Humana/Mays Healthcare Analytics Case Competition

Modeling Covid Vaccine Hesitancy

October 10, 2021

Executive Summary

The current Covid-19 pandemic has provided unprecedented growth and acceleration across a multitude of industries yet its adverse impact will be felt across humanity for generations to come. This paper delves into vaccine hesitancy regarding Covid-19 and recommends potential solutions to approach this nuanced issue within society. With this in mind, our goal is to develop a classification model to predict vaccine hesitancy across society that is representative of society. In addition, our model takes artificial intelligence fairness, equity, and inclusivity into consideration as key tenets for analysis.

Firstly, we identified the target variable *'covid_vaccination'* (1 if the member is vaccinated and 0 if the member is not) and discovered that the dataset was imbalanced. Consequently, we cleaned the data, explored a few sampling techniques, and proceeded to use weights to balance the data.

Second, we quickly iterated different classification models and found ensemble models have better predictive performance. Through further hyperparameter fine-tuning and evaluation, LightGBM model with a set of optimal hyperparameters yielded the best accuracy with Area Under Curve (AUC) of 0.6869 and the best AI Fairness with a Disparity Score of 0.9988.

Then, we categorized the top features from our predictive model into demographic features, financial conditions, lifestyle features, and health condition. These categorizations help generalize significant features at a higher level and derive insights from there. The age of an individual provided significant insights into whether that individual is likely to get the vaccine or not. Furthermore, lifestyle is informative with regards to the type of lifestyle a person might use. For example, whether that individual spends less money on fast food, takes medication on-time, or does not have hyperlipidemia. Financial conditions allude to the wealth an individual might have. An individual with less wealth would have more barriers to getting the vaccine, such as transportation or other time-commitments.

In response to these discoveries, we recommend that Humana forms strategic partnerships with key community stakeholders, increases vaccination awareness with representative, diverse, and generational messages, and expands health initiatives throughout diverse communities. A wholistic and comprehensive effort is required to reach the diverse members of the society we

live in. Critically, this is an issue that Humana cannot afford not to be a part of. Essentially, the cost of hospitalizations alone for a subset of the population provided result in exorbitant costs for all stakeholders involved. For example, for every person that Humana is able to help get the vaccine it would save in hospitalization costs, administrative expenses, and potential insurance fraud schemes.

In conclusion, we believe Humana is ideally placed to be a key stakeholder in combating this pandemic and our model will help Humana reduce costs by knowing which segments of the population to target with tailored strategies to ensure the most effectiveness. This would lessen advertising costs associated with the future outreaches that Humana will be engaging in.

Table of Contents

1. Background

Vaccine hesitancy is a significant concern in society; especially, with the current Covid-19 pandemic. The lingering pandemic poses concerns that will have far reaching consequences for years to come, let alone this current moment. With this in mind, it is important to target hesitant members or those members, of society, that are resistant to getting the vaccine. Conceivably, there are various reasons why someone might be opposed to the vaccine such as inaccurate or false information. This case explores the reasons why members of society might be hesitant to receive a Covid-19 vaccination shot and provides recommendations to improve the vaccination rate among vulnerable and underserved members in society. It is not a simple endeavor but something that requires the wisdom of the crowd to achieve. Furthermore, as acknowledged by this year's problem statement, 'existing disparities in health equity have become increasingly apparent during the vaccine to Covid-19.'

Humana is an important stakeholder in the Health Care industry and delivering actionable insights would contribute to the organization's value creation. A critical enhancement of this year's challenge included recognizing an algorithm's fairness score. In other words, the insights discovered through our analysis are representative and provide equitable solutions to everyone in society.

2. Data Preparation

2.1. Data Understanding

Our goal is to identify members most likely to be hesitant to the covid vaccine. The target variable is *'covid_vaccination'* which is a binary variable.

To perform our analysis, we were given a training dataset of 974,842 rows and 368 columns. The data was Humana MAD member level data with information provided on Covid-19 Vaccination status, medical claims features, pharmacy claims features, lab claims features, demographic and consumer data, credit data features, clinical Condition related features, CMS (The Centers for Medicare & Medicaid Services) Member data elements, and others.

The vaccination data is from the 3rd week of March 2020 to the 3rd week of March 2021. Demographic and RWJF features are collected as of July 2020.

Following are some of the data nuances that we observed during processing the data:

- The data was highly unbalanced with the target variable '*covid_vaccination'* in the ratio of 805,389 (83%) for class 0, i.e. people who do not get vaccination; and 169,453 (17%) for class 1, i.e. people who get vaccinated.
- The histogram of member age skewed to the left, indicating most members are seniors with an average age of 71.18. We assume that all members are eligible for getting Covid-19 Vaccination.
- According to Humana, the *'zip_cd'* feature is totally synthetic. In fact, more than 70% of zip code are invalid according to USPS database, the valid zip code could also be inaccurate. We understand this is to protect sensitive patient information, but it also prevented us from collecting useful geographical information and further improve our model.
- 27 columns have a variance of 0 indicating they will not influence the target variable.

- 148 columns have missing values. 74% '*lang_spoken_cd*' values and 64% '*mabh_seg*' values are missing. Other than that, the proportion of missing values is less than 32.1% for all other columns.
- 32 pairs of columns are highly correlated with an absolute correlation > 0.8.
- Transformation is needed for categorical variables and ordinal variables.

2.2. Data Preprocessing

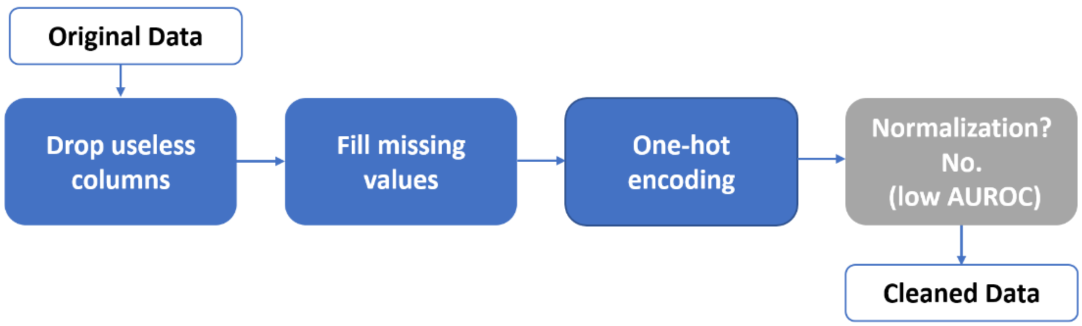

Figure 1: Data Preprocessing

We explored and extracted those feature columns where the variance was greater than zero. Columns that have a variance of zero do not influence the target features of '*transportation_issues*'. Hence, we removed the 27 feature columns, including '*auth_3mth_hospice*', '*auth_3mth_post_acute_rsk*', etc.

