

Salesforce Predictive Modelling

Equipo WorkingData



UNIVERSITYHACK 2018[®] DATAATHON



wefferent

Deloitte.

minsait
by Indra

VIEWNEXT
AN IBM SUBSIDIARY

Hewlett Packard
Enterprise

kabel



ugr

Universidad
de Granada

¿Quiénes somos?



**DAVID JIMÉNEZ
PAREDES**

PREDICTIVA



RAÚL LÓPEZ JIMÉNEZ

LUCA

Telefonica DATA UNIT



**FRANCISCO JAVIER
VÍLCHEZ TORRALBA**

raona

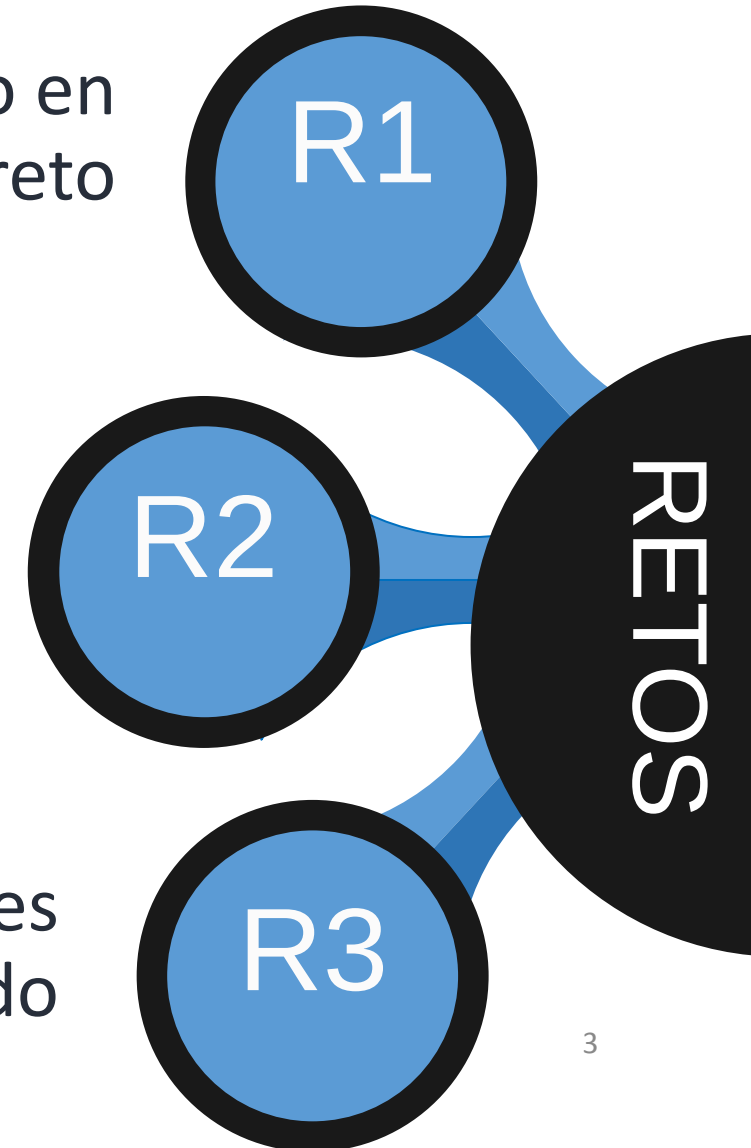
¿Por dónde empezar?



