



Universidad de Buenos Aires
Facultad de Ciencias Económicas



Aprendizajes: Supervisados, Semi supervisados, No Supervisados

AR Tema extractado del libro "**Análisis Funcional de Sistemas y Tecnologías de la Información**" de Aníbal M. Mazza Fraquelli - ISBN 978-987-26981-3-3

Presentación del Tema

El aprendizaje automático (Machine Learning – ML, Aprendizaje Automático) constituye una subdisciplina de la Inteligencia Artificial (Artificial Intelligence – AI, Inteligencia Artificial) orientada a desarrollar sistemas capaces de identificar patrones en datos y mejorar su desempeño a partir de la experiencia. En el ámbito organizacional, estos sistemas son fundamentales para la analítica avanzada, la automatización de decisiones y la generación de ventajas competitivas basadas en información.

Desde la perspectiva de las Tecnologías de la Información (TI), el aprendizaje automático no debe ser entendido únicamente como una técnica estadística sofisticada, sino como un componente estratégico de la arquitectura de datos y de los sistemas de soporte a decisiones (Decision Support Systems – DSS, Sistemas de Soporte a Decisiones). Su implementación impacta en la infraestructura tecnológica, la gobernanza de datos (Data Governance, Gobernanza de Datos), la seguridad informática y el cumplimiento normativo.

Existen tres enfoques principales dentro del aprendizaje automático: aprendizaje supervisado (Supervised Learning), aprendizaje no supervisado (Unsupervised Learning) y aprendizaje semi-supervisado (Semi-supervised Learning). Cada uno responde a diferentes tipos de problemas empresariales, estructuras de datos y niveles de disponibilidad de información etiquetada. Comprender sus diferencias es esencial para los futuros licenciados en administración que deberán interactuar con equipos técnicos, evaluar proyectos de analítica avanzada y tomar decisiones estratégicas basadas en modelos algorítmicos.

Desarrollo

1. Marco general del aprendizaje automático en la organización

El aprendizaje automático se basa en la construcción de modelos matemáticos que, a partir de datos históricos, permiten predecir comportamientos futuros o identificar estructuras ocultas en los datos. Estos modelos se entrenan utilizando conjuntos de datos (datasets) que pueden contener variables explicativas (features, características) y, en algunos casos, resultados conocidos (labels, etiquetas).

Desde el punto de vista organizacional, el aprendizaje automático se integra en:

- Sistemas de Business Intelligence (BI, Inteligencia de Negocios).
- Plataformas de Big Data (Grandes Volúmenes de Datos).
- Sistemas de Gestión Empresarial (Enterprise Resource Planning – ERP).
- Plataformas de Gestión de Relaciones con Clientes (Customer Relationship Management – CRM).

Su correcta implementación requiere:

- Calidad de datos (Data Quality).
 - Infraestructura de almacenamiento (Data Warehouses, Almacenes de Datos; Data Lakes, Lagos de Datos).
 - Seguridad y protección de datos.
 - Capacidad de cómputo adecuada (Cloud Computing, Computación en la Nube).
-

2. Aprendizaje supervisado

El aprendizaje supervisado es aquel en el cual el modelo se entrena con datos previamente etiquetados. Esto significa que cada registro del conjunto de entrenamiento incluye tanto las variables de entrada como el resultado esperado.

El objetivo del modelo es aprender la relación entre entradas y salidas para poder predecir resultados en datos nuevos.

Existen dos grandes tipos de problemas en aprendizaje supervisado:

1. Clasificación (Classification): el resultado es una categoría.
2. Regresión (Regression): el resultado es un valor numérico continuo.

Ejemplo de clasificación:

Un banco utiliza datos históricos de clientes (edad, ingresos, historial crediticio) junto con la etiqueta "cumplió / incumplió". El modelo aprende a clasificar nuevos solicitantes como de alto o bajo riesgo.

Ejemplo de regresión:

Una empresa de seguros estima el monto probable de un siniestro futuro en función de características del asegurado.

Desde la mirada de TI, el aprendizaje supervisado exige:

- Datos históricos confiables y correctamente etiquetados.
- Procesos de limpieza y transformación de datos (ETL – Extract, Transform, Load; Extraer, Transformar y Cargar).
- Validación del modelo mediante métricas de desempeño (accuracy, precisión; recall, exhaustividad; mean squared error, error cuadrático medio).

