Forest Fires in Portugal

João Teixeira - up200705307, Nuno Peixoto - up200804621

25-01-2015

Introduction

- Monitoring and forecasting forest fires in Portugal;
- The several variables may influence the burnt area;
- In 2003, Portugal faced the worst forest fire losing 8.6% of of the total area;
- Elevation, slope or density are some of the specifications of the data set;

Objective : Explore and predicte the data of the forest.

Exploratory analysis of the data

- Global Summary
- Main Variables
- Target Variable

Global Summary

- Number of Columns: 81.
- Number of Rows: 990.
- Number of Data: 80190.
- Target Value: 1 (TotalBurntArea) - Numeric variable.
- Number of Unknown Values: 0.

Global Summary (cont.)

Climate Variables - The climatic conditions may affect the probability of a fire to occur;

Landscape Variables - The landscape has been extensively associated with fire occurrence;

Socio-economic Variables - Human have impact in historical fire patterns;

Topographic Variables - The topographic features may influence the fire ignitions;

Main Variables

In the following table we have the **TOP5** main variables:

attr_importance	attribute
0.2037	ELEV_MAX
0.1962	bio1
0.1926	ELEV_MEAN
0.1898	bio7
0.1844	DensPop01

Main Variables (Number of outliers)

- ELEV_MAX: 8 (0.81%)
- Bio1: 21 (2.12%)
- ELEV_MEAN: 9 (0.91%)
- Bio7: 1 (0.1%)
- DensPop01: 132 (13.33%)

Main Variables (Standard Deviation)

- ELEV_MAX: 339.100654
- Bio1: 14.710837
- ELEV_MEAN: 251.9971412
- Bio7: 30.9059137
- DensPop01: 1222.3683295

Target Variable

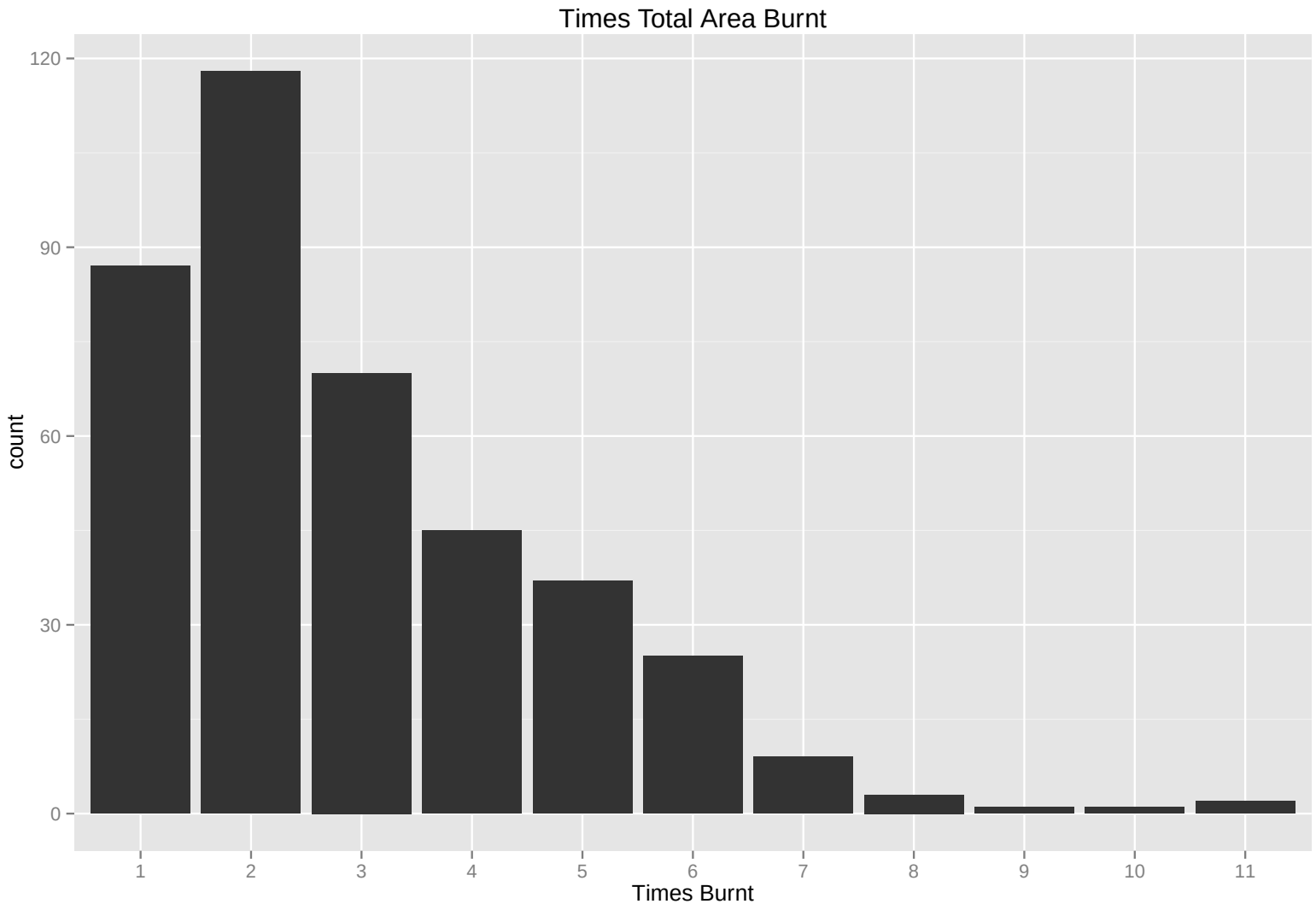
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0	609	2550	2752	68981

Target Variable (Number of outliers)

- TotalBurntArea: 106 (10.71%)

We can see that more than 10% of the total burnt area values are considered outliers.

Target Variable (Total Area vs. Total Burnt Area)



Data Pre-Processing

- Remove None importance Variables
- Normalizing Value

Remove None importance Variables

attr_importance	attribute
NaN	TCI_STD
NaN	LPI
NaN	ED
NaN	FRAC_SD
NaN	IJI
NaN	ENN_AM
NaN	eucalipto_AREA_perc
NaN	outfolhosas_AREA_perc

- This pre-processing that does not prejudice the information

Normalizing Value

- Data normalization pre-processing we will use for the analisys;

```
##   ELEV_MAX ELEV_MEAN ELEV_STD SLOPE_MAX SLOPE_MEAN SLOPE
## 1      396  168.5080  76.7385   44.9590   18.87750  11.9
## 2      706  604.4890  42.7725   39.1152    8.99396   6.0
## 3       88   34.2032  23.7021   14.3287    2.32026   1.7
```

```
##      ELEV_MAX ELEV_MEAN   ELEV_STD  SLOPE_MAX SLOPE_MEAN
## 1 -0.2493703 -0.569479  0.2047242  0.5829093  1.0775585
## 2  0.6648126  1.160624 -0.4090371  0.2295771 -0.3401967
## 3 -1.1576551 -1.102441 -0.7536369 -1.2690826 -1.2975129
```