

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

Alumno: Marta Cruz

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.

Objetivos específicos

Analizar el dataset histórico de casos de VIH por jurisdicción, sexo, tipo de relación y edad mediana, explorar la correlación entre las variables predictoras y la tasa de diagnóstico, entrenar y evaluar un modelo de regresión lineal múltiple para estimar la tasa de diagnóstico a partir de las variables independientes seleccionadas. Interpretar los coeficientes del modelo para determinar qué variables tienen mayor impacto en la tasa de diagnóstico

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.