

Tarea para SAA02.

Título de la tarea: Clasificación de casos prácticos según algoritmo a utilizar.

Ciclo formativo y módulo: Curso especialización en Inteligencia Artificial y Big Data - Sistemas de Aprendizaje Automático

Curso académico: 2022-2023

¿Qué te pedimos que hagas?

- **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).

Según la descripción del caso, aplicaría un modelo de **Aprendizaje Automático Supervisado**, dado los datos específicos, contrastados y un objetivo definido. Para la obtención de resultados, me decanto por utilizar por una parte, un algoritmo de árbol de decisión (**CART**), que trabaja perfectamente con datos de clasificación que irán dividiendo nuestro muestreo en grupos homogéneos para llegar al objetivo. Y Complementando lo anterior, un algoritmo de **Regresión Logística** para aprovechar la selección y cribado de valores realizada durante el desarrollo de la BD, y la discriminación de casos positivos o no.

- **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, me decanto también por un modelo de **Aprendizaje Automático Supervisado**, donde haremos las predicciones como nos apuntan desde el temario, mediante instancias cuyos datos son valores reales, utilizando un algoritmo de **Regresión Lineal** que utilice esas variables de entrada para utilizar como aprendizaje y conseguir la predicción del valor objetivo que se solicita para el precio de venta.

- **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.

En este caso, **por un lado** creo que se debería aplicar un modelo de **Aprendizaje Automático Supervisado** y el uso de algoritmos **CART** y de **Regresión Logística**, por las mismas razones que en el caso 1.

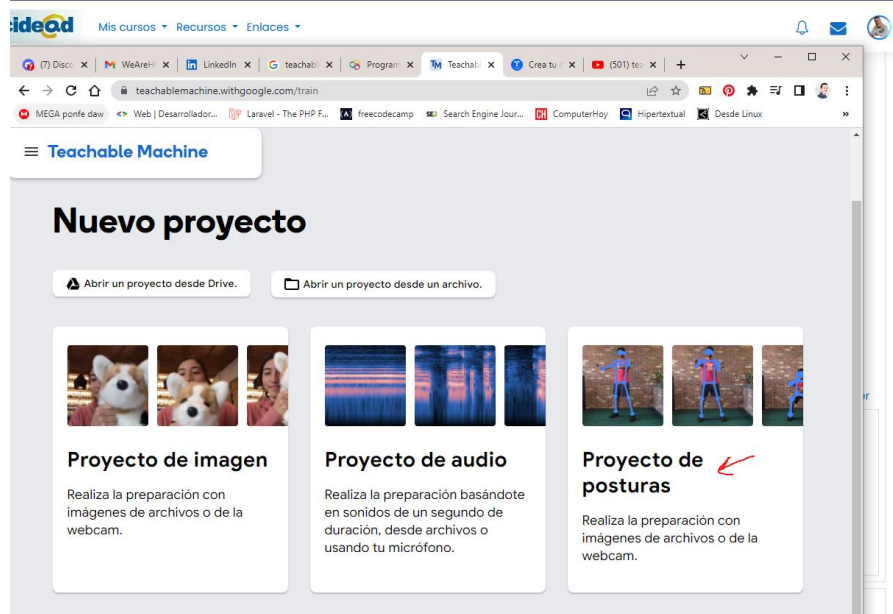
Sin embargo, creo que es **perfeccionable** si se combina con **Algoritmos de Clustering**, dentro de un modelo de **Aprendizaje Automático No supervisado**, para poder encontrar grupos de características comunes entre las instancias, que puedan ayudar a encontrar patrones dentro de los casos de fraude.

- **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.

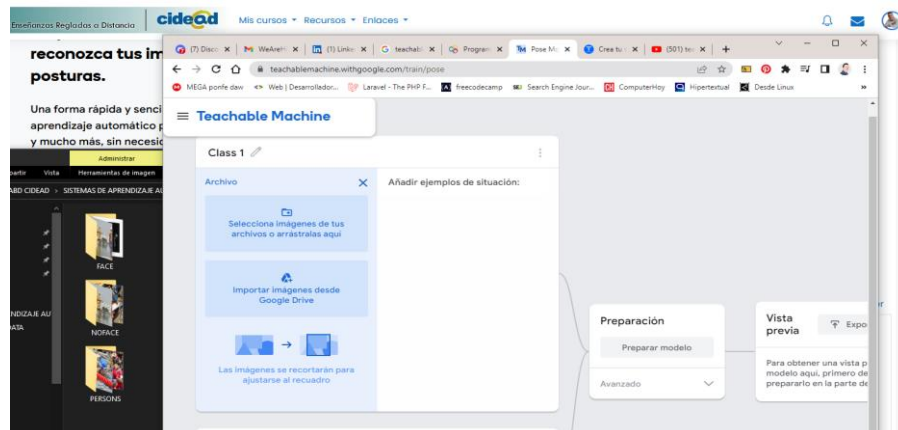
Para este último caso, vuelvo a pensar en **Aprendizaje Automático Supervisado** y el uso de **máquinas de vector soporte**, dado el amplio uso de este tipo de algoritmo para trabajos con clasificación de imágenes y que se encargará de separar en grupos, según como estén etiquetadas, las distintas imágenes clasificadas para la creación del modelo.

