Information Management and Business Review (ISSN 2220-3796)
Vol. 16, No. 3S(a), pp. 366-378, Oct 2024

Personal Bankruptcy Prediction Using Logistic Regression Model

*Sharifah Heryati Syed Nor[1], Shafinar Ismail[2], Yap Bee Wah[3]
[1] Faculty of Business Management, Universiti Teknologi Mara, Puncak Alam Campus, Malaysia
[2] Faculty of Business Management, Universiti Teknologi Mara, Melaka Campus, Malaysia
[3]Unitar International University, Malaysia
*heryati879@uitm.edu.my, shafinar@uitm.edu.my, bee.wah@unitar.my
Corresponding Author: Sharifah Heryati Syed Nor

Abstract: According to the Insolvency Department of Malaysia, as of December 2023, 233,483 Malaysians are currently involved in bankruptcy cases due to their defaults on hire purchase loans, credit card loans, personal loans, housing loans, and business loans. This is indeed a critical issue because the growing number of personal bankruptcy cases will hurt the Malaysian economy as well as society. From an individual's economic perspective, bankruptcy minimizes their chances of getting a job. Apart from that, their accounts will be frozen, they will lose control of their properties and assets, and they will not be allowed to start any business or be a part of any company's Board of Directors. Bankrupts also will be rejected from any loan application. This paper examines this problem by developing a personal bankruptcy prediction model using the logistic regression technique. This paper defines "bankrupt" as terminated members who failed to settle their loans. The sample comprised 24,546 cases with 17% settled cases and 83% terminated cases. The data included a dependent variable, i.e., bankruptcy status (Y=1(bankrupt), Y=0(non-bankrupt)), and 12 predictors. Upon completion, this paper succeeds in coming out with a reliable personal bankruptcy prediction model and significant variables of personal bankruptcy. The findings of this paper are very beneficial and significant to creditors, banks, the Malaysia Department of Insolvency, potential borrowers, members of AKPK, and society in general in raising awareness of personal bankruptcy risks and such information may help them to take preventive measures in minimizing the number of personal bankruptcy cases.

Keywords: *Bankruptcy, business loans, loan Malaysian Economy*

1. Introduction and Background

Personal bankruptcy has been a topic of extensive interest because personal bankruptcy filings are an important indicator of household financial problems nationally (International Monetary Fund, 2017; Garrido et al., 2020). In Malaysia, personal bankruptcy cases have been on an upward trend since 2007. However, from 2014 until 2018, the trend has been up and down with not less than twelve thousand cases per year (Malaysia Department of Insolvency, 2019). In addition, from 2020 to 2022, the trend has decreased, with 7221 cases in 2020, 6554 cases in 2021, and 5695 cases in 2022. (Malaysia Department of Insolvency, 2020,2021,2022). This was due to the amendment of the bankruptcy threshold amount. In Malaysia, a debtor is declared bankrupt, under an Adjudication Order made by the High Court against the debtor if he is unable to pay his debts of at least RM30,000.00 (Malaysia Department of Insolvency, 2017). On 29th of March 2017, the Bankruptcy (Amendment) Bill 2016 was passed in the Dewan Rakyat. Among the major changes to the Bankruptcy Act 1967 were substituting the term ‘bankruptcy’ in all proceedings to ‘insolvency’, the threshold to be declared as bankrupt increased to RM50k debt amount, automatically releasing bankrupts after three years and social guarantors cannot be declared bankrupt. Again in 2020, the Parliament sought to amend the Bankruptcy Act 1967 by increasing the minimum debt threshold for the presentation of a bankruptcy petition, from RM50,000 to RM100,000 (Malaysia Department of Insolvency, 2020)

As of December 2023, the Malaysia Department of Insolvency reported that the total accumulated number of bankrupt individuals was 233,483. These bankruptcy cases are mainly due to defaults in personal loans, business loans, housing loans, hire purchases, and credit cards (Malaysia Department of Insolvency, 2023). This is alarming because if the number of personal bankruptcy cases continues to increase, it will hurt the Malaysian economy and society. From the aspect of individuals’ economy, bankruptcy minimizes their chances of securing a job.

As one of the efforts taken to curb the increasing household debt which mainly leads to personal bankruptcy, Bank Negara Malaysia has set up a debt management agency, Agensi Kaunseling and Pengurusan Kredit. This agency is an avenue for potential individual borrowers and distressed borrowers to acquire assistance and seek advice in managing their debts and finances. Data used in this paper were provided by this agency. Thus, this paper illustrates the application of data mining techniques to determine the conditional probability of a borrower belonging to a class (bankrupt or non-bankrupt) using the logistic regression technique. The findings from this paper are useful for various parties to make decisions and take action, such as the management level of insolvency departments, banks, debt management agencies, hire-purchase companies, and credit companies. These actions are important to avoid or prevent default payments, bad debts, and personal bankruptcy. Therefore, the objectives of this paper are to identify the significant variables and to determine the conditional probability of a borrower belonging to a class (bankrupt or non-bankrupt) using the logistic regression technique.

2. Literature Review

Bankruptcy is a legal declaration of financial insolvency for an individual or firm that allows certain legal protections to both debtors and creditors (Cohen, 2015). It is crucial to have a better understanding of the causes of personal bankruptcy to provide practical solutions to this financial problem. Korol (2022) mentioned that forecasting consumer bankruptcy risk has received increasing scientific and public attention. In addition, Brygala and Korol (2024), indicated that it has become important to have an early prediction model that provides accurate assurance for users about the financial situation of consumers.

According to Papana and Spyridou (2020), research on bankruptcy prediction is of the utmost importance as it aims to build statistical models that can distinguish healthy firms from financially distressed ones. Meanwhile, Smiti et. Al., (2024) added that bankruptcy prediction is considered one of the most important research topics in the field of finance and accounting. The rapid increase of data science, artificial intelligence, and machine learning has led researchers to build an accurate bankruptcy prediction model.

