

18 octubre 2023

Modelos Predictivos

Procedimiento AutoML Abanca

Modelos Predictivos. Inteligencia de Clientes

AutoML: Origen modelo AutoML

Objetivo:

Creación de un procedimiento que de manera interactiva e intuitiva que permita cubrir todos los pasos necesarios para la creación y puesta en producción de un modelo de calidad con el fin de agilizar y eficientar la construcción de éstos.

<https://plotly.com/r/>



<https://pkgs.rstudio.com/flexdashboard/articles/examples.html>



AutoML: Ciclo de Vida



Lectura del tablón, estudio univariante de las variables, **eliminación de variables correlacionadas y poco importantes**.



Entrenamiento y selección automatizada de los modelos ML con los datos depurados.



Creación de un .html con la explicabilidad del modelo.

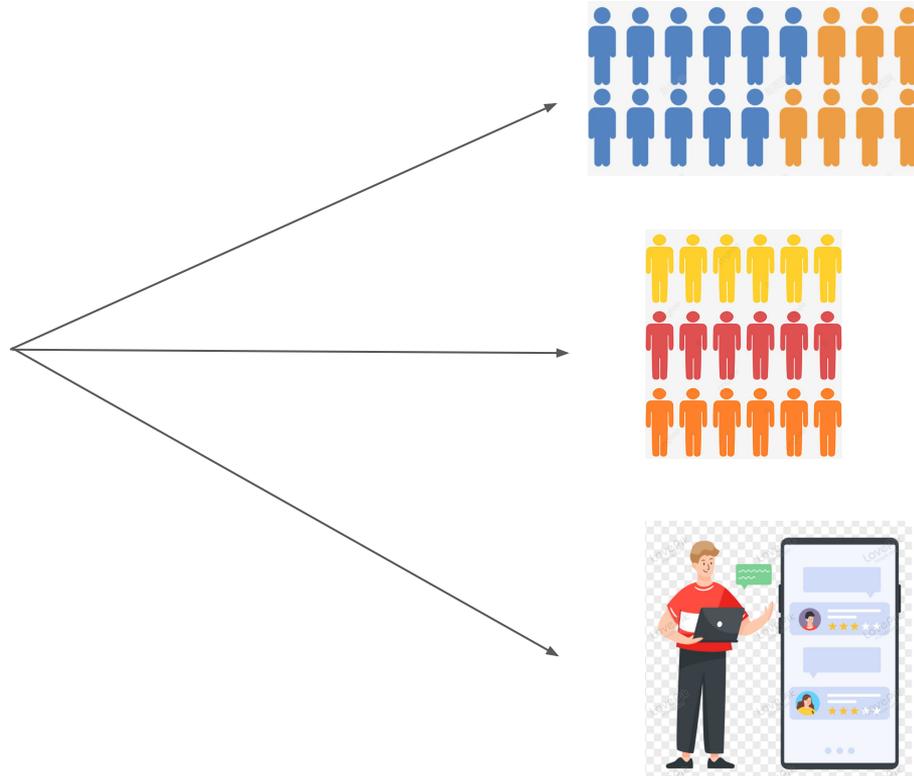


Puesta en producción del modelo seleccionado.



Seguimiento de los modelos que tenemos **en producción**.

AutoML: 3 tipos de variable objetivo



AutoML: Ejemplo introducir parámetros

```
---  
title: Ejecucion de los modelos  
editor_options:  
  chunk_output_type: console  
output:  
  flexdashboard::flex_dashboard:  
    orientation: rows  
    vertical_layout: fill  
params:  
  texto: 'CR7'  
  password:  
    value: "Siuuuuu"  
    input: password  
  rueda:  
    label: "Numero de pruebas"  
    value: 10  
    input: slider  
    min: 0  
    max: 500  
  seleccion:  
    label: "Entrenamiento del modelo:"  
    value: 'Si'  
    input: select  
    choices: ['Si', 'No']  
  
  tick : FALSE  
  
  insercion_fichero:  
    label: "Seleccion de hiperparámetros GBM:"  
    value:  
    input: file  
---
```

