75.06/95.58 Organización de Datos
Segundo Cuatrimestre de 2018
Trabajo Práctico 2: Enunciado

El segundo trabajo práctico es una competencia de Machine Learning en donde cada grupo debe intentar determinar, para cada usuario presentado, cuál es la probabilidad de que ese usuario realice una conversión en Trocafone en un periodo determinado.

La competencia se desarrollará en la plataforma de Kaggle. En el siguiente link se provee la siguiente información para la competencia:
https://drive.google.com/file/d/1kQujhvOKAU4EzhaYDbgquf_XKFt9sHMy/view?usp=sharing

En este podrán encontrar los siguientes archivos:

`events_up_to_01062018.csv`
`labels_training_set.csv`

El archivo `events_up_to_01062018.csv` contiene en el mismo formato utilizado en el TP1 información de eventos realizado en la plataforma para un conjunto de usuarios hasta el 31/05/2018.

Por otro lado el archivo `labels_training_set.csv` indica para un subconjunto de los usuarios incluidos en el set de eventos `events_up_to_01062018.csv` si los mismos realizaron una conversión (columna label = 1) o no (columna label = 0) desde el 01/06/2018 hasta el 15/06/2018.

```
conversiones_train_pub = pd.read_csv('Data/Trocafone/Pub/labels_training_set.csv')
conversiones_train_pub.head()
```

	person	label
0	0566e9c1	0
1	6ec7ee77	0
2	abe7a2fb	0
3	34728364	0
4	87ed62de	0

La información de estos archivos debe ser utilizada para entrenar un modelo de Machine Learning, de tal forma de poder indicar la probabilidad de que conjunto seleccionado de usuarios realice una conversión desde el 01/06/2018 al 15/06/2018.

Se pedirá indicar esa probabilidad de conversión para usuarios que no se encuentran en el archivo `labels_training_set.csv`, pero para los cuales se cuenta con información en `events_up_to_01062018.csv`

El listado de estas personas será provisto en el archivo `trocafone_kaggle_test.csv`

```
personas_a_predecir = pd.read_csv('Data/Trocafone/trocafone_kaggle_test.csv')
personas_a_predecir.head()
```

	person
0	4886f805
1	0297fc1e
2	2d681dd8
3	cccea85e
4	4c8a8b93

El link a la competencia es https://www.kaggle.com/t/f477d6cc8ce34161a33bcc02ad055912

Los grupos deberán probar distintos algoritmos de Machine Learning para predecir cuál es la probabilidad de conversión del conjunto de usuarios seleccionados de Trocafone para la competencia en el periodo descrito. A medida que los grupos realicen pruebas deben realizar el correspondiente submit en Kaggle para evaluar el resultado de los mismos.

Al finalizar la competencia el grupo que mejor resultado tenga obtendrá 10 puntos para cada uno de sus integrantes que podrán ser usados en el examen por promoción o segundo recuperatorio.

Requisitos para la entrega del TP2:

- El TP debe ser programado en Python o R.

- Debe entregarse una carpeta con el informe de algoritmos probados, algoritmo final utilizado, transformaciones realizadas a los datos, feature engineering, etc.
- La entrega debe incluir también un link a github con el informe presentado en pdf, y todo el código.
- El grupo debe presentar el TP en una computadora en la fecha indicada por la cátedra, el TP debe correr en un lapso de tiempo razonable (inferior a 1 hora) y generar un submission válido que iguale el mejor resultado obtenido por el grupo en Kaggle.

El TP2 se va a evaluar en función del siguiente criterio:

- Cantidad de trabajo (esfuerzo) del grupo: ¿Probaron muchos algoritmos? ¿Hicieron un buen trabajo de pre-procesamiento de los datos y feature engineering?
- Resultado obtenido en Kaggle (obviamente cuanto mejor resultado mejor nota)
- Presentación final del informe, calidad de la redacción, uso de información obtenida en el TP1, conclusiones presentadas.
- Performance de la solución final.