Spring 2018

A Comparison of Machine Learning Algorithms for Prediction of Past Due Service in Commercial Credit

Liyuan Liu M.A, M.S.
Analytics and Data Science, lliyuan@students.kennesaw.edu

Jennifer Lewis Priestley Ph.D.
Analytics and Data Science, jpriestl@kennesaw.edu

A Comparison of Machine Learning Algorithms for Prediction of Past Due Service in Commercial Credit

Liyuan Liu
Analytics and Data Science
Kennesaw State University
1000 Chastain Road
Kennesaw, Georgia 30144
Email: lliyuan@students.kennesaw.edu

Jennifer Lewis Priestley
Analytics and Data Science
Kennesaw State University
1000 Chastain Road
Kennesaw, Georgia 30144
Email: jpriestl@kennesaw.edu

***Abstract*—Credit risk modeling has carried a variety of research interest in previous literature, and recent studies have shown that machine learning methods achieved better performance than conventional statistical ones. This study applies decision tree which is a robust advanced credit risk model to predict the commercial non-financial past-due problem with better critical power and accuracy. In addition, we examine the performance with logistic regression analysis, decision trees, and neural networks. The experimenting results confirm that decision trees improve upon other methods. Also, we find some interesting factors that impact the commercials' non-financial past-due payment.**

***Index Terms*—Past Due; Credit Risk Model; Logistic Regression; Decision Trees; Neural Networks;**

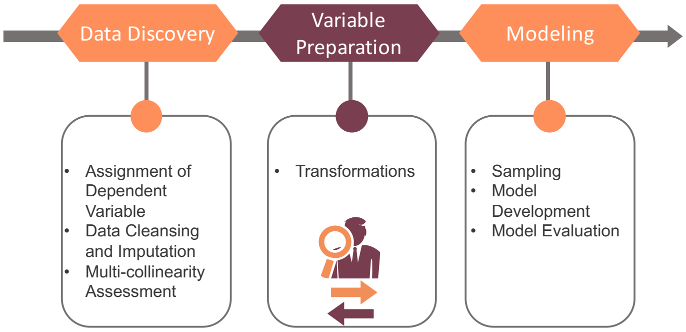


Fig. 1. Workflow of Modeling Process

I. Introduction

Commercial loan is a phrase commonly used to indicate a loan not ordinarily maintained by either the real estate or consumer loan departments. In asset distribution, commercial or business loans comprise one of the most critical assets of a bank [1]. The volume of the commercial loan is exceptionally high; according to Real Capital Analytics(RCA) report, the total global amount of commercial loan was $826 billion in 2016. Nowadays, the loan default still happened usually in the commercial lender. Based on the Federal Reserve senior loan officer opinion survey report, oil and gas companies defaulted on $39 billion in 2016, and the high yield bond default rate for the energy sector peaked at 18.8% during the year. In fact, the loan defaults happened in every industry. Therefore, the commercial credit risk prediction is a critical research part that helps to protect the economic environment.

In this study, we used a real-world dataset provided by Equifax and compare different machine learning algorithms including logistic regression, decision trees, and neural networks. The modeling process is described in Figure 1. At first, we described the process of data imputation, transforming, and selecting model variables from messy and sparse data. Secondly, we illustrated the logistic regression, decision tree, and neural network algorithms. Finally, we discussed the results, compared the models using ROC index and Kolmogorov-Smirnov statistic and found the most important factors that impact commercial credit.

II. Related Work

There are many related works in credit risk modeling research. Nargundkar and Priestley examined and compared the most prevalent modeling techniques in the credit industry [2]. Laitinen predicted corporate credit analyst's risk estimate by the weighted logistic and linear regression analyses [3]. The datasets included 35 variables from 3200 observations. Atiya developed a neural network model to predict bankruptcy [4]. First, the author reviewed the topic of bankruptcy prediction, with emphasis on neural-network (NN) models. Second, the author developed an NN bankruptcy prediction model. Huang *et al.* employed support vector machines and neural networks to credit rating analysis [5]. The authors used backpropagation neural network (BNN) as a benchmark and obtained prediction accuracy around 80% for both BNN and SVM methods for the United States and Taiwan markets. The dataset included 74 cases with bank credit rating and 21 financial variables, which covered 25 financial institutes from 1998 to 2002. Wang *et al.* used the fuzzy support vector machine to evaluate credit risk [6]. Authors provided a new fuzzy SVM to evaluate the credit risk of consumer lending. The dataset contains 30 failed and 30 non-failed firms. Twelve variables are used as the firms' characteristics. Lin *et al.* published a survey that reviewed 130 related journal papers from the period between 1995 and 2010, focusing on the development of state-of-the-art machine-learning techniques, including hybrid and ensemble classifiers in financial crisis prediction research [7]. Pinches *et al.* utilized multiple discriminant analysis (MDA) to bear a linear discriminant function relating a set of independent variables to a dependent variable to better suit the ordinal nature of bond-rating data and increase classification accuracy. Other researchers also utilized logistic regression analysis and probit analysis [8], [9], [10]. Recently, artificial intelligence techniques and machine learning techniques such as neural networks and decision trees have been used to support such analysis. Neural networks emerged in Hagan *et al.*'s research and are widely used in many different domains, including credit risk modeling [11].

