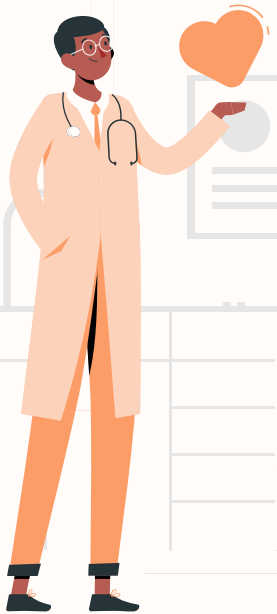


COVID-19 Test Predictability

Sidni Johnson
General Assembly

Problem Statement

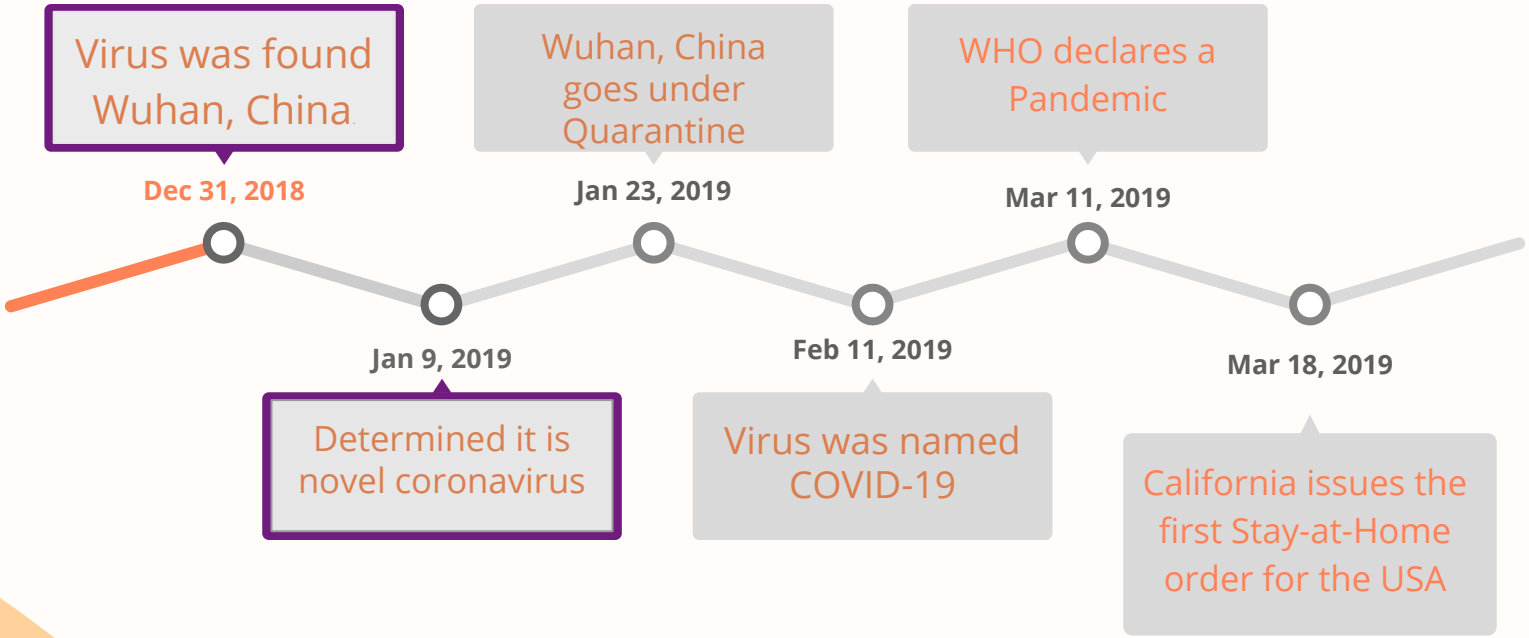
Predicting COVID-19 test results could be easy, with a simple questionnaire. Just asking a patient a few poignant questions could save the hospitals millions of dollars. We at InfoMedics, are attempting to devise a method that could do such that. With our questionnaire, we are trying to predict whether or not a person will have a positive or negative test from a few simple questions.

