MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

@twitterhandle

Artificial Intelligence Approach for Severe Dengue Early Warning System

Dina Nur A.N., Ph.D.

Assistant Professor

Universitas Negeri Semarang, Indonesia

@TheInstituteDH #MEDINFO23

AIDH IMIA

Abstract

OBJECTIVE

we developed an artificial intelligence model with spatiotemporal to predict Dengue outbreak and incidence case which is ready to be implemented into an early warning system application

METHODS

- Dengue cases, climatological and meteorological, spatiotemporal data in Semarang, Indonesia, from 2014-2021: 7208 samples (80% training, 20% testing)
- Machine learning and LSTM

RESULTS

Extra Trees Classifier to predict outbreak: accuracy 90.71 % (AUC = 95.71%)

CatBoost Regressor to predict

incidence case: MAE 0.6402, RMSE

1.1069, MSE 1.2252, R^2 0.5214

Introduction

DENGUE

- Mosquito-borne viral infection (DEN1-4)
- Flu-like illness that occasionally develops into lethal complication (severe Dengue)

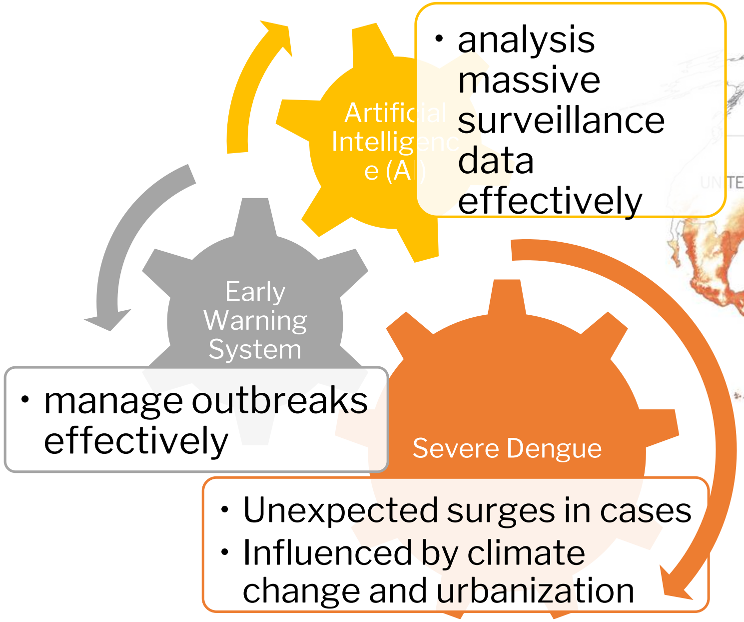
BURDEN

- > 50% world's population is now at risk
- Dengue vector also transmits Chikungunya, Yellow Fever and Zika infection

