

Taller Machine learning

Presentado por:

Alexander Rendón Londoño.

Braian Estrada

Presentado a:

Juan Camilo

Programa:

Ingeniería Informática

Universidad Autónoma Latinoamericana
(UNAULA)

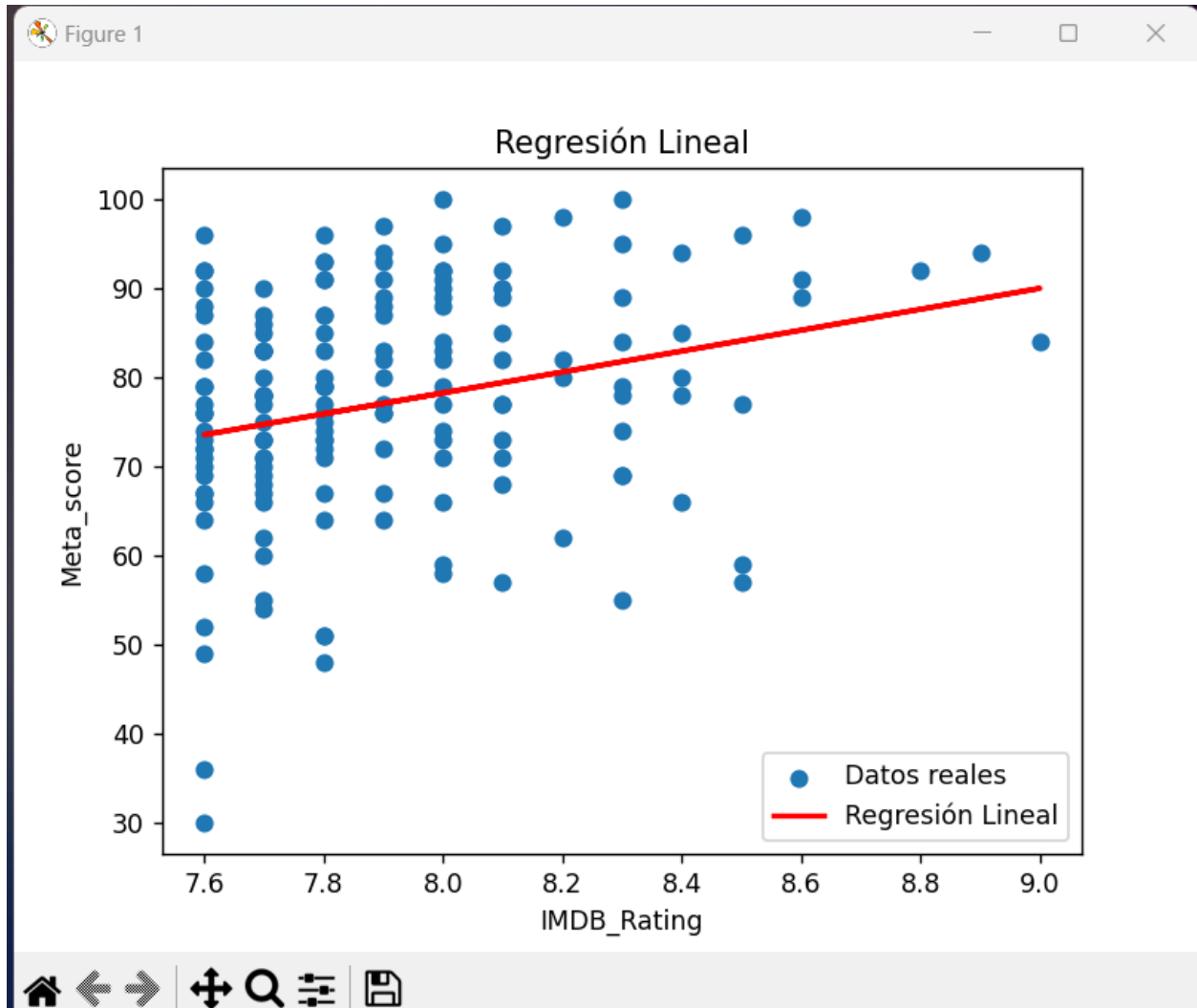
2023-2

Interpretaciones:

Regresión Lineal:

- Coeficiente R^2 de Regresión Lineal: El coeficiente R^2 es una medida de la bondad del ajuste del modelo de regresión lineal. En este caso, tiene un valor de aproximadamente 0.0685. Esto significa que el modelo de regresión lineal no explica la variabilidad en los datos de Meta_score de manera efectiva, ya que R^2 es bastante bajo. Puede deberse a que la relación entre IMDB_Rating y Meta_score no es lineal.
- Error Cuadrático Medio (MSE) de Regresión Lineal: El MSE es una medida del error promedio entre los valores predichos por el modelo y los valores reales. Con un valor de aproximadamente 154.08, indica que el modelo tiene un alto error cuadrático medio, lo que sugiere que las predicciones no son precisas en general. Esto podría deberse a la naturaleza compleja y no lineal de los datos.
- Gráfico de Regresión Lineal: El gráfico muestra los datos reales (puntos dispersos) y la línea de regresión lineal (en rojo). La dispersión de los puntos alrededor de la línea indica que el modelo no se ajusta bien a los datos y no puede capturar la relación entre las variables de manera efectiva.

```
Seleccione un modelo:
1. Regresión Lineal
2. Árboles de Decisión
3. K-Means
4. Salir
Ingrese el número de la opción que desea: 1
Coeficiente  $R^2$  de Regresión Lineal: 0.06846863387898594
Error cuadrático medio (MSE) de Regresión Lineal: 154.08418284060935
```



Árboles de Decisión:

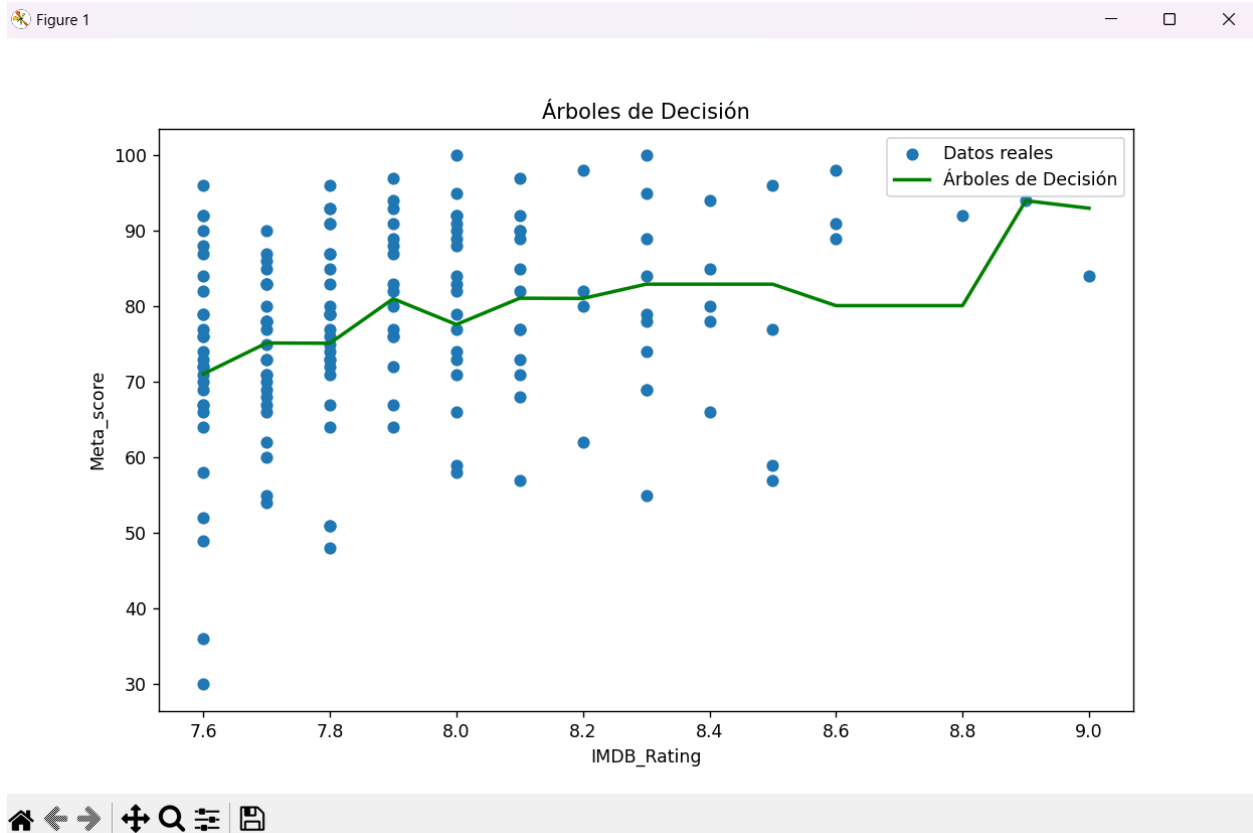
- Coeficiente R^2 de Árboles de Decisión: El coeficiente R^2 para el modelo de Árboles de Decisión tiene un valor de aproximadamente 0.0643, similar al de la regresión lineal. Esto indica que, al igual que la regresión lineal, el modelo de Árboles de Decisión no explica la variabilidad en los datos de Meta_score de manera efectiva.
- Error Cuadrático Medio (MSE) de Árboles de Decisión: El MSE para Árboles de Decisión es de aproximadamente 154.77, lo que también indica un alto error y que las predicciones no son precisas. Esto puede deberse a que el modelo no es lo suficientemente complejo o

