American Journal on Intellectual and Developmental Disabilities

Using Machine Learning to Predict Patterns of Employment and Day Program Participation

--Manuscript Draft--

Manuscript Number:	AJIDD-D-20-00144R2
Article Type:	Research Report
Keywords:	machine learning; employment; predictive modeling
Corresponding Author:	Michael Broda Virginia Commonwealth University Richmond, VA UNITED STATES
First Author:	Michael Broda
Order of Authors:	Michael Broda
	Matthew Bogenschutz
	Parthenia Dinora
	Seb Prohn
	Sarah Lineberry
	Erica Ross
Manuscript Region of Origin:	UNITED STATES
Abstract:	In this article, we demonstrate the potential of machine learning approaches as inductive analytic tools for exploring data on individuals with intellectual and developmental disabilities (IDD). Using data from the National Core Indicators In-Person Survey (NCI-IPS), a nationally-validated survey of more than 20,000 people with IDD that was developed by the National Association of State Directors of Developmental Disabilities Services (NASDDDS) and the Human Services Research Institute (HSRI), we fit a series of classification tree and random forest models to predict individuals' employment status and day activity participation as a function of their responses to all other items on the NCI-IPS. The most accurate model, a random forest classifier, predicted employment and day participation outcomes of adults with IDD with an accuracy of 89 percent on the testing sample, and 80 percent on the holdout sample. The most important variable in this prediction was whether or not community employment was a goal in this person's service plan. These results suggest the potential machine learning tools to examine other valued outcomes used in evidence-based policymaking to support people with IDD.

Using Machine Learning to Predict Patterns of Employment and Day Program Participation

Evidence-based policy and programs to support people with intellectual and developmental disabilities (IDD) rely, in part, on the availability of carefully collected data and rigorous analysis. However, methodological innovations are rare, and it is often not feasible to use emerging analytic approaches to study IDD programs and policies because datasets either do not adequately account for IDD (Havercamp et al., 2019) or are not robust enough to conduct analyses that require a large sample (Wagner, Kim, & Tasse, 2019). This paper takes an innovative approach to the examination of employment and day program participation, a commonly studied topic in IDD research, by applying machine learning methods. To our knowledge this is among the first applications of machine learning in IDD policy studies, and clinical application of machine learning in this population has been limited.

Background

Applications of Machine Learning

Methodological innovations in IDD research have the potential to transform how researchers and advocates can use large datasets to inform evidence-based policymaking. For instance, Wagner and colleagues (2019) identified the possibilities for merging smart home and wearable technology data with existing Medicaid data to better understand health outcomes for people with IDD. The need to create and utilize more merged datasets is supported by other researchers as a potential way of gaining more complete and nuanced understanding about how systems work to affect outcomes for people with IDD (Dinora et al., 2020; Havercamp, 2019). In addition to using methodological innovation to construct more robust datasets, researchers have also identified the need to use analytic innovation to, for example, build algorithms to find

people with IDD within large administrative datasets (Lin et al., 2013), spark innovations in using state and local-level administrative data (Bonardi et al., 2019), and to use artificial intelligence in the disability arena (e.g.: Bertoncelli et al., 2019; Maenner et al., 2016).

Quantitative research, including most contemporary research in the IDD field, takes a deductive approach to data analysis, whereby the researcher determines, *a priori,* combinations of variables to examine based on past research findings, existing theory, or the values that guide the field, and then uses data to test the relationships between the selected variables. By contrast, machine learning is an inherently inductive process, in which a large dataset is used to train an algorithm which then explores all possible explanations for a selected outcome, based on the data available (Breiman, 2001b). In this way, the inductive process of machine learning is designed to create pathways toward the generation of empirically-derived theories of how a particular outcome occurs (Murdoch et al., 2020). Often, a researcher may then use established research, theory, or values to aid interpretation of machine learning output, which can sometimes serve as a form of triangulation for previous research findings.

Machine learning, a form of artificial intelligence that originally emanated from computer science, is steadily becoming more common in application to human services (Santiago & Smith, 2019). Criminal justice was among the first fields to extensively embrace machine learning methods to address topics such as determining the terms of bail and assessing recidivism risk (e.g.: Berk, 2019; Rudin & Ustun, 2019). Other human service fields, notably healthcare administration (e.g.: Kavakiotis, et al, 2017; Rudin & Ustun, 2019) and education (Chung & Lee, 2019), have also begun to use machine learning more extensively in recent years.

Despite the proliferation of machine learning in human services, relatively few studies that use machine learning to understand policy-relevant outcomes for people with disabilities

currently exist. Specific to people with disabilities, machine learning applications have largely been clinical in nature. For instance, autism researchers have identified a number of clinical applications of machine learning (Hyde et al., 2019), and studies have investigated topics such as detection and surveillance of autism spectrum disorders (Maenner et al., 2016; Thabtah & Peebles, 2020) and prediction of lifetime health (Bishop-Fitzpatrick et al., 2018). Though use of machine learning has been increasingly useful in clinical applications for people with a range of disabilities, use in the IDD field has been rare, and we are unaware of studies that have used machine learning to investigate policy-relevant topics pertaining to IDD. In this paper, we use employment and day activity, a widely studied and highly valued outcome for people with IDD, to demonstrate how machine learning can be used to inform IDD policy and system advancement.

Overview of Day Activities for People with IDD

Daytime Activities. Integrated, competitive employment is often seen as an important outcome for people with IDD, and recent evidence suggests that state investments in employment for people with IDD can have positive effects on employment outcomes (Nord et al., 2020). Despite a shift in policies for expanding competitive, integrated employment, (i.e., working at part-time or more, at above minimum wage, alongside people without disabilities), most adults with IDD participate in facility-based work (i.e., supervised settings where the majority of people have disabilities) and day activities (non-work, day services). Of the estimated 642,000 adults with IDD who utilized supports for employment or other daytime activities in 2017, Winsor and colleagues (2019) found that about 20% were engaged in competitive, integrated employment, while the remaining 80% spent at least part of their day in facility-based activity. The last decade has seen overall declines in use of facility-based day

activities, but a rapid increase in community-based non-work activities (i.e., services that support people in community activities where the majority of people do not have disabilities; Winsor et al, 2019).

Employment has also been reinforced in recent years as a primary outcome for people with IDD that is supported through policy. For instance, Employment First is a framework for state disability service agencies that prioritizes integrated community employment for people with disabilities (Association of People Supporting Employment First [APSE], N.D.; U.S. Department of Labor, N.D.). Employment First begins with the assumption that people with IDD are capable of working and support to find a job in the community is offered before more segregated options (APSE, N.D.). As of January 2020, 40 states had legislation or an official policy on Employment First as the preferred model for service provision (APSE, 2020), reinforcing the high value placed on competitive employment among people with IDD, their advocates, and policymakers.

