Alan Johnson

HON 4355

Final Report

Executive Summary

This report examines the factors that influence the employment of disabled individuals across the United States, with a particular interest in the effectiveness of Vocational Rehabilitation (VR) services. In the process of determining the most significant factors, there are several statistical models that are considered to predict the employment of disabled individuals. The main body of the report is structured as follows:

1. **Motivation**: Overview of the research purpose and goals.
2. **Data and Metadata**: Explanation of the data sources, variables included in the dataset, and filtering process to modify the raw data. Reasoning for methodological choices to compile the dataset.
3. **Analysis**: Summary of key insights, summary statistics, and graphs. Reasoning for methodological choices to create visualizations.
4. **Results and Interpretation**: Summary of key insights and evidence to support interpretation. Reasoning for methodological choices to analyze the data.
5. **Conclusion**: Final thoughts and key take away messages.

For readers seeking detailed data or extended discussions, appendices provide supplemental materials, including:

- Repository hosting source code
- Pruning the decision tree

- Neural Network architecture

1. Motivation

According to the U.S. Bureau of Labor Statistics only 40% of the working-age disabled individuals are employed. Vocational Rehabilitation (VR) services such as the Texas Workforce Commission (TWC) aims to provide support, skills, and opportunities for those with disabilities. However, these programs have struggled to improve employment as seen by the recent TWC Overview of Vocational Rehabilitation Service Needs and Strategies. “55% of working-age individuals with disabilities were not in the labor force during 2018, compared to 20% of those without a disability.” Judging from this disparity, the need for quality VR services is evident [1]. To address the lack of employment opportunities, the research question this report aims to answer is “What are the factors that increase the employment opportunities for disabled individuals in the United States?”

By comparing federal data on state-level employment outcomes, this report will highlight the relative shortcomings of VR services and explore how addressing these gaps could improve employment opportunities for the population of interest. Disabled individuals, their families, employers, and policymakers are directly impacted by the success of VR Programs. Therefore, it is important to keep VR services accountable by determining if they significantly impact the employment of disabled individuals and in doing so, reveal other factors which VR services should consider when tackling employment for this population.

This research is critical to understanding and dismantling the barriers to workforce participation for this demographic, a step essential for fostering inclusivity and equity in society. Ultimately, this report is intended to inspire policy changes that not only enhance rehabilitation services but also benefit the workforce and society by enabling greater participation and productivity. Key stakeholders to this project are state and federal government agencies, special education departments, non-profits, and advocacy groups working with disabled individuals.

2. Data and Metadata

The source used to create the dataset is the [Current Population Survey, July 2019: Disability Supplement]() from the Census Bureau [2]. Since the original dataset contains over 400 variables, only variables related to disability and general demographic variables were considered for this report. The original dataset was filtered to the following 29 input variables:

- how_disability_affects_ability_to_work: Describes how the individual's disability impacts their ability to work (e.g., limits the type or amount of work they can do)
- used_vocational_rehabilitation_agencies: Indicates whether the individual has used vocational rehabilitation (VR) services
- used_one_stop_career_centers: Indicates whether the individual has accessed One-Stop Career Centers for employment support
- used_the_ticket_to_work_program: Indicates whether the individual has participated in the Ticket to Work program
- used_assistive_technology_act_prog: Indicates whether the individual has utilized programs under the Assistive Technology Act
- used_ctr_for_indpt_living_for_ind_w_dis: Indicates whether the individual has used services from Centers for Independent Living
- used_the_client_assistance_program: Indicates whether the individual has accessed services from the Client Assistance Program

- used_any_other_employment_assistance_program: Indicates whether the individual has utilized any other employment assistance programs
- how_helpful_vocational_rehab_agency: Measures the perceived helpfulness of VR agency services
- how_helpful_one_stop_career_centers: Measures the perceived helpfulness of One-Stop Career Center services
- the_ticket_to_work_program_helpfulness: Measures the perceived helpfulness of the Ticket to Work program
- ctr_for_indpdt_living_for_ind_w_dis_helpful: Measures the perceived helpfulness of Centers for Independent Living services
- other_employment_assist_program_helpful: Measures the perceived helpfulness of other employment assistance programs
- does_this_person_have_any_of_these_disability_conditions: Identifies the types of disability conditions the individual experiences
- barrier_lack_of_education_or_training: Indicates if lack of education or training is a barrier to employment
- barrier_lack_of_job_counseling: Indicates if lack of job counseling is a barrier to employment
- barrier_lack_of_transportation: Indicates if lack of transportation is a barrier to employment
- barrier_loss_of_government_assistance: Indicates if fear of losing government assistance is a barrier to employment

