

MACHINE LEARNING

Integrantes:

Samuel Restrepo

Heidy Puerta

Juan José Suarez

Daniel Guarín

Erica Lopez



- El ‘machine learning’ –aprendizaje automático– es una rama de la inteligencia artificial que **permite que las máquinas aprendan sin ser expresamente programadas** para ello. Una habilidad indispensable para hacer sistemas capaces de identificar patrones entre los datos para hacer predicciones. Esta tecnología está presente en un sinfín de aplicaciones como las recomendaciones de Netflix o Spotify, las respuestas inteligentes de Gmail o el habla de Siri y Alexa.

¿QUÉ ES?

APRENDIZAJE SUPERVISADO

- El aprendizaje supervisado es una rama del machine learning (aprendizaje automático) en la que se entrena nodelos utilizando ejemplos etiquetados. En este enfoque, un algoritmo de aprendizaje supervisado analiza un conjunto de datos de entrenamiento que contiene ejemplos de entrada (características) junto con sus correspondientes salidas deseadas (etiquetas).



CARACTERÍSTICAS

- I. Datos etiquetados: El aprendizaje supervisado requiere datos de entrenamiento que estén correctamente etiquetados, es decir, que cada instancia tenga una salida conocida asociada a ella.
- I. Objetivo de generalización: El objetivo principal es aprender una función que generalice correctamente los datos de entrenamiento para predecir o clasificar correctamente nuevas instancias no vistas.
- I. Retroalimentación explícita: Durante el proceso de entrenamiento, se proporciona una retroalimentación explícita al modelo sobre sus predicciones en comparación con las salidas reales.



ALGORITMOS Y TÉCNICAS

- I. Regresión lineal: Utiliza una función lineal para predecir una variable continua.
- I. Regresión logística: Utilizada para clasificar datos en dos o más clases.
- I. Árboles de decisión: Utilizan una estructura de árbol para realizar decisiones de clasificación.
- I. Máquinas de vectores de soporte (SVM): Clasifican datos utilizando hiperplanos en un espacio de alta dimensión.
- I. Redes neuronales: Modelos inspirados en el cerebro que utilizan capas de neuronas artificiales para realizar predicciones.

1. Validación cruzada: Divide los datos en múltiples conjuntos de entrenamiento y prueba, y realiza iteraciones para evaluar el modelo en diferentes divisiones de los datos.
2. División de datos de entrenamiento y prueba: Se separan los datos en un conjunto de entrenamiento y un conjunto de prueba, y se evalúa el modelo en el conjunto de prueba.
3. Métricas de evaluación: Se utilizan medidas como precisión, exhaustividad, puntuación F1, exactitud y matriz de confusión para evaluar el rendimiento del modelo.

MÉTODOS DE EVALUACIÓN Y VALIDACIÓN

ÉTICA Y CONSIDERACIONES LEGALES

1. Privacidad y protección de datos: Asegurarse de que los datos utilizados para el entrenamiento del modelo sean anónimos y cumplan con las regulaciones de privacidad.
2. Sesgos y discriminación: Los modelos pueden aprender sesgos o discriminación si los datos de entrenamiento contienen sesgos inherentes. Es necesario abordar estos problemas y garantizar la equidad.
3. Transparencia y explicabilidad: Algunos algoritmos de aprendizaje supervisado, como las redes neuronales, pueden ser cajas negras difíciles de interpretar. Es importante comprender y explicar cómo se toman las decisiones.

HERRAMIENTAS Y BIBLIOTECAS



Python: scikit-learn,
TensorFlow, Keras, PyTorch.



R: caret, randomForest,
glmnet.



Weka: una plataforma para
el aprendizaje automático.

APRENDIZAJE SEMI-SUPERVISADO

- El aprendizaje semi-supervisado es un enfoque de la inteligencia artificial que combina elementos del aprendizaje supervisado y no supervisado. En lugar de contar únicamente con datos etiquetados, como en el aprendizaje supervisado, el aprendizaje semi-supervisado utiliza una combinación de datos etiquetados y no etiquetados para entrenar un modelo.



CARACTERÍSTICAS

- I. Datos etiquetados y no etiquetados: Se utilizan datos etiquetados para entrenar el modelo, pero también se aprovecha un conjunto más amplio de datos no etiquetados.
- I. Aprovechamiento de información no etiquetada: El objetivo es utilizar la información contenida en los datos no etiquetados para mejorar la capacidad de generalización y rendimiento del modelo.
- I. Ahorro de costos en etiquetado de datos: Al utilizar datos no etiquetados, se reduce la necesidad de etiquetar manualmente grandes volúmenes de datos, lo que puede resultar costoso y consumir mucho tiempo.



ALGORITMOS Y TÉCNICAS

- I. Propagación de etiquetas: Se propaga la información de las etiquetas conocidas a los ejemplos no etiquetados a través de la estructura de los datos.
- I. EM semi-supervisado: Se utiliza el algoritmo Expectation-Maximization (EM) en combinación con datos etiquetados y no etiquetados para estimar los parámetros del modelo.
- I. Aprendizaje generativo: Se modela la distribución conjunta de las variables de entrada y las etiquetas, utilizando datos etiquetados y no etiquetados.

MÉTODOS DE EVALUACIÓN Y VALIDACIÓN

1. Validación semi-supervisada: Se utilizan datos etiquetados y no etiquetados para evaluar el rendimiento del modelo. Se puede utilizar una combinación de conjuntos de validación etiquetados y no etiquetados.
2. Evaluación basada en consultas de oráculos: Se selecciona un subconjunto de ejemplos no etiquetados y se solicita la etiqueta a un experto humano para evaluar el rendimiento del modelo.

ÉTICA Y CONSIDERACIONES LEGALES

- Al igual que en el aprendizaje supervisado, el aprendizaje semi-supervisado también plantea consideraciones éticas y legales similares, como la privacidad y la protección de datos, la equidad y la transparencia. Además, en el aprendizaje semi-supervisado, es importante garantizar que los datos no etiquetados utilizados cumplan con las regulaciones de privacidad y no se utilicen de manera indebida.

HERRAMIENTAS Y BIBLIOTECAS



scikit-learn: Proporciona implementaciones de varios algoritmos de aprendizaje semi-supervisado, como la propagación de etiquetas y EM semi-supervisado.



TensorFlow: Aunque TensorFlow es conocido principalmente por el aprendizaje profundo, también incluye funciones y técnicas para el aprendizaje semi-supervisado.



PyTorch: Al igual que TensorFlow, PyTorch es una biblioteca popular para el aprendizaje profundo, pero también ofrece herramientas y algoritmos relevantes para el aprendizaje semi-supervisado.