

Aprendizaje Automático aplicado a procesos industriales

Machine Learning applied to industrial processes

Rosa Jazmín Trasviña Osorio⁵
rosa.to@purisima.tecnm.mx
<https://orcid.org/0000-0003-0895-2009>



José Ruiz Tamayo⁶
jose.rt@purisima.tecnm.mx
<https://orcid.org/0000-0002-4837-3696>

Erick Rojas Mancera⁷
Erick.rm@purisima.tecnm.mx
<https://orcid.org/0000-0002-5965-9120>

Recibido: 29/11/2024; Aceptado: 9/1/2025

Resumen

En una perspectiva práctica, recolectar una muestra pequeña de datos tiene dos argumentos relevantes: su almacenamiento no es costoso o inconveniente y el procesamiento estadístico es asequible. Sin embargo, actualmente se recolectan datos masivos los cuales requieren del procesamiento de datos mediante algoritmos para su adecuada interpretación. El Aprendizaje Automático (conocido como Machine Learning) relacionado al campo estadístico, consiste en extraer, procesar, reconocer patrones y tendencias a partir de los datos. En este sentido, los algoritmos permiten procesar información útil cuando se generan y obtienen datos masivos. En el presente estudio, se realizó una investigación documental con alcance descriptivo. Es importante reconocer que, además del campo estadístico, el Machine Learning cada vez está siendo más aplicado en campos del conocimiento heterogéneos tales como Negocios, Ingenierías Biomédica, Química y Agricultura. La necesidad de resolver problemas del entorno requiere la incorporación de las metodologías y técnicas de la Inteligencia Artificial para que puedan ser usadas en el campo emergente de la ciencia de los datos en diferentes industrias.

Palabras clave: procesamiento de datos, aprendizaje automático, algoritmos.

Abstract

In a practical perspective, collecting a small sample of data has two relevant arguments: its storage is not expensive or inconvenient, and statistical processing is affordable. However, massive data requires processing using algorithms for proper interpretation. Machine Learning, related to the statistical field, consists of extracting, processing, and recognizing patterns and trends from data. In this sense, algorithms allow us to process useful information when massive data is generated and obtained. In the present study, documentary research was carried out with a descriptive scope. It is important to recognize that, in addition to the statistical field, Machine Learning is increasingly being applied in heterogeneous fields of knowledge such as Business, Biomedical Engineering, Chemistry and Agriculture.

5 Maestría en Ciencias y Tecnología en Ingeniería Industrial y Manufactura, Tecnológico Nacional de México / ITS de Purísima del Rincón, México

6 Doctorado en Ciencias de la Ingeniería, Tecnológico Nacional de México / ITS de Purísima del Rincón, México

7 Doctorado en Ciencias en Robótica y Manufactura Avanzada, Tecnológico Nacional de México / ITS de Purísima del Rincón, México

The need to solve environmental problems requires the incorporation of Artificial Intelligence methodologies and techniques so they can be used in the emerging field of data science in different industries.

