

Análisis de factores que generan la propagación del Dengue



Universidad
Industrial de
Santander

Santiago Angulo Flórez
Jean Carlos Portilla

Implementos:



NumPy



pandas



scikit
learn



#LaUISqueQueremos





Propósito y pregunta a responder



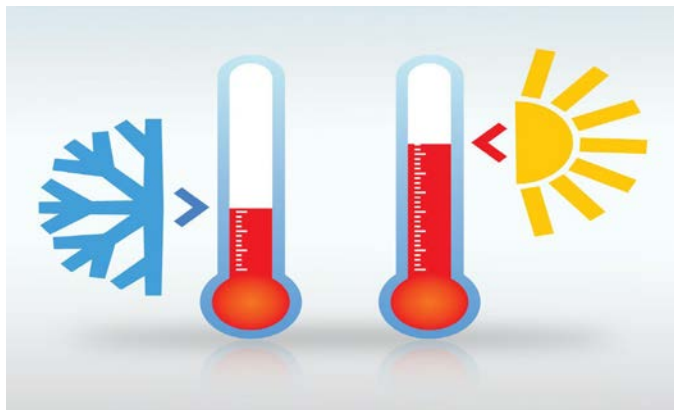
El objetivo de este proyecto es predecir la cantidad de casos de dengue en una ciudad, según sus factores, en base a esto surge la pregunta principal.

¿Qué factores del ambiente son los que se deben tener en cuenta para predecir la propagación del dengue?





¿Qué contiene nuestro dataset?



Temperatura

- Unidades: °C, Kelvin.
- Rango de valores: 21.4 a 30.6°C, 294.89 a 302.61 Kelvin



Humedad

- Unidades: Gramo por kilogramo de aire [g*kg], relativa (%).
- Rango de valores: 11.71 a 20 [g*kg] , 57.78 a 98.61%



Precipitación

- Unidades: milímetros [mm]
- Rango de valores: 0 a 390 [mm]

Hay 1119 registros para cada factor climatológico

¿Qué contiene nuestro dataset?



NDVI

-Rango de valores: -0.45 a 0.54



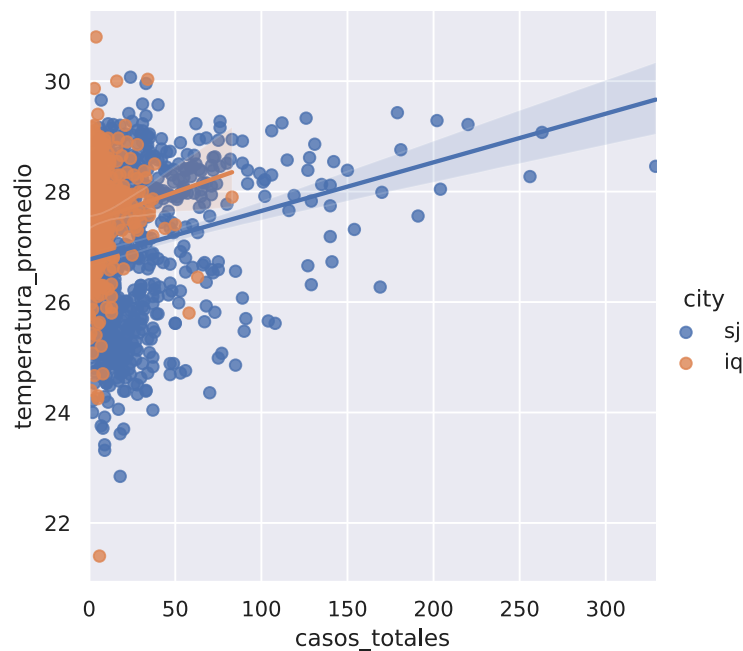
Total de casos de dengue por semana

-Rango de valores: 0 a 329

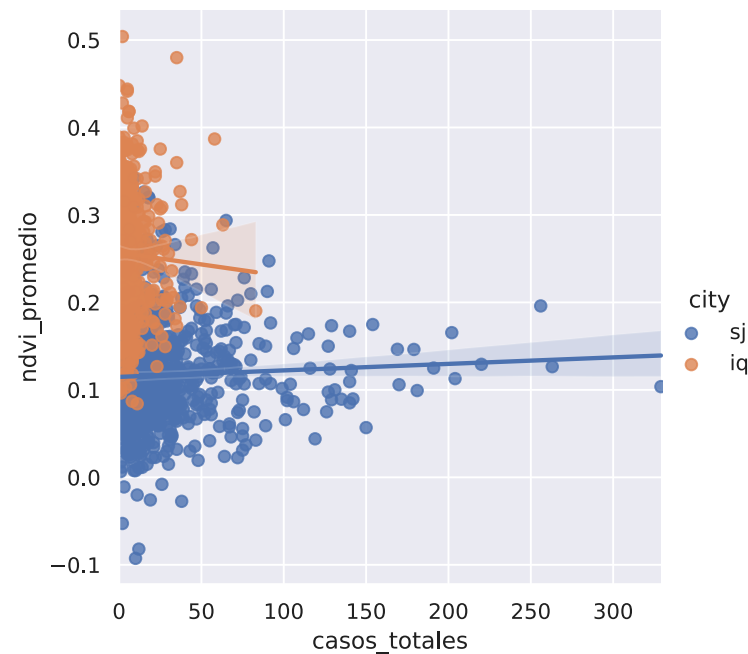
Todo esto para dos ciudades: San Juan de puerto rico (sj) e Iquitos en Perú (iq)



Gráfica de regresión lineal



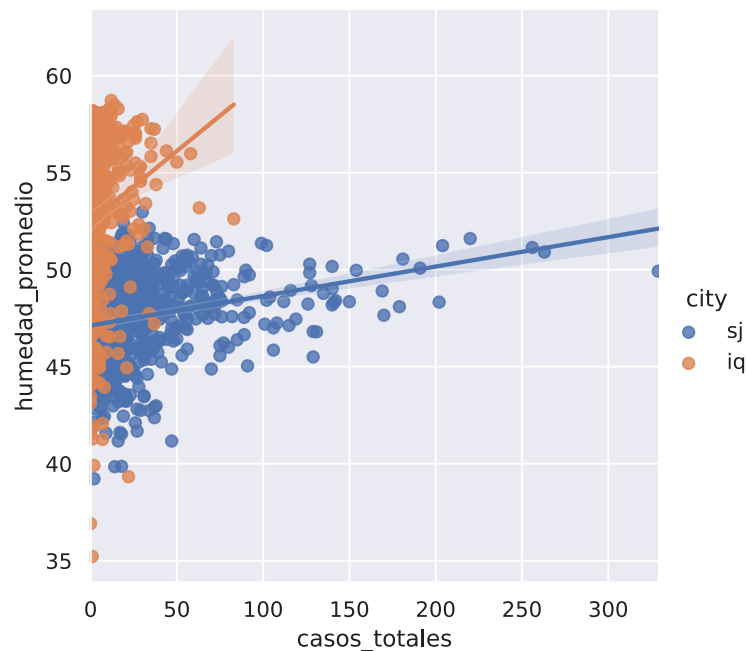
De acuerdo con la gráfica a mayor temperatura, mayor cantidad de casos.



