Modelos Predictivos Basados en Regresión Lineal Múltiple

El EDA confirmó que las variables seleccionadas en cada *dataset* presentan una relación lineal adecuada para la RLM, especialmente en el modelo del Dólar.

Hallazgos Clave

* Dólar: Se confirmó una correlación extremadamente fuerte entre día y precio\_Dolar. Las variables económicas como la Inflación y la tasa\_interes también mostraron una correlación, aunque menor, lo que implica que el factor temporal dominará la predicción.
* Glucosa: La relación entre IMC y Nivel\_Glucosa fue la más significativa, siendo positiva.
* Energía: La Temperatura resultó ser la variable más influyente linealmente. Las variables Hora y Día\_Semana exhibieron patrones cíclicos de consumo, aunque su correlación lineal directa fue más moderada.

3. Análisis de Resultados de Modelado y Coeficientes

Predicción del Precio del Dólar

Desempeño del Modelo

|  |  |  |
| --- | --- | --- |
| Métrica | Valor | Interpretación |
| MSE | 186.29 | Error promedio muy bajo, indicando que el modelo se ajusta de manera excelente a los datos de prueba. |
| R^2 | 0.9998 | El modelo explica el 99.98% de la variabilidad en el Precio del Dólar. Es un ajuste casi perfecto, impulsado por la fuerte tendencia de la variable Día. |

Interpretación de Coeficientes

|  |  |  |
| --- | --- | --- |
| Variable | Coeficiente (β^​) | Interpretación de Impacto |
| Día | +4.83 | Por cada día adicional, el Precio del Dólar aumenta en 4.83 unidades. |
| Inflación | +1225.80 | Por cada unidad de aumento en la Inflación, el precio aumenta en 1225.80. |
| Tasa\_interes | -22.50 | Por cada unidad de aumento en la Tasa de Interés, el precio disminuye en 22.50, mostrando un efecto ligeramente regulador. |

Variable con mayor impacto fue el día, debido a que su coeficiente es el más influyente en la predicción al aplicarse sobre valores grandes.

Ejercicio 2: Predicción de Nivel de Glucosa en Sangre

Desempeño del Modelo

|  |  |  |
| --- | --- | --- |
| Métrica | Valor | Interpretación |
| MSE | 130.41 | El error es moderado, mostrando que el modelo tiene cierta precisión. |
| R^2 | 0.8037 | El modelo explica el 80.37% de la variabilidad en el Nivel de Glucosa, lo que se considera un buen ajuste en un modelo de salud. |

Interpretación de Coeficientes

|  |  |  |
| --- | --- | --- |
| Variable | Coeficiente (β^​) | Interpretación de Impacto |
| Edad | +1.34 | Por cada año de edad, el Nivel de Glucosa aumenta en 1.34$mg/dL. |
| IMC | +2.09 | Por cada punto de IMC, el Nivel de Glucosa aumenta en 2.09 mg/dL. |
| Actividad\_Fisica | -0.98 | Por cada hora semanal de ejercicio, el Nivel de Glucosa disminuye en 0.98 mg/dL. |

Variable con mayor impacto fue el IMC, ya que su coeficiente absoluto de 2.09 es el más alto entre los predictores, lo que subraya la importancia de la masa corporal.

Ejercicio 3: Predicción del Consumo de Energía Eléctrica

Desempeño del Modelo

|  |  |  |
| --- | --- | --- |
| Métrica | Valor | Interpretación |
| RMSE | 27.99 | El error de predicción promedio es de $27.99$ kWh, indicando una precisión aceptable. |
| R^2 | 0.8170 | El modelo explica el 81.70% de la variabilidad en el Consumo de Energía, un ajuste muy bueno. |

Interpretación de Coeficientes

|  |  |  |
| --- | --- | --- |
| Variable | Coeficiente (β^​) | Interpretación de Impacto |
| Temperatura | +11.83 | Por cada aumento de 1 centigrado, el Consumo de Energía aumenta en 11.83 kWh. |
| Hora | -0.26 | Efecto muy bajo. Por cada hora que pasa, el consumo disminuye en 0.26 kWh esto podría reflejar una pequeña caída promedio por la noche. |
| Día\_Semana | +2.01 | Por cada día que avanza en la semana por ejemplo de lunes a lartes, el Consumo aumenta en $2.01$ kWh. |

Variable con mayor impacto fue la temperatura, con un coeficiente de 11.83 es por mucho el principal motor del consumo.

4. Persistencia y Despliegue de los Entregables

Para asegurar la funcionalidad de los modelos en una aplicación externa, se realizaron los siguientes pasos:

1. Exportación de Modelos: Los tres modelos de Regresión Lineal entrenados fueron serializados y guardados en archivos .joblib. Esto garantiza que los coeficientes se conserven y puedan cargarse sin necesidad de reentrenar.
2. Interfaz Web Funcional: Se desarrolló una aplicación web básica utilizando Streamlit (app.py). Esta interfaz permite al usuario seleccionar cualquiera de los tres ejercicios e ingresar los valores de las variables predictoras, utilizando los archivos .joblib cargados para generar y mostrar la predicción instantáneamente.

5. Conclusiones

Los modelos de Regresión Lineal Múltiple cumplieron con el objetivo de predecir las variables dependientes con un alto grado de fiabilidad, especialmente el modelo del Dólar y Energía ambos con R^2 > 0.80.

* Modelo del Dólar: El factor Día fue la variable más influyente, confirmando que la tendencia temporal domina la predicción de esta serie.
* Modelo de Glucosa: El IMC +2.09 y la Edad +1.34 son los factores biológicos más importantes para el aumento de la glucosa, mientras que la Actividad\_Fisica tiene el efecto corrector más débil.
* Modelo de Energía: La Temperatura es el factor de consumo más dominante con un impacto de $+11.83 kWh.