Predicting E-Cigarette Smoking (Vaping) Trends Using CDC 2017 BRFSS Data

Stephanie Halsing, Rick Morales, and Erika Torkildsen

**Each author contributed equally to the design, coding & development, analysis, and writing of this project*

May 11, 2025

Synopsis

This project leverages data from the 2017 Behavioral Risk Factor Surveillance System (BRFSS) to examine the associations between demographic, behavioral, and clinical variables and e-cigarette (vaping) use status among the diagnosed asthmatic adult population. By focusing on this subpopulation, we aim to understand predictors of e-cigarette (vaping) use that may help shape health messaging and prevention strategies for respiratory-vulnerable populations.

Introduction

Background

E-cigarette (vaping) usage continues to rise in the United States (U.S.), creating public health concerns due to potential respiratory effects and associated behaviors in alcohol use. As the tobacco industry has evolved over time, e-cigarette smoking, or vaping, has become a popular alternative to conventional smoking. This rise in popularity for vaping has encouraged researchers to conduct studies to examine the vaping rates as well as the health conditions and illnesses believed to have arisen from it. According to the results of the Behavioral Risk Factor Surveillance System (BRFSS) 2016-2018, e-cigarette use has been on the rise from 4.5% in 2016 to 5.4% in 2018 with 2017 being a mostly stable year at 4.4% (Obisesan et al., 2020). In one particular study, researchers surveyed asthmatics who vaped and non-asthmatics who transitioned from smoking to vaping and found the transition to have no effect on pulmonary function tests (Solinas et al., 2020). Another study examining the state of e-cigarette or vaping use-associated lung injury (EVALI) in a lifelong tobacco smoker who switched to e-cigarettes found the patient’s case of asthma to be triggered by irritants found in vape liquid (Roberts et al., 2021). It is imperative to gain further knowledge and understanding on the effects of vaping to support the conclusions from past research and discover new insights. The results of this study will help support efforts in executing public health interventions in communities most impacted by asthma, along with partnerships to develop future health guidance.

The Centers for Disease Control and Prevention (CDC) has a vested interest in this research to gain further knowledge and understanding of tobacco product trends. Specifically within the CDC, the Office on Smoking and Health (OSH) organizes efforts to intervene through prevention techniques, aid for those who want to quit, and methods of reducing exposure to secondhand smoke. As this study is targeted towards adult asthmatics, the National Asthma Control Program (NACP) will also be informed since it aims to combat the chronic health condition through research that leads to interventions and treatment options offered to asthmatic communities.

While our primary focus is on e-cigarette usage among adult asthmatics, we recognize the need to clarify terminology throughout the report. We use the term “vaping” and “e-cigarette smoking” interchangeably to reflect the behavior of using electronic nicotine delivery systems (ENDS), including tobacco. This terminology is consistent with common usage in BRFSS documentation and aligns with how participants are likely to interpret survey questions.

Question

“Can we predict e-cigarette smoking (vaping) of tobacco and marijuana among United States adult asthmatics in 2017 after accounting for demographics, conventional smoking status, and alcohol usage?”

Hypotheses

1. Adult asthmatics who vape will be more likely to be past smokers because advertisers promoted vaping as a healthier alternative to smoking in the years leading up to 2017.

2. Adult asthmatics who vape are more likely to engage in other risky lifestyle choices, like high alcohol consumption, because individuals can have multiple addictions, or they may have lifestyles in which they do these activities together.

Predictions

1. We predict that vaping will be relatively higher among adult asthmatics who smoke or have a history of smoking as compared to those who do not smoke tobacco cigarettes.

Data

Data Acquisition

For the dataset, the 2017 BRFSS data, which is a cross-sectional survey conducted by the CDC that collects health-related data from a nationally representative sample of non-institutionalized U.S. adults. This analysis has been filtered to only include respondents who reported ever having been diagnosed with asthma. The target variable, ECIG_COMP, categorized e-cigarette usage into five values. These were mapped into the following VapeStatus labels: 1 = "Current User - Every Day", 2 = "Current User - Some Days", 3 = "Former User", 4 = "Never Used", and 9 = "No Response". For the modeling portion, the value 9 was treated as missing. We also created a binary version, ECIG_BINARY, that grouped values 1 and 2 as "Current User" and values 3 and 4 as "Other". The predictor variables included conventional smoking status, alcohol use, body mass index (BMI), healthcare access indicators, mental health status, and demographic factors such as age, income, education, and sex. These variables were selected based on prior literature and relevance to the target behavior. All variables are included in the codebook located in Appendix A.

The BRFSS 2017 dataset contains self-reported survey data, which introduces potential for recall bias and missingness in sensitive topics such as tobacco, alcohol, and health history.

Some variables had to be dropped due to high missingness or encoding issues ("Refused" or "Don't Know" values coded as 7 or 9). Despite this, key predictors were retained, and all missing categorical entries were handled through filtering or recoding, while numeric predictors were imputed using MICE (Multiple Imputation by Chained Equations).

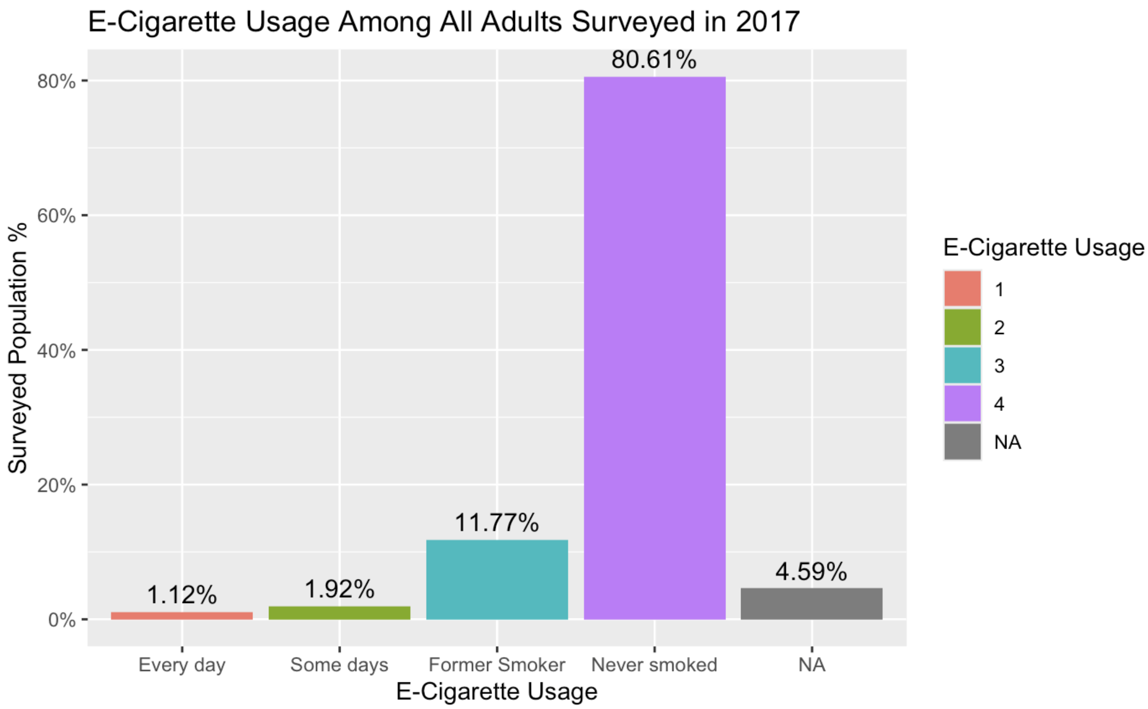
Data Cleaning

The first step of the data cleaning process included the removal of irrelevant or redundant variables and re-coding of invalid responses (e.g., 9 = Refused/Don't Know). The original dataset included 358 variables and 450,016 observations. Variables were removed if there was a high percentage of missing values or related variables, such as height and weight, that are associated with BMI. The purpose of renaming variables was to provide clarification and understanding of each variable. Once the data cleaning was complete, the final dataset included 62 variables and 12,553 observations.

Data Exploration

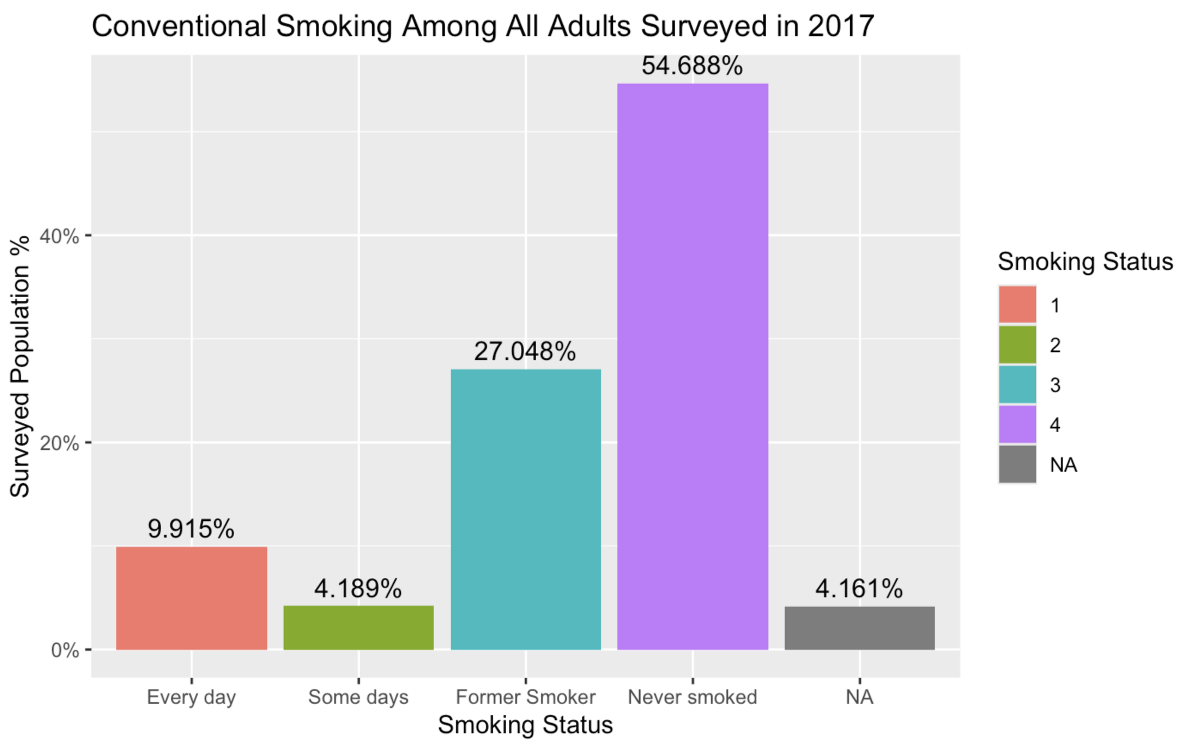
Visualizations were created using ggplot() and include bar plots with the target variable (ECIG_BINARY) levels factored to categorize and store the data as factor levels with distinct colors per category. The percentage of counts within each category is computed using prop.table(stat(count()). A continuous scale is specified for counts on the y-axis and a discrete scale used for categories on the x-axis.

