Humana-Mays Healthcare Analytics

2022 Case Competition

Predicting Members with Housing Insecurity

&

Proposing Solutions and Recommendations

Table of Contents

1. INTRODUCTION

1.1 Background

Housing insecurity encompasses several dimensions of housing problems people may experience, including affordability, safety, quality, insecurity, and loss of housing. Housing Insecurity in all its forms can negatively impact human health making it harder to access health care.

In a study that analysed data from the 2011 Washington State Behaviour Risk Factor Surveillance System, approximately one third of the housing insecure respondents reported delaying doctor visits due to the costs. 26.9% of the housing insecure respondents were current smokers and 26.3% had poor or fair health. It may happen due to high housing costs relative to income, poor housing quality, unstable neighbourhoods, overcrowding, and, but may not include homelessness.[2] Due to a limited rental market with few affordable vacancies, people with the lowest incomes may be forced to rent substandard housing that exposes them to health and safety risks, such as vermin, mold, water leaks, and inadequate heating or cooling systems. They may also be forced to move in with others, potentially resulting in overcrowding. It affects mental health, stress levels, relationships, sleep, and it may increase the risk of infectious disease. Also, research has shown that renters who are forced to move are more likely to relocate to poorer and higher-crime neighbourhoods compared to those who move voluntarily. This increased stress levels have resulted in suicide rates having doubled between 2005 and 2010, when the United States experienced historically high rates of foreclosures, including foreclosures on rental properties.[1] When people move 3 or more times in a year, it is considered as ‘multiple moves’. Children who move frequently are more likely to have chronic conditions and poor physical health. They may also be less likely to have consistent health insurance coverage.

These issues can be resolved by creating awareness among the people about the potential problems as listed above. Also, increasingly, state Medicaid agencies are focusing on addressing housing-related needs of their enrolees through their managed care Contracts.[3] Non-profit organizations, health insurance providers along with government are together offering measures to reduce housing insecurity by monitoring existing housing quality, working with owners to address code violations, supporting redevelopment and affordable housing of the distressed public, helping older homeowners make necessary repairs and modifications, providing coverage mindfulness practises and therapy, etc.

1.2 The Business Problem

Humana is a leading healthcare company that offers a wide array of insurance products and health and wellness services. It serves around 17.1 million members nationwide. The aim of this analysis is to help Humana to identify Medicare members most likely to be struggling with housing insecurity issues and establish key indicators to inform Humana's business decisions.

To understand the magnitude of the problem that Humana is tackling, it is worthwhile to look at a few statistics. Around 37.1 million American households are "Housing cost-burdened," and 32.7% of older adult households have severe housing problems (as provided by Humana). It is reported that America's precarious housing situation might cost the healthcare around ~$ 111 billion [as of July 2022] in next 10 years which makes it extremely pertinent for this issue to be tackled. Resulting health problems can range from allergies to neurological to heart damage. Therefore, it is necessary for advancements to be made to proactively identify individuals who are at a higher risk of housing insecurity for tracking and appropriate servicing. This will not only help improve the well-being of the members but also reduce the cost burden on the individual as well as Humana.

1.3 Key Performance Indicators

We have identified the following key performance indicators to evaluate the business problem:

i. **Health Outcomes**: We have designed solutions keeping in mind Humana's Bold Goal to improve patient's access to healthcare and improve health outcomes.

ii. **Cost of Care**: To ensure that the cost of care for members is affordable.

2. DATA PREPARATION

2.1. Analytics Tools Used

This project/case was realized using Python and implemented in Jupyter Notebooks. We used Pandas and Numpy libraries for the exploratory data analysis, feature engineering, and feature selection, Scikit-learn, and Keras for building and testing different machine learning models, and Matplotlib for visualization.

2.2. Exploratory Data Analysis

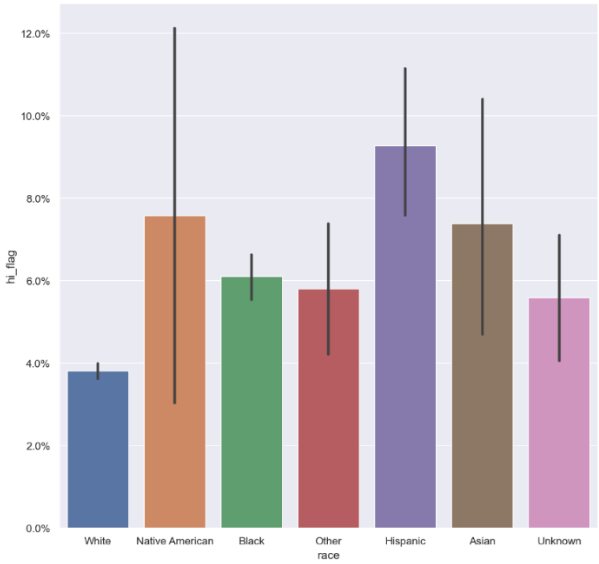
2.2.1 Understanding Variables

The training dataset had 880 features and 48300 data points. The features can be broadly classified into the following categories –

- Medical claims features
- Pharmacy claims features
- Demographic/Consumer data
- Credit features
- Clinical condition features
- CMS features
- County-level SDOH features
- Revenue features
- Outreach point features

Each feature belonged to one of the following data type classes – numeric, categorical, and binary indicators.

After the first round of quick high-level analysis of the data we performed an exploratory data analysis (EDA) to understand the data and find interesting patterns.
Figure 1 and Figure 2 show how percentage of members with housing insecurity are distributed across different demographic categories.



