Master of Engineering in Artificial Intelligence
College of Engineering
University of the Philippines - Diliman

Climatologically-Driven Temporal Predictive Modeling of Dengue Cases in Philippine Locations

PRESENTED BY

DELA CRUZ, Joshua

DELA CRUZ, Mary Natalie

June 2024

[redacted]

1 **INTRODUCTION**
Problem Being Solved, Related Works, Objectives

2 **METHODOLOGY**
Dataset, Features, Implementation, Model Evaluation

3 **RESULTS AND DISCUSSION**
Outcome of Experiments

4 **CONCLUSION AND RECOMMENDATION**
Future for the study

1

INTRODUCTION

Background of the Study

Point 1

Endemic in over 128 countries, **dengue fever continues to persist as a significant public health concern** in the Philippines

Point 2

Even with the availability of prevention methods such as vaccine, **there is a concerning decline in vaccine confidence**

Point 3

To achieve the target of zero dengue deaths and minimal cases, it is imperative to **establish policies** through multi-sectoral collaboration

1 INTRODUCTION
Related Works

Recent local studies [1][2] have employed and confirmed the **significance of climate factors** such as rainfall, temperature, and relative humidity in predicting dengue cases.

Classical machine learning models with different feature combinations[3] were tested for predicting monthly dengue cases in five local communities with highest cases, result shown that RF achieved the best performance (**R-squared = 0.5654**) [3]

Exploring **climate-based classical machine learning models for predicting dengue cases** specific to the Philippines is essential due to the influence of climatic variables on mosquito behavior.[4][5]

[1] Carvajal, T.M., Viacrusis, K.M., Hernandez, L.F.T. et al. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. BMC Infect Dis 18, 183 (2018). https://doi.org/10.1186/s12879-018-3066-0
[2] Marigmen, J. L. D. C., & Addawe, R. C. (2022). Forecasting and on the influence of climatic factors on rising dengue incidence in Baguio City, Philippines. Journal of Computational Innovation and Analytics, 1(1), 43-68. https://doi.org/10.32890/jcia2022.1.1.3
[3] Addawe, Jozelle & Caro, Jaime & Juayong, Richelle Ann. (2022). Machine Learning Methods for Modeling Dengue Incidence in Local Communities. 10.1007/978-3-031-17601-2_38.
[4] Yavari Nejad, Felestin, and Kasturi Dewi Varathan. "Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction." BMC Medical Informatics and Decision Making 21.1 (2021): 141.
[5] Sánchez López, Brenda Sofía, et al. "Traditional Machine Learning based on Atmospheric Conditions for Prediction of Dengue Presence." Computación y Sistemas 27.3 (2023): 769-777.

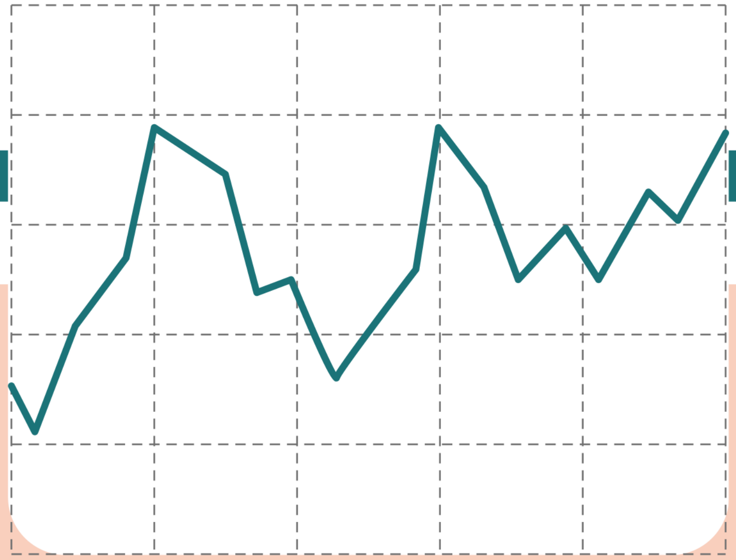
1

INTRODUCTION

Main Objectives

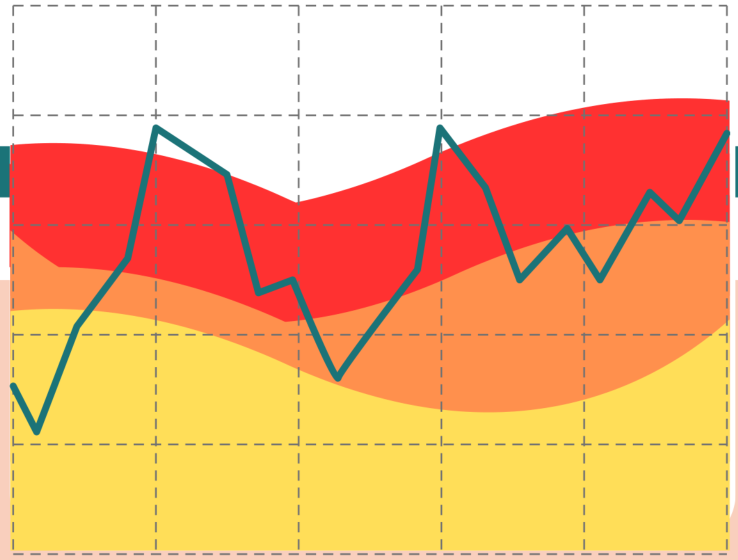
Develop a **predictive model for dengue cases within the same cluster** of cities in the Philippines

1



Conduct **time-series analysis** with the application of **machine learning and climate data**

2



Using the forecasted dengue cases, utilize established guidelines to create **outbreak zones and predict cities that may experience outbreak**