Figure 1. E-Cigarette Usage Among All Adults Surveyed in 2017



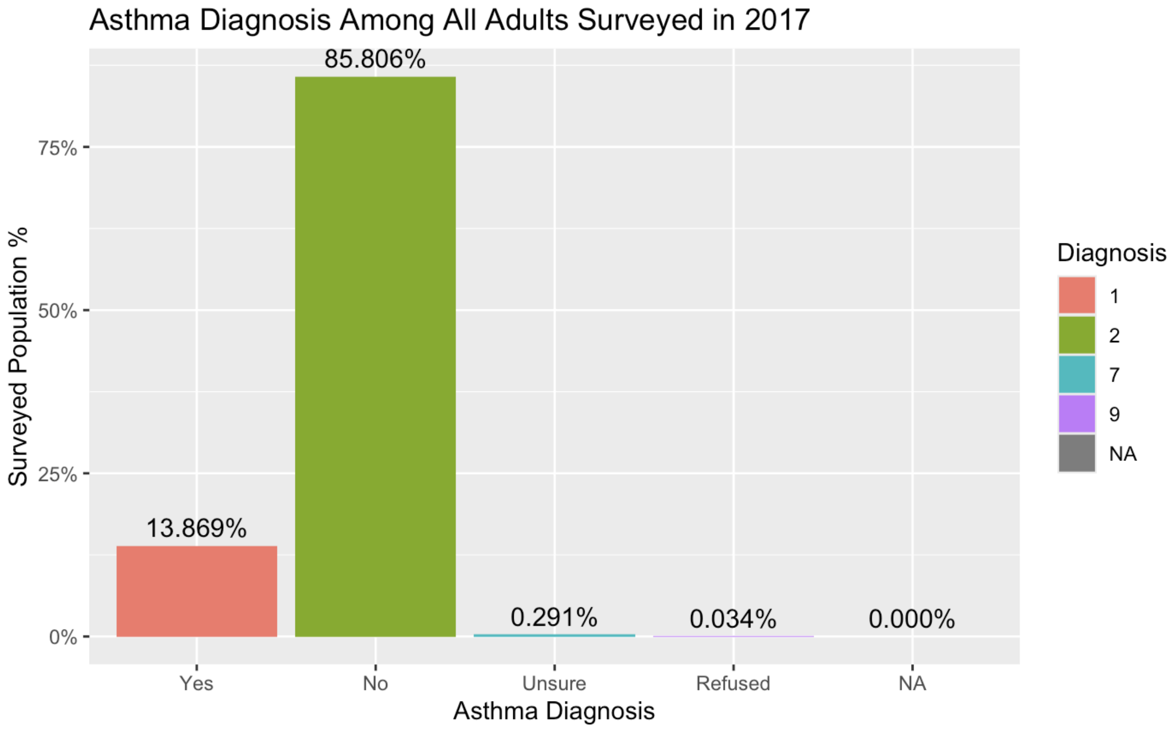
Prior to pre-processing and feature engineering, bar plots were made to visualize relationships between predictors and the target variable and provide the team with greater understanding of initial observations. To examine proportions of surveyed adults within the target variable, a plot for e-cigarette usage was made and it is found that nearly 80% of all surveyed adults in 2017 never smoked e-cigarettes according to the BRFSS dataset.

Figure 2. Conventional Smoking Among All Adults Surveyed in 2017



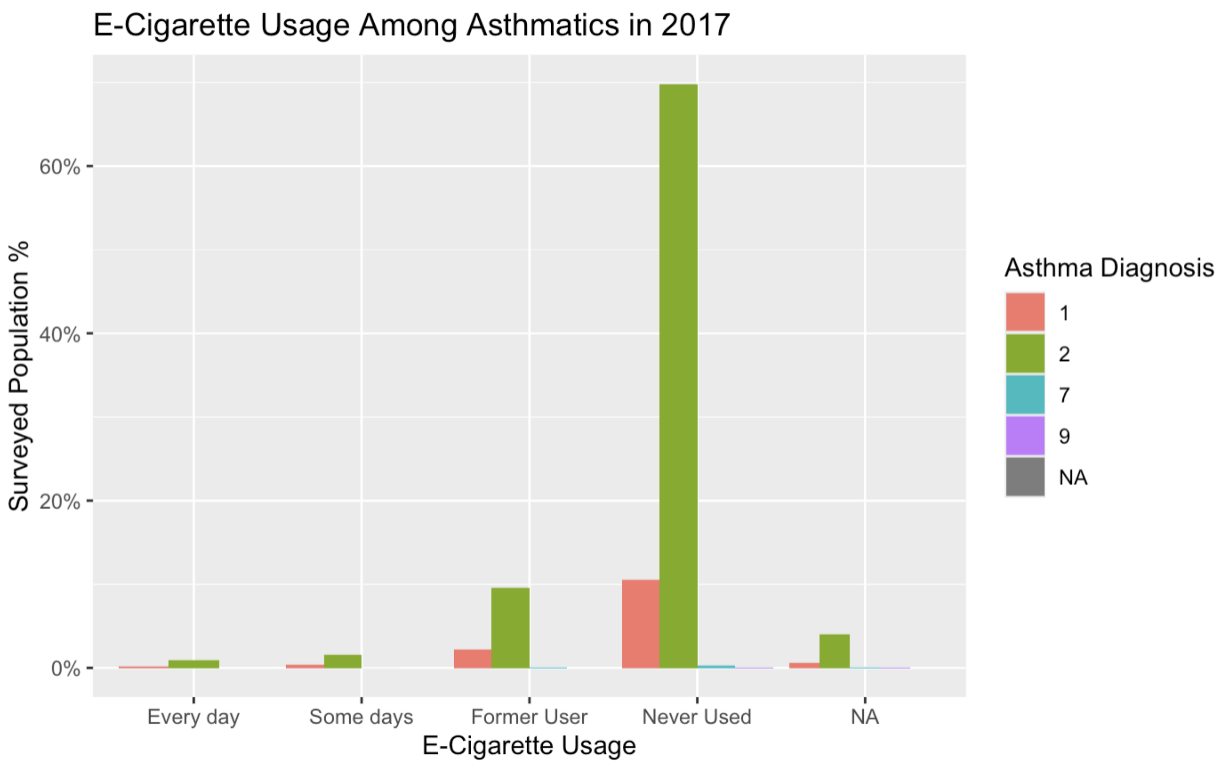
To examine another predictor and the percentages in each predictor category within the population, a plot was made to visualize the proportion of adults in each smoking status category. It is discovered that the greatest proportion of adults, nearly 55% of adults surveyed in 2017, have never smoked tobacco. This is followed by the second highest proportion with roughly 27% of surveyed adults identifying as former smokers.

Figure 3. Asthma Diagnosis Among All Adults Surveyed in 2017



A third plot visualizes the proportion of surveyed adults with and without asthma in addition to those who were unsure and those who refused to respond. A significant majority of participants say that they have never been told that they have asthma.

Figure 4. Asthma Diagnosis Among All Adults Surveyed in 2017



In figure 4, this bar chart shows the e-cigarette usage among adult asthmatics and the data is categorized by the different responses: 1 = “Yes”, 2 = “No”, 7 = “Don’t know/Not Sure”, and 9 = “Refused”. The majority of non-asthmatics never used e-cigarettes with a small portion of them being former users. Those with asthma also appeared to never vape with a small portion being former e-cigarette users.

Figure 5. E-Cigarette Usage Among Conventional Smokers in 2017

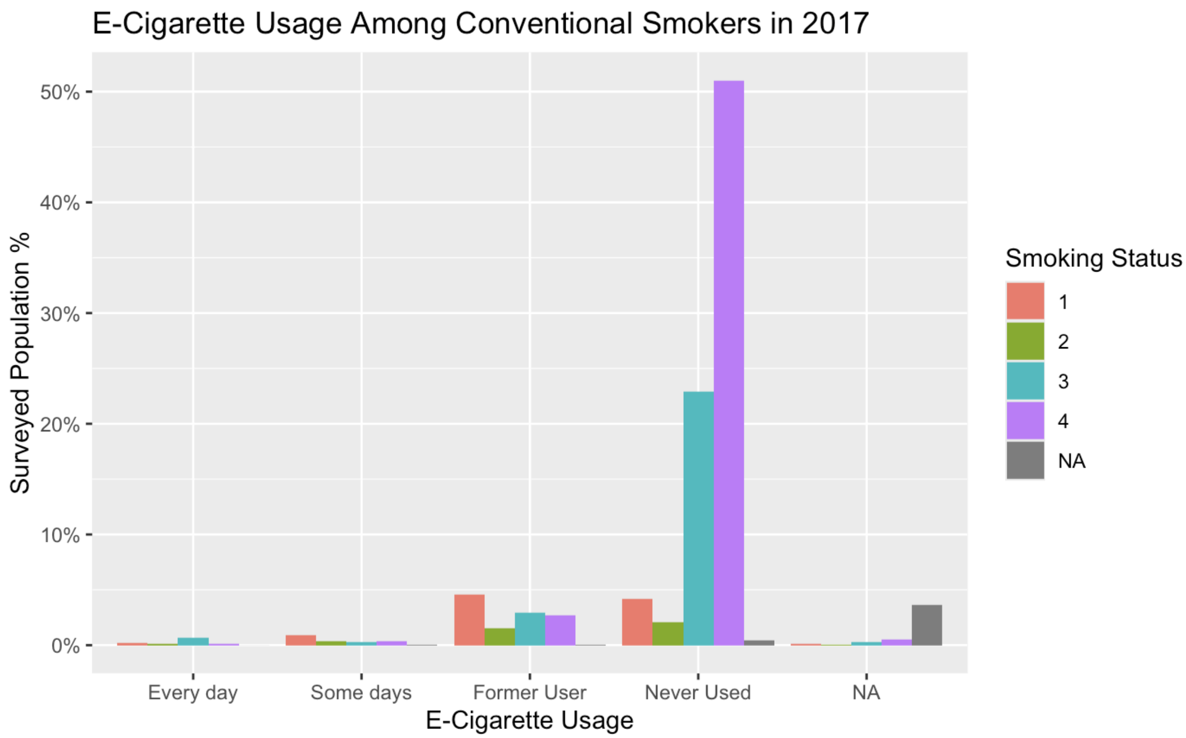
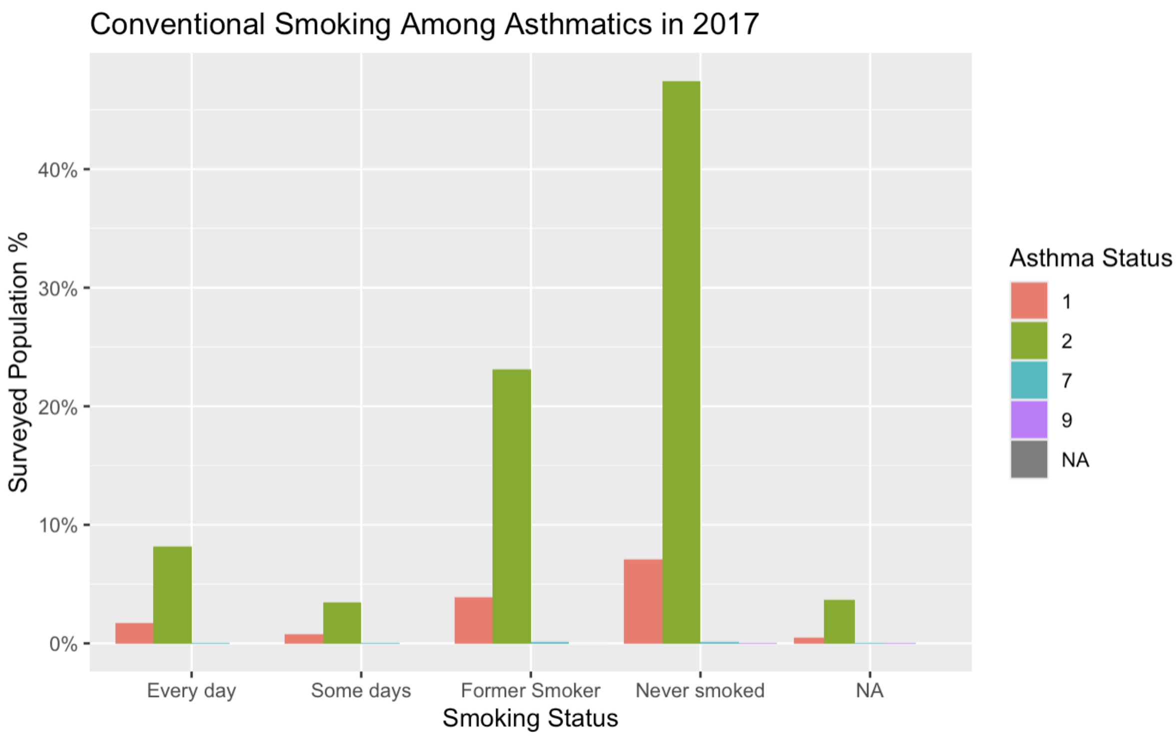


