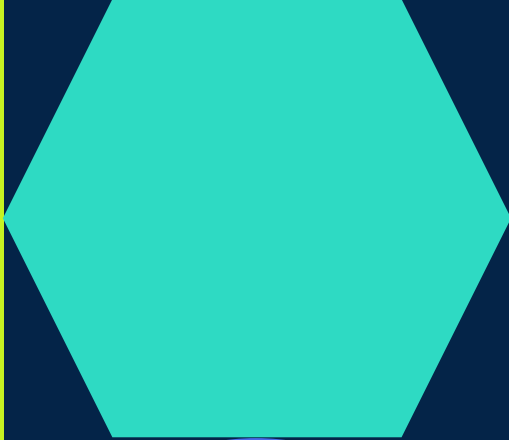
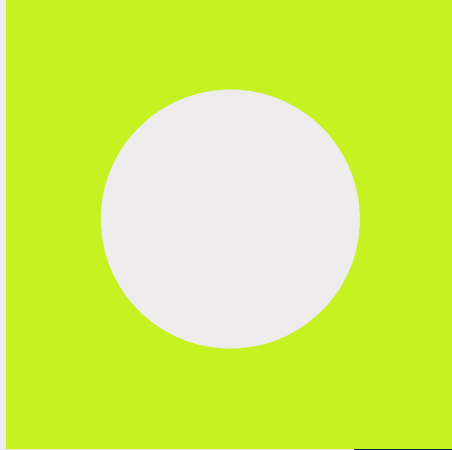




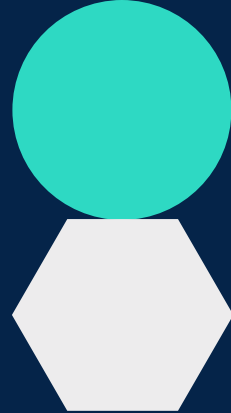
# Time-Series Forecasting for COVID-19 Death Prediction & Decision-Making

MIT Sloan - 15.072

*Alexandre Berkovic  
Max Petruzzi  
Sri Reddy  
Osho Yonzon*



# What is the problem ?



## Epidemic Spread

- Diseases can **rapidly spread** in population
- Understanding the **factors** that can help mitigate the spread is essential

## Staff shortage

- In times of crisis, **human resources** is the most valuable asset
- This needs to be **optimized** to help overcome the crisis quickly

## Unpreparedness

- Hospitals have shown signs of **weakness** during the pandemic
- Understanding which factors come into play is crucial

# Agenda

01.

Data overview

02.

Prediction

03.

Decision-making

04.

