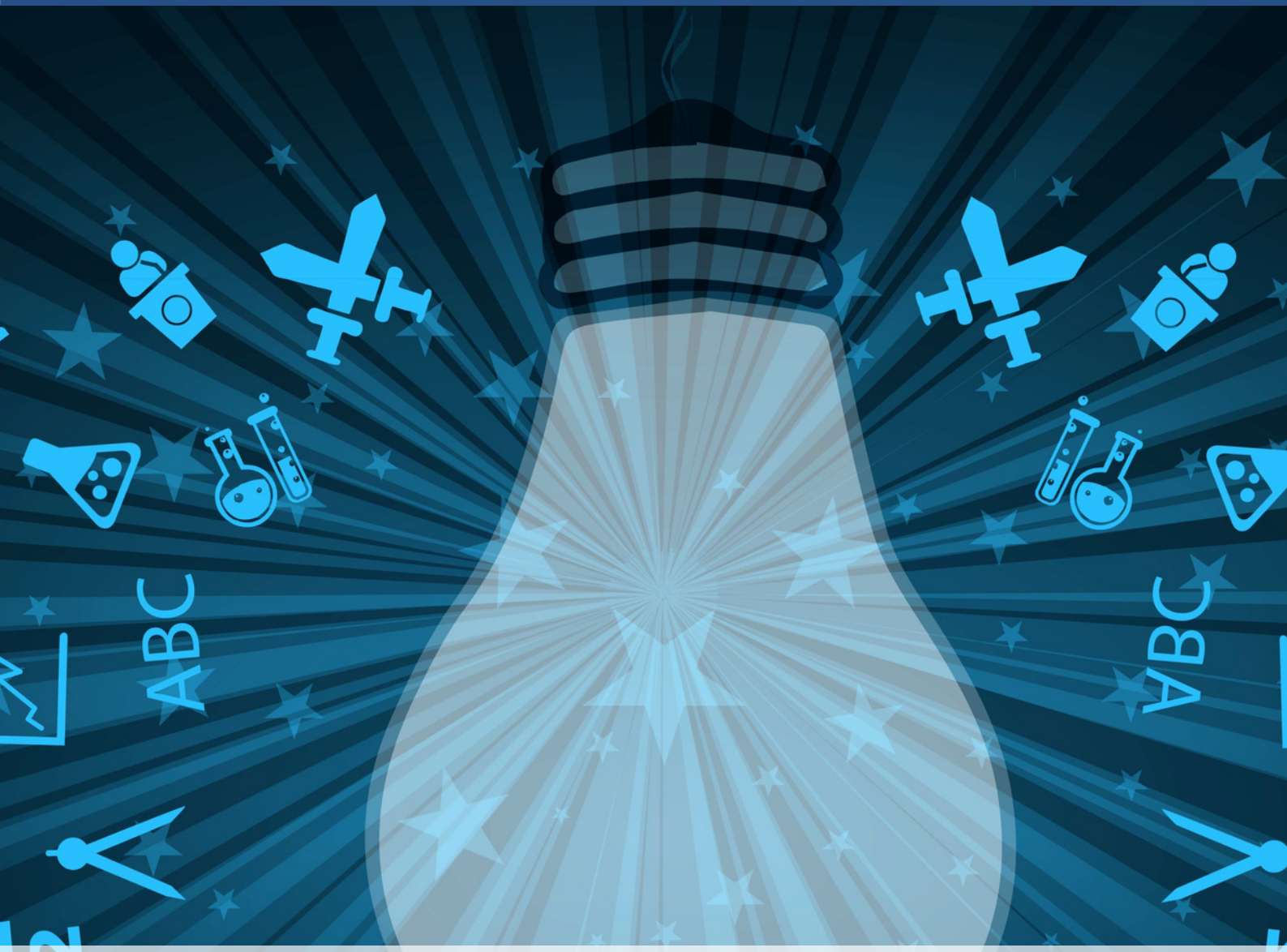
COLOMBIA POTENCIA DE LA VIDA

TIC



Aprendizaje semisupervisado

Aplicación en situaciones con etiquetas limitadas

TALENTO TECH | AZ PROYECTOS EDUCATIVOS | UTP Universidad Tecnológica de Pereira

Aprendizaje semisupervisado: aplicación en situaciones con etiquetas limitadas

El aprendizaje semisupervisado es un paradigma de aprendizaje automático que se sitúa entre el aprendizaje supervisado y no supervisado. En este enfoque, se utilizan conjuntos de datos que contienen tanto datos etiquetados como no etiquetados para entrenar modelos; a diferencia del aprendizaje supervisado tradicional, donde se requieren etiquetas para cada entrada, y del aprendizaje no supervisado, donde no hay etiquetas; el aprendizaje semisupervisado busca aprovechar la información valiosa contenida en los datos no etiquetados para mejorar el rendimiento del modelo.