Employment Predictors. This section provides background on frequently identified employment indicators for the population with IDD as context for our analysis. At the level of the individual, level of IDD has often been found to predict employment outcomes, with people with mild or moderate IDD being most likely to secure competitive, integrated employment (Bush & Tasse, 2017; Nord et al., 2018; Park & Bouck, 2018). People with strong independent living skills (Carter et al., 2012; Chan et al., 2018), people who communicate verbally (Carter et al., 2012), and people with few challenging behaviors (Shogren & Shaw, 2016; Simonsen & Neubert, 2013) have all been found to be more likely to engage in competitive, integrated employment.

Competitive, integrated employment can also be supported by system-driven factors. For instance, previous analyses have found that having an employment goal in the individual service plan of a person with IDD predicts positive employment outcomes (Butterworth et al., 2015; Nord, et al., 2018). More specifically, Butterworth and colleagues (2015) noted that only 14% of people with IDD who were not employed had a goal for competitive employment, although 47% of people with IDD who did not have a competitive job stated a desire for one. Nord and colleagues (2020) also found that state level investments in employment support services help to moderate the effects of other factors that predict employment outcomes, such as the age of the job seeker with IDD. People who reside independently are more likely to be competitively employed than are people who reside in institutional settings (Butterworth et al., 2015), a finding that may be related to Bush and Tasse's (2017) finding that people with IDD who were able to exercise choice in their daily lives were more likely to be competitively employed.

It is also important to note that individual and environmental factors often interact, and that the nature of these interactions may be important in determining whether a person with IDD achieves successful community employment or whether they use other day supports. For example, Nord and colleagues (2020) noted that state funding for employment services interacts with a person's age to help predict employment outcomes, and Butterworth's (2015) work highlighted the relationships between living setting, level of ID, and employment outcomes. Gaining better understanding of other interactions, for example, between intensity of supports, presence of employment goals, level of ID, and demographic characteristics may be particularly important in helping us to understand pathways to employment moving forward.

Research Questions

The current study addresses two central research questions: Based on a randomly selected sample of adults with IDD who use publicly-funded services in the United States: 1) Can a machine learning model be built to accurately and consistently predict employment and day program participation?; and 2) If such a model is possible, what factors in that model most contribute to the accurate prediction of employment and day program participation status?

Methods

Sample

The data from this study came from the National Core Indicators In-Person Survey (NCI-IPS) a nationally-validated survey of people with IDD that was developed by the National Association of State Directors of Developmental Disabilities Services (NASDDDS) and the Human Services Research Institute (HSRI). This survey was administered in 35 states and the District of Columbia in 2017/2018.

The NCI-IPS is conducted as a face-to-face interview with people with IDD aged 18 and older who use at least one public IDD service in addition to case management/support coordination. The survey consists of three parts. The background information, which is typically completed using administrative records, details demographic, service utilization, and basic health status information. The survey portion of the IPS is administered directly with the person with IDD by a trained interviewer. Section I of the IPS must be answered by the person with IDD directly, while Section II permits proxy response, if needed.

Measures

Consistent with the goals of this research, we sought to keep as many variables from the NCI as possible in our analytic dataset. We excluded participant identifiers, as well as metadata pertaining to the date, time, etc., of the NCI interview. We also excluded all variables from the

section on employment, with the exception of PAIDFACWORK ("Person was in paid work performed in facility-based setting during typical 2-week period"), PAIDCOMMJOBGRP15 ("Person was in paid small-group job in community-based setting during typical 2-week period"), PAIDCOMMJOBIND15 ("Person was in paid individual job in community-based setting during typical 2-week period"), UNPAIDCOMMACT ("Person was in unpaid activity in community-based setting during typical 2-week period") and UNPAIDFACACT ("Person was in unpaid activity in facility-based setting during typical 2-week period"), which were used to identify clusters of participants based on their employment and program/activity participation. We also removed any variables from the remaining sections that contained the word "JOB" and all open-ended qualitative responses.

The remainder of the variables available in the IPS were all included as predictors in our analysis. Because the IPS is a large and sweeping survey, it is impossible to explain all variables in the space of this manuscript, however, it is important to provide some background. The variables used as predictors came from the background section (typically completed by a case manager; containing data about demographics, medical, mental health, and behavioral conditions, medications, and services used), Part I of the in-person interview (respondent must reply via self-response; variables in domains such as satisfaction with supports, choices, and rights) and Part II of the in-person interview (respondents can either self-report or have a proxy respondent; variables include those related to relationships, community participation, etc.). There were also variables related to who responded to questions and the interviewer's perception of the quality of responses. For instance, a variable specifically identified whether self-report or proxy response was used in Part II, and this variable was entered into our analysis as a predictor.

Readers who wish to gain a full understanding of the scope of variables in the IPS are directed to the National Core Indicators website.

Feature Engineering and Data Preprocessing

We used the *tidymodels* package (Kuhn et al., 2020) in R (R Core Team, 2020) to prepare data for analysis. All continuous variables (AGE, WEIGHT, BMI, and HEIGHTFT) were normalized to have a mean of 0 and standard deviation of 1. Median imputation was used to recode any missing values for these variables. All nominal variables were converted to separate dummy variables for each category. Here, missingness was treated as a separate category and the participant was given a dummy variable indicating such.

Data Analysis

In this paper, we use two machine learning techniques - classification trees and random forests (Breiman, 2001a; Strobl et al., 2009). Classification trees are highly flexible recursive partitioning models that are most useful when examining large numbers of predictors with potentially nonlinear or other complex associations not typically captured by linear modeling (e.g., higher order interactions). Unlike, for example, logistic or multinomial logistic regression, which require that a model and hypothesis be specified beforehand, classification tree algorithms search for linear, nonlinear, and interactive effects associated with a given set of predictors and a given outcome, which can be binary, ordinal, or in our case, multi-categorical (Loh, 2014). In doing so, these models can "learn" which effects to include to best predict the outcome. Thus, classification trees can often expose relationships that are unexplored and/or undetectable by regression-based methods.

Essentially, classification trees seek to predict the value of an outcome variable by splitting (or partitioning) a dataset into several chunks based on values of the input variable(s).

