

UNIVERSITYHACK 2020®  
DATAATHON



Data ttack

RETO MINSAIT LAND CLASSIFICATION



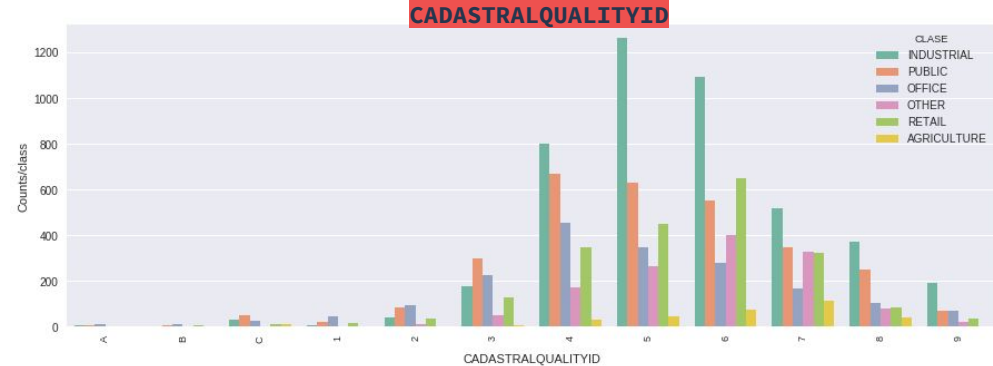
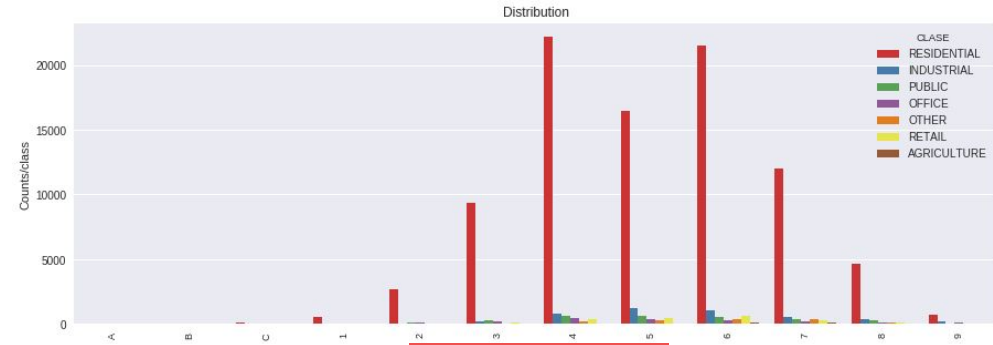
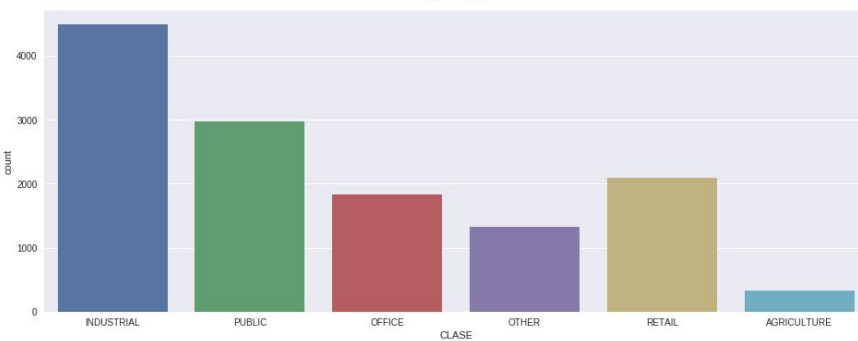
VANESSA NAVARRO