- barrier_need_for_special_features: Indicates if the need for special features or accommodations is a barrier to employment
- barrier_employer_or_coworker_attitudes: Indicates if negative employer or coworker attitudes are a barrier to employment
- barrier_your_difficulty_with_disability: Indicates if personal difficulties associated with a disability are a barrier to employment
- barrier_other: Indicates if there are other unspecified barriers to employment
- barrier_none: Indicates if the individual faces no barriers to employment
- demographics_race_of_respondent: Indicates the race or ethnicity of the respondent.
- demographics_sex: Indicates the respondent's gender
- demographics_age: Indicates the respondent's age
- household_total_family_income_past_12_months: Indicates the total household income in the past 12 months
- state: The state of residence of the respondent
- county: The county of residence of the respondent

and one output variable:

- labor_force_employment_status: Binary classification indicating whether the individual is "Employed-At Work" (1) or not employed (0)

The reasoning behind the methodological decisions to include and exclude variables in this dataset were partly due to practicality and due to modeling causal chains. Since the original dataset had far too many variables to analyze, features had to be cut down to the most relevant (related to disability and employment). A causal diagram (Directed Acyclic Graph) was created to analyze the potential impact of these variables to the output variable. The variables that were added to this diagram became the final 29 candidates for the report.

In addition to filtering the number of variables, further data processing steps were undertaken to refine the dataset for analysis. First, records that did not fall within the universe of valid labor_force_employment_status values were removed. Columns with names starting with "Unnamed" were also excluded to clean up the data. Next, the dataset was filtered to include only records of individuals with a disability, ensuring the analysis focused on the target population. Finally, the labor_force_employment_status variable was recategorized into binary values, with 1 representing "Employed-At Work" and 0 representing all other statuses. These preprocessing steps helped streamline the dataset for meaningful analysis.

3. Analysis

In order to answer the research question, this report considered four predictive models that are trained to fit the data and evaluate the feature importance. An 80-20 training and testing split was used for all four models. The first model that was trained was a logistic regression model, because it is often used for binary classification problems. Logistic regression models the probability of a binary outcome using a linear combination of input gestures transformed by the sigmoid (logistic) function [3]. This model achieved a macro average of 91% precision, 92% recall, and 92% f1-score.

Since the precision (which is the ratio of true positives over total predicted positives) was relatively low for the logistic regression, the next model that was considered was a decision tree. A decision tree recursively splits data into subsets based on feature values, until reaching leaf nodes that provide the final classification. Since many of the feature variables were categorical, a decision tree seemed like a good candidate because they naturally handle splitting nodes based on feature values in an intuitive, rule-based decision making process [4]. However, decision trees are notorious for overfitting the data; thus, the model requires a pruning process which is described in detail in the appendix. The pruned tree achieved a macro average of 95% precision, 91% recall, and 93% f1-score.

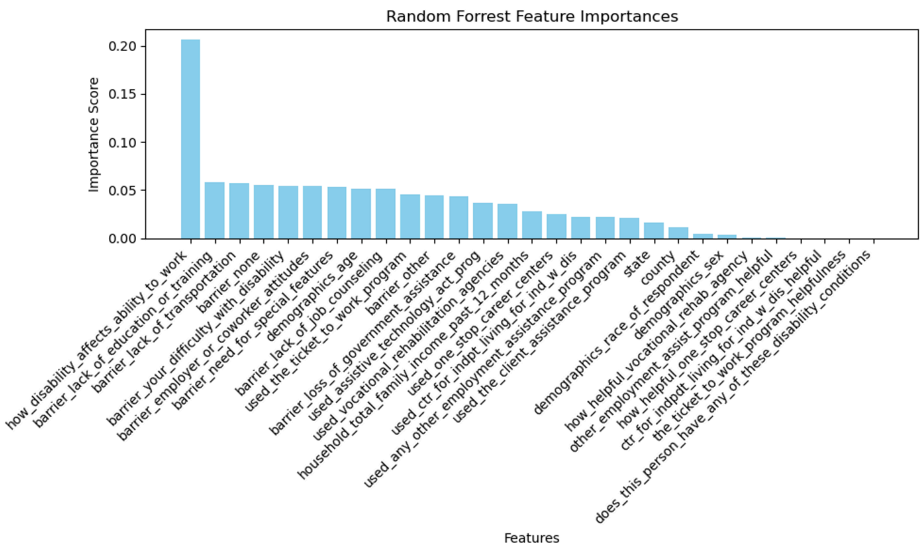

Figure 1. Feature importances (average reduction of entropy) in a Random Forrest Model

