Disruptions in Health Care due to COVID-19

MIGUEL SANTANA, MBA

MAY 9TH, 2021

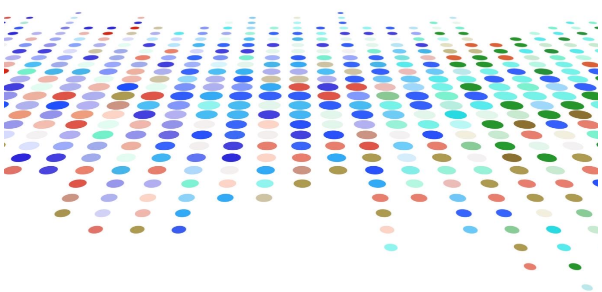
Introduction

- METHODOLOGY & DATA
- DATA PROCESSING
- MODEL SELECTION
  - Validation
- INTERPRET RESULTS
- CONCLUSION
  - Limitations & Future Work
- THANK YOU

Methodology & Data

- COVID-19 CAUSED MAJOR DISRUPTIONS ALL OVER THE WORLD. IN ORDER TO WORK TOWARD NORMALCY, WE MUST IDENTIFY AND UNDERSTAND PAIN POINTS BEFORE BUILDING TOOLS AND RESOURCES THAT HELP VULNERABLE POPULATIONS MANAGE THE NEW CHALLENGES OF THE WORLD.
    - This analysis will explore two datasets in order to gain a deeper understanding of how COVID-19 affected the availability of health care as well as the availability of mental health in the United States.
- **DATASETS**
    - Premise General Population COVID-19 Health Services Disruption Survey 2020
        - Bill and Melinda Gates Foundation
    - Household Pulse Survey (USA) | Mental Health Care (last 4 weeks)
        - Kaggle

Data Processing



OSEMN FRAMEWORK

Obtain, Scrub, Explore, Model, Interpret

- KEY DECISIONS
  - Reducing Datasets
    - Individuals who 'needed' access to health care
      - Specifying whether they did/did not receive it
  - Filling Missing Values
    - Leveraging pre/post COVID-19 habits to fill data

Model Selection

Preliminary models were run to narrow down our selection

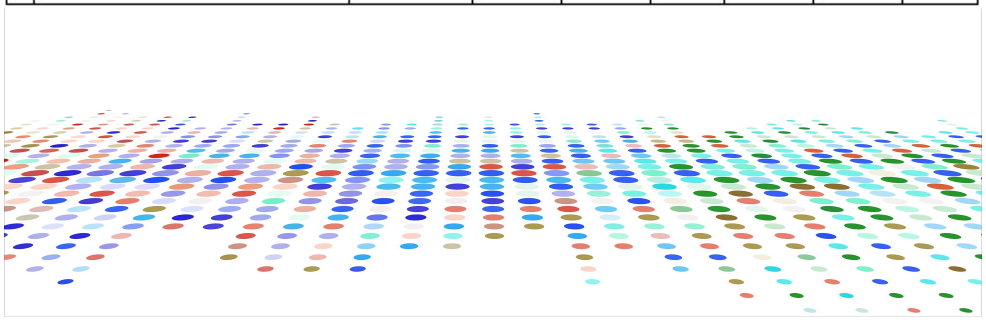
- FACTORS
  - AUC
  - Accuracy
  - Precision
  - Recall
- SELECTION
  - Gradient Boosting Classifier

