CS699 Data Mining Project

Predicting Difficulty Living Independently

Prepared by

* Walter Stock
* Arfaan Jeddy

Institution

* Boston University, MET College

Course

* Data Mining
* CS699 A1, Spring 2025
* Instructor: Jae Young Lee

Date

* April 7th, 2025

Table of Contents

[Summary 2](#_Toc194744659)

[Datamining Tools 2](#_Toc194744660)

[Algorithms 2](#_Toc194744661)

[Datamining Procedure 2](#_Toc194744662)

[Datamining Results 2](#_Toc194744663)

# Summary

F

# Datamining Tools

F

# Algorithms

F

# Datamining Procedure

This is a detailed overview of the procedure used to complete this project. The project was completed with a single R file, ***full\_project.R***, that is included with the submission. There are 5 sections: Preprocess, Split, Balance, Select, Model.

## Preprocessing

Preprocessing includes loading the data provided for the project, ***project\_data.csv***, and then the following steps.

### Preprocessing – 1. Removal of Columns without Value

Columns without value are immediately removed to provide the greatest benefit. These columns all have the same value for all records and include: STATE, REGION, DIVISION, ADJINC, and RT.

### Preprocessing – 2. Data Types

Information about each column of data will be important throughout the project. Here we create a data frame, ***df\_columns\_info***, with every column name and the class of the column after import. All columns are reviewed and classified as Integer, Factor, Factor Levels, Logical.

### Preprocessing – 3. Data Dictionary

To assist with analysis and communication the data dictionary for the data being analyzed is downloaded, ***PUMS\_Data\_Dictionary\_2023.csv,*** and loaded into data frame, ***data\_dict\_df***.

# Datamining Results