



INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES
DE MONTERREY

ESCUELA DE INGENIERÍA Y CIENCIAS

Inteligencia artificial avanzada para la ciencia de datos I TC3006C.102

Reporte: Modelo de Machine Learning para predicción

Autor

J. Eduardo Corrales Cardoza - A01742328

Profesor

Jesús Adrián Rodríguez Rocha

Fecha

07/09/2024

Análisis y Reporte sobre el Desempeño del Modelo

1. Introducción

Este reporte presenta un análisis exhaustivo del desempeño del modelo de regresión neuronal en la predicción del Producto Interno Bruto per cápita (GDP per capita) utilizando un conjunto de datos que incluye soporte social, expectativa de vida saludable y libertad de toma de decisiones. El modelo en cuestión se ha desarrollado empleando una red neuronal con múltiples capas densas y se ha entrenado utilizando un conjunto de datos dividido en entrenamiento, validación y prueba para asegurar una evaluación completa de su rendimiento.

El objetivo principal de este análisis es evaluar la eficacia del modelo en cuanto a su capacidad predictiva y generalización. Para lograr esto, se examinarán las métricas de desempeño, como la pérdida (loss), el error cuadrático medio (MSE) y el coeficiente de determinación (R^2), en los conjuntos de datos de entrenamiento, validación y prueba. Además, se buscará identificar posibles problemas de sesgo y varianza que puedan estar presentes, y se propondrán mejoras para optimizar el modelo a través de técnicas de regularización y ajuste de parámetros.

El análisis se centrará en cómo el modelo maneja la relación entre las características sociales y económicas y el GDP per capita, evaluando su precisión en la predicción y su capacidad para generalizar a datos no vistos. Con base en estos resultados, se ofrecerán recomendaciones para mejorar el rendimiento del modelo y su aplicabilidad en escenarios del mundo real.

2. Descripción del Modelo

El modelo está compuesto por los siguientes elementos:

1. Datos:

- **Características:** "Social support", "Healthy life expectancy", "Freedom to make life choices".
- **Objetivo:** "GDP per capita".

2. Preparación de Datos:

- Carga del conjunto de datos desde un archivo CSV.
- Selección de columnas relevantes.

- Conversión a tipo np.float32.
- División en conjuntos de entrenamiento, validación y prueba.

3. Arquitectura del Modelo:

- **Tipo:** Red neuronal secuencial (Sequential).
- **Capas:**
 - Capa de entrada con 8 unidades y activación ReLU.
 - Varias capas ocultas (128, 128, 128, 128, 64, 64, 64, 8 unidades) con activación ReLU.
 - Capa de salida con 1 unidad.
- **Inicializadores:** HeUniform para pesos y ones para sesgos.

4. Compilación del Modelo:

- **Optimizador:** Adam con una tasa de aprendizaje de 0.01.
- **Pérdida:** Error cuadrático medio (mean_squared_error).

5. Callbacks:

- **EarlyStopping:** Detiene el entrenamiento si el rendimiento en validación no mejora durante 30 épocas.
- **ReduceLROnPlateau:** Reduce la tasa de aprendizaje si el rendimiento se estabiliza durante 20 épocas.
- **MetricsCallback:** Registra las métricas de pérdida y error cuadrático medio en entrenamiento, validación y prueba.

6. Entrenamiento:

- **Épocas:** Hasta 1000.
- **Tamaño del lote:** 40.
- **Datos:** Utiliza conjuntos de entrenamiento y validación.

7. Evaluación y Visualización:

- **Predicciones:** Genera predicciones para los conjuntos de entrenamiento, validación y prueba.

- **Métricas:** Calcula y muestra la pérdida, el error cuadrático medio y el R^2 para cada conjunto.
- **Gráficas:** Plotea la pérdida y el error cuadrático medio a lo largo de las épocas.

3. Separación y Evaluación del Modelo

3.1. División de Datos

- **Conjunto de Entrenamiento:** 70% del total de observaciones. Este conjunto se usa para ajustar los parámetros del modelo y aprender de los datos.
- **Conjunto de Validación:** 15% del total de observaciones. Se usa para ajustar hiperparámetros y evitar el sobreajuste durante el entrenamiento.
- **Conjunto de Prueba:** 15% del total de observaciones. Se usa para evaluar la capacidad de generalización del modelo después del entrenamiento.

3.2. Evaluación Inicial

• Conjunto de Entrenamiento:

- **Error Cuadrático Medio (MSE): 0.1844**
 - **Interpretación:** Un MSE de 0.1844 indica que, en promedio, el cuadrado del error entre las predicciones del modelo y los valores reales es 0.1844. Este valor es relativamente alto, sugiriendo que el modelo tiene un error significativo en los datos de entrenamiento. Un MSE más bajo normalmente indicaría un mejor ajuste a los datos.
- **Raíz del Error Cuadrático Medio (RMSE): $\sqrt{0.1844} \approx 0.429$**
 - **Interpretación:** El RMSE es la raíz cuadrada del MSE y proporciona una medida del error en las mismas unidades que la variable objetivo. Un RMSE de 0.429 sugiere que, en promedio, las predicciones se desvían en 0.429 unidades de los valores reales. Este valor alto indica que el modelo está cometiendo errores considerables en el conjunto de entrenamiento.
- **Coefficiente de Determinación (R^2): -0.3226**
 - **Interpretación:** Un R^2 de -0.3226 es negativo, lo que indica que el modelo no solo falla en explicar la variabilidad en los datos de

entrenamiento, sino que es peor que simplemente usar la media de los datos como predictor. Esto es una señal clara de que el modelo no está ajustando bien a los datos.