AutoML: Ejemplo introducir parámetros

Knit with Parameters

texto
CR7

password

Numero de pruebas
10 500
0 50 100 150 200 250 300 350 400 450 500

Entrenamiento del modelo:
Si

tick

Selección de hiperparámetros GBM:
Browse... No file selected

Cancel Knit

AutoML: Introducción parámetros procedimiento

Knit with Parameters

tabla
MKTG_USR.dataset_prueba

directorio
//cgefiles95/apps5/cge9/cag-dep4/KRSS/KRSS001D_RStudio/Modelo_prueba

usuario
12345

password

modelo
prueba

variable_objetivo
DIABETES

Numero de pruebas
0 50 100 150 200 250 300 350 400 450 500

Cancel Knit

AutoML: Introducción parámetros procedimiento

Knit with Parameters

Lectura de datos:

Ejecucion graficas:

Selección de Variables:

Entrenamiento del modelo:

- GBM
- RF
- MLP
- XGBoost
- LightGBM
- Bayes
- AdaBoost

AutoML: Introducción parámetros procedimiento

Knit with Parameters

Balanceo del modelo:
No

Selección de hiperparámetros GBM:
Browse... No file selected

Selección de hiperparámetros RF:
Browse... No file selected

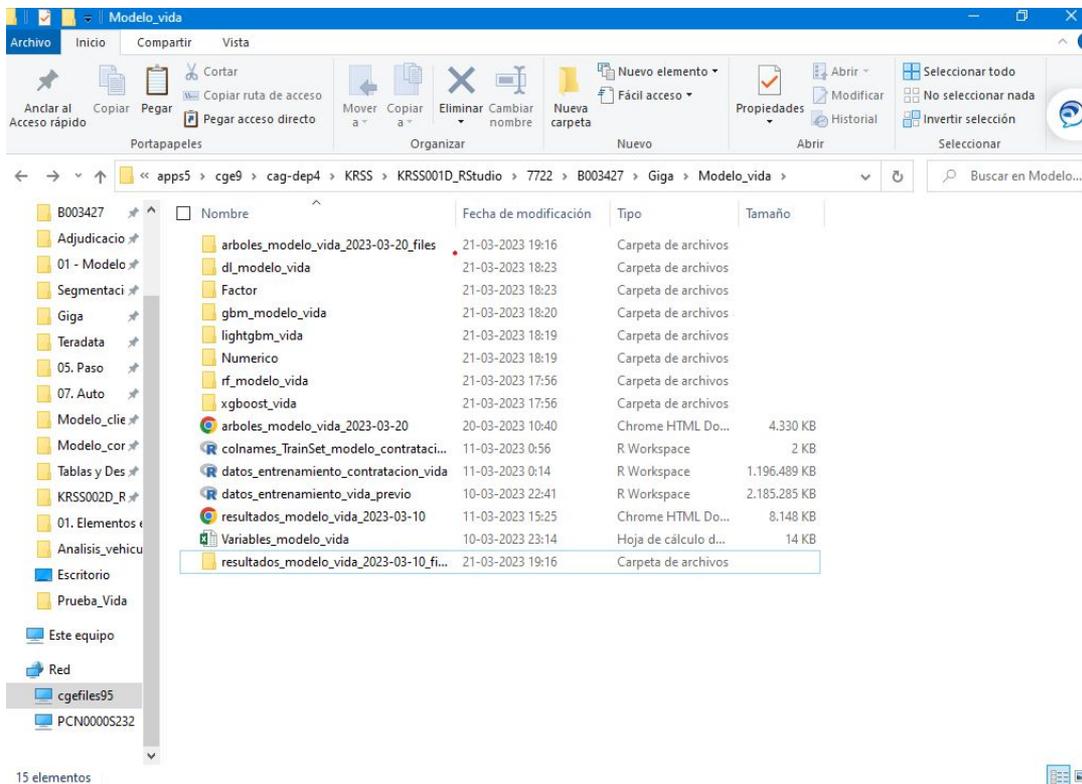
Selección de hiperparámetros MLP:
Browse... No file selected

