

Resumen

Este proyecto buscó determinar los factores que más influyen en la propagación del dengue. Tomando un conjunto de datos con distintos factores con su valor y el total de casos por con distintos factores climatológicos. Además, según el valor de cada factor, se generaron grupos para verificar donde se generaban más casos.

De los cuatro factores trabajados, los cuales son humedad, precipitación, NDVI y temperatura, y se busca determinar cuál de ellos es más preciso.

Introducción

El dengue es un problema que ha azotado a varios países por mucho tiempo, y recientemente esta problemática ha perdido relevancia. El dengue se transmite por un mosquito tropical, por ciertos ambientes climatológicos como la humedad o la temperatura se pueden generar nidos. Si se identifican adecuadamente estos factores, se conseguirá predecir los contagios.

Reconociendo las zonas de alto riesgo, 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.



Figura 1. Mosquito *Aedes aegypti*

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.

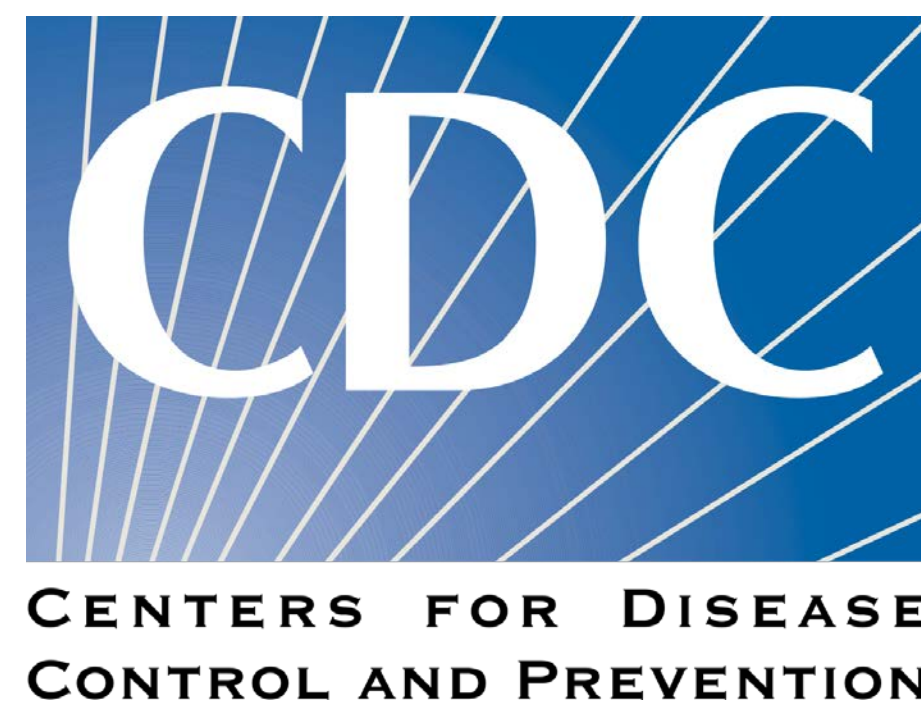


Figura 2. Logo de la entidad gubernamental CDC



Figura 3. Logo de la entidad gubernamental NOAA

Resultados

Se realizó una regresión lineal con el fin de determinar el comportamiento esperado del total de casos de dengue para ciudad de acuerdo con cada factor.

Se predijo el total de cosas para cada ciudad, usando Random Forrester Regression y Decision Tree Regression con todos los factores juntos y posteriormente individualmente, los estimadores se entrenaban con un 80% de entrenamiento y un 20% de prueba.

Haciendo uso de aprendizaje no supervisado con KMeans, 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. Regresión lineal del total de casos y la humedad.