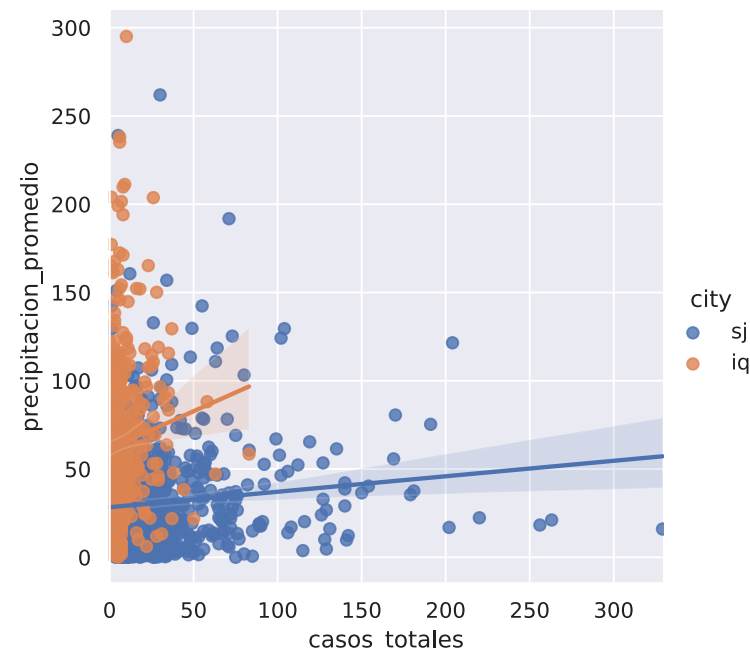
De acuerdo con la gráfica el NDVI cercanos 0.1 ocasiona más casos.



Gráfica de regresión lineal



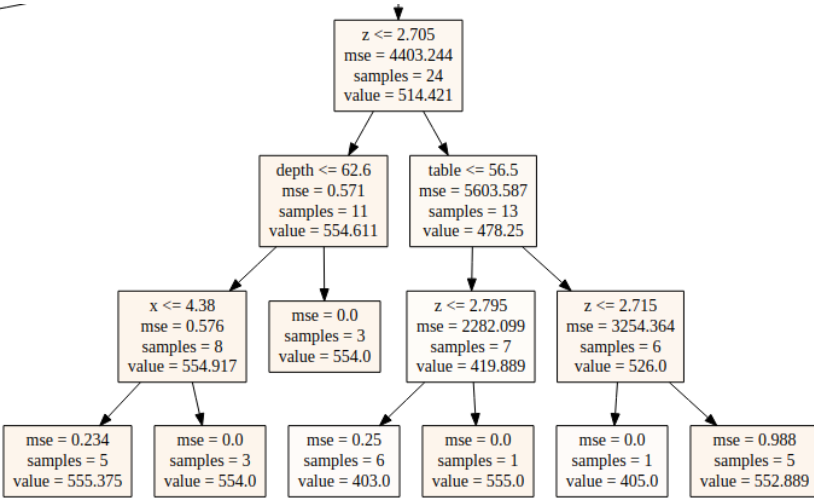
De acuerdo con la gráfica, a mayor humedad mayor cantidad de casos.



De acuerdo con la gráfica, en precipitaciones de 100 hay mayor cantidad de casos.

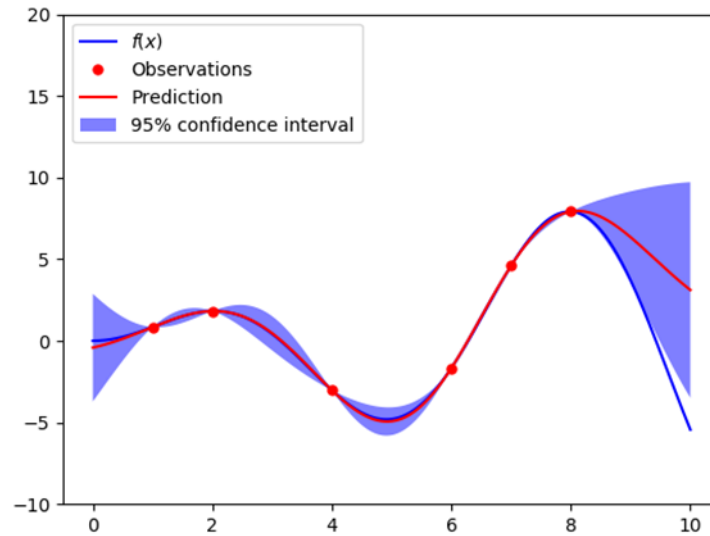


Aprendizaje supervisado



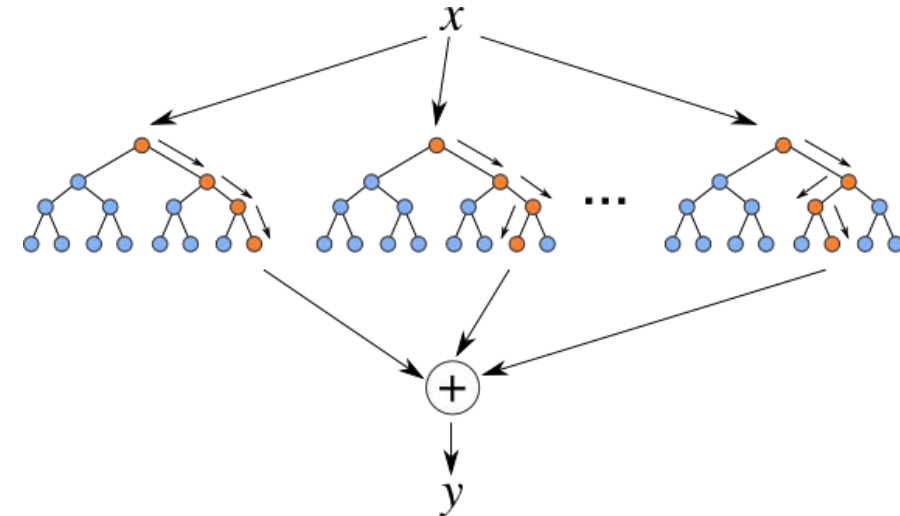
DecisionTreeRegressor (DTR)

No es necesario tener
normzalizados los datos.



GaussianProcessRegression(GPR)

Es muy versátil para usarla
en entrenamientos.



