

Anàlisi de la relació entre les emissions de Co2 i l'augment de la temperatura, mitjançant tècniques i tecnologies de Big Data. 1r Informe de Seguiment

Miquel Freixes Faya



1 EVOLUCIÓ DELS OBJECTIUS

L'OBJECTIU principal d'aquest treball segueix sent el mateix que el esmentat anteriorment, analitzar el canvi de temperatura al llarg dels anys. El que si que ha canviat són parametres amb els que es compararà la temperatura. Enlloc de fer-ho amb la petjada ecològica i les emissions de Diòxid de Carboni (CO₂), com s'havia plantejat en un inici, es farà amb les emissions de tres gasos hivernacles. També s'ha decidit reduir l'àrea d'on s'agafaran les dades, enlloc de fer-ho de tota la Terra es focalitzarà en el continent d'Europa, concretament en els països dins de la Unió Europea.

El primer canvi s'ha decidit després d'analitzar els *DataSets* amb més detall i veure que tenia més sentit comparar els canvis de temperatura amb les principals causes que podien provocar-los. Per tant s'ha decidit agafar només les dades sobre les emissions de CO₂, les de Metà (N₂O) i les de La comparació inicial també complicava bastant el treball, degut a la complexitat de les fórmules i conceptes involucrats en el *DataSet* de la petjada ecològica. Com el treball està enfocat en el tractament de les dades i al seu anàlisi, s'ha trobat més coherent agafar conceptes menys complicats per poder dedicar més temps en el seu processament.

El segon té una justificació semblant, agafant els països dins de la Unió Europea s'obtenen unes 22000 files d'informació i agafar més països hagués produït un augment del temps d'entrenament de l'algorisme Això causaria una extensió de la planificació. Com s'ha esmentat anteriorment, com l'objectiu es dins de l'àmbit informàtic, hi ha suficient amb un gruix de dades de més de 20000 files per asolir-lo, per tant es descartaran la resta de països. A més a Europa hi ha una combinació de països curiosa, ja que hi ha alguns que han reduït dràsticament les emissions mentres d'altres que les han anat pujant o mantingut al llarg dels anys.

- E-mail de contacte: m.freixes.faya@gmail.com
- Menció realitzada: Tecnologies de la Informació
- Treball tutoritzat per: Jordi Casas Roma (departament)
- Curs 2018/19

2 EVOLUCIÓ DE LA PLANIFICACIÓ

La planificació s'està seguint tal com es va acordar al informe inicial. Hi ha un retard d'aproximadament una setmana, degut a que no es va considerar la redacció dels informes. A la figura 1 es mostra el Diagrama de Gantt actualitzat amb les tasques dels informes afegides.

3 EVOLUCIÓ DE LA METODOLOGIA

A la figura 2 es pot veure com s'ha aplicat el mètode Kanban en aquesta primera tasca. S'ha decidit seguir la metodologia per cada fase i no de tot el projecte a la vegada. Així es pot mantenir un seguiment més concís de les tasques de cada una de les fases. Per fer-ho simple s'han creat tres columnes de tasques: Les que s'han de fer, les que s'estàn fent i les que estàn fetes. Així es pot veure tota la feina feta i la que queda per fer i planejar-se segons avanci el projecte. També es pot veure el repositori on de moment només hi ha penjat els arxius dels informes, fonts d'informació i els *DataSets*. Això és degut a que tot el codi s'ha escrit per la interfície on-line de DataBricks, el qual ja té un repositori incorporat, i s'ha trobat innecessari replicant els commits que es feien al repositori local.

