

Proyecto de Aprendizaje Automático: Predicción de Tasas de VIH en Argentina

Descripción del Proyecto

Este proyecto tiene como objetivo aplicar técnicas de Aprendizaje Automático supervisado, para analizar y predecir la tasa de diagnósticos de VIH por cada 100.000 habitantes en las provincias argentinas. A partir de datos históricos provenientes del Ministerio de Salud de la Nación (2018–2023), se busca cuantificar el impacto de variables como el sexo, la edad mediana, tipo de relación y los casos reportados influyen de manera significativa en la variación de la tasa de diagnóstico entre jurisdicciones y años. El enfoque permite generar estimaciones continuas y precisas que pueden ser utilizadas en la planificación de políticas públicas de salud.

Objetivo general

Desarrollar y evaluar un modelo de Aprendizaje Automático que permita predecir la tasa de diagnóstico de VIH por 100.000 habitantes en las provincias argentinas, utilizando variables demográficas y epidemiológicas registradas entre 2018 y 2023.

Preguntas de investigación

¿Qué variables (sexo, edad mediana, tipo de relación, casos reportados, año, jurisdicción) influyen más en la tasa de diagnóstico de VIH?

¿Existen diferencias significativas en la tasa de diagnóstico entre provincias o a lo largo de los años?

¿Es posible predecir de forma precisa la tasa de VIH por 100.000 habitantes en una jurisdicción dada, usando únicamente datos históricos?

¿Qué patrones temporales o geográficos emergen del análisis de los coeficientes del modelo?

Contexto del problema y relevancia

siendo un desafío de salud pública en Argentina, con variaciones notables en la tasa de diagnóstico entre provincias y grupos poblacionales, entre 2018 y 2023, los registros del Ministerio de Salud muestran diferencias asociadas a factores como el sexo, el tipo de relación sexual reportada y la edad mediana de los casos.

El análisis mediante regresión lineal múltiple permite cuantificar el impacto de estas variables sobre la tasa de diagnóstico, generando evidencia útil para orientar políticas de prevención y focalización territorial de intervenciones sanitarias.

Se Elimina espacios en los nombres de columnas:

Renombra una columna: TIPO DE RELACION

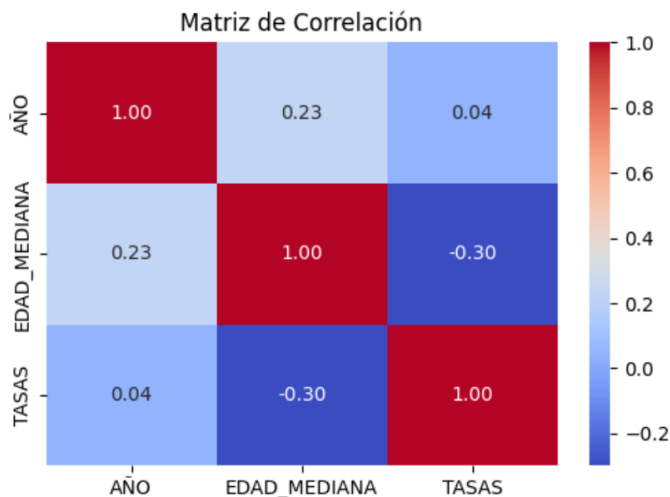
Corrige el formato decimal en la columna "TASAS":

Convierte columnas a numéricas: EDAD MEDIANA, AÑOS

Elimina filas con datos faltantes: EDAD MEDIANA, AÑOS, TASAS, CASO_VIH

Se eliminan filas que tengan valores nulos en las columnas clave y se reinicia el índice.

Hay una matriz de correlación



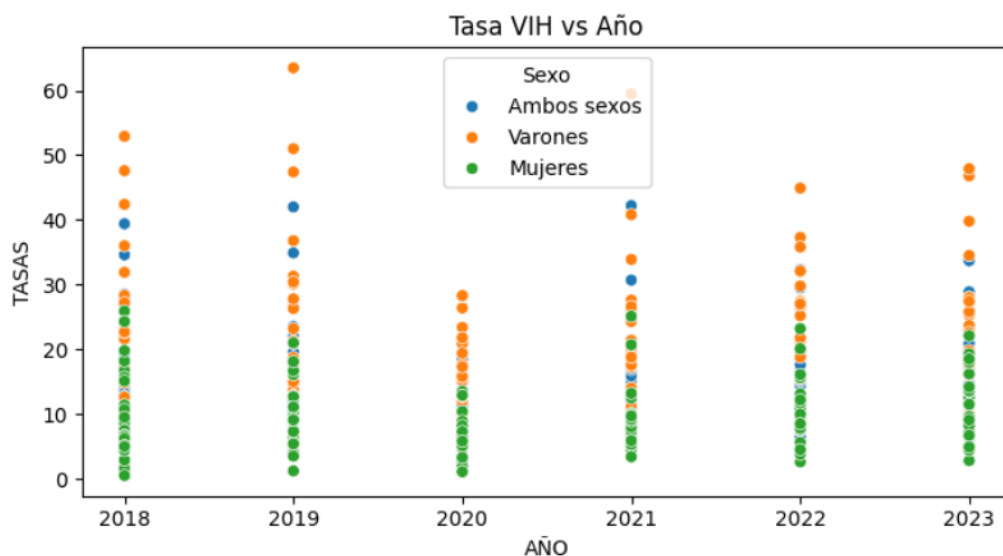
Este gráfico es una Matriz de Correlación, que muestra cómo se relacionan entre sí tres variables: AÑO, EDAD_MEDIANA y TASAS.