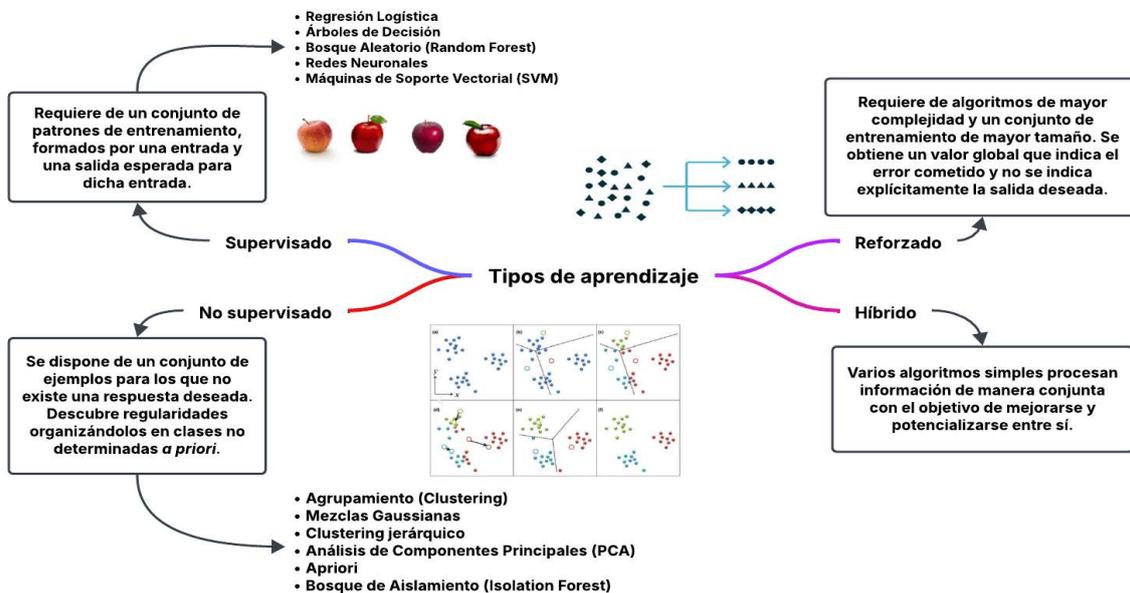
Keywords: data processing, machine learning, algorithms.

Introducción

El Machine Learning (ML) tiene sus cimientos en áreas como las matemáticas, estadísticas y computación (Maisueche, 2019), y hoy en día puede ser considerada como un conjunto de técnicas que pertenecen al campo emergente de la ciencia de los datos. El ML es una rama de la Inteligencia Artificial (IA) que permite la articulación de bases de datos de gran volumen (Big data) y algoritmos para su procesamiento y análisis de la información (Osorio & Enerieth, 2020).

El ML puede clasificarse con diferentes enfoques, tales como el aprendizaje supervisado, no supervisado, por refuerzo e híbrido, en donde el modelo más simple corresponde a un modelo de aprendizaje supervisado. Este tipo de modelo aprende a partir de datos previamente clasificados, la identificación de patrones y toma de decisiones.

Figura 1. Tipos de aprendizaje



Fuente: Elaboración propia

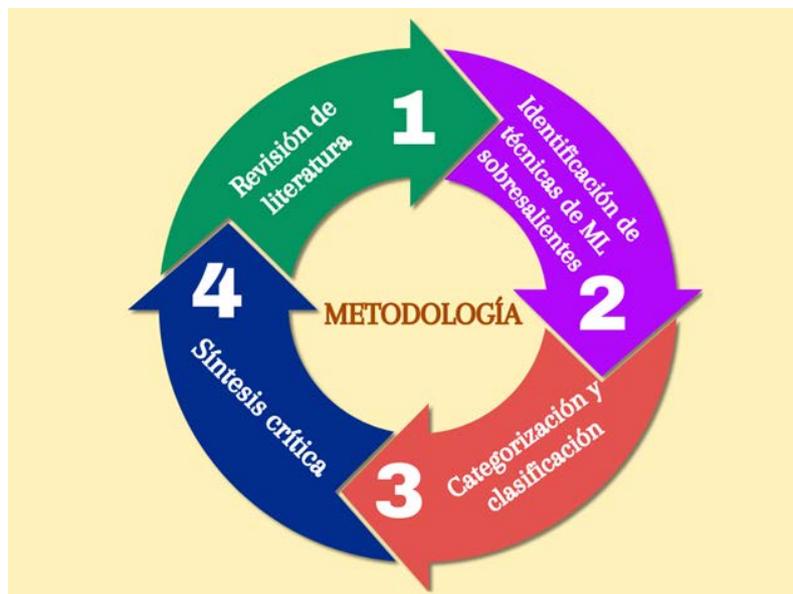
La resolución de problemas utilizando ML ha crecido de manera exponencial en diferentes tipos de industrias en los últimos años (Rojas, 2020). Profesionistas de diversas áreas han mostrado interés en el aprendizaje de ML, lenguajes de programación y su aplicación para generar innovaciones tecnológicas. Existen técnicas de ML que permiten generar datos para obtener más información de un problema de estudio. Algunas de estas técnicas permiten adquirir conocimiento mediante predicciones e identificación de patrones, tales como: la regresión lineal, polinómica, logística, árboles de decisión, redes neuronales artificiales (RNA), redes bayesianas y cadenas de Markov (Rojas, 2020).