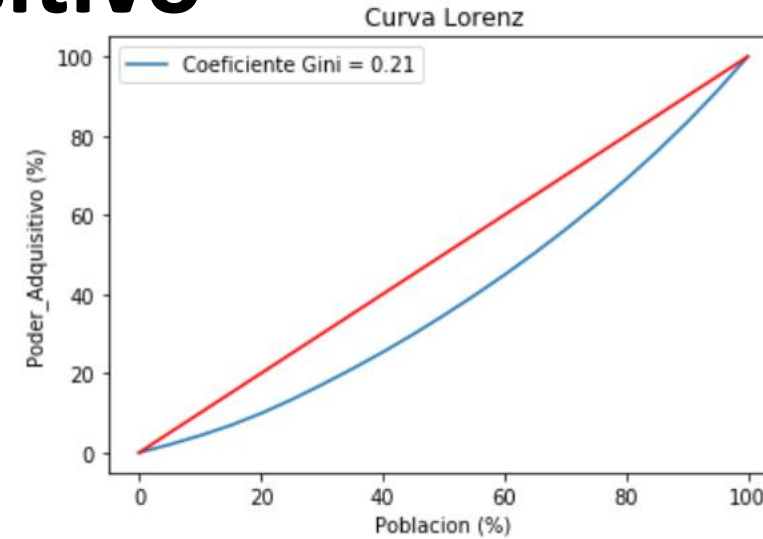
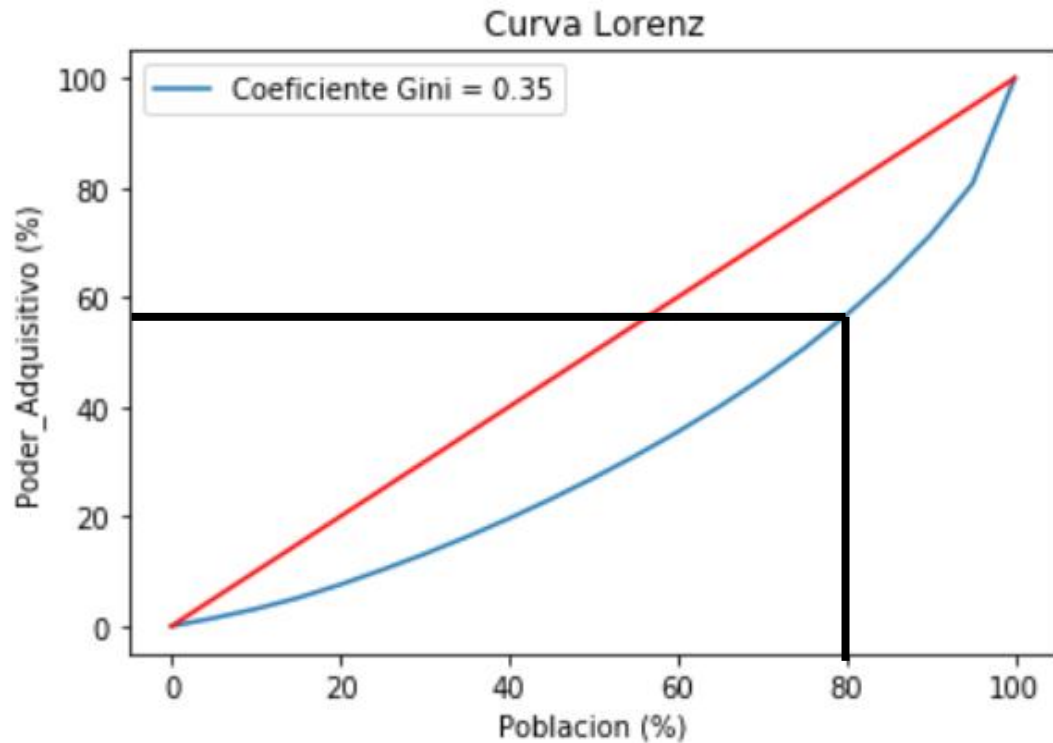
Desconocimiento teórico en
el ámbito asociado al reto

Conjunto de datos con una
elevada dimensionalidad

Contexto de las variables
desconocido



Distribución del poder adquisitivo



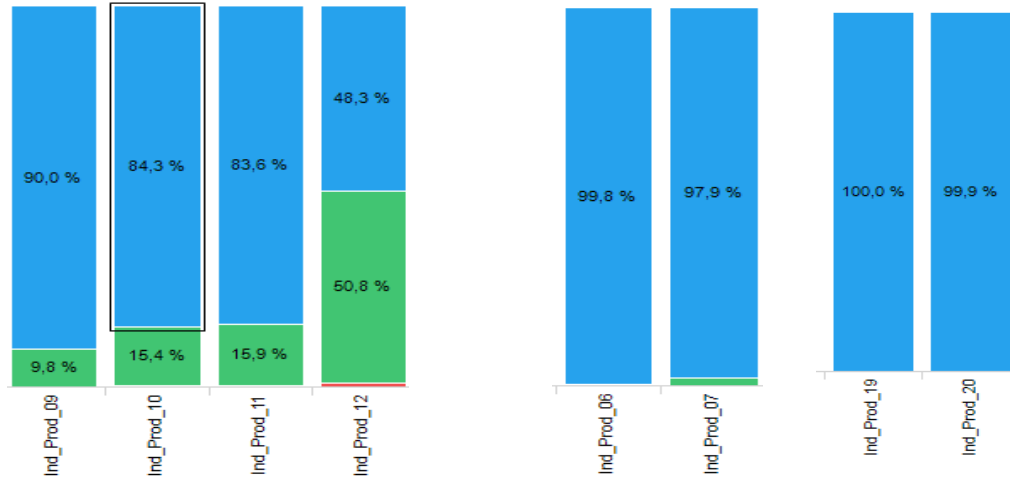
Clase Baja



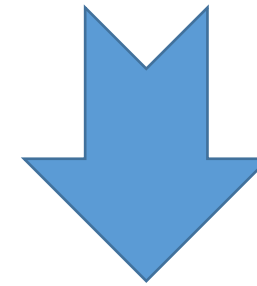
Clase Alta

¿Afecta la tenencia de productos?