Next, we removed the columns that more than 70% values are missing, thus, '*lang_spoken_cd*' column was removed.

Then, we segregated the features as numerical, categorical, and ordinal features and handled each in a different way.

- For the numerical features, such as '*est_age*' and '*rwjf_uninsured_child_pct*', we replaced their missing values with the Mean value.
- For the categorical variables like '*sex_cd*' and '*race_cd*', we filled the missing values with the Mode and transformed them into numerical variables using One-hot encoding.

- For the 38 ordinal variables like *‘astotal_bh_copay_pmpm_cost_t_9-6-3m_b4’* and *‘mcc_ano_pmpm_ct_t_9-6-3m_b4’*, considering their order matters, we assign weights to objects to reflect a rank or ordering on an attribute.

Table 1: Ordinal Variable Values and Corresponding Assigned Values

Ordinal variable values	**Assigned value**
'New’	10
‘Inc_over_8x’	8
'Inc_4x-8x’,	6
'Inc_2x-4x’	3
Inc_1x-2x’	1.5
'No Activity’	0.5
'No_Change’	0
'Dec_1x-2x’	-1.5
'Dec_2x-4x’	-3
'Dec_4x-8x’,	-6
'Dec_over_8x’	-8
‘Resolved’	-10

Then we explored if data normalization is necessary. Comparing the prediction performance between normalized data and original data, we found that normalization almost always led to lower AUROC for tree-based models, thus, we decided not to normalize the dataset for all tree-based models. However, normalize for non-tree-based models is still necessary.

2.3. Data Sampling

Considering the data is unbalanced (83% for class 0 and 17% for class 1), we tried three different sampling methods to handle the unbalanced data (83% for class 0 and 17% for class 1). The results showed oversampling method led to lower prediction performance, the undersampling method would result in model overfitting which made model generalization more difficult. Weight balance during the hyperparameters tuning brought the best performance. Balancing the class weights is necessary for a better generalization of the model.

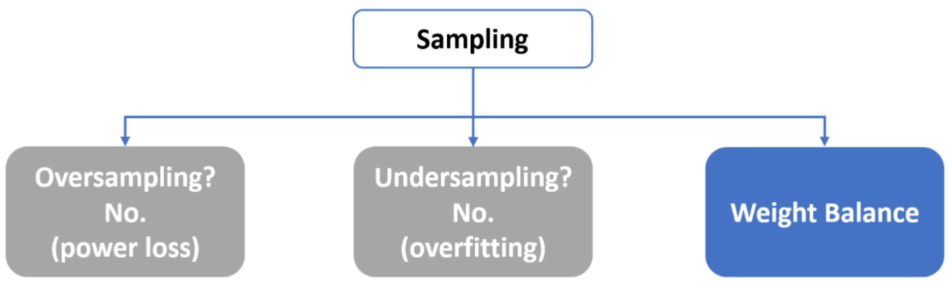

Figure 2: Sampling Methods Selection

2.4. Data Partition: Train, Validation, and Test Set

We partitioned data into 75% training data, 12.5% validation data, and 12.5% test data. Then we followed three steps for model building: model building, hyperparameter tuning, and model evaluation.

3. Modeling

Now that we get training data prepared, the next step was modeling. Since the problem is a binary classification problem, we were considering quickly iterating different models. We considered AUROC to measure prediction accuracy and Disparity Scores to evaluate algorithm fairness which will discuss more in section 3.2. Computational efficiency is also being taken into consideration.

3.1. Model Building

We explored various classification models, including Logistic Regression, Random Forest, Decision Tree, etc. The AUROC results showed the following four ensemble models have better performance compared to any of the constituent learning algorithms alone.

- **Random Forest**: Random Forest models are formed by a large number of uncorrelated decision trees, which joint together constitute an ensemble. In Random Forest, each decision tree makes its own prediction, and the overall model output is selected to be the prediction that appears most frequently.
- **Gradient Boosting:** Gradient Boosting trains many models in a gradual, additive and sequential manner. Subsequent trees help us to classify observations that are not well classified by the previous trees. Predictions of the final ensemble model are therefore the weighted sum of the predictions made by the previous tree models.
- **Extreme Gradient Boosting (XGBoost),** XGBoost is an efficient and high-speed implementation of Gradient Boosting framework that provides a parallel tree boosting that is useful for many machine learning problems.
- **Light Gradient Boosted Machine (LightGBM).** LightGBM is a gradient boosting framework that uses tree-based learning algorithms and is designed to be fast and efficient with lower memory usage, better accuracy, and capable of handling large-scale data through parallelization.

3.2. Model Fine Tuning and Evaluation

According to the previous analysis, we focused our study on these four models and manually chose some important model hyperparameters based on our experience to improve model performance. We then trained the model, evaluated its accuracy, and started the process again. This loop was repeated until a satisfactory accuracy was scored.

Then we used the AUROC to evaluate model accuracy since our problem is a binary classification problem, and applied Disparity Scores[1] to measure algorithm fairness. The higher the AUROC and the Disparity Scores, the better the performance of the model at distinguishing between the positive and negative classes. Specifically, AUC = 1 means the classifier perfectly distinguished all the Positive and the Negative cases correctly. Disparity Ratio = 1 means the predicted positive rate in each Race and Sex group is perfectly proportional to that of the ground truth positive rate.

