**Projet Kaggle : Telstra**

**But : Prédire la sévérité des perturbations du service sur leur réseau en utilisant une base de données avec plusieurs attributs différents.**

**Prédire si c’est une perturbation momentanée ou une interruption de la connectivité.**

**Goal:** The goal of the problem is to predict Telstra network's fault severity at a time at a particular location based on the log data available.

Each row in the main dataset (train.csv, test.csv) represents a *location and a time point*. They are identified by the "id" column, which is the key "id" used in other data files.

Fault severity has 3 categories: 0,1,2 (0 meaning no fault, 1 meaning only a few, and 2 meaning many).

Different types of features are extracted from log files and other sources: *event\_type.csv*, *log\_feature.csv*, *resource\_type.csv*, *severity\_type.csv*.

Note: “severity\_type” is a feature extracted from the log files (in severity\_type.csv). Often this is a severity type of a warning message coming from the log. "severity\_type" is categorical. It does not have an ordering. “fault\_severity” is a measurement of actual reported faults from users of the network and is the target variable (in train.csv).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| severity\_type.csv [Categorie] | resource\_type.csv  [] | log\_feature.csv  [] | event\_type.csv  [] | sample\_submission  [] |
| * id * severity\_type   ex : 6597,severity\_type 2 | * id * resource\_type   ex : 6597,resource\_type 8 | * id * log\_feature * volume   ex : 6597,feature 68,6 | * id * event\_type   ex : 6597,event\_type 11 | * id * predict\_0 * predict\_1 * predict\_2   ex : 11066,0,1,0 |

|  |  |
| --- | --- |
| Train.csv | Test.csv |
| * id * location * **fault\_severity (variable à prédire)**   ex : 14121,location 118,1 | * id * location   ex : 11066,location 481 |

**File descriptions**

train.csv - the training set for fault severity (measurement of actual reported faults from users)

test.csv - the test set for fault severity

sample\_submission.csv - a sample submission file in the correct format

event\_type.csv - event type related to the main dataset

log\_feature.csv - features extracted from log files

resource\_type.csv - type of resource related to the main dataset

severity\_type.csv -  severity type of a warning message coming from the log