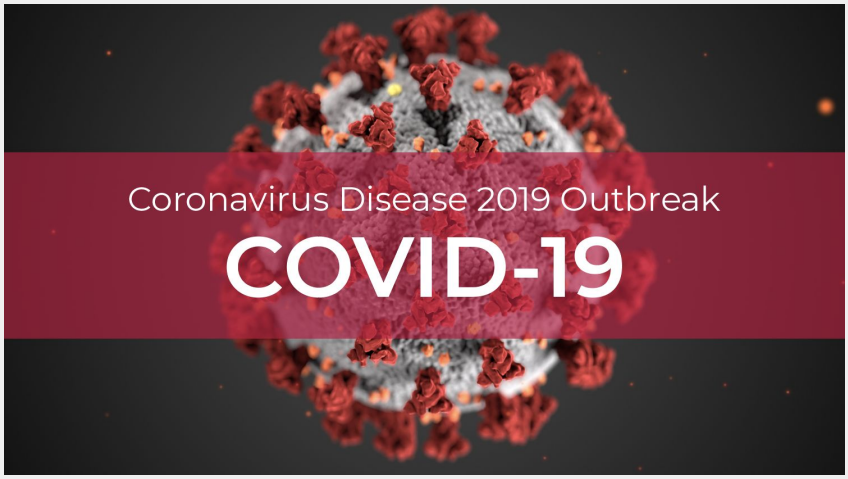


Final Project

Business Case

In this report, three topics regarding covid-19 are covered:

- The covid-19 pandemic test, case, recovery, death, and vaccination totals are covered followed by time-series modeling to determine the progression of the pandemic.
- A logit model is used to determine the log odds relationship between government covid-19 regulations and cases of SARS-CoV-2.
- SARS-CoV-2 symptoms, lifecycle, testing, and vaccines are explained followed by a recurrent neural network that is used to predict the next nucleotide in the sequence of a SARS-CoV-2 genome.

Forecasting The Pandemic

Active Cases of Virus

	Country	Active Cases
1	United States	6948028
2	France	4197252
3	Brazil	1371216
4	Belgium	793295
5	Italy	562832
6	India	553874
7	Poland	388235
8	United Kingdom	379848
9	Ukraine	323448
10	Russia	282382

The United States has cumulatively had, and still has, the most cases of the virus.

Total Tests for Virus

	Country	Tests
1	United States	401946739
2	India	242650025
3	China	160000000
4	United Kingdom	124452321
5	Russia	119900000
6	France	63999096
7	Italy	49551436
8	Germany	48979281
9	Spain	42707830
10	Turkey	38338045

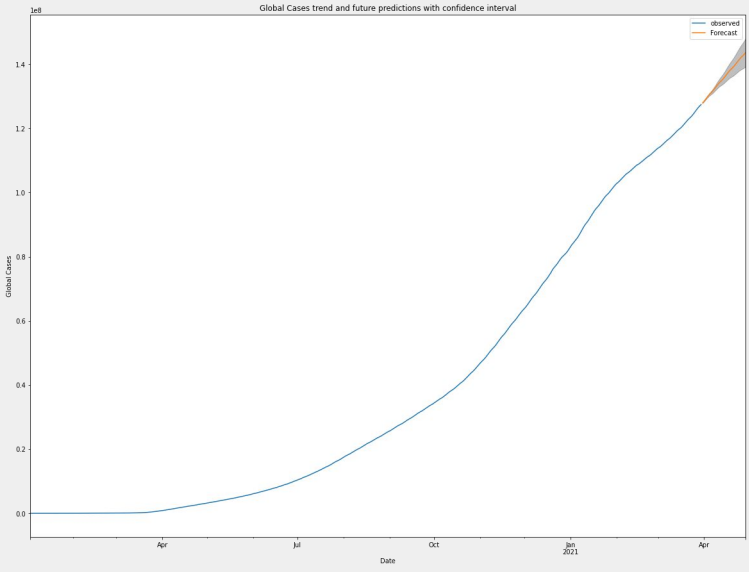
The United States likely has the most cases because it has tested the most.