Because it is critical to develop an effective bankruptcy prediction model, a few researchers discovered that logistic regression is the most commonly used statistical method for assessing consumer credit risk (Paleologo et al., 2010; Finlay, 2009). According to Li and Zhong (2012), logistic regression can forecast an applicant's default likelihood and identify the characteristics associated with his or her default behavior. Furthermore, Brygala (2022) employed logistic regression to anticipate household consumer bankruptcy and discovered that the logit model with a balanced sample has better predictive performance. In addition, Sahiq et al. (2021) employed the logistic regression model and effectively found that the primary factors of personal bankruptcy filings were demographic indicators, socioeconomic status indicators, debt indicators, financial indicators, social stigma indicators, behavioral indicators, and macroeconomic indicators.

3. Research Methodology

This study involved a dataset obtained from an authorized debt management agency. The data consisted of settled members and terminated members. Settled members were those who managed to settle their loans, while terminated were those who were unable to pay their loans. There were 4,174 settled members and 20372 terminated members. The total sample size was 24,546 with 17.0% (4,174) settled and 82.99% (20,372) terminated cases. It is noted here that the negative instances belong to the majority class (terminated) and the positive instances belong to the minority class (settled); an imbalanced dataset. According to Akosa (2017), the most commonly used classification algorithms dataset (e.g. scorecard, logistic regression, and decision tree) do not work well for imbalanced datasets. This is because the classifiers tend to be biased towards the majority class and therefore perform poorly on the minority class. He added, that to improve the performance of the classifiers or model, down-sampling or up-sampling techniques can be used. This paper deployed the random undersampling technique. The random undersampling technique is considered a basic sampling technique in handling imbalanced datasets (Yap, Rahman, He & Bulgiba, 2016). Random undersampling (RUS), also known as down-sampling, excludes the observations from the majority class to balance with the number of available observations in the minority class. The RUS was used by randomly selecting 4,174 cases from the 20,372 terminated cases. This RUS process was done using the IBM Statistical Package for the Social Science (SPSS) software. Therefore, the total sample size was 8,348 with 50% (4,174) representing settled cases and 50% (4,174) representing terminated cases for the balanced dataset. This paper used both sample sizes for further analysis to see the differences in the results of the statistical analysis.

The data covered the period from 1st January 2010 to 31st October 2015, which were received in Excel files. Data cleaning was the first step to remove outliers and redundant data. Once the data cleaning process was completed, the Excel data file was converted into a SAS file using SAS 9.4 software. The logistic regression model was run using the SAS Enterprise Miner 14.1 software.

Logistic regression is an extensively used statistical modeling technique in which the probability of a dichotomous outcome (Y=0 or Y=1) is related to a set of potential predictor variables. The objective of a logistic regression model in personal bankruptcy prediction is to determine the probability of an individual belonging to a class (bankrupt or non-bankrupt), given the values of the independent variables of that individual (Hosmer & Lemeshow, 2013). For this study, logistic regression was used to model the event Y= 1(bankrupt).

The logistic regression model is written as:

$$\log\left[\frac{pi}{1-pi}\right] = \boldsymbol{\beta 0 + \beta_1 X_{j1} + \beta_2 X_{j2} + \ldots\ldots + \beta_k X_{jk}}$$

Where
p_i = P(Y=1) for each case or observation, i = 1,2,…,n
β_0 = intercept of the logistic regression equation
β_j = the coefficient of the predictor variables, j= 1,2…..k
X_j = the predictor variable, j = 1,2,…k

In this logistic regression model, maximum likelihood is used to estimate parameters β_1 to β_k. These parameter estimates measure the rate of change of logit for one unit change in the input variable (adjusted for other inputs), that is, they are the slopes of the regression line between the target and their respective input variables X_1 to X_k. The parameters are dependent on the unit of the input (Siddiqi, 2006).
Whereas, Wald statistics is an alternative test that is commonly used to test the significance of individual logistic regression coefficient resulting in identifying the variables that influence the dependent variable (bankruptcy status):

H_o: $\beta_1 = 0$ (the independent variable does not affect the dependent variable)
H_1: $\beta_1 \neq 0$ (the independent variable affects the dependent variable)

The Wald test calculates z statistics, z =$(\frac{B}{SE})$ Where B is the regression coefficient estimation and SE is the standard error of estimation. The Z value is then squared where the Wald statistic follows the Chi-square distribution (z^2~x12) with the Wald statistic now:
Wald Statistic = $Z^2 = (\frac{B}{SE})^2$
Reject H_o if Wald statistic > $x^2_{\infty,1}$ or p-value< 0.05 ((Katsaragakis et. Al., 2005)

4. Results

This study used the stepwise method in developing the logistic regression model. The stepwise method is useful and intuitively appealing in that it sequentially builds models and it allows for the examination of a collection of models which might not otherwise have been examined (Hosmer et al., 2013). This method was involved in adding and removing characteristics (variable) dynamically from the model in each step until the best combination of characteristics (variable) was achieved. The significance level used in this study was 0.05. Therefore, the best combination of characteristics (variable) is whereby when the p-value of individual characteristics is less than 0.05. Statistical measures like Chi-square or Standardized estimates were used to measure the strength of the predictive model (Siddiqi, 2006). The results showed that the logistic regression models based on imbalanced and balanced datasets were statistically significant. The logistic regression model (imbalanced dataset) was statistically significant with a likelihood ratio of 2542.7031, degree of freedom of 30, and p-values < 0.0001 (Table 1). Meanwhile, the logistic regression model based on a balanced dataset was also statistically significant with a likelihood ratio of 1588.3079, degree of freedom of 27, and p-values <0.0001(Table 2).

Table 1: Logistic Regression Model Results (Imbalanced Dataset)

	Likelihood Ratio Test	**DF**	**Likelihood Ratio Chi-Square**	**P-value**
	Logistic Regression model	**30**	**2542.7031**	**<.0001**
No	**Variables**	**DF**	**Wald Chi-Square**	**P-value**
1	Age	4	19.2614	0.0007
2	Gender	1	10.3994	0.0013
3	Race	2	123.4855	<.0001
4	Marital Status	3	18.1206	0.0004
5	Number of Children	3	53.4015	<.0001
6	Employment Status	4	117.5712	<.0001
7	Household Monthly Income	3	164.7161	<.0001
8	Household Monthly Expenses	3	16.8579	0.0008
9	Outstanding Loan	5	395.2122	<.0001
10	Number of Loans	2	306.5579	<.0001
11	Location of Residence	-	-	-
12	Experience of adversities	-	-	-

Table 2: Logistic Regression Model Results (Balanced Dataset)

	Likelihood Ratio Test	**DF**	**Likelihood Ratio Chi-Square**	**P-value**
	Logistic Regression model	**27**	**1588.4307**	**<.0001**
No	Variables	DF	Wald Chi-Square	P-value
1	Age	4	16.3079	0.0026
2	Gender	1	9.6735	0.0019
3	Race	2	64.9646	<.0001
4	Marital Status	-	-	-
5	Number of Children	3	35.3943	<.0001
6	Employment Status	4	90.9931	<.0001
7	Household Monthly Income	3	87.1208	<.0001
8	Household Monthly Expenses	3	14.0243	0.0029
9	Outstanding Loan	5	243.7229	<.0001
10	Number of Loans	2	204.5720	<.0001
11	Location of Residence	-	-	-
12	Experience of adversities	-	-	-

Based on Table 1 and Table 2, for logistic regression model that is based on imbalanced dataset, the best combination of characteristics (variables) resulting in (10) significant variables (age, race, gender, number of children, employment status, household monthly income, household monthly expenses, outstanding loan, number of loans and marital status) and nine (9) significant variables (age, race, gender, number of children, employment status, household monthly income, household monthly expenses, outstanding loan and number of loans) for logistic regression model that is based on balanced dataset.

Table 3: Analysis of Maximum Likelihood Estimates (Imbalanced Dataset)

Variable	Category	DF	Estimate	P-value	Odd ratio (Exp(Est))
Intercept	-	1	3.6192	<.0001	37.308
Age	1 (20-29)	1	-0.0899	0.1394	0.914
Age	2 (30-39)	1	-0.00791	0.8736	0.992
Age	3 (40-49)	1	0.1636	0.0037	1.178
Age	4 (50-59)	1	-0.1456	0.0224	0.864
Employment status	PRS (Private)	1	-0.1051	0.0666	0.900
Employment status	PUS (Public)	1	-0.4826	<.0001	0.617
Employment status	R (Retired)	1	-0.1884	0.1892	0.828
Employment status	SE (Self-employed)	1	0.4726	<.0001	1.604
Gender	F (Female)	1	-0.0783	0.0013	0.925
Household Monthly Expenses	1 (<RM1k)	1	0.0614	0.2735	1.063
Household Monthly Expenses	2 (RM1,001-RM2,000)	1	0.0464	0.2490	1.048
Household Monthly Expenses	3 (RM2,001-RM3,000)	1	0.1740	0.0010	1.190
Household Monthly Income	1 (<RM1k)	1	0.5488	<.0001	1.731
Household Monthly Income	2 (RM1,001-RM2,000)	1	0.2053	<.0001	1.228
Household Monthly Income	3 (RM2,001-RM3,000)	1	-0.0900	0.0698	0.914
Outstanding Loan	1(≤RM29.9k))	1	-1.2892	<.0001	0.275
Outstanding Loan	2(RM30k-RM59.9k)	1	-0.3551	<.0001	0.701
Outstanding Loan	3(RM60k-RM89.9k)	1	-0.1406	0.1239	0.869
Outstanding Loan	4(RM90k-RM119.9k)	1	0.4967	0.0006	1.643
Outstanding Loan	5(RM120k-RM149.9k)	1	0.3658	0.0297	1.442
Marital Status	D(Divorce)	1	0.0848	0.3414	1.088
Marital Status	M(Married)	1	-0.0511	0.3734	0.950
Marital Status	S(Single)	1	-0.2919	<.0001	0.747
Number of Children	1 (0)	1	-0.2669	<.0001	0.766
Number of Children	2 (1-3)	1	-0.2455	<.0001	0.782
Number of Children	3 (4-6)	1	0.1430	0.0095	1.154
Number of loans	1 (1-4)	1	-1.1387	<.0001	0.320
Number of loans	2 (5-8)	1	-0.0458	0.5493	0.955
Race	C (Chinese)	1	-0.4235	<.0001	0.655
Race	I (Indian)	1	0.3221	<.0001	1.380

Table 4: Analysis of Maximum Likelihood Estimates (Balanced Dataset)

Variable	Category	DF	Estimate	P-value	Odd ratio (Exp(Est))
Intercept	-	1	1.9033	<.0001	6.708
Age	1 (20-29)	1	-0.0724	0.3800	0.930
Age	2 (30-39)	1	0.0802	0.2355	1.084
Age	3 (40-49)	1	0.2090	0.0055	1.232
Age	4 (50-59)	1	-0.1817	0.0307	0.834
Employment status	PRS (Private)	1	-0.2016	0.0080	0.817
Employment status	PUS (Public)	1	-0.6682	<.0001	0.513
Employment status	R (Retired)	1	0.0411	0.8275	1.042
Employment status	SE (Self-employed)	1	0.4632	<.0001	1.589
Gender	F (Female)	1	-0.1025	0.0019	0.903
Household Monthly Expenses	1 (<RM1k)	1	0.0619	0.4331	1.064
Household Monthly Expenses	2 (RM1,001-RM2,000)	1	0.0886	0.1170	1.093

Household Monthly Expenses	3 (RM2,001-RM3,000)	1	0.2293	0.0021	1.258
Household Monthly Income	1 (<RM1k)	1	0.4759	0.0006	1.609
Household Monthly Income	2 (RM1,001-RM2,000)	1	0.2309	0.0003	1.260
Household Monthly Income	3 (RM2,001-RM3,000)	1	-0.0347	0.5947	0.966
Outstanding Loan	1(≤RM29.9k))	1	-1.2809	<.0001	0.278
Outstanding Loan	2(RM30k-RM59.9k)	1	-0.3006	0.0005	0.740
Outstanding Loan	3(RM60k-RM89.9k)	1	0.0222	0.8406	1.022
Outstanding Loan	4(RM90k-RM119.9k)	1	0.3701	0.0260	1.448
Outstanding Loan	5(RM120k-RM149.9k)	1	0.0914	0.6159	1.096
Number of Children	1 (0)	1	-0.2820	0.0002	0.754
Number of Children	2 (1-3)	1	-0.2004	0.0009	0.818
Number of Children	3 (4-6)	1	0.2295	0.0010	1.258
Number of loans	1 (1-4)	1	-1.1474	<.0001	0.317
Number of loans	2 (5-8)	1	-0.0744	0.3764	0.928
Race	C (Chinese)	1	-0.4237	<.0001	0.655
Race	I (Indian)	1	0.3661	<.0001	1.442