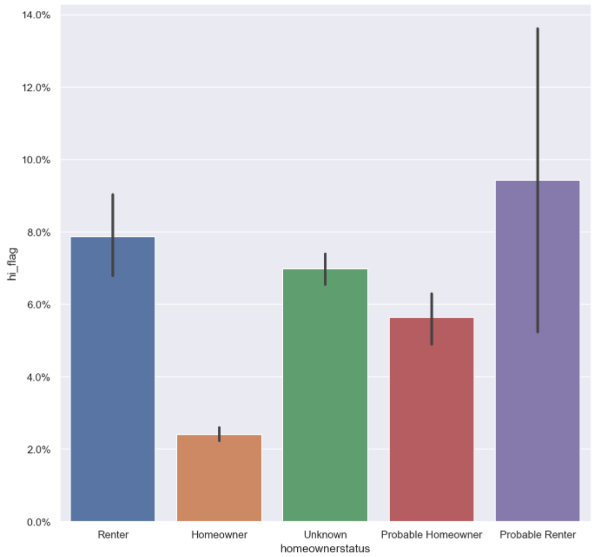
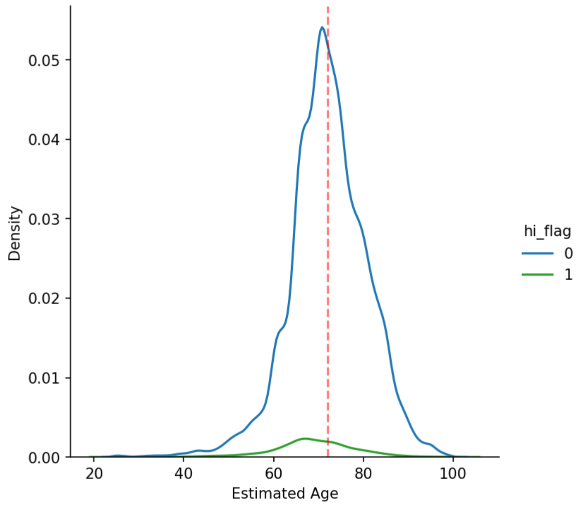
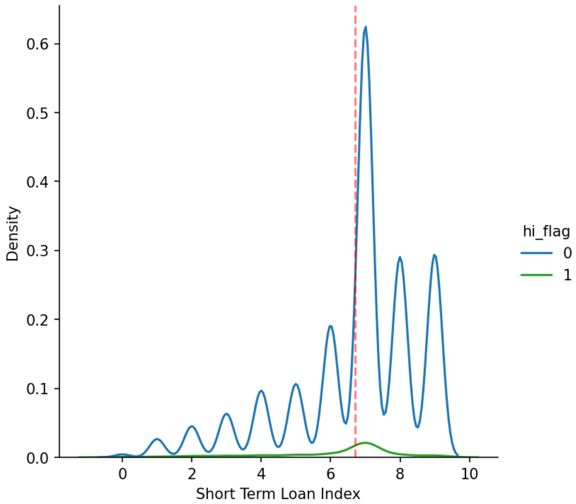


Figure 1 & Figure 2: Percentage of Members with Housing Insecurity across demographic categories

From Figure 1 & 2 it can be observed that Hispanic population has a higher proportion of members facing housing insecurity and that the members who are probably renting and members who are definitely renters have high proportion of members facing housing insecurity.

There are factors other than demographic factors that play an important role in understanding who is impacted by housing insecurity and what are their characteristics.





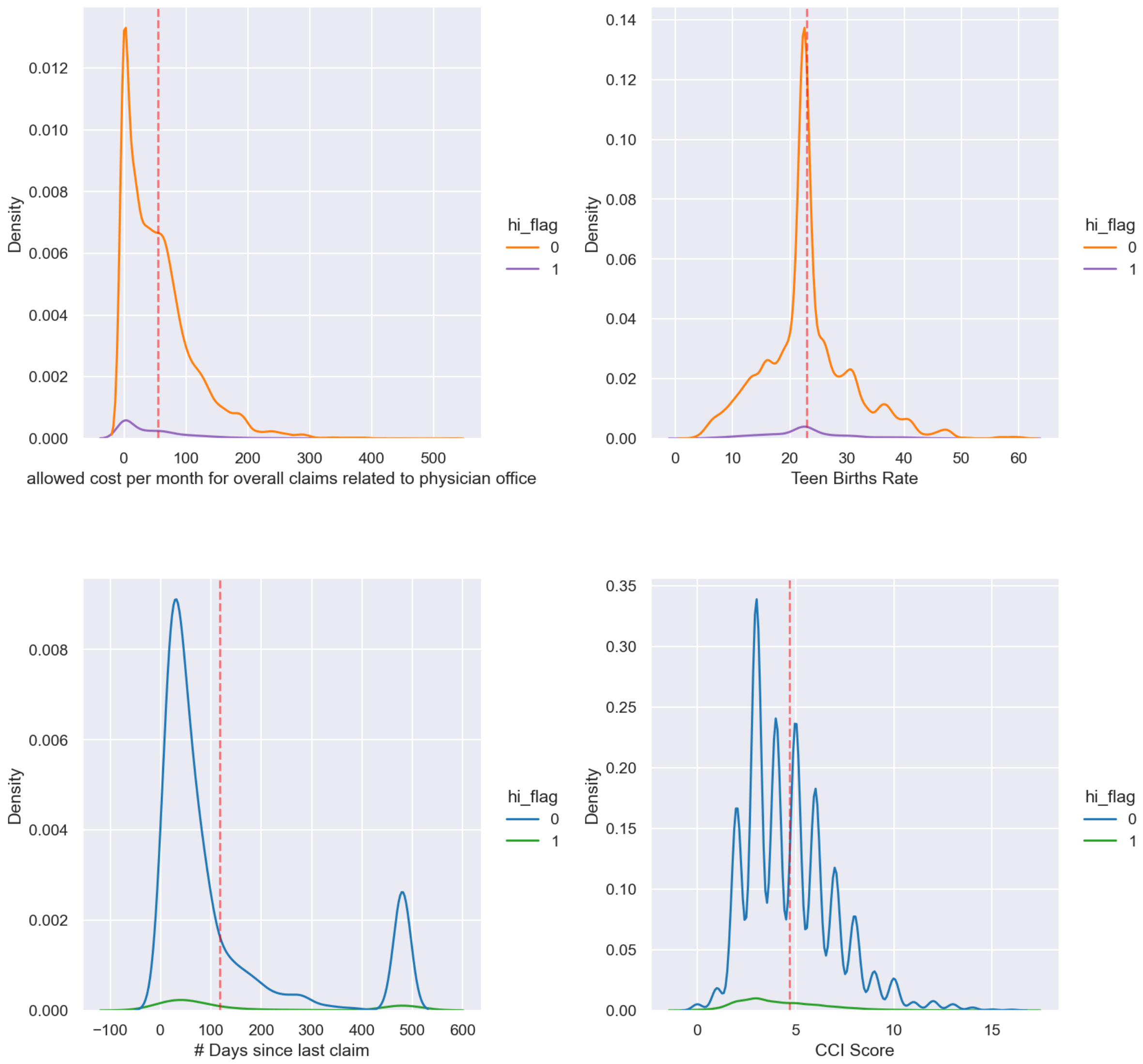


Figure 3: Distribution of members in dataset against various factors.

From the different distribution graphs in Figure 3, we can note that there is a difference between the characteristics of members who face housing insecurity and members who don't. We see that members who rank high on the Short-Term Loan Index are prone to housing insecurity. And those who have high # of Days since last claim.

Although we see a significant difference in patterns in the data for members with and without housing insecurity, it is difficult to pin point the underlying drivers that lead to it. Hence, using a machine learning model would be the best way to identify the same.

