



UCLM - Escuela Superior de Informática - Ciudad Real

Machine Learning Lab Book

Made by:
Benjamín Cádiz de Gracia
Iván García Herrera
Dídimo Javier Negro Castellanos

Date: October, 2018.

Índice

1. Monday, 15 October 2018	3
1.1. Steps	3

1. Monday, 15 October 2018

1.1. Steps

- First, we have created a *cookiecutter* based directory structure.
- Then, we have revised the dataframe.
- We have deleted non numerical columns, it means, the columns that contain string values.
- We have deleted the columns which had all **NaNs**. If we don't have care enough and we try to delete rows with **NaNs** directly, we will delete the complete dataset, because there are some columns with all null values.
- We have removed duplicate rows, that is, two or more rows with exactly the same values. We have seen that in this dataset there are not duplicate rows.
- We have deleted the rows containing not a number (**NaNs**) values.
- We have tried to identify some outliers in some columns but we haven't have success.