Limitations & Next steps





**99,100,000**

TOTAL CONFIRMED COVID-19  
CASES IN THE USA



**50,000**

Data points of hospitals  
across the US

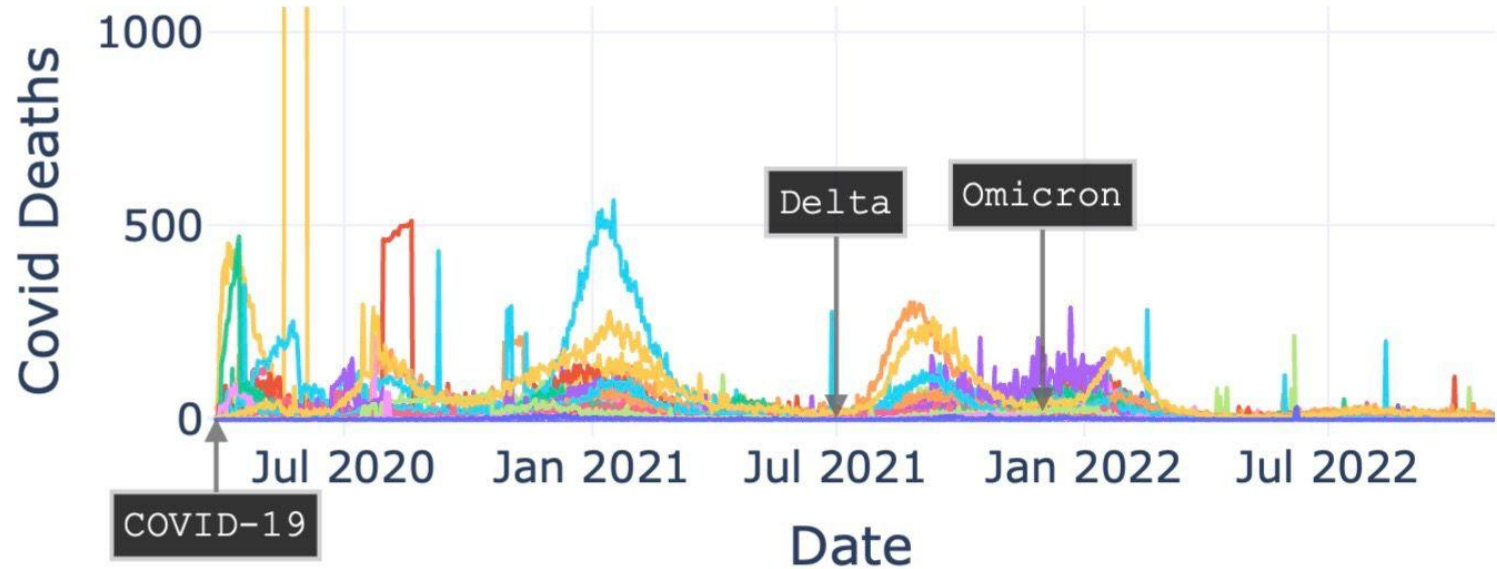
**135**

Features

**$\approx 3$**

Years of data between  
January 2020  
and October 2022

# Covid Deaths by State





# **Feature Selection**

# VARIANCE INFLATION FACTOR

$$VIF_i = \frac{1}{1 - R_i^2}$$



Feature	VIF
critical_staffing_shortage_today_not_reported	5.653332e+07
critical_staffing_shortage_anticipated_within_week	5.651664e+07
inpatient_beds_utilization_utilization	4.223471e+04
inpatient_beds_used	4.125496e+04
inpatient_beds	3.512014e+04
adult_icu_bed_utilization_utilization	1.690436e+04
staffed_adult_icu_bed_occupancy	1.139998e+04
total_staffed_adult_icu_beds	8.717810e+03
critical_staffing_shortage_anticipated_within_week	7.401827e+03
total_adult_patients_hospitalized_confirmed_covid	7.015022e+03
inpatient_bed_covid_utilization_utilization	6.811808e+03
total_adult_patients_hospitalized_confirmed_and	6.517941e+03
critical_staffing_shortage_today_no	5.777808e+03
staffed_icu_adult_patients_confirmed_and_suspected	5.063853e+03
staffed_icu_adult_patients_confirmed_covid	4.597697e+03
percent_of_inpatients_with_covid_utilization	4.408997e+03
previous_day_admission_adult_covid_confirmed	1.947843e+03
inpatient_beds_used_covid	1.410311e+03
adult_icu_bed_covid_utilization_utilization	1.068796e+03
critical_staffing_shortage_anticipated_within_week	1.067663e+03
critical_staffing_shortage_today_yes	7.705780e+02
previous_day_admission_adult_covid_confirmed_60-69	6.961388e+02
previous_day_admission_adult_covid_confirmed_70-79	5.727656e+02
previous_day_admission_adult_covid_suspected	5.207502e+02
previous_day_admission_adult_covid_confirmed_50-59	4.231249e+02
previous_day_admission_adult_covid_suspected_60-69	3.237052e+02
previous_day_admission_adult_covid_suspected_70-79	2.963050e+02
previous_day_admission_adult_covid_confirmed_80+	2.858520e+02
previous_day_admission_adult_covid_suspected_80+	2.407732e+02
previous_day_admission_adult_covid_suspected_50-59	2.111917e+02
previous_day_admission_adult_covid_confirmed_30-39	2.107643e+02
previous_day_admission_adult_covid_confirmed_40-49	2.072935e+02
previous_day_admission_adult_covid_suspected_40-49	1.293927e+02
previous_day_admission_adult_covid_suspected_20-29	1.118458e+02
previous_day_admission_adult_covid_suspected_30-39	1.083590e+02
previous_day_admission_adult_covid_confirmed_20-29	9.016365e+01
hospital_onset_covid	4.228266e+01
previous_day_admission_adult_covid_confirmed_unknown	2.159156e+01
previous_day_admission_adult_covid_suspected_18-19	1.816665e+01
previous_day_admission_adult_covid_suspected_unknown	6.605897e+00
previous_day_admission_adult_covid_confirmed_18-19	1.303630e+00

