1. Implementar algoritmos de árboles binarios

La interpretación del árbol de decisión te permite entender cómo el modelo toma decisiones basadas en las características del dataset. Las métricas de rendimiento y la visualización del árbol proporcionan una visión clara de la precisión del modelo y de las reglas de decisión subyacentes.

Cada camino desde la raíz hasta una hoja representa una regla de decisión. Por ejemplo:

Si TotalCases > 50,000 y TotalDeaths > 1,000, entonces ActiveCases es High.

Si TotalCases <= 50,000, entonces ActiveCases es Low.

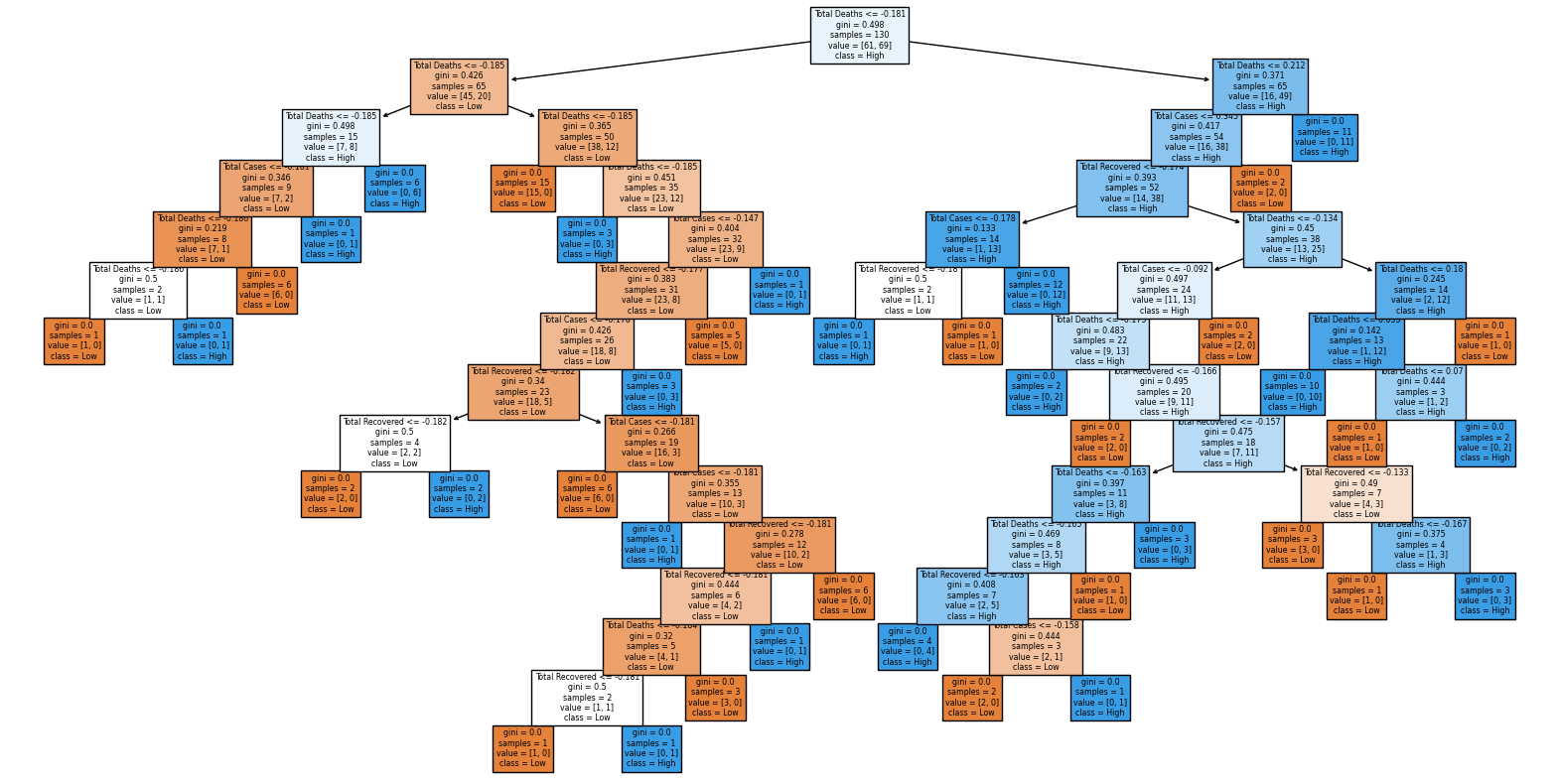
Interpretación:

Si TotalCases > 50,000 y TotalDeaths > 1,000, entonces ActiveCases es High.

Si TotalCases > 50,000 y TotalDeaths <= 1,000, entonces ActiveCases es Low.

Si TotalCases <= 50,000, entonces ActiveCases es Low.

La profundidad del árbol indica cuántos niveles de decisión se tomaron. Un árbol más profundo puede capturar más complejidades, pero también corre el riesgo de sobreajustarse a los datos de entrenamiento.



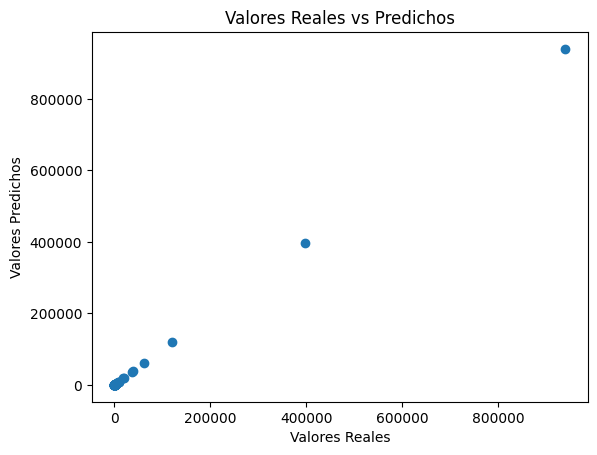
1. Implementar algoritmos de regresión lineal

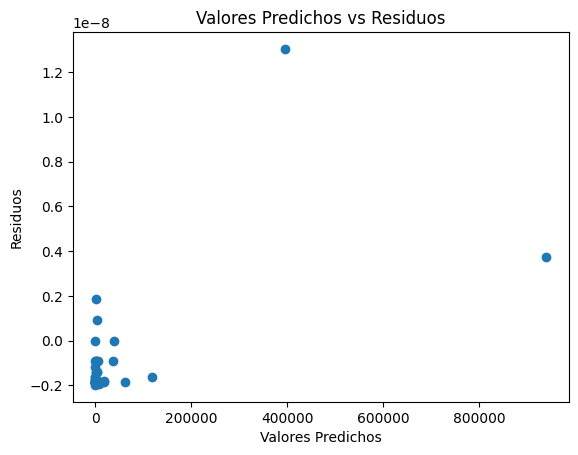
Interpretar los resultados de una regresión lineal implica analizar tanto las métricas de rendimiento como los gráficos de validación. Un buen R² y un bajo MSE indican un modelo bien ajustado, mientras que los gráficos de residuos ayudan a identificar cualquier problema en las predicciones del modelo.

Interpretación de los Gráficos

* Valores Reales vs. Valores Predichos: Si los puntos están cerca de la línea y=xy=x, indica buenas predicciones. Si hay mucha dispersión, el modelo tiene errores significativos.
* Gráfico de Residuos: Si los residuos están distribuidos aleatoriamente alrededor de cero, es una señal de que el modelo está bien ajustado. Patrones en los residuos pueden sugerir problemas como:
  + Tendencias: Podrían indicar relaciones no lineales no capturadas por el modelo.
  + Heterocedasticidad: Si la variabilidad de los residuos cambia con los valores predichos, puede ser necesario transformar las variables o usar un modelo más complejo.

Este gráfico muestra los residuos (diferencias entre los valores reales y predichos) en función de los valores predichos. Ayuda a identificar patrones en los errores que pueden indicar problemas con el modelo, como heterocedasticidad (variabilidad no constante) o falta de ajuste en ciertas regiones.





1. Implementar algoritmos de regresión logística

La regresión logística es un algoritmo de aprendizaje supervisado utilizado para tareas de clasificación binaria. La regresión logística modela la relación entre las características independientes (XX) y una variable dependiente binaria (yy) usando la función sigmoide:

P(y=1∣X)=11+e−(β0+β1X1+β2X2+⋯+βnXn)P(y=1∣X)=1+e−(β0​+β1​X1​+β2​X2​+⋯+βn​Xn​)1​

Aquí, P(y=1∣X)P(y=1∣X) es la probabilidad de que yy sea 1 dado XX, y ββ son los coeficientes del modelo.

Durante el entrenamiento, el algoritmo ajusta los coeficientes ββ minimizando la función de pérdida logarítmica, también conocida como pérdida logarítmica binaria.

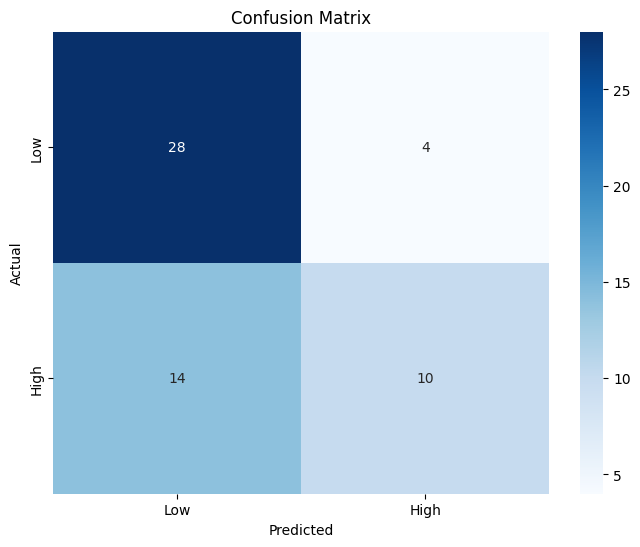
Las predicciones se realizan calculando la probabilidad de la clase positiva y asignando la clase basada en un umbral (por defecto 0.5). Es decir:

1 & \text{si } P(y=1|X) > 0.5 \\

0 & \text{si } P(y=1|X) \leq 0.5

\end{cases} \]

La matriz de confusión proporciona una visualización clara de los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Esto es útil para identificar qué tipo de errores está cometiendo el modelo.



1. Implementar algoritmos de clustering

Los clusters identificados permiten categorizar a los países según su situación respecto a la pandemia. Esta información puede ser útil para:

* Análisis Comparativo: Comparar la eficacia de las políticas de salud pública entre diferentes grupos de países.
* Toma de Decisiones: Identificar países con situaciones similares para adoptar estrategias comunes.
* Estudios Epidemiológicos: Realizar estudios más detallados sobre los factores que contribuyen a las diferencias en la propagación y recuperación de COVID-19.

Interpretación

* Visualización: Observa cómo se agrupan los países en el gráfico de dispersión.
* Estadísticas Descriptivas: Examina las características promedio de cada cluster.
* Comparación: Identifica patrones y diferencias entre los clusters.
* Aplicaciones Prácticas: Utiliza la información para análisis y toma de decisiones.

Interpretación de la Separación de Clusters

La separación de los clusters en el gráfico de dispersión y las diferencias en las estadísticas descriptivas. Los países se agrupan en diferentes clusters según la gravedad y la gestión de la pandemia. Si un cluster tiene más puntos (países) que otro, indica que esa situación particular (por ejemplo, un número moderado de casos) es más común entre los países. Los clusters muestran diferencias claras en las estadísticas, lo que ayuda a identificar patrones en la propagación y gestión de COVID-19.