Figure 5 shows the e-cigarette usage among conventional smokers and this is categorized by five different responses: 1 = “Current smoker - every day”, 2 = “Current smoker - some days”, 3 = “Former smoker”, 4 = “Never smoked”, and 9 = “Don’t know/Refused/Missing”. The majority of the entire population has never used e-cigarettes, but the data showed that current everyday smokers were former e-cigarette users. According to the bar chart, a small percentage of former smokers switched to using e-cigarettes as their everyday form of smoking.

Figure 6. Conventional Smoking Among Asthmatics in 2017



In Figure 6, conventional smoking status is displayed among adult asthmatics. The majority of non-asthmatics either have never smoked or are former smokers. Many diagnosed asthmatics tend to never smoke, but some are former or everyday smokers.

Models

Pre-processing and Dimensionality Reduction

Pre-processing and feature engineering were used to evaluate predictors of vaping status. First, no encoding was needed for the variable, STATE, since it is a single, numeric variable, and this format was kept from the original dataset. No other categorical variables existed in the original dataset. Next, the missing values were imputed using Multiple Imputation by Chained Equations (MICE) with Classification and Regression Trees (CART). The dataset was then split using stratified sampling to maintain proportional representation of classes and applied the Synthetic Minority Oversampling Technique (SMOTE) within cross-validation to address class imbalance.

Multicollinearity was checked using the Variance Inflation Factor (VIF) as shown in Figure 7, and then the exclusion or combination of variables was added when needed. In Figure 7, the VIF shows that the variances would be between 0% to 20% higher than expected if no multicollinearity were present. Then, the application of the Principal Component Analysis (PCA) and k-means clustering during exploratory analysis to understand variable groupings and natural clusters. K-means clustering is executed using the kmeans() package in R. A value of 3 is specified for the number of clusters after determining this value to be optimal from the elbow curves computed in the Appendix, Figures 25 - 27. The total within the cluster sum of squares decreases as the number of clusters increases. Specifying an mstart of 25 generates 25 initial random centroids. An fviz_cluster() visualizes results and allows inputs for data point size and color per cluster.

The optimal train-test ratio was discovered using random values and a custom function. The random values method resulted in a 0.5 split ratio while the custom function resulted in a 0.8

split ratio. Therefore, it was decided to use the median of the two results and round up to a 0.7 split ratio.

Figure 7. Variance Inflation Factor (VIF)

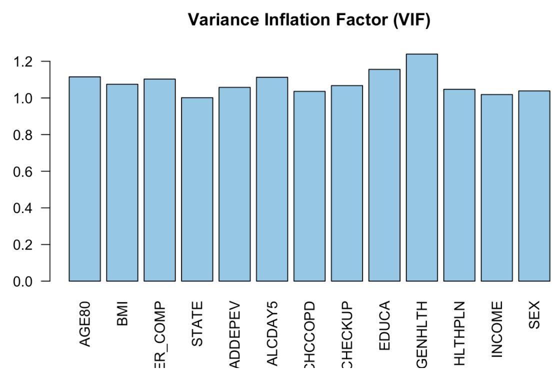
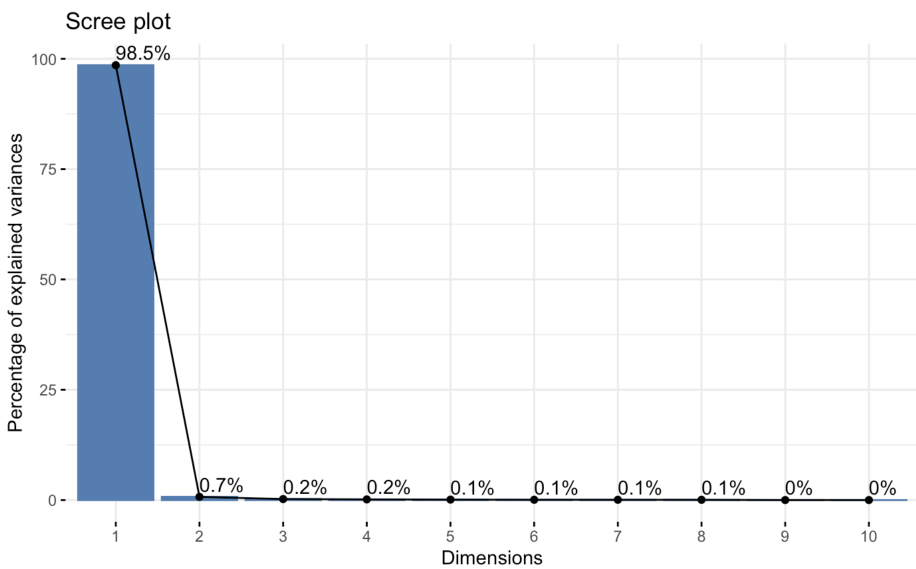
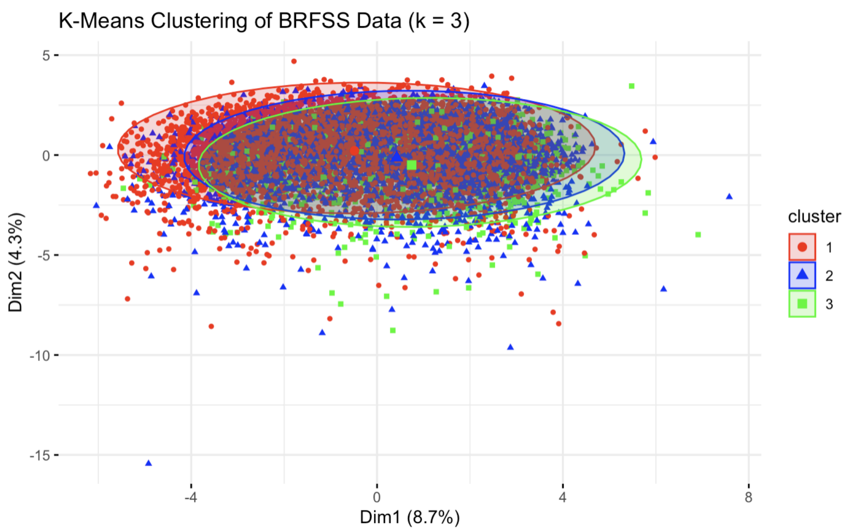


Figure 8: Explained Variance per Principal Component (Model 3)



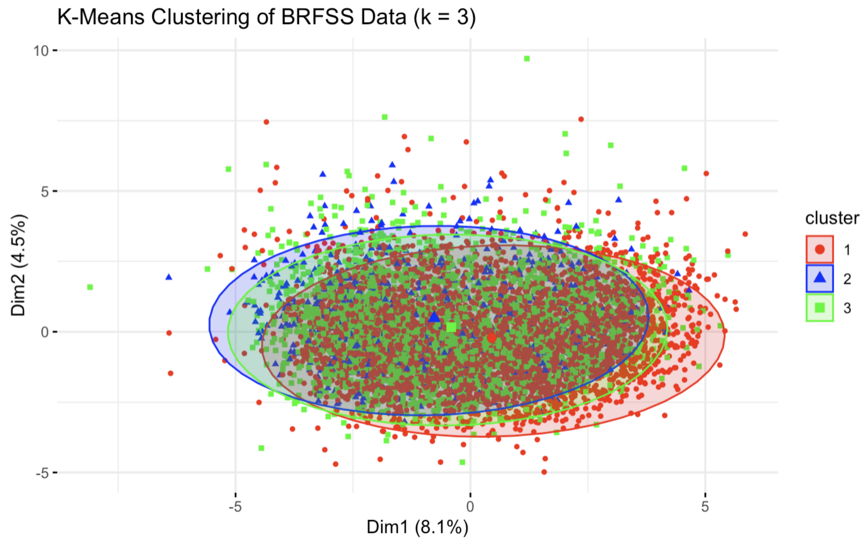
A scree representing eigenvalues, also known as principal components (PC), was made and the percentage of variance drops off significantly from the first principal component to the second. This indicates that 99.2% of the variance is explained by just two principal components. Despite ensuring NA values are not present in the data, the percentage of explained variance is not balanced across components. In order to address this issue in further studies, techniques such as standardization and additional efforts to transform the data will be used as a potential avenue to explore the cause of high variance in one PC.