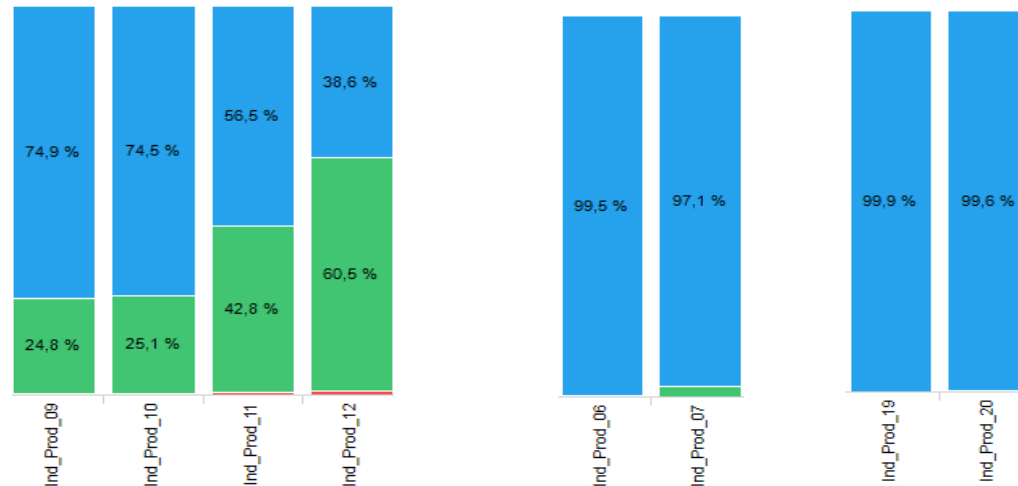
CLASE ALTA



Misma distribución de tenencia de productos en ambas clases económicas

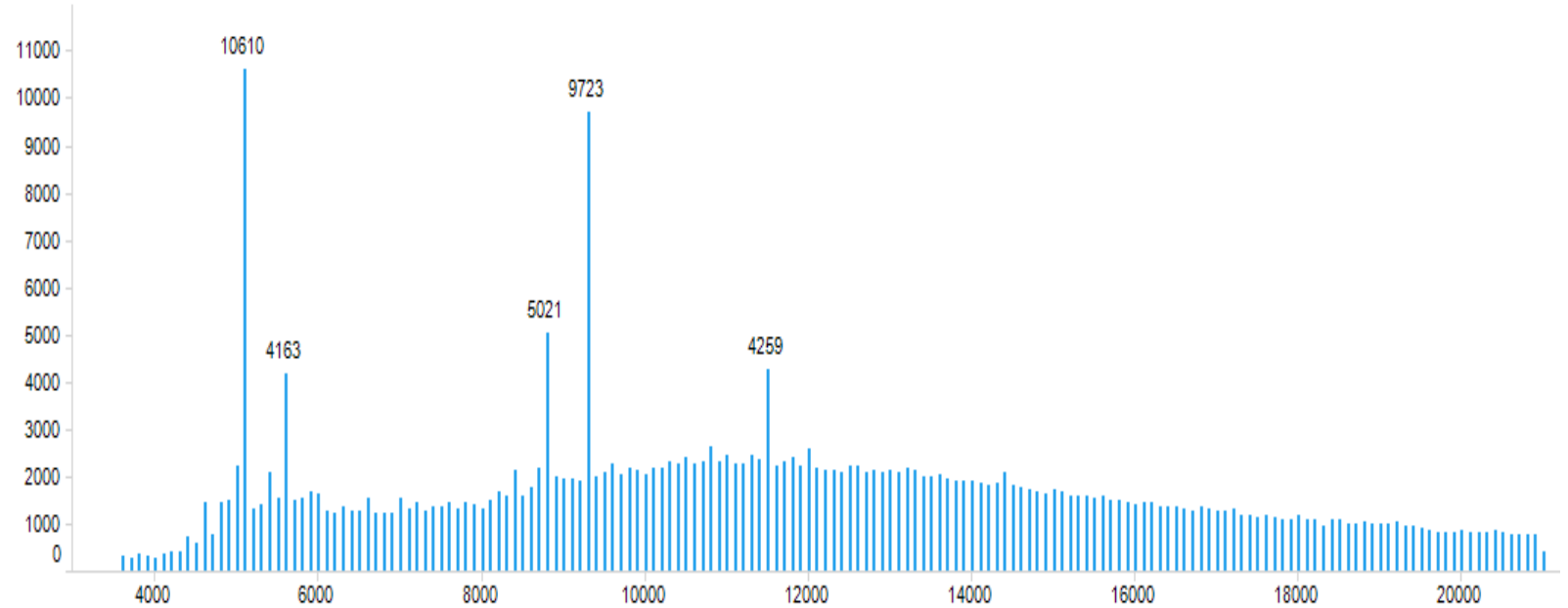


CLASE BAJA



SE PUEDEN ELIMINAR,
REDUCCIÓN DE
DIMENSIONALIDAD

Trabajando con clase “Baja”



**¿QUÉ SON
ESOS PICOS?**



¿VARIAS POBLACIONES JUNTAS EN EL DATASET?



¿TIENEN ALGO EN COMÚN?

Modelos

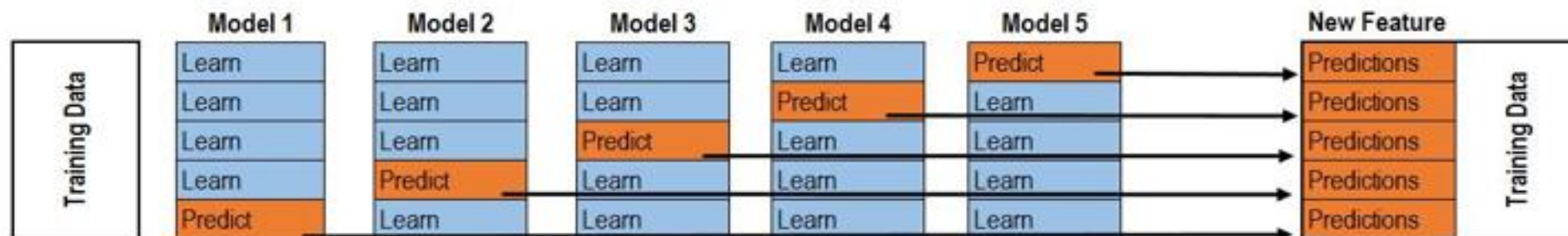
Regresión lineal *Lasso*



Light GBM, Gradient Boosting Regression, XGBoost...



Modelos de apilamiento/*Stacking Models*



