



Proyecto de consolidación NLP

Proyecto de consolidación

Se trata de un problema de clasificación multietiqueta en el que tenéis que etiquetar un conjunto de tweets de acuerdo a 11 sentimientos ('anger', 'anticipation', 'disgust', 'fear', 'joy', 'love', 'optimism', 'pessimism', 'sadness', 'surprise', 'trust').

ID	Tweet	anger	anticipatio	disgust	fear	joy	love	optimism	pessimism	sadness	surprise	trust
2018-Es-06697	No me pienso perder la pelea de McGregor contra Mayweather	False	True	False	False	False	False	False	False	False	False	False
2018-Es-05385	Yo preocupada pensando que mi papá ya había llegado y me estaba llamando para decirme que el vuelo todavía no había salido 👍	False	True	False	True	False	False	False	False	False	False	False
2018-Es-03777	Pucha ya no me sirven todos los carros a Rondizzoni.... #ironía jajajajaja #unañomas	False	False	False	False	True	False	False	False	False	False	False
2018-Es-02175	Si estar contigo es un delito, hago cien años en prision 😊	False	False	False	False	True	True	False	False	False	False	False
2018-Es-00726	@macacifuentesC @sergmujica Perfect... Y de una mansa ni que terrible PLR	False	False	True	False	False	False	False	False	False	False	False

Disponéis de un conjunto de 3561 tweets para entrenar el modelo o modelos necesarios, en el archivo 'sem_eval_train_es.csv'.



Proyecto de consolidación

Cada grupo tiene que validar sus modelos sobre un conjunto ciego de test con el nombre **'sem_eval_test_grupo_N.csv'** donde N es el nº asignado al grupo. Cada Tweet tiene un ID (primera columna) que tenéis que mantener en vuestra solución. Debéis aplicar vuestro modelo entrenado a este conjunto y guardarlo en un archivo CSV de nombre **'soluciones_grupo_N.csv'**. El archivo a entregar cargado en Pandas debe tener esta estructura:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 679 entries, 0 to 678
Data columns (total 12 columns):
#      Column          Non-Null Count  Dtype
---  -
0      ID                679 non-null   object
1      anger             679 non-null   bool
2      anticipation      679 non-null   bool
3      disgust           679 non-null   bool
4      fear              679 non-null   bool
5      joy               679 non-null   bool
6      love              679 non-null   bool
7      optimism          679 non-null   bool
8      pessimism         679 non-null   bool
9      sadness           679 non-null   bool
10     surprise          679 non-null   bool
11     trust             679 non-null   bool
dtypes: bool(11), object(1)
```

Entrega:
Cada equipo deberá entregar por correo electrónico a joan.vila@uv.es tanto el archivo CSV de las predicciones en test como el código en Python utilizado tanto para el entrenamiento de los modelos como la inferencia en el conjunto ciego de test.



Proyecto de consolidación

Podéis evaluar el rendimiento de vuestra solución en <http://febe.uv.es:8501>

Evaluación del proyecto de consolidación

Máster IA

Módulo NLP

Selecciona un archivo

Drag and drop file here

Limit 200MB per file

Browse files

soluciones_grupo_5.csv 54.9KB

×

Rendimiento

Cargada tabla con 679 valores

	valor
accuracy	0.1281
f1_score	0.2409
precision	0.6497

☐ Mostrar datos