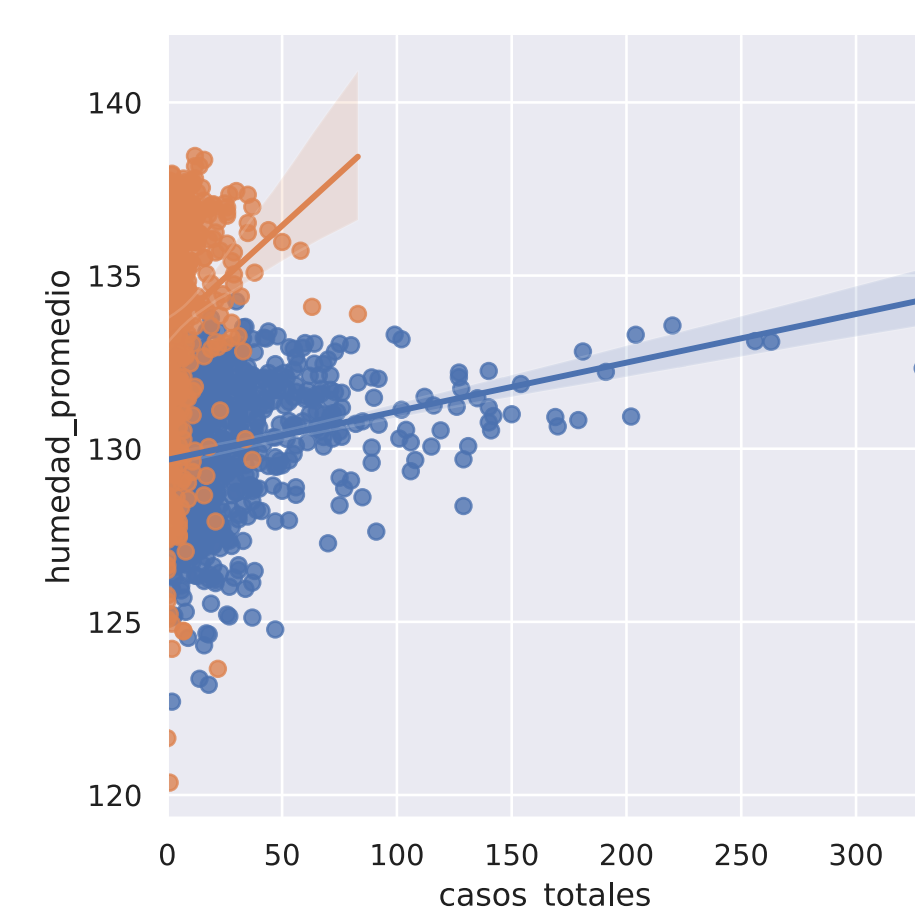


Gráfico 2. Comparativa de casos obtenidos y esperados

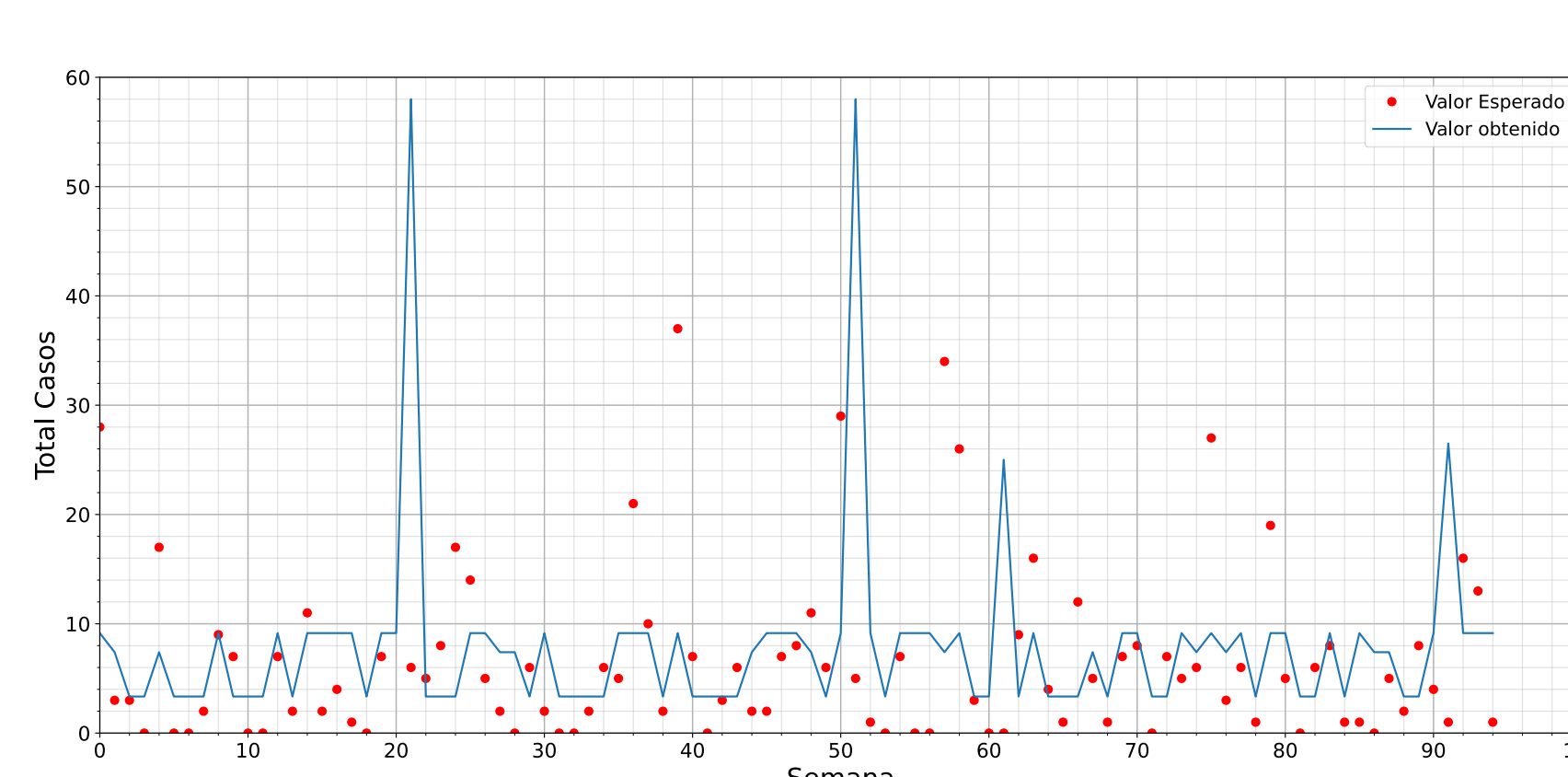
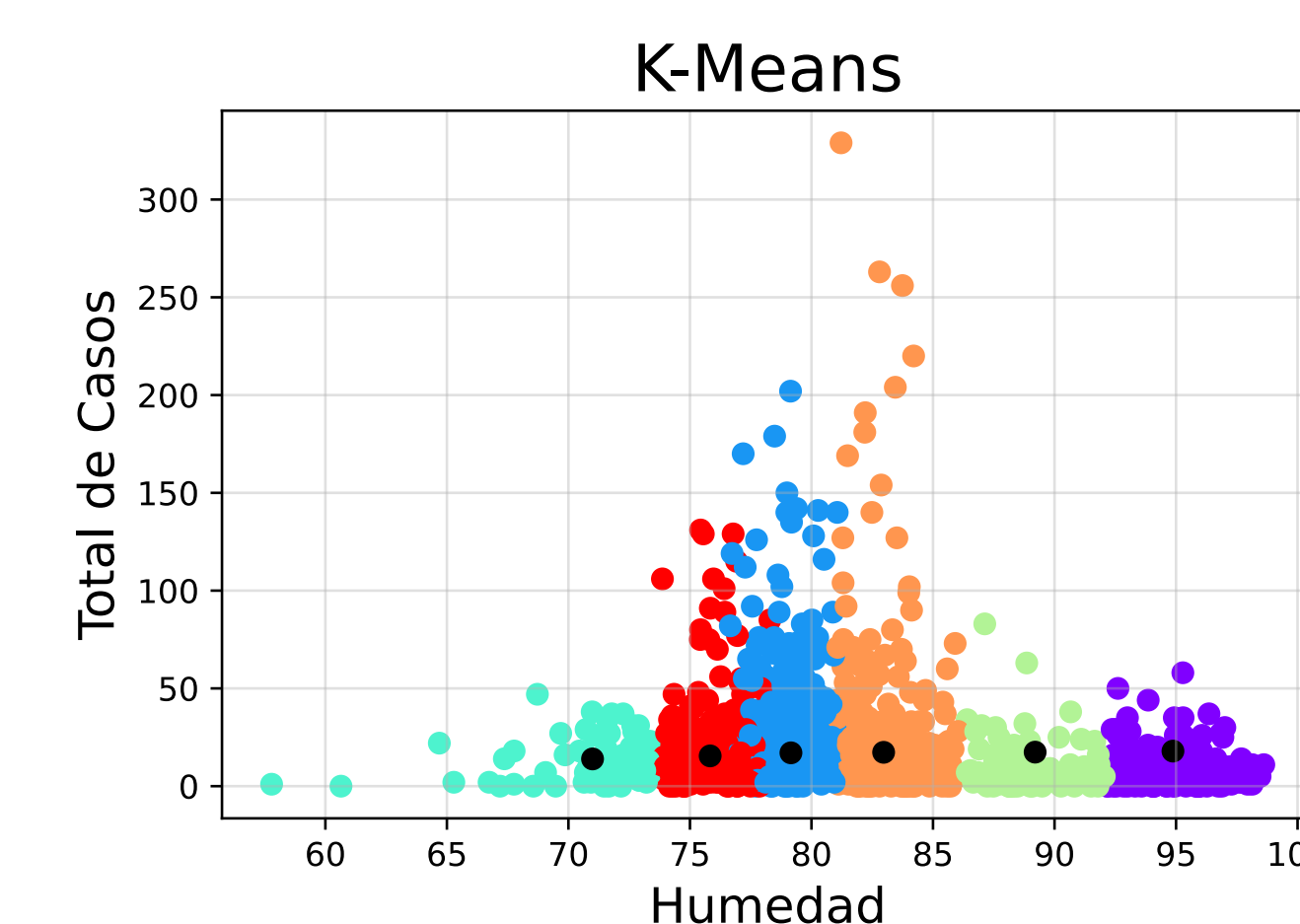


Gráfico 3. Generación de grupos con aprendizaje no supervisado



Conclusiones

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

• Los algoritmos implementados para predecir el numero 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 números de casos.