III. Data Discovery

This large-scale study examined 36 separate datasets, with each dataset drawing quarterly commercial loan related information from 2006 to 2014. Each dataset contains 11,787,287 observations and 305 explanatory attributes from different commercials. These explanatory variables include five categories: non-financial account, telecommunication account, utility account, service account, and industry account. This study focuses on predicting commercial past due activities in their service accounts using SAS9.4.

A. Assignment of Dependent Variable

We assigned "totNFPDAmt12mon" as target variable of interested which represent the total non-financial past due amount in last 12 months. There are two main reseasons why we chose this variable as the target variable. Our first goal was to predict if the commerce would have non-financial past due in the future. Usually, the financial loan is the most prioritizes by the commerce because of the significant cost of the late payment. However, sometimes, even though the commerce doesn't have financial past dues, it can still have the non-financial past dues. Therefore, predicting commerces non-financial past dues is more accurate than predicting commerces financial past dues to estimate if the commerce will default. Second, we chose a dependent variable that did not contain more than 80% coded values. We did this to ensure our variable had enough information for accurate prediction. Figure 2 shows the distribution of the selected target variable; the variable has 60% coded value which fits our rules. Besides, based on the previous research, the typical performance window of credit risk model is 12-24 months [12]. Ultimately, we decided to choose the total non-financial past due amount in last 12 months as the dependent variable. We created a new variable "GOODBAD" as the target variable. Figure 3 shows the example of the new target variable creation step. The value of "totNFPDAmt12mon" represents the commercial non-financial past due amount in last 12 months. If the "totNFPDAmt12mon" value is 0, we define the "GOODBAD" value as 0. Otherwise, "GOODBAD" is defined as 1.

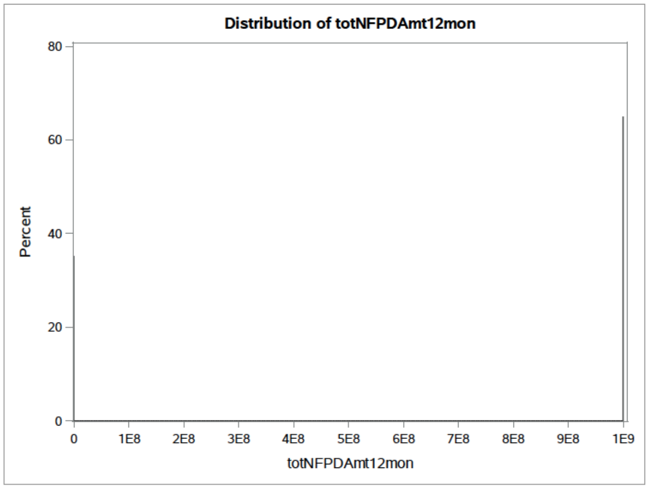
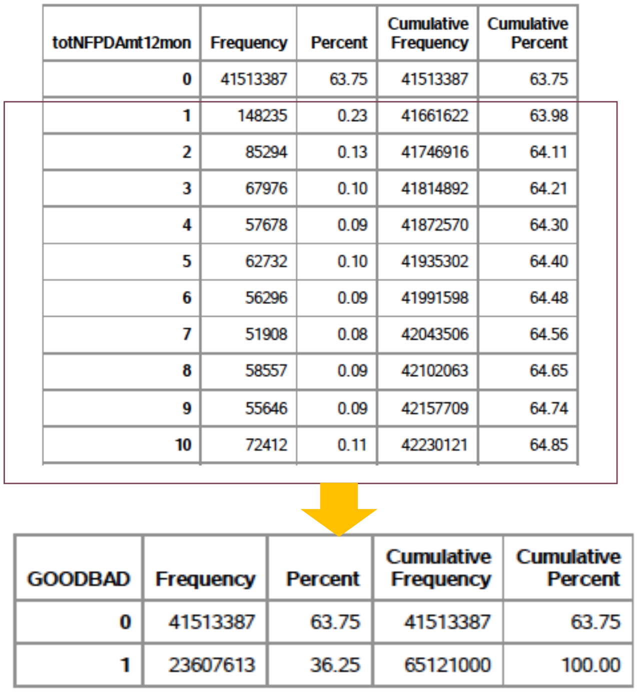


Fig. 2. Distribution of Selected Dependent Variable



totNFPDAmt12mon	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	41513387	63.75	41513387	63.75
1	148235	0.23	41661622	63.98
2	85294	0.13	41746916	64.11
3	67976	0.10	41814892	64.21
4	57678	0.09	41872570	64.30
5	62732	0.10	41935302	64.40
6	56296	0.09	41991598	64.48
7	51908	0.08	42043506	64.56
8	58557	0.09	42102063	64.65
9	55646	0.09	42157709	64.74
10	72412	0.11	42230121	64.85

GOODBAD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	41513387	63.75	41513387	63.75
1	23607613	36.25	65121000	100.00

Fig. 3. Distribution of the Binary Dependent Variable GOODBAD in the Merged Dataset

B. Independent Variables Cleansing and Imputation

There were three steps of the cleansing and imputation of explanatory variables.

