


Machine Learning Models to Predict the Deaths Associated with COVID-19

Contributors:

Seung Woo Choi, Eric Rodriguez, and
Muhammad Tahir

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.

Contents:

- Problem Statement
 - Data Source
 - Baseline, Ridge and Lasso Modelling
 - Production Model
 - Conclusion and Future Steps
- 
- A large, dark blue, curved shape that starts from the bottom left and extends diagonally upwards towards the right, covering the bottom half of the slide.

Problem Statement

- **Context:** Data science consultants for the CDC
- **Goal:** Build a regression model to predict COVID-19 deaths and to identify important features that can be applied towards a time series forecasting model

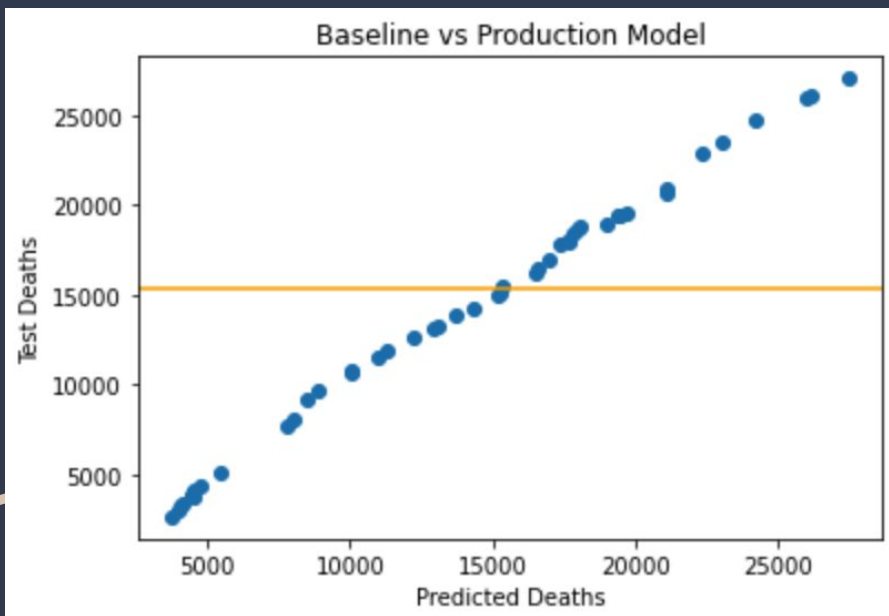
Data Source

- COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University
- Data from July 2020 to December 2020 was used

Baseline, Ridge, and Lasso Models

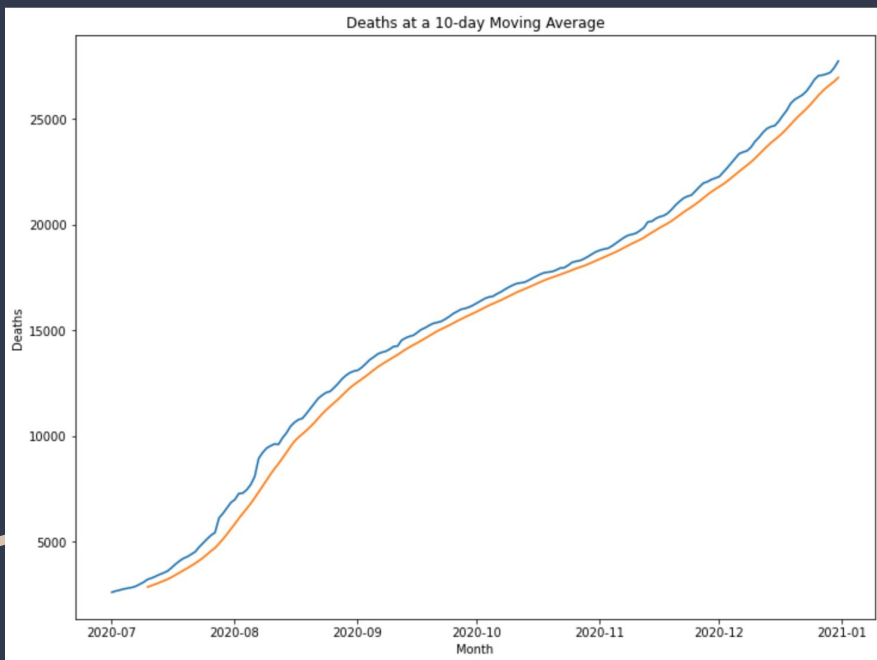
- Baseline model:
 - R-squared: 0
 - RMSE: 6936.42
- Ridge regression model:
 - R-squared: 99.5%
 - RMSE: 480.43
- Lasso regression model:
 - R-squared: 99.5%
 - RMSE: 494.51

The Production Model



- Model Type: Ridge Regression
- Features:
 - Confirmed COVID-19 cases
 - Recovered COVID-19 cases
 - Active COVID-19 cases
 - COVID-19 Incident rate
 - COVID-19 Testing rate
- R-squared:
 - Training: 99.2%
 - Testing: 99.5%
- RMSE:
 - Training: 621.03
 - Testing: 480.43

Conclusion & Future Steps



- **Conclusion:**
 - Ridge regression
 - Confirmed cases and Active cases
- **Future Steps:** Time series modeling to predict future COVID-related deaths

Thank You!