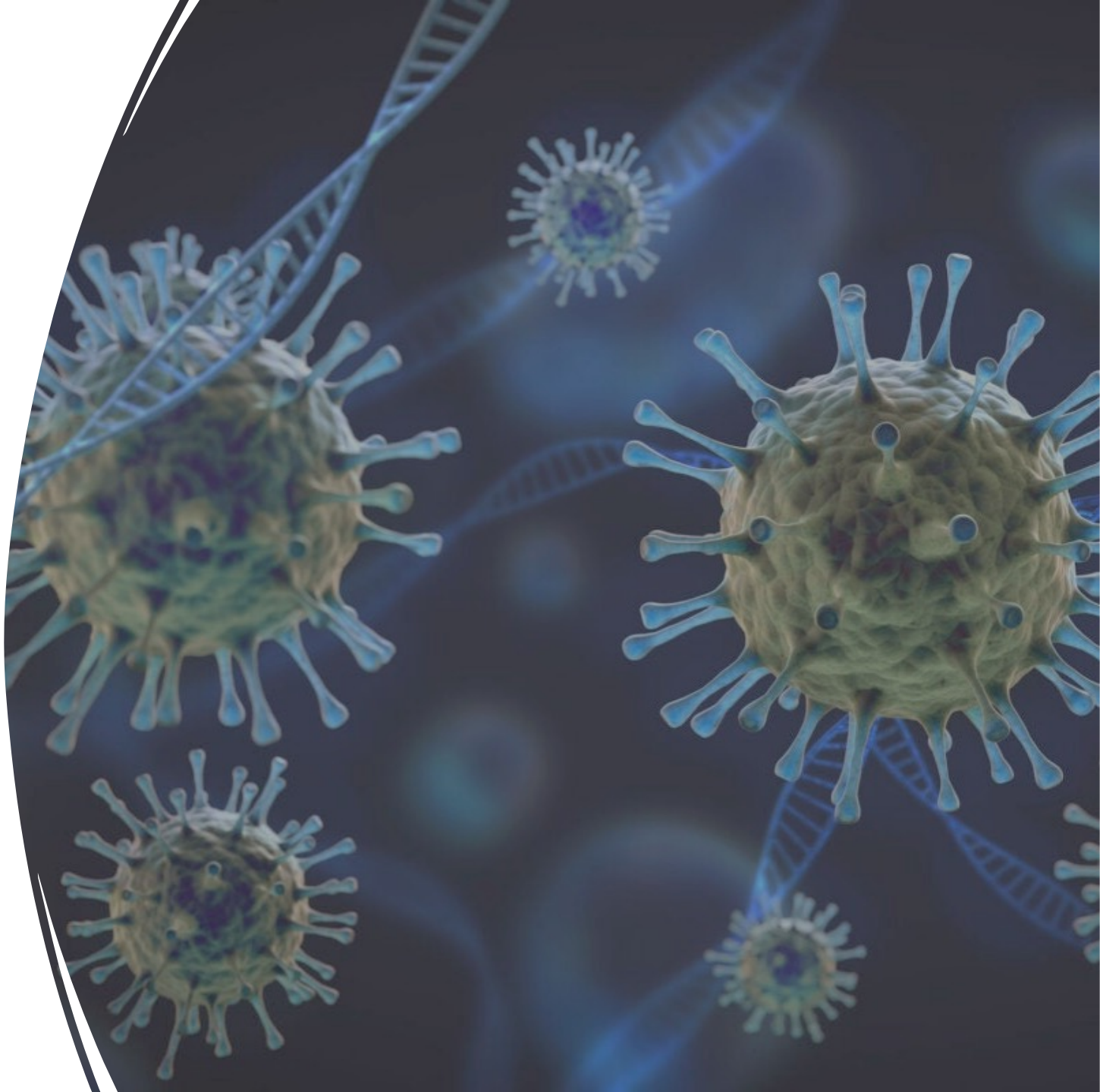
Improve COVID-19 Outbreak Forecasting using Machine Learning Algorithms