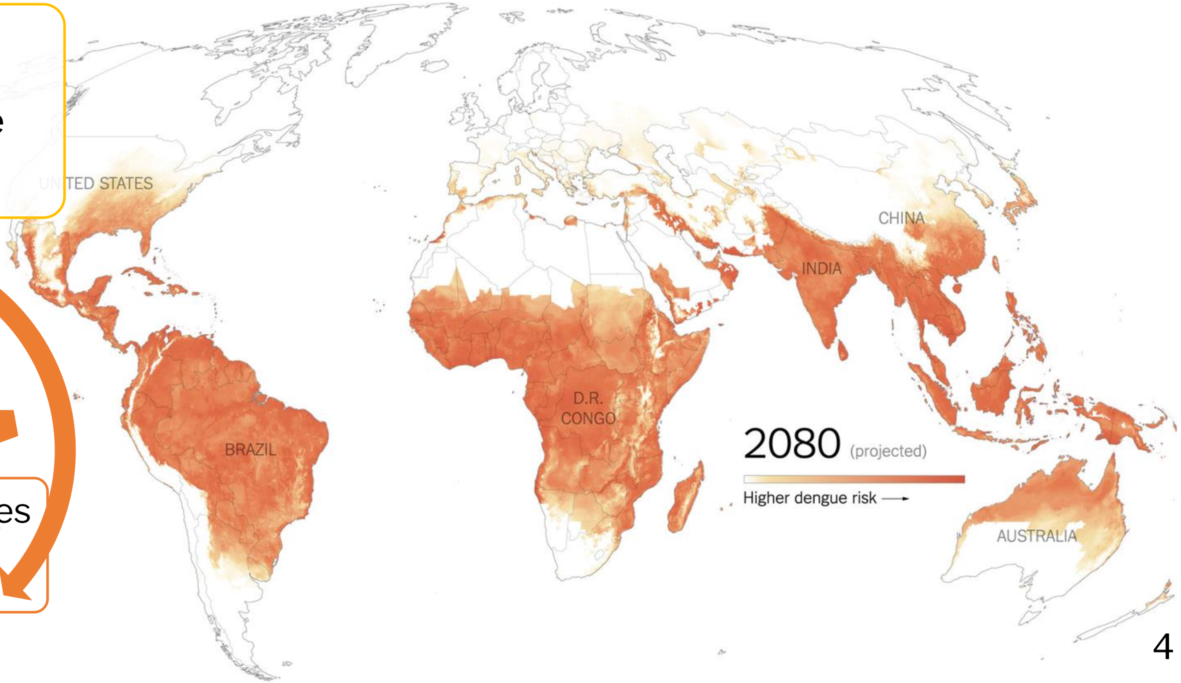
Prevention and Control

- Early detection & access to proper medical care
- Depends on effective vector control measures

Current Situation



Under a moderate warming scenario, 2.25 billion more people could be at risk for dengue fever by 2080.



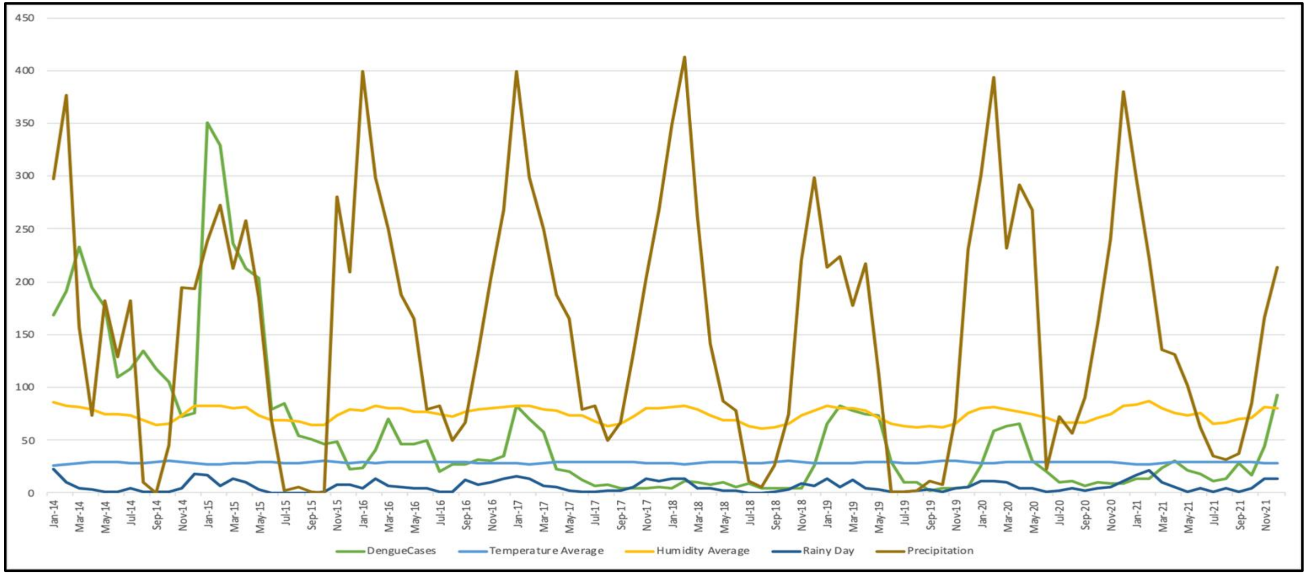
Objective

Using AI model with the spatio-temporal data to predict Dengue outbreak and incidence cases that is feasible to be implemented in early warning system application