The looping results indicate LightGBM outperforms other models due to higher AUROC, faster training speed, and higher efficiency. In terms of Disparity Ratio, all models have a score of around 0.98 which is pretty high and does not distinguish models. The optimal threshold for classifying members who are willing to get vaccinated and who are not is also calculated for each model and marked in the Area Under the ROC Curve plots. The values are all close to 0.5.

Table 2: Compare Four Ensemble Models Performance

Model	Training AUROC	Test AUROC	Disparity Ratio
Random Forest	0.8559	0.6676	0.9863
Gradient Boosting	0.7897	0.6774	0.9855
XGBoost	0.6909	0.6800	0.9826
LightGBM	0.6944	**0.6824**	**0.9810**

[1] For each group within a sensitive variable (RACE, SEX), disparity ratio DR is defined as: DR = S_n/S_0, Where S_n is the scoring metric (AUC, positive rate, etc) for each class and S_0 is the scoring metric for the reference group. The reference group is defined as the privileged group (i.e. for Medicare data Race: White, Sex: Male) within a sensitive variable class.

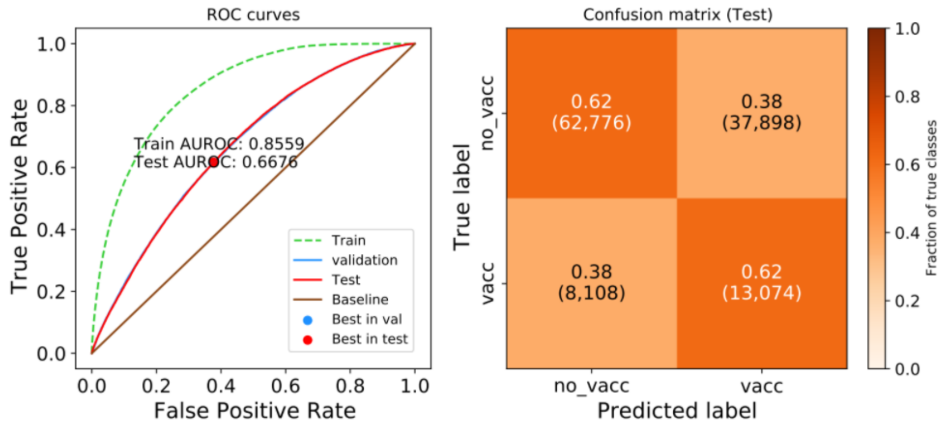

Figure 3: Random Forest ROC Curve and Confusion Matrix

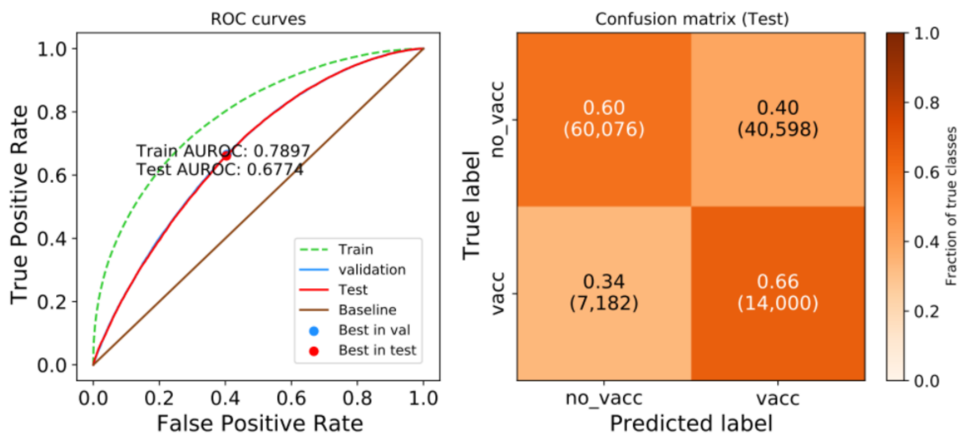

Figure 4: Gradient Boosting ROC Curve and Confusion Matrix

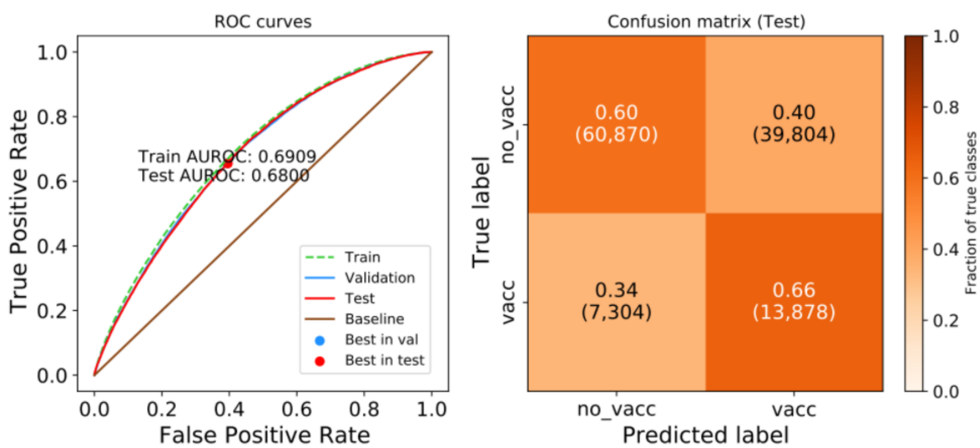

Figure 5: XGBoost ROC Curve and Confusion Matrix

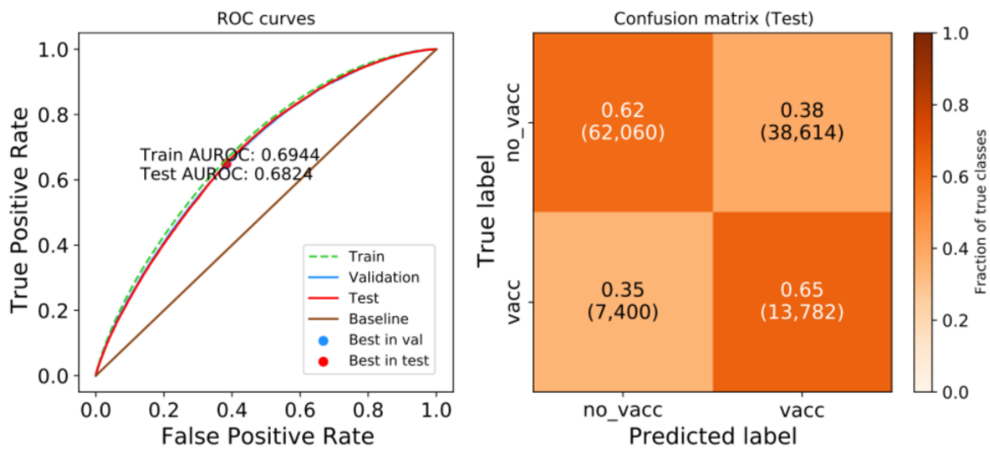

Figure 6: LightGBM ROC Curve and Confusion Matrix

Next, we used Optuna, an automatic hyperparameter optimization software framework that can help find the optimal hyperparameter combination, to improve LightGBM model performance further.

