

## FAQ Reforestación

- ¿Cómo funciona el proceso para obtener las predicciones a grandes rasgos ?

Se utilizan los datos con las características previamente compartidas para entrenar el modelo de elevación, distancia a cuerpos de agua entre otras, internamente se hace un preprocesamiento y se ajusta el modelo y se deposita la información en Athena. Una vez en Athena es posible actualizar el dashboard.

- ¿Qué servicios de amazon se utilizan en el uso del modelo ?

Se utilizan los siguientes servicios de aws: S3, Sagemaker Endpoint, Athena, Quick Sight

- ¿Cómo se describen los inputs y los outputs ?

El input son los datos con las características previamente compartidas para entrenar el modelo en csv. Como elevación, distancia a cuerpos de agua entre otras, internamente.

Existen dos outputs la tabla de Athena que contiene todas las características compartidas más una columna extra que contiene la predicción y el segundo output es la visualización del dashboard de las predicciones.

- ¿Cómo se utilizan los outputs ?

Se puede acceder a la tabla de Athena por consultas en sql y los resúmenes analíticos extras dependen del equipo del parque.

El output de la visualización del dashboard de las predicciones pueden ser usadas para tomar decisiones estratégicas que también dependen de las asociaciones.

Por ejemplo, si en las predicciones se observa una mayor proporción de la especie B propensa a crecer una posible decisión estratégica sería comprar semillas de la especie B en lugar de las demás.

- ¿Cada cuánto se pueden usar las predicciones ?

Las predicciones pueden usarse en cada campaña y el número de veces que se utilice impactará en el costo

- ¿Cuál es el costo estimado anual ?

Suponiendo que se tienen 2 reforestaciones al año, con un total de 60,000 registros a predecir entonces se estima un gasto de 2 dolares por año. En dado caso que se necesiten más registros el costo puede variar.

¿Cuánto tiempo toma calcular las predicciones ?

Para calcular predicciones se necesitan aproximadamente 10 minutos dependiendo del volumen de datos . Esto es a que primero se debe prender un endpoint para poder hacer uso del modelo y posteriormente apagarlo

- ¿La tabla de Athena almacena información histórica?

Para reducir costos cada campaña sobrescribirá datos en la tabla , es decir no existirá un registro de campañas históricas en la tabla más que la última campaña a la cual se le calculó la predicción

- ¿El modelo se puede usar de por vida ?

Al ser un modelo de ML puede requerir un re entrenamiento en caso de que se cuente con un nuevo set de datos, se recomienda monitorear el comportamiento a través del tiempo aunque para medir el éxito del modelo tiene que pasar una cantidad considerable de tiempo se puede evaluar con estas opciones propuestas:

- ❖ Hacer muestreos de la data evaluada y etiquetar manualmente cada determinado periodo de tiempo y así poder evaluar el performance del modelo
- ❖ Recolectar datos del comportamiento temprano de la semilla que se usó para reforestar para usar
- Cual es la periodicidad con el que el dashboard será actualizado

Se actualizará en cada campaña y tampoco se guardarán vistas históricas