

1º

FEATURE  
ENGINEERING

(CAT BOOST)

(BUEN IN  
DESCR./MT.)

→ NUEVAS FEATURES

¿Que hacemos con  
los NAN?

→ Si es numerica? → \*  
→ Si es categorica? →  
→ Si es boolean? → \*

ANIMOS PROMEDIO!!  
EL

\* { ANTIGÜEDAD → AVG. por ciudad.  
HABITACIONES → AVG. por ciudad.  
GARAGE → 0  
BAÑOS → 1  
METROS CUADRADOS → AVG. x piso.  
METROS TOTALES → AVG. x ciudad.

TP2

"Regla de ruta"

3º

TUNEO HIPER-PARAMETROS

→ GRID SEARCH

→ RANDOM SEARCH

≡ HYPER OPT

K-FOLD-CROSS-VALIDATION!

2º

ALGORITMOS

- KNN - "the KNN test"
- Random forest - una primera aproximación.

- XGBOOST - la puta,  
"muchos xgboost chiquitos son  
mejores que un xgboost grande"  
"ensemble??"

CATEGORIAS NAN

CIUDAD

PROVINCIA

TIPO DE PROPIEDAD

→ BORRAMOS

SOOFILAS  
APPOX.



# COLUMNS:

HOJA N°

FECHA

- ID → Si, para identificación propiedad
- TÍTULO → Si, PARSARLO (y buscar features)
- DESCRIPCION → Si, PARSARLO (y buscar features)
- DIRECCION → NO
- CIUDAD → Si, CATEGORICA
- PROVINCIA → Si, CATEGORICA? por ahí alcanza con ciudad.
- LAT → NO
- LONG → NO
- ID ZONA → NO
- TIPO DE PROPIEDAD → Si, CATEGORICA
- METROS TOTALES → Si, NUMERICA
- METROS CUBIERTOS → Si, NUMERICA
- ANTIGÜEDAD → Si, NUMERICA
- HABITACIONES → Si, NUMERICA
- GARAJES → Si, NUMERICA
- BAÑOS → Si, NUMERICA
- FECHA → Si, solo el año. **PROBAR CATEGORICA!**
- GIMNASIO → Si, BOOLEAN
- USOS MÚLTIPLES → Si, BOOLEAN
- PISCINA → Si, BOOLEAN
- ESCUELAS CERCANAS → Si, BOOLEAN
- CENTROS COMERCIALES CERCANOS → Si
- PRECIO → **SI, EL LABEL!**