



# Tecnológico de Monterrey

**Campus Querétaro**

**Entrega 1.**

Presenta

Rodrigo Antonio Benítez De La Portilla || A01771433

Inteligencia artificial avanzada para la ciencia de datos I

TC3006C.101

Catedrático Benjamín Valdés Aguirre

Fecha de entrega 31/08/25

El Machine Learning es una rama de la inteligencia artificial que permite a los algoritmos aprender automáticamente a partir de los datos, mejorando su desempeño sin necesidad de ser programados explícitamente para cada tarea. Existen diferentes tipos de Machine Learning, entre los cuales destaca el Aprendizaje Supervisado. Este enfoque utiliza datos etiquetados para entrenar modelos capaces de hacer predicciones sobre nuevos datos. En este proyecto, se aplicó el Aprendizaje Supervisado para desarrollar un modelo que predice el precio de cierre del oro, utilizando información histórica y variables relevantes del mercado. (Coursera,2023)

Antes de ponernos a programar el algoritmo, lo primero es entender bien los datos con los que vamos a trabajar. No se trata solo de ver qué información tenemos, sino también de prepararla para que el modelo pueda aprender de la mejor manera posible. Por eso, aplicamos el proceso de ETL, que significa Extracción, Transformación y Carga. Primero, extraemos los datos de la fuente (en este caso, un archivo CSV). Luego, transformamos los datos: limpiamos los valores faltantes, seleccionamos las columnas que nos interesan y normalizamos los valores para que todas las variables estén en una escala similar. Finalmente, cargamos estos datos ya listos en el modelo para que pueda entrenar y hacer predicciones de manera más precisa. (IBM, 2021)

Para entrar en contexto, el archivo CSV que utilicé se llama “financial\_regression.csv” y contiene 3905 registros históricos con información sobre el precio del oro, la plata y otros metales preciosos. Entre las columnas disponibles están los precios de apertura, máximo, mínimo, cierre y el volumen de transacciones para cada metal. Para este proyecto, me enfoqué únicamente en las columnas relacionadas con el oro: 'gold open', 'gold high', 'gold low' y 'gold volume' como variables de entrada, y 'gold close' como variable objetivo a predecir. Otras columnas que no eran relevantes para el modelo las eliminé para simplificar el análisis. Además, convertí los datos seleccionados a listas numéricas y los normalicé, de modo que todas las variables tuvieran una escala similar y el modelo pudiera aprender de manera más eficiente.

El código desarrollado tiene como objetivo predecir el precio de cierre del oro utilizando técnicas de machine learning, específicamente una regresión lineal múltiple implementada desde cero. El proceso inicia con la importación de las librerías necesarias: “numpy” para operaciones matemáticas, “pandas” para la manipulación y análisis de datos, y “matplotlib.pyplot” para la visualización de resultados.

A continuación, se realiza la carga de los datos desde el archivo financial\_regression.csv, el cual contiene información histórica sobre precios y volúmenes de metales preciosos. Se eliminan las filas con valores faltantes para asegurar que el modelo trabaje únicamente con datos completos y confiables. Posteriormente, se seleccionan las columnas relevantes para el análisis, enfocándose en los precios de apertura, máximo, mínimo y el volumen del oro como variables independientes, mientras que el precio de cierre del oro es la variable objetivo que se

busca predecir. Para que el modelo pueda aprender de manera eficiente, se normalizan todas las variables, es decir, se transforman para que tengan media cero y desviación estándar uno. Esto es fundamental porque evita que variables con escalas muy diferentes, como el volumen frente a los precios, dominen el proceso de aprendizaje y permite que el gradiente descendente funcione de manera más estable y rápida.

Una vez preparados los datos, se inicializan los parámetros del modelo: los coeficientes o pesos para cada variable independiente, el sesgo o bias, y la tasa de aprendizaje, que determina cuánto se ajustan los parámetros en cada iteración. Se definen tres funciones clave: la función de hipótesis, que calcula la predicción del modelo para una entrada dada; la función de costo, que mide el error cuadrático medio entre las predicciones y los valores reales; y la función de actualización, que ajusta los parámetros del modelo utilizando el método de gradiente descendente para reducir el error en cada iteración.

El entrenamiento del modelo se realiza a lo largo de un número determinado de épocas, durante las cuales se calcula el error, se actualizan los parámetros y se almacena el historial de errores para su posterior análisis. El proceso se monitorea imprimiendo el error cada cierto número de épocas, lo que permite observar cómo mejora el modelo con el tiempo.

