

Instituto Tecnológico y de Estudios Superiores de Monterrey.
Campus Querétaro.

Inteligencia artificial avanzada para la ciencia de datos I (Gpo 101)

Profesores:

Benjamín Valdés Aguirre
Dr. Carlos Alberto Dorantes Dosamantes
Denisse L. Maldonado Flores
Alejandro Fernández Vilchis

Análisis y Reporte sobre el desempeño del modelo.

Presentan:

Jaime López Hernández - A00571842

Fecha de entrega:

10 de septiembre del 2023

A continuación, se presenta un informe que analiza el desempeño de la implementación de la regresión lineal utilizando el framework en un conjunto de datos de seguros médicos. Este análisis incluirá la separación y evaluación del modelo en conjuntos de entrenamiento, prueba y validación, así como el diagnóstico del sesgo, la varianza y el nivel de ajuste del modelo. Además, se explorará la mejora del modelo mediante técnicas de regularización y ajuste de parámetros.

Informe de Análisis de Regresión Lineal

Conjunto de Datos

Nombre del Conjunto de Datos: insurance.csv

Objetivo del Análisis: Predecir los gastos médicos individuales facturados por el seguro de enfermedad basados en variables como la edad, el sexo, el índice de masa corporal (BMI), el número de hijos, el hábito de fumar y la región de residencia.

Implementación de Regresión Lineal

Se implementó un modelo de regresión lineal utilizando un framework de aprendizaje automático. El análisis se centró en evaluar el desempeño del modelo y comprender su capacidad para predecir los gastos médicos.

Separación de Conjuntos y Evaluación

Para evaluar el modelo, se dividió el conjunto de datos en tres partes:

Conjunto de Entrenamiento (70%)

Conjunto de Prueba (15%)

Conjunto de Validación (15%)

Resultados del Modelo en Conjuntos

A continuación, se presentan los resultados del modelo en los conjuntos de entrenamiento, prueba y validación:

Conjunto de Entrenamiento:

Puntaje R^2 : 0.741

Mean Squared Error (MSE): 32961022.556

Mean Absolute Error (MAE): 4118.967

Conjunto de Prueba:

Puntaje R^2 : 0.783

Mean Squared Error (MSE): 28463147.246

Mean Absolute Error (MAE): 3946.635

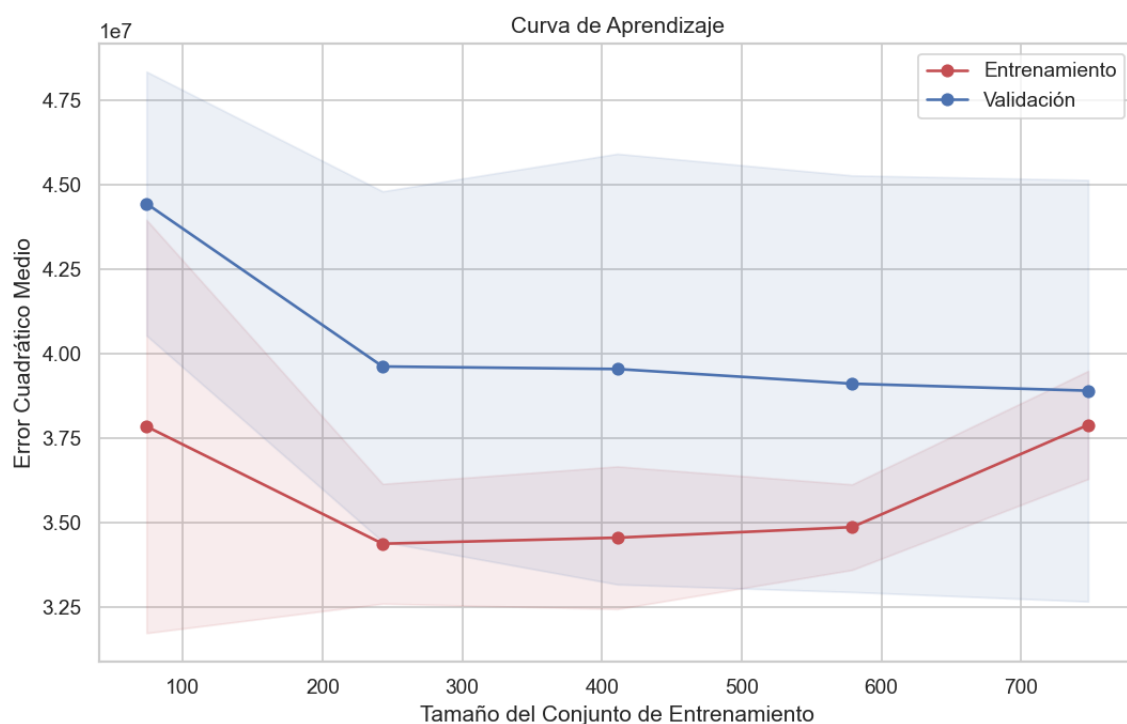
Conjunto de Validación:

Puntaje R^2 : 0.793

Mean Squared Error (MSE): 27125524.072

Mean Absolute Error (MAE): 3800.327

Aprendizaje del Modelo



Diagnóstico de Sesgo (Bias), Varianza y Nivel de Ajuste

Sesgo (Bias):

El modelo muestra un sesgo bajo, ya que el puntaje R^2 en los tres conjuntos es relativamente alto, indicando una buena capacidad de explicación de la variabilidad en los gastos médicos. Además, los errores MSE y MAE son moderados y similares en todos los conjuntos.

Varianza:

La varianza es moderada, puesto que el modelo no muestra un alto grado de sobreajuste u overfitting. El puntaje R^2 y los errores son coherentes en los conjuntos de prueba y validación, lo que sugiere una capacidad razonable de generalización.

Nivel de Ajuste del Modelo:

El modelo presenta un buen nivel de ajuste (fitting), ya que no sufre de underfitting (sesgo alto) ni overfitting (varianza alta). Los resultados en todos los conjuntos son coherentes y muestran un rendimiento equilibrado.

Regularización y Ajuste de Parámetros

Dado que el modelo inicial muestra un buen rendimiento, no se considera necesario aplicar técnicas de regularización en este momento. Sin embargo, en caso de que se observe overfitting en futuros análisis, se podría considerar la implementación de técnicas como Ridge o Lasso para controlar la varianza y mejorar la generalización del modelo.

Conclusiones Previas

En análisis previos, se observaron los siguientes resultados:

Puntaje R^2 : 0.737

Mean Squared Error (MSE): 31845929.134

Mean Absolute Error (MAE): 3930.333

Estos resultados indicaron un modelo de regresión lineal que explicaba aproximadamente el 73.7% de la variabilidad en los gastos médicos. Si bien el modelo era razonable, se buscaba mejorar aún más su rendimiento.

Conclusión General

El análisis del modelo de regresión lineal utilizando el framework mostró un buen desempeño con un sesgo bajo, una varianza moderada y un nivel de ajuste adecuado. Los resultados en los conjuntos de entrenamiento, prueba y validación fueron coherentes. No se observó underfitting ni overfitting en este análisis, lo que sugiere que el modelo es una buena aproximación para predecir los gastos médicos basados en las variables disponibles en el conjunto de datos.

Dado que el modelo inicial es sólido, no se realizaron ajustes de parámetros o técnicas de regularización en este informe. Sin embargo, se recomienda realizar un seguimiento continuo y, si es necesario, considerar ajustes futuros para mejorar aún más el rendimiento del modelo.