American Journal of Data Mining and Knowledge Discovery
2018; 3(1): 1-12
http://www.sciencepublishinggroup.com/j/ajdmkd
doi: 10.11648/j.ajdmkd.20180301.11

Naïve Bayesian Classifier and Classification Trees for the Predictive Accuracy of Probability of Default Credit Card Clients

NH Niloy*, MAI Navid

Department of Science, Ruhea College, Rangpur, Bangladesh

Email address:
niloynh1997@gmail.com (NH N.)
*Corresponding author

To cite this article:
NH Niloy, MAI Navid. Naïve Bayesian Classifier and Classification Trees for the Predictive Accuracy of Probability of Default Credit Card Clients. *American Journal of Data Mining and Knowledge Discovery*. Vol. 3, No. 1, 2018, pp. 1-12. doi: 10.11648/j.ajdmkd.20180301.11

Received: October 17, 2017; **Accepted**: November 7, 2017; **Published**: January 10, 2018

Abstract: Decision Trees use a decision support tool that utilizes tree like graph model and make decisions. Naïve Bayesian classifier is a binary classifier to get yes/no from the data and it is a very primitive method of finding true or false classification from a dataset. Both algorithms can be used as a predictive model in machine learning and data-mining. Here, a comparative analysis between these two machine learning algorithms is done. The data we have is used to classify if the client is the default credit card holder or not. In the perspective of risk management, the result can be used to accurately get the result of classifying credible or non-credible clients.

Keywords: Machine Learning, Naïve Bayesian Classifier, Decision Trees, Predictive Model

1. Introduction

Many statistical methods, including discriminant analysis, logistic regression, Bayes classifier, and nearest neighbor, have been used to develop models of risk prediction [1]. With the evolution of artificial intelligence and machine learning, artificial neural network and classification trees were also employed to forecast credit risk [2], Credit risk here means the probability of a delay in the repayment of the credit granted. At the same time, most cardholders, irrespective of their repayment ability, overused credit card for consumption and accumulated heavy credit and cash-card debts, The crisis cause the below to consumer finance confidence and it is a big challenge for both banks and cardholder. In a well-developed financial system, crisis management is on the downstream and risk prediction is on the upstream. The major purpose of risk prediction is to use financial information, such as business financial statement, customer transaction and repayment records, etc., to predict business performance or individual customer's credit risk and to reduce the damage and uncertainty. From the perspective of risk control, estimating the probability of default will be more meaningful than classifying customers into the binary results – risky and non-risky. Therefore, whether or not the estimated probability of default produced from data mining methods can represent the real probability of default is an important problem. To forecast probability of default is a challenge facing practitioners and researchers, and it needs more study [1, 3-5].

2. Literature Review

Data mining techniques

Right now, data mining is an indispensable tool in decision support system and plays a key role in market segmentation, customer services, fraud detection, credit and behavior scoring, and benchmarking. In the era of information explosion, individual companies will produce and collet huge volume of data every day. Discovering useful knowledge from the database and transforming information into actionable results is a major challenge facing companies. Data mining is the process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules [6]. The pros and cons of the naïve Bayesian classifier and classification trees employed in our study are reviewed as follows [7-10].

Naïve Bayesian classifier (NB)

The naïve Bayesian classifier is based on Bayes theory and assumes that the effect of an attribute value on a given class is independent of the other attributes. This assumption is called class conditional independence. This assumption is called conditional independence. Bayesian classifiers are useful is that they provide a theoretical justification for other classifiers that do not explicitly use Bayes theorem. The major weakness of NB is that the predictive accuracy is highly correlated with the assumption of class conditional independence. This assumption simplifies computation. In practice, however, dependences can exist between variables.

Classification trees (CTs)

The top-most node in a tree is the root node. In a classification tree structure, each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes. CTs are applied when the response variable is qualitative or quantitative discrete. Classification trees perform a classification of the observation on the basis of all explanatory variables and supervise by the presence of the response variables. The segmentation process is typically carried out using only one explanatory variable at a time. CTs are based on minimizing impurity, which refers to a measure of variability of the response values of the observations. CTs can result in simple classification rules and can handle the nonlinear and interactive effects of explanatory variables. It is difficult to take a tree structure designed for one context and generalize it for other contexts. However, their sequential nature and algorithmic complexity can make them depends on the observed data, and even a small change might alter the structure of the tree.

3. Related Works

Credit scoring is the term used to describe formal statistical methods which are used for classifying applicants for credit into "good" and "bad" risk classes [1]. Such methods have become increasingly important with the dramatic growth in consumer credit in recent years. A wide range of statistical methods has been applied, though the literature available to the public is limited for reasons of commercial confidentiality. Many static and dynamic models have been used to assist decision making in the area of consumer and commercial credit. The decision of interest includes whether to extend credit, how much credit to extend, when collections of delinquent accounts should be initiated, and what action should be taken. They surveyed the use of discriminant analysis, classification trees, and expert systems for static decisions, and dynamic programming, linear programming and Markov chains for dynamic decision models. Bayesian methods, coupled with Markov Chain Monte Carlo computational techniques, could be successfully employed in the analysis of highly dimensional complex dataset, such as those in credit scoring and benchmarking. Paolo employs conditional independence graphs to localize model specification and inferences, thus allowing a considerable gain in flexibility of modeling and efficiency of the computations. It was found that, based on eight real-life credit scoring data sets, both the LS-SVM and neural network classifiers yield a very good performance, but also simple classifiers such as logistic regression and liner discriminant analysis perform very well for credit scoring [4]. It was explored the performance of credit scoring by integrating the back propagation neural networks with the traditional discriminant analysis approach [11]. The proposed hybrid approach converges much faster than the conventional neural networks model. Moreover, the credit scoring accuracy increases in terms of the proposed methodology and the hybrid approach outperforms traditional discriminant analysis and logistic regression.

4. Our Works

4.1. Attribute Information

This research employed a binary variable, default payment (Yes = 1, No = 0), as the response variable.

This study reviewed the literature and used the following 23 variables as explanatory variables:

LIMIT_BAL: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.

SEX: Gender (1 = male; 2 = female). EDUCATION: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others). MARRIAGE: Marital status (1 = married; 2 = single; 3 = others). AGE: Age (year). (X6-X11) PAY_0 - PAY_6: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows:

X6 = the repayment status in September, 2005;

X7 = the repayment status in August, 2005; . . .;

X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -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.