El objetivo de este artículo es presentar una descripción de los algoritmos utilizados de ML, así como sus aplicaciones industriales en diferentes ramas o campos de estudio. A continuación, se presenta la metodología sistemática llevada a cabo para realizar el presente estudio.

Metodología

Para la revisión de la literatura del presente artículo, se utilizó una metodología con marco descriptivo. Se recolectaron fuentes secundarias de bases de datos con acceso abierto de: Science Direct, Scopus, SciELO y Latindex. Los pasos para llevar a cabo el estudio fueron los siguientes:

Figura 2. Metodología de la revisión de la literatura



Fuente: Elaboración propia

De manera inicial, se revisó la literatura del estado del arte para la recopilación de fuentes relacionadas al ML y su aplicación en industrias. Posteriormente, se identificaron las técnicas sobresalientes y mayormente aplicadas de ML, así como su explicación y de resultados obtenidos. En seguida, se aplicaron criterios de categorización y/o clasificación en el aprendizaje automático para la organización de la información. Y finalmente, se realizó una síntesis crítica donde se detallaron fortalezas, limitaciones y relaciones de las fuentes secundarias.

Los hallazgos encontrados dentro del estado del arte contribuyeron a la investigación con el objetivo de establecer un sendero que facilite el reconocimiento de las técnicas y aplicación del aprendizaje automático.

Resultados y discusión

Se entiende que un dato por sí solo no representa un contexto, solo un mero reporte. En este sentido, la información <<conformada por un conjunto de datos>> sí brinda un contexto sobre algún tópico en particular. Una vez analizada la información, con cierto sentido, dará paso a la generación de conocimiento.

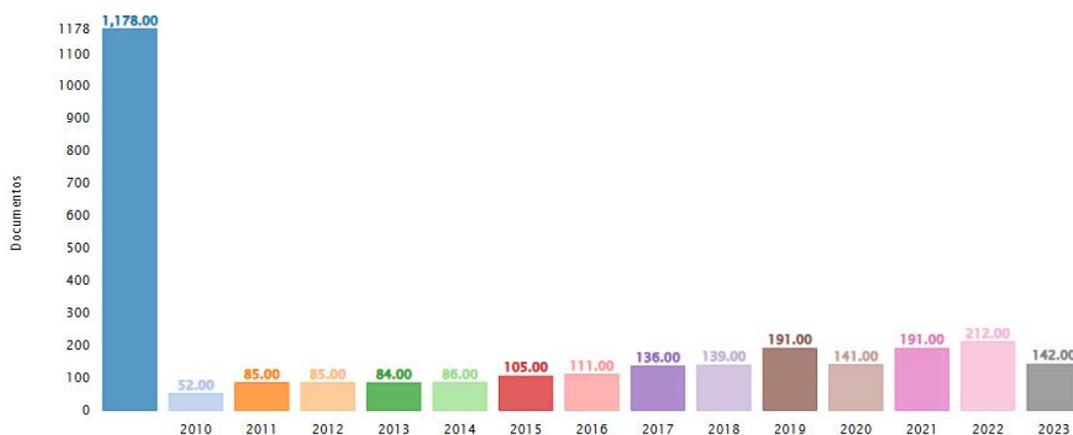
El término de IA es utilizado a sistemas computacionales en analogía a la inteligencia humana por lo que, la cualidad base de la IA, es la creación de algoritmos que imiten funciones cognitivas de los seres humanos.

