

Ficha Técnica

Proyecto: “Observatorio de datos para descubrimientos de patrones Sociales-EspacioTemporales en Salud, Movilidad y Calidad del Aire ”

Registro: 7051

1. Objetivo general del proyecto	Definir las bases de interoperabilidad para homologar la información de las bases de datos de salud , y obtener datos precisos y confiables, que permitan tener series históricas y generar la trazabilidad de los pacientes, en especial de aquellos que padecen enfermedades crónicas no transmisibles, así como la construcción de indicadores de salud y su relación con la exposición a contaminantes atmosféricos.
2. Propósito del registro	<p>generación automatizada de modelos predictivos del número y severidad de las hospitalizaciones por enfermedades y categorías de enfermedades</p> <p>Mediante el análisis de datos de egresos hospitalarios, se ha desarrollado un sistema automatizado para la generación de modelos predictivos del número y la severidad de hospitalizaciones por diversas enfermedades y categorías, seleccionando los códigos CIE o grupos de códigos a modelar. Para tal fin, se han empleado técnicas estadísticas, tales como la regresión logística y la regresión binomial negativa.</p> <p>No obstante, es importante destacar que los modelos generados no tienen como finalidad la predicción, ni están destinados a ser utilizados con ese propósito, sino que su objetivo es interpretar las variables que resultan relevantes a la hora de examinar la severidad y el número de hospitalizaciones. En este sentido, los modelos consideran el efecto de múltiples variables, tales como factores socioeconómicos, exposición a contaminantes ambientales, lugar de residencia, peso, edad, sexo, densidad de población, derechohabiencia, fecha de ingreso (con el fin de identificar patrones temporales) y se permite especificar el rango de años a utilizar. Por lo tanto, el interés radica en comprender cómo estas variables influyen en el modelo y la magnitud de su efecto.</p> <p>Además, para mantener la interpretabilidad del modelo, se incluyen únicamente aquellas variables que contribuyen de manera significativa al mismo, a fin de que éste sea lo más parsimonioso posible y se reduzca el problema de multicolinealidad. Este procedimiento se lleva a cabo a través de un algoritmo ad-hoc que incluye penalizaciones en el proceso de selección de variables, y que toma en cuenta tanto la multicolinealidad como la paradoja de Simpson. En caso de que exista un efecto</p>

	<p>diferente al esperado (por ejemplo, de un análisis de correlación), se mantiene únicamente si su contribución al modelo resulta significativa.</p> <p>Cabe destacar que, **por motivos de privacidad, la base de datos original no se comparte**. No obstante, se ha generado una versión con **datos sintéticos** que permite probar los algoritmos y adaptarlos para su uso en otros proyectos.</p>
3. Autor	Carlos Minutti Martinez
4. Recolección / obtención de los datos	<p>[Censo de Población y Vivienda 2020 (INEGI)](https://www.inegi.org.mx/programas/ccpv/2020/)</p> <p>[Datos de calidad de aire de la Ciudad de México](http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBhnmM=%27)</p>
5. Periodo de recolección / obtención de los datos	2020
6. Variables incluidas	20
7. Estrategia de aseguramiento para la protección de datos sensibles / personales	Eliminación y modificación de elementos de información identificable que permite cumplir con las regulaciones en materia de protección de datos personales. En lugar de utilizar la edad, se generó la variable categoría de eda.
8. Fecha de última actualización	1 junio de 2023
9. Mencionar el tipo de controles para la validación / verificación de la captura / registro de los datos.	No se realizó imputación de datos.
10. Otras plataformas donde se encuentren disponibles estos recursos	
11. Otras fuentes de Financiamiento	No aplica

12. Seguimiento de la Cohorte en estudio	No aplica
13. Publicaciones	Ninguna hasta el momento