Table 3 and Table 4 display the Analysis of Maximum Likelihood Estimates results. In these results, the Odds ratio (OR) is used to measure the strength of association between the predictor variable and the predicted event (Wuensch, 2015; Hailpern & Visintainer, 2003). The predicted event in this study was bankrupt or non-bankrupt. From the odds ratio, it can be determined which category is more or less likely to be bankrupt. Guidelines for interpreting the odds ratio are as per below (Yap et al.,2011):

- If the ratio of category A vs R (reference category) is greater than one, it indicates that those in category A are more likely to be bankrupt.
- If the ratio of category A vs R (reference category) is less than one, it indicates that those in category A are less likely to be bankrupt.
- If the ratio of category A vs R (reference category) is equal to one, it indicates that cases in both A and R are equally likely to be bankrupt.

For reference category, SAS Enterprise E-miner automatically creates (number of categories; c - 1) dummy variables for categorical variables with c levels. The dummy variables for categorical variables with c levels are depicted in Table 5.

Table 5: Dummy Variables for Categorical Variables

Independent Variable	Dummy variables
Age (5 categories) - 20-29, 30-39, 40-49 50-59, 60 & above	4 dummy variables (5-1 = 4) 1,0,0,0 (20-29), 0,1,0,0(30-39), 0,0,1,0(40-49), 0,0,0,1 (50-59), ***0,0,0,0(60 & above)***
Employment status (5 categories) - Private sector (PRS), Public sector (PUS), Retiree (R), Self-employed (SE), Unemployed (U)	4 dummy variables (5-1 = 4) 1,0,0,0 (PRS), 0,1,0,0(PUS), 0,0,1,0(R), 0,0,0,1 (SE), ***0,0,0,0(U)***
Gender (2 categories)- Female, Male	1 dummy variable (2-1 = 1) 1(Female), 0 (Male)
Household Monthly Expenses (4 categories) - <RM1,000, RM1,001-RM2,000, RM2,001-RM3,000, >RM3,000	3 dummy variables (4-1=3) 1,0,0, (<RM1000), 0,1,0,(RM1001-RM2000), 0,0,1,(RM2001-RM3000), ***0,0,0 (>RM3,000)***
Household Monthly Income (4 categories) <RM1,000, RM1,001-RM2,000, RM2,001-RM3,000, >RM3,000	3 dummy variables (4-1=3) 1,0,0, (<RM1000), 0,1,0,(RM1001-RM2000), 0,0,1,(RM2001-RM3000), ***0,0,0 (>RM3,000)***
Outstanding Loan (6 categories) <RM29.9k, RM30k-RM59.9k, RM60k-RM89.9k, RM90k- RM119.9k,	5 dummy variables (6-1=5)

RM120k-RM149.9k, >RM150,000	1,0,0,0,0(<RM29.9k), 0,1,0,0,0(RM30k-RM59.9k), 0,0,1,0,0(RM60k-RM89.9k)), 0,0,0,1,0 (RM90k-RM119.9K), 0,0,0,0,1(RM120k-RM149.9k), ***0,0,0,0,0 (>RM150k)***
Marital status (4 categories) – Divorced, Married, Single, Widow	3 dummy variables (4-1=3) 1,0,0 (Divorced), 0,1,0(Married), 0,0,1(Single), ***0,0,0 (Widow)***
Number of Children (4 categories) 0, 1-3, 4-6, 7 & above	3 dummy variables (4-1=3) 1,0,0 (0), 0,1,0(1-3), 0,0,1(4-6), ***0,0,0 (7& above)***
Number of Loans (3 categories) 1-4, 5-8, 9 & above	2 dummy variables (3-1=2) 1,0 (1-4), 0,1(5-8), ***0,0(4-6)***
Race (3 categories) – Chinese, Indian, Malay	2 dummy variables (3-1=2) 1,0 (Chinese), 0,1(Indian), ***0,0(Malay)***

*The reference category for each variable is bold and italicized

Referring to Table 3 for the imbalanced dataset, the significant variables are those variables with a p-value less than 0.05. Below are the interpretations of the odds ratio for the significant variables.

<u>Age</u>

(OR=1.178): Borrowers who are between 40 to 49 years old are slightly more likely to be bankrupt compared to borrowers aged 60 and above.

(OR=0.864): Borrowers who are between 50 to 59 years old are less likely to be bankrupt compared to borrowers aged 60 and above.

<u>Race</u>

(OR=0.655): Borrowers who are Chinese are less likely to be bankrupt compared to those who are Malay.

(OR=1.380): Borrowers who are Indian are 1.4 times more likely to be bankrupt compared to borrowers who are Malay.

<u>Gender</u>

(OR=0.925): Female borrowers are less likely to be bankrupt compared to male borrowers.

<u>Number of Children</u>

(OR=0.766): Borrowers who have no child are less likely to be bankrupt compared to borrowers with children 7 and above.

(OR=0.782): Borrowers who have 1 to 3 children are less likely to be bankrupt compared to those with children 7 and above.

(OR=1.154): Borrowers who have 4 to 6 children are slightly more likely to be bankrupt compared to borrowers with children 7 and above.

<u>Marital Status</u>

(OR=0.747): Single Borrowers are less likely to be bankrupt compared to borrowers who are widows.

<u>Employment Status</u>

(OR=0.617): Borrowers who are attached to the Public sector are less likely to be bankrupt compared to Unemployed borrowers.

(OR=0.1604): Self-employed Borrowers are 1.6 times more likely to be bankrupt compared to Unemployed borrowers.

<u>Household Monthly Income</u>

(OR=1.731): Borrowers who earn RM1,000 and below are 1.7 times more likely to be bankrupt compared to borrowers who earn RM3k and above.

(OR=1.228): Borrowers who earn RM1,001-RM2,000 are slightly more likely to be bankrupt compared to borrowers who earn RM3k and above.

