

Descripción de la tarea.

Caso práctico



Max, una estudiante de Inteligencia Artificial en prácticas está revisando los proyectos que desde diferentes departamentos de la empresa han enviado al de desarrollo de Aprendizaje Automático. Tiene que hacer una primera evaluación del tipo de herramienta o aplicación que piden, qué cantidad y tipo de datos serían precisos y finalmente qué tipo de algoritmos tendrían que utilizar.

Entre todas las peticiones hay cinco que le han llamado especialmente la atención: Un detector de casos de diabetes, una calculadora del posible precio de venta de inmuebles sanitarios, un detector de posibles usos fraudulentos de la tarjeta sanitaria, y un sistema de seguridad en hospital basado en reconocimiento de imagen.

Apartado 1: Analiza los siguientes casos prácticos de proyectos en los que se quiere aplicar modelos de Aprendizaje Automático (Machine Learning) y determina:

- **Qué tipo de Aprendizaje habría que aplicar (supervisado, no supervisado o por refuerzo).**
- **Qué tipo de algoritmo o algoritmos se podría/n utilizar para obtener resultados óptimos.**

Caso 1: Para un hospital.

- Queremos obtener un modelo que ayude a detectar posibles casos de diabetes entre mujeres de cualquier edad.
- Nuestra base de datos consta de 156.374 fichas con diferentes casos ya estudiados.
- Tenemos 8 campos de datos numéricos:
 - Embarazos.
 - Glucosa plasmática.
 - Presión de la sangre.
 - Grosor de la piel en el tríceps.
 - Nivel de insulina.
 - BMI ().
 - Pedigrí de diabetes.
 - Edad.
- Y también conocemos para cada caso si tuvieron Diabetes o no (dato tipo categoría Sí/No).

Para este caso usaría un **tipo de aprendizaje automático supervisado** dado a la cantidad de valores contrastados que tenemos y también un objetivo definido.

El algoritmo que usaría, sería **regresión logística** ya que se usa para casos en los que se desea predecir la presencia o ausencia de una característica.

También usaría un **árbol de decisión** ya que estos algoritmos trabajan bien tanto con datos del tipo de regresión como de clasificación.

Caso 2: Para una empresa del sector inmobiliario.

- Queremos obtener un modelo que prediga el precio de venta de un inmueble.
- Nuestra base de datos consta de 239.751 inmuebles vendidos en el último año en todo el territorio nacional.
- Contamos con la siguiente información por cada inmueble:
 - Metros cuadrados habitables.
 - Número de habitaciones.
 - Categoría de inmueble (piso, casa, adosado).
 - Código postal donde se encuentra el inmueble.
 - Valor según catastro.
 - Año de construcción.
 - Número de reformas realizadas hasta la fecha.
 - Precio de venta.

Para este caso usaría también un **tipo de aprendizaje automático supervisado**.

El algoritmo que usaría, sería **Regresión Lineal** ya que se usa para estimar valores reales de variables con distribución continua, obteniendo como objetivo un valor de salida, en este caso el precio de venta. Dada la cantidad de datos que tenemos como categoría, metros cuadrados... Podemos generar una línea que se ajuste lo mejor posible a la distribución de resultados.

Caso 3: Para una entidad bancaria.

- Queremos obtener un modelo que ayude a detectar posibles usos fraudulentos en tarjetas de crédito.
- Nuestra base de datos cuenta con 2.351.298 datos de transacciones en los últimos nueve meses.
- Para cada transacción contamos con estos datos:
 - Fecha.
 - Hora.
 - Localización (código postal) donde se ha realizado la transacción.
 - Cantidad de dinero que se ha gastado en la transacción.
 - Tipo de dispositivo en el que se ha realizado la transacción (cajero, comercio de productos, comercio de servicios).
 - Saldo medio del propietario de la tarjeta en el último año.
 - Saldo actual del propietario de la tarjeta en el momento de hacer la transacción.

Dado que nos dicen que ayude a detectar posibles usos fraudulentos, entiendo que no determinamos ningún campo objetivo, como si sucede en los casos anteriores. Entonces en primer lugar diría un **tipo de aprendizaje automático no supervisado**.

El algoritmo que usaría, sería **Clustering**, ya que se utiliza para agrupar datos existentes de los que, no intuimos sus características en común. Estas agrupaciones son útiles para explorar los datos e identificar anomalías en ellos y poder realizar predicciones, a fin de detectar correlaciones inesperadas.

Aunque por otro lado también veo un **tipo de aprendizaje supervisado**, ya que el algoritmo de **árbol de decisión** se basa en distinguir grupos en función de los valores de diferentes atributos. Este algoritmo podría ayudar fácilmente debido a las variables de entrada. El algoritmo más popular es el CART.

Caso 4: Para un fabricante de teléfonos móviles.

- Queremos incorporar en la cámara de fotos del móvil una tecnología que reconozca caras para que en esos casos la cámara las enfoque prioritariamente al resto de elementos que aparezcan en el encuadre.
- Contamos con un banco de más de 2.000.000 de imágenes clasificadas y etiquetadas en las que se indica en cuáles hay rostros humanos, y, en las que sí los hay, en qué zonas de la imagen se encuentran.

Para el reconocimiento de imágenes usaría un **tipo de aprendizaje automático supervisado**.

El algoritmo que usaría dado a la cantidad de imágenes que disponemos, sería **máquinas de vector soporte** ya que se aplican a problemas de clasificación y en este caso en especial al reconocimiento de imágenes.

Descarto K-Means (Clustering) dado que tenemos las imágenes clasificadas y etiquetadas.

NOTAS:

Dado el comentario en el foro de que no es necesario usar Teachable Machine en esta tarea, aquí finaliza esta tarea.