RandomForestRegressor(RFR)

No es afectado por el
sobreajuste de datos

Tabla comparativa de métodos

Datos tomados de la ciudad de Iquitos

Método	Temperatura (C)	NDVI	Humedad	Precipitación
RFR	RMSE 3.253 (+/- 1.12629)	RMSE 3.164 (+/- 1.59908)	RMSE 3.173 (+/- 1.21401)	RMSE 3.193 (+/- 1.27601)
DTR	RMSE 3.111 (+/- 2.06094)	RMSE 3.176 (+/- 1.72913)	RMSE 3.166 (+/- 1.93092)	RMSE 3.129 (+/- 1.84861)
GPR	RMSE 93.632 (+/- 8346.54323)	RMSE 3.278 (+/- 1.41717)	RMSE 94.573 (+/- 6597.20752)	RMSE 3.429 (+/- 2.06630)



Tabla comparativa de métodos

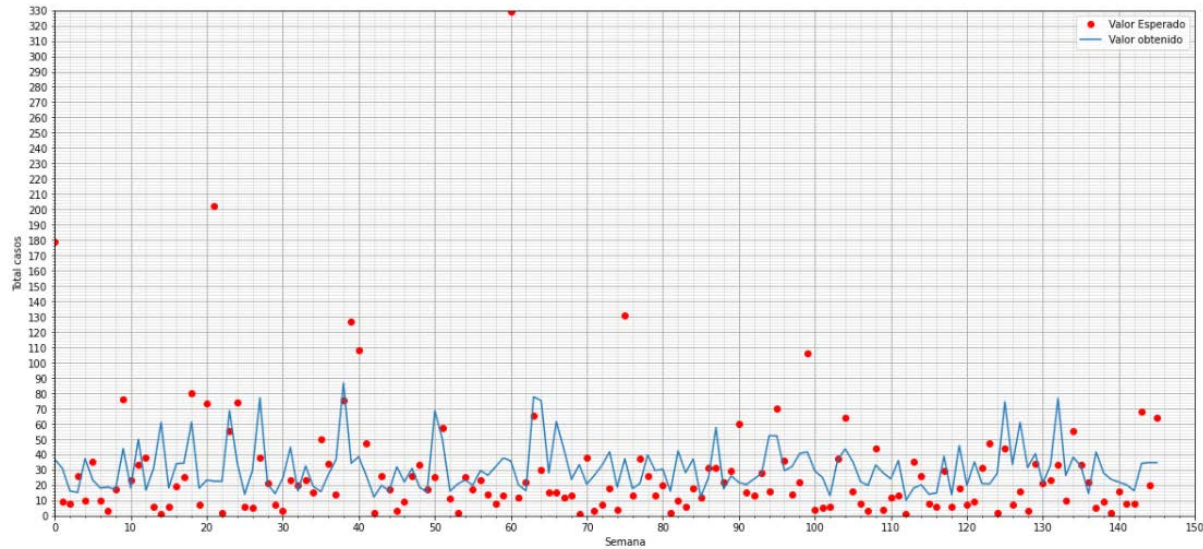
Datos tomados de la ciudad de San Juan

Método	Temperatura (C)	NDVI	Humedad	Precipitación
RFR	RMSE 6.336 (+/- 8.34354)	RMSE 6.250 (+/- 3.71698)	RMSE 6.068 (+/- 4.57751)	RMSE 6.332 (+/- 3.20392)
DTR	RMSE 6.187 (+/- 4.82246)	RMSE 6.124 (+/- 7.08360)	RMSE 6.195 (+/- 6.05332)	RMSE 6.342 (+/- 1.96056)
GPR	68.519 (+/- 2387.99180)	RMSE 6.463 (+/- 5.40583)	RMSE 119.684 (+/- 14186.46925)	RMSE 17.231 (+/- 219.73383)