2.2.2 Data Cleaning

260 features of the 880 in the dataset had null values. Dropping these columns/rows was not a feasible solution as it would result in a very small dataset. To remedy this, we replaced the missing data with median of the respective column.

There were some categorical features which had wildcard symbols such as '*' which we attributed to the Unknown categories for the respective column.

In addition to this during the exploratory data analysis we noted that there are 221 features in the dataset that have zero variance i.e., all the values in the column are same. Using such features for model training will only increase the size of the dataset and thereby increasing the training time while adding no incremental value to the model performance. So, we have gone ahead and dropped these features from the dataset.

2.3. Feature Engineering

2.3.1 One-Hot Encoding

A machine can only understand numbers and cannot deal with text directly. In Humana's dataset there are several such variables/features like – sex, language spoken, race, homeowner status and different risk factors just to name a few.

One-hot encoding is a technique of translating categorical features into a sequence of binary indicators in order to register the same information in a numerical manner.

For instance, a sex feature having two types of entries (M, F) would be translated into two different columns: one binary indicator (1s and 0's) to represent the Male population and another to represent the Female population (Figure 4).

SEX_CD
M
F
M
M
F
F

SEX_CD_M	SEX_CD_F
1	0
0	1
1	0
1	0
0	1
0	1

Figure 4. Categorical Feature pre and post one-hot encoding.

One drawback of using this technique is expansion of the feature list, one new feature to represent all unique values of all categorical variables. This results in dataset that is very sparse in nature leading to some performance issues.
In Humana's dataset there exist a total of 8 categorical variables which were translated into a total of new 24 binary indicators.

2.4. Feature Selection

We noticed that the credit related data features had null values for more than 90% of the members. So, initially we decided to drop these columns for model training because we believed that they would not improve the performance significantly.
But after some background research on housing insecurity, we understood that credit data played a significant role in determining housing insecurity outcomes. As such, we looked at Pearson correlation between the credit features and the target variables and/or other features. We concluded that credit related information had a considerable correlation with the target variable and thus we have finally decided to go ahead and use it in the final model training. Additionally, we looked at the correlation between all the features and the target variable, some of which are visualized in Figure 5.

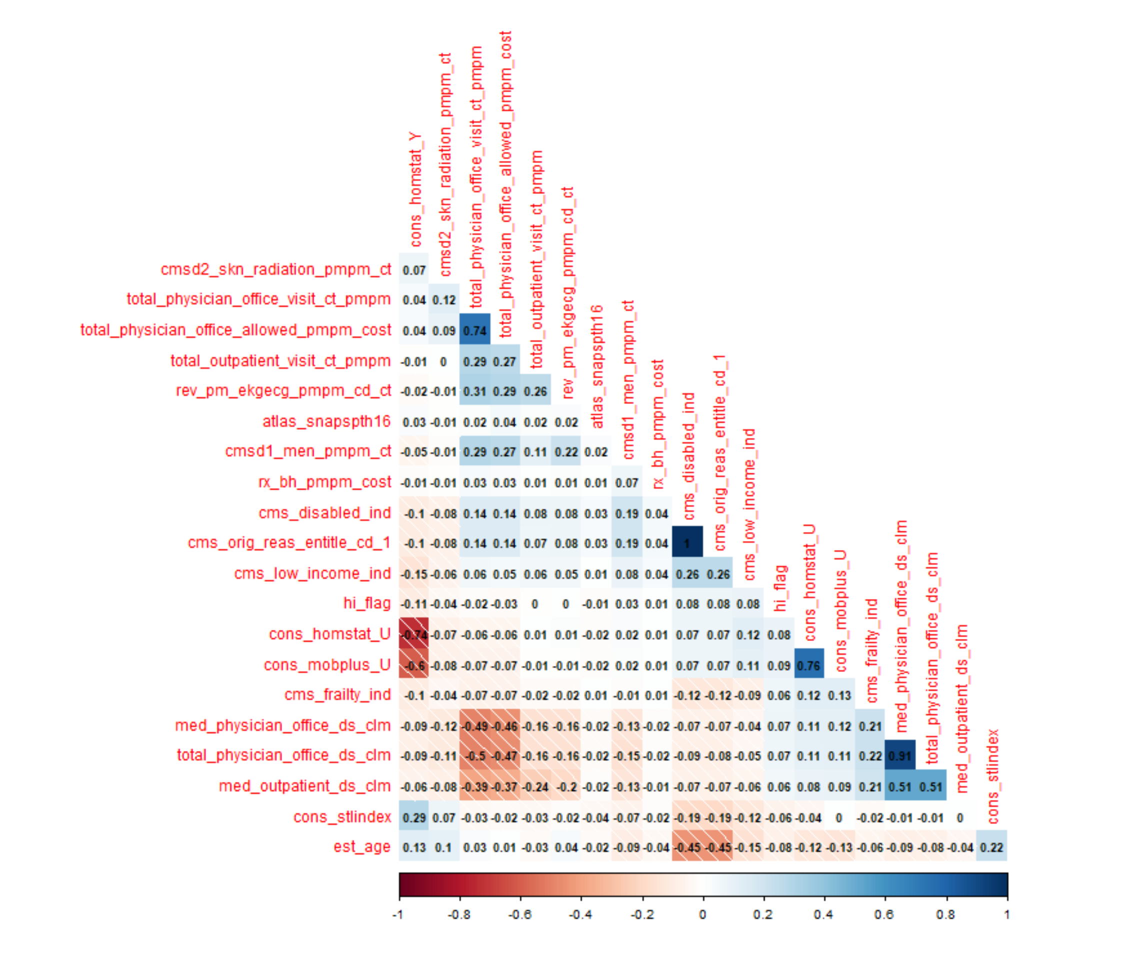


Figure 5: Correlation Heatmap for highly correlated features.

From the correlation matrix we observed that there are combinations of features that have very high correlations, including both does not add any incremental value to the model/analysis. For instance –

- atlas_totalpopacs (Total population, 5-year average) and atlas_totalpopest2016 (Population Size) are highly correlated (correlation coefficient is greater than 0.99).
- And so are atlas_totalocchu (Total number of occupied housing units) and rwjf_population (Demographics – Population) (correlation coefficient is greater than 0.98).

So, to avoid multicollinearity and/or use of highly correlated features, we set a threshold of 0.9 on the correlation coefficient. We have dropped one feature from any pair of correlated features having coefficient greater than 0.9.

