Evaluation of Multi-class Classification Models for Census Mindsets

Mary H. Mulry, Steven Scheid, Darcy Steeg Morris and Nancy Bates[1]
U.S. Census Bureau, Washington, DC 20233

Abstract
The U.S. Census Bureau fielded the 2020 Census Barriers, Attitudes and Motivators Study (CBAMS) sample survey in 2018 to collect data on attitudes and knowledge about the U.S. Census. Data from over 17,000 respondents was used to cluster individuals into one of six psychographic profiles referred to as Census "mindsets." In social marketing campaigns, mindsets are constructed to reflect an individual's knowledge, attitudes and opinions toward a topic. The mindsets are then used in developing messages with a call to action. In our case, the requested action is responding to the 2020 Census. We recently re-administered the 2020 CBAMS to a sample selected from an online panel in anticipation of the 2020 Census, but a new classifier is needed to assign new respondents to existing mindsets. We compare several methods for fitting multi-class classification models to assign the mindsets and examine methods for model evaluation and selection. We include the evaluation method for multi-class classification models developed by Hand and Till (2001), which is a generalization of the area under the curve (AUC) for receiver operating curves (ROC) for binary classification models.

Key Words: multi-class AUC, 2020 CBAMS sample survey, 2020 Census Communications Campaign

1. Background

The U.S. Census Bureau incorporated a social marketing campaign into the 2020 Census to encourage self-response. The first communications campaign to encourage Census response occurred in 1950 when the U.S. Census Bureau began a partnership with the Advertising Council, a pro bono group of advertising agencies that create marketing campaigns for non-profit causes. For the 2000 Census, the U.S. Census Bureau decided to invest hundreds of millions of dollars in a paid advertising campaign. The success of achieving a 67 percent mail response rate when the predicted mail response rate was 61 percent led to a paid advertising campaign in the 2010 Census. The 2010 Census's 64 percent mail response rate was again higher than the predicted mail response rate of 61 percent. Experts attributed the high mail response rates to the paid advertising campaign (Bates 2017; Evans, Douglas, Datta, and Yan 2014; Williams, Bates, Lotti, and Wroblewski 2014).

As the 2020 Census approached, the agency planned for a 60.5 percent self-response rate of all modes: Internet, mail, and telephone (U.S. Census Bureau 2017). Although the COVID-19 pandemic has delayed tabulating the mail and telephone self-responses, it is apparent that the final tabulation will show an overall self-response rate that is higher than projected.

To prepare for the 2020 Census Communications Campaign, the Census Bureau conducted the 2020 Census Barriers, Attitudes and Motivators Study survey (CBAMS). The purpose of the CBAMS was to inform the 2020 Census Integrated Partnership and Communications program, specifically the messaging for the

[1] Disclaimer: Any views expressed are those of the authors and not those of the U.S. Census Bureau. Data approved by CBDRB-FY18-422 (August 13, 2018).

campaign and the development of mindsets. The CBAMS used a nationwide sample and interviews conducted in 2018. The survey achieved about 17,500 respondents. Using the answers to the CBAMS survey, the Census Bureau identified six distinct Census mindsets that were used in developing targeted messages about the Census to encourage response, particularly self-response.

Because of the importance of communications campaigns in the previous and current censuses in encouraging self-response, the U.S. Census Bureau is conducting an evaluation of the 2020 Census Communications Campaign. The evaluation consists of several studies, each focusing on different aspects of the campaign. The results of the evaluation studies are important for understanding influence the campaign had during the 2020 Census and for informing the planning of the 2030 Census.

The research described here is a component of a larger evaluation study called the Mindset Shift Study. The Mindset Shift Study is designed to examine whether the 2020 Census Communications Campaign was able to shift attitudes toward the Census in a positive direction by moving respondents from their original mindsets to mindset that is more likely to respond. Such a shift is viewed as an indicator of success of the campaign. To collect data for the analyses, the Mindset Shift Study selected new samples and administered the CBAMS questionnaire twice: (1) before campaign started in Dec 2019 and (2) during height of campaign in April 2020.

The Mindset Shift Study needs a method to assign each of the new respondents to 1 of the 6 mindsets. However, the clustering method used for assigning mindsets to the CBAMS respondents in 2018 cannot be used to classify new respondents into mindsets.

This paper describes our attempt to develop a re-classifier to assign the new respondents to mindsets. Section 2 provides background on the construction of the Census mindsets that were used in developing messages to promote Census response. Sections 3 and 4 focus on our research to find a re-classifier for the mindsets that can be applied to classify the respondents in the evaluation sample into mindsets. The final section contains a summary.

2. Background

2.1 Description of Communications Campaign

The goal of the 2020 Census Integrated Communications Campaign (ICC) is to encourage self-response in the 2020 Census through a research-based communications campaign. The Census Bureau engaged in similar social marketing campaigns that included a paid advertising campaign in both the 2000 and 2010 Census. Such campaigns include paid advertising (television, radio, print, digital, etc.) as well as earned media (such as newspaper articles and news segments) and a more community-based outreach using Partnership Specialists who partner with local elected officials, community activists, leaders and advocates to raise awareness of the Census and encourage participation.