(X12-X17)

BILL_AMT1 - BILL_AMT6**:** Amount of bill statement (NT dollar).

X12 = amount of bill statement in September, 2005;

X13 = amount of bill statement in August, 2005; . . .;

X17 = amount of bill statement in April, 2005.

(X18-X23)

PAY_AMT1 - PAY_AMT6**:** Amount of previous payment (NT dollar).

X18 = amount paid in September, 2005;

X19 = amount paid in August, 2005;

X23 = amount paid in April, 2005.

The dataset contains 30,000 observations and has 23 variables.

```
## 'data.frame':    30000 obs. of  25 variables:
##  $ ID                        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ LIMIT_BAL                 : int  20000 120000 90000 50000 50000 50000 500000 100000 140000 20000 ...
##  $ SEX                       : int  2 2 2 2 1 1 1 2 2 1 ...
##  $ EDUCATION                 : int  2 2 2 2 2 1 1 2 3 3 ...
##  $ MARRIAGE                  : int  1 2 2 1 1 2 2 2 1 2 ...
##  $ AGE                       : int  24 26 34 37 57 37 29 23 28 35 ...
##  $ PAY_0                     : int  2 -1 0 0 -1 0 0 0 0 -2 ...
##  $ PAY_2                     : int  2 2 0 0 0 0 0 -1 0 -2 ...
##  $ PAY_3                     : int  -1 0 0 0 -1 0 0 -1 2 -2 ...
##  $ PAY_4                     : int  -1 0 0 0 0 0 0 0 0 -2 ...
##  $ PAY_5                     : int  -2 0 0 0 0 0 0 0 0 -1 ...
##  $ PAY_6                     : int  -2 2 0 0 0 0 0 -1 0 -1 ...
##  $ BILL_AMT1                 : int  3913 2682 29239 46990 8617 64400 367965 11876 11285 0 ...
##  $ BILL_AMT2                 : int  3102 1725 14027 48233 5670 57069 412023 380 14096 0 ...
##  $ BILL_AMT3                 : int  689 2682 13559 49291 35835 57608 445007 601 12108 0 ...
##  $ BILL_AMT4                 : int  0 3272 14331 28314 20940 19394 542653 221 12211 0 ...
##  $ BILL_AMT5                 : int  0 3455 14948 28959 19146 19619 483003 -159 11793 13007 ...
##  $ BILL_AMT6                 : int  0 3261 15549 29547 19131 20024 473944 567 3719 13912 ...
##  $ PAY_AMT1                  : int  0 0 1518 2000 2000 2500 55000 380 3329 0 ...
##  $ PAY_AMT2                  : int  689 1000 1500 2019 36681 1815 40000 601 0 0 ...
##  $ PAY_AMT3                  : int  0 1000 1000 1200 10000 657 38000 0 432 0 ...
##  $ PAY_AMT4                  : int  0 1000 1000 1100 9000 1000 20239 581 1000 13007 ...
##  $ PAY_AMT5                  : int  0 0 1000 1069 689 1000 13750 1687 1000 1122 ...
##  $ PAY_AMT6                  : int  0 2000 5000 1000 679 800 13770 1542 1000 0 ...
##  $ default.payment.next.month: int  1 1 0 0 0 0 0 0 0 0 ...
```

4.2. Experiment the Data

1. Fix what appears to be a typo in the field header PAY_0

```
names(credit_data_df)[names(credit_data_df) == "PAY_0"] <- "PAY_1"
```

2. Change codes to values for sex, education, and marriage. Any observations associated with undocumented code values will be removed.

[1] "Female" "Male"
[1] University Graduate School High School
Levels: High School University Graduate School
[1] "Married" "Single"

3. Rename the columns in order to use tidyr to convert the dataset from wide to long format.

```
##  [1] "ID"                       "LIMIT_BAL"
##  [3] "SEX"                      "EDUCATION"
##  [5] "MARRIAGE"                 "AGE"
##  [7] "PAY_1"                    "PAY_2"
##  [9] "PAY_3"                    "PAY_4"
## [11] "PAY_5"                    "PAY_6"
## [13] "BILLAMT_1"                "BILLAMT_2"
## [15] "BILLAMT_3"                "BILLAMT_4"
## [17] "BILLAMT_5"                "BILLAMT_6"
## [19] "PAYAMT_1"                 "PAYAMT_2"
## [21] "PAYAMT_3"                 "PAYAMT_4"
## [23] "PAYAMT_5"                 "PAYAMT_6"
## [25] "DEFAULT_NEXT_PAYMENT_IND"
```

4. Convert from wide to long using tidyr

```
tidy_credit_data_df <- credit_data_df_s %>%
                       gather(col, "VAL", -ID:-AGE, -DEFAULT_NEXT_PAYMENT_IND) %>%
                       separate(col, c("L1", "MTH"), sep="_") %>%
                       spread(L1, VAL) %>%
                       select(ID:MTH, BILLAMT, PAYAMT, PAY)
```

The resulting dataset looks like this:

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	DEFAULT_NEXT_PAYMENT_IND	MTH	BILLAMT	PAYAMT	PAY
1	20000	Female	University	Married	24	1	1	3913	0	2
2	120000	Female	University	Single	26	1	1	2682	0	-1
3	90000	Female	University	Single	34	0	1	29239	1518	0
4	50000	Female	University	Married	37	0	1	46990	2000	0
5	50000	Male	University	Married	57	0	1	8617	2000	-1
6	50000	Male	Graduate School	Single	37	0	1	64400	2500	0
7	500000	Male	Graduate School	Single	29	0	1	367965	55000	0
8	100000	Female	University	Single	23	0	1	11876	380	0
9	140000	Female	High School	Married	28	0	1	11285	3329	0
10	20000	Male	High School	Single	35	0	1	0	0	-2
11	200000	Female	High School	Single	34	0	1	11073	2306	0
12	260000	Female	Graduate School	Single	51	0	1	12261	21818	-1
13	630000	Female	University	Single	41	0	1	12137	1000	-1
14	70000	Male	University	Single	30	1	1	65802	3200	1
15	250000	Male	Graduate School	Single	29	0	1	70887	3000	0
17	20000	Male	Graduate School	Single	24	1	1	15376	3200	0
18	320000	Male	Graduate School	Married	49	0	1	253286	10358	0