(i) The first step was dimensionality reduction by removing variables with the high ratio of coded or missing values. There are many pieces of research about how to remove the variables that including missing values. Based on the book *Discovering Knowledge in Data: An Introduction*

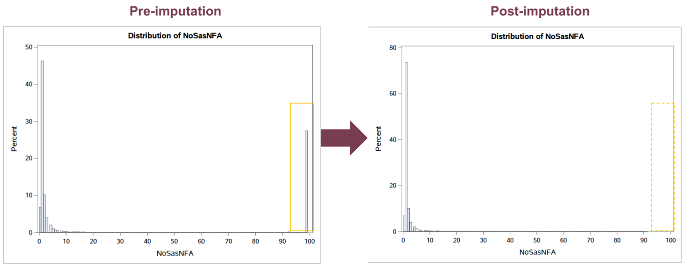
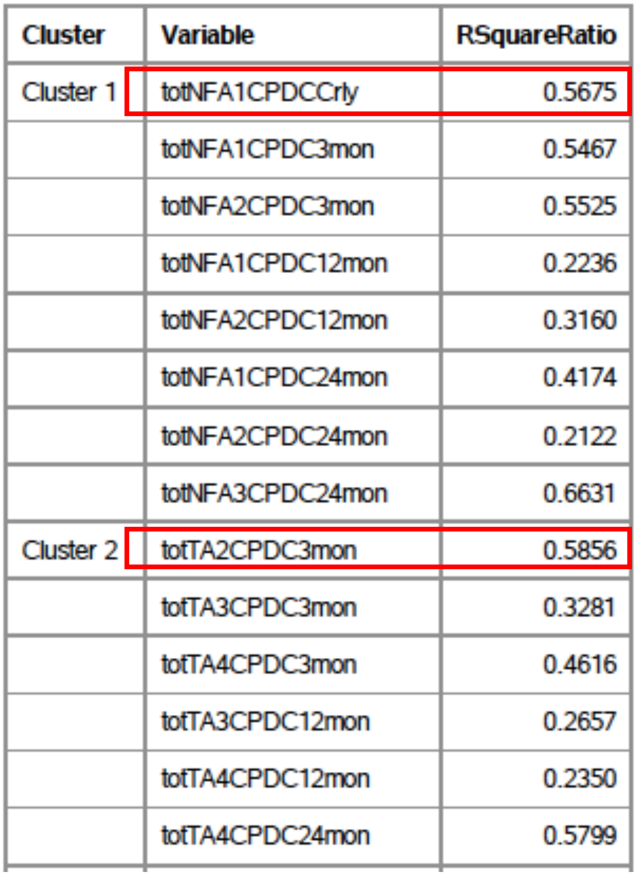


Fig. 4. Example of Variable Distribution with Pre-imputation and Post-imputation

to Data Mining, variables that contain more than 90% missing or coded values should be removed [13]. In addition, based on Allison's research, variables should be deleted when they include 70% missing values or higher [14]. In this case, we employed Allison's research by deleting the independent variables containing more than 70% missing or coded values. We removed the categorical variables which only contain one level. We also removed the categorical variables that indicate specific information such as zip code and state name.

(ii) The second step was imputing the missing or coded values by median and most frequency. Median imputation is a method replacing any missing or coded value with the median of that variable for all other cases, which has the benefit of not changing the sample median for that variable. Most frequency imputation is an imputation method of the categorical variable that is replacing missing or coded value with the most frequency. In Bennett's research, he points out some different methods to impute the missing value such as last value carried forward, mean substitution, regression methods, hot-deck imputation and cold-deck imputation [15]. In this study, since the numeric data are continuous and the dataset is extensive, we chose the median imputation for numeric variables and the most frequency imputation for categorical variables. Median imputation is a method in which the median of the available cases replaces the missing value on a particular variable. This method maintains the sample size and is easy to use [16]. This imputation method will help the variables avoid a skewed distribution. Figure 4 shows the example of a numeric variable's distribution pre-imputation and post-imputation. From Figure 4, we can see the median imputation of the valid values is more stable in highly skewed data. After the first and second data cleaning steps, there are 109 variables, 96 numeric variables, and 13 categorical variables remaining. Before going to the next step, we ran he c-test and found that there was no significant improvement of models between including and excluding the categorical variables. Therefore, before going to the next step, we removed the categorical variables.

(iii) The third step was dimensionality reduction by variable clustering. Dimension reduction is a process of reducing



Cluster	Variable	RSquareRatio
Cluster 1	totNFA1CPDCCrly	0.5675
	totNFA1CPDC3mon	0.5467
	totNFA2CPDC3mon	0.5525
	totNFA1CPDC12mon	0.2236
	totNFA2CPDC12mon	0.3160
	totNFA1CPDC24mon	0.4174
	totNFA2CPDC24mon	0.2122
	totNFA3CPDC24mon	0.6631
Cluster 2	totTA2CPDC3mon	0.5856
	totTA3CPDC3mon	0.3281
	totTA4CPDC3mon	0.4616
	totTA3CPDC12mon	0.2657
	totTA4CPDC12mon	0.2350
	totTA4CPDC24mon	0.5799