Because the 2020 Census must count every single person living in the U.S. on April 1st, the communications and advertising campaign must be designed and implemented in a way that enables it to reach all segments of the diverse U.S. population. To help with creative message development, social marketing campaigns commonly develop psychographic profiles of the population (known as “mindsets”) according to their knowledge, attitudes, and practices towards a particular product (or in our case the 2020 Census). The 2020 Census advertising contractor, Young and Rubicam (Y&R), used data from the CBAMS to produce six such mindsets that reflect shared patterns of attitudes, behaviors, and motivators toward the 2020 Census (See Kulzick et al 2019).

The Census Bureau administered the 2020 CBAMS survey between February 20, 2018 and April 17, 2018 to 50,000 housing units in all 50 states and the District of Columbia. The survey contained questions designed to measure the public's attitudes, knowledge, and opinions regarding the 2020 Census. The results were primarily for the purposes of developing the creative platform and messaging for the 2020 Census Communications Campaign.

The sample design for the survey included stratifying the U.S. population into eight strata based on a Census tract's racial and ethnic makeup as well as characteristics related to internet response in the American Community Survey. Each household in the sample received a prepaid incentive and up to five mailings inviting them to participate by mail or Internet in either English or Spanish (for more information on this methodology, see McGeeney et al., 2019). In all, approximately 17,500 adults responded to the survey, and survey weights were constructed so that the weighted distribution of the respondents matched the distribution of all householder adults in the U.S. The details of the construction of the survey weights, including the nonresponse adjustment and the variables used in raking, may be found in Appendix B of McGeeney et al. (2019). The final, weighted response rate was 39.4 percent and was calculated using a modified version of the American Association for Public Opinion Research (AAPOR) RR3 (AAPOR 2016).

2.2 Methodology of the Mindsets

Specific details about the methodology that produced the mindsets are available in a report by Kulzick et al. (2019), but we summarize the three basic steps in creating the mindsets here. The first step was dimension reduction because the team realized that the knowledge, attitudes, barriers, and motivators measured in CBAMS sample survey reflected a smaller number of underlying factors. The team used principal component analysis (PCA) with a varimax rotation to reduce the CBAMS survey variables (45+) to eight factors that captured most of the information in the responses. In the second step, the team created candidate mindset solutions using a clustering algorithm to group respondents based on underlying similarities in these eight factors. This process identified three sets of solutions that were feasible to implement. The final selection process identifies the mindset solution that is actionable for the communications team. The criteria used by the team was whether a candidate solution provided guidance for the effective specification of marketing instruments (Wedel and Kamakura 2000). The review of the three candidate solutions involved the identification of the CBAMS survey questions with the most distinctive set of responses for each potential mindset to aid in differentiating the types of messages to target to each mindset. The process identified six distinct mindsets:

Eager Engagers are the most civically engaged mindset and have the highest knowledge about the census, as well as intent to respond. This mindset also comprises the highest percentage of college-educated people and the highest household incomes.

Fence Sitters are the largest mindset in number of CBAMS respondents. They do not have major concerns about taking the census and are less civically active than Eager Engagers, but they are still highly inclined to respond. This mindset is the least diverse and has the highest percentage of males.

Confidentiality Minded are most concerned that their answers to the census will be used against them, but they believe their answers matter and are still fairly likely to respond. This mindset is the most diverse and has the highest percentage of foreign-born people.

Head Nodders are most likely to give affirmative answers to all knowledge questions and demonstrate significant knowledge gaps in specific areas. This mindset has the highest percentage of people 18-34 years old and above average percentage of foreign-born people.

Wary Skeptics are skeptical of the government, as shown by their high distrust of the government, and are, therefore, reluctant to participate in the census. This mindset has the highest percentage of Black/African-Americans and below average education attainment.

Disconnected Doubters do not use or have access to the internet, do not believe that their response matters, and are the least likely to respond to the census. This mindset has the highest percentage of people 65 years or older, and has the lowest levels of education.

The six mindsets were structured around the following four questions:

- o Who are they?
- o Do they intend to respond, and how do they think about the census?
- o What are their potential barriers to participation?
- o What are their potential motivators for participation?

Figure 1 is a visual dashboard of key characteristics of the final six mindsets, which, in order of the percentage who intend to respond, are: (1) Eager Engagers, (2) Fence Sitters, (3) Confidentiality Minded, (4) Head Nodders, (5) Wary Skeptics, and (6) Disconnected Doubters (McGeeney et al 2019).

