

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.



Anonización de datos para proteger datos personales.

Acotación de alcance

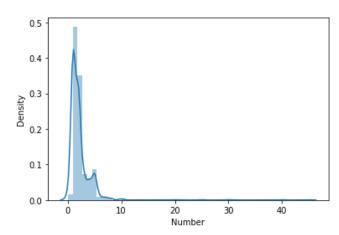
En la anterior entrega no dimos claridad específica para la optimización de asignaciones de cupos.

Métrica

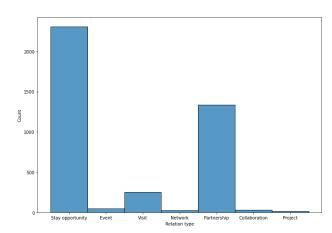
Mejorar el kpi de asignación de cupos mediante el modelo de regresión lineal.

Análisis exploratorio

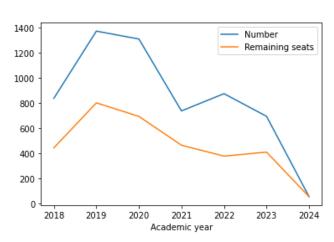




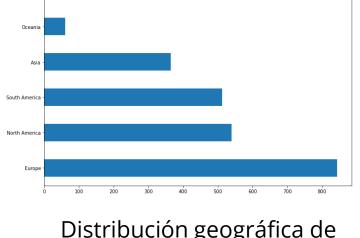
Distribución de cupos



Convenios / Oportunidades de intercambio



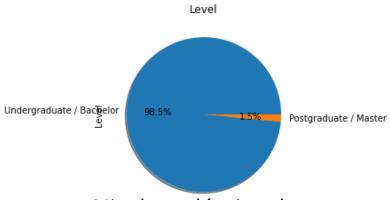
Asignación de cupos



Distribución geográfica de instituciones



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.

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

Total registros: 635

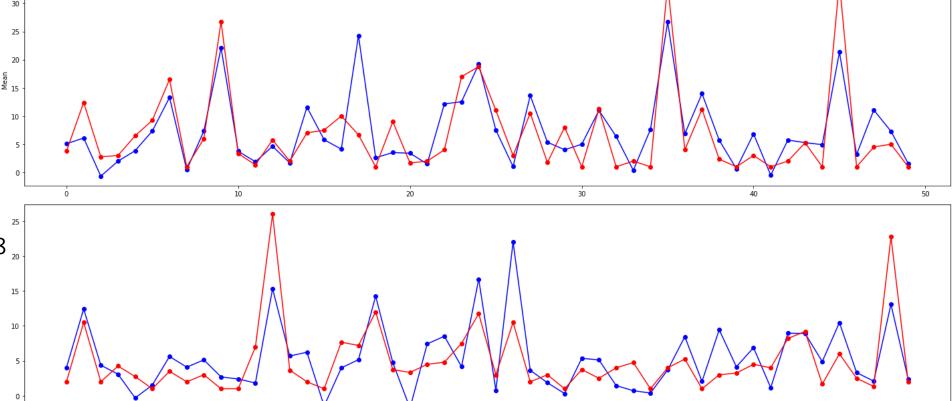
Datos entrenados: 20%

R2 entrenamiento: 0.51018

R2 prueba: 2.17651

Overtiffing

Nota: Los valores de R2 son los aiustados.



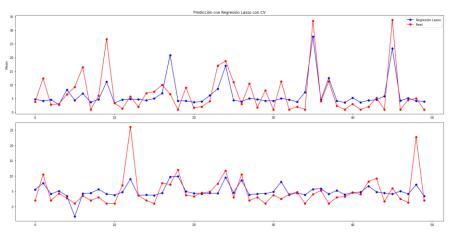
Predicción con Regresión Lineal

Modelos experimentales

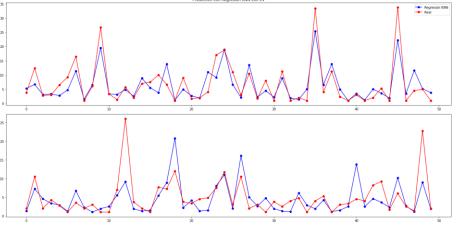


Usamos Grid search para encontrar los mejores parámetros.

Polinomial Lasso



Nearest Neighbor



R2 entrenamiento: 0.15831

R2 prueba: 2.16543

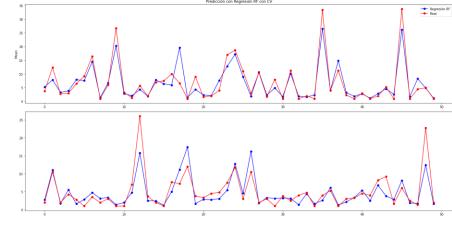
Overtiffing

R2 entrenamiento: 0.49260

R2 prueba: 2.44011

Overtiffing

Random Forest



R2 entrenamiento: 0.75043

R2 prueba: 1.78340

Mejor ajuste

Conclusiones preliminares



MODELO BASE

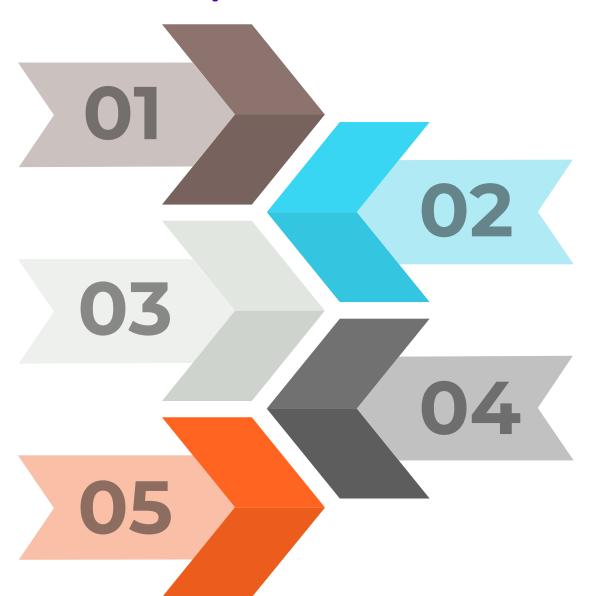
No tiene buena predicción sobre los datos

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 reducer el set de datos para procesar.

LECCIÓN APRENDIDA

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