

Table of Contents

Section No	Content	Page No
1	Introduction	3
2	Dataset	3
3	Data Processing	4
4	Feature Scaling	5
5	Data Splitting	5
6	Models Used	5
7	Conclusion	7

Introduction

Dengue fever is an illness transmitted by mosquitoes and commonly found in tropical and subtropical regions. In less severe instances, its symptoms resemble those of the flu, including fever, rash, muscle and joint discomfort. In more serious cases, dengue fever can lead to substantial bleeding, low blood pressure, and even fatality.

As mosquitoes are carriers of the disease, the spread of dengue is linked to climate factors such as temperature and precipitation. Although the connection with climate is intricate, an increasing number of experts suggest that alterations in climate patterns due to global warming could result in shifts in its distribution, posing significant global health concerns.

.In this analysis we used a series of features collected from two cities and the features relate to environmental factors which helped to predict the number of weekly dengue cases.

Dataset

There are 25 features and 1456 data points. It is a regression problem since we have to predict the number of dengue cases each week (in each location) based on environmental variables describing changes in temperature, precipitation, vegetation, using environmental data collected by various U.S. Federal Government agencies—from the Centers for Disease Control and Prevention to the National Oceanic and Atmospheric Administration in the U.S. Department of Commerce.

Feature explanation:

- NDVI (Normalized Difference Vegetation Index): NDVI values are indicators of vegetation health and can give insights into the environment. Different regions might have varying levels of vegetation, which can affect mosquito breeding habitats and thus dengue spread.
- Precipitation Amount: The amount of rainfall during the week. Rainfall can lead to stagnant water, providing breeding grounds for dengue-carrying mosquitoes.
- Reanalysis Air Temperature: This indicates the air temperature at different times of the day. Temperature can affect mosquito breeding and virus replication rates.

- Reanalysis Dew Point Temperature: The temperature at which air becomes saturated with moisture. This is linked to humidity and can influence mosquito activity.
- Reanalysis Precipitation Amount: Similar to "Precipitation Amount," this can give insights into the amount of rainfall.
- Reanalysis Relative Humidity: This is the amount of moisture in the air relative to its capacity. High humidity can be favorable for mosquito breeding.
- Reanalysis Specific Humidity: The amount of water vapor present in the air. It is related to temperature and can affect mosquito activity.
- Reanalysis TDTR (Temperature Diurnal Range): The difference between the maximum and minimum temperatures in a day. It can influence mosquito behavior and virus replication.
- Station Temperature: The temperature observed at a specific station. This can provide a localized view of the temperature impact on dengue cases.
- Station Diurnal Temperature Range: The temperature range within a day at a specific station.
- Station Precipitation: The amount of precipitation at a specific station.
- Reanalysis Precipitation Amount (kg/m²): This feature represents the amount of precipitation in kilograms per square meter. It quantifies the total rainfall, which can impact mosquito breeding and the overall environment conducive to dengue transmission.
- Reanalysis Saturation Precipitation Amount (mm): This refers to the amount of precipitation as measured by saturation. It's another representation of rainfall, which can contribute to mosquito breeding habitats.
- Reanalysis Specific Humidity (g/kg): Specific humidity measures the actual water vapor content in the air per kilogram of air. It's a crucial parameter to assess atmospheric moisture and its influence on dengue transmission.

All the features are categorical except for two cities. We separate the data into two subsets for the two cities San Juan and Iquitos.

Dataset Preprocessing

We first found out all the summations of the null values present in each column and replaced them by the mean values of those column values. For some data, we deleted the rows.

Following the process of filling in missing data, we decided to remove the 'week_start_date' variable. This choice is made because we are uncertain about its potential contribution to predicting the total instances of dengue cases in the two cities. Then correlations between features were checked using a heatmap, and it was observed that temperature variables were strongly correlated.