Prepared by: Meaad Alrehaili

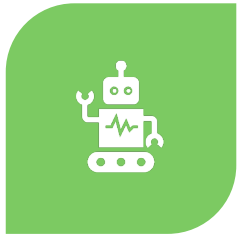


Outline:

INTRODUCTION

DESIGN

DATA DESCRIPTION



ALGORITHMS

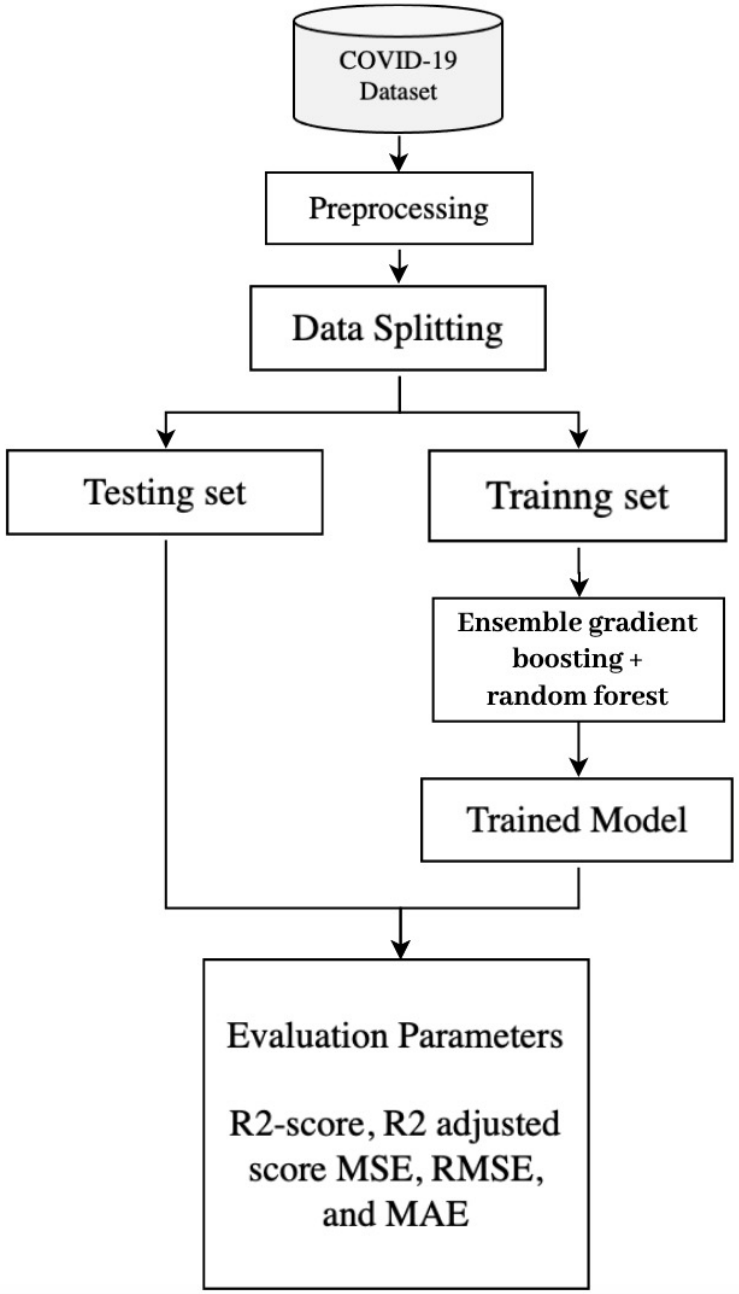
RESULTS

CONCLUSION

Introduction

- Forecasting a pandemic's movements is extremely important for governmental or organizational actions such as those that need to be taken by the governments and organizations.
- Problem: Forecasting a new confirmed cases, recovered cases, and deaths cases of the upcoming 10 days using the Covid-19 dataset.

Design


COVID-19
Dataset
Preprocessing
Data Splitting
Testing set
Trainng set
Ensemble gradient
boosting +
random forest
Trained Model
Evaluation Parameters
R2-score, R2 adjusted
score MSE, RMSE,
and MAE

Data Description

- The COVID-19 dataset is obtinaed from Johns Hopkins University (JHU).
- The data sourced from governments, national agencies across the world.
- The data features include the state, region, date for , number of confirmed, deaths, and number of recovered cases. Total features are 49 and the total samples around 7000+.