5. With the dataset in long format, create some derived fields:

Field Name	Description
AMT_OWED	Running or cumulative sum of bill amount - payment amount for each individual
AVG_6MTH_BAL	Mean value of AMT_OWED over a 6 month period
MISSED_PAYMENTS	Maxium number of missed payments recorded for the individual
BALANCE_TO_LIMIT_RATIO	Average 6 month balance divided by the individual's credit limit; note anything <= .3 is considered good
AGE_RANGE	Groups individuals into 10 year age grouping [20-29] through [70-79]

The resulting dataset looks like this: s:

ID	CREDIT_LIMIT	SEX	EDUCATION	MARITAL_STATUS	DEFAULT_NEXT_PAYMENT_IND	AVG_6MTH_BAL	MISSED_PAYMENTS	BALANCE_TO_LIMIT_RATIO	AGE_RANGE
1	20000	Female	University	Married	1	6383.167	2	0.319	20-29
2	120000	Female	University	Single	1	6905.333	2	0.058	20-29
3	90000	Female	University	Single	0	59605.833	0	0.662	30-39
4	50000	Female	University	Married	0	143223.833	0	2.864	30-39
5	50000	Male	University	Married	0	13195.500	0	0.264	50-59
6	50000	Male	Graduate School	Single	0	164519.667	0	3.290	30-39
7	500000	Male	Graduate School	Single	0	1388642.500	0	2.777	20-29
8	100000	Female	University	Single	0	10754.667	0	0.108	20-29
9	140000	Female	High School	Married	0	37143.000	2	0.265	20-29
10	20000	Male	High School	Single	0	-223.167	-1	-0.011	30-39

6. Using the dataset just created and stored in credit_data_individual, create an aggregate dataset for the different group combinations of Sex, Age Range, Marital Status, and Education.

Aggregation Level	Description
LVL1	Top Level - Aggregates to Age Range
LVL2	Aggregates to Age Range and Sex
LVL3	Aggregates to Age Range, Sex, Marital Satus
LVL4	Bottom Level - Aggregates to Age Range, Sex, Marital Satus, Education

Visualize the groups using the data. tree package:
Field Definition:

Aggregation Level	Description
n	Count of individuals in the groups
Avg Limit	Average Credit Limit of the group
% Good Credit	Percentage representation of how many individuals in the group have a <= .3 limit-to-balance ratio
% Default	Percentage representation of how many individuals in the group are predicted to default

Level 1: Summarized to Age Range

```
##   levelName      n Avg Limit % Good Credit % Default
## 1 All           NA                       NA        NA
## 2  ¦--20-29   9435  $124,157           0.32      0.23
## 3  ¦--30-39  10981  $197,420           0.46      0.21
## 4  ¦--40-49   6197  $182,750           0.42      0.23
## 5  ¦--50-59   2222  $165,729           0.37      0.25
## 6  ¦--60-69    303  $186,832           0.42      0.29
## 7  °--70-79     25  $218,800           0.40      0.28
```

Level 2: Summarized to Age Range and Sex

```
##            levelName    n Avg Limit % Good Credit % Default
## 1  All                 NA                       NA        NA
## 2   ¦--20-29           NA                       NA        NA
## 3   ¦   ¦--Female 6219  $131,269           0.34      0.22
## 4   ¦   °--Male   3216  $110,404           0.28      0.24
## 5   ¦--30-39           NA                       NA        NA
## 6   ¦   ¦--Female 6501  $203,130           0.50      0.19
## 7   ¦   °--Male   4480  $189,134           0.39      0.23
## 8   ¦--40-49           NA                       NA        NA
## 9   ¦   ¦--Female 3534  $184,020           0.47      0.22
## 10  ¦   °--Male   2663  $181,066           0.37      0.25
## 11  ¦--50-59           NA                       NA        NA
## 12  ¦   ¦--Female 1181  $159,670           0.40      0.23
## 13  ¦   °--Male   1041  $172,603           0.33      0.27
## 14  ¦--60-69           NA                       NA        NA
## 15  ¦   ¦--Female  141  $160,355           0.43      0.31
## 16  ¦   °--Male    162  $209,877           0.41      0.27
## 17  °--70-79           NA                       NA        NA
## 18      ¦--Female   12  $213,333           0.50      0.25
## 19      °--Male     13  $223,846           0.31      0.31
```

Level 3: Summarized to Age Range, Sex, and Marital Status

```
##                 levelName    n Avg Limit % Good Credit % Default
## 1  All                      NA                      NA        NA
## 2   ¦--20-29                NA                      NA        NA
## 3   ¦   ¦--Female           NA                      NA        NA
## 4   ¦   ¦   ¦--Married 1154  $122,626             0.27      0.27
## 5   ¦   ¦   °--Single  5065  $133,238             0.36      0.22
## 6   ¦   °--Male             NA                      NA        NA
## 7   ¦       ¦--Married  295  $117,085             0.26      0.27
## 8   ¦       °--Single  2921  $109,730             0.29      0.24
## 9   ¦--30-39                NA                      NA        NA
## 10  ¦   ¦--Female           NA                      NA        NA
## 11  ¦   ¦   ¦--Married 3510  $191,148             0.47      0.20
## 12  ¦   ¦   °--Single  2991  $217,192             0.53      0.17
## 13  ¦   °--Male             NA                      NA        NA
## 14  ¦       ¦--Married 1916  $196,200             0.42      0.26
## 15  ¦       °--Single  2564  $183,853             0.37      0.21
## 16  ¦--40-49                NA                      NA        NA
## 17  ¦   ¦--Female           NA                      NA        NA
## 18  ¦   ¦   ¦--Married 2645  $184,021             0.48      0.22
## 19  ¦   ¦   °--Single   889  $184,016             0.42      0.21
## 20  ¦   °--Male             NA                      NA        NA
## 21  ¦       ¦--Married 1931  $195,831             0.41      0.25
## 22  ¦       °--Single   732  $142,114             0.25      0.24
## 23  ¦--50-59                NA                      NA        NA
## 24  ¦   ¦--Female           NA                      NA        NA
## 25  ¦   ¦   ¦--Married  880  $165,250             0.42      0.25
## 26  ¦   ¦   °--Single   301  $143,355             0.36      0.19
## 27  ¦   °--Male             NA                      NA        NA
## 28  ¦       ¦--Married  821  $188,197             0.37      0.28
```