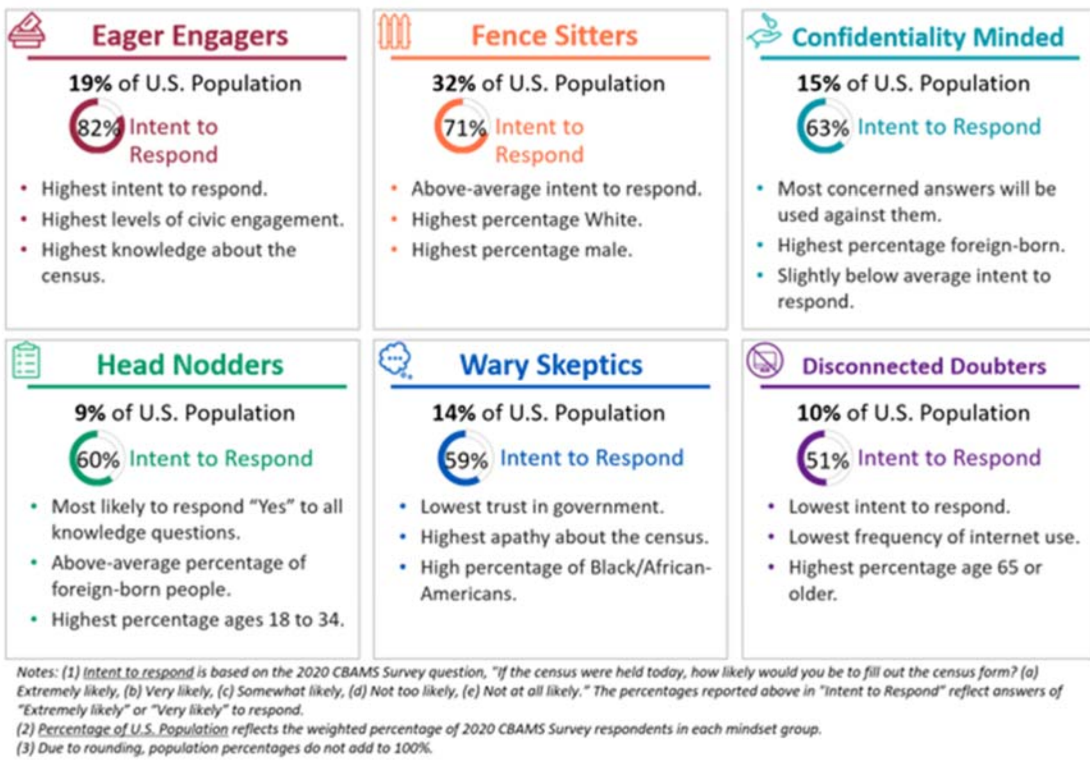

Figure 1. Overview of Mindsets (McGeeney et al 2019)

Of the six mindsets, Eager Engagers and Fence Sitters have a positive attitude toward the 2020 Census although some of their characteristics are different. Both the Confidentiality Minded and the Head Nodders are neutral towards the 2020 Census, but they have different concerns and attitudes which required the campaign to use different types of messaging to address their concerns. The last pair, Wary Skeptics and Discontented Doubters, are somewhat negative toward the Census but have different reasons and concerns which influenced the types of messages the campaign used to encourage their response.

2.3 The Mindset Shift Study and the need for a re-classifier

The goal of the Mindset Shift Study is to measure whether (and how) mindsets may have shifted over the course of the campaign. One way to measure the success of the campaign is to gauge whether the portion of the population with mindsets less inclined to participate in the census shrank over the course of the campaign while the portion of the population with mindsets more inclined to participate grew larger. This approach is similar to one used in an evaluation of the 2010 Census Communications Campaign (see Bates and Mulry, 2012).

To collect data for the Mindset Shift Study, the U.S. Census Bureau purchased a sample from the Knowledge Panel (Ipsos Knowledge Networks). The Knowledge Panel is a probability-based nationally representative online survey. This particular panel has been used to evaluate other public information campaigns (see Vallone et al., 2015; McAffee, et al. 2013; and Bigsby, Cappella and Seitz, 2013). The evaluation administered the CBAMS survey questionnaire to the sample from the Knowledge Panel in two waves – one *prior* to the campaign startup and one at the *height* of the campaign. The goal in this approach is to determine whether respondents' mindsets shift between the two waves, presumably as a result of the campaign. For example, if the Eager Engager mindset expands and the Disconnected Doubters shrinks, we have some insights into whether the campaign worked to change knowledge, attitudes and opinions around the census.

The U.S. Census Bureau also administered the CBAMS questionnaire to a second fresh sample from the Knowledge Panel that coincided with the second wave. The fresh sample will be used to evaluate potential panel conditioning effects on the respondents in the other sample who received the CBAMS questionnaire twice.

A mindset re-classifier, which is the focus of the current research, is required because each respondent in the evaluation samples needs to be assigned to a mindset. The clustering method used for the CBAMS respondents prevents us from easily defining mindsets and classifying new cases. We need to develop a new classification model using the CBAMS data to be able to assign mindsets to new respondents. For this reason, the evaluation sample respondents received the same questionnaire that was used in the CBAMS and the data processing was as similar as possible.

Several models considered for the re-classifier are explored in Sections 3 and 4. We use the CBAMS Public Use Microdata Sample (PUMS) dataset (U.S. Census Bureau 2018) in fitting the models considered for the re-classifier. However, after we select the type of model to use for the re-classifier, we will fit the chosen type of model using the CBAMS dataset to construct the model that will be applied to the evaluation samples.

2.4 Research questions

The focus of our research is on answering the following questions:

1) What type of multi-class classification model produces the best classifier for assigning mindsets to the evaluation samples (described in Section 2.3)?

2) Is the multi-class AUC, an evaluation method for multi-class classification models developed by Hand and Till (2001), helpful in identifying the new classifier for assigning mindsets?

3) What other evaluation method(s) are helpful in identifying the new classifier?

3. Methodology

3.1 Strategy for finding re-classifier for mindsets

Our strategy for identifying a re-classifier includes applying five types of multi-class classification models using the CBAMS PUMS file. Then we compare several quality measures for each of the classification methods to select a well-performing re-classifier. The CBAMS PUMS data was transformed into eight factors using Principal Components Analysis with the same survey variables used to create the mindsets in order to replicate the formation of the mindsets as close as possible. We use the 8 factors as independent variables when fitting the candidates for the re-classifier model and do not include variable selection.