3

Skip to Results and Discussion

2 METHODOLOGY
Implementation

1 DATASET

Collection

Dengue Cases - Readily Available
Climate Data - Web-scrape

Wrangling

Standardized Temporal Granularity, Imputation

Exploratory Analyis

Prelimenary Feature Selection, Stationary Check

2 FOCUS of Study

TSNE + K-Means Clustering

3 FEATURES

Anomaly Detection (DBScan)

Feature Engineering

Anomalies, Fourier Transformation, Moving Average

Selection

Correlation Threshold (95%)

4 MODEL

TPot

Survey of models, best models:

SGDRegressor

XGBoost

SARIMA/SARIMAX

5 FORECAST

Model Evaluation

R-squared and MSE
AIC

Outbreak Classification

WHO Moving Threshold

Hyperparameter Tuning Technique

- Grid Search
- Optuna
- Auto Arima

2

METHODOLOGY

Dataset

Back to Implementation Overview

- **Philippine Dengue Cases and Deaths** [1] recorded by Department of Health-Epidemiology Bureau in the Philippines from 2016 to 2020.
- **Climate Data** [2] from NASA *Prediction Of Worldwide Energy Resources* (POWER) project

Hydrological Variables

Evaporation Land
Evapotranspiration Energy Flux
Profile Soil Moisture
Root Zone (Surface) Soil Wetness
Precipitation
Relative (Specific) Humidity at (2 Meters, 10 Meters)

Solar Radiation

All Sky Insolation Clearness Index
All Sky Surface Longwave Downward Irradiance
All Sky Surface Shortwave Diffuse Irradiance
All Sky Surface Shortwave Downward Direct Normal Irradiance
All Sky Surface Shortwave Downward Irradiance
All Sky Surface UV Index

Cloud Cover

Cloud Amount
Cloud Amount at Daytime
Cloud Amount at Nighttime

Temperature

Temperature at (Surface, 2 Meters, 10 Meters)
(Mean, Maximum, Minimum, Range)

Windspeed

Wind Speed at (2 Meters, 10 Meters, 50 Meters)
(Mean, Maximum, Minimum, Range)

[1] Retrieved from https://data.humdata.org/dataset/philippine-dengue-cases-and-deaths, 2024
[2] National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) Prediction of Worldwide Energy Resource (POWER) Project funded through the NASA Earth Science/Applied Science Program, retrieved from the POWER Project's Hourly 2.x.x version on 2024/05/10.