Method: Data source

Indonesia: Semarang

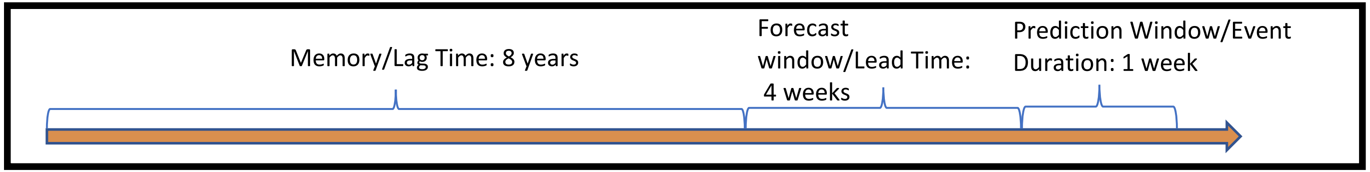
- 1st week (January) 2014 until 53rd weeks (December) 2021:
- Spatiotemporal and Dengue surveillance epidemiology data: Semarang City Regional Health Office (16 district)
- Meteorological and climatological data: Indonesian Meteorological, Climatological, and Geophysical Agency



Methods: Data Preprocessing

Output:

- Dengue outbreak events (outbreak or non outbreak) district/week:
  - there is an increase in the number of dengue patients two or more times in a one week/month time compared to the previous week/month or the same month last year.
- Dengue incidence cases:
  - number of dengue hospitalized confirmed cases (laboratory confirmation) consist of Dengue Hemorrhagic Fever (DHF), and Dengue Shock Syndrome (DSS).



Methods: Data Preprocessing

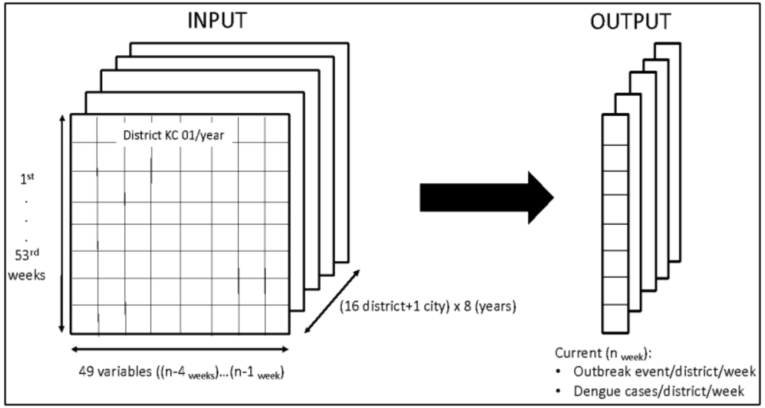
Input (49 variables):

Month (catgorical): 12 variables	Week (numeric): 1 riables: 1-53	Distrct (categorical): 17 variables
Incidence cases/week/district (numeric): 4 variables	Temperature Average/week (numeric): 3 variables	Humidity/week (numeric): 3 variables
Air Pressure/week (numeric): 3 variables	Rainy day/week (numeric): 3 variables	Precipitation/week (numeric): 3 variables