Selección de hiperparámetros XGBoost:
Browse... No file selected

Selección de hiperparámetros LightGBM:
Browse... No file selected

Cancel Knit

AutoML: Generación de archivos



AutoML: Ejemplo creación de un html

```
---
title: Ejemplo de prueba
editor_options:
  chunk_output_type: console
output:
  flexdashboard::flex_dashboard:
    orientation: rows
    vertical_layout: fill
---

<!-- Añadimos sección / pestaña mediante el = -->

Sección 1
=====

<!-- Añadimos filas mediante Row -->

Row
-----

<!-- Nombre del gráfico mediante ### e inserción de código mediante ```{r} -->

### Gráfico de prueba

```{r}
x<-c(1,2,3)
y<-c(1,2,3)
plot(x,y)
```

Row {data-height=100}
-----

### Gráfico de prueba lateral

```{r}
x<-x+1
plot(x,y,col="blue")
```

### Gráfico de prueba contiguo

```{r}
y<-y+1
plot(x,y,col="red")
```

Sección 2
=====

Row
-----
```

Gráfico de prueba

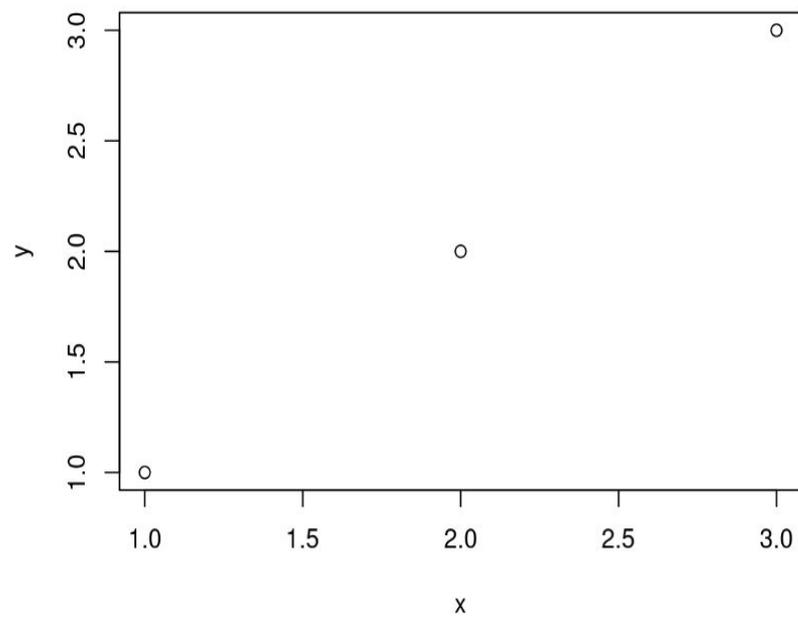


Gráfico de prueba lateral

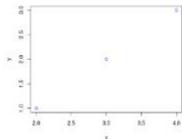
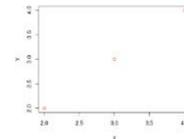


