



# Módulo VIII: Deployando un modelo

## Clase 30: Monitoreo de modelos en producción (parte II)





**¿Ponemos a grabar  
el taller?**

## ¿Con qué practicaremos?



- Conceptos de data drift, model drift y concept drift.

- ¿Cómo desarrollo un portfolio?



# ¿Repasamos?

## Puesta en producción



- Predecir nuevos datos
- Comunicar los resultados
- Integrar los resultados con aplicación

## Machine Learning Operations

El objetivo de **MLOps** es reducir la fricción técnica para poder llevar la idea de un modelo de ML a su puesta en producción en el **menor tiempo posible** con el **menor riesgo posible**.

## Desarrollo de ML con MLOps

- Se acelera el proceso de creación de valor.
- Se reduce el riesgo durante la validación de modelos
- Se simplifica la implementación de soluciones más complejas
- Procesos de mejora de modelos más rápidos y con menos errores gracias a la automatización de procesos.
- Monitoreo y actualización continua de los modelos para que evolucionen con los datos.

## ¿Qué elementos monitorear?

### **Datos entrantes del modelo:**

Identificar cambios en la distribución de los datos (**Data Drift**) y de diferencias entre la distribución de datos de entrenamiento y de producción

Monitoreo operativo de fallas evidentes: ej: imágenes que son muy oscuras, mucho contraste, etc

Monitoreo estadísticas comparadas con el set de entrenamiento: media, desvío estándar, correlaciones



## ¿Qué elementos monitorear?

### Resultados del modelo



Monitoreo del desempeño y observar si cambia la distribución de las predicciones a través del tiempo  
(**concept drift**)



Monitoreo de la salud de los pipelines y de que no sufran fallos inesperados



Monitoreo de que el desempeño cumpla con los requisitos del usuario  
(ej: Velocidad de respuesta)



# Model Drift

## Model Drift

Model drift se define como la **caída en el poder predictivo** de un modelo debido a cambios en el ambiente donde el modelo se desempeña. Puede tener múltiples desencadenantes, como por ejemplo, cambios en el ambiente digital o en la relación entre las variables.

## Model Drift

**1**

Data Drift

**2**

Concept Drift

?



# Data Drift

## Data Drift

Data drift se define como el **cambio** repentino e inesperado de las características de los datos causado por la actualización y/o mantenimiento del sistema que produce los **datos que ingresan** al modelo de ML.

## Data Drift

Para poder detectarlo debemos **observar los datos entrantes** al modelo en **producción** y **compararlos** con los datos que se usaron durante el **entrenamiento** del mismo modelo. Si el dataset de producción y de entrenamiento no tienen el mismo formato o distribución es un indicio de que el sistema está experimentando data drift.

## ¿Cómo detectar Data Drift?

- Test de Kolmogorov-Smirnov (K-S)
- Population stability Index (PSI)