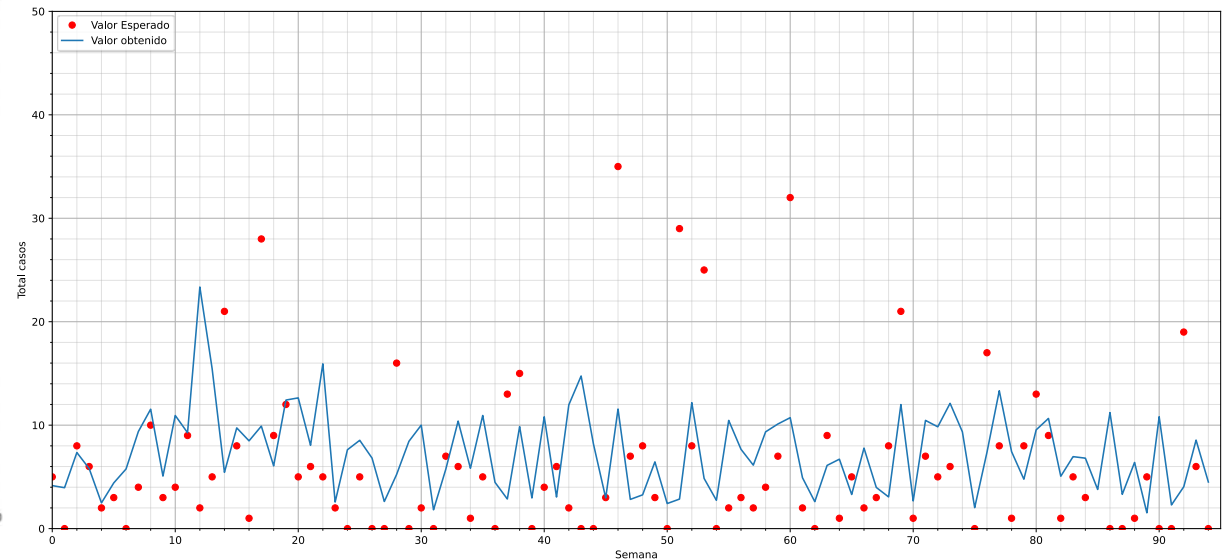


Comparativa ciudades

San Juan



Iquitos

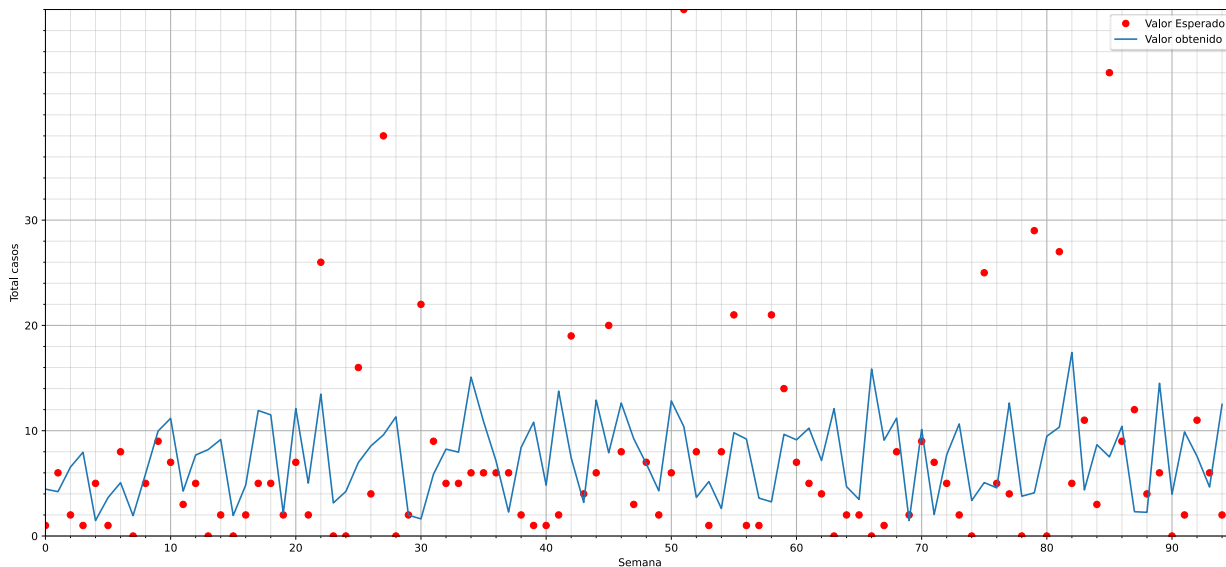
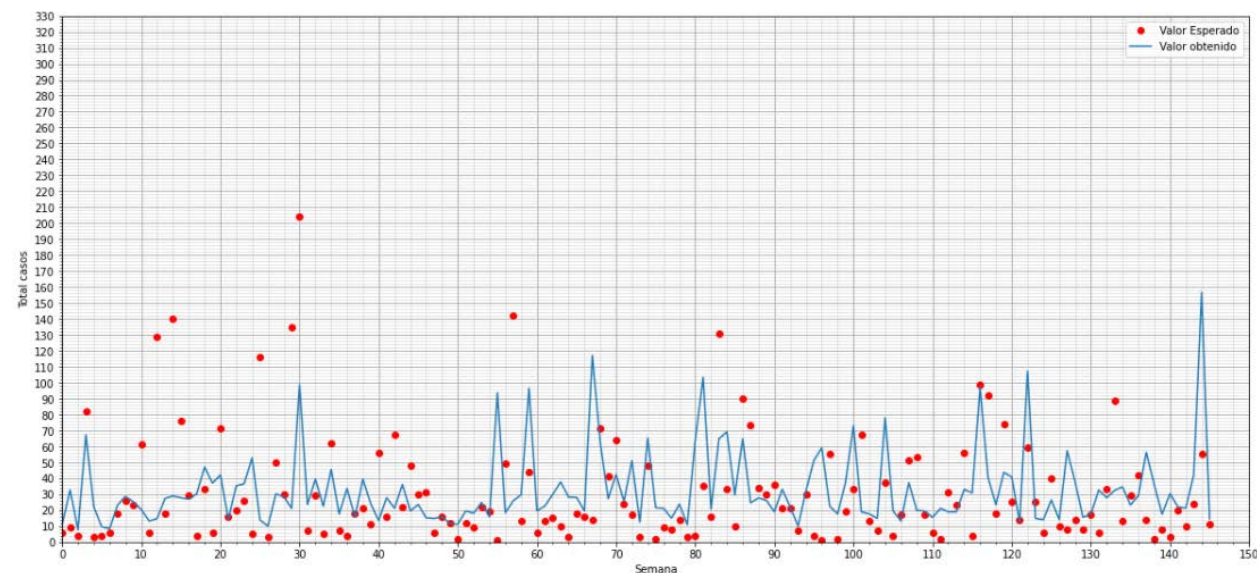


Menor numero de casos en Iquitos a diferencia de San Juan dando una idea de los errores



Comparativa de Factores

- Temperatura K

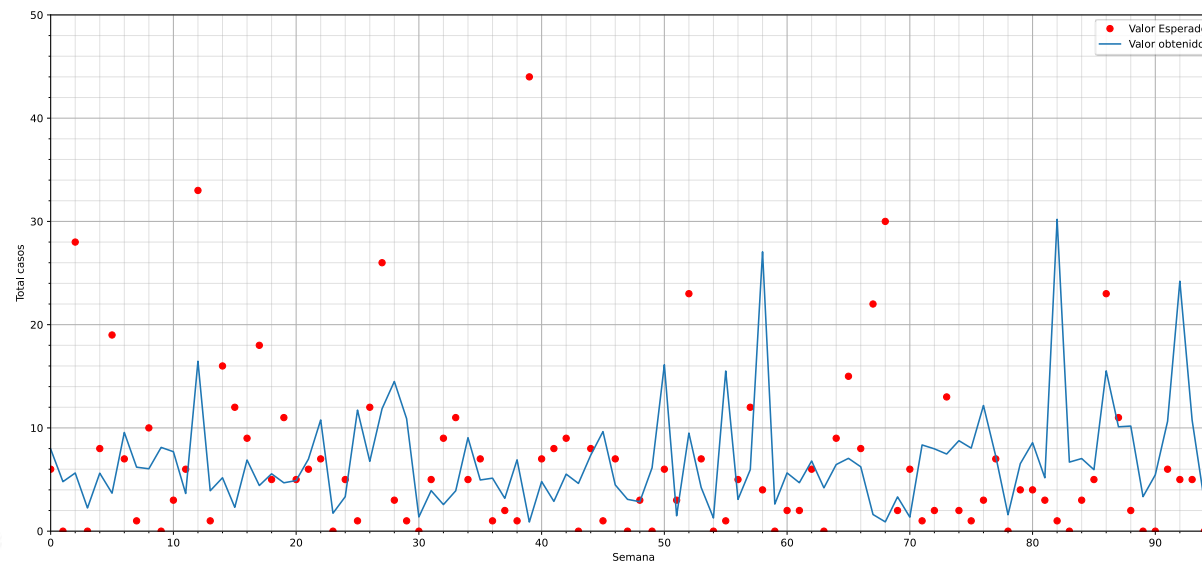
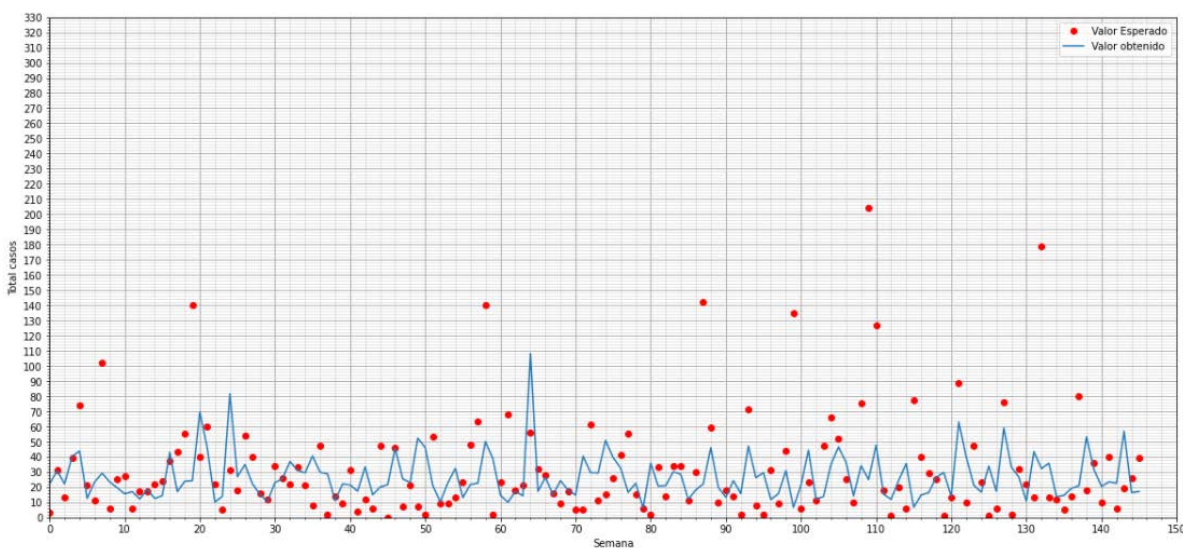