Evaluación del modelo

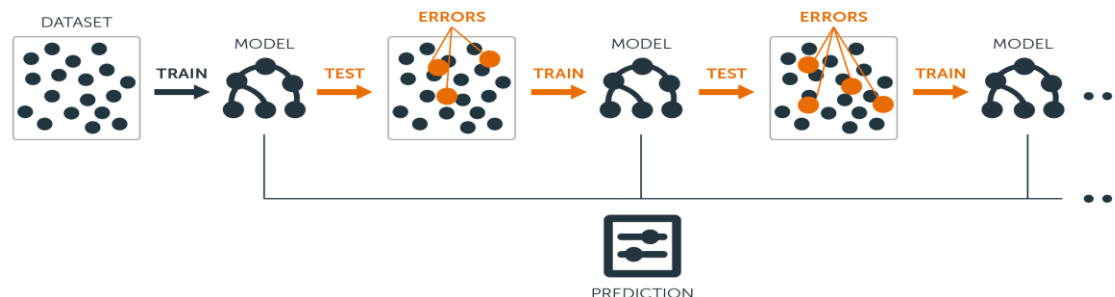
Medida de evaluación → *Root Mean Squared Logarithmic Error* (RMSLE)

Validación Cruzada → K_fold = 5

Tuning parámetros por cada modelo

Mejor modelo → Modelo de apilamiento

$\text{RMSLE}(\text{Train}) = 0.2579$



Herramientas Utilizadas

TIBCO
Spotfire®

dmlc
XGBoost

pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

scikit
learn
contrib



NumPy

matplotlib

Bokeh



UNIVERSITYHACK 2018®
DATA**THON**

La competición de analítica de datos
más grande de toda España.

Muchas gracias.