Disaster-Relief-Project-Part-1

Mahin Ganesan, Wyatt Priddy, and Taylor Tucker

2024-03-07

Introduction

INTRO TEXT...

Data

TEXT

Table 1: Missing Values in Haiti Data Set

Class	Red	Green	Blue	BlueTarp
0	0	0	0	0

Table 2: Correlation Matrix of Predictor Variables

	Red	Green	Blue
Red	1.000	0.982	0.936
Green	0.982	1.000	0.965
Blue	0.936	0.965	1.000

Description of Methodology

TEXT

Results

Conclusion

TEXT

Appendix

```r
knitr::opts_chunk$set(echo=FALSE)
knitr::opts_chunk$set(cache=TRUE, autodep=TRUE)
knitr::opts_chunk$set(fig.align="center", fig.pos="tbh")
# Set up Parallel Processing
library(doParallel)

cl <- makePSOCKcluster(parallel::detectCores(logical = FALSE))

registerDoParallel(cl)

# Load Libraries
library(tidyverse)
library(corrplot)
library(tidymodels)

# Read in Data
haiti <- read_csv('https://gedeck.github.io/DS-6030/project/HaitiPixels.csv', show_col_types=FALSE) %>%
          mutate(BlueTarp= factor(ifelse(Class=="Blue Tarp", "Yes", "No"),labels=c("No", "Yes")))

# Look for Missing Values
haiti %>%
  summarise_all(~sum(is.na(.))) %>% knitr::kable(caption="Missing Values in Haiti Data Set")
# View Correlation of Variables
haiti %>%
  select_if(is.numeric) %>%
  cor() %>%
  knitr::kable(digits=3, caption='Correlation Matrix of Predictor Variables')
# Set seed
set.seed(81718)

# Create initial split for 80/20 with stratified sampling on Hazardous
haiti_split <- initial_split(haiti, prop=.8, strat=BlueTarp)

# Create training data set
train_data <- training(haiti_split)

# Create test data set
test_data <- testing(haiti_split)

# Set up 10-fold cross-validation
resamples <- vfold_cv(train_data, v=10, strata=BlueTarp)

# Set settings for control resamples
cv_control <- control_resamples(save_pred=TRUE)
```