• Por ultimo el factor que mas 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.

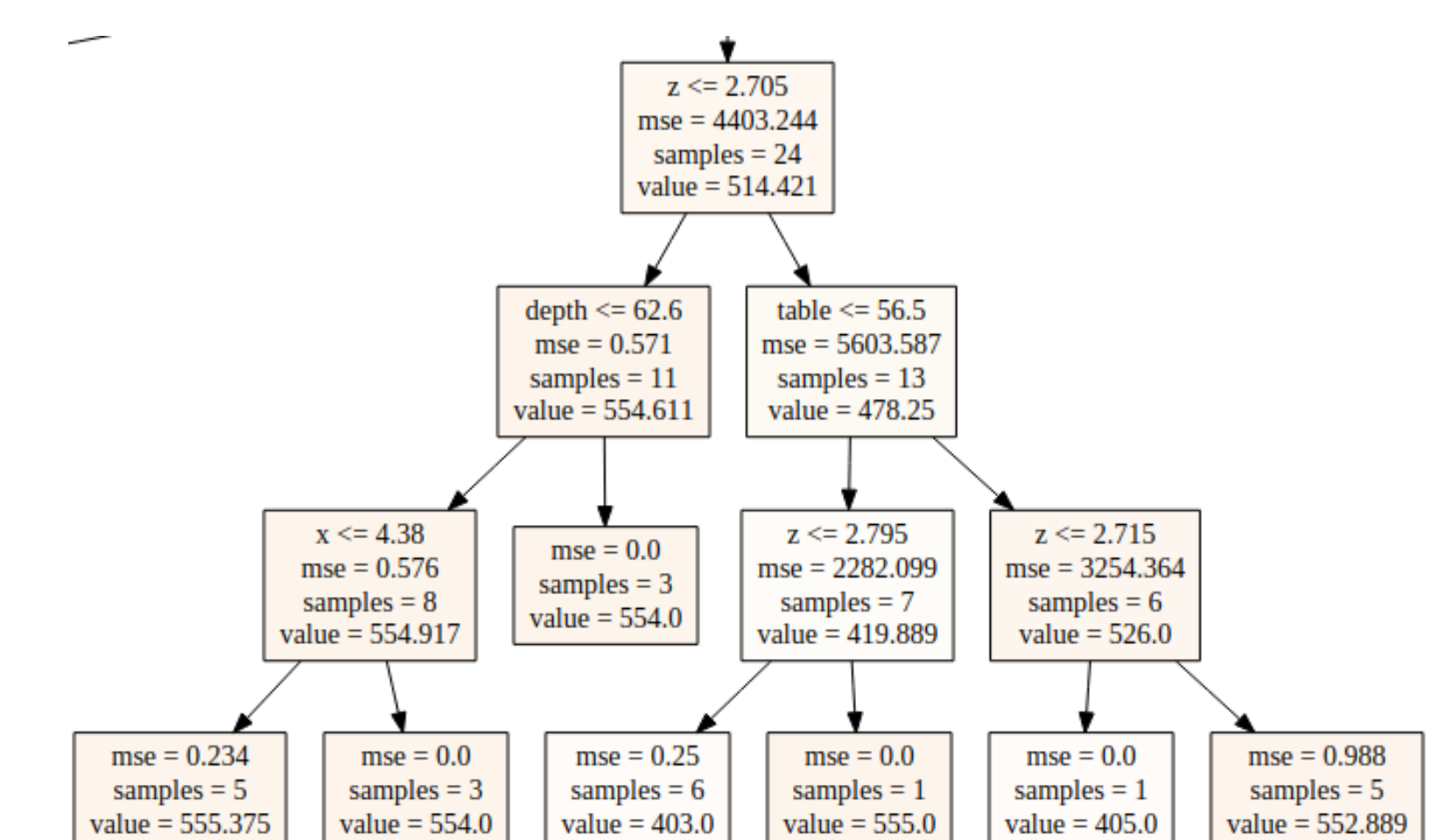


Figura 4. Ejemplo de Decision Tree Regressor

Trabajo Futuro

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

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/>