

01

Presentación

Retención de baja en los siguientes 2 meses

El dataset contiene 155 features y datos desde ene-19 a jul-21



02

Modelo

De árboles a LGBM



03

Optimización

De grid search a bayesian optimization (66 iteraciones, converge)
5-fold CV, jun-21 val y jul-21 test.

Parámetros relevantes:
-num_iterations
-learning_rate
-feature_fraction
-num_leaves
-min_data_in_leaf



04

Primera Aproximación

BO con LGBM de modelo sencillo
6 meses, lags 1, 3, 6

Tratar data drifting por inflación?
No...

Ganancia: cerca de \$110mill.



05

Datos + Procesamiento

Más Meses: Desde 2019, exc pandemia

Feature Engineering Histórico: -Lags 1,3,6

-Delta 1

-Prom. 6 meses

-Tend. 7 meses

-Agrup. saldo caja

-Quitar col. >50% nulos



07

Semillero

Mismo hiperparámetro de la mejor BO de la semilla 1
Promedio de scores de 100 semillas distintas (5 días...)



06

Ganancia del nuevo modelo

Mejora de ganancia: cerca de \$125m
Duración de BO: 6 días con 256 gb de RAM

Mejores hiperparámetros de Semilla 1:
-num_iterations: 1213
-learning_rate: 0.0272068173298842
-feature_fraction: 0.130761948380116
-num_leaves: 3427
-min_data_in_leaf: 43108



08

Conclusiones

Datos+Procesamiento+Reducción de Varianza

Limitaciones: Tiempo... ¿Undersampling?



U1

Presentación

Retención de baja en los siguientes 2 meses

El dataset contiene 155 feautres y datos
desde ene-19 a jul-21



U2

Modelo

De árboles a LGBM



LightGBM

U3

Optimización

De grid search a bayesian optimization (66 iteraciones, converge)
5-fold CV, jun-21 val y jul-21 test

Parámetros relevantes:

- num_iterations
- learning_rate
- feature_fraction
- num_leaves
- min_data_in_leaf



