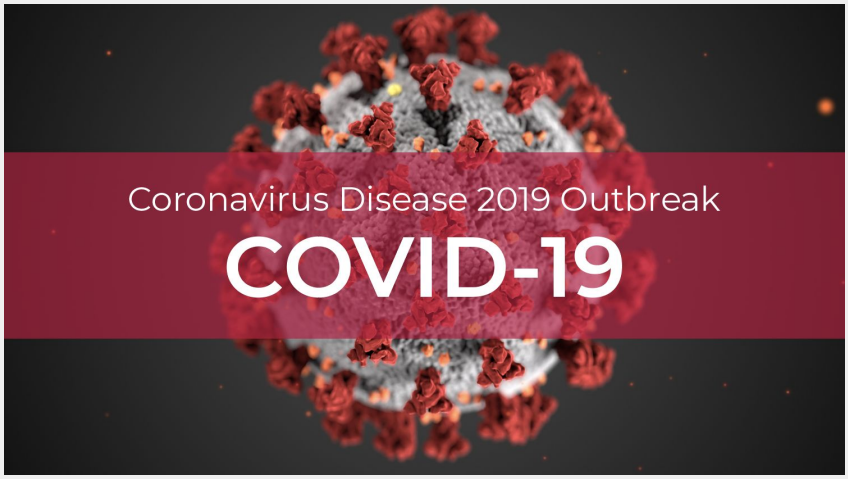
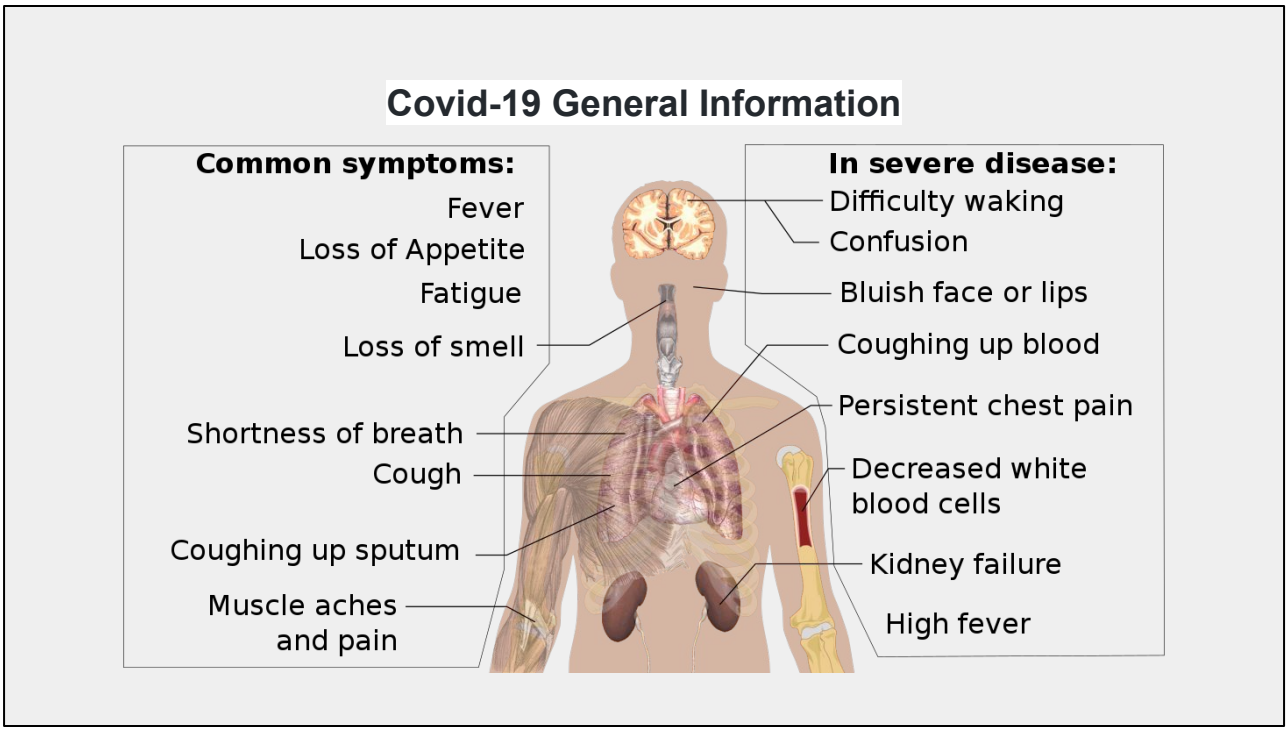


