THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

Climate-Based Forecasting of Dengue Dynamics

Abdullah Al Helal, Haridas Kumar Das, Chun-hao Chen, Feng Zhu

A Deep Learning Bootcamp project
Erdős Institute's May-Summer 2024 Cohort

Overview

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

Question: How can we use climate, epidemiological, and other data to forecast the dynamics of infectious diseases, e.g. dengue?

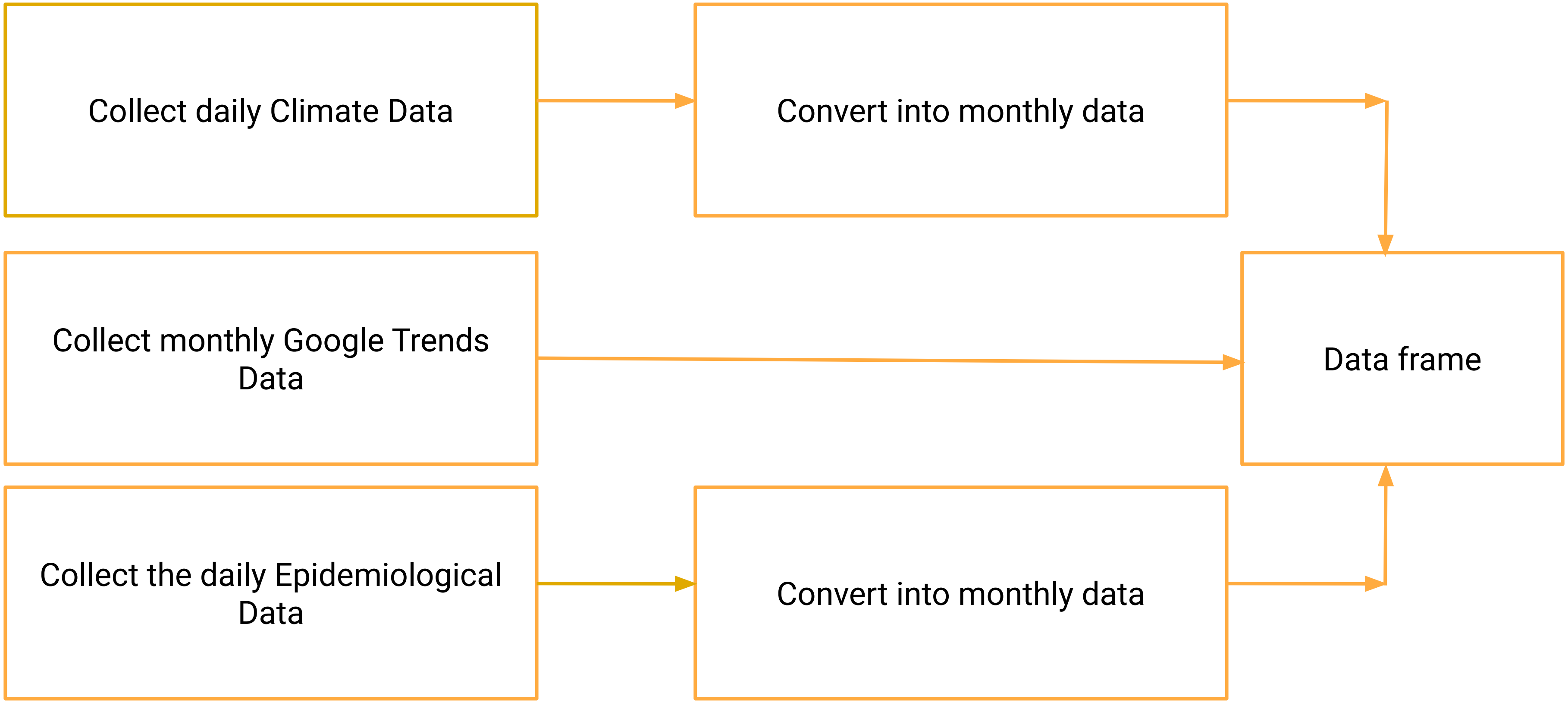
Goal: We use deep learning algorithms to analyze climate, epidemiological, and social data in order to forecast dengue dynamics, focusing on the case of Bangladesh.