AÑO vs EDAD_MEDIANA: Correlación positiva débil (0.23) a medida que avanza el año, la edad mediana tiende a aumentar ligeramente.

AÑO vs TASAS: Correlación casi nula (0.04) No hay relación significativa entre el año y las tasas.

EDAD_MEDIANA vs TASAS: Correlación negativa moderada (-0.30) A mayor edad mediana, las tasas tienden a disminuir (o viceversa).

La única relación notable es que a medida que aumenta la edad mediana, las tasas tienden a bajar. El año no parece tener una influencia clara en las otras dos variables.

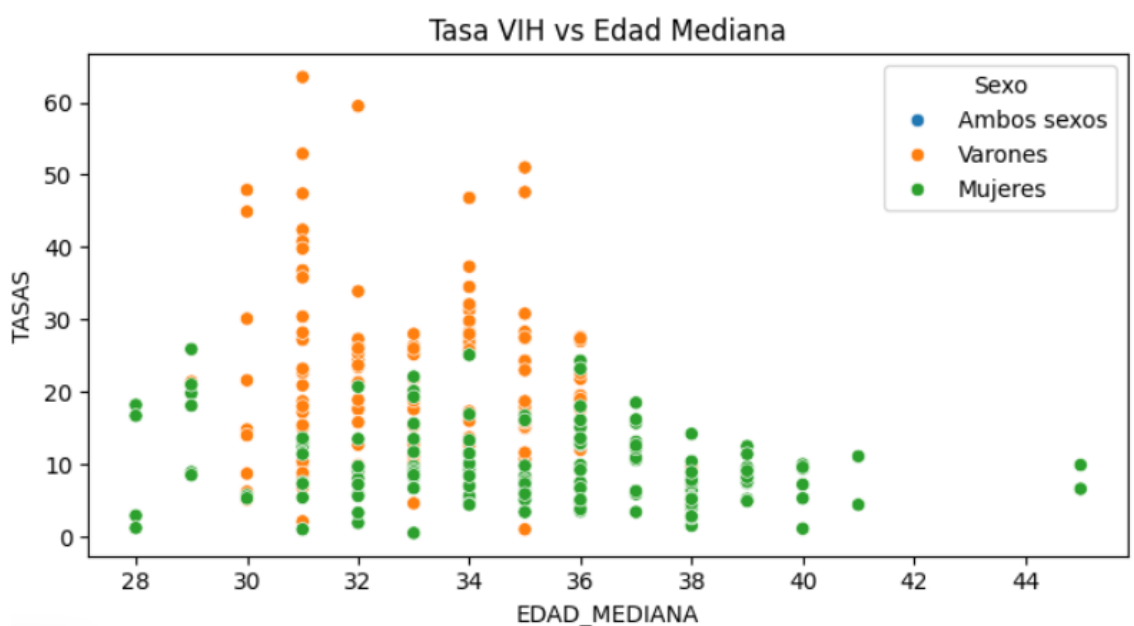


Este gráfico de dispersión muestra cómo han variado las tasas de VIH por año (2018–2023), diferenciadas por sexo:

● Ambos sexos ● Varones ● Mujeres

Observaciones clave:

Se observa que en todos los años las tasas en varones son consistentemente más altas que en mujeres, mientras que las tasas de ambos sexos se mantienen intermedias. Además, hay una cierta disminución en las tasas después de 2019, en 2020 hubo una caída notable en todas las categorías, posiblemente relacionada con la pandemia, en 2023 las tasas vuelven a subir especialmente en varones, esto sugiere una tendencia general descendente con variaciones por año y por sexo.



Las tasas más altas se concentran en personas con edad mediana entre 30 y 36 años, especialmente en varones.

A medida que la edad mediana aumenta (más allá de 38 años), las tasas tienden a disminuir.

Las mujeres tienen tasas generalmente más bajas que los varones en todos los rangos de edad.

No hay una relación lineal perfecta, pero sí una tendencia clara: las tasas son más altas en edades más jóvenes.

Las tasas de VIH están más concentradas en grupos con edad mediana entre 30 y 36 años, principalmente en varones. Conforme la edad aumenta, las tasas disminuyen, lo que sugiere que los casos se focalizan en poblaciones más jóvenes.