Figure 9: K-Means Clusters for Asthmatics with History of Vaping



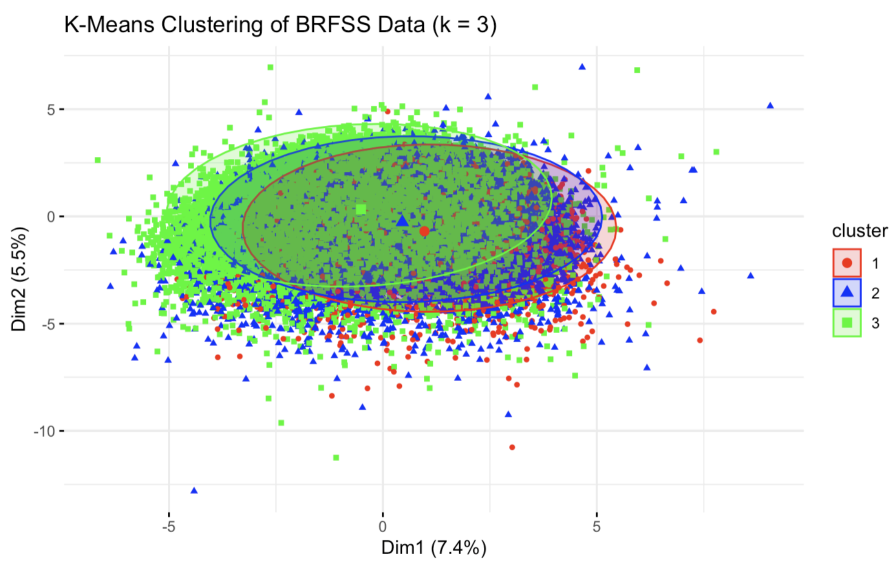
Computed clusters for asthmatics with a history of e-cigarette usage contain centroids that are relatively close in distance to one another, which indicates similar averages per datapoint cluster. Points are widely spread out from the centroids, which indicates less than ideal clustering performance. There are a fairly large number of outliers that fall outside of the three clusters. There is not a clear distinction between the clusters because they overlap significantly. Points are extremely close to one another, which indicates homogeneity and shared qualities.

Figure 10: K-Means Clusters for Asthmatics with History of Vaping & Cigarette Smoking



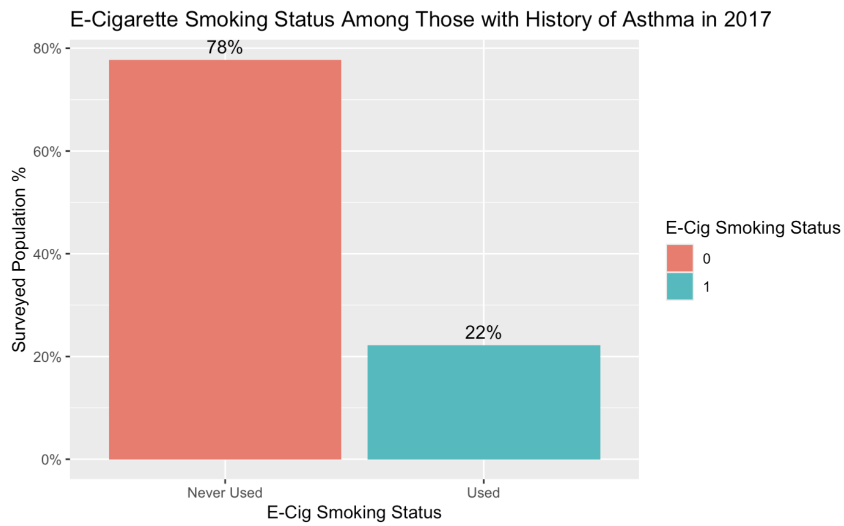
Similar to the K-means clusters computed in Model 1, clusters in Model 2 exhibit the same characteristics of homogeneity and centroids relatively close in distance to each other. There are a significant number of outliers outside of each cluster's border, seen by the ellipse outlines. It is very difficult to distinguish between the three clusters.

Figure 11: K-Means Clusters for Asthmatics



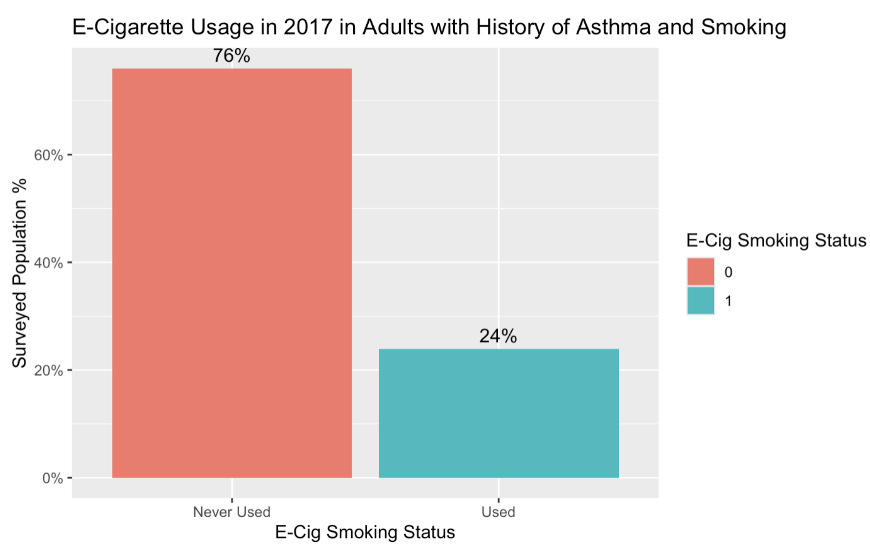
Results from K-Means clustering in Model 3 are consistent with those in Models 1 and 2 with all three centroids close in distance to each other. Many outliers for cluster 3 are found outside.

Figure 12: Class Imbalance Check for Asthmatics with History of Vaping



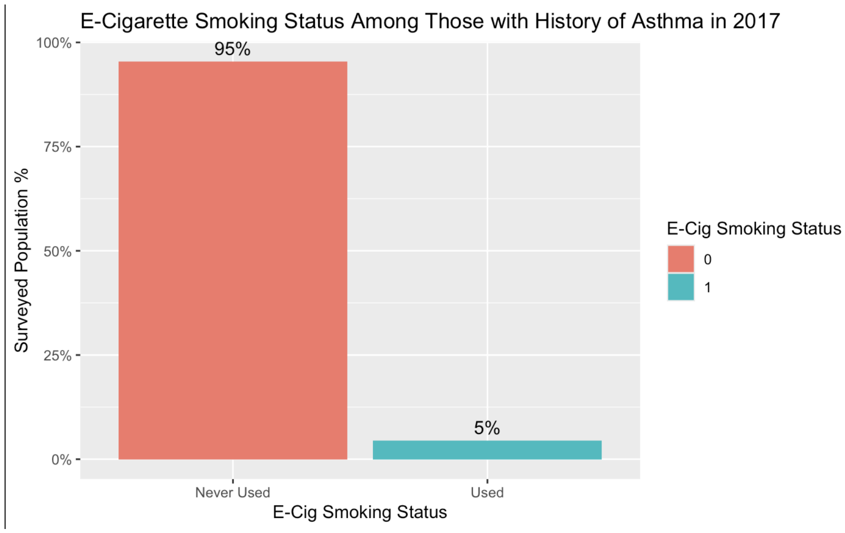
Following imputation of NAs and checks for multicollinearity, a check was performed to see if the model remained imbalanced and it was, with 78% of asthmatics having never used e-cigarettes according to the Model 1 training data.

Figure 13: Class Imbalance Check for Asthmatics with History of Vaping & Smoking



Similarly, for Model 2, significant class imbalance is found with 76% of surveyed adults asthmatics in 2017 having never used e-cigarettes.

Figure 14: Class Imbalance Check for Asthmatics



Model 3 contains the most class imbalance, with 95% of surveyed adult asthmatics having never used e-cigarettes. Imputation did not solve for imbalance, but rather it increased the number of non-NA observations. Class imbalance has the potential to cause bias in the model classifier, and this occurrence will be an observation of interest to be discussed further in the study

Algorithm Selection

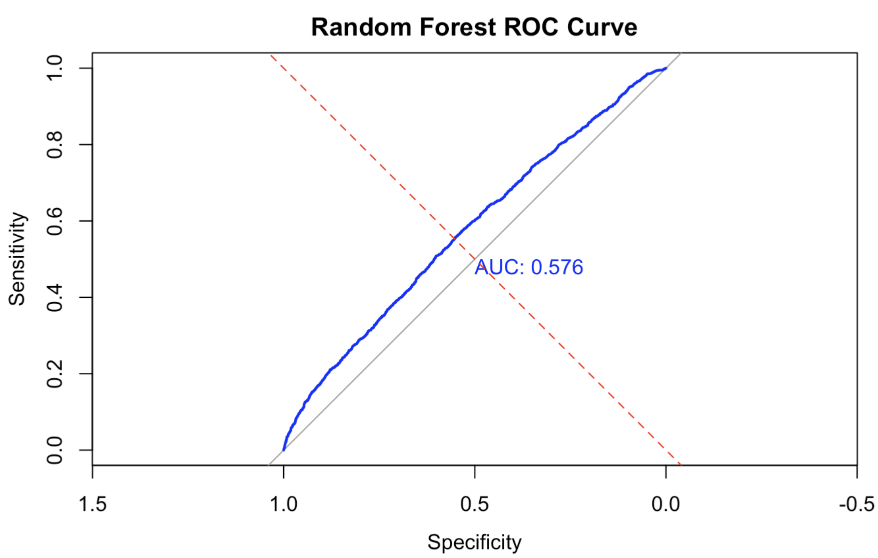
Two primary models were evaluated: (1) an ordinal logistic regression using the polr function to reflect the ordered nature of ECIG_COMP, and (2) a Random Forest classifier. For the latter, both the multiclass (single test case) and binary target versions (three test cases) were tested. All models were evaluated using the stratified 5-fold cross-validation, and SMOTE was applied during training for class imbalance, as the target variable is ordinal.

The ordinal logistic model offered some interpretability benefits but performed poorly in prediction, with low prAUC and inconsistent classification of minority classes. In contrast, the

Random Forest model achieved an AUC of 0.833 after tuning hyperparameters (ntree = 1000, mtry = 2, nodesize = 4). It outperformed the logistic regression model across ROC, accuracy, and PR-AUC metrics. Feature importance analysis showed that SMOKER_COMP, AGE80, and ALCDAY5 were consistently the strongest predictors of the target variable.

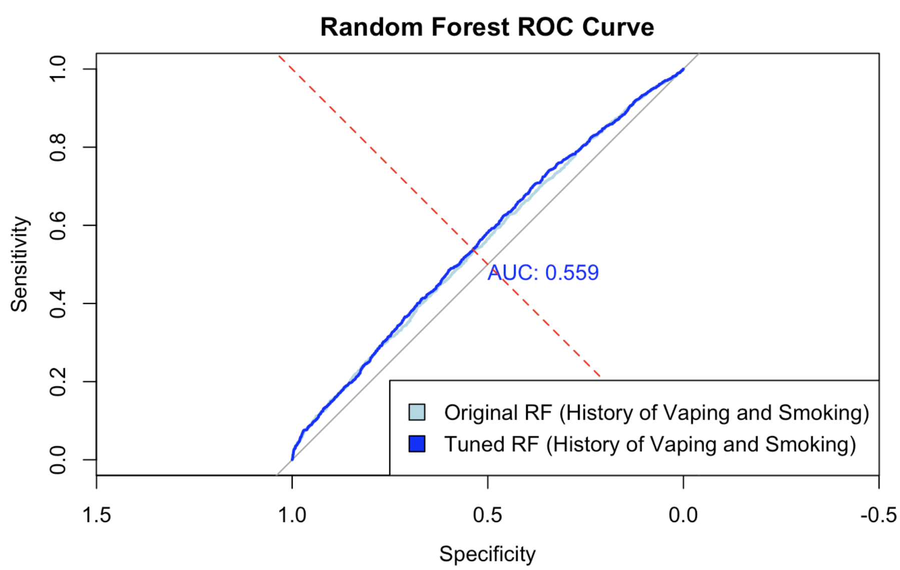
Model assumptions were addressed throughout the analysis. For logistic regression, we assessed residuals, checked multicollinearity, and confirmed model convergence. For Random Forest, generalization was ensured by applying stratified cross-validation and tuning key parameters. Calibration and classification errors were explored both visually and numerically. Stratified cross-validation was used to address overfitting in the model.