<u>Household Monthly Expenses</u>

(OR=1.190): Borrowers who spend between RM2,001 to RM3,000 are slightly more likely to be bankrupt compared to borrowers who spend RM3k and above.

<u>Outstanding Loan</u>

(OR=0.275): Borrowers with outstanding loans of RM29,999 and below are less likely to be bankrupt compared to borrowers with outstanding loans of more than RM150k.

(OR=0.701): Borrowers with outstanding loans between RM30,000 to RM59,999 are less likely to be bankrupt compared to borrowers with outstanding loans of more than RM150k.

(OR=1.643): Borrowers with outstanding loans between RM90,000 to RM119,999 are 1.6 times more likely to be bankrupt compared to borrowers with outstanding loans of more than RM150k.
(OR=1.442): Borrowers with outstanding loans between RM120,000 to RM149,999 are 1.4 times more likely to be bankrupt compared to borrowers with outstanding loans of more than RM150k.
Number of Loans
(OR=0.320): Borrowers who have 1 to 4 loans are less likely to be bankrupt compared to borrowers with 9 loans and above.

Next, based on Table 4, below are the interpretations of odd ratios for the significant variables for the balanced dataset. The significant variables are those variables with a p-value less than 0.05.
Age
(OR=1.232): Borrowers who are between 40 to 49 years old are slightly more likely to be bankrupt compared to borrowers aged 60 and above.
(OR=0.834): Borrowers who are between 50 to 59 years old are less likely to be bankrupt compared to borrowers aged 60 and above.
Race
(OR=0.655): Borrowers who are Chinese are less likely to be bankrupt compared to those who are Malays.
(OR=1.442): Borrowers who are Indian are 1.4 times more likely to be bankrupt compared to borrowers who are Malay.
Gender
(OR=0.903): Female borrowers are less likely to be bankrupt compared to male borrowers.
Number of Children
(OR=0.754): Borrowers who have no child are less likely to be bankrupt compared to borrowers with 7 children and above.
(OR=0.818): Borrowers who have 1 to 3 children are less likely to be bankrupt compared to borrowers with 7 children and above.
(OR=1.258): Borrowers who have 4 to 6 children are slightly more likely to be bankrupt compared to borrowers with 7 children and above.
Employment Status
(OR=0.817): Borrowers who are attached to the Private sector are less likely to be bankrupt compared to Unemployed borrowers.
(OR=0.513): Borrowers who are attached to the Public sector are less likely to be bankrupt compared to Unemployed borrowers.

(OR=1.589): Self-employed Borrowers are 1.5 times more likely to be bankrupt compared to unemployed borrowers.
Household Monthly Income
(OR=1.609): Borrowers who earn RM1,000 and below are 1.6 times more likely to be bankrupt compared to borrowers who earn RM3k and above.
(OR=1.206): Borrowers who earn RM1,001-RM2,000 are slightly more likely to be bankrupt compared to borrowers who earn RM3k and above.
Household Monthly Expenses
(OR=1.258): Borrowers who spend between RM2,001 to RM3,000 are slightly more likely to be bankrupt compared to borrowers who spend RM3k and above.
Outstanding Loan
(OR=0.278): Borrowers with outstanding loans of RM29,999 and below are less likely to be bankrupt compared to borrowers with outstanding loans of more than RM150k.
(OR=0.470): Borrowers with outstanding loans between RM30,000 to RM59,999 are less likely to be bankrupt compared to borrowers with outstanding loans of more than RM150k.
(OR=1.448): Borrowers with outstanding loans between RM90,000 to RM119,999 are 1.4 times more likely to be bankrupt compared to borrowers with outstanding loans of more than RM150k.
Number of Loans
(OR=0.317): Borrowers who have 1 to 4 loans are less likely to be bankrupt compared to borrowers with 9 loans and above.

Therefore, based on Table 3 and Table 4, the estimated logistic regression for the imbalanced dataset and the balanced dataset is written as below:

Logistic Regression model (Imbalanced Dataset)

$\log\left[\frac{pi}{1-pi}\right]$ = 3.6192 – 0.0899(Age 1) – 0.00791(Age 2) + 0.1636(Age 3) – 0.1456(Age 4) – 0.1051(PRS) – 0.4826(PUS) – 0.1884(R) + 0.4726(SE) – 0.0783(F) + 0.0614 (HME 1) + 0.0464(Household Monthly Expenses 2) + 0.1740(Household Monthly Expenses 3) + 0.5488(Household Monthly Income 1) + 0.2053 (Household Monthly Income 2) – 0.0900(Household Monthly Income 3) – 1.2892(Outstanding loan 1) – 0.3551(Outstanding loan 2) – 0.1406(Outstanding loan 3) + 0.4967(Outstanding loan 4) + 0.3658 (Outstanding loan 5) + 0.0848(D) – 0.0511(M) – 0.2919(S) – 0.2669(No of child 1) – 0.2455(No of child 2) + 0.1430(No of child 3) – 1.1387(No of loans 1) – 0.0458(No of loans2) – 0.4235(Chinese) + 0.3221(Indian).

Logistic Regression model (Balanced Dataset)

$\log\left[\frac{pi}{1-pi}\right]$ = 1.9033 – 0.0724(Age 1) + 0.0802(Age 2) + 0.2090(Age 3) – 0.1817(Age 4) – 0.2016(PRS) – 0.6682(PUS) + 0.0411(R) + 0.4632(SE) – 0.1025(F) + 0.0619(HME 1) + 0.0886(HME 2) + 0.2293(HME 3) + 0.4759(HMI 1) + 0.2309(HMI 2) – 0.0347(HMI 3) – 1.2809(Outstanding loan 1) – 0.3006 (Outstanding loan 2) + 0.0222(Outstanding loan 3) + 0.3701(Outstanding loan 4) + 0.0914(Outstanding loan 5) – 0.2820(No of child 1) – 0.2004(No of child 2) + 0.2295(No of child 3) – 1.1474(No of loan 1-4) – 0.0744(No of loan 5-8) – 0.4237(Chinese) + 0.3661(Indian).