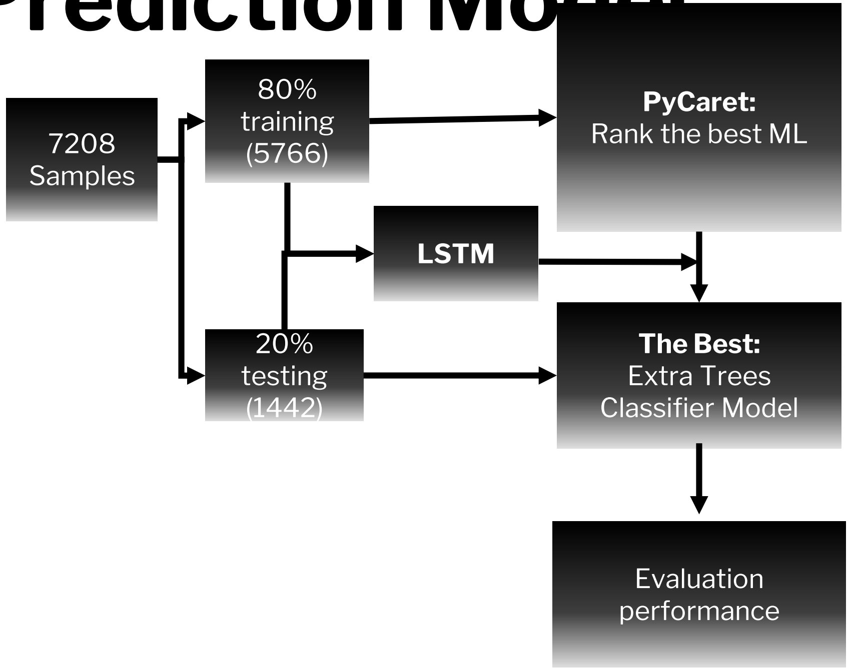
Prediction output:

- Outbreak status per week/district of Dengue from (binary)
- Incidence Dengue cases/week/district (numeric)