Ventajas organizacionales:

- Alta precisión cuando se dispone de datos de calidad.
- Aplicabilidad directa a procesos de decisión estructurados.

Riesgos:

- Sesgo en los datos históricos (Algorithmic Bias, Sesgo Algorítmico).

- Sobreajuste (Overfitting), cuando el modelo se adapta excesivamente a los datos de entrenamiento y pierde capacidad de generalización.

Para un administrador, el desafío consiste en evaluar la calidad de los datos y comprender que el modelo refleja patrones del pasado, no necesariamente realidades futuras en contextos cambiantes.

3. Aprendizaje no supervisado

El aprendizaje no supervisado se utiliza cuando los datos no están etiquetados. El objetivo no es predecir un resultado específico, sino descubrir estructuras, patrones o agrupaciones dentro del conjunto de datos.

Uno de los métodos más comunes es el clustering (agrupamiento), que identifica grupos homogéneos dentro de una población.

Ejemplo organizacional:

Una empresa de telecomunicaciones segmenta su base de clientes en grupos según comportamiento de consumo, frecuencia de uso y tipo de servicio contratado. No existe una etiqueta previa; el modelo identifica segmentos con características similares.

Otro ejemplo:

Detección de anomalías en transacciones financieras para identificar posibles fraudes sin necesidad de un registro completo de casos etiquetados.

Desde el punto de vista de TI, este enfoque requiere:

- Grandes volúmenes de datos.
- Capacidad de procesamiento intensivo.
- Visualización avanzada de datos (Data Visualization).

Ventajas:

- Permite descubrir oportunidades de negocio no evidentes.
- Útil en etapas exploratorias.

Riesgos:

- Interpretación subjetiva de los resultados.

- Segmentaciones que pueden carecer de sentido estratégico si no se contextualizan.

El administrador debe comprender que el aprendizaje no supervisado es una herramienta exploratoria que requiere interpretación estratégica y validación posterior.

4. Aprendizaje semi-supervisado

El aprendizaje semi-supervisado combina ambos enfoques. Se utiliza cuando solo una parte del conjunto de datos está etiquetada, mientras que la mayoría no lo está.

Este escenario es frecuente en organizaciones donde etiquetar datos implica costos elevados (por ejemplo, revisión manual de documentos, validación humana de imágenes o transacciones).

Ejemplo:

Una empresa tecnológica posee millones de registros de comportamiento de usuarios, pero solo una pequeña parte ha sido clasificada como "abandono" o "no abandono". El modelo utiliza los datos etiquetados para aprender y luego aprovecha la estructura de los datos no etiquetados para mejorar su precisión.

Desde la perspectiva de TI, este enfoque implica:

- Integración entre técnicas supervisadas y no supervisadas.
- Gestión eficiente del almacenamiento masivo.
- Optimización de recursos computacionales.

Ventajas:

- Reduce costos de etiquetado.
- Mejora desempeño respecto al uso exclusivo de datos etiquetados limitados.

Riesgos:

- Propagación de errores si las etiquetas iniciales son incorrectas.
- Complejidad técnica superior.

Para la gestión organizacional, el aprendizaje semi-supervisado representa una solución intermedia que equilibra eficiencia económica y calidad predictiva.

5. Comparación estratégica de los tres enfoques

Desde la perspectiva administrativa y tecnológica, la elección del enfoque depende de:

1. Disponibilidad de datos etiquetados.
2. Objetivo del análisis (predicción vs exploración).
3. Recursos tecnológicos disponibles.
4. Nivel de riesgo asociado a la decisión.

El aprendizaje supervisado es apropiado cuando se requiere precisión predictiva en decisiones críticas (por ejemplo, crédito, seguros, precios).

El aprendizaje no supervisado es útil para descubrimiento de patrones, segmentación y análisis exploratorio.

El aprendizaje semi-supervisado se utiliza cuando los datos etiquetados son escasos pero se desea mantener un enfoque predictivo.