Level 4: Summarized to Age Range, Sex, Marital Status, and Education

```
##                         levelName    n Avg Limit % Good Credit % Default
## 1  All                              NA                      NA        NA
## 2   ¦--20-29                        NA                      NA        NA
## 3   ¦   ¦--Female                   NA                      NA        NA
## 4   ¦   ¦   ¦--Married              NA                      NA        NA
## 5   ¦   ¦   ¦   ¦--High School     192  $ 99,688             0.21      0.29
## 6   ¦   ¦   ¦   ¦--University      814  $118,206             0.25      0.28
## 7   ¦   ¦   ¦   °--Graduate School 148  $176,689             0.45      0.16
## 8   ¦   ¦   °--Single               NA                      NA        NA
## 9   ¦   ¦       ¦--High School     365  $110,219             0.31      0.24
## 10  ¦   ¦       ¦--University     2337  $119,718             0.28      0.24
## 11  ¦   ¦       °--Graduate School 2363  $150,165            0.45      0.19
## 12  ¦   °--Male                     NA                      NA        NA
## 13  ¦       ¦--Married              NA                      NA        NA
## 14  ¦       ¦   ¦--High School      53  $103,962             0.23      0.32
## 15  ¦       ¦   ¦--University      197  $110,457             0.25      0.26
## 16  ¦       ¦   °--Graduate School  45  $161,556             0.36      0.27
## 17  ¦       °--Single               NA                      NA        NA
## 18  ¦           ¦--High School     325  $ 70,062             0.22      0.30
## 19  ¦           ¦--University     1462  $ 91,826             0.20      0.27
## 20  ¦           °--Graduate School 1134  $144,180            0.41      0.19
## 21  ¦--30-39                        NA                      NA        NA
## 22  ¦   ¦--Female                   NA                      NA        NA
## 23  ¦   ¦   ¦--Married              NA                      NA        NA
## 24  ¦   ¦   ¦   ¦--High School     500  $141,260             0.34      0.22
## 25  ¦   ¦   ¦   ¦--University     1948  $173,850             0.40      0.21
## 26  ¦   ¦   ¦   °--Graduate School 1062  $246,365            0.67      0.19
## 27  ¦   ¦   °--Single               NA                      NA        NA
```

4.3. Result

Which group has the highest average credit limit?

SEX	MARITAL_STATUS	EDUCATION	AGE_RANGE	FORMATTED_AVG_CREDIT_LIMIT
Male	Married	Graduate School	60-69	$288,600

Which group has the lowest average credit limit?

SEX	MARITAL_STATUS	EDUCATION	AGE_RANGE	FORMATTED_AVG_CREDIT_LIMIT
Male	Single	High School	20-29	$ 70,062

Which group is comprised of highest percentage of people who have a balance-to-limit rating less than or equal to 30%?

SEX	MARITAL_STATUS	EDUCATION	AGE_RANGE	PCT_IN_GROUP_GOOD_CREDIT
Female	Married	Graduate School	50-59	0.73

Which group has the lowest utilization or balance-to-limit rating?

SEX	MARITAL_STATUS	EDUCATION	AGE_RANGE	PCT_IN_GROUP_GOOD_CREDIT
Male	Single	University	60-69	0.15

Which group is the most likely to predicted to default?

SEX	MARITAL_STATUS	EDUCATION	AGE_RANGE	PCT_IN_GROUP_DEFAULT_RATE
Male	Married	University	60-69	0.45

Which group has the highest amount of debt, is the most likely to default, and is the most likely to miss a payment?

SEX	MARITAL_STATUS	EDUCATION	AGE_RANGE
Female	Single	High School	60-69

Which group has the lowest amount of debt, is the least predicted to default, and is not likely to miss a payment?