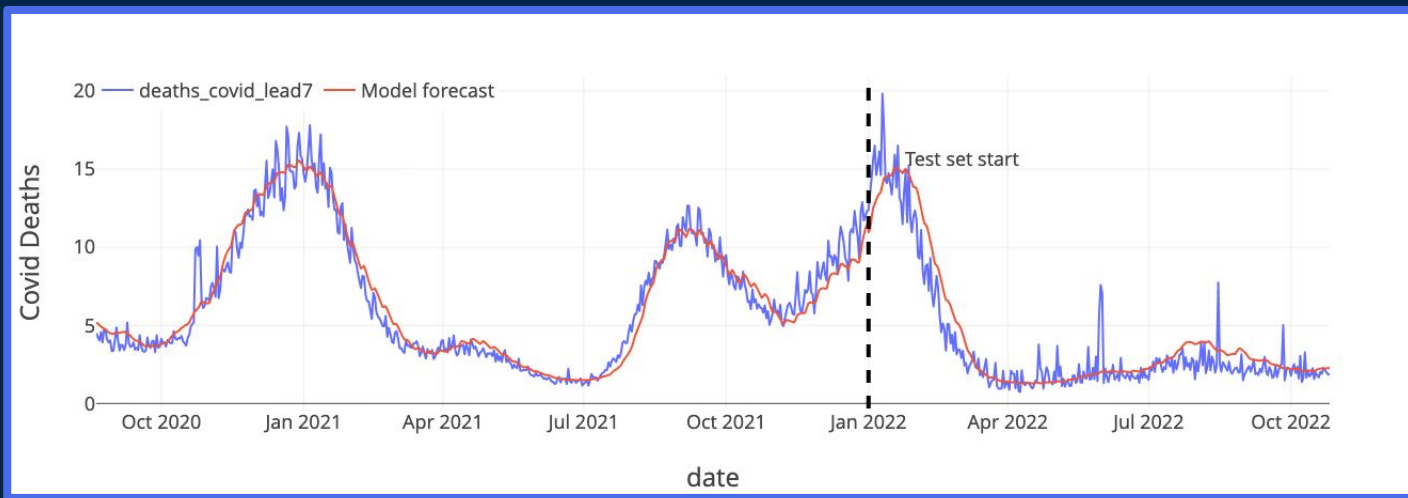


Feature	VIF
inpatient_beds_used_covid	113.996285
total_adult_patients_hospitalized_confirmed_and...	78.190769
adult_icu_bed_covid_utilization_utilization	48.977481
hospital_onset_covid	17.429036
critical_staffing_shortage_anticipated_within-w...	7.738703
deaths_covid	7.660202