Figure 15: ROC Curve for Model 1 (Original)



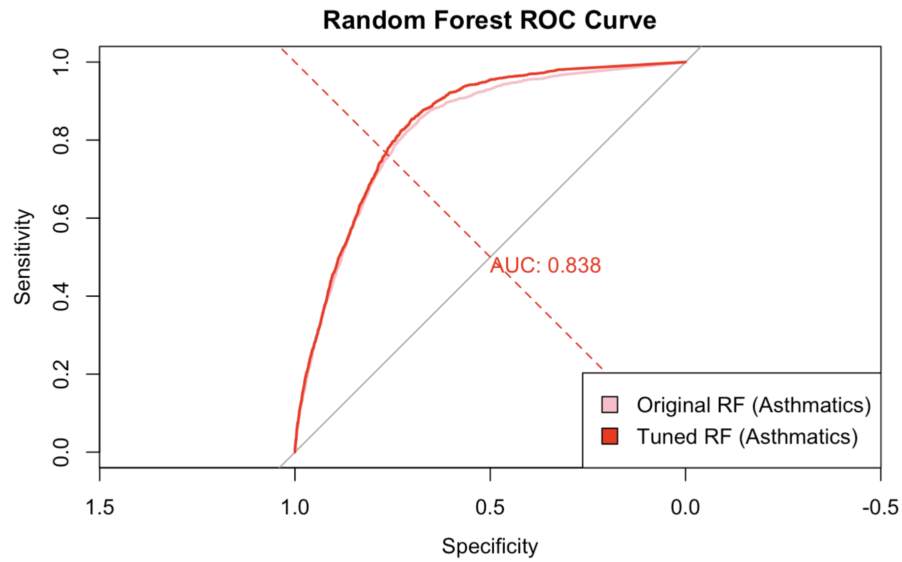
Receiver Operator (ROC) curves are computed using calculated predictions made on the random forest model. The obline() function creates a diagonal dotted line in red that represents an AUC of 50% for reference to a random guess. Sensitivity (recall) is understood to be the true positive rate of classification, while 1 minus specificity represents the false positive rate of classification. Above, an ROC curve is computed from a random forest model with a configuration of 1000 trees which was specified as a hyperparameter tuned from the original # of 500 trees to increase accuracy and reduce variance. To find the optimal value for mtry, which is taken to be the number of variables randomly sampled as candidates at each split after the ntreeTry hyperparameter is given a value of 1000, the tuneRF() function is used. The optimal mtry value has the lowest out-of-bag error. An mtry value of 3 is found to be optimal, which was already specified originally prior to running tuneRF(). An AUC of 57.6% indicates poor performance, which is slightly better than a random guess. It corresponds to an out-of-bag estimate of error rate of 22.21%. This metric indicates the number of misclassified instances, which is significant.

Figure 16: ROC Curve for Model 2 (Original and Tuned)



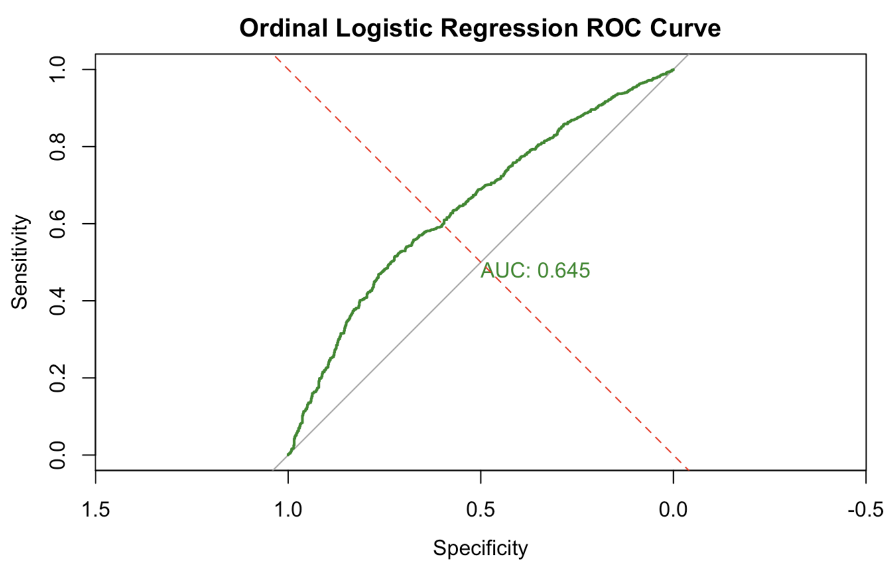
In the above plot, the original RF model with an mtry of 3 is plotted beside the tuned model with an mtry of 2. The tuned model performs better with a slightly higher AUC of 55.9% compared to 55.3% in the original RF model (Appendix, Figure 28). Despite tuning the model's hyperparameters to increase accuracy and better performance, the model did not significantly improve, and prediction accuracy can be comparable to a coin toss since it falls near 50%.

Figure 17: ROC Curve for Model 3 (Original and Tuned)



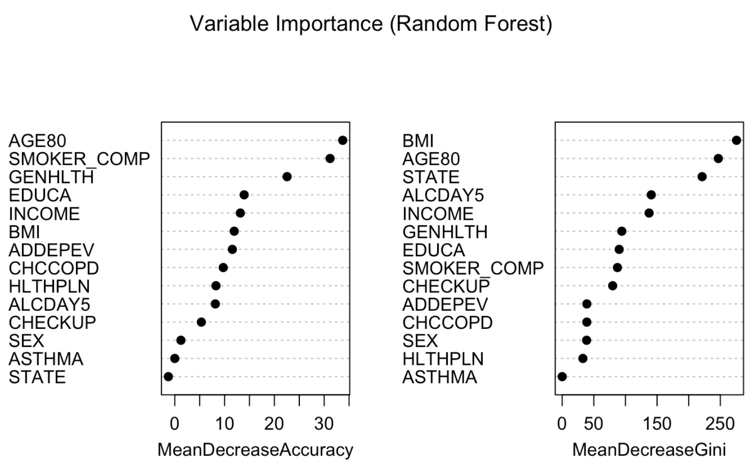
In the above plot, the original RF model with an mtry of 3 is plotted beside the tuned model with an mtry of 4, which is found through the computation of tuneRF(). This model is the best of the three, exhibiting the best performance with an AUC of 83.8%. Tuning the mtry model brings the AUC from 82.6% (Appendix, Figure 29) to 83.8%. The random forest model does very well at classifying the survey data.

Figure 17: ROC Curve for Model 4 (Original)



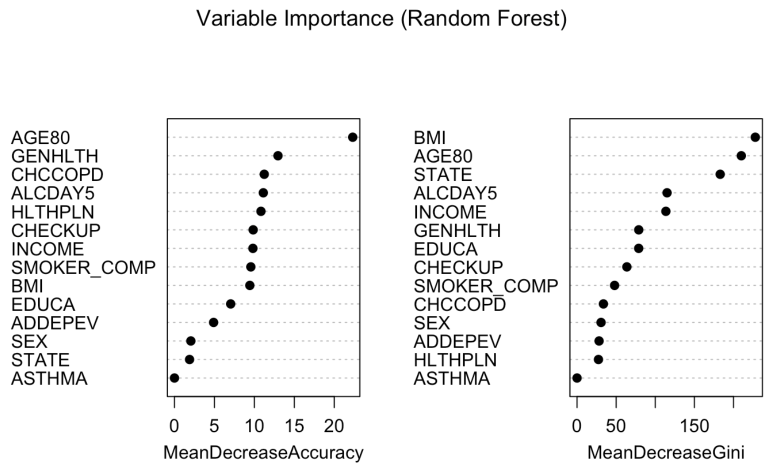
For a fourth model computed above for our multi-class target with an ordinal logistic regression, an AUC of 64.5% is found. While the study will continue analysis with the binary class target, the multi-class target results are described to compare performance across class types. Ultimately, the binary target models prove the best path forward for this study due to the simplicity and lower complexity when training the model. It is considered the best choice for an initial study but future studies might be well suited to build off of current research and focus on a multi-class target if resources allow for greater complexity and longer training time.

Figure 18: Feature Importance in Model 1 (Asthmatics with History of Vaping)



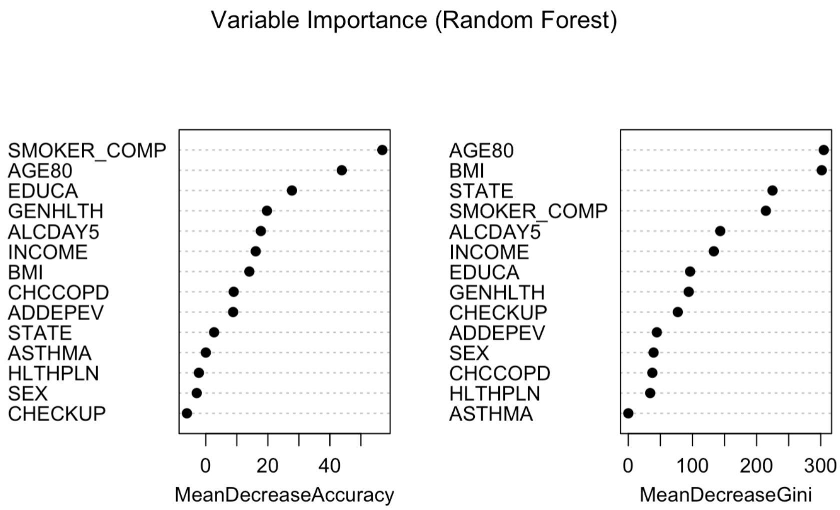
Following random forest computation and making receiver operating curves, feature importance plots are created using varImpplot(). The figure on the left shows how much accuracy the model loses by excluding each variable. Variables are shown with the highest levels of importance at the top and descending in importance. Respondents' age is found to hold the greatest amount of importance, followed by conventional smoking status and general health status. The graph to the right shows importance in contributing successfully to hyperparameters in the model with BMI contributing the most, followed by age and state of residence.

Figure 19: Feature Importance in Model 2 (Asthmatics with History of Vaping & Smoking)



For surveyed adult asthmatics with a history of vaping and smoking, respondents' age is found to hold the greatest amount of importance. But in contrast to previous feature results from Model 1, Model 2's feature with the second highest importance has a lower meanDecreaseAccuracy value. This feature represents general health status. COPD status has the third highest feature importance. BMI contributes the most to hyperparameters followed by age and state of residence which mirrors the same results from Model 1.