To give a basic example, a model predicting whether or not a person still uses a landline telephone at home might first split the data by participants' age, and then produce separate predictions of the likelihood of using a landline for older participants and younger participants. This simple tree only partitions the data once, and it does so using values of a single variable, though in the case of a continuous variable, classification trees iteratively test splitting the data at every possible value (e.g., splitting those older and younger than 40 versus splitting those older and younger than 50). However, classification trees can partition data many times over, resulting in trees with depths of 10, 15, 20, or more and resulting in a unique prediction of the outcome for each terminal node in the tree. Furthermore, trees can partition data using binary, continuous, ordinal, and nominal variables (Strobl et al., 2009). For example, after splitting by age, the previous tree predicting landline usage might further split older participants based on whether or not they own a mobile phone.

While useful, classification trees have several limitations, including high sensitivity to the order in which variables are selected by the algorithm, inconsistent predictive performance, and a tendency to overfit the model to the data (Kuhn & Johnson, 2013). In these models, slight changes in the data can affect the order of variable selection, which can result in large changes in a final classification tree solution. To counter some of the limitations of classification trees, we also fit random forest models (Breiman, 2001a). Random forests are a type of "ensemble learning" approach in which numerous (e.g., hundreds or thousands) of classification trees are fit using randomly-chosen subsets of the input variables and bootstrap sampling of observations (Breiman, 2001a; Kuhn & Johnson, 2013). The results of these trees are then averaged together to obtain final predictions of the outcome value. This algorithm greatly improves the accuracy of predictions of random forests (when compared to single-tree models) by minimizing correlations

between trees and greatly reducing the potential for any overfitting (Breiman, 2001b; Strobl et al., 2009). Furthermore, this approach allows for variables to be introduced in many different combinations, resulting in models that are able to explain more variance in a given outcome, have greater predictive power in an external sample, and have greater flexibility for capturing nonlinearity and interactive effects (Ziegler & König, 2014).

Analytic Process

Our analysis proceeded in several steps. First, we generated frequencies for all possible combinations of our five employment and day participation outcomes (PAIDFACWORK, PAIDCOMMJOBGRP15, PAIDCOMMJOBIND15, UNPAIDCOMMACT and UNPAIDFACACT). Then, we limited our outcome to the nine most common patterns of employment and day participation, all of which were reported by at least one percent of the overall NCI sample. Next, we applied the feature engineering and data preprocessing procedures described above to create a version of the NCI data that was conducive to use in ML models. This included the application of the SMOTE algorithm (Chawla et al., 2020), which used synthetic oversampling to create equal size clusters for each of our eight employment and day participation outcome categories. This approach is able to create equal group sizes, which is optimal for prediction, while preserving the unique qualities of participants in each distinct group.

Then, we fit a classification tree model to the training data for our outcome. Next, we fit a random forest model with 1,000 trees to the same training data. This approach enables us to compare the performance of both algorithms, with performance here defined as the model classification accuracy, to one another. We would expect the random forest to outperform a single classification tree (as it is an aggregation of many classification trees). Next, we

performed a 90/10 cross validation procedure with both the classification tree and the random forest models with 25 replications on the training data. This means models were repeated 25 times, each time training on a random 90% of the sample and testing on the remaining 10%. Finally, we used the hold-out, or test data, to examine each model's performance on a new sample that has not been previously tested.

All analyses were conducted in *RStudio* (2020). Classification tree and random forest models were all fit using the *tidymodels* package (Kuhn et al., 2020), which provides a unified set of tools for developing predictive models in R. When fitting classification trees, we used the algorithm found in the *rpart* package (Therneau & Atkinson, 2018); and when fitting random forests, we used the approach found in the *ranger* package (Wright & Ziegler, 2017).

Model Validation Procedure

When fitting the classification tree and random forest models, we used multiple approaches to cross validate our results. As mentioned previously, cross validation provides a transparent method for identifying how well a given statistical model will perform on another sample not used in the initial fitting of the model (James et al., 2013), and it is particularly important to cross validate non-parametric models to avoid overfitting. Cross validation can be performed by first randomly dividing the full sample into two separate datasets: a training dataset and a test dataset, with the training dataset comprising a considerably larger proportion of the initial data than the test dataset (e.g., 80% of the total observations in the data). Models are then fit and adjusted using the training data and, once the researchers arrive at a final model, the performance of this model is then tested using the test dataset. This approach guards against overly-optimistic interpretations of the model's generalizability.

An alternative to splitting data *a priori* into train and test samples is to use 10-fold cross-validation or a similar resampling approach. In 10-fold cross-validation, the full dataset is randomly divided into 10 equally-sized subsets (or folds). A model is then fit using data from nine of the sets, and the performance of the model is tested using the set that was held out from the initial model fitting. This process is then repeated using a different set as the hold-out sample and so on until each of the 10 sets has been held out and used for validation. The performances of the models are then averaged together. In repeated 10-fold cross-validation, this entire process is repeated multiple times so that the observations composing each of the 10 folds will differ between repetitions (Kuhn & Johnson, 2013; Molinaro, Simon, & Pfeiffer, 2005). We repeated this process 25 times in this analysis.

Measuring Variable Importance

We use multiple approaches for determining the relative importance of the variables included in our models. For our classification tree models, variable importance is represented graphically. Variables that the model splits on earlier (i.e., those closer to the "root" or top of the tree) are more important for predicting the outcome than are those that the model splits on later. In other words, these variables are more useful for partitioning the data into distinct subsets.

Finally, variable importance in the random forest is represented by the percent increase in node purity, which provides an indication of the extent to which splitting a tree on a given variable decreases the residual sum of squares. For each variable, increase in node purity is calculated across all splits on that variable across all trees, with higher values corresponding to greater variable importance (Grömping, 2009). We present variable importance in a series of dot plots with node purity on the x-axis and variable names on the y-axis.

Results

Choosing Clusters of Employment and Day Participation

Our initial analysis revealed 190 distinct patterns of employment and day participation[1]. However, many of these patterns, or clusters, of participants were quite small, containing less than five participants. On closer inspection, balancing the need to keep as many participants in the data set as possible with the need for parsimony, we found that nine clusters, which each contained more than 1% of the sample, were sufficient to explain nearly all the meaningful variability in employment and day program participation. We arrived at these nine clusters, shown in Table 1, by eliminating all clusters that contained missing (NA) or "Don't Know" (99) response options. This eliminated about 4,000 participants from the analysis - the foregoing results do not generalize to people in those clusters.

The bulk of the sample (about 73%) consists of people who participate only in day programming (43.4%) or people who do not participate in day programming or any type of employment (29.9%). The remaining five clusters represent people who spend their time in some combination of independent work in the community, group work in the community, facility work (paid or unpaid), and/or day program participation.

