**TAREAS Y PROCESOS**

1. **Clasificación**: Toma como entrada de un modelo una entidad y como salida dice a qué clase pertenece. Las clases pueden ser excluyentes.

Ejemplo: Para un cierto producto queremos saber si un cliente lo comprará o no.

Clases: 1-> Lo comprará, 2 -> No lo comprará. En este caso solo habría dos clases y serían excluyentes.

Similares: *scoring*, estimación de probabilidad de clase.

1. **Regresión** (“estimación de valor”): Dado una entidad intentará estimar o predecir el valor numérico de alguna de sus variables.

Ejemplo: Queremos saber para un cliente el uso que le va a dar a un servicio. Para ello podemos comparar con gente que sea similar y mirar sus histórico de uso de ese servicio.

**Nota**: Aunque la clasificación está relacionada con la regresión, la clasificación nos dirá **si algo ocurrirá o no** mientras que la regresión nos dirá **cuánto** ocurrirá.

1. **Clustering** (“agrupación): Intenta agrupar una población por su similitud, pero no por un propósito en concreto.

Ejemplo: Se puede utilizar para una exploración preliminar de clientes para ver por ejemplo que productos ofrecemos o desarrollamos.

1. **Coincidencia de similitud**: Teniendo el conocimiento de una entidad se intenta encontrar otra similar.

Ejemplo: IBM está interesado en encontrar compañías similares a las que le compran sus productos. Usan sus datos “*firmográficos*” (tamaño, el valor de los activos, el número de empleados, área donde actúa esa empresa).

Usos: La similitud se usa en tareas de minería de datos como clasificación, regresión y *clustering*(agrupación).

1. **Agrupación de co-ocurrencia**: Intenta encontrar asociaciones entre entidades basándose en las transacciones que las involucran.

Ejemplo: ¿Qué productos se compran juntos de manera frecuente? Quizá el libro A se suele comprar a la vez que el libro B o cierto tipo de carne se comprar junto a cierto tipo de salsa.

Usos: Para análisis de bolsa de mercado o para mejorar los sistemas de recomendación.

Sinónimos: Minería de objetos frecuentes, descubrimiento de reglas de asociación, análisis de cesta de mercado.

1. **Profiling** (perfilar o descripción de comportamiento): Intenta caracterizar el comportamiento típico de un individuo, grupo o población.

Ejemplo: ¿Cuál es el tipo de móvil que usa un determinado segmento de clientes? Es complejo porque podría ver cuánto tiempo emplea escribiendo, cuánto viaja, uso internacional, etc.

Usos: Se puede usar para detectar fraude, monitorizar intrusos en sistemas informáticos, mirando por ejemplo el horario de accesos o si hace alguna compra anormal.

1. **Predicción de conexiones**: Sugiere conexiones entre objetos de datos y estima la fuerza de ese enlace.

Ejemplo: Viendo que Karen y tú tenéis 10 amigos en común, ¿te gustaría ser su amiga? También se utiliza para los sistemas de recomendación mirando el grafo de las películas que un cliente ha visto o puntuado bien.

1. **Reducción de datos:** Se planea reemplazar un conjunto muy grande de datos por otro más pequeño que contenga información más importante, así es más sencillo de procesar.

Ejemplo: Se quiere saber las preferencias de un consumidor con respecto a las películas que prefiere ver. Se puede aplicar una reducción de datos filtrando por las preferencias que tiene en base al género de las películas.

1. **Modelado causal**: Intenta arrojar luz a que eventos o acciones influyen realmente en otros. Siempre debe acabar con una conclusión causal con las suposiciones exactas hechas para una conclusión causal vigilando el “efecto placebo” de pasar por alto ciertas suposiciones.

Ejemplo: Tras elegir un público para presentarles un anuncio las ventas suben. ¿Ha sido por el anuncio o por elegir bien el público? Para saberlo hay que invertir en métodos sofisticados.

Técnicas: Experimentos aleatorios controlados como los test A/B. Se basan en un análisis contrafactual el cual intenta entender que diferencias puede haber entre una situación y otra.