4 PREPROCESSAMENT DE DADES

Per començar es va importar els dos *DataSets* des del seu csv corresponent:

- El primer està format de quatre columnes, tal com es mostra a la taula 1 hi ha una amb la data, en format String amb dies mesos i anys. El segon valor es la temperatura mesurada amb un marge d'error del 95%, serà el valor que es donarà per vàlid al llarg del treball. El tercer es el marge d'error que pot tenir la temperatura de la segona columna i l'última el nom del país al que s'està mesurant les temperatures.
- El segon està compost per les mateixes columnes. Com es pot veure a la taula 2 es bastant semblant a

l'anterior. La primera es el nom del país on s'està mesurant, la segona es l'any on s'ha mesurat, aquest cop en format de enter. En la tercera tenim el valor de la mitjana de Kilotones de gas produït per capita en el país i en l'última tenim el tipus de gas que s'ha mesurat.

REFERÈNCIES

- [1] <https://www.kaggle.com/unitednations/international-greenhouse-gas-emissions>
- [2] <https://www.kaggle.com/footprintnetwork/national-footprint-accounts-2018>
- [3] "Ecological Footprint", *WWF*. [En línia]. Disponible a: https://wwf.panda.org/knowledge_hub/teacher_resources/webfieldtrips/ecological_balance/eco_footprint/. [Accedit Octubre 4, 2018].
- [4] "Sample Research", *The Standish Group*. [En línia]. Disponible a: https://www.standishgroup.com/sample_research [Accedit Octubre 4, 2018].
- [5] "Scrum Reference Card now available in English and Spanish", *Agile Methodology*. [En línia]. Disponible a: <http://agilemethodology.org/>. [Accedit Octubre 4, 2018].
- [6] "The Complete Guide to Understanding Agile Testing", *QA Symphony*. [En línia]. Disponible a: <https://www.qasymphony.com/blog/agile-methodology-guide-agile-testing/>. [Accedit Octubre 4, 2018].
- [7] A.Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow*. Sebastopol, CA: O'Reilly Media, 2017.
- [8] "Supervised Learning", *Scikit-Learn*. Disponible a: http://scikit-learn.org/stable/supervised_learning.html#supervised-learning. [Accedit Octubre 5, 2018].

5 ANNEX

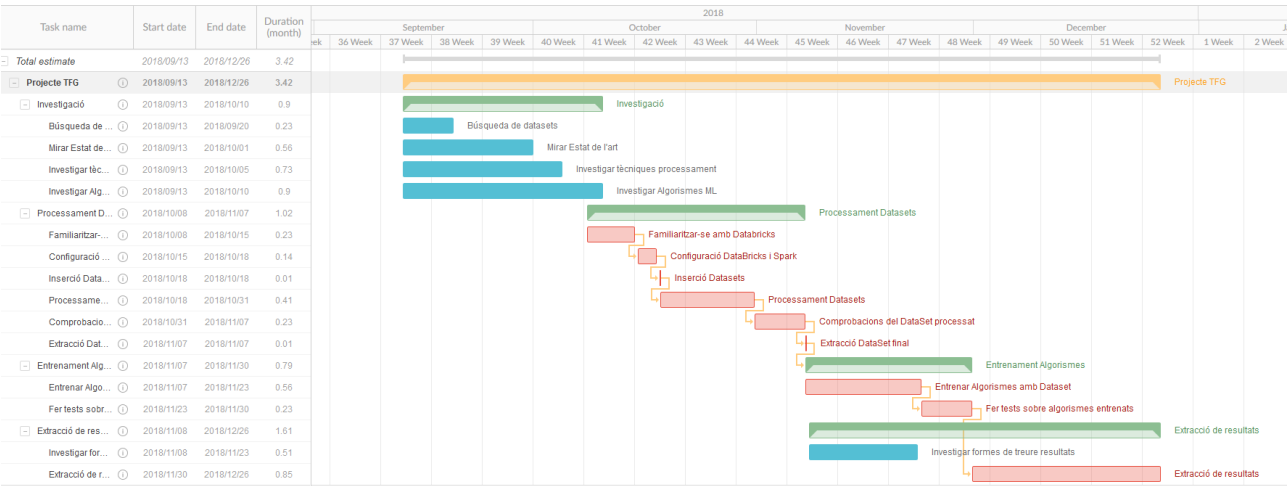


Fig. 1: Diagrama de Gantt