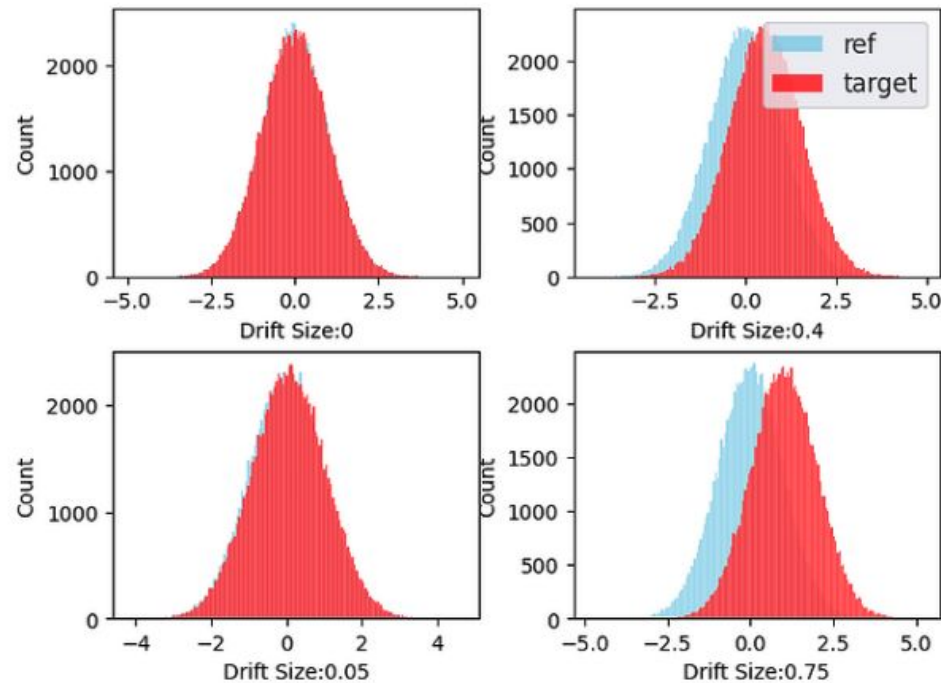
?



## ¿Cómo detectar Data Drift?

- Test de Kolmogorov-Smirnov (K-S)

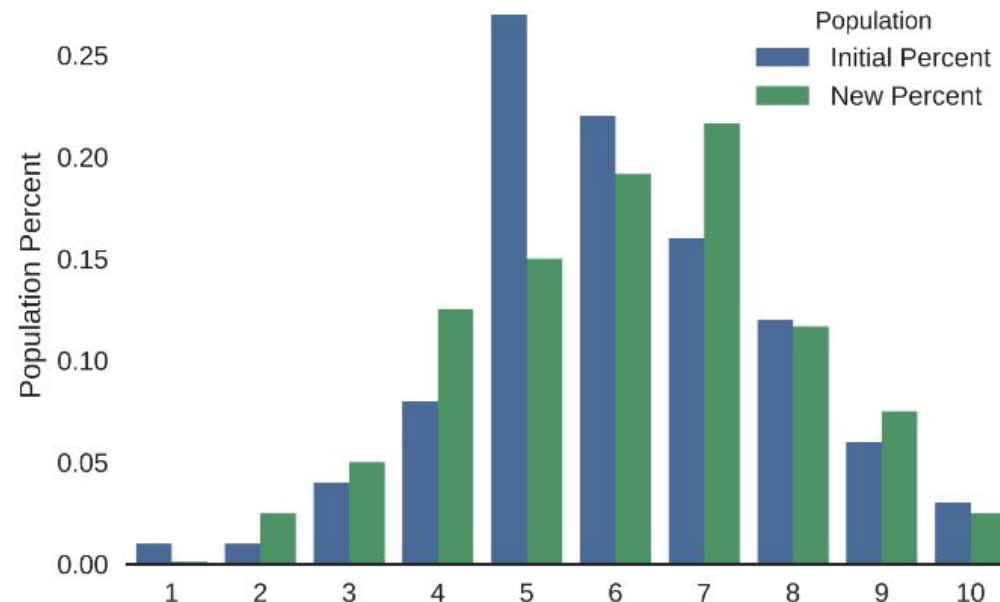
Es un test para evaluar la igualdad entre dos distribuciones de una dimensión.



## ¿Cómo detectar Data Drift?

- Population stability Index (PSI)

Es una métrica para evaluar cuánto una variable ha cambiado en su distribución entre dos muestras en el tiempo. Se divide la distribución en decilos, se calcula el % en cada uno y se obtiene un índice, que si es mayor a 0.2 significa que hubo un cambio significativo.