que la relación entre las variables no se puede capturar de manera óptima mediante un árbol de decisión con una profundidad limitada.

- Gráfico de Árboles de Decisión: El gráfico muestra los datos reales y la curva generada por el modelo de Árboles de Decisión. Aunque la curva sigue una tendencia general, no se ajusta perfectamente a los datos, lo que sugiere que el modelo no puede modelar con precisión la relación entre las variables.

```

Seleccione un modelo:
1. Regresión Lineal
2. Árboles de Decisión
3. K-Means
4. Salir
Ingrese el número de la opción que desea: 2
Coeficiente R^2 de Árboles de Decisión: 0.06429931833852864
Error cuadrático medio (MSE) de Árboles de Decisión: 154.7738284559288
    
```

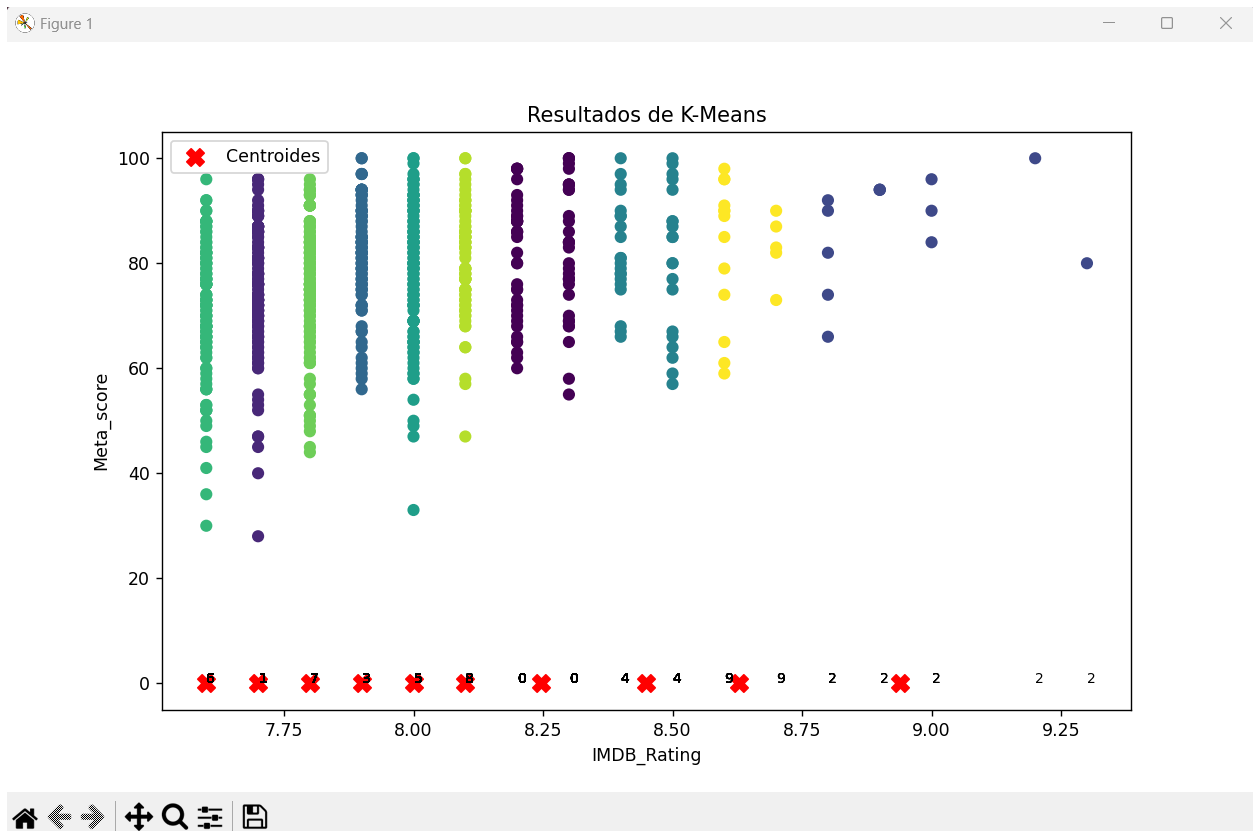


K-Means:

- Puntuación de Silueta de K-Means: La puntuación de silueta es una medida de cuán cerca están los puntos de un cluster entre sí en comparación con los otros clusters. Con un valor de aproximadamente 0.927, indica que los puntos están muy cerca de sus respectivos clusters y relativamente lejos de otros clusters. Esto sugiere que el algoritmo K-Means ha agrupado los datos de manera efectiva en 10 clusters basados en la característica IMDB_Rating.