TABLE OF CONTENTS

1

Background

BACKGROUND- World Health Organization



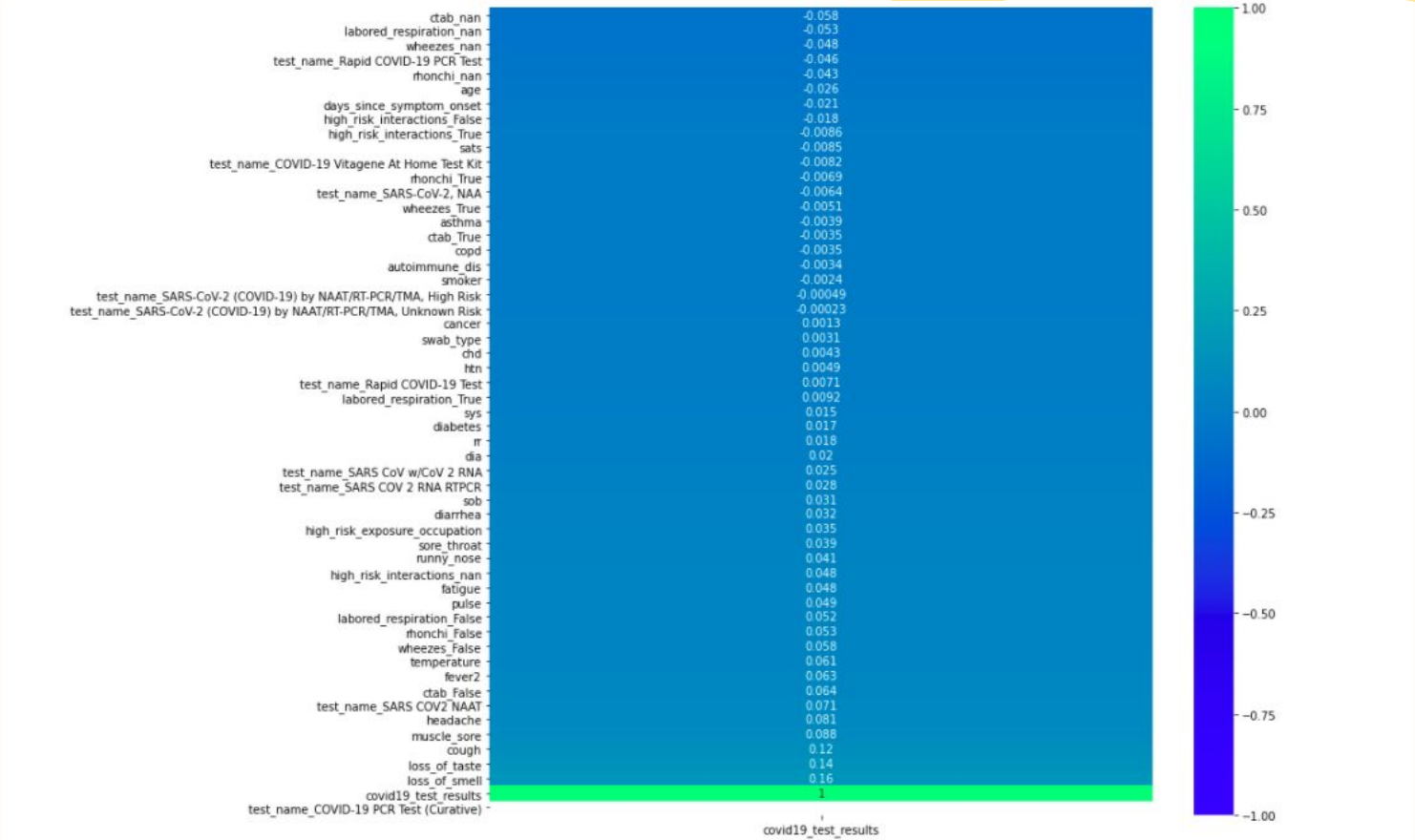
2,131,000+

Deaths worldwide from COVID-19

DATA: CovidClinicalData.org

	Rows	Columns
Original DataFrame	93995	46
Cleaned DataFrame	73369	55

Correlation in relation to COVID19 test results



TABLEAU

Links at the end of the presentation

REMOVED INFORMATION

Before modeling, we removed any data that dealt with swab type or testing method.

