Name: Shashish Shoda

Roll number: m2021anlt032

Title: **Default of credit card clients**.

Data-Set Information:

This dataset contains information on default payments, credit data, history of payments, and bill statements of credit card clients in Taiwan from April 2005 to September 2005.

Attribute Information:

1. ID: ID of each client
2. LIMIT_BAL: Amount of given credit in NT dollars (includes individual and family/supplementary credit
3. SEX: Gender (1=male, 2=female)
4. EDUCATION: (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
5. MARRIAGE: Marital status (1=married, 2=single, 3=others)
6. AGE: Age in years
7. PAY_0: Repayment status in September, 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months, … 8=payment delay for eight months, 9=payment delay for nine months and above)
8. PAY_2: Repayment status in August, 2005 (scale same as above)
9. PAY_3: Repayment status in July, 2005 (scale same as above)
10. PAY_4: Repayment status in June, 2005 (scale same as above)
11. PAY_5: Repayment status in May, 2005 (scale same as above)
12. PAY_6: Repayment status in April, 2005 (scale same as above)
13. BILL_AMT1: Amount of bill statement in September, 2005 (NT dollar)
14. BILL_AMT2: Amount of bill statement in August, 2005 (NT dollar)
15. BILL_AMT3: Amount of bill statement in July, 2005 (NT dollar)
16. BILL_AMT4: Amount of bill statement in June, 2005 (NT dollar)
17. BILL_AMT5: Amount of bill statement in May, 2005 (NT dollar)
18. BILL_AMT6: Amount of bill statement in April, 2005 (NT dollar)
19. PAY_AMT1: Amount of previous payment in September, 2005 (NT dollar)
20. PAY_AMT2: Amount of previous payment in August, 2005 (NT dollar)
21. PAY_AMT3: Amount of previous payment in July, 2005 (NT dollar)
22. PAY_AMT4: Amount of previous payment in June, 2005 (NT dollar)

23. PAY_AMT5: Amount of previous payment in May, 2005 (NT dollar)
24. PAY_AMT6: Amount of previous payment in April, 2005 (NT dollar)
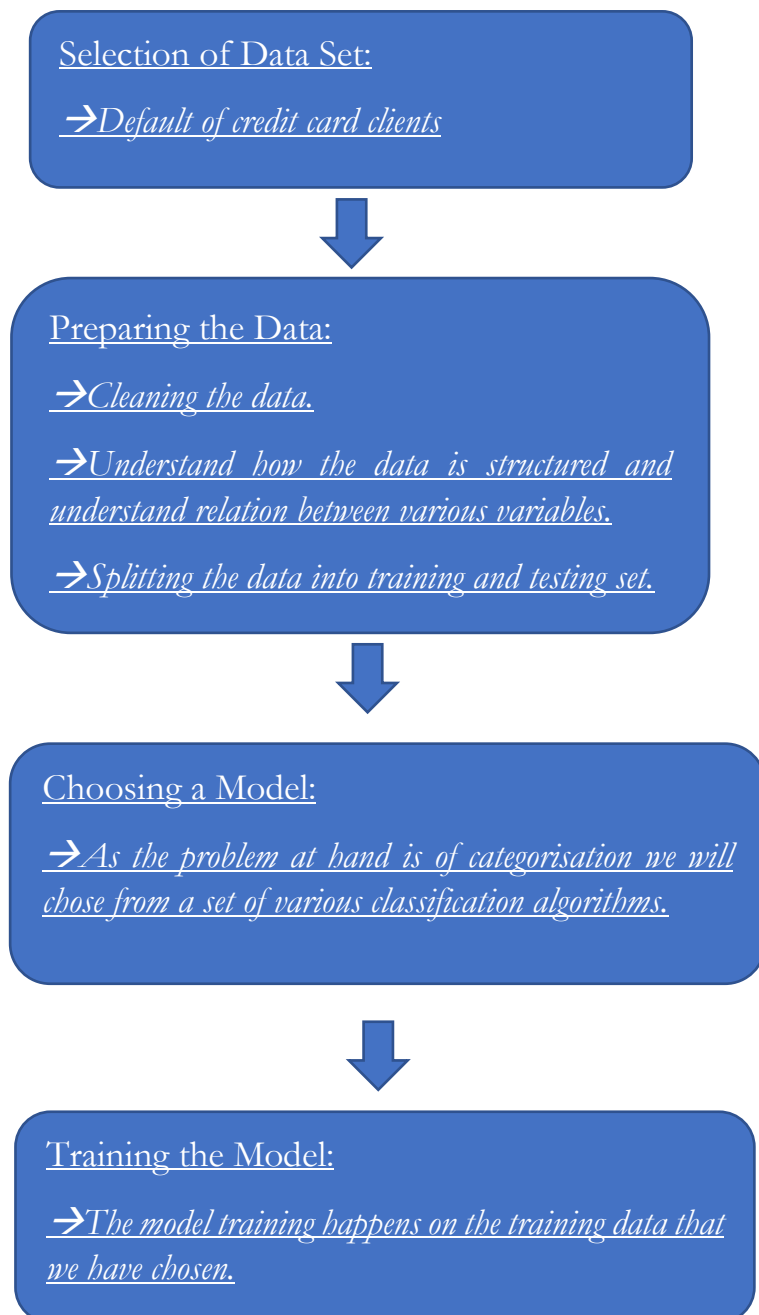25. default payment next month: Default payment (1=yes, 0=no)

Data Set Characteristics: → Multivariate

Attribute Characteristics: → Integer, Real

Associated Tasks: → Classification

Number of Instances: → 30000

Analytical Frame:

Selection of Data Set:

→*Default of credit card clients*

Preparing the Data:

→*Cleaning the data.*

→*Understand how the data is structured and understand relation between various variables.*

→*Splitting the data into training and testing set.*

Choosing a Model:

→*As the problem at hand is of categorisation we will chose from a set of various classification algorithms.*

Training the Model:

→*The model training happens on the training data that we have chosen.*

Evaluating the Model:

→*The model evaluation happens on the test data that we have chosen.*

Tuning the parameters:

→*Tuning of parameters is done to check if the accuracy of the given model can be improved in any way .*

Making prediction:

→*Using our model on unseen data to check prediction accuracy.*

Final model selection:

→*We will choose the model that gives us the best prediction accuracy.*

References:

Lichman, M. (2013). UCI Machine Learning Repository [https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients]. Irvine, CA: University of California, School of Information and Computer Science.

Source:

Name: I-Cheng Yeh
email addresses: (1) icyeh '@' chu.edu.tw (2) 140910 '@' mail.tku.edu.tw
institutions: (1) Department of Information Management, Chung Hua University, Taiwan. (2) Department of Civil Engineering, Tamkang University, Taiwan.
other contact information: 886-2-26215656 ext. 3181