- Gráfico de K-Means: El gráfico muestra los datos de IMDB_Rating y Meta_score coloreados por cluster. Los puntos rojos representan los centroides de cada cluster. Los datos están bien agrupados alrededor de los centroides, lo que confirma la efectividad de K-Means en la agrupación de datos en este espacio de características unidimensional.

```

Seleccione un modelo:
1. Regresión Lineal
2. Árboles de Decisión
3. K-Means
4. Salir
Ingrese el número de la opción que desea: 3
Puntuación de silueta de K-Means: 0.9269140240892139
  
```



Diferencia de Precisión:

- La diferencia en la precisión entre los modelos de Regresión Lineal y Árboles de Decisión es mínima y ambos tienen un bajo R^2 y un alto MSE. Esto sugiere que ninguno de los dos modelos es adecuado para predecir `Meta_score` en función de `IMDB_Rating`. Ambos modelos no pueden capturar la relación entre estas dos variables de manera efectiva debido a su simplicidad o linealidad.
- En contraste, K-Means tiene una puntuación de silueta bastante alta, lo que sugiere que ha agrupado efectivamente los datos en clusters basados en `IMDB_Rating`. Sin embargo, no está diseñado para predecir valores numéricos como `Meta_score`, sino para agrupar datos. Por lo tanto, la diferencia en la precisión se debe a que K-Means aborda un tipo diferente de problema y no se puede comparar directamente con los modelos de regresión aun así vemos que en cuanto a precisión el k-means lo hace de manera más adecuada.