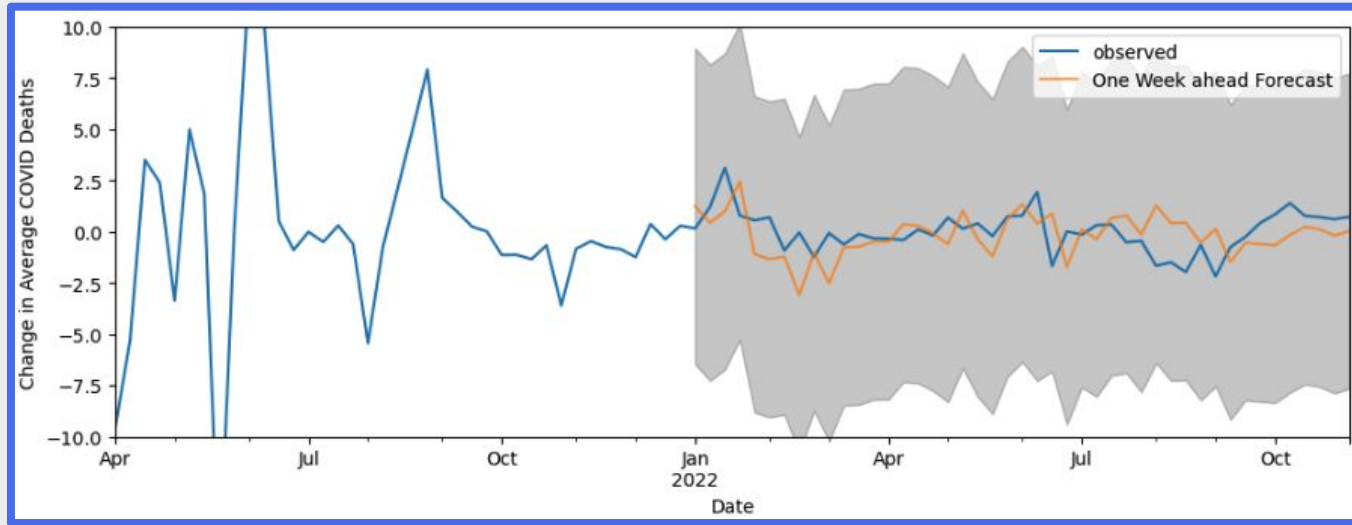
# **Prediction of COVID Deaths**

# LSTM



7 days forecast:  $\text{MSE} = 0.114$   
14 days forecast:  $\text{MSE} = 0.142$

# SARIMA



7 days forecast: MSE = 1.83

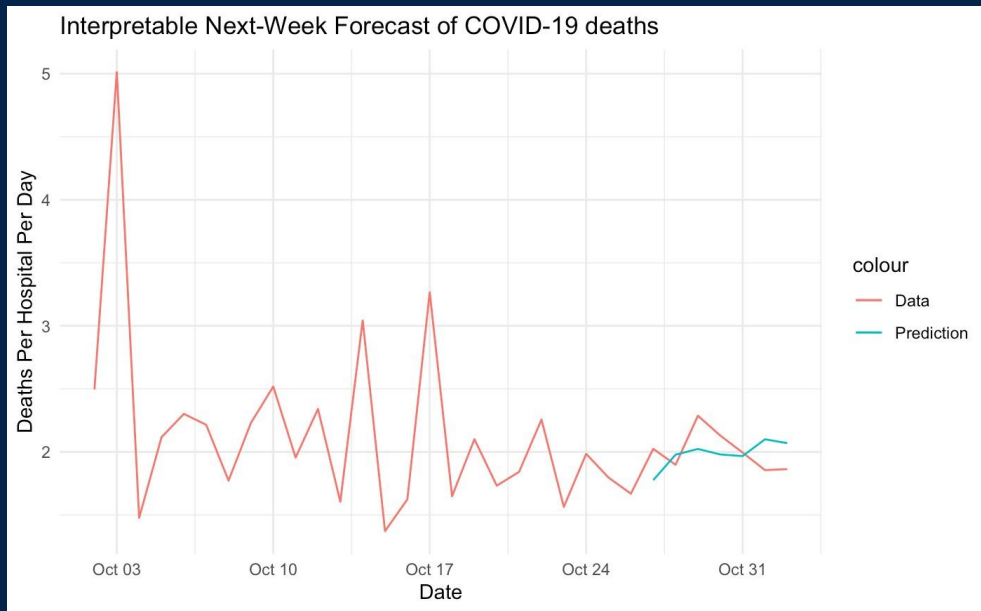
Model parameters:  $(0,1,0) \times (0,1,1,52)$