Use: Policymakers can use our model to analyze or predict dengue dynamics, in order to better allocate resources and implement more effective public health measures.

Data: Climate Data , Epidemiological Data and Google Trends Data

Data preprocessing

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry



Exploratory data analysis

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 192 entries, 0 to 191
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   date            192 non-null    object 
 1   temp_2m_mean    192 non-null    float64
 2   appar_temp_mean 192 non-null    float64
 3   rain_sum        192 non-null    float64
 4   goog_trends     192 non-null    int64  
 5   case            192 non-null    int64  
dtypes: float64(3), int64(2), object(1)
memory usage: 9.1+ KB
```

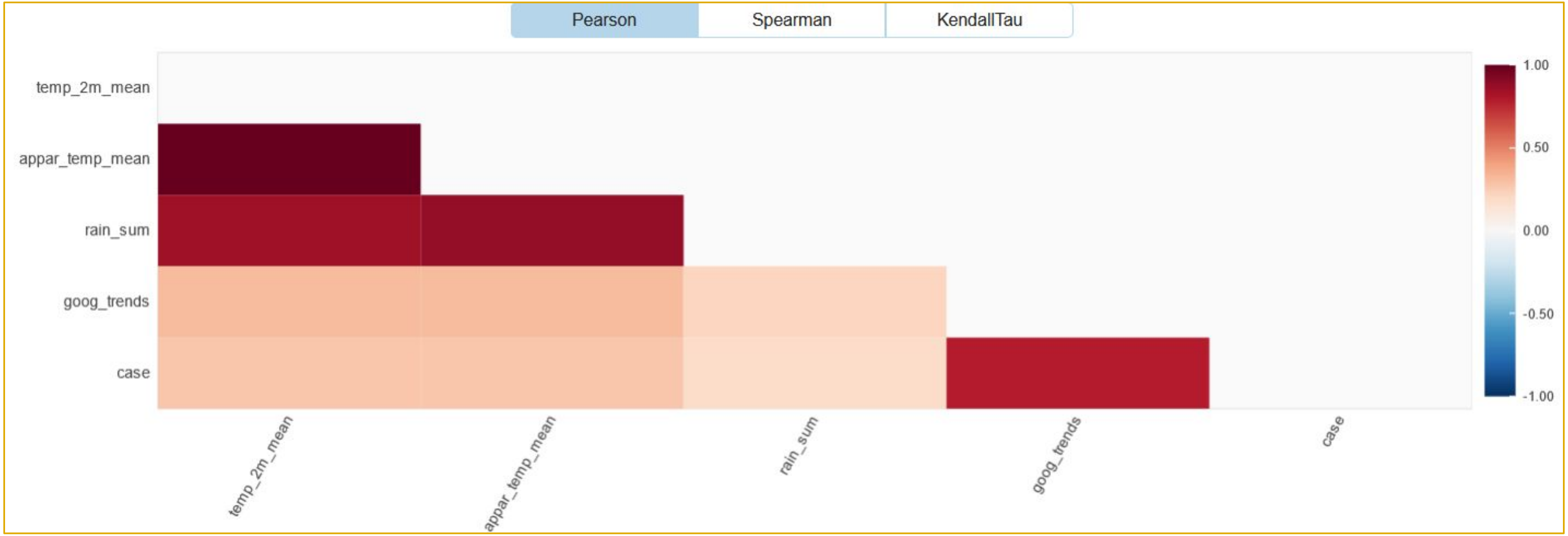
Exploratory data analysis


```
[ ] df.corr()
```

	date	temp_2m_mean	appar_temp_mean	rain_sum	goog_trends	case
date	1.000000	0.007473	0.031027	0.031935	0.377799	0.356004
temp_2m_mean	0.007473	1.000000	0.992262	0.845130	0.315251	0.278008
appar_temp_mean	0.031027	0.992262	1.000000	0.880395	0.316291	0.275526
rain_sum	0.031935	0.845130	0.880395	1.000000	0.216992	0.180182
goog_trends	0.377799	0.315251	0.316291	0.216992	1.000000	0.783459
case	0.356004	0.278008	0.275526	0.180182	0.783459	1.000000