Gráfico de prueba contiguo



362

Clientes totales en el entrenamiento



65

Número de Éxitos



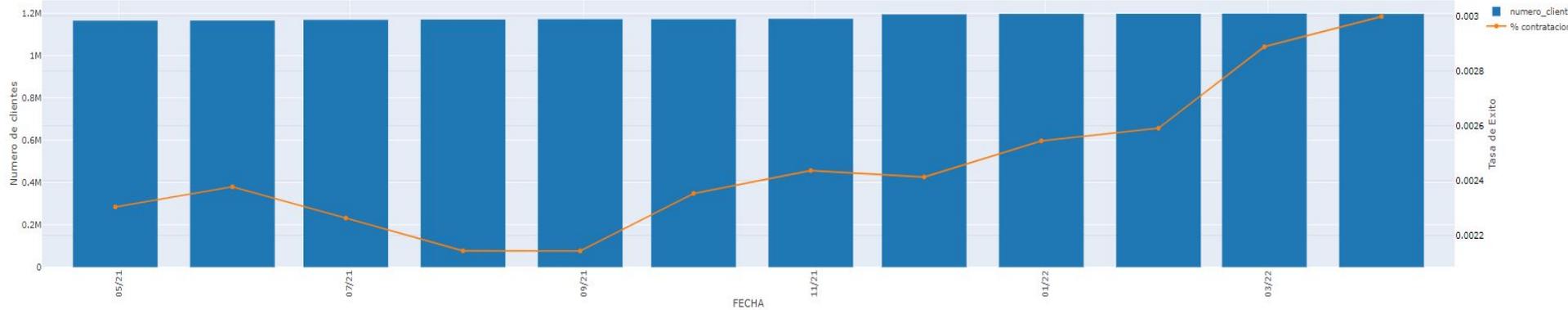
296

Número de Variables

Resumen de los datos

| Fecha | Número de clientes | Número de éxito | Tasa de éxitos |
|------------|--------------------|-----------------|----------------|
| 2021-05-31 | 302 | 93 | 11 |
| 2021-06-30 | 230 | 03 | 12 |
| 2021-07-31 | 834 | 23 | 12 |
| 2021-08-31 | 839 | 17 | 11 |
| 2021-09-30 | 018 | 01 | 13 |
| 2021-10-31 | 193 | 90 | 12 |
| 2021-11-30 | 839 | 02 | 11 |
| 2021-12-31 | 394 | 30 | 10 |
| 2022-01-31 | 938 | 21 | 11 |
| 2022-02-28 | 394 | 89 | 12 |
| 2022-03-31 | 948 | 49 | 11 |
| 2022-04-30 | 738 | 12 | 12 |

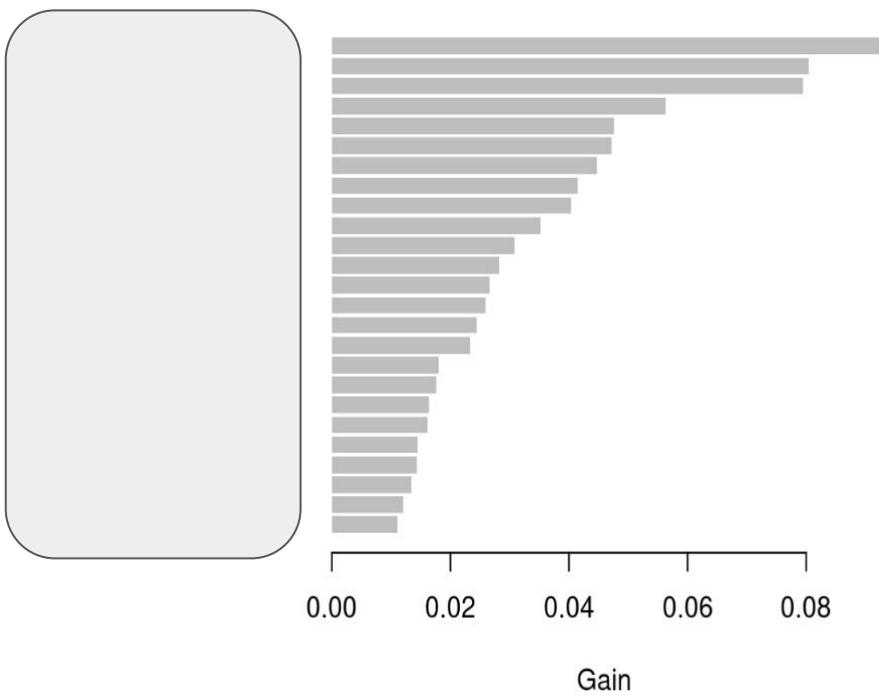
Evolución de los datos





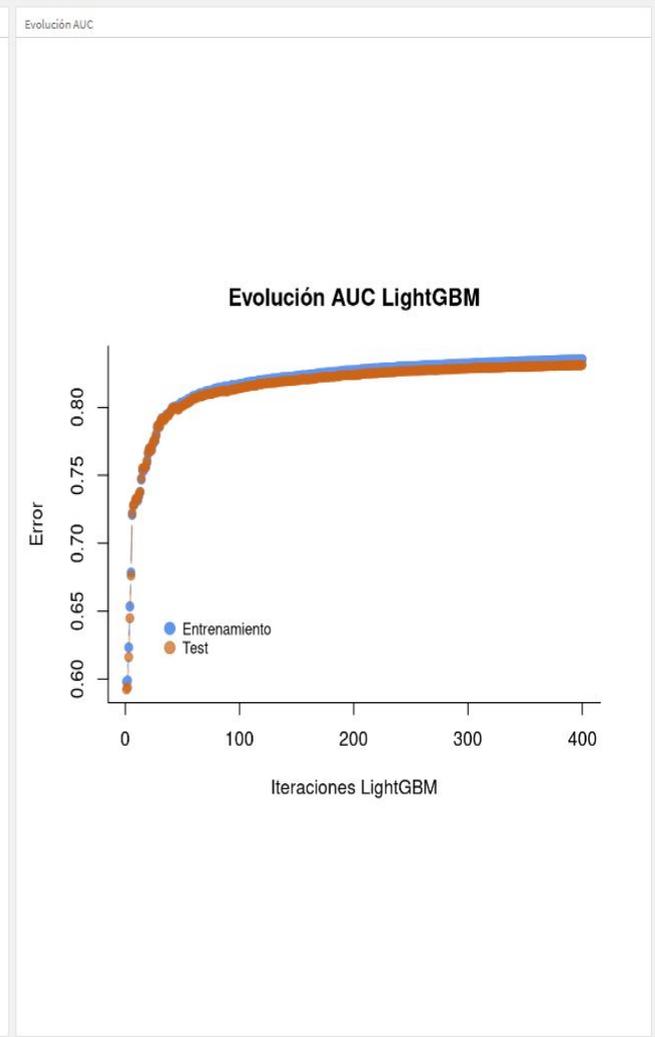
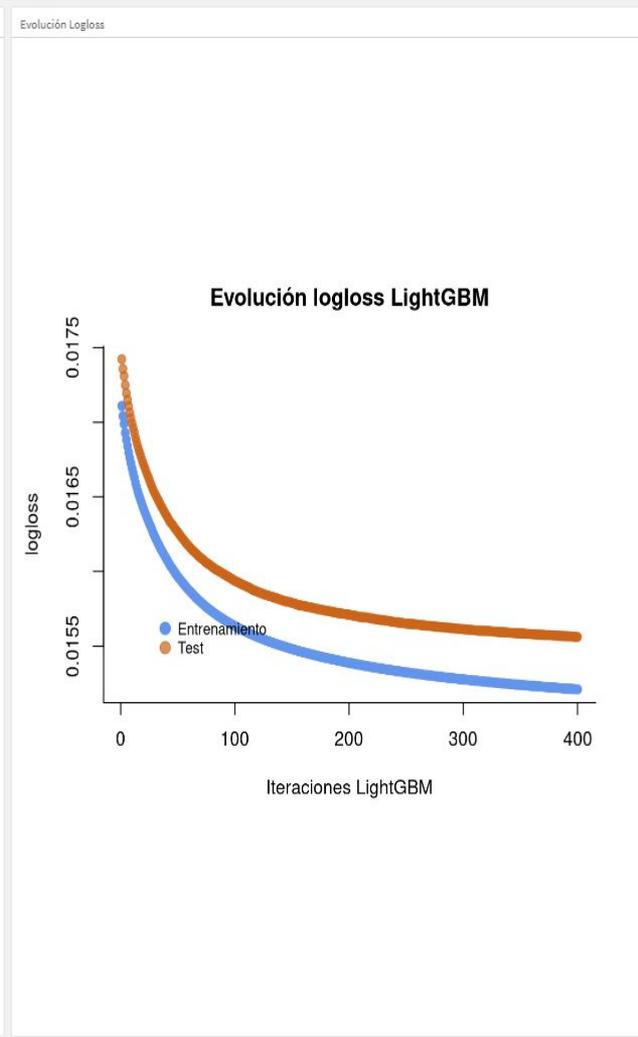
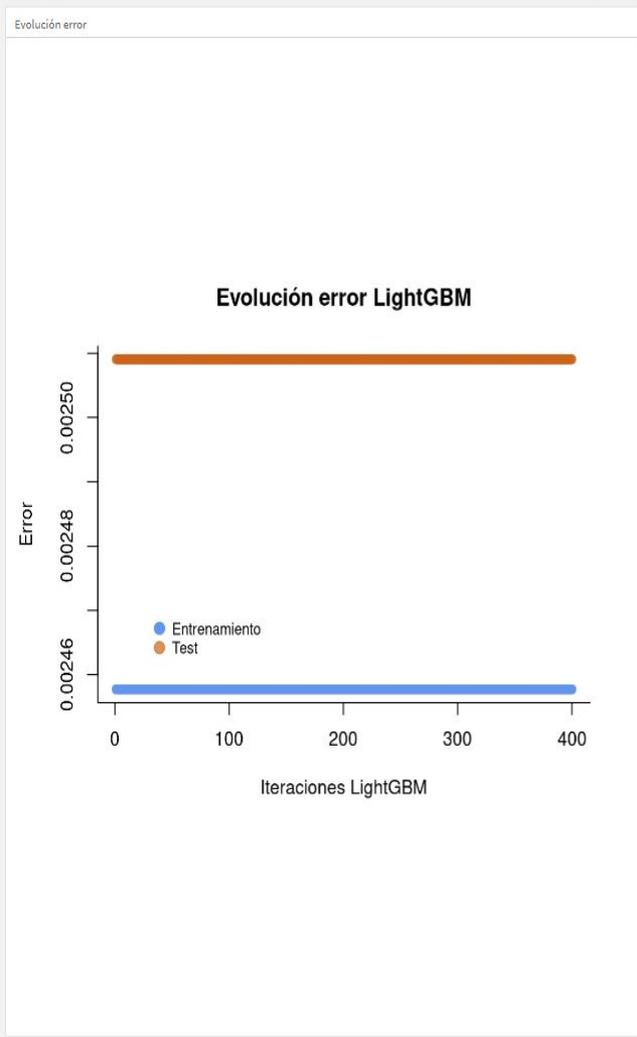
Importancia de las variables

Feature Importance

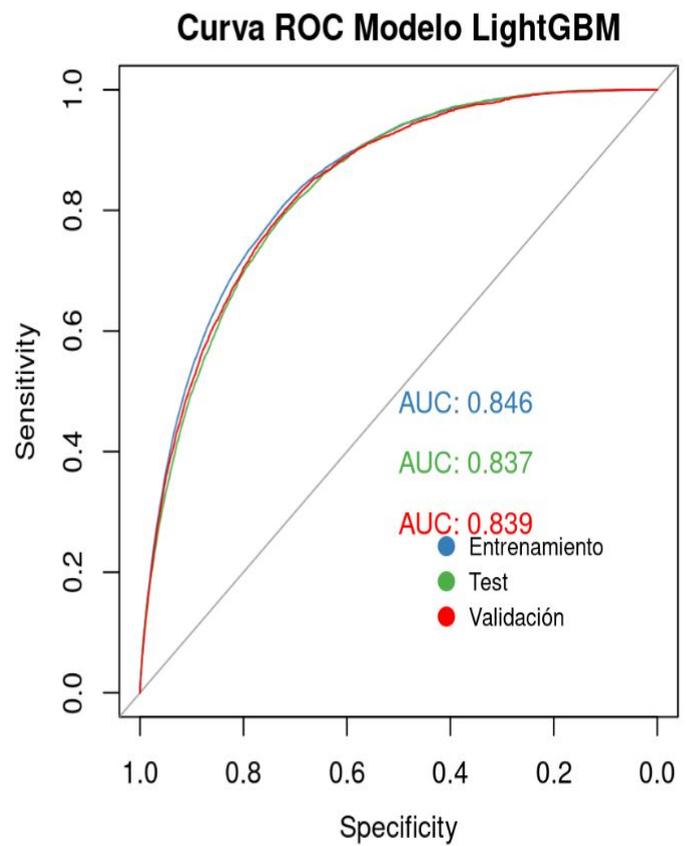


| Hiperparámetros mejor | |
|-----------------------|--------------------|
| num_iterations | 132 |
| learning_rate | 0.0172230931935829 |
| min_gain_to_split | 10.58348202391310 |
| num_leaves | 24140 |
| lambda_l1 | 5.92841924617410 |
| lambda_l2 | 4.47618801845238 |
| max_depth | 3 |
| feature_fraction | 0.501430464356672 |
| colsample_bylevel | 0.849237202685326 |