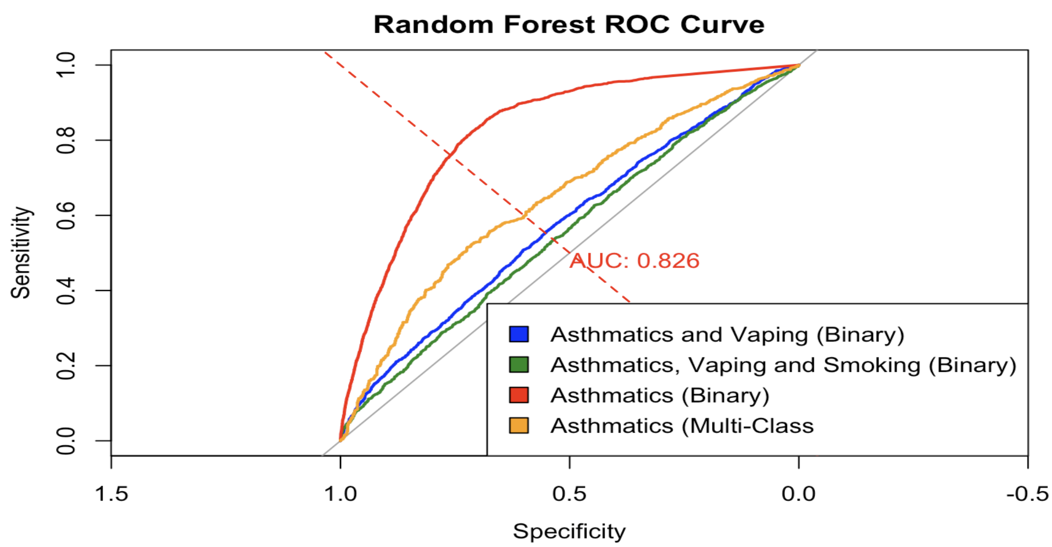
Figure 20: Feature Importance in Model 3 (Asthmatics)



For surveyed adult asthmatics,, smoking status is found to hold the greatest amount of importance followed by age and years of education, as with models 1 and 2, age, BMI, and state of residence contribute the most successfully to hyperparameters.

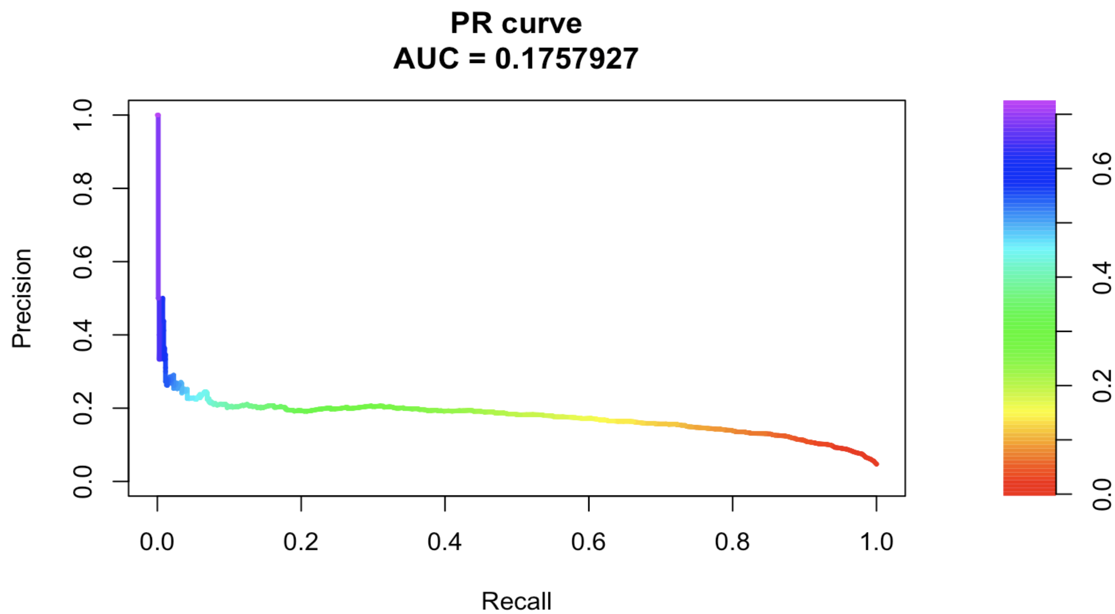
Final Model

Figure 21: ROC Curve Comparison Across Models



After considering results across the three models for e-cigarette usage as a binary target and the fourth model for e-cigarette usage as a multi-class target, the third model for asthmatics using the binary target is chosen as the best of the four in terms of performance. Feature importance in Model 3 also signifies this model is the best fit for the research question and hypothesis which future research can investigate further. This is proven by the best OOB estimate of error and area under the curve signifying the greatest accuracy across the models. However, it is known that ROC AUC can be overly optimistic. According to a research study by Dr. Jesse Davis and Dr. Mark Goadrich, highly skewed datasets may need additional tools for analysis beyond the receiver operator curve. They argue that precision-recall (shown below) gives a more informative glimpse of a model's performance. The precision-recall auc for this final model is 17.6% which is inconsistent with receiver operator auc results. Therefore, it can be concluded that this study serves as a baseline for researchers to build off of and experiment with further tuning techniques.

Figure 22: Precision Recall Curve for Model 3 (Asthmatics)



Discussion & Next Steps

This analysis affirms that e-cigarette usage among adult asthmatics in 2017 is strongly associated with current or past tobacco smoking and alcohol use, meaning both hypotheses are supported. The prediction was also supported through the results of the analysis. It was discovered that age and general health indicators were meaningful contributions to the classification. While the ordinal regression offered valuable insights into predictor directionality, the binary Random Forest model outperformed in predictive accuracy and generalization. The caveats and limitations include that the model is limited to only the 2017 dataset and does not include behavioral factors such as anxiety and stress. The third model, filtered on asthmatics with a binary target, has a relatively high accuracy rate but is not ready for deployment and should be further evaluated. Improvements should be made using advanced sampling techniques and other modeling methods.

Considering caveats and limitations of survey data helps to understand the results while accounting for factors that may explain patterns or specific behaviors found during analysis. For example, data values are imputed using the mice technique because it assumes the data is missing at random. Participants may have chosen to end their survey participation after answering a survey question or they may have chosen not to respond to select questions. This dataset contains a significant amount of missing data and it is important to consider why this occurs. An attempt to find the root cause of missingness can be made but it may tough to find this answer. Comparing 2017 BRFSS data against surveys from other years in addition to working directly with the source to find ways for improvement in data collection may prove a valuable route going forward.

The recommendations include incorporating additional behavioral variables (mental health frequency, substance use) and exploring temporal BRFSS data to assess trends after 2017. Another suggestion is to evaluate calibration metrics and deploy Shapley Additive Explanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME) for deeper model interpretability. Addressing class imbalance through alternative methods, such as stacking or cost-sensitive learning, could improve the classification of minority cases.

Although we identified clear predictors of vaping behavior among asthmatics, we caution against using this model for clinical decision-making without further validation. Instead, it may serve as a tool for public health surveillance or early targeting of risk-prone populations. Next steps include retraining on multi-year datasets, adding mental health predictors, and exploring fairness metrics across race and income subgroups.

Code Availability

https://github.com/rickmorales7693/Team-Sigma

Appendix A

Codebook

Original Name	New Name	Data Type	Description	Values / Labels
_AGE80	AGE80	Numeric	Imputed age value collapsed above 80	Age in years (capped at 80)
_BMI5	BMI	Numeric	Body Mass Index (BMI)	1 or greater
_ECIGSTS	ECIG_COMP	Factor	Four-level e-cigarette smoker status	Everyday e-cigarette user, Someday e-cigarette user, Former e-cigarette user, Non-e-cigarette user
_SMOKER3	SMOKER_COMP	Factor	Conventional smoking status	Current User, Former User, Never Used
_STATE	STATE	Factor	Standard FIPS code used to identify geographic US states and territories	FIPS code for each individual state
ADDEPEV2	ADDEPEV	Factor	Ever told you had a depressive disorder	Yes, No
ALCDAY5	ALCDAY5	Integer	Days alcohol consumed in past 30 days	1-30 Days, 88 (none), 77/99 (refused/don’t know)-recoded
ASTHMA_NOW	ASTHMA_NOW_COMP	Factor	Currently has asthma	Yes = 1, No = 2, Missing = 9
ASTHMA3	ASTHMA	Factor	Ever told you had asthma	Yes, No
CHCCOPD1	CHCCOPD	Numeric	Ever told you had COPD	1 = Yes, 2 = No
CHECKUP1	CHECKUP	Factor	Length of time since last routine checkup	Within past year, Within past 2 years, Within past 5 years, 5 or more years ago
ECIG_COMP	ECIG_BINARY	Factor	Three-level e-cigarette smoker status	Current User, Former User, Other

EDUCA	EDUCA	Factor	Education Level	Never attended/kindergarten, Grades 1-8, Grades 9-11, GED/HS Grad, Some College, College Grad
GENHLTH	GENHLTH	Factor	General health status	Excellent, Very Good, Good, Fair, Poor
HLTHPLN1	HLTHPLN	Factor	Health care coverage	Yes, No
INCOME2	INCOME	Factor	Annual household income (raw)	01-08 codes, labeled
SEX	SEX	Factor	Respondent's sex	Male, Female
VapeStatus	VapeStatus	Factor	Vaping status based on _ECIGSTS	Current User, Former User, Never Used, No Response

Appendix B

Additional Figures

Figure 23: Explained Variance per Principal Component (Model 1)

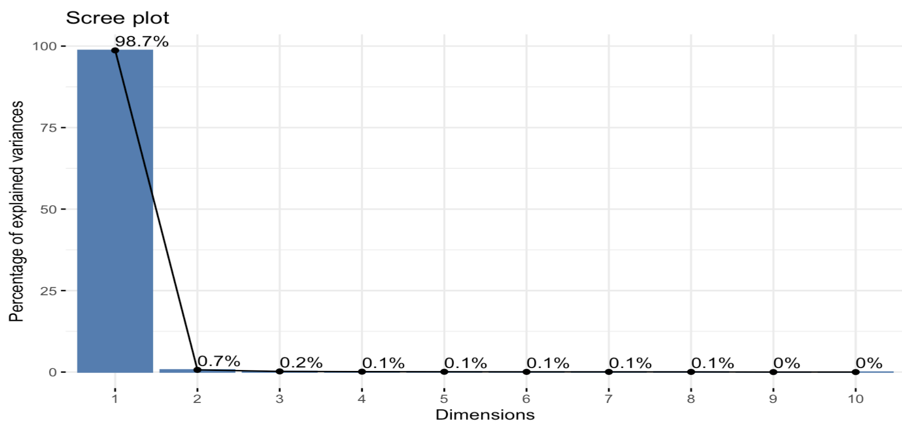


Figure 24: Explained Variance per Principal Component (Model 2)