SEX	MARITAL_STATUS	EDUCATION	AGE_RANGE
Female	Single	Graduate School	30-39

4.4. Result in Graph

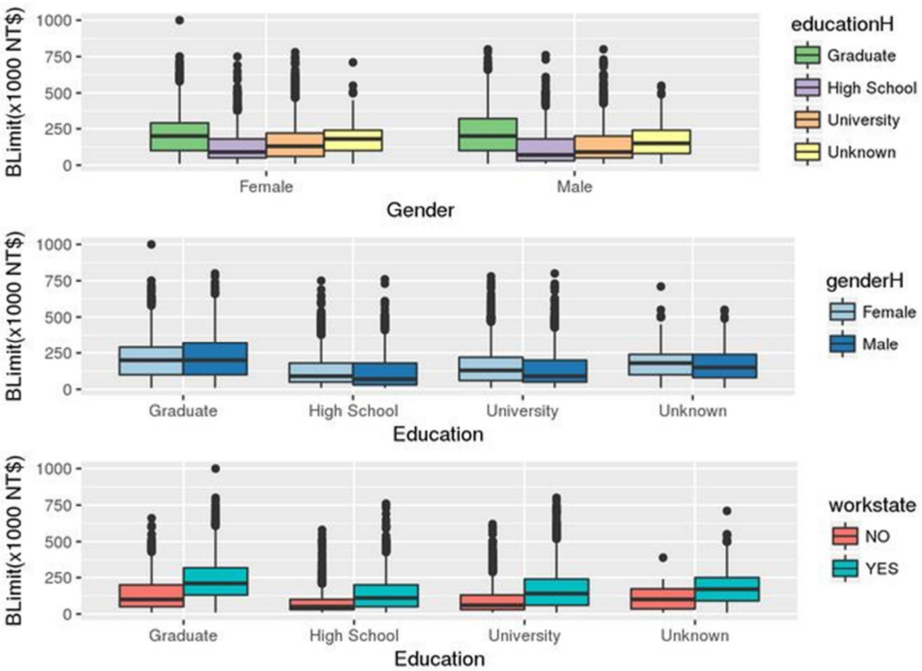

Figure 1. *Balance Limit by Gender, Education, Work State.*

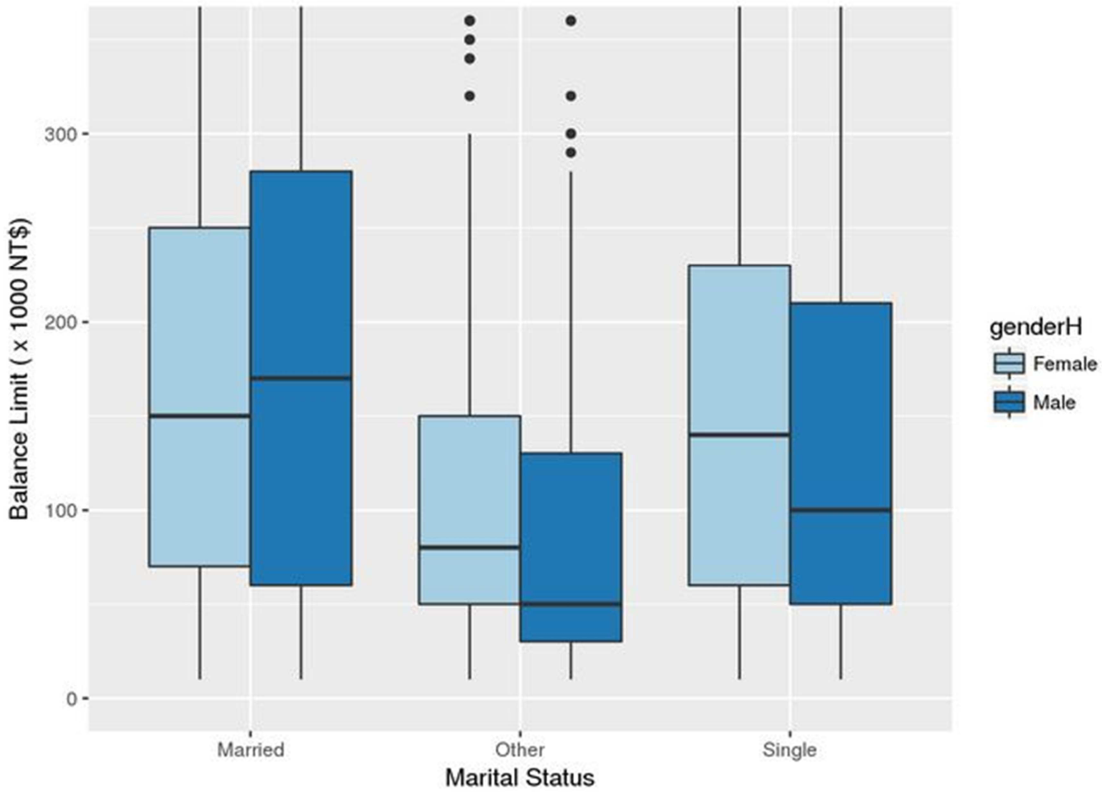

Figure 2. *Relation between Marital Status & Balance Limits by Gender.*

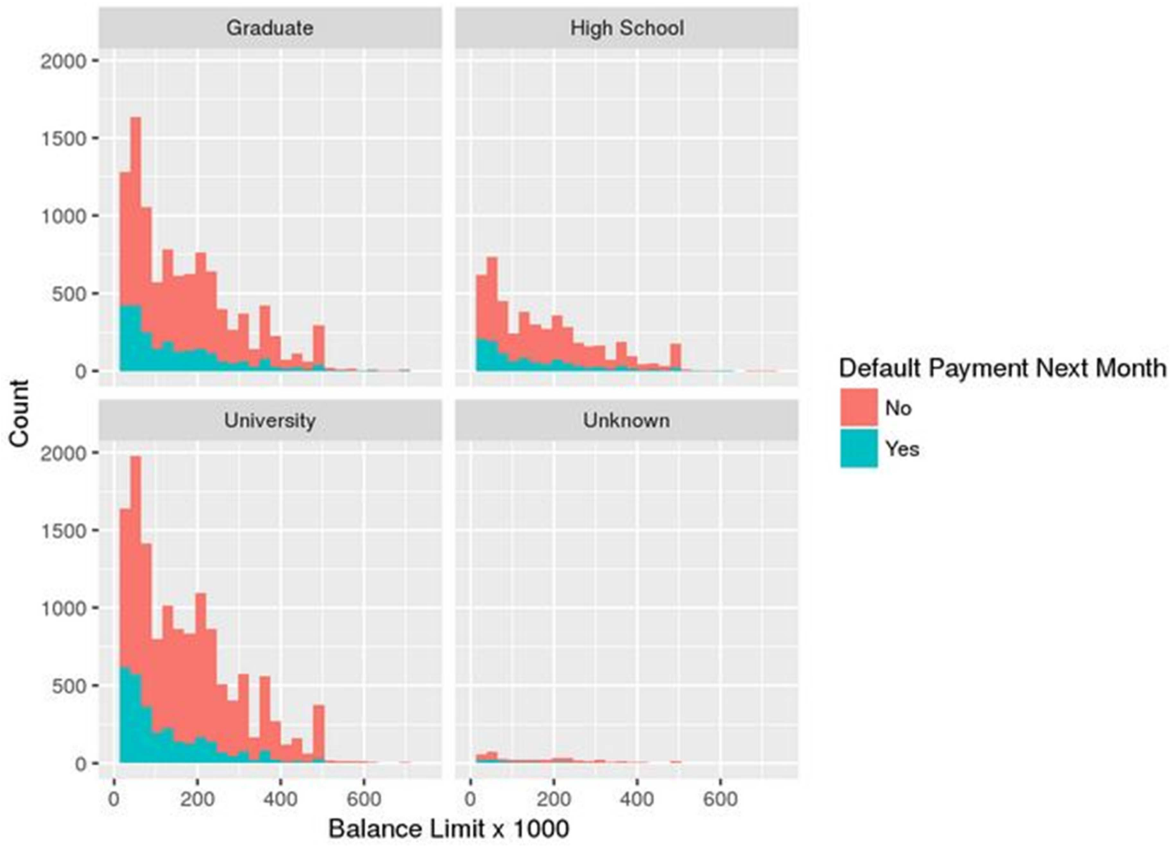

Figure 3. *Histogram of Limit Balance & Default Payment.*

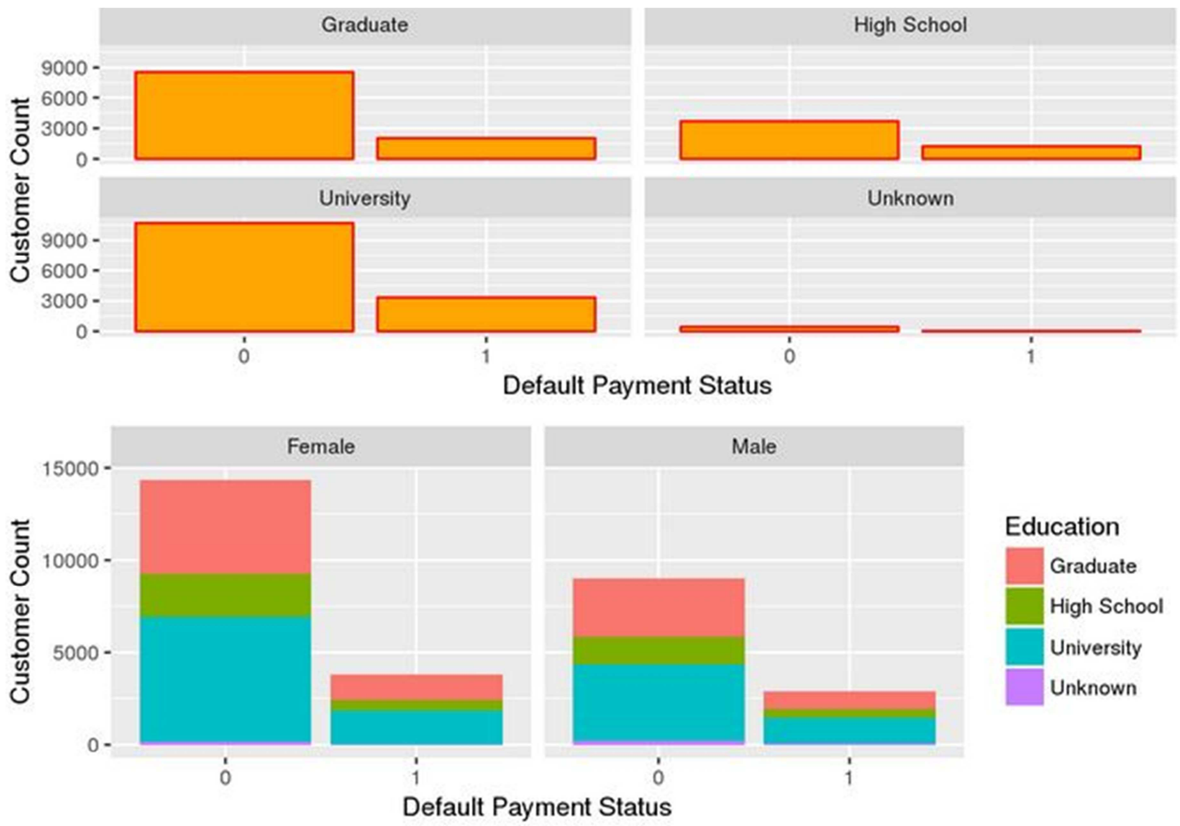

Figure 4. *Relation between Education & Default Payment.*

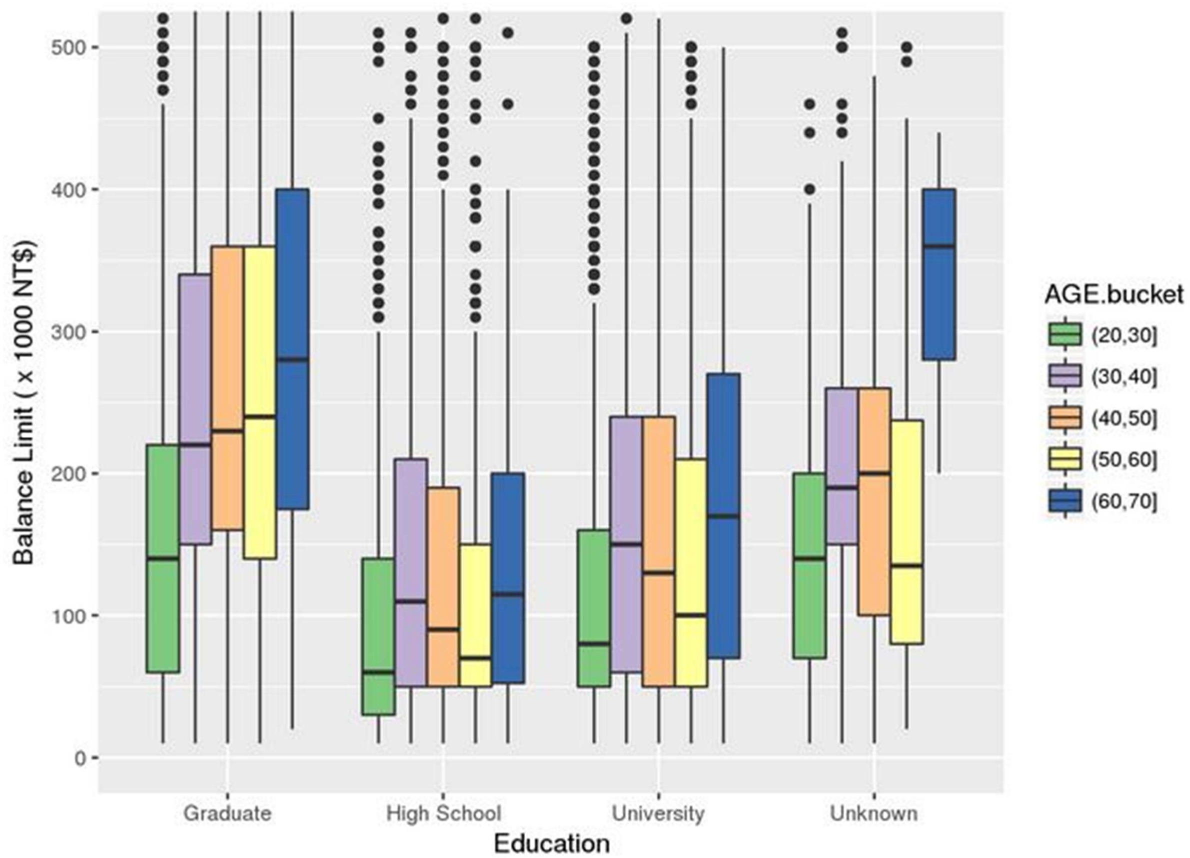

Figure 5. *Balance Limits by Age Groups & Education.*

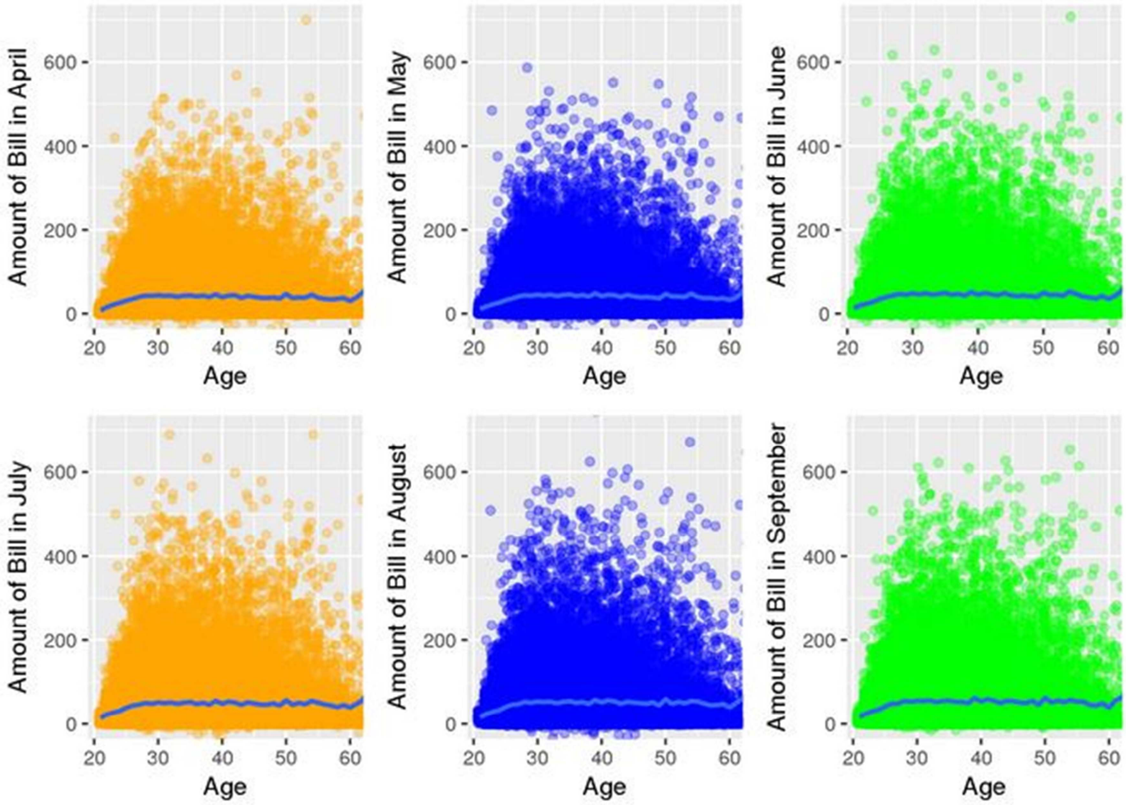

Figure 6. *Expenditure by Months.*

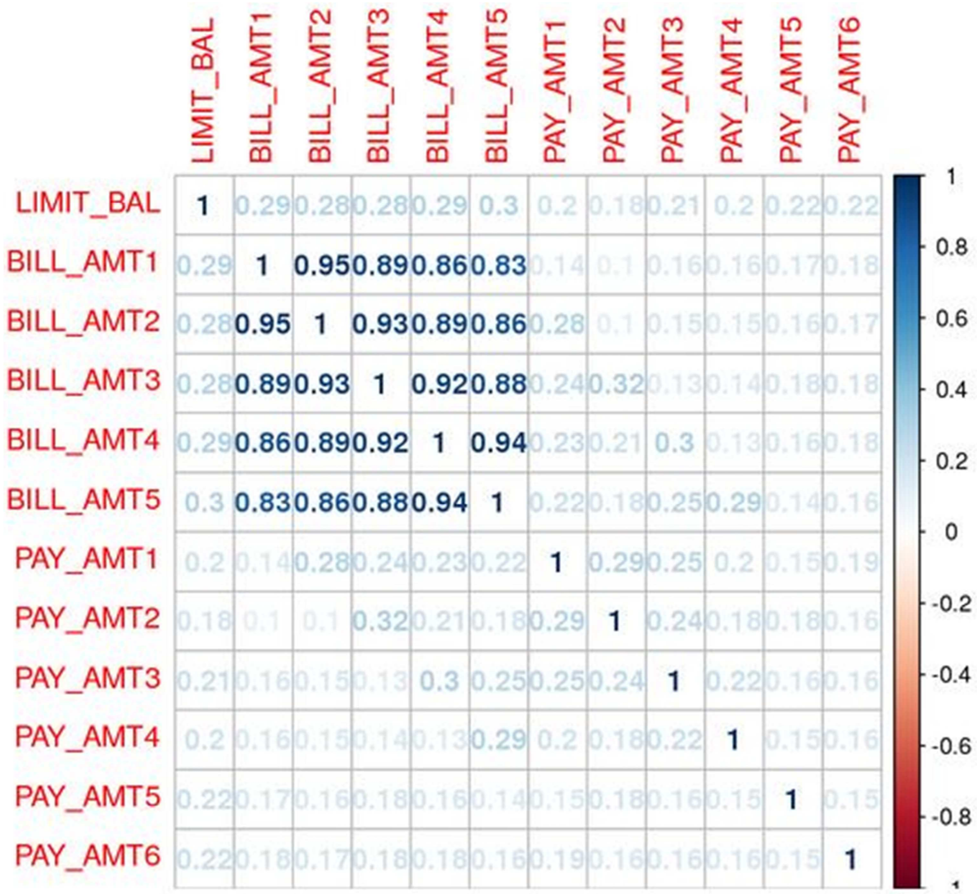

Figure 7. *Correlations between Limit Balance, Bill Amounts & Payments.*