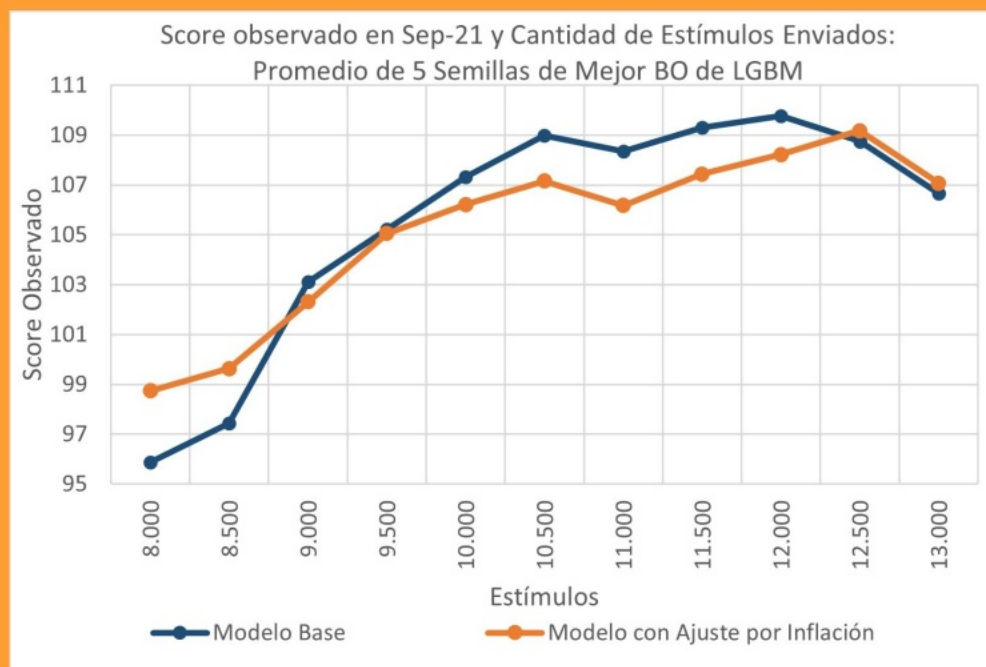
U4

Primera Aproximación

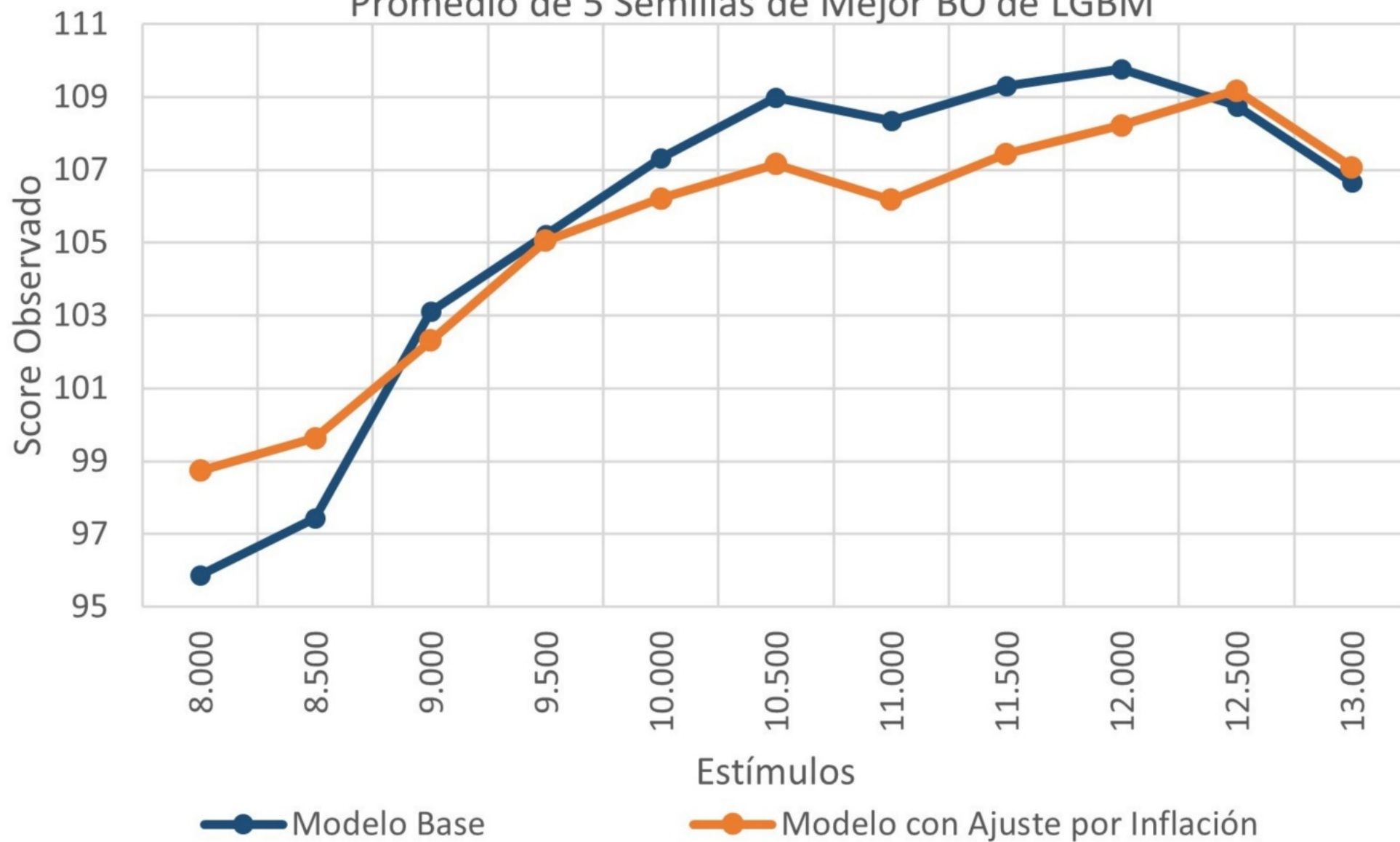
BO con LGBM de modelo sencillo
6 meses, lags 1, 3, 6

Tratar data drifting por inflación?
No...