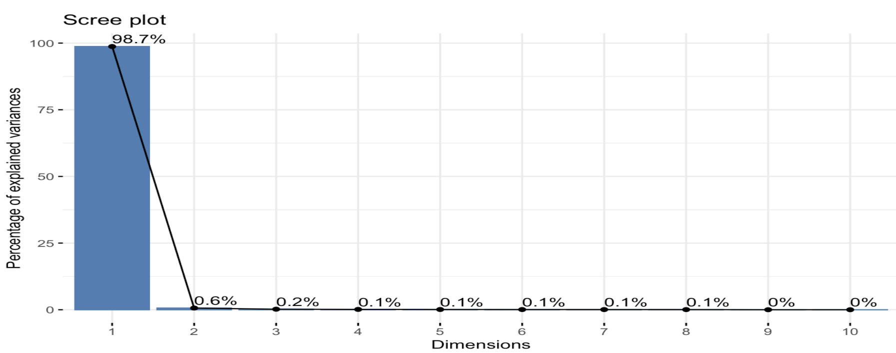


Figure 25: Explained Variance per Principal Component (Model 1)

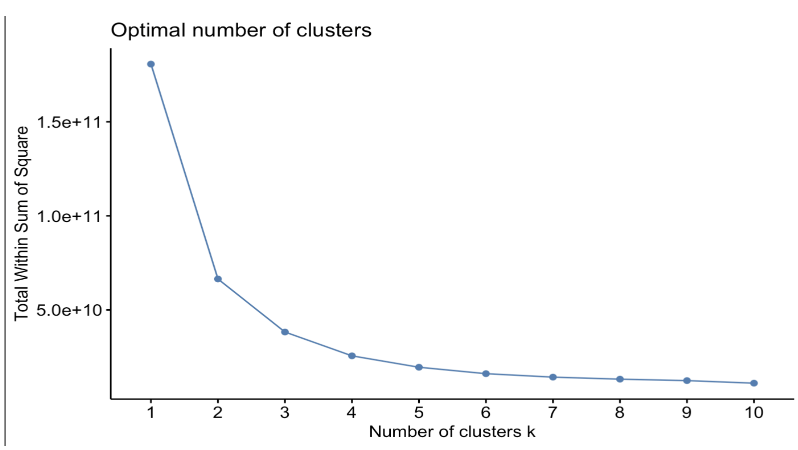


Figure 26: Explained Variance per Principal Component (Model 2)

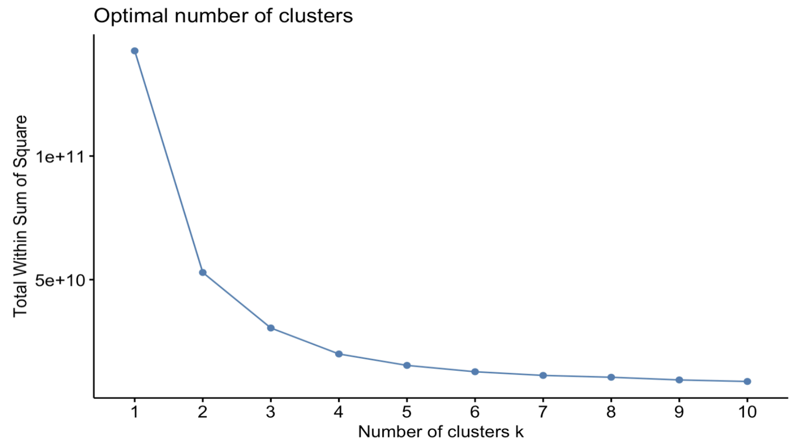


Figure 27: K-Means Clusters on Unfiltered BRFSS 2017 Data

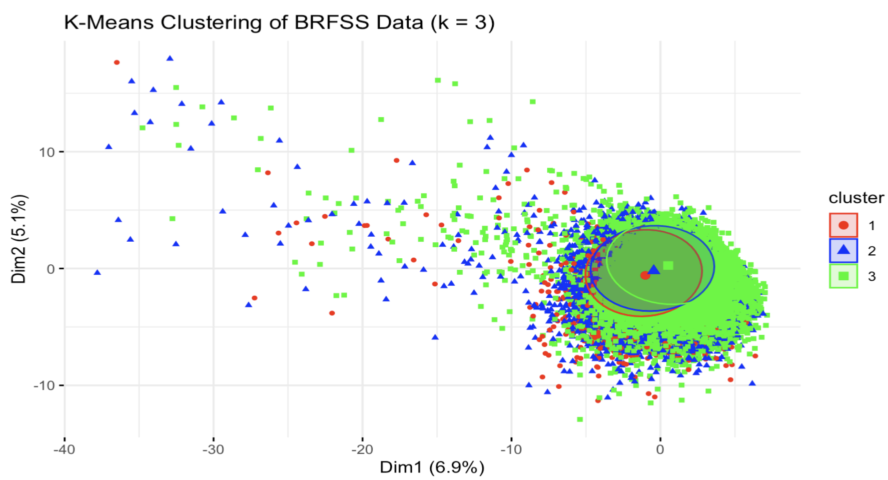


Figure 28: ROC Curve for Asthmatics with History of Vaping and Smoking (Untuned)

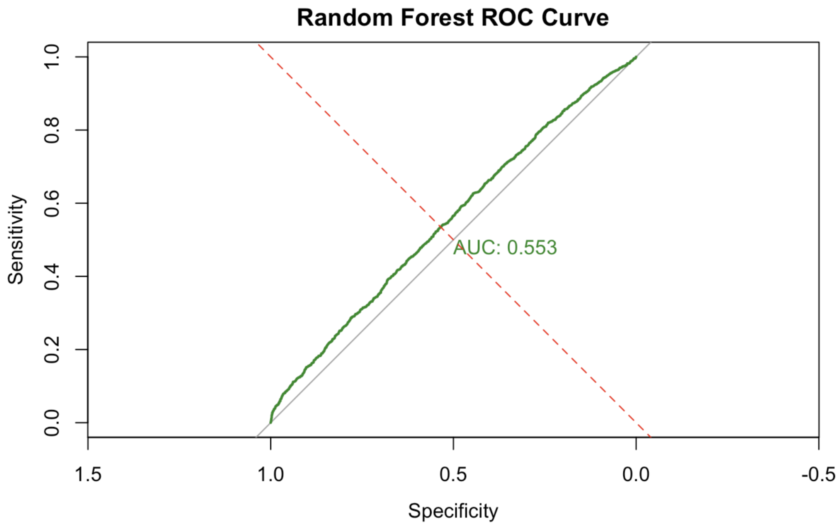


Figure 29: ROC Curve for Asthmatics (Untuned)

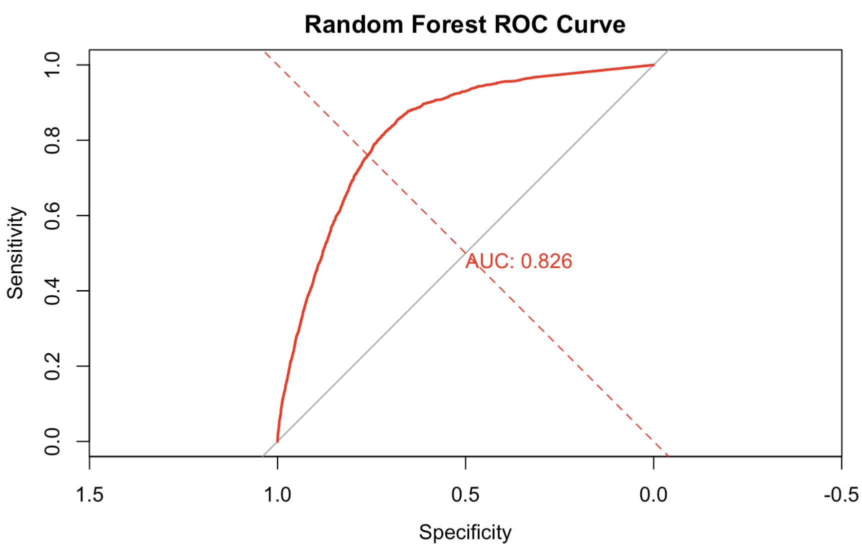


Figure 30: Out of Bag Error Estimate for Asthmatics with History of Vaping

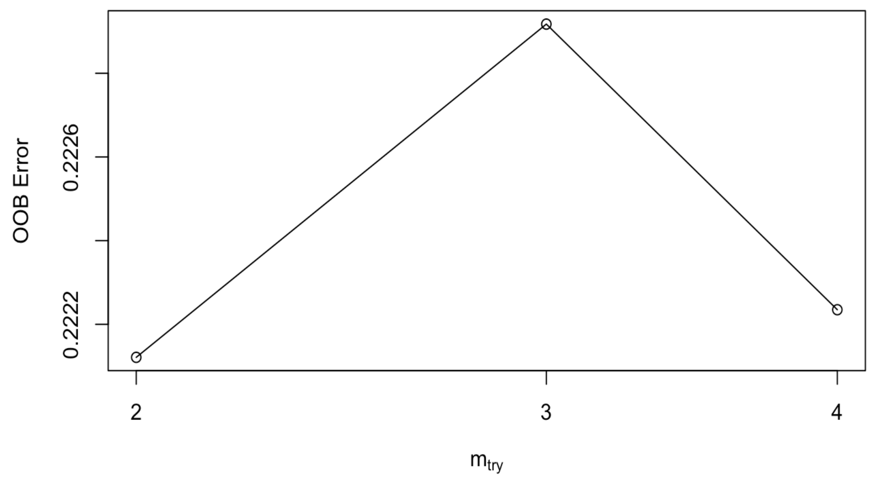


Figure 31: Out of Bag Error Estimate for Asthmatics with History of Vaping and Smoking

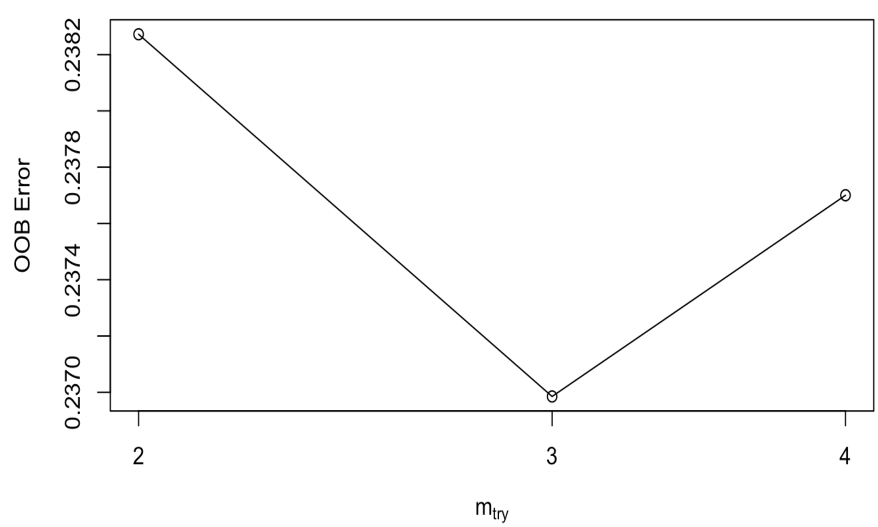


Figure 32: ROC Line Plot with 5-Fold Cross Validation and SMOTE Applied (Asthmatics with History of Vaping)

