Group Coursework Submission Form

Specialist Masters Programme

Please list all names of group members: (Surname, first name) 1. Arora, Akshay 2. Chawla, Sarthak 3. Gaur, Siddharth	4. Li, Jiaying 5. Wu, Ziyan	
	GROUP NUMBER:	**2**

MSc in:
Business Analytics

Module Code:
SMM636

Module Title:
Machine Learning

Lecturer: Zhu, Rui	**Submission Date:** 03/03/2023

Declaration:

By submitting this work, we declare that this work is entirely our own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work we acknowledge that we have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. We also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.

Marker's Comments (if not being marked on-line):

Deduction for Late Submission: **Final Mark:**   **%**

Group 2. R mainly includes the following parts:
- A user interaction (UI) object

- A server functions

This study uses R built-in mtcars dataset to develop shiny app. Through this app, data visualization analysis, decision tree model construction, random forest model construction, etc. are realized, and the importance of variables is visually analyzed according to the results of the classification model. By publishing the Shiny application developed in this research on Shinyapps, it can also be provided to more people remotely. An example dataset is shown below.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

The relationship between vs and mpg can be visualized through box diagram.
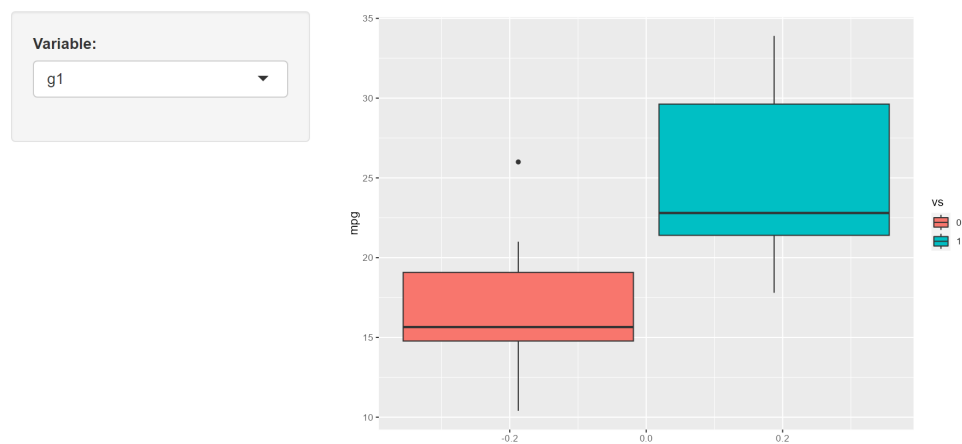
Visualize the relationship between disp and vs through box diagram.
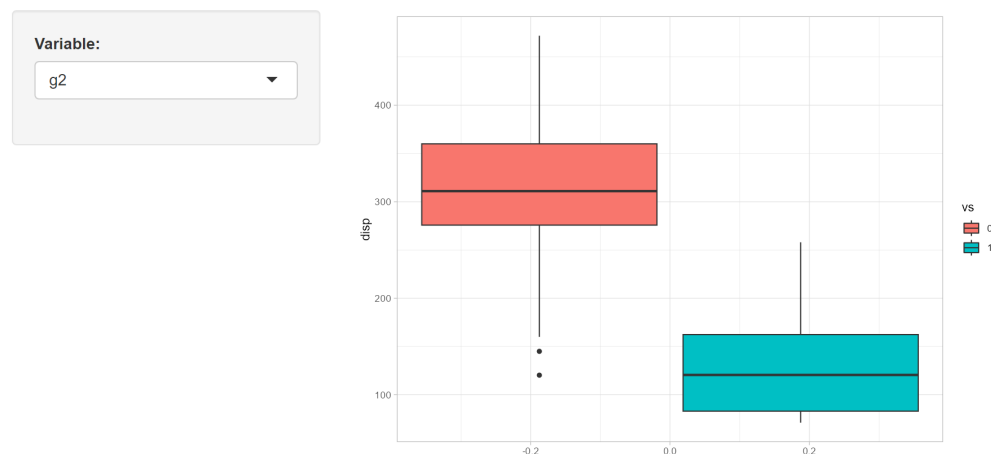

Through statistical analysis of the data through the box chart, the chart can obtain the minimum, maximum, quantile, and other indicator information. It can reflect the characteristics of distribution in original data and is also able to compare between different data distributions. Through this figure, the data is analyzed simply and explosively.

With variable vs (engine type) as the dependent variable, a decision tree model is built to display the model results. Build a decision tree model and visualize the model. The results are as follows.

