

Homework, Households, and Hurdles: The Unexpected Drivers of Student Graduation Perceptions Supplementary Text

Daniel Alhassan Zahra Fatah Priscilla Mansah Codjoe Caroline Bena Kuno
Dorcas Ofori-Boateng

S1 Hyperparameters and ranges explored for each model

The table below displays the range of hyperparameters explored for each of the seven machine learning algorithms described in the original text.

Table S1: Hyperparameters and ranges explored for each model

Model	Range of hyperparameters explored
Logistic Regression	Not applicable
Decision Trees	cost_complexity: [0.001, 0.01], min_n: [1, 10]
Random Forest	trees: [50, 100], min_n: [1, 10]
Neural Networks	penalty: [0.001, 0.1], epochs: [50, 100]
Gradient Boosting	trees: [50, 100], tree_depth: [1, 7], learn_rate: [0.01, 0.1]
KNN	neighbors: [1, 20]
AdaBoost	trees: [50, 100]

S2 Results from Statistical Analysis of Training and Testing Sample Data

The following output represents the results from the preliminary statistical analysis of training and testing sample data performed to ensure a similar distribution of features in both training and testing samples.

Figure S2: These plots show the distributions of features 9–16 in the training and testing samples with chi-square statistics and the respective p-values.

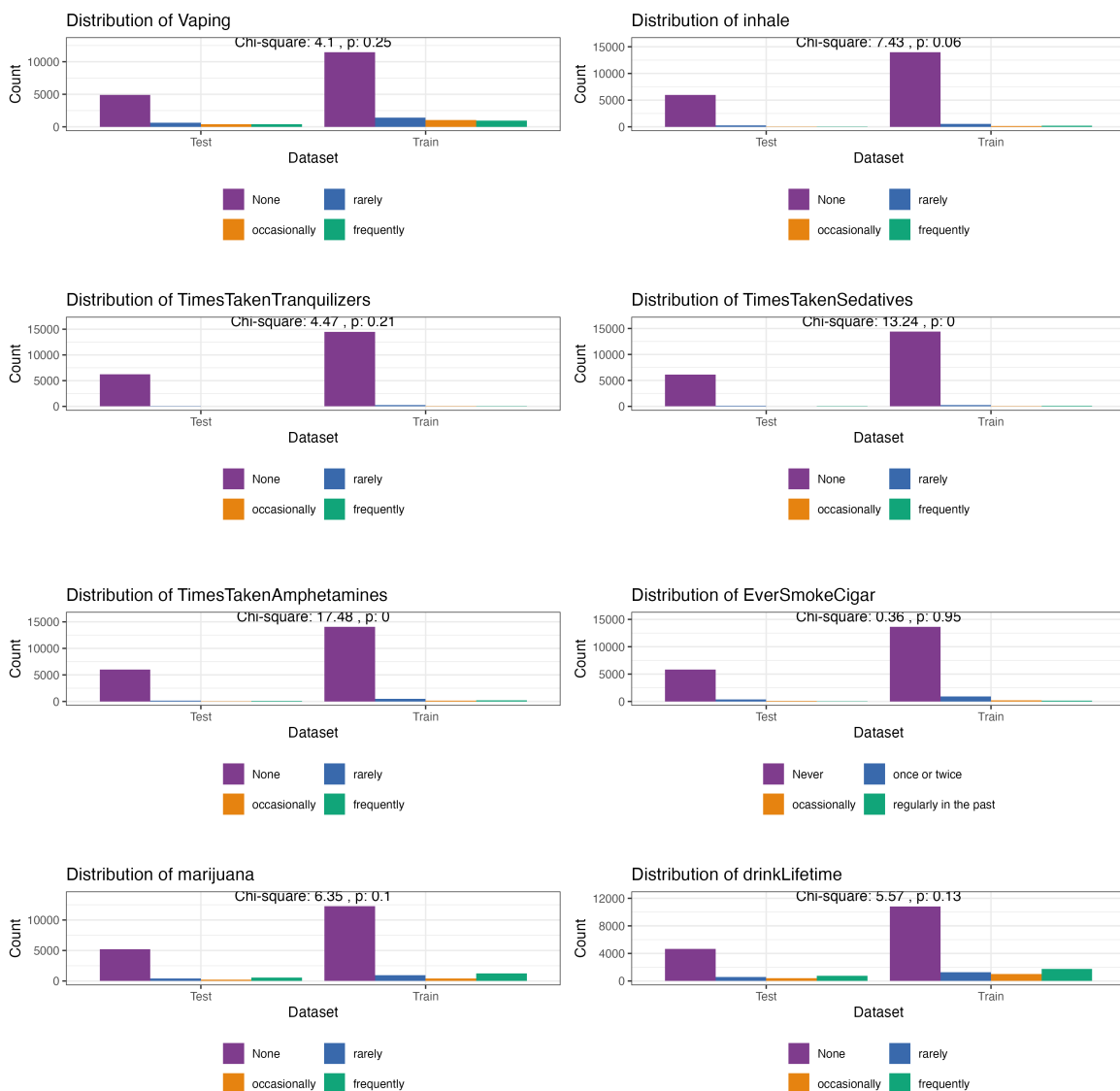


Figure S3: These plots show the distributions of features 17–24 in the training and testing samples with chi-square statistics and the respective p-values.

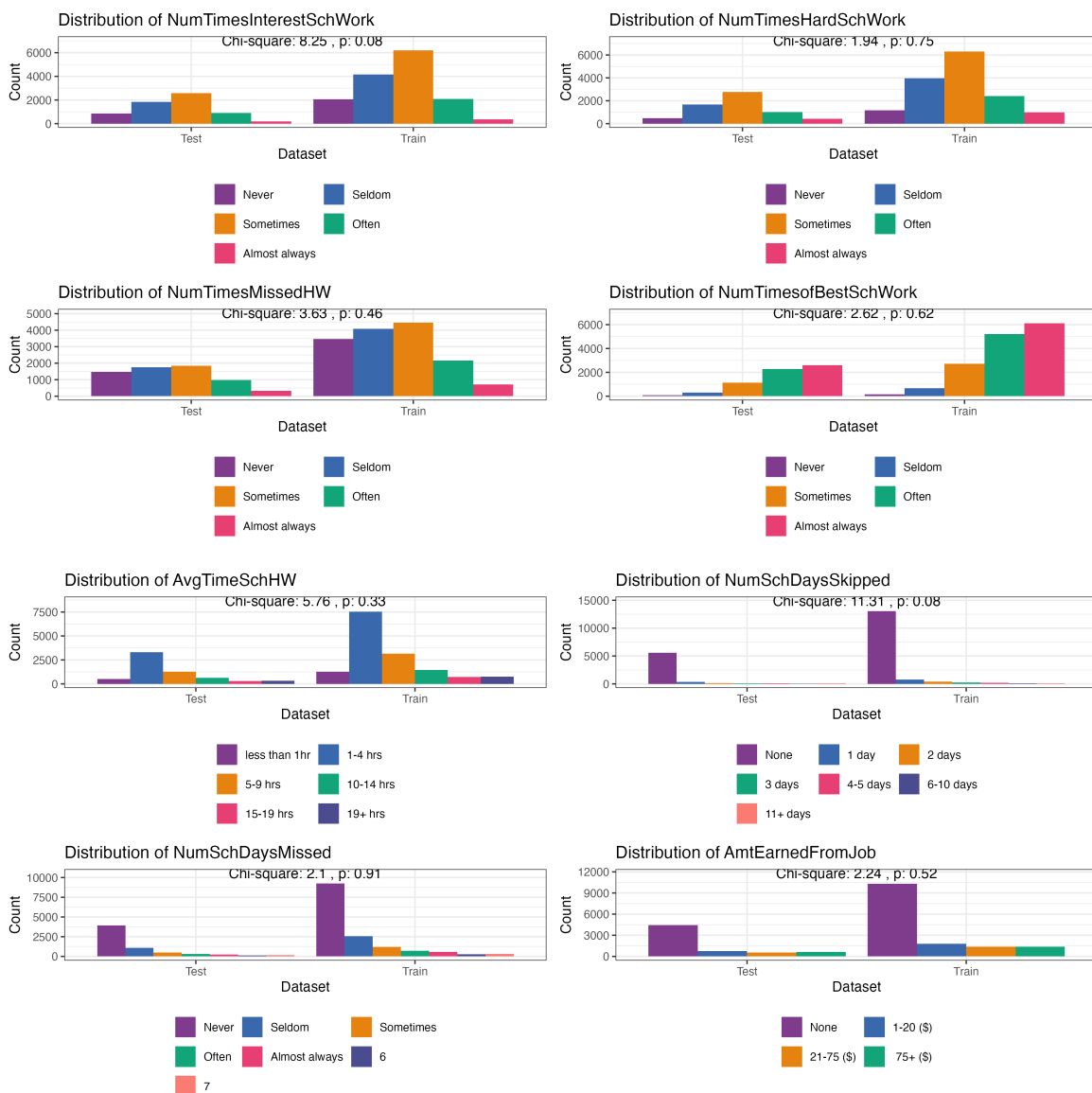


Figure S4: These plots show the distributions of the last feature in the training and testing samples with chi-square statistics and the respective p-values.

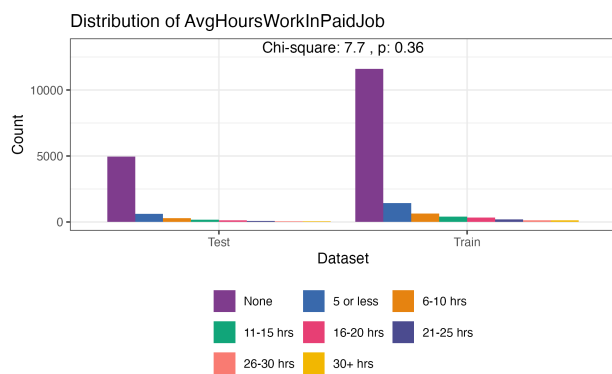


Figure S1: These plots show the distributions of the first eight features in the training and testing samples, with chi-square statistics for categorical features, t statistics for numeric features, and the respective p-values.