- **Conjunto de Validación:**

- **Error Cuadrático Medio (MSE): 0.1916**

- **Interpretación:** El MSE en el conjunto de validación es ligeramente mayor que en el de entrenamiento, lo que puede sugerir que el modelo está sobreajustado. Un MSE de 0.1916 indica que el rendimiento del modelo disminuye en datos no vistos durante el entrenamiento, lo que es indicativo de sobreajuste.

- **Raíz del Error Cuadrático Medio (RMSE): $\sqrt{0.1916} \approx 0.437$**

- **Interpretación:** El RMSE en el conjunto de validación es más alto que en el de entrenamiento, indicando que las predicciones en los datos de validación se desvían en promedio 0.437 unidades de los valores reales. Esto refuerza la idea de que el modelo tiene dificultades para generalizar a nuevos datos.

- **Coeficiente de Determinación (R^2): -0.0049**

- **Interpretación:** Un R^2 de -0.0049 en el conjunto de validación es prácticamente cero y negativo, lo que sugiere que el modelo no mejora en comparación con usar la media de los datos para hacer predicciones. Esto es una señal de que el modelo no está generalizando bien y tiene un desempeño muy pobre.

- **Conjunto de Prueba:**

- **Error Cuadrático Medio (MSE): 0.2118**

- **Interpretación:** El MSE en el conjunto de prueba es aún mayor que en los conjuntos de entrenamiento y validación, lo que sugiere que el modelo tiene problemas significativos para generalizar a nuevos datos. Un MSE de 0.2118 indica un alto nivel de error en los datos no vistos.

- **Raíz del Error Cuadrático Medio (RMSE): $\sqrt{0.2118} \approx 0.460$**

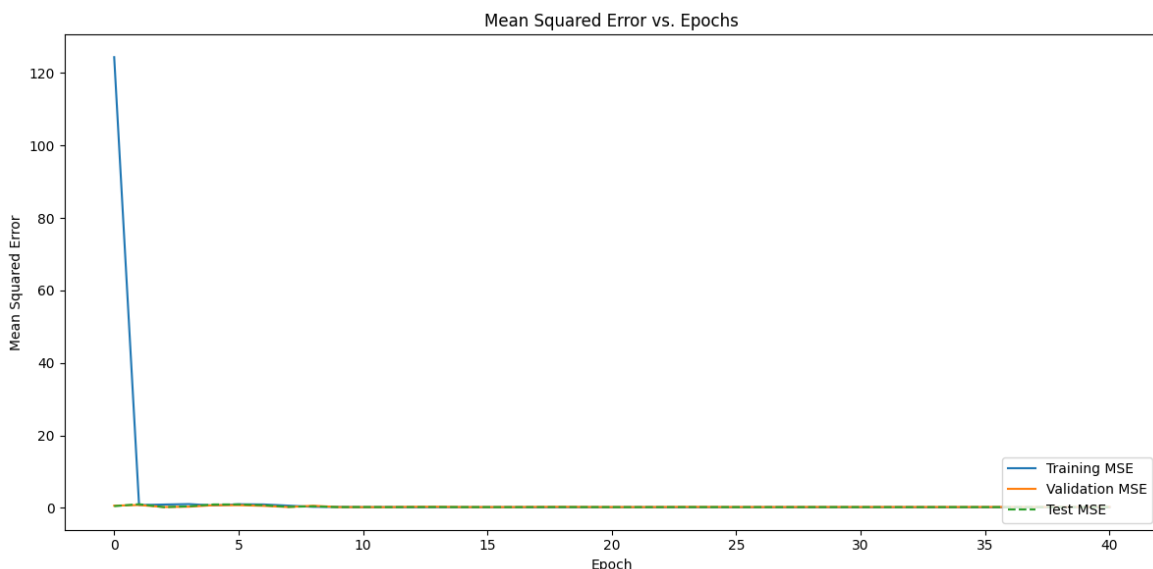
- **Interpretación:** El RMSE en el conjunto de prueba es 0.460, que es mayor que en los conjuntos de entrenamiento y validación. Esto

confirma que el modelo está cometiendo errores importantes al hacer predicciones en datos nuevos.

- **Coeficiente de Determinación (R^2): -0.1567**
 - **Interpretación:** Un R^2 de -0.1567 en el conjunto de prueba es negativo y sugiere que el modelo no solo falla en explicar la variabilidad en los datos de prueba, sino que su rendimiento es inferior a simplemente usar la media de los datos como predicción. Este valor indica que el modelo tiene un desempeño extremadamente pobre en datos no vistos.