Fig. 5. Example of Chosen Representative Variable

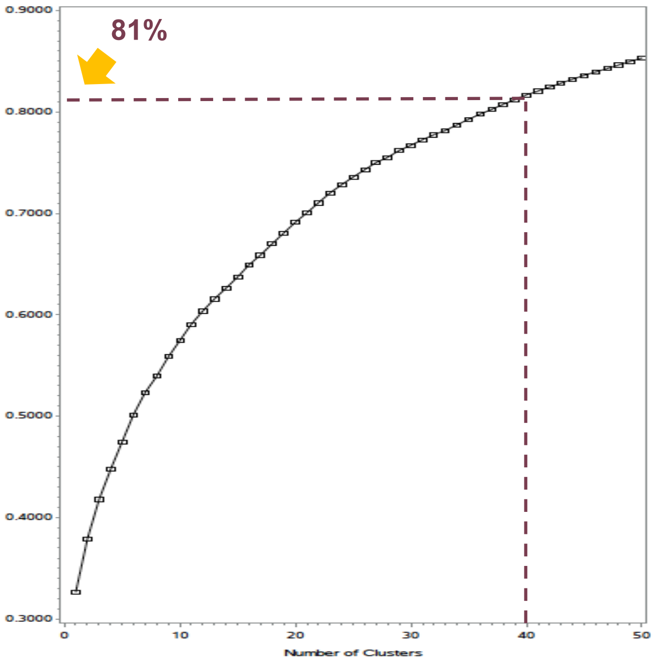


Fig. 6. Proportion of Variation Explained by Clusters

the number of random variables under consideration by obtaining a set of essential variables. It is a critical data preprocessing technique for large-scale and streaming data classification tasks. In order to speed up the modeling process, the predictor variables should be grouped into similar clusters. A few variables can then be selected from each cluster-this way the analyst can quickly reduce the number of variables and speed up the modeling process [17]. In this study, we found the cluster of variables that were highly correlated among themselves and not correlated with variables in other groups. Chose the variable which had the highest R^2 ratio in each cluster, thus reducing the dimension of the dataset. The R^2 ratio can be calculated by:

$$R^2 Ratio = \frac{R^2 OwnCluster}{R^2 NextClosest} \tag{1}$$

Next, the cluster representatives are put into the predictive model. Figure 5 shows the example of how to choose the representative variable in each cluster. Figure 6 shows that 40 clusters explained approximately 81% of the variability in the data, so we chose the best 40 variables to consider as predictors in our model.

C. Multicollinearity Assessment

Multicollinearity may have several opposing impacts on estimated coefficients in classification regression analysis. Consequently, it is essential that researchers should focus on detecting its existence. Analyzing latent roots and latent vectors of the correlation matrix and the variance inflation factors (VIF) is necessary for the analysis process. A VIF quantifies how much the variance is inflated. The VIF of K^th predictor can be calculated by:

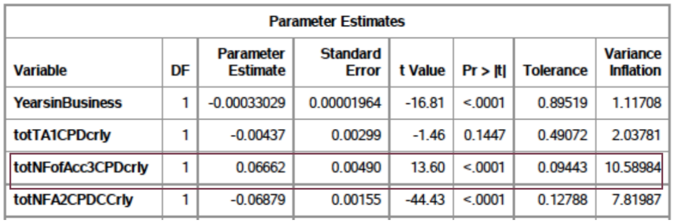
$$VIF_K = \frac{1}{1 - R^2{}_K} \tag{2}$$

The VIF of 1 indicates that there is no correlation between the K^th predictor and the remaining predictor variables. The general rule of thumb is that VIFs exceeding 4 warrant further investigation, while VIFs exceeding 10 are signs of severe multicollinearity requiring correction [18]. In this study, we tested VIFs of the 40 variables. Figure 7 shows the example of VIFs of a part of variables. From the results, there are two variables in which the VIF values are greater than 10, so we removed these two variables because it will cause the multicollinearity problem. After checking the multicollinearity of the 40 variables, we found some variables represent same meanings with the target variables. Based on the financial knowledge and the data description of the codebook, we removed these variables. Finally, there were 16 independent variables that remained in the dataset.

IV. Method

A. Logistic Regression

Logistic regression is an important machine learning algorithm. The goal is to model the probability of a random variable Y is 0 or 1 given experimental data. The brief explanation of logistic regression concepts is below [19]. The generalized linear model function of logistic regression can be defined with the parameter θ.



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
YearsinBusiness	1	-0.00033029	0.00001964	-16.81	<.0001	0.89519	1.11708
totTA1CPDcrly	1	-0.00437	0.00299	-1.46	0.1447	0.49072	2.03781
totNFofAcc3CPDcrly	1	0.06662	0.00490	13.60	<.0001	0.09443	10.58984
totNFA2CPDCCrly	1	-0.06879	0.00155	-44.43	<.0001	0.12788	7.81987

Fig. 7. Example of VIFs of Variables

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \tag{3}$$

The likelihood function is below, that assuming all the samples are independent.

$$\begin{aligned} L(\theta|x) &= Pr(Y|X;\theta) \\ &= \prod_i Pr(y_i|x_i;\theta) \\ &= \prod_i h_\theta(x_i)^{y_i}(1 - h_\theta(x_i))^{1-y_i} \end{aligned} \tag{4}$$

Typically, the log likelihood is maximized with a normalizing factor N^{-1}. This maximized is used in gradient descent.

$$N^{-1} log L(\theta|x) = N^{-1} \sum_{i=1}^{N} log Pr(y_i|x_i;\theta) \tag{5}$$

Assuming the (x,y) pairs are drawn uniformly from the underlying distribution, then in the limit of large N, The H(Y—X) is the conditional entropy and the $D_K L$ is the Kullback-Leibler divergence.

