Supplement to Yufei Wu’s MADA Project

# 1. Overview

You can find all the supplemetary materials (figures and tables) in the supplement.

# 2. Additional results

## 2.1 Exploratory/Descriptive analysis

[Table 1](#tbl-sum_prev_wage) shows a summary table for prevailing wage.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Summary Table for Prevailing Wage.   | count | mean\_wage | median\_wage | sd\_wage | min\_wage | max\_wage | | --- | --- | --- | --- | --- | --- | | 3023857 | 102611.2 | 97261 | 34807.18 | 15080 | 760202 | |

[Figure 1](#fig-bottom_wage_by_soc_title) shows a bar plot of bottom 10 job titles by lowest median wage. [Figure 2](#fig-bottom_wage_by_employer) shows a bar plot of bottom 10 employers by lowest median wage. [Figure 3](#fig-bottom_wage_by_state) shows a bar plot of bottom 10 states by lowest median wage.

|  |
| --- |
| Figure 1: Bottom 10 Job Titles by Lowest Median Wage. |

|  |
| --- |
| Figure 2: Bottom 10 Employers by Lowest Median Wage. |

|  |
| --- |
| Figure 3: Bottom 10 States by Lowest Median Wage. |

## 2.2 Model Fitting

[Table 2](#tbl-rf_combined_accuracy_cv_default) to [Table 4](#tbl-combined_accuracy_cv_svm) show the model accuracy in both training and test data of the three models: Random Forest, XGBoost, and SVM.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Accuracy of Random Forest.   | dataset | .metric | .estimator | .estimate | | --- | --- | --- | --- | | Train | accuracy | multiclass | 0.8202899 | | Test | accuracy | multiclass | 0.7398844 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Accuracy of XGBoost.   | dataset | .metric | .estimator | .estimate | | --- | --- | --- | --- | | Train | accuracy | multiclass | 0.7376812 | | Test | accuracy | multiclass | 0.7254335 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: Accuracy of SVM.   | dataset | .metric | .estimator | .estimate | | --- | --- | --- | --- | | Train | accuracy | multiclass | 0.8862319 | | Test | accuracy | multiclass | 0.7167630 | |

## 2.3 Full analysis

[Figure 4](#fig-cv_accuracy_distribution_panel) shows CV accuracy distribution for different models.

|  |
| --- |
| Figure 4: CV accuracy distribution for different models. |