2

METHODOLOGY

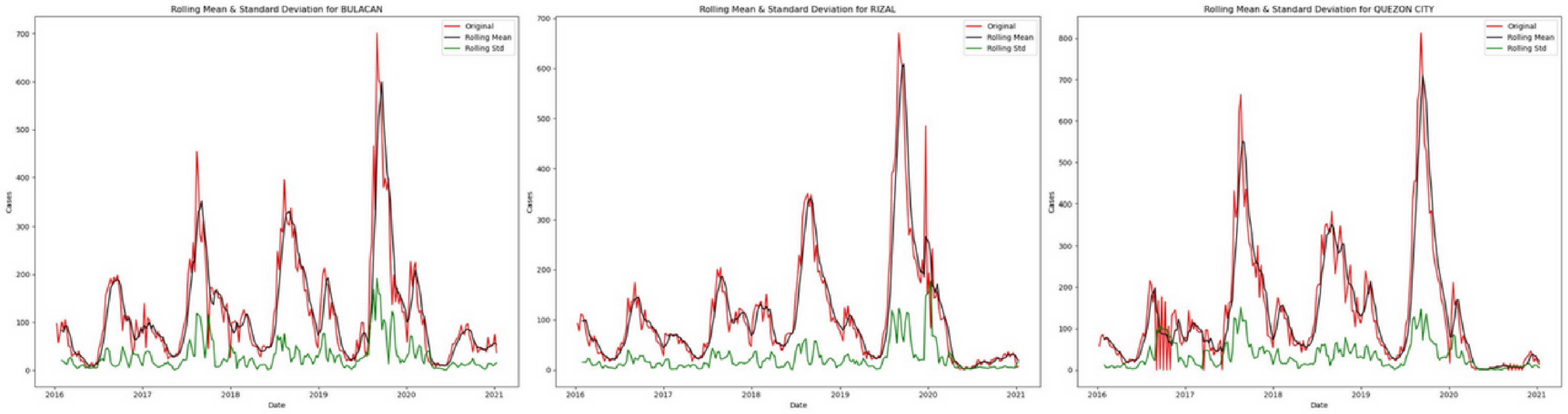
Dataset

Back to Implementation Overview

Webscraping per location and per time

Aggregation of mean, max and min per and filling of null values

Stationarity Check



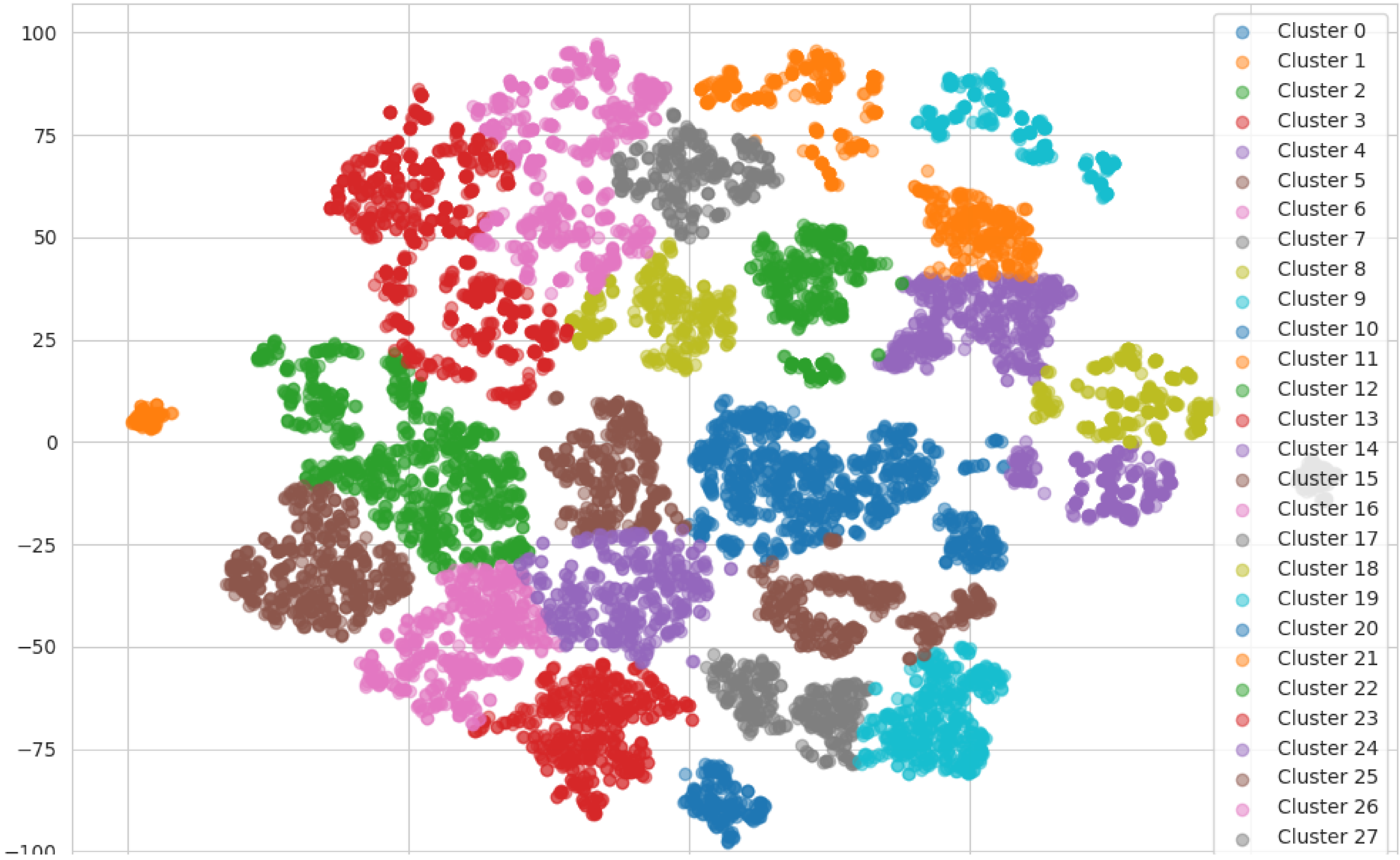

```
Results for: BULACAN
ADF Test statistic:  -4.687209028908886
p-value:  8.89340327049059e-05
Critical Values: {'1%': -3.4562572510874396, '5%': -2.8729420379793598, '10%': -2.5728461399461744}

Results for: RIZAL
ADF Test statistic:  -3.961942028539276
p-value:  0.0016234425953196826
Critical Values: {'1%': -3.4562572510874396, '5%': -2.8729420379793598, '10%': -2.5728461399461744}

Results for: QUEZON CITY
ADF Test statistic:  -3.846997454400206
p-value:  0.002461458093812482
Critical Values: {'1%': -3.456360306409983, '5%': -2.8729872043802356, '10%': -2.572870232500465}
```

[Back to Implementation Overview](#)