Since the bulk of the sample is in one of the first two categories, we needed to develop an empirical strategy for handling the smaller group sizes of clusters 3-8. If not, any predictive model (linear, nonlinear, algorithm-based, etc.) would maximize its prediction by placing nearly all participants into one of the first two groups. Thus, we would end up with a model that predicts those groups very well, but does a poor job of explaining why people might fall into the remaining categories, which in this case have real substantive importance.

[1] See our online methodological appendix available via the Open Science Framework [link removed for peer review]. A table with all 91 categories is available at [link removed for peer review].

Our solution was to apply the synthetic non-majority over-sampling technique (SMOTE; Chawla et al., 2020), an algorithm that balances group sizes by creating new, or "synthetic" versions of cases in small groups by using their nearest neighbors. This approach is able to create equal group sizes, which is optimal for prediction, while preserving the unique qualities of participants in each distinct group.

Classification Tree Results and Interpretation

The tree drawn above presents results for our classification model predicting our eight categories of employment and day participation. More important variables are found at the top, with each split sending observations along different paths to lower "nodes" on the tree. The tree results in a total of 9 terminal nodes, which are the rounded rectangles found at the bottom of the tree. The number at the top of the terminal node represents the predicted class for participants in this node (e.g., in the terminal node furthest to the top left, participants would have been predicted to be in class one, or no employment or activity participation). The percentage at the bottom of the node represents the proportion of the total sample found in that node (e.g., the node above contains 11% of the total sample). The eight proportions listed in the middle of the node represent the proportion of the node that is made up of participants who are actually in a given cluster (the "true" value). So, again using the node furthest to the top left, 42% of the participants here actually were in cluster 1, while 13% were in cluster 2, and so on. Ideally, most of the participants would end up in a terminal node that matches their "true" value; deviations from this pattern represent misfit in the model prediction.

This particular classification tree had an overall accuracy rate of about 32%, meaning in this case that 32% of participants were accurately classified into their work/day cluster. This corresponds to an area under receiver operator characteristic curve (AUC_ROC) of .73. Thus, a

single regression tree, while readily interpretable, does not do an optimal job of accurately classifying participants into employment clusters. One option is to grow many trees (in this case 1,000), and average their performance, creating a random forest classifier, which we describe below.

Random Forest Results

Overall, the random forest model did an excellent job classifying participants into the eight categories. On the training data, using a sample of 25 cross-validation samples (with 90% used for training, 10% for testing, sampling with replacement), we found that the model accurately classified 89.0% of all cases, with an area under receiver operator characteristic curve (AUC_ROC) of .99. A full confusion matrix listing the predicted and actual class membership for all participants can be found in our online methodological appendix at [link removed for peer review].

We also tested the model's performance on a holdout sample of 30% of the total sample. This data was not included in the previous models and was coded and prepared according to the same procedures described above in data preprocessing and feature engineering. The model again performed well, with a classification accuracy of 80%.

Variable Importance Analysis

Having established the predictive ability of the random forest model, we also calculated variable importance measures to determine which variables made the most contribution to predictive power of the model. Figure 2 plots the 10 most important variables in decreasing order of importance, where importance, as described above, represents the relative decrease in accuracy if that variable was excluded from the model. The NCI variable name is listed first, followed by an underscore, and then the response category. The most important variable for

prediction was the variable IEGOAL ("Is community employment a goal in this person's service plan?"), specifically those participants for whom the response was "Yes" (Category 2) to this question. This variable is about 30 precent more important than the next highest variable, VOLUNT15 ("Do you volunteer?"). Table 2 includes the 10 most important variables, along with the item stems and specific response categories, as well as the additional variables identified by the classification tree above.

Discussion

In this study, we explored the potential of using machine learning to develop a predictive model of employment and day participation outcomes for adults with IDD. This predictive model used classification trees and random forests and performed well in predicting employment and day program participation for adults with IDD.

The performance metrics presented show that this-model predicts employment and day participation outcomes of adults with IDD with an accuracy of 89.0 percent. The model's performance was further tested on a holdout sample, and performed well again, with an accuracy of 80 percent. These results suggest that machine learning may have predictive utility when used with a large, representative sample of people with IDD, and suggests the potential for applications of machine learning to examine other valued outcomes. Results indicate the potential contribution that machine learning could make toward evidence-based policymaking to support people with IDD.

In our findings, the strongest predictor of employment and day participation outcomes was having an employment goal in the person with IDD's service plan, a finding consistent with past research that found correspondence between having an employment goal and competitive employment (Butterworth et al., 2015; Nord, et al., 2018). Although it is possible that there may

be a bias for people who are more likely to secure employment to have an employment goal, this finding may suggest the importance of prioritizing employment, and offering employment services as part of a standard approach to service planning. This is consistent with recommended Employment First initiatives (APSE, n.d.), in addition to supporting the informed choice around individual and family contributions to the decision-making processes about employment (Hoff & Holz, 2020). Despite this evidence, nationally, only about 29% of people with IDD had a formal goal for competitive integrated employment in 2018 (HSRI & NASDDDS, 2018).

After employment goals, volunteerism was the second most important predictor of employment. Commonly the benefits of volunteerism, such empowerment, improved social skills, and verbal communication have been viewed as innately valuable but indirectly related to employment (Miller et al., 2002). Wicki and Meier (2016) supported volunteerism but only with concurrent employment and never as an alternative or precursor to employment. Trembeth and colleagues (2010) concluded that volunteerism's benefits are unlikely to extend to employment anyway, stating that employment training was a preferred alternative to volunteerism that could increase the likelihood of employment. Our results indicate that people with IDD would benefit from either. As the number of individuals with community-based (non-employment) day services grows nationally, high-quality volunteerism and employment training remain viable mechanisms for states to invest in human capital and provide opportunities for career exploration and development (Windsor et al., 2019).

Among the top ten predictors of employment to come out of our analysis, factors related to everyday choice making (Bush & Tasse 2017) and having choice about what to do during the day (Shogren & Shaw, 2016) were consistent with prior literature. These findings effectively triangulated findings from deductively-driven analyses with the inductive empirically-driven

approach inherent in machine learning (Breiman, 2001). Despite these consistencies, many factors that have been identified as important predictors of employment outcomes in recent literature did not appear in the list of top predictors based on this machine learning analysis of employment and day participation outcomes. For example, many of the personal characteristics that have been identified in prior literature, such as race (Kaya, 2018; Sannicandro et al., 2018; Simonsen & Neubert, 2013), communication method (Carter et al., 2012), gender (Carter et al., 2012; Kaya, 2018), and level of IDD (Bush & Tassé, 2017; Nord et al., 2018; Park & Bouck, 2018) had relatively low predictive power in our machine learning models using the NCI.