Although overfitting in the decision tree was reduced via pruning, the final tree was still quite large (Figure 6). To create a more generalizable model, a Random Forest model was chosen because it builds multiple decision trees during training, each using a random subset of features to reduce overfitting [5]. The final prediction is made by aggregating the outputs of individual trees. A Random Forest model was trained using the entropy criterion and achieved a macro average of 96% precision, 91% recall and 93% f1-score. Since this is the most accurate model thus far, the feature importances of the Random Forest model was visualized in Figure 1. A bar graph was chosen for this visualization because it most clearly answers the research question by showing the most significant factors that affect the employment of disabled individuals in order.

This final model considered was a Neural Network in order to maximize the precision of the predictive models. A multi layer perceptron with one hidden layer and ReLU activation was trained on 1000 epochs. It utilizes cross-entropy loss and the Adam Optimizer [6]. A more detailed explanation of the architecture of the Neural Network can be found in the appendix. This model achieved a macro average of 98% precision, 90% recall, and 93% f1-score which can be seen in the classification report in Figure 2.

	precision	recall	f1-score	support
0	0.96	1.00	0.98	1957
1	1.00	0.79	0.88	434
accuracy			0.96	2391
macro avg	0.98	0.90	0.93	2391
weighted avg	0.96	0.96	0.96	2391

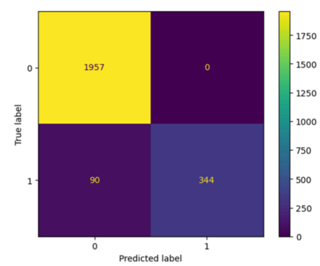

Figure 2. Classification Report and Confusion Matrix for Neural Network

Lastly, a choropleth map was created to better compare the rate of employment across states. As shown in Figure 3, this map represents the percentage of disabled working-age individuals who are in the universe of employment that were employed as of 2019. This map is only based on the dataset used for this report, therefore it may not be representative of the true population [2].

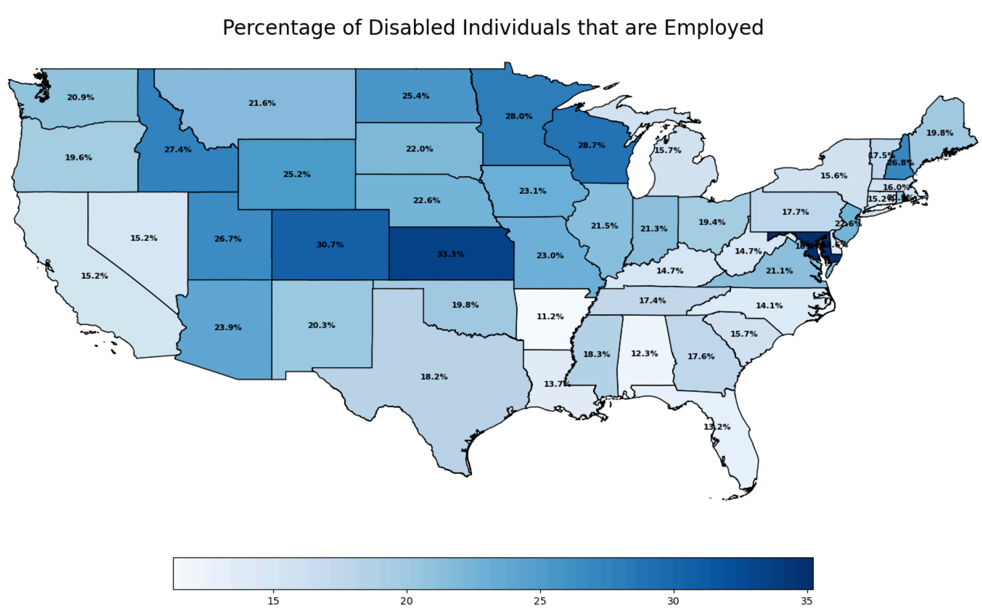

Figure 3. Percentage of Disabled Working-Age Individuals that are Employed

4. Results and Interpretation

Figure 1 demonstrates that some of the most significant factors that influence the employment of disabled individuals are the severity at which an individual's disability affects their ability to work, barriers due to lack of education or training, barriers due to lack of transportation and barriers due to the need for special accommodations or workplace features. It is important to note that the use of vocational rehabilitation services (used_vocational_rehabilitation_agencies and how_helpful_vocational_rehab_agency) is not one of the top ten most significant features that affects the employment of the disabled individuals in the dataset.