En Iquitos no intenta predecir un numero de casos similar a las semanas mas afectadas, a diferencia de San Juan



Comparativa de Factores

- Temperatura C

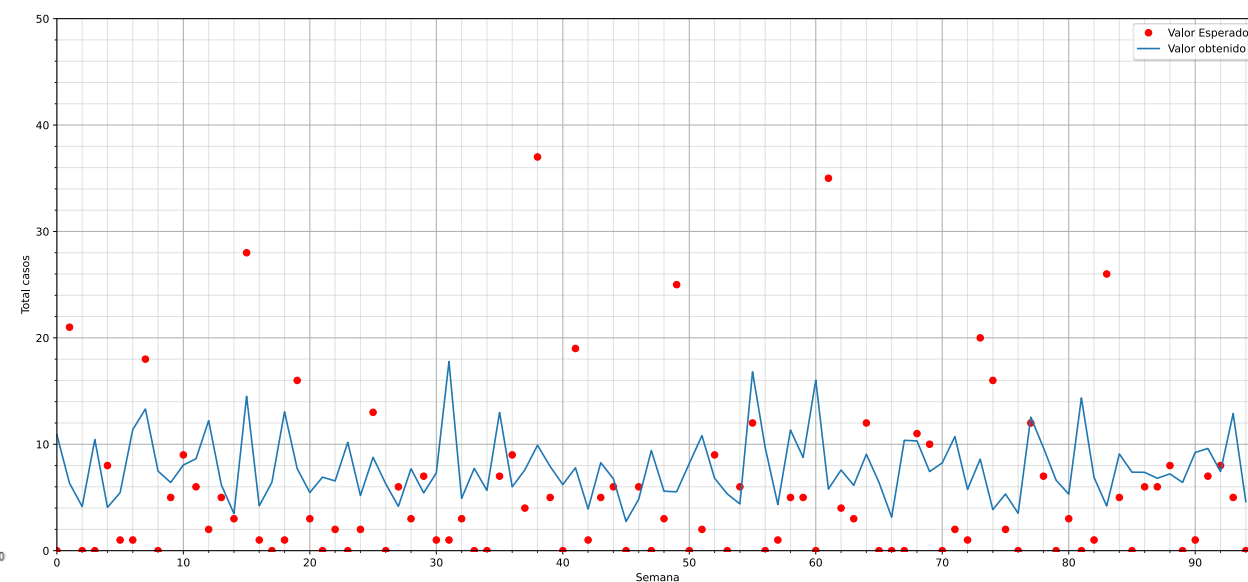
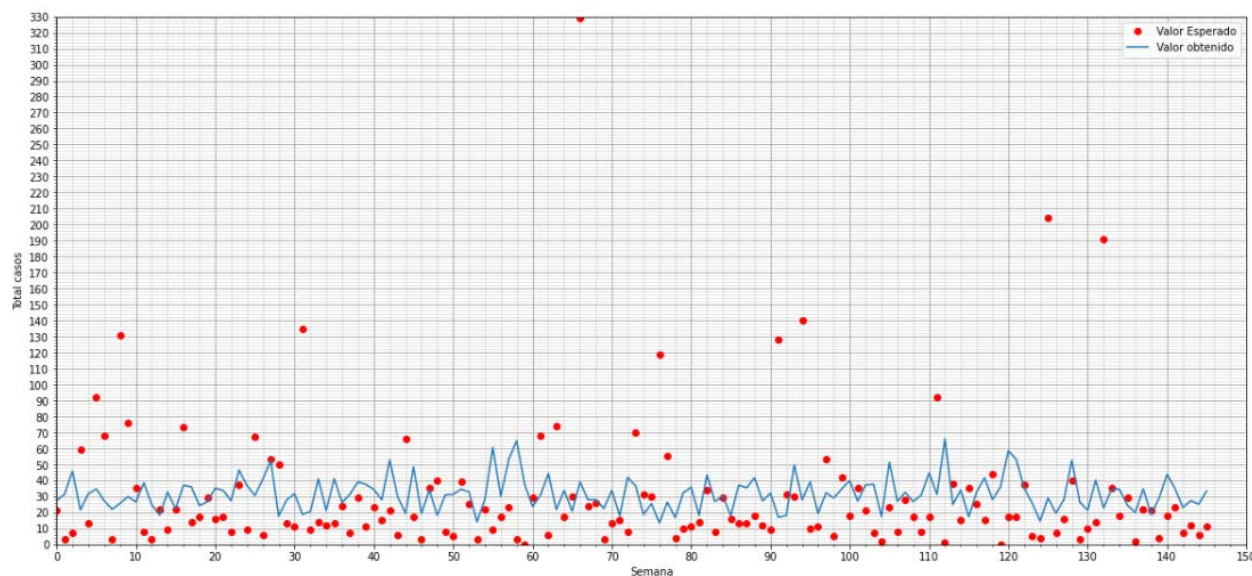


A diferencia de los grados K, en San Juan no intenta predecir un numero de casos similar a las semanas mas afectados, a diferencia de Iquitos