Taken together, our results may suggest the power of policy in improving employment and day participation outcomes for people with IDD. Most of the strongest predictors of employment and day participation identified in our analyses are factors that are often aligned with state regulatory and policy approaches, in conjunction with informed individual and family choice. For instance, recommending employment as a first option among day activities is supported through Employment First, and requirements of service providers to support choice making are supported through the HCBS Final Rule. Transportation was another important predictor of employment and day participation to emerge from our analysis, and accessible public transportation is mandated via the Americans with Disabilities Act. Such findings could prove powerful in the policymaking process, since they suggest that employment and day participation outcomes may be improved by changes in factors that are amenable to policy change, and that factors that are immutable, such as many personal characteristics, seem to hold less predictive importance with regard to employment and day participation outcomes of people with IDD.

Since people with IDD who are engaged in competitive, integrated employment tend to have higher quality of life than their peers who attend day programs or go to sheltered workshops (Beyer, et al., 2010), these findings are encouraging, since there is an indirect suggestion that the lives of people with IDD may be substantially improved with the will to enact policies that prioritize employment as the preferred daytime activity for people with IDD.

Potential and Cautions for Machine Learning

Machine learning has a wealth of potential for use in IDD research, including in the policy domain, where it may serve as a compliment to deductive approaches to research. As policymakers seek to create more responsive service systems with limited budgets, machine learning may have potential to help predict where expenditures will have the greatest impact on health and wellness, social inclusion, and safety, among other important outcomes. Machine learning may help in designing efficient policy tools that will effectively target improvements in desired outcomes, making machine learning methods a powerful emerging tool for the IDD field to embrace.

Use of machine learning in other fields, most notably criminal justice and healthcare, has not been without controversy, and lessons from those fields may be instructive for researchers in IDD. Most notably, researchers and advocates have pointed to the potential for discriminatory bias and unfair decision making from machine learning that uses historical data to train computer algorithms, since such data may have been based on biased and discriminatory assumptions. Authors have called for improvements in the interrogation of source data that informs machine learning (Veale & Binns, 2017), since reliance on extant data can implicitly discriminate against some groups of people who have historically experienced oppression, including people with disabilities (Trewin, 2018). While the current study did not rely on historical data to create our

algorithm, caution should be exercised if using other datasets that would employ machine learning to predict outcomes for people with disabilities.

Limitations & Conclusion

In addition to the general word of caution about potential bias in machine learning, as noted above, there are other important limitations to the methods used in this study. As discussed previously, classification trees are sensitive to the order in which variables are introduced and may be susceptible to overfitting. Although random forest models are less susceptible to overfitting, they have other limitations. For example, random forests do not provide beta coefficients that describe the relationship between a specific predictor and the outcome. They do provide some metrics describing variable importance, but these metrics are not as interpretable as traditional beta coefficients in a more conventional regression model. Similarly, the approach used here does not involve the pre-specification of specific variable interactions that can be tested using traditional hypothesis testing techniques. By the incorporation of variables in random order across many different trees, this approach does capture nonlinear or interactive contributions of individual variables, but it does not provide guidance on which specific nonlinearities or interactions may be most important.

In addition, the NCI-IPS includes responses only from adults with IDD who use state-funded services, and it is important that findings from this study not be extrapolated to people who do not receive such supports. Though sampling for the NCI-IPS is generally random and representative of people with IDD across the United States, different states construct their samples in slightly different ways, and sometimes record participant responses in different ways, introducing some potential for minor discrepancies in the reliability of the NCI as a measure. Finally, though these results are a valid representation of the national NCI-IPS dataset, caution

should be exercised in assuming that the precise findings presented here apply to any particular state. State IDD systems vary considerably, and observations that hold true based on a national dataset may look differently at different ecological levels.

Despite these limitations, the current study breaks new ground in the IDD field by being the first of its kind to use IDD-specific data with policy implications in a machine learning application, thereby introducing a fresh analytic perspective in the IDD policy field. By using a novel machine learning approach with a nationally-representative sample of adults with IDD, we have discovered important predictors of employment and day participation outcomes that are amenable to change via targeted policy, including designing policy that reinforces choice making, prioritizes individualized planning for employment, and making transportation widely accessible for potential employees with IDD. By utilizing machine learning to examine other important outcome domains for people with IDD, policymakers may be able to craft cost-effective, targeted service plans that will support increased quality of life for people with IDD.

References

Association of People Supporting Employment First, (N.D.). APSE fact sheet: Employment First. https://www.apse.org/wp-content/uploads/docs/Employment%20First%20-%20Legislator%20Fact%20Sheet.pdf

Association of People Supporting Employment First (2020). Employment First map. https://apse.org/employment-first-map/

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J.A. Larusson & B. White**,** (Eds.), *Learning analytics: From research to practice* (pp. 61-75). New York: Springer.

Berk, R. (2019). *Machine learning risk assessments in criminal justice settings*. New York: Springer.

Bertoncelli, C.M., Altamura, P., Vieira, E.R.; Bertoncelli, D., & Solla, F. (2019). Using artificial intelligence to identify factors associated with autism spectrum disorder in adolescents with cerebral palsy. *Neuropediatrics, 50*(3), 178-187. http://doi.org/10.1055/s-0039-1685525

Beyer, S., Brown, T., Akandi, R., & Rapley, M. (2010). A comparison of quality of life outcomes for people with intellectual disabilities in supported employment, day services and employment enterprises. *Journal of Applied Research in Intellectual Disabilities, 23*(3), 290-295. https://doi.org/10.1111/j.1468-3148.2009.00534.x

Bishop-Fitzpatrick, L., Movaghar, A., Greenberg, J.S., Page, D., DeWalt, L.S., Brilliant, M.G., & Mailick, M. (2018). Using machine learning to identify patterns of lifetime health problems in decedents with autism spectrum disorder. *Autism Research, 11*, 1120-1128. https://doi.org/10.1002/aur.1960

Bonardi, A., Krahn, G., Morris, A., & the National Workgroup on State and Local Health Data (2019). *Enriching our knowledge: State and local data to inform health surveillance of the population with intellectual and developmental disabilities.* Washington, DC: Administration on Intellectual and Developmental Disabilities. https://doi-org.proxy.library.vcu.edu/10.1352/1934-9556-57.5.390