Our approach to constructing the multi-class classification model requires randomly partitioning the CBAMS PUMS data set into two equally sized datasets: one for training and one for testing the model. Then we use the "Train" dataset to create one model of each type and validate the stability of the results using a cross-validation. Some types of models that we consider require identifying the optimal values of tuning parameters as part of finding the best model of its type for our data. Next, we apply the best model in its category to the "Test" dataset to estimate how it will perform on new data. We calculate several performance characteristics, which are described in Section 3.2, for use in a comparison with the other types of models. We identify the model with the best performance on the Test dataset using those criteria. The characteristics used in the comparison of the models focus on the multi-class AUC, the correct classification rate (CCR) and the range of the correct classification rates across the mindsets. For additional insight, we also construct a matrix that compares the distribution of the classifications in the Test dataset to the original classifications.

3.2 Quality measures for classification models

The assessment of the quality of the models under consideration uses three criteria: the multi-class AUC, the correct classification rate, and the range of the correct classification rates across the classes. These quality measures are defined below.

- **Multi-class AUC**

The multi-class area under the curve, which we call the multi-class AUC (Hand and Till 2001) provides an indicator of the overall performance of a classification rule in separating the classes. The algorithm uses the probabilities produced for the mindsets by a classifier model to calculate the multi-class AUC, which has an interpretation similar to the ROC curve for binary classification.

The following is a brief definition of the multi-class AUC, but for a complete discussion, see Hand and Till (2001). Suppose we have *c* classes labeled *0, 1, 2, …, c-1, where c > 1*. For $i \neq j$, define $\widehat{A}\,(i|j)$ as follows:

$\widehat{A}\,(i|j)$= the probability that a randomly drawn member of class j has a lower estimated probability of belonging to class j than a randomly drawn member of class i.

Then, $\widehat{A}\,(i|j)$ and $\widehat{A}\,(j|i)$ may be used to define a measure $\hat{A}(i,j)$ that reflects the ability of the classifier model to separate class *i* from class *j* as follows:

$$\hat{A}(i,j) = \frac{\hat{A}(i|j)+\hat{A}(j|i)}{2}.$$

Subsequently, a measure that reflects the ability of the classifier model to distinguish all the classes from each other, called the *Multi-class AUC*, maybe be defined as follows:

$$\textit{Multi-class AUC} = \frac{2}{c(c-1)}\sum_{i<j}\hat{A}(i,j)$$

We use two software programs to implement the multi-class AUC: the R package `HandTill2001` (2001) in R and the SAS Macro `MultAUC` (SAS 2019) in SAS.

- **Correct classification rate**

The correct classification rate (*CCR*) is a measure of the accuracy of a classifier model. The definition below shows that it is the proportion of the records that are classified properly.

Suppose we have *c* classes and a dataset of size *n*. Let the number of records correctly classified in class *i*, where *i = 1, 2,….c*, be denoted by *cc(i)*. Suppose the size of class $c(i) = n_i$, Then define the *CCR* as follows:

$$CCR = \frac{\sum_{i=1,..n} cc(i)}{\sum_{i=1,..n} n_i}$$

Note that the sum of the class sizes n_i equals the size of the dataset.

- **Range of misclassification rates across classes**

The range of the correct classification rates across the classes indicates the overall effectiveness of the model. The definition of the correct classification rate for class *i* (CCR_i) is

$$CCR_i = \frac{cc(i)}{n_i}$$

We want to avoid a situation where the model has high correct classification rates for some classes and low classification rates for others. In particular, lopsided situations where the correct classifications are concentrated in the larger classes leaving the smaller classes with much lower correct classification rates are undesirable.

- **Cross-validation**

Cross-validation is a method of assessing how a model will perform on a new dataset. The technique requires partitioning the user's data set into a specified number, say *k*, subsets. For subset *i*, the model is fit using all the data except i^{th} subset and then applied to the i^{th} subset. These steps are repeated for the remaining *k-1* subsets. We illustrate the subsequent steps for the correct classification rate but these steps apply to other measures of accuracy. To obtain the cross-validated value of the correct classification rate, we first calculate the correct classification rate for the mindsets predicted for each of the *k* subsets. The average correct classification rate for the predicted mindsets over all the *k* subsets is the cross-validated correct classification rate. A cross-validation that uses a partition of the dataset into *k* subsets is called a "k-fold cross-validation."

3.3 Types of models

We considered the following 5 types of models in our search for a re-classifier.

3.3.1 *Linear discriminant analysis* produces a set of linear combinations of the quantitative variables that reveal the differences among the classes. We used SAS `PROC DISCRIM` with the option that `METHOD=NORMAL` that assumes a parametric method based on a multivariate normal distribution within each class is used to derive a linear discriminant function.

3.3.2 *Quadratic discriminant analysis* is similar to linear discriminant analysis except that the covariance matrix can be different for each class and the discriminant function contains second order terms. We used SAS `PROC DISCRIM` with the option `METHOD=NORMAL`, that assumes a parametric method based on a multivariate normal distribution within each class is used to derive a quadratic discriminant function.

3.3.3 *Multinomial logistic regression* is used to model nominal outcome variables, in which the log odds of the outcomes are modeled as a linear combination of the independent variables. These results can be converted to probabilities for each observation that indicate how likely the observation is to be in each class. We used SAS `PROC LOGISTIC` with the option `UNEQUAL SLOPES` and the R package `nnet` (Venables & Ripley 2002) to fit multinomial logistic regression models.