2

MODEL

Neural Networks

RECURRENT NEURAL NETWORK

Black box model

PROBLEMS WITH REAL WORLD DATA

Imbalanced-learn Library

IMBALANCED-LEARN

UNDERSAMPLING

Remove data from majority class in order to balance the classes

OVERSAMPLING

Generates new samples of the under-represented class

SMOTE

Synthetic Minority Oversampling Technique

3

ANALYSIS

PERFORMANCE MEASURES

ACCURACY

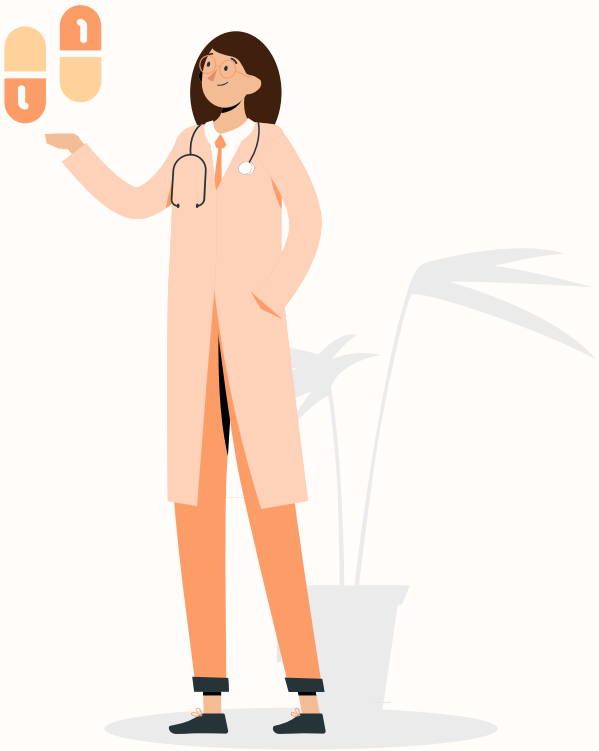
rate of correct classification

RECALL/ Sensitivity

the percentage of total relevant results correctly classified

PRECISION

the percentage of your results which are relevant



.9820

Baseline Score

EVALUATING THE MODEL

	Accuracy	Recall	Precision
Undersampling	.9800	.003	0.025
Oversampling	.8080	0.7386	0.066
SMOTE	.9834	0.201	0.611

	Predicted Class		
Actual Class		1	0
	1	243	86
	0	3435	14579

Conclusion

- Oversampling the data has worked best for neural networks.
- Recall is the best metric to quantify imbalanced classes
- Since we used a neural network there is not much interpretability.

Future Steps

- Find a balance between accuracy and recall
- I would like to create an interactive questionnaire that runs a neural network in the background, so that within a few seconds you knew whether you should quarantine without taking a psychical test.

RESOURCES

https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline?gclid=CjwKCAiA9bmABhBbEiwASb35V5XzjGvljqGnIyw1lFlpEkUKDMhM_8bsjNQRw_ZHYMEzy3gTfx74RBoCCAcQAvD_BwE#event-91

https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020

https://www.google.com/search?sxsrf=ALeKk030EXaNzMLm7c6W4mDmQm9WuWFFZA%3A1611589252796&ei=hOYOYPiUMISu5wLyxpOYCQ&q=total+deaths+from+covid+19+worldwide&oq=total+deaths+worldwide&gs_lcp=CgZwc3ktYWIQARgAMgYIABAHEB4yBggAEAcQHjICCAAyAggAMgYIABAHEB4yBggAEAcQHjICCAAyAggAMgYIABAHEB4yAggAOgQIABBHOgQIABANUNMeWMolYI8yaABwAngAgAFZiAGWAZIBATKYAQCgAQGqAQdnd3Mtd2l6yAEEwAEB&sclient=psy-ab

https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/#:~:text=80%25%20accurate.&text=Precision%20%2D%20Precision%20is%20the%20ratio,the%20total%20predicted%20positive%20observations.&text=F1%20score%20%2D%20F1%20Score%20is,and%20false%20negatives%20into%20account.

Tableau links

- https://public.tableau.com/profile/sidni.johnson#!/vizhome/AgesintheDataFrame/AgesintheDataFrame?publish=yes
- https://public.tableau.com/profile/sidni.johnson#!/vizhome/COVID19resultsbyBinnedAges/COVID19resultsbyBinnedAges?publish=yes
- https://public.tableau.com/profile/sidni.johnson#!/vizhome/ColumnsOverlap/ColumnsOverlap?publish=yes