Total Vaccinations for Virus

	Country	Active Cases	Vaccinations
1	United States	6948028	145812835
2	China	173	110962000
3	India	553874	61113354
4	United Kingdom	379848	34119095
5	Brazil	1371216	18082153

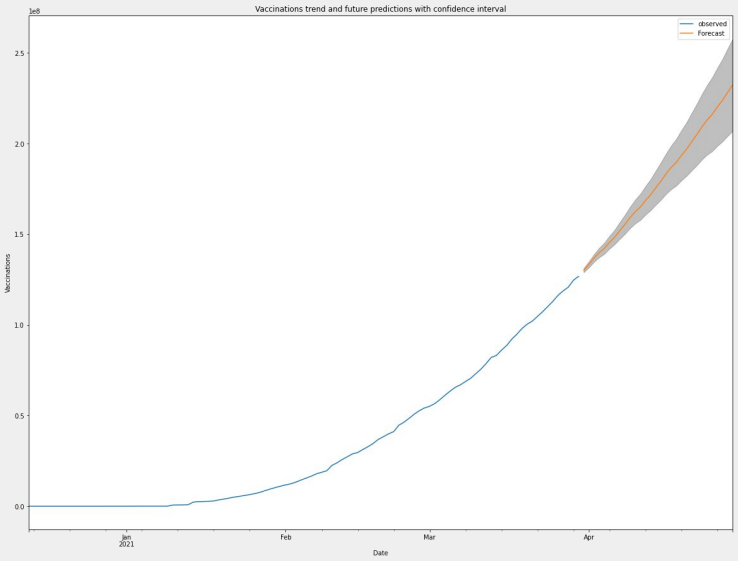
The United States leads in vaccinations.

Virus Cases Forecast



The cumulative cases of the virus are projected to increase by 12.8% over the next month.

Virus Vaccinations Forecast



The cumulative virus vaccinations are projected to increase by 83.3% over the next month.

Government Regulations In Response To Covid-19

Government Regulation Relationship With Virus Cases

Regulations that decreased log odds of virus cases:

- school closures
- travel restrictions
- state of emergency declarations
- wage support
- tax credits
- interest rate lowering
  interest rate lowering

Government regulations such as disallowing public gatherings and mandating wearing masks did not decrease the log odds of the presence of virus cases.