El uso de técnicas de IA estriba en la necesidad de resolver problemas convencionales de manera alternativa. La dinámica global requiere la modificación de perspectivas con enfoque tradicional. Por lo que, la toma de decisiones con ML ofrece este panorama para predecir las oportunidades emergentes del mercado y tomar mejores decisiones.

Ahora bien, es posible encontrar diferentes productos o modelos derivados de estos programas mediante algoritmos con parámetros previamente definidos, por ejemplo: ecuaciones matemáticas, grafos, una serie de reglas lógicas, etc. En cuanto al aprendizaje automático, algunos algoritmos que se encuentran en la literatura son las redes neuronales artificiales (RNA), Random Forest y Support Vector Machine (SVM). La generación de programas computacionales en el aprendizaje automático facilita la determinación de diferentes métricas al momento de evaluar diferentes tipos de datos.

A continuación, en la siguiente Figura 3, se aprecia el comportamiento de publicaciones relacionadas con ML de Intelligo4Repositorios, el cual genera mapas de resúmenes de palabras clave de diferentes textos. El resultado de la Figura 3 fue generado con la palabra clave "Machine Learning".

Figura 3. Resultados de publicaciones relacionadas a ML.



Fuente: <http://repos.explora-intelligo.info>

En los últimos tiempos se han desarrollado modelos híbridos a base de principios físicos y técnicas de aprendizaje de ML. Y aunque inicialmente el control predictivo estaba basado principalmente en redes neuronales para el modelado de procesos, mediante el aprendizaje profundo (por sus siglas en inglés, Deep Learning) ha transicionado a arquitecturas más complejas de memoria a largo plazo y al uso de redes neuronales convolucionales (Chango, et al., 2024).

La aplicación de ML en las industrias reduce los costos mediante mantenimiento predictivo. A manera de ejemplo, la optimización de los flujos de producción y transporte, así como la reducción de inventarios, se puede presentar como una respuesta cambiante de los clientes en el mercado. Tan solo a nivel operacional, las necesidades específicas del mercado demandan habilidades y necesidades tecnológicas en profesionistas con perfil tecnológico o de ciencias de los datos.

Los algoritmos de ML en la industria farmacéutica pueden ser utilizados para la creación de moléculas potenciales candidatos a fármacos y posteriormente a escala industrial como nuevos medicamentos. Las propiedades de las estructuras químicas pueden caracterizarse mediante modelos predictivos y pueden ser utilizados para predecir la actividad biológica o farmacológica, toxicidad e interacciones. Para la validación de estos modelos, se realizan procesos de supervisión y seguimiento de pruebas in vitro e in vivo, así como de ensayos clínicos (Blasco, et al., 2023).

En otro caso de estudio, para la determinación de la calidad del acero, se depende principalmente de su composición química, procesos de fundición, laminado y de tratamiento térmico. Los defectos de algunas piezas de acero, como las bobinas, puede generarse durante los procesos de recalentado y laminado dando lugar a la aparición de impurezas de óxidos superficiales. El procedimiento para retirar este tipo de defecto es mediante un proceso químico con ácido clorhídrico. Para este caso, se han aplicado modelos de Máquina de Vectores de Soporte (por sus siglas en inglés, SVM) para evaluar el grado de limpieza (González-Marcos, 2017).

En la industria metalúrgica se han presentado estudios relacionados a procesos de mecanizado y predicción de la composición de los productos. Tan solo las reacciones químicas, la naturaleza de las variables de estudio y otros factores aparentemente controlables como la transferencia de masa y calor, hacen complejo el análisis debido a la caracterización no lineal de la transformación de materiales (Maisueche, 2019).

Los temas relacionados a ML no son familiares para todos los profesionistas ya que las técnicas pertenecen al campo de la informática. A continuación, se muestra de manera sintetizada el camino de ML para gente no experta en el tema que tengan la intención de atender una necesidad para cualquier tipo de industria:

Figura 4. Desarrollo de un modelo de ML.



