

Conceptos avanzados a monitorear

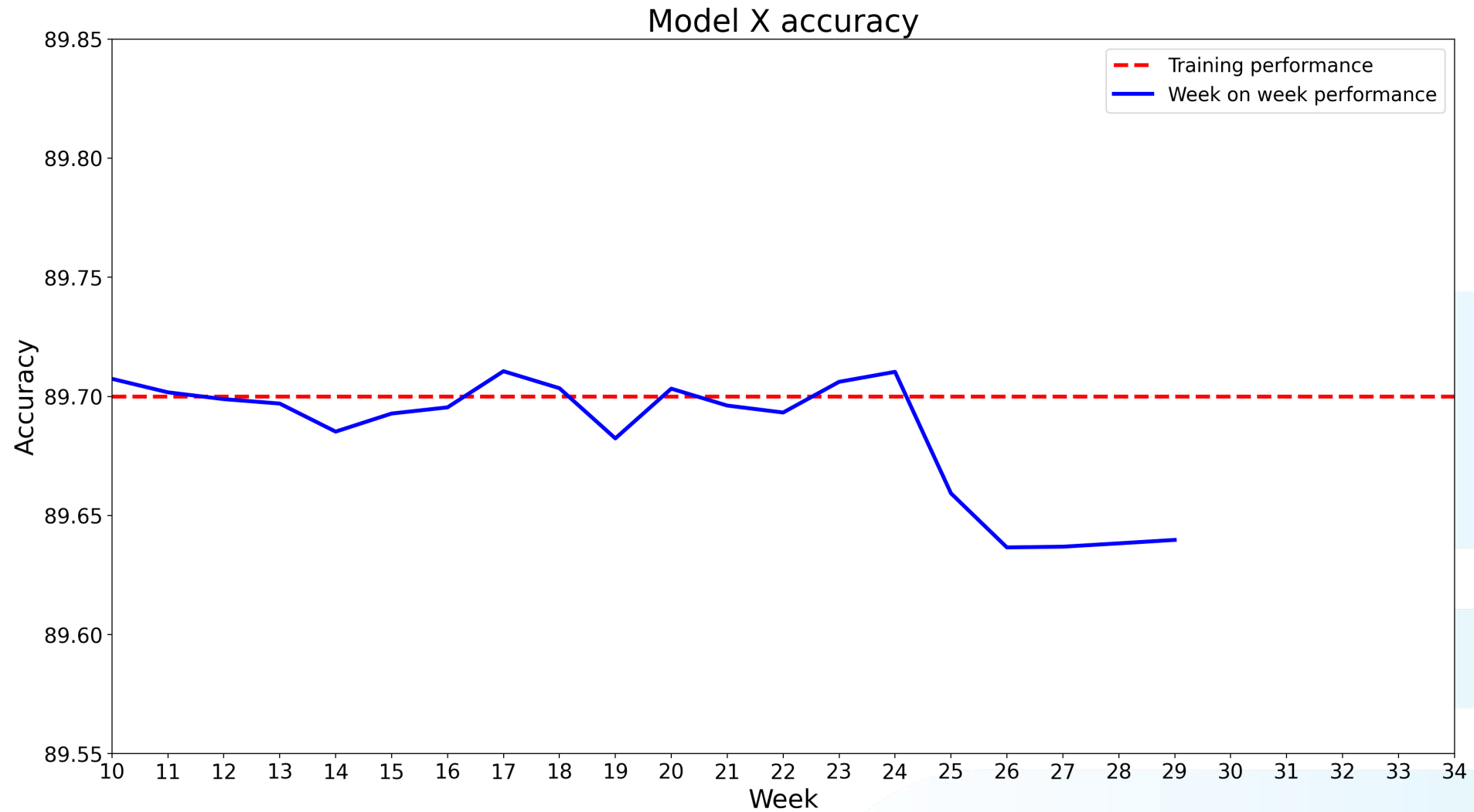


Model decay

- Es un término comúnmente utilizado en la documentación y literatura.
- Consiste en comparar las métricas utilizadas durante el desarrollo con las predicciones reales en producción para determinar si el modelo sigue siendo relevante y útil.

