Data 622 - Homework 3

Leticia Salazar

December 3, 2023

- Read the following articles:
  - https://www.hindawi.com/journals/complexity/2021/5550344/
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8137961/
- Search for academic content (at least 3 articles) that compare the use of decision trees vs SVMs in your current area of expertise.
- Perform an analysis of the dataset used in Homework #2 using the SVM algorithm.
- Compare the results with the results from previous homework.
- Answer questions, such as:
  - Which algorithm is recommended to get more accurate results?
  - Is it better for classification or regression scenarios?
  - Do you agree with the recommendations?
  - Why?

Format * Essay (minimum 500 word document) * Write a short essay explaining your selection of algorithms and how they relate to the data and what you are trying to do * Analysis using R or Python (submit code + errors + analysis as notebook or copy/paste to document) * Include analysis R (or Python) code.

Load Libraries: Below are the libraries used to complete this assignment

```
library(tidyverse) # data prep
```

```
FALSE -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
FALSE v dplyr     1.1.3     v readr     2.1.4
FALSE v forcats   1.0.0     v stringr   1.5.0
FALSE v ggplot2   3.4.3     v tibble    3.2.1
FALSE v lubridate 1.9.3     v tidyr     1.3.0
FALSE v purrr     1.0.2
FALSE -- Conflicts --------------------------------------- tidyverse_conflicts() --
FALSE x dplyr::filter() masks stats::filter()
FALSE x dplyr::lag()    masks stats::lag()
FALSE i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become er
```

```r
library(skimr) # data prep
library(rpart) # decision tree package
library(rpart.plot) # decision tree display package
library(knitr) # kable function for table
library(tidyr) # splitting data
library(ggplot2) # graphing
library(hrbrthemes) # chart customization
```

```
FALSE NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.
FALSE       Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and
FALSE       if Arial Narrow is not on your system, please see https://bit.ly/arialnarrow
```

```r
library(gridExtra) # layering charts
```

```
FALSE
FALSE Attaching package: 'gridExtra'
FALSE
FALSE The following object is masked from 'package:dplyr':
FALSE
FALSE     combine
```

```r
library(stringr) # data prep
library(tidymodels) # predictions
```

```
FALSE -- Attaching packages --------------------------------------- tidymodels 1.1.1 --
FALSE v broom        1.0.5     v rsample      1.2.0
FALSE v dials        1.2.0     v tune         1.1.2
FALSE v infer        1.0.5     v workflows    1.1.3
FALSE v modeldata    1.2.0     v workflowsets 1.0.1
FALSE v parsnip      1.1.1     v yardstick    1.2.0
FALSE v recipes      1.0.8
FALSE -- Conflicts ------------------------------------------ tidymodels_conflicts() --
FALSE x gridExtra::combine() masks dplyr::combine()
FALSE x scales::discard()    masks purrr::discard()
FALSE x dplyr::filter()      masks stats::filter()
FALSE x recipes::fixed()     masks stringr::fixed()
FALSE x dplyr::lag()         masks stats::lag()
FALSE x dials::prune()       masks rpart::prune()
FALSE x yardstick::spec()    masks readr::spec()
FALSE x recipes::step()      masks stats::step()
FALSE * Dig deeper into tidy modeling with R at https://www.tmwr.org
```

```r
library(corrplot) # correlation plot
```

```
FALSE corrplot 0.92 loaded
```

```r
library(randomForest) # for the random forest
```

```
FALSE randomForest 4.7-1.1
FALSE Type rfNews() to see new features/changes/bug fixes.
```

```
FALSE
FALSE Attaching package: 'randomForest'
FALSE
FALSE The following object is masked from 'package:gridExtra':
FALSE
FALSE     combine
FALSE
FALSE The following object is masked from 'package:dplyr':
FALSE
FALSE     combine
FALSE
FALSE The following object is masked from 'package:ggplot2':
FALSE
FALSE     margin
```

```
library(caret) # confusion matrix
```

```
FALSE Loading required package: lattice
FALSE
FALSE Attaching package: 'caret'
FALSE
FALSE The following objects are masked from 'package:yardstick':
FALSE
FALSE     precision, recall, sensitivity, specificity
FALSE
FALSE The following object is masked from 'package:purrr':
FALSE
FALSE     lift
```

Load Data: The data chosen is from []. The data set is included in my GitHub and read into R.

The Data:

Data Exploration:

Data Preparation:

Model Building:

Conclusion: