Conclusión

De acuerdo a la hipótesis del trabajo ¿se pueden predecir las enfermedades no transmisibles de acuerdo a los factores de riesgos? Teniendo en cuenta las condiciones demográficas y de salud de la población se evaluó la posibilidad de desarrollar la diabetes mellitus tomando dos variables la primera es la prevalencia de diabetes y la otra es el índice FINDRISC, que es la posibilidad de desarrollar diabetes mellitus en los próximos 10 años. A través de la comparación de tres modelos

**Resultados para la variable prevalencia de diabetes**

* **Regresión Lineal Múltiple (LinearRegression)**:
  + MSE: 3.65296385088139e-31
  + R²: 1.0
* **Árbol de Decisión (DecisionTreeRegressor)**:
  + MSE: 0.0
  + R²: 1.0
* **Máquinas de Soporte Vectorial (SVR)**:
  + MSE: 0.011486195530000362
  + R²: 0.9106538076273543
* **Support Vector Regressor (SVR) final ajuste de hiperparámetros:**
* MSE : 0.0039
* R²: 0.963

Los modelos de **Regresión Lineal** y **Árbol de Decisión** mostraron un ajuste perfecto en los datos de entrenamiento, con un R² de 1.0 y un MSE extremadamente bajo para el árbol de decisión. En consecuencia, estos resultados sugieren un posible sobreajuste, Esto conduce a un rendimiento deficiente con datos nuevos y desconocidos

El modelo SVR mostró un rendimiento ligeramente inferior pero aún muy alto, con un R² de aproximadamente 0.91 y un MSE muy bajo. Este modelo podría ser más robusto y consistente frente a nuevos datos. Sin embargo, al ajustar los hiperparámetros, el modelo alcanzó un MSE muy bajo y un R² de 0.963, explicando el 96.3% de la variabilidad. Esto indica un alto nivel de precisión y capacidad explicativa.

**Resultados para la variable FINDIRSC**

* **Regresión Lineal Múltiple (LinearRegression)**:
  + MSE: 0.6964197080816318
  + R²: 0.43905246886028326
* **Árbol de Decisión (DecisionTreeRegressor)**:
  + MSE: 2.090909090909091
  + R²: -0.6841715976331362
* **Máquinas de Soporte Vectorial (SVR)**:
  + MSE: 0.5995618596581396
  + R²: 0.5170688867102707

**Regresión Lineal Múltiple**: Mostró un rendimiento moderado con un R² de 0.439 y un MSE de 0.696. Esto indica que el modelo puede explicar aproximadamente el 44% de la variabilidad en la prevalencia de diabetes. Lo que indica que el modelo tiene un porcentaje muy bajo para poder predecir los nuevos datos.

**Árbol de Decisión**: Mostró un rendimiento pobre con un R² negativo de -0.684 y un MSE alto de 2.091. Un R² negativo sugiere que el modelo no es adecuado para los datos.

**SVR**: Mostró el mejor rendimiento con un R² de 0.517 y un MSE de 0.600. Esto indica que el modelo SVR puede explicar aproximadamente el 52% de la variabilidad en la prevalencia de diabetes, siendo el mejor entre los tres modelos evaluados, aunque sigue siendo muy bajo el porcentaje de variabilidad.

Por último decidí evaluar todas las provincias que tenga similar condiciones climáticas para evidenciar si el aumento de casos podía mejorar el modelo de aprendizaje y así tener mejores predicciones para la diabetes mellitus.

**Resultados para la Patagonia**

* **Regresión Lineal Múltiple (LinearRegression)**:
  + MSE: 1.348568245892349e-30
  + R²: 1.0
* **Árbol de Decisión (DecisionTreeRegressor)**:
  + MSE: 0.0
  + R²: 1.0
* **Máquinas de Soporte Vectorial (SVR)**:
  + MSE: 0.005119718267134216
  + R²: 0.9505124105671228

Los modelos de **Regresión Lineal** y **Árbol de Decisión** mostraron un ajuste perfecto en los datos de entrenamiento, con un R² de 1.0 y un MSE extremadamente bajo. Sin embargo, estos resultados indican un posible sobreajuste.

El modelo **SVR** mostró un rendimiento ligeramente inferior pero aún muy alto, con un R² de aproximadamente 0.95 y un MSE muy bajo. Este modelo podría ser más robusto frente a nuevos datos.

**Conclusión Final**

En resumen, el proyecto logró desarrollar modelos predictivos con un alto grado de precisión para la prevalencia de diabetes. Aunque los modelos de Regresión Lineal y Árbol de Decisión mostraron un ajuste perfecto en los datos de entrenamiento y esto puede evidenciar que hay un sobreajuste, el modelo SVR demostró ser una opción más robusta y confiable para predicciones futuras con un buen porcentaje de variabilidad en el primer caso con el 96% con el modelo SRV.

En consecuencia, este modelo podría ayudar a la población en anticipar la probabilidad de tener diabetes en el futuro y así modificar aquellos factores de riesgo para mejorar su calidad de vida.