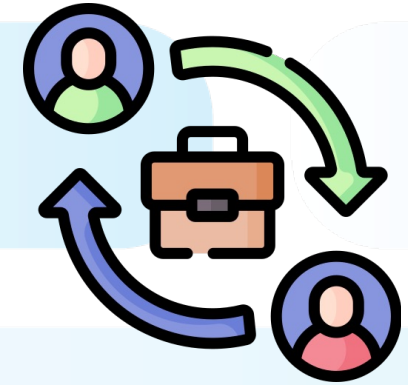


Model decay



Feature drift

- Significa que los datos de entrada a un modelo han cambiado, y que el modelo dejó de ser relevante para esta nueva versión del mundo que le está mandando estas nuevas señales.
- Ejemplo: un modelo entrenado con rangos salariales entre 1000 y 2000 comienza a recibir nuevos rangos entre 3000 y 5000



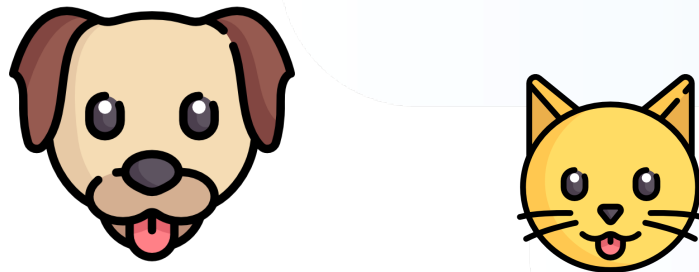
Label drift

- Se presenta cuando la distribución de las variables de entrada es la misma que siempre, pero la distribución de respuestas ha cambiado
- Ejemplo: antes de cierto evento en 2020, 20% de clientes cancelaban, después de marzo 2020, esta cifra aumentó a 80%



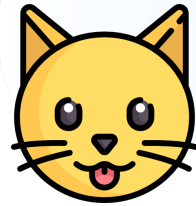
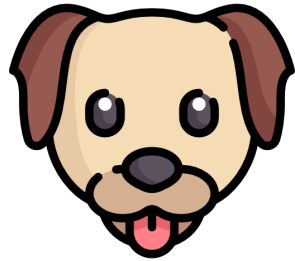
Concept drift

- Significa que la relación entre los datos de entrada y la salida esperada han cambiado.
- A veces es seguro asumir que la relación entre estas dos variables se va a mantener estática:



Siempre será un perro y un gato

Concept drift



Siempre será un perro y un gato

- Esta presunción no siempre se va a mantener para todos los modelos...

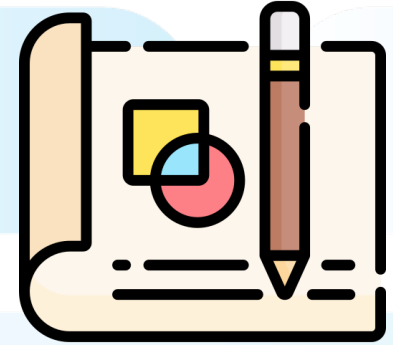
Training-Serving skew

- Se refiere a las diferencias entre las características de entrada en producción y las vistas durante el entrenamiento.
- No se debe a cambios en el mundo real, sino a errores en el modelado:
 - Generación sintética de información, uso de datos desactualizados o errores en la ingeniería de características.