$$\begin{aligned} &\lim_{N \to +\infty} N^{-1} \sum_{i=1}^{N} log Pr(y_i|x_i;\theta) \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in Y} Pr(X = x, Y = y) log Pr(Y = y|X = x;\theta) \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in Y} Pr(X = x, Y = y) Pr(X = x, Y = y) \\ &(-log \frac{Pr(Y = y|X = x)}{Pr(Y = y|X = x;\theta)}) + log Pr(Y = y|X = x) \\ &= -D_{KL}(Y||Y_\theta) - H(Y|X) \end{aligned} \tag{6}$$

We used stepwise selection method for variables screening at the significant level of 0.05. After we generated the classification table, a cutoff value of P was selected for prediction purpose. P is calculated by the probability that each observation will belong to one of the two classes. The math formula of this regression, where B denotes the models parameters and X refers to the input parameters:

$$P = \frac{1}{1 + e^{-(B_0 + B_1 X_1 + \ldots + B_n X_n)}} \tag{7}$$

When making predictions on the validation set, observations with the appropriate P value exceed the cutoff value are predicted as 1 while those with the fitted value of P smaller than the cutoff value are predicted as 0.

B. Decision Tree

The decision tree is currently one of the most interesting supervised learning algorithms [20]. In this algorithm, an empirical tree expresses a classification of the data that is created by applying a series of simple rules. These algorithms produce set of rules which can be employed for prediction through the repeated process of splitting. The chi-squared automatic interaction detection (CHAID), classification and regression trees (CART), C4.5 and C5.0 are some of the most common tree methods. Information gain and entropy are used to create the trees [21]. Information gain estimates how well a given attribute separates the training examples according to their dependent variable. The measure is used to choose among the candidate variables at each step while growing the tree. Information gain (S, X) of a variable X relative to a collection of examples S, is defined as:

$$Gain(S,X) = Entropy(S) - \sum_{v \in value(X)} \frac{|S_v|}{|s|} Entropy(S_v) \tag{8}$$

Entropy is a measure of homogeneity that can be defined as below, c denotes to the number of classes, P_i denotes to the proportion of sample S that belong to class “i”.

$$Entropy(S) = \sum_{i=1}^{c} -P_i log_2 P_i \tag{9}$$

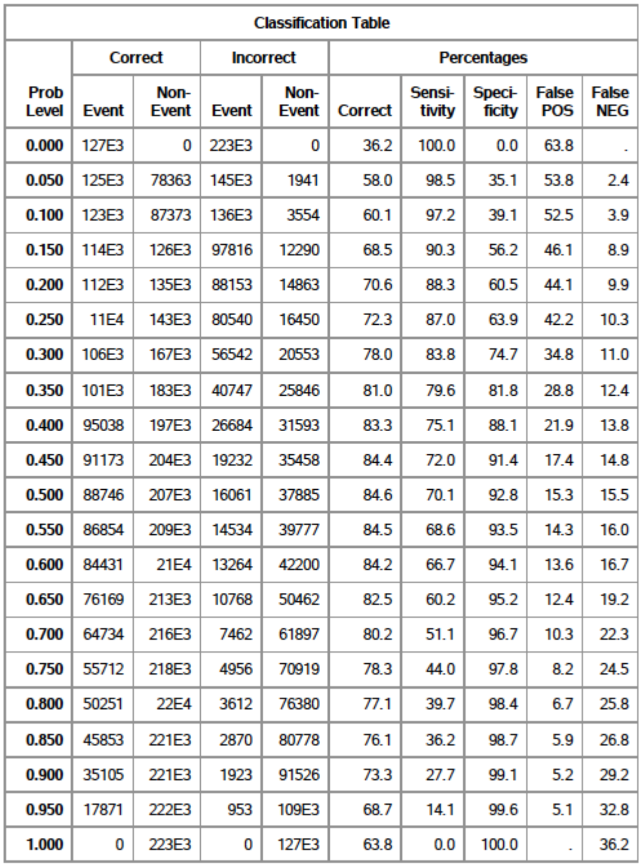
In this study, the class i=2 because our dependent variable is binary.

C. Neural Networks

The neural network is another general method of regression and classification. Neural networks were initially developed by researchers trying to simulate the neurophysiology of the human being brain. The feedforward neural network is the most manageable and most popular application. Training a neural network is the process of setting the best weights on the inputs of each of the units, and backpropagation (back-drop) is the most common method for computing the error gradient for a feedforward network [22]. Back-propagation is a supervised learning technique used in neural networks and is most suitable for diagnostic and predictive problems. Back-propagation neural network involves multi-layer topology that includes an input layer, a hidden layer, and an output layer [23]. In this study, we used linear combination functions in the hidden layer and sigmoid function in the output layer. The formula for the sigmoid is below:

$$sigmoid(x) = \frac{1}{1+e^{(-x)}} \tag{10}$$

The number of hidden layers applied is defined by evaluating the generalization error of each network.



Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	127E3	0	223E3	0	36.2	100.0	0.0	63.8	.
0.050	125E3	78363	145E3	1941	58.0	98.5	35.1	53.8	2.4
0.100	123E3	87373	136E3	3554	60.1	97.2	39.1	52.5	3.9
0.150	114E3	126E3	97816	12290	68.5	90.3	56.2	46.1	8.9
0.200	112E3	135E3	88153	14863	70.6	88.3	60.5	44.1	9.9
0.250	11E4	143E3	80540	16450	72.3	87.0	63.9	42.2	10.3
0.300	106E3	167E3	56542	20553	78.0	83.8	74.7	34.8	11.0
0.350	101E3	183E3	40747	25846	81.0	79.6	81.8	28.8	12.4
0.400	95038	197E3	26684	31593	83.3	75.1	88.1	21.9	13.8
0.450	91173	204E3	19232	35458	84.4	72.0	91.4	17.4	14.8
0.500	88746	207E3	16061	37885	84.6	70.1	92.8	15.3	15.5
0.550	86854	209E3	14534	39777	84.5	68.6	93.5	14.3	16.0
0.600	84431	21E4	13264	42200	84.2	66.7	94.1	13.6	16.7
0.650	76169	213E3	10768	50462	82.5	60.2	95.2	12.4	19.2
0.700	64734	216E3	7462	61897	80.2	51.1	96.7	10.3	22.3
0.750	55712	218E3	4956	70919	78.3	44.0	97.8	8.2	24.5
0.800	50251	22E4	3612	76380	77.1	39.7	98.4	6.7	25.8
0.850	45853	221E3	2870	80778	76.1	36.2	98.7	5.9	26.8
0.900	35105	221E3	1923	91526	73.3	27.7	99.1	5.2	29.2
0.950	17871	222E3	953	109E3	68.7	14.1	99.6	5.1	32.8
1.000	0	223E3	0	127E3	63.8	0.0	100.0	.	36.2

Fig. 8. Classification Table

V. Model Development and Comparision

A. Logistic Regression

Before testing the model, we split the dataset with the 16 explanatory variable as 70% training data and 30% validation data. We employed the stepwise selection and generated a classification table based on the training data. Figure 8 shows the classification table, and it suggests the cutoff value of P=0.5 should be selected since this value can reach a relatively high accuracy(84.6%) and relative low false negative (15.5%) on the training dataset. False negative is a test result indicates a condition does not hold, while in fact it does. The false negative rate and accuracy calculated by:

$$ACC = \frac{TP+TN}{FN+FP+TP+TN} \tag{11}$$

$$FNR = \frac{FN}{TP+FN} \tag{12}$$

This cutoff value was then used to make predictions on the validation set. Figure 9 demonstrates the ROC curve of logistic regression based on the validation dataset. The ROC AUC of logistic regression algorithm is 88.55%. The coefficient estimations and the P value for these 16 variables are shown in Figure 10. The variable with larger Wald Chi-Square values

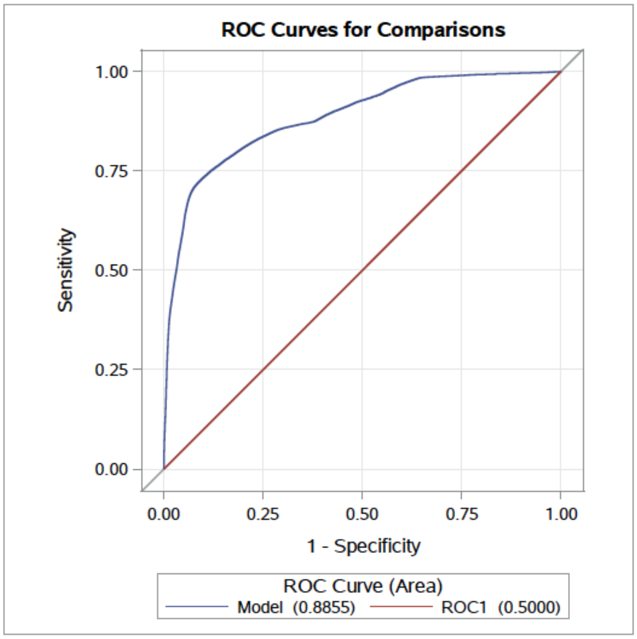
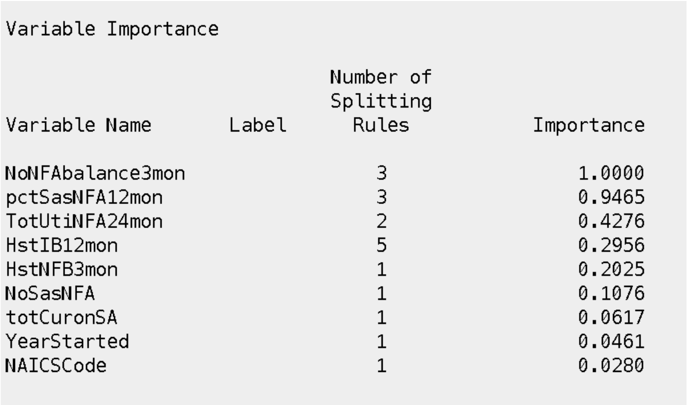


Fig. 9. ROC of Logistic Regression

are considered to be more important in making predictions. The KolmogorovSmirnov statistic is a method to evaluate the models' performance. The KolmogorovSmirnov statistic quantifies a distance between the experimental distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples [24]. Figure 11 shows the KS test result of logistic regression. The D value of the KS test is 0.634 which shows the model has a good performance.

B. Decision Trees and Neural Networks

Figure 12 presents the variable importance of the decision tree. Figure 13 shows the ROC curve of the decision tree and neural networks. The neural networks get the highest accuracy, and both neural network and decision tree's ROC AUC are higher than logistic regression.