3.3.4 *Random forest* is a classification method that consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest estimates a class prediction and the class with the most votes becomes the random forest class prediction. We used the R package `randomForest` (Liaw & Wiener 2002) in our analyses.

3.3.5 *Support vector machine* (SVM) is a machine learning method designed for finding the optimal separating hyperplane between two classes by maximizing the margin between the closest points in the two classes. The method has been extended to multi-class classification by fitting binary sub-classifiers and finding the correct class by a voting procedure. In our analyses, we used the R package `e1071` (Meyer et al. 2017), which functions as an interface to another R package `libsvm` (Chang and Lin 2001). We used the option `cross=10` that produces a 10-fold cross-validation.

4. Results for Five Classification Models

4.1 Tuning Parameter Selection for SVM and Random Forest

We use CCR and multi-class AUC to choose the optimal tuning parameters for the SVM and random forest methods with a 10-fold cross-validation on the training data. For SVM, we use the default kernel option of a radial basis function in the R package `e1071`. The fit of the SVM depends on the choice of tuning parameters *cost* – a penalty parameter that controls penalty of misclassification – and *gamma* – a kernel parameter that controls the radius of the area of influence. Both cost and gamma take values that range between 0 and 1. To find the optimal value of both parameters, we jointly vary *cost* = (.5, 1, 2, 3, 4, 5) and *gamma* = (.25, .5, .75, 1, 2), where large values of both may lead to overfitting the data. Figure 2 shows the value of CCR and multi-class AUC for all combinations of the two tuning parameters. We find that the highest multi-class AUC (0.970) is attained at *cost = 2* and *gamma = 0.5* and the highest CCR (0.851) is attained at *cost = 2* and *gamma = 0.75*. Multi-class AUC and CCR agree on the optimal *cost* tuning parameter. We decided to select *gamma = 0.75* because it produces the highest value of CCR when *cost = 2* and the multi-class AUC is not very sensitive to changes in the neighborhood of *gamma=0.75* and *cost=2*. Therefore, the SVM model with *cost = 2* and *gamma = 0.75* is used in the rest of our analyses.

For random forest, we use the default node characteristics and cutoff in the R package `randomForest`. The random forest algorithm further depends on the choice of tuning parameters *ntree* – the number of trees – and *mtry* – the number of predictors in each tree. We jointly vary *ntree* from 100 to 2000 by 100 and *mtry* from 1 to 8 by 1. Figure 3 shows the value of CCR and multi-class AUC for all combinations of the two tuning parameters.

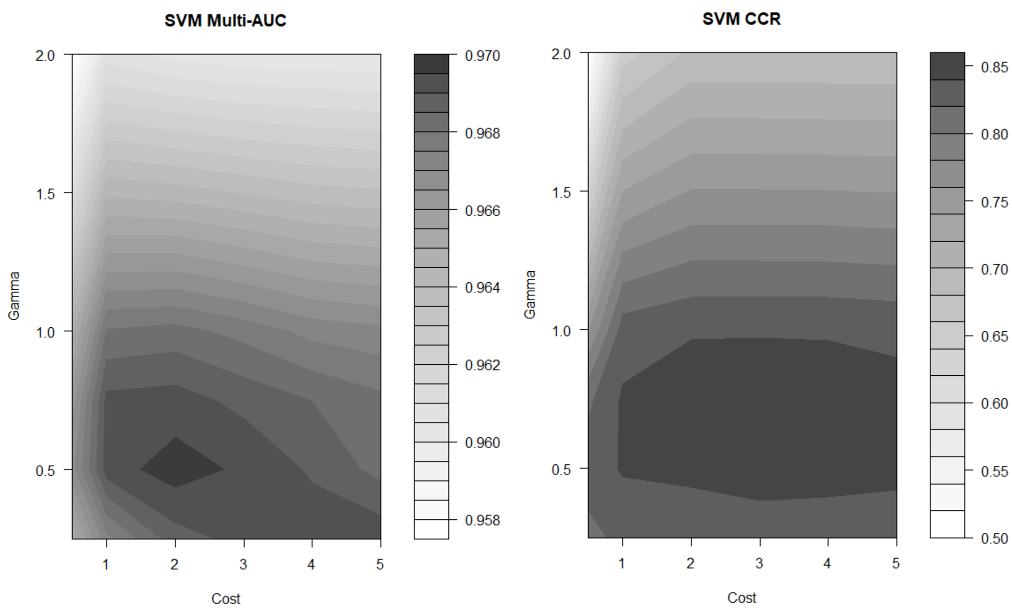

Figure 2: Multi-class AUC and CCR for all SVM tuning parameters.

We find that the highest multi-class AUC (0.964) is attained at *mtry = 2* and *ntree = 2000* and the highest CCR (0.798) is attained at *mtry = 2* and *ntree = 1600*. Multi-class AUC has similar values across the range of *ntree*, but decreases as the *mtry* increases after two. CCR also decreases as the *mtry* increases after two, but varies more across *ntree*. Multi-class AUC and CCR agree on the *mtry=2* optimal value, and we will use the CCR optimal value of *ntree = 1600* because multi-class AUC is not very sensitive to the choice of this parameter. The random forest model with tuning parameters *ntree = 1600* and *mtry=2* is used in the rest of the analysis.