3. MODELING

3.1. Model Selection & Performance Evaluation

We have looked at a total of 6 different binary classification algorithms: Logistic Regression, SVM, KNN, Random Forests, Neural Networks, and Gradient Boosting Trees, specifically XGBOOST. We assess every model by calculating probabilities of housing insecurity for each member and then calculating AUC scores using the labelled training data.
Because the dataset is relatively small, we decided to use a 5-fold Stratified Cross Validation to maximize training data. And hence we compare the mean AUC across the 5 folds across to different models. Stratified Cross Validation ensures that there is an equal balance of positive and negative target variables across all folds.

We use baseline models of each classification algorithm and look at the performance of their “vanilla” versions before engaging in hyper parameter tuning. SVM, KNN, and Neural Networks face some training issues due to the high sparse nature of the dataset and hence result in sub-standard AUC scores. The Logistic Regression, Random Forest and XGBOOST models are thus selected for hyperparameter tuning.

3.2. Hyperparameter Tuning

For each model, a Grid Search algorithm is used where a wide variety of different parameters are used to train the model. This algorithm systematically uses all the combinations of the specified parameters to train the model.
Once a set of broad ranges of the parameters were narrowed down. We then used another round of Grid Search algorithm to narrow down a much more granular set of potential parameters.

Finally, a selection of parameters was then used for 5-fold Stratified Cross-Validation finalizing the parameters. The optimized hyperparameters and AUC scores for each model can be found in Table 1.

Model	Tuned Hyperparameters	Average AUC Score ± SD (5-fold Stratified CV)
Logistic Regression	solver = 'liblinear' class_weight = 'balanced' C = 4	0.734 ± 0.01
Random Forest	n_estimators = 1000 max_features = 'sqrt' max_depth = 3 criterion = 'gini' class_weight = {0: 1, 1: 21} n_jobs = -1	0.719 ± 0.008
XGBOOST	learning_rate = 0.01 n_estimators = 2500 max_depth = 2 scale_pos_weight = 10 max_delta_step = 1 objective = 'binary:logistic' alpha = 0 reg_lambda = 0.2 min_child_weight = 1 gamma = 0 subsample = 0.9 colsample_bytree = 0.3	0.749 ± 0.007

Table 1: Hyperparameters and AUC scores of Tuned Models.

The ROC Curves for the mean of 5-fold Stratified CV AUC scores mentioned in Table 1 are visualized in Figure 6.

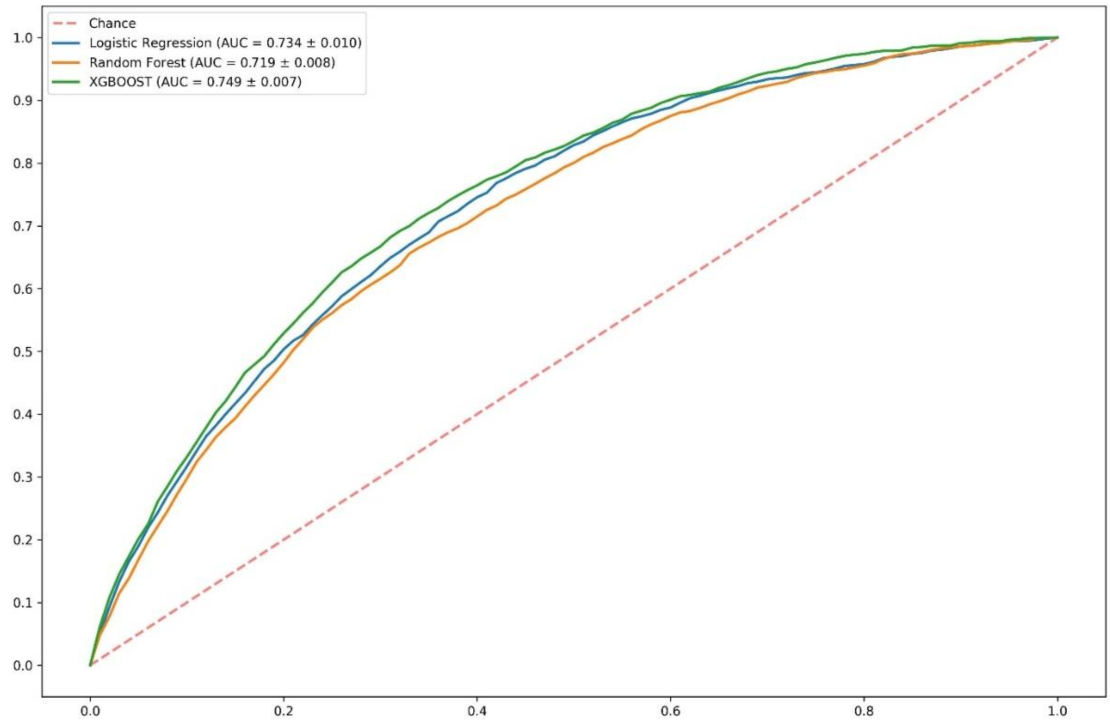


Figure 6: Mean ROC Curve for Fine Tuned Models for Comparison

3.3. Final Model

The XBOOST model was identified as the best performing model and was thus selected for predicting housing insecurity outcomes for the holdout data. The ROC curves for the 5-fold Stratified CV with their respective AUC scores are visualized in Figure 7.