Después de entrenar el modelo de regresión lineal, se obtienen varios resultados que permiten evaluar su desempeño. Uno de los principales indicadores es el error cuadrático medio (MSE), que muestra qué tan lejos, en promedio, están las predicciones del modelo respecto a los valores reales. Un MSE bajo indica que el modelo está haciendo buenas predicciones, mientras que un MSE alto sugiere que el modelo necesita mejoras, ya sea en los datos, en la selección de variables o en los parámetros de entrenamiento.

En las gráficas generadas, se puede observar cómo las predicciones del modelo se comparan con los valores reales del precio de cierre del oro. Si los puntos de la gráfica de predicciones vs valores reales se alinean cerca de la línea ideal (la línea roja), significa que el modelo está prediciendo correctamente. Además, la curva de error muestra cómo el MSE disminuye a medida que avanza el entrenamiento, lo que indica que el modelo está aprendiendo y mejorando con cada época. El histograma de errores absolutos permite ver la distribución de los errores: si la mayoría de los errores son pequeños, el modelo es preciso; si hay muchos errores grandes, el modelo podría necesitar ajustes.

## Evaluación del modelo de regresión

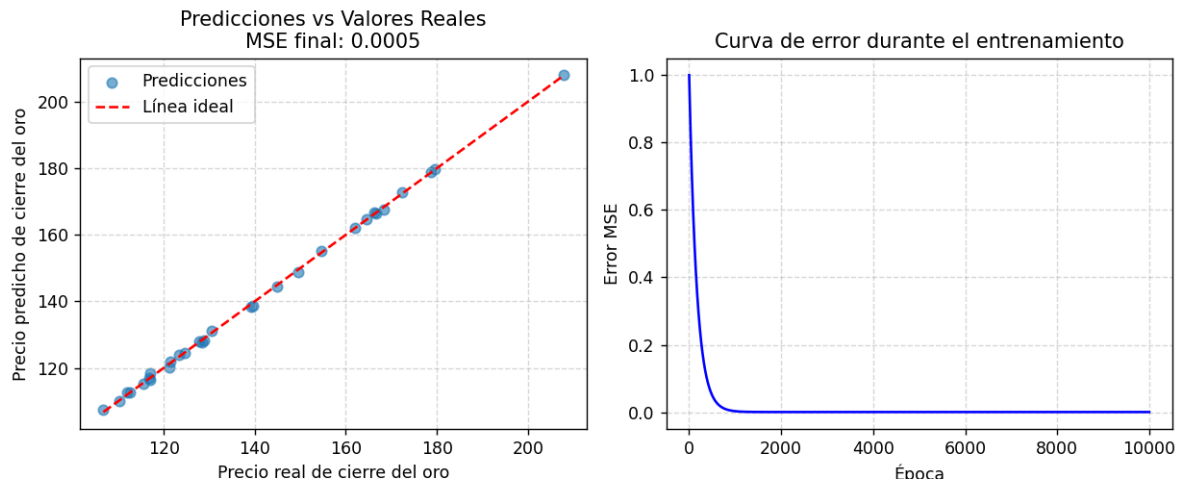


Ilustración 1 Gráficas de resultados

Por último, pero no menos importante, en cuanto a la media (mean) y la desviación estándar (std), estos valores se utilizan durante la normalización de los datos. La media representa el valor promedio de una variable, mientras que la desviación estándar indica cuánto varían los datos respecto a ese promedio. Normalizar los datos usando la media y la desviación estándar asegura que todas las variables tengan una escala similar, lo que facilita el aprendizaje del modelo y evita que variables con valores mucho mayores influyan desproporcionadamente en el resultado. Al final del proceso, tanto las predicciones como los valores reales se desnormalizan utilizando estos mismos valores, para que los resultados sean interpretables en la escala original de los precios del oro.

## Referencias

3 tipos de machine learning que debes conocer. (2023). Coursera.

<https://www.coursera.org/mx/articles/types-of-machine-learning>

IBM. (2021, October 4). ETL. Ibm.com. [https://www.ibm.com/es-](https://www.ibm.com/es-es/think/topics/etl#:~:text=ETL%20(que%20significa%20extraer%2C%20transformar,base%20de%20datos%20de%20destino)

[es/think/topics/etl#:~:text=ETL%20\(que%20significa%20extraer%2C%20transformar,base%20de%20datos%20de%20destino](https://www.ibm.com/es-es/think/topics/etl#:~:text=ETL%20(que%20significa%20extraer%2C%20transformar,base%20de%20datos%20de%20destino)