# 75.06/95.58 Organización de Datos
# Segundo Cuatrimestre de 2019
# Trabajo Práctico 2: Enunciado

El segundo TP es una competencia de Machine Learning en donde cada grupo debe intentar determinar, para cada propiedad presentada, cuál es su valor de mercado.

La competencia se desarrolla en la plataforma de Kaggle, se proveen una serie de archivos en:

- Entrenamiento
- Evaluación
- Ejemplo de Respuesta

El dataset consta de propiedades en venta en México entre los años 2012 y 2016, valuadas en pesos mexicanos. El archivo *train.csv* tiene 240K filas y 22 columnas, el archivo *test.csv* tiene 60K filas y 21 columnas. El equipo de Navent preparó un tutorial, en formato de jupyter notebook.

- *id*: Un id numérico para identificar la propiedad
- *titulo*: El título de la propiedad publicada
- *descripcion*: La descripción de la propiedad publicada
- *direccion*: La dirección de la propiedad
- *ciudad*: La ciudad de la propiedad
- *provincia*: La provincia donde está localizada la propiedad
- *lat*: Latitud
- *lng*: Longitud
- *tipodepropiedad*: El tipo de propiedad (Casa, departamento, etc)
- *metrostotales*: Metros totales de la propiedad
- *metroscubiertos*: Metros cubiertos de la propiedad
- *antiguedad*: Antigüedad de la propiedad
- *habitaciones*: Cantidad de habitaciones
- *garages*: Cantidad de garages
- *banos*: Cantidad de baños
- *fecha*: Fecha de publicación
- *gimnasio*: Si el edificio o la propiedad tiene un gimnasio
- *usosmultiples*: Si el edificio o la propiedad tiene un SUM
- *piscina*: Si el edificio o la propiedad tiene un piscina
- *escuelascercanas*: Si la propiedad tiene escuelas cerca
- *centroscomercialescercanos*: Si la propiedad tiene centros comerciales cerca
- *precio*: Valor de publicación de la propiedad en pesos mexicanos

*En el siguiente link pueden acceder a la competencia.*

Los submits con el resultado deben tener el formato:

Id: Un id numérico para identificar la propiedad
Valor: Precio estimado para la propiedad.

Siendo los Ids los correspondientes a las propiedades incluidas en el archivo Evaluación.

Los grupos deberán probar distintos algoritmos de Machine Learning para predecir cuál es el valor de mercado de cada propiedad. 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 programarse 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.
- En la entrega debe adjuntarse la carpeta presentada en el TP1.
- La entrega debe incluir también un link a github con el informe presentado en pdf, y todo el código (no hay que imprimir código, solo debe incluirse el link al repositorio en donde se encuentre el código subido).
- 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.