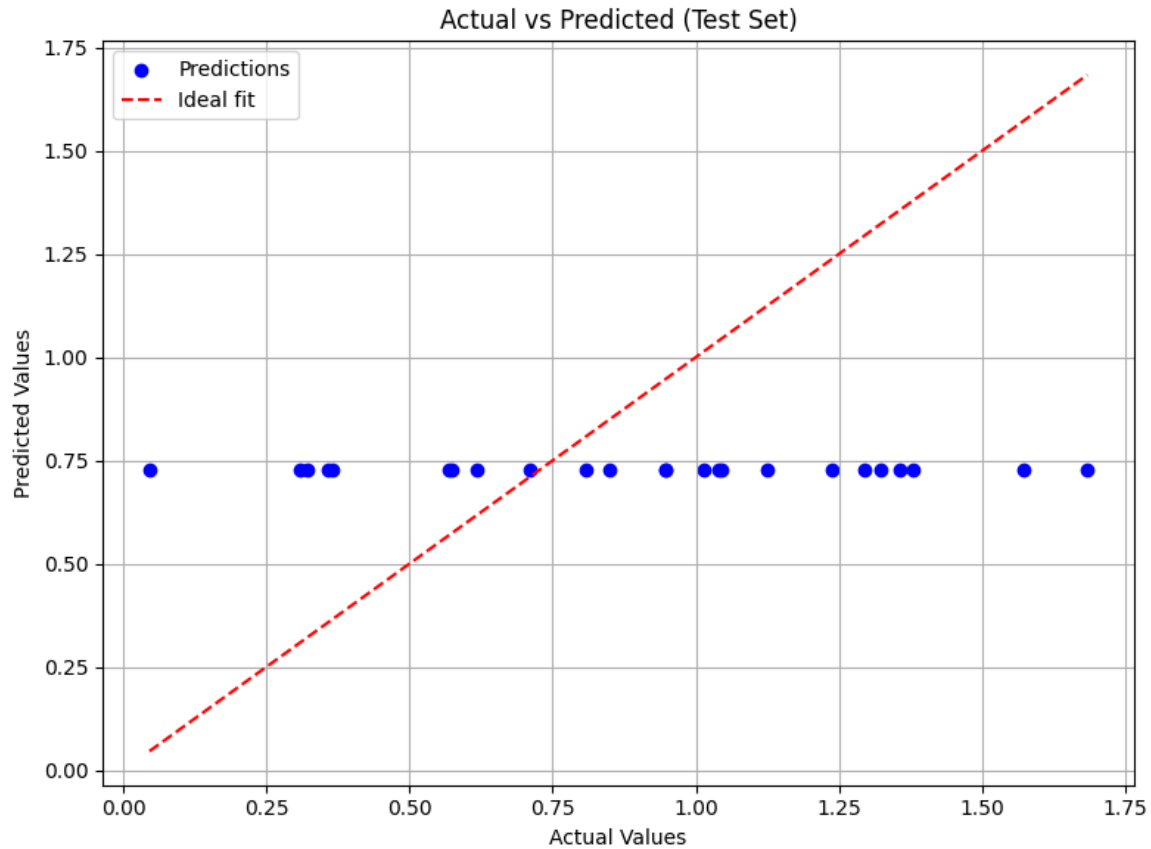
3.3. Gráficas Comparativas

- **Curvas de Aprendizaje:**
 - Las gráficas de MSE y RMSE durante el entrenamiento y validación permiten visualizar cómo el error del modelo cambia con el tiempo. Un descenso en MSE y RMSE durante el entrenamiento indica que el modelo está aprendiendo y ajustando bien a los datos. La comparación entre entrenamiento y validación puede revelar sobreajuste si el MSE en validación se estabiliza o aumenta mientras el MSE en entrenamiento continúa disminuyendo.



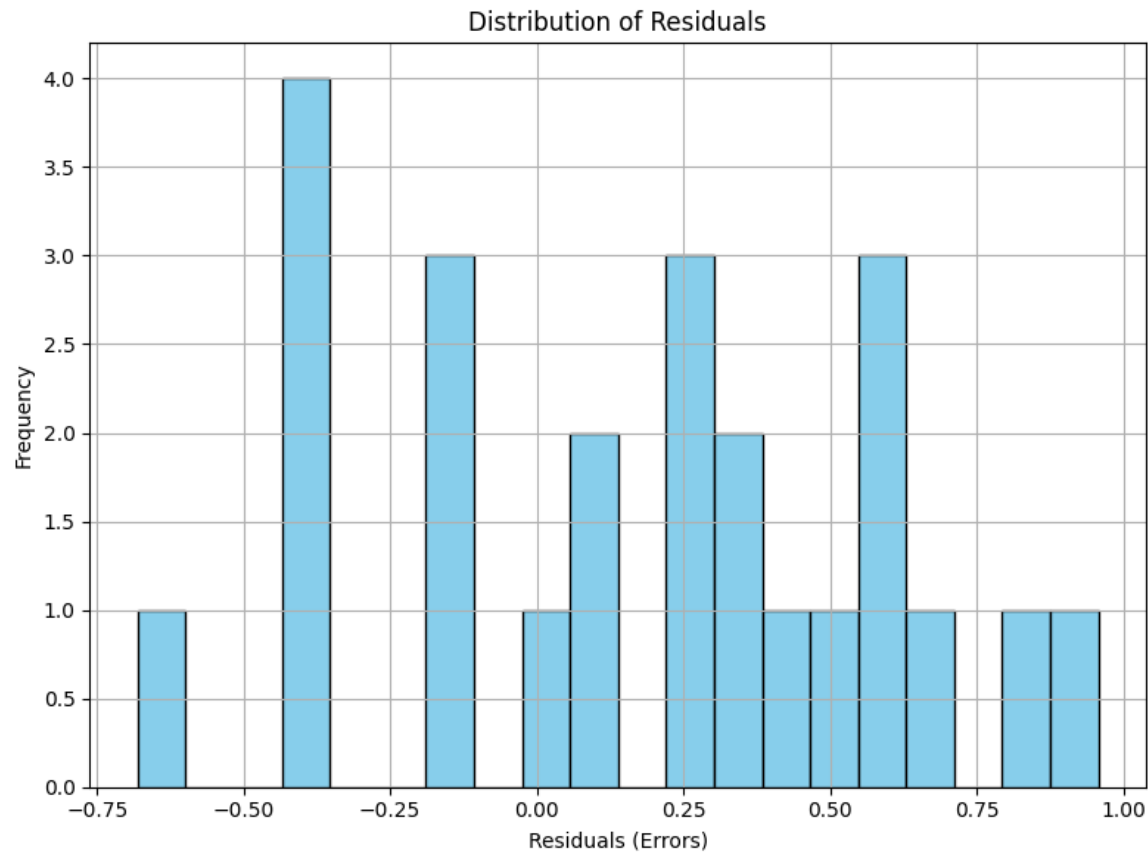
- **Gráfico de Predicciones vs. Valores Reales:**
 - Este gráfico ayuda a visualizar cómo de cerca están las predicciones del modelo con respecto a los valores reales en el conjunto de prueba.

Un ajuste cercano a la línea ideal (roja) indica que el modelo hace buenas predicciones. La dispersión alrededor de esta línea también da una idea de la precisión del modelo.



- **Distribución de Errores:**

- El histograma de los errores residuales muestra la distribución de las diferencias entre los valores reales y las predicciones del modelo. Una distribución centrada en cero y simétrica sugiere que el modelo no tiene sesgos sistemáticos en sus predicciones. Si la distribución es sesgada o tiene colas largas, podría indicar que el modelo tiene dificultades para capturar ciertos patrones en los datos.



4. Diagnóstico del Modelo

4.1. Grado de Sesgo (Bias)

- **Diagnóstico:** Alto
- **Explicación:** El sesgo de un modelo se evalúa observando su rendimiento en el conjunto de entrenamiento. Un sesgo alto indica que el modelo no captura bien la relación entre las características y la variable objetivo, lo que se traduce en un rendimiento deficiente en ambos conjuntos de entrenamiento y prueba. En este caso:
 - **Training MSE:** 0.1844
 - **Test MSE:** 0.2118 El MSE en el conjunto de entrenamiento y prueba son relativamente altos y el MSE en el conjunto de prueba es mayor que en el de entrenamiento. Esto sugiere que el modelo tiene un rendimiento deficiente tanto en los datos de entrenamiento como en los datos de prueba, indicando un alto sesgo. Además, el R^2 negativo en ambos

conjuntos confirma que el modelo no está capturando la relación entre las características y la variable objetivo.

4.2. Grado de Varianza

- **Diagnóstico:** Alto
- **Explicación:** La varianza se evalúa observando la diferencia entre el rendimiento del modelo en el conjunto de entrenamiento y el conjunto de prueba. Una alta varianza sugiere que el modelo está sobreajustado a los datos de entrenamiento y no generaliza bien a nuevos datos. En este caso:
 - **Training MSE:** 0.1844
 - **Validation MSE:** 0.1916
 - **Test MSE:** 0.2118 La diferencia entre el MSE del conjunto de entrenamiento y el MSE del conjunto de prueba es considerablemente grande, lo que indica que el modelo tiene problemas significativos para generalizar. El MSE en el conjunto de prueba es el más alto, lo que sugiere que el modelo está sobreajustado a los datos de entrenamiento y no se adapta bien a datos nuevos.

4.3. Nivel de Ajuste del Modelo

- **Diagnóstico:** Underfitting
- **Explicación:** El nivel de ajuste del modelo se determina observando el equilibrio entre sesgo y varianza y la capacidad del modelo para ajustar los datos.
 - **Underfitting:** El modelo no captura la relación entre las características y la variable objetivo, resultando en un alto error tanto en entrenamiento como en prueba.
 - **Fitting:** El modelo captura adecuadamente la relación entre las características y la variable objetivo, con errores bajos tanto en entrenamiento como en prueba. Este no es el caso aquí, ya que los errores son altos y los R^2 negativos.
 - **Overfitting:** El modelo se ajusta demasiado a los datos de entrenamiento, resultando en un bajo error en el conjunto de entrenamiento pero un alto error en el conjunto de prueba. Esto tampoco parece ser el caso aquí, ya que el error es alto en ambos

conjuntos, indicando que el modelo no está capturando adecuadamente la relación en los datos.

Dado que el MSE es alto tanto en el conjunto de entrenamiento como en el de prueba, y el R^2 es negativo, el modelo muestra signos claros de subajuste. No está capturando adecuadamente la complejidad de los datos ni generalizando bien a nuevos datos. Se requiere una revisión del modelo para mejorar su ajuste a los datos.