Logit Metrics

Precision Score: 0.8072260328601053

Recall Score: 0.8241521110703962

F1 Score: 0.8156012651852449

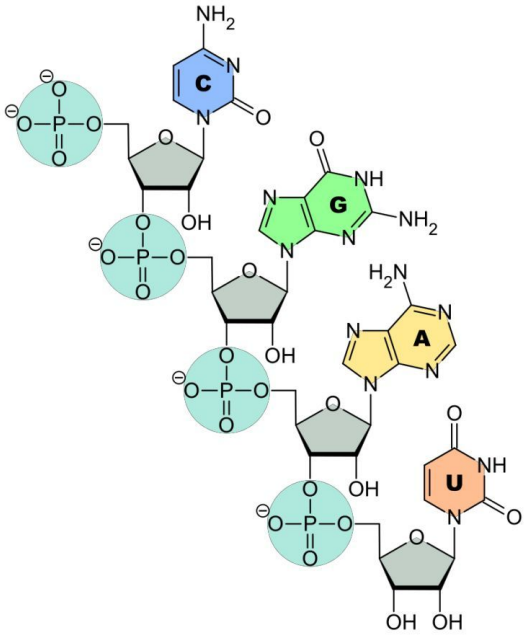
Accuracy Score: 0.817523923444976

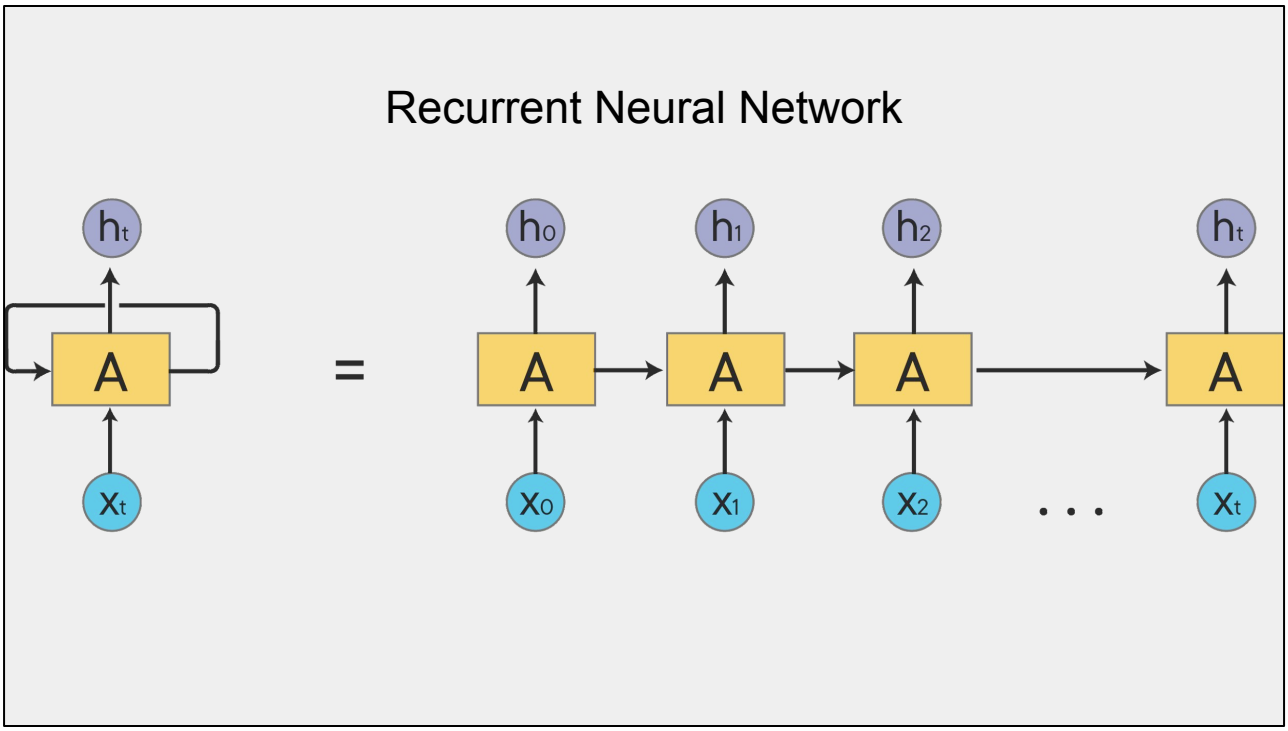
Specificity Score: 0.8278185297400733

ROC AUC: 0.8176583048418117

Sars-CoV-2 Genome

RNA





Recurrent Neural Network is a neural network that passes its output, h, from a layer back into itself as input for the next layer. An RNN is used to predict the next nucleotide in the genome.

RNN Metrics

Loss 40

Accuracy 82

Precision 94

Recall 73

Conclusion

- Predicting the next nucleotide in a genome will be useful in reconstructing fragmented genomes for curing genetic diseases and restoring extinct species.
- The virus spread is slowing and will decrease more as more people get vaccinated.
- Instead of wearing masks and preventing gatherings, carrying handkerchiefs in which people could sneeze or cough and sanitizing areas were people gather would be sufficient in preventing the spread of SARS-CoV-2.

Future Work

Using neural networks to predict sequences of nucleotides or amino acids.

Thank You

Sources

https://www.ncbi.nlm.nih.gov/

https://datarepository.wolframcloud.com/

https://machinelearningmastery.com/

https://towardsdatascience.com/

https://medium.com/

https://stackoverflow.com/