kaggle

minsaIt  
by Indra

idealista/data



# 1. CLASE

# 2. VARIABLES



1. Clase



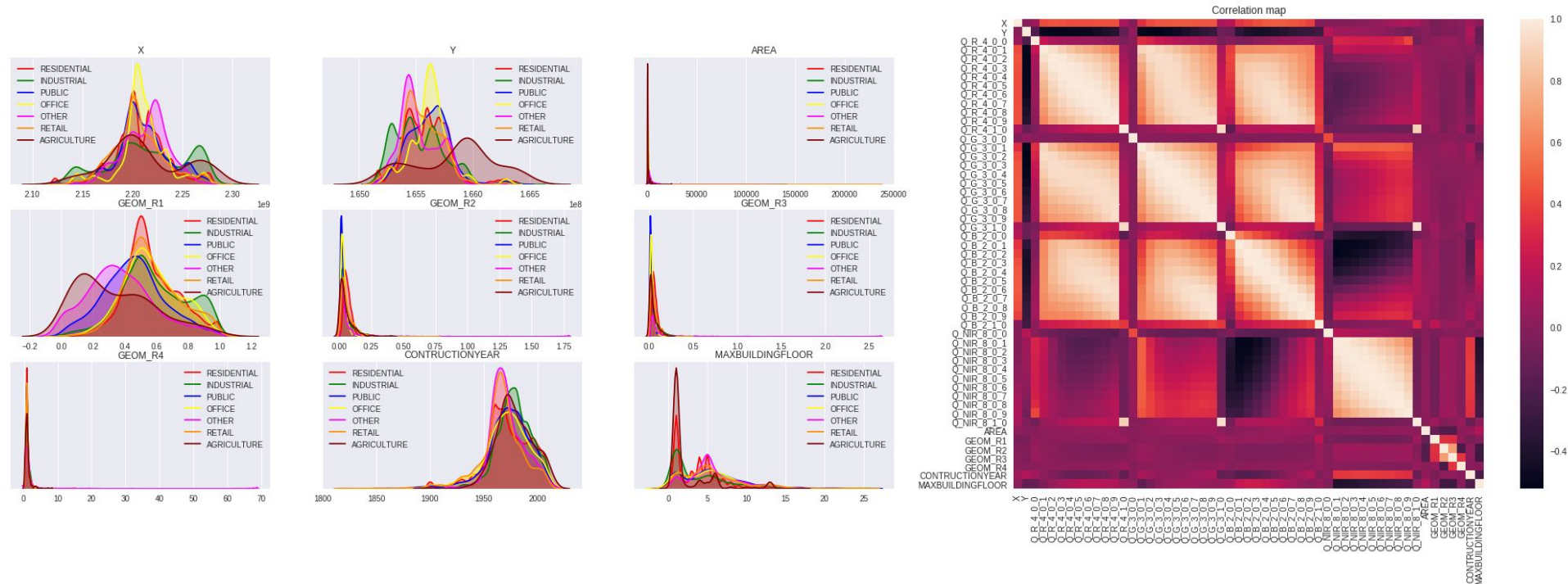
2. Variables



3. Modelos



4. Resultados



## 2. VARIABLES

1. Clase

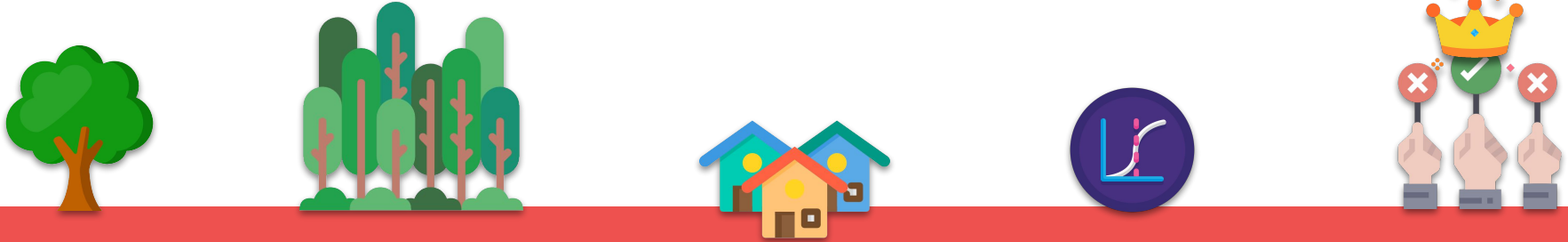
2. Variables

3. Modelos

4. Resultados

- Tratamiento de valores perdidos.
- Holdout para Validación.
- Pipelines:
  - SimpleImputer
  - StandardScaler
  - OneHotEncoder
  - ColumnTransformer
- GridSearchCV -> Mejores combinaciones de hiperparámetros.
- SelectKBest -> Mejores subconjuntos de variables.

PREPROCESAMIENTO



## 3. MODELOS



FAIL

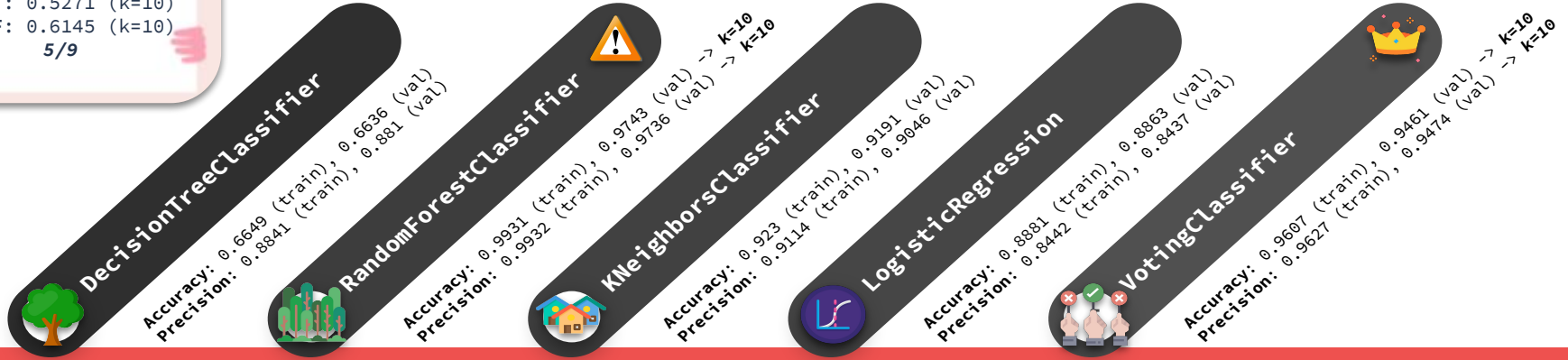
#### TEST

Entrega intermedia:

- DT: 0.5271 (k=10)
  - RF: 0.6145 (k=10)
- 5/9

## Conclusiones:

- El conjunto de estimación es muy diferente al de modelado en cuanto a la distribución de la variable clase.
- “Lo más sencillo es lo más inteligente.”
- Trabajo futuro: NN.



## 4. RESULTADOS

1. Clase

2. Variables

3. Modelos

4. Resultados

THE END