APP function realization

Variable:

g3
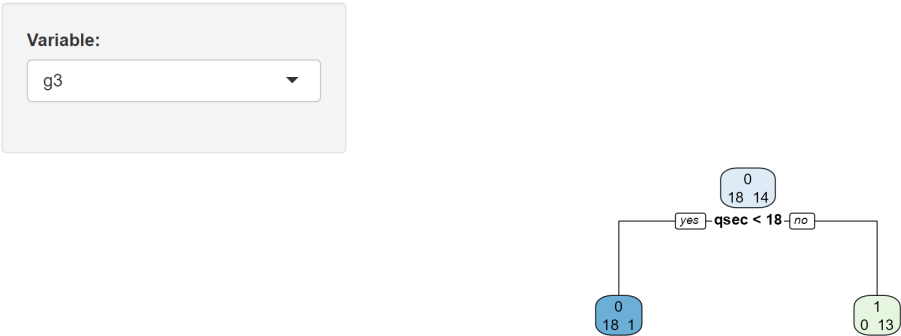
This figure intuitively shows the judgment process of the decision tree model.

According to the importance diagram of variables obtained from the decision tree model, the influence of variable qsec (measuring the starting acceleration ability) on the dependent variable ranks first, followed by the variable hp (total horsepower)

APP function realization

Variable:

Decision tree model
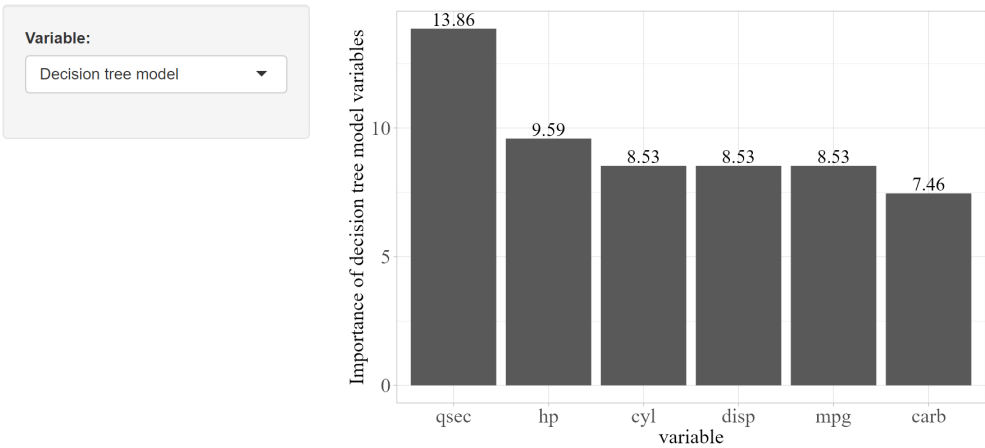
Draw the importance diagram of variables obtained from the random forest model. According to the diagram, the variable qsec (measuring the starting acceleration ability) has the greatest impact on the dependent variable, followed by the variable mpg (how many miles per gallon of oil can run)
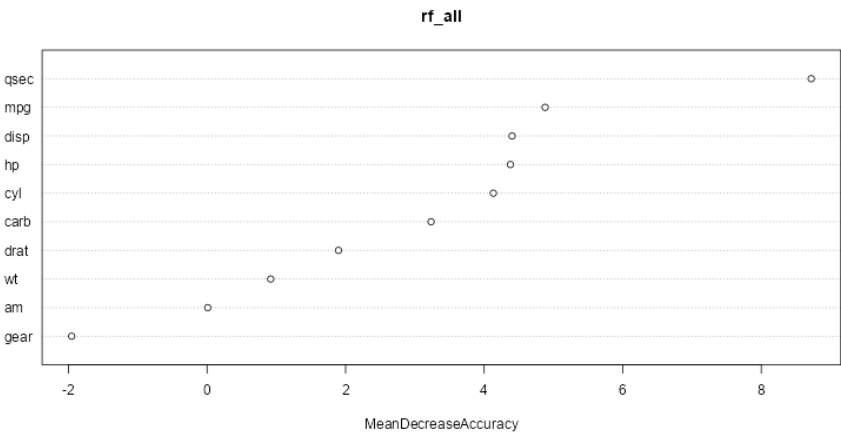
The following two figures show the results of the random forest model, including the average reduction accuracy chart and the average reduction gini coefficient chart. The average decline accuracy chart shows the decline of the prediction accuracy of the random forest model. The larger the index, the greater the importance of the independent variables in the model; The average reduction of gini coefficient graph determines how each variable's impact on the variation of the observed values at each node of the classification tree is calculated. The higher the value, the variable is more important.

APP function realization