

# Modelamiento predictivo para asignación de cupos de MoveOn

Alejandro García Flores, Brayann Y. A. Quevedo Benito, Fabián C. Castellanos Pinto

{a.garcia13, by.quevedo, fc.castellanos}@uniandes.edu.co

# Contenido

- Análisis exploratorio.
- Construcción del modelo base.
- Modelos experimentales.
- Conclusiones preliminares.

# Pautas principales

## Contexto

Los cupos no se asignan correctamente en los convenios por baja o alta demanda.

## Objetivo

Construir un modelo de RL que prediga el número óptimo de cupos de convenios en un semestre dado.

## Protección de datos

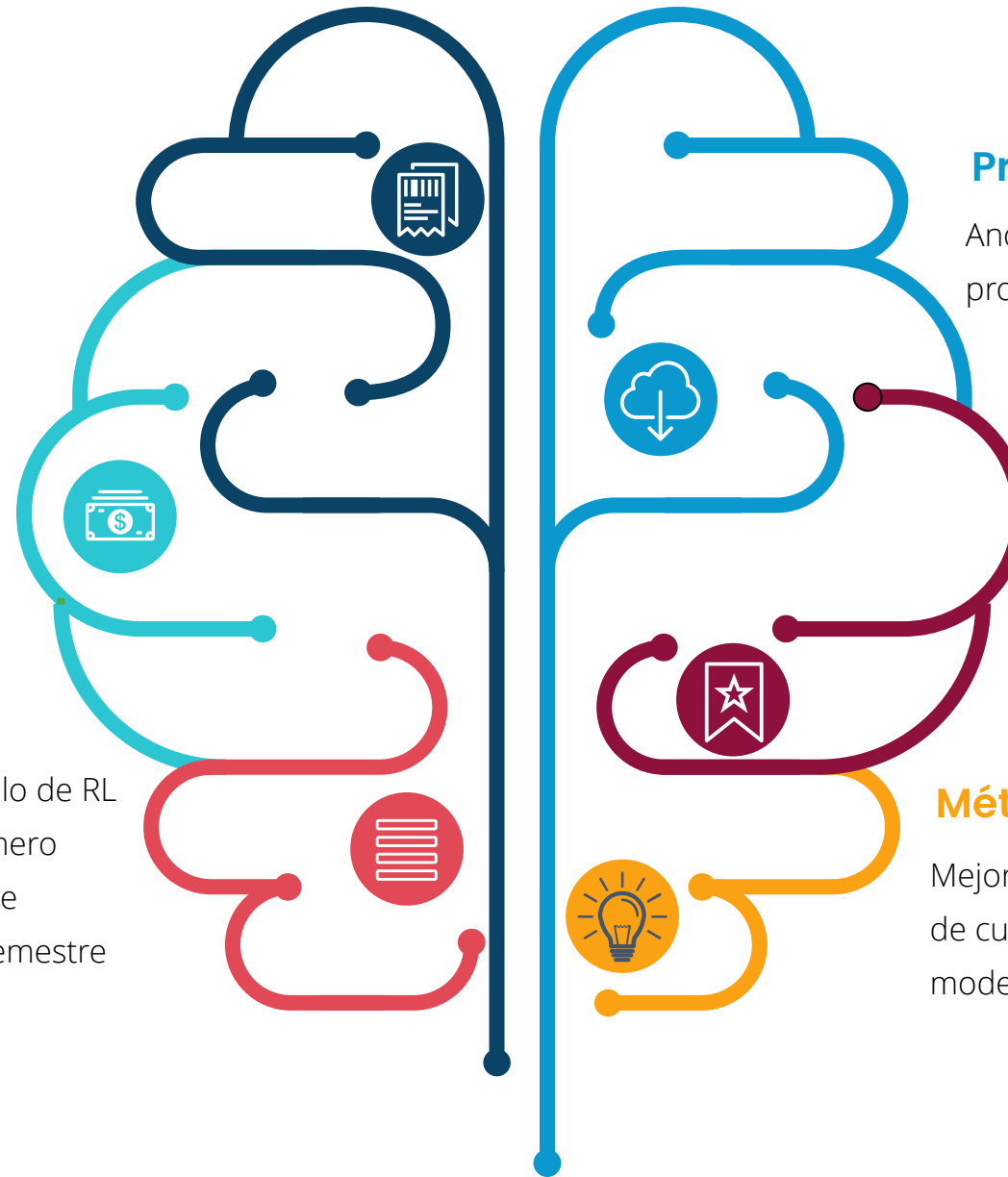
Anonización de datos para proteger datos personales.

## Acotación de alcance

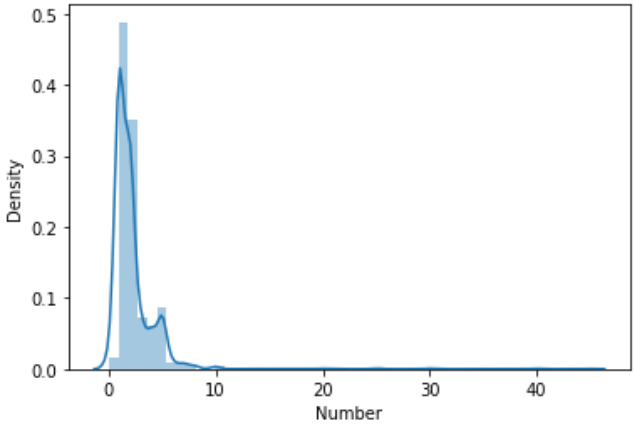
En la anterior entrega no dimos claridad específica para la optimización de asignaciones de cupos.

## Métrica

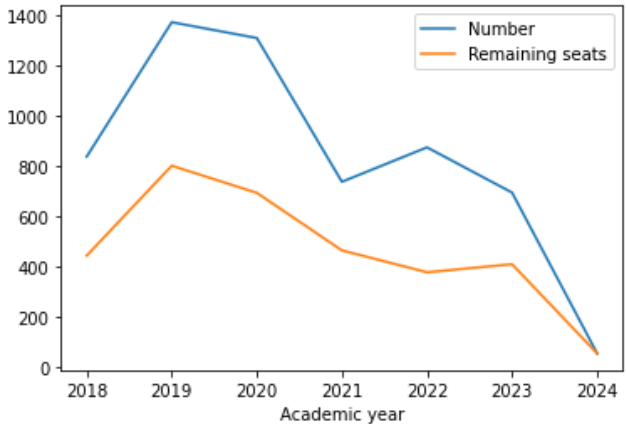
Mejorar el kpi de asignación de cupos mediante el modelo de regresión lineal.



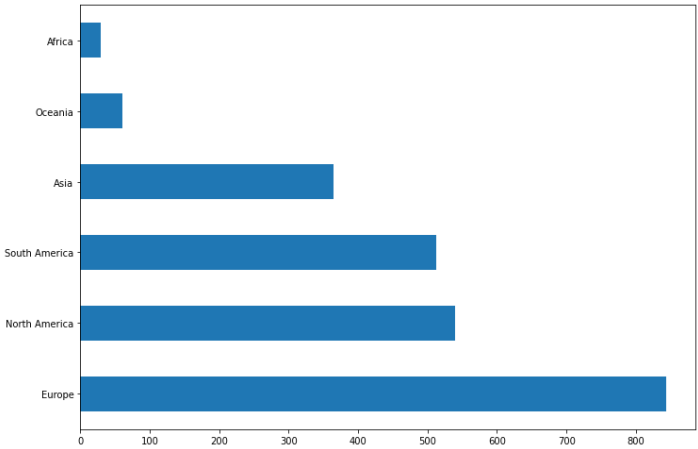
# Análisis exploratorio



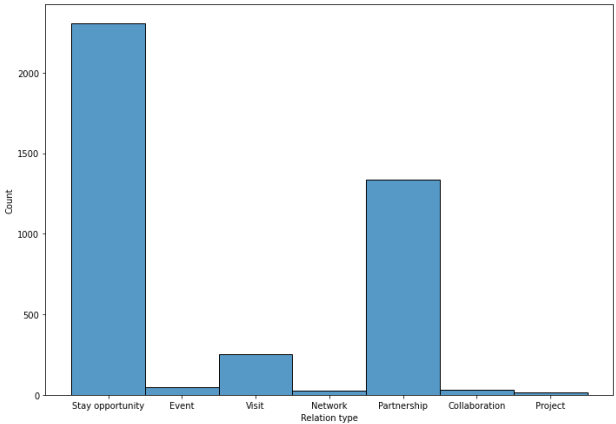
Distribución de cupos



Asignación de cupos



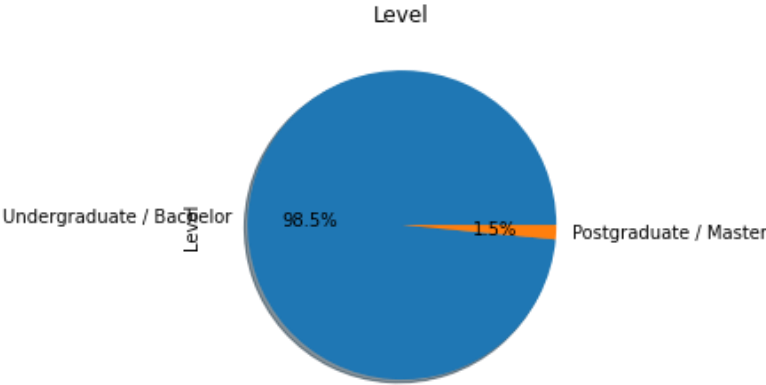
Distribución geográfica de instituciones



Convenios / Oportunidades de intercambio

	Academic year	Number	Remaining seats	Relation: Relation ID
Academic year	1.000000	-0.045528	-0.204352	-0.358787
Number	-0.045528	1.000000	0.792317	0.011644
Remaining seats	-0.204352	0.792317	1.000000	0.050139
Relation: Relation ID	-0.358787	0.011644	0.050139	1.000000

Correlación de variables para asignaciones (Seats)



Nivel académico de aspiraciones (Saty Wishes)

# Construcción del modelo base

## Variables

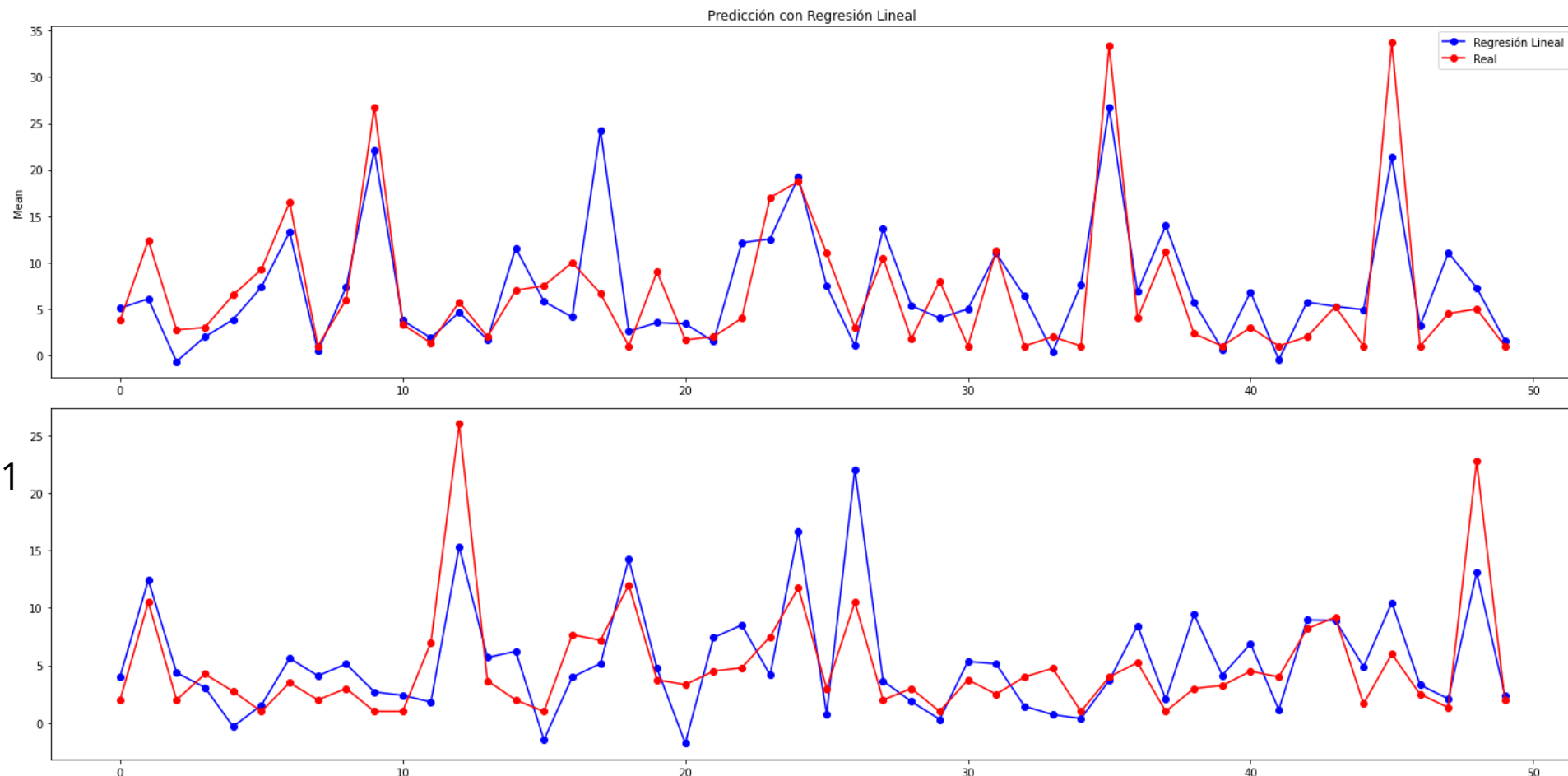
- Country.
- Main External Institutions.
- External institutions.
- Institution Name.
- Relation Type.
- Frameworks.
- Start Date.
- End Date.
- Status.
- Correlation Content.
- Last renewal.
- Degree programme.
- Internal institutions.
- Direction.
- Parent Relation.
- Level.
- Reach

¿Regularización?: No  
One Hot Encoder?: Si  
Variable a predecir: Mean

Total registros: 430  
Datos prueba: 20%

R2 entrenamiento: 0.69011  
R2 prueba: 0.43250

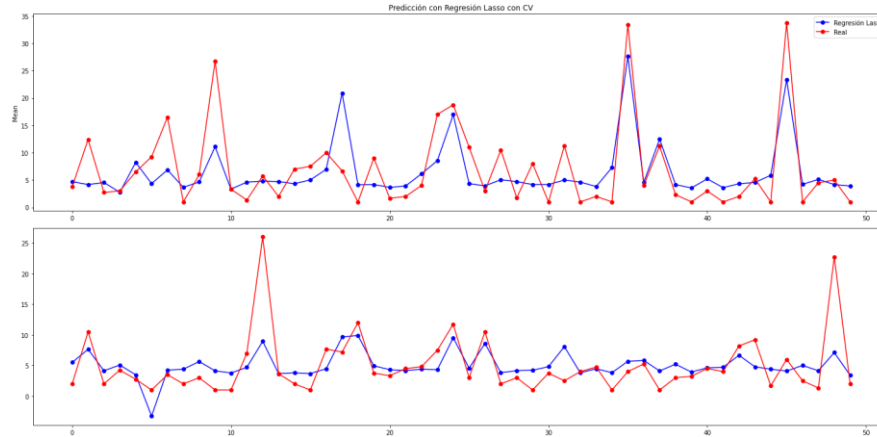
*Overtiffing*



# Modelos experimentales

Usamos Grid search para encontrar los mejores parámetros.

*Polinomial Lasso*

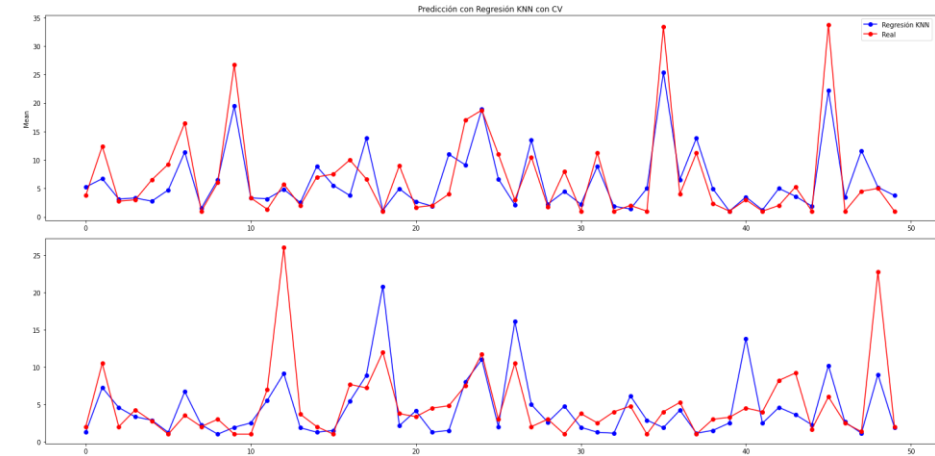


**R2 entrenamiento: 0.69011**

**R2 prueba: 0.43250**

***Overtiffing***

*Nearest Neighbor*

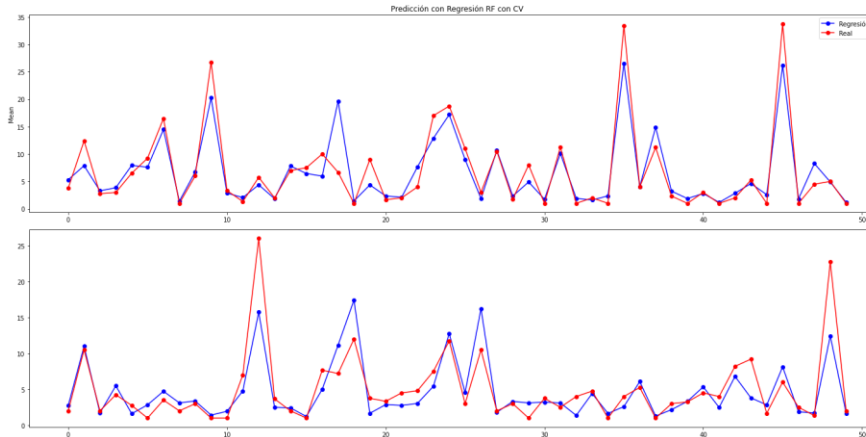


**R2 entrenamiento: 0.46750**

**R2 prueba: 0.43785**

***Overtiffing***

*Random Forest*



**R2 entrenamiento: 0.84211**

**R2 prueba: 0.62212**

***Mejor ajuste***

# Conclusiones preliminares

## MODELO BASE

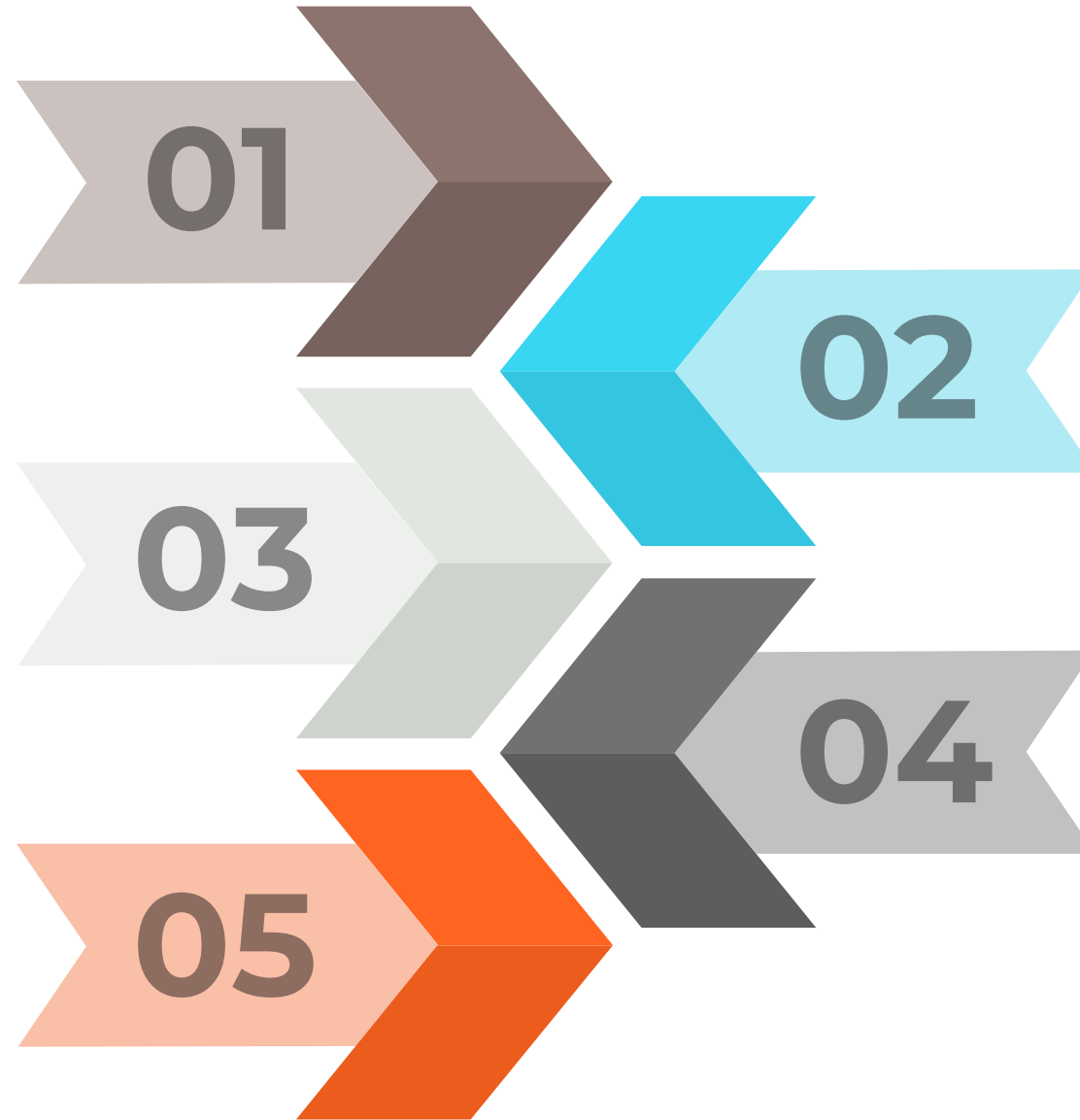
No tiene buena predicción sobre los datos

## AGRUPACIONES

Fue necesario incluir valores que permitieran un mejor procesamiento a la información, p.e. Continente

## Resultados esperados

Las predicciones de los modelos experimentales particularmente en random forest se acercan al objetivo planteado del Proyecto.



## CARDINALIDAD

La institución es una variable importante en la predicción, su cardinalidad es alta, fue necesario reducir el set de datos para procesar.

## LECCIÓN APRENDIDA

Definición de alcance, preguntas de negocio con respecto a la anterior entrega.