Figure 7: Model Selection Process

We tried more than 1000 combinations of hyperparameters, including 'boosting_type', 'num_boost_round', 'lambda_l1', 'lambda_l2', 'bagging_fraction', 'max_depth', 'num_leaves' 'feature_fraction', 'bagging_freq', 'min_data_in_leaf' and 'min_child_samples'. At last, we finalized our model with a set of optimal hyperparameters (listed in the Appendix). Comparing to manually tuning LightGBM models, the test AUROC and Disparity Ratio of our final model are both increased significantly. The test AUROC increase to 0.6869 compared to 0.6824 without significant overfitting. A 0.178 rise in the Disparity Ratio makes its value close to 1.

Table 3: Compare Model Tuning Method and Corresponding Model Performance

Model	Model Tuning Method	Training AUROC	Test AUROC	Disparity Ratio
LightGBM	Manually looping	0.6944	0.6824	0.9810
LightGBM	Optuna	0.6997	**0.6869**	**0.9988**

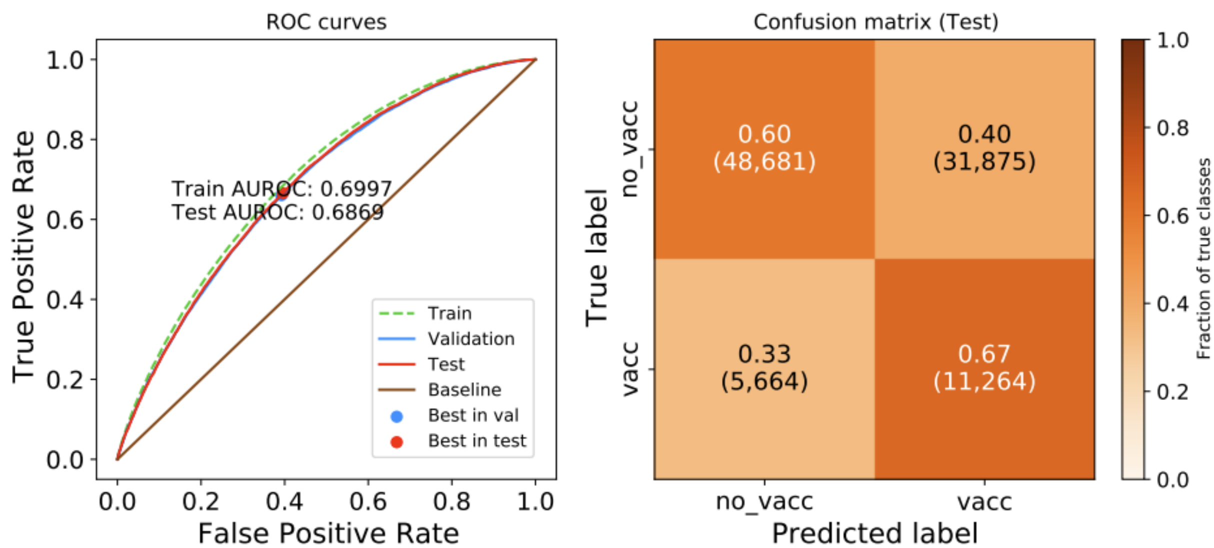

Figure 8: LightGBM with Optimal Hyperparameters ROC Curve and Confusion Matrix

4. Key Performance Indicators

In order to interpret the effects of the different features on the probability of getting vaccinations, we decided to make feature importance related plots using SHAP, a game-theoretic approach that can explain the output of any machine learning model.