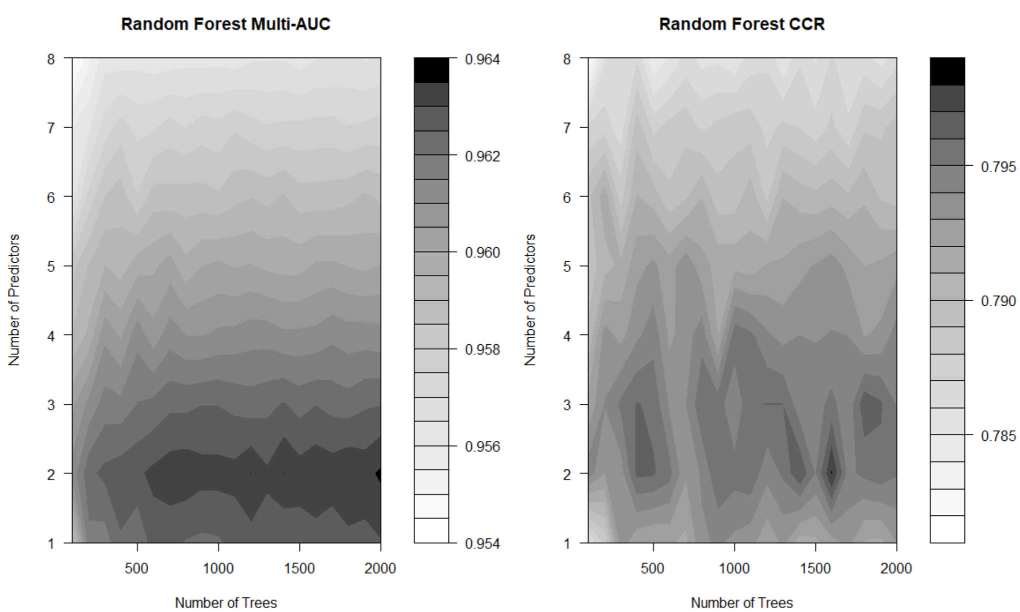

Figure 3: Multi-class AUC and CCR for all random forest tuning parameters.

4.2 Model Comparison on Training Data with Cross-Validation

We fit each model listed in Section 3.3 on the training data with 10-fold cross-validation. That is, we fit each model on 9/10 of the Train data, evaluate the predictions on the other 1/10 of the Train data, and repeat this process ten times. Table 2 shows the cross-validated (CV) values of the multi-class AUC, CCR, and CCR_i for the five models. These measures are the average across the ten cross-validation datasets. We use these cross-validated values as the basis for selecting what we consider the best model. We see that the linear discriminant analysis has the lowest value of both the multi-class AUC and CCR. The SVM model has the highest values of both the multi-class AUC and CCR as well as the highest minimum and maximum CCR_i. This leads us to conclude that the SVM is the best re-classifier for the CBAMS PUMS data.

Table 2. Comparison of models fit on the training dataset with cross-validation (8,614 records)

Model	CV Multi-class AUC	CV CCR	Range of CV CCR_i
linear discriminant analysis	0.9396	0.7171	0.609 – 0.795
quadratic discriminant analysis	0.9510	0.7582	0.619 – 0.826
multinomial logistic regression	0.9415	0.7251	0.632 – 0.789
random forest	0.9635	0.7945	0.711 – 0.851
support vector machine	0.9693	0.8507	0.771 – 0.894

4.3 Model Comparison on Test Data

We validate our model selection findings by applying each of the five selected models to the Test dataset. We compare multi-class AUC, CCR, and CCR_i to assess the quality of the classifications produced by the models. The results for these evaluation criteria are shown in Table 3. The SVM model has the highest values of the performance measures with a CCR of 0.8573 and a multi-class AUC of 0.9722. These results based on the test data are consistent with the finding from cross-validation with the training data.

Table 3. Comparison of models fit on the Test dataset (8,669 observations)

Model	Multi-class AUC	CCR	CCR_i
linear discriminant analysis	0.9422	0.7296	0.601 – 0.813
quadratic discriminant analysis	0.9531	0.7643	0.623 – 0.828
multinomial logistic regression	0.9441	0.7340	0.619 – 0.801
random forest	0.9669	0.8009	0.715 – 0.842
support vector machine	0.9722	0.8573	0.821 – 0.889

The cross-tabulation of the observed mindsets by the SVM-predicted mindsets in Table 4 provides insight about the pattern of the differences between the observed and predicted mindsets. The correct classification rates for the individual mindsets (CCR_i) along the diagonal range from 82.1 percent for Wary Skeptics to 88.9 percent for Fence Sitters. The correct classification rate for the SVM is over 85 percent for three mindsets.

Table 4. Cross-tabulation and row percent of observed mindsets by the SVM-predicted mindsets for the test dataset (8,669 observations)

	SVM Predictions						
Observed	Eager Engagers	Conf Minded	Fence Sitters	Head Nodders	Wary Skeptics	Discon Doubters	Total
Eager Engagers	1516 86.4%	12 0.7%	151 ***8.6%***	19 1.1%	50 2.8%	6 0.3%	1754
Conf Minded	33 2.6%	1046 82.6%	110 ***8.7%***	22 1.7%	40 3.2%	16 1.3%	1267
Fence Sitters	150 ***5.2%***	66 2.3%	2564 88.9%	25 0.9%	59 2.0%	20 0.7%	2884
Head Nodders	21 2.8%	15 2.0%	69 ***9.1%***	625 82.7%	18 2.4%	8 1.1%	756
Wary Skeptics	36 3.3%	37 3.4%	83 ***7.7%***	18 1.7%	885 82.1%	19 1.8%	1078
Discon Doubters	12 1.3%	23 2.5%	43 4.6%	19 2.0%	37 4.0%	796 85.6%	930

Note: The five largest row misclassification percentages are in bold.