Ganancia: cerca de \$110mill.



Score observado en Sep-21 y Cantidad de Estímulos Enviados:
Promedio de 5 Semillas de Mejor BO de LGBM



U5

Datos + Procesamiento

Más Meses: Desde 2019, exc pandemia

Feature Engineering Histórico: -Lags 1,3,6

-Delta 1

-Prom. 6 meses

-Tend. 7 meses

-Agrup. saldo caja

-Quitar col. >50% nulos



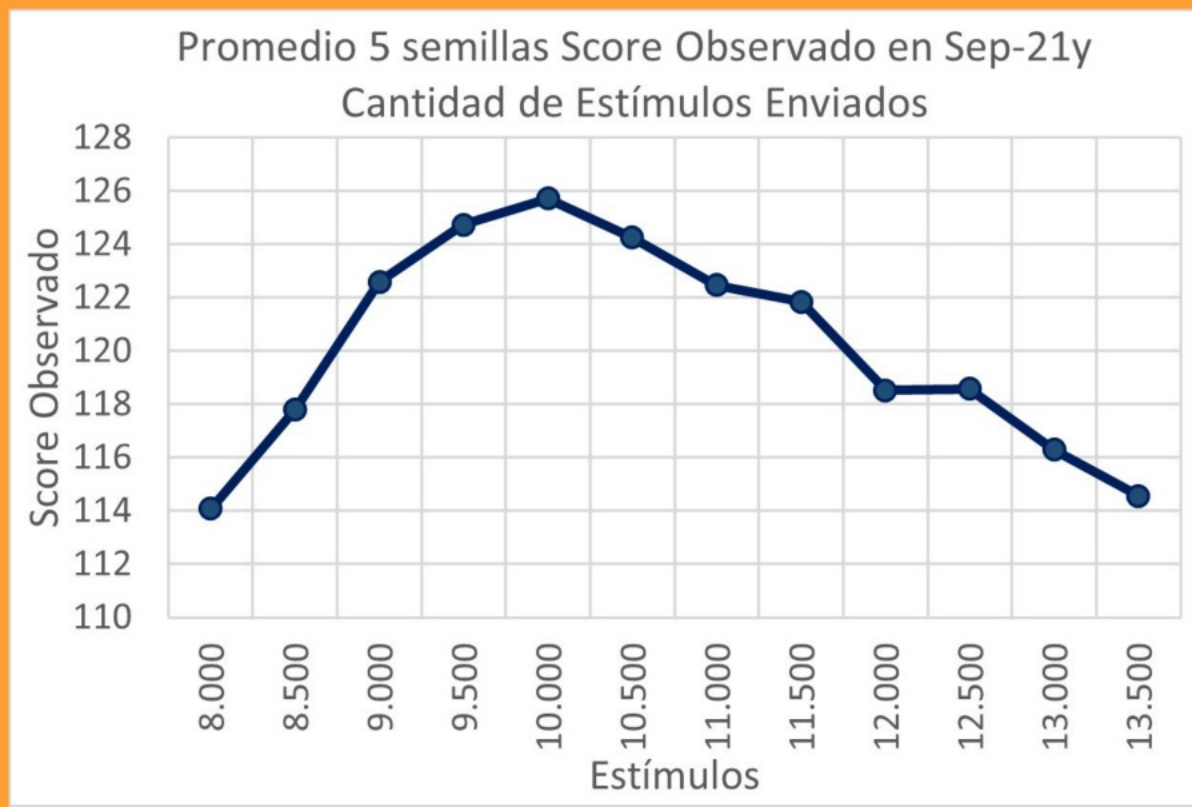
U6

Ganancia del nuevo modelo

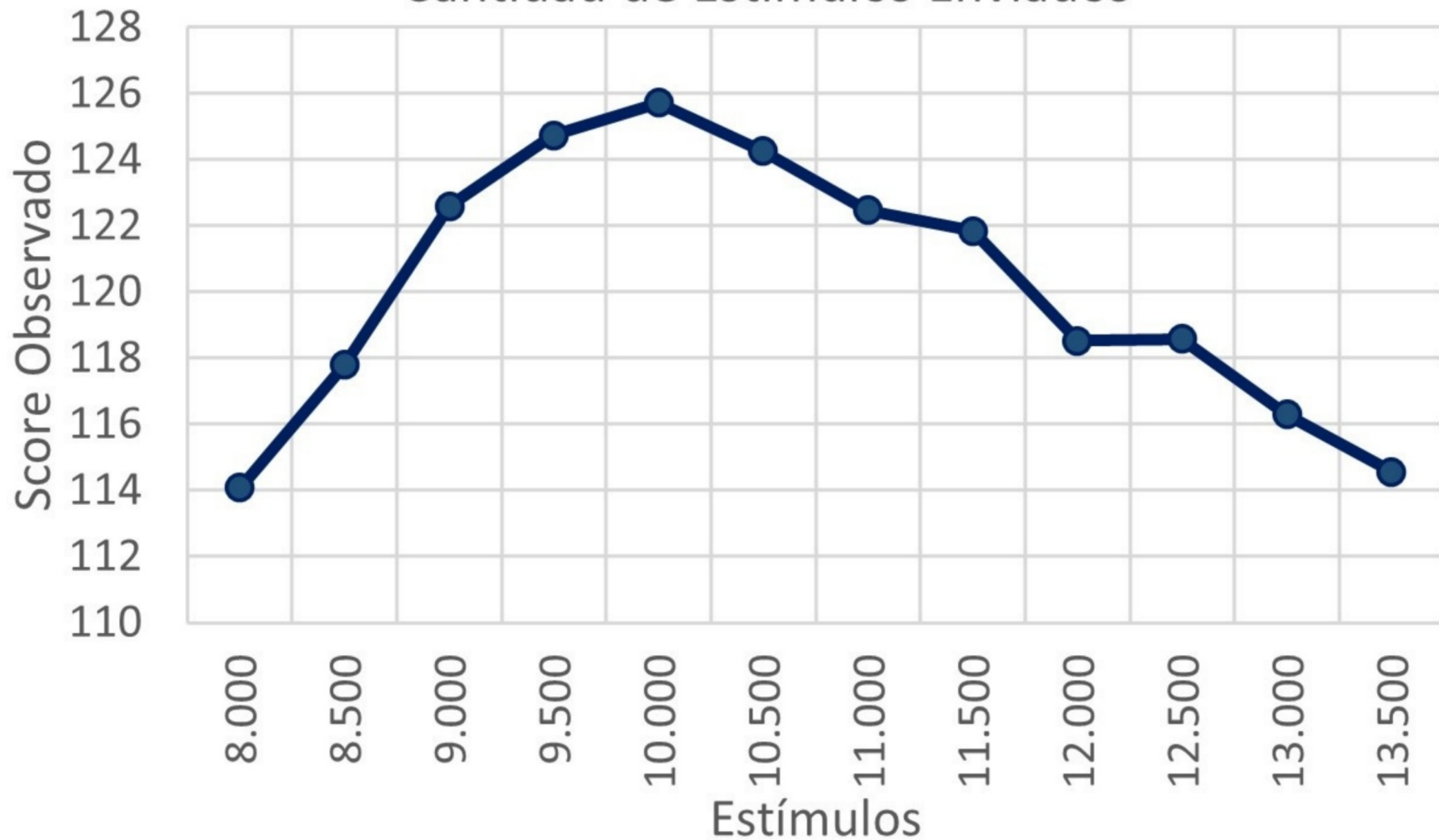
Mejora de ganancia: cerca de \$125m
Duración de BO: 6 días con 256 gb de RAM

