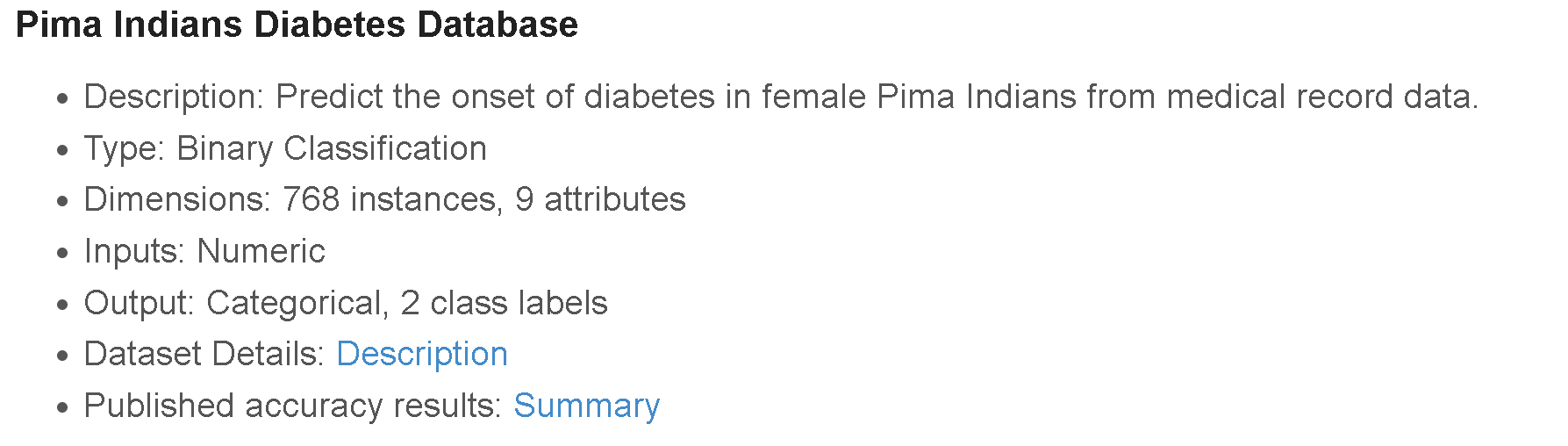
精算與大數據專題 (期末練習)

install.packages("mlbench")

library(mlbench)

library(help = "mlbench")

We will use ‘Pima Indians Diabetes Database’



data(PimaIndiansDiabetes)

dim(PimaIndiansDiabetes)

levels(PimaIndiansDiabetes$diabetes)

head(PimaIndiansDiabetes)

#Training/Prediction

set.seed(1)

index = createDataPartition(PimaIndiansDiabetes$diabetes, p = .8)

traindata = PimaIndiansDiabetes[index[[1]],]

preddata = PimaIndiansDiabetes[-index[[1]],]

Now, Complete the table and handout the result.

* Note all the evaluation index are calculated based on prediction dataset.
* The final model is the best tuned model.

**I. Without Data preprocessing**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Accuracy | Precision | Recall |
| CART |  |  |  |
| XgBoost |  |  |  |
| CatBoost |  |  |  |
| Random Forest |  |  |  |

**II. With Data preprocessing (Standardized + Remove highly correlated variables)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Accuracy | Precision | Recall |
| CART |  |  |  |
| XgBoost |  |  |  |
| CatBoost |  |  |  |
| Random Forest |  |  |  |

請同學們交上以上的表格 & Coding，貼在word檔後繳交。