Escola Universitària
Politècnica de Mataró

Centre adscrit a:

UNIVERSITAT POLITÈCNICA
DE CATALUNYA

Grau en enginyeria informàtica

Machine Learning with H2o and R: A Big Data problem.

Feasability Study

Ricard Gardella Garcia
TUTOR: Dr. Xavier Font Aragonès

Spring 2017

TecnoCampus
Mataró-Maresme

Index.

Figure Index.

Table Index.

Term Index.

TFG	Final Degree Project
“	Inches
MB	Megabytes
GHZ	Giga Hertz

1. Schedule.

1.1. Initial schedule

The tasks to perform for the developing of this TFG are:

Task Name	Start Date	End Date	Duration
Theory framework research	2/5/17	13/5/17	11
Research R and H2o.	2/5/17	7/5/17	5
Research for data.	2/5/17	4/5/17	2
Data understanding	5/5/17	6/5/17	1
Descriptive analytics	7/5/17	12/5/17	5
Deep Learning	13/5/17	25/5/17	12
Documentation	20/5/17	30/5/17	10

Table 1.1. Initial schedule

1.2. Gantt Diagram

The figure 1.1 shows the Gantt diagram based on the tasks described in the point 1.1 of this document.

The critical tasks are: Research for data, data understanding and descriptive analytics.

- Research for data: Without any data, it is impossible to work, so, this task was crucial to be completed before the starting of the next tasks.
- Data understanding: Before starting to work with the data, was necessary a study of it.
- Descriptive analytics: Before starting the Deep Learning task, I considered that the descriptive analytics tasks must be finished. Descriptive analytics give a more clear vision of the problem.

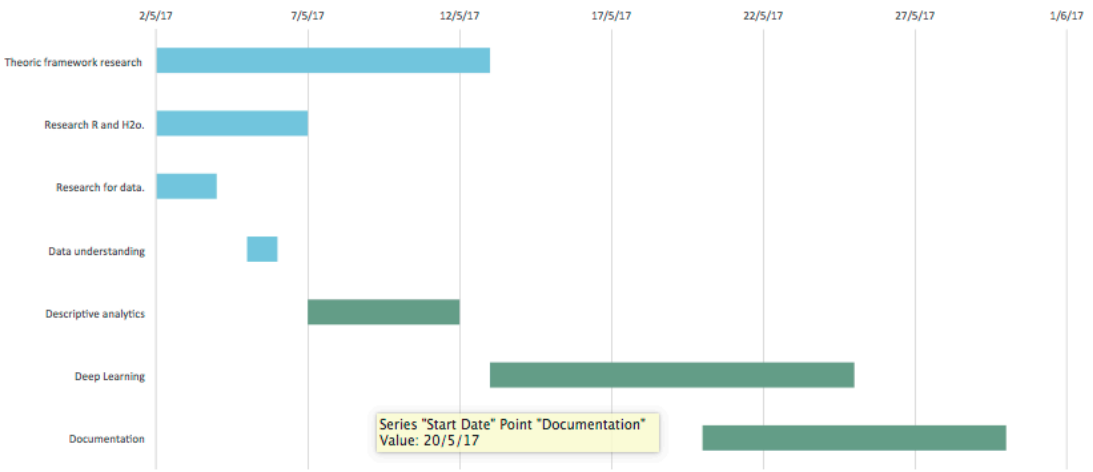

Figure 1.1. Gantt diagram

1.3. Deviations

I experimented these deviations in the developing of this project

1. The framework research took less time than expected; instead of eleven days, it took eight and it was started earlier.
2. The descriptive analytics tasks took a lot of more time because I was not taken into account the research of libraries for better data representation. Instead of five days, it took 8.
3. The descriptive analytics task was improved, so, two more dedication days were needed.
4. The documentation tasks was started earlier on time.

1.4. Professional profile

The profile of the author of this project is a computer engineer

2. Viability analysis.

2.1. Technical resources.

The only technical resources needed for the project are a computer.

The computer used for the project is a MacBook Pro Retina display 15" early 2013. The details of the computer are showed in the figure 2.1.

Figure 2.1. MacBook Pro details.

2.2. Infrastructure resources.

There are no infrastructure resources used for this project.

3. Economic viability.

3.1. Estimate cost.

These are the parameters that are being took into account in the estimate cost:

- Number of hours: Every day approx. eight hours since 2/5/2017, a total of approx. 240 hours.
- Price/h: The price/h is established at 30€/h.
- Computer cost: The MacBook Pro used costs 2500€, the project lasted 2 months, so, the proportional cost for this project is 140€.
- Other costs: 5€ in concept of electricity and 100€ in concept of transport.

That makes 7445€.

4. Legal aspects.

This TFG is under the creative commons licence.

- H2o uses a Apache 2.0 open source licence.
- RStudio is under the GNU Affero General Public Licence v3.
- Leaflet is under a BSD-2-Clause.