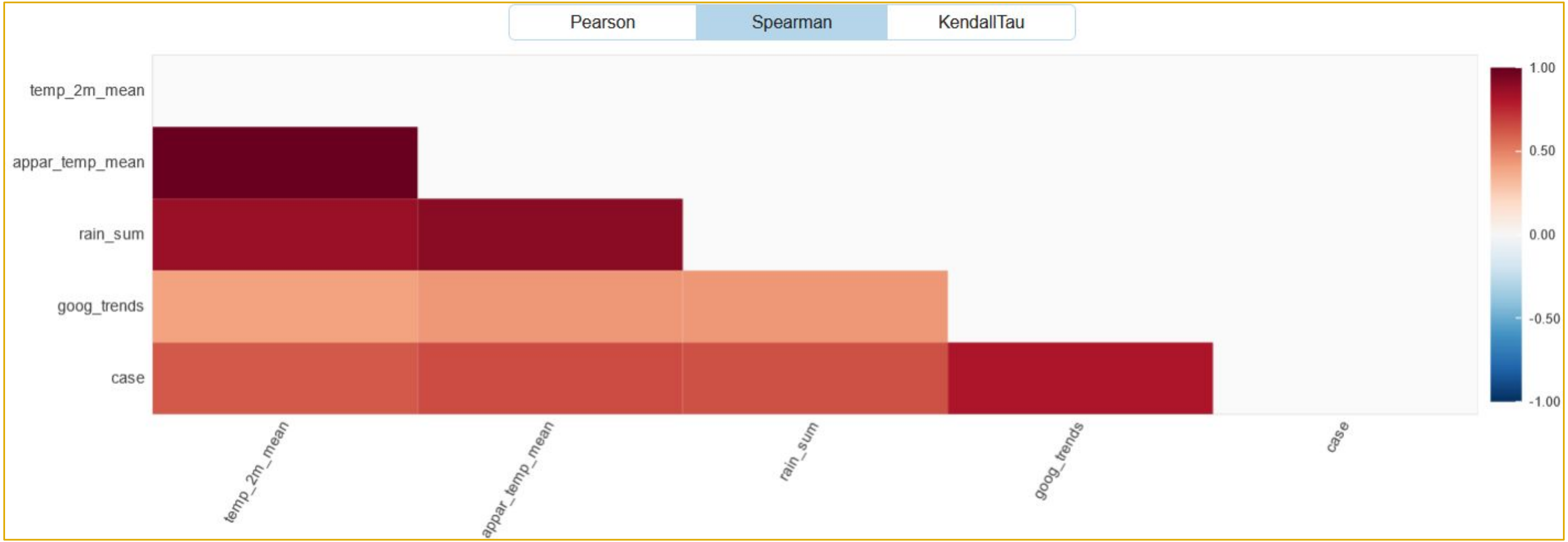
Exploratory data analysis

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry



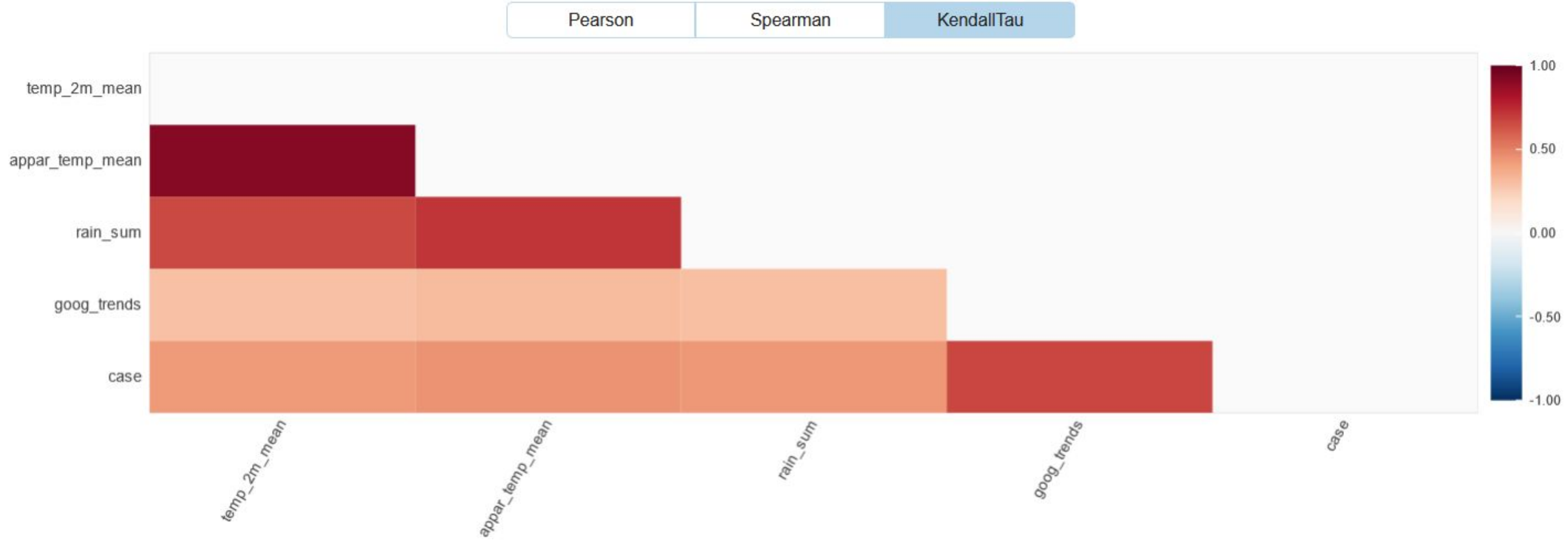
Exploratory data analysis

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

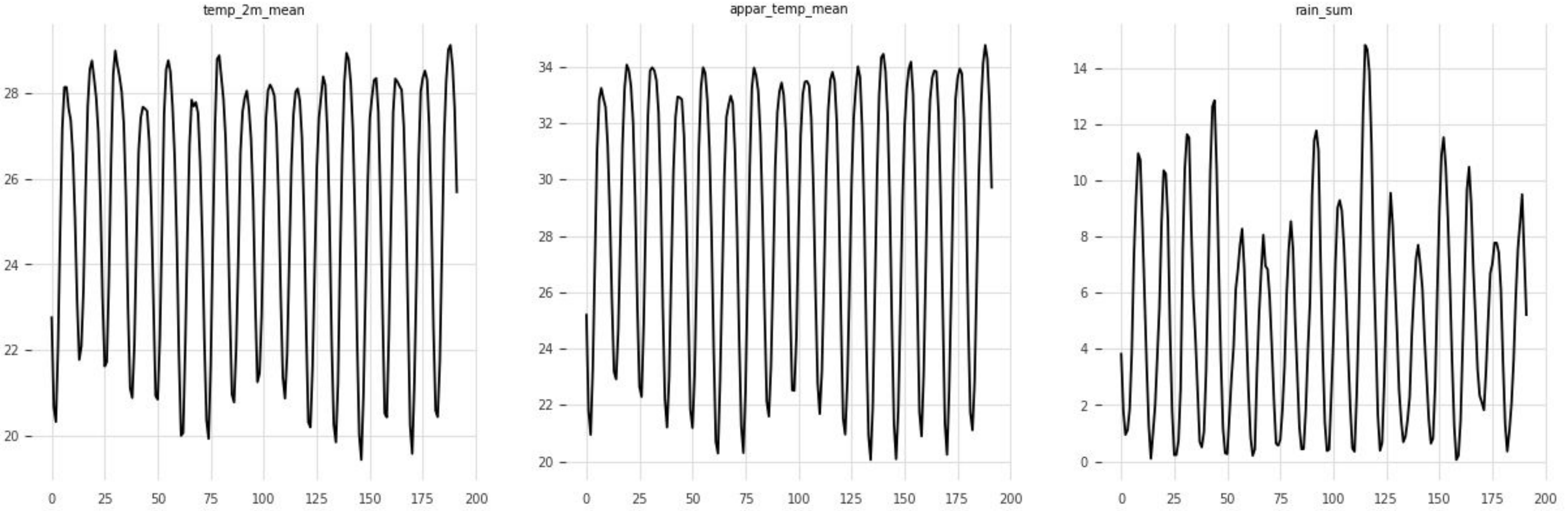


Exploratory data analysis

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry



Visualizing the data

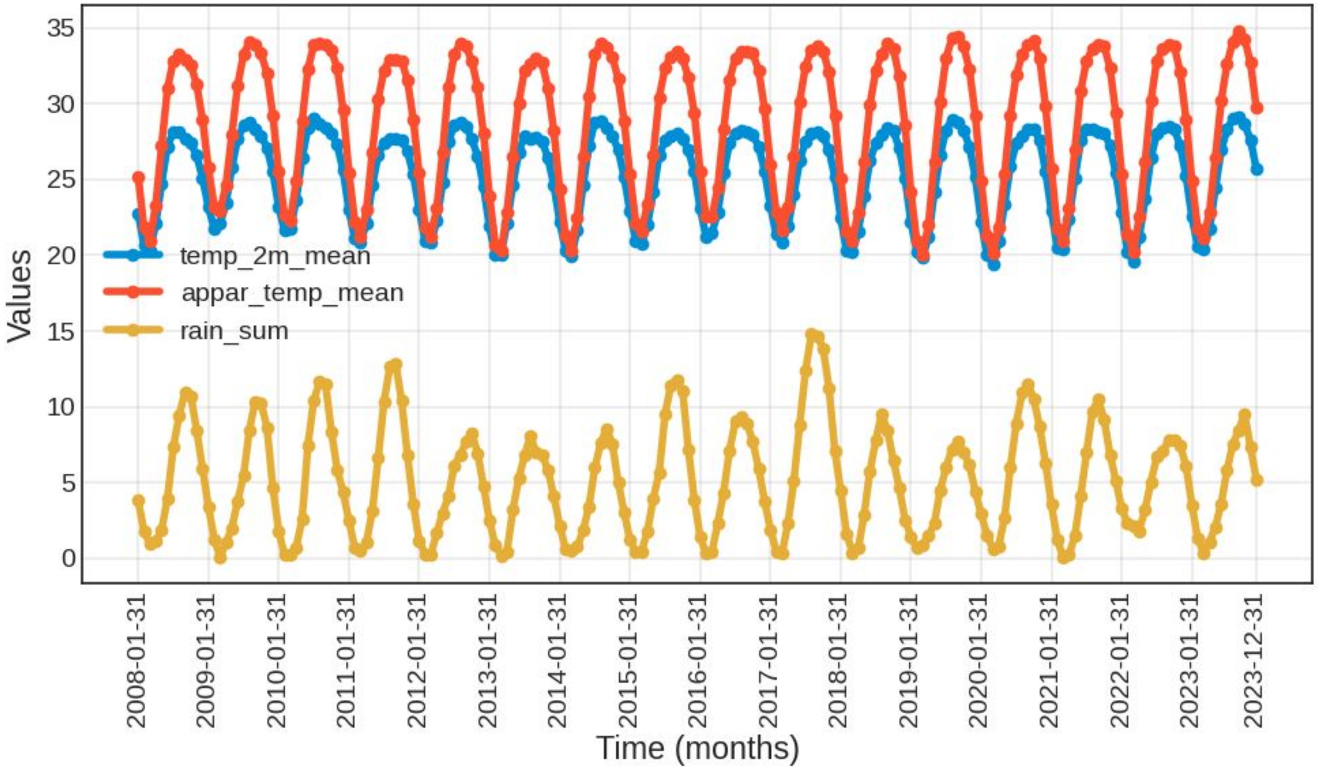


Visualizing the data

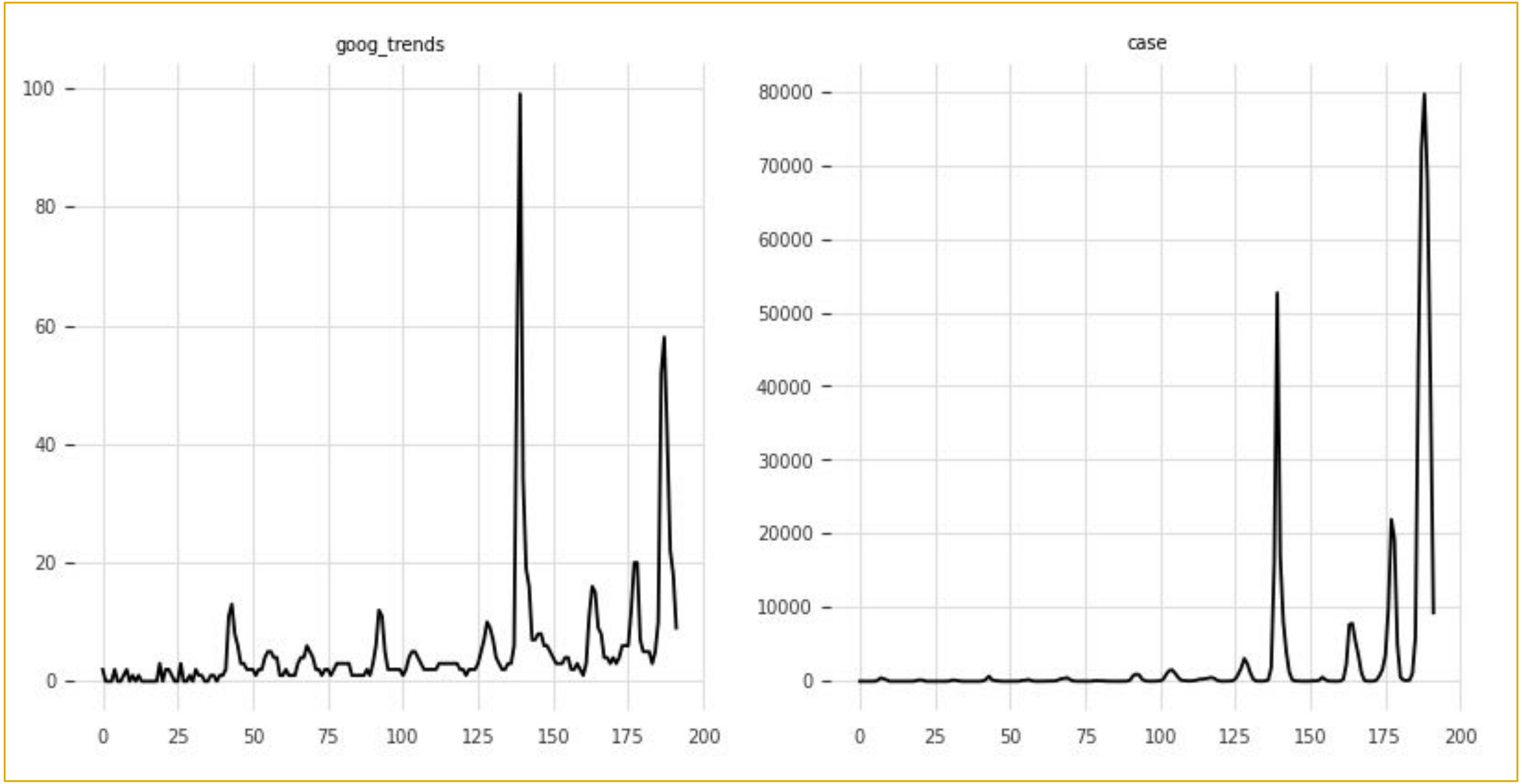
Temperature (2m)

Relative Humidity (%)

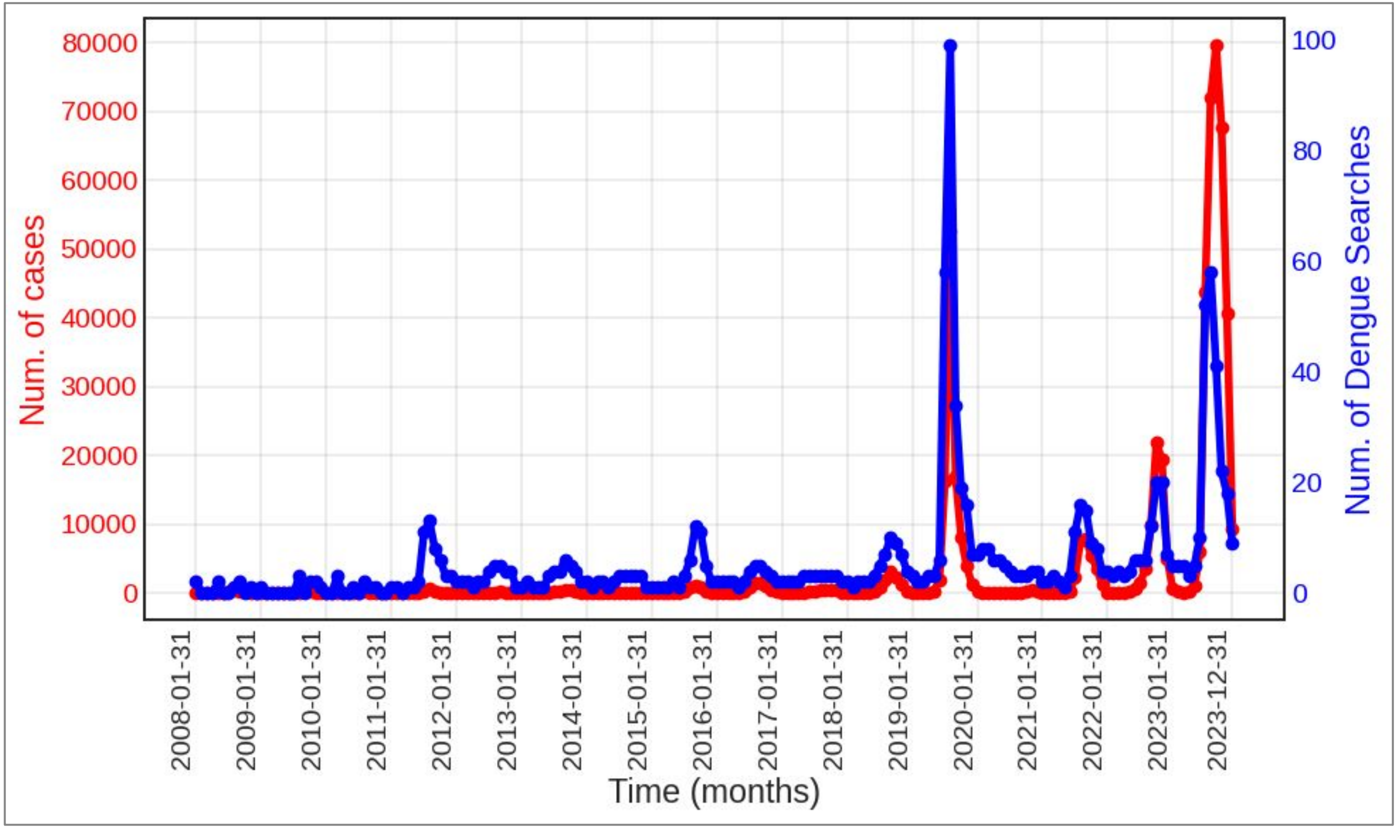
dew point 2m (degre C)



Visualizing the data



Visualizing the data



Some aspects of model training