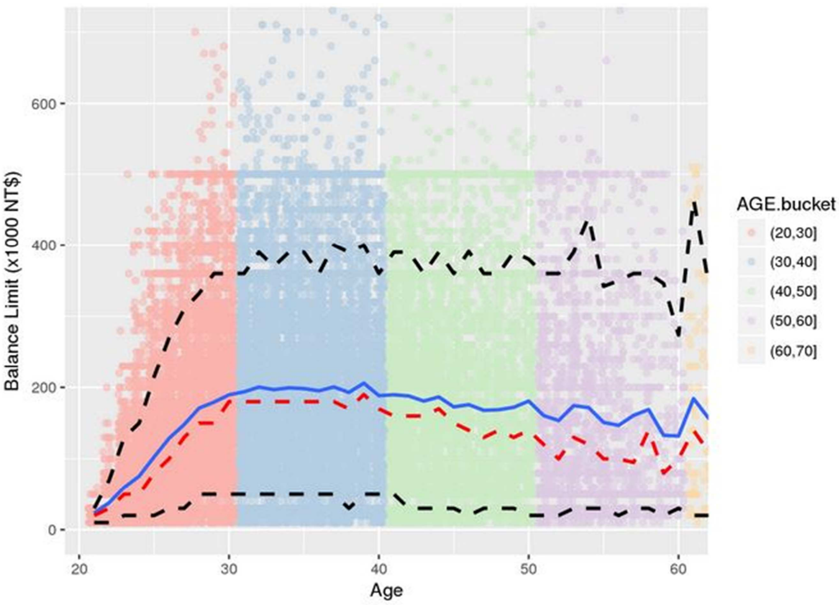

Figure 8. *Personal Balance Limits Probabilities & Given Limits by Age.*

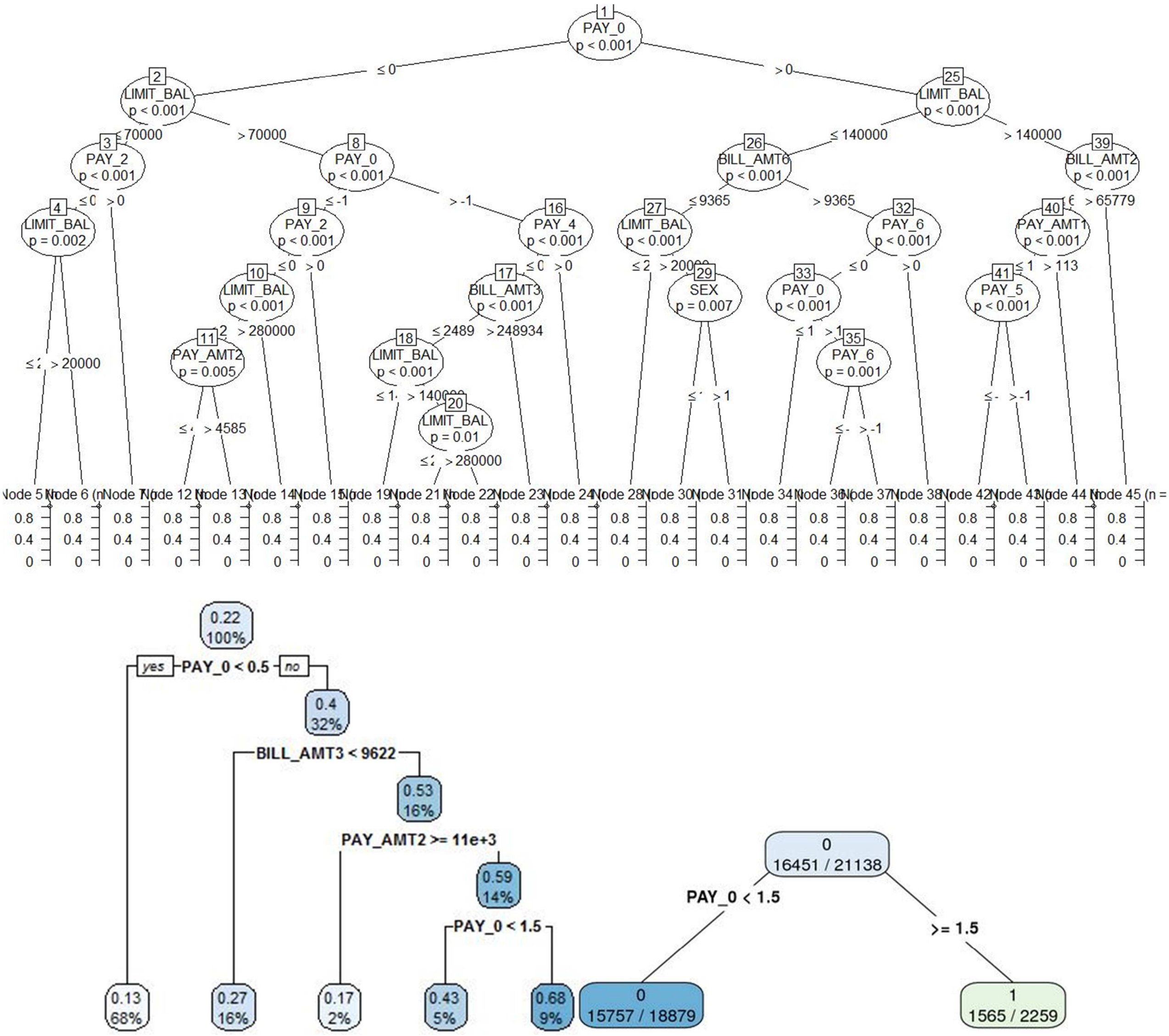

Figure 9. *Decision Tree.*

5. Conclusion

In the classification accuracy between the two data mining techniques, this result shows that there are little differences in error rates between two methods. However, there are relatively big differences in area ratio between two techniques. This paper we examines the two major classification techniques of Naïve Bayesian classifier and Classification tress for the performance of classification predictive accuracy. Naïve Bayesian performs classification more accurately than classification trees. Therefore, it can be concluded that the classifier is most important to measure the classification accuracy of models.

