Motivation

The aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorological and other data.

Number of Instances: 517, **Number of Attributes: 1**3

Attribute Information:
1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9
2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9
3. month - month of the year: 'jan' to 'dec'
4. day - day of the week: 'mon' to 'sun'
5. FFMC - FFMC index from the FWI system: 18.7 to 96.20
6. DMC - DMC index from the FWI system: 1.1 to 291.3
7. DC - DC index from the FWI system: 7.9 to 860.6
8. ISI - ISI index from the FWI system: 0.0 to 56.10
9. temp - temperature in Celsius degrees: 2.2 to 33.30
10. RH - relative humidity in %: 15.0 to 100
11. wind - wind speed in km/h: 0.40 to 9.40
12. rain - outside rain in mm/m2 : 0.0 to 6.4
13. area - the burned area of the forest (in ha): 0.00 to 1090.84

Code Snippet

```
In [37]: lr = LinearRegression()
         lrModel = lr.fit(trainingDF)

In [38]: type(lrModel)

Out[38]: pyspark.ml.regression.LinearRegressionModel

In [39]: predictionsAndLabelsDF = lrModel.transform(testDF)

In [40]: eval = RegressionEvaluator()

In [41]: eval.setMetricName("rmse").evaluate(predictionsAndLabelsDF)

Out[41]: 22.808595864541225
```

Performed Linear Regression to predict the area of the fire. The rmse obtained = 22.8

Visualization