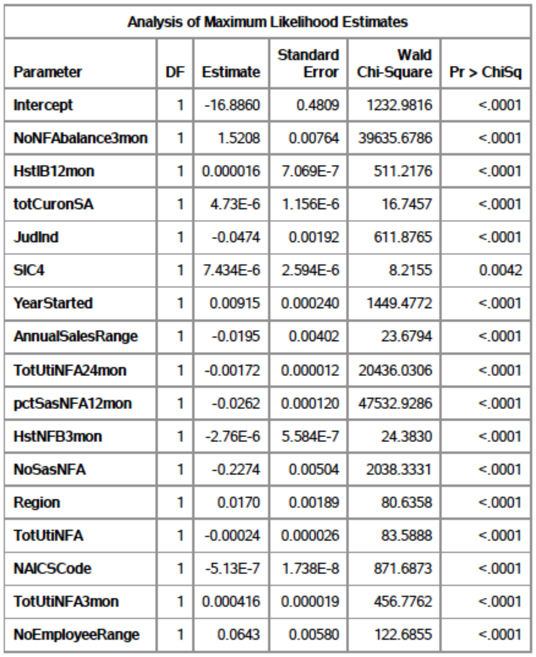


```
Variable Importance


                                Number of
                                Splitting
Variable Name        Label        Rules              Importance

NoNFAbalance3mon                    3                    1.0000
pctSasNFA12mon                      3                    0.9465
TotUtiNFA24mon                      2                    0.4276
HstIB12mon                          5                    0.2956
HstNFB3mon                          1                    0.2025
NoSasNFA                            1                    0.1076
totCuronSA                          1                    0.0617
YearStarted                         1                    0.0461
NAICSCode                           1                    0.0280
```

Fig. 12. The Most Important Variables Generated by Decision Trees



Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-16.8860	0.4809	1232.9816	<.0001
NoNFAbalance3mon	1	1.5208	0.00764	39635.6786	<.0001
HstIB12mon	1	0.000016	7.069E-7	511.2176	<.0001
totCuronSA	1	4.73E-6	1.156E-6	16.7457	<.0001
JudInd	1	-0.0474	0.00192	611.8765	<.0001
SIC4	1	7.434E-6	2.594E-6	8.2155	0.0042
YearStarted	1	0.00915	0.000240	1449.4772	<.0001
AnnualSalesRange	1	-0.0195	0.00402	23.6794	<.0001
TotUtiNFA24mon	1	-0.00172	0.000012	20436.0306	<.0001
pctSasNFA12mon	1	-0.0262	0.000120	47532.9286	<.0001
HstNFB3mon	1	-2.76E-6	5.584E-7	24.3830	<.0001
NoSasNFA	1	-0.2274	0.00504	2038.3331	<.0001
Region	1	0.0170	0.00189	80.6358	<.0001
TotUtiNFA	1	-0.00024	0.000026	83.5888	<.0001
NAICSCode	1	-5.13E-7	1.738E-8	871.6873	<.0001
TotUtiNFA3mon	1	0.000416	0.000019	456.7762	<.0001
NoEmployeeRange	1	0.0643	0.00580	122.6855	<.0001

Fig. 10. Parameter Estimations for the Logistic Regression

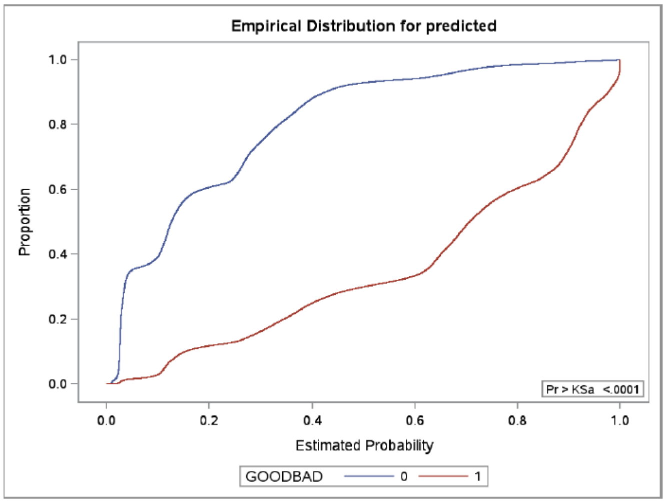


Fig. 11. KS Test Result of Logistic Regression

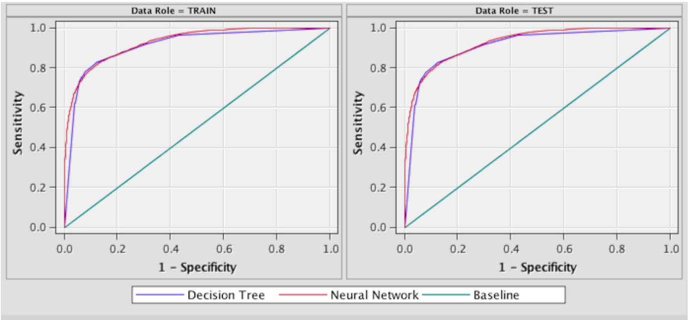


Fig. 13. ROC of Decision Tree and Neural Networks

Models	Accuracy	KS-test(D)
Logistic Regression	88%	0.63
Decision Tree	91%	0.70
Neural Networks	93%	0.69

TABLE I
MODELS'S PERFORMANCE COMPARISON

Table I shows the accuracy and D value of KS test for each model. From the table, we can find that neural networks have the best accuracy and the decision tree has the best D value in KS test. It shows the new machine learning algorithms have better performance compare with the logistic regression.

