

Resumen

Este proyecto buscó determinar los factores que más influyan en la propagación del dengue. Tomando un conjunto de datos con distintos factores con su valor y el total de casos en por semana en dos ciudades, se hicieron predicciones tomando juntos los factores y cada uno por separado, esto para cada ciudad, con el fin de determinar cuál predecía mejor el número de casos. También se generaron grupos por cada factor para determinar cuáles eran más vulnerables de acuerdo con los contagios que han presentado.

De los cuatro factores trabajados, los cuales son humedad, precipitación, NDVI y temperatura, se pudo determinar cuáles de ellos fueron más precisos a la hora de predecir los casos.

Introducción

La motivación de este trabajo nace debido a que el dengue es un problema que ha azotado a varios países por mucho tiempo, y recientemente esta problemática ha perdido relevancia. Partiendo del hecho que el dengue se transmite por un mosquito tropical, además de los cuidados caseros como no dejar recipientes con agua estancada, a nivel general, se deben tomar medidas para no permitir nidos naturales que se pueden generar por altos niveles de humedad u otros factores. Si se identifican adecuadamente los factores que facilitan su propagación, se conseguirán disminuir los contagios, lo cual es muy positivo teniendo en cuenta que no existe un tratamiento específico para esta enfermedad.

Si se orientarán recursos a prevenir esta enfermedad identificando las zonas de alto riesgo, y no a tratarla cuando ya está propagada, conllevaría a un ahorro de recursos al estado y una reducción en la carga laboral de los médicos, permitiéndolos enfocarse en otras problemáticas que actualmente son más difíciles de prever.

Proceso y método

Este problema es un estudio de casos de dengue, donde se busca determinar qué variables climatológicas son más precisas para la predicción de casos en una ciudad. La información fue extraída de los habitantes infectados en dos ciudades las cuales son San Juan, en Puerto Rico, e Iquitos, en Perú. Esta información proviene de una recopilación de datos de varias entidades gubernamentales de los Estados Unidos, como por ejemplo el CDC la información sobre el número de casos de dengue y la NOAA la información climatológica de esas zonas.

Para analizar estos datos se empleó la versión 1.0.5 de pandas En un entorno de notebook con Python 3. Con pandas, se observó del conjunto de datos y se identificaron dos ciudades, las cuales se usaron de referencia para dividir los datos de identificar los factores que estaban asociadas a ellas.



Figura 1. Logo de la entidad gubernamental CDC

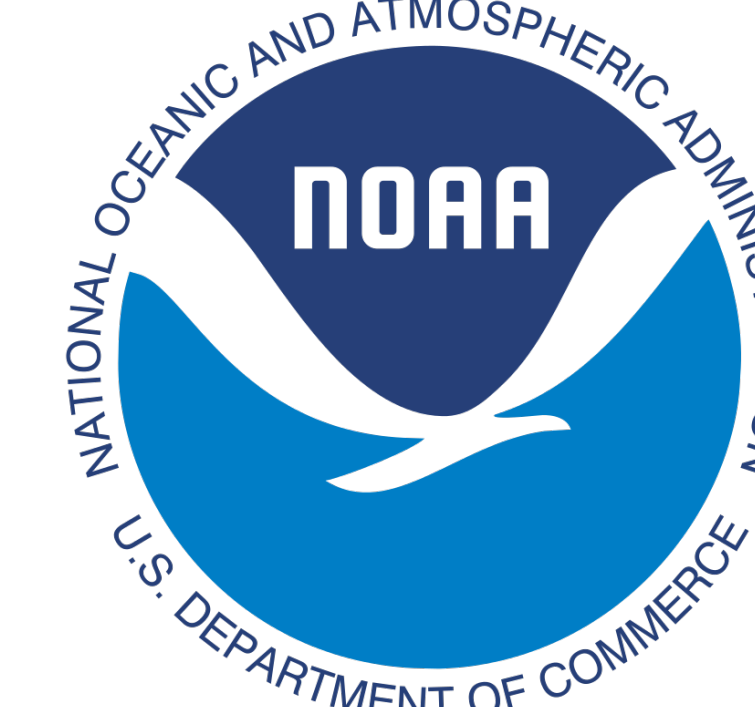


Figura 2. Logo de la entidad gubernamental NOAA

Resultados

Se estudió el conjunto de datos donde se identificaron cuatro factores, los cuales son: la temperatura, el NDVI, la humedad, y la precipitación. Además la medida en que estos influyen sobre la cantidad de casos de dengue.

Se predijo el total de casos para cada ciudad, usando todos los factores juntos y posteriormente individualmente, los estimadores se entrenaban con un 80% de entrenamiento y un 20% de prueba, y usando la función de predicción, con el fin de ver el número de casos esperado y el número de casos obtenido.

Haciendo uso de aprendizaje no supervisado, se hicieron grupos para cada factor climatológico con el fin de identificar grupos donde hubo mayor número de casos de acuerdo con el valor de el factor.