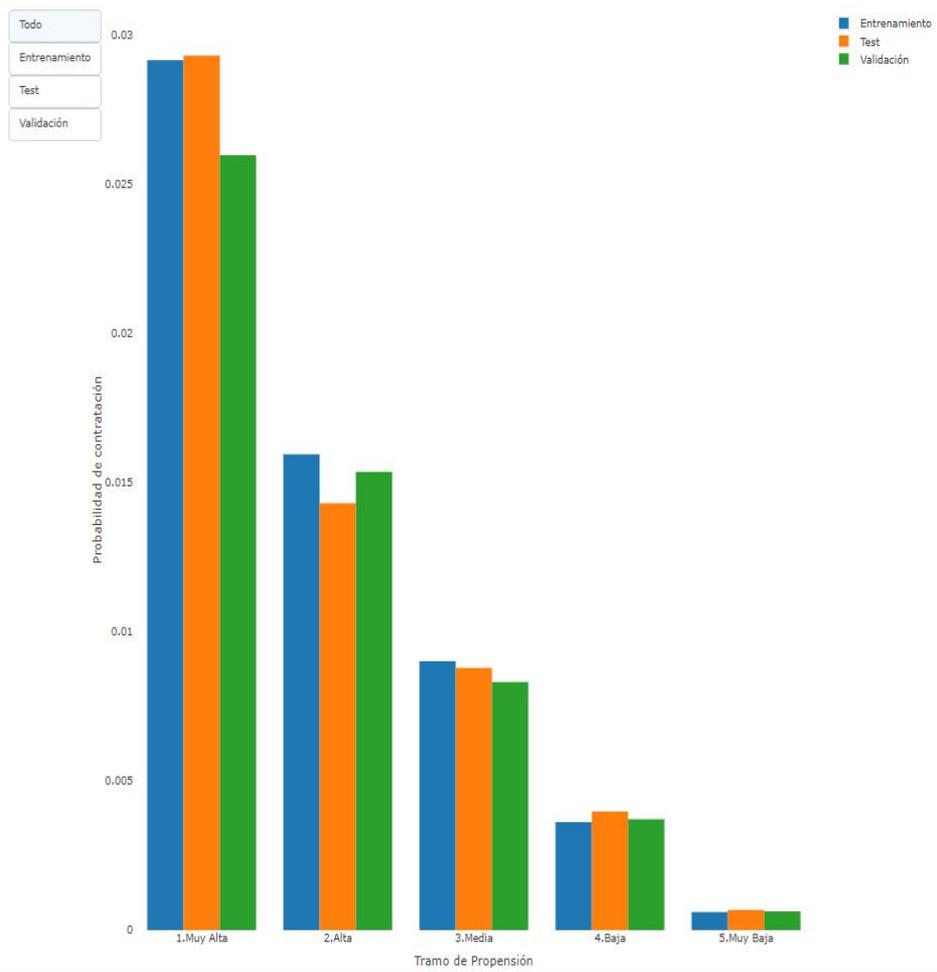
| Resultados mejor | | | | | |
|------------------|-----------|------------------|----------------------|---------------------|------------------|
| | AUC | Tasa_Exito_nueva | Tasa_Exito_Propensos | Tasa_Exito_Conjunto | Eficiencia_nueva |
| train_resultado | 0.8496243 | 0.01135621 | 0.01236719 | 0.003264023 | 5.230452 |
| test_resultado | 0.8432982 | 0.0135271 | 0.0112342 | 0.002509059 | 5.112396 |
| valid_resultado | 0.848872 | 0.01225515 | 0.01363831 | 0.00211237 | 5.108097 |



