


This project aims to predict the number of deaths and the Case Fatality Rate (CFR) due to Ebola. The dataset was first cleaned and preprocessed, followed by model building for predictions. The Random Forest Single-Task Model was selected for its superior performance in predicting these metrics.

## DataSet Used

-  **Training Data:** [Google Sheets - Training Data](#)
-  **Test Data:** [Google Sheets - Test Data](#)

## Directory Structure

```

CB-project/
|
├── Codes/                                # 📁 Contains all code files
|   ├── images/                          # 🖼️ Contains image files used in the
project
|   ├── models/                          # 🤖 Contains model-related scripts
for training and evaluation
|   └── data/                             # 📊 Contains all data-related folders
|       ├── raw/                         # 📄 Raw data files (e.g., original
dataset)
|       ├── interim/                     # 🔄 Interim processed data before
final cleaning
|       └── final/                       # ✅ Final cleaned and preprocessed
data ready for modeling

```

## Project Description

This project focuses on predicting the number of deaths and Case Fatality Rate (CFR) due to Ebola. The dataset used was first cleaned to ensure accuracy and consistency. Subsequently, a series of models were trained and evaluated to determine the best approach for these predictions.

## Model Selection and Evaluation

After training and evaluating various models such as AdaBoost, Decision Tree, and others, we finalized the Random Forest Single-Task Model based on its superior performance. The table below summarizes the evaluation metrics for all models, including MAE, MSE, and  $R^2$  scores:

Model	Deaths_MAE	Deaths_MSE	Deaths_R <sup>2</sup>	CFR_MAE	CFR_MSE	CFR_R <sup>2</sup>
Linear Regression	43.26389	2625.766	0.004933	1.006661	123.8043	0.00341
Random Forest	13.02522	284.476	0.892194	0.338237	26.52387	0.78649
SVR	41.12771	2504.951	0.050717	0.729752	124.6278	0.00322
Gradient Boosting	33.4908	1697.821	0.356589	0.442241	0.47705	0.99616

Model	Deaths_MAE	Deaths_MSE	Deaths_R <sup>2</sup>	CFR_MAE	CFR_MSE	CFR_R <sup>2</sup>
Decision Tree	0	0	1	0	0	1
K-Nearest Neighbors	28.56804	1348.983	0.488786	0.734198	100.1167	0.19409
AdaBoost	39.70689	2198.235	0.166951	0.675545	0.88858	0.99284

## Working of the Random Forest Single-Task Model

The Random Forest Single-Task Model is designed to predict two key metrics: Deaths and Case Fatality Rate (CFR). It consists of two separate models:

1. **Deaths Prediction Model** – Trained to predict the number of deaths due to Ebola in a given region.
2. **CFR Prediction Model** – Trained to predict the Case Fatality Rate (CFR), which is the proportion of confirmed deaths among confirmed cases of Ebola.

Both models are trained and evaluated based on their performance metrics, with R<sup>2</sup> (coefficient of determination) and MAE (Mean Absolute Error) being the key evaluation criteria. For each model, we selected the best-performing model based on the R<sup>2</sup> score, as it provides the best explanation of variance in the predictions.

## Model Selection Process

- **Deaths Prediction Model:** Multiple models were trained, and the one with the highest R<sup>2</sup> score was selected for maximum accuracy in predicting the number of deaths.
- **CFR Prediction Model:** Similarly, the model with the best R<sup>2</sup> score was selected to predict the case fatality rate with the highest reliability.

## Final Prediction Process

For new predictions, we use the model with the best R<sup>2</sup> score for both metrics (Deaths and CFR) to ensure accurate results. By applying these two separate models, we calculate the final Confirmed Deaths by combining the outputs from the Death Prediction Model and the CFR Prediction Model, following the given formulation. This method leverages the strengths of Random Forest in handling complex, non-linear relationships between features (latitude, longitude, etc.) and predicted outcomes (Deaths and CFR). It ensures precise and reliable predictions.