2 METHODOLOGY
Focus of the Study



TSNE + K-Means Clustering

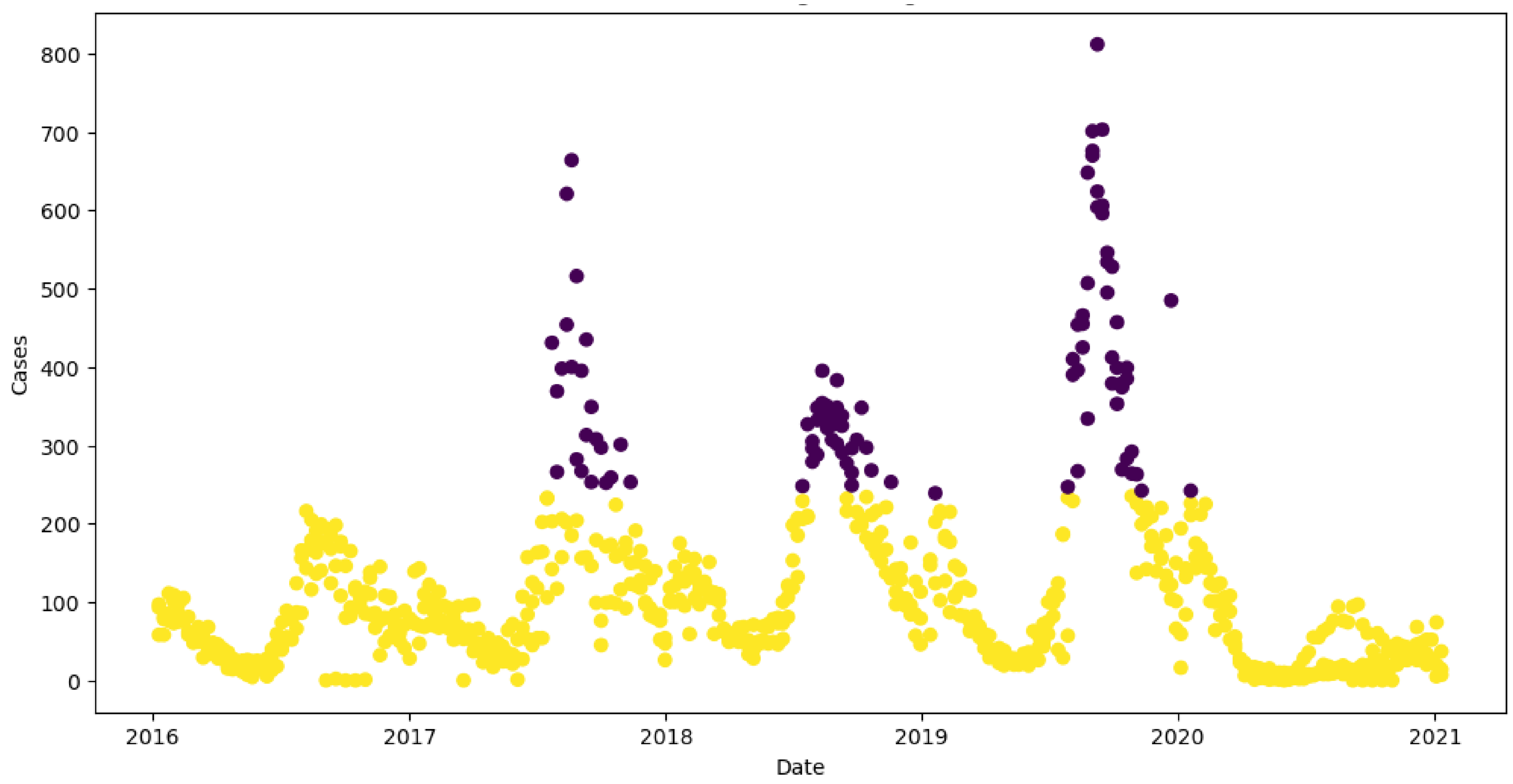
Silhouette Score: 0.42

TSNE Parameters	Value
n components	2
perplexity	39.14
learning rate	527.07

K-Means Parameters	Value
number of clusters	28
initialization method	k-means++
initialization runs	23

