**UNIVERSIDAD DE CUNDINAMARCA**

**FACULTAD DE INGENIERÍA**

**Clasificación Iris**

**801**

**Francisco Antonio Vargas**

**Juan Manuel Pastor**

**Estudiantes**

**Machine Learning**  
**Alexander Espinosa**

**Docente**

**Facatativá**

**16/09/2025**

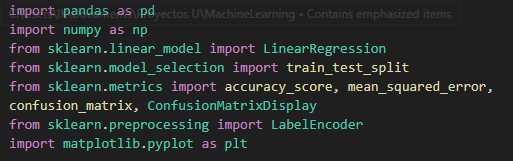
**Clasificación Iris**

**Introducción**

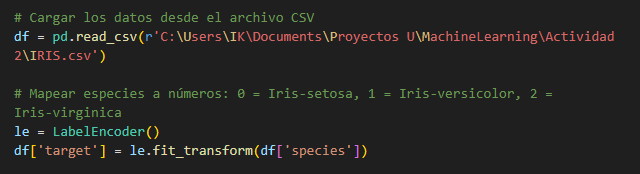
Dentro del avance de la tecnología el aprendizaje de la maquina ha sido el enfoque principal de la última década, ya que en la actualidad la inteligencia artificial se convirtió en una herramienta útil y accesible que facilita la vida de las personas, sin embargo, dentro de esta área se presentan diversos conflictos a resolver, por ejemplo, la interpretación y el análisis para la toma de decisiones, son cosas que una maquina no es capaz de hacer, o por lo menos no sin la instrucción correcta y el dataset apropiado.

En este documento se pondrán en práctica conceptos y técnicas como la regresión lineal para la clasificación de la flor de iris con el fin de mostrar la diferencia entre una clase y otra dependiendo del alto y ancho de los pétalos.

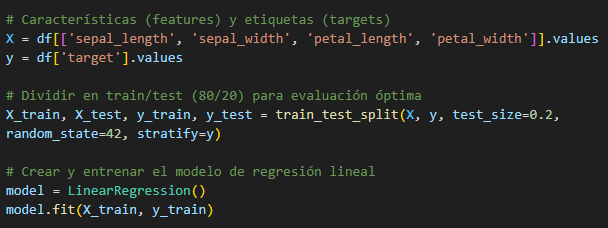
**Desarrollo**



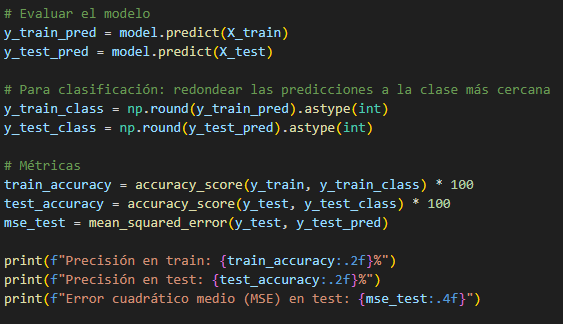
A la hora de elaborar el programa lo primero y mas importante es importar todas las librerías que se planea usar para el funcionamiento del mismo, entre estas se encuentra pandas para el manejo de los datos, también se encuentra numpy para cálculos numéricos rápidos, de scikit learn se necesitara la regresión lineal el cual es el modelo que se va a manejar para aprender los patrones de los datos, además se utilizara el train\_test\_split el cual se encargara de separar una parte de los datos para entrenamiento y la otra para poder evaluar si aprendió bien, se incluirá también diversas métricas para asegurarse de que tan bueno es el modelo, y por ultimo se utilizara matplotlib para poder crear y mostrar las gráficas.



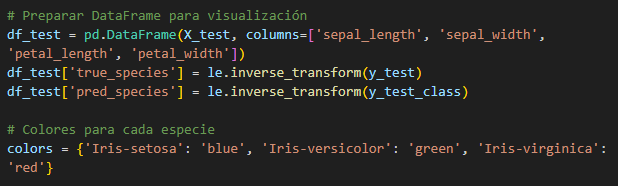
En la siguiente sección se buscara cargar el dataset IRIS de su respectiva ubicación en el directorio correspondiente, primero se lee y luego se transforma en un dataframe, este dataset tiene 150 filas con 4 medidas y la especie como etiqueta. Posteriormente se convertiran los diferentes tipos de especies en numeros para que de esta forma sea posible realizar calculos mas adelante, es decir, que luego de encontrar los tipos de especies como la iris-setosa la transformara en un 0, la iris-versicolor en un 1, y la iris-virginia en un 2 para agregar una nueva columna al dataframe llamada target.



A continuación lo que se debe realizar es una separacion entre las entradas y las salidas, para ello se toman las variables X e Y en donde X reflejara las entradas que en este caso serian las medidas y las convierte en un array, luego la Y reflejara la salida que es la columna target anteriormente creada. Con esto es posible dividir los datos en una proporcion 80/20, es decir, que simulando un escenario real se tomaran unos valores para entrenamiento y otros para pruebas, luego de preparar los datos se entrenara el modelo con el uso de la funcion de regresion lineal quien asume que los valores esperados en Y se pueden predecir con un linea recta que separe los datos de X.



Para evaluar el modelo se utilizan los datos previamente separados tanto para X como para Y, se aplica la formula y esto devuelve valores decimales, a los que se les aplica una función de redondeo para poder determinar a que clase pertenecen, es decir, si el valor resultante es de 0.7 que aproxime a 1, transformando así la regresión continua a una discreta. Para comprobar el nivel de precisión en el entrenamiento y en la evaluación posterior se tomaron en cuenta el porcentaje de predicción correcto contrastando la salida y las entradas obtenidas junto con las deseadas, y tambien se calculo el error cuadratico medio que en el caso de ser bajo indicaria un buen resultado.



Luego se decide acomodar únicamente los datos de la evaluación realizada para la visualización en las gráficas. Se tomo la decisión de realizar 5 graficas, la primera compara la longitud de los pétalos con el ancho de los pétalos, mostrando como la especie real y la predicción realizada se superponen, la segunda grafica es similar a la primera, con la diferencia que esta contrasta el ancho y la longitud de los sépalos, para la tercer grafica se realizo una matriz en donde se evidencia el 100% de efectividad de la evaluación debido a que en todos los casos la predicción acierta con la realidad, como penúltima grafica se realizo un diagrama entre las predicciones y valores reales, en donde cada una de los puntos azules (reales en X, y predicciones en Y) se acerca a la línea roja que representaría la perfección, y como ultima grafica se puede observar un histograma de los residuos o las diferencias entre el valor predicho y el real centrado en el 0 lo que quiere decir que no tiene sesgo.

**Resultados**