Gráfico 1. Comparativa de casos obtenidos y esperados

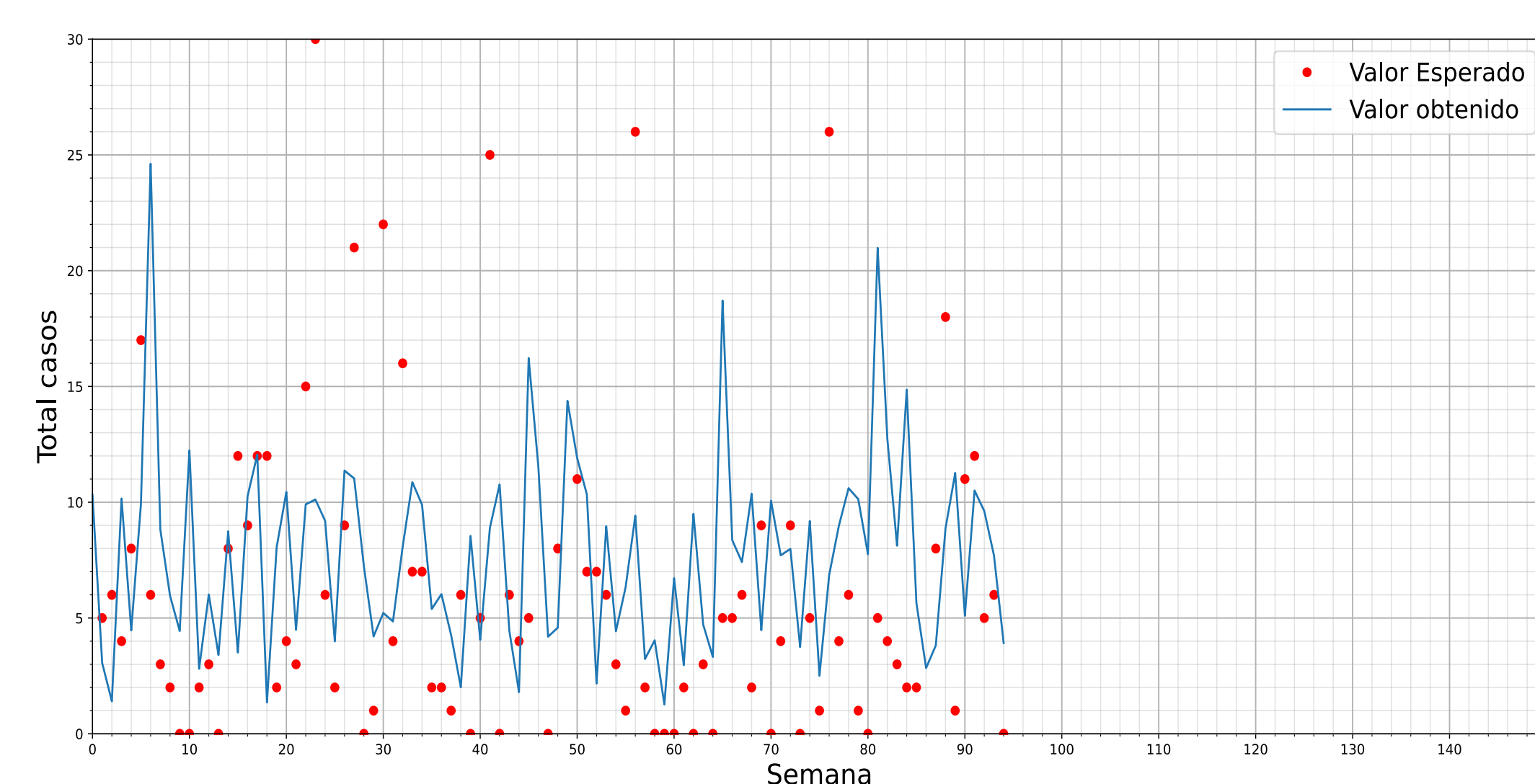
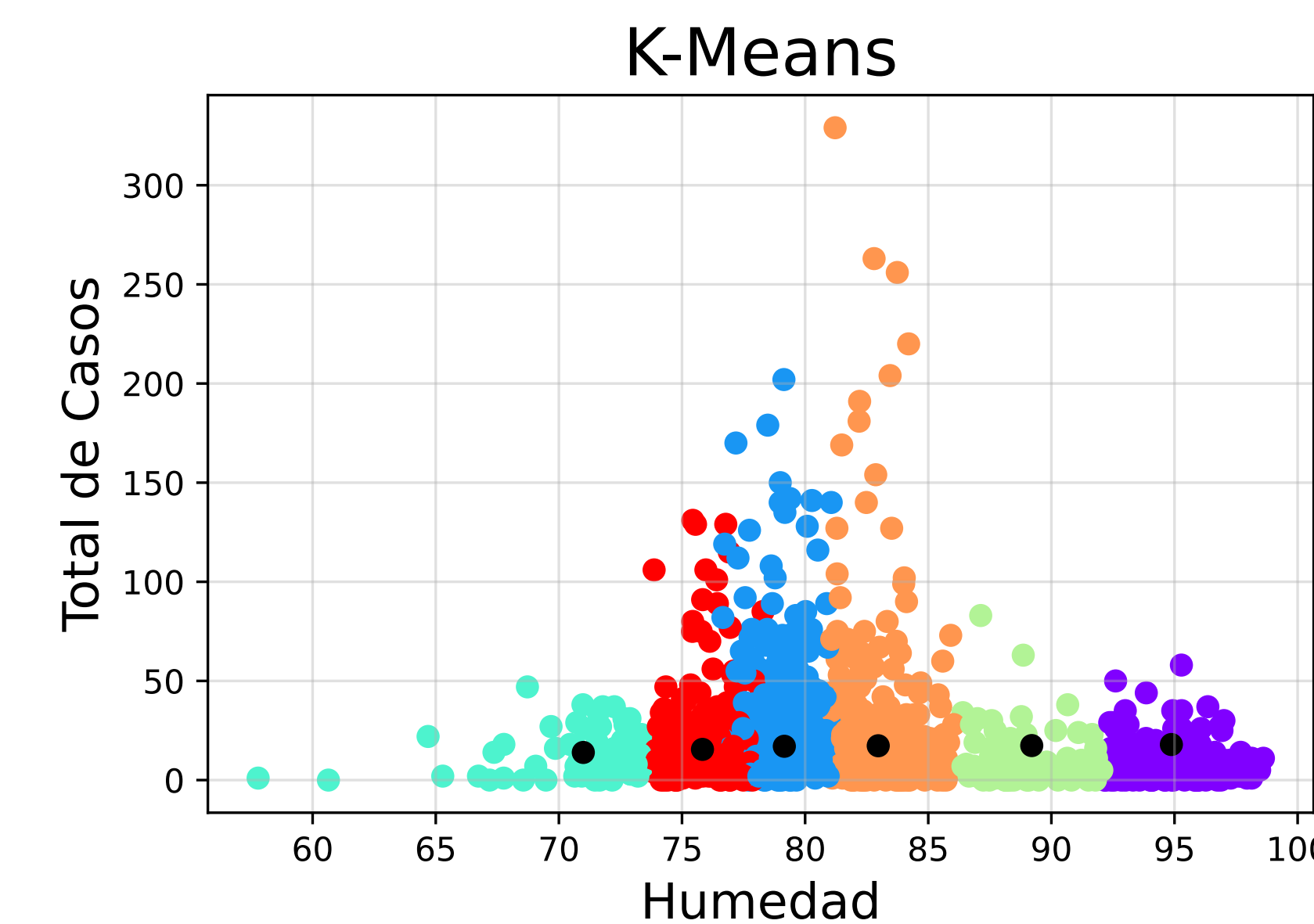