En todos los casos, la organización debe considerar:

- Gobernanza de datos.
- Cumplimiento normativo (por ejemplo, regulaciones de protección de datos).
- Seguridad de la información.
- Explicabilidad del modelo (Explainable AI – XAI, Inteligencia Artificial Explicable).

6. Implicancias para la arquitectura de sistemas

La implementación de modelos de aprendizaje automático requiere:

- Infraestructura en la nube (Cloud Infrastructure).
- Integración con sistemas transaccionales.
- Monitoreo continuo del desempeño del modelo.

- Auditoría tecnológica.

Además, se debe contemplar el ciclo de vida del modelo (Model Lifecycle Management), que incluye:

1. Diseño.
2. Entrenamiento.
3. Validación.
4. Implementación.
5. Monitoreo.
6. Actualización.

El modelo no es un producto estático; requiere mantenimiento continuo.

7. Dimensión ética y de gobernanza

El uso de aprendizaje automático implica riesgos éticos y regulatorios. Entre ellos:

- Discriminación algorítmica.
- Uso indebido de datos personales.
- Falta de transparencia en decisiones automatizadas.

El administrador debe garantizar:

- Controles internos.
- Auditorías independientes.
- Documentación de modelos.
- Políticas de uso responsable de la IA.

La tecnología no sustituye la responsabilidad humana.

Conclusión

El aprendizaje supervisado, no supervisado y semi-supervisado representan tres enfoques fundamentales del aprendizaje automático que permiten a las organizaciones transformar datos en conocimiento estratégico.

Desde la mirada de las Tecnologías de la Información, su implementación no se limita al desarrollo de modelos matemáticos, sino que implica decisiones sobre arquitectura de datos, infraestructura tecnológica, gobernanza, seguridad y cumplimiento normativo.

El futuro profesional en administración debe comprender que estos modelos influyen directamente en la toma de decisiones empresariales. La clave no reside únicamente en la precisión algorítmica, sino en la adecuada integración entre tecnología, estrategia y responsabilidad organizacional.

La ventaja competitiva no surge simplemente de utilizar inteligencia artificial, sino de implementarla con criterio estratégico, calidad de datos y sólidos marcos de control.

Preguntas de autoevaluación

1. ¿Cuál es la diferencia principal entre aprendizaje supervisado y no supervisado en términos de estructura de datos?
 2. ¿En qué situaciones organizacionales resulta más conveniente aplicar aprendizaje semi-supervisado?
 3. ¿Qué riesgos pueden surgir si los datos utilizados en el entrenamiento contienen sesgos históricos?
 4. ¿Por qué la gobernanza de datos es crítica en proyectos de aprendizaje automático?
 5. ¿Qué responsabilidades debe asumir un administrador frente a la implementación de modelos basados en Inteligencia Artificial?
-

Material de Clases

Compilado por **Aníbal M. Mazza Fraquelli** Doctor de la Universidad de Buenos Aires para el uso de sus clases en la Facultad de Ciencias Económicas de la Universidad de Buenos Aires.

Contenidos de esta página

Los contenidos **aquí incluidos integran desarrollos y escritos propios del autor, así como materiales de terceros (documentos, textos, fragmentos, conceptos, imágenes, esquemas, definiciones u otros recursos)**, los cuales son utilizados a título ilustrativo, explicativo o formativo, respetando la normativa vigente en materia de derechos de autor y citando las fuentes cuando corresponde.

La selección, organización, adaptación pedagógica y contextualización de los contenidos constituye un trabajo original del autor, orientado a facilitar los procesos de enseñanza y aprendizaje.

Este material no persigue fines comerciales y su reproducción, total o parcial, queda limitada al ámbito educativo, debiendo preservarse siempre la mención de la autoría y las fuentes originales.

Autorización de uso

Se permite la reproducción, comunicación pública, distribución y utilización total o parcial de los contenidos de su material, en formato físico o digital, con fines exclusivamente educativos, académicos o de divulgación, siempre que se respete la integridad del contenido y se incluya la correspondiente referencia a la fuente y a la autoría.

Las ideas, opiniones e interpretaciones contenidas en este material corresponden exclusivamente al autor.

Queda expresamente excluido cualquier uso con fines comerciales.