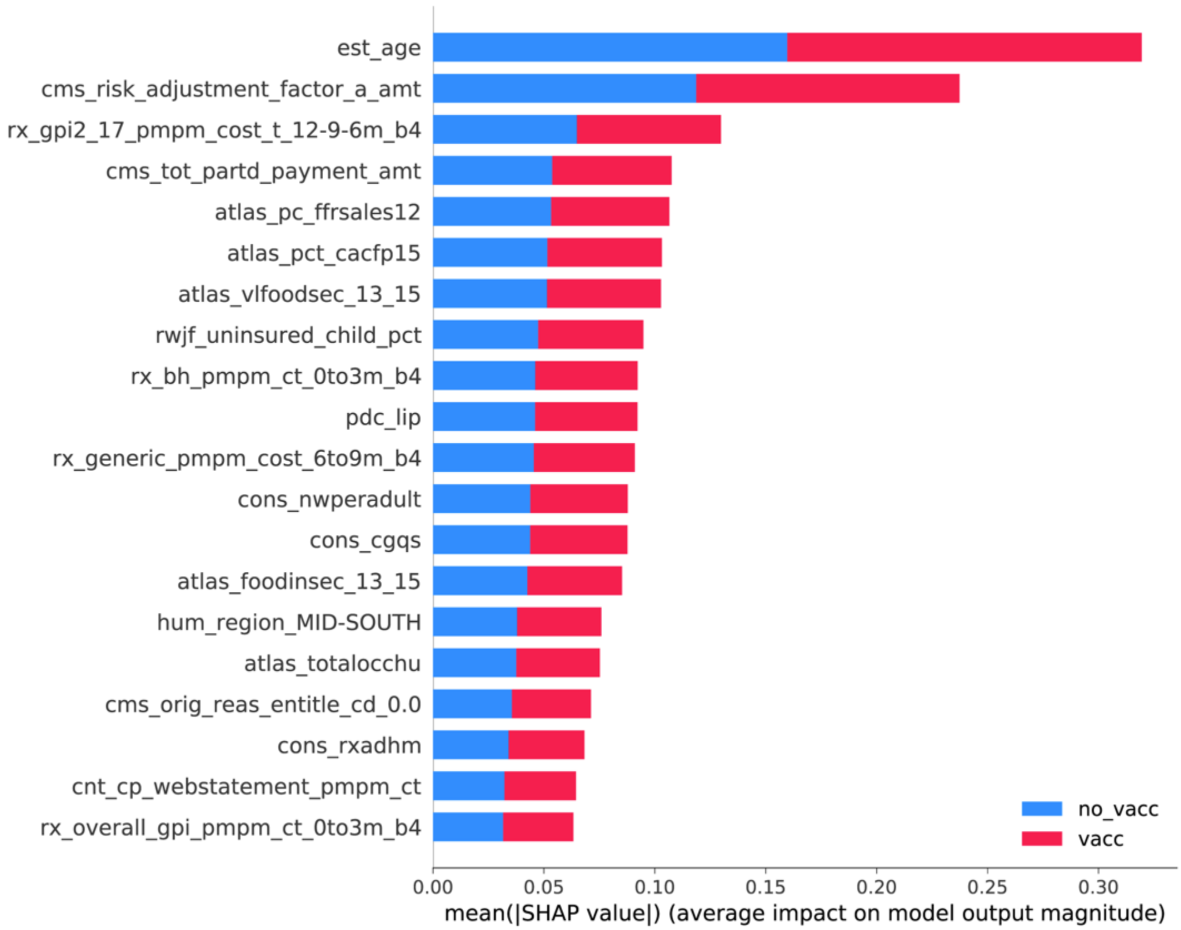

Figure 9: Top Twenty Features

In the SHAP summary plot, we found the twenty most influential variables ranked in descending order for our model in relation to the response variable *'covid_vaccination'*. The horizontal location shows whether the effect of that value is associated with a higher or lower prediction (higher prediction indicates that observation is more likely to get vaccinated). We categorized the top twenty features into four categories. This could be used to segment the data in a stratified way for further target analysis and recommendations in relation to the covid-19 vaccine.

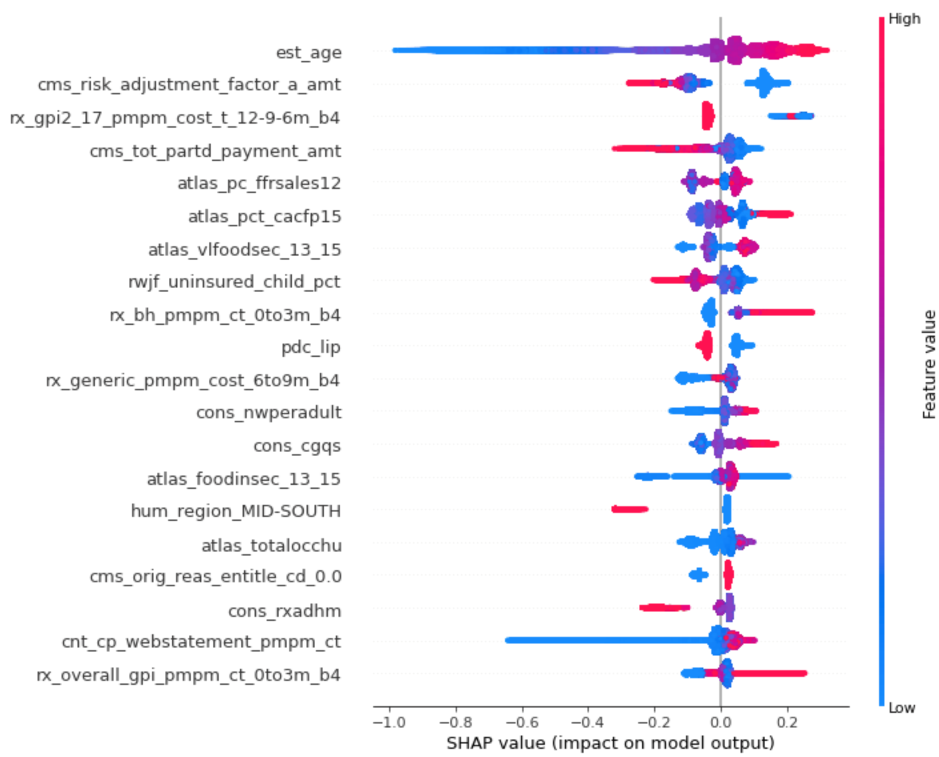

Figure 10: SHAP Summary Plot - Top 20 Features

4.1. Demographic Features

Figure 10 showed that *'est_age'* is the most important factor that influences the target variable. A high level (red part) of the *'est_age'* has a high and positive impact on the outcome, suggesting older people are more likely to get vaccinated. Seniors are among the most vulnerable groups of people in the pandemic so that they are eligible to get vaccines in the first phase. However, we cannot conclude younger people are more hesitant in getting vaccinations, because they may not be eligible to the vaccination at the data collection time.

4.2. Financial Condition

Variable *'Atlas_pct_cacfp15'*, *'rwjf_uninsured_child_pct'*, and *'cons_nwperadult*' are grouped into financial condition categories. These variables suggests that people that are less wealthy or have higher medical payments with Medicare programs are more hesitant in getting vaccinated. This observation is supported by New York Times, it indicates the Wealthy are getting more vaccinations, even in poorer neighborhoods.

The potential reasons can be the following: less wealthy people are busy with work so that they do not have enough time to wait for busy registration phone lines and navigate websites. They may lack transportation or time off from jobs to get to appointments.

Both variable *'atlas_vlfoodsec_13_15'* and *'atlas_foodinsec_13_15'* suggest individuals who experience less food insecurity have stability and might not be as concerned about where their next meal might come from. They are also more hesitant to get vaccines.