Principales características del aprendizaje semisupervisado

Conozcamos algunas de estas características:

Conjunto de datos mixto

Combina datos etiquetados y no etiquetados para el entrenamiento.

Objetivo de optimización dual

Busca minimizar el error tanto en los datos etiquetados como en los no etiquetados.

Transferencia de conocimiento

Permite que el modelo generalice patrones aprendidos de datos etiquetados a datos no etiquetados.

Mejora de la generalización

Puede conducir a modelos más robustos y con mejor capacidad de generalización, especialmente en situaciones donde las etiquetas son escasas.

Aplicación en situaciones con etiquetas limitadas

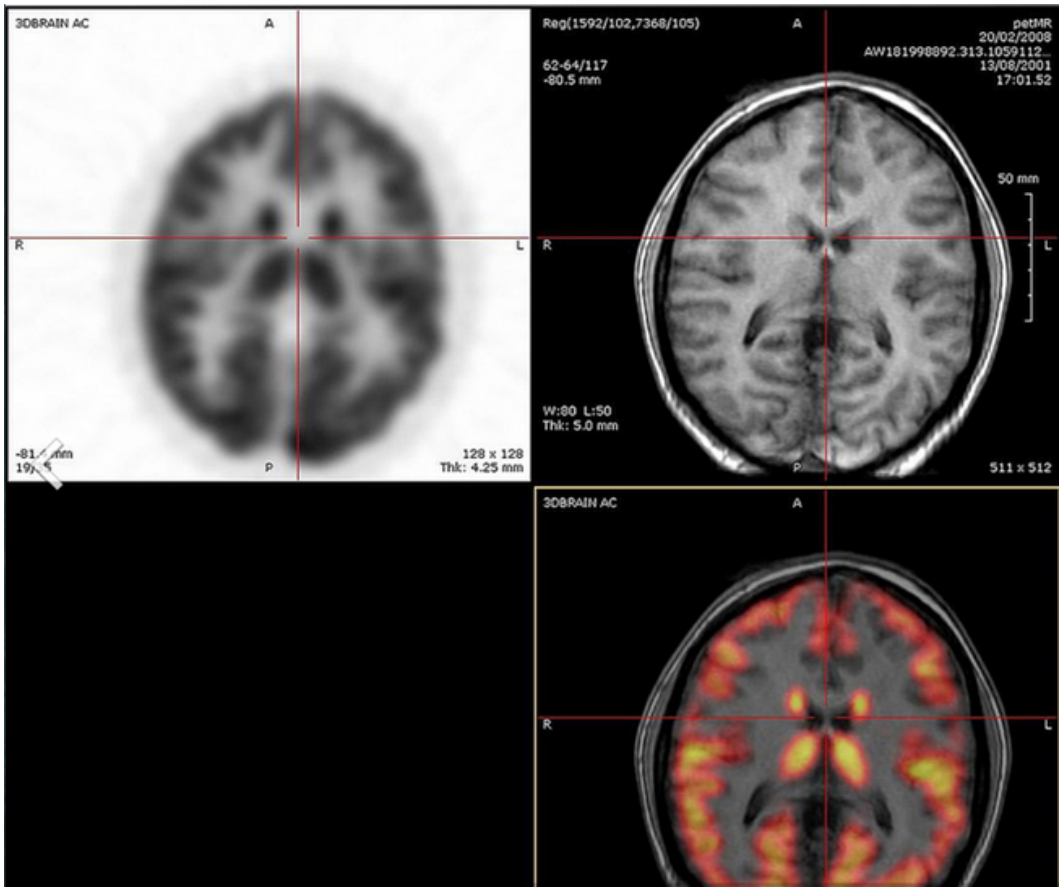
Conozcamos algunas de las situaciones:

Clasificación de Documentos En tareas como clasificación de documentos, donde la etiqueta puede ser costosa de obtener, se pueden usar técnicas semisupervisadas para mejorar la precisión del modelo.	**Imagen Médica** En diagnósticos médicos, donde la obtención de etiquetas puede ser difícil y costosa, el aprendizaje semisupervisado puede aprovechar datos no etiquetados para mejorar la precisión.
Detección de Anomalías En situaciones donde la identificación de anomalías es crítica, se puede utilizar aprendizaje semisupervisado para entrenar modelos con datos normales y detectar patrones anómalos.	**Traducción Automática** En tareas de traducción automática, donde la disponibilidad de pares de traducción es limitada, el aprendizaje semisupervisado puede mejorar la calidad de las traducciones.

El aprendizaje semisupervisado ofrece una solución valiosa en casos donde obtener etiquetas es difícil o costoso, permitiendo el aprovechamiento de la información contenida en datos no etiquetados para mejorar el rendimiento del modelo.

Estudio de casos que ilustran cómo el aprendizaje semisupervisado mejora el rendimiento del modelo

- **Estudio de Caso 1: clasificación de Imágenes Médicas**



Imaginen un escenario en el campo de la medicina, donde se busca clasificar imágenes radiológicas para la detección temprana de enfermedades. Obtener etiquetas precisas para cada imagen es costoso y requiere la experiencia de profesionales médicos. Sin embargo, se dispone de un conjunto limitado de imágenes etiquetadas y un conjunto más extenso de imágenes no etiquetadas.

En este caso, se utiliza un modelo de red neuronal para la clasificación de imágenes médicas, se entrena inicialmente con el conjunto de datos etiquetados. Luego, aprovechando el aprendizaje semisupervisado, se utiliza el conjunto de datos no etiquetados para mejorar la capacidad del modelo para generalizar patrones. La red se ajusta de manera que los conocimientos adquiridos de las imágenes etiquetadas se transfieran de manera efectiva a las imágenes no etiquetadas.

El modelo, mediante el aprendizaje semisupervisado, logra una mejora significativa en la precisión de la clasificación. Las características aprendidas a partir de las imágenes etiquetadas se generalizan con éxito a las imágenes no etiquetadas, permitiendo una detección más precisa de condiciones médicas.

Estudio de Caso 2: Análisis de Sentimientos en Redes Sociales

En el ámbito del análisis de sentimientos en redes sociales, se enfrenta el desafío de etiquetar grandes cantidades de datos para entrenar modelos efectivos. Un conjunto de datos contiene una fracción de tweets etiquetados con sentimientos, pero la mayoría no está etiquetada.

Se utiliza un modelo de procesamiento del lenguaje natural (NLP) para analizar el sentimiento de los tweets. Inicialmente, se entrena con el conjunto etiquetado; luego, mediante el aprendizaje semisupervisado, se incorporan los tweets no etiquetados. El modelo ajusta sus pesos para comprender patrones más sutiles en el lenguaje y mejorar su capacidad de clasificación. El modelo, al integrar información de tweets no etiquetados, muestra una mayor precisión en la identificación de sentimientos, la capacidad del modelo para comprender expresiones lingüísticas más diversas se ve ampliamente beneficiada por el aprendizaje semisupervisado.

Estos casos, ilustran cómo el aprendizaje semisupervisado puede ser una estrategia efectiva para mejorar el rendimiento de los modelos de red neuronal en situaciones donde las etiquetas son escasas o costosas de obtener. La capacidad de transferir conocimientos de datos etiquetados a no etiquetados resulta fundamental para abordar problemas del mundo real.