VI. CONCLUSION

In this study, we used the real world data provided by Equifax and compared the different machine learning algorithms including logistic regression, decision tree, and neural network. We found that, even though neural networks returned the best accuracy, it is harder to explain due to the calculations hidden layers and variety nodes, especially for the risk models. The decision tree is our recommended model since it has superior performance and it is easy to interpret. The most critical factors that affect the commerces' past due are: "NoNFAbalance3mon", the Number of Non-Financial Accounts with Balance Reported in the Last 3 Months; "TotUtiNFA3mon", the Total Utilization on Non-Financial Accounts; "totCuronSA", the Total Current Balance on Service Accounts and "YearStarted", the Commerce Year.

In summary, this study first uses the real world large and unique datasets to build models. Secondly, this study's explored how to predict if enterprises will have past due in next 12 months of their non-financial account which researchers have seldom focused. Thirdly, in this study, we compared three different models: logistic regression, decision tree, and neural networks. We found decision tree is the best model since it has better performance and is easy to interpret. The last but not the least, we discovered some interesting factors that can impact the enterprises' credit score.

VII. LIMITATIONS

Our study provides an in-depth illustration of the potential benefits that machine-learning techniques can bring to commercial credit prediction. However, neural networks are not interpretable since they contain hidden layers and nodes. Thus, how to interpret neural networks in commercial credit prediction will be our future research.

VIII. ACKNOWLEDGMENT

I would like to thank Equifax for providing the datasets and Dr. Jennifer Lewis Priestley for providing the classroom framework and fantastic class instruction for completing this study.

REFERENCES

[1] Comptroller's Handbook, "Commercial Loans," (online), 1998, https://www.occ.treas.gov/publications/publications-by-type/comptrollers-handbook/commercial-loans/pub-ch-commercial-loans.pdf.

[2] S. Nargundkar and J. L. Priestley, "Assessment of evaluation methods for prediction and classifications of consumer risk in the credit industry," *Neural Networks in Business Forecasting*, p. 266, 2004.

[3] E. K. Laitinen, "Predicting a corporate credit analyst's risk estimate by logistic and linear models," *International review of financial analysis*, vol. 8, no. 2, pp. 97–121, 1999.

[4] A. F. Atiya, "Bankruptcy prediction for credit risk using neural networks: A survey and new results," *IEEE Transactions on neural networks*, vol. 12, no. 4, pp. 929–935, 2001.

[5] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision support systems*, vol. 37, no. 4, pp. 543–558, 2004.

[6] Y. Wang, S. Wang, and K. K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 6, pp. 820–831, 2005.

[7] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial crisis prediction: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 421–436, 2012.

[8] G. E. Pinches and K. A. Mingo, "A multivariate analysis of industrial bond ratings," *The journal of Finance*, vol. 28, no. 1, pp. 1–18, 1973.

[9] L. H. Ederington, "Classification models and bond ratings," *Financial review*, vol. 20, no. 4, pp. 237–262, 1985.

[10] J. A. Gentry, D. T. Whitford, and P. Newbold, "Predicting industrial bond ratings with a probit model and funds flow components," *Financial Review*, vol. 23, no. 3, pp. 269–286, 1988.

[11] M. T. Hagan, H. B. Demuth, M. H. Beale *et al.*, *Neural network design*. Pws Pub. Boston, 1996, vol. 20.

[12] E. Huang and C. Scott, "Credit risk scorecard design, validation and user acceptance," 2007.

[13] D. T. Larose and C. D. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014.

[14] P. D. Allison, *Missing data*. Sage publications, 2001, vol. 136.

[15] D. A. Bennett, "How can i deal with missing data in my study?" *Australian and New Zealand journal of public health*, vol. 25, no. 5, pp. 464–469, 2001.

[16] R. J. Little and D. B. Rubin, "Single imputation methods," *Statistical Analysis with Missing Data, Second Edition*, pp. 59–74, 2002.

[17] B. D. Nelson, "Variable reduction for modeling using proc varclus," in *Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference, Cary, NC, SAS Institute*, 2001.

[18] R. M. O?brien, "A caution regarding rules of thumb for variance inflation factors," *Quality & Quantity*, vol. 41, no. 5, pp. 673–690, 2007.

[19] A. Ng, "Cs229 lecture notes," *CS229 Lecture notes*, vol. 1, no. 1, pp. 1–3, 2000.

[20] P. Geurts, A. Irrthum, and L. Wehenkel, "Supervised learning with decision tree-based methods in computational and systems biology," *Molecular Biosystems*, vol. 5, no. 12, pp. 1593–1605, 2009.

[21] Q. R. Wang and C. Y. Suen, "Analysis and design of a decision tree based on entropy reduction and its application to large character set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 4, pp. 406–417, 1984.

[22] M. J. Berry and G. Linoff, *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., 1997.

[23] C.-L. Chang and C.-H. Chen, "Applying decision tree and neural network to increase quality of dermatologic diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 4035–4041, 2009.

[24] H. W. Lilliefors, "On the kolmogorov-smirnov test for normality with mean and variance unknown," *Journal of the American statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.