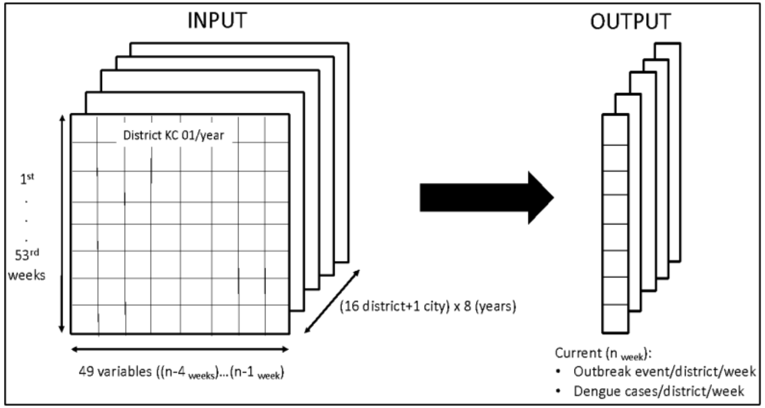
Method: Outbreak Prediction Model



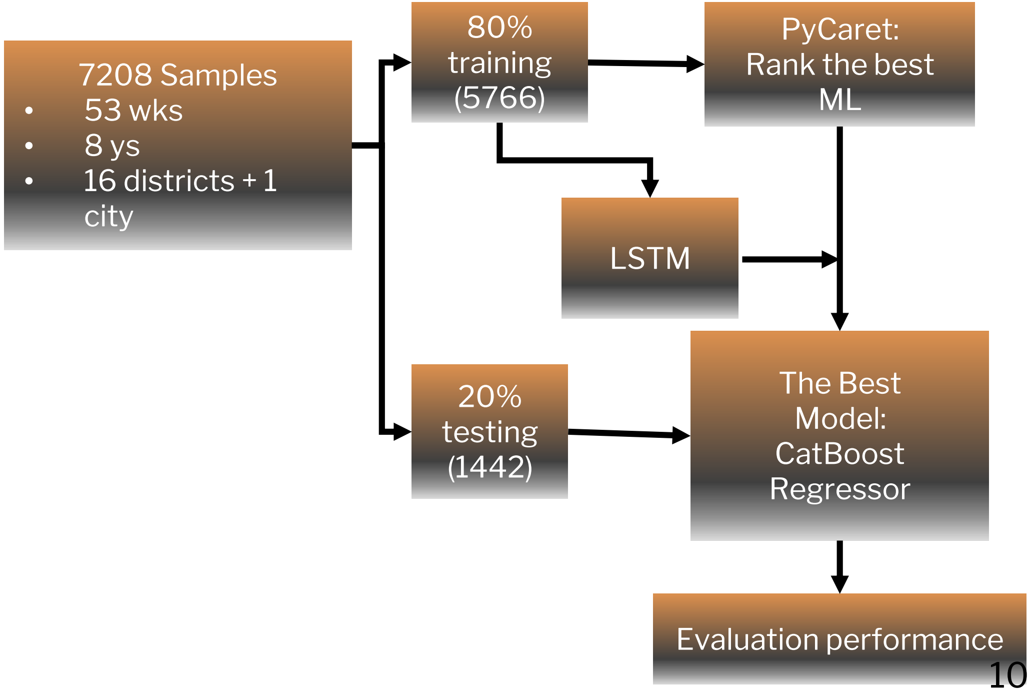
Excel MS Office & SAS



Method: Cases Prediction Model



Excel MS Office & SAS



Method: Evaluation

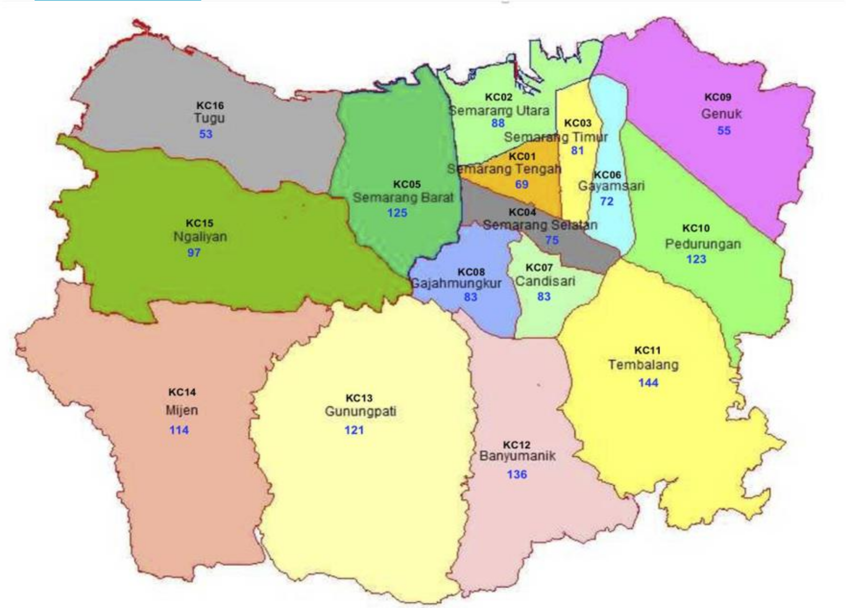
Dengue outbreak:

- Recall and precision
- Receiving operating characteristic (ROC) curve
- F1-score (weighted average of the precision and sensitivity)

Dengue incidence cases:

- MAE
- RMSE
- R^2

Results: Outbreak Distribution 2014-2021



7208 samples:

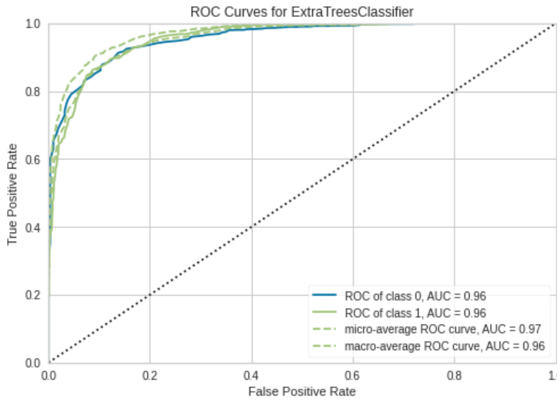
- 5599 (77.68%) non outbreak
- 1609 (22.32%) outbreak