Table 4 shows that four mindsets have between 7.7 percent and 9.1 percent of their observations misclassified as Fence Sitters, which is the largest mindset. We note that the two largest misclassification cells are the Eager Engagers misclassified as Fence Sitters at 151 and the Fence Sitters misclassified as Eager Engagers at 150. The counts in these misclassification cells are almost equal, but the size difference between the two mindsets creates disparity in impact. For the Eager Engagers, the misclassifications as Fence Sitters account for 8.6 percent of its observations while for Fence Sitters, only 5.2 percent of its observations are misclassified as Eager Engagers.

We examined the cross-tabulations of the observed mindset assignments by the predicted mindsets for each of the other models, and saw the same troubling misclassification pattern between Eager Engagers and Fence Sitters. As the largest mindset, Fence Sitters appeared to be pulling in more than its share of misclassifications of the other mindsets. To further understand any patterns in this misclassification across models, we examine the cross-tabulation of predictions from the best performing models, random forest and SVM, in Table 5 to see where their classifications agree and disagree.

The cell percentages in Table 5 show that the predicted mindsets from random forest (RF) and SVM agree for about 85 percent of the observations in the Test dataset. Disagreements between three pairs of mindsets account for over half (8.1 percent) of the remaining 15 percent. The three pairs are Fence Sitters and Eager Engagers (2.2 + 1.8 = 3.0 percent), Fence Sitters and Confidentiality Minded (1.1 + 1.1 = 2.2 percent), and Fence Sitters and Wary Skeptics (0.9 + 1.0 = 1.9 percent). The symmetry between the three pairs of misclassification cells may be an indicator that the observations in each pair are ‘close,’ meaning that one or both of RF and SVM have difficulty in distinguishing between them. A supporting factor is that the remaining disagreement cells in Table 5 have cell percentages of 0.8 percent or less and do not exhibit symmetry in the size of the pairs of disagreement percentages. To illustrate the lack of symmetry in the remaining misclassification cells, the cell percentage of the observations that RF classifies as Head Nodders but SVM classifies Fence Sitters is 0.8 percent. However, the cell percentage for the observations that RF classifies as Fence Sitters and SVM classifies as Head Nodders is only 0.4 percent.

Table 5. Cross-tabulation and cell percent for random forest mindset predictions by SVM mindset predictions for the test dataset (8,669 observations)

Random Forest Predictions	SVM Predictions					
	Eager Engagers	Conf Minded	Fence Sitters	Head Nodders	Wary Skeptics	Discon Doubters
Eager Engagers	1491 17.2%	16 0.2%	157 ***1.8%***	15 0.2%	61 0.7%	5 0.1%
Conf Minded	20 *0.23%*	1012 11.7%	97 ***1.1%***	17 0.2%	34 0.4%	12 0.1%
Fence Sitters	194 ***2.2%***	96 **1.1%**	2587 29.8%	32 0.4%	79 **0.9%**	46 0.5%
Head Nodders	23 0.2%	18 0.2%	72 0.8%	651 7.5%	31 0.4%	14 0.2%
Wary Skeptics	30 0.3%	52 0.6%	83 ***1.0%***	7 0.1%	840 9.7%	28 0.3%
Discon Doubters	6 0.1%	5 0.1%	24 0.3%	6 0.1%	44 0.5%	760 8.8%

Note: The six largest disagreement cell percentages are in bold.

5. Summary

Our analyses focused on identifying the best type of multi-class classification model to use in assigning each respondent in the samples selected for the Mindset Shift Study to one of the six mindsets developed for the 2020 Census Communications Campaign. The Mindset Shift Study is part of the evaluation of the 2020 Census Communications Campaign. The mindsets were used in developing messages to persuade the public to participate in the 2020 Census.

The goal of the Mindset Shift Study is to assess whether the campaign was able to shift attitudes toward responding to the 2020 Census in a positive direction. The evaluation administered the CBAMS survey questionnaire to samples selected from a panel in late December 2019 before the campaign began and again at the height of the campaign in late April 2020. The purpose is to see if the distribution of the mindsets shifted in a positive direction over the course of the campaign.

Our search for a multi-class classifier to use in assigning mindsets to the evaluation samples led us to consider five types of models: linear discriminant analysis, quadratic discriminant analysis, multinomial logistic regression, random forest, and support vector machine (SVM). Our analyses indicated that SVM with optimal parameters demonstrated the best performance. It had the highest value of each of the following three evaluation criteria over all the models considered: the highest overall correct classification rate, the multi-class AUC and the range of the correct classification rates for the individual mindsets.

The chosen SVM model with the optimal parameters does have a weakness in that the rate of misclassification between the two positive mindsets (Eager Engagers and Fence Sitters) is higher than we would prefer – a weakness shared by the other models as well. However, the main goal is to determine whether the four non-positive mindsets (two negative and two neutral) shifted to a more positive mindset. Having some ambiguity about whether the shift was to Eager Engagers or Fence Sitters is less important than determining there was a shift toward a positive attitude about the 2020 Census.