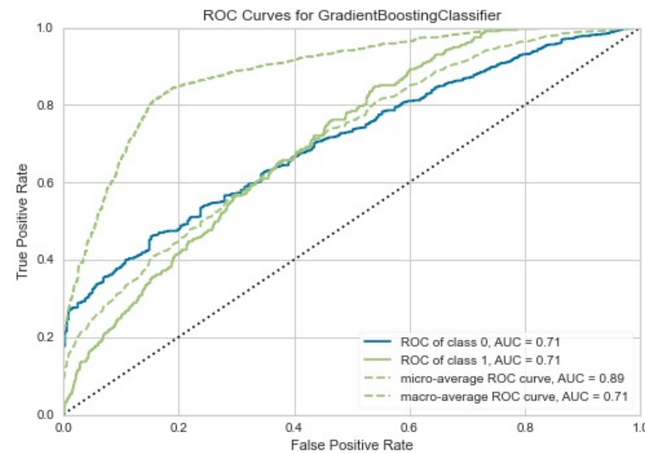
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
0	Logistic Regression	0.8269	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0148
1	Extreme Gradient Boosting	0.8243	0.7056	0.0422	0.4433	0.0759	0.0466	0.0881	0.2835
2	Gradient Boosting Classifier	0.8224	0.7038	0.0395	0.3752	0.0709	0.0397	0.0720	1.0360
3	Random Forest Classifier	0.8186	0.6138	0.0653	0.3801	0.1091	0.0610	0.0917	0.1123
4	Light Gradient Boosting Machine	0.8160	0.6874	0.0884	0.3769	0.1427	0.0814	0.1095	0.3079
5	CatBoost Classifier	0.8144	0.7000	0.0748	0.3432	0.1218	0.0635	0.0875	4.0240
6	Ada Boost Classifier	0.8090	0.6913	0.1007	0.3384	0.1541	0.0801	0.0998	0.2927
7	Extra Trees Classifier	0.8090	0.6602	0.0776	0.2984	0.1229	0.0559	0.0721	0.2492
8	SVM - Linear Kernel	0.7576	0.0000	0.0694	0.1290	0.0895	-0.0350	-0.0373	0.4705
9	Decision Tree Classifier	0.7187	0.5363	0.2571	0.2264	0.2403	0.0689	0.0692	0.0314
10	Ridge Classifier	0.6005	0.0000	0.7020	0.2591	0.3780	0.1678	0.2138	0.0106
11	Linear Discriminant Analysis	0.5995	0.6958	0.7020	0.2586	0.3775	0.1669	0.2129	0.0520
12	K Neighbors Classifier	0.5779	0.4931	0.3524	0.1644	0.2241	-0.0156	-0.0177	0.0150
13	Quadratic Discriminant Analysis	0.3157	0.5796	0.9823	0.1998	0.3320	0.0622	0.1686	0.0171
14	Naive Bayes	0.1736	0.4896	0.9986	0.1731	0.2950	-0.0002	-0.0028	0.0052