Based on the above result, the following is the discussion on the results of significant categories. The significant categories were those categories with p-values less than 0.05 (Table 1 and Table 2). In terms of age, both logistic regression models found that the age category that was more likely to be bankrupt was between 40 to 49 years as compared to 60 years and above. The finding on the age category between 40 to 49 years was the more likely to bankrupt fell within part of the range of age group of bankrupts (35 to 44) as reported by the Malaysia Department of Insolvency (2017-2023) and Hospodka (2015). They reported that the highest percentage of bankruptcies and the most vulnerable age group is between 35 to 44 years old. This model further identified that borrowers aged between 50 to 59 years old were less likely to be bankrupt as compared to borrowers aged between 60 and above.

Both logistic regression models found that the male category was more likely to be bankrupt compared to the female category. This finding is similar to Othman et al. (2015), Eaw et al. (2014, 2015), and Jullamon (2012) where they indicated that the majority of bankrupts are male. Malaysia Department of Insolvency also reported that the highest percentage of bankrupt is male (Malaysia Department of Insolvency, 2017-2023).
For employment status, the researcher found that borrowers who were attached to the public sector were less likely to be bankrupt compared to unemployed borrowers. This finding is similar to both Desai (2016) and Zhu (2011) who pointed out that the percentage of bankruptcy filing probability is higher for the unemployed filers compared to the employed filers. On the other hand, self-employed borrowers were 1.6 times more likely to be bankrupt as compared to unemployed borrowers. According to the Malaysia Department of Insolvency, self-employed are among the employment statuses declared by the bankrupts (Malaysia Department of Insolvency, 2017- 2023)

Hospodka et al. (2015) found that a higher percentage of lower-income debtors file for bankruptcy compared to higher-income debtors. Meanwhile, Agarwal et al. (2011) showed that a 1% increase in income could lead to a decrease of 20% in bankruptcy filing. In terms of income, the results showed that borrowers who earned RM1,000 and below were 1.7 times more likely to be bankrupt compared to borrowers who earned RM3k and above and borrowers who earned RM1,001-RM2,000 were slightly more likely to be bankrupt as compared to borrowers who earned RM3k and above. These findings support Hospodka (2015) and Agarwal et al. (2011) and it is an extension to the literature where the results indicated the specific range of income of bankrupt.

As for expenses, the results indicated that borrowers who spent between RM2001 and RM3000 were more likely to be bankrupt. This finding also furthers the literature as the result showed the exact range of expenses that the borrowers spent every month. Whereas, Zhu (2011), Agarwal et al. (2011), Azaizeh (2010), and Rhee (2001) only indicated that high expenses increase the likelihood of bankruptcy filing.

The outcome of the outstanding loan category showed that borrowers with outstanding loans of more than RM150k (high outstanding balance) were most likely to be bankrupt. This is similar to the findings of Hospodka et al. (2015), Dawsey (2014), Gross and Souleles (2002), and Stavins (2000) which indicated that households who had filed for bankruptcy carried high loan balances. The logistic regression models also found that borrowers with outstanding loans between RM90,000 to RM119,999 were 1.6 times more likely to be bankrupt as compared to borrowers with outstanding loans of more than RM150k.

The results indicated that married and divorce were an insignificant category and single borrowers were less likely to be bankrupt compared to borrowers who were widows. This finding supports the results found by Jullamon (2012) who indicated that 8.6% of the total respondents who filed for bankruptcy were widows. According to Zhu (2011), borrowers file for bankruptcy soon after the divorce or death of a spouse because of household income issues. Divorce or the death of a spouse puts household members into a difficult situation because they do not have much earning power or well-established credit. As a result, borrowers could file for bankruptcy after getting divorced or being a widow.

Borrowers with no child or 1 to 3 children were less likely to be bankrupt compared to those with 7 children and above. The finding above is consistent with Creswell (2014) who found that borrowers with larger families had a higher probability of bankruptcy. Zhu (2011) also found that the more children the borrowers have, the higher the probability of bankruptcy filing. In addition, the results also found that borrowers who had 4 to 6 children were slightly more likely to be bankrupt as compared to borrowers with 7 children and above. In terms of number of loans, borrowers who had 1 to 4 loans were less likely to be bankrupt as compared to borrowers with 9 loans and above. This result supports Desai (2016) and Dawsey (2014) who pointed out that the higher the number of loans, the more likely the borrowers file for bankruptcy.

Lastly, for the race category, borrowers who were Chinese were less likely to be bankrupt as compared to Malays. This is similar to the report from the Malaysia Department of Insolvency in 2009 and 2013 where they found that the Malay race is the majority of bankrupts in Malaysia (Malaysia Department of Insolvency, 2010, 2014). From 2014 onwards, the Malaysia Department of Insolvency does not report on the race category which the researcher believes is due to the issue of racial sensitivity. For Indian borrowers, they were 1.4 times more likely to be bankrupt as compared to Malay borrowers. One likely reason for this outcome could be that Indian households are earning low income as compared to Malay households. This result is in line with Khalid (2011) who found that the Indian has the least ownership in wealth where 23.7% of them did not have wealth compared to 14.7% of the Bumiputera and 10.5% of the Chinese.

In addition, the predictive performance of the logistic regression model can be evaluated by using performance measure which consists of misclassification, accuracy, sensitivity, and specificity rate. Misclassification is the probability that the model has wrongly predicted bankrupt as non-bankrupt and non-bankrupt as bankrupt. Accuracy means the probability of the model correctly predicted bankrupt and non-bankrupt. Sensitivity is the probability that the model can correctly predict bankruptcy and specificity means the probability that the model can correctly predict non-bankrupt. A good model consists of a lower misclassification rate and higher accuracy and sensitivity rate (Akosa, 2017; Brown, 2014).

Table 6 displays the performance measure results. The results showed that the validation classification accuracy and specificity for imbalanced datasets are 83.17% and 8.38% respectively. The specificity rate is low, less than 10% and the sensitivity rate is 98.49% (high rate, almost perfect) which indicates that the personal bankruptcy prediction model was affected by the imbalanced data. Then, undersampling was performed by randomly selecting 4,174 cases from the 20,372 terminated cases and re-evaluating the model using the balanced sample of 8,348 cases. The validation classification accuracy decreased slightly to 71.144% but the sensitivity rate increased to 78.43%.