Comparativa de Factores

- NDVI

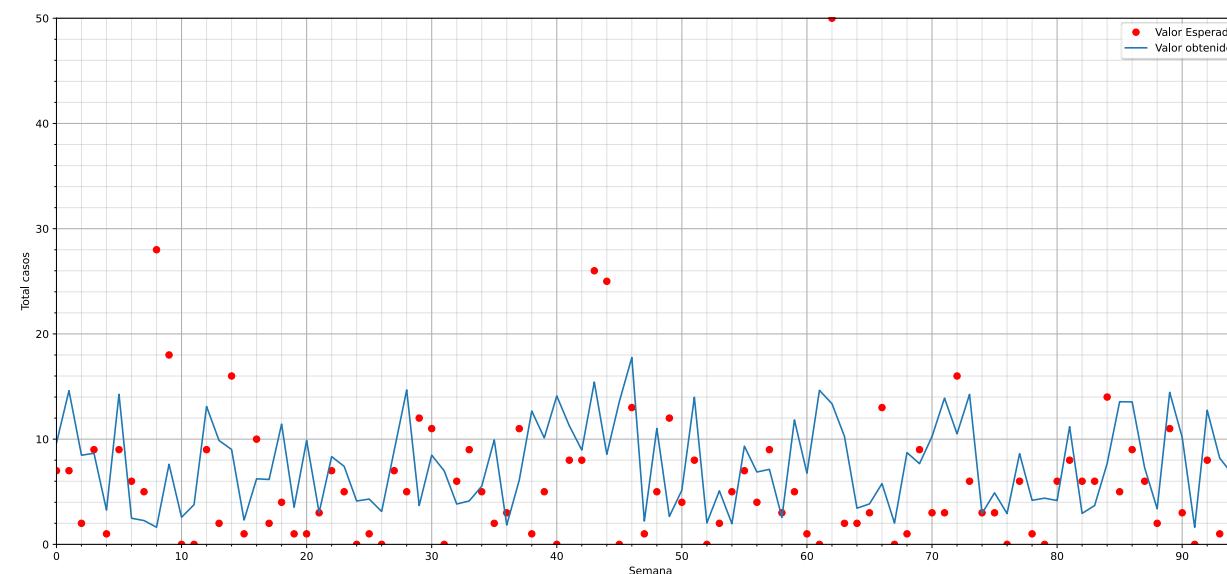
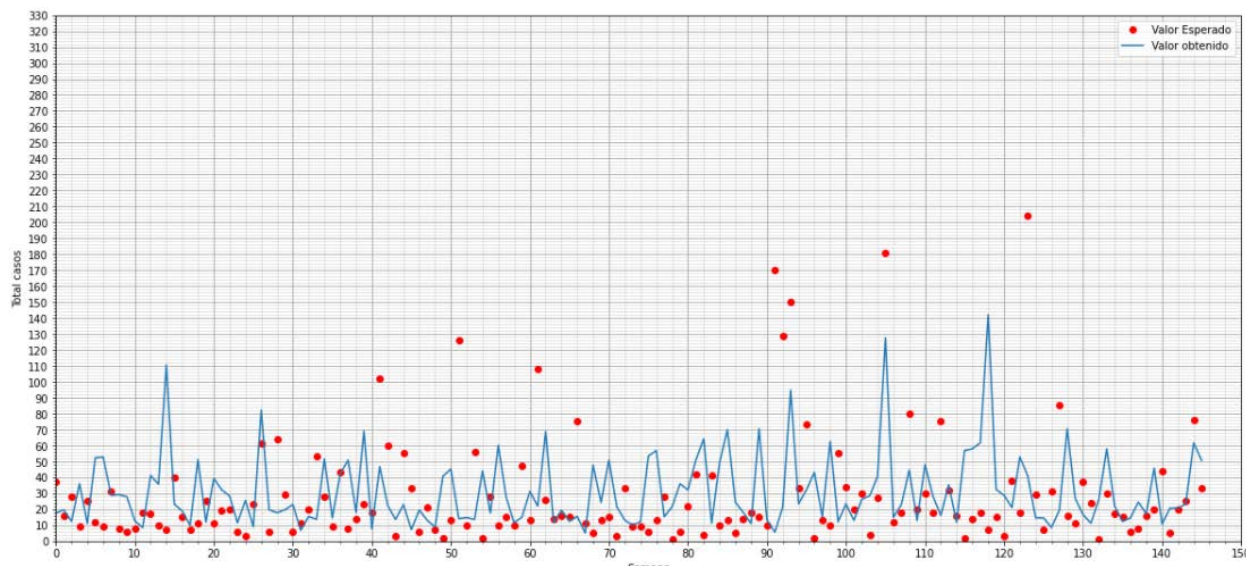


En San Juan no superan los 70 casos, comportamiento un tanto plano, en Iquitos los 20



Comparativa de Factores

- Humedad

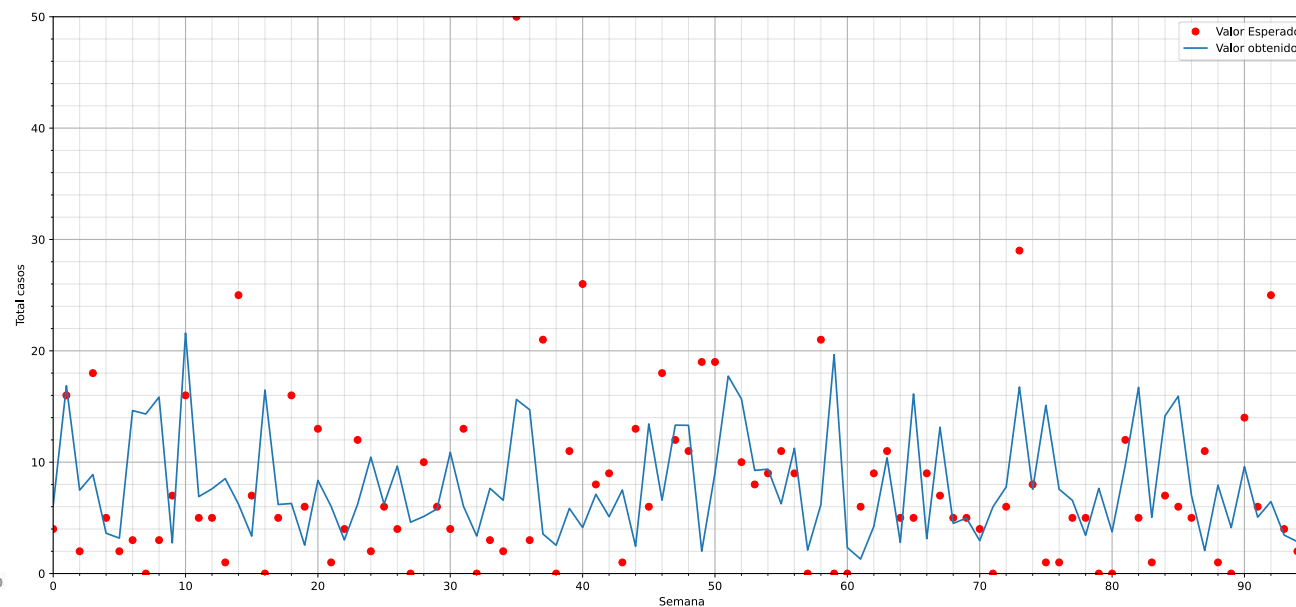
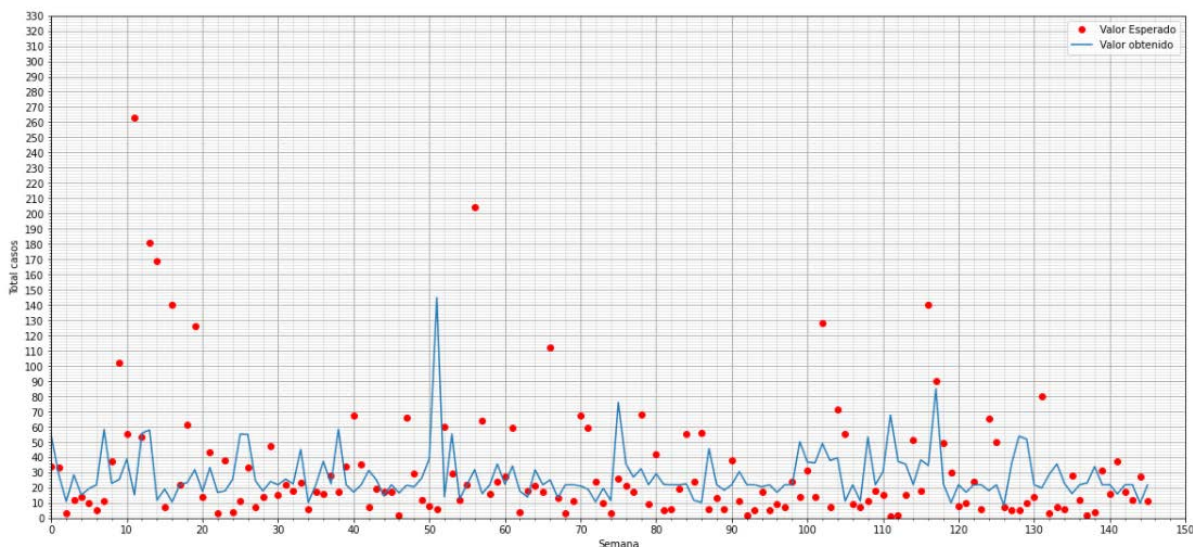