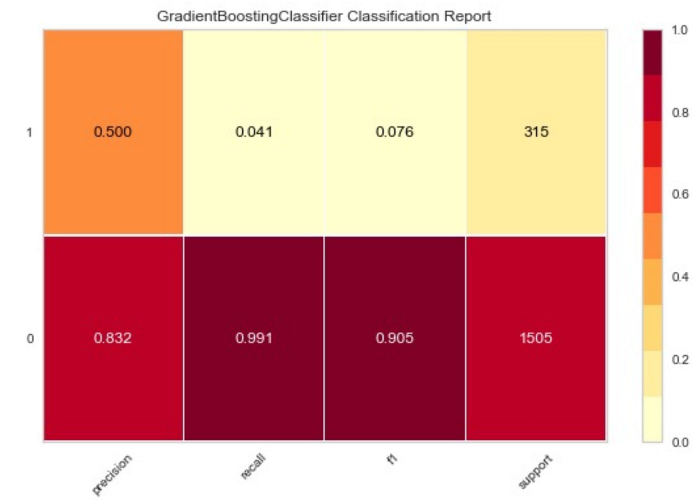
Validation

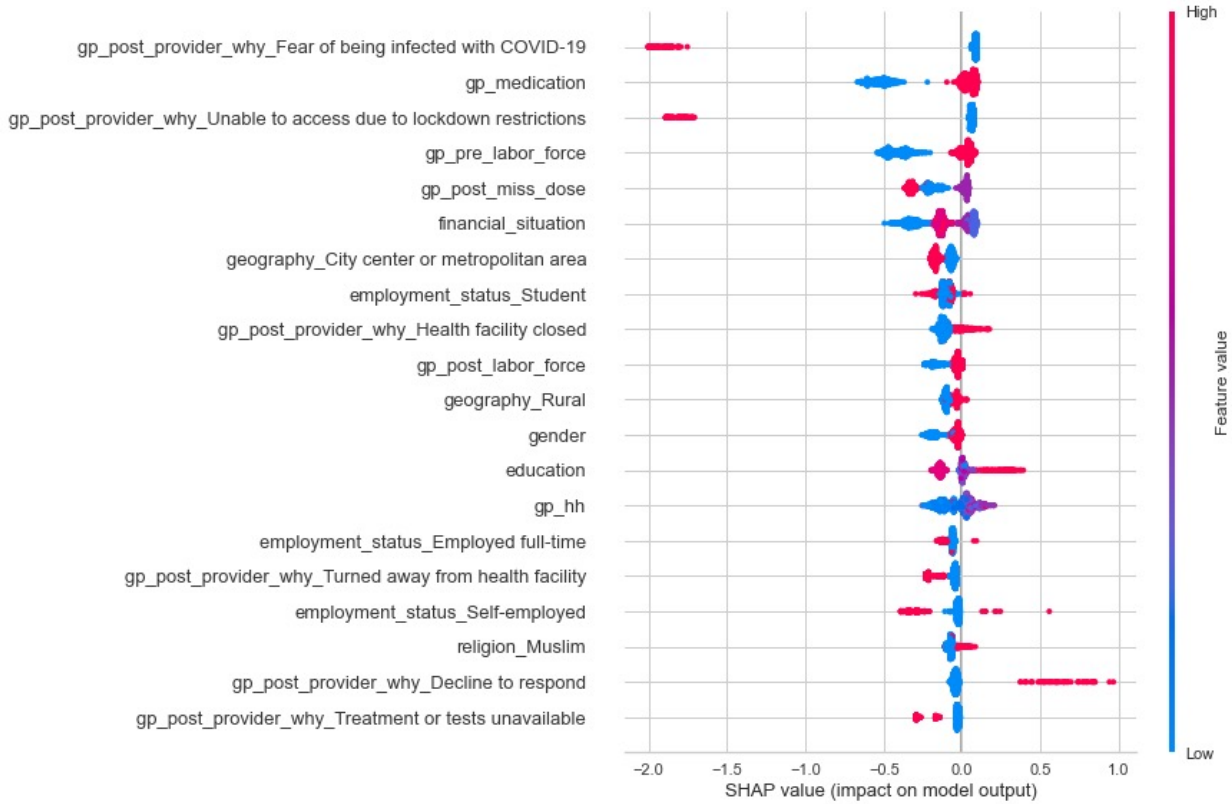
Gradient Boosting Classifier with Cross Validation (5-Fold)

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Gradient Boosting Classifier	0.8269	0.7064	0.0413	0.5	0.0762	0.0512	0.104









Interpret Results

SHAP ANALYSIS

Conclusion

Our analysis is focused on dataset classifiers that negatively affect our target variable. Lower target variables represent limited (or no health care) received by individuals who explicitly needed health care pre/post COVID-19.

Individuals affected by limited (or no health care) have one or more of the following qualities:

- **Reason for Missing Dr. Appointment | Fear of contracting COVID-19**
- **Do not currently take medication**
- **Live in countries with lock down restrictions**
- **Not part of the labor force pre or post COVID-19**
- **Students**
- **Women**
- **Self Employed**

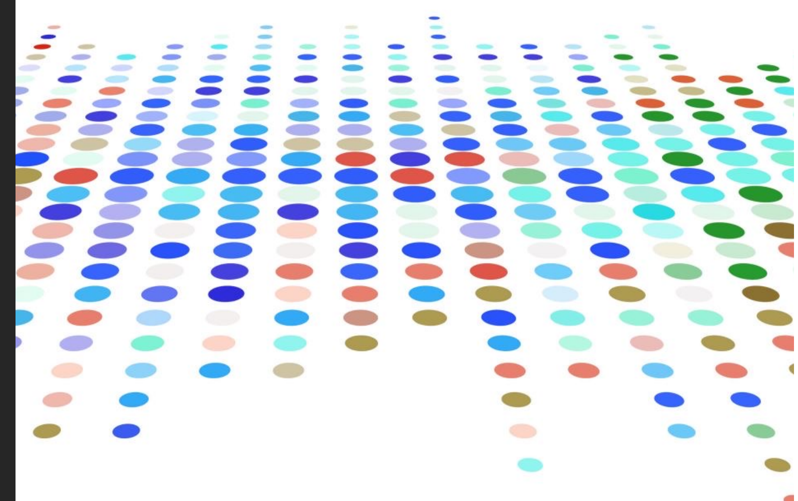
Limitations & Future Work

Limitations

- Online survey responses naturally carry a bias due to the fact that certain groups of individuals are not as tech savvy as others. As a result, the representation of these groups is limited.
- From a cultural perspective, some individuals (based on country) are less willing to be honest on surveys (regulations, etc).

Future Work

- Future work should include resource development for individuals in the 20-30 age category with an emphasis on college students. The specific demographic seems especially affected with respect to limited (or no) healthcare.



THANK YOU

QUESTIONS?

CONTACT: SANTANA2.MIGUEL@GMAIL.COM

ADDITIONAL PROJECTS CAN BE FOUND ON GITHUB

USERNAME: MIGUELANGELSANTANA