We found the multi-class AUC to be a helpful supporting measure in evaluating the models. Alongside the overall correct classification rate and the range of the correct classification rates for the individual mindsets, the multi-class AUC aided in weighing the strengths and weaknesses for each model.

The next step in the Mindset Shift Study will be to fit and tune a SVM model using the CBAMS survey data collected in 2018. The resulting SVM will be used to assign mindsets to the newly collected data from evaluation survey respondents for use in determining whether the mindsets shifted during the 2020 Census Communications Campaign. The results also will inform planning for the 2030 Census and other programs at the U.S. Census Bureau.

7. References

AAPOR (2016), "Standard Definitions Report, Ninth Edition," AAPOR. Available at https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf (Accessed September 2020).

Bates, N. (2017), "The Morris Hansen Lecture. Hard-to-Survey Populations and the U.S. Census: Making Use of Social Marketing Campaigns," *Journal of Official Statistics*, Vol. 33, No. 4, 2017, 873–885. doi:10.1515/jos-2017-0040.

Bates, N., and M. H. Mulry, (2012). "Did the 2010 Census Social Marketing Campaign Shift Public Mindsets?" *Proceedings of 2012 AAPOR conference in the Proceedings of the American Statistical Association Survey Research Methods Section.* Alexandria, VA. 5257-5271. Available at http://www.asasrms.org/Proceedings/y2012f.html (Accessed September 2020).

Bigsby, E., Joseph N. Cappella, J.N., & Seitz, H.H. (2013). Efficiently and Effectively Evaluating Public Service Announcements: Additional Evidence for the Utility of Perceived Effectiveness. *Communication Monographs, 80*(1): 1–23. doi:10.1080/03637751.2012.739706.

Chang, C.C. and Lin, C.J. (2001) LIBSVM: A library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, detailed documentation (algorithms, formulae, . . .) available at http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.ps.gz (Accessed September 2020).

Evans, W., A. Douglas, R. Datta, and T. Yan (2014), "Use of paid media to encourage 2010 Census participation among the hard to count," in *Hard-to-Survey Populations*, eds. R. Tourangeau, B. Edwards, T. P. Johnson, K. Wolter, and N. Bates, pp. 519-540, Cambridge, UK: Cambridge University Press.

Hand, D. J. and R. J. Till (2001). "A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems." *Machine Learning,* 45, 171-186. Available at https://link.springer.com/content/pdf/10.1023/A:1010920819831.pdf (Accessed September 2020).

Kulzick,R., Kail, L., Mullenax, S., Kriz, B., Shang, H., Walejko, G., Vines, M., Bates, N. Scheid, S. and García Trejo, Y. (2019). *2020 Census Predictive Models and Audience Segmentation Report*. 2020 Census Research Memorandum Series. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/programs-surveys/decennial-census/2020-census/research-testing/communications-research/2020-tract-segments.html (Accessed September 2020).

Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

McAfee, T., Davis, K.C. Alexander, R.L. Pechacek, T.F., and Bunnell, R. (2013). Effect of the first federally funded US antismoking national media campaign. *The Lancet*, *382*(9909), 2003-2011. doi:10.1016/S0140-6736(13)61686-4.

McGeeney, K., Kriz, B., Mullenax, S., Kail, L., Walejko, G., Vines, M., Bates, M., & García Trejo, Y. (2019). *2020 Census Barriers, Attitudes, and Motivators Study Survey Report*. Suitland, MD: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-cbams-study-survey.pdf (Accessed September 2020).

Meyer, D. (2019) Support Vector Machines, the interface to libsvsm in package e1071. https://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf (Accessed September 2020).

Meyer, D, E. Dimitriadou, K. Hornik, A. Weingessel, F.Leisch (2017). *Package 'e1071'*. Cran. https://CRAN.R-project.org/package=e1071 (Accessed September 2020).

SAS (2019). *Sample 64029: Area under the ROC curve measure (AUC) for multinomial models.* https://support.sas.com/kb/64/029.html#hist (Accessed September 2020).

U.S. Census Bureau (2017). 2020 Census Life-Cycle Cost Estimate Executive Summary. Version 2.0." Washington, DC: U.S. Census Bureau.
https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/life-cycle-cost-estimate_v2.pdf (Accessed September 2020).

Vallone, D.M., Ilakkuvan, V., Xiao, H., Cantrell, J., Rath, J. and Hair, E. (2015). Contextual Information and Campaign Awareness Among Young Adults: Evidence from the National truth® Campaign. *Behavioral Medicine*, *41*(3), 155-163. doi: 10.1080/08964289. 2015.1036832.

Venables, W. N. and Ripley, B. D. (2002.). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Williams, J. D., N. Bates, M. A. Lotti, and M. J. Wroblewski (2014), "Marketing the 2010 Census: Meeting the Challenges of Persuasion in the Largest-Ever Social Marketing Campaign," in *The Handbook of Persuasion and Marketing*, ed. D. W. Stewart, pp. 117–154, Santa Barbara, CA: Praeger.

Wedel, M. and W. A. Kamakura (2000). *Market segmentation: Conceptual and Methodological Foundations*. NewYork, NY: Springer US.