Breiman, L. (2001a). Random forests. *Machine Learning*, *45*(1), 5-32. http://dx.doi.org/10.1023/A:1010933404324

Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3), 199-231. http://dx.doi.org/10.1214/ss/1009213726

Bush, K. L., & Tassé, M. J. (2017). Employment and choice-making for adults with intellectual disability, autism, and Down Syndrome. *Research in Developmental Disabilities*, *65*, 23-34. https://doi.org/10.1016/j.ridd.2017.04.004

Butterworth, J., Hiersteiner, D., Engler, J., Bershadsky, J., & Bradley, V. (2015). National Core Indicators©: Data on the current state of employment of adults with IDD and suggestions for policy development. *Journal of Vocational Rehabilitation, 42*(3), 209-220. https://doi.org/10.3233/JVR-150741

Carter, E. W., Austin, D., & Trainor, A. A. (2012). Predictors of postschool employment outcomes for young adults with severe disabilities. J*ournal of Disability Policy Studies, 23*(1), 50-63. https://doi.org/10.1177/1044207311414680

Chan, W., Smith, L. E., Hong, J., Greenberg, J. S., Lounds Taylor, J., & Mailick, M. R. (2018). Factors associated with sustained community employment among adults with autism and co-occurring intellectual disability. *Autism, 22*(7), 794-803. https://doi.org/10.1177/1362361317703760

Chung, J.Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review, 96*, 346-353. https://doi.org/10.1016/j.childyouth.2018.11.030

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *The Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd Ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Dinora, P., Bogenschutz, M., & Broda, M. (2020). Identifying Predictors for Enhanced Outcomes for People with Intellectual and Developmental Disabilities. *Intellectual and Developmental Disabilities*, *58*(2), 139–157. https://doi.org/10.1352/1934-9556-58.2.139

Dressel, J. & Farid, H. (2019). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*(1), eaao5580. http://doi.org/10.1126/sciadv.aao5580

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. In *Advances in Neural Information Processing Systems* (pp. 155-161).

Fox, J. (2015). *Applied regression analysis and generalized linear models*. (6th Ed.) Thousand Oaks, CA: Sage Publications.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics, 29*(5), 1189-1232. https://www.jstor.org/stable/2699986

Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*, *63*(4), 308-319. http://dx.doi.org/10.1198/tast.2009.08199

Havercamp, S. (2019). Improving data collected on people with intellectual and developmental disabilities. https://thehill.com/opinion/healthcare/465524-improving-data-on-people-with-intellectual-and-developmental-disabilities

Havercamp, S. M., Krahn, G. L., Larson, S. A., Fujiura, G., Goode, T. D., Kornblau, B. L., & National Health Surveillance for IDD Workgroup. (2019). Identifying people with intellectual and developmental disabilities in national population surveys. *Intellectual and Developmental Disabilities, 57*(5), 376-389. https://doi.org/10.1352/1934-9556-57.5.376

Hawkins, D., Basak, S., & Mills, D. (2003). Assessing model fit by cross-validation. J*ournal of Chemical Information and Computer Sciences, 43*(2), 579-586. http://dx.doi.org/10.1021/ci025626i

Hoff, D. & Holz, N. (2020). Employment and employment supports: A guide to ensuring informed choice for individuals with disabilities. *Institute on Community Inclusion Tools for Inclusion,* 31. Retrieved from https://www.communityinclusion.org/pdf/TO31_F.pdf.

Hvitfeldt, E. (2020). themis: Extra Recipes Steps for Dealing with Unbalanced Data. R package version 0.1.1. https://CRAN.R-project.org/package=themis

Hyde, K.K., Novack, M.N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D.R., & Linstead, E. (2019). Applications of supervised machine learning in autism spectrum disorder research: A review. *Review Journal of Autism and Developmental Disorders, 6*, 128-146. http://doi.org/10.1007/s40489-019-00158-x

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York: Springer.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and*

Structural Biotechnology Journal, 15, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005.

Kaya, C. (2018). Demographic variables, vocational rehabilitation services, and employment outcomes for transition- age youth with intellectual disabilities. *Journal of Policy and Practice in Intellectual Disabilities, 15*(3), 226-236. https://doi.org/10.1111/jppi.12249

Krumm, A., Means, B., & Bienkowski, M. (2018). *Learning analytics goes to school: A collaborative approach to improving education*. London: Routledge.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer.

Kuhn, M., & Wickham, H. (2020). *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles*. https://www.tidymodels.org.

Larson, S.A., Lakin, K.C., Anderson, L., Kwak Lee, N., Lee, J.H., & Anderson, D. (2001). Prevalence of mental retardation and developmental disabilities: Estimates from the 1994-1995 National Health Interview Survey Disability Supplement. *American Journal of Mental Retardation, 106*(3), 321-252. http://dx.doi.org/10.1352/0895-8017(2001)106%3C0231:POMRAD%3E2.0.CO;2

Lin, E., Balogh, R., Cobigo, V., Ouellette-Kuntz, H., Wilton, A., & Lunsky, Y. (2013). Using administrative health data to identify individuals with intellectual and developmental disabilities: a comparison of algorithms. *Journal of Intellectual Disability Research, 57*(5), 462-477. http://doi.org/10.1111/jir.12002.

Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, *82*(3), 329-348. https://doi.org/10.1111/insr.12016

Maenner, M.J., Yeargin-Allsopp, M., Van Naarden Braun, K., Christensen, D.L., & Achieve, L.A. (2016). Development of a machine learning algorithm for the surveillance of autism spectrum disorder. *PLoS ONE, 11*(12), e0168224. http://doi.org/journal.pone.0168224.

Miller, K. D., Schleien, S. J., Rider, C., Hall, C., Roche, M., & Worsley, J. (2002). Inclusive volunteering: Benefits to participants and community. *Therapeutic Recreation Journal*, *36*(3), 247-259.

Molinaro, A. M., Simon, R., & Pfeiffer, R. M. (2005). Prediction error estimation: A comparison of resampling methods. *Bioinformatics, 21*(15), 3301-3307. http://dx.doi.org/10.1093/bioinformatics/bti499

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences of the United States of America, *116*(44), 22071–22080. https://doi.org/10.1073/pnas.1900654116

Nord, D., Grossi, T., & Andresen, J. (2020). Employment equity for people with IDD across the lifespan: The effects of state funding. *Intellectual and Developmental Disabilities, 58*(4), 288-300. https://doi.org/10.1352/1934-9556-58.4.288

Nord, D., Hamre, K., Pettingell, S., & Magiera, L. (2018). Employment goals and settings: Effects of individual and systemic factors. *Research and Practice for Persons with Severe Disabilities, 43*(3), 194-206. https://doi.org/10.1177/1540796918785352

Park, J., & Bouck, E. (2018). In-school service predictors of employment for individuals with intellectual disability. *Research in Developmental Disabilities, 77*, 68-75. https://doi.org/10.1016/j.ridd.2018.03.014

RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA. URL http://www.rstudio.com/.