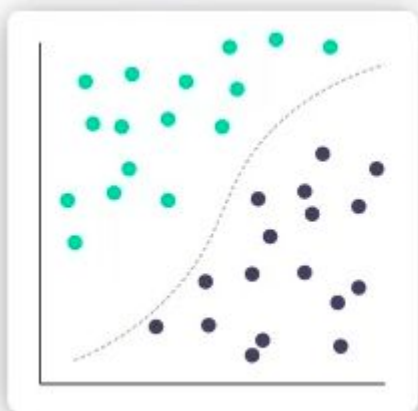
# Concept Drift

## Concept Drift

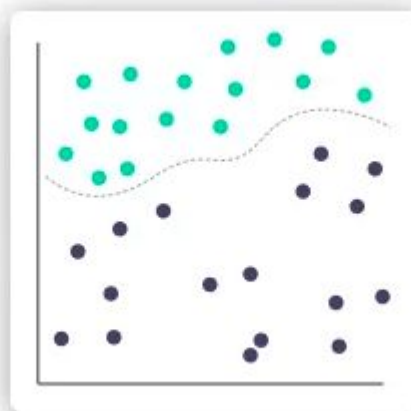
Concept drift ocurre cuando las características estadísticas de la **variable a predecir** cambian. El significado de los datos en los que el modelo fue entrenado cambia, como el modelo no lo sabe no puede predecir los nuevos valores correctamente.

# Concept Drift

**Datos originales**

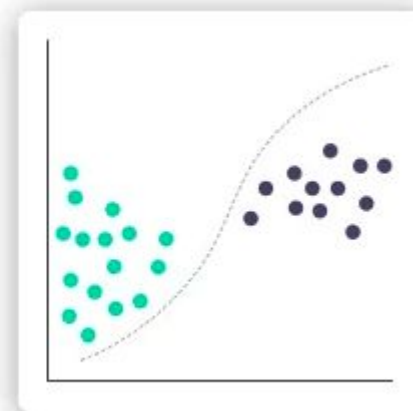


**Concept Drift**



Cambia la variable a  
predecir

**Data Drift**



Cambian los datos  
de entrada pero no la  
variable a predecir

## ¿Cómo prevenir Concept Drift?

- Detectando cambios en la probabilidad de predicción, o sea de los datos de salida, usando los datos de producción o un dataset holdout
  - ADWIN algorithm

## ¿Cómo prevenir Concept Drift?

- Model monitoring
  - Revela si el desempeño del modelo decayó en el tiempo y puede indicar concept drift
- Estrategia basada en el tiempo
  - El modelo es re-entrenado periódicamente luego de cierta cantidad de tiempo donde se considera no es aceptable el desempeño del mismo
- Aprendizaje Online
  - El modelo se entrena cada vez que hay nuevos datos disponibles, en lugar de acumular un dataset masivo y luego re-entrenar.



# Repasamos en Kahoot



## Preparando la entrega final

En la **última clase** deberán presentar un resumen de lo trabajado durante el curso. Para ello podrán preparar una **presentación** en PowerPoint, Slides o Canva.

Tendrán 15 minutos para contar a sus compañeras la síntesis de lo aprendido.

Además, para recibir el certificado, deberán adjuntar en el aula virtual un link a su repositorio. El mismo debe incluir los ítems detallados en la siguiente hoja.



## Preparando entrega final

Ítems a incluir en el repositorio:

- Readme (resumen, no más de un párrafo por ítem. En el mismo pueden contar de dónde obtuvieron dataset, resumen proyecto, qué objetivo se propusieron y qué descubrieron en análisis exploratorio, detalle de los modelos ajustados)
- Dataset
- Notebook con análisis exploratorio (correspondiente a pre-entrega 2)
- Notebooks donde apliquen modelos de aprendizaje supervisado y no supervisado y donde se observen resultados obtenidos (correspondientes a pre-entrega 3 y 4)
- Será valorado que cuente una historia con esos datos: que puedan demostrar cualitativamente qué preguntas se hicieron, conclusiones que sacaron con los datos, decisiones tomadas.





**¿DUDAS?**

FUNDACIÓN  
**YPF**

¡Muchas gracias!

