CoronaCaster:
Predicting New COVID-19 Cases

STAT 390 Group 2
Cindy Ha, Willie Xie, Erica Zhang

TABLE OF CONTENTS

01

Overview

What is our project about?

Through this project, we will:

- Utilize a regularly updated and comprehensive COVID-19 dataset
- Employ a regression analysis approach to create models that can accurately predict new cases
- Apply the implications of the forecasts to guide individuals and public health systems

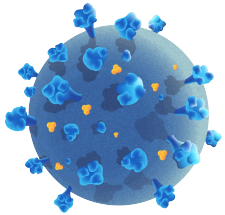
02

Objectives

What do we hope to achieve?

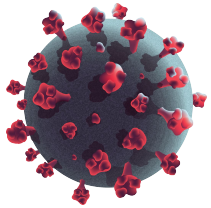
GOAL

- Forecast the number of confirmed daily new cases of COVID-19 in the US
- Measure the degree of spread of COVID-19



APPLICATION

- Advise healthcare systems on potential spikes and when to take preventative measures



03

Data

Data & Initial Preprocessing

The initial dataset contains:

- 67 variables: 1 response = new_cases 66 possible predictors
- 341,408 observations

Some issues addressed:

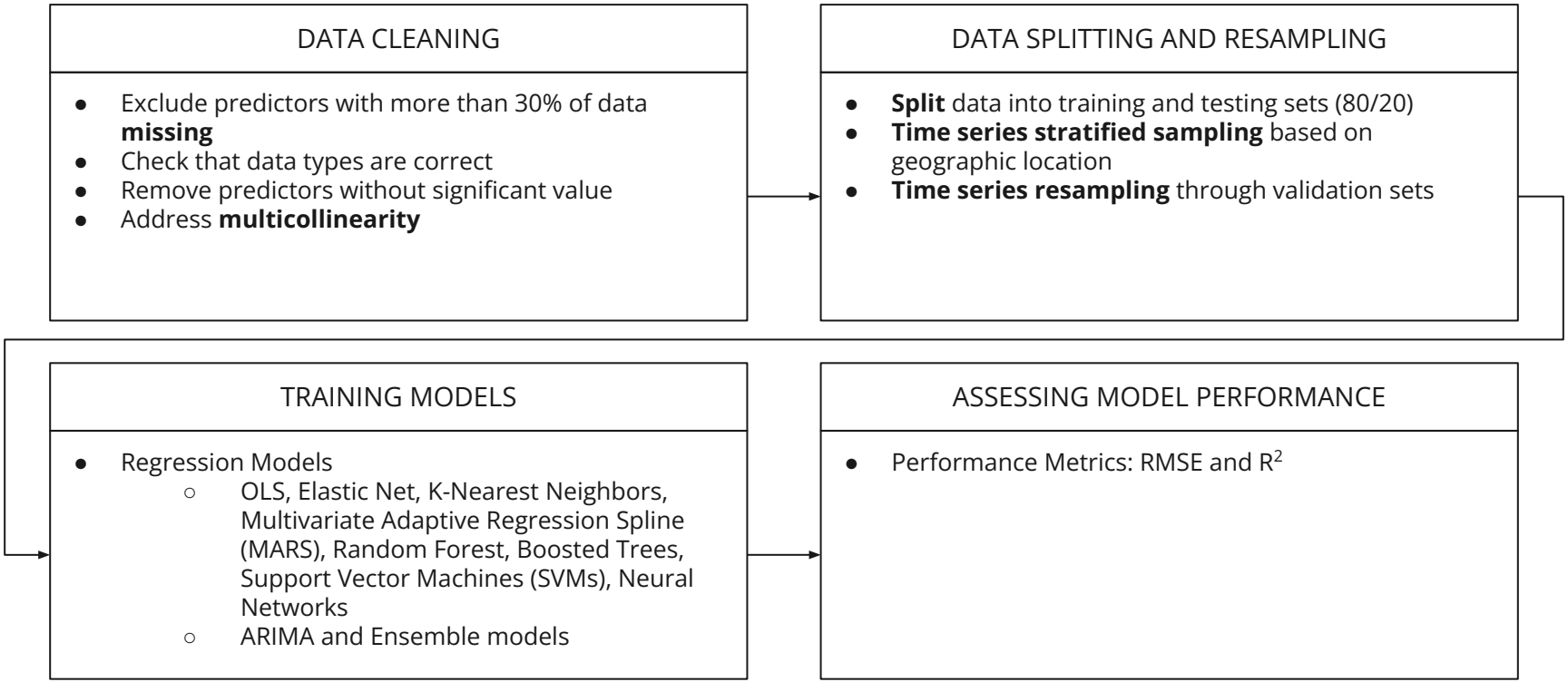
- Missingness
  - Around 41-ish predictors have at least 30% missingness
    Note some variables are not updated daily or recently
- Insignificant variables
  - Some variables hold no variable (aka country codes)
- Multicollinearity
  - Some variables are running averages of other variables and others are scaled by population size
  - Only 3 predictors have large multicollinearity issues with many other variables: population, life_expectancy, human_development_index

Our World in Data. (2023). Our World in Data - COVID-19 [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/6559049

04

Methodology

Methodology



05

References

Useful Articles & Books

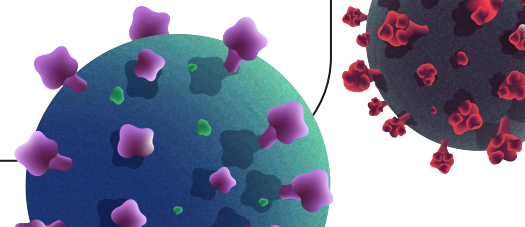
Below are starting papers and books that describes how to approach time series predicting and different approaches to predicting COVID-19.

Predicting Time Series Data

- Alassafi M., Jarrah M, Alotaibi R. (2022). Time series predicting of COVID-19 based on deep learning.
- Long, J. (JD), & Teetor, P. (2019). *R cookbook, 2nd edition*. 14 Time Series Analysis.

Predicting COVID-19

- Painuli, D., Mishra, D., Bhardwaj, S., & Aggarwal, M. (2021). Forecast and prediction of COVID-19 using machine learning.
- Smita Rath, Alakananda Tripathy, Alok Ranjan Tripathy. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model.



06

Conclusion

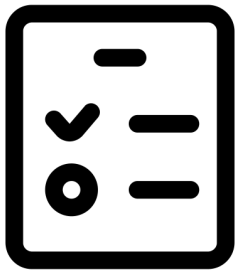
Our Plan

What We Did

1. Set a **project goal**: predict the daily count of new confirmed COVID-19 cases in the US
2. Determine our **regression models**
3. Determine performance metrics **(RMSE, R^2)**
4. Started data preprocessing to narrow down predictor variables

What We Plan to Do

1. Continue working with the data
2. Build and run models
3. Apply our insights!



Thank You!