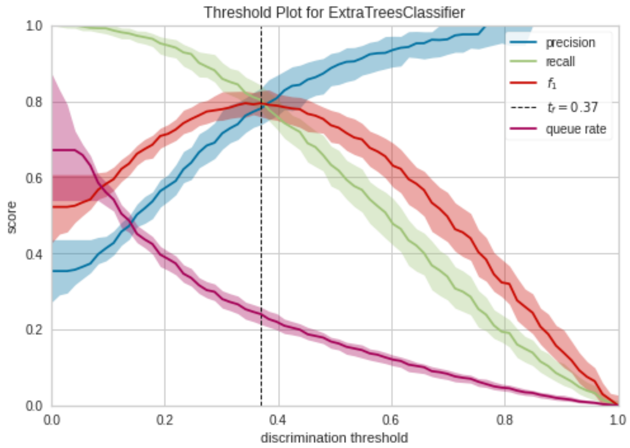
Results: Outbreak models performance

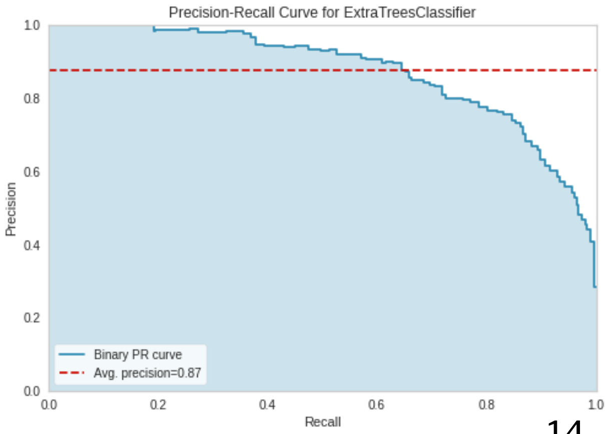
Model	Accuracy	AUC	Recall	Precision	F1 score
Extra Trees Classifier	**0.8925**	**0.9529**	**0.6117**	**0.8880**	**0.7238**
CatBoost Classifier	0.8561	0.8997	0.5049	0.7994	0.6174
Extreme Gradient Boosting	0.8514	0.8929	0.5493	0.7406	0.6299
Light Gradient Boosting Machine	0.8453	0.8793	0.5034	0.7439	0.5990
LSTM (Training)	0.7836				
(Testing)	0.6400	0.6500	0.5700	0.3200	0.4100

Results: Outbreak prediction performance

Model	Sample	Accuracy	AUC	Recall	Precision	F1 score
Extra Trees Classifier	Training					
	Testing	0.9071	0.9571	0.6442	0.8973	0.7500