2 METHODOLOGY
Feature Processing

Back to Implementation Overview



DBSCAN

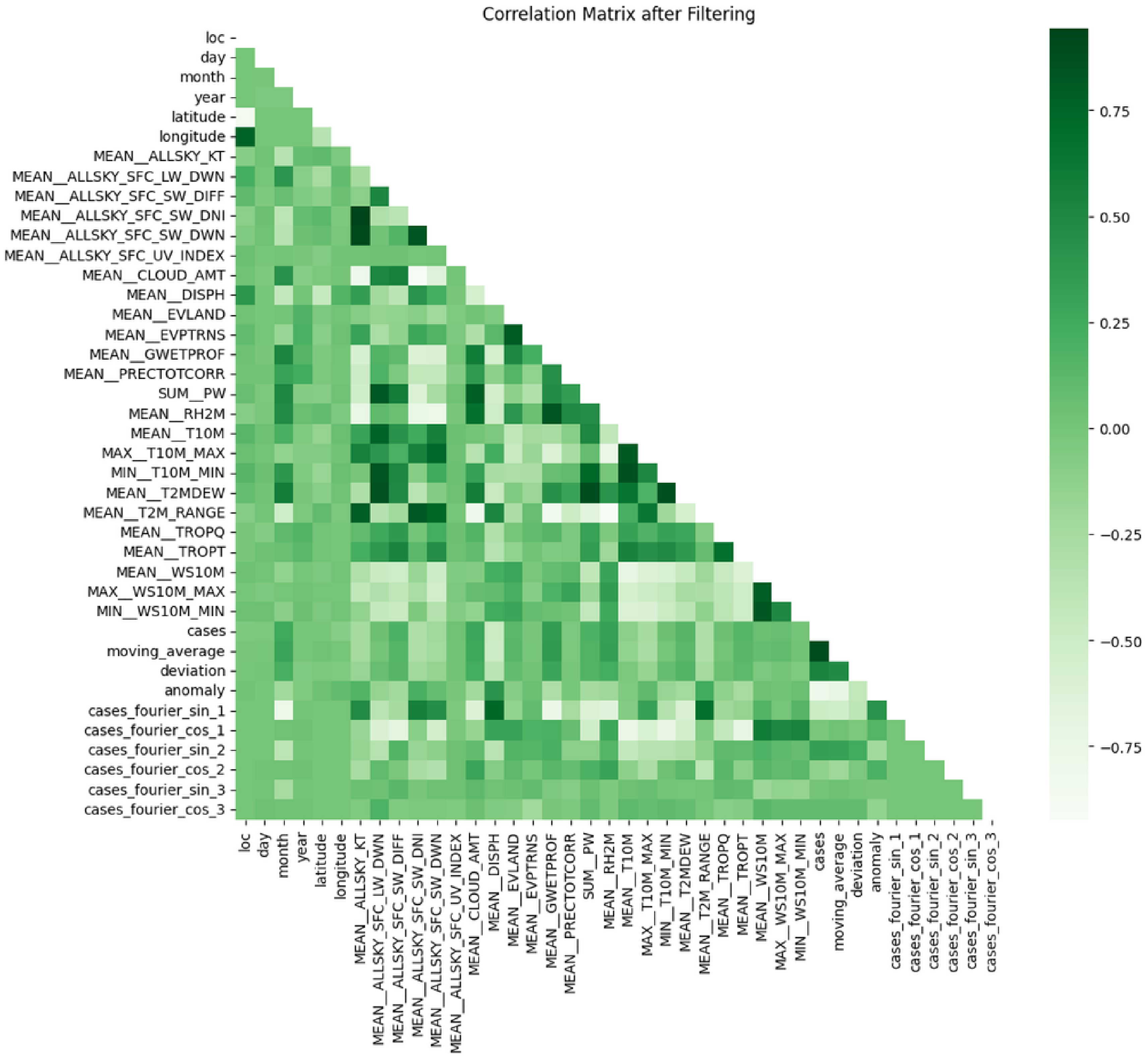
Silhouette Score: 0.72

Parameters	Value
epsilon	10
minimum samples	20

Anomalies = **14%**

2 METHODOLOGY
Feature Processing

Back to Implementation Overview



UNIVARIATE CORRELATION CHECK

Threshold: 0.95

Before Filtering: 59
After Filtering: 40

METHODOLOGY

Model Selection

Back to Implementation Overview

Auto-ARIMA

MODEL	PARAMETERS TUNED
SARIMA	**p** = 0–5, **d** = 0–2, **q** = 0–5, **P** = 0–5, **D** = 0–2, **Q** = 0–5, **m** = 4
SARIMAX	**p** = 0–5, **d** = 0–2, **q** = 0–5, **P** = 0–5, **D** = 0–2, **Q** = 0–5, **m** = 4

2

METHODOLOGY

Model Selection

Back to Implementation Overview

TPOT-SURVEYED MODELS

MODEL	PARAMETERS TUNED
SGDRegressor	**loss, penalty, alpha, learning_rate, fit_intercept, l1_ratio., eta0, power_t**
GradientBoostingRegressor	**n_estimators, loss, learning_rate, max_depth, min_samples_split, min_samples_leaf, subsample, max_features, alpha**
LinearSVR	loss, dual, tol, C, epsilon
ElasticNetCV	l1_ratio, tol
RandomForestRegressor	**n_estimators, max_features, min_samples_split, min_samples_leaf, bootstrap**
XGBRegressor	**n_estimators, max_depth, learning_rate, subsample, min_child_weight, n_jobs, verbosity, objective**
AdaBoostRegressor	n_estimators, learning_rate, loss
DecisionTreeRegressor	max_depth, min_samples_split, min_samples_leaf
RidgeCV	N/A
ExtraTreesRegressor	n_estimators, max_features, min_samples_split, min_samples_leaf, bootstrap
KNeighborsRegressor	n_neighbors, weights, p

2 METHODOLOGY
Forecasting

Back to Implementation Overview

WHO Standard[1]

Let:

- $\text{cases}(t)$: Number of dengue cases in week t.
- $MA_4(t)$: 4-week moving average ending in week t.
- $MA_{3\times 4}(t)$: Moving average of three consecutive 4-week periods ending in week t.
- $SD_4(t)$: Standard deviation of the dengue cases over the 4 weeks prior to week t.
- $\text{Threshold}(t)$: Outbreak threshold for week t.

4-week Moving Average

$$MA_4(t) = \frac{1}{4}\sum_{i=0}^{3} \text{cases}(t-i)$$

Moving Average of Three 4-week Periods

$$MA_{3\times 4}(t) = \frac{1}{3}\sum_{j=0}^{2} MA_4(t-4j)$$

Standard Deviation of Previous 4 Weeks

$$SD_4(t) = \sqrt{\frac{1}{4}\sum_{i=0}^{3}(\text{cases}(t-i) - MA_4(t))^2}$$

Outbreak Threshold

$$\text{Threshold}(t) = MA_{3\times 4}(t) + 2 \times SD_4(t)$$

Determining an Outbreak

An outbreak is determined if the number of cases in week t exceeds the outbreak threshold:

$$\text{Outbreak}(t) = \begin{cases} 1 & \text{if } \text{cases}(t) > \text{Threshold}(t) \\ 0 & \text{otherwise} \end{cases}$$

[1]World Health Organization. Technical handbook for dengue surveillance, dengue outbreak prediction/detection and outbreak response ("model contingency plan"). WHO Library Cataloguing-in-Publication Data, 1–92. http://www.who.int/tdr/news/2016/ handbook-dengue-outbreak/en/. Accessed 12 June 2024.

