

# Organizacion de Datos - TP1

May 21, 2020

# Contents

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Introducción</b>                            | <b>3</b> |
| 1.1      | Objetivos . . . . .                            | 3        |
| 1.2      | Set de datos . . . . .                         | 3        |
| <b>2</b> | <b>Análisis del set de datos</b>               | <b>4</b> |
| 2.1      | Análisis general . . . . .                     | 4        |
| 2.2      | Analisis de las palabras que se usan . . . . . | 4        |
| 2.3      | . . . . .                                      | 4        |

# 1 Introducción

## 1.1 Objetivos

El presente informe contiene un análisis exploratorio como primer trabajo practico grupal de la materia, hecho sobre un set de datos obtenido a partir de Twitter, con sus respectivas visualizaciones y conclusiones.

En dicho set, las entradas se obtienen a partir de una lista de tweets que podrían o no hablar sobre una catástrofe, junto con otros atributos que los caracterizan. Buscaremos, luego de un análisis de los datos, si existe alguna relación directa entre las variables que nos presentan o si se puede obtener alguna conclusión interesante sobre los datos que podemos extraer de dichas variables.

## 1.2 Set de datos

El set de datos dado para el análisis consta de registros que representan cada uno de los tweets, con la siguiente estructura:

- id - identificador unico para cada tweet
- text - el texto del tweet
- location - ubicación desde donde fue enviado (podría no estar)
- keyword - un keyword para el tweet (podría faltar)
- target - en train.csv, indica si se trata de un desastre real (1) o no (0)

## **2    Análisis del set de datos**

### **2.1    Análisis general**

### **2.2    Análisis sobre el texto del tweet**

#### **2.2.1    Largo de los tweets**

#### **2.2.2    Largo de palabras**

#### **2.2.3    Palabras en común**

#### **2.2.4    Hashtags**

#### **2.2.5    Etiquetas**

#### **2.2.6    Links**

#### **2.2.7    Fechas y horarios**

### **2.3    Análisis sobre palabras clave**

### **2.4    Análisis sobre la localización de los tweets**

### **2.5    Otros Análisis**

## **3    Conclusiones**