

index

Welcome

This site presents an analysis of the **Diabetes Health Indicators** dataset from the 2015 BRFSS survey. The main outcome of interest is `Diabetes_binary`, which indicates whether a respondent has diabetes or not.

This project has two main parts:

1. **Exploratory Data Analysis (EDA)** – understanding the data, distributions, and relationships.
 2. **Modeling** – building and comparing predictive models (classification tree and random forest) using the `tidymodels` framework.
-

Pages

1. Exploratory Data Analysis

The EDA page:

- Describes the dataset and key variables.
- Recodes variables into meaningful factor levels.
- Explores diabetes prevalence overall and across groups (example., sex, BMI, general health).
- Summarizes numeric variables and examines correlations.

[The EDA Page](#)

2. Modeling

The Modeling page:

- Splits the data into training (70%) and test (30%) sets with stratification on `Diabetes_binary`.
- Defines a common **recipe** for pre processing predictors.
- Fits and tunes:
 - A **classification tree**
 - A **random forest**
- Uses **5 fold cross validation** with **log loss** and **accuracy** to select the best model in each family.
- Compares the final models on the test set and selects an overall "winner" for deployment as an API.

[The Modeling Page](#)

Notes

- The final selected model from the Modeling page is used in a separate `.R` file to define a **Plumber API**, which is then packaged in a **Docker** image for deployment.