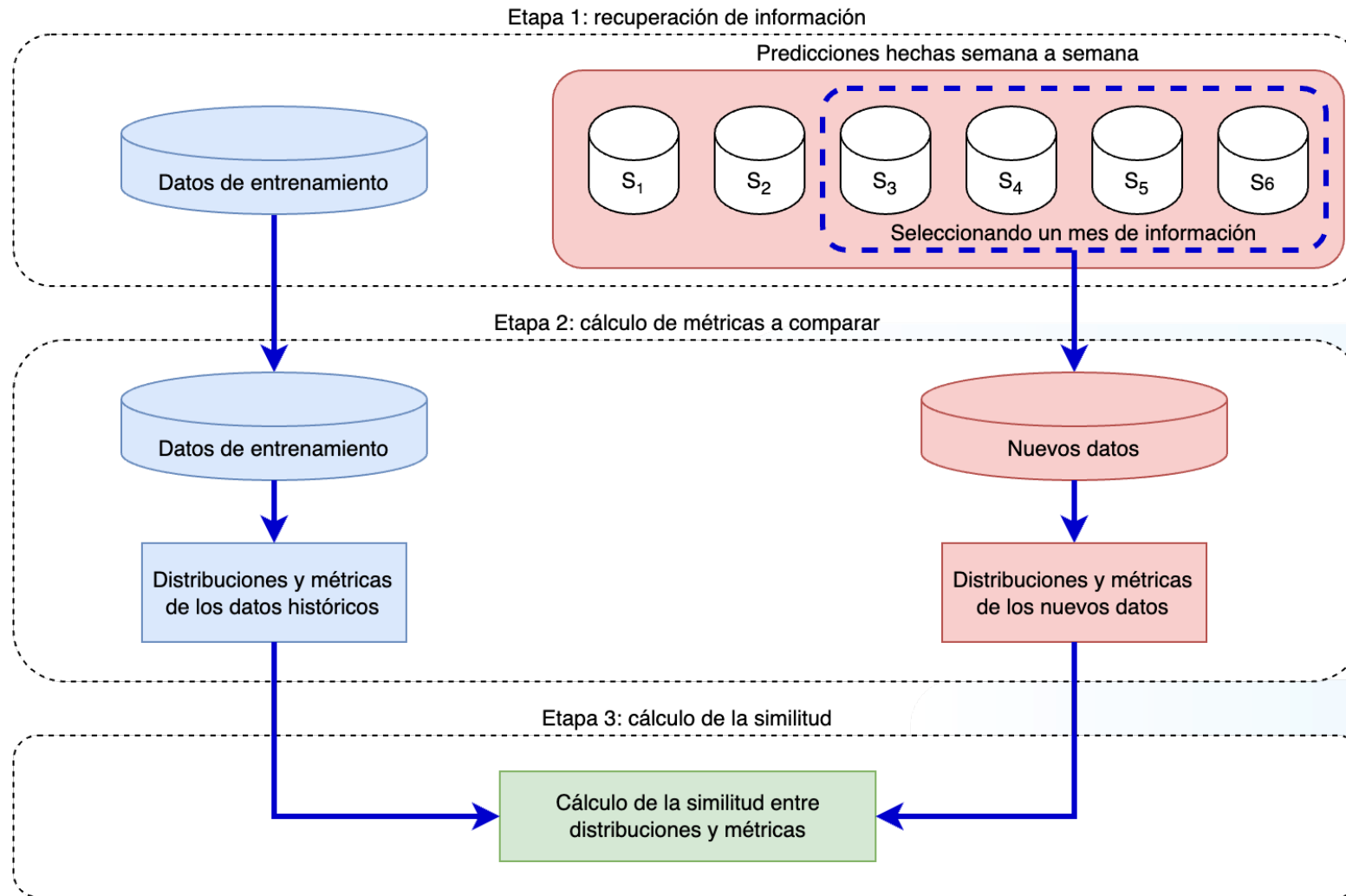


Schema skew

- Se da cuando nuestro modelo recibe variables distintas a las que fue entrenado.
- Ejemplo: una variable que al momento de entrenamiento contenía solamente números flotantes pero al momento de inferencia recibe números enteros.

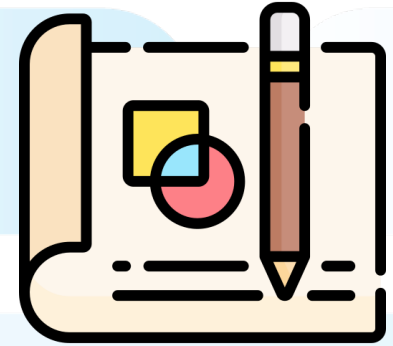


Detección de drifts



Detección de schema skew

- Un simple chequeo de tipos de datos nos bastará para descubrir si un tipo de dato en una columna ha cambiado o si nuevas columnas han sido añadidas a nuestro dataset.



Conclusión

- No es una tarea sencilla.
- Sin embargo, es una tarea crucial que nos ayuda a mantener nuestro modelo fresco y en producción.