Comportamiento sobresaliente para San Juan, mientras que en Iquitos predice casos con diferencias con poco error.



Comparativa de Factores

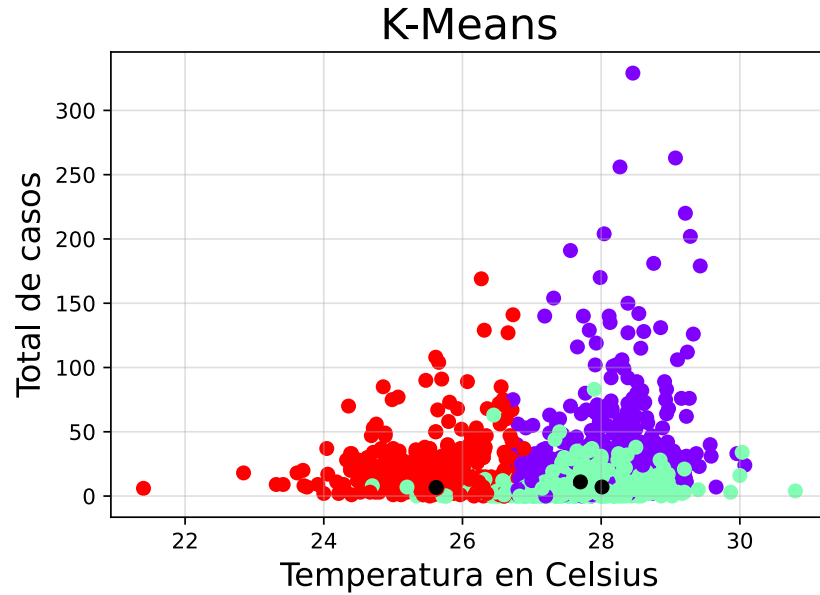
- Precipitación



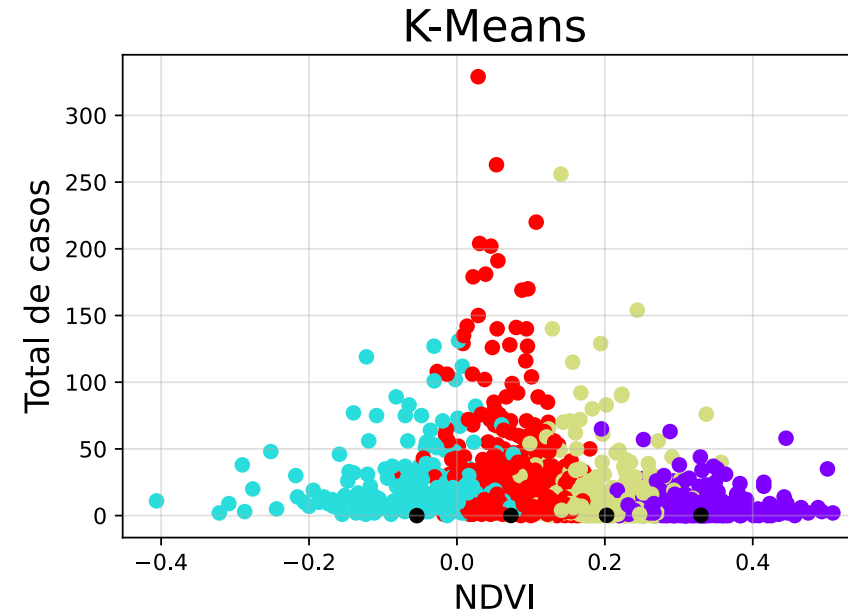
Buen Comportamiento en Iquitos, y en San Juan en las semanas de casos menores.



Aprendizaje no supervisado



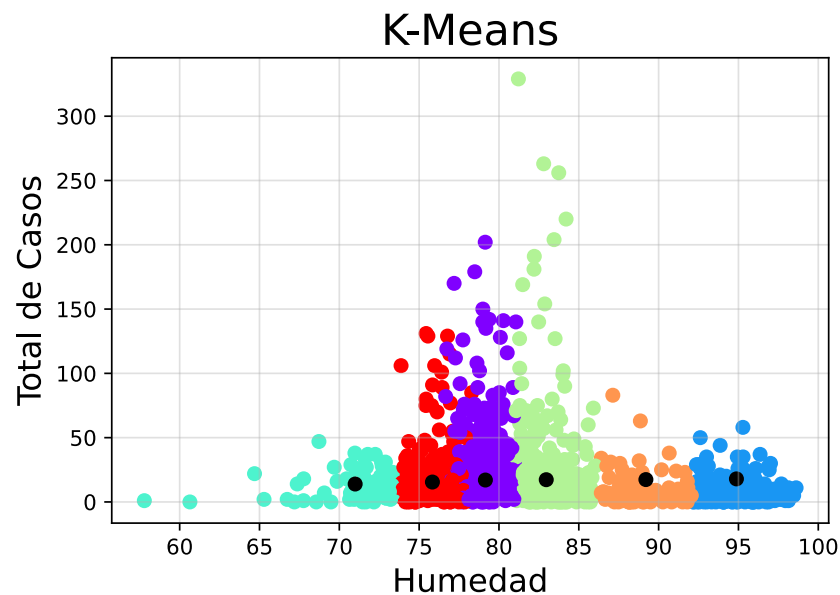
En temperaturas de 28 a 30 °C
tiende a haber más casos.



