region located in the northwest of Algeria. There are 122 instances for each region and the instances are recorded for the period from June 2012 to September 2012.

The dataset includes 11 attributes and 1 output attribute (class)

Goal: In this challenge, you must build a predictive model using Decision Tree classifier that is able to predict fire based on the meteorological data corresponding to the critical weather elements that influence the forest fire occurrence, namely temperature, relative humidity, wind speed, etc

1. Load forest fire prediction dataset from https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++

2. Do the exploratory analysis of the dataset to determine the importance of each feature:
    - Perform univariate analysis by plotting various charts like: bar charts, distribution plots, boxplots.
    - Perform multivariate analysis
3. Impute the missing values and remove any undesirable feature from the dataset.
4. Check for the outliers in the columns and treat the outliers if present.
5. Split the dataset into train and test.
6. Construct a fully grown Decision tree model to predict the fire. Check whether your model is an overfitted model.
7. Construct Decision tree models using different hyper-parameter values to restrict the growth of the tree.
8. Compare the performance of different decision tree models constructed in step 6 and 7, using various evaluation metrics.
9. Look for real world applications where you can apply Decision Tree classification model.