Fuente: <http://repos.explora-intelligo.info>

Para la descripción de la Figura 4, se detallará el estudio de (Burgos-Naranjo, et al., 2021) en el sector alimenticio para la estandarización en la formulación de mermeladas:

- **Etapas 1:** Recolección de datos históricos con una dimensionalidad de 574 filas y 22 columnas y se seleccionan variables de interés de la base de datos, tanto de entrada y salida.
- **Etapas 2:** Los datos fueron normalizados para cada uno de los sabores de las mermeladas.
- **Etapas 3:** Los datos atípicos de la base de datos fueron descartados, al igual que aquellos duplicados. Para el caso de datos faltantes, se sustituyeron por el valor de la mediana de las variables correspondientes (como nota, véase la fuente la fuente para el criterio de imputación).
- **Etapas 4:** Se seleccionaron algoritmos de modelos estadísticos multivariados, lineales generalizados y de aprendizaje automático. Para los casos donde no se pudo normalizar a las variables de respuesta, se adaptaron modelos predictivos como de árboles de decisión y de k vecinos más cercanos (por sus siglas en inglés, KNN).
- **Etapas 5:** Los autores (Burgos-Naranjo, et al., 2021), evaluaron los rendimientos de los algoritmos mediante criterios de análisis estadístico.

En la siguiente Tabla 1, se muestran los resultados obtenidos de los autores para el caso de estudio:

Tabla 1. Error absoluto medio (MAE) de los algoritmos de árboles de decisión y k vecinos más cercanos

Sabor	MAE (DT)	MAE (kNN)
Frutilla	0.91	0.77
Frutimora	0.71	0.58
Mora	0.63	0.55
Piña	0.47	0.59
Guayaba	0.53	0.51

Fuente: Burgos-Naranjo, et al., 2021

Como parte de las problemáticas del sistema agroalimentario, (Ruiz-Tamayo, et al., 2024) evaluaron redes neuronales convolucionales para la detección y clasificación de enfermedades foliares en plantas de maíz (*Zea mays* L.) tales como el VSM, Plaga y Roya, con el objetivo de que los agricultores se beneficien con diagnósticos más rápidos y precisos para aumentar el rendimiento de las cosechas. Los autores utilizaron condiciones controladas en una cabina portátil para tomar muestras de imágenes de alta calidad. Dichas condiciones permitieron capturar información visual de utilidad para la generación de la base de datos. El estudio presentó resultados con precisiones del 90 y 95% para dos tipos de redes neuronales convolucionales. Sin embargo, como parte de las limitaciones, debido a que la base de datos de imágenes estuvo en condiciones específicas de luz y entorno, esto limita a la generalización de los resultados para otras áreas o climas específicos.

El desarrollo de modelos compuestos por diferentes herramientas <<algoritmos>> tiene por objetivo mejorar la predicción de problemas de caso de estudio, para generalizar soluciones e implementar estas metodologías a otros problemas. En el sector de la educación, existen variables complejas que no se pueden medir de manera directa.

(De La Hoz, et al., 2019) desarrollaron una metodología para clasificar y predecir a los usuarios en el desempeño de los exámenes de acuerdo con la interacción en plataforma, para finalmente clasificarlos en función del nivel de conocimiento. Las variables de interés fueron las actividades realizadas por los estudiantes en plataforma: frecuencia de las visitas, tiempo activo y desempeño en exámenes. Los autores señalan que el proceso fue complejo debido a que la plataforma no les permitió obtener la información de manera depurada. Y como parte de las limitaciones, conllevó a la recolección individual de la información y depuración de datos atípicos para posteriormente obtener una base de datos candidata al análisis.

Los resultados de (De La Hoz, et al., 2019) arrojaron que el tiempo del estudiante en plataforma no necesariamente conlleva a que obtengan buen desempeño o mejores notas. Y aunque sus resultados arrojaron precisiones del 91% para KNN, clasificador discriminante lineal del 85% y, 84% para una cuadrática, concluyen con la importancia del papel del docente en identificar los perfiles de los estudiantes para favorecer los procesos de enseñanza-aprendizaje.

A manera de resumen, la aplicación de ML a diferentes problemas favorece el análisis de la información y la evaluación del desempeño para concluir de manera más precisa con la información de entornos específicos. A medida que se presentan restricciones en los datos, la conjunción de algoritmos ayuda a superar los desafíos en diferentes escenarios para garantizar resultados satisfactorios y tomar mejores decisiones.

Conclusiones

El aprendizaje automático se ha consolidado como una herramienta esencial para resolver problemas complejos en diversas áreas, como la salud, las finanzas, el transporte y la tecnología. Su capacidad para analizar grandes volúmenes de datos, identificar patrones y realizar predicciones precisas ha transformado la forma en que las organizaciones toman decisiones estratégicas.

Sin embargo, el éxito de los modelos de ML depende de factores clave, como la calidad de los datos, la selección de características relevantes y el diseño cuidadoso del modelo. De igual importancia, ofrece una amplia gama de oportunidades en la industria, permitiendo optimizar procesos, mejorar la toma de decisiones y desarrollar productos.

- Finalmente, puntualizando con los ejemplos mencionados anteriormente, se concluye lo siguiente:
- En la industria farmacéutica, se acelera el descubrimiento de medicamentos y personaliza tratamientos.
- Se puede predecir la demanda con mayor precisión para ajustar inventarios y reducir desperdicios.
- Es posible realizar diagnósticos más precisos la identificación de enfermedades en cultivos para aumentar el rendimiento de las cosechas.
- Se desarrollan metodologías compuestas para mejorar la predicción y clasificación de la información de las bases de datos.

Referencias bibliográficas

- Blasco, R. F., Ferrándiz, Á. B., Barbero, A., & Barbero, B. (2023). La profesión farmacéutica frente al reto de la inteligencia artificial.
- Chango, D. A. C., Anchatipán, A. D. P., & Bedón, F. R. R. (2024). Análisis del Uso de Machine Learning para Sistema de control predictivo a nivel industrial. *Polo del Conocimiento*, 9(7), 1023-1040.
- De La Hoz, E. J., De La Hoz, E. J., & Fontalvo, T. J. (2019). Metodología de aprendizaje automático para la clasificación y predicción de usuarios en ambientes virtuales de educación. *Información tecnológica*, 30(1), 247-254.
- González-Marcos, A., & Alba-Elías, F. (2017). Machine learning en la industria: el caso de la siderurgia. *Economía industrial*, 55.
- Osorio, A., & Enerieth, N. (2020). El derecho de autor en la Inteligencia Artificial de machine learning (Copyright Law in the Artificial Intelligence of Machine Learning). *La propiedad inmaterial*, (30).
- Rojas, E. M. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E28), 586-599.
- Maisueche Cuadrado, A. (2019). Utilización del Machine Learning en la industria 4.0.
- Naranjo, Á. I. B., Játiva, D. S. V., & Chávez, D. O. N. (2021). Aplicaciones de la industria 4.0 en la estandarización del proceso productivo de las mermeladas. *Ingeniería Investigación y Desarrollo: I2+ D*, 21(1), 39-46.
- Ruiz-Tamayo, J., Antonio-Godínez, Z., Becerra-Mendoza, F. P., Rojas-Mancera, E., Trasviña-Osorio, R. J., Villanueva-Jiménez, L. F., y Almanza-Rangel, S. (2024). Evaluación Comparativa de Redes Neuronales Convolucionales ResNet-50 y MobileNet-v2 para el Diagnóstico de Enfermedades Foliares en Cultivos de Maíz. *Revista Multidisciplinaria de Ciencia Básica, Humanidades, Arte y Educación*, 2(8), 47-54.