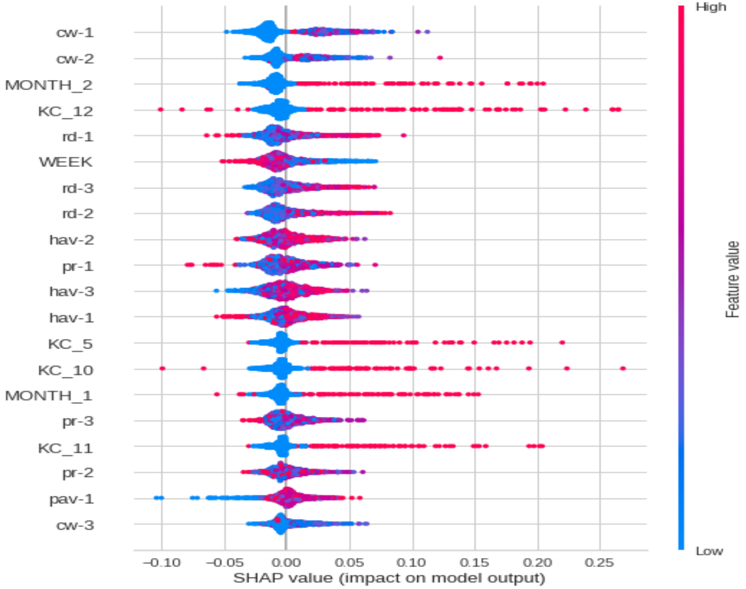






Resuls: Outbreak model SHAP values

1 week previous cases
2 weeks previous cases
February (month)
KC 12 (district)
1 week previous rainy days
Weeks in the year
3 weeks previous rainy days
2 weeks previous rainy days
2 weeks previous humidity average
1 week previous precipitation
3 weeks Previous humidity average
1 week previous humidity average
KC 05 (district)
KC 10 (district)
January (month)
3 weeks previous precipitation
KC 11 (district)
2 weeks previous precipitation
1 week previous atmospheric pressure average
3 weeks previous cases

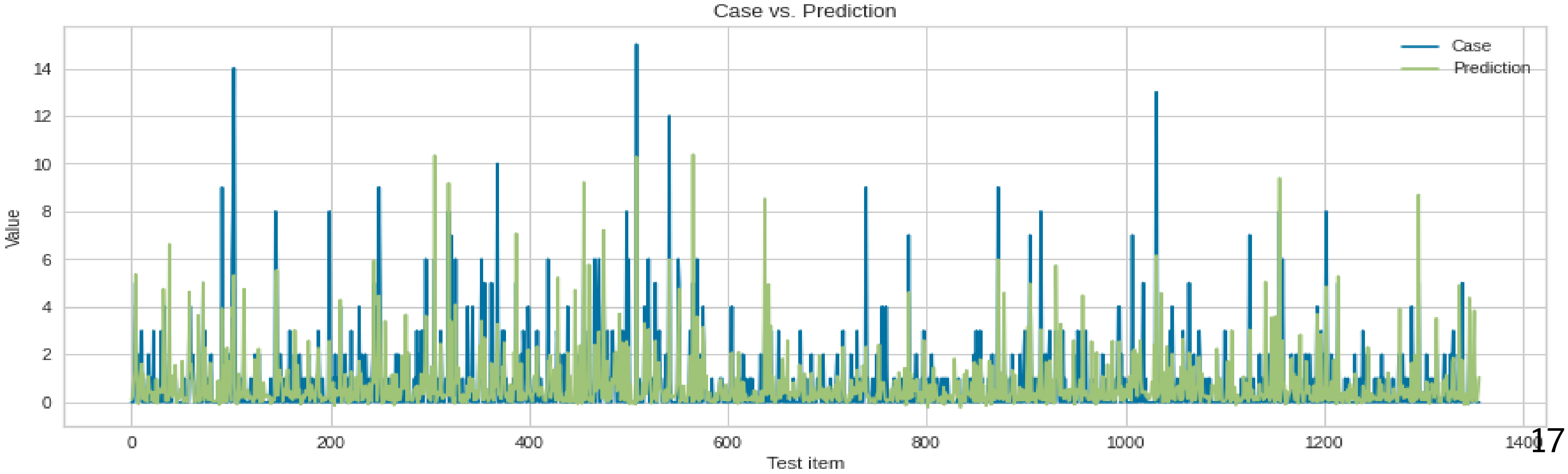


Results: Case models performance

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE
CatBoost Regressor	0.6304	1.1997	1.0891	0.5621	0.4053	0.5996
Gradient Boosting Regressor	0.6271	1.1935	1.0871	0.5609	0.4031	0.5798
Orthogonal Matching Pursuit	0.6410	1.2291	1.1044	0.5472	0.4209	0.5747
Huber Regressor	0.6202	1.2581	1.1158	0.5410	0.4062	0.5836