Gráfico 2. Generación de grupos con aprendizaje no supervisado



Conclusiones

Después de las distintas pruebas que se realizaron en el conjunto de datos respecto a cada uno de los factores en las dos distintas ciudades se puede concluir que:

- Los dos estimadores que muestran mejores resultados más cercanos a los casos de dengue son los estimadores de Random Forrester Regression y Decision Tree Regression.

- Los algoritmos implementados para predecir el número de casos tuvieron un mejor funcionamiento en la ciudad de Iquitos, a diferencia de San Juan.

- La implementación de Kmeans permite en los factores de propagación identificar grupos en un rango de valores para identificar donde se puede mayor número de casos.

- Por último el factor que más ayudó para las distintas pruebas de machine learning supervisado y no supervisado fue la humedad por tener la menor distancia media cuadrática mínima (RMSE), y por identificar de mejor manera grupos con el Kmeans.

Trabajo Futuro

Para obtener mejores resultados se podría estudiar más a fondo el uso de deep learning, ya que las redes neuronales tienen mayor capacidad que las técnicas de machine learning.

También, se podría probar combinar los factores en duplas o ternas, y observar si se obtuvo un error más bajo.

Información de contacto

Santiago Angulo Flórez, Email: santiago02170130@correo.uis.edu.co
Jean Carlos Portilla Mora, Email: jean2171452@correo.uis.edu.co

Docente: Fabio Martinez Carrillo, famarcas@saber.uis.edu.co

Referencias Bibliográficas (en formato APA)

1. Organización Mundial de la Salud (OMS). Artículo informativo, Dengue y Dengue grave, recuperado en 24 de junio de 2020, de <https://www.who.int/es/news-room/fact-sheets/detail/dengue-and-severe-dengue>
2. DengAI - How to approach Data Science competitions (2020), artículo, recuperado en 31 de julio de 2020, de <https://erdem.pl/2020/07/deng-ai-how-to-approach-data-science-competitions-edu>
3. Driven Data, DengAI : Predicting Disease Spread, de <https://www.drivendata.org/competitions/44/dengai-predicting-disease-spread/>
4. Centers for Disease Control and Prevention, Areas with risk of dengue, de <https://www.cdc.gov/dengue/areaswithrisk/index.html>
5. Earth observing system, NDVI, de <https://eos.com/ndvi/es/>