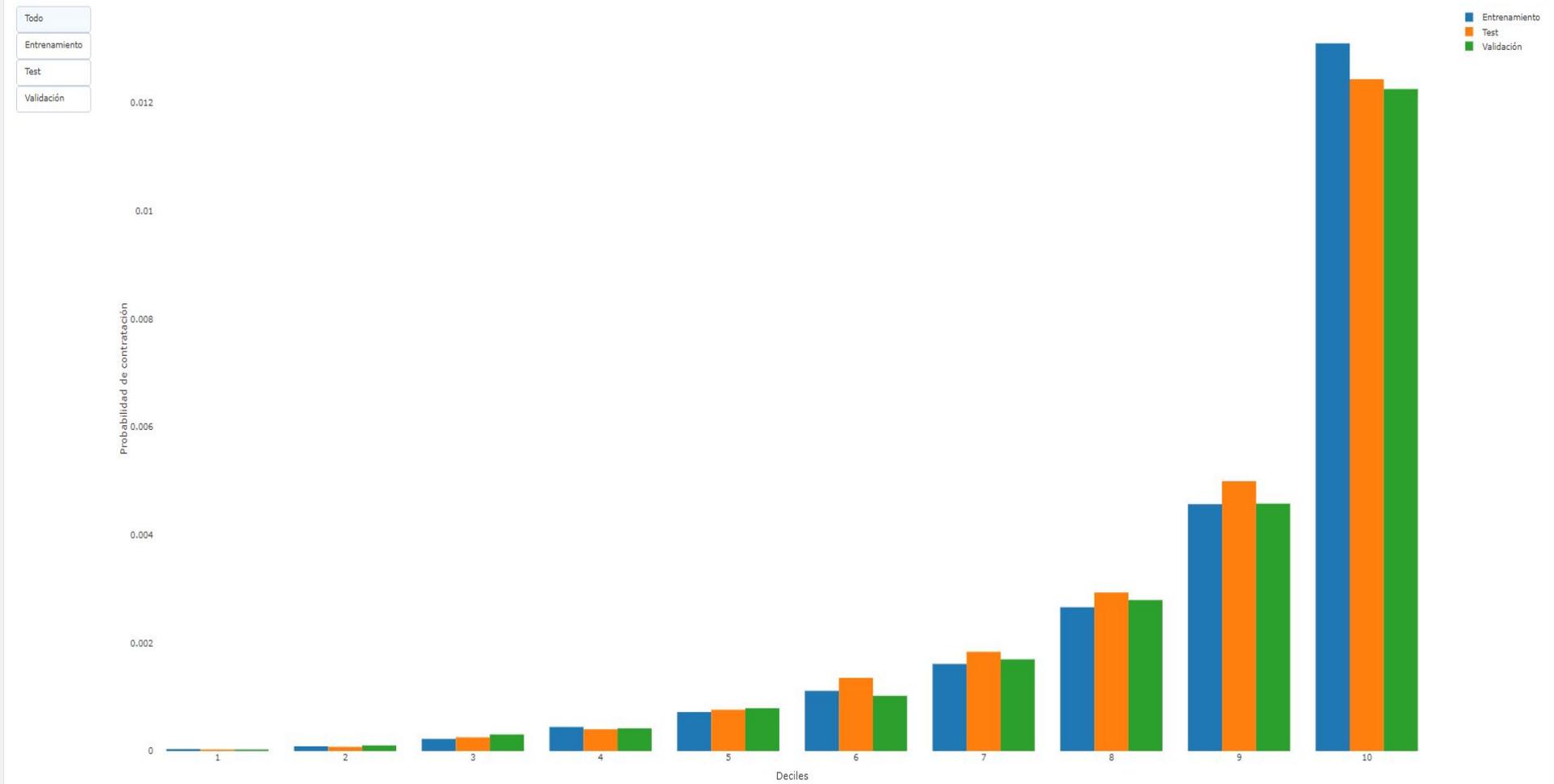
Curva ROC



Tramos de propensión



Deciles



Ventajas:

1. Herramienta que nos permite **agilizar sustancialmente el tiempo de construcción** de los modelos predictivos.
2. **Fácil de usar y de entender** para el resto de usuarios.
3. Al estar integrada con Gitlab podemos **programar las ejecuciones** de este para que se ejecute en momentos **donde la carga computacional es menor**.
4. Es una herramienta modular y fácilmente modificable para futuras necesidades, permitiendo la **mejora continua** reduciendo con ello el tiempo de espera ante posibles mejoras/soluciones.
5. Permite la **governabilidad de los modelos** de modo que cuando se note un deterioramiento en el modelo se pueda reentrenar de manera automática.
6. Gran **adaptabilidad a multitud de casos**, no sólo al ámbito bancario.
7. Estas librerías permiten la **creación de reportes automáticos y dinámicos usando R**.

¡Muchas Gracias!