4.3. Lifestyle Features

Lifestyle gives a more comprehensive picture of the member and their respective community. *'Atlas_pc_ffrsales12', 'pdc_lip', 'cons_cgqs', 'cons_rxadhm'*, and suggest people have a healthier lifestyle. For example, spend less money on fast food, do not have hyperlipidemia, take medicines on time are having higher hesitancy getting vaccines. They may think it is unnecessary to get vaccines because they take good care of themselves so that the risk of getting covid is low.

Another inference we can draw from *'hum_region*' and *'atlas_totalocchu'* is people who live in less-density communities have lower willingness to get the vaccines. The interactions they have with others may be limited due to the isolation which makes them believe get vaccination is unnecessary.

Variable *'cnt_cp_webstatement_pmpm_ct'* indicates people with lower online behaviors are more hesitant to get vaccines. The barriers preventing members from receiving a vaccine could be a lack of sufficient information or access to the Covid vaccine.

4.4. Health Condition

Four variables, *'Cms_risk_adjustment_factor_a_amt', 'rx_gpi2_17_pmpm_cost_t_12-9-6m_b4', 'cms_tot_partd_payment_amt'* and *'cms_orig_reas_entitle_cd'* showed individuals with higher health risks, higher healthcare spending are more hesitant. People may worry about the side effects of the Covid-19 Vaccines. As many health practitioners have repeated, the risks of severe side effects from a vaccine are tiny in comparison to the risk of the disease itself. So, addressing conspiracy theories, misinformation, and disinformation through constant and consistent truth-telling and myth-busting public health campaigns is particularly important.

4.5. Key Observations

To conclude, we identified the following groups of people as our target audience:

- Less wealthy people with limited access to vaccines due to impropriate time, location, appointment failure, and concern about potential spending.
- People who live a healthy lifestyle are confident they are not risky under the virus attack.
- People who live in remote areas that have limited access to vaccine-related information and Covid vaccines.
- People who have preexisting diseases are worried about the side effects.

5. Recommendations

Essentially, this is an unprecedented crisis that is unlike any other and will remain that way unless a comprehensive wholistic plan is developed. The purpose of this case is to target unpresented and or marginalized communities across society, and it is our hope that this paper will lessen the steps required to reach that point. Vaccine hesitancy is a significant concern for Humana, and we believe that this problem is not Humana's to bear alone but society. Therefore, with the results of our model and independent research, we believe that tackling the problem of misinformation regarding the vaccination, increasing accessibility, and tailored messaging will be crucial in challenging vaccine hesitancy within society. Our recommendations will explore these points through the following categories finance, lifestyle, and health.

Consider the following hypothetical:

Given 974,842 records and Humana provides services to 40% of this population, which is approximately 389,937 patients. From this group, approximately 25% get Covid-19 requiring hospitalization for a period of time results in 97,485 patients that Humana serves visiting a hospital. According to Amin and Cox at the Health System's Tracker, who explored the hospitalization costs regarding Covid-19, discovered that the average hospitalization cost for Covid-19 to be $20,000 per patient. Therefore, extrapolating this cost out would result in a expenditures of $1,949.7 billion USD. Although, this is a simplistic calculation we believe that it sheds some light on the exorbitant costs that insurance providers might face as a result of individuals needing hospitalization on average. The average hospitalization $20,000 might not account for outlier cases where patients might require longer stays in hospital and future treatment because of Covid-19.

The lingering question is what does this mean for Humana? The current value proposition for Humana is cost and risk mitigation while building goodwill across society. If Humana spends $10 million or $100 million on tailored advertising or outreach campaigns the net gain is exponential to mitigate future losses or expenses as a result of a patient getting hospitalized. Also, the other part of the equation is that research has indicated that individuals who have received the vaccine and are hospitalized as a result have shorter hospital stays than if the individual is unvaccinated. This is critical to future cost savings.

5.1 Strategic Partnerships with Local Stakeholders

Our model highlighted socioeconomic inequality across society. Essentially, financial condition play a role in whether an individual will receive a vaccine or not. For example, barriers to vaccination could include the lack of transportation to vaccination sites, inability to take sick leave to attend vaccination appointments, difficulty in scheduling appointments, or difficult websites to navigate to schedule appointments. Of the features mentioned above food insecurity emphasizes the importance of accessibility for a member to potentially get the vaccine. The hesitancy might stem from job insecurity or lack of transportation to vaccination sites.

For those who lived in how density areas or have limited accessed to technology and information, partnership with local stakeholders is even more important for reaching out to those targets. We can organize special clinical day for those people to have easy access to information and vaccines.

This is an opportunity for Humana to create value in communities across society. For example, approaching community leaders in underrepresented communities on how the community members would be receptive to receiving the vaccine and places to hold future clinics to increase accessibility in these communities. Furthermore, engaging with community leaders fosters goodwill and develops community relationships that would prove pivotal to future business. Through discussions with diverse communities, Humana would be able to help health officials provide targeted and tailored information regarding vaccine acceptance and information. This would ensure that segments of society receive information in the means that they use for communication, whether this is the news, radio, community center, or social media. Society is interconnected in more ways than one and these positive messages are uplifting for communities.

5.2. Increase Vaccination Awareness

Our model also indicated that certain personal lifestyle factors influence a member's hesitancy to get the vaccine. For those people who live a healthy lifestyle are confident they are not risky under the virus attack, we believe that a solution to tackle vaccine hesitancy would be increasing the awareness and availability of the vaccine in these areas and expanding vaccination access at the community level through various services. We envision a comprehensive and personalized messaging campaign and social media ads to increase awareness and knowledge regarding the vaccine and its impacts on a person's health. These conversations require empathy as the impact of the pandemic has been disproportionate across society.

5.3. Expanding Health Initiatives Across Diverse Communities

One of the more critical initiatives would be challenging vaccine misinformation within society. This would include addressing conspiracy theories regarding the pandemic, especially for those people are too concerned about the side effect of Covid vaccine. For example, engaging in a truth-telling campaign where individuals in society can speak to educated health officials regarding side effects, vaccine efficacy or other concerns that they may have with regards to this vaccine. Furthermore, this is not a one-size fits all solution. Diverse communities need to be heard and feel seen during this discussion surrounding health regarding potential side effects to the vaccine. Therefore, messaging needs to be tailored with specific communities in mind to be most effective.