Final Project



A new type of coronavirus, SARS-CoV-2, started in December 2019 and has caused the COVID-19 global pandemic. SARS-CoV-2 originated in a Chinese city called Wuhan. Coronaviruses are common in animals, and scientists hypothesize that SARS-CoV-2 may have transferred to humans, zoonosis, through the Wuhan food market. SARS stands for severe acute respiratory syndrome. In 2003, there was an outbreak of SARS in China that spread to other countries and then ended in 2004. SARS-CoV-2 has spread faster than the 2003 SARS. SARS-CoV-2 is spread when an infected person coughs or sneezes droplets of saliva or mucus with the virus into the air. The respiratory droplets usually do not go further than a few feet, are airbourne for a few moments, and then land on a surface. SARS-CoV-2 has an incubation period of 2 to 14 days.

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

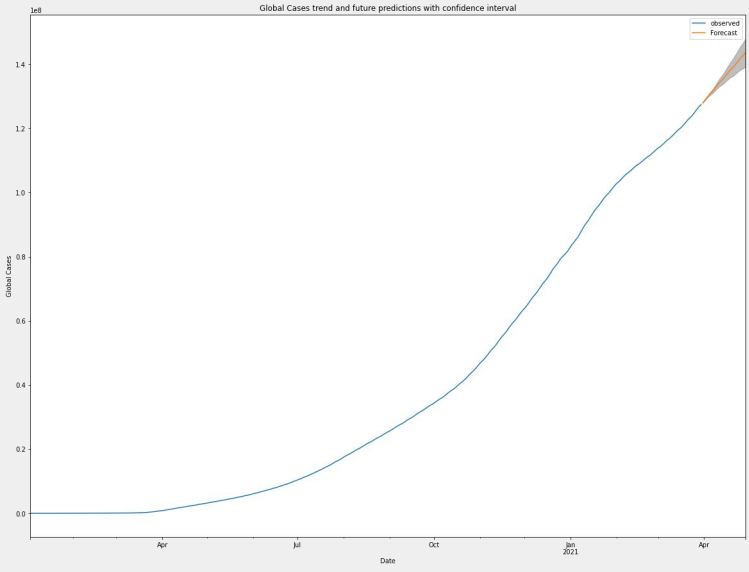
The United States likely has the most cases because it has tested the most.

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

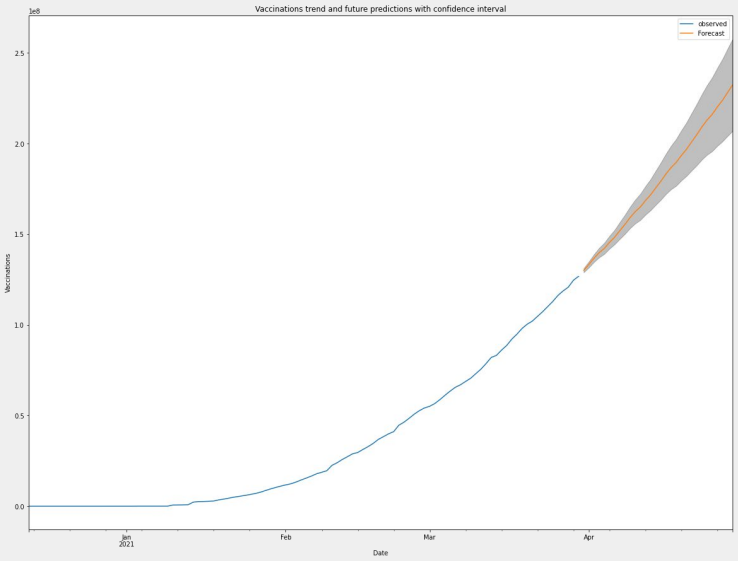
The United States leads in vaccinations.