Rudin, C. & Ustun, B. (2019). Optimizing scoring systems: Toward trust in machine learning for healthcare and criminal justice. *Informs Journal on Applied Analytics, 48*(5), 449-466. https://doi.org/10.1287/inte.2018.0957

Sannicandro, T., Parish, S. L., Fournier, S., Mitra, M., & Paiewonsky, M. (2018). Employment, income, and SSI effects of postsecondary education for people with intellectual disability. *American Journal on Intellectual and Developmental Disabilities, 123*(5), 412-425. http://doi.org/10.1352/1944-7558-123.5.412

Santiago, A.M. & Smith, R. (2019). What can "Big Data" methods offer human services research on organizations and communities? *Human Service Organizations: Management, Leadership & Governance, 43*(4), 344-356. http://doi.org/10.1080/23303131.2019.1674756.

Shogren, K. A., & Shaw, L. A. (2016). The role of autonomy, self-realization, and psychological empowerment in predicting outcomes for youth with disabilities. *Remedial and Special Education, 37*(1), 55-62. http://doi.org/10.1177/0741932515585003

Simonsen, M. L., & Neubert, D. A. (2013). Transitioning youth with intellectual and other developmental disabilities: Predicting community employment outcomes. *Career Development and Transition for Exceptional Individuals, 36*(3), 188-198. http://doi.org/10.1177/2165143412469399

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323. http://dx.doi.org/10.1037/a0016973

Thalbah, F., & Peebles, D. (2020). A new machine learning model based on induction of rules for autism detection. *Health Informatics Journal, 26*(1), 264-286. https://doi.org/10.1177/1460458218824711

Therneau, T., & Atkinson, B. (2018). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-13. https://CRAN.R-project.org/package=rpart

Titterington, M. (2010). Neural networks. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(1), 1-8. http://doi.org/10.1002/wics.50

Trembath, D., Balandin, S., Stancliffe, R. J., & Togher, L. (2010). Employment and volunteering for adults with intellectual disability. *Journal of Policy and Practice in Intellectual Disabilities*, *7*(4), 235-238. https://doi.org/10.1111/j.1741-1130.2010.00271.x

Trewin, S. (2018). AI Fairness for people with disabilities. Retrieved from https://arxiv.org/abs/1811.10670.

U.S. Department of Labor (N.D.). Employment First. https://www.dol.gov/agencies/odep/initiatives/employment-first

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society, 4*(2), 1-17. https://doi.org/10.1177/2053951717743530

Wagner, J. B., Kim, M., & Tassé, M. J. (2019). Technology tools: Increasing our reach in national surveillance of intellectual and developmental disabilities. *Intellectual and Developmental Disabilities, 57*(5), 463-475. https://doi.org/10.1352/1934-9556-57.5.463

Wicki, M. T., & Meier, S. (2016). Supporting volunteering activities by adults with intellectual disabilities: an explorative qualitative study. Journal of Policy and Practice in Intellectual Disabilities, 13(4), 320-326. https://doi.org/10.1111/jppi.12207

Wiens, J., & Shenoy, E. (2018). Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases, 66*(1), 149-153. http://doi.org/10.1093/cid/cix731

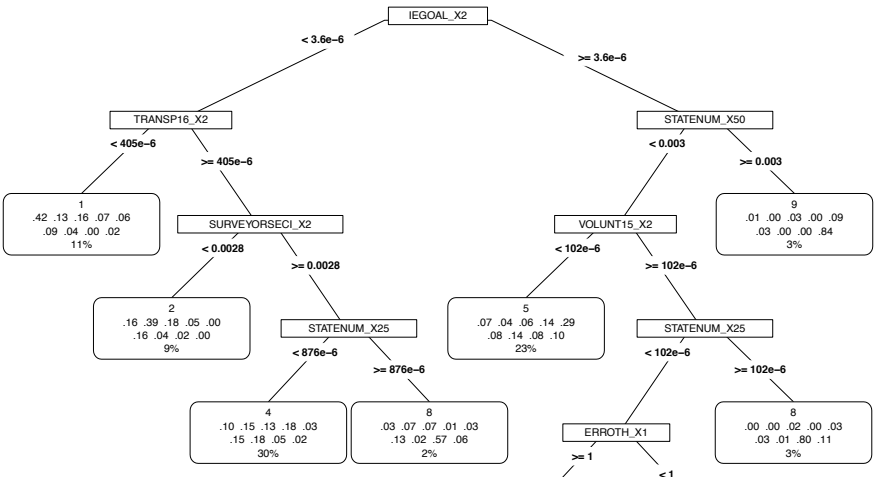
Winsor, J., Timmons, J., Butterworth, J., Migliore, A., Domin, D., Zalewska, A., & Shepard, J. (2019). StateData: The national report on employment services and outcomes. Boston, MA: University of Massachusetts Boston, Institute for Community Inclusion.

Wright, M.N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software, 77*(1), 1-17. http://doi.org/10.18637/jss.v077.i01

Ziegler, A., & König, I. R. (2014). Mining data with random forests: current options for real-world applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4*(1), 55-63. http://doi.org/10.1002/widm.1114

Figure 1. Classification Tree Results Predicting Employment and Day

Click here to access/download;Figure;preview



Codes for Terminal Node Labels:
1=No Participation
2=Unpaid Facility Only
3=Unpaid Community Only
4=Paid Facility Only
5=Paid Independent Community Only
6=Unpaid Only
7=Paid Facility and Unpaid Facility and Community
8=Paid Community Only
9=Paid Independent Community and Unpaid Community

Figure 2. Ten Most Important Variables for Accurate Prediction of Employment and Day Participation. Variable importance determined here via random forest permutation.