Ensemble Algorthims

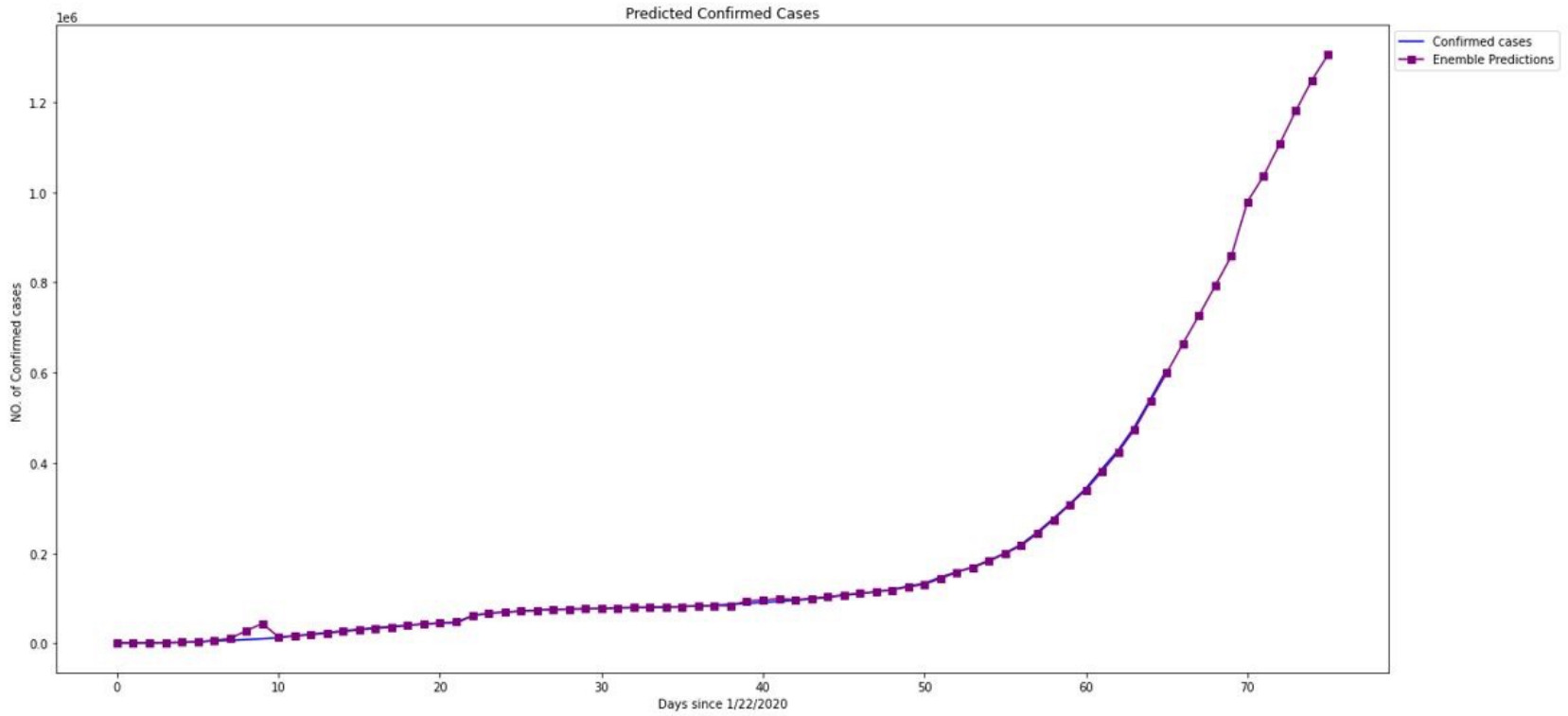
- Ensemble of gradient boosting and random forest were used in this project to minimize the burden of data collection and to improve its efficiency.

