Artwork/Project Title

Stroke Prediction Using Logistic Regression

Year Accomplished

2022

Role/Position

Data Scientist

Publication Link

https://rpubs.com/belindamutiara/929798

Artwork/Project Description

As part of my second-semester exam project, I successfully implemented logistic regression and decision tree algorithms in R language to predict stroke. The project encompassed various stages, including data overview, data preparation, data mining, model building, and evaluation. By leveraging my strong analytical and technical skills, I achieved an accuracy rate of 83% using logistic regression and an impressive 95% accuracy using a decision tree.

```
## 'data.frame':    5110 obs. of  12 variables:
##  $ id               : int  9046 51676 31112 60182 1665 56669 53882 10434 27419 60491 ...
##  $ gender           : chr  "Male" "Female" "Male" "Female" ...
##  $ age              : num  67 61 80 49 79 81 74 69 59 78 ...
##  $ hypertension     : int  0 0 0 0 1 0 1 0 0 0 ...
##  $ heart_disease    : int  1 0 1 0 0 0 1 0 0 0 ...
##  $ ever_married     : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ work_type        : chr  "Private" "Self-employed" "Private" "Private" ...
##  $ Residence_type   : chr  "Urban" "Rural" "Rural" "Urban" ...
##  $ avg_glucose_level: num  229 202 106 171 174 ...
##  $ bmi              : chr  "36.6" "N/A" "32.5" "34.4" ...
##  $ smoking_status   : chr  "formerly smoked" "never smoked" "never smoked" "smokes" ...
##  $ stroke           : int  1 1 1 1 1 1 1 1 1 1 ...
```

Figure 1 Data description

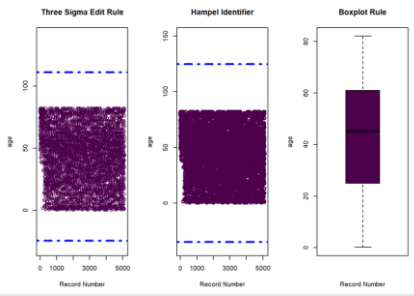

Figure 2 Outlier checking

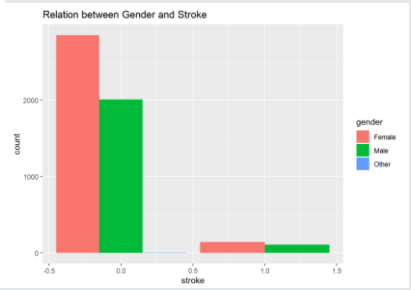

Figure 3 Data visualization

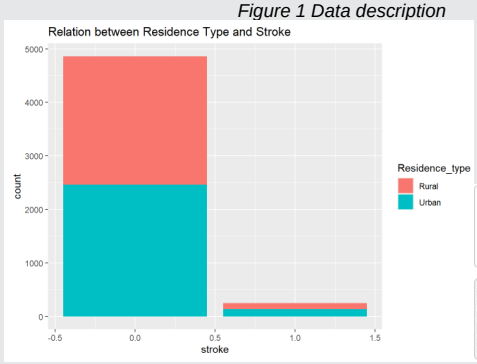

Figure 4 Data visualization

```
I separated the data into two groups with an 80:20 ratio. Of course, the t
raining set was bigger than the validation set. The training set contain o
f 4088 observations while validation set contain of 1021 observations.
```

```
LogisticModel <- glm(stroke~., data= trainingset, family='binomial'(link
="logit"))
summary(LogisticModel)
```

Figure 5 Model Building

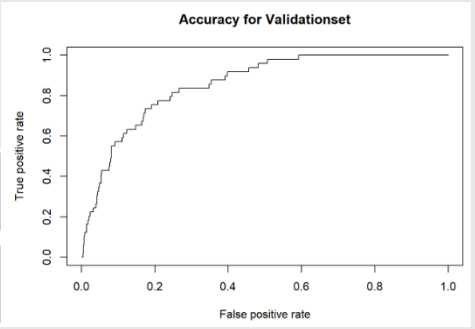

Figure 6 Evaluation on validation set

Figure 7 Evaluation on training set

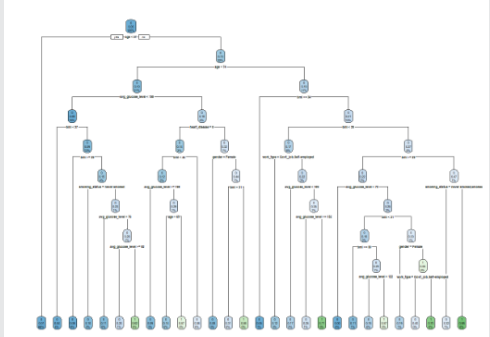

Figure 8 Decision Tree Model

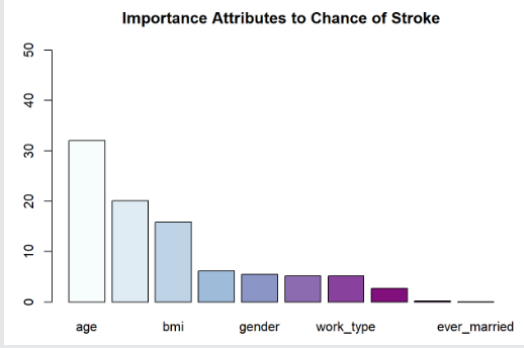

Figure 9 Feature Importance

Belinda Mutiara

BINUS Students 5th Semester

belinda.mutiara@binus.ac.id

phone: 0812 8308 7870
line: belinda.mutiara
linkedin: www.linkedin.com/in/belinda-mutiara

Portfolio Submission for

BINUS Internship Track 2024