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

- cases: binary variable equal to 1 if there were cases of SARS-CoV-2 and 0 otherwise;
- school: binary variable equal to 1 if schools were closed and 0 otherwise;
- domestic: binary variable equal to 1 if there was a domestic lockdown and 0 otherwise;
- travel: binary variable equal to 1 if travel restrictions were implemented and 0 otherwise;
- travel_dom: binary variable equal to 1 if travel restrictions within the country (e.g. inter-region travels) were implemented and 0 otherwise;
- curf: binary variable equal to 1 if a curfew was implemented and 0 otherwise;
- mass: binary variable equal to 1 if bans on mass gatherings were implemented and 0 otherwise;
- elect: binary variable equal to 1 if some elections were postponed and 0 otherwise;
- sport: binary variable equal to 1 if bans on sporting and large events were implemented and 0 otherwise;
- rest: binary variable equal to 1 if restaurants were closed and 0 otherwise;
- testing: binary variable equal to 1 if there was a public testing policy and 0 otherwise;
- surveillance: binary variable equal to 1 if mobile app or bracelet surveillance was implemented and 0 otherwise;
- masks: binary variable equal to 1 if the obligations to wear masks in public spaces was implemented and 0 otherwise;
- state: binary variable equal to 1 if the state of emergency is declared and 0 otherwise;
- cash: binary variable equal to 1 if cash transfers are implemented and 0 otherwise;
- wage: binary variable equal to 1 if wage support is implemented and 0 otherwise;
- credit: binary variable equal to 1 if credit schemes are implemented and 0 otherwise;
- taxc: binary variable equal to 1 if tax credits are implemented and 0 otherwise;
- taxd: binary variable equal to 1 if tax delays are implemented and 0 otherwise;
- export: binary variable equal to 1 if supports to importers or exporters are implemented and 0 otherwise;
- rate: binary variable equal to 1 if the Central Bank lowered the interest rates and 0 otherwise;

Government Regulation Log Odds Relationship With Virus Cases

Feature	coef	std err	z	P>|z|	[0.025	0.975]
school	-0.2940	0.025	-11.625	0.000	-0.344	-0.244
domestic	0.5552	0.036	15.518	0.000	0.485	0.625
travel	-0.5417	0.032	-17.017	0.000	-0.604	-0.479
travel_dom	0.1073	0.033	3.226	0.001	0.042	0.173
curf	0.1619	0.032	5.057	0.000	0.099	0.225
mass	0.3236	0.035	9.146	0.000	0.254	0.393
elect	0.5072	0.032	15.606	0.000	0.443	0.571
sport	0.1403	0.034	4.137	0.000	0.074	0.207
rest	0.4030	0.031	13.109	0.000	0.343	0.463
masks	0.7515	0.030	25.031	0.000	0.693	0.810
surveillance	0.3802	0.041	9.341	0.000	0.300	0.460
state	-0.1474	0.028	-5.255	0.000	-0.202	-0.092
cash	0.7581	0.033	23.305	0.000	0.694	0.822
wage	-0.0819	0.032	-2.555	0.011	-0.145	-0.019
credit	0.0879	0.032	2.733	0.006	0.025	0.151
taxc	-0.3738	0.029	-12.740	0.000	-0.431	-0.316
taxd	0.4360	0.031	14.210	0.000	0.376	0.496
export	0.1784	0.036	5.006	0.000	0.109	0.248
rate	-0.0819	0.030	-2.756	0.006	-0.140	-0.024

The logit model suggests that the government regulations of school closures, travel restrictions, state of emergency declarations, wage support, tax credits, and interest rate lowering decreased the log odds of the presence of virus cases. Government regulations such as disallowing public gatherings and mandating wearing masks did not decrease the log odds of the presence of virus cases.

