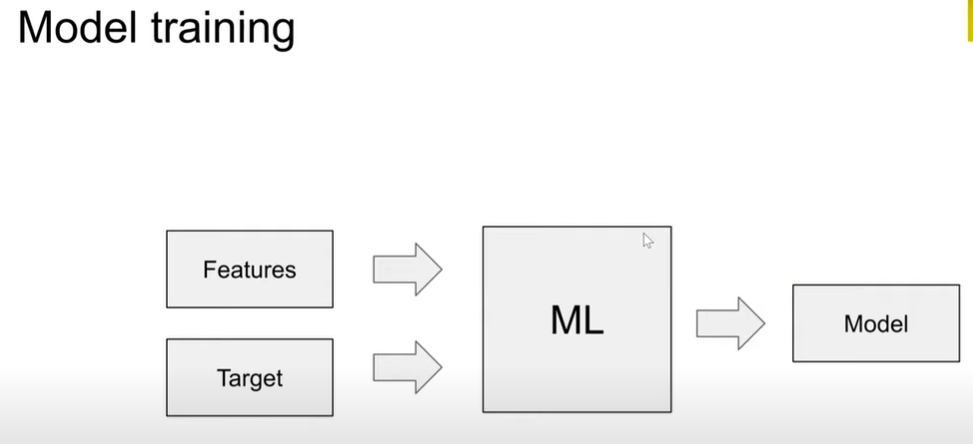
**Week 1**

Intro ML: Es el proceso de extraer patrones desde los datos. Normalmente hay dos tipos de datos: features (características), que es información histórica sobre el objeto que queremos predecir. Y target (objetivo), que es lo que queremos predecir acerca del objeto.

La salida del entrenamiento del modelo es el modelo, que es un objeto que encapsula esos patrones.

Diagrama

Descripción generada automáticamenteNuevas características (features) + el modelo encapsulado, generamos predicciones.

**Sistemas basados en reglas vs Sistemas de ML:**

La diferencia entre los sistemas ML y los basados en reglas se explica con el ejemplo de un filtro antispam.

Los sistemas tradicionales basados en reglas se basan en un conjunto de características (palabras clave, longitud del correo electrónico, etc.) que identifican un correo electrónico como spam o no. Como los mensajes de spam cambian con el tiempo, el sistema debe actualizarse, lo que hace inviable el proceso debido a la complejidad del mantenimiento del código a medida que crece el sistema.

Se puede utilizar ML para resolver este problema con los siguientes pasos:

1. Obtener datos

Los correos electrónicos de la carpeta de spam y de la bandeja de entrada del usuario ofrecen ejemplos de spam y no spam.

2. Definir y calcular características

Las reglas/características de los sistemas basados en reglas pueden utilizarse como punto de partida para definir las características del modelo ML. El valor de la variable objetivo para cada correo electrónico puede definirse en función de la procedencia del correo (carpeta de spam o bandeja de entrada).

Cada correo electrónico puede codificarse (convertirse) con los valores de sus características y objetivo.

3. Entrenar y utilizar el modelo

A continuación, se puede aplicar un algoritmo de aprendizaje automático a los correos codificados para construir un modelo que pueda predecir si un nuevo correo electrónico es spam o no. Las predicciones son probabilidades, y para tomar una decisión es necesario definir un umbral para clasificar los correos electrónicos como spam o no spam.

No saltas directamente de un modelo basado en reglas a uno de ML. Sino que puedes usar esas reglas como punto de partida para tu modelo de ML, usándolas como características (features).