Con un correcto cribado y clasificación de grupos, podrían utilizarse (aunque no se exige para la tarea) herramientas como Teachable Machine para el desarrollo y la mejora del modelo.

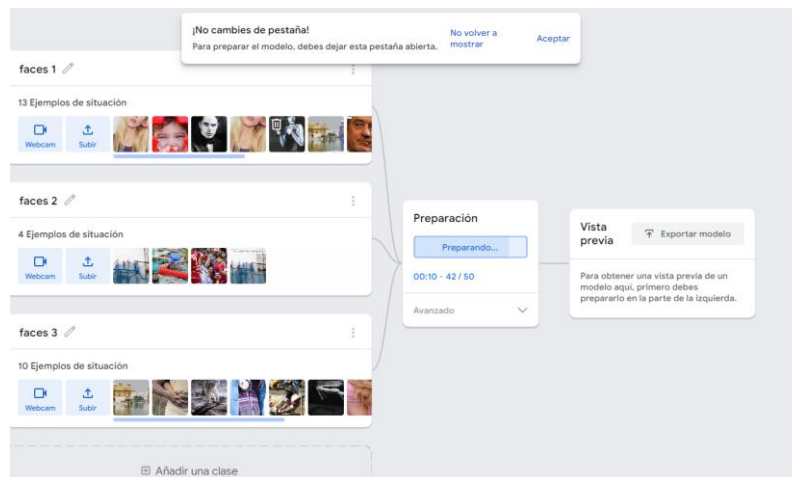
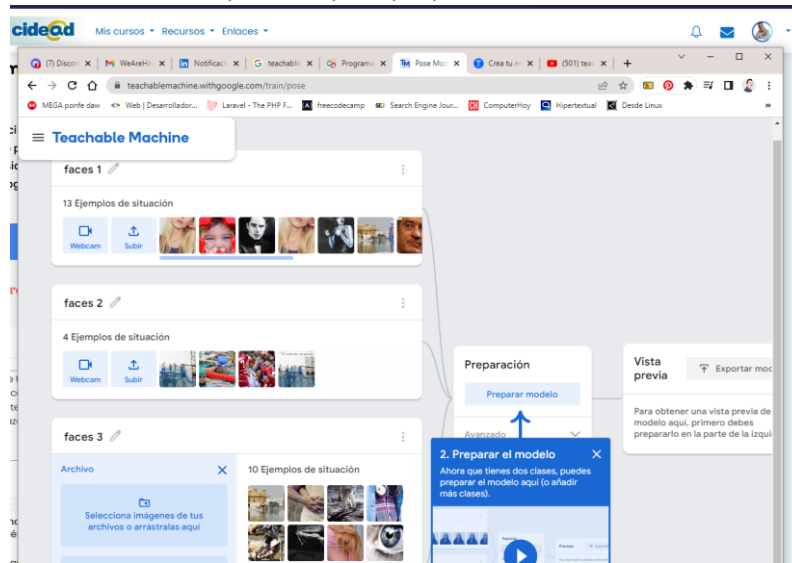
He aprovechado una pequeña muestra de imágenes de stock que tengo en mi equipo (ya que no dispone este de cámara) para intentar simular el proceso:



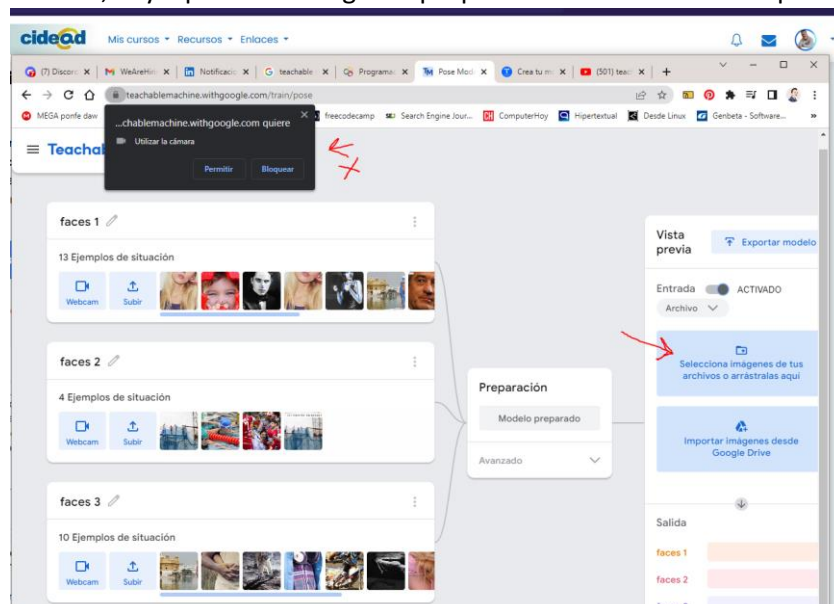
Al crear el proyecto, he separado las imágenes donde aparecen rostros por separado, grupos de personas y otro donde no es posible el reconocimiento facial.



Creo las clases respectivas para preparar el modelo:



Si tenemos cámara, podemos aprovechar y captar de forma directa la imagen. En mi caso, voy a probar con algunas propias de entre mis archivos para probarla.



A pesar de ser un banco bastante pequeño, parece seleccionar de forma bastante adecuada, lo que permitiría avanzar en un posible proyecto para el caso propuesto.