Logit Model Metrics

Train:

Precision Score: 0.8072260328601053

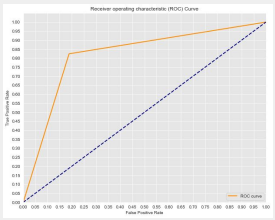
Recall Score: 0.8241521110703962

F1 Score: 0.8156012651852449

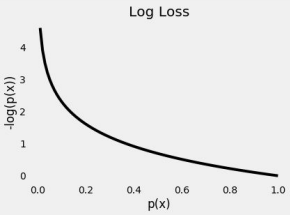
Accuracy Score: 0.817523923444976

Specificity Score: 0.8278185297400733

ROC AUC: 0.8176583048418117



Mean Cross Validation of Cost Function Negative Log Loss Score:
-0.42917685692963775



Test:

Precision Score: 0.8076553491351698

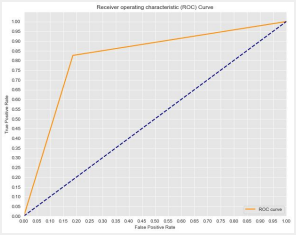
Recall Score: 0.8259089420242385

F1 Score: 0.8166801619433198

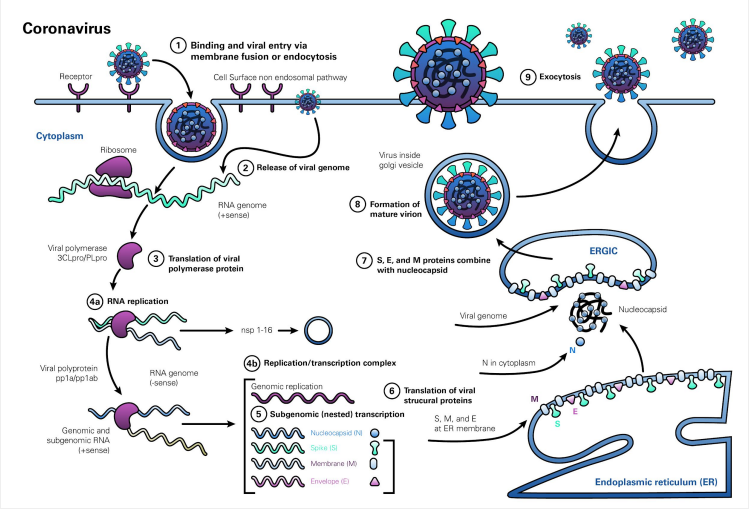
Accuracy Score: 0.8194577352472089

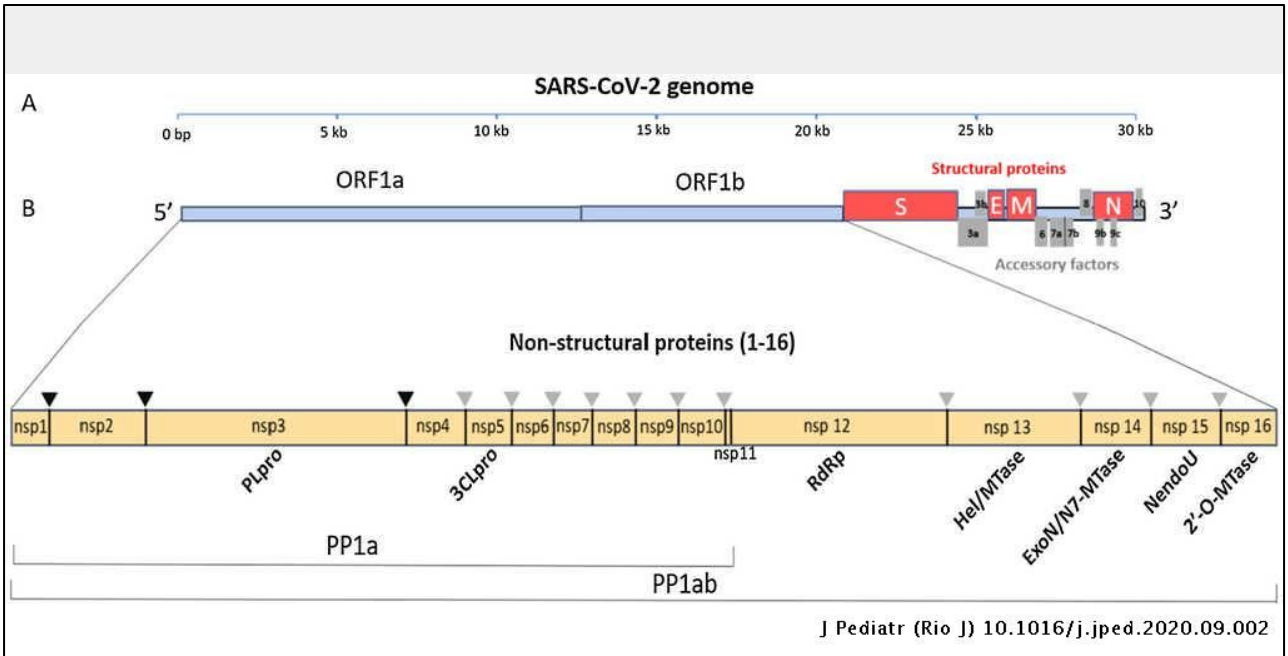
Specificity Score: 0.8311626429479034

ROC AUC: 0.8196221738408417

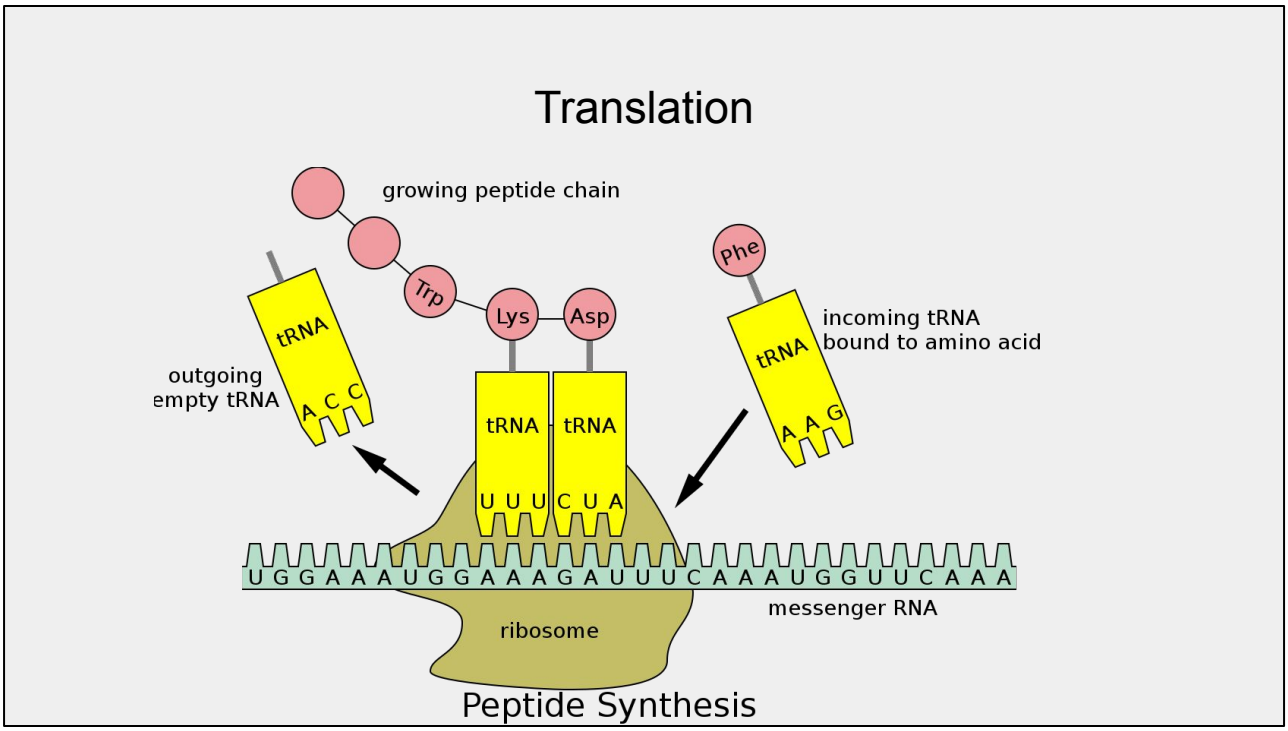


Sars-cov-2 lifecycle



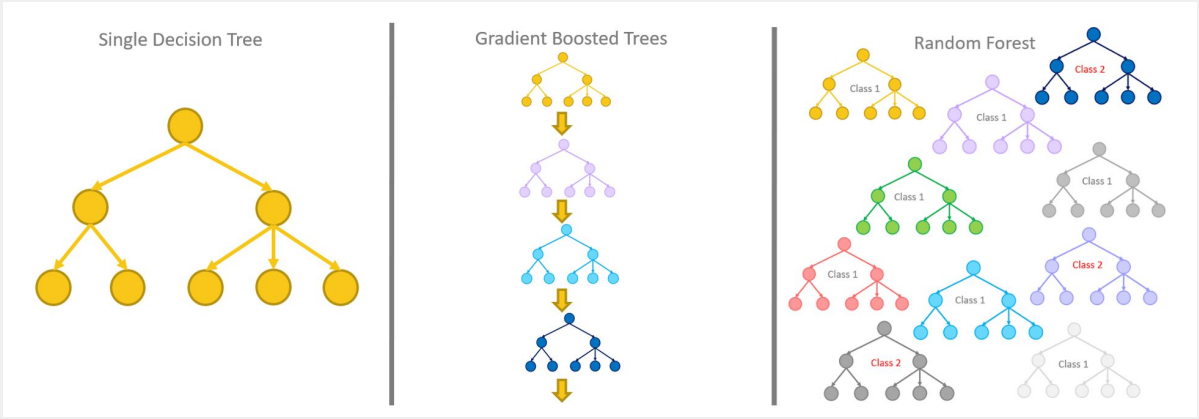


J Pediatr (Rio J) 10.1016/j.jped.2020.09.002



Machine learning is used to replicate translation.

Machine Learning Models



Machine Learning Metrics

Train:

Precision Score: 1.0

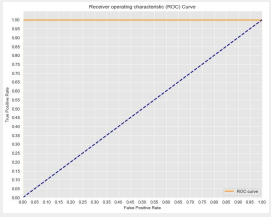
Recall Score: 1.0

F1 Score: 1.0

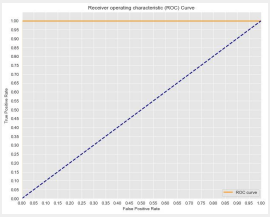
Accuracy Score: 1.0

Specificity Score: 1.0

ROC AUC: 1.0



Cross Validated ROC AUC score: 1.0



Test:

Precision Score: 1.0

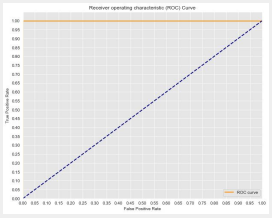
Recall Score: 1.0

F1 Score: 1.0

Accuracy Score: 1.0

Specificity Score: 1.0

AUC: 1.0



The the machine learning translation algorithms, the decision tree, random forest, and gradient boost all predicted the train and test set with scores of 100.

Conclusion

- 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/