Table 6: Performance Measure Results (Accuracy, Precision, Specificity and Sensitivity)

		Imbalanced data (n=24546)		Balanced data (n=8348)	
Model		**Training (%)**	**Validation (%)**	**Training (%)**	**Validation (%)**
Logistic Regression	Accuracy	83.35	83.17	71.71	71.14
	Misclassification	16.64	16.82	28.29	28.86
model	Specificity	9.75	8.38	78.16	78.43
	Sensitivity	98.43	98.49	65.25	63.84

5. Managerial Implications and Recommendations

This paper presents some important ideas to policy maker or managerial level of commercial banks and credit companies based on the statistical analysis results. They should give more attention to borrowers with the following characteristics: male, aged between 30-39 years old, Malay, married, number of children between 1 – 3 children, work in the private sector, monthly household income RM3,000 and above, monthly household expenses between RM1,001-RM2,000, number of loans between 1- 4, outstanding loan RM30,000 and above, never experience adversities and live in the city. This is because policymakers at or managerial level can develop strategies that can help to reduce defaulters. The fewer the defaulters, the lesser will be the bankrupts. Furthermore, wrong credit assessment leads to an increase in the number of defaulters and as a sequence could drive financial institutions toward bankruptcy (Kambal et al., 2013). In addition, policymakers maker or managerial levels can also develop strategies which meet the organisation requirements and potential borrowers' needs. The strategies include providing short financial management courses to potential loan applicants or borrowers, educating on personal bankruptcy impact through a personal bankruptcy awareness campaign, tightening loan application procedures, offering low loan amounts, shorter loan periods, and regular follow-up on default accounts.

Conclusion: This paper discussed the improvements in the prediction of personal bankruptcy using random undersampling to correct the imbalanced data. The application of the logistic regression technique in this study showed that the specificity rate increased after the random undersampling strategy was applied. In conclusion, the predictive performance of the personal bankruptcy model based on the balanced dataset is more reasonable compared to the imbalanced dataset. According to Yap et al. (2011), in practical applications, classification methods that are easy to understand such as decision trees and scorecards are more appealing to users (Yap et al.,2011). For future research, the researcher intends to consider a Decision Tree, Scorecard, Support Vector Machine and Naïve Bayes model.

Acknowledgments: The authors would like to express their gratitude to Agensi Kaunseling dan Pengurusan Kredit (AKPK), Kuala Lumpur for their continuous support and cooperation.

References

Agarwal, S., Chomsisengphet, S., & Liu, C. (2011). Consumer bankruptcy and default: The role of individual social capital. Journal of Economic Psychology, 32(4), 632–650. https://doi.org/10.1016/j.joep.2010.11.007

Akosa, J. S. (2017). Predictive accuracy : A misleading performance measure for highly imbalanced data classified negative. SAS Global Forum, 1–12. Retrieved from http://support.sas.com/resources/papers/proceedings17/0942-2017.pdf

Azaizeh, S. Y. (2010). Essays on household demand for credit cards, bankruptcy and overspending, (Available from ProQuest Dissertations & Theses Global. (763225189). Retrieved from http://search.proquest.com.ezaccess.library.uitm.edu.my/docview/763225189?accountid=42518).

Brown, I. (2014). Developing Credit Risk Models Using SAS Enterprise Miner and SAS/STAT. Cary, NC: SAS Institute Inc

Brygała, M. (2022). Consumer bankruptcy prediction using balanced and imbalanced data. Risks, 10(2), 24.

Brygała, M., & Korol, T. (2024). Personal bankruptcy prediction using machine learning techniques. Economics and Business Review, 10(2).

Cohen, E. K. (2015). Definition of bankruptcy. In Wiley Encyclopaedia of Management (pp. 1–2).

https://doi.org/10.1002/9781118785317.weom080051

Creswell, P. D. (2014). Personal bankruptcy and the health of families and children, (Available from ProQuest Dissertations & Theses Global. (1612415829). Retrieved from http://search.proquest.com.ezaccess.library.uitm.edu.my/docview/1612415829?accountid=42518).

Dawsey, A. E. (2014). Externalities among creditors and personal bankruptcy. Journal of Financial Economic Policy, 6(1), 2–24. https://doi.org/10.1108/JFEP-09-2013-0037

Desai, C. A. (2016). Mortgage credit and bankruptcy chapter choice: a case of US counties. Managerial Finance, 42(7), 680–705. http://doi.org/10.1108/MF-05-2015-0155

Eaw, H. C., Khong, K. W., Rajagopalan, U., & A Hamid, B. (2014). Causality of personal bankruptcy in Malaysia. Emerging Trends in Scientific Research, 363–367.

Eaw, H. C., Rajagopalan, U., Hamid, B. A., Khong, K. W., & Ahmad, R. (2015). Psychographic factors : Does it influence personal bankruptcy In Malaysia ? Journal of Business and Economic Studies, 1(2), 72–78.

Finlay, S. (2009). Are we modelling the right thing? The impact of incorrect problem specification in credit scoring. Expert Systems with Applications, 36(5), 9065–9071. https://doi.org/10.1016/j.eswa.2008.12.016

Garrido, J., Nadeem, S., Riad, N., Delong, C., Rendak, N., & Rosha, A. (2020). Tackling private over-indebtedness in Asia : Economic and legal aspects. *IMF Working Papers*, *WP/20/172*(August), 1–36.

Gross, D. B., & Souleles, N. S. (2002). An empirical analysis of Personal Bankruptcy and delinquency. The Review of Financial Studies, 15(1), 319–347.

Hailpern, S. M., & Visintainer, P. F. (2003). Odds ratios and logistic regression : further examples of their use and interpretation. The Stata Journal, (3), 213–225.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Introduction to the Logistic Regression Model. In Applied Logistic Regression (Vol. 398, pp. 1–6). John Wiley & Sons. Retrieved from http://resource.heartonline.cn/20150528/1_3kOQSTg.pdf.