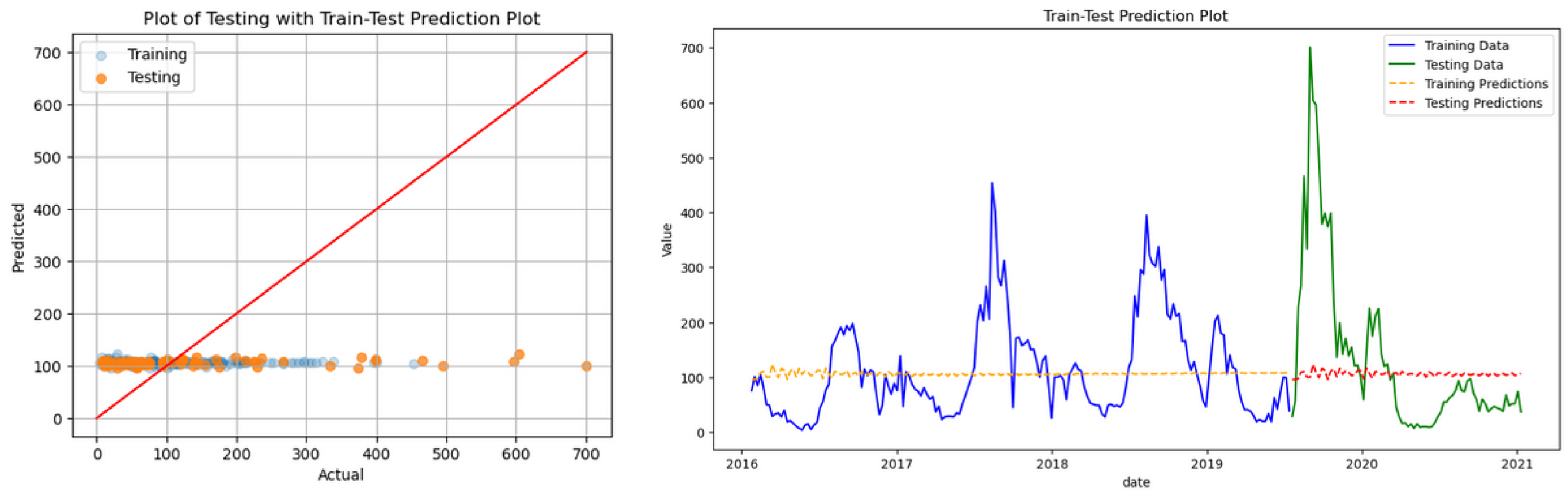
3 RESULTS AND DISCUSSION

Hyperparameter Optimization and Test Result for Bulacan

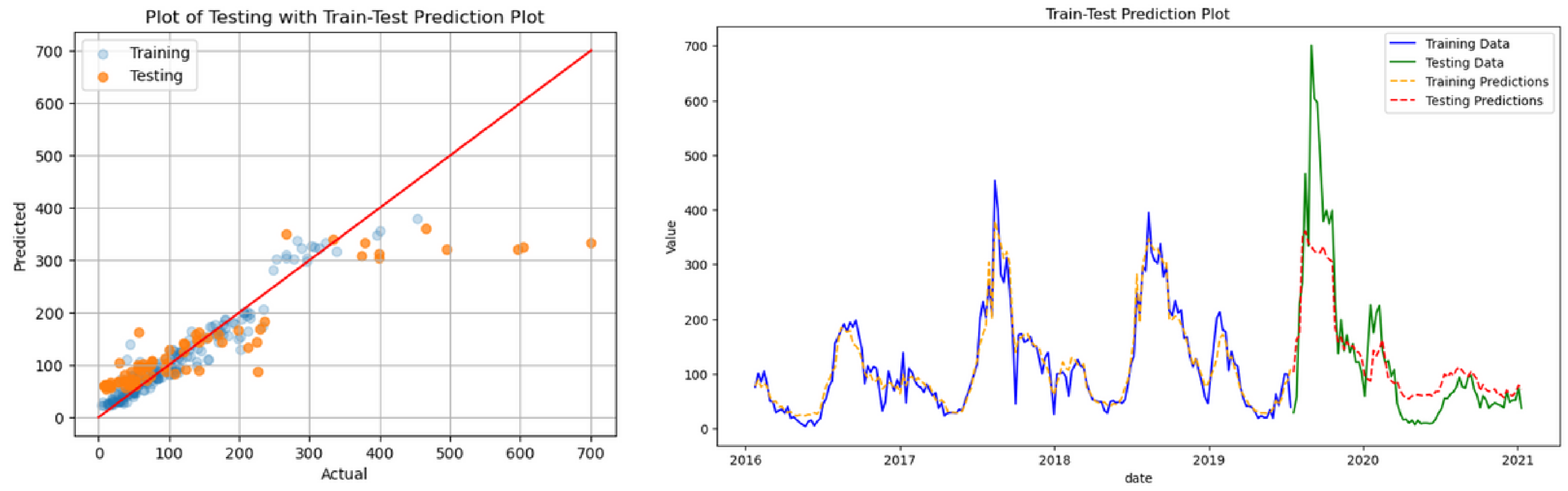
Model	Tuned Hyperparameters	R-squared	RMSE
SARIMA	**p** = 2, **d** = 1, **q** = 3, **P** = 3, **D** = 0, **Q** = 2, **m** = 4	0.02%	151.75
SARIMAX	**p** = 1, **d** = 0, **q** = 1, **P** = 0, **D** = 0, **Q** = 0, **m** = 4	79.38%	68.29
XGBoost	‘**n_estimators’**=164, **max_depth**'= 4, ' **learning_rate**'= 0.0283, '**subsample**'= 0.452, '**min_child_weight**'= 14	72.93%	78.23
SGDRegressor	‘**alpha**’=0.0304, **fit_intercept**=True, **l1_ratio**=0.8563, **loss**='huber', **penalty**='elasticnet', **power_t**=0.2232	85.10%	52.07

