

LAB 2. Machine Learning. Introducción y conceptos principales de aprendizaje basado en datos

Caso 1. Una empresa de marketing quiere explorar un nuevo mercado de usuarios potenciales. Dicho mercado es un nicho para explotar y la empresa no tiene conocimiento previo del mismo, por lo que ha comprado a otra empresa una base de datos de usuarios con muchos campos de información que describen, a priori, sus hábitos de consumo, pero no indican qué tipo de usuarios son. Esta empresa de marketing necesita demostrar que este nicho de mercado es importante y evidenciar la presencia de diferentes usuarios/targets a partir de los cuales puedan desarrollar una campaña. Sobre este escenario que te hemos descrito, responde a las siguientes preguntas:

- ¿Qué clase de modelo y/o tarea dentro del machine learning le recomendarías usar?
 - Teniendo en cuenta este escenario, lo mejor sería usar **aprendizaje no supervisado**. De esta forma se analizaría la base de datos y se podrían descubrir patrones en los usuarios. Estos patrones serían los diferentes targets de la campaña de marketing. Teniendo en cuenta que utilizamos aprendizaje no supervisado para identificar patrones, el tipo de tarea más conveniente sería **clustering**. Con esta técnica agruparíamos datos similares en conjuntos, conocidos como 'clusters'.
- ¿Qué es lo que podrían llegar a conseguir con él o qué impacto tendría en su negocio?
 - Los beneficios que encuentro de tener los datos agrupados y segmentados son:
 1. **Identificar segmentos de mercado**, para comprender mejor las necesidades y el comportamiento de cada sector.
 2. **Capacidad de personalizar más las campañas**, siendo capaz de entender mejor a cada grupo de usuarios y ofreciendo lo más demandan/necesitan.
 3. **Optimizar el gasto en publicidad**, ya que no van a necesitar hacer sondeos en grupos donde no esté su target y tienen agrupados su público objetivo.
- ¿Cómo medirías, según tu criterio (sin utilizar una métrica concreta), el desempeño del modelo de machine learning?
 - El desempeño del modelo machine learning es inmejorable. Va a permitir a la empresa interpretar mejor los datos y segmentar a los usuarios, de una manera consistente en el tiempo y escalable. Además de mejorar su rendimiento económico y mejorar sus campañas de marketing.

Caso 2. Una fábrica quiere mejorar la detección de un fallo que se ha producido pocas veces en un conjunto de máquinas, pero que ha tenido un impacto relevante en su producción. Al producirse sin una regularidad fija y en determinadas ocasiones, disponen de un conjunto de datos etiquetados pequeño, los cuales recopilan información de las máquinas –a través de sensores, etc. – e indican si fallaron, o no, en un determinado momento. Al mismo tiempo, como las máquinas se encuentran monitorizadas continuamente, disponen de un gran conjunto de datos de estas, pero sin etiquetar. Entendido el escenario, responde a estas tres cuestiones:

- ¿Qué clase de modelo y/o tarea dentro del machine learning le recomendarías usar?
 - Teniendo en cuenta este escenario, lo mejor sería usar **aprendizaje semisupervisado**. He elegido este tipo debido a que la fábrica, por el momento, maneja una cantidad de datos pequeña y podríamos usar esos datos para entrenar un modelo inicial de detección de fallos, al que en un futuro le podríamos añadir más conjuntos de datos.
- ¿Qué es lo que podrían llegar a conseguir con él o qué impacto tendría en su negocio?
 - Con este tipo de aprendizaje podríamos obtener:
 1. **Mejorar la capacidad de detección de fallos**, aprovechamos datos tanto etiquetados como no etiquetados, para que el modelo pueda identificar patrones.
 2. **Ahorro de costes**, ya que una detección temprana de los errores ayudaría a evitar errores mayores a futuro.
 3. **Mejoraría la productividad de la fábrica**, ya que corregir estos errores a tiempo evitarían el impacto en la producción que estos tienen.
- ¿Cómo medirías, según tu criterio (sin utilizar una métrica concreta), el desempeño del modelo de machine learning?
 - En este ejemplo veo como el desempeño del modelo permite adaptarse a los diferentes cambios posibles en los errores, tiene una gran capacidad de interpretación de datos, tiene un uso práctico muy útil que permite mejorar la eficiencia de la fábrica y el tiempo de detección de errores es mayor al actual, por lo que detectará los fallos de manera preventiva y permitirá solucionarlos.

Caso 3. Un periódico online quiere agilizar su proceso de revisión de artículos antes de publicarlos. Su objetivo es tener una herramienta que pueda, al mismo tiempo, sumar y completar texto que previamente haya escrito un periodista. Ten en cuenta que es un periódico que posee muchos años de existencia, por lo que disponen de multitud de artículos previos digitalizados.

- ¿Qué clase de modelo y/o tarea dentro del machine learning le recomendarías usar?
 - Teniendo en cuenta este escenario, lo mejor sería usar un **modelo generativo**. Este tipo de modelo tienen la capacidad de generar texto de una forma coherente y teniendo en cuenta el contexto. Además, pueden ser entrenados con los datos del periódico para adaptarse al estilo del periodista.
- ¿Qué es lo que podrían llegar a conseguir con él o qué impacto tendría en su negocio?
 - Implantar una herramienta como esta, basada en el procesamiento del lenguaje natural (NLP), traería al periódico diferentes beneficios:
 1. Mejorar la productividad de los periodistas, al tener un lenguaje predictivo, serían capaces de escribir más artículos.
 2. Mejora en la calidad de los artículos, debido a las sugerencias contextualizadas y coherentes. Además, añade corrección automática de gramática.
 3. Se aprovecharían los datos históricos, por lo que la esencia del contenido prevalecería intacta y se lograría mantener la calidad y estilo propio del periódico.
- ¿Cómo medirías, según tu criterio (sin utilizar una métrica concreta), el desempeño del modelo de machine learning?
 - En este ejemplo veo como el desempeño del modelo permite generar sugerencias coherentes y consistentes, que prometen un impacto positivo. Además, se preservaría el estilo del periódico. Se optimizaría el trabajo de los periodistas y esto conllevaría un impacto económico positivo también para el periódico.