Feature Scaling

The following features were scaled:

'weekofyear', 'ndvi_se','precipitation_amt_mm', 'reanalysis_air_temp_k',
'reanalysis_avg_temp_k', 'reanalysis_dew_point_temp_k',
'reanalysis_max_air_temp_k', 'reanalysis_min_air_temp_k',
'reanalysis_precip_amt_kg_per_m2',
'reanalysis_relative_humidity_percent', 'reanalysis_sat_precip_amt_mm',
'reanalysis_specific_humidity_g_per_kg',
'station_avg_temp_c', 'station_max_temp_c', 'station_min_temp_c'

Data Splitting:

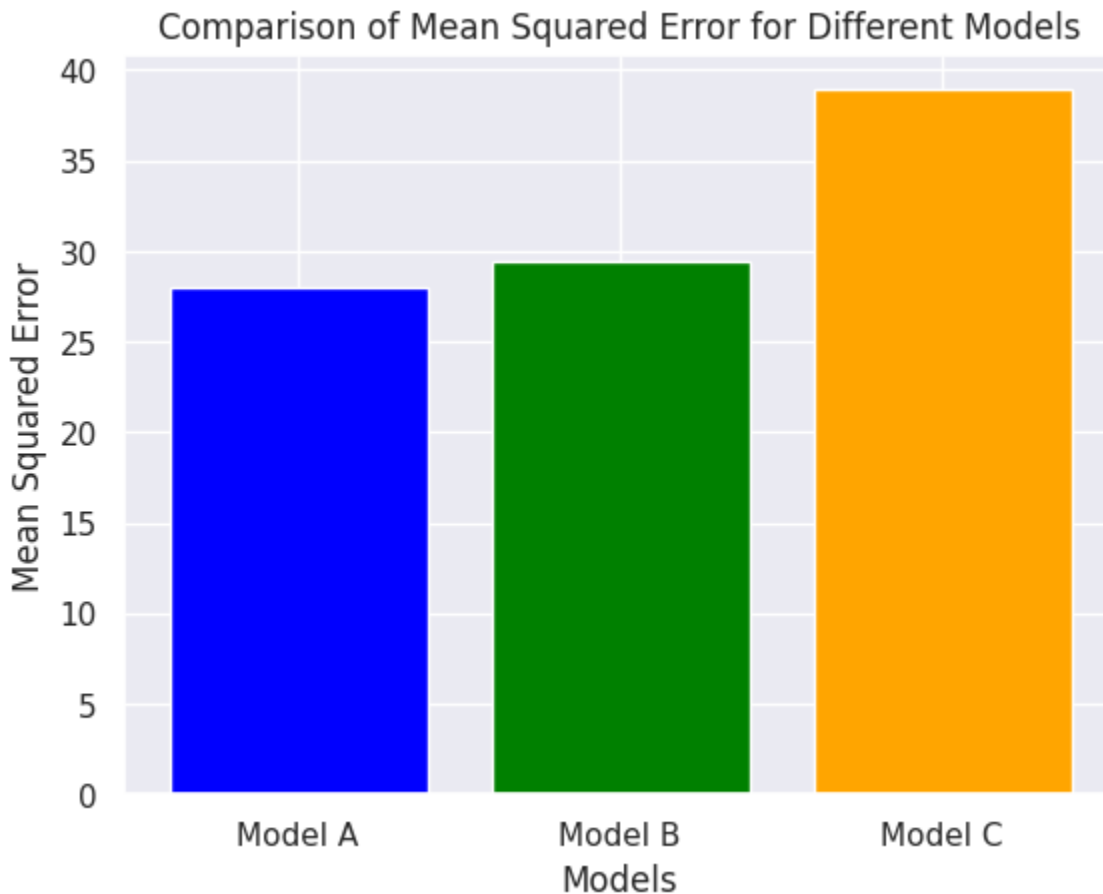
- Train: 80%
- Test: 20%

Models Used

- Linear Regression
- KNN model
- Decision Tree

Each of the models were applied on the train and test values and Mean Squared Error of the predicted y values and test y values were found.

It was found that for both of the cities, Decision Tree gave the most error when compared to the other two.



Conclusion

Using environmental variables, we attempted to predict weekly dengue cases in this study. We sought to learn more about the potential effects of climate on disease transmission by analyzing a wide range of datasets that included variables linked to climate. Preprocessing, feature scaling, and model evaluation were all part of our investigation. This initiative highlights the value of data-driven insights in public health and lays the groundwork for future studies into disease prediction. While models are useful tools, tackling the complex nature of disease transmission calls for ongoing improvement and complete understanding.