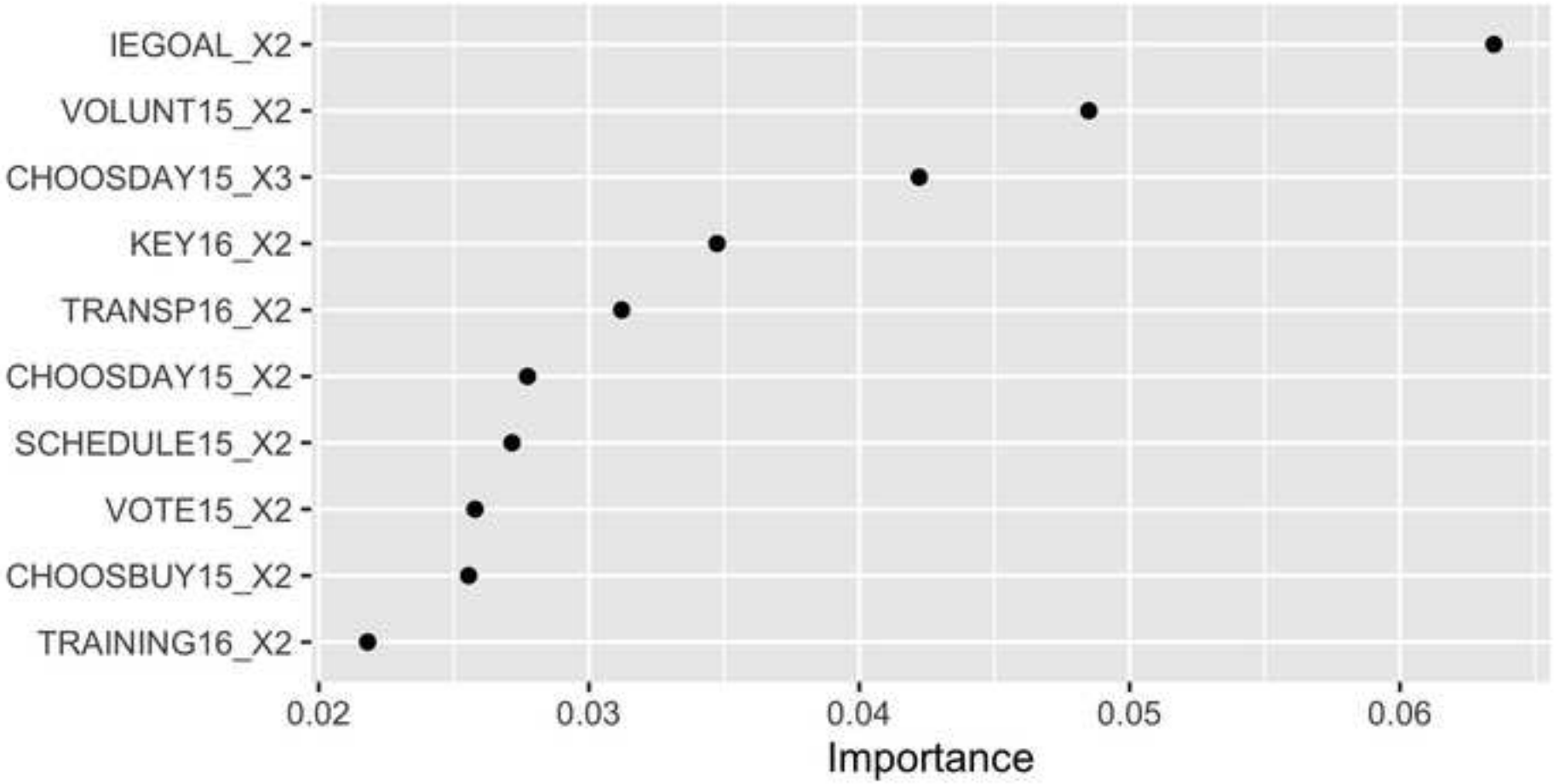


Table 1

Reduced Set of 9 Most Common Work and Day Participation Patterns

Cluster	Paid Comm. Group Job	Paid Comm. Ind. Job	Paid Facility Work	Unpaid Comm. Activity	Unpaid Facility Activity	Count	Proportion
1 "No Participation"	1 [No]	1 [No]	1 [No]	1 [No]	1 [No]	7239	0.337
2 "Unpaid Facility Only"	1 [No]	1 [No]	1 [No]	1 [No]	2 [Yes]	5844	0.272
3 "Unpaid Community Only"	1 [No]	1 [No]	1 [No]	2 [Yes]	1 [No]	2100	0.098
4 "Paid Facility Only"	1 [No]	1 [No]	2 [Yes]	1 [No]	1 [No]	1589	0.074
5 "Paid Independent Community Only"	1 [No]	2 [Yes]	1 [No]	1 [No]	1 [No]	1582	0.074
6 "Unpaid Only"	1 [No]	1 [No]	1 [No]	2 [Yes]	2 [Yes]	1505	0.070
7 "Paid Facility and Unpaid Facility and Community"	1 [No]	1 [No]	2 [Yes]	2 [Yes]	2 [Yes]	717	0.033
8 "Paid Community Only"	2 [Yes]	1 [No]	1 [No]	1 [No]	1 [No]	500	0.023
9 "Paid Independent Community and Unpaid Community"	1 [No]	2 [Yes]	1 [No]	2 [Yes]	1 [No]	416	0.019

Notes. N = 21,492. Comm. = Community; Ind. = Independent. Numbers represent value code from NCI, words in brackets represent value labels.

Table 2

Further Description of Ten Most Important Variables from Random Forest and Variables Identified in Regression Tree

Variable Name	Question Stem	Category	Label
	Variables Identified by Random Forest		
IEGOAL	Is community employment a goal in this person's service plan?	2	Yes
VOLUNT15	Do you volunteer?	2	Yes
CHOOSDAY15	Who chose where you go during the day?	3	Person had some input
KEY16	Do you have a key to your home?	2	Yes
TRANSP16	Which services/supports funded by the state (or county) agency does this person receive? (Transportation)	2	Yes
CHOOSDAY15	Who chose where you go during the day?	2	Person made the choice
SCHEDULE15	Who decides your daily schedule?	2	Person made the choice
VOTE15	Have you voted? (In a local state or federal election)	2	Yes
CHOOSBUY15	Do you choose what you buy with your spending money?	2	Yes
TRAINING16	Do you take classes, training or do something to help you get a job, a better job or do better at the job you have now?	2	Yes
	Additional Variables Identified in Classification Tree, but Not in Random Forest		

ERROTH	Who did you usually [run errands] with- others not listed	1	No
STATENUM	State Number	25	Massachusetts
STATENUM	State Number	50	Vermont
SURVEYORSEC1	In your opinion, did the individual appear to understand most questions, and did they answer in a consistent manner?	2	Yes