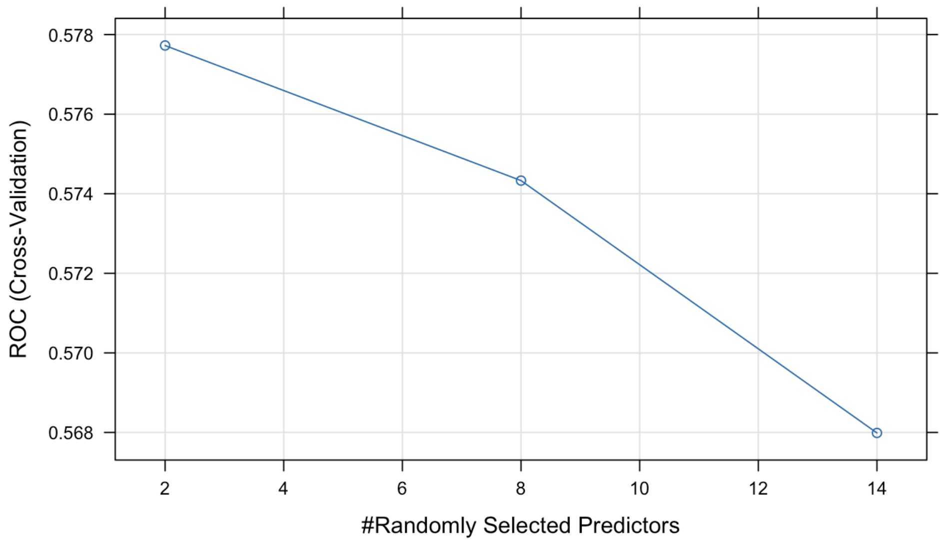


Figure 33: ROC Line Plot with 5-Fold Cross Validation and SMOTE Applied (Asthmatics with History of Vaping and Smoking)

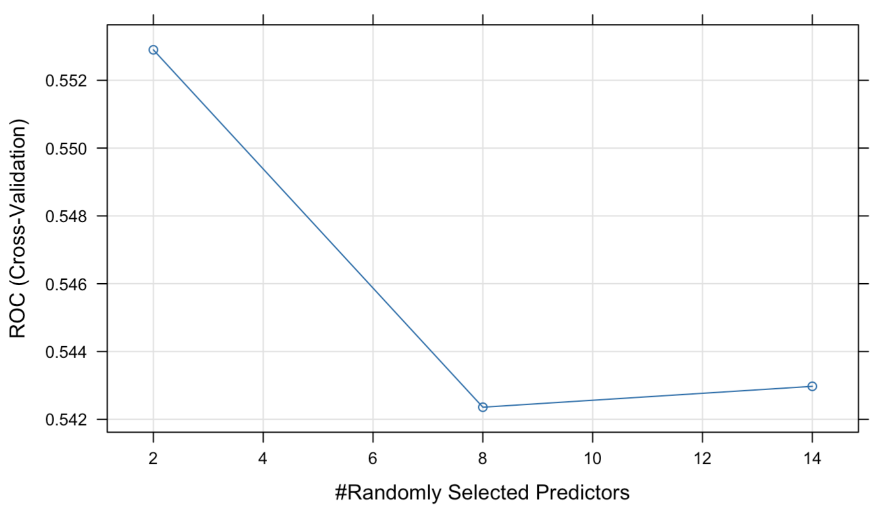


Figure 34: ROC Line Plot with 5-Fold Cross Validation and SMOTE Applied (Asthmatics)

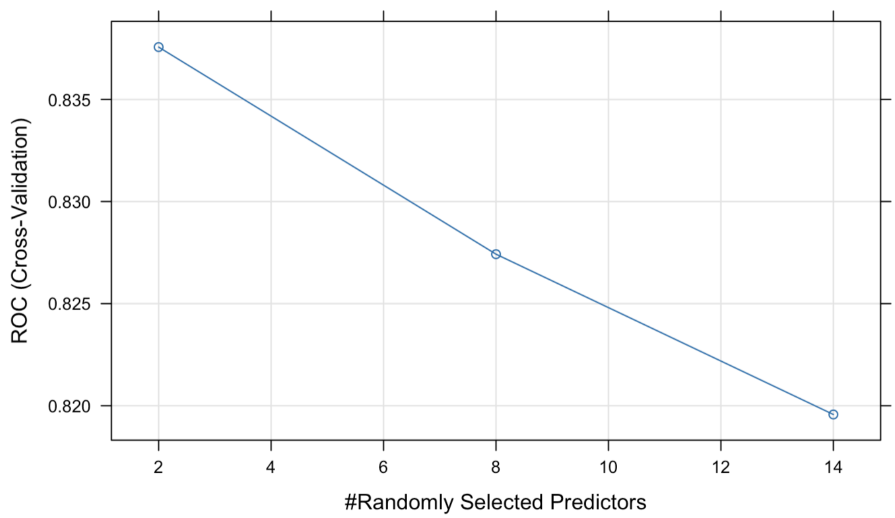


Figure 35: Precision Recall Curve for Model 1 (Asthmatics and History of Vaping)

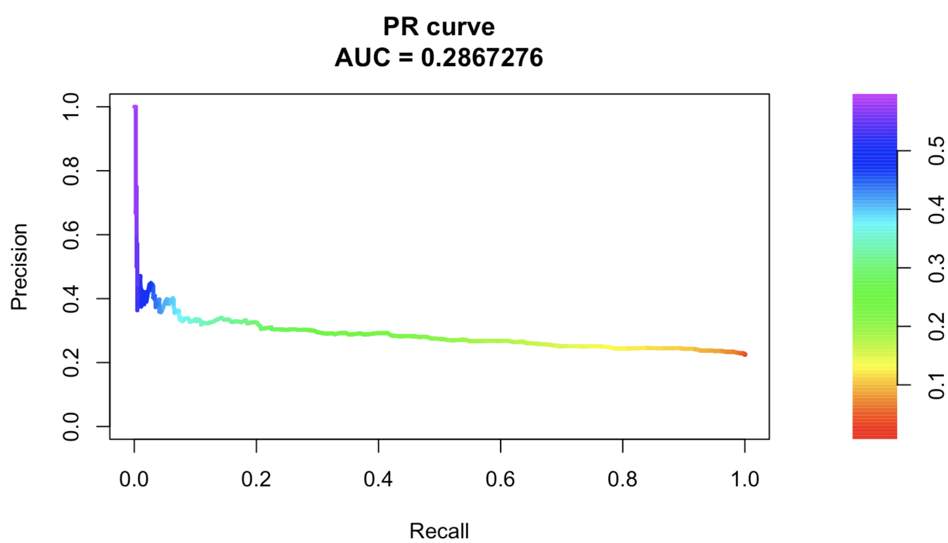


Figure 36: Precision Recall Curve for Model 2 (Asthmatics w/ History of Vaping & Smoking)

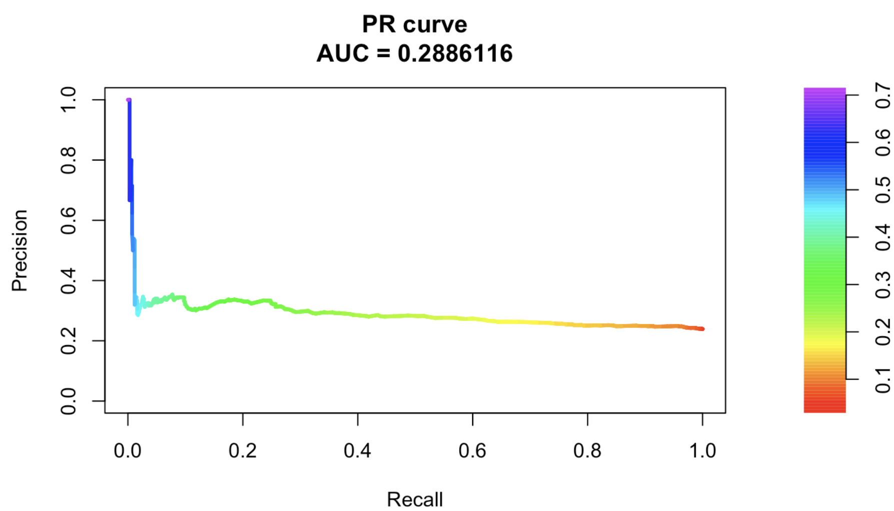
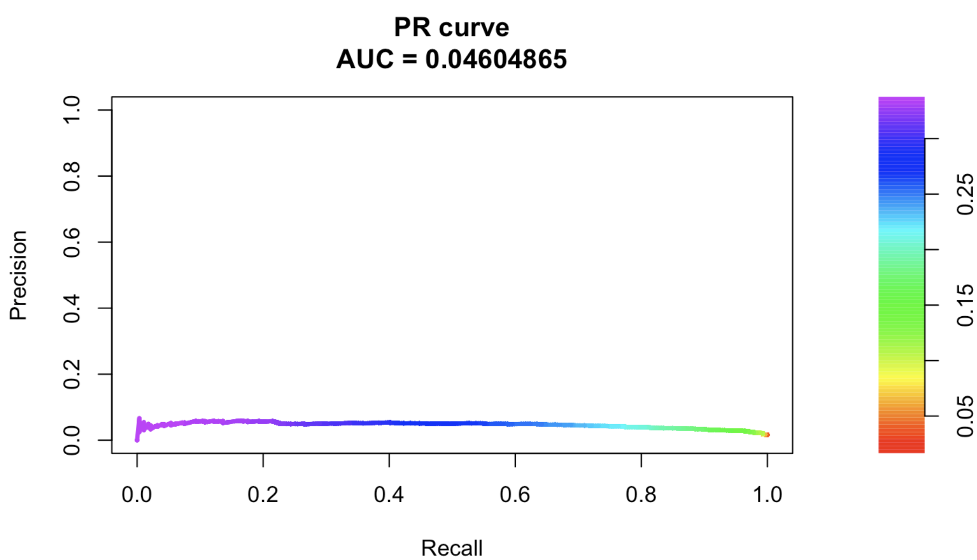


Figure 37: Precision Recall Curve for Model 4 (Multiclass, filtered on Asthmatics)



References

Jesse Davis and Mark Goadrich. 2006. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning (ICML '06). Association for Computing Machinery, New York, NY, USA, 233–240. https://doi.org/10.1145/1143844.1143874

Obisesan, O. H., Osei, A. D., Uddin, S. M., Dzaye, O., Mirbolouk, M., Stokes, A., & Blaha, M. J. (2020). Trends in e-cigarette use in adults in the United States, 2016-2018. *JAMA Internal Medicine*, *180*(10), 1394. https://doi.org/10.1001/jamainternmed.2020.2817

Roberts, J., Chow, J., & Trivedi, K. (2021). Adult-onset asthma associated with e-cigarette use. *Cureus*. https://doi.org/10.7759/cureus.19190

Solinas, A., Paoletti, G., Firinu, D., Di Pino, M., Tusconi, M., Mura, J. F., Del Giacco, S., & Marongiu, F. (2020). Vaping effects on asthma: Results from a web survey and clinical investigation. *Internal and Emergency Medicine*, *15*(4), 663–671. https://doi.org/10.1007/s11739-019-02247-5