# **Decision-Making for Hospitals**

# Random Forest

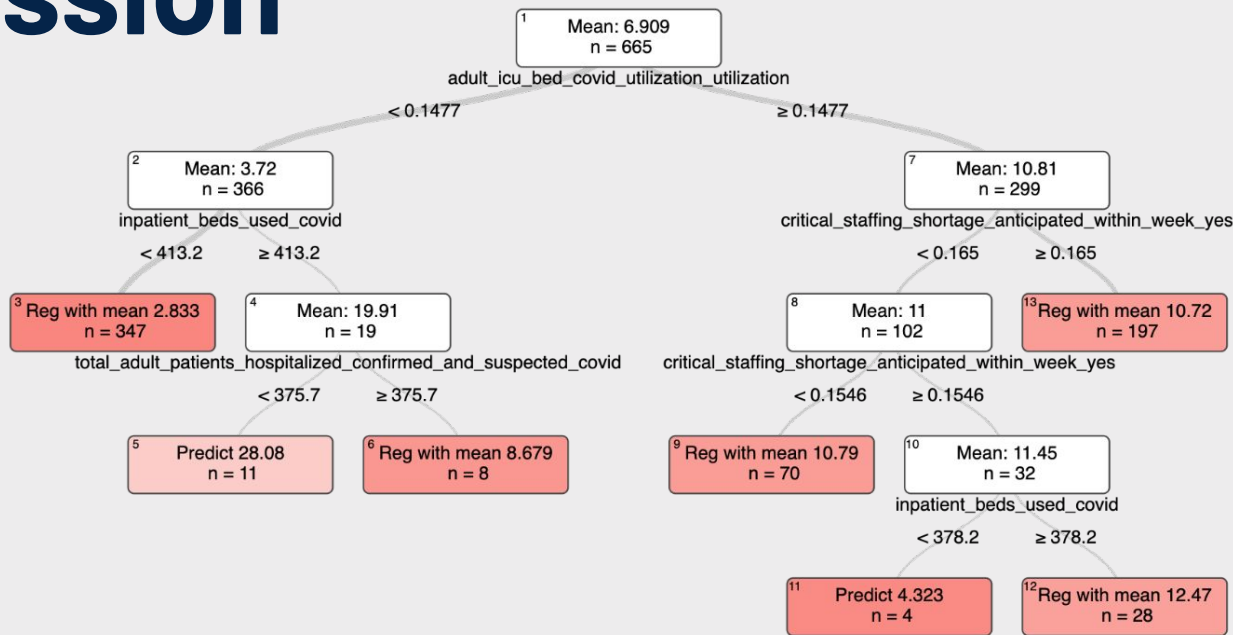
7-Day Forecast  
MSE: 0.028



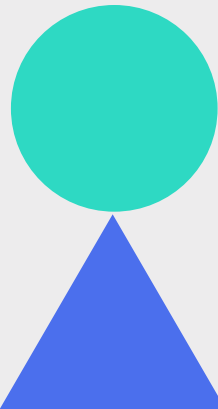
Feature Name	Average Importance Rating	Average Rank
Is Critical Staffing Shortage Expected	494.36	1
Number of Inpatient Beds Used	14.4	2
Number of COVID Deaths	8.50	3.14
Adult COVID ICU Bed Utilization	2.52	4.71
Previous Day's Admission of Suspected COVID (80+ yrs old)	0.85	6

# Optimal Regression Trees

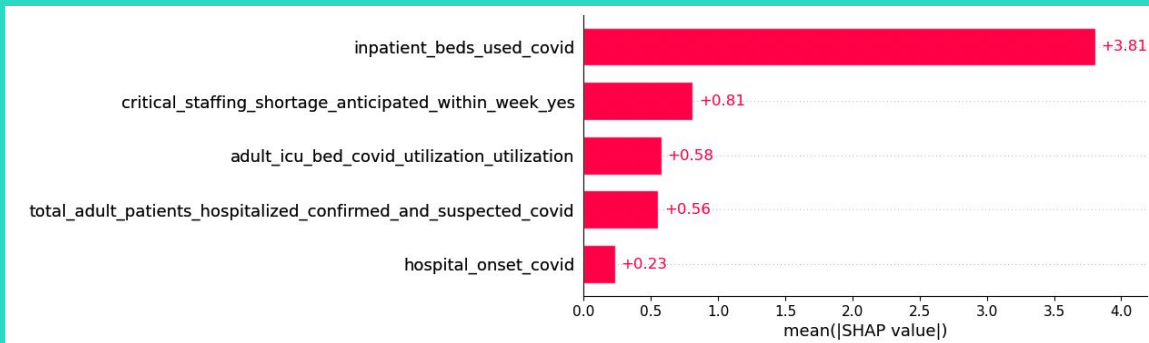
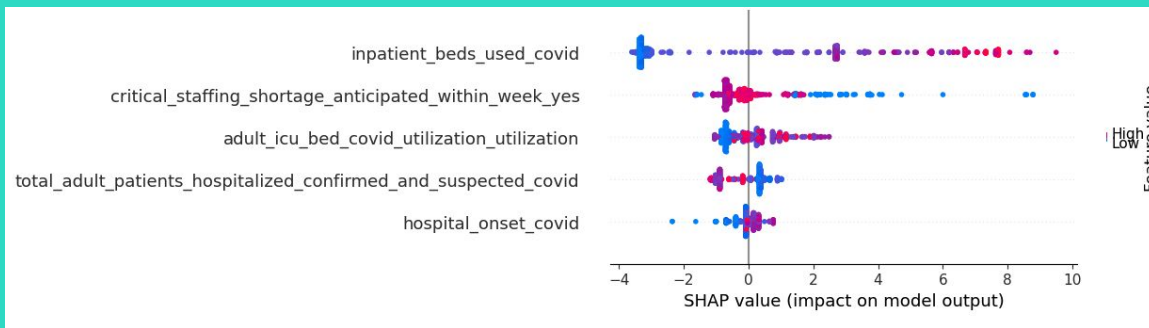
Linear Prediction  
Implemented in the Leaves