Results

GB-RF	R^2-score	Adjusted R^2	MAE	MSE	RMSE
Confirmed cases	0.99	0.99	69138158.83	4092.88	8314.94
Recovered cases	0.99	0.99	1682739.6	711.0	1297.2
Death cases	0.99	0.99	200297.36	215.2	447.55

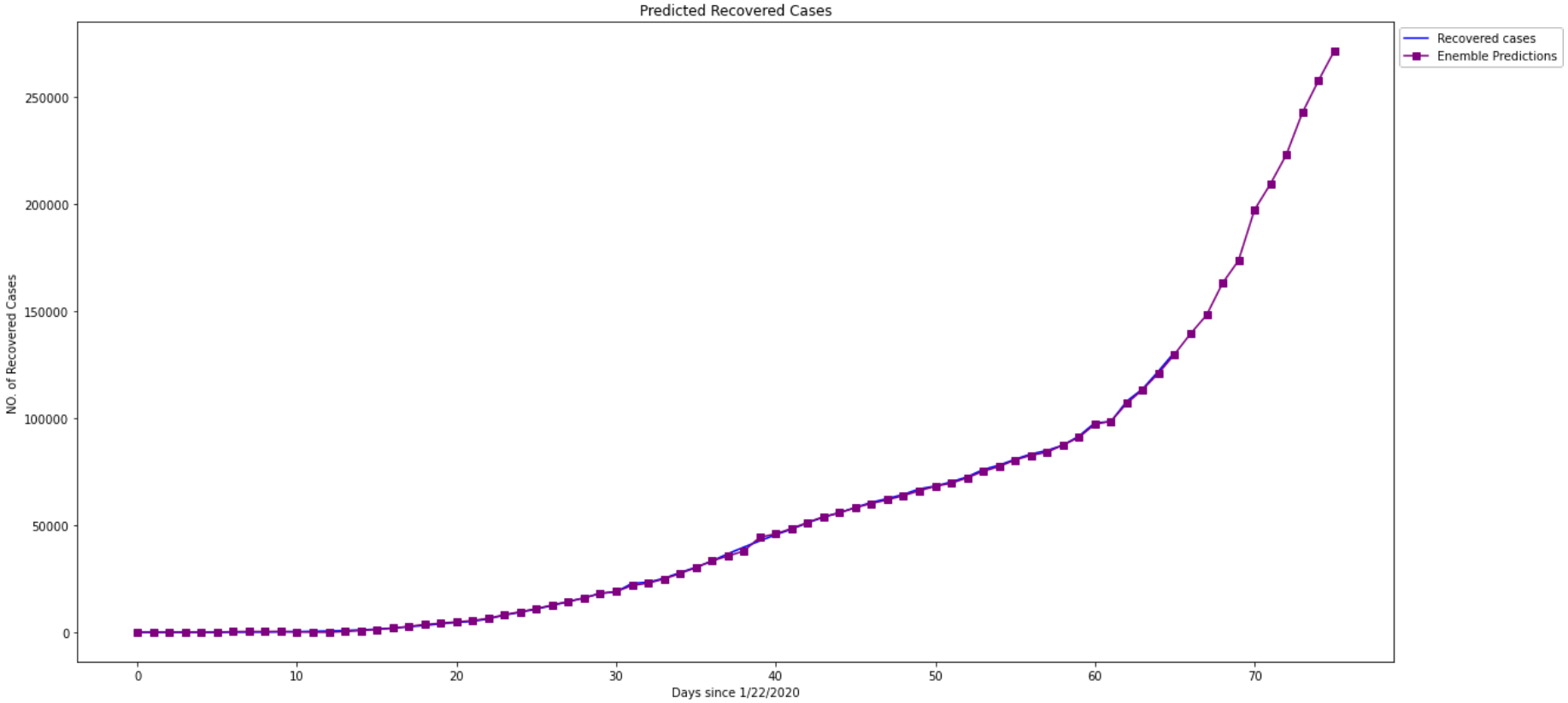
Results

- New infected cases for the upcoming 10 days using Gradient Boosting and Random Forest