Figure 1 is appropriate to analyze the data because the Random Forest model has the best balance of precision and recall. Therefore, there is strong evidence that VR services did not play a significant role in the employment of the studied population. These results clearly answer the research question by visualizing the most critical factors that affect the employment of disabled individuals.

Although Figure 3 only visualizes the filtered data from the dataset, it still provides strong evidence that Texas falls behind on enabling working age disabled individuals to find employment opportunities. Unlike states in the Midwest such as Kansas and Colorado, only 18.2% of disabled individuals were employed in Texas as of 2019.

Despite these results, some limitations to the dataset is that there is a class imbalance of individuals who are unemployed vs employed. As shown by Figure 2, the number of disabled individuals that are unemployed in the testing split greatly outweighs the number of employed individuals. This could cause a bias toward the majority class and may not generalize well to real-world scenarios correctly. Additionally, since this dataset collected information from a single year, these results cannot prove the long term of effectiveness of VR services. Utilizing longitudinal studies, such as National Longitudinal Transition Study (NLTS), to closely research the effectiveness of VR services in youth transition would help to better answer the research question.

5. Conclusion

Current VR services have struggled to effectively assist disabled working-age individuals in securing stable employment due to inconsistent service delivery across states, a lack of resources, and a limited focus on long-term career development. While some programs offer job training and placement support, they often fail to address the diverse needs of individuals with disabilities, such as specialized accommodations or long-term career counseling.

The results from this report highlight the need to further investigate why vocational rehabilitation services have limited impact and explore strategies to enhance their effectiveness in supporting disabled individuals' employment. This report hopes to inspire policy, especially in the state of Texas, to address the most significant barriers to employment which were discussed in the previous section.

The biggest takeaway from this report is that society is still a long way from truly inclusive employment practices that empower individuals with disabilities to reach their full professional potential. Despite incremental progress, systemic barriers continue to limit opportunities, creating a persistent employment gap that requires comprehensive, innovative approaches. The challenges extend beyond mere job placement, demanding a holistic reimagining of vocational rehabilitation services that center the unique strengths, aspirations, and individual needs of the disabled working-age population. By recognizing these complexities and committing to transformative solutions, policymakers and service providers can begin to dismantle the structural obstacles that have historically marginalized disabled workers.

References

[1] "2020 Comprehensive Statewide Needs Assessment Report." An Overview of Vocational Rehabilitation Services Needs and Strategies, Texas Workforce Commission, www.twc.texas.gov/sites/default/files/vr/docs/2020-comprehensive-statewide-needs-assessment-twc.pdf.

[2] https://api.census.gov/data/2019/cps/disability/jul.html

[3] https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

[4] https://scikit-learn.org/stable/modules/tree.html

[5] https://scikit-learn.org/1.5/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[6] Marquez-Soto, Pedro. "Create a ML Classification Model with Pytorch." Pedro's Tech Blog, 3 Oct. 2022, pedromarquez.dev/blog/2022/10/pytorch-classification.

Appendix A: Repository with complete source code

Here is a link to the repository which contains the data dictionary, DAG model, and the complete source code for all the models discussed in this report: notebook.ipynb

Appendix B: Decision Tree Cost Complexity Pruning

Figure 4 demonstrates that the preliminary decision tree was too large and severely overfitted the data. To optimize the tree based on the results of the classification report, pruning was necessary.

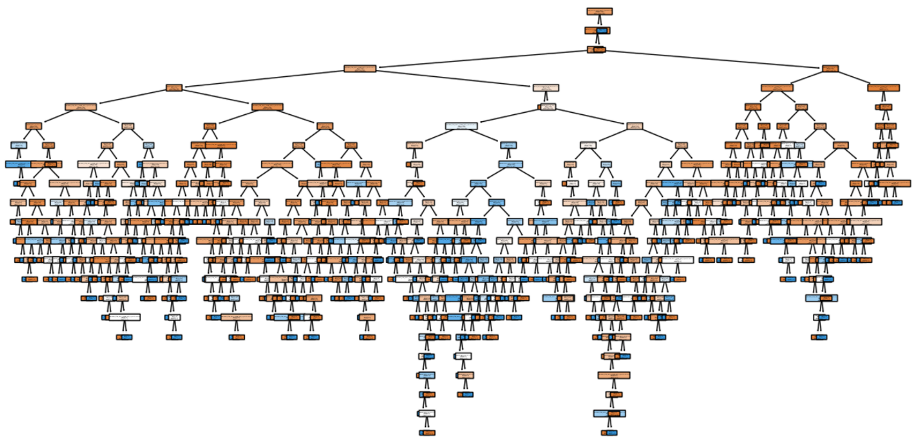

Figure 4. Unpruned Decision Tree

For decision trees, the pruning parameter alpha controls how little or how much pruning takes place. One way to find the optimal value for alpha is to plot the accuracy of the tree as a function of different alpha values. This is known as cost complexity pruning. Additionally, cross validation is used to ensure that the optimal alpha value is not sensitive to a particular dataset. The results of cross complexity pruning are shown in Figure 5