3 RESULTS AND DISCUSSION

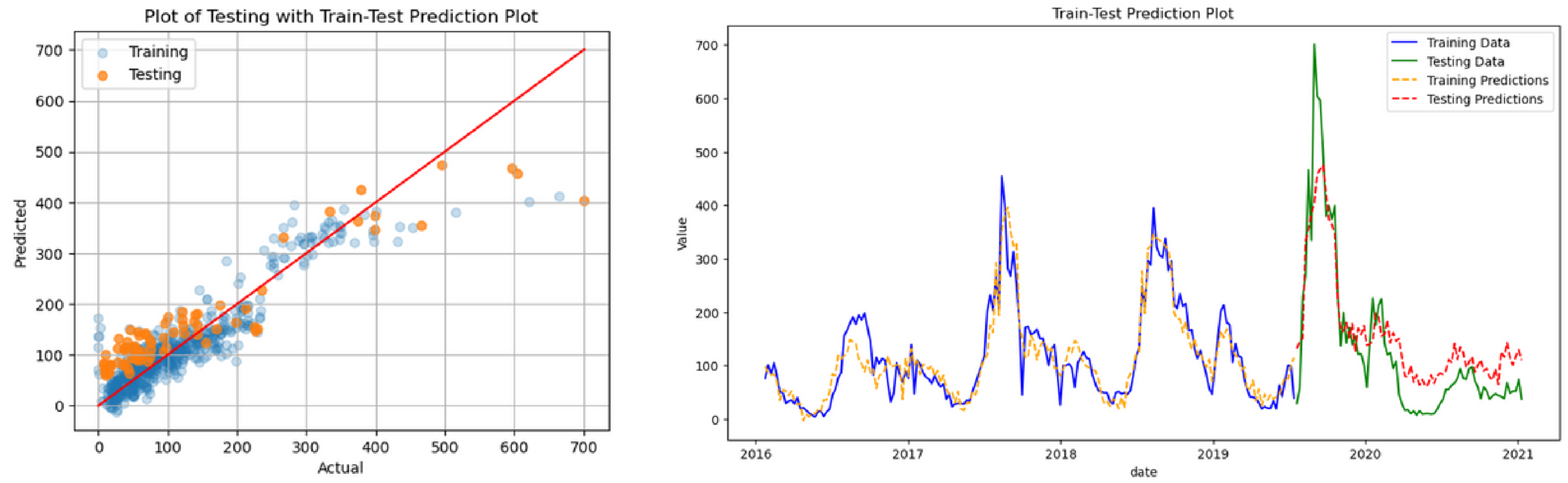
Plot Results for **Bulacan**



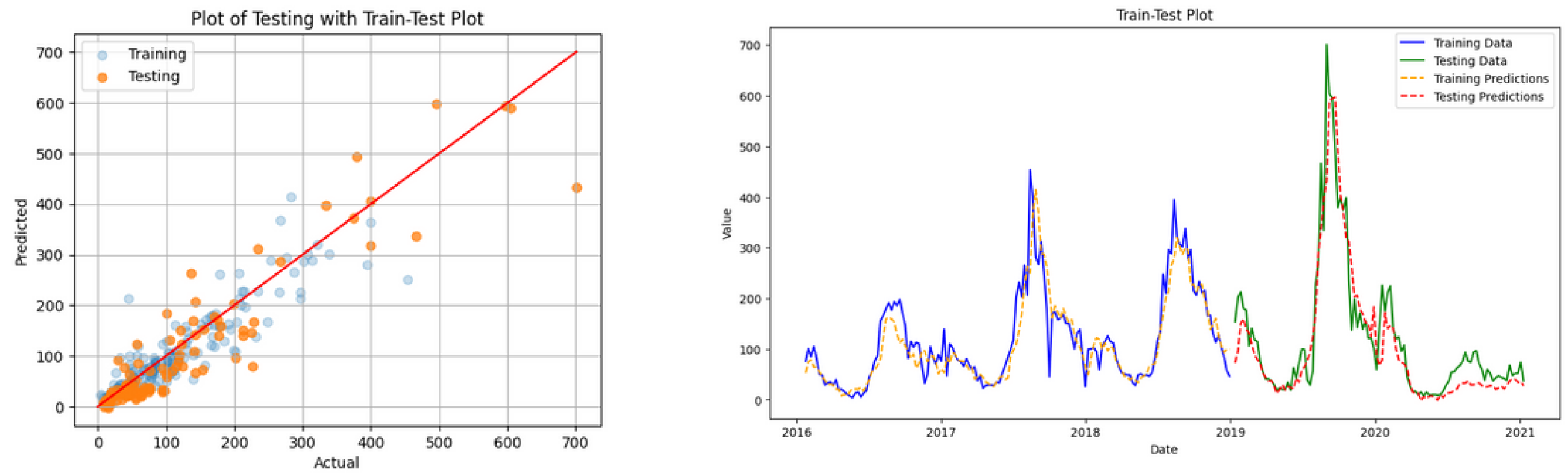
SARIMA



XGBoost



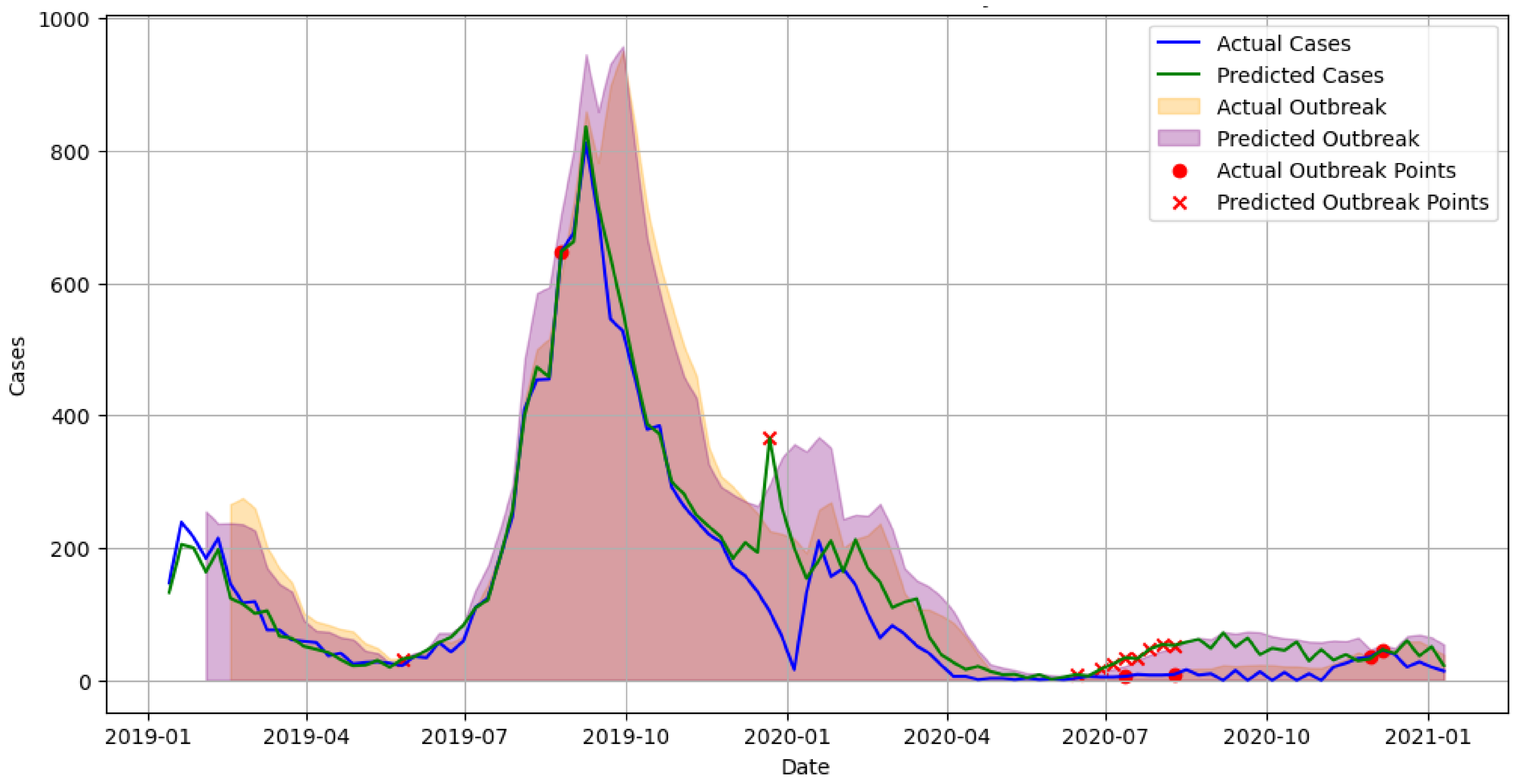
SARIMAX



SGDRegressor

3 RESULTS AND DISCUSSION

Outbreak Prediction based on WHO STANDARD



UNIVARIATE CORRELATION CHECK

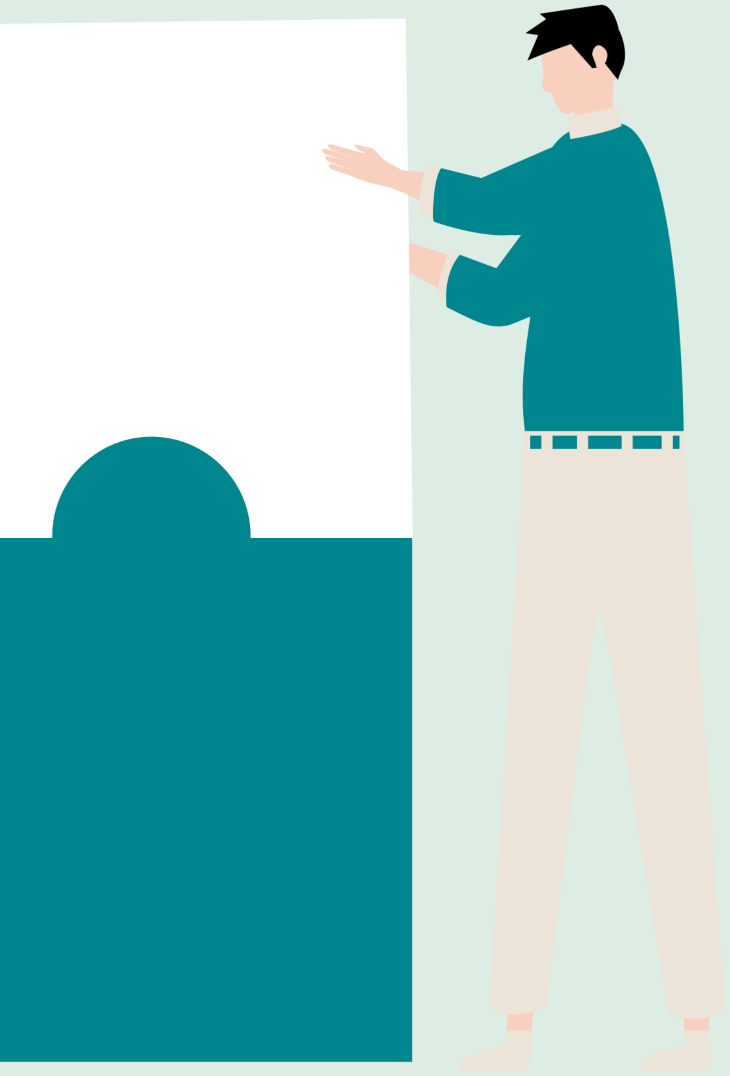
Actual outbreaks: 5
Predicted outbreaks: 10

4

CONCLUSIONS

Conclusion 1

Developed a climate-based model to predict dengue cases within the same cluster of locations in the Philippines



Conclusion 02

Found that the SGDRegressor yielded the best performance of R-squared of **85.10%** and RMSE of **52.07** for predicting the number of dengue cases in Bulacan.

Conclusion 03

Utilized the data obtained from this study and applied existing guidelines to forecast the the possibility of outbreak in locations.

4

RECOMMENDATIONS

Future of the Study

01. Consider more features such population, demographics, access to healthcare, city index.

02. Explore spatial dependencies.

03. Machine learning for dengue outbreak prediction and hotspots detection

Thank you very much!

github link here soon!