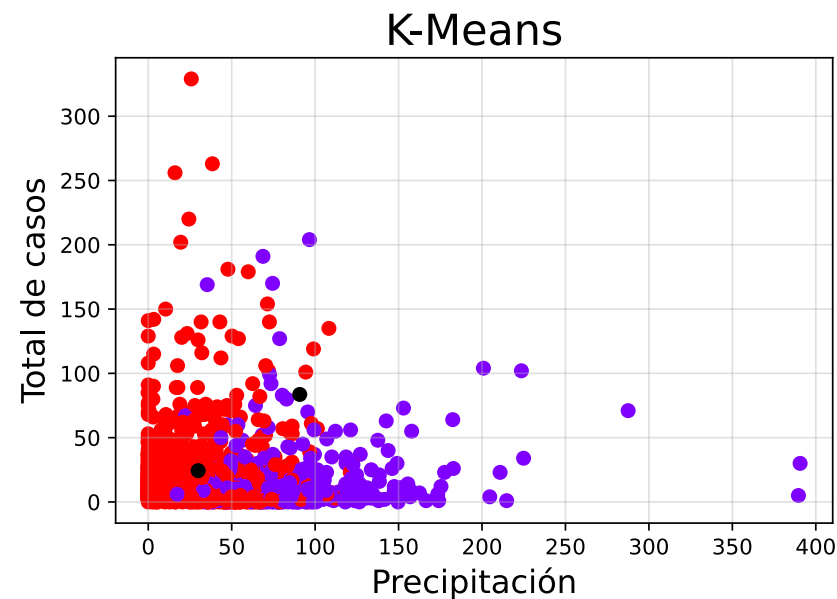
Cuando el NDVI equivale a
0,1 suelen haber más casos.



Aprendizaje no supervisado



Se presentan más casos en niveles de humedad de 80-85%.

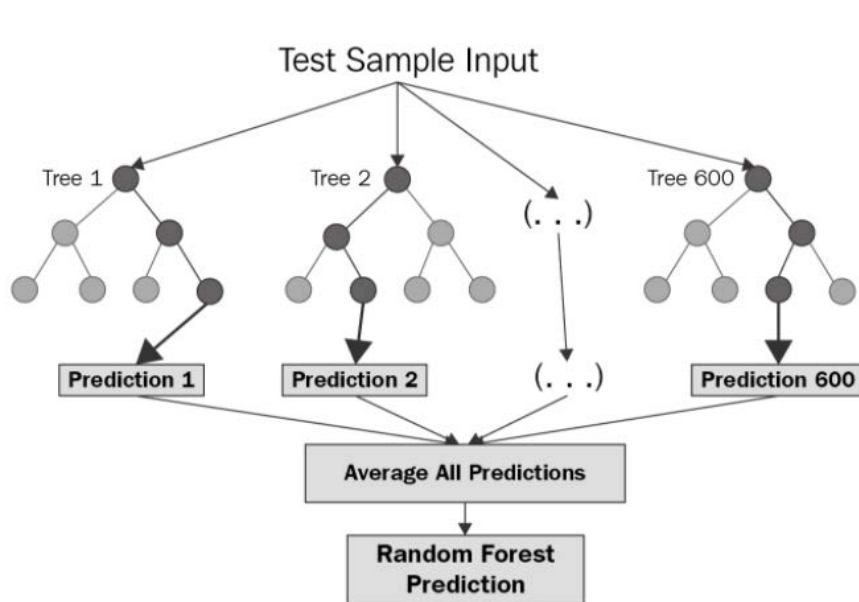


Se presentan más casos en precipitaciones menores a 100 [mm].



CONCLUSIONES

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



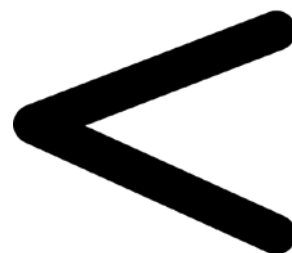


CONCLUSIONES

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



IQUITOS



RMSE



SAN JUAN



CONCLUSIONES

- La implementación de Kmeans permite en los factores de propagación identificar grupos en un rango de valores donde hay mayor números de casos.





CONCLUSIONES

- 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.





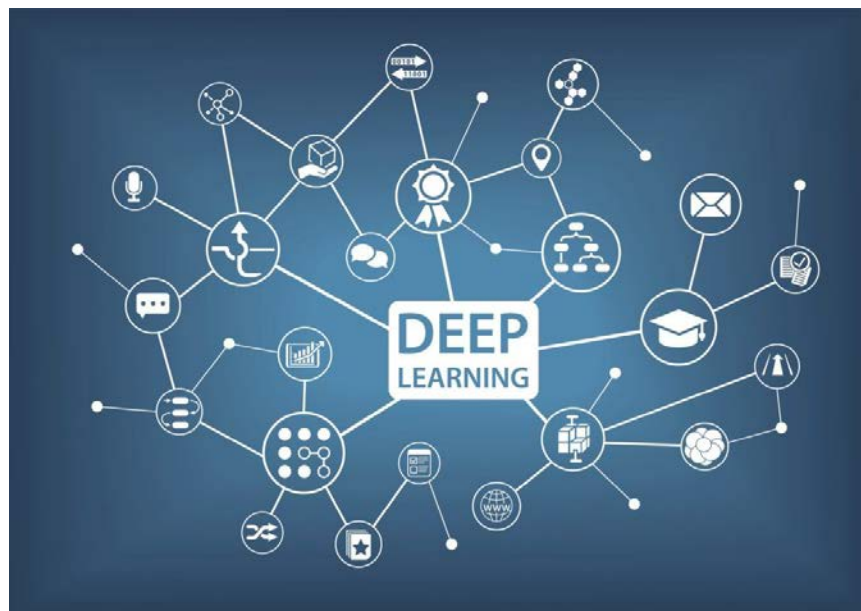
Trabajo futuro y ambición

DengAI: Predicting Disease Spread

HOSTED BY DRIVENDATA



Universidad
Industrial de
Santander



Más Información del concurso:



mejor escenario
de creación e innovación.

www.uis.edu.co



Universidad
Industrial de
Santander

#LaUISqueQueremos

iGracias!