*Out-of-sample  $R^2 = 0.826$*



# Shapley + XGBoost





# Insights



## LSTM = prediction & ORT = decision-making

- LSTMs out performed the SARIMA model by a **1.716 MSE**
- **ORTs** clearly point out the synergies between variables leading to an estimated number of deaths



## Identified contributor to COVID deaths

- **Inpatient Beds Used** and **Critical Staffing Shortage** were the biggest contributors to COVID deaths.



## Hospital Unpreparedness Acted as Catalyst for COVID Deaths

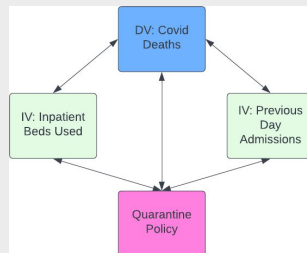
- **Inpatient Beds Used** and **Critical Staffing Shortage** were the biggest contributors to COVID deaths.

# Limitations



## Confounding Variables

- During the pandemic there were factors affecting both hospitals and the number of COVID deaths.
- These variables “confound” true result, skewing our predictions and interpretation.



## Data

- Though we found some annual seasonality in the data, the limited number of observations prevented us from modeling this accurately.



## Variables

- We trimmed down to a total of five variables due to collinearity. If we were to get more diverse features, we could obtain stronger predictions and clearer interpretability.

# Next Steps



## SIR Modeling

Susceptible, Infectious and Recovered models used for epidemic spread



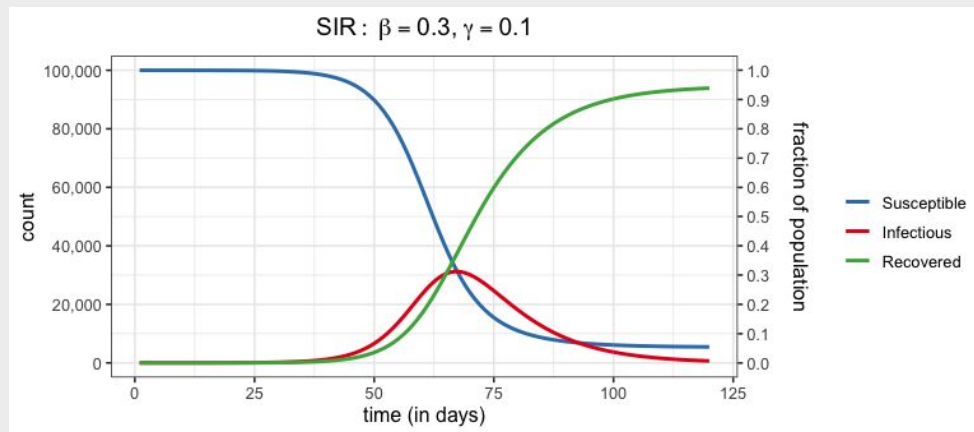
## Markov Chains

Probabilistic perspective on infectious disease spread in a population



## Confounding Variables

Examine data regarding many external factors to COVID to see all that contributed to COVID deaths





**Thank you!**