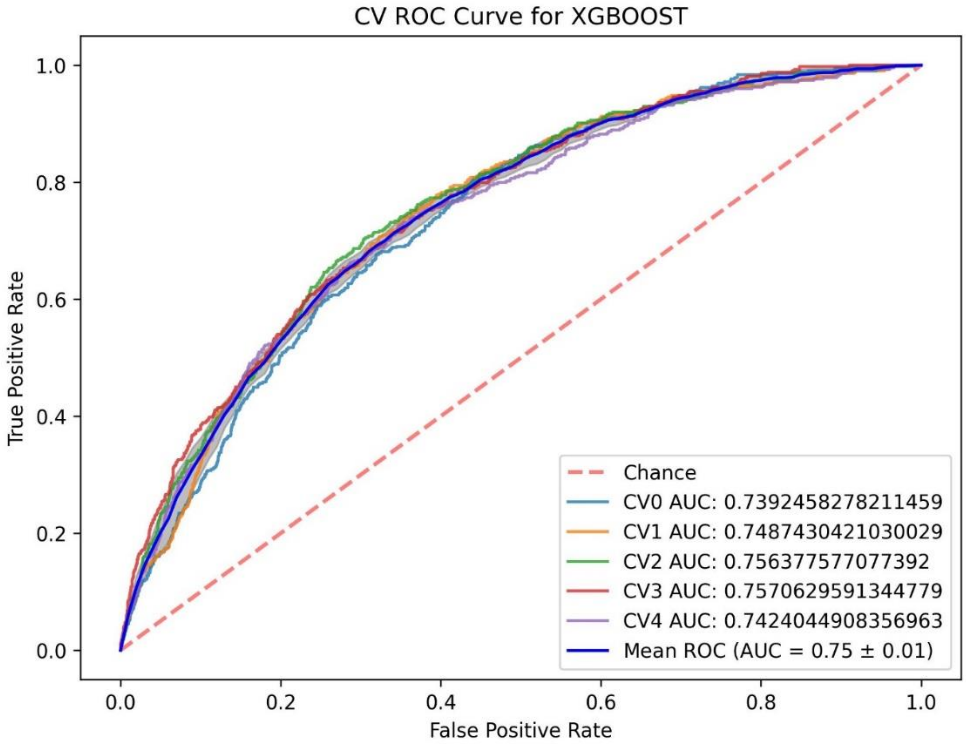


Figure 7: ROC Curves for 5-fold Stratified CV for XGBOOST

The Confusion Matrix for the best performing CV Fold for XGBOOST is as show in Figure 8.

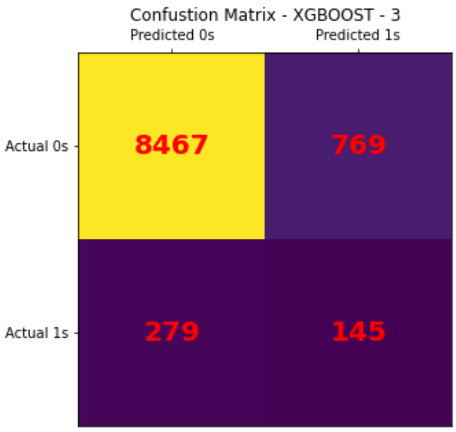


Figure 8: Confusion Matrix for best performing CV fold for tuned XGBOOST Model

3.4. Post-Modelling Analysis

In order to extract maximum value and actionable insights from the Machine Learning model it is essential to understand which features used to train the model are more important and have higher impact on housing insecurity. To do so, we have plotted the top 20 influential features (as per XGBOOST model) based on SHAP value in Figure 9.

A feature having a higher SHAP value has a greater influence on the overall output of the target variable.

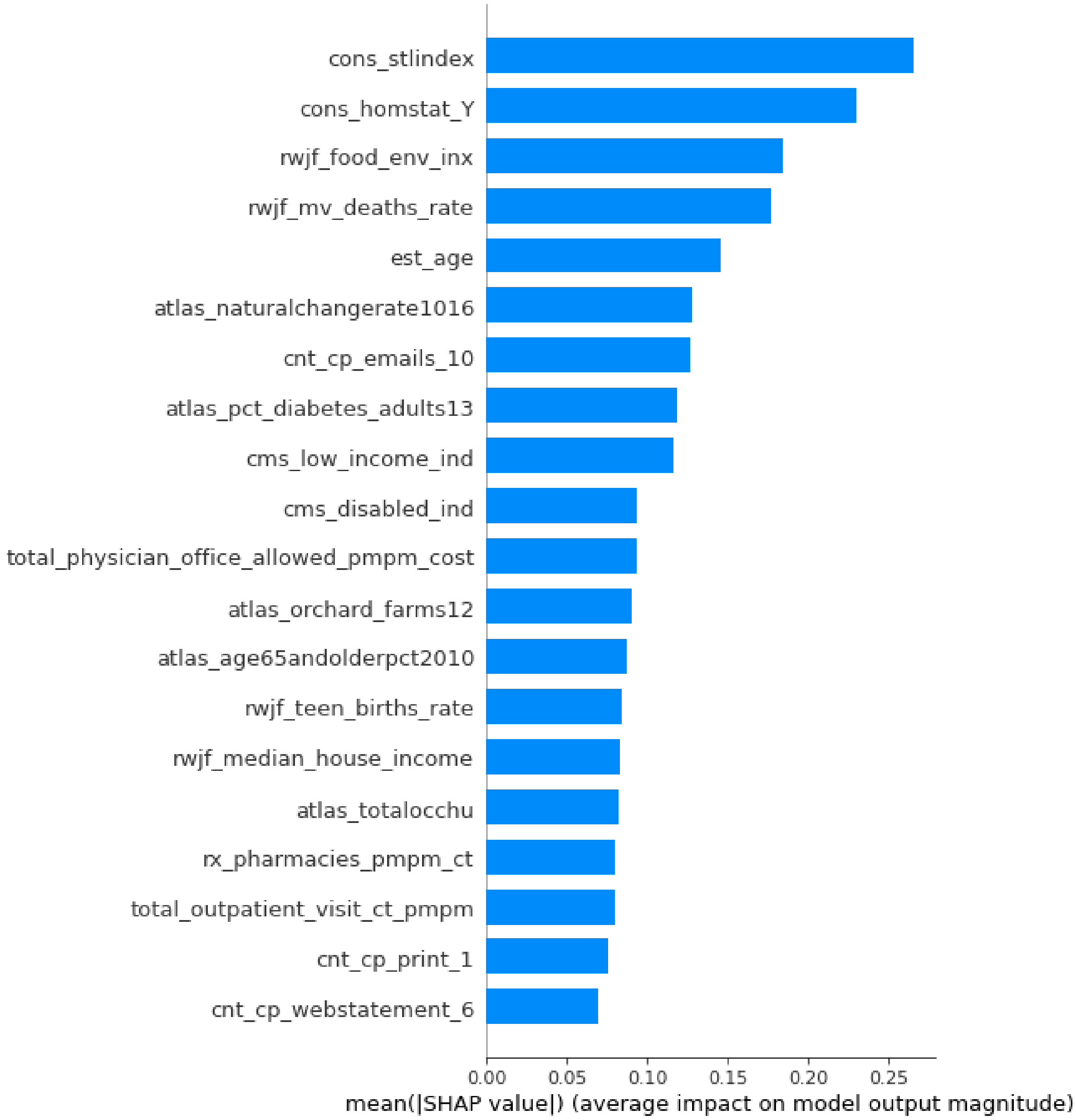


Figure 9: Top 20 features in the tune XGBOOST Model

These top 20 features are classified into broader categories: homeowner status, disability, age, hospital visit history – outpatient and physician office visits, financial situation and mental health/non-mental health related claims.

The SHAP value of a feature also varies with the relative value of that feature. For instance, in Figure 10 cms_low_income_ind i.e., the Low Income Index has a higher SHAP value for higher index value. Which means that higher the value of low income index higher is the impact of the variable on the housing insecurity outcome. And the concentration of the markers in the image represents the number of members meeting that criterion. For instance, cons_homstat_Y which is a homeownership binary indicator has a bimodal distribution, lower the value (0 meaning do not own a home) higher is the SHAP value i.e., higher is the impact on housing insecurity outcome and higher the value (1 meaning owns a home) lower is the impact on the outcome.