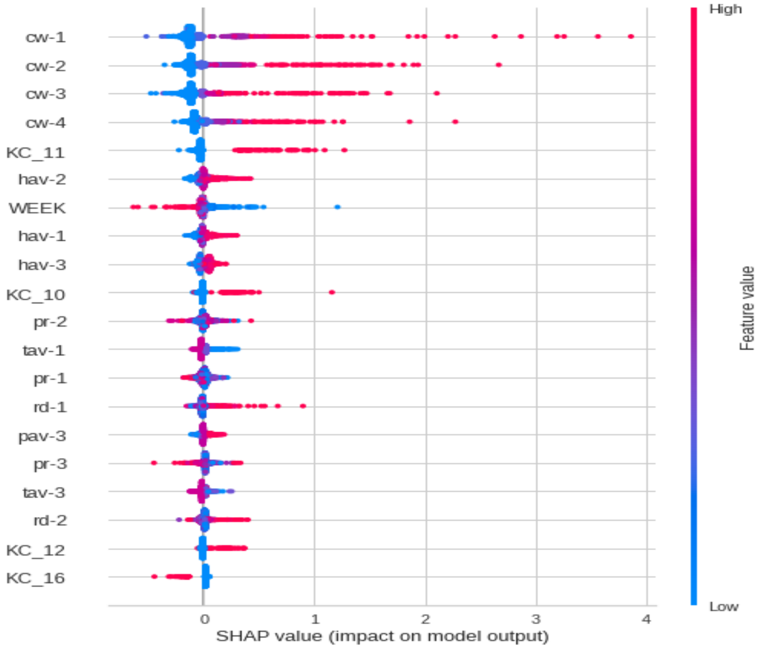
Results: Case prediction performance

Model	Sample	MAE	MSE	RMSE	R2	RMSLE	MAPE
CatBoost Regressor	Training	0.6304	1.1997	1.0891	0.5621	0.4053	0.5996
	Testing	0.6402	1.2252	1.1069	0.5214	0.4024	0.6174



Results: Case model SHAP values

1 week previous cases
2 weeks previous cases
3 weeks previous cases
4 weeks previous cases
KC 11 (district)
2 weeks previous humidity average
Weeks in the year
1 week previous humidity average
3 weeks Previous humidity average
KC 10 (district)
2 weeks previous precipitation
1 week temperature average
1 week previous precipitation
1 week previous rainy day
3 weeks previous atmospheric pressure average
3 weeks previous precipitation
3 weeks temperature average
2 weeks previous rainy days
KC 12 (district)
KC 16 (district)



Discussion

Outbreak

(2021) Malaysia (district, weekly):

SVM Linear:

- Accuracy 70.12%,
- Recall 14.40%
- Precision 56.25%

Incidence Cases

(2017) Semarang City (January 2010-April 2015):

- classical SIR Model

(2021) Semarang City (January 2010-April 2015):

- Autoregressive Distributed Lag (ARDL)

Discussion

Variables influencing the prediction

Extreme weather related
- Rainy days
- Humidity
- Precipitation
- atmospheric

Spatial data

features affecting outbreak prediction
Districts with frequently outbreak numbers (KC 11, KC 12, KC 10, and KC 05)

Temporal Data

Important features for prediction
Granularity data per week, sequence of weeks in a year, and some months which has higher rainy days or precipitation

Deploying into Semarang Dengue EWARS

Comprehensive variables
- Meteorological
- Climatological
- extreme weather
- Epidemiology
- spatio-temporal

Updated over time

Detail geographical unit (district)

Regional speciific model (high-risk area)

Limitation and future research

Deep learning should be used in future studies

The ML approach does not allow transfer learning
which is necessary for updating the model in Dengue EWARS based on new data

- variables that reflect each region should be used,
- data from wearable devices/IoT smart homes should be collected to provide detailed climatological or meteorological data for each area

No district climatological and meteorological stations in Semarang
The granularity of climatological and meteorological data is confined to the city region

Conclusion

Using AI model with the spatio-temporal data :

- Extra Trees Classifier could predict Dengue outbreak and outperform prev. study (ACC 90.71 %, AUC = 95.71%)
- CatBoost Regressor could predict incidence cases with MAE 0.6402, RMSE 1.1069, MSE 1.2252, R^2 0.5214
- AI model with spatiotemporal data feasible to implement in Semarang City Dengue early warning system application