There are business opportunities for Humana given the pandemic. The value of grassroots initiatives to communicate with diverse community leaders will hopefully lead to decreased costs related to hospital bills if a larger portion of the population receives this vaccine. According to the FAIR Health, 'the average charge per Covid-19 patient requiring a hospital stay to be $73,300.' This would be for a 'patient seeing an-out-of-network provider whose health plan has no out-of-network benefit.' Humana would be indirectly affected as a result of this hospital stay. A patient covered by Humana might be impacted in other ways as a result of a patient being hospitalized for Covid-19. Also, given that a significant percentage of the dataset is hesitant to get the vaccine for various reasons, Humana is ideally placed to mitigate future costs by focusing resources on increasing vaccination rates across society.

Third, community initiatives enhance Humana's current investments into well-being and health initiatives across the United States. According to John Barger, Humana Medicaid president, 'Housing insecurity can disproportionally affect Medicaid-eligible patients and lead to inadequate health outcomes' (EastaBrook). This acknowledges the need for wholistic solutions when facing global challenges.

In conclusion, tailored incentives showed in Figure 11 for different members might increase the likelihood that an individual gets the vaccine. According to a survey conducted by the Kaiser Family Foundation in June 2021, individuals listed and recognized some of the incentives that would play a role in their likelihood to get the Covid-19 vaccine. It is important as this indicates a receptiveness to keeping an open mind.

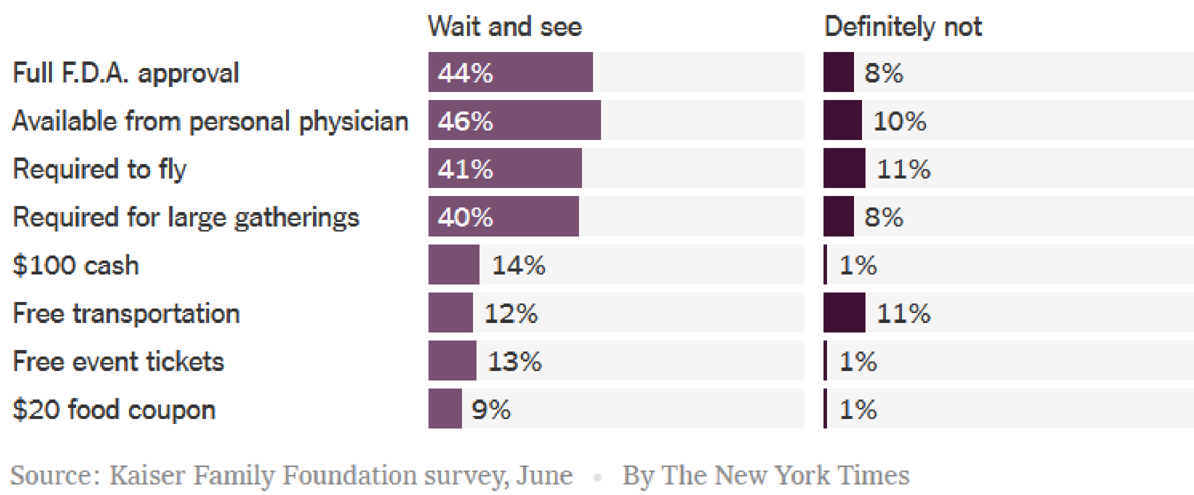

Figure 11: Vaccine Incentives (source: New York Times)

6. Conclusion

According to an interview conducted by the American Society for Microbiology with Shonta Chambers, the messaging for the Covid-19 vaccine needs to improve for underrepresented communities. Society has changed and traditional vaccination sites like clinics and hospitals might not be places where some minorities would feel comfortable receiving the vaccine. Perhaps, society needs to evolve and enable community leaders to discuss vaccination sites with health officials where the community is heard and seen in these discussions. Essentially, giving more communities a say in how they interact with government and health officials.

A reoccurring theme across our research suggests that 'availability does not equate to accessibility.' In other words, there are other factors such as income and lived experiences in relation to healthcare that could influence a member's hesitancy in relation to this vaccine. Therefore, we would recommend that Humana and other insurance companies establish grassroots efforts to play a role in decreasing the spread of misinformation in relation to vaccinations; alongside other initiatives that decrease the barriers or costs associated for minorities or underrepresented individuals to get the vaccine. Some of these costs might relate to subsidizing transportation costs, increased vaccine availability during grassroots initiatives and appropriately targeting individuals who might be living in remote areas. For instance, individuals who require financial assistance are hesitant to get the vaccine due to financial barriers. Therefore, healthcare officials need to reach out to communities in a way that these messages will be receptive to the targeted participant. Essentially, these initiatives should be tailored to the areas or segments of society that Humana wishes to target.

To conclude, the pandemic affects everyone across society yet, its impact is disproportional to different stakeholders within it. A collective effort is required to challenge the spread of misinformation and ensure that communities feel heard. We believe that Humana is ideally placed to align their Environmental, Social, and Governmental Targets with messaging related to Covid-19. The lasting impact for the business need not be financial. In this current market, the reputation such as goodwill will resonate for years to come in the communities that Humana reaches and makes a difference within.

7. Limitations

Given the manner in which the Covid-19 vaccine was distributed in society, we need to acknowledge that the dataset is imbalanced with regards to age, indicating our modeling may not generalize to younger groups of people. Different States in the United States had different roll-out plans for the vaccine which would possibly change the impact of age information within the dataset. Not to mention that young adults, ages 20-25, were part of the last age group that was eligible to get the vaccine.

Further constraints include policy restrictions or changes across different states for the period would influence people's willingness to take the vaccine.

Zip code information provided is synthetic thus cannot be converted to actual geographical locations. If accurate zip code information was provided it would allow for more targeted analysis.