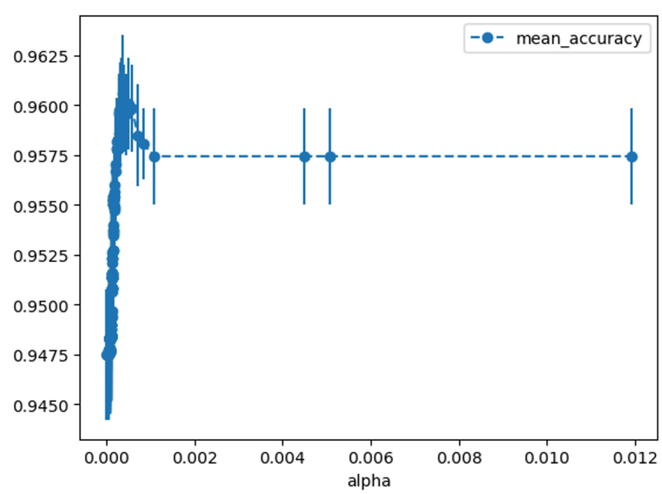

Figure 5. Calculating the ideal value for alpha parameter

As shown by Figure 5, the ideal value for alpha is in between 0.000 and 0.002. By splicing the data within this range and sorting the value in decreasing order of mean_accuracy, the ideal alpha value was found to be 0.000373. The final tree using the ideal alpha value from cross complexity pruning is shown in Figure 6.

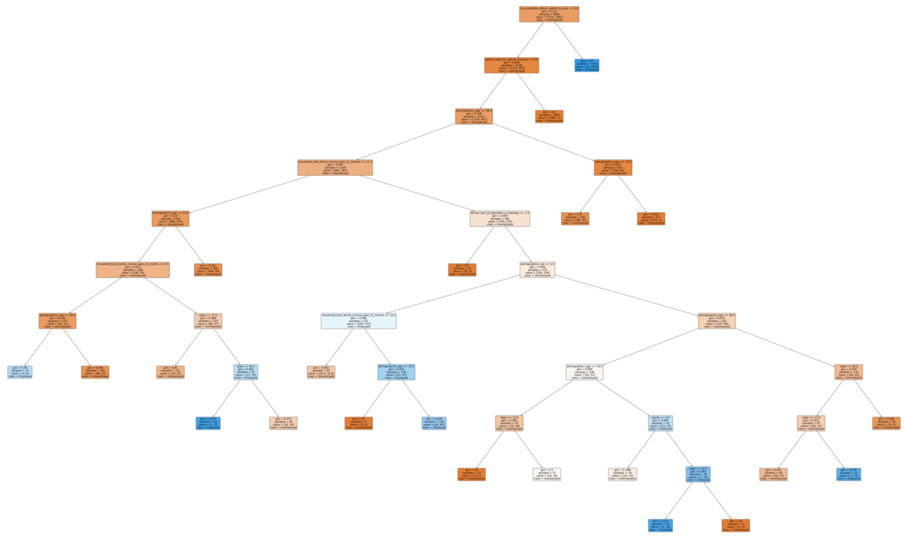

Figure 6. Pruned Decision Tree Predicting Employment Likelihood (Orange=0, Blue=1)

Appendix C: Neural Network Architecture

The neural network is a fully connected binary classification model designed to predict labor force employment status. Its architecture features two primary layers: a hidden layer with 270 neurons that transform input features, and an output layer with two neurons representing the binary classification outcome. To enhance learning, the network incorporates ReLU activation for introducing non-linearity and a dropout rate of 0.1 to prevent overfitting by randomly deactivating a small percentage of neurons during training.

The training methodology leverages the Adam optimizer with a learning rate of 0.01, which adjusts gradient descent to find optimal model parameters. Cross-entropy loss function is employed to measure the divergence between predicted and actual probabilities, providing precise feedback for model refinement. The network processes data in mini-batches of 50 samples, with data shuffled at each of the 1,000 training epochs to prevent overfitting.