

Predictor de Total de Ingresos Respiratorios en zonas de la Ciudad de México mediante Redes Neuronales LSTM

Resumen-Planeación

Violeta Benítez Rodrigo Trejo

March 27, 2024

Presentation Overview

① Resumen

② Plan de Trabajo

Objetivo

Diseñar un modelo de Aprendizaje Profundo para predecir mediante la detección de patrones en una serie de tiempo, los futuros ingresos hospitalarios de enfermedades respiratorias en algunas zonas de la Ciudad de México, con el fin de observar la relación que existe entre los casos de hospitalización por enfermedades respiratorias y variables que evalúan la calidad del aire.

Datos

Se cuenta con una base de datos que registra los lugares de la Ciudad de México donde ocurrieron la mayor cantidad de casos de enfermedades respiratorias que requirieron hospitalización en el periodo de enero de 2009 a diciembre de 2022, en donde encontramos datos asociados a los antecedentes de los pacientes de enfermedades respiratorias, así como datos del proceso de su enfermedad.

Además, por cada registro (según su localidad), se tienen datos atmosféricos asociados con los que se pretende realizar la predicción.

Observaciones respecto a los datos

A cada fila existente en la base de datos, se le asocia una `clave_res` que indica ***el lugar específico*** en donde se registra la cantidad de personas que sufrieron de alguna enfermedad respiratoria.

Esto sugiere que quizá la cantidad de muestras no alcance para realizar una predicción ***global*** (*a nivel Ciudad de México*) con precisión adecuada.

Observaciones respecto a los datos

Cada alcaldía de la Ciudad de México se encuentra representada con al menos una localidad lo largo de la serie de tiempo.

El set de datos se encuentra des-balanceado espacialmente (*en Tlalpan hay más localidades registradas que en Iztacalco, por ejemplo*).

Esto sugiere que para algunas zonas la predicción puede no ser representativo a lo que esté sucediendo realmente

Como se cuentan con datos que abarcan solamente hasta el año 2022, y el objetivo es realizar predicciones para años como 2024 y 2025, se necesitarán realizar estimaciones de los datos faltantes que corresponden al periodo de enero a diciembre de 2023. Lo que puede afectar la precisión de la predicción del modelo.

Metas

Contamos con un aproximado de 5 semanas en las que se pretenden lograr los siguientes objetivos específicos:

Análisis exploratorio, limpieza y preparación de los datos.

Construcción, entrenamiento y prueba de los modelo LSTM para predecir los datos atmosféricos faltantes que corresponden al periodo de enero a diciembre de 2023 y hasta el momento en que se desee obtener la predicción de los ingresos respiratorios.

*Por cada variable atmosférica que se desee utilizar, se debe entrenar un modelo diferente (**$n_variables$**).*

*Por cada zona de estudio se debe entrenar un modelo diferente (**m_zonas**).*

Entonces se tendrán un total de **$n_variables \times m_zonas$** modelos específicos.

Metas

Construcción del conjunto de *train* y *test* para cada predictor de ingresos respiratorios.

Construcción, entrenamiento y prueba de los modelo LSTM para predecir los ingresos respiratorios por cada zona de interés de estudio.

Cronograma

Semana	Temática
25-29/03	Resumen, Plan de trabajo y preparación del set de datos
01-05/04	Introducción a las Redes Neuronales Artificiales, Redes LSTM, y creación de los modelos de las variables ambientales
08-12/04	Construcción, entrenamiento y prueba del modelo de predicción de ingresos hospitalarios

Table: Planificación del curso por semana

Cronograma

Semana	Temática
15-19/04	Modificaciones y variantes del modelo principal. Elección del mejor modelo
22-26/04	Resolución y ajustes finales

Table: Planificación del curso por semana

Gracias