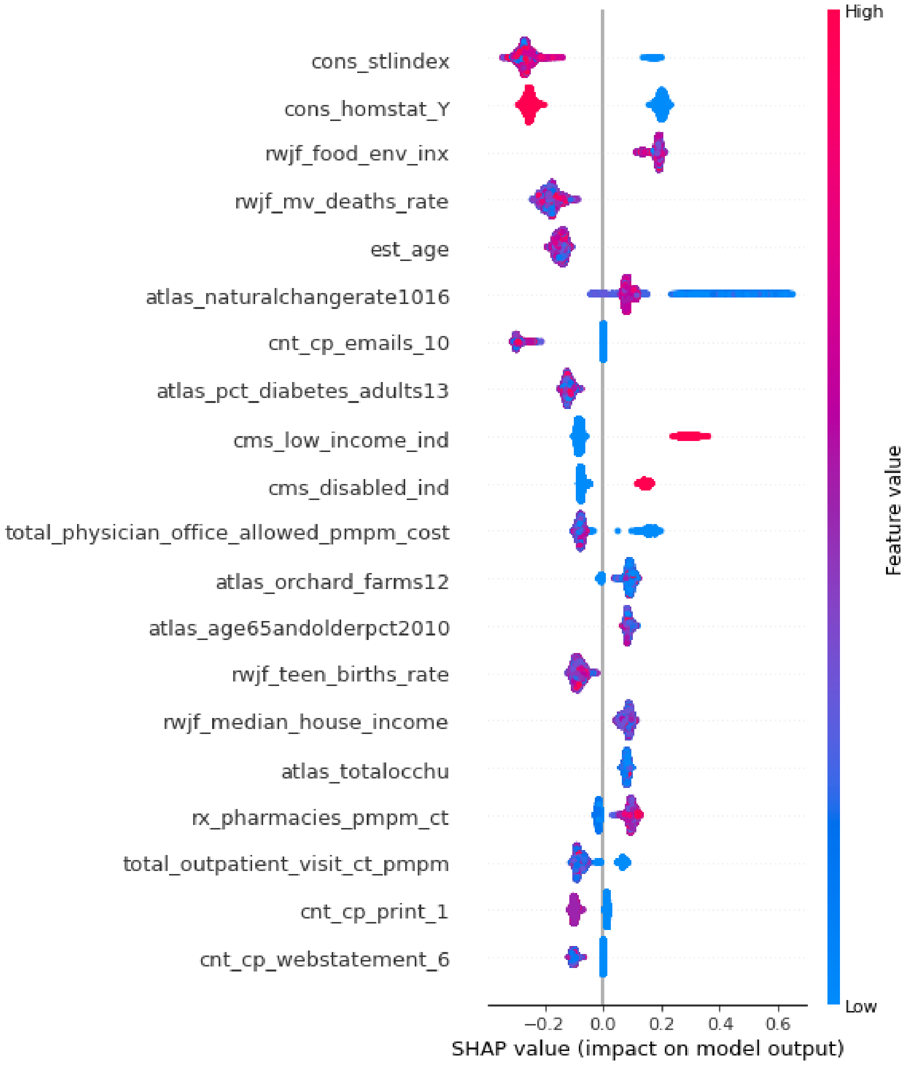


Figure 10: SHAP value as a function of relative feature value for top 20 important features.

4. RECOMMENDATIONS & ACTIONABLE INSIGHTS

4.1. Contributors of Housing Insecurity

After identifying the 20 top features of the predictive model, we wanted to analyse further how these features explained housing insecurity among Humana MAPD members. For this, we decided to perform Principal Component Analysis (PCA) to group the features into condensed categories. The principal components explain 54.83% of the whole data and allowed us to understand the different characters that make up the entire feature list that impact the housing insecurity outcomes and propose solutions that could help improve housing conditions and reduce housing insecurity.

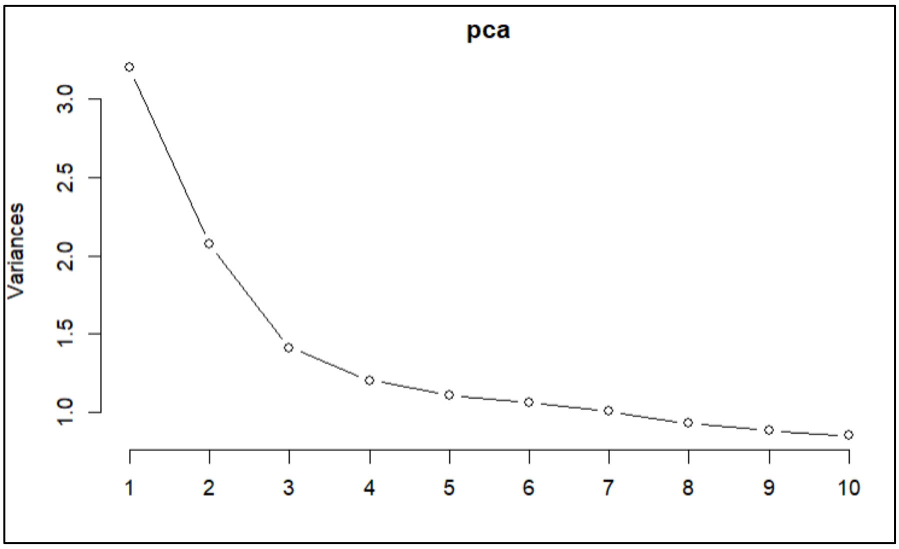


Figure 11: Variance explained by Principal Component Analysis

Principal Component 1:

This component contains features that include information on physician visits, especially frequency of visits per month and days since last claim related to all (including non-behavioural) health claims. There is high negative correlation with days since last claims in this component, which tells us that if a member has a high score on this principal component, then it is unlikely that they went through a claim, even if they visited a physician.

Principal Component 2:

This component contains features that indicate a strong relationship between disabilities and lack of homeownership. There is a strong negative correlation with homeowner status and a positive correlation explaining that members entered Medicare due to disabilities reasons. This is corroborated by the SHAP values as well, which indicate that if a member is disabled, he or she is more likely to be at risk of housing insecurity. We also understand from this that members with a high score of the component are mostly disabled and renters.