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

- Data was scaled before training for better performance
- For deep learning algorithms, hyperparameter optimization was done using a grid search or a Bayesian method (using the Optuna library)

Model building

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

1. CNN
2. LSTM
3. GRU
4. Random Forest
5. XGBoost

Statistical Model: ARIMA

Model building

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

1. CNN
2. LSTM
3. GRU
4. Random Forest
5. XGBoost

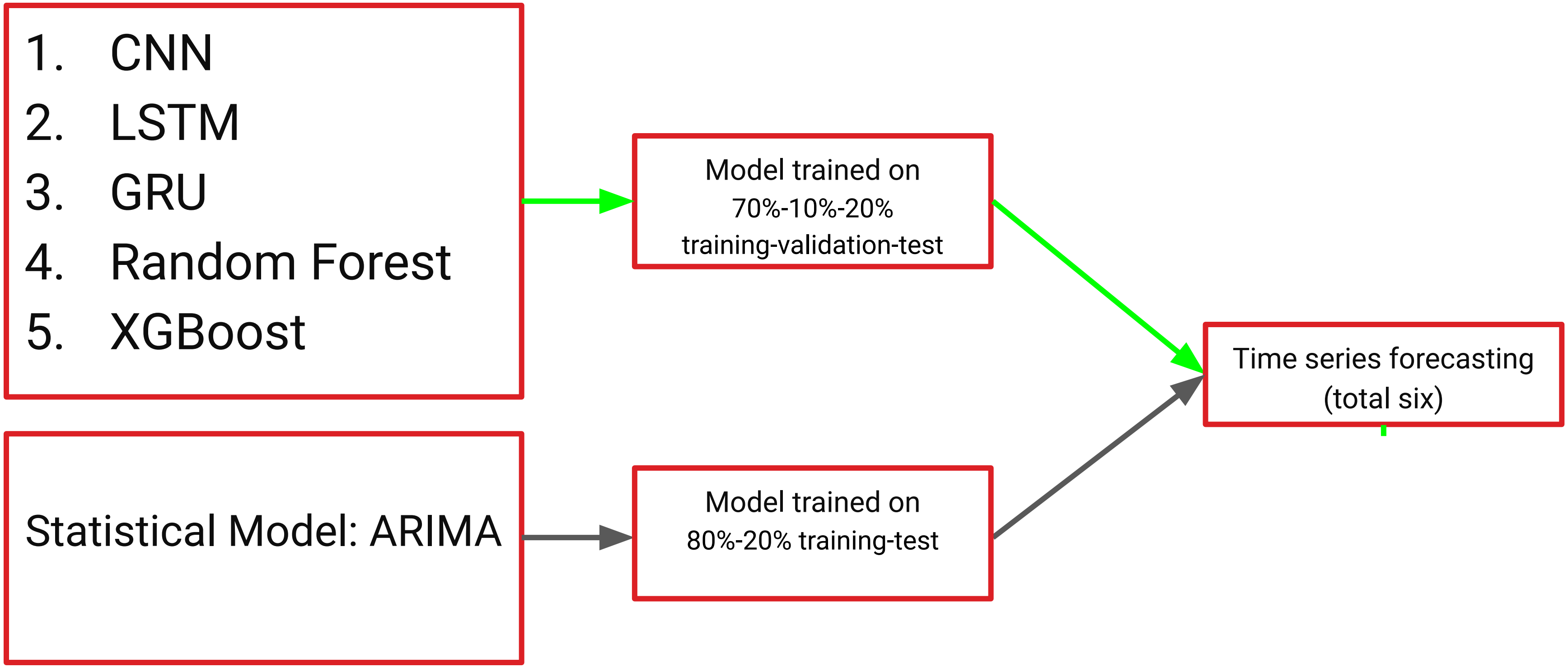
→ Model trained on 70%-10%-20% training-validation-test

Statistical Model: ARIMA

→ Model trained on 80%-20% training-test

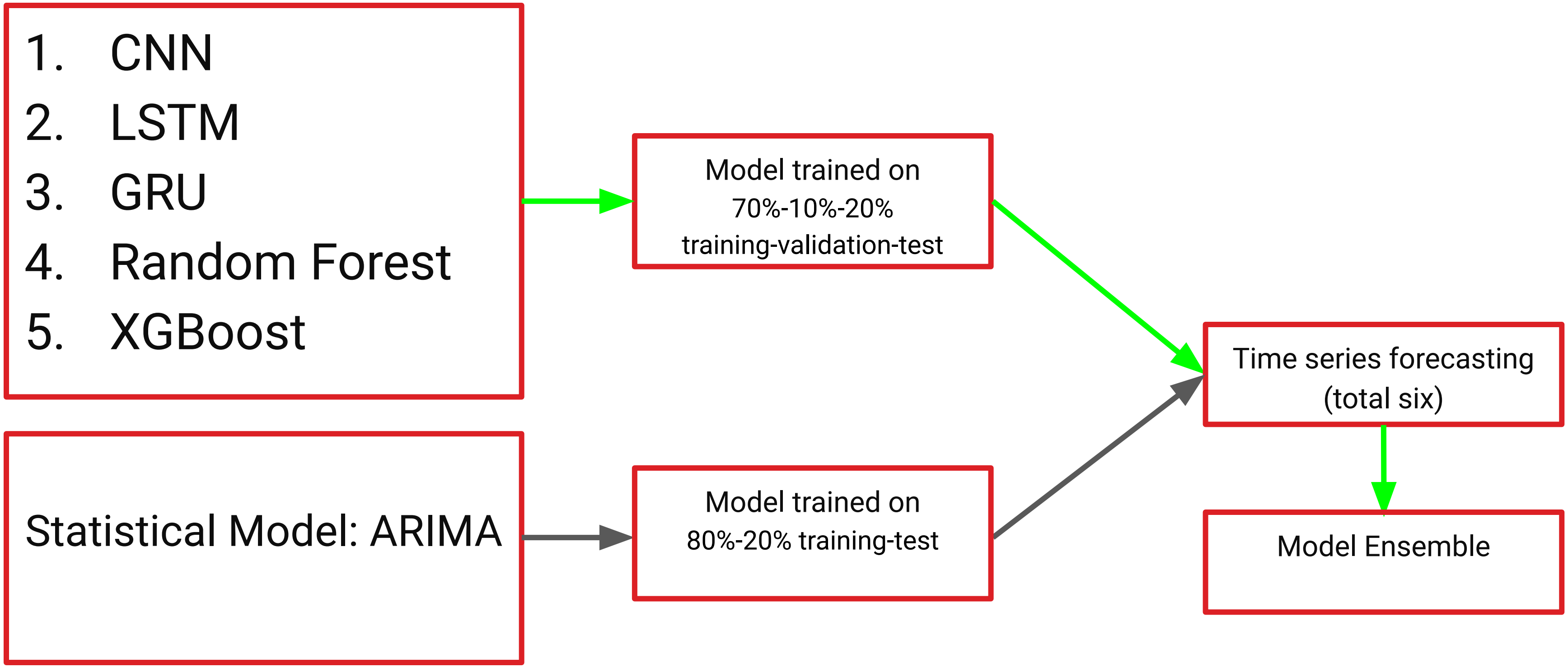
Model building

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

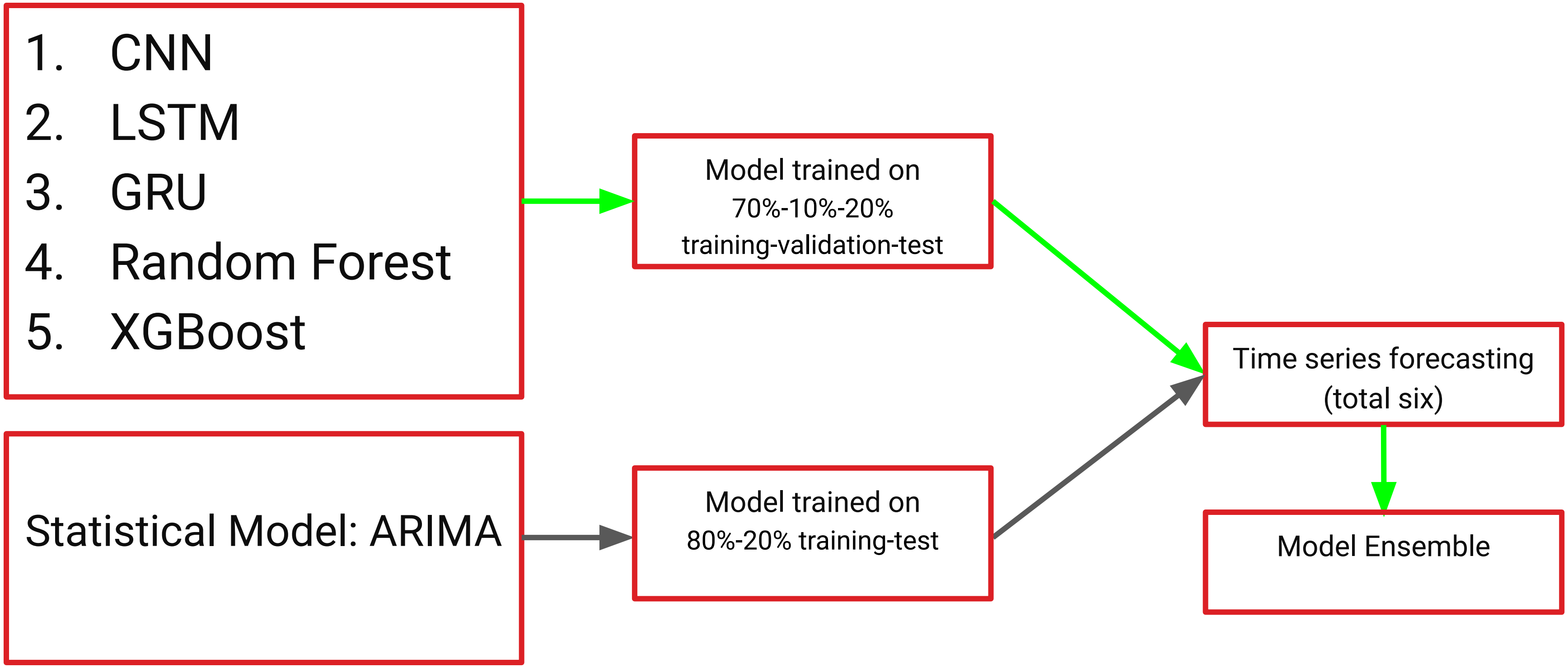


Model building

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry



Model building



Ensemble model

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

Model Predictions Using Ensemble Approach

Consider that we have RMSE scores (R_i) from M different modes. We define the penalty for each model as: $\hat{w}_i = \frac{1}{R_i}$.

We normalize these penalties to obtain the final weights w_i, which are calculated as $w_i = \frac{\hat{w}_i}{\sum_{i=1}^{M} \hat{w}_i}$.

Note that the sum

$$\sum_{i=1}^{M} w_i = \sum_{i=1}^{M} \frac{\hat{w}_i}{\sum_{j=1}^{M} \hat{w}_j} = \frac{\sum_{i=1}^{M} \hat{w}_i}{\sum_{j=1}^{M} \hat{w}_j} = \frac{\sum_{i=1}^{M} \hat{w}_i}{\sum_{i=1}^{M} \hat{w}_i} = 1$$

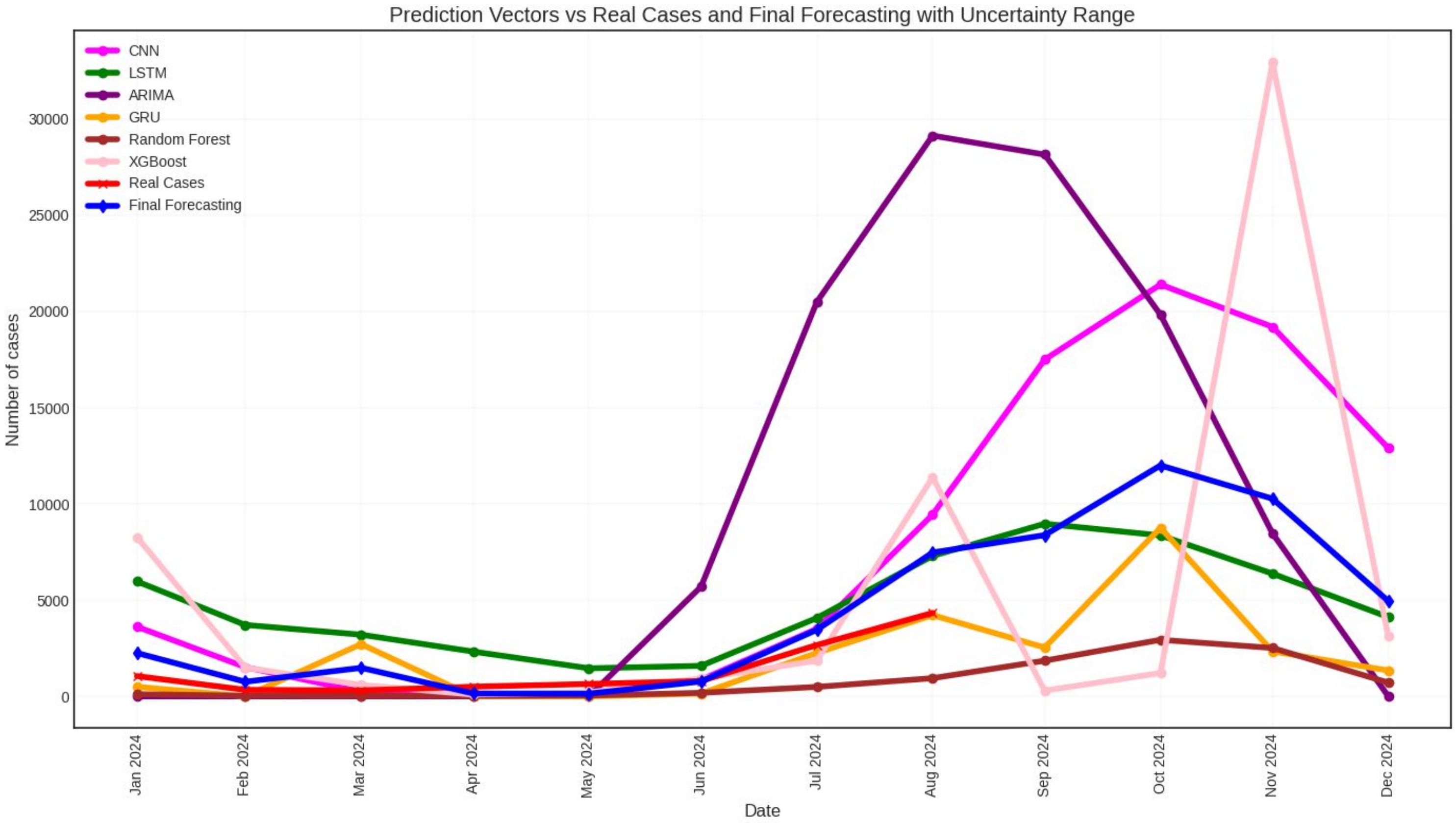
\subsection*{2. Compute Ensemble Predictions}

To compute the final ensemble prediction:

Let y_i represent the prediction vector from the i-th model. The final ensemble prediction vector y_{ensemble} is calculated as:
$y_{\text{ensemble}} = \sum_{i=1}^{M} w_i \mathbf{y}_i$.

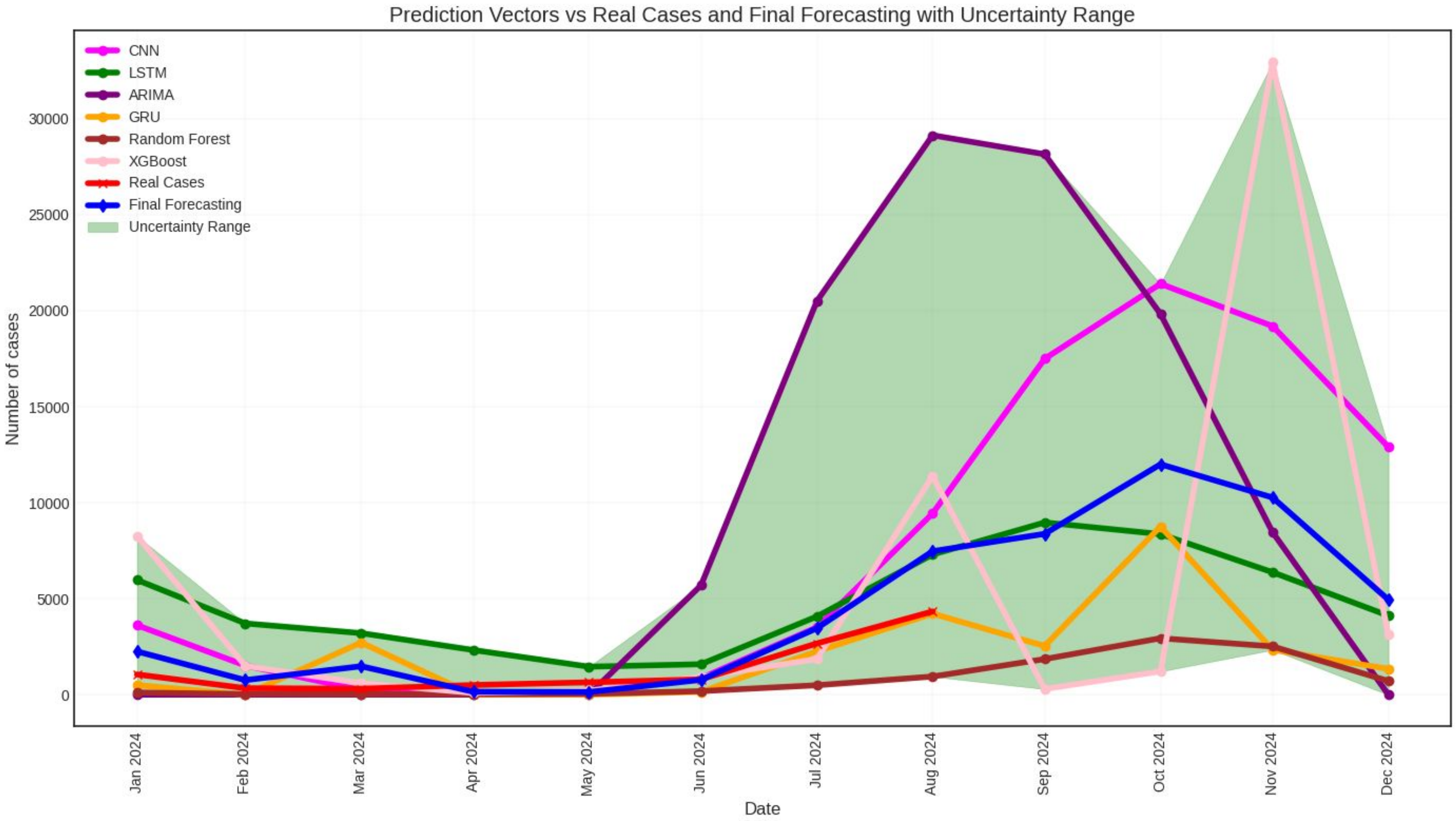
Ensemble model: Forecasting

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

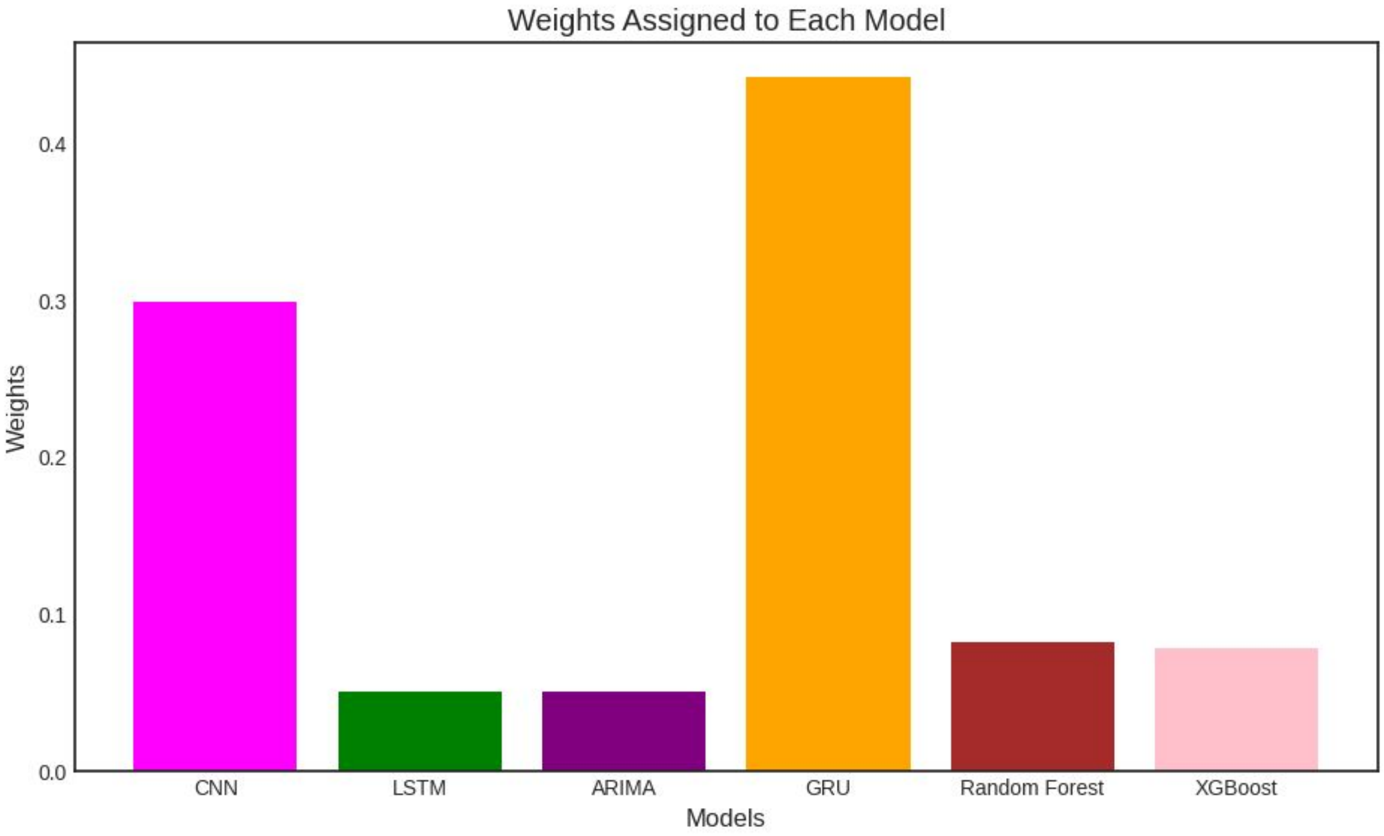


Ensemble model: Forecasting

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry



Model performance



Conclusion

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

- Six deep learning and statistical models were deployed to forecast dengue dynamics in Bangladesh.
- Difficult to forecast the magnitude of the peak, but more reasonable to forecast when the peak will be and general trends
- Two models—GRU and CNN—reasonably forecast dengue cases for 2024, in terms of both test error and visual trend of forecasting in the unknown future.

Future directions

THE ERDŐS INSTITUTE
Revolutionary Collaborations in Academia and Industry

- It remains to be explored if separating seasonal and trend components in the data improves model performance.
- Forecasting can be improved by using a bigger dataset.
- Future work will involve implementing hybrid deep learning models and sophisticated probabilistic time series forecasting algorithms.

Thank you for joining!

ahelal@okstate.edu
Department of Mathematics
Oklahoma State University

haridas.das@okstate.edu
Department of Mathematics
Oklahoma State University

chen4co@mail.uc.edu
MS in Information Systems
University of Cincinnati

fzhu52@wisc.edu
Department of Mathematics
University of Wisconsin–Madison