References

[1] Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. Journal of the Royal Statistical Society, Series A – Statistics in Society, 160(3), 523–541.

[2] Koh, H. C., & Chan, K. L. G. (2002). Data mining and customer relationship marketing in the banking industry. Singapore Management Review, 24(2), 1–27.

[3] Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, 49(3), 312–329.

[4] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 54(6), 627–635.

[5] Desai, V. S., Crook, J. N., & Overstreet, G. A. A. (1996). Comparison of neural networks and linear scoring models in the credit union environment. European Journal of Operational Research, 95(1), 24–37.

[6] Berry, M., & Linoff, G. (2000). Mastering data mining: The art and science of customer relationship management. New York: John Wiley & Sons, Inc. Chou, M. (2006). Cash and credit card crisis in Taiwan. Business Weekly, 24–27.

[7] Han, J., & Kamber, M. (2001). Data mining: Concepts and techniques. San Fransisco: Morgan Kaufmann.

[8] Hand, D. J., Mannila, H., & Smyth, P. (2001). Data mining: Practical machine learning tools and techniques. Cambridge: MIT Press.

[9] Paolo, G. (2001). Bayesian data mining, with application to bench marking and credit scoring. Applied Stochastic Models in Business and Society, 17, 69–81.

[10] Witten, I. H., & Frank, E. (1999). Data mining: Practical machine learning tools and techniques with java implementations. San Fransisco: Morgan Kaufman.

[11] Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. Expert Systems with Applications, 23(3), 245–254.