Principal Component 3:

This component summarizes features that indicate disabilities (behavioural and non-behavioural), but it is different from the previous component as it positively correlates with homeownership. We believe that if a homeowner member has a high score on the principal component 3, they have disabilities and face housing insecurity issues due to unsafe living conditions. This is corroborated by the SHAP values, which indicate that homeownership is an important metric in housing insecurity, potentially due to failure to make mortgage payments.

Principal Component 4: Healthy but at potential risk

This component does not show features that describe housing insecurity currently, but these members could be under risk of housing insecurity in the near future. This particular component could be important for Humana as it would help identify potential members under risk in advance and take preventive measures that can reduce the impact of by dampening extreme effects of unsafe housing issues. Helping members understand the meaning of housing insecurity and raising awareness on Humana's solutions to housing security could benefit these members significantly.

4.2. Segmentation

Post interpreting the principal components and reasons behind housing insecurity, we wanted to identify and differentiate Humana's members that face housing insecurity. For this, we performed a k-means clustering analysis to deep-dive and narrow down the reasons for housing insecurity so we could propose targeted solutions. We found that the 4 principal components segmented and separated the members into 4 groups.

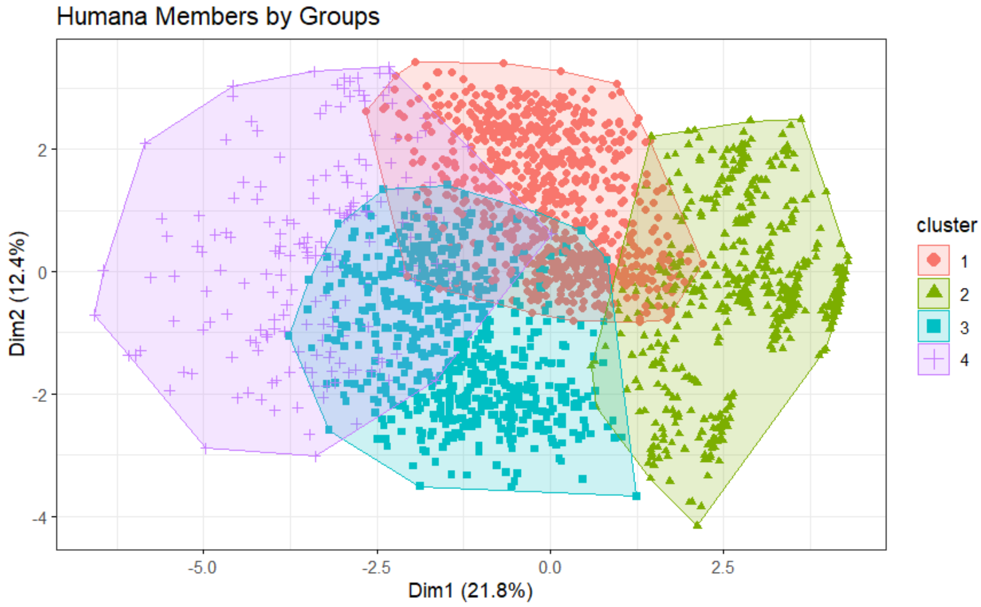


Figure 12: Results of k-means clustering

Group 1: Members with health issues and irregular claims

This group comprises 35.36% of Humana's housing-insecure members, with a large percentage of the members entering Medicare due to disabilities. This group also has the highest average age of 73.75 years compared to other groups. Additionally, this group has the least visits per month to physicians and has a high number of days since their last claim, which tells us that either this segment of members hesitates from reaching out to physicians for help or does not have easy access to physicians. According to the data collected by the U.S. Census Bureau through the Household Survey Pulse, people with health issues, especially disabilities, are disproportionately more like to pay more than 30% of their income on rent and face eviction. With low employment rates and low-income jobs for people with disabilities and severe health issues, this group of people is often trapped in a loop where housing insecurity exacerbates poor health and vice-versa.

Group 2: Members who do not own homes

This group comprises 30.55% of Humana's housing-insecure population. The housing problems faced by these members are not health-related. Instead, this group faces housing insecurity due to factors related to housing safety and neighbourhood safety. There have also been reported cases of instability in housing from renters because of homeowner negligence and harassment. The U.S. Department of Housing and Urban Development has reported that homeowners and landlords sometimes display subtle forms of discrimination based on racial background and ethnicity by refusing to help with maintenance and raising rent without notice. This often leads to unsafe living conditions, which can degrade the self-care and independence of renters.

Group 3: Members who have access to physicians

This group comprises 25.17% of Humana's housing-insecure population. In the data, we notice that this group of members visit doctors often and have high claims for behavioural and non-behavioural issues. This tells us that this group has access to easy healthcare but also faces health issues, given the frequent monthly visits. The average age of this group is lower than that of group 1, which also faces health issues, and is around 69.47 years, but it has a higher average of low-income indication and receives financial aid from the Centres for Medicare & Medicaid Services. A survey conducted on over 40,000 renters showed that low-income renters often face difficulties in paying rent and sacrifice other basic needs like food, water, education, and transportation to avoid eviction and homelessness.

Group 4: Members for whom we have limited information

This small group comprises 8.92% of Humana's housing-insecure population. For many of these members, we did not have information on their general demographics, like homeowner status, claims information, physician visits, and health-related information. The lack of information for this group of members can lead to wrongly identifying people who are facing housing insecurity. It can also result in not identifying members who are facing housing insecurity at all, which can have drastic consequences on the quality of life of members as they do not get the right resources when required.

4.3. Recommendations and Solutions

In this section, we are focussing on potential practical solutions to improve housing conditions for every segment of the group described above. We want to focus on recommendations that are easy to implement for Humana, cost-effective, and practical.

Group 1: Members with health issues and irregular claims

For this group which comprises mostly of people with disabilities, Humana does offer coverage without having to pay no premium for hospital coverage. But, to access doctor visits, lab tests and preventive services (Medicare Part B), the members are expected to pay a monthly premium of ~ $165 which may not be very affordable to all as members with disabilities would want to have more easy access to doctor visits than having hospital coverage. We recommend Humana to propose monthly / yearly coverages with lower premiums, so that it would include this group where people have low incomes, have disabilities and are facing housing insecurities.

Also, this group has the least visits per month to physicians and has a high number of days since their last claim, indicating that members availed services which may not be covered by Humana, or the members have mental health issues which is causing them not to have frequent doctor visits. We recommend Humana to broaden the coverage when it comes to disabilities, dental, behavioural coverages. With respect to mental health, the member must pay a co-payment of between $25 - $100 per visit with their in-network therapist to access services and higher out-of-network payments.[6] Hence, considering the high fees, members who are disabled and face mental health issues may not be able to afford. Also, with respect to the coverages, Humana provides online mental health coverage but is planning to rescind the same post-covid. It also does not cover career and life coaching, and only a few plans cover couple's therapy.[7] We advise Humana to continue to provide online mental-health coverage by creating awareness amongst the members and to include other coverages, as mentioned above. Humana can also partner with local non-profit organizations to reach out better and have regular check-in with the members, increasing physician visits. There is a strong need to create more awareness, education, and training for the members of this group.