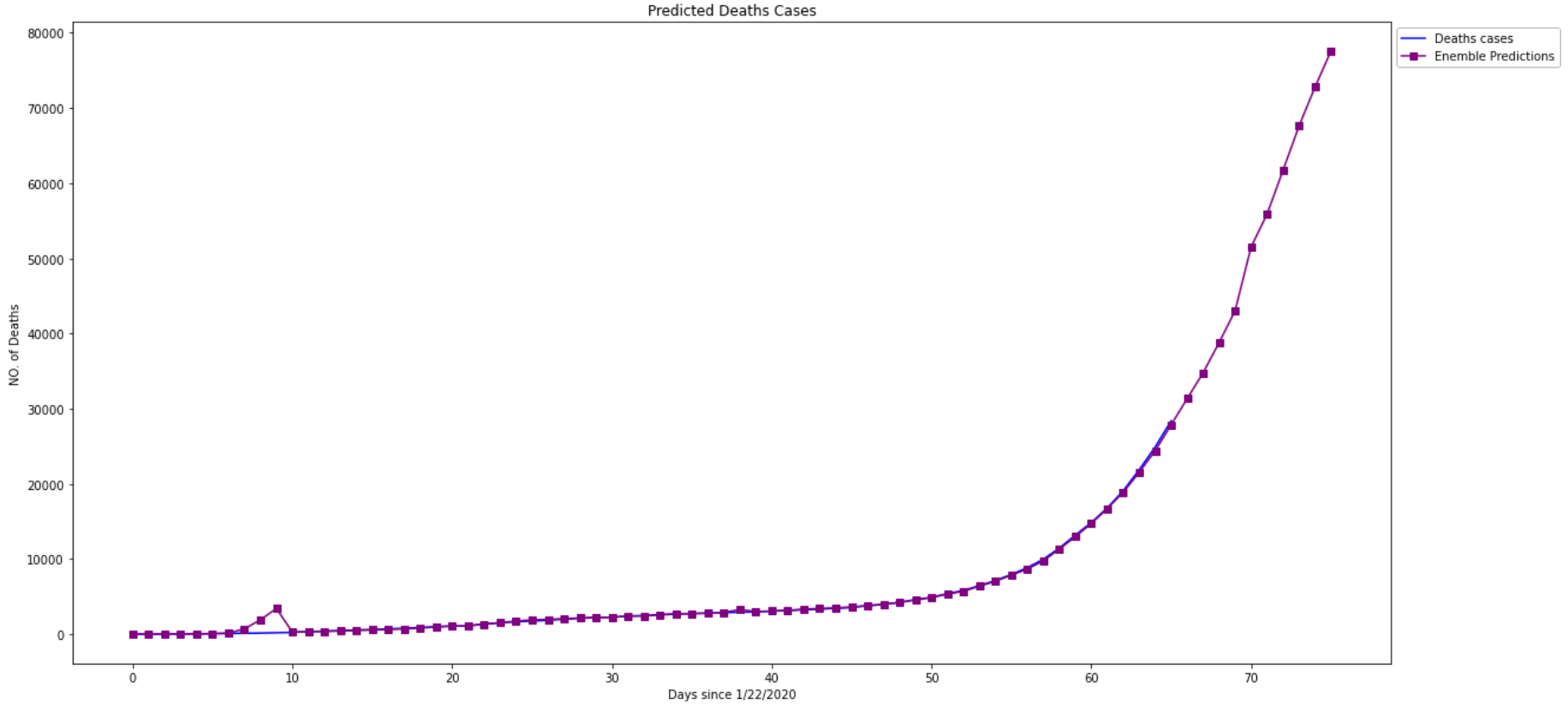
Results

- New recovered cases for the upcoming 10 days using Gradient Boosting and Random Forest



Results

- New death cases for the upcoming 10 days using Gradient Boosting and Random Forest



Conclusion

- ML-based prediction models have been proposed to predict the risks of COVID-19 outbreak.
- The results showed that the best accuracy was obtained by the ensemble of Gradient Boosting and Random Forest (GBRF) in forecasting newly infected cases, recovery cases, and death cases.