Mejores hiperparámetros de Semilla 1:

- num_iterations: 1213
- learning_rate: 0.0292608373298842
- feature_fraction: 0.136781998386116
- num_leaves: 342
- min_data_in_leaf: 43158

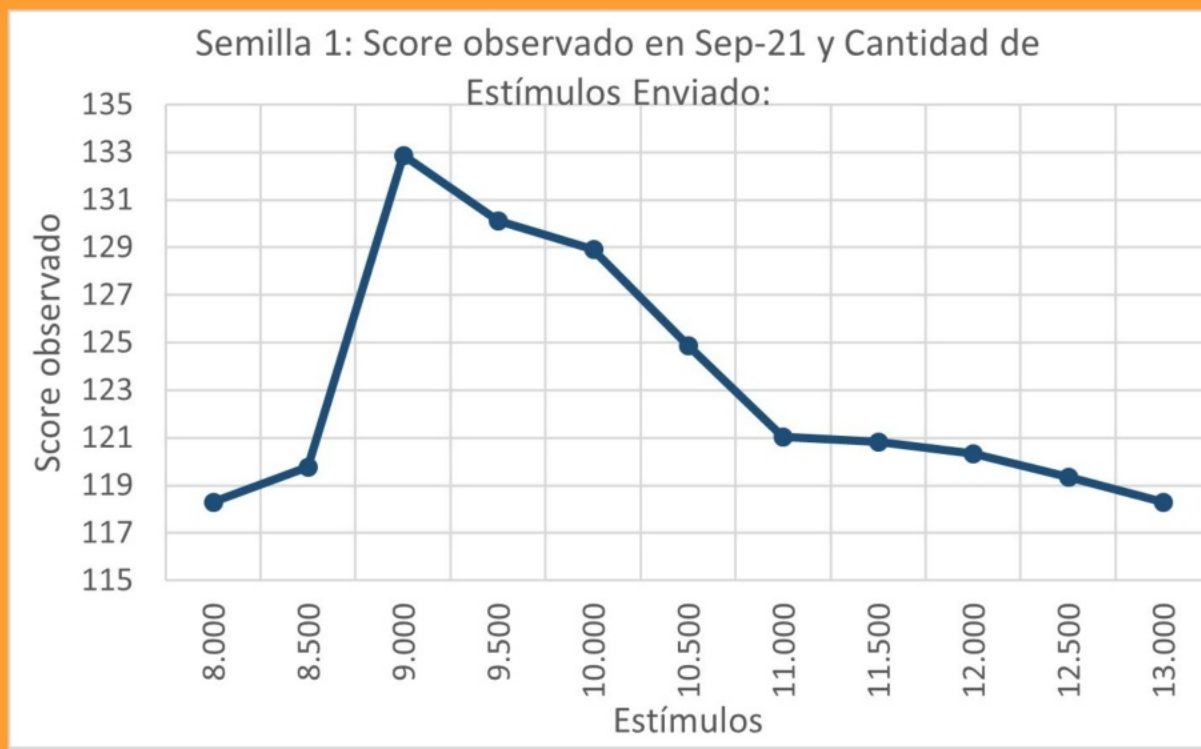


Promedio 5 semillas Score Observado en Sep-21y
Cantidad de Estímulos Enviados



Semillerío

Mismo hiperparámetro de la mejor BO de la semilla 1
Promedio de scores de 100 semillas distintas (5 días...)





Conclusiones

Datos+Procesamiento+Reducción de Varianza
Limitaciones: Tiempo... ¿Undersampling?