Group 2: Members who do not own homes

Renters can often face hardships due to lack of help from landlords. Humana offers services that help maintain cordial relationships between tenants and owners, but we suggest taking it one step forward by collaborating with local governments and businesses like they do with Louisville's Unite Community program.[4] Often such community services provided by local governments are more reactive and can ease the process when it comes to handling problems related to evictions and lack of housing maintenance (repairs and hygiene). Adding such information to Humana's website that shares local resources on how Humana can help with intervention between tenants and homeowners. It will help members live in safe conditions and can potentially avoid homelessness. Furthermore, teaming with local communities would help save resources spent by Humana.

Group 3: Members who have access to physicians

For this group that has access to healthcare but also faces health issues, we suggest Humana partner with local hospitals to raise awareness on services that Humana offers in terms of housing insecurity. Currently, on Humana's website, there is one flyer that gives information about Housing to Seniors, which might not be accessible to many of Humana's elderly population.[5] We suggest that Humana should have representatives go to hospitals and make administration, nurse practitioners, and doctors aware of the existing services. In addition, Humana should send out regular emails and newsletters with this information to members suspected to be facing housing insecurity, as it would help drastically reduce the reaction time in terms of providing help to the members.

Group 4: Members for whom we have limited information

The lack of information on MAPD members is hindering the identification of a few members who may be facing housing insecurity. Increasing the capture rate of data will not only allow us to predict with accuracy if a member could face housing instability, but it would also help Humana in improving upon the reason for housing insecurity and the segments. Apart from capturing data of registered MAPD members, we suggest that Humana send out pulse check surveys every quarter to their members. The surveys should include detailed questions about income, living conditions, employment, education, and neighbourhood conditions. Collecting data at regular intervals would help Humana forecast in advance time if the housing insecurity of a member is increasing or decreasing. However, it should be ensured that data is collected from members everywhere equitable and the surveys are easily accessible to everyone. This would also allow for better management of resources as Humana would be able to proactively reach out to members and increase awareness of available services.

Not only would this be an effective way to collect data and check up on the members, it would also be a cost effective way to do so while still being able to scale, because sending out survey to collect data is next to nothing.

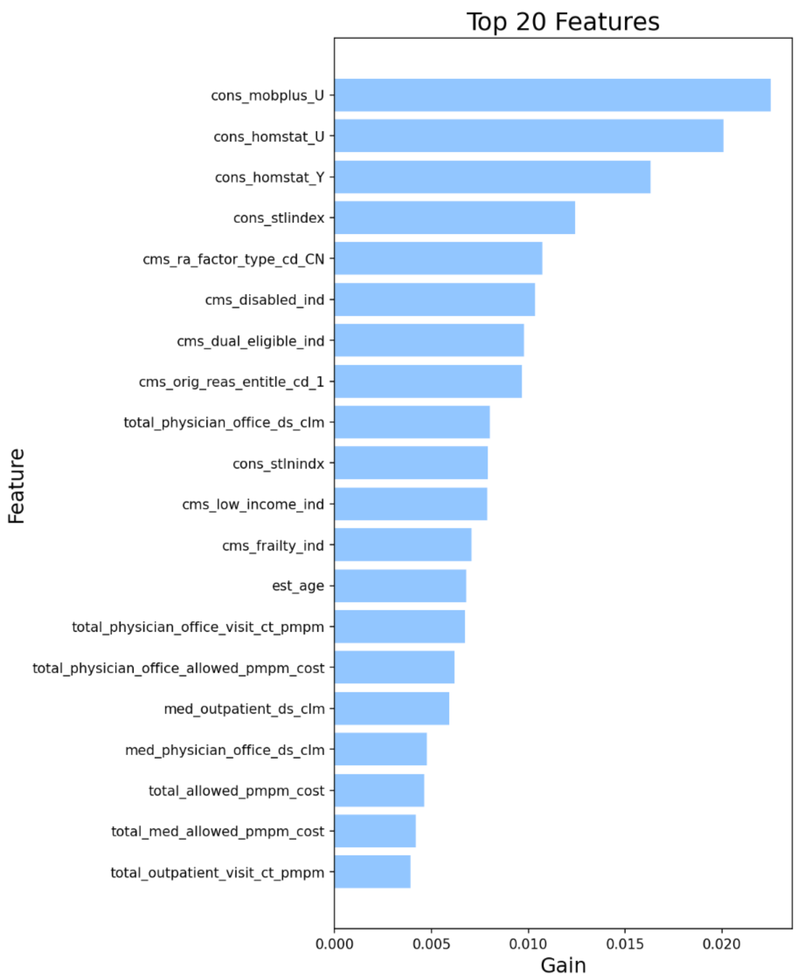
5. CONCLUSION

In predicting which of Humana's Medicare members were most likely to be experiencing Housing Insecurity, we filtered 870 features down to 668 features for use in training our machine learning models, ultimately determining that the XGBOOST model, with an AUC score of 0.75, could predict housing insecurity issues most accurately and appropriately. By identifying individuals with housing insecurity and noting the key drivers of these outcomes, we generated business proposals such as collaborating with local communities and hospitals, increasing awareness of Humana's housing services, and collecting more quality data. All these strategies are recommended while keeping in mind both the cost of implementation and effectiveness in combatting housing insecurity while also being able to scale these effectively.

For future considerations, we further recommended accurately identifying potential members that could be at risk of housing insecurity as a preventive measure while also directing resources toward members that are currently facing housing issues.

6. APPENDIX

Top 20 features based on Gain in XGBOOST. These top 20 features are selected based on the number of times the feature is used when creating a classification tree.



7. REFERENCES

1. https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/housing-instability
2. https://www.huduser.gov/portal/pdredge/pdr-edge-trending-053122.html#:~:text=Housing%20insecurity%20in%20all%20its,mental%20and%20physical%20health%20problems.
3. https://www.nashp.org/how-states-improve-housing-stability-through-medicaid-managed-care-contracts/
4. Recognizing and Addressing Housing Insecurity for Disabled Renters - Center for American Progress
5. Rental Insecurity: The Threat of Evictions to America's Renters (apartmentlist.com)
6. https://zencare.co/health-insurance/humana
7. https://www.humana.com/medicare/medicare-resources/medicare-and-disability