8. Appendix

Hyperparameters	**Optimal Value**
boosting	gbdt
objective	binary
metric	auc
tree_learner	serial
device_type	gpu
num_iterations	10000
learning_rate	0.1
num_leaves	3
num_threads	0
deterministic	0
force_col_wise	1
force_row_wise	0
histogram_pool_size	-1
max_depth	11
min_data_in_leaf	6420
min_sum_hessian_in_leaf	0.001
bagging_fraction	0.719156
pos_bagging_fraction	1
neg_bagging_fraction	1
bagging_freq	1
bagging_seed	3
feature_fraction	0.508113
feature_fraction_bynode	1
feature_fraction_seed	2
extra_trees	0
extra_seed	6
early_stopping_round	100
first_metric_only	0
max_delta_step	0

lambda_l1	7.62E-05
lambda_l2	0.000357594
linear_lambda	0
min_gain_to_split	0
drop_rate	0.1
max_drop	50
skip_drop	0.5
xgboost_dart_mode	0
uniform_drop	0
drop_seed	4
top_rate	0.2
other_rate	0.1
min_data_per_group	100
max_cat_threshold	32
cat_l2	10
cat_smooth	10
max_cat_to_onehot	4
top_k	20
monotone_constraints_method	basic
monotone_penalty	0
refit_decay_rate	0.9
cegb_tradeoff	1
cegb_penalty_split	0
path_smooth	0
verbosity	-1
saved_feature_importance_type	0
linear_tree	0
max_bin	255
min_data_in_bin	3
bin_construct_sample_cnt	200000
data_random_seed	1

is_enable_sparse	1
enable_bundle	1
use_missing	1
zero_as_missing	0
feature_pre_filter	1
pre_partition	0
two_round	0
header	0
precise_float_parser	0
objective_seed	5
num_class	1
is_unbalance	1
scale_pos_weight	1
sigmoid	1
boost_from_average	1
reg_sqrt	0
alpha	0.9
fair_c	1
poisson_max_delta_step	0.7
tweedie_variance_power	1.5
lambdarank_truncation_level	30
lambdarank_norm	1
multi_error_top_k	1
num_machines	1
local_listen_port	12400
time_out	120
machines	
gpu_platform_id	0
gpu_device_id	0
gpu_use_dp	0
num_gpu	1

9. References

1. Diane E., “Humana’s New $25M Investment Could Help More Seniors Age in Place”, *McKnights Senior Living,* Sep. 13, 2021 https://www.mcknightsseniorliving.com/home/news/home-care-daily-news/humanas-new- 25m-investment-could-help-more-seniors-age-in-place/
2. David R., “Why Some People Don’t Want a Covid-19 Vaccine”, *BBC,* Jul. 22, 2021 https://www.bbc.com/future/article/20210720-the-complexities-of-vaccine-hesitancy
3. Julie B., Jan H., Margot S.K., Tim A., “Who Are the Unvaccinated in America? There’s No One Answer.”, *The New York Times,* Aug. 4, 2021 https://www.nytimes.com/2021/07/31/us/virus-unvaccinated-americans.html
4. “Ways Health Department Can Help Increase Covid-19 Vaccinations”, *Centers for Disease Control and Prevention,* Aug. 16, 2021 https://www.cdc.gov/vaccines/covid-19/health-departments/generate-vaccinations.html
5. Florian F. S., Kirstin W. S., Wilson N., Kusum M., Lucy S., Sydney F., Nasma B., Alex S., Ashwin G., Rama A.S., “Identifying Strategies to Boost Covid-19 Vaccine Acceptance in the United States”, *RAND Corporation, 2021* https://www.rand.org/pubs/research_reports/RRA1446-1.html
6. Anna M., “What Are the Most Effective Ways to Improve Vaccination Rates?”, *GAVI The Vaccine Alliance*, May 5, 2021 https://www.gavi.org/vaccineswork/what-are-most-effective-ways-improve-vaccination-rates
7. “Child and Adult Care Food Program”, *Food and Nutrition Service* https://www.fns.usda.gov/cacfp
8. “Covid-19 Vaccinations in Underrepresented Communities”, *American Society for Microbiology,* Feb. 22, 2021 https://asm.org/Articles/2021/February/Covid-19-Vaccinations-in-Underrepresented-Communit

9. "Explain Your Model with the SHAP Values", *Towards Data Science,* Sep. 13, 2019 https://towardsdatascience.com/explain-your-model-with-the-shap-values-bc36aac4de3d
10. "Medicare Risk Adjustment", *Foresee Medical* https://www.foreseemed.com/medicare-risk-adjustment
11. Juliette C., Anthony D., Tricia N., "10 Things to Know about Medicare Part D coverage and Costs in 2019", *KFF,* Jun. 04, 2019 https://www.kff.org/medicare/issue-brief/10-things-to-know-about-medicare-part-d-coverage-and-costs-in-2019/
12. "Geographic Overview", *Census Bureau* https://www2.census.gov/geo/pdfs/reference/GARM/Ch2GARM.pdf
13. "Food Security in the U.S.", *Economic Research Service U.S. Department of Agriculture* https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/
14. Tim K., "In-Depth: The Cost of Covid-19, Health Insurance, and Hospitalization", *ABC Action News,* Aug. 26, 2021 https://www.abcactionnews.com/news/coronavirus/in-depth-the-cost-of-covid-19-health-insurance-and-hospitalization
15. "Costs for a Hospital Stay for Covid-19", *Fair Health, Apr. 16, 2020* https://www.fairhealth.org/article/costs-for-a-hospital-stay-for-covid-19
16. Gary L., "Tuning Hyperparameters with Optuna", *Towards Data Science*, Mar. 08 https://towardsdatascience.com/tuning-hyperparameters-with-optuna-af342facc549
17. Alvira S., "CatBoost vs. Light GBM vs. XGBoost", *Towards Data Science*, Mar. 13, 2018 https://towardsdatascience.com/catboost-vs-light-gbm-vs-xgboost-5f93620723db
18. Pranjal K., "Which Algorithm Takes the Crown: Light GBM vs XGBOOST?", *Analytics Vidhya,* Jun. 12, 2017 https://www.analyticsvidhya.com/blog/2017/06/which-algorithm-takes-the-crown-light-gbm-vs-xgboost/
19. Krutika A., Cynthia C., "Unvaccinated Covid-19 Hospitalizations Cost Billions of Dollars", Sep. 14, 2021

https://www.healthsystemtracker.org/brief/unvaccinated-covid-patients-cost-the-u-s-health- system-billions-of-dollars/