Hospodka, J., Buben, O., Randáková, M., & Bokšová, J. (2015). Personal bankruptcy in the Capital City region and South Bohemian region in the Czech Republic. Procedia Economics and Finance, 25(182), 41–52. https://doi.org/10.1016/S2212-5671(15)00711-X

International Monetary Fund. (2017). *Global Financial Stability Report 2017 - Chapter Two: Household Debt and Financial Stability* (Issue October 2017). https://www.imf.org/en/Publications /GFSR/Issues/2017/09/27/global-financial-stability-report-october-2017

Jullamon, K. (2012). Consumer bankruptcy in Thailand, (Available from ProQuest Dissertations & Theses Global. (1562209086). Retrieved from http://search.proquest.com.ezaccess.library.uitm.edu.my/docview/1562209086?accountid=42518).

Kambal, E., Osman, I., Taha, M., Mohammed, N., & Mohammed, S. (2013). Credit scoring using data mining techniques with particular reference to Sudanese banks. 2013 International Conference on Computing, Electrical and Electronic Engineering (ICCEEE), 378–383. https://doi.org/10.1109/ ICCEEE.2013.6633966

Katsaragakis, S., Koukouvinos, C., Stylianou, S., Theodoraki, E.-M., & Theodoraki, E.-M. (2005). Comparison Of Statistical Tests In Logistic Regression: The Case of Hypernatremia. Journal of Modern Applied Statistical Methods, 4(2), 514–521. https://doi.org/10.22237/jmasm/1130804100

Khalid, M. A. (2011). Household wealth in Malaysia: Composition and inequality among ethnic groups. Jurnal Ekonomi Malaysia, 45(1), 71-80.

Korol, T., & Fotiadis, A. (2022). Implementing artificial intelligence in forecasting the risk of personal bankruptcies in Poland and Taiwan. Oeconomia Copernicana, 13, 407-438.)

Li, X.-L., & Zhong, Y. (2012). An overview of personal credit scoring: Techniques and future work. International Journal of Intelligence Science, 02(24), 182–190. https://doi.org/10.4236/ijis.2012.224024

Malaysia Department of Insolvency, 2009 Annual Report. (2010). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrieved from http://www.bheuu.gov.my/portal/pdf

Malaysia Department of Insolvency, 2010 Annual Report. (20111). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrieved from http://www.bheuu.gov.my/portal/pdf

Malaysia Department of Insolvency, 2013 Annual Report. (2014). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrieved from http://www.bheuu.gov.my/portal/pdf

Malaysia Department of Insolvency, 2014 Annual Report. (2015). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrieved from http://www.bheuu.gov.my/portal/pdf

Malaysia Department of Insolvency, 2017 Annual Report. (2018). Bahagian Hal Ehwal Undang-Undang, Jabatan

Perdana Menteri, Retrived from http://www.bheuu.gov.my/portal/pdf.

Malaysia Department of Insolvency, 2018 Annual Report. (2019). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrived from http://www.bheuu.gov.my/portal/pdf.

Malaysia Department of Insolvency, 2019 Annual Report. (2020). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrived from http://www.bheuu.gov.my/portal/pdf.

Malaysia Department of Insolvency, 2020 Annual Report. (2021). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrived from http://www.bheuu.gov.my/portal/pdf.

Malaysia Department of Insolvency, 2021 Annual Report. (2022). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrived from http://www.bheuu.gov./portal/pdf.

Malaysia Department of Insolvency, 2022 Annual Report. (2023). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrived from http://www.bheuu.gov.my/portal/pdf.

Malaysia Department of Insolvency, 2023 Annual Report. (2024). Bahagian Hal Ehwal Undang-Undang, Jabatan Perdana Menteri, Retrived from http://www.bheuu.gov.my/portal/pdf.

Malaysia Department of Insolvency (MdI). (2020). Personal Bankruptcy. http://www.mdi.gov.my/index.php/about-us/core-business/core-business/personal-bankruptcy

Othman, M. A., Abdul Rahim, H., & Sabri, M. F. (2015). Differences in financial information and financial status among Malaysian bankrupts. Proceedings of the Australasian Conference on Business and Social Sciences 2015, 525–531. Retrieved from https://www.aabss.org.au/system/files/published/000959-published-acbss-2015-sydney.pdf

Paleologo, G., Elisseeff, A., & Antonini, G. (2010). Subagging for credit scoring models. *European Journal of Operational Research*, *201*(2), 490–499. https://doi.org/10.1016/j.ejor.2009.03.008

Papana, A., & Spyridou, A. (2020). Bankruptcy prediction: the case of the Greek market. Forecasting, 2(4), 505-525.

Rhee, E. (2001). Determinants of consumer bankruptcy: Evidence from the Panel Study of Income Dynamics(PSID) 1984-1995. Arbor Ciencia Pensamiento Y Cultura. https://doi.org/10.16953/deusbed.74839

Sahiq, A. N. M. (2021). Towards Achieving Long-Term Debt Sustainability: A Systematic Review of the Key Determinants of Personal Bankruptcy. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(3), 1305-1317.

Siddiqi, N. (2006). Credit Risk Scorecards : Developing and Implementing Intelligent Credit Scoring. John Wiley & Sons, Inc., 1, 1–210.

Smiti, S., Soui, M., & Ghedira, K. (2024). Tri-XGBoost model improved by BLSmote-ENN: an interpretable semi-supervised approach for addressing bankruptcy prediction. *Knowledge and Information Systems*, 1-38.

Stavins, J. (2000). Credit card borrowing, delinquency, and personal bankruptcy. New England Economic Review, 2000(July), 15–30. Retrieved from http://www.bostonfed.org/economic/neer/neer2000/neer400b.pdf

Wuensch, K. L. (2015). Logistic Regression With SAS, (http://core.ecu.edu/psyc/wuenschk /MV/multReg/Logistic-SAS.pdf), 1–6.

Yap, B. W., Rahman, H. A. A., He, H., Bulgiba, A., Salleh, S., Aris, N., ... Lee, M. H. (2016). Handling imbalanced datasets using SVM and k-NN approach. AIP Conference Proceedings, 1750, 20023). AIP Publishing.

Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using Data Mining to Improve Assessment of Credit Worthiness via Credit Scoring Models. Expert Systems with Applications, 38(10), 13274–13283. https://doi.org/10.1016/j.eswa.2011.04.147

Zhu, N. (2011). Household consumption and personal bankruptcy. Journal of Legal Studies, 40(1), 1–37.