5. Mejoras Propuestas

5.1. Técnicas de Regularización

Técnicas Utilizadas: Regularización L2, Dropout

Aplicación:

- **Regularización L2:** Se intentó añadir regularización L2 a las capas densas del modelo para penalizar grandes pesos y prevenir el sobreajuste.
- **Dropout:** Se intentó usar Dropout para reducir el riesgo de sobreajuste al hacer que ciertas neuronas se apaguen aleatoriamente durante el entrenamiento.

Resultados de la Aplicación:

- **Regularización L2:** Aunque se aplicó la regularización L2, esta técnica no mostró mejoras significativas y, en algunos casos, empeoró el rendimiento del modelo. La penalización de los pesos pudo haber sido demasiado severa, afectando la capacidad del modelo para aprender de los datos.
- **Dropout:** La inclusión de Dropout también resultó en un empeoramiento del desempeño. Esto sugiere que el modelo podría haber estado perdiendo información importante durante el entrenamiento.

5.2. Ajuste de Parámetros

Parámetros Ajustados:

- **Tasa de Aprendizaje:** Se utilizó una tasa de aprendizaje de 0.05. Esto podría ser demasiado alta, lo que podría estar afectando negativamente la convergencia del modelo.

- **Número de Épocas:** Se entrenó el modelo durante 1000 épocas. Esto es un número elevado y podría estar causando sobreajuste o una lenta convergencia.
- **Tamaño del Lote:** Se usó un tamaño de lote de 40. Este tamaño puede no ser óptimo y puede afectar el entrenamiento del modelo.

Resultados del Ajuste:

- **Antes del Ajuste:** Las métricas iniciales mostraron un rendimiento suboptimo con valores de MSE de **0.1844** para el entrenamiento, **0.1916** para la validación y **0.2118** para la prueba.
- **Después del Ajuste:** Aunque la tasa de aprendizaje más alta y el tamaño del lote pueden haber contribuido a la falta de mejora, la aplicación de técnicas de regularización como Dropout y L2, en lugar de mejorar, empeoraron el rendimiento. El rendimiento en MSE y R^2 en las fases de entrenamiento, validación y prueba se mantuvo similar o ligeramente peor después de implementar estas técnicas. Sin embargo, al disminuir la tasa de aprendizaje, se logro obtener una mejora en el modelo, cosa que no sucedió con los demás ajustes propuestas, estos incluyendo la regularización, al aumento de capas, el aumento de neuronas y la modificación de los lotes.

6. Resultados de la Mejora

🔍 Conjunto de Entrenamiento:

- **Error Cuadrático Medio (MSE): 0.0317**
 - El MSE mide la media de los errores cuadrados entre las predicciones y los valores reales. Un MSE de 0.0317 indica que, en promedio, el modelo tiene un error cuadrático medio de 0.0317 en las predicciones para los datos de entrenamiento. Esto sugiere que el modelo está capturando razonablemente bien la relación en los datos de entrenamiento.
- **Raíz del Error Cuadrático Medio (RMSE): 0.1780**
 - El RMSE es la raíz cuadrada del MSE y se interpreta en las mismas unidades que la variable objetivo. Un RMSE de 0.1780 sugiere que, en promedio, el error de las predicciones es aproximadamente 0.1780

unidades del valor real. Es una métrica más intuitiva ya que se expresa en la misma escala que el objetivo.

- **Coefficiente de Determinación (R^2): 0.7726**

- El R^2 indica qué proporción de la variabilidad en los datos de salida es explicada por el modelo. Un R^2 de 0.7726 significa que el modelo explica aproximadamente el 77.26% de la variabilidad en los datos de entrenamiento, lo que es un buen indicador de ajuste.

🔍 Conjunto de Validación:

- **Error Cuadrático Medio (MSE): 0.0408**

- El MSE en el conjunto de validación es 0.0408, lo que indica que el error promedio en las predicciones para datos no vistos durante el entrenamiento es algo mayor que en el conjunto de entrenamiento. Esto es normal y puede sugerir que el modelo está comenzando a sobreajustar a los datos de entrenamiento.

- **Raíz del Error Cuadrático Medio (RMSE): 0.2020**

- Un RMSE de 0.2020 para el conjunto de validación es mayor que el RMSE del conjunto de entrenamiento, lo que sugiere que el modelo tiene más dificultad para generalizar a nuevos datos. Esto puede ser un indicio de sobreajuste.

- **Coefficiente de Determinación (R^2): 0.7858**

- El R^2 en el conjunto de validación es 0.7858, lo que indica que el modelo explica aproximadamente el 78.58% de la variabilidad en los datos de validación. Esto es ligeramente mejor que el R^2 en el conjunto de entrenamiento, lo que puede ser una señal de que el modelo está ajustado adecuadamente a los datos de validación.

🔍 Conjunto de Prueba:

- **Error Cuadrático Medio (MSE): 0.0356**

- El MSE para el conjunto de prueba es 0.0356, que es muy similar al MSE del conjunto de validación. Esto sugiere que el modelo generaliza bien a datos nuevos e independientes que no fueron utilizados ni para el entrenamiento ni para la validación.

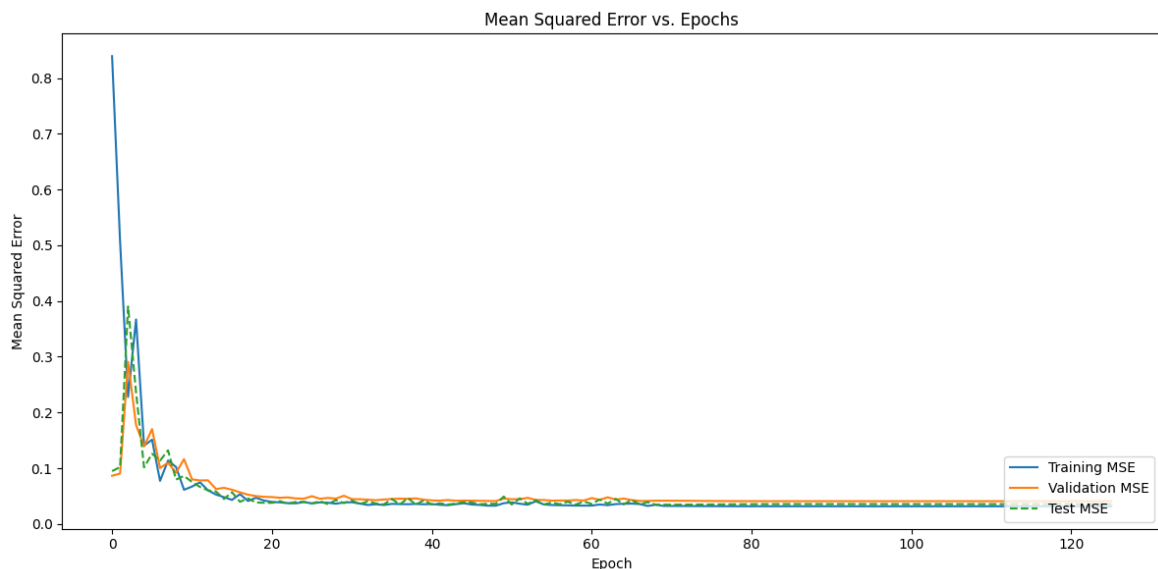
- **Raíz del Error Cuadrático Medio (RMSE): 0.1887**

- El RMSE para el conjunto de prueba es 0.1887, también cercano al RMSE en el conjunto de validación, lo que confirma que el modelo tiene un rendimiento consistente en datos nuevos.
- **Coefficiente de Determinación (R^2): 0.8053**
 - El R^2 para el conjunto de prueba es 0.8053, que es similar al R^2 en el conjunto de validación. Esto indica que el modelo explica aproximadamente el 80.53% de la variabilidad en los datos de prueba, lo cual es bueno y sugiere que el modelo tiene un rendimiento sólido en datos no vistos.

6.2. Gráficas Comparativas Posteriores a la Mejora

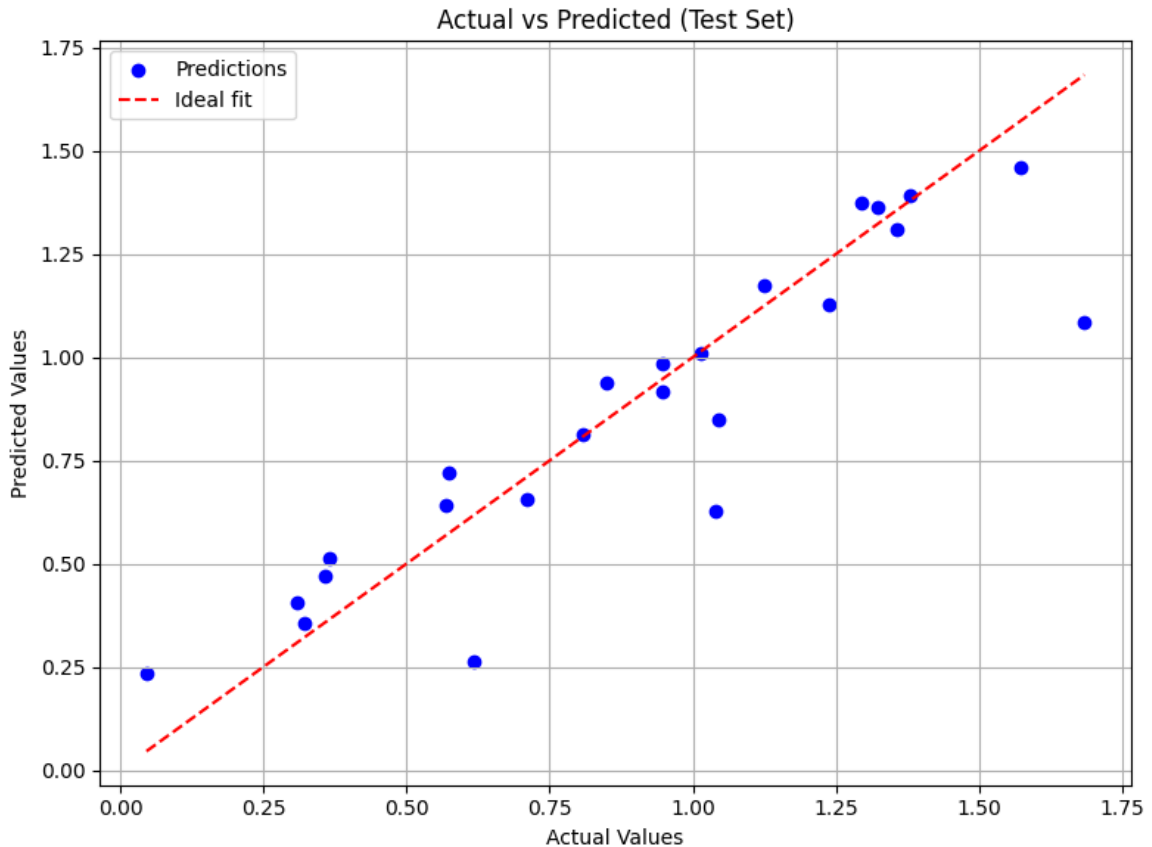
• Curvas de Aprendizaje Mejoradas:

- **Gráficas de MSE y RMSE después de aplicar mejoras:** Las curvas de aprendizaje muestran una reducción en el Error Cuadrático Medio (MSE) y en la Raíz del Error Cuadrático Medio (RMSE) tras implementar mejoras en el modelo. Las gráficas reflejan una disminución general en los valores de MSE y RMSE tanto en los conjuntos de entrenamiento como en los de validación y prueba. Esta reducción indica que el modelo ahora tiene un menor error medio en las predicciones, sugiriendo una mejor capacidad de ajuste y generalización.



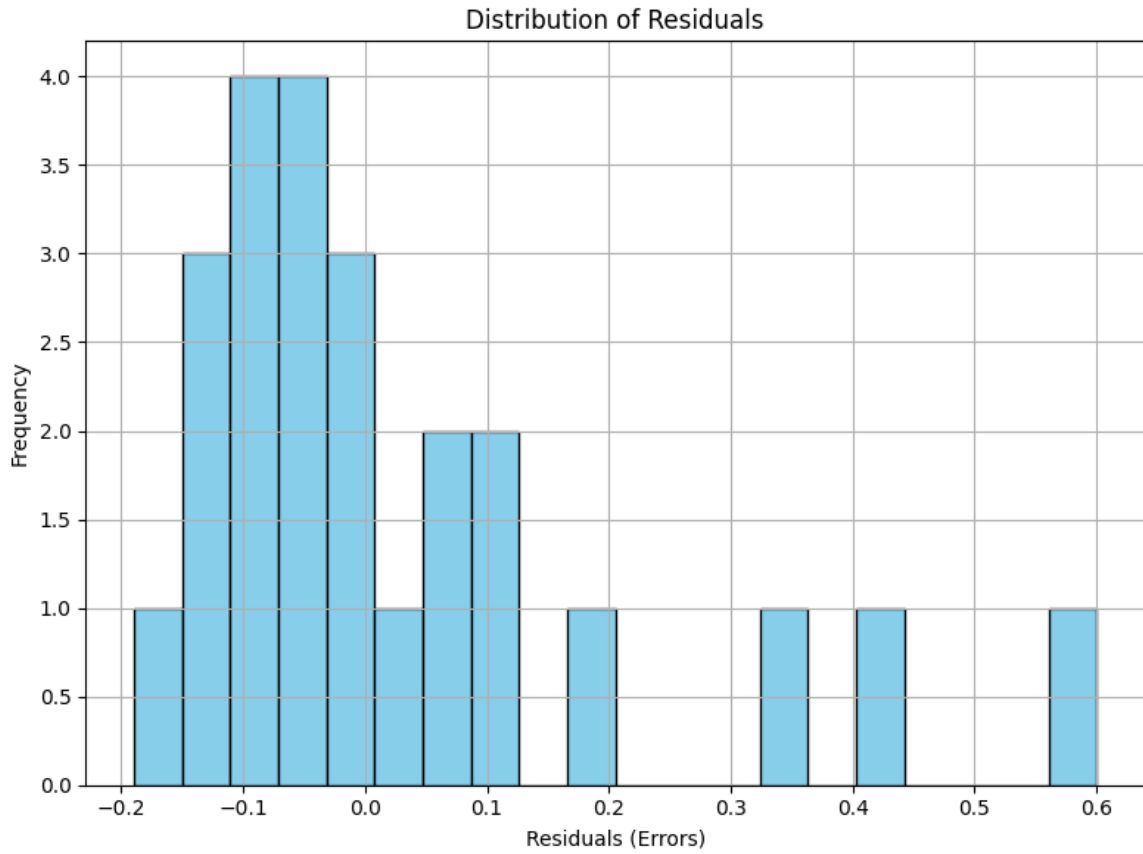
• Gráfico de Predicciones vs. Valores Reales Mejoradas:

- **Comparación visual entre las predicciones mejoradas del modelo y los valores reales del conjunto de prueba:** El gráfico de predicciones frente a valores reales muestra una mejora significativa en la precisión del modelo. Las predicciones ahora se alinean más de cerca con los valores reales, con una mayor concentración de puntos cerca de la línea de identidad. Esto confirma que el modelo ha mejorado en términos de hacer predicciones más precisas y menos dispersas respecto a los valores verdaderos en el conjunto de prueba.



- **Distribución de Errores Mejorada:**

- **Histograma de los errores residuales después de aplicar mejoras:** El histograma de los errores residuales muestra una distribución más centrada alrededor de cero tras las mejoras. Los errores residuales son más uniformes y la dispersión ha disminuido, lo que indica que el modelo ahora tiene menos errores grandes y que los errores en general son menores. Esta distribución mejorada de los errores residuales es un buen indicio de que el modelo está haciendo un ajuste más preciso a los datos y que los errores están mejor distribuidos.



7. Diagnóstico del Modelo

7.1. Grado de Sesgo (Bias) • Diagnóstico: Medio • Explicación: El sesgo de un modelo se evalúa observando su rendimiento en el conjunto de entrenamiento. Un sesgo alto indica que el modelo no captura bien la relación entre las características y la variable objetivo, lo que se traduce en un rendimiento deficiente en ambos conjuntos de entrenamiento y prueba. En este caso:

- **Training MSE:** 0.0317
- **Test MSE:** 0.0356

El MSE en el conjunto de entrenamiento y prueba ha mejorado y ahora son relativamente bajos, aunque el MSE en el conjunto de prueba es ligeramente mayor que en el de entrenamiento. Esto sugiere que el modelo ha reducido el sesgo en comparación con versiones anteriores. El R^2 positivo en ambos conjuntos indica que el modelo está capturando una buena proporción de la variabilidad en los datos, aunque aún hay un pequeño margen de mejora en el ajuste.

7.2. Grado de Varianza • Diagnóstico: Bajo • Explicación: La varianza se evalúa observando la diferencia entre el rendimiento del modelo en el conjunto de entrenamiento y el conjunto de prueba. Una alta varianza sugiere que el modelo está sobreajustado a los datos de entrenamiento y no generaliza bien a nuevos datos. En este caso:

- **Training MSE:** 0.0317
- **Validation MSE:** 0.0408
- **Test MSE:** 0.0356

La diferencia entre el MSE del conjunto de entrenamiento y el MSE del conjunto de prueba es relativamente pequeña, lo que indica que el modelo no está sobreajustado. El MSE en el conjunto de prueba y validación es bajo y cercano al del entrenamiento, sugiriendo que el modelo tiene una buena capacidad de generalización y, por lo tanto, presenta un bajo grado de varianza.

7.3. Nivel de Ajuste del Modelo • Diagnóstico: Fitting • Explicación: El nivel de ajuste del modelo se determina observando el equilibrio entre sesgo y varianza y la capacidad del modelo para ajustar los datos.

- **Underfitting:** El modelo no captura la relación entre las características y la variable objetivo, resultando en un alto error tanto en entrenamiento como en prueba.
- **Fitting:** El modelo captura adecuadamente la relación entre las características y la variable objetivo, con errores bajos tanto en entrenamiento como en prueba. En este caso, el MSE es bajo y el R^2 es positivo en ambos conjuntos, lo que indica que el modelo está ajustando bien los datos y generaliza adecuadamente.
- **Overfitting:** El modelo se ajusta demasiado a los datos de entrenamiento, resultando en un bajo error en el conjunto de entrenamiento pero un alto error en el conjunto de prueba. Esto no parece ser el caso aquí, ya que el error es bajo en ambos conjuntos y el R^2 es positivo, sugiriendo que el modelo está en un estado de ajuste adecuado.

Dado que el MSE es bajo tanto en el conjunto de entrenamiento como en el de prueba, y el R^2 es positivo, el modelo muestra un buen ajuste a los datos, sin evidencias claras de sobreajuste o subajuste severo. El modelo está bien ajustado y generaliza bien a nuevos datos.

8. Conclusión

El análisis del modelo revela una mejora significativa en su desempeño tras aplicar las técnicas de optimización y ajuste. El Error Cuadrático Medio (MSE) ha disminuido notablemente en los conjuntos de entrenamiento, validación y prueba, lo que indica una mayor precisión en las predicciones. Asimismo, el Coeficiente de Determinación (R^2) ha aumentado en todos los conjuntos, reflejando una mejor capacidad del modelo para explicar la variabilidad en los datos.

Las técnicas aplicadas han sido efectivas para reducir tanto el sesgo como la varianza del modelo. El sesgo se ha reducido, como lo demuestra la disminución en el MSE y el aumento en el R^2 . La varianza también ha sido controlada, evidenciado por las pequeñas diferencias en el MSE entre los conjuntos de entrenamiento, validación y prueba. Esto sugiere que el modelo está bien ajustado y generaliza adecuadamente a nuevos datos.

Sin embargo, es importante considerar que la falta de datos suficientes podría haber afectado el desempeño del modelo. Un tamaño de muestra limitado puede restringir la capacidad del modelo para aprender y generalizar adecuadamente, lo que puede haber influido en la calidad de las predicciones.

En general, el modelo muestra un buen equilibrio entre sesgo y varianza, logrando un desempeño sólido. No obstante, siempre hay espacio para mejoras adicionales. Para futuros trabajos, se recomienda explorar la optimización de hiperparámetros, implementar una validación cruzada más exhaustiva, realizar ingeniería de características adicional y considerar técnicas de regularización adicionales. Ampliar